A Course in Quantum Information Theory

Ofer Shayevitz

Spring 2007

Based on lectures given at the Tel Aviv University

Edited by Anatoly Khina

Version compiled January 9, 2010

Contents

Chapter 1

Preliminaries

Summary by Yuval Regev.

I A Very Brief Introduction

I.1 Quantum Information Theory

Quantum information theory combines two distinct theories – *(classical) information theory* and *quantum mechanics*. One might wonder, what does a theory of information have to do with a physical theory? In classical information theory we are usually not very concerned with this relation, as the theory is abstract enough and does not depend on any specific physical media. However, information is still a physical quantity, as it is stored, measured, transmitted and received using physical devices. The most basic element of classical information theory is the (classical) **bit**. A bit is represented in practice by some classical physical quantity such as voltage or magnetization, and the statistical laws governing the way it can be processed are derived from the relevant (classical) physical theory. In the information theoretic regime, these statistical laws are all that is required in order to determine the fundamental limits of information processing.

The basic element in quantum information theory is the *quantum bit*, or **qubit**. Similarly to a classical bit, the qubit is a physical quantity. In contrast however, it obeys the laws of quantum mechanics. As we shall see, information under the laws of quantum mechanics behaves in a markedly different way than classical information, which cannot be captured using the classical tools. One example is the renowned *no cloning Theorem*, which states that (unlike a bit) a qubit cannot be accurately duplicated. The following example demonstrates an even more peculiar aspect of that behavior. We will return to this example at the end of the course.

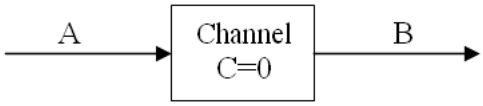


Figure 1.1: Channel with zero capacity

Example I.1. *Consider a channel with zero quantum capacity, namely a channel over which qubits cannot be reliably transmitted (we will define all this formally later on in the course). The channel is depicted in Fig. 1.1. Now, suppose we also have a similar zero-capacity channel in the opposite direction, as depicted in Fig. 1.2. In a purely classical setting, namely when the channel has zero (classical) capacity, adding a zero-capacity backwards channel is useless, the capacity in both directions would still be zero. In the quantum regime however, things may be significantly different. Although the unidirectional quantum capacities are zero, using both channels together may allow reliable communication of qubits back and forth, making the quantum capacity nonzero! We will return to this example in Section VI.1, Chapter 8.*

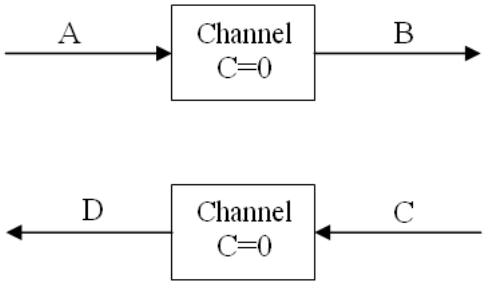


Figure 1.2: Two channels with zero capacity

I.2 Quantum Computing

The basic elements of a "classical computer" are bits, and the basic operations performed over these bits are boolean operations. The basic elements of a "quantum computer" are qubits, and the basic operations performed over qubits are quantum transformations and measurements, which will be defined in the sequel. While in a classical computer the state of the system is uniquely defined and can be precisely measured or copied, a quantum computer may be in a superposition os states, which cannot be reliably measured or copied. This makes quantum computation a much more difficult task, but also holds a great promise - there exist quantum algorithms that can take advantage of this superposition to significantly boost performance relative to their classical counterparts. One prominent example is Shor's algorithm for inte-

ger factorization which runs in polynomial time on a quantum computer, in contrast to the exponential complexity of the best known classical algorithms.

II Preliminaries: Linear Algebra

II.1 Notations

- $\mathcal{H}_n$ - *Hilbert* space of dimension n (a.k.a "inner product space").
- $|\psi\rangle$ - *ket* - vector in $\mathcal{H}_n$ (column vector when representing $\mathcal{H}_n$ w.r.t. an orthonormal base).
- $\langle\psi|$ - *bra* - conjugate transpose of $|\psi\rangle$.
- $\langle\psi|\phi\rangle$ - *braket* - inner product between $|\psi\rangle$ and $|\phi\rangle$.

Property II.1 (Inner Product Properties)**.** *1. Conjugate Symmetry:* $\langle\psi|\phi\rangle = \langle\phi|\psi\rangle^*$.

2. Linearity in the second vector: $\langle\phi|\left(a|\psi_1\rangle + b|\psi_2\rangle\right) = a\langle\psi|\phi_1\rangle + b\langle\psi|\phi_2\rangle$.

3. Non-negativity: $\langle\psi|\psi\rangle \geq 0,\ \langle\psi|\psi\rangle = 0 \Leftrightarrow |\psi\rangle = 0$.

II.2 Definitions and Basic Properties

Definition II.1 (Linear Operator)**.** *A linear operator can always be represented by a corresponding matrix with respect to some orthonormal space. We shall use operators and their corresponding matrices interchangeably.*

Definition II.2 (Adjoint Opeartor)**.** *The operator $A^\dagger$ is the adjoint operator of A, if it satisfies $\langle\psi|A^\dagger|\phi\rangle = \langle\phi|A|\psi\rangle^*$ for every $\phi, \psi \in \mathcal{H}$. If we represent the operator A by a matrix according to some orthonormal base, then the conjugate transpose of this matrix is the representation matrix of $A^\dagger$.*

Definition II.3 (Outer Product)**.** *The outer product between $|\psi\rangle$ and $|\phi\rangle$ is the operator: $|\psi\rangle\langle\phi|$. $\langle\phi|$ is an operator from $\mathcal{H}_n$ to $\mathbb{C}$(projects onto the state $|\phi\rangle$), and $|\psi\rangle$ is an operator from $\mathbb{C}$ to $\mathcal{H}_n$.*

Example II.1. $\left(|w\rangle\langle v|\right)\left(|u\rangle\right) = |w\rangle \underbrace{\langle v|u\rangle}_{scalar} = \langle v|u\rangle|w\rangle$.

Lemma II.1. *If $\{|i\rangle\}_{i=1}^n$ is an orthonormal basis of $\mathcal{H}_n$ then $\sum_{i=1}^n |i\rangle\langle i| = I$.*

Proof. Let v be any vector in $\mathcal{H}$. Then, it has a unique representation according to the orthonormal basis $\{|i\rangle\}_{i=1}^n$, $\sum_j \alpha_j |j\rangle$. Hence

$$\left(\sum_i |i\rangle\langle i|\right)|v\rangle = \left(\sum_i |i\rangle\langle i|\right)\left(\sum_j \alpha_j |j\rangle\right) = \sum_{i,j} \alpha_j |i\rangle \underbrace{\langle i|j\rangle}_{\delta_{ij}} = \sum_j \alpha_j |j\rangle = |v\rangle.$$

□

Lemma II.2. *If $\{|i\rangle\}_{i=1}^k$ is an orthonormal basis of the sub-space $\mathcal{G}_k$ of the Hilbert space $\mathcal{H}_n$ ($k \leq n$), then the operator $P = \sum_i^k |i\rangle\langle i|$ is a projection operator onto $\mathcal{G}_k$. Namely, the following properties hold:*

Property II.2 (Projector)**.**

(i) $P = P^\dagger$.

(ii) $P = P^2$.

Remark II.1. *Note that $I - P$ is a projector onto the orthogonal subspace $\mathcal{G}_k^\perp$.*

Lemma II.3 (Spectral Decomposition)**.** *Let A be a Hermitian matrix, namely $A = A^\dagger$. Then A can be unitarily diagonalized , i.e., there exists an orthonormal basis composed of its eigenvectors $\{|i\rangle\}_{i=1}^n$ of A, such that*

$$A = \sum_{i=1}^n \lambda_i |i\rangle\langle i|,$$

where $\{\lambda_i\}_{i=1}^n$ are the corresponding (real) eigenvalues.

Thus, a Hermitian operator can always be represented by some linear combination of projection operators.

Definition II.4 (Positive Opeartor)**.** *An operator A is Positive (resp. Positive Definite) if all of its eigenvalues λ_i are real and non-negative (resp. positive), or equivalently if $\langle\psi|A|\psi\rangle \geq 0$ (resp. > 0) for all $|\psi\rangle$ (resp. $|\psi\rangle \neq 0$).*

Note that a Positive operator (matrix) over $\mathcal{H}_n$ is necessarily Hermitian (this statement is not true over $\mathbb{R}^n$, namely not any Positive operator (matrix) is Symmetric).

Lemma II.4. *The matrix $A^\dagger A$ is positive for every A.*

Definition II.5 (Function of Hermitian Operator)**.** *Let A be a hermitian operator and $f : \mathbb{C} \to \mathbb{C}$ some scalar function. The operation of f on A is defined by applying f to the eigenvalues of A:*

$$f(A) \triangleq \sum_{i=1}^{n} f(\lambda_i)|i\rangle\langle i| \,.$$

Example II.2. *If A is positive, then $\sqrt{A} \triangleq \sum_{i=1}^{n} \sqrt{\lambda_i}|i\rangle\langle i|$.*

Lemma II.5. *A matrix U is unitary, i.e., $U^\dagger U = I$, iff there exist two orthonormal bases $\{|v_i\rangle\}$ and $\{|w_i\rangle\}$ such that*

$$U = \sum_i \underbrace{|w_i\rangle}_{\substack{\text{multiplies the} \\ \text{respective element} \\ \text{of the second basis}}} \underbrace{\langle v_i|}_{\substack{\text{projects on} \\ \text{the first basis}}} .$$

The unitary operator U can be thought of a transformation from a representation according the orthonormal basis $\{|v_i\rangle\}$ to a representation according the orthonormal basis $\{|w_i\rangle\}$.

II.3 The Trace Operator

The trace operator is widely used in the sequel. Therefore a brief review of its definition and some of its properties are given below.

Definition II.6 (Trace)**.** *The trace of a $m \times m$ matrix A is defined as the sum of its main diagonal elements:*

$$tr(A) = \sum_{i=1}^{m} A_{ii}.$$

Property II.3.

(i) Linearity: $tr(A + B) = tr(A) + tr(B)$

(ii) $tr(AB) = tr(BA)$.

(iii) For every unitary operator U: $trA = tr(UAU^\dagger)$ (prove using the previous property).

(iv) For any state $|\psi\rangle$: $\langle\psi|A|\psi\rangle = tr(A|\psi\rangle\langle\psi|)$.

(v) For any orthonormal basis $\{|i\rangle\}$: $trA = tr(A\sum_i |i\rangle\langle i|) = \sum_i tr(A|i\rangle\langle i|) = \sum_i \langle i|A|i\rangle$.

III Postulates of Quantum Theory

III.1 The State Space

Postulate I. *A **closed** physical system is described by a **state vector**, which is a unit vector in a Hilbert Space called **the state space** of the physical system.*

Example III.1. *The smallest non-trivial state space is $\mathcal{H}_2$. Let $\{|0\rangle, |1\rangle\}$ be a basis of $\mathcal{H}_2$, with a vector representation*

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

This basis is occasionally called the Computational Basis. A general state (unit vector) in $\mathcal{H}_2$ can be expressed as (see also Fig. 1.3):

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

with $\alpha, \beta \in \mathbb{C}$ satisfying $|\alpha|^2 + |\beta|^2 = 1$.

Definition III.1. *A state vector in $\mathcal{H}_2$ is called a **qubit**.*

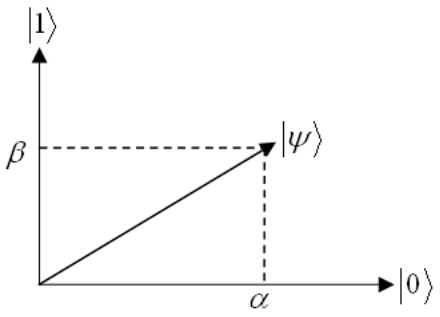


Figure 1.3: Pictorial representation of a qubit

Measuring a qubit. According to the laws of quantum mechanics (to be explicitly given soon), if we measure a qubit in the computational basis we would get the result $|0\rangle$ with probability $|\alpha|^2$ and the result $|1\rangle$ with probability $|\beta|^2$. Moreover, after the measurement, the qubit "collapses" to the state measured (it becomes equal, with probability 1, to the outcome of the measurement). Hence, in this experiment, the qubit acts like a regular bit, with probability distribution $\left\{|\alpha|^2, |\beta|^2\right\}$. The difference is (informally), that 1) the qubit can have values that are "superpositions of 0 and 1", 2)

the measurement process changes the qubit's state, and 3) measuring in different bases results in different probability distributions. To further clarify, consider the following example.

Example III.2 (Measuring in different bases)**.** *Define the two (orthonormal) states in* $\mathcal{H}_2$*:*

$$|+\rangle = \tfrac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right),$$
$$|-\rangle = \tfrac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right).$$

These states can be written in a vector notations as (see also Fig. 1.4):

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad |+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad |-\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

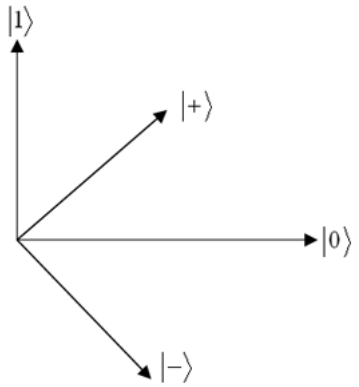


Figure 1.4: Pictorial representation of $|0\rangle, |1\rangle, |+\rangle$ and $|-\rangle$

Suppose that the initial state of the system is $|+\rangle$*, and a measurement is performed. The measurement can be made is different bases:*

1. ***Measurment in the computational base****: We get one of the results* $\{0, 1\}$ *with equal probability. The system collapses to one of the states* $\{|0\rangle, |1\rangle\}$*, depends on the result of the measurment.*

2. ***Measurment in the basis*** $\{|+\rangle, |-\rangle\}$*: We get the result* $+$ *with probability* 1*. The state does not change.*

3. ***Measurment in the basis*** $\{|+\rangle, |-\rangle\}$ ***after measuring in the computational base****: We get one of the resutls* $\{+, -\}$ *with equal probability. After the first measurement (in the computational base) the qubit collapsed to either* $|0\rangle$ *or* $|1\rangle$*, and after the second measurement (the one in* $\{|+\rangle, |-\rangle\}$ *we get* $\{+, -\}$ *with equal probabilities.*

Thus the act of measuring itself, effects the state of the system. We will elaborate on the measurement process when we get to Postulate IV.

III.2 Composite Systems and the Tensor Product

Let us continue our informal introduction, and discuss a system of more than one qubit.

Example III.3 (Two qubit system)**.** *As we have seen before, one qubit is a normalized vector in a two-dimensional space (with analogy to the classical bit, which can take one of two possible values). Following this logic, two qubits are defined as a normalized vector in a four-dimensional space (two classical bits correspond to four distinct values).*

$$|00\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |01\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad |10\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad |11\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{1.1}$$

A general state in this four-dimensional space is described by a lienar combination of the vectos in (1.1)*, i.e., by* $|\psi\rangle = \sum_{i,j\in\{0,1\}} \alpha_{ij}|ij\rangle$ *for some* $\{\alpha_{ij}\}$ *satisfying* $\sum_{i,j\in\{0,1\}} |\alpha_{ij}|^2 = 1$*. Again,* $|\alpha_{ij}|^2$ *is the probability of measuring* $|ij\rangle$*, when performing a measurement in the computational base* (1.1)*.*

Example III.4 (Classicsal analogy)**.** *Let* $X_0 \sim \mathrm{Ber}(p)$ *and* $X_1 \sim \mathrm{Bern}(q)$ *be two independent binary random variables.*

The joint distribution of $(X_1, X2)$ *is*

$$(X_0, X_1) \sim ((1-p)(1-q), (1-p)q, p(1-q), pq)\,.$$

Note that the joint distribution vector is a ***tensor product*** *of the marginal distribution vectors. Following that, one may expect that a two qubit system*

$$|\psi_1\rangle = \alpha_1|0\rangle + \beta_1|1\rangle,$$
$$|\psi_2\rangle = \alpha_2|0\rangle + \beta_2|1\rangle$$

would admit a joint state given by

$$|\psi\rangle = |\psi_1\psi_2\rangle = \alpha_1\alpha_2|00\rangle + \alpha_1\beta_2|01\rangle + \alpha_2\beta_1|10\rangle + \beta_1\beta_2|11\rangle. \tag{1.2}$$

This is indeed the case. The state (1.2) *is called a tensor product of the two qubit states.*

Definition III.2 (Tensor Product)**.** *Let* V *and* W *be two Hilbert spaces of dimensions* m, n *respectively. The* ***tensor product*** *of* V *and* W*, denoted by* $V \otimes W$*, is a Hilbert space of dimension* $m \cdot n$*, in which each state is a linear combination of states of the form* $|v\rangle \otimes |w\rangle = |v\rangle|w\rangle = |vw\rangle$*, for every* $|v\rangle \in V, |w\rangle \in W$*.*

The natural inner product in the space $V \otimes W$ (induced by the inner products of the spaces V and W) is:

$$\left(\sum_i a_i |v_i w_i\rangle \, , \, \sum_j b_j |v_j' w_j'\rangle \right) \triangleq \sum_{i,j} a_i^* b_j \langle v_i | v_j' \rangle \langle w_i | w_j' \rangle .$$

Property III.1 (Properties of tensor product)**.**

1. *Linearity in both variables.*
2. *Let $\{|i\rangle\}_{i=1}^n$ and $\{|j\rangle\}_{j=1}^m$ be two orthonormal bases of the space V and W, respectively. Then $\{|ij\rangle\}$ is an orthonormal bas of the space $V \otimes W$.*

Example III.5 (Tensor product of $\mathcal{H}_2$ spaces)**.** *Let $V = W = \mathcal{H}_2$ and consider the respective states*

$$\begin{aligned} |\psi_v\rangle &= \tfrac{1}{\sqrt{2}} \left(|0\rangle + |1\rangle \right) \\ |\psi_w\rangle &= \tfrac{1}{\sqrt{2}} \left(|0\rangle - |1\rangle \right) . \end{aligned}$$

Using Property III.1 we have:

$$|\psi_v \psi_w\rangle = \frac{1}{2} \left(|0\rangle + |1\rangle \right) \left(|0\rangle - |1\rangle \right) = \frac{1}{2} \left(|00\rangle + |10\rangle - |01\rangle - |11\rangle \right) ,$$

where $|\psi_v \psi_w\rangle \in V \otimes W$. Had we chosen the orthonormal basis $\{|+\rangle, |-\rangle\}$ for W , we would have gotten:

$$|\psi_v \psi_w\rangle = \frac{1}{\sqrt{2}} \left(|0-\rangle + |1-\rangle \right) = |+-\rangle \, .$$

Example III.6 (Inner product in tensor product space)**.** *Continuing the previous example, we define two states in the tensor product space $V \otimes W$:*

$$\begin{aligned} |\psi_1\rangle &= \tfrac{1}{\sqrt{2}} \left(|00\rangle + |11\rangle \right) \quad \textit{(EPR state)} \\ |\psi_2\rangle &= \tfrac{1}{\sqrt{2}} \left(|0\rangle + |1\rangle \right) |0\rangle = \tfrac{1}{\sqrt{2}} \left(|00\rangle + |10\rangle \right) . \end{aligned}$$

The inner product of these two states is:

$$\begin{aligned} \langle \psi_1 | \psi_2 \rangle &= \frac{1}{2} \left(\langle 00|00\rangle + \langle 00|10\rangle + \langle 11|00\rangle + \langle 11|10\rangle \right) \qquad (1.3) \\ &= \frac{1}{2} \left(\underbrace{\langle 0|0\rangle}_{1} \underbrace{\langle 0|0\rangle}_{1} + \underbrace{\langle 0|1\rangle}_{0} \underbrace{\langle 0|0\rangle}_{1} + \underbrace{\langle 1|0\rangle}_{0} \underbrace{\langle 1|0\rangle}_{0} + \underbrace{\langle 1|1\rangle}_{1} \underbrace{\langle 1|0\rangle}_{0} \right) \\ &= \frac{1}{2} \, . \end{aligned}$$

where first equality follows from Property III.1. An alternative way of deriving the same result, is recalling that (1.1) *forms an orthonormal basis in $\mathcal{H}_4$, and therefore only the term $\langle 00|00\rangle$, in* (1.3) *differs from* 0*.*

We are now ready for the next postulate.

Postulate II. *The state space of a "composite" system (a system which is composed of several sub-systems) is the tensor product of the state spaces of the sub-systems. Furthermore, if the k-th ($1 \leq k \leq n$) sub-system is in state $|\psi_k\rangle$, then the composite system is in state $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$.*

Remark III.1. *The converse is not necessarily true: not every state of the composite system can be decomposed into a tensor product of states of the sub-systems composing it.*

Remark III.1 leads us to the notion of entanglement.

Entanglement

Definition III.3 (Entangled state)**.** *A state of a composite system, which cannot be written as a tensor product of states of the sub-systems composing it, is said to be **entangled**. The two sub-systems are also called entangled.*

Example III.7 (EPR state)**.** *The state $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ (one of the EPR states) cannot be decomposed into the form $|\psi\rangle = |a\rangle|b\rangle$, and thus is entangled.*

In an entangled state, it is not possible to define the state of each of the sub-systems. Each sub-system is found in a state which is a combination of several states, with a certain probability distribution.

III.3 Unitary Evolution

Postulate III. *The evolution of a **closed** quantum system is described by a **unitary operator** (unitary transformation). Namely, if $|\psi_1\rangle$ and $|\psi_2\rangle$ are the states of the system in times t_1 and t_2, then*

$$|\psi_2\rangle = U|\psi_1\rangle,$$

where U is unitary and depends only on t_1 and t_2.

Remark III.2.

(i) The transformation must be unitary in order to preserve the norm - the resulting state must be of unit length at all times.

(ii) The evolution of a closed physical system is reversible - it is possible to go from state $|\psi_2\rangle$ to state $|\psi_1\rangle$ by applying the inverse operator $U^{-1} = U^\dagger$.

Example III.8 (Pauli matrices)**.**

- *Quantum NOT (bit flip):* $|0\rangle \longrightarrow |1\rangle, \quad |1\rangle \longrightarrow |0\rangle$

  $$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad XX^\dagger = I$$

  $$X\left(\alpha|0\rangle + \beta|1\rangle\right) = X \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \beta \\ \alpha \end{pmatrix} = \beta|0\rangle + \alpha|1\rangle$$

- *Phase flip:* $|0\rangle \longrightarrow |0\rangle, \quad |1\rangle \longrightarrow -|1\rangle$

  $$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad ZZ^\dagger = I$$

- $|0\rangle \longrightarrow i|1\rangle, \quad |1\rangle \longrightarrow -i|0\rangle$

  $$Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad YY^\dagger = I$$

  The matrices $\{I, X, Y, Z\}$ *are called* Pauli Matrices*, and they form a basis of the complex* 2×2 *matrix space.*

Definition III.4 (Hadamard matrix)**.**

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

The Hadamard matrix transforms the states $\{|0\rangle, |1\rangle\}$ *to the states* $\{|+\rangle, |-\rangle\}$*:*

$$H|0\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) = |+\rangle$$

$$H|1\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right) = |-\rangle\,.$$

Example III.9 (Quantum operations on two qubits)**.**

- ***Controlled-NOT (CNOT) Gate:*** *The CNOT gate (operator) is described by the following unitary matrix (see Fig. 1.5 for a schematic representation):*

$$\left(\begin{array}{cc|cc} 1 & 0 & \multicolumn{2}{c}{\multirow{2}{*}{0}} \\ 0 & 1 & & \\ \hline \multicolumn{2}{c|}{\multirow{2}{*}{0}} & 0 & 1 \\ & & 1 & 0 \end{array}\right) = \left(\begin{array}{c|c} I & 0 \\ \hline 0 & X \end{array}\right)$$

The operation of the CNOT gate in the computational basis is given by

$$
\begin{aligned}
|00\rangle &\longrightarrow |00\rangle\,, \\
|01\rangle &\longrightarrow |01\rangle\,, \\
|10\rangle &\longrightarrow |11\rangle\,, \\
|11\rangle &\longrightarrow |10\rangle\,.
\end{aligned}
$$

Or using dirac notations, $|ij\rangle \overset{CNOT}{\longrightarrow} |i\rangle|i \oplus j\rangle$*. The second qubit is flipped if the first qubit is 1.*

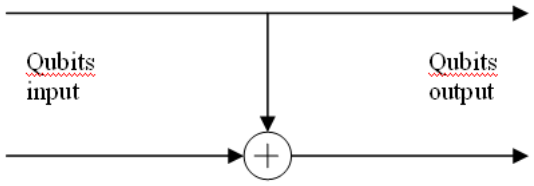


Figure 1.5: Schematic representation of a CNOT gate

- ***Controlled-U Gate:*** *Similarly to the CNOT gate, the first qubit controls whether the U operation will be performed on the remaining qubits (see also Fig. 1.6).*

$$
\left(\begin{array}{c|c} I & 0 \\ \hline 0 & U \end{array}\right)
$$

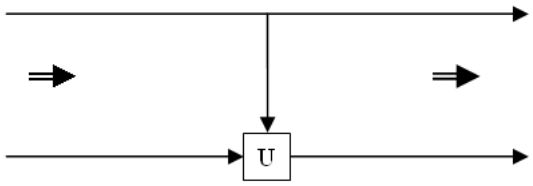


Figure 1.6: Schematic representation of a controlled-U gate

Definition III.5. ***Local operations*** *are operation which are applied to only a sub-system, whereas the remaining sub-systems remain untouched.*

Example III.10 (Local operations)**.** *Fig. 1.7) schematically depicts an operation* X *applied to the first qubit, and* H *to the second qubit. This has the following meaning:*

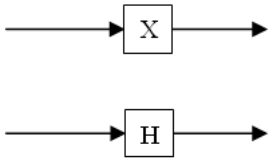


Figure 1.7: X applied to the first qubit and H to the second one

$$
\begin{aligned}
&|00\rangle \xrightarrow{X_1} |10\rangle \xrightarrow{H_2} |1+\rangle = |1\rangle \cdot \tfrac{1}{\sqrt{2}} \left(|0\rangle + |1\rangle\right) = \tfrac{1}{\sqrt{2}} \left(|10\rangle + |11\rangle\right) \\
&|01\rangle \xrightarrow{X_1} |11\rangle \xrightarrow{H_2} |1-\rangle = \tfrac{1}{\sqrt{2}} \left(|10\rangle - |11\rangle\right) \\
&|10\rangle \xrightarrow{X_1} |00\rangle \xrightarrow{H_2} |0+\rangle = \tfrac{1}{\sqrt{2}} \left(|00\rangle + |01\rangle\right) \\
&|11\rangle \xrightarrow{X_1} |01\rangle \xrightarrow{H_2} |0-\rangle = \tfrac{1}{\sqrt{2}} \left(|00\rangle - |01\rangle\right) ,
\end{aligned}
$$

where, in general, G_i means applying with the operator G on the ith qubit. The matrix describing the operation on the whole system is:

$$
\frac{1}{\sqrt{2}} \left(\begin{array}{cc|cc} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \\ \hline 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{array}\right) = X \otimes H \,,
$$

where in the last equality we used the following definition.

Definition III.6 (Matrix tensor product)**.** *The tensor product between matrix $A \in \mathbb{C}^{m \times n}$ and matrix $B \in \mathbb{C}^{p \times q}$ is a matrix of size $pm \times qn$, which has the form:*

$$
A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{m1}B & \cdots & \cdots & a_{mn}B \end{pmatrix} .
$$

Assertion III.1. *After applying a tensor product of two operators $U \otimes V$ on a composite state $|\psi_1\psi_2\rangle$, the resulting state is the tensor product of the sub-states after applying the local operators to them:*

$$
|\psi_1\psi_2\rangle \xrightarrow{U_1 V_2} (U \otimes V) |\psi_1\psi_2\rangle = \left(U|\psi_1\rangle\right) \left(V|\psi_2\rangle\right)
$$

Example III.11 (Tensor product of vectors)**.**

$$|0\rangle \otimes |1\rangle = |01\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$|-\rangle \otimes |+\rangle = |-+\rangle = \tfrac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \otimes \tfrac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \tfrac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

Property III.2 (Some simple tensor product properties, prove!)**.**

(i) Linearity in both variables.

(ii) Associativity: $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.

(iii) $(A \otimes C) \cdot (B \otimes D) = (A \cdot B) \otimes (C \cdot D)$*, whenever the right hand side is defined.*

(iv) The tensor product preserves Unitarity, Hermiticity, Positivity and the Projection property.

(v) if v_A, v_B *are vectors consisting of all the the eigenvalues of* A, B *respectively, then* $v_A \otimes v_B$ *is a vector consisting of all the eigenvalues of* $A \otimes B$*.*

(vi) $tr(A \otimes B) = tr(A) \cdot tr(B)$, $rank(A \otimes B) = rank(A) \cdot rank(B)$

III.4 Measurements

Postulate IV. *A general quantum measurement is described by a collection* $\{M_m\}$ *of* ***"measurement operators"*** *which act on the state space, and satisfy the "completeness equation":*

$$\sum_m M_m^\dagger M_m = I.$$

The index m is the outcome index of the measurement. If the state, prior to the measurement, is $|\psi\rangle$, then the probability of measuring m is:

$$P(m) = \langle \psi | M_m^\dagger M_m | \psi \rangle$$

and the state after the measurement, given that m was measured is:

$$|\psi_m\rangle = \frac{M_m |\psi\rangle}{\sqrt{P(m)}}.$$

Note that the "completeness equation" is equivalent to the requirement that the sum of all probabilities is equal to 1:

$$\sum_m P(m) = \sum_m \langle\psi|M_m^\dagger M_m|\psi\rangle = \langle\psi| \left(\sum_m M_m^\dagger M_m\right) |\psi\rangle = 1, \qquad \forall\psi \in \mathcal{H}$$

.

Remark III.3. *The number of measurement operators can be smaller, larger or equal to the dimension of the state space.*

Example III.12 (Measuring a qubit)**.** *To perform a measurement in the computational base, the projectors on the axes are used:*

$$M_0 = |0\rangle\langle 0|\,,$$
$$M_1 = |1\rangle\langle 1|\,.$$

Using the property of projection operators $P^2 = P$, one notes that these projectors satisfy the completeness equation:

$$M_0^\dagger M_0 + M_1^\dagger M_1 = M_0^2 + M_1^2 = M_0 + M_1 = |0\rangle\langle 0| + |1\rangle\langle 1| = I\,.$$

Measuring the qubit $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, using the projectors above, we get the result 0 with probability

$$P(0) = \langle\psi|M_0^\dagger M_0|\psi\rangle = (\alpha^*\langle 0| + \beta^*\langle 1|)\underbrace{|0\rangle\langle 0|}_{M_0}(\alpha|0\rangle + \beta|1\rangle) = \alpha^*\alpha = |\alpha|^2$$

and the result 1 with probability $P(1) = |\beta|^2$. Given that the result of the measurement was 0 (for example), the state collapses to the new state

$$|\psi_0\rangle = \frac{M_0|\psi\rangle}{\sqrt{P(0)}} = \frac{\alpha|0\rangle}{|\alpha|} = e^{j\theta_\alpha}|0\rangle = |0\rangle\,.$$

The last transition requires an explanation.

Remark III.4. *The postulates can be modified such that a state vector is defined up to a global phase, i.e., up to a multiplication by any $e^{j\theta}$. This does not change anything, since a global phase has no influence on measurements or their outcomes/probabilities. Therefore, we can always disregard a global phase.*

In a similar manner, to perform a measure in the basis $\{|+\rangle, |-\rangle\}$, one needs to use the measurement operators $M_0 = |+\rangle\langle +|$, $M_1 = |-\rangle\langle -|$. We can also perform the measurement by first "rotating" the state using the operator H, measuring in the computational base, and then "rotating" the state back using $H^\dagger = H$.

Example III.13 (Measuring one of two qubits)**.** *Let*

$$|\psi\rangle = \sum_{i,j\in\{0,1\}} \alpha_{ij}|ij\rangle$$

be the states of some composite system AB. We want to measure only the first qubit, in the computational base. The measurement operators which achieve this are:

$$M_0 = (|0\rangle\langle 0|) \otimes I = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) = \left(\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right)$$

$$M_1 = (|1\rangle\langle 1|) \otimes I = \ldots = \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & I \end{array}\right) .$$

The probabilty of the possible outcomes are:

$$\begin{array}{l} P(0) = \langle\psi|M_0|\psi\rangle = |\alpha_{00}|^2 + |\alpha_{01}|^2 \\ P(1) = \langle\psi|M_1|\psi\rangle = |\alpha_{10}|^2 + |\alpha_{11}|^2 \end{array} .$$

After measuring the result 0*, the state collapses to the state:*

$$|\psi_0\rangle = \frac{\alpha_{00}|00\rangle + \alpha_{01}|01\rangle}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}} = \frac{1}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}}|0\rangle\left(\alpha_{00}|0\rangle + \alpha_{01}|1\rangle\right) .$$

Definition III.7 (POVM)**.** *"Positive Opeartor Valued Measure" (POVM) is a measurement which is described by a set of* ***positive operators*** $\{E_m\}$ *which satisfy* $\sum_m E_m = I$*. The probability of the result* m *is then* $P(m) = \langle\psi|E_m|\psi\rangle$*.*

Remark III.5. *Definition III.7 is a special case of Postulate IV, taking* $E_m = M_m^\dagger M_m$*. We therefore use POVM when we care only about the probabilities of the measurement outcome, and not about the state after the measurement. It is sometimes more convenient to work with. The POVM used in Example III.12 is* $E_0 = M_0^\dagger M_0 = M_0$ *and* $E_1 = M_1^\dagger M_1 = M_1$*.*

Definition III.8 (Orthogonal Projection Operators)**.** *Orthogonal projectors are a set of operators* $\{P_m\}$*, which satisfy:*

$$\begin{cases} P_m = P_m^\dagger \\ P_m P_k = P_m \delta_{mk} \end{cases}$$

Measurement performed using sets composed of only orthogonal projectors are called ***Von-Neumann (VN) measurements****.*

Remark III.6. *Orthogonal measurements are a special case of the general measurement operation, described in Postulate IV. Furthermore, in this case the POVM and the measurement operators are exactly the same, as in Example III.12.*

Definition III.9 (Observable)**.** *Von-Neumann measurement can be described by a hermitian operator termed observable:*

$$M = \sum_m \underbrace{m}_{\substack{Measurement\ outcomes \\ =\ Eigenvalues\ of\ M}} \underbrace{P_m}_{\substack{Orthogonal \\ projectors}}$$

The exepcted value of the outcome of the measurement is:

$$\begin{aligned} < M > &= E(M) = \sum_m mp(m) \\ &= \sum_m m\langle\psi|P_m|\psi\rangle = |\psi\rangle \left(\sum_m mp(m)\right)|\psi\rangle \\ &= \langle\psi|M|\psi\rangle = tr\left(M|\psi\rangle\langle\psi|\right), \end{aligned} \tag{1.4}$$

where tr stands for trace (see Section II.3, Chapter 2).

Example III.14.

- *The observable* $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ *describes a measurement in the computational basis* $\{|0\rangle, |1\rangle\}$ *with the measurement outcomes (eigenvalues)* $\{\pm 1\}$*.*
- *The observable* $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ *describes a measurement in the basis* $\{|+\rangle, |-\rangle\}$ *with the measurement outcomes (eigenvalues)* $\{\pm 1\}$*. If we calculate the expected value of the outcome, when measuring the state* $|0\rangle$*, we have*

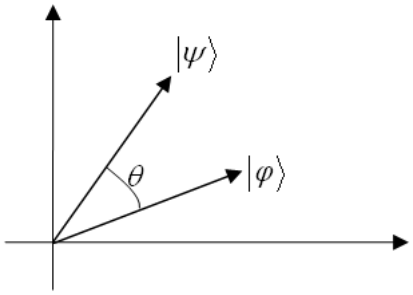
$$tr\left(X|0\rangle\langle 0|\right) = tr\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right) = tr\left(\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right) = 0.$$

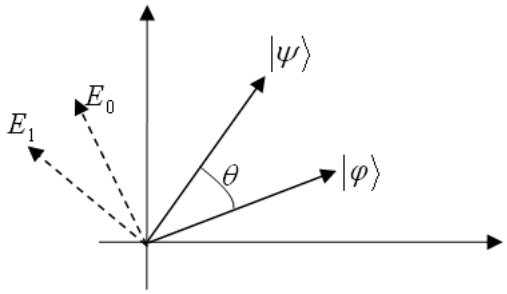
since we get $|+\rangle$ *and* $|-\rangle$ *with equal probabilities.*

Definition III.10 (Distinguishability)**.** *Two states are called seperable or distinguishable if there exists a measurement that can distinguish between the two states with certainty (w.p.* 1*).*

Example III.15. *A qubit is in a state* $|\varphi\rangle$ *or* $|\psi\rangle$*, and* $\langle\varphi|\psi\rangle \neq 0$*. We already know these states cannot be reliably distinguished. Suppose however we add an "erasure option", namely after measuring we can announce the state is* $|\varphi\rangle$ *or* $|\psi\rangle$*, but we can also announce we cannot decide. Of course, we cannot use an orthogonal measurement to that end, as that will yield only two results. Can we find another set of measurements that will give us a zero probability for a wrong detection? Consider the POVM*

$$\begin{aligned} E_0 &= a\left(I - |\varphi\rangle\langle\varphi|\right) \\ E_1 &= b\left(I - |\psi\rangle\langle\psi|\right) \\ E_2 &= I - E_0 - E_1 \end{aligned}$$

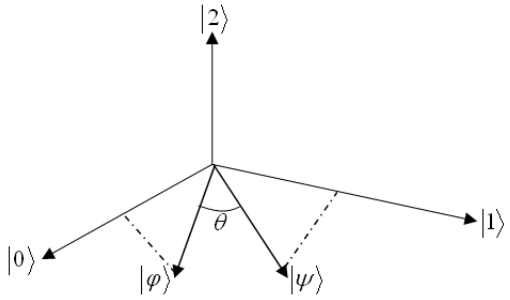




where $a, b \geq 0$ (it's possible to show that for small enough a and b, E_2 is also positive). The sum of the operators is 1 by definition. Using the following table we can tell the state the qubit is in by the measurement outcome:

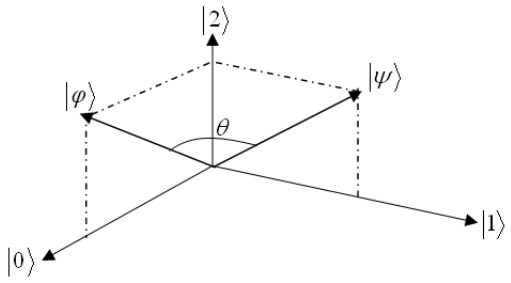
outcome	state
0	$\|\psi\rangle$
1	$\|\varphi\rangle$
2	*?*

"Physical" Interpretation: Adding an additional dimension, the states are now described in a three dimentional space:



After applying a unitary transformation in $\mathcal{H}_3$, we obtain:

Now, a VN measurement is performed onto the axes $\{|0\rangle, |1\rangle, |2\rangle\}$.



Theorem III.1 (Neumark Theorem)**.** *A general measurement can be realized by a VN measurement on a larger system, of a dimension equal to the number of measurement operators. The extra dimensions are called* ***"ancilla"****.*

The proof of this theorem is given as part of the proof of Theorem V.1, Chapter 2. We therefore immediately see that general measurements cannot give us any advantage in terms of distinguishability.

Theorem III.2. *Two quantum states are distinguishable iff they are orthogonal.*

Corollary 1.1. *A single qubit can contain no more than one classical bit of information (even on average over a block of qubits).*

Remark III.7. *There is a way of "squeezing" more information into a single qubit via entanglement, see Chapter 2.*

Chapter 2

Basic Communication Protocols and Mixed States

Summary by Amir Ingber.

I Superdense Coding

Suppose Alice wished to transmit two (classical) bits to Bob by using only a single qubit. As was claimed in Corollary 1.1, Chapter 1, this is not possible unless Alice and Bob share entangled states. In superdense coding (SDC), we assume that Alice and Bob share one of the following states:

$$\begin{aligned} |\beta_{00}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \\ |\beta_{01}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \\ |\beta_{10}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \\ |\beta_{11}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \end{aligned}$$

commonly known as *Bell* or *EPR states* (Einstein, Podolsky and Rosen). One may easily verify that these states form an orthonormal basis for $\mathcal{H}_4$ and that the two qubits, forming each of these states, are entangled.

An important property of the Bell states is that each state is reachable from any other state by a local operation (see Definition III.5, Chapter 1), in our case , a local operation in Alice's side only. For example, if we start at $|\beta_{00}\rangle$:

$$
\begin{aligned}
|\beta_{00}\rangle &\xrightarrow{I_1} |\beta_{00}\rangle,\\
|\beta_{00}\rangle &\xrightarrow{Z_1} |\beta_{01}\rangle,\\
|\beta_{00}\rangle &\xrightarrow{X_1} |\beta_{10}\rangle,\\
|\beta_{00}\rangle &\xrightarrow{X_1 Z_1} |\beta_{11}\rangle,
\end{aligned}
$$

where the operations I, Z, and X, are the Pauli matrices defined in Example III.8, Chapter 1.

Algorithm I.1 (SDC)**.** *Alice activates one of the four operations on her qubit according to her two classical bits. She then sends her qubit to Bob. Bob performs a measurement, using the Bell basis, and reconstructs the bits with probability* 1*.*

Remark I.1.

- *We see clearly that entanglement is an information theoretic resource (reminds of classical common randomness, only "stronger").*
- *A Bell state is commonly called an* entangled bit*, or* ebit*. The resource relation stemming from the SDC protocol is:*

$$1qubit + 1ebit \overset{\geq}{\Longrightarrow} 2cbits. \tag{2.1}$$

- *It turns out that* (2.1) *is tight, namely no more than* 2 *classical bits can be conveyed using a shared Bell state. If the qubits are not fully entangled, then the information that can be conveyed (on average) is between 1 and 2 cbits.*
- *Information security: if the transmitted qubit falls into the hands of eavesdroppers they can extract no information from it regarding the classical bits that Alice is trying to send: it can be shown that the result of any measurement applied to this qubit is independent of the cbits.*

II Quantum Teleportation

We saw that entanglement can be used in order to convey more classical information via transmission of qubits. We now show how entanglement allows the conveying of qubits using classical transmission.

Consider the case where Alice wishes to transmit a qubit $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$, by sending only classical information (bits) through a classical channel.

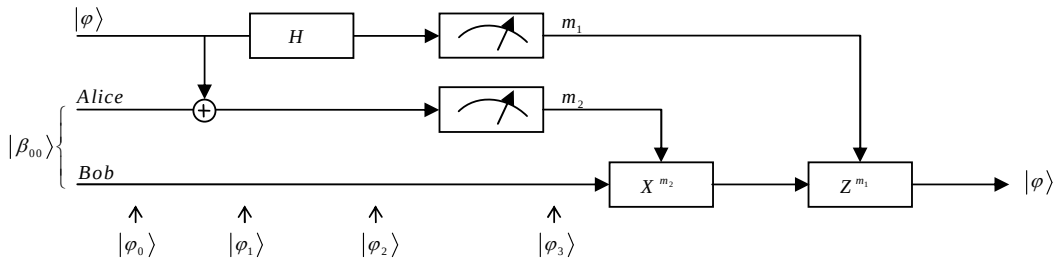


Figure 2.1: Quantum Teleportation

Note that α and β are not known to Alice (she cannot measure the qubit precisely), and even if they were known, it would require infinitely many bits to describe them exactly, as they can take any complex values as long as $|\alpha|^2 + |\beta|^2 = 1$ (any value on the unit sphere).

The protocol that enables this is called "quantum teleportation", and is described as follows:

Algorithm II.1 (Quantum Teleportation)**.** *As in the superdense coding protocol, Alice and Bob share an EPR state:*

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

The protocol is best explained in Fig. 2.1. We track the states as they are marked in the figure:

0 The state of the whole system is:

$$|\varphi_0\rangle = |\varphi\rangle|\beta_{00}\rangle = \frac{1}{\sqrt{2}}\left[\alpha|0\rangle(|00\rangle + |11\rangle) + \beta|1\rangle(|00\rangle + |11\rangle)\right],$$

where $|\varphi\rangle$ is the qubit that Alice tries to convey and $\beta_0 0$ is the shared Bell state.

1. *After a CNOT operation by Alice (local operation where the control qubit is $|\varphi\rangle$), the state of the system is:*

$$|\varphi_1\rangle = \frac{1}{\sqrt{2}}\left[\alpha|0\rangle(|00\rangle + |11\rangle) + \beta|1\rangle(|10\rangle + |01\rangle)\right].$$

2. *This operation is followed by a Hadamard operation (see Definition III.4, Chapter 1) on $|\varphi\rangle$:*

$$\begin{aligned} |\varphi_2\rangle &= \frac{1}{2}\left[\alpha(|0\rangle + |1\rangle)(|00\rangle + |11\rangle) + \beta(|0\rangle - |1\rangle)(|10\rangle + |01\rangle)\right] = \\ &= \frac{1}{2}\left[|00\rangle(\alpha|1\rangle + \beta|0\rangle) + |01\rangle(\alpha|1\rangle + \beta|0\rangle) + |10\rangle(\alpha|1\rangle - \beta|0\rangle) + |11\rangle(\alpha|1\rangle - \beta|0\rangle)\right]. \end{aligned}$$

3. *Finally Alice measures both of her qubits in the computational basis, and sends the result (2 classical bits) to Bob.*

4. *Bob perform a local operation on his qubit according to the following:*

$$\begin{aligned} 00 : \alpha|0\rangle + \beta|1\rangle &\rightarrow \quad \textit{this is } |\varphi\rangle, \\ 01 : \alpha|1\rangle + \beta|0\rangle &\rightarrow \quad \textit{use } X \textit{ and get } |\varphi\rangle, \\ 10 : \alpha|0\rangle - \beta|1\rangle &\rightarrow \quad \textit{use } Z \textit{ and get } |\varphi\rangle, \\ 11 : \alpha|1\rangle - \beta|0\rangle &\rightarrow \quad \textit{use } ZX \textit{ and get } |\varphi\rangle. \end{aligned}$$

Remark II.1.

- *Again, we see that entanglement is an information theoretic resource. Here we see that:*

$$2cbits + 1ebit \stackrel{\geq}{\Rightarrow} 1qubit. \tag{2.2}$$

- *No information was transmitted at a speed greater than the speed of light, since the classical information is required. If Bob measured his part of the ebit in the computational basis (or any other basis), before getting the bits from Alice, he would get a uniform distribution.*

- *If Alice does not know the value of $|\varphi\rangle$, then the protocol is optimal in the sense of the tradeoff of* (2.2) *(but not the only one).*

  In the case where Alice knows the value of $|\varphi\rangle$, only one classical bit is required (on average) for each qubit:

$$1cbits + 1ebit \Rightarrow 1qubit \quad (|\varphi\rangle \textit{ known})$$

  This is achieved by the use of large blocks. This case, in which Alice knows the value of $|\varphi\rangle$ is called "remote state preparation".

  - Alice's state $|\varphi\rangle$ is destroyed in the protocol (by the measurement), which is necessary for the teleportation. Hence no cloning is possible this way, or any other way, as explained in detail in the following section.

III No Cloning

Theorem III.1. *An unknown quantum state cannot be copied ("cloned") exactly.*

Unitary Opeartions. We now show that it is impossible to clone a (pure) quantum state by using a unitary operation (this is also true for general operations and mixed states).

Let us assume that there exists a unitary operator U such that for all $|\psi\rangle$:

$$U(|\psi\rangle|s\rangle) = |\psi\rangle|\psi\rangle,$$

where $|s\rangle$ is the 'slot' to which we wish to copy $|\psi\rangle$.

Now let us consider the cloning operation of two qubits, $|\psi\rangle$ and $|\phi\rangle$:

$$U(|\psi\rangle|s\rangle) = |\psi\rangle|\psi\rangle,$$

$$U(|\varphi\rangle|s\rangle) = |\varphi\rangle|\varphi\rangle.$$

By applying the inner product to both equations:

$$(\langle s|\langle\varphi|)U^\dagger U(|\psi\rangle|s\rangle) = (\langle s|\langle\varphi|)(|\psi\rangle|s\rangle) = \langle s|s\rangle\langle\varphi|\psi\rangle = \langle\varphi|\psi\rangle,$$

$$(\langle s|\langle\varphi|)U^\dagger U(|\psi\rangle|s\rangle) = (\langle\varphi|\langle\varphi|)(|\psi\rangle|\psi\rangle) = \langle\varphi|\psi\rangle^2,$$

which implies that either $\langle\varphi|\psi\rangle = 1$ or $\langle\varphi|\psi\rangle = 0$ must hold. Thus a cloning device can only clone states which are orthogonal to one another[1], and therefore cloning in general cannot be performed. □

IV Mixed States Density Matrix

The quantum states that we discussed so far are all *pure states*. A *mixed state* is an ensemble of (pure) states with some probability distribution: $\{p_i, |\psi_i\rangle\}$.

A simple example for a mixed state is when someone performs a measurement on a qubit, and does not tell us the resulting state. The resulting state in our point of view is an ensemble of the possible states, with respective probabilities. We shall restrict out attention to ensembles of countable (and even finite) number of states.

Let us see what happens to a mixed state when it is measured by the set of operators $\{M_m\}$. Since we have an ensemble of states, an expectation with respect to the states composing the ensemble (ψ_i, p_i) is required (according to the "law of total expectation")

$$p(m) = \sum_i p_i p(m|i) = \sum_i p_i \langle\psi_i|M_m^\dagger M_m|\psi_i\rangle = \sum_i p_i tr(M_m^\dagger M_m|\psi_i\rangle\langle\psi_i|) =$$

[1] Which is the case in classical physics

$$= tr(M_m^\dagger M_m \sum_i p_i |\psi_i\rangle\langle\psi_i|).$$

By defining $\rho \triangleq \sum_i p_i |\psi_i\rangle\langle\psi_i|$, we get the mixed state version of the 4th axiom:

$$p(m) = tr(M_m^\dagger M_m \rho) = tr(M_m \rho M_m^\dagger).$$

ρ is called the *density matrix* (d.m.) of the ensemble.

Given the result m, the measured state is now:

$$|\psi_i^m\rangle = \frac{M_m|\psi_i\rangle}{\sqrt{\langle\psi_i|M_m^\dagger M_m|\psi_i\rangle}}.$$

The density matrix is now:

$$\rho_m \triangleq \sum_i p(i|m)|\psi_i^m\rangle\langle\psi_i^m| = \sum_i p_i \frac{p(m|i)}{p(m)} \frac{M_m|\psi_i\rangle\langle\psi_i|M_m^\dagger}{\langle\psi_i|M_m^\dagger M_m|\psi_i\rangle} =$$
$$= \sum_i p_i \frac{M_m|\psi_i\rangle\langle\psi_i|M_m^\dagger}{tr(M_m\rho M_m^\dagger)} = \frac{M_m\rho M_m^\dagger}{tr(M_m\rho M_m^\dagger)}.$$

A mixed state goes through a unitary transformation, results in:

$$\rho \xrightarrow{U} \sum_i p_i U|\psi_i\rangle\langle\psi_i|U^\dagger.$$

This is an analog to the third axiom:

$$\rho \xrightarrow{U} U\rho U^\dagger.$$

Another analogy to the pure state properties is that for a number of independent systems in states ρ_i, the total state of the system is

$$\rho = \rho_1 \otimes \rho_2 \otimes ... \otimes \rho_n.$$

Nevertheless, in the following property, a real difference is observed. Let us define $\rho = |\psi\rangle\langle\psi|$ for a pure state.[2] It is known that for a pure state $tr(\rho^2) = 1$, whereas it can be shown that, for mixed states, $tr(\rho^2) < 1$.

Assertion IV.1. *ρ is a d.m. (for some ensemble) iff the following conditions hold:*

[2]This can be thought of as a mixed state with a degenerate randomness.

- $tr(\rho) = 1$.
- ρ *is positive.*

Proof. Direct: Let ρ be a d.m. for the ensemble $\{p_i, |\psi_i\rangle\}$. Then:

$$tr(\rho) = \sum_i p_i tr(|\psi_i\rangle\langle\psi_i|) = \sum_i p_i \langle\psi_i|\psi_i\rangle = \sum_i p_i = 1,$$

and

$$\langle\varphi|\rho|\varphi\rangle = \sum_i p_i \langle\varphi|\psi_i\rangle\langle\psi_i|\varphi\rangle = \sum_i p_i |\langle\varphi|\psi_i\rangle|^2 \geq 0.$$

Converse: Let ρ be some positive matrix satisfying $tr(\rho) = 1$. Then ρ can be decomposed (orthogonal diagonalization) into $\rho = \sum \lambda_i |i\rangle\langle i|$, where $\{|i\rangle\}$ is an orthonormal basis for ρ, and λ_i are the eigenvalues of ρ. Since $\sum_i \lambda_i = tr(\rho) = 1$, ρ can be viewed as the density matrix of the ensemble $\{|i\rangle, \lambda_i\}$. □

Measuring an Observable

For a pure state (see (1.4), Chapter 1):

$$< M >= \langle\psi|M|\psi\rangle = tr(M|\psi\rangle\langle\psi|).$$

Thus, for a mixed state, we have:

$$< M >= \sum_i \langle\psi_i|M|\psi_i\rangle = \sum_i p_i tr(M|\psi_i\rangle\langle\psi_i|) = tr(M\rho).$$

Remark IV.1.

- *There is an infinite number of ensembles with the same density matrix. Therefore it is impossible to tell what was the exact generating ensemble.*
- *If we perform a measurement on ρ with the operators $\{M_m\}$, then after the measurement, the density matrix (*without knowing the result*) is:*

$$\rho' = \sum_m p(m)\rho_m = \sum_m M_m \rho M_m^\dagger.$$

V The Reduced Density Matrix

Suppose we have an entangled state: $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. This state cannot be decomposed into a tensor product of two qubits - $|a\rangle \otimes |b\rangle$ (a bipartite system). However, there exists a 'marginal distribution' for each of the qubits. Thus, our aim is to answer the following question.

Question: Assume a composite system $A \otimes B$ in a state ρ^{AB}. does a d.m. ρ^A exists, s.t. for all observable M on the system A, the following holds:

$$tr(M\rho^A) = tr((M \otimes I_B)\rho^{AB})?$$

Note that asking the question on observables is sufficient since they give the statistics of each measurement (generalizations are possible).

Solution: the matrix ρ^{AB} is composed of vectors of the form $|ij\rangle\langle kl|$. Since the trace operator is linear, it is sufficient to examine the trace operation, when performed on the basis elements only:

$$\begin{aligned} tr((M \otimes I_B)|ij\rangle\langle kl|) &= tr((M|i\rangle)|j\rangle\langle kl|) = \langle kl|(M|i\rangle)|j\rangle = \langle k|M|i\rangle \cdot \langle l|j\rangle \\ &= tr(M|i\rangle\langle k|) \cdot tr(|j\rangle\langle l|) = tr(M \underbrace{(tr(|j\rangle\langle l|) \cdot |i\rangle\langle k|)}_{\text{an element of } \rho^A}). \end{aligned} \tag{2.3}$$

This leads to the following definition.

Definition V.1 (Partial Trace)**.** *Partial trace, performed on a sub-system B of a composite system AB, is a linear operator which satisfies:*

$$tr_B(|ij\rangle\langle kl|) \triangleq |i\rangle\langle k| \cdot tr(|j\rangle\langle l|).$$

This definition, along with (2.3), allows us to fully answer the question.

Definition V.2 (Reduced d.m.)**.** *The reduced density matrix (r.d.m.) of a sub-system A of a composite system AB is defined as:*

$$\rho^A \triangleq tr_B\left(\rho^{AB}\right)$$

Remark V.1.

- *Performing partial trace on B, is punned "tracing out" the system B.*
- *It can be shown that the matrix ρ^A is indeed a density matrix.*

Example V.1. *Consider the singlet state* $|\psi\rangle = \beta_{11} = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. *The density matrix of both qubits is a pure state:*

$$\rho = |\psi\rangle\langle\psi| = \frac{1}{2}(|01\rangle - |10\rangle)(\langle 01| - \langle 10|) = \frac{1}{2}\left[|01\rangle\langle 01| - |01\rangle\langle 10| - |10\rangle\langle 01| + |10\rangle\langle 10|\right].$$

The r.d.m. of the first qubit is given by:

$$\begin{aligned}
\rho^{(1)} &= tr_2(\rho) = \frac{1}{2}\left[|01\rangle\langle 01| - |01\rangle\langle 10| - |10\rangle\langle 01| + |10\rangle\langle 10|\right] = \\
&= \frac{1}{2}\left[|0\rangle\langle 0|tr(|1\rangle\langle 1|) - |0\rangle\langle 1|tr(|1\rangle\langle 0|) - |1\rangle\langle 0|tr(|0\rangle\langle 1|) + |1\rangle\langle 1|tr(|0\rangle\langle 0|)\right] = \\
&= \frac{1}{2}\left[|0\rangle\langle 0|\langle 1|1\rangle - |0\rangle\langle 1|\langle 0|1\rangle) - |1\rangle\langle 0|\langle 1|0\rangle + |1\rangle\langle 1|\langle 0|0\rangle\right] = \\
&= \frac{1}{2}\left[|0\rangle\langle 0| + |1\rangle\langle 1|\right] = \frac{1}{2}I.
\end{aligned}$$

This state is called "completely mixed state", since for any pure state $|\psi\rangle$ *we get* $\langle\psi|\frac{I}{2}|\psi\rangle$. *Hence, it is impossible to tell the difference between states.*

Corollary 2.1. *A sub-system of a system in a pure state will be in a mixed state iff there is entanglement between the parts of the system.*

Example V.1 was an example for the direct part of Corollary 2.1. The next example demonstrates the opposite direction.

Example V.2 (Product state)**.**

$$tr_2\left(|00\rangle\langle 00|\right) = |0\rangle\langle 0|.$$

In general, if a system is in a product state: $\rho = \rho^A \otimes \rho^B$, *then after tracing out one part of the system, we are left with the state of the other sub-system:*

$$tr_B(\rho) = {}_{TB}(\rho^A \otimes \rho^B) = \rho^A tr(\rho^B) = \rho^A,$$

.

Theorem V.1. Any *measurement performed on system A can be implemented by a* ***unitary*** *operation on the system AR, followed by the discarding (tracing out) of system R.*

Meaning that a measurement is equivalent to performing a *unitary* operation on a larger system and tracing out the complement half of the system.

Proof. Assume some measurement of A using the operators $\{M_m\}$. The state of the system, after measurement, is (see Remark IV.1): $\rho^{A'} = \sum_m M_m \rho^A M_m^\dagger$.

Define the system R with some orthonormal basis $\{|j\rangle\}$ with a one-to-one correspondence to the measurement operators $\{M_m\}$.

We decompose ρ^A using the "spectral decomposition": $\rho^A = \sum_i \lambda_i |a_i\rangle\langle a_i|$, where $\{|a_i\rangle\}$ is an orthonormal set.

Next we define the operator U in the following way:

$$U|a_i\rangle|j\rangle = \begin{cases} \sum_m M_m|a_i\rangle|m\rangle\,, & j=0 \\ \text{completion to orthonormal basis,} & j\neq 0 \end{cases}$$

that is, when system R in state $|0\rangle$, run over all the measurement operators in the first system and "store" all the possible results in system R ($|m\rangle$).

Note that U is indeed unitary, since it maps one orthonormal basis to another one:

$$\begin{aligned}\langle 0|\langle a_{i'}|U^\dagger U|a_i\rangle|0\rangle &= \sum_{m,m'}\langle a_{i'}|M_m^\dagger M_m|a_i\rangle\langle m'|m\rangle = \sum_m \langle a_{i'}|M_m^\dagger M_m|a_i\rangle \\ &= \langle a_{i'}|\sum_m M_m^\dagger M_m|a_i\rangle = \langle a_{i'}|a_i\rangle = \delta_{i,i'}.\end{aligned}$$

Place system R in state $|0\rangle\langle 0|$. Hence, the state of the system prior to measuring is $\rho^{AR} \triangleq \rho^A \otimes |0\rangle\langle 0|$.

After applying U, we have:

$$\begin{aligned}\rho^{A'R'} &= U\left(\rho^{AR}\right)U^\dagger = U\left(\rho^A\otimes|0\rangle\langle 0|\right)U^\dagger \\ &= U\left(\sum_i \lambda_i a_i\otimes|0\rangle\langle 0|\right)U^\dagger = \sum_i \lambda_i U\left(a_i\otimes|0\rangle\langle 0|\right)U^\dagger \\ &= \sum_i \lambda_i\left(\sum_m M_m|a_i\rangle|m\rangle\right)\left(\sum_{m'}\langle m'|\langle a_i|M_{m'}^\dagger\right) \\ &= \sum_{m,m'} M_m\left(\sum_i \lambda_i|a_i\rangle\langle a_i|\right)M_{m'}^\dagger\otimes|m\rangle\langle m'| = \sum_{m,m'} M_m\rho^A M_{m'}^\dagger\otimes|m\rangle\langle m'|.\end{aligned}$$

After tracing out the system R':

$$\rho^{A'} = tr_{R'}(\rho^{A'R'}) = \sum_{m,m'} tr_{R'}\left(M_m\rho^A M_{m'}^\dagger\otimes|m\rangle\langle m'|\right) =$$

$$= \sum_{m,m'} M_m\rho^A M_{m'}^\dagger\otimes\langle m'|m\rangle = \sum_m M_m\rho^A M_{m'}^\dagger,$$

as required.

We next show that Neumark's theorem (Theorem III.1, Chapter 1) stems from these arguments. When performing the measurements using the operators $\{I_{A'} \otimes |m\rangle\langle m|\}$ (measuring R' only), the result is m w.p.:

$$p(m) = tr(I_{A'} \otimes |m\rangle\langle m| \rho^{A'R'} I_{A'} \otimes |m\rangle\langle m|) = tr(M_m \rho^A M_m^\dagger \otimes |m\rangle\langle m|) = tr(M_m \rho^A M_m^\dagger),$$

and the state, given the result m is

$$\rho_m^{A'R'} = M_m \rho^a M_m^\dagger \otimes |m\rangle\langle m| / p(m),$$

which, after tracing out R', results in,

$$\rho_m^{A'} = \frac{M_m \rho^a M_m^\dagger}{p(m)},$$

as required. Note that $\{I_{A'} \otimes |m\rangle\langle m|\}$ is a Von-Neumann measurement, since both I'_A and $\{|m\rangle\langle m|\}$ are projectors, and thus also their tensor product. Moreover, $U\{I_{A'} \otimes |m\rangle\langle m|\}$ is a VN measurement as well, since U is unitary. Thus the proofs of both, Theorem V.1 and Theorem III.1, Chapter 1, are established. □

VI Remote State Preparation

Problem: Alice wishes to convey to Bob a quantum state $|\psi\rangle$ that is *known only to her*, using classic communication only.

Algorithm VI.1 (Basic algorithm)**.**

- *Suppose Alice shares with Bob the EPR state* $|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.

  Alice measures her ebit using the basis $\{|\psi\rangle, |\psi^\perp\rangle\}$. *The probability for* 0 *(*$|\psi\rangle$*) as the outcome of the measurement is* $\langle\psi|\rho^A|\psi\rangle = \langle\psi|\frac{I}{2}|\psi\rangle = \frac{1}{2}$, *which means that Bob has a* 50% *chance of getting* $|\psi\rangle$, *and* 50% *for* 1 *(*$|\psi^\perp\rangle$*). This is not sufficient for transmitting the qubit, because even if Alice uses her classical bit to convey the result of the measurement (*0 *or* 1*), in the case where Bob gets* $|\psi^\perp\rangle$, *there is no way to tell* $|\psi\rangle$ *from that.*

- *This issue is circumvented by the working with large blocks, similarly to what is done in CIT algorithms: Assume Alice wants to convey n qubits to Bob,* $|\psi_1\rangle, ..., |\psi_n\rangle$, *and that there is no limit on the number of ebits.*

  Alice performs for each *state* $|\psi_i\rangle$ *measurements on* $m = 2^{n+\log n}$ *ebits in her side, using the basis* $\{|\psi_i\rangle, |\psi_i^\perp\rangle\}$. *In total she gets mn results.*

The probability that all the jth measurements (for each of the states $\{|\psi_i\rangle\}$) of all the qubits will result in 0 ($|\psi_i\rangle$ and not $|\psi_i^{\perp}\rangle$) is clearly 2^{-n}. Therefore the probability that at least one series of measurements j from the m measurement series will contain only 0s (successes) for all the qubits is

$$p = 1 - (1 - 2^{-n})^m = 1 - (1 - 2^{-n})^{2^{n+\log n}} \underset{n\to\infty}{\longrightarrow} 1.$$

Therefore with high probability there will be one measurement in which all results are 0, and all that remains for Alice to do is to transmit to Bob, using classic bits, exactly which measurement is the successful one. This requires $\log_2 m = n + \log n$ bits. In total, we transmitted n qubits using $n + \log n$ bits, or $1 + \frac{\log n}{n}$ bits per qubit, with high probability.

- *Is it possible to achieve success with probability one, by using the quantum teleportation protocol if no measurement was sucessful. This will result in $p\left(1 + \frac{\log n}{n}\right) + 2(1 - p)$ bits per qubit. Since $p \to 1$, the mean number of bits per qubit approaches 1 when $n \to \infty$.*

Remark VI.1. *The number of ebits that were used here is exponential in the number of qubits. In 2004, a protocol that solves this issue was found, and allows a rate of one ebit (and one cbit) per qubit. The idea is that instead of performing many simple measurements, a single measurement is performed and it has $\approx 2^{n+\log n}$ results. Then it is shown that there exists a set of measurement operators (which dependence on $|\psi\rangle$!) on Alice's side, and a universal set of operations that Bob uses to reconstruct the qubit at his side.*

VII The EPR Paradox and Bell Inequalities

Einstein: "I, at any rate, am convinced that He does not throw dice"[3]
to what Max Born answered:
Born: "Do not tell God what to do".

In 1935, Einstein Podolsky and Rosen (EPR) authored a paper which claimed that quantum mechanics theory was *incomplete*. The main idea was the following.

Suppose Alice and Bob share a pair of qubits in the state

$$|\psi\rangle = |\beta_{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \tag{2.4}$$

If Alice performs a measurement in the basis $\{|0\rangle, |1\rangle\}$, she can predict that Bob would get the opposite results if he measured using the same basis. This might still sound reasonable from a

[3]Commonly paraphrased "God does not play dice".

classic point of view, since you might think that the bits were in opposite direction in advance, and the measurement only revealed that. However the following property is problematic in this context.

Property VII.1. *For each orthonormal basis of $\mathcal{H}_2$, $\{|v\rangle, |v^\perp\rangle\}$, the following holds:*

$$|\beta_{11}\rangle = \frac{1}{\sqrt{2}}(|vv^\perp\rangle - |v^\perp v\rangle),$$

up to a global phase.

Therefore, for each basis in which Alice performs the measurement, she knows that afterwards Bob will measure the opposite, using the same basis!

The members of EPR have found this issue problematic. Why? In their view, a physical theory must obey the "local realism" paradigm:

Reality every measurable quantity must have a single value that does not depend on performing the measurement itself.

Locality performing a measurement in some place cannot immediately change the results of another measurement in another distant location.

Suppose we agree with EPR, and these two properties indeed hold. If Alice and Bob share the EPR pair (2.4), then Alice can measure her qubit in the basis $\{|0\rangle, |1\rangle\}$, and Bob can measure his qubit in the basis $\{|+\rangle, |-\rangle\}$ at the same time. But due to Property VII.1, this means that Alice and Bob together know the measurement result of (say) Alice's qubit in two different bases, which stands in contradiction to the postulates of quantum mechanics! Unless of course, there is no fixed value to a measurable quantity before a measurement is performed (no reality), or measurement in one place can instantly effect a measurement result in a remote place (no locality). Therefore, either that one of the assumptions of the local realism paradigm is wrong, or the quantum mechanics theory is not complete. Various experiments were designed and carried out to resolve this issue. The basic idea behind them is to put some constraints (a.k.a. Bell inequalities; see Section VII) on the statistical result of the experiment under the local realism paradigm, and then find a way to do better with a quantum protocol using a shared entangled state (e.g., the EPR state). These experiments were clearly on the side of quantum mechanics.

Let us first analyze this from an information theoretic perspective: did Alice's measurement convey information to Bob?

We saw earlier that Bob's r.d.m. is given by

$$\rho^{(2)} = tr_1(\rho) = I/2.$$

What is $\hat{\rho}^{(2)}$ - Bob's d.m. after Alice's measurement?

Alice measures $|v\rangle$ w.p. $p(v) = 1/2$, and Bob will be in state $|v^\perp\rangle$. Therefore Bob's d.m. after Alice measures v is

$$\hat{\rho}^{(2)}_{v^\perp} = |v^\perp\rangle\langle v^\perp|,$$

and after Alice measures $v^\perp$:

$$\hat{\rho}^{(2)}_{v} = |v\rangle\langle v|,$$

and in total:

$$\hat{\rho}^{(2)} = \frac{1}{2}\hat{\rho}^{(2)}_{v^\perp} + \frac{1}{2}\hat{\rho}^{(2)}_{v} = \frac{1}{2}\left[|v^\perp\rangle\langle v^\perp| + |v\rangle\langle v|\right] = I/2 = \rho^{(2)},$$

exactly as before the measurement! Hence, from an information theoretic perspective nothing has happened at Bob's side as a result of Alice's measurement!

This idea can be generalized further:

Theorem VII.1 (Locality)**.** *Let AB be a composite system in state ρ^{AB}. Then any quantum operation (including unitary and measurements), performed on the system A alone, will not change the r.d.m. $\rho^B = tr_A(\rho^{AB})$ of system B.*

Remark VII.1.

- *For each measurement in system A, we get a different ensemble in system B, but all such ensembles have the same d.m. - the r.d.m. of B before the measurement.*
- *It can be deduced from the above that if there is no communication between A and B, we may always assume that system A was measured in some basis, from system B's point of view. This is called the* Principle of Implicit Measurements.
- *In general an operation in system A does not convey information to system B, and therefore does not imply communication at a speed greater than the speed of light.*

A Bell Inequality

Consider the following **experiment**: Alice and Bob get the (classical) bits a and b respectively. These bits are selected uniformly in an i.i.d. manner. Alice and Bob are each required to declare a bit, so that the following holds true:

$$\begin{cases} a = b = 1, & \text{Alice and Bob declare the same bit value,} \\ o.w. & \text{Alice and Bob declare } \textit{different} \text{ bit values.} \end{cases}$$

It is not difficult to see that in the classical case they cannot succeed with probability greater than 75% (even if they use common randomness). This bound is *a Bell inequality* (one of

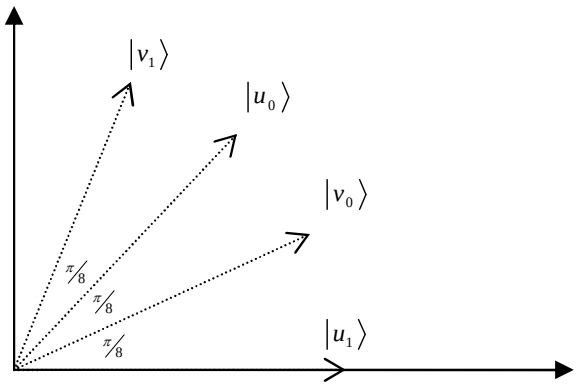


Figure 2.2: Measuring bases for Alice and Bob

several), and it turns out that it can be violated when using quantum states, as demonstrated by the following algorithm.

Algorithm VII.1. *Assume that Alice and Bob share the singlet state* $|\beta_{11}\rangle \triangleq \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. *We define two measurement bases for Alice and Bob, depicted in Fig. 2.2.*

Alice: $\{|0\rangle\langle u_0|, |1\rangle\langle u_0^\perp|\}, \{|0\rangle\langle u_1|, |1\rangle\langle u_1^\perp|\}$;

Bob: $\{|0\rangle\langle v_0|, |1\rangle\langle v_0^\perp|\}, \{|0\rangle\langle v_1|, |1\rangle\langle v_1^\perp|\}$.

Alice and Bob choose what measurement to use and what bit to declare, according to:

- *Alice measures her qubit using the basis* $\{|0\rangle\langle u_a|, |1\rangle\langle u_a^\perp|\}$, *according to the value of her cbit* a.
- *Bob measures his qubit using the basis* $\{|0\rangle\langle v_b|, |1\rangle\langle v_b^\perp|\}$, *according to the value of his cbit* b.
- *Alice and Bob declare on their measurement result:* 0 *if* $|0\rangle$ *was the outcome of the measurement, and* 1 *otherwise.*

Probability of success analysis:
We shall analyze the probability of success for all possible a, b *values.*

<u>$a = b = 0$:</u> *In this case, success is achieved when Alice and Bob declare different bits. Assume Alice measured* 0 ($|u_0\rangle$). *Bob is now at state* $|u_0^\perp\rangle$. *The chances for success is the probability that Bob measures* 1 ($|v_0^\perp\rangle$). *The probability for that is*

$$|\langle u_0^{\perp}|v_0^{\perp}\rangle|^2 = \cos^2(\frac{\pi}{8}) \approx 0.85.$$

(The probability of success, in case Alice measures 1 *(*$|u_0^{\perp}\rangle$*), is identical)*

$\underline{\{a=0, b=1\}, \{a=1, b=0\}}$*: The calculations in these cases are similar and yield the same result.*

$\underline{a=b=1}$*: Here, the goal is to output the same bit. If Alice measured* 0 *(*$|u_1\rangle$*), then the chance for success is again,*

$$|\langle v_1|u_1^{\perp}\rangle|^2 = \cos^2(\frac{\pi}{8}) \approx 0.85.$$

(Again, the probability of success, in case Alice measures 1 *(*$|u_1^{\perp}\rangle$*), is identical)*

We see here that the mentioned Bell inequality is violated, since the described strategy results in chances of $85\% > 75\%$ *for success.*

Remark VII.2. *The measurement of Alice did not convey any information to Bob or vice versa, but correlation was present between their measurement results, due to their common ebit.*

Corollary 2.2. *At least one of the principles, reality or locality, does not hold. This is because the coexistence of both principles brings us back to classical physics, under which Bell's inequalities hold. These inequalities can be violated however, in the framework of quantum mechanics.*

Chapter 3

Quantum Compression

Summary by Shai Machnes.

I The Quantum Compression Problem

Definition I.1 (Discrete Memoryless Quantum Source)**.** *A Discrete Memoryless Quantum Source is an ensemble* $\{|\psi_i\rangle, p_i\}_{i=1}^m$ *in a d-dimensional Hilbert space* H_d *with d.m.* $\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i|$.

Remark I.1. *A source is defined by an ensemble and not by a density matrix. However, as we discuss later in Theorem I.1, the source can be equivalently defined via the d.m. alone.*

In this chapter we address the basic question of quantum compression: How many qubits are needed (on average) per input state dimension ($\log_2(d)$), to represent the the source "reliably"? First, we need to define a notion of reliability, or fidelity, in the quantum setting.

I.1 Fidelity

Definition I.2 (Fidelity)**.** *The fidelity between two pure quantum states,* $|\psi\rangle$ *and* $|\phi\rangle$*, is defined as*

$$F(|\psi\rangle, |\phi\rangle) \triangleq |\langle\phi|\ \psi\rangle|^2 .$$

Suppose we compress the state $|\psi\rangle$, and get the reconstructed state $|\phi\rangle$. Then, after preforming a Von-Neumann measurement with the operators $\{|\psi\rangle\langle\psi|, \mathcal{I} - |\psi\rangle\langle\psi|\}$, we reproduce the source state $|\psi\rangle$ with probability $0 \leq F(|\psi\rangle, |\phi\rangle) = |\langle\phi|\ \psi\rangle|^2 \leq 1$.

Definition I.3 (Fidelity between pure and mixed states)**.** *The fidelity between a pure state* $|\psi\rangle$ *and a mixed state* ω *with orthonormal decomposition* $\omega = \sum_i q_j |\varphi_j\rangle \langle\varphi_j|$ *is defined as*

$$F(|\psi\rangle, \omega) \triangleq \sum_j q_j |\langle\psi|\varphi_j\rangle|^2 = \sum_j q_j \langle\psi|\varphi_j\rangle\langle\varphi_j|\psi\rangle = \langle\psi|\omega|\psi\rangle .$$

This definition suits the case where a single pure qubit $|\psi\rangle$ is reconstructed by an ensemble/mixed state ω. We generalize the definition of fidelity further, to match the case of an ensemble of pure states $\{|\psi_i\rangle\}$ that is reconstructed by an *ensemble of corresponding mixed states* $\{\omega_i\}$.

Definition I.4 (Average Fidelity)**.** *The average fidelity between the ensemble of pure states* $\{|\psi_i\rangle, p_i\}$ *and the* corresponding *mixed states* $\{\omega_i\}$ *is*

$$F^{(avg)} \triangleq \sum_i p_i \langle\psi_i| \omega_i |\psi_i\rangle .$$

Property I.1 (Fidelity properties)**.** *The following properties are true for all the three aforementioned definitions of fidelity (Definition I.2, Definition I.3, Definition I.4):*

- $0 \leq F \leq 1$.
- F *is* 1 *for identical states.*
- F *is* 0 *for orthogonal states.*
- $F \to 1$ *implies that any measurement performed on the reconstructed system will result in a probability distribution approaching (in total variation) that obtained by performing the same measurement on the original system, making the two systems "arbitrarily close" to being indistinguishable.*

I.2 Quantum Coding Scheme

A quantum coding scheme of rate R (qubits/state) and block length n for a quantum (memoryless) source over $\mathcal{H}_d$ is denoted $\mathcal{C}(\mathcal{H}_d, n, R, E, D)$.

Where the encoder E is a mapping

$$E : \mathcal{H}_d^{\otimes n} \longrightarrow \Delta(\mathcal{H}_{2^{nR}})$$

and the decided decoder D is a mapping

$$D : \Delta(\mathcal{H}_{2^{nR}}) \longrightarrow \Delta\left(\mathcal{H}_d^{\otimes n}\right) ,$$

and where $\Delta(\cdot)$ is the set of all density matrices over its argument Hilbert space.

Remark I.2. *E and D must satisfy some conditions we will define in the sequel, in order to be feasible by the laws of quantum mechanics*

Definition I.5 (Fidelity of Coding Scheme)**.** *Given an alphabet A and encoding scheme $\mathcal{C}$, the fidelity of the coding scheme is defined as*

$$F\left(A,\mathcal{C}\right)\triangleq\sum_{\{(\psi_i,p_i)\}\in A^{\otimes n}}p_iF\Big(\psi_i,D\circ E\left(\left|\psi_i\right\rangle\right)\Big)=\sum_{\{(\psi_i,p_i)\}\in A^{\otimes n}}p_i\langle\psi_i|D\circ E\left(\left|\psi_i\right\rangle\right)\left|\psi_i\right\rangle,$$

where $A^{\otimes n}$ is an ensemble of tensor products of n identical copies of the source A, with the associated multiplicative probabilities.

Definition I.6 (Achievable Rate)**.** *A rate R is achievable if there exists a sequence of coding schemes of fixed rate R such that*

$$\lim_{n\to\infty}F\left(A,\mathcal{C}_n\right)=1\,.$$

As usual, we will be interested in the infimum over all achievable rates.

Example I.1. *Consider an ensemble of $\{|0\rangle,|+\rangle\}$ with equal probabilities. From a classically point of view, it seems no compression is possible. However, note that these states are non-orthogonal and hence are indistinguishable, according to Theorem III.2, Chapter 1. Intuitively, this fact can be used for compression, as the decoder is not required to distinguish between the two.*

One can show that the best "guess", without information being conveyed, is $|\tilde{0}\rangle=\cos\left(\frac{\pi}{8}\right)|0\rangle+\sin\left(\frac{\pi}{8}\right)|1\rangle$, whereas the worst guess is $|\tilde{1}\rangle=\sin\left(\frac{\pi}{8}\right)|0\rangle-\cos\left(\frac{\pi}{8}\right)|1\rangle$. see Fig. I.1. Note that a guess of a constant qubit corresponds to $R=0$.

To evaluate an achievable rate for this scheme, when $n\to\infty$, let us look on the set

$$B=\left\{|\tilde{0}\rangle,|\tilde{1}\rangle\right\}^{\otimes n}$$

and express the ensemble states with respect to this basis:

$$\begin{aligned}|0\rangle&=\cos\left(\frac{\pi}{8}\right)|\tilde{0}\rangle+\sin\left(\frac{\pi}{8}\right)|\tilde{1}\rangle\\|+\rangle&=\cos\left(\frac{\pi}{8}\right)|\tilde{0}\rangle-\sin\left(\frac{\pi}{8}\right)|\tilde{1}\rangle\,.\end{aligned}$$

Now we can easily express a source block of length n, $|\psi\rangle\in A^{\otimes n}$ ($|\psi\rangle=|\psi_{i_1}\rangle|\psi_{i_2}\rangle\ldots|\psi_{i_n}\rangle$), in this basis:

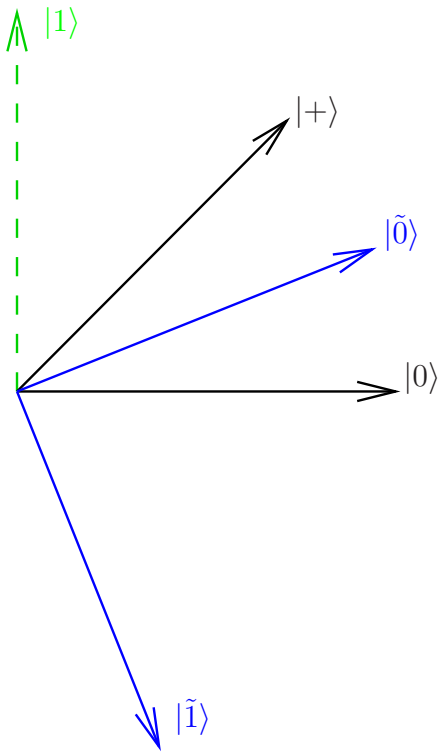


Figure 3.1: Best and worst compressions of single qubit with $R = 0$

$$
\begin{aligned}
|\psi\rangle &= \left(\cos\left(\frac{\pi}{8}\right)|\tilde{0}\rangle \pm \sin\left(\frac{\pi}{8}\right)|\tilde{1}\rangle\right) \otimes \cdots \otimes \left(\cos\left(\frac{\pi}{8}\right)|\tilde{0}\rangle \pm \sin\left(\frac{\pi}{8}\right)|\tilde{1}\rangle\right) \\
&= \sum_{\phi\in\{\tilde{0},\tilde{1}\}^n} \pm \sin\left(\frac{\pi}{8}\right)^{n_1(\phi)} \cos\left(\frac{\pi}{8}\right)^{n-n_1(\phi)} |\phi\rangle\,,
\end{aligned}
$$

where $n_1(\phi)$ denotes the number of $\tilde{1}$s in ϕ, where ϕ runs over all sequences of $\{|\tilde{0}\rangle, |\tilde{1}\rangle\}$.

Remark I.3. *All the source sequences have equal probabilities. If we would have taken non-equal probabilities for $|0\rangle$ and $|+\rangle$ we would have seen typical and non-typical sequences here, unlike this case, in which all sequences are typical.*

The fidelity between $|\psi\rangle$ and some $|\phi\rangle$ is

$$
|\langle\psi|\phi\rangle|^2 = \left(\sin^2\left(\frac{\pi}{8}\right)\right)^{n_1(\phi)} \left(\cos^2\left(\frac{\pi}{8}\right)\right)^{n-n_1(\phi)} = \lambda^{n_1(\phi)}(1-\lambda)^{n-n_1(\phi)}\,,
$$

*where $\lambda \triangleq \sin^2\left(\frac{\pi}{8}\right)$. Hence, the projection over $|\phi\rangle$ can be viewed as the probability of a **classical** binary source $\tilde{A}$ sequence with probabilities $\tilde{A} \sim (\lambda, 1-\lambda)$, consisting of $n_1(\phi)$ ones and $(n - n_1(\phi))$ zeros. From the AEP, there are $\sim 2^{nh_b(\lambda)}$ typical sequences of $\tilde{A}$ with probability approaching 1, for $n \to \infty$.*

Corollary 3.1. *There are* $\sim 2^{nh_b(\lambda)}$ *vectors in* B *which span a typical subspace, with the property that each source sequence has a projection on this subspace with length (arbitrarily) close to* 1*, i.e., fidelity close to* 1*. Hence this scheme give rise to the achievable rate:*

$$R = h_b(\lambda)$$

Remark I.4. *It can be shown that the following equality holds:* $\rho = |\psi\rangle\langle\psi| = (1-\lambda)|\tilde{0}\rangle\langle\tilde{0}| + \lambda|\tilde{1}\rangle\langle\tilde{1}|$*, meaning we were able to compress the source to the (classical) entropy of the eigenvalues of its d.m.. This is true for any quantum source, as we see next in Theorem I.1.*

I.3 AEP - Asymptotic Equipartition Property

Definition I.7 (Von-Neumann Entropy)**.** *The Von-Neumann (quantum) entropy of a quantum source with d.m.* ρ *is*

$$S(\rho) \triangleq -\sum_i \lambda_i \log \lambda_i = -tr(\rho \log \rho) \ , \tag{3.1}$$

where $\{\lambda_i\}$ *are the eigenvectors of* ρ*.*

Proof. The last equality holds since ρ is a positive symmetric matrix, and hence can be represented as $\rho = UDU^\dagger$, for some diagonal matrix D with a non-negative diagonal and unitary matrix U. Thus, $\log \rho = U \log DU^\dagger$ and $tr(\rho \log \rho) = tr(UD \log DU^\dagger) = tr(D \log D) = \sum_i \lambda_i \log \lambda_i$ □

Remark I.5.

- *The eigenvalues* $\{\lambda_i\}$ *always constitute a probability vector.*
- $S(\rho) = H(\{\lambda_i\}) \leq H(\{p_i\})$*, where* $\{p_i\}$ *are the ensemble probabilities; equality holds iff the ensemble states are orthogonal, since in this case we get a classical source. The difference between the two flanks may serve as a measure for the amount of separability between the ensemble states.*
- $S(\rho) = H(\{\lambda_i\}) \leq \log d$*, where* d *is the dimension of the system.*

Theorem I.1 (The Quantum Coding Theorem, Schumacher 1995 [12])**.** *Let* A *be a quantum source with density matrix* ρ *and let* $\epsilon, \delta > 0$*.*

1. ***Direct:*** *For every large enough* n*, there exists a coding scheme of rate* $R < S(\rho) + \delta$*, and fidelity* $F > 1 - \epsilon$*.*

2. ***Converse:*** *For every large enough n, the fidelity of any coding scheme with $R < S(\rho) - \delta$ satisfies $F < 1 - \epsilon$.*

Proof. 1. ***Achievability:***
ρ is a positive matrix with trace equal to 1 and hence the following orthogonal decomposition

$$\rho = \sum_i \lambda_i |v_i\rangle \langle v_i| , \tag{3.2}$$

for some orthonormal basis $\{v_i\}$ and where all eigenvectors are non-negative and satisfy: $\sum_i \lambda_i = 1$. Define a classical source $\tilde{A}$, such that

$$\tilde{A} = \{\lambda_i, |v_i\rangle\} .$$

This source has a (classical) entropy of

$$H\left(\tilde{A}\right) = H(\{\lambda_i\}) = S(\rho) .$$

Let us now examine the source

$$\tilde{A}_n \equiv \tilde{A}^{\otimes n}$$

with the associated density matrix $\rho^{\otimes n}$.

According to the AEP, for every $\epsilon, \delta > 0$ there exists a large enough n such that there is a subspace of $2^{n\left(H\left(\tilde{A}\right)+\delta\right)} = 2^{n(S(\rho)+\delta)}$ typical sequences with an overall fidelity of at least $1 - \epsilon$.

Therefore, there are $2^{n(S+\delta)}$ *typical* orthonormal eigenvectors of $\rho^{\otimes n}$, whose corresponding eigenvalues sum is equal or greater than $1 - \epsilon$. Denote this typical subspace by $T = T(n, \epsilon, \delta)$.

Encoding Scheme:
The encoding scheme acts on A_n.

1. Project the data onto the subspace T, using the POVM $\{P_T, I - P_T\}$, where P_T is the projector on T.

2. If the result is not in T (i.e., belongs to $T^{\perp}$ - the "non-typical" case), then encode using some predefined state $|err\rangle \in T$. This way an "error" is declared. Denote the state after these two stages by $|\phi\rangle$.

3. Apply the following unitary operator, which maps the typical subspace into $n(S + \delta)$ qubits:

$$U|\phi\rangle = \begin{cases} |\psi_\phi\rangle |0_{rem}\rangle & |\phi\rangle \in T \\ \text{unitary completion of } U & \text{otherwise} \end{cases}$$

where $|\psi_\phi\rangle$ is a state composite of $n(S+\delta)$ qubits, whereas $|0_{rem}\rangle = |0\ldots0\rangle$ makes the remaining $n - n(S+\delta)$ qubits.

Note: This is simply an enumeration of the elements in T.

4. Send the first $n(S+\delta)$ qubits of the resulting state.

Decoding Scheme:

1. Pad the state $|\psi_\phi\rangle$ with $|0_{rem}\rangle$.

2. Apply $U^\dagger \equiv U^{-1}$.

Performance Analysis The encoding rate is clearly $R = S(\rho) + \delta$. We shall now prove that $F \to 1$.

Assume some state $|a\rangle \in A_n$. This state can be decomposed into a sum of a vector $|t_a\rangle$ belonging to the typical subspace T and a vector $|t_a^\perp\rangle$ belonging to $T^\perp$:

$$|a\rangle = \alpha_a |t_a\rangle + \beta_a \left|t_a^\perp\right\rangle$$

.

After measuring with the POVM $\{P_T, I - P_T\}$, and replacing the resulting state with the predefined state $|err\rangle$, in case it belongs to $T^\perp$, we are left with the mixed state

$$\omega_a = |\alpha_a|^2 |t_a\rangle\langle t_a| + |\beta_a|^2 |err\rangle\langle err|$$

, which is also the state, recovered at the decoder. By calculating the fidelity between the (original) source state $|a\rangle$ and the quantized state ω_a, we arrive to the following lower bound:

$$\begin{aligned} F(|a\rangle, \omega_a) &= \langle a|\omega_a|a\rangle \\ &= |\alpha_a|^2 |\langle a|t_a\rangle|^2 + |\beta_a|^2 |\langle a|err\rangle|^2 \\ &\geq |\alpha_a|^2 |\alpha_a|^2 = |\alpha_a|^4 \geq 2|\alpha_a|^2 - 1 \\ &= 2\langle a|P_T|a\rangle - 1 = 2tr(P_T|a\rangle\langle a|) - 1, \end{aligned}$$

where the second inequality follows from the fact that $|\alpha_a|^4 - 2|\alpha_a|^2 + 1 = (|\alpha_a|^2 - 1)^2 \geq 0$.

To evaluate the fidelity of the coding scheme, we average over the ensemble:

$$
\begin{aligned}
F &\geq \sum_{(|a\rangle,p_a)\in A_n} p_a \left(2tr\left(P_T \left|a\right\rangle \left\langle a\right|\right) - 1\right) \\
&= 2 \sum_{(|a\rangle,p_a)\in A_n} tr\left(P_T p_a \left|a\right\rangle \left\langle a\right|\right) - 1 \\
&= 2tr\left(P_T \sum_{(|a\rangle,p_a)\in A_n} p_a \left|a\right\rangle \left\langle a\right|\right) - 1 \\
&= 2tr\left(P_T \rho^{\otimes n}\right) - 1 \\
&\geq 2\left(1-\epsilon\right) - 1 \\
&= 1 - 2\epsilon\,,
\end{aligned}
$$

where the second inequality holds since $\rho^{\otimes n}$ and P_T are diagonal w.r.t. the same basis, due to the definition of T, where P_T has $2^{n(S+\delta)}$ ones on its diagonal (and $2^{n(1-(S+\delta))}$ zeros), which correspond to a set of (the largest $2^{n(S+\delta)}$) eigenvectors, on the diagonal of $\rho^{\otimes n}$, which sum up to $(1-\epsilon)$ or more.

Remark I.6.

- *The suggested scheme provides a good fidelity without knowing the exact source state, but rather only the d.m. of the ensemble. It is optimal for all sources with the same density matrix. Hence, it is convenient to think of sources as density matrices.*

- *Nonetheless, high fidelity is maintained w.r.t. the true states emitted by the source, although these are not known and could not be generally determined.*

- *Suppose that it is given that $S(\rho_i) < S$, where $\{\rho_i\}$ are possible sources which satisfy $\rho_i\rho_j = \rho_j\rho_i$, i.e., simultaneously diagonizable. Then, it is possible to construct a "universal" scheme, which allows to quantize all the sources with (arbitrarily) good fidelity by considering a (polynomially) larger space, which consists of the intersection of the typical sets of all the sources.*[1]

2. ***Upper Bound:***
We shall restrict our attention to unitary decoders only, since the proof for the general case is much more complicated and lacks intuition. It can be found in [1].

[1]This is similar to the classical case, in which we consider all sources with bounded entropy, by taking all types with smaller entropy, whose number is polynomial in n, and thus, does not produce any penalty

Denote by $\Omega_a \in \Delta(\mathcal{H}_{2^{nR}})$ the d.m. of the state $|a\rangle$, of the source A_n, after encoding: the encoded source state:

$$\begin{aligned} \Omega_a &= E(|a\rangle) \in \Delta\left(\mathcal{H}_{2^{nR}}\right) , \\ \dim \Omega_a &\leq 2^{nR} . \end{aligned}$$

Unitary Decoder: Denote by ω_a the reconstructed state at the decoder:

$$\omega_a = D(\Omega_a) = U \left(\underbrace{\Omega_a}_{\dim \leq 2^{nR}} \otimes \underbrace{|0_{rem}\rangle\langle 0_{rem}|}_{\dim=1} \right) U^\dagger ,$$

for some unitary operator U. Note that $\dim(\omega_a) = \dim(\Omega_a) \leq 2^{nR}$, and therefore there exists some subspace Λ_n of $H_d^{\otimes n}$, of dimension 2^{nR}, such that:

$$\begin{aligned} \forall |a\rangle \in A_n : \quad \text{support}(\omega_a) &\subseteq \Lambda_n , \\ \dim(\Lambda_n) &= 2^{nR} . \end{aligned}$$

Hence, ω_a has an orthonormal decomposition with a basis $\left\{ |\xi_1^{(a)}\rangle, ..., |\xi_{2^{nR}}^{(a)}\rangle \right\}$ laying within the subspace Λ_n:

$$\omega_a = \sum_{j=1}^{2^{nR}} q_j^{(a)} |\xi_j^{(a)}\rangle\langle\xi_j^{(a)}| ,$$

where $\{q_j^{(a)}\}$ are non-negative and sum to 1.

The fidelity of $|a\rangle$ and its corresponding "quantized" state ω_a is

$$\begin{aligned} F\left(|a\rangle, \omega_a\right) &= \langle a|\omega_a|a\rangle = \sum_{j=1}^{2^{nR}} q_j^{(a)} \langle a|\xi_j^{(a)}\rangle\langle\xi_j^{(a)}|a\rangle = \sum_{j=1}^{2^{nR}} q_j^{(a)} \langle\xi_j^{(a)}|a\rangle\langle a|\xi_j^{(a)}\rangle \\ &\leq \sum_{j=1}^{2^{nR}} \langle\xi_j^{(a)}| \left(|a\rangle\langle a|\right) |\xi_j^{(a)}\rangle = \sum_{j=1}^{2^{nR}} \langle\xi_j^{(a)}|P_a|\xi_j^{(a)}\rangle = \sum_{j=1}^{2^{nR}} tr\left(P_a |\xi_j^{(a)}\rangle\langle\xi_j^{(a)}|\right) \\ &= tr\left(P_a \sum_{j=1}^{2^{nR}} |\xi_j^{(a)}\rangle\langle\xi_j^{(a)}|\right) = tr(P_a P_{\Lambda_n}) , \end{aligned}$$

where the inequality holds true since $q_j^{(a)} \leq 1$, and P_a and P_{Λ_n} are the projectors on $|a\rangle$ and the subspace Λ_n, respectively.

Remark I.7. *$|a\rangle$ does not necessarily lay inside Λ_n.*

The fidelity of the coding scheme can be upper bounded by:

$$F = \sum_{(p_a,|a\rangle)\in A_n} p_a \langle a|\omega_a|a\rangle \leq \sum_{(p_a,|a\rangle)\in A_n} p_a tr(P_a P_{\Lambda_n}) = tr\left(P_{\Lambda_n} \sum_{(p_a,|a\rangle)\in A_n} p_a |a\rangle\langle a|\right) = tr\left(P_{\Lambda_n}\rho^{\otimes n}\right). \tag{3.3}$$

If we denote by $\left\{|e_i\rangle_{i=1}^{d^n}\right\}$ the basis of the eigenvectors of $\rho^{\otimes n}$, with the corresponding eigenvalues $\{\mu_i\}$ which satisfy, w.l.o.g, we can rewrite the upper bound of (3.3) into:

$$F \leq tr\left(P_{\Lambda_n}\rho^{\otimes n}\right) = \sum_{i=1}^{d^n} \mu_i tr\left(P_{\Lambda_n}|e_i\rangle\langle e_i|\right) = \sum_{i=1}^{d^n} \mu_i \langle e_i|P_{\Lambda_n}|e_i\rangle.$$

Now observe the following two properties:

$$0 \leq \langle e_i|P_{\Lambda_n}|e_i\rangle \leq 1$$
$$\sum_{i=1}^{d^n} \langle e_i|P_{\Lambda_n}|e_i\rangle = \sum_{i=1}^{d^n} tr\left(P_{\Lambda_n}|e_i\rangle\langle e_i|\right) = tr(P_{\Lambda_n} I_n) = tr(P_{\Lambda_n}) = 2^{nR}.$$

Hence, there are at most 2^{nR} non-zero (strictly positive) summands in (3.3), meaning that the coding scheme fidelity can be further bounded by:

$$F \leq \sum_{i=1}^{2^{nR}} \mu_i < \epsilon, \tag{3.4}$$

where the last inequality holds for large enough n and stems from the classical AEP: if we consider a classical source $\tilde{A}$ with probabilities equal to the eigenvalues of ρ, then one sees that $\{\mu_i\}$ correspond to the probabilities of sequences of $\tilde{A}$ of length n. Since we assumed $R < S(\rho) \equiv H(\tilde{A})$, according to the (classical) AEP, the sum in (3.4) goes to zero, as exponentially less than $2^{nH(\tilde{A})}$ sequences are being summed. See [4, ch. 3]. □

Remark I.8.

- *No assumption was made on the encoding process. Hence, in the case that the encoder is aware of the exact source state ("visible coding"), the optimal achievable rate cannot improve over the case in which the encoder is ignorant of the exact state to be sent ("blind coding"). This holds true for general decoders as well.*
- *While proving Theorem I.1, we proved the quantum AEP, stated in Chapter 4, Section I.*

- *A more general decoder cannot improve the achievable rate. Nevertheless, it may improve the fidelity, at least for visible coding.*

- *The proof assumed block coding; however, there exists a converse for variable-length coding as well.*

- ***Quantum variable-length coding:*** *Much more involved due to the entanglement that might exist between the lengths of the code word and the information itself.*

  Example I.2. *Given an orthonormal decomposition* $\rho = \sum_i d_i |e_i\rangle$*, one might want to represent* $|e_i\rangle$*, using* $\log \frac{1}{d_i}$ *qubits. However, a quantum source state is, in general, a superposition of* $\{|e_i\rangle\}$*, and hence its length would be a quantum quantity, i.e., a superposition of values (which is not necessarily integer!). Furthermore, a measurement of the length may effect the source state. A solution to this problem was given by Schumacher and Westmoreland.*

- *Variable-length coding does not proivde* $F = 1$ *("lossless"), unless a classical channel exists, on which the lengths of the codewords can be conveyed and the source is visible to the encoder (which is aware of the code word lengths, without measuring). In this case it is possible to achieve rate beneath* $S(\rho)$*.*

Chapter 4

Quantum AEP and Von-Neumann Entropy Properties (Part I)

Summary by Yuval Kochman and Anatoly Khina.

I Quantum Asymptotic Equipartition Property

Theorem I.1 (Quantum AEP)**.** *Let $\epsilon, \delta > 0$.*
Direct: *for any n large enough there exists a typical linear subspace in $\mathcal{H}_d^{\otimes n}$ $T = T(n, \epsilon, \delta)$ of dimension $2^{nS(\rho)+\delta}$, which satisfies:*

$$tr\left(\rho^{\otimes n} P_T\right) \geq 1 - \epsilon. \tag{4.1}$$

Converse: *for every subspace $\widetilde{T} = \widetilde{T}(n, \epsilon, \delta)$ in $\mathcal{H}_d^{\otimes n}$ of dimension $2^{nS(\rho)-\delta}$, there exists n large enough, such that:*

$$tr\left(\rho^{\otimes n} P_{\widetilde{T}}\right) < \epsilon. \tag{4.2}$$

Remark I.1. *This theorems implies that $\rho^{\otimes n}$ contained in a space which is almost entirely spanned by a subspace of dimension $nS(\rho)$.*

II Purification and Schmidt Decomposition

II.1 Purification

Proposition II.1. *Suppose a state ρ^A of a quantum system A. It is possible to introduce another system R and define a pure state $|AR\rangle$ for the joint system AR, such that:*

$$\rho^A = tr_R\left(|AR\rangle\langle AR|\right). \tag{4.3}$$

$|AR\rangle$ is the "purification" of ρ^A.

Proof. Suppose ρ^A has orthonormal decomposition $\rho^A = \sum_i p_i |i^A\rangle\langle i^A|$. Introduce a system R with the same state space as system A and with an orthonormal basis $\left\{|i^R\rangle\right\}$. Define the state

$$|AR\rangle \triangleq \sum_i \sqrt{p_i}|i^A\rangle|i^R\rangle$$

in composite system AR. Tracing out system B, we obtain:

$$\begin{aligned} tr_R\left(|AR\rangle\langle AR|\right) &= tr_R\left(\sum_{i,j}\sqrt{p_i p_j}|i^A i^R\rangle\langle j^A j^R|\right) = \sum_{i,j}\sqrt{p_i p_j}|i^A\rangle\langle j^A| tr_R\left(|i^R\rangle\langle j^R|\right) \\ &= \sum_{i,j}\sqrt{p_i p_j}|i^A\rangle\langle j^A|\underbrace{\langle i^R|j^R\rangle}_{\delta_{i,j}} = \sum_i p_i |i^A\rangle\langle i^A| = \rho^A. \end{aligned}$$

Thus $|AR\rangle$ is a purification of ρ^A. □

Remark II.1.

1. *According to this proposition one can suggest that only pure states exist, and view mixed states as pure states where only part of the system is visible to us. This is a valid (and sometimes useful) point of view.*

2. *The converse of this proposition was already seen earlier Chapter 2, Section IV: measuring a pure state gives a mixed state if the result of the measurement is not known.*

3. *There is an infinite number of possible purifications of some (specific) mixed state, even when using the same system R. This can be easily seen from the construction used in the proof, which allows to use any orthonormal basis in R (there is an infinite number of such bases). See home assignment.*

4. *Let $|AR_1\rangle$ and $|AR_2\rangle$ be two purifications of a state ρ^A to a composite system AR. Then there exists a unitary transformation U_R acting on system R, such that:*

$$|AR_2\rangle = \left(I_A \otimes U_R\right)|AR_1\rangle$$

, where I_A is the identity transformation.

5. *The dimension of R is greater or equal to $rk\left(\rho^A\right)$. And, as seen from the proof, no higher dimension than that is needed.*

II.2 Schmidt Decomposition

Theorem II.1. *Let $|\psi\rangle$ be a pure state of a composite system AB. Then there exist two orthonormal sequences $\left\{|e_i^A\rangle\right\}, \left\{|f_i^B\rangle\right\}$ of systems A, B respectively, such that:*

$$|\psi\rangle = \sum_i \lambda_i |e_i^A\rangle |f_i^B\rangle,$$

where $\{\lambda_i\}$ are real none-negative numbers satisfying $\sum_i \lambda_i^2 = 1$, known as "Schmidt coefficients" and the number of strictly positive such coefficient is known as "Schmidt Number" and is denoted by $Sch(\psi)$.

Proof. Let $\left\{|j^A\rangle\right\}, \left\{|k^B\rangle\right\}$ be two orthonormal bases of systems A, B respectively. Hence $|\psi\rangle$ can be represented, according to these bases, as $\sum_{j,k} \alpha_{jk} |j^A\rangle |k^b\rangle$.

For the rest of the proof we shall introduce a new definition, a simple result that applies to it and use the SVD.

Definition II.1 (Representation Operator)**.** *The representation operator (or matrix) of $|\psi\rangle$ with respect to bases $\left\{|j^A\rangle\right\}, \left\{|k^B\rangle\right\}$, is defined as*

$$[|\psi\rangle] = [\psi] \triangleq \sum_{j,k} |j^A\rangle\langle k^B|.$$

Remark II.2. *Note that $[\psi]$ and $|\psi\rangle$ are isomorphic.*

Proposition II.2. *For every pair of unitary operators U, V acting on systems A, B respect., the following is true:*

$$[(U \otimes V)|\psi\rangle] = U\,[|\psi\rangle]\,V^\dagger. \tag{4.4}$$

Proof of Prop. II.2.

$$(U \otimes V)|\psi\rangle = \sum_{j,k} \alpha_{jk} \left(U|j^A\rangle\right) \left(V|k^B\rangle\right).$$

Hence, according to Definition II.1, we have

$$[(U \otimes V)|\psi\rangle] = \sum_{j,k} \alpha_{jk} U|j^A\rangle\langle k^B|V^\dagger = U \left(\sum_{j,k} \alpha_{jk}|j^A\rangle\langle k^B|\right) V^\dagger = U\,[\psi]\,V^\dagger,$$

Which concludes the proof. □

Theorem II.2 (Singular Value Decomposition)**.** *For every matrix A there exist unitary matrices U, V and a diagonal matrix Λ with a none-negative diagonal, satisfying:*

$$A = U\Lambda V^{\dagger}. \tag{4.5}$$

Using Theorem II.2, we can decompose $[\psi]$ into

$$[\psi] = U\Lambda V^{\dagger} \tag{4.6}$$

for some unitary matrices U, V and a diagonal matrix Λ with a none-negative diagonal, the last elements of whom we denote by $\{\lambda_i\}$. This allows us to define the pure state

$$|\varphi\rangle = \sum_i \lambda_i |i^A\rangle |i^B\rangle. \tag{4.7}$$

Using Definition II.1, we have

$$[\varphi] = \Lambda \tag{4.8}$$

and by combining (4.6) and (4.8) we have

$$[\psi] = U\,[\varphi]\,V^{\dagger} = \left[(U \otimes V)|\varphi\rangle\right],$$

where in the last equality we used Prop. II.2. Now using the isomorphism, mentioned in Remark II.2 we obtain

$$|\psi\rangle = (U \otimes V)|\varphi\rangle = \sum_i \lambda_i \underbrace{U|i^A\rangle}_{|e_i^A\rangle} \underbrace{V|i^B\rangle}_{|f_i^B\rangle} = \sum_i \lambda_i |e_i^A\rangle |f_i^B\rangle.$$

Finally note that the sequences $\left\{|e_i^A\rangle\right\}, \left\{|f_i^B\rangle\right\}$ are orthonormal since U, V are unitary. □

Corollary 4.1. *The following equalities hold true:*

$$\rho^A = tr_B\left(|\psi\rangle\langle\psi|\right) = \sum_i \lambda_i^2 |e_i^A\rangle\langle e_i^A|$$
$$\rho^B = tr_B\left(|\psi\rangle\langle\psi|\right) = \sum_i \lambda_i^2 |f_i^B\rangle\langle f_i^B|.$$

Hence, ρ^A and ρ^B have the same eigenvalues*, which differ from zero, λ_i^2 and the same dimension $Sch(\psi) = \dim tr_B\left(|\psi\rangle\langle\psi|\right)$.*

Remark II.3.

1. *A and B are not necessarily of the same dimension.*

2. *The sequences* $\left\{|e_i^A\rangle\right\}, \left\{|f_i^B\rangle\right\}$ *are, in general, state-dependent. That is, they can vary, depending on the state* $|\psi\rangle$.[1]

3. *The Schmidt coefficients* (λ_i) *(and hence also the Schmidt number) are invariant to local unitary transformations.*

4. *Corollary 4.1 implies that if a system AB is in pure state, then both of its subsystems A and B have the same VN entropy.*

5. *Purification can be thought of as a "correct" Schmidt basis selection. The state* $|AR\rangle$ *is chosen such that its Schmidt basis in system A, viz.* $|e_i^A\rangle$*, is the diagonalizing basis of* ρ^A*. Moreover, the corresponding Schmidt coefficients are the square-roots of the eigenvalues of* ρ^A*.*

6. $|AR_1\rangle$ *and* $|AR_2\rangle$ *are purifications of* ρ^A *in the system AR if and only if there exists a Unitary* U_R *acting on the R system such that* $|AR_2\rangle = (I_A \otimes U_R)|AR_1\rangle$*.*

7. *In purification, rank*$(R) \geq \rho^A$*, and the minimal rank is always sufficient.*

III Fidelity between Mixed States

In Definition I.3 we defined the fidelity between a pure state and a mixed state as

$$F\left(|\psi\rangle, \rho\right) \triangleq \langle\psi|\rho|\psi\rangle.$$

However a straightforward generalizing this definition to *two mixed states* would fail. Hence, a different fidelity definition is needed.

Definition III.1. *Let* ρ *and* σ *be two mixed states of system A. Then the fidelity between these states is defined as*

$$F(\rho,\sigma) \triangleq \min_{\substack{|\psi\rangle\in AR \\ tr_R(|\psi\rangle\langle\psi|)=\rho}} \max_{\substack{|\varphi\rangle\in AR \\ tr_R(|\varphi\rangle\langle\varphi|)=\sigma}} |\langle\psi|\phi\rangle|^2 \tag{4.9}$$

where ψ *and* φ *are some purification of* ρ *and* σ *to system AR, respectively.*

Remark III.1.

1. *One can prove that*

$$F(\rho,\sigma) = \max_{|\varphi\rangle} |\langle\psi|\varphi\rangle|^2 \tag{4.10}$$

holds for any fixed *purification of* ρ*, and the maximization is over all purifications of* σ*.*

[1] This can be seen in the proof of Theorem II.1, since the matrices U, V of the SVD depend on the exact matrix being decomposed.

2. *Note that the maximization in* (4.10) *is essential (unlike the outer minimization of* (4.9)*) since if we, for instance, were to use* min *instead of* max *in* (4.10)*, we would approach* 0 *by taking* $\dim(R)$ *to infinity.*

3. *Since the exact purification of* ρ *ins not important (as long as we fix it and maximize with respect to* $|\varphi\rangle$*), we can replace* (4.9) *with*

$$F(\rho,\sigma) \triangleq \max_{|\psi\rangle,|\varphi\rangle\in AR} |\langle\psi|\phi\rangle|^2 , \tag{4.11}$$

where the maximization is over all purifications $|\psi\rangle, |\varphi\rangle$ *of* ρ, σ *respectively.*

4. (4.11) *suggests that fidelity is symmetric in its arguments.*

Theorem III.1 (Uhlmann's Theorem)**.** *Suppose* ρ *and* σ *are two (mixed) states of the same system A. Then the fidelity between these two states satisfies:*

$$F(\rho,\sigma) = \left[tr\left(\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}}\right)\right]^2 . \tag{4.12}$$

Remark III.2. (4.12) *is reduced to the previously defined fidelity between a pure state and a mixed state (see Definition I.3, Chapter 3)*

$$F\left(|\psi\rangle,\rho\right) = \left[tr\left(\sqrt{\langle\psi|\rho|\psi\rangle|\psi\rangle\langle\psi|}\right)\right]^2 = \langle\psi|\rho|\psi\rangle .$$

Theorem III.2. *Suppose* ρ *and* σ *are two (mixed) states of the same system A. Then the fidelity between these two states satisfies:*

$$F(\rho,\sigma) = \min_{\{E_m\}} \sum_m \sqrt{p_m q_m} ,$$

where

$$p_m \triangleq tr(\rho E_m)$$
$$q_m \triangleq tr(\sigma E_m)$$

and the minimization is taken over all possible POVMs $\{E_m\}$*.*

The proofs of Theorem III.1 and Theorem III.2 can be found in [11].

IV Quantum Source Coding of an Ensemble of Mixed States

Definition IV.1 (i.i.d. Mixed Source)**.** *An ensemble* $\{\rho_i, p_i\}_{i=1}^M$ *in a d-dimensional Hilbert space* $\mathcal{H}_d$*, whose probabilities* $\{p_i\}$ *are drawn in an i.i.d. manner.*

If we denote by σ_n a source block of length n, i.e., $\sigma_n = \rho_{i_1} \otimes \rho_{i_2} \otimes \cdots \otimes \rho_{i_n}$, then we can define the average fidelity between this ensemble and a reconstruction state $\tilde{\sigma}_n$ as

$$F \triangleq \sum_{i_1,i_2,\ldots,i_n=1}^{M} P(\sigma_n) F(\sigma_n, \tilde{\sigma}_n) \ . \tag{4.13}$$

Remark IV.1. *This definition reduces to Chapter 3,Definition I.4, if all $\{\rho_i\}$ are pure.*

Theorem IV.1. *Let $\{\rho_i, p_i\}_{i=1}^{M}$ be an i.i.d. mixed source. Then for any $\epsilon > 0$, there exists n_0, such that for every $n > n_0$ there exists a coding scheme with rate*

$$R \geq S(\rho) - \sum_i p_i S(\rho_i) \tag{4.14}$$

and an average fidelity that satisfies $F > 1 - \epsilon$, where $\rho = \sum_i p_i \rho_i$.

Remark IV.2.

1. *The right-hand side of* (4.14) *is known as the Holevo quantity and is denoted by χ.*
2. *Since VN entropy is none-negative (Shannon entropy of the eigenvalues) we have $S(\rho) \geq \chi$ with equality iff all $\{\rho_i\}$ are pure. Hence, in the case of an ensemble of mixed states, a rate of $S(\rho)$ is achievable but not optimal, whereas when all $\{\rho_i\}$ are pure, the bound of* (4.14) *reduces to Chapter 3, Theorem I.1.*
3. *The lower bound of* (4.14) *is not tight.*
4. *In the case of a pure-state source, $S(\rho)$ was the optimal achievable rate both for visible and blind coding, i.e., both when the encoder knows or is ignorant of the state he compresses. In the case of mixed-states, there are cases in which a visible encoder achieves strictly better rates than a blind one.*

Example IV.1 ($S(\rho)$ is not tight)**.** *Suppose we want to compress an ensemble composed of only one* mixed *state $\{\rho_1\}$. Obviously, one can achieve a rate $R = 0$ with fidelity $F = 1$. However, $S(\rho)$ is strictly positive in this case due to the "impurity" of ρ_1. On the other hand, $\chi(\rho) = 0$ in this case, implying the bound* (4.14) *being tight.*

V Information Quantities and Properties

Definition V.1 (Quantum Relative Entropy)**.** *Let ρ and σ be two density matrices of the same system A. Then the "quantum relative entropy" between ρ and σ is defined as:*

$$S(\rho \| \sigma) \triangleq tr(\rho \log \rho) - tr(\rho \log \sigma) .$$

Theorem V.1 (Klein's Theorem)**.**

$$S\left(\rho\|\sigma\right) \geq 0, \tag{4.15}$$

with equality iff $\rho = \sigma$*.*

Proof. Decompose rho and σ to their orthonormal decompositions:

$$\rho = \sum_i p_i |i\rangle\langle i| \ , \quad \sigma = \sum_j q_j |j\rangle\langle j|. \tag{4.16}$$

Note that $\{|i\rangle\}$ and $\{|j\rangle\}$ are not (necessarily) the same basis. Hence we can rewrite the relative entropy as

$$S\left(\rho\|\sigma\right) = \sum_i p_i \log p_i - \sum_i \langle i|\rho \log \sigma|i\rangle,$$

using the identity $tr(A) = \sum_i \langle i|A|i\rangle$.

The decomposition in (4.16) implies $\langle i|\rho = p_i\langle i|$. Using this, we further the relative entropy:

$$\begin{aligned} S\left(\rho\|\sigma\right) &= \sum_i p_i \log p_i - \sum_i p_i \langle i| \log \sigma |i\rangle \\ &= \sum_i p_i \log p_i - \sum_i p_i \langle i| \left(\sum_j \log(q_j) |j\rangle\langle j| \right) |i\rangle \\ &= \sum_i p_i \left[\log p_i - \sum_j C_{ij} \log q_j \right], \end{aligned}$$

where $C_{ij} \triangleq \langle i|j\rangle\langle j|i\rangle = |\langle i|j\rangle|^2$ satisfies:

$$\begin{aligned} &\sum_i C_{ij} = \sum_i \langle i|j\rangle\langle j|i\rangle = tr\left(\sum_i |i\rangle\langle i||j\rangle\langle j| \right) = tr\left(I|j\rangle\langle j|\right) = \langle j|j\rangle = 1, \\ &\sum_j C_{ij} = 1, \end{aligned}$$

thus constituting a doubly-stochastic matrix. This enables us to define the probability vector $r_i \triangleq \sum_j C_{ij} q_j$ in the sequel

$$\begin{aligned} S\left(\rho\|\sigma\right) &= \sum_i p_i \left[\log p_i - \sum_j C_{ij} \log q_j \right] \overset{(a)}{\geq} \sum_i p_i \left[\log p_i - \log \left(\overbrace{\sum_j C_{ij} q_j}^{r_i} \right) \right] \\ &= \sum_i p_i \big[\log p_i - \log r_i \big] = \sum_i p_i \log \frac{p_i}{r_i} = D\left(\underline{p}\|\underline{r}\right) \overset{(b)}{\geq} 0 \, . \end{aligned}$$

(a) is due to Jensen's inequality and is held with equality iff C is a permutation matrix, i.e., ρ and σ have the same eigenvectors. Equality in (b) holds iff $p_i = q_i$ for every i, viz. identical eigenvalues. Thus equality in (4.15) holds iff $\rho = \sigma$. □

Definition V.2. *Let system AB be a quantum system in a state with entropy $S(AB) \equiv S(A,B)$. Then the* ***conditional VN entropy*** *is defiend as*

$$S(A|B) \triangleq S(A,B) - S(B), \tag{4.17}$$

and the ***quantum mutual information*** *as*

$$S(A;B) \triangleq S(A) - S(A|B) = S(A) + S(B) - S(A,B). \tag{4.18}$$

Remark V.1. *The quantum quantities of Definition V.2 are a natural generalization of the parallel classical quantities. However, in the quantum case, the meaning of these quantities is less apparent.*

Example V.1. *Let us look look on a Bell state $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ of a composite system AB, where the first qubit is of system A whereas the second one - of system B. Since $|\psi\rangle$ is a pure state, its VN entropy is $S(A,B) = 0$. However its reduced density matrices $\rho^A = \rho^B = \frac{I}{2}$ posses an entropy of $S(A) = S(B) = 1$, giving rise to the following, somewhat strange equalities:*

$$S(A|B) = S(B|A) = -1 < 0\,.$$

At first, it seems as though the conditional VN entropy (and consequently quantum mutual information) has no operative meaning. Luckily this is not the case. It can be shown that:

1. $\underline{S(B|A) \geq 0}$ *: Alice needs to transmit $S(B|A)$ qubits, so that Bob will be able to reproduce the mutual state (asymptotically with $F \to \infty$) - as in the classical case.*

2. $\underline{S(B|A)}$ *: Bob can produce the composite state by himself and* in addition *$|S(B|A)|$ ebits for for future communication (assuming there exists a* classical *channel with unlimited capacity).*

V.1 Von-Neumann Entropy Properties

Property V.1.

$$0 \leq S(A) \leq \log\left(\dim(A)\right)$$

where equality in the left inequality holds for a pure state and in the right one - for a maximally mixed state.

Proof. Define the dimension of A by $d \triangleq \dim(A)$. The left inequality and the corresponding equality condition are obvious. Let us prove the right inequality by looking at the following

relative entropy and applying Klein's Theorem:

$$0 \leq S\left(\rho^A \middle\| \frac{I}{d}\right) = -S(\rho^A) - tr\left(\rho^A \log \frac{I}{d}\right)$$
$$= -S(\rho^A) - \log \frac{1}{d} \cdot tr(\rho^A) - tr(\rho^A \underbrace{\log I}_{=0}) = -S(\rho^A) + \log d\,,$$

where in the last equality we used the fact that $tr(\rho^A) = 1$ for any density matrix.

Hence $S(\rho^A) \geq \log d$ with equality iff $\rho^A = \frac{I}{d}$ as indicated by (4.15). □

Property V.2. *If AB is in a pure state, then $S(A) = S(B)$.*[2]

Property V.3 (Additivity of Product States)**.** $S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$.

Property V.4 (Projective Measurements Increase Entropy)**.** *Let $\{P_i\}$ be some set of orthonormal projectors and ρ some density matrix. Then the entropy of the state after the measurement $\rho' \triangleq \sum_i P_i \rho P_i$ is at least as great as the original entropy, before measuring:*

$$S(\rho') \geq S(\rho),$$

with equality iff $\rho' = \rho$

Proof. By considering the relative entropy between ρ' and ρ and applying Klein's Theorem to it, we have

$$0 \leq S(\rho' \| \rho) = -S(\rho) - tr\left(\rho \log \rho'\right).$$

Let us now prove that $-tr(\rho \log \rho') = S(\rho')$. Using the relations of projective operators $P_i^2 = P_i$ and the completeness relation $\sum_i P_i = \sum_i P_i^2 = I$, we obtain

$$-tr\left(\rho \log \rho'\right) = -tr\left(\overbrace{\sum_i P_i^2}^{=I} \rho \log \rho'\right) = -\sum_i tr\left(P_i^2 \rho \log \rho'\right) = -\sum_i tr\left(P_i \rho \log \rho' P_i\right)$$
$$= -tr\left(\sum_i P_i \rho \log \rho' P_i\right).$$

Note that $\{P_j\}$ commute ρ' and hence with $\log \rho'$ as well. This is true due to $P_i P_j = P_i \delta_{i,j}$:

$$\rho' P_j = \left(\sum_i P_i \rho'\right) P_j = P_j \rho P_j = P_j \rho'.$$

Thus

$$-tr\left(\rho \log \rho'\right) = -tr(\rho' \log \rho') = S(\rho').$$

Thus we proved that $S(\rho') \geq S(\rho)$, with equality iff $\rho = \rho'$ □

[2]This was already stated and proved in Remark II.3.

Remark V.2. *A none-projective measurement might decrease entropy.*

Property V.5 (Subadditivity)**.** *Let A and B be distinct quantum systems in a joint state ρ^{AB}. The the following inequality holds:*

$$S(A,B) \leq S(A) + S(B) \tag{4.19}$$

with equality iff $\rho^{AB} = \rho^A \otimes \rho^B$.

Proof. We shall look at the entropy between the joint state $\rho^A B$ and the tensor state $\rho^A \otimes \rho^B$ and apply Klein's Theorem:

$$0 \leq S\left(\rho^{AB} \middle\| \rho^A \otimes \rho^B\right) = -S(\rho^{AB}) - tr\left(\rho^{AB} \log(\rho^A \otimes \rho^B)\right).$$

Let us consentrate on the second elemnt on the right-hand side:

$$\begin{aligned} tr\left(\rho^{AB} \log(\rho^A \otimes \rho^B)\right) &= tr\left(\rho^{AB} \log\left\{(\rho^A \otimes I)(I \otimes \rho^B)\right\}\right) \\ &= tr\left(\rho^{AB} \log(\rho^A \otimes I)\right) + tr\left(\rho^{AB} \log(I \otimes \rho^B)\right), \end{aligned}$$

where the last equality is true since the identity operator can be decomposed into an orthonormal decomposition by any orthonormal basis.

Now using the fact that $tr(\bullet) = tr_A\big(tr_B(\bullet)\big)$, we come to

$$tr\left(\rho^{AB} \log(\rho^A \otimes I)\right) + tr\left(\rho^{AB} \log(I \otimes \rho^B)\right) = tr\left(\rho^A \log \rho^A\right) + tr\left(\rho^B \log \rho^B\right) = -S(A) - S(B).$$

Combining all results and leads to the desired result. □

Remark V.3. *Note that in the course of the proof we also proved the following identity:*

$$S(A;B) = S(A) + S(B) - S(A,B) = S(\rho^{AB} \| \rho^A \otimes \rho^B).$$

Chapter 5

Von-Neumann Entropy Properties (Part II)

Summary by Yael Ben−Haim.

I Further Properties of the Von-Neumann Entropy

The triangle inequality is the quantum analogue of the classical inequality $H(X,Y) \geq H(X)$. It states the following.

Property I.1 (Triangle (Araki-Lieb) inequality)**.** *For any pair of systems A and B,*

$$S(AB) \geq |S(A) - S(B)| \,.$$

Proof. Introduce a system R that purifies the system AB. The subadditivity property (Property V.5, Chapter 4) implies that

$$S(R) + S(A) \geq S(AR) \,.$$

Since the system ABR is pure, by Property V.2, Chapter 4,

$$S(R) = S(AB)$$

and

$$S(AR) = S(B) \,.$$

Substituting (I) and (I) in (I), we obtain

$$S(AB) \geq S(B) - S(A) \,.$$

Symmetrical arguments show that

$$S(AB) \geq S(A) - S(B),$$

which completes the proof. □

Property I.2. *Let $\{p_i, \rho_i\}$ be an ensemble of states. Classical information theory is obtained when the states ρ_i are pure and orthogonal. Otherwise, the incapability to distinguish between states increases the uncertainty. This is made precise in the following theorem.*

Let $\{p_i, \rho_i\}$ be an ensemble of states. Then

$$S(\sum_i p_i \rho_i) \leq \underbrace{\sum_i p_i S(\rho_i)}_{\substack{\text{penalty for} \\ \text{indistinguishability}}} + H(\{p_i\}),$$

with equality if and only if the states ρ_i are orthogonal, i.e., they have supports on orthogonal spaces.

Proof. We prove here only that orthogonality implies equality. The rest is proved in the home assignment.

Let $\left\{\lambda_i^j\right\}_j$ and $\left\{|e_i^j\rangle\right\}_j$ be the eigenvalues and orthogonal eigenvectors of ρ_i. The orthogonality of the states ρ_i implies that $\left\{p_i\lambda_i^j\right\}_{i,j}$ are the eigenvalues of $\sum_i p_i\rho_i$ with orthonormal eigenvectors $\left\{|e_i^j\rangle\right\}$. Hence

$$S\left(\sum_i p_i\rho_i\right) = -\sum_{i,j}\left(p_i\lambda_i^j\right)\log\left(p_i\lambda_i^j\right) .$$

Since $\sum_j \lambda_i^j = 1$ for all i, we obtain

$$S\left(\sum_i p_i\rho_i\right) = -\sum_i p_i \log p_i - \sum_i p_i \sum_j \lambda_i^j \log \lambda_i^j \tag{5.1}$$

$$= H(\{p_i\}) + \sum_i p_i S(\rho_i) . \tag{5.2}$$

□

Property I.3 (Concavity)**.** *Let $\{p_i, \rho_i\}$ be an ensemble of states. Then*

$$S(\sum_i p_i\rho_i) \geq \sum_i p_i S(\rho_i) .$$

Proof. Let A be the system of the states $\{\rho_i\}$. Introduce a system B whose state space has an orthonormal basis $|i\rangle$, with one-to-one correspondence to the states ρ_i and define the composite state:

$$\rho^{AB} = \sum_i p_i \rho_i \otimes |i\rangle\langle i| \,.$$

Then

$$S(A) = S\left(\sum_i p_i \rho_i\right)$$

and

$$S(B) = S\left(\sum_i p_i |i\rangle\langle i|\right) = H(\{p_i\}) \,.$$

Remark I.1. *The states $\{\rho_i\}$ are not necessarily orthogonal, whereas the $\{|i\rangle\langle i|\}$ - are. Hence, the states $\{\rho_i \otimes |i\rangle\langle i|\}$ are orthogonal as well.*

Since the states $\rho_i \otimes |i\rangle\langle i|$ are orthogonal, we have, according to Property I.2

$$S(AB) = S(\rho^{AB}) = \sum_i p_i S(\rho_i \otimes |i\rangle\langle i|) + H(\{p_i\}) \,.$$

By Property V.3, Chapter 4,

$$S(AB) = \sum_i p_i S(\rho_i) + \sum_i p_i S(|i\rangle\langle i|) + H(\{p_i\}) \,.$$

Since $|i\rangle$ is a pure state, we have $S(|i\rangle\langle i|) = 0$, and

$$S(AB) = \sum_i p_i S(\rho_i) + H(\{p_i\})$$

follows. Applying Property V.5, Chapter 4: $S(AB) \leq S(A) + S(B)$, we obtain

$$\sum_i p_i S(\rho_i) \leq S(\sum_i p_i \rho_i) \,.$$

□

So far we have encountered two properties, Property I.1 and Property I.3, whose proofs involved the introduction of an auxiliary system. This is a useful method in quantum information theory. In Property I.3, B allows us to store the knowledge of the index i, which is unknown in A. In other words, one can regard B as the system of the creator of the state of ρ^A.

Property I.4 (Strong subadditivity)**.** *For any triplet of systems A, B, C:*

$$S(ABC) + S(B) \leq S(AB) + S(BC) \,.$$

Remark I.2.

- *All the known proofs to Theorem I.4 are very difficult, and none of them will be given here. See [11, ch. 11.4] for a proof and references. In fact, strong subadditivity was an open conjecture for five years (from 1968 until 1973).*
- *The conditions for equality are also involved. See [9].*
- *Atrong subadditivity is trivial in the classic case and holds with equality iff $A \to B \to C$, i.e., iff A, B, C form a Markov chain.*

Property I.5 (Conditioning reduces entropy)**.** *For any triplet of systems A, B, C:*

$$S(A|BC) \leq S(A|B)\,.$$

Proof. By definition,

$$S(ABC) = S(A|BC) + S(BC)$$

and

$$S(B) = S(AB) - S(A|B)\,.$$

The claim follows by substituting (I) and (I) in (I.4). □

Corollary 5.1. *Property I.5 implies that quantum conditional mutual information is non-negative.*

The following property follows easily from Property I.4.

Property I.6 (Discarding a system does not increase mutual information)**.** *For any triplet of systems A, B, C:*

$$S(A;B) \leq S(A;BC)\,.$$

Property I.7 (Unitary operators preserve entropy)**.** *For any state ρ and unitary operator U:*

$$S(\rho) = S(U\rho U^\dagger)\,.$$

Proof. Since U is unitary, the set of eigenvalues of ρ is equal to the set of eigenvalues of $U\rho U^\dagger$. The claim follows from the definition of entropy. □

The following property follows immediately from Property I.7 and the definition of mutual information (Definition V.2, Chapter 4).

Property I.8 (Local unitary operators preserve mutual information)**.** *Let U_A, U_B be unitary operators on systems A and B, respectively. Denote by A', B' the systems obtained from applying $U_A \otimes U_B$ to AB. Then*

$$S(A;B) = S(A';B')\,.$$

Property I.9 (Local measurements do not increase mutual information)**.** *Suppose that systems A, B are measured locally and denote by A', B' the systems after these local measurements. Then*

$$S(A';B') \leq S(A;B)\,,$$

if the outcomes of the measurements are not known.

Proof. A measurement is equivalent to an introduction of an auxiliary system, followed by a unitary operation and the discarding of the auxiliary system, afterwards. By Property I.8, the unitary operation preserves mutual information; by Property I.6, discarding the auxiliary system does not increase mutual information, which completes the proof. □

The following three properties are difficult to prove and therefore are state without such.

Property I.10 (Quantum relative entropy is jointly convex)**.** *Let $\{\rho_i\}$ and $\{\sigma_i\}$ be some state sets and p_i - a set of numbers. Then,*

$$S\left(\sum_i p_i\rho_i || \sum_i p_i\sigma_i\right) \leq \sum_i p_i S(\rho_i||\sigma_i)\,.$$

Property I.11 (Monotonicity of relative entropy)**.** *Let ρ^{AB}, σ^{AB} be any two states in a composite system AB. Then*

$$S(\rho^A||\sigma^A) \leq S(\rho^{AB}||\sigma^{AB})\,.$$

Property I.12 (Concavity of conditional entropy)**.** *Let ρ^{AB} be any state in a composite system AB. Then the conditional entropy $S(A|B)$ is concave in ρ^{AB}.*

II Accessible information

We have seen that it is impossible to distinguish between non-orthogonal states (or clone such). Hence, we would like to quantify how much, two states, are distinguishable.

Consider the following scenario: a classical source X with probability distribution $(p_0, \ldots, p_{n-1})$ is encoded to the quantum mixed states $(\rho_0, \ldots, \rho_{n-1})$, respectively. Alice chooses an index $x \in \{0, \ldots, n-1\}$ with probability p_i, and sends the corresponding state ρ_i to Bob. Bob performs a measurement, resulting in a classical random variable Y, whose possible values are

in the range $\{0,\ldots,m-1\}$. The amount of information that Bob acquires about X is $I(X;Y)$. Obviously, Bob should choose the POVM that maximizes $I(X;Y)$, as the (classical) MI is a good measure for the information acquired by measurement: X can be perfectly reconstructed iff $I(X;Y)=H(X)$; according to Fano's inequality:

$$h_b(p_e)+p_e\log(n-1)\geq \underbrace{H(X|Y)}$$
$$I(X;Y)=H(X)-\underbrace{H(X|Y)}\,.$$

This brings us to the following definition.

Definition II.1 (Accessible information)**.** *With the setting described above, the accessible information is defined as*

$$I_{acc}\triangleq \max_{\{E_y\}} I(X;Y),$$

where the maximization is over all possible POVMs $\{E_y\}$*.*

Our task is to estimate the amount of accessible information. We first note that $I_{acc}\leq H(X)$, with strict inequality when the states ρ_i are not orthogonal. We proceed to the introduction of a more sophisticated estimate.

Theorem II.1 (The Holevo bound)**.** *With the setting described above, we have*

$$I_{acc}\leq \chi,$$

where $\rho=\sum_x p_x\rho_x$ *and* $\chi\triangleq S(\rho)-\sum_x p_x S(\rho_x)$ *is the "Holevo quantity"; equality holds iff* $\{\rho_x\}$ *commute.*

Proof. Let us denote the measurement operators by $\{M_y\}_{y=0}^{m-1}$, i.e., $E_y=M_y^\dagger M_y$. Recall that the state after the measurement is

$$\rho'=\sum_y M_y\rho M_y^\dagger\,.$$

Define the following three systems:

- P ("preparation system") - a system with orthonormal basis $\{|x\rangle\}_{x=0}^{n-1}$, in correspondence to the values of X.
- Q ("storing system") - the system that Alice gives to Bob, i.e., the system of the states ρ_x.
- R ("results accumulating system") - a system with orthonormal basis $\{|y\rangle\}_{y=0}^{m-1}$, in correspondence to the values of P.

Alice and Bob prepare the following state:

$$\rho^{PQR} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x \otimes |0\rangle\langle 0| \,.$$

Remark II.1. *The part $p_x|x\rangle\langle x| \otimes \rho_x$ represents the choice and encoding of x by Alice, while $|0\rangle\langle 0|$ represents the initialization of R, done by Bob. Tracing-out PR, we receive exactly the system that is sent by Alice to Bob.*

Define unitary operators on the system R:

$$T_y|i\rangle = |i+y\rangle, \tag{5.3}$$

where addition is modulo m; $\{|i\rangle\}$ orthonormal basis; note that (5.3) implies $T_y^\dagger|i\rangle = |i-y\rangle$.

Assertion II.1. *The set $\{M_y \otimes T_y\}$ is a set of measurement operators on QR.*

Proof. (Assertion)
Indeed, it is easy to check that $M_y \otimes T_y \geq 0$, and

$$\sum_y (M_y \otimes T_y)(M_y \otimes T_y)^\dagger = \sum_y M_y M_y^\dagger \otimes T_y T_y^\dagger \tag{5.4}$$

$$\stackrel{(a)}{=} \left(\sum_y M_y M_y^\dagger\right) \otimes I_R \tag{5.5}$$

$$\stackrel{(b)}{=} I_Q \otimes I_R = I_{QR}, \tag{5.6}$$

where (a) is true because T_y is unitary for all y, and (b) is because $\{M_y\}$ is a set of measurement operators, and thus $\sum_y M_y M_y^\dagger = I_Q$ due to the completeness equation (Postulate IV, Chapter 4). □

Thus, we can perform a measurement of ρ^{PQR} with the set $\{I \otimes M_y \otimes T_y\}$, to obtain the state

$$\rho^{P'Q'R'} = \sum_x \sum_y p_x |x\rangle\langle x| \otimes M_y \rho_x M_y^\dagger \otimes |y\rangle\langle y| \,.$$

The rationale behind this measurement is to measure Q with the set $\{M_y\}$ and store the result in R.

Remark II.2. *Note that this measurement is consistent with a direct measurement of ρ:*

$$\rho^{Q'} = tr_{P'Q'}\left(\rho^{P'Q'R'}\right) = ... = \sum_y < M_y \rho M_y^\dagger \,.$$

To complete the proof, we show the following chain:

$$I(X;Y) \overset{(1)}{=} S(P';R') \overset{(2)}{\leq} S(P;Q) \overset{(3)}{=} S(\rho) - \sum_x p_x S(\rho_x)\,.$$

1. We elaborate on $\rho^{P'R'}$:

$$\begin{aligned}\rho^{P'R'} &= \sum_{x,y} \mathrm{tr}(M_y \rho_x M_y^\dagger) p_x |x\rangle\langle x| \otimes |y\rangle\langle y| \\ &= \sum_{x,y} p_{y|x} p_x |x\rangle\langle x| \otimes |y\rangle\langle y| \\ &= \sum_{x,y} p_{x,y} |x\rangle\langle x| \otimes |y\rangle\langle y|\,.\end{aligned}$$

   It follows that $P'R'$ is a classical system, and hence:

$$S(P';R') = I(X;Y)\,.$$

2. We have

$$S(P';R') \leq S(P';Q'R') \leq S(P;QR) = S(P;Q),$$

   where the first inequality follows from Property I.6, the second inequality follows from Property I.9, and the final equality follows from the fact that R is in a product state with PQ, see the home assignment.

3. We elaborate on ρ^{PQ}:

$$\rho^{PQ} = \mathrm{tr}_R(\rho^{PQR}) = \sum_x p_x |x\rangle\langle x| \otimes \rho_x\,.$$

   By Property 7,

$$S(PQ) = H(X) + \sum_x p_x S(\rho_x)$$

   Now,

$$S(P) = H(X),$$

   since

$$\rho^P = \sum_x p_x |x\rangle\langle x|,$$

   and

$$S(Q) = S(\rho),$$

   since

$$\rho^Q = \sum_x p_x \rho_x\,.$$

Hence

$$S(P;Q) = S(P) + S(Q) - S(PQ) = S(\rho) - \sum_x p_x S(\rho_x)\,. \tag{5.7}$$

□

Remark II.3.

1. *By Property I.3, χ is always non-negative.*

2. *As mentioned in the theorem, equality in the Holevo bound holds if and only if the states ρ_x commute, i.e., they have the same set of eigenvectors. In this case, the bound is attained using a Von-Neumann measurement in the joint basis.*

3. *An optimal measurement that attains the accessible information is generally not known. Nevertheless, it is known [6] that there exists an optimal measurement of the form*

$$E_y = \alpha_y |\psi_y\rangle\langle\psi_y|,$$

where $y = 0, \ldots, m-1$ and

$$dim\mathcal{H} \leq m \leq (dim\mathcal{H})^2\,.$$

4. *Sometimes the bound can be far from tight. For example, if the states $\{\rho_x\}$ are pure, then the Holevo quantity is $S(\rho)$, i.e., independent of the ensemble. Nevertheless, tighter (and more involved) upper and lower bounds exist.*

III The pretty good measurement (PGM)

For an ensemble $\{p_x, \rho_x\}_0^{n-1}$, the PGM is given by the POVM:

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

where $\rho \triangleq \sum_x p_x \rho_x$.

Let us now show that this is indeed a POVM. Denote by $\mathcal{H}$ the support of ρ. one easily verifies that $E_x \geq 0$ and

$$\sum_x E_x = \rho^{-1/2} \sum_x p_x \rho_x \rho^{-1/2} = \rho^{-1/2} \rho \rho^{-1/2} = I\,,$$

in $\mathcal{H}$. if ρ is singular, we shall introduce an additional opeartor E_n to the POVM: E_n is nullified on $\mathcal{H}$ and is equal to the identity operator outside it, i.e., E_n is the completion to I outside $\mathcal{H}$.

Property III.1.

- *If $\{\rho_x\}$ are pure, then the probability of success, for the optimal measurement possible, satisfies [5]:*

$$P_C^2(opt) \leq P_C(PGM),$$

  which is pretty good! $\ddot{\smile}$

- *If $\{\rho_x\}$ are pure and orthogonal, then the PGM is reduces to the optimal VN measurement.*

- *There are more scenarios in which the PGM is optimal.*

Chapter 6

Quantum Channels and Classical Capacity

Summary by Yuval Kochman and Anatoly Khina.

I Quantum Channels

Definition I.1 (Quantum Channel)**.** *A quantum channel is a mapping from* $\Delta\left(\mathcal{H}_{d_{in}}\right)$ *to* $\Delta\left(\mathcal{H}_{d_{out}}\right)$*:*

$$\mathcal{E}: \quad \Delta\left(\mathcal{H}_{d_{in}}\right) \longrightarrow \Delta\left(\mathcal{H}_{d_{out}}\right),$$

which satisfies the property:

$$\mathcal{E}\left(\rho^{AB}\right) = tr_{BD}\Big(U\big(\rho^{AB} \otimes |0_{CD}\rangle\langle 0_{CD}|\big)U^{\dagger}\Big), \tag{6.1}$$

where $\Delta\left(\mathcal{H}\right)$ *is defined as the set of all possible d.m. in* $\mathcal{H}$*.*

Remark I.1.

1. (6.1) *suggests taking the system (AB) with a certain environment (CD), performing some (unitary) interaction (U) between the two and "throwing" away part of the environment (D) along with part of the system (B).*

2. *The ancilla state of the environment (*$|0_{CD}\rangle\langle 0_{CD}|$*) is assumed to be a pure state, since otherwise one could convert it into such by applying purification to it, in the process of which - enlarging the dimensions of the environment.*

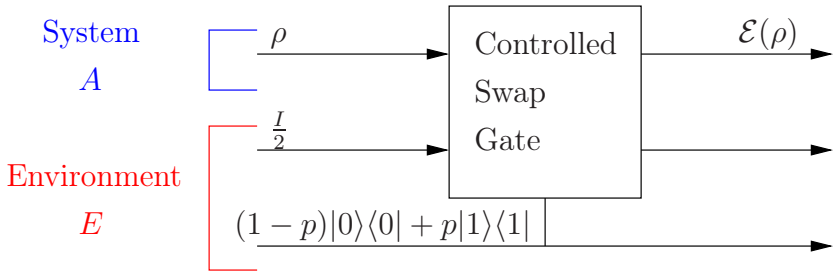


Figure 6.1: Depolarizing Channel.

3. *A common special case is the case of quantum operations of the form*

$$\mathcal{E}(\rho^A) = tr_E\Big(U\big(\rho^A \otimes |0_E\rangle\langle 0_E|\big)U^\dagger\Big).$$

Note that in this case the output and input spaces are identical.

Example I.1 (Depolarizing Channel)**.** *This channel is the quantum generalization of the classical BSC: its single qubit input is depolarized with probability p, i.e., replaced by the completely mixed state $I/2$ and with probability $(1-p)$ - left untouched. The depolarizing channel operation can be conveniently described as*

$$\mathcal{E}(\rho) = p\frac{I}{2} + (1-p)\rho. \tag{6.2}$$

Note that the environment of the channel is described as the tensor product of two mixed states: a controlling state $(1-p)|0\rangle\langle 0| + p|1\rangle\langle 1|$ which swaps the two inputs of the controlled gate with probability p and does no effect with probability $(1-p)$; the second environment mixed state is a completely mixed state $\frac{I}{2}$, which is swapped (or not) with the "system state" ρ. Hence this channel can be described in the following way:

$$\begin{aligned}
\mathcal{E}(\rho) &= tr_E\left(U\left[\rho \otimes \frac{I}{2} \otimes \Big((1-p)|0\rangle\langle 0| + p|1\rangle\langle 1|\Big)\right]U^\dagger\right) \\
&= tr_E\left((1-p)\rho \otimes \frac{I}{2} \otimes |0\rangle\langle 0| + p\frac{I}{2} \otimes \rho \otimes |1\rangle\langle 1|\right) \\
&= (1-p)\rho + p\frac{I}{2}\ , \qquad \checkmark
\end{aligned}$$

for $U = I$.

Remark I.2. *This is the quantum parallel of $BSC\big(\frac{p}{2}\big)$, since with probability p we pick at* random.

I.1 Classical Capacity of a Quantum Channel

The aim of the encoder is to transmit as much *classical* information as possible to the decoder over a quantum channel. Hence, this problem is close in mind to the classical (Shannon) channel coding problem, only now the channel is quantum.

The most general (block-code) coding scheme, with block length n and rate R has the following form:

Encoder: $E: \quad \left\{1,\ldots,2^{nR}\right\} \longrightarrow \Delta\left(\mathcal{H}_{d_{in}}^{\otimes n}\right)$.

Decoder: $D: \quad \{E_m\}_{m=1}^{2^{nR}}$.

This scheme takes the classical message, transforms it into a quantum channel state and measures it at the output. Our aim is, of course, to maximize the rate R of the scheme, such that the probability of error would go to zero with the block-length n:

$$P_e = \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} P\Big(D \circ \mathrm{E}\Big(E(m)\Big) \neq m \Big) \longrightarrow \infty \,.$$

Proposition I.1 (Classical-Classical Capacity)**.** *The classical capacity of a quantum channel $\mathcal{E}$ is*

$$C_{CC} = \max_{p_i,\rho_i} I_{acc}\Big(\Big\{ p_i, \mathcal{E}(\rho_i) \Big\} \Big) \tag{6.3}$$

when restricting the use to a product-state at the encoder:

$$E : \left\{1,\ldots,2^{nR}\right\} \longrightarrow \bigotimes_{i=1}^{n} \Delta\left(\mathcal{H}_{d_{in}}\right) \tag{6.4}$$

and a decoder of the form

$$D : \{E_m\}\,, \qquad E_m = \bigotimes_{i=1}^{n} E_{m,i}\,. \tag{6.5}$$

Remark I.3. *Constraints* (6.4),(6.5) *mean that no entanglement is exploited, not at the input of the channel, nor in its output. Making this approach to the problem very "classical".*

Proof Sketch. For every channel instance, the maximum that can be achieved is I_{acc} and there is no entanglement that can be exploited. Hence the effective channel is a classical channel whose input can be "shaped" by selecting different ensembles $\{p_i, \rho_i\}$. □

Remark I.4. *From The Holevo bound Theorem II.1, (Chapter 5, the following upper bound on C_{cc} holds*

$$\begin{aligned} C_{CC} &\leq \max_{\{p_i,\rho_i\}} \chi\Big(\mathcal{E}, \Big\{ p_i, \rho_i \Big\} \Big) \\ \chi\Big(\mathcal{E}, \Big\{ p_i, \rho_i \Big\} \Big) &\triangleq S\Big(\mathcal{E}\Big(\sum_i p_i \rho_i \Big) \Big) - \sum_i p_i S\Big(\mathcal{E}(\rho_i) \Big). \end{aligned} \tag{6.6}$$

Note that in general this bound is not tight Theorem II.1, Chapter 5. Nevertheless, by allowing a general measurement, we can achieve the Holevo upper bound, as suggested by the following theorem.

Theorem I.1 (HSW)**.** *The classical capacity of a quantum channel $\mathcal{E}$, under a product-state constraint* (6.4) *at the encoder, is:*

$$C_{CQ} = \max_{\{p_i,\rho_i\}} \chi\left(\mathcal{E}, \left\{p_i, \rho_i\right\}\right), \tag{6.7}$$

when χ is defined in (6.6)*.*

Remark I.5. *This capacity can be strictly greater then the capacity of Theorem I.1, since in general, there are additional degrees of freedom that can be exploited when using a general measurement and not measuring sequential d-dimensional states as in* (6.5)*. - Resembles the gain of soft decoding over hard decoding.*

Before turning to the formal proof, we shall start with the idea of the achievability proof.

Achievability Proof Idea. Generate 2^{nR} words according to $\{p_i, \rho_i\}$ in an i.i.d. manner. Hence the mth codeword has the form

$$\rho^{(m)} = \rho_{m_1} \otimes \rho_{m_2} \otimes \cdots \otimes \rho_{m_n}.$$

Since the the channel is "memoryless" the corresponding output is of the form

$$\sigma^{(m)} = \mathcal{E}^{\otimes n}\left(\rho^{(m)}\right) = \sigma_{m_1} \otimes \sigma_{m_2} \otimes \cdots \otimes \sigma_{m_n}.$$

The average channel entry and channel output of this ensemble are

$$\begin{aligned} \rho &= \sum_i p_i \rho_i \\ \sigma &= \sum_i p_i \mathcal{E}\left(\rho_i\right) = \sum_i p_i \sigma_i \ . \end{aligned}$$

Denote the average entropy conditioned on the message and the entropy of the average channel input state by

$$\begin{aligned} \bar{S} &\triangleq \sum_i p_i S(\sigma_i) \\ S &\triangleq S(\sigma). \end{aligned}$$

- The average output d.m. is $\sigma^{\otimes n}$ and it has approximately 2^{nS} orthonormal eigenvectors and eigenvalue 2^{-nS}.

- A random codeword induces an output state which has, with high probability, the form

$$\sigma^{(m)} = \bigotimes_i \sigma_i^{\otimes np_i}$$

(up to permutation)
and using the additivity of product states property (Property V.3), we have

$$S\left(\sigma^{(m)}\right) \approx \sum_i np_i S(\sigma_i) = n\bar{S} \, .$$

Hence $\sigma^{(m)}$ has a typical subspace of dimension $\sim 2^{n\bar{S}}$.

- The probability of an *intersection* between an output state induced by a random codeword and the output is, on average

$$\frac{2^{n\bar{S}}}{2^{nS}} = 2^{-n\chi} \ .$$

- Using a measurement which "separates" between typical subspaces, the error probability is

$$P_e \approx 2^{nR} \cdot 2^{-n\chi} \xrightarrow[n\to\infty]{R<\chi} 0 \, .$$

However, finding such a measurement is far from being trivial.

Proof. Achievability: we shall denote by P the projector onto the typical subspace of $\sigma^{\otimes n}$, and by P_m - the projector on the typical subspace of $\sigma^{(m)}$.

Using the quantum AEP (Theorem I.1), for any $\epsilon, \delta > 0$, taking a large enough n, we have

$$tr\left(\sigma^{\otimes n} P\right) > 1 - \epsilon \tag{6.8a}$$

$$\mathbb{E} tr\left(\sigma^{(m)} P_m\right) > 1 - \epsilon \tag{6.8b}$$

$$\mathbb{E} tr(P_m) \leq 2^{n(\bar{S}+\delta)} \ , \tag{6.8c}$$

where the expectations are over all codebooks. Using the POVM

$$E_m \triangleq \left(\sum_k P P_k P\right)^{-\frac{1}{2}} P P_m P \left(\sum_k P P_k P\right)^{-\frac{1}{2}}$$

one may show that the average error probability, averaged over all codewords, satisfies the following inequality [10]

$$\begin{aligned} P_e &= \frac{1}{2^{nR}} \sum_m \left(1 - tr\left(\sigma^{(m)} E_m\right)\right) \\ &\leq \frac{1}{2^{nR}} \sum_m \left[3tr\left(\sigma^{(m)}(I-P)\right) + \sum_{k\neq m} tr\left(P\sigma^{(m)} P P_k\right) + tr\left(\sigma^{(m)}(I-P_m)\right)\right] . \end{aligned}$$

Now averaging over all codebooks we obtain

$$\begin{aligned}\mathbb{E}P_e &\overset{(a)}{\leq} \overbrace{3tr\left(\sigma^{\otimes n}(I-P)\right)}^{\substack{(6.8a)\\ \Downarrow \\ \leq 3\epsilon}} + \left(2^{nR}-1\right) tr\left(P\sigma^{\otimes n}P\mathbb{E}(P_1)\right) + \overbrace{\mathbb{E}tr\left(\sigma^{(1)}(I-P_1)\right)}^{\substack{(6.8b)\\ \Downarrow \\ \leq \epsilon}} \\ &\overset{(b)}{=} \left(2^{nR}-1\right) tr\left(\sigma^{\otimes n}P\mathbb{E}(P_1)\right) + 4\epsilon\,, \end{aligned} \tag{6.9}$$

where in (a) is due to the linearity of the trace and the expectation functional, the equality $\mathbb{E}\sigma^{(m)} = \sigma^{\otimes n}$, the fact that P_k and $\sigma^{(m)}$ are independent and the symmetry between P_1 and P_k due to the expectation over all codebooks; (b) holds since P is a projector and hence $P\sigma^{\otimes n}P$ holds true.

To upper-bound the remaining element of the r.h.s. in (6.9) we shall use the following theorem.

Theorem I.2. *Let A, B be two positive Hermitian matrices. Then, the following inequality holds*

$$tr(AB) \leq \sum_i \lambda_i(A)\lambda_i(B), \tag{6.10}$$

where the eigenvalues $\{\lambda_i(A)\}$ and $\{\lambda_i(B)\}$ of A and B respect., are ordered in a non-increasing order.

According to the quantum AEP (Theorem I.1), the number of the non-zero eigenvalues of $\mathbb{E}P_1$ is no greater than $2^{n(\bar{S}+\delta)}$, whereas the eigenvalues of $\sigma^{\otimes n}P$ are not less than $2^{-n(S-\delta)}$. Combining these results with the result of Theorem 6.10 and applying them to (6.9) gives rise to

$$\mathbb{E}P_e \leq 2^{nR}\cdot 2^{n(\bar{S}+\delta)}\cdot 2^{-n(S-\delta)} \xrightarrow[n\to\infty]{R<\chi-2\delta} 4\epsilon\,.$$

Finally taking ϵ and δ to zero achieves the desired result.

Converse: denote by d the dimension of the input space. Then the rate cannot exceed $R \leq \log d$, since the dimension of the Hilbert space is the maximal number of distinguishable states, as indicated by Theorem III.2, Chapter 1. Moreover, WLOG, the optimal POVM will contain at most d^n elements [6]. Hence, using the *Fano* inequality we have

$$\begin{aligned} &h(P_e) + P_e \log\left(d^n - 1\right) \geq H(M|Y) \\ &nP_e \log d \geq H(M) - I(M;Y) - h(P_e)\,, \end{aligned} \tag{6.11}$$

where M is the information codeword and hence satisfies $H(M) = nR$. Note that for every message value $w \in \left\{1,\ldots,\lceil 2^{nR}\rceil\right\}$, the corresponding output state is σ_w (since the input state

to the channel is ρ_w in this case). Hence we have the following chain on inequalities:

$$I(M;Y) \overset{(a)}{\leq} S\left(\frac{1}{2^{nR}}\sum_m \sigma_{m_1}\otimes\sigma_{m_2}\otimes\cdots\otimes\sigma_{m_n}\right) - \frac{1}{2^{nR}}\sum_m S\Big(\sigma_{m_1}\otimes\sigma_{m_2}\otimes\cdots\otimes\sigma_{m_n}\Big) \quad (6.12)$$

$$\overset{(b)}{\leq} \sum_{j=1}^{n}\left[S\left(\frac{1}{2^{nR}}\sum_m \sigma_{m_j}\right) - \frac{1}{2^{nR}}\sum_m S\Big(\sigma_{m_j}\Big)\right] \overset{(c)}{=} \sum_{j=1}^{n}\left[\chi\left(\epsilon,\left\{\frac{1}{2^{nR}},\sigma_{m_j}\right\}\right)\right] \overset{(s)}{\leq} nC_{CQ}\,, \quad (6.13)$$

where in (a) we used the Holevo bound, which can be directly applied here since we can refer to $\left\{\frac{1}{2^{nR}}\sigma_i\right\}$ as the ensemble of Theorem II.1, Chapter 5; (b) is true due to Property V.5 and Property V.3 of Chapter 4; in (c) we use once again the bound of Theorem II.1, Chapter 5, only now we apply it to the ensemble $\left\{\frac{1}{2^{nR}},\sigma_{m_j}\right\}_m$ (for every $j\in\{1,\ldots,n\}$), and finally in (d) we use the definition of C_{CQ}. Combining the results of (6.11), (6.13) we obtain

$$\begin{aligned} nP_e\log d &\geq nR - nC_{CQ} - h(P_e) \\ P_e &\geq \frac{R-C_{CQ}}{\log d} + \mathcal{O}\left(\frac{1}{n}\right). \end{aligned} \quad (6.14)$$

Hence, when working at rates greater then $R > C_{CQ}$, the error probability is bounded from below. □

Remark I.6.

- *C_{CQ} can always be achieved by an ensemble of pure states, and no more than d_{in}^2 such states are needed.*
- *There are channels for which C_{CQ} is achieved* only *by non-orthogonal states ([8]).*

Example I.2 (Depolarizing Channel)**.** *Consider again the depolarizing channel introduced in Example I.1. According to Remark I.6 we can restrict our attention to ensembles of pure states, i.e., of the form $\{p_i,|\psi\rangle_i\}$ and C_{CQ} is given by*

$$C_{CQ} = \max_{\{p_i,|psi\rangle\}}\chi\left(\mathcal{E},\{p_i,|psi\rangle\}\right) = \max_{\{p_i,|\psi\rangle\}}\left[S\left(\mathcal{E}\left(p_i|\psi\rangle\langle\psi|\right)\right) - \sum_i p_i S\left(\mathcal{E}\left(|\psi\rangle\langle\psi|\right)\right)\right].$$

Note that for any pure state, the channel output is of the form

$$\mathcal{E}\left(|\psi\rangle\langle\psi|\right) = (1-p)|\psi_i\rangle\langle\psi_i| + p\frac{I}{2}.$$

This state has two eigenvectors: $|\psi_i\rangle$ and $|\psi_i^{\perp}\rangle$ with eigenvalues $\left(1-\frac{p}{2}\right)$ and $\frac{p}{2}$ respectively. Hence, the VN entropy of any *pure state is $S\Big(\mathcal{E}\big(|\psi_i\rangle\langle\psi_i|\big)\Big) = h\left(\frac{p}{2}\right)$. Hence our aim is to*

maximize the entropy of the output state. This is achieved by taking the ensemble $\{|0\rangle, |1\rangle\}$ *with equal probabilities. The corresponding entropy is*

$$\left(\mathcal{E}\left(|\psi_i\rangle\langle\psi_i|\right)\right) = h\left(\frac{p}{2}\right).$$

Hence,

$$C_{CQ} = 1 - h\left(\frac{p}{2}\right),$$

which fits with the observation of Remark I.2.

Proposition I.2 (Classical Capacity)**.** *The classical capacity of a quantum channel* $\mathcal{E}$ *is given by*

$$C = \lim_{n\to\infty} \frac{1}{n} C_{CQ}\left(\mathcal{E}^{\otimes n}\right) \geq C_{CQ}(\mathcal{E}).$$

Remark I.7. *Prop. I.2 is trivial. The non-trivial part is whether the limit is necessary or equivalently, whether the inequality holds with equality. This however is an open question. Another equivalent form of this question is whether the following inequality holds true:*

$$\chi_{\max}\left(\mathcal{E}_1 \otimes \mathcal{E}_2\right) \stackrel{?}{=} \chi_{\max}\left(\mathcal{E}_1\right) + \chi_{\max}\left(\mathcal{E}_2\right)$$

($\leq$ is obvious but $=$ is not).

Corollary 6.1. *For every quantum channel which is not constant (two input states exist for which the output states are different), the classical capacity is* strictly *positive.*

Proof. Since the channel is not constant thee exist two states the corresponding channel output of whom differ. Moreover one may prove that if these states are mixed, then there also exist two pure state which satisfy the property above. Denote these pure states by $|\psi\rangle\langle\psi|$ and $|\varphi\rangle\langle\varphi|$. Then by using the ensemble that is composed of these channels with equal probabilities, we can lower-bound the classical capacity as follow:

$$C \geq C_{CQ} \geq S\left(\frac{1}{2}\mathcal{E}\left(|\psi\rangle\langle\psi|\right) + \frac{1}{2}\mathcal{E}\left(|\varphi\rangle\langle\varphi|\right)\right) - \frac{1}{2}S\big(\mathcal{E}\left(|\psi\rangle\langle\psi|\right)\big) - \frac{1}{2}S\big(\mathcal{E}\left(|\varphi\rangle\langle\varphi|\right)\big) > 0,$$

where the last inequality is true due to concavity of VN entropy (Property I.3, Chapter 5). □

Example I.3 (Quantum Erasure Channel)**.** *This channel is described by*

$$\mathcal{E}(\rho) = (1-p)\rho \otimes |0\rangle\langle 0| + p\frac{I}{2} \otimes |1\rangle\langle 1|,$$

whether the second qubit indicates whether an erasure has occurred.

Using the ensemble $\{|0\rangle, |1\rangle\}$ with equal probabilities and measuring in the corresponding base ($\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$) the first qubit ($\rho$) gives rise to the classical BEC with erasure probability of p (where the second qubit is measured as well to identifies erasures). The capacity of this classical channel is $(1-p)$, hence $C \geq 1-p$. On the other hand, after n uses of the channel, the decoder has $\approx n(1-p)$ non-erased qubits. Since every qubit can yield one classical bit at most: $C \leq 1-p$, and hence

$$C = 1 - p.$$

Special Cases for which $\mathbf{C = C_{CQ}}$:

1. Entanglement Breaking Channels: channels consisting of a measurement followed by sending a *classical* message.

2. Unital Qubit Channels: channels satisfying $\mathcal{E}\left(\frac{I}{2}\right) = \frac{I}{2}$, such as the depolarizing channel [King 2002].

Chapter 7

Entanglement-Assisted Capacity and Entanglement Quantification

Summary by Judy Kupferman.

I Entanglement-Assisted Capacity

Example I.1 (Noiseless Channel)**.** *In a noiseless channel, using superdense coding, we managed to transmit* 2 *bits using a single qubit, i.e.,*

$$\underbrace{C_E}_{\substack{\text{Capacity assuming}\\ \text{entanglement}\\ \text{avaliable is}}} = \underbrace{2C}_{\substack{\text{Classical capacity}\\ \text{without}\\ \text{entanglement}}} .$$

We would like to derive similar results for other (noisy) channels. The operation of the channel $\mathcal{E}$ on a quantum system Q can be thought of as a channel of the form $\mathcal{E} \otimes I_R$, applied to a composite system QR in a pure state (where R is a purification system). W e will later interpret the output or the "purification part" as a noiseless channel used to create the entanglement prior to trnasmission. Thus, the input state $|\varphi_\rho\rangle$ is a pure state (see Fig. 7.1):

$$\begin{aligned} &\phi_\rho = |\varphi_\rho\rangle \langle\varphi_\rho| \,, \\ &tr_R(\phi_\rho) = \rho \,. \end{aligned}$$

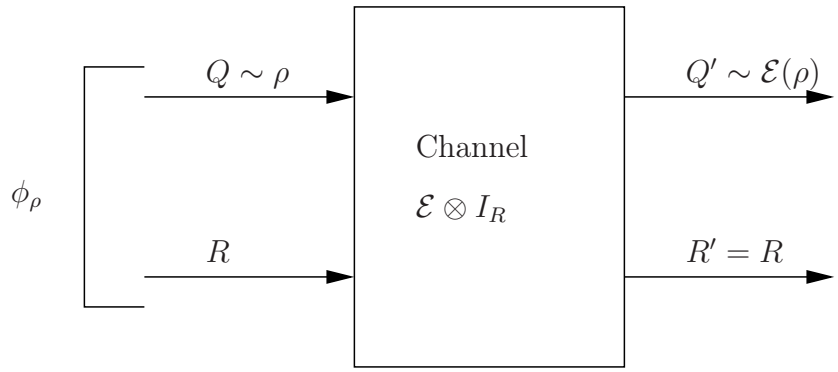


Figure 7.1: Equivalent quantum channel with entanglement

Theorem I.1 (Bennett et al. 2002, [3])**.** *The classical capacity of a quantum channel $\mathcal{E}$, in the presence of unlimited entanglement, is given by*

$$C_E = \max_{\rho \in Q} S\left(Q'; R'\right) . \tag{7.1}$$

Proof. First we define terms (see Fig. 7.1):

$$\begin{aligned} S\left(Q'\right) &= S\left(\mathcal{E}\left(\rho\right)\right) && (7.2)\\ S\left(R'\right) &= S\left(R\right) = S\left(\rho\right) && (7.3)\\ S\left(Q'R'\right) &= S\left(\left(\mathcal{E} \times I\right)\phi_\rho\right) \triangleq S\left(\rho, \mathcal{E}\right) , && (7.4) \end{aligned}$$

where (7.2) is the entropy of the output state; (7.3) holds since ϕ_ρ is a pure state, Property V.2, and R goes through a noiseless channel (see also Remark II.3, Chapter 4, whereas (7.4) is the "entropy-exchange" to be defined next.

Definition I.1 (Entropy-exchange)**.** *The entropy-exchange of quantum state ρ and channel $\mathcal{E}$ is defined as*

$$S\left(\rho, \mathcal{E}\right) \triangleq S\left(\left(\mathcal{E} \times I\right)\phi_\rho\right) .$$

Note that if we denote the environment, corresponding to the channel $\mathcal{E}$, by E, then, by the Schmidt decomposition, the entropy that "passes" to the environment is exactly the entropy-exchange:

$$S(E') = S(Q', R') = S(\rho, \mathcal{E}) .$$

Hence, the capacity formula (7.1) can be written as:

$$C_E = \max_{\rho \in Q} \Big[S\left(\rho\right) + S\left(\mathcal{E}\left(\rho\right)\right) - S\left(\rho, \mathcal{E}\right) \Big] .$$

We shall break the proof into 3 stages:

1. Acheiving the rate $S(\rho) + S(\mathcal{E}(\rho)) - S(\rho, \mathcal{E})$ for $\rho = \frac{I}{d}$, where $d = \dim Q$.

2. Generalization for $\rho = \frac{P}{tr(P)}$, for some projection matrix P.

3. Generalization for all ρ.

Stage 1: $\rho = \frac{I}{d}$.

Generalization of superdense coding to general dimension:

Transmitter and receiver share state $\left|\psi^{AB}\right\rangle$ such that $d = \dim A = \dim B$ and $tr_A\left\{\left|\psi^{AB}\right\rangle\left\langle\psi^{AB}\right|\right\} = I/d$, where A denotes the transmitter system and B - the system of the receiver.

Remark I.1. *The state $|\psi^{AB}\rangle\langle\psi^{AB}|$ is maximally entangled, as $\rho \equiv \rho^A = \frac{I}{d}$, similarly to the Bell state, see Section VII, Chapter 2.*

By Theorem II.1, Chapter 4 (Schmidt decomposition) we can decompose $|\psi^{AB}\rangle$ as

$$\left|\psi^{AB}\right\rangle = \frac{1}{\sqrt{d}}\sum_{j=1}^{d}|j_A\rangle\,|j_B\rangle$$

with $\{|j_A\rangle\}$, $\{|j_B\rangle\}$ being some orthonormal bases of A, B respectively. Applying a unitary operator U to A gives rise to the state

$$(U \otimes I)\left|\psi^{AB}\right\rangle = \frac{1}{\sqrt{d}}\sum_{j=1}^{d}U|j_A\rangle\,|j_B\rangle\ ,$$

which is also maximally entangled, as local unitary operations cannot change the r.d.m..

In superdense coding for $d = 2$ we were able to produce 4 orthogonal states, i.e., an orthonormal basis of AB, using local unitary operators (Pauli matrices), applied to system A only. This can be generalized to any (natural) d: it is possible show there exists a set of d^2 local unitary operators $\{U_i\}_{i=1}^{d^2}$ of A, such that $\left\{(U_i \otimes I)\left|\psi^{AB}\right\rangle\right\}_{i=1}^{d^2}$ constitutes an orthonormal basis of AB.

Corollary 7.1. *If a clean ("noiseless") channel exists between transmitter and receiver, then using SDC of dimension d, i.e., to transmit $2\log d$ classical bits (cbits) using a system of dimension d.*

For the noisy channel, we shall use $\left|\psi^{AB}\right\rangle$ and take $d = \dim Q$.

We may think of $\left|\psi^{AB}\right\rangle$ as a state that was created at the transmitter, where B passes through a clean channel. Hence

$$C_E(\mathcal{E}) \geq C'(\mathcal{E} \otimes I) \geq C'_{CQ}(\mathcal{E} \otimes I)\ ,$$

where C' and C'_{CQ} stand for the "regular" capacities, i.e., *without* entanglement, subject to a constraint of an input ensemble of the form $\left\{|\psi^i\rangle = (U_i \otimes I)\left|\psi^{AB}\right\rangle\right\}$, where the subsystem A is transmitted through the channel $\mathcal{E}$, and the subsystem B is a fixed state at the transmitter which passes through a clean channel I. Hence, we choose an input ensemble of the form: $\left\{\psi_i^{AB}\right\} \triangleq \left\{(U_i \otimes I)\left|\psi^{AB}\right\rangle\right\}_{i=1}^{d^2}$, with equal probabilities $p_i = \frac{1}{d^2}$, where $|\psi^{AB}\rangle$ is some fixed maximally entangled state.

The mixes input state is, therefore,

$$\sum_{i=1}^{d^2} \frac{1}{d^2}\Big(U_i \otimes I\Big)\left|\psi^{AB}\right\rangle\left\langle\psi^{AB}\right|\Big(U_i^\dagger \otimes I\Big) = \frac{I_{d^2}}{d^2} = \frac{I_d}{d} \otimes \frac{I_d}{d} = \rho \otimes \rho$$

, and the output state is

$$(\mathcal{E} \otimes I)\left(\frac{I}{d} \otimes \frac{I}{d}\right) = \mathcal{E}(\rho) \otimes \rho\,.$$

To find the output state given an input state (or equivalently i), note that each input state $\psi_i = (U_i \otimes I)\,\psi\psi^{AB}$ is a purification of I/d, and hence, according to Property V.2, Chapter 4:

$$\rho = tr_B\left((U_i \otimes I)\left|\psi^{AB}\right\rangle\right) = \frac{I}{d} = \rho\,.$$

Hence the output state given i may be written as $(\mathcal{E} \otimes I)\,\phi_\rho^i$, with $\phi_\rho^i (\equiv \psi_i)$, being a purification of $\rho = \frac{I}{d}$.

Remark I.2. *$S\left((\mathcal{E} \otimes I)\,\phi_\rho^i\right)$ does not depend on the exact purification, and hence is constant for all i.*

Combining all the results above, we arrive at

$$\begin{aligned} C_E(\mathcal{E}) \geq C'_{CQ}(\mathcal{E} \otimes I) &\geq S(\mathcal{E}(\rho) \otimes \rho) - \frac{1}{d^2}\sum_i S\left((\mathcal{E} \otimes I)\,\phi_\rho^i\right) \\ &= S(\mathcal{E}(\rho)) + S(\rho) - \underbrace{S((\mathcal{E} \otimes I)\,\phi_\rho)}_{S(\rho,\mathcal{E})}, \end{aligned}$$

where the second inequality is true for our choice of the ensemble and the equality is due to Remark I.2.

Stage 2: $\rho = \frac{P}{tr(P)}$, for some projection matrix P.

If we limit the input to the subspace spanned by P, of dimension $d' = tr(P)$, we get an effective channel $\mathcal{E}'$. The proof of first step then applies, by taking the input to be $\rho' = \frac{I}{d'}$.

Stage 3: generalization to any ρ.

Denote by P_n the projection operator on the typical subspace of $\rho^{\otimes n}$, and define the state $\Pi_n = \frac{P_n}{tr(P_n)}$. We shall use Π_n, as it is asymptotically equal to $\rho^{\otimes n}$, and therefore $dim\Pi_n \approx 2^{nS(\rho)}$.

Note that all the eigenvalues of $\Pi_n$are almost *identical* (uniform distribution over the typical subspace). Hence, according to Theorem I.1, Chapter 6, we have

$$\lim_{n\to\infty}\frac{1}{n}S\left(\Pi_n\right)=\lim_{n\to\infty}\frac{1}{n}\log\dim\Pi_n=S\left(\rho\right)\ .$$

Define by E the environment and by $\mathcal{N}$ the effective channel which the environment is passing through. W.l.o.g, we assume the state of the composite system QRE to be pure, where the state of the sub-system E is known, and can be pre-defined to be $|0_E\rangle$ (otherwise one may perform purification). See Fig. 7.2.

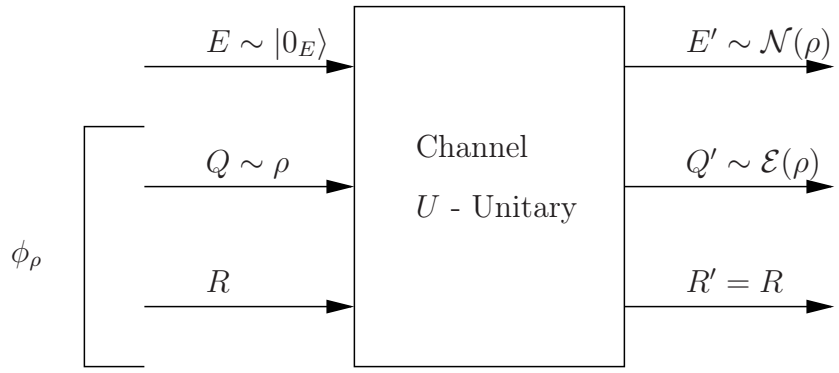


Figure 7.2: General quantum channel with "purified" input state and environment

Hence, since for a composite (bi-partite) system in a pure state both parts have the same entropy (see remark Remark II.3, Chapter 4, the entropy-exchange of the channel is equal to the entropy of the environment output state, as discussed earlier:

$$S\left(N(\rho)\right)=S\left(\left(\mathcal{E}\otimes I\right)\phi_\rho\right)=S\left(\rho,\mathcal{E}\right)$$

Consider the channel $\mathcal{E}^{\otimes n}$. Using the result of Stage 2, for the input state Π_n, we achieve the rate:

$$R=\frac{1}{n}\left[S\left(\Pi_n\right)+S\left(\mathcal{E}^{\otimes n}\left(\Pi_n\right)\right)-S\left(N^{\otimes n}\left(\Pi_n\right)\right)\right]\ . \tag{7.5}$$

Looking at the addend in (7.5), in the limit of $n\to\infty$:

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}S\left(\Pi_n\right)&=S\left(\rho\right)\\ \lim_{n\to\infty}\frac{1}{n}S\left(\mathcal{E}^{\otimes n}\left(\Pi_n\right)\right)&=\lim_{n\to\infty}\frac{1}{n}S\left(\mathcal{E}^{\otimes n}\left(\rho^{\otimes n}\right)\right)=S\left(\mathcal{E}(\rho)\right)\\ \lim_{n\to\infty}\frac{1}{n}S\left(N^{\otimes n}\left(\Pi_n\right)\right)&=S(N(\rho))=S\left(\rho,\mathcal{E}\right)\ .\end{aligned}$$

Thus,

$$C_E\geq S\left(\mathcal{E}\left(\rho\right)\right)+S\left(\rho\right)-S\left(\rho,\mathcal{E}\right)\ ,$$

for any admissible ρ, or equivalently

$$C_E \geq \max_{\rho \in Q} S\left(\mathcal{E}\left(\rho\right)\right) + S\left(\rho\right) - S\left(\rho, \mathcal{E}\right) .$$

The converse stems from a similar construction using the connections between C_E and C_{CQ}, given in Section I.1, and from the additivity of C_E mentioned below. □

Remark I.3.

- *For classical channels, entanglement obviously cannot increase capacity, i.e.,* $C_E = C$.
- $C_E\left(\mathcal{E}_1 \otimes \mathcal{E}_2\right) = C_E\left(\mathcal{E}_1\right) + C_E\left(\mathcal{E}_2\right)$.
- *The entanglement-assisted capacity of the depolarizing channel, defined in Example I.1, Chapter 6, is* $C_E = 2 - H\left(1 - \frac{3}{4}p, \frac{1}{4}p, \frac{1}{4}p, \frac{1}{4}p\right)$, *achieved by* $\rho = \frac{I}{2}$, *which is strictly greater than* $C = 1 - h_b\left(\frac{p}{2}\right)$ *for* $p > 0$. *See also Fig. I.3.*

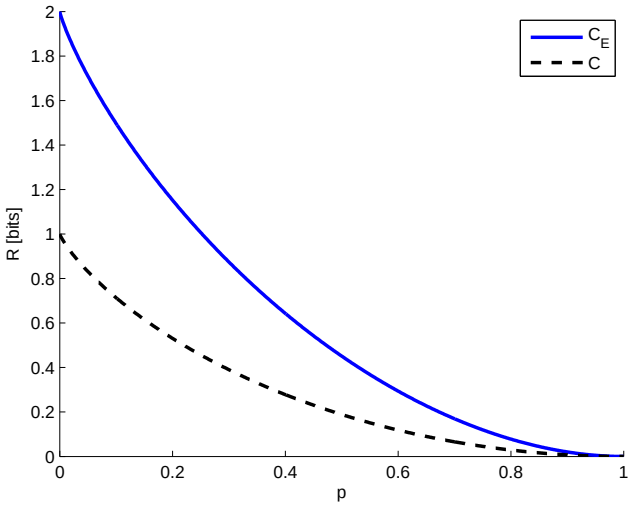


Figure 7.3: Capacity and entanglement-assisted capacity of the depolarizing channel

- *The entanglement-assisted capacity of the erasure channel, defined in Example I.3, Chapter 6, is* $C_E = 2(1-p) = 2C$. *Superdense coding turns the channel into two parallel quantum erasure channels (with no entangelement).*
- *Cost of protocol [13]:* $S\left(\rho\right)$ *ebits (EPR pairs) per channel use suffice to achieve* C_E.

I.1 Relation between C_E and C_{CQ}

Consider the system QRE of the proof of Theorem I.1. Using the Schmidt decomposition (Theorem II.1, Chapter 4) of ϕ_ρ we can write the input state of the system as:

$$|\psi\rangle = \sum_x \sqrt{p_x}|x^R\rangle|x^Q\rangle|0^E\rangle\,.$$

Note that $S(\rho) = H(\{p_x\})$. The output state of the system is therefore

$$\sum_x \sqrt{p_x}|x^R\rangle|\psi_x^{Q'E'}\rangle\,.$$

Having in mine the result of Theorem I.1, we evaluate $S(Q';R')$.

Measuring R (or R') according to the basis $\left\{\left|x^R\right\rangle.\right\}$ we get the mixed output state

$$\sum_x p_x|x\rangle\langle x|^R \otimes |\psi_x\rangle\langle\psi_x|^{Q'E'}\,.$$

Thus, the mutual information of interest is

$$\begin{aligned} S(Q';R') &= \underbrace{H(\{p_x\})}_{S(R')\text{ after measure}} + \underbrace{S(\mathcal{E}(\rho))}_{\substack{\text{did not change,}\\ \text{since measurement is}\\ \text{unitary and local}}} - \underbrace{\left[H(p_x) + \sum p_x S(\mathcal{E}(\rho_x))\right]}_{\text{entropy-exchange}} \\ &= S(\mathcal{E}(\rho)) - \sum p_x S(\mathcal{E}(\rho_x)) = \chi(\{p_x,\rho_x\})\,, \end{aligned}$$

which is exactly C_{CQ}.

Corollary 7.2. *C_{CQ} is achieved by measuring the system R (alone) $\Leftrightarrow$ destruction of entanglement.*

II Quantifying Entanglement

We have seen that ebits (EPR pairs) are very useful. Hence, we would like to be able to quantify the amount of ebits available.

Assertion II.1 (Entanglement Distillation of Pure States)**.** *Assume Alice and Bob share n copies of $|\psi\rangle$ and n is "large". Define $\rho = tr_A(|\psi\rangle\langle\psi|)$. Then Alice and Bob can* convert *these copies into $\approx nS(\rho)$ ebits using only local operations.*

Proof. Using the Schmidt decomposition (Theorem II.1, Chapter 4) for each of the shared states, we can rewrite the shared block as

$$(|\psi\rangle)^{\otimes n} = \left(\sum_{i=1}^{d} \sqrt{\lambda_i} \left|i^A\right\rangle \left|i^B\right\rangle\right)^{\otimes n} = \sum_{i_1,i_2,\ldots,i_n} \underbrace{(\lambda_{i_1}\ldots\lambda_{i_n})^{\frac{1}{2}}}_{\substack{\text{The same coeff.}\\ \text{for all summands}\\ \text{of the same type}}} \left|i_1^A\ldots i_n^A\right\rangle \left|i_1^B\ldots i_n^B\right\rangle ,$$

where $\{\lambda_i\}$ are the eigenvalues of ρ.

Denote by T_k the k^{th} type for some fixed enumeration. Then,

$$|\psi\rangle^{\otimes n} = \sum_k \alpha_k \sum_{(i_1,\ldots,i_n)\in T_k} \left|i_1^A\ldots i_n^A\right\rangle \left|i_1^B\ldots i_n^B\right\rangle ,$$

where $\alpha_k \triangleq (\lambda_{i_1}\ldots\lambda_{i_n})^{\frac{1}{2}} |T_k|$ and $|T_k|$ is the cardinality of T_k.

Remark II.1. *The analysis here is similar to the one made for quantum compression (Theorem I.1, Chapter 3): in the quantum compression theorem proof, we were only interested in ("weak") typicality. Here we are interested in the exact type ("strong typicality").*

Define the following projectors:
P_k^A - the projectors on the subspace of $T_k$in system A;
P_k^B - the projectors on the subspace of $T_k$in system B.

Alice measures her state halves using the POVM $\left\{P_k^A\right\}$, whereas Bob - using $\left\{P_k^B\right\}$. Alice and Bob will measure *the same type*![1]

The shared state, if result k (type T_k) was measured, is:

$$|\varphi_k\rangle = \frac{1}{\sqrt{|T_k|}} \sum_{(i_1,\ldots i_n)\in T_k} \left|i_1^A\ldots i_n^A\right\rangle \left|i_1^B\ldots i_n^B\right\rangle ,$$

and its probability is $p_k = |\alpha_k|^2 = \lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_n} |T_k|^2$. In fact, we got this way a uniform distribution over a subspace of dimension T_k. In case $|T_k| = 2^l$ ($l \in \mathbb{N}$), there are 2^l terms in the Schmidt decomposition of $|\varphi_k\rangle$, which are in fact ebits.

If $|T_k|$ is not a power of 2, then we can repeat the procedure for m blocks and suffer a *loss* which vanishes for $m \to \infty$.

Corollary 7.3. *Using local unitary transformation, it is possible to establish any orthogonal basis, in particular a basis that is a product of l ebits.*

[1]This is in fact a generalization of the EPR paradox (Section VII, Chapter 2).

AEP suggests,

$$\mathbb{E}_k \log |T_k| = nS(\rho)$$
$$n(S(\rho) - \delta) \leq \log T_k \leq n(S(\rho) + \delta)$$

w.p. 1.

Hence $nS(\rho)$ ebits may be generated from n copies of $|\psi\rangle$, when n goes to infinity, using local operations only. □

Remark II.2.

- *The operation, described in Assertion II.1, is called "entanglement distillation" and denoted by E_D. Assertion II.1 suggests that*

$$E_d(|\psi\rangle) \geq S(\rho) .$$

- *The reverse operation, which transforms ebits into a block of a tensor product of the same ((less entangled)) shared state is termed "entanglement dilution". The quantity of how many ebits are needed, per copy of a quantum state, is called "entanglement of formation" (denoted by E_F). It is possible to prove that, for a pure state $|\psi\rangle$:*

$$E_F(|\psi\rangle) \geq S(\rho) ,$$

  with high fidelity (which approaches 1 for $n \to \infty$).

  This is achieved by Alice preparing the states $(|\psi\rangle)^{\otimes n}$, keeping part, compressing the other part using $\approx nS(\rho)$ qubits with high fidelity, and teleporting these qubits to Bob. Of course, classical communication is required for this protocol.

- LOCC model *(local operation classical communication) Working in the LOCC framework, and assuming a pure state $|\psi\rangle$, we have*

$$E_d(|\psi\rangle) = E_F(|\psi\rangle) = S(\rho) ,$$

  where $\rho \triangleq tr_A(|\psi\rangle\langle\psi|)$.

- *If the shared state is mixed, only bounds on $E_d\left(\rho^{AB}\right)$ and $E_F\left(\rho^{AB}\right)$ are known.*

  - *Entanglement states exist from which it is impossible to create ("distill") ebits, i.e., $E_d\left(\rho^{AB}\right) = 0$ but $E_F\left(\rho^{AB}\right) > 0$. This means that one can take ebits, form mixed states out of them, and not to be able to "distill" any ebits back ("irreversible").*

– *The "One-shot entanglement of formation" is defined as*

$$E_F^{(1)}(\rho) = \min\left\{\sum_i p_i E_F(|\psi_i\rangle) : \rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|\right\} .$$

Clearly $E_F(\rho) \leq E_F^{(1)}(\rho)$*. It is not known whether*

$$E_F^{(1)}(\rho_1 \otimes \rho_2) = E_F^{(1)}(\rho_1) + E_F^{(1)}(\rho_2) ,$$

that is, whether it is additive or just $\leq$ *. Nevertheless, Paz, Silva and Rains claim to have proved this. If this indeed to be true, many other relations could be derived.*

Chapter 8

Further Notions of Capacity

Summary by Ohad Barak.

I Operator-Sum Representation

We present an equivalent characterization of quantum operations to that of Definition I.1, Chapter 6.

Theorem I.1 (Kraus)**.** *$\mathcal{E}(\rho)$ describes a quantum channel if and only if it can be written as*

$$\mathcal{E}(\rho) = \sum_{k=1}^{K} E_k \rho E_k^\dagger \,,$$

where the channel operators $\{E_k\}$ are mappings from the input Hilbert space to the output Hilbert space,

$$E_k : H_{d_{in}} \to H_{d_{out}}$$

and satisfy the completeness equation:

$$\sum_{k=1}^{K} E_k^\dagger E_k = I \tag{8.1}$$

and K, the "Completeness Rank", satisfies

$$1 \leq K \leq d_{in} d_{out} \,.$$

We shall not present the proof, as it is merely technical.

Example I.1 (Closed system)**.** *According to Postulate III, Chapter 1, the evolution of a closed quantum system is described by some unitary operator U, i.e.,*

$$\mathcal{E}(\rho) = U\rho U^\dagger .$$

In this case there is a single channel operator $E_1 = U$.

Example I.2 (Measurement)**.** *A ("generalized") measurement on a system is described by a collection of "measurement operators" $\{M_m\}$, which satisfy the completeness equation, as suggested by Postulate IV, Chapter 1 and Section IV, Chapter 2:*

$$\mathcal{E}(\rho) = \sum_m M_m \rho M_m^\dagger .$$

We see that the channel operators $\{E_m\}$ are exactly the measurement operators $\{M_m\}$!

As indicated by Theorem III.1, Chapter 1, a measurement can be described by adding an ancilla system, carrying a unitary operator and discarding. Theorem I.1 suggests the opposite direction as well: a unitary operation on a system and its environment, followed by the discarding of the environment is equivalent to a measurement.

Remark I.1.

1. *The completeness requirement (8.1) is equivalent to the requirement that $tr(\mathcal{E}(\rho)) = 1$.*
2. *The matrix representation of the operators E_k is not necessarily square (when $d_{in} \neq d_{out}$).*
3. *Every channel can be represented by many different sets of channel operators. These sets correspond to different choices of the basis of the environment.*
4. *The set of channel operators is also called a* superoperator*. The operators themselves are called* Kraus operators*.*

Theorem I.2. *$\mathcal{E}(\rho)$ is a quantum channel if and only if all the following conditions hold:*

1. *$tr(\mathcal{E}(\rho)) = 1 \quad \forall \rho$.*
2. *For every input ensemble $\{p_i, \rho_i\}$: $\mathcal{E}\left(\sum_i p_i \rho_i\right) = \sum_i p_i \mathcal{E}(\rho_i)$.*
3. *$\mathcal{E}$ is completely positive: if $\mathcal{E}$ operates on a system A, then for every system R we have $(\mathcal{E} \otimes I_R)(\rho^{AR}) \geq 0$.*

Remark I.2. *Complete positiveness is a stronger requirement than definite positiveness.*

Due to items 1 and 3 a quantum channel is referred to as *Completely Positive Trace Preserving* mapping, commonly abbreviated *CPTP*.

Example I.3 (Depolarizing channel)**.** *Recall the depolarizing channel Example I.1, Chapter 6:*

$$\mathcal{E}(\rho) = (1-\rho)\rho + \rho\frac{I}{2}\,.$$

This is not an operator-sum representation! To find the opeartor-sum representation, we shall use the following identity, that holds for every ρ:

$$I = \frac{1}{2}[\rho + X\rho X + Y\rho Y + Z\rho Z]\,,$$

where X, Y, Z are the Pauli matrices, defined in Example III.8, Chapter 1. This identity is proved in the home assignment. Thus,

$$\mathcal{E}(\rho) = \left(1 - \frac{3}{4}p\right)\rho + \frac{\rho}{4}[X\rho X + Y\rho Y + Z\rho Z]\,.$$

and the channel operators are, therefore

$$\left\{\sqrt{1-\frac{3p}{4}}I, \frac{\sqrt{p}}{2}X, \frac{\sqrt{p}}{2}Y, \frac{\sqrt{p}}{2}Z\right\}\,.$$

Hence, an alternative of looking on the depolarizing channel is that the channel has no effect on the input state w.p. $(1-\frac{3p}{4})$ and activates one of the operators X, Y or Z with equal probabilities otherwise.

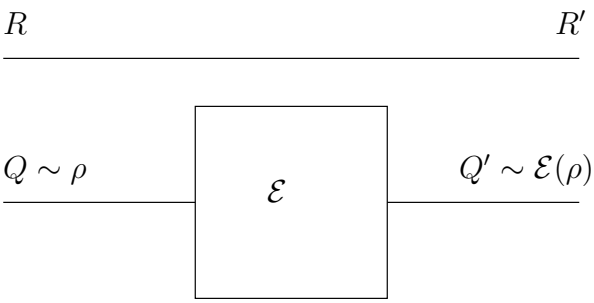
Remark I.3. *An equivalent representation would be*

$$\left\{\sqrt{1-p}I, \sqrt{\frac{p}{3}}X, \sqrt{\frac{p}{3}}Y, \sqrt{\frac{p}{3}}Z\right\}\,,$$

which is obtained by replacing $\frac{3}{4}P$ with P.

II Entanglement Fidelity

So far we discussed the fidelity between the input and output states of a quantum operation. However, this does not tell us how well is the entanglement between the input state and its environment is preserved by the quantum operation. Hence, $F(\rho, \mathcal{E}(\rho))$ is not a suitable fidelity measure for this case. Instead we examine the fidelity between the purification of the input state and its corresponding output state, where the added system represents the entanglement.



Definition II.1 (Entanglement fidelity)**.** *The entanglement fidelity of a channel $\mathcal{E}$ with input state $\rho \in Q$ is defined by:*

$$F_e(\rho, \mathcal{E}) \triangleq F(RQ, R'Q') = \langle RQ| \left[(I_R \otimes \mathcal{E})(|RQ\rangle\langle RQ|)\right] |RQ\rangle ,$$

where $|RQ\rangle$ is some purification of the state ρ of Q.

Property II.1.

1. $0 \leq F_e(\rho, \mathcal{E}) \leq 1$. *Furthermore, $F_e(\rho, \mathcal{E}) = 1$ if and only if*

$$\mathcal{E}(|\psi\rangle\langle\psi|) = |\psi\rangle\langle\psi| \quad \forall \quad |\psi\rangle \in supp(\rho) .$$

2. *F_e does not depend on the purification preformed.*

3. *If $\mathcal{E}$ is described by a set of Kraus operators $\{E_k\}$, then*

$$F_e(\rho, \mathcal{E}) = \sum_k |tr(\rho E_k)|^2 .$$

4. $$F_e(\rho, \mathcal{E}) \leq F(\rho, \mathcal{E}(\rho)).$$
   This means that entanglement fidelity is a more stringent measure than "regular" fidelity. Thus, "good" entanglement fidelity guarantees "good" "regular" fidelity as well.

5. *Concavity:* $$F_e\left(\sum_i p_i\rho_i, \mathcal{E}\right) \leq \sum_i p_i F_e(\rho_i, \mathcal{E}).$$

6. *By combining 4,4 we have that for any decomposition $\rho = \sum_i p_i\rho_i$ the following inequality holds:*

$$F_e(\rho, \mathcal{E}) \leq \sum_i p_i F(\rho_i, \mathcal{E}(\rho_i)) = \bar{F}.$$

7. *$S(\rho)$ is still an achievable rate under entanglement fidelity criterion.*

III Coherent Communication

Define the following systems:
A - Alice's system (transmitter).
B - Bob's system (receiver).
E - Eve's system (environment).

Definition III.1 (qubit channel (noisless))**.** *A (noiseless) qubit channel is the ability to perform the mapping*

$$|x^A\rangle \longrightarrow |x^B\rangle|\theta^E\rangle$$

for $|x\rangle \in \{|0\rangle, |1\rangle\}$ *and* $|\theta^E\rangle$ *some fixed state of the environment.*

A qubit (**qu**antum **bit**) channel allows to convey *any* qubit:

$$\alpha|0^A\rangle + \beta|1^A\rangle \longrightarrow (\alpha|0^A\rangle + \beta|1^A\rangle) \otimes |\theta^E\rangle \xrightarrow{\mathrm{tr}_E} \alpha|0^B\rangle + \beta|1^B\rangle\,.$$

Definition III.2 (cbit channel (noiseless))**.** *a cbit (**c**lassical **bit**) channel is the ability to perform the mapping*

$$|x^A\rangle \longrightarrow |x^B\rangle|x^E\rangle\,.$$

A cbit channel allows to convey the states $\{|0\rangle, |1\rangle\}$ perfectly, but not other state (superpositions of these two states):

$$\alpha|0^A\rangle + \beta|1^A\rangle \longrightarrow \alpha|00^{BE}\rangle + \beta|11^{BE}\rangle \xrightarrow{\mathrm{tr}_E} |\alpha|^2|0\rangle\langle 0|^B + |\beta|^2|1\rangle\langle 1|^B\,.$$

That is to say that if we try to pass a qubit through this "classical" channel, we get on its output a statistical mixture of the states $|0\rangle, |1\rangle$. This is equivalent to measuring the input with the basis $\{|0\rangle, |1\rangle\}$ and sending the outcome over a *classical* channel. This phenomenon, in the process of which quantum superposition turns into statistical mixture, is called *"decoherence"*.

Definition III.3 (cobit channel (noiseless))**.** *A cobit (**co**herent **bit**) channel is the ability to perform the mapping*

$$|x^A\rangle \longrightarrow |x^A\rangle|x^B\rangle|\theta^E\rangle\,.$$

With a cobit channel at hand we can implement a cbit channel, but not a qubit channel ("no cloning"). Nevertheless, a cobit channel, unlike a cbit one, allows to attain entanglement:

$$\alpha|0^A\rangle + \beta|1^A\rangle \longrightarrow \left(\alpha|00\rangle^{AB} + \beta|11\rangle^{AB}\right) \otimes |\theta^E\rangle \xrightarrow{\mathrm{tr}_E} \alpha|00\rangle^{AB} + \beta|11\rangle^{AB}\,.$$

Note that we can imitate a cobit channel with a qubit channel, by using a unitary operator that satisfies:

$$U|0^A\rangle|0^B\rangle = |0^A\rangle|0^B\rangle$$
$$U|1^A\rangle|0^B\rangle = |1^A\rangle|1^B\rangle\,,$$

or equivalently

$$U|x^A\rangle|0^B\rangle = |x^A\rangle|x^B\rangle\,.$$

Such an operator can be easily constructed, for instance:

$$U = |00\rangle\langle 00| + |11\rangle\langle 10| + |01\rangle\langle 01| + |10\rangle\langle 11|\,.$$

The aforementioned relations are summarized in the form of *resource inequalities*, as follows:

$$\text{qubit} \geq \text{cobit} \geq \text{cbit} \tag{8.2}$$
$$\text{cobit} \geq \text{ebit} \tag{8.3}$$

where '$\geq$' means that the right-hand side resource can be implemented by the resource on the other flank, but not vice versa.

III.1 Replacing Classical Operations by Coherent Opeartions

It is sometimes possible to turn classical communication into coherent communication, with or without loss in rate. This is done by:

- Turning destructive operations like measurements into unitary operations + environment and keeping the environment.
- Turning classically-dependent operations into quantum-dependent ones.

  Example III.1. *Suppose that some protocol uses a classical switch that chooses between the quantum operations (Pauli matrices) I and X, according to the value of some cbit. This can be turned into a quantum controlled-X gate.*

Example III.2 (Superdense coding)**.** *We have seen in Section I, Chapter 2, that by using superdense coding we have: $1 qubit + 1 ebit \geq 2 cbits$. Turning it into a coherent protocol gives rise to the following relation*

$$1 qubit + 1 ebit \geq 2 cobits\,.$$

Example III.3 (Teleportation)**.** *We have seen in Section II, Chapter 2, that by using quantum teleporatation coding we have:* $2cbits + 1ebit \geq 1qubit$*. Turning it into a coherent protocol gives rise to the following relation*

$$2cobits + 1ebit \geq 1qubit + 2ebits\,.$$

IV Private / Secret-Key Classical Capacity

Assume a quantum channel $\mathcal{E}$ from Alice to Bob.

Goal: Transmission of *private information* between Alice and Bob, such that a potential eavesdropper (Eve) could recover and information of what was transmitted.

Environment is the part that Alice and Bob are ignorant of, and therefore functions as "noise" from their point of view.

Strict Assumption: The Eavesdropper (Eve) has access to all the environment E that corresponds to the channel $\mathcal{E}$, and can apply any quantum operation to it.

Definition IV.1 (Private communication)**.** *Private communication is the ability to convey information of some (classical) rate* R *between* A *and* B*, such that asymptotically:*

$$I_{acc}(E; B_T) \equiv I_{acc}(Environment, Message) \overset{n\to\infty}{\longrightarrow} 0\,,$$

where B_T *is the transmitted system.*

Example IV.1 (Coherent communication)**.** *A special case (and* not *equivalent) of Definition IV.1 is the ability to hold coherent communication of rate* R *between* A *and* B*:*

$$|x_A^{nR}\rangle \longrightarrow |x_A^{nR}\rangle|x_B^{nR}\rangle|\theta_E\rangle$$

or, in other words, to perform coherent communication with rate R *from* A *to* B*.*

Definition IV.2 (Private capacity)**.** *The private capacity of channel* $\mathcal{E}$*, denoted by* $C_P(\mathcal{E})$*, is the highest achievable private communication for this channel.*

Remark IV.1. *Theorem I.1 cannot be used directly here, i.e., in general* $C_p(\mathcal{E}) \neq C_{CQ}(\mathcal{E})$*.*

For instance, one notes that C_{CQ} *does not provide coherent communication, since the environment still depends on the message at the end of the protocol.*

Let $\mathcal{E}$ be some quantum channel and define (see also Fig. 8.1): A - The system that stays at the encoder throughout the whole process (Alice). This system is assumed to be in pure state

(since otherwise purification can be performed).
Q - The quantum system at the input to the channel, that is sent to Bob.
E - The environment of the input channel state (of Q).
B - The system at the output of the channel (Bob).
E' - The environment system at the output of the channel (Eve).
$U_{\mathcal{E}}$ - A unitary expansion of $\mathcal{E}$.

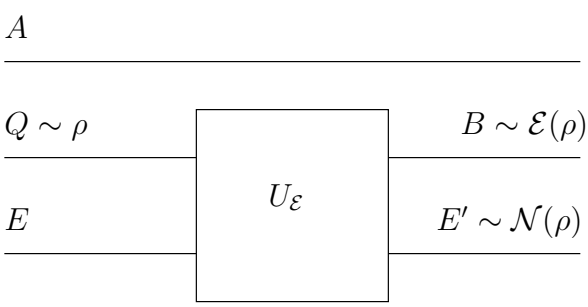


Figure 8.1: Secret Key

Definition IV.3 (Coherent information)**.** *The coherent information of channel $\mathcal{E}$ with input ρ is defined as the difference between the output entropy and the entropy-exchange:*

$$I_c(\rho,\mathcal{E}) \triangleq S(\mathcal{E}(\rho)) - S(\rho,\mathcal{E})\,.$$

Definition IV.4 (Private information)**.** *The private information of channel $\mathcal{E}$ with input ρ is defined as*

$$I_p(\rho,\mathcal{E}) \triangleq I_c(\rho,\mathcal{E}) - \min_{\{p_x,\rho_x\}} \left\{ \sum_x p_x I_c(\rho_x,\mathcal{E}) \middle| \sum_x p_x\rho_x = \rho \right\}\,.$$

Theorem IV.1 (Devetak 2005, [7])**.** *The private classical capacity of channel $\mathcal{E}$ is given by*

$$C_p(\mathcal{E}) = \lim_{n\to\infty} \frac{1}{n} \max_{\rho\in H_{in}^{\otimes n}} I_p\left(\rho,\mathcal{E}^{\otimes n}\right)\,.$$

Sketch of direct. We shall use the input ensemble $\{p_x,\rho_x\}$, $\rho = \sum p_x\rho_x$.

The composite input system AQE is in state

$$\sum_x p_x |x\rangle\langle x|^A \otimes \rho_x \otimes |0\rangle\langle 0|^E\,.$$

The composite output system is in state

$$\sum_x p_x |x\rangle\langle x|^A \otimes \phi_x^{BE'}\,,$$

where,

$$\mathrm{tr}_{E'}(\phi_x^{BE'}) = \mathcal{E}(\rho_x)\,,$$

is the channel from Alice to Bob, and

$$\mathrm{tr}_{B}(\phi_x^{BE'}) = \mathcal{N}(\rho_x)\,,$$

is the channel from Alice to Eve.

According to Theorem I.1, Chapter 6, the rate

$$\chi\left(\mathcal{E}, \{p_x, \rho_x\}\right) = S(\mathcal{E}(\rho)) - \sum p_x S(\mathcal{E}(\rho_x))\,,$$

is achievable. Moreover, according to (5.7), we have

$$S(A;B) = \chi\left(\mathcal{E}, \{p_x, \rho_x\}\right).$$

Applying the same arguments to the "channel" $\mathcal{N}$ from Alice to Eve, we have:

$$S(A;E') = S(\mathcal{N}(\rho)) - \sum p_x S(\mathcal{N}(\rho_x)) \geq 0\,.$$

Coding scheme:

- Block-length n.
- Alice and Bob use approximately $2^{nS(A;B)}$ codewords, meaning Bob may reliably decode (with high probability) $nS(A;B)$ bits.
- Eve can match the message to one of approximately $2^{nS(A;E')}$ distinct sets (with high probability, i.e., to decode approx. $nS(A;E')$ bits.
- This is analogous to the situation in which Eve can recover the $nS(A;E')$ MSBs (with high prob.) while having no effect or access to any knowledge of the rest. Moreover, using random codes it can be shown that such a mapping exists.

Corollary 8.1. *In order to hold* private communication *between Alice and Bob, information ought to be sent only over the left $n\left[S(A;B) - S(A;E')\right]$ LSBs. Thus, the following private rate is achievable:*

$$\begin{aligned} C_p(\mathcal{E}) &\geq S(A;B) - S(A;E') = \\ &= S(\mathcal{E}(\rho)) - S(\mathcal{N}(\rho)) - \sum_x p_x \left[S(\mathcal{E}(\rho_x)) - S(\mathcal{N}(\rho_x))\right] = \\ &= S(\mathcal{E}(\rho)) - S(\rho,\mathcal{E}) - \sum_x p_x \left[S(\mathcal{E}(\rho_x)) - S(\rho_x,\mathcal{E})\right]\,, \end{aligned}$$

where $S(\mathcal{N}(\rho)) = S(\rho,\mathcal{E})$ is the entropy-exchange of Definition I.1, Chapter 7.

Finally taking the limit of $n \to \infty$, we achieve the desired rate.

The converse will not be given here. □

Remark IV.2.

1. *One can show that:* $I_p(\rho, \mathcal{E}) \geq I_c(\rho, \mathcal{E})$. *Equality holds, for instance, when all* $\{\rho_x\}$ *are pure, i.e.,* $\{\rho_x = |\psi_x\rangle\langle\psi_x|^A\}$*: w.l.o.g. the state of the environment (Eve) can be assumed to be some constant pure state* $|0^E\rangle$. *Thus the received joint state of Bob and Eve can be written as:* $|\phi^{SB}\rangle = U_{\mathcal{E}}|\psi_x^A\rangle|0^E\rangle$, *where* $U_{\mathcal{E}}$ *is a unitary operator, as described by Theorem III.1, Chapter 1. Hence, this joint state is pure as well and thus the entropies of its reduced density matrices are equal:* $S(\mathcal{E}(\rho_x)) = S(\mathcal{N}(\rho_x))$.

   There are examples for which $I_p > I_c > 0$, *but it is* not known *whether there is a finite gap between the two informations under maximization, in the limit of* $n \to \infty$.

2. I_p *and* I_p *may take* negative *values, but not under maximization over* ρ, *since for* $\rho = \rho_x = I/d_{in}$, *they are equal to* 0, *as is the case for the choice* $\rho = \rho_x = |\psi\rangle\langle\psi|$, *i.e., some pure state, as explained in Property V.1.*

3. *Similarly to the definition of private communication, one can consider the problem of private/secret-key distribution.*

   Definition IV.5 (Secret-key capacity)**.** *The secret-key capacity, denote by* $K(\mathcal{E})$, *is the maximal rate of common "key" generation, between Alice and Bob, where all communication held between the two is via a quantum channel* $\mathcal{E}$, *such that Eve cannot attain any information of this key sequence.*

   It is evident that $K(\mathcal{E}) \geq C_p(\mathcal{E})$, *since Alice can simply choose a random sequence in an i.i.d. manner according to* $Ber(1/2)$ *and send it, using the private communication scheme. However, one could suggest that since the* exact *key sequence is not interesting as long as it is random, known to both Alice and Bob and no information of it can be recovered by Eve, the secret-key capacity may be strictly larger than the private capacity, at least for some channels. Nonetheless, it turns out that*

$$K(\mathcal{E}) = C_p(\mathcal{E})$$

   for all quantum channels $\mathcal{E}$.

4. *The presence of a classical infinite-capacity* public *channel from Alice to Bob cannot increase* $C_p(\mathcal{E})$ *or* $K(\mathcal{E})$.

5. *It is possible to perform the private communication protocol absolutely coherently only for ensembles for which* $I_p = I_c$, *for instance, ensembles of pure states.*

V Entanglement-Generating Capacity

Definition V.1 (Entanglement-generating capacity)**.** *The Entanglement-generating capacity of a quantum channel $\mathcal{E}$, denoted by $E(\mathcal{E})$, is the maximal possible rate of generating EPR pairs, shared by the transmitter and receiver.*

Theorem V.1. *The Entanglement-generating capacity is upper bounded by the secret-key and private capacities:*

$$E(\mathcal{E}) \leq K(\mathcal{E}) = C_p(\mathcal{E}) \,.$$

Proof. We shall prove that the secret-key capacity upper bounds the entanglement-generating capacity. Out of every ebit, shared by Alice and Bob, they can produce a random bit, using local measurements with respect to the Schmidt basis, as explained in the proof of Theorem I.1. The resulting bits are independent of any information that could be gained by the environment (Eve), since the ebits cannot have and (quantum or classical) dependency on the environment, or else we would get a mixed-state, instead of a pure one, in the joint system of Alice and Bob. □

Theorem V.2. *The entanglement-generating capacity is given by*

$$E(\mathcal{E}) = \lim_{n\to\infty} \frac{1}{n} \max_{\rho\in {H_{in}}^{\otimes n}} I_c(\rho, \mathcal{E}^{\otimes n}) \,.$$

Proof. The converse stems from Theorem V.1. For the achievability we consider the *coherent* key generation protocol, which is possible for $I_p = I_c$, as suggested by Remark IV.2. Hence, using Theorem IV.1, we may convey cobits with rate I_c. Now, using (8.3), we achieve the desired result. □

Property V.1 (I_c, $S(\rho, \mathcal{E})$)**.**

- *W.l.o.g., we take system E to be a purification of AQ, viz., the state of the joint system AQE is some pure state $|AQE\rangle$. Thus the input state of the composite system AQE is pure, as is the output state $|ABE'\rangle$, since it is merely the state $|AQE\rangle$ after passing through the unitary operation $I^A \otimes U_{\mathcal{E}}^{QE}$.*

- *Since $|ABE'\rangle$ is a pure state, $S(E') = S(AB)$, according to Property V.2, Chapter 4. Thus,*

$$I_c(\rho, \mathcal{E}) \triangleq S(\mathcal{E}(\rho)) - S(\mathcal{N}(\rho)) = S(B) - S(E') = S(B) - S(AB) = -S(A|B) \,.$$

- *As mentiond in Remark IV.2, I_c may indeed be negative. However, $E(\mathcal{E}) \geq 0$, since for the choice $\rho = |\psi\rangle\langle\psi|$ we always have $I_c(|\psi\rangle\langle\psi|, \mathcal{E}) = 0$, since for this choice QE and thus also BE' are in pure states, and hence*

$$S(B) = S(E') \Leftrightarrow I_c(\rho, \mathcal{E}) = S(B) - S(E') = 0\,.$$

- *The following notation is often used:*

$$I_c(\rho, \mathcal{E}) \equiv I_c(A > B)\,.$$

  Note that, in general, $I(A > B) \neq I(B > A)$.

- *The coherent information obeys a data processing inequality, as suggested by the following theorem.*

  Theorem V.3 (Quantum data processing inequality)**.** *Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be some two quantum channels and ρ some quantum state. Then*

$$S(\rho) \geq I_c(\rho, \mathcal{E}_1) \geq I_c(\rho, \mathcal{E}_1 \circ \mathcal{E}_2)\,,$$

  with equality iff it is possible to perfectly reverse the operation $\mathcal{E}_1$, i.e., iff there exists a quantum operation $\mathcal{R}$ such that $F_e(\rho, \mathcal{R} \circ \mathcal{E}_1) = 1$.

  This is one of the reasons that for a long time I_c was "suspected" to be the quantum channel capacity, which turned out to be true, see Theorem VI.1.

- *The entropy-exchange $S(\rho, \mathcal{E})$ satisfies the quantum Fano inequality.*

  Theorem V.4 (Quantum Fano inequality)**.** *Let ρ be a quantum state and $\mathcal{E}$ a quantum channel. Then*

$$S(\rho, \mathcal{E}) \leq h_b(F_e(\rho, \mathcal{E})) + (1 - F_e(\rho, \mathcal{E})) \log(d^2 - 1)\,,$$

  where d is the dimension of the input space, and $h_b(\bullet)$ is the binary entropy.

  Thus in order to get high fidelity the entropy exchange ought to be low.

- *Unfortunately, the maximum over the coherent information I_c is* not *additive. Thus in general, we may not remove the "*lim*" in Theorem V.2.*

- *The entropy-exchange $S(\rho, \mathcal{E})$, can be expressed as a VN entropy of a state W, where W depends on ρ and the Kraus operators of $\mathcal{E}$, in the following way:*

$$W_{ij} \triangleq tr(E_i \rho E_j{}^\dagger)\,,$$
$$S(\rho, \mathcal{E}) \equiv S(W) \triangleq -tr(W \log W)\,.$$

VI The Quantum Channel Capacity

Define the following quantum channel and coding scheme:
n - coding block-length.
H - Hilbert space of the information source (increases with n).
H_{in} - channel input Hilbert space.
H_{out} - channel output Hilbert space.
$\mathcal{E}$ - quantum channel (CPTP); $\quad \mathcal{E}: \Delta(H_{\text{in}}) \to \Delta(H_{\text{out}})$.
E - encoder (CPTP); $\quad E: \Delta(H) \to \Delta\left(H_{\text{in}}^{\otimes n}\right)$.
D - decoder (CPTP); $\quad D: \Delta\left(H_{\text{out}}^{\otimes n}\right) \to \Delta(H)$.

The fidelity of the encoding scheme is given by

$$F = \min_{|\psi\rangle \in H} F(|\psi\rangle, (D \circ \mathcal{E}^{\otimes n} \circ E)(|\psi\rangle\langle\psi|)) .$$

Definition VI.1 (Achievable quantum rate)**.** *A rate*

$$R = \frac{1}{n} \log \dim H \left[\frac{qubits}{channel\ use} \right] ,$$

is said to be achievable if there exists a sequence of coding schemes such that the fidelity F tends to 1 as $n \to \infty$.

Definition VI.2 (Quantum capacity)**.** *The quantum capacity of a quantum channel $\mathcal{E}$, denoted by $Q(\mathcal{E})$, is the supremum over all achievable rates of this channel.*

Definition VI.3 (Quantum capacity (alternative definition))**.** *Define the fidelity of the channel $\mathcal{E}$ as $F(\rho) = F_e(\rho, D \circ \mathcal{E}^{\otimes n} \circ E)$ and the rate of the scheme as $R = S(\rho)$. Then the quantum channel capacity, $Q(\mathcal{E})$, is the supremum of all achievable rates, where optimization is done over over all admissible d.m. ρ and D and E.*

Remark VI.1. *Definition VI.2 and Definition VI.3 are equivalent capacity-wise [2].*

Definition VI.4 (Entanglement-assisted quantum capacity)**.** *The Entanglement-assisted quantum capacity is defined as the quantum capacity in the presence of* unlimited *entanglement, and is denoted by $Q_E(\mathcal{E})$.*

Assertion VI.1. *The quantum capacity of a channel $\mathcal{E}$ is upper-bounded by its classical capacity: $Q(\mathcal{E}) \le C(\mathcal{E}) \le C_E(\mathcal{E})$.*

Proof. The second inequality is trivial as the use of entanglement can only increase classical capacity. The first inequality follows from (8.2), since we can transmit 1 cbit using a single qubit. □

Assertion VI.2. *In the presence of (unlimited) entanglement, the quantum capacity is equal to half the classical capacity of this channel:*

$$Q_E(\mathcal{E}) = \frac{1}{2} C_E(\mathcal{E}) \,.$$

Proof. Achievability: Perform quantum teleportation (see Section II, Chapter 2). Hence, according to (2.2), we have $2cbits + 1ebit \geq 1qubit$.

Converse: Perform Superdense coding (see Section I, Chapter 2). Hence, according to (2.1), we have $1qubit + 1ebit \geq 2cbits$.

Thus we establish the equality, in the presence of entanglement:

$$1qubit \overset{1ebit}{\rightleftharpoons} 2cbits \,.$$

Which concludes the proof. □

Assertion VI.3. *The quantum capacity of the erasure channel (see Example I.3, Chapter 6) is* $Q = \max(1 - 2p, 0)$.

Proof. To prove that we first show that for $p = 0.5$ the quantum capacity is $Q = 0$: Define a composite channel from Alice to Bob and Charley, which is composed of erasure channels with $p = 0.5$ to Bob and Charley, such that whenever an erasure occurs at Bob's side, Charley receivers the qubit sent by Alice with fidelity 1 and vice versa. If the capacity of the erasure channel with $p = 0.5$ were strictly positive, then cloning could be performed, as both Bob and Charley would be able to recover the same state, in contradiction to Theorem III.1, Chapter 2.

Hence the quantum capacity of the erasure channel with $p = 0.5$ is $Q = 0$.

An erasure channel with $p < 0.5$ can be "simulated" by time-sharing between a clean channel (with capacity equal to $1qubit$) and an erasure channel with $q = 0.5$ above channel. Therefore, the weighted average of the capacities of the aforementioned channels $1 - 2p$, is achievable. The converse, i.e., that no rate larger than $1 - 2p$, is proved in the home assignment. See also Example VI.2. □

Theorem VI.1 (Devetak 2005, [7]; Bennett et al. 2002, [3])**.** *The quantum channel capacity is given by*

$$Q(\mathcal{E}) = E(\mathcal{E}) = \lim_{n\to\infty} \frac{1}{n} \max_{\rho \in H_{in}^{\otimes n}} I_c(\rho, \mathcal{E}^{\otimes n}) \,.$$

To prove this theorem we define three more quantum capacities of channels in the presence of different classical channel resources.

Definition VI.5 (Feedforward quantum capacity)**.** *The feedforward quantum capacity of a quantum channel $\mathcal{E}$, denoted by $Q_{\rightarrow}(\mathcal{E})$, is the supremum over all achievable rates of this channel in the presence of an unlimited classical channel from the* transmitter *to the* receiver*.*

Definition VI.6 (Feedback quantum capacity)**.** *The feedback quantum capacity of a quantum channel $\mathcal{E}$, denoted by $Q_{\leftarrow}(\mathcal{E})$, is the supremum over all achievable rates of this channel in the presence of an unlimited classical channel from the* receiver *to the* transmitter*.*

Definition VI.7 (Classically assisted quantum capacity)**.** *The classically assisted quantum capacity of a quantum channel $\mathcal{E}$, denoted by $Q_{\leftrightarrow}(\mathcal{E})$ or Q_2, is the supremum over all achievable rates of this channel in the presence of an unlimited* bidirectional *classical channel, i.e., both a channel from the* transmitter *to the* receiver *and a channel from the* receiver *to the* transmitter*.*

Definition VI.8 (Feedforward entanglement-generating capacity)**.** *The feedforward entanglement-generating capacity of a quantum channel $\mathcal{E}$, denoted by $E_{\rightarrow}(\mathcal{E})$, is the maximal possible rate of generating EPR pairs, shared by the transmitter and receiver, in the presence of un unlimited classical channel from the* transmitter *to the* receiver*.*

Remark VI.2. *The feedback and classical assisted entanglement-generating capacity can be defined in a similar manner.*

Proof. Converse: Combining the results of (8.2) (8.3), we see that $1qubit \geq 1ebit$. Thus, $Q(\mathcal{E}) \leq E(\mathcal{E})$.

Achievability: We break the achievability proof into two stages.

Stage 1: We show that $E(\mathcal{E}) \leq Q_{\rightarrow}(\mathcal{E})$.

A presence of a classical channel between transmitter and receiver can only increase the entanglement-generating capacity of this channel. Thus $E(\mathcal{E}) \leq E_{\rightarrow}(\mathcal{E})$.

But having generated $E_{\rightarrow}(\mathcal{E})$ ebits per channel use, we may perform quantum teleportation of $E_{\rightarrow}(\mathcal{E})$ qubits (with fidelity 1), by making use of the classical channel available. Thus

$$E(\mathcal{E}) \leq E_{\rightarrow}(\mathcal{E}) \leq Q_{\rightarrow}(\mathcal{E})\,.$$

Stage 2: We show (not rigourously) that $Q_{\rightarrow}(\mathcal{E}) = Q(\mathcal{E})$, for any CPTP $\mathcal{E}$, i.e., that the presence of a forward classical channel (*alone*) *does not* increase quantum capacity.

The most general protocol for conveying qubits is the following:

- Alice chooses one one of the encoders $\{E_k\}$ according to some strategy (for instance, randomly w.p.t. some measurement result), when k can be random.

- Alice conveys the value of k to Bob over the classical channel and the encoded state over the quantum channel.
- Bob opearstes with the decoder D_k.

We need this scheme to have a good fidelity. Thus, to achieve a certain rate R, we need the following to be satisfied: for every $\epsilon > 0$ there exist $n, \{D_k\}, \{E_k\}$ such that

$$\mathbb{E}_k \left\{ \min_{|\psi\rangle} F(|\psi\rangle, D_k \circ \mathcal{E}^{\otimes n} \circ E_k(|\psi\rangle\langle\psi|) \right\} \geq 1 - \epsilon \,.$$

Note that k cannot depend on $|\psi\rangle$ (at least not in the limit of $n \to \infty$), since otherwise it would mean applying a "destructive measurement" of the state to be transmitted, which would prevent its perfect reconstruction at the decoder ($F \to 1$).

Therefore, there exists some $k = j$, which does not depend on $|\psi\rangle$, such that

$$\min_{|\psi\rangle} F(|\psi\rangle, D_j \circ \mathcal{E}^{\otimes n} \circ E_j(|\psi\rangle\langle\psi|)) \geq 1 - \epsilon \,.$$

Thus, using E_j and D_j, we obtain the same rates without the need of any classical communication, meaning:

$$Q_{\rightarrow}(\mathcal{E}) = Q(\mathcal{E}) \,.$$

□

Example VI.1 (Depolarizing channel)**.** *An explicit expression for the quantum capacity of the depolarizing channel (see Example I.1, Chapter 6) is not known in general. Nevetheless, it is lower bounded by*

$$Q \geq 1 - h(\frac{3}{4}p) - \frac{3}{4}p \cdot \log_2 3$$

meaning $Q > 0$ for $p < 0.254$.

On the other hand, it can be shown that $Q = 0$ for $p > 1/3$.

Example VI.2 (Erasure channel)**.** *The choice $\rho = \frac{I}{2^n}$, achieves capacity asymptotically:*

$$Q == \lim_{n\to\infty} \frac{1}{n} I_C \left(\frac{I}{2^n}, \mathcal{E}^{\otimes n} \right) = \max(1 - 2p, 0) \,.$$

VI.1 Classically Assisted Quantum Capacity

As was shown in the proof of Theorem VI.1, classical *feedforward* does not increase quantum capacity. Nevertheless classical *feedback*, all the more so bidirectional classical communication, *can* increase quantum capacity, as shown in the following example.

Example VI.3 (Feedback quantum capacity of erasure channel)**.** *The feedback quantum capacity of the erasure channel is equal to its classical capacity:*

$$Q_{\leftrightarrow} = 1 - p = C\,,$$

as is proved in the home assignment. Note that this is strictly larger than the its quantum capacity (without any classical assistance), which is equal to $Q = 1 - 2p < 1 - p = Q_{\leftrightarrow}$.

Remark VI.3. *There are examples of channels for which* $Q < Q_{\leftrightarrow} < C$.

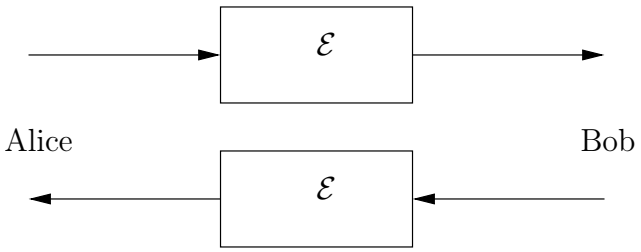
Open Question: Is there a quantum channel for which $Q_{\leftrightarrow} > C$?

Open Question: TO find expressions for $Q_{\leftrightarrow}$ and $Q_{\leftarrow}$.

The Example from the Beginning of the Semester

Recall Example I.1, Chapter 1.

Proposition VI.1. *There are quantum channels* $\mathcal{E}$ *for which* $Q(\mathcal{E})$*, while the quantum bidirectional system, consisting of a channel* $\mathcal{E}$ *for each side (see Fig. VI.1) has a positive capacity for quantum information.*



Sketch. Assume a non-fixed mapping (CPTP) $\mathcal{E}$, i.e., that does not map all admissible states to the same d.m.. The classical capacity of such a channel is strictly positive, according to Remark 6.1, Chapter 6. We further assume that $Q_{\leftrightarrow}(\mathcal{E}) > 0$. This is possible, as demonstrated by Example VI.3 (even if $Q(\mathcal{E}) = 0$).

Since $C(\mathcal{E}) > 0$, we can use both channels for classical information exchange to obtain some (striclty positive) portion of $Q_{\leftrightarrow}(\mathcal{E})$. Thus a strictly positive rate of quantum communication is achievable using this system.

This protocol can be materialized using the following protocol.

1. Alice prepares n copies of some state ρ^{AB}.

2. Alice transmits the B parts to Bob over the channel $\mathcal{E}$.

3. Alice and Bob share n copies of the state $\sigma^{AB} = (I \otimes \mathcal{E})\rho^{AB}$.

4. Assuming that $E_d(\sigma^{AB}) > 0$ (true for at least some state since $\mathcal{E}$ is non-fixed), Alice and Bob can prepare $nE_d(\sigma^{AB})$ ebits out of many copies of σ^{AB}, working under the LOCC model (see Section II, Chapter 7).

5. Alice and Bob perform quantum teleportation (Section II, Chapter 2), using the distilled ebits and the classical communication.

□

Bibliography

[1] H. Barnum, C. A. Fuchs, R. Josza, and B. Schumacher. General fidelity limit for quantum channels. *Phys. Rev. A*, 54:4707, 1996.

[2] H. Barnum and E. Knill M. A. Nielsen. On quantum fidelities and channel capacities. *IEEE Trans. Information Theory*, IT-46:1317–1329, July 2000.

[3] C. H. Bennett, P. W. Shor J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Trans. Information Theory*, IT-48:2637–2655, Oct. 2002.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

[5] E. Davies. Reversing quantum dynamics with near-optimal quantum and classical fidelity. *Journal of Mathematical Physics*, vol. 43:2097–2106, May 2002.

[6] E. Davies. Information and quantum measurement. *IEEE Trans. Information Theory*, IT-24:596–599, Sep. 1978.

[7] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Information Theory*, IT-51:44–55, 2005.

[8] C. Fuchs. Nonorthogonal quantum states maximize classical information capacity. *Phys. Rev. Lett.*, 79:1162–1165, 1997.

[9] P. Hayden, R. Jozsa, D. Petz, and A. Winter. Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Communications in Mathematical Physics*, vol. 246, num. 2:359–374, Feb. 2004.

[10] A. S. Holevo. Statistical problems in quantum physics. In *Proc. of the Second Japan-USSR Symposium on Probability Theory, Springer-Verlag, Berlin, 1973.*

[11] M. A. Nielsen and I. J. Chuang. *Quantum Computation and Quantum Information*. Cambridge Univ. Press, Cambridge, 2000.

[12] B. Schumacher. Quantum coding. *Phys. Rev. A*, 51:2738–2747, 1995.

[13] P. W. Shor. The classical capacity achievable by a quantum channel assisted by limited entanglement. pages 2637–2655, submitted, 2004.