Chapter 8

Capacity of quantum channels

8.1 Overview

In this chapter we consider the problem of transmission of information through quantum channels. A sender wants to communicate a message to a receiver through a quantum noisy channel As we will see in more detail later, a noisy quantum channel can be reasonably modeled by a *linear completely positive map* from the space of density matrices (input) to the space of density matrices (output)

$$\mathcal{N} : \rho_{input} \to \mathcal{N}(\rho_{input}) = \rho'_{output}$$

In the quantum setting even the problem of point to point communication is much richer than in the classical case and questions of principle such as the analogs of Shannon's theorem are not completely solved. We start the chapter with an overview of the main settings that hav been explored.

In general the encoder will input in the channel "words" of length N. The words of length N are states of a Hilbert space $\mathcal{H}^{\otimes N}$ where the single letter space $\mathcal{H}$ has $\dim \mathcal{H} = d$. Here we will always have $d = 2$ in mind (so the channels are "binary-input", i.e their inputs are Qbits, for example polarization states of photons). As the following discusion will make clear there are many possible settings that one can imagine. Let us briefly descibe three broad situations have been discussed in the literature. In the sequel we treat in more detail only the first one.

Sending classical messages through quantum channels. This situation is the closest possible to the classical one and is depicted on figure 8.1. Classical messages from a set $\{1, ..., M\}$ of M possible messages are to be sent by Alice to Bob and should be reproduced faithfully by Bob. Alice encodes

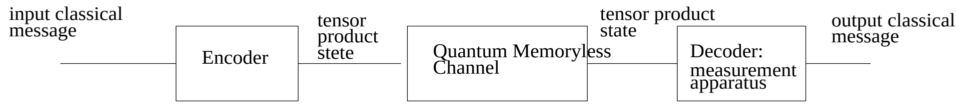

Figure 8.1: Model of classical-quantum communication

them as input code words that are N fold tensor product states:

$$\rho_{x_1} \otimes .... \otimes \rho_{x_N}$$

where each ρ_x is chosen (according to some encoding rule) in the 2×2 density matrices of $\mathcal{H}$. Strictly speaking the codewords are guenuinely quantum, except if the states of terms in the tensor product are mutualy orthogonal (and thus distinguishable). These are transmitted over a quantum channel (photons are sent through an optic fiber) and the output is in general a quantum state of $\mathcal{H}^{\otimes N}$. If the channel is memoryless (and binary-output say) the output state will again be a tensor product:

$$\rho'_{x_1} \otimes .... \otimes \rho'_{x_N}$$

In order to decode, the decoder has to measure the state.

Consider for a moment the (seemingly) simplest situation where he uses measurements that act on each term of the tensor product separately. In other words he observes the output Qbits "one by one". Moreover suppose each measurement is done independently from previous measurement outcomes. Shannon's classical theory implies that the capacity[1] of this encoding-decoding scheme is given by

$$\max_{p_x, \rho_x} \sup_{\{P_y\}} I(X;Y), \qquad \rho'_x = \mathcal{N}(\rho_x) \tag{8.1}$$

The p_x in the optimization can be interpreted as the probability distribution according to which the encoder chooses the Qbits ρ_x. We recognize in this expression the "accessible information" (chap 7) is

$$Acc(X; \sum_x p_x \rho'_x) = \sup_{\{P_y\}} I(X;Y) \tag{8.2}$$

The maximum accesible information (8.1) has been called by some authors the $C_{1,1}$ capacity of the channel. The first 1 means that we allow only tensor product states and the second 1 means that we allow only measurements

[1]that is the maximum achievable rate $R = \frac{1}{N} \log_2 M$ for reliable communication. The reliability criterion is given in later section

that act on one received Qbit at a time. As we learned in chap 7 there is no known information theoretic expression for the optimzation problem (8.2) but at least we have the Holevo upper bound[2] $Acc(X;\mathcal{N}(\rho)) \leq \chi(X;\mathcal{N}(\rho))$ which implies

$$C_{1,1} \leq \max_{p_x,\rho_x} \chi(X;\mathcal{N}(\rho)) = \sup_{p_x,\rho_x}\left[S(\mathcal{N}(\sum_x p_x\rho_x) - \sum_x p_x S(\mathcal{N}(\rho_x))\right]$$

Clearly the capacity may only increase if we increase the space of possible measurements. The $C_{1,\infty}$ capacity denotes the maximum achievable rate when there is no restriction on the type of measurements of the output state. More precisely, the measurement basis acts on the whole of $\mathcal{H}^{\otimes N}$ (and as we will see we let $N \to +\infty$). It turns out that when we allow general measurements the Holevo bound is achieved (as we will show). This is the content of the Holevo-Schumacher-Westmoreland (HSW) theorem and this capacity is commonly called the *classical-quantum* or *product state* or $C_{1,\infty}$ capacity. We have

$$C_{1,\infty} = \max_{p_x,\rho_x}\left[S(\mathcal{N}(\sum_x p_x\rho_x) - \sum_x p_x S(\mathcal{N}(\rho_x))\right] = \sup_{p_x,\rho_x} \chi(X;\mathcal{N}(\rho))$$

This formula can be seen as the quantum analog of Shannon's

$$C = \max_{p_x}[H(Y) - H(Y|X)] = \max_{p_x} I(X;Y)$$

But does this *really* give the maximum achievable rate of transmission of a classical message through a quantum channel ? The answer is not known in general. Indeed one could allow to encode classical messages into entangled states (not just tensor products as above). Then one sends the Qbits still one by one through the quantum channel (remember the Qbit is embodied by a photon which is simply sent through the optic fiber say). The output state is still more or less entangled in general and we allow general measurements (see figure 8.1). The maximum achievable rate in this situation is called $C_{\infty,\infty}$ and it has been conjectured, but not known, that $C_{\infty,\infty} = C_{1,\infty}$. It is non trivial to exclude that entanglement of the inputs do not help increasing the capacity. In fact this conjecture has been shown to be equivalent to the additivity conjecture for the composition of two channels

$$C_{1,\infty}(\mathcal{N}_1 \otimes \mathcal{N}_2) = C_{1,\infty}(\mathcal{N}_1) + C_{1,\infty}(\mathcal{N}_2)$$

Note that superadditivity is known($\geq$); it is the $\leq$ inequality that is really hard. For some class of channels there are proofs though.

[2] here $\rho = \sum_x p_x\rho_x$ and by linearity of the channel map $\mathcal{N}(\rho) = \sum_x p_x\mathcal{N}(\rho_x)$

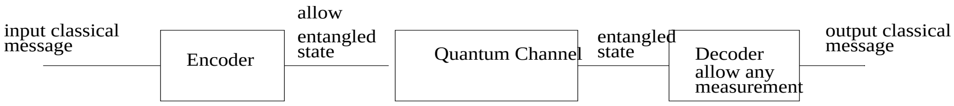

Figure 8.2: More general model of classical-quantum communication: one would like to know the $C_{\infty,\infty}$ capacity

Entanglement assisted communication. As argued above it is conjectured that entanglement of the input dos not help increase the capacity. Is there another way to use entanglement ? When we introduced dense coding (also called superdense coding in the litterature) we saw that Alice can transmit two bits of classical information to Bob through a noiseless channel, by first sharing an EPR pair with him and then sending her particle (one Qbit). This was a noiseless situation where the capacity is twice the unassisted one. This suggests that the transmission rate for classical messages can be increased by sharing entanglement between the sender and the receiver. A situation analyzed recently by Bennett-Shor-Smolin-Thapliyal is that of transmitting classical messages through a quantum channel allowing the sender and receiver to share an unlimited amount of EPR pairs (figure 8.1). Assume that beforehand, Alice and Bob share an unlmited number of entangled pairs thanks to some noiseles medium. Alice encodes classical messages by performing operations *on her Qbits in her lab* and then sends them through a noisy quantum channel. This time there is not limitation to tensor product states in the encoding map of Alice. As before Bob can do any measurement on his side. In other words this is a kind of ∞, ∞ situation and the channel is $\mathcal{N} \times I$. The maximum achievable rate has the singe letter characterization

$$C_E = \max_{\rho}[S(\rho) + S(\mathcal{N}(\rho)) - S((\mathcal{N} \otimes I)|\Psi_\rho\rangle)]$$

where the maximum is taken over density matrices (here think of 2×2 matrices) in the Hilbert space (of letters) of Alice $\mathcal{H}_A$. The state $|\Psi_\rho\rangle$ is a purification of ρ in the bigger space $\mathcal{H}_{Alice} \otimes \mathcal{H}_{Bob}$. In other words $Tr_{Bob}|\Psi_\rho\rangle\langle\Psi_\rho| = \rho$. One can show that the third term does not depend on the purification that is choosed[3]. The bracket $[-]$ is a well defined functional of ρ only.

The proof of this formula uses Holevo's formula for blocks and the channel $\otimes_{i=1}^n(\mathcal{N} \otimes I)$. The optimization can then be reduced to the single letter characterization by the use of dense coding and an additivity property

$$C_E(\mathcal{N}_1 \otimes \mathcal{N}_2) = C_E(\mathcal{N}_1) + C_E(\mathcal{N}_2)$$

[3] in the achievability part of the proof this purification is fixed at the outset by the shared EPR pairs: the encoding map acts only on Alice's particle

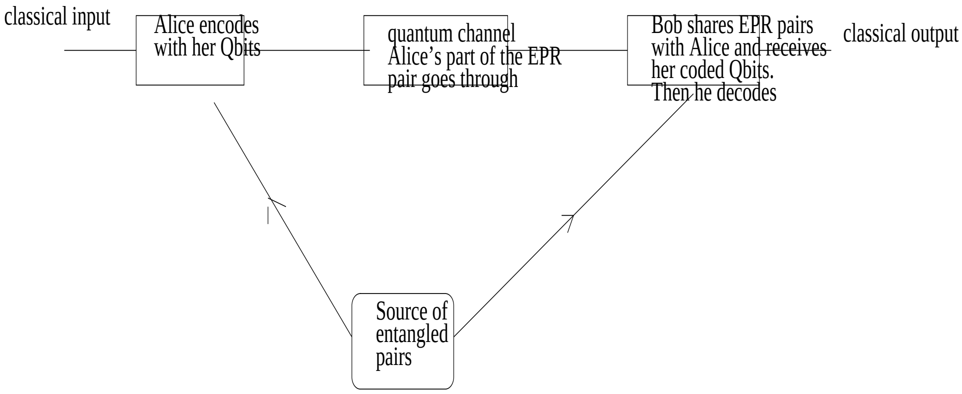

Figure 8.3: Alice and Bob share entanglement in order to communicate

We will not prove the capacity formula in the course, but just mention that it can be guessed by noting that it has a close ressemblance to Shannon's

$$C = \max_{p_x}[H(X) + H(Y) - H(X;Y)] = \max_{p_x} I(X;Y)$$

To see the analogy first note that $S(\rho)$ and $S(\mathcal{N}(\rho))$ correspond to $H(X)$ and $H(Y)$. But why is it that $S((\mathcal{N} \otimes I)|\Psi_\rho\rangle)$ corresponds to $H(X,Y)$? This is because once Bob receives Alice's particles he holds in his lab the whole state $(\mathcal{N} \otimes I)|\Psi_\rho\rangle$ which has partial traces

$$Tr_{Alice}|\Psi_\rho\rangle\langle\Psi_\rho| = \rho, \qquad Tr_{Bob}|\Psi_\rho\rangle\langle\Psi_\rho| = \mathcal{N}(\rho)$$

Here it is important to stress that it is thanks to the shared entangement that both states ρ and $\mathcal{N}(\rho)$ "exist at the same time" in the full received state. In the HSW situation of the previous paragraph these states do not exist at the same time ! Once ρ is sent, $\mathcal{N}(\rho)$ is received but ρ has disapeared. Because of the no-cloning theorem Alice cannot copy ρ before sending it. In classical information theory it is always possible to copy X (this is hat happens when you send a fax) so that X and Y "can exist at the same time". These remarks perhaps explain why classicaly the two mutual informations $H(Y)-H(Y|X)$ and $H(X)+H(Y)-H(X,Y)$ are equal but quantum mechanically become $C_{1,\infty}$ and C_E which are different.

Sending quantum messages through quantum channels. Suppose that the source produces quantum messages. These are quantum states and it is meaningful to linearly combine them (superposition principle). We define such a set of quantum messages as an M dimensional Hilbert space. So taking a basis of M (say othogonal) states the source produces the $span\{|1\rangle, |2\rangle, ..., |M\rangle\}$. Note that this span is isomorphic to the tensor

Figure 8.4: Transmitting quantum messages: model of quantum-quantum communication

product $\mathcal{H}^{\otimes \log_2 M}$ (to see this think of the dimension of this space) where $\mathcal{H}$ is a single Qbit space. Alice has to send any message in this span reliably to Bob. She encodes the message by a map from the message space to a Hilbert space $\mathcal{H}^{\otimes N}$ where $\mathcal{H}$ is a single Qbit space. The rate of the code is defined as $\frac{1}{N}\log_2 M$. Bob decodes by a map from $\mathcal{H}^{\otimes N}$ to $span\{|1\rangle, |2\rangle, ..., |M\rangle\}$. The reliability criterion is that the overlap (inner product) between the original and decoded state is cose to one. Note that the decoder should not observe the quantum states in order not to destroy them (this does not mean that this communication is useless. The sender might want to transmit a quantum state to a receiver that needs it to perform a specific task even if he does not observe this state). The maximum achievable rate for reliable communication, sometimes called the quantum-quantum capacity, has been shown to be equal to (Shor)

$$Q = \lim_{N\to\infty} \frac{1}{N} \sup_{\rho_N} [S(\mathcal{N}(\rho_N)) - S((\mathcal{N}\otimes I)|\Psi_{\rho_N}\rangle)]$$

where $|\Psi_{\rho_N}\rangle$ is a purification of ρ_N in a bigger space $\mathcal{H}^{\otimes N}\otimes\mathcal{R}$. Again, one can show that this quantity does not depend on the particular purification. However this formula is not a "single letter expression" (it is not additive) and the optimization is intractable (even for $N = 1$ it is hard). This expression does not have a clear classical analog, and has been called "coherent information".

Comparison of various capacities.

It is a general fact that

$$C_{1,\infty} \leq C_{\infty,\infty} \leq C_E$$

Moreover Q generaly vanishes above some noise threshold above which quantum coherence (linear superpositions) cannot be maintained faithfuly.

It is in general difficult to compute the quantum capacities, but there are a certain number of situations where this has been done exactly. The simplest situation is perhaps the case of the quantum erasure channel. In the classical case the erasure channel transmit the input from some alphabet $\{1, 2, ..., d\}$ intact with probability $1-p$ and outputs an erasure symbol E with probability p. Shannon's capcity is $C = (1-p)\ln d$. The quantum version is as follows. A state vector from a d dimensional Hilbert space $\mathbf{C}^d$ is transmitted intact with probability $1-p$ and outputs a specific extra state

$|E\rangle \perp \mathbf{C}^d$ with probability p (so the output Hilbert space is the direct sum $\mathbf{C}^d \oplus \mathbf{C}$). The result is:

$$\begin{aligned}&C_{1,\infty} = C_{\infty,\infty} = (1-p)\ln d\\ &C_E = 2(1-p)\ln d\\ &Q = (1-2p)\ln d,\ \ 0 \le p \le \frac{1}{2} \qquad Q = 0\ \ p \ge \frac{1}{2}\end{aligned}$$

For further examples we refer the reader to the litterature.

8.2 Noise in open quantum systems

At the fundamental level for an isolated system there is no such thing as noise; indeed the dynamics is purely deterministic and given by a unitary operator U acting on states of $\mathcal{H}_S$. When we observe the system and make measurements the outcomes are described by the measurement postulate which does not involve any noisy element. In fact this is also true classicaly except that the dynamics is given by Newton's law and the measurements are purely deterministic.

Noise is a phenomenological concept that is useful to describe the influence of the environnement on an open system. let $\mathcal{H}_S$ be the Hilbert space of the system of interset which interacts with some environnement described by the space $\mathcal{H}_E$. the total space is $\mathcal{H}_S \otimes \mathcal{H}_E$. We will suppose that initialy the two systems are decoupled and have state $\rho_S \otimes \rho_E$. The evolution operator U acts nontrivialy on the tensor product and couples the two systems: after some time the total system's state is

$$U(\rho_S \otimes \rho_E)U^\dagger = \rho_{total}$$

If we wnat to describe only the dynamics of $\mathcal{H}$ we use the partial density matrix

$$\rho_S' = Tr_E \rho_{total}$$

For simplicity let us suppose that the environnement is initialy in a pure state $\rho_E = |E\rangle\langle E|$. Then

$$\rho_S' = \sum_k \langle k|U\rho \otimes |E\rangle\langle E|U^\dagger|k\rangle = \sum_k \langle k|U|E\rangle \rho_S \langle E|U^\dagger|k\rangle$$

This is of the form

$$\rho_S' = \sum_k N_k \rho_S N_k^\dagger, \qquad N_k = \langle k|U|E\rangle$$

Here N_k are operators $\mathcal{H}_S \to \mathcal{H}_S$ (because $U : \mathcal{H}_S \otimes \mathcal{H}_E \to \mathcal{H}_S \otimes \mathcal{H}_E$). These operators satisfy a completeness relation

$$\sum_k N_k^\dagger N_k = \sum_k \langle E|U^\dagger|k\rangle\langle k|U|E\rangle = \langle E|U^\dagger U|E\rangle = I$$

A model of noise. We arrive at the conclusion that although for the whole system the evolution is unitary, for the open part $\mathcal{H}$ it is described by a set of operators $\{N_k : \mathcal{H} \to \mathcal{H}\}$,

$$\rho \to \rho' = \sum_k N_k \rho N_k^\dagger, \qquad \sum_k N_k^\dagger N_k = I \tag{8.3}$$

The reader can easily verify that $\rho' \geq 0$ and that $Tr\rho' = 1$. The evolution of an open quantum system is not necessarily unitary. But at least it is linear and it maps density matrices into density matrices. The operators $\{N_k\}$ describe the noise. Their choice defines what we call a model of quantum noise.

Analogy with a transition probability matrix. The density matrices ρ and ρ' have spectral decompositions which allow to interpret them as classical mixtures (this is a mathematical interpretation, it does not mean the system is prepared in a classical state). This allows then to interpret the map $\rho \to \rho'$ as a transition probability matrix. Indeed let $\rho = \sum_x p_x|x\rangle\langle x|$ and $\rho' = \sum_y p'_y|y\rangle\langle y|$. We have

$$\rho' = \sum_x p_x N_k|x\rangle\langle x|N_k^\dagger$$

Inserting to closure relations, we get

$$\rho' = \sum_{z,z'}(|z\rangle\langle z'|)(\sum_x p_x\langle z|N_k|x\rangle p_x\langle x|N_k^\dagger|z^\dagger\rangle)$$

In other words ρ' has (zz') matrix elements

$$\sum_x p_x\langle z|N_k|x\rangle\langle x|N_k^\dagger|z^\dagger\rangle$$

Let D be the matrix that diagonalises it. We must have

$$p(y) = \sum_x p_x \sum_{z,z'} D_{yz}\langle z|N_k|x\rangle\langle x|N_k^\dagger|z^\dagger\rangle D_{z'y}$$

We see on this formula that the map between ρ and ρ' can be described by a transition probability matrix

$$p(y|x) = \sum_{z,z'} D_{yz}\langle z|N_k|x\rangle\langle x|N_k^\dagger|z^\dagger\rangle D_{z'y}$$

Superoperators. A superoperator is defined as a map of the form (8.3). There are a few fundamental properties that such maps obey.

Property 1. Any superoperator has a unitary representation on a higher dimensional Hilbert space. In other words any superoperator may be viewed as coming from the effect of some environnement on a quantum system. Indeed given $\rho' = \sum_{k=1}^{M} N_k \rho N_k^\dagger$ consider a Hilbert space $\mathcal{H}_E$, $\dim \mathcal{H}_E = M$ and define $U : \mathcal{H} \otimes \mathcal{H}_E \to \mathcal{H} \otimes \mathcal{H}_E$ such that $|\phi\rangle \otimes |E\rangle \to \sum_{k=1}^{M} (N_k|\phi\rangle) \otimes |E \oplus k\rangle = U(|\phi\rangle \otimes |E\rangle)$ where $\oplus$ means the sum $mod\, M$. One can easily check that U preserves the inner product so that it is unitary. Moreover one can check that $\rho' = Tr_E(U(\rho \otimes |E\rangle\langle E|)U^\dagger)$.

Property 2. Superoperators are linear maps from density matrices to density matrices (check that). In fact the converse is also true. This is the content of the Kraus representation theorem that we do not prove here.

Theorem 1 (Kraus). *Let $\mathcal{N} : \rho \to \mathcal{N}(\rho)$ be a map satisfying*

- $\mathcal{N}(\rho) \geq 0$ if $\rho \geq 0$
- $Tr\mathcal{N}(\rho) = 1$ if $Tr\rho = 1$
- $\mathcal{N}(\alpha\rho_1 + \beta\rho_2) = \alpha\mathcal{N}(\rho_1) + \beta\mathcal{N}(\rho_2)$
- $\mathcal{N}(\rho)^\dagger = \mathcal{N}(\rho) if \rho^\dagger = \rho$

Then the map $\mathcal{N}$ can be represented as a superoperator for some set of $\{N_k\}$. This also means that $\mathcal{N}(\rho)$ can also be viewed as the reduced density matrix of a larger system evolving unitarily.

With this theorem you see that if you accept that state of a quantum system are density matrices (as advocated by von Neumann) you are also inevitably led to the fact that quantum evolution is described by unitary operators (at the expense of enlarging and/or purifying the system).

Examples of superoperators.

- A unitary evolution $\rho' = U\rho U^\dagger$
- A measurement that does not record the measurement results $\rho' = \sum_i P_i \rho P_i$, $\{P_i\}$ a measurement basis.
- The channel models of next section.

8.3 Models of noisy quantum channels

Physicaly a quantum channel is a medium, supporting the propagation or storage of quantum states (pure or mixed), which is subject to interaction with the environnement. The preceding analysis of open quantum system makes it quite natural to model a noisy Q-channel by a superoperator (figure ??). If for example the input and the output are Qbits (2×2 matrices in the Bloch sphere) we may speak of a binary input-binaryoutput channel. (But note that the Hilbert space of the input and that of the output need not have the same dimension, which means that N_k are not square matrices.) In the course we consider only memoryless channels. This means that

$$\mathcal{N}(\rho_1 \otimes ... \otimes \rho_N) = \mathcal{N}(\rho_1) \otimes ... \otimes \mathcal{N}(\rho_N)$$

Moreover we will restrict ourselves to the binary-input binary-output case. In the later case the most general input state is (one term of the tensor product above) of the form

$$\rho = \frac{1}{2}(I + \mathbf{a} \cdot \Sigma), \qquad |\mathbf{a}| \leq 1$$

The superoperator has to be linear and map this density matrix to another density matrix of the form

$$\rho' = \frac{1}{2}(I + \mathbf{a}' \cdot \Sigma), \qquad |\mathbf{a}'| \leq 1$$

It is posssible to show that necessarily

$$\mathbf{a}' = M\mathbf{a} + \mathbf{c}, \qquad M = OS$$

were O is a real orthogonal matrix and S is real symmetric. The channel has the effect of deforming and rotating the Bloch sphere. We now review the most common examples of channels.

Bit flip channel.

$$\rho' = N_0 \rho N_0^\dagger + N_1 \rho N_1^\dagger = p\rho + (1-p) X \rho X$$

with

$$N_0 = \sqrt{p} Id, \qquad N_1 = \sqrt{1-p} X$$

This is anlogous to the BSC(p) channel as can be seen from the unitary representation in a larger Hilbert space

$$U(\alpha|0\rangle + \beta|1\rangle) \otimes |0,1\rangle = \sqrt{p}(\alpha|0\rangle + \beta|1\rangle) \otimes |0,1\rangle + \alpha|1\rangle + \beta|0\rangle) \otimes |1,0\rangle$$

Phase flip channel.

$$\rho' = N_0 \rho N_0^\dagger + N_1 \rho N_1^\dagger = p\rho + (1-p) Z\rho Z$$

with

$$N_0 = \sqrt{p} Id, \qquad N_1 = \sqrt{1-p} Z$$

This is anlogous to the BSC(p) channel as can be seen from the unitary representation in a larger Hilbert space

$$U(\alpha|0\rangle + \beta|1\rangle) \otimes |0,1\rangle = \sqrt{p}\alpha|0\rangle + \beta|1\rangle) \otimes |0,1\rangle + \alpha|1\rangle - \beta|0\rangle) \otimes |1,0\rangle$$

The reader can check that this is a bit flip in the rotated basis $\{H|0\rangle, H|1\rangle\}$ or X basis.

Bit-Phase flip channel. This cahnnel is a sucession of bit and phase flips. In fact it is a bit flip in the Y basis. More formaly

$$\rho' = N_0 \rho N_0^\dagger + N_1 \rho N_1^\dagger = p\rho + (1-p) Z\rho Z$$

with

$$N_0 = \sqrt{p} Id, \qquad N_1 = \sqrt{1-p} Y = \sqrt{1-p} XZ$$

Depolarizing channel. In this case a Qbit is completely depolarized with probability p and left untouched with probability (1-p). This is the analog of the classical binary erasure channel (BEC) where the bit value is erased with probability p and left untouched with probability $1-p$. Here

$$\rho' = \frac{p}{2} I + (1-p)\rho$$

It is not immediately clear that this compatible with the superoperator represemntation (but from the Kraus theorem it has to be !). We notice that

$$\frac{I}{2} = \frac{1}{4}(\rho + X\rho X + Y\rho Y + Z\rho Z)$$

for any ρ (exercises). Thus

$$\rho' = (1 - \frac{3p}{4})\rho + \frac{p}{4}(X\rho X + Y\rho Y + Z\rho Z)$$

The elements of this superoperator are $\{N_k\} = \{\sqrt{1 - \frac{3p}{4}}, \frac{\sqrt{p}}{2} X, \frac{\sqrt{p}}{2} Y, \frac{\sqrt{p}}{2} Z\}$. This last representation is not unique, indeed

$$\rho' = (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z)$$

A unitary representation is

$$U|\phi\rangle \otimes |0,1,2,3\rangle = \sqrt{1-p}|\phi\rangle \otimes |0,1,2,3\rangle + \sqrt{\frac{p}{3}}X|\phi\rangle \otimes |1,2,3,0\rangle + \sqrt{\frac{p}{3}}Y|\phi\rangle \otimes |2,3,0,1\rangle + \sqrt{\frac{p}{3}}Z|\phi\rangle \otimes |3,0,1,2\rangle$$

Amplitude damping channel.

Phase damping channel.

8.4 Product State Capacity: HSW theorem

In this paragraph we adress the classical-quantum communication problem and prove the HSW theorem. A sender has M messages $m \in \{1, ..., M\}$. These messages can be described by $\log_2 M$ bits and we assume that the source produces the messages with uniform probability (say some compression has been done beforehand). Each message is encoded as strings of N Qbits in states that are of tensor product form. The alphabet for the channel is $\rho_x \in \{\rho_1, ..., \rho_D\}$ (has size D) where we restrict (this is not loss of generality) to 2×2 matrices. Each message is encoded as a block of length N

$$\mathcal{E} : m \to \rho_{x_1(m)} \otimes \rho_{x_2(m)} \otimes ... \otimes \rho_{x_N(m)}$$

where the code introduces redundancy because we take $N > \log_2 M$. This map defines the code C. The rate of the code is $R = \frac{\log_2 M}{N}$. The code word is sent through a quantum memoryless channel which produces the output

$$\mathcal{N}(\mathcal{E}(m)) = \rho'_{x_1(m)} \otimes \rho'_{x_2(m)} \otimes ... \otimes \rho'_{x_N(m)}$$

The receiver performs a measurement thanks to a basis of $\mathbf{C}^{\otimes N}$. So this basis consists in a set of orthogonal projectors spanning the Hilbert space. We need M projectors in order to potentialy obtain the M messages and an extra one corresponding to outcomes not in the set of messages. So we take a measurement apparatus described by $\{Q_0, Q_1, ..., Q_M\}$ with

$$Q_0 + Q_1 + ... + Q_M = I$$

In summary the decoder $\mathcal{D}$ is a measurement apparatus: if the outcome of the measurement is in subspace Q_j, $j = 1, ..., M$ the decoder declares that the message sent was $\widehat{m} = j$; if the outcome is in subspace Q_0 the decoder declares an error $\widehat{m} = 0$.

The reliability of the $(\mathcal{E},\mathcal{D})$ scheme is measured by an average probability of error. Given that m is sent the probability of error is

$$P_{error}^{(m)} = Prob(\widehat{m} \neq m | m = input) = Tr(I - Q_m)\rho'_{x(1)} \otimes \rho'_{x(2)} \otimes ... \otimes \rho'_{x(N)}$$

Since the source produces messages uniformly at random the average probability of error is

$$P_{error} = \frac{1}{M}\sum_{m=1}^{M} P_{error}^{(m)}$$

Suppose that we have shown that there exist $(\mathcal{E},\mathcal{D})$ for which the average probability $P_{error} < \epsilon$. Then we can use the standard trick of "expurgation" to deduce that there exist $(\mathcal{E},\mathcal{D})$ with $P_{error}^{(m)} < 2\epsilon$ for all m in the new code book.

The following theorem is the analog of Shannon's famous channel coding theorem. In its present form it was proven separately by Holevo and Schumacher-Westmoreland.

Theorem 2 (**HSW theorem.**). *Fix ϵ and δ small positive. Set*

$$C_{1,\infty} = \max_{p_x,\rho_x}\big[S(\mathcal{N}(\sum_x p_x\rho_x)) - \sum_x p_x S(\mathcal{N}(\rho_x))\big]$$

(the Bloch sphere is compact so max *is attained).*

- **Achievability.** *Let $R < C_{1,\infty} - \epsilon$. Then there exists, for N and M large enough, a scheme $(\mathcal{E},\mathcal{D})$ such that $P_{error} \leq \delta$.*

- **Converse.** *Let $R > C_{1,\infty} + \epsilon$. Then for any scheme $(\mathcal{E},\mathcal{D})$ we have $P_{error} \geq \delta$*

Proof of the converse. The proof combines the Holevo bound with the classical Fano inequality. If we think of the classical case this means that in some sense the Holevo bound (which boils down to strong subadditivity) takes care at the same time of the data processing and subadditivity steps. The receiver has a mixture $\{\otimes_{i=1}^N \rho'_{x_i(m)}; \frac{1}{M}\}$ that is to be interpreted as a preparation of a quantum state according to the random variable $\mathcal{M}$ with $Prob(\mathcal{M} = m) = \frac{1}{M}$. The measurement outcome is labelled by a classical variable $\widehat{\mathcal{M}}$ with values in $\{0, 1, ..., M\}$. Recall $P_{error} = \frac{1}{M}\sum_{m=1}^M Prob(\widehat{m} \neq m)$.

$$h_2(P_{error}) + P_{error}\log_2(2^{NR} + 1 - 1) \geq H(\mathcal{M}|\widehat{\mathcal{M}})$$

where h_2 is the binary entropy function. So

$$P_{error} \geq \frac{1}{NR} H(\mathcal{M}|\widehat{\mathcal{M}}) - \frac{1}{N} h_2(P_{error})$$

On the other hand by Holevo's bound and $H(\mathcal{M}) = \log_2 M = NR$,

$$H(\mathcal{M}|\widehat{\mathcal{M}}) = H(\mathcal{M}) - I(\mathcal{M}|\widehat{\mathcal{M}}) \geq NR - \chi\Big(\frac{1}{M}; \sum_{m=1}^{M} \frac{1}{M} \otimes_{i=1}^{N} \rho'_{x_i(m)}\Big)$$

Now from the subadditivity of von Neumann's entropy,

$$S\big(\frac{1}{M}\sum_{m=1}^{M} \otimes_{i=1}^{N} \rho'_{x_i(m)}\big) \leq \sum_{i=1}^{N} S\big(\frac{1}{M}\sum_{m=1}^{M} \rho'_{x_i(m)}\big)$$

(on the right hand side we took all N partial traces leaving out only the 1 particle subsystems intact). Therefore using (additivity of entropy for independent subsystems)

$$S\big(\otimes_{i=1}^{N} \rho'_{x_i(m)}\big) = \sum_{i=1}^{N} S(\rho'_{x_i(m)})$$

we get

$$\chi\big(\frac{1}{M}; \rho'^{\otimes N}\big) \leq \sum_{i=1}^{N} \Big(S\big(\frac{1}{M}\sum_{m=1}^{M} \rho'_{x_i(m)}\big) - \frac{1}{M}\sum_{m=1}^{M} S(\rho'_{x_i(m)}) \Big)$$

We recognize on the r.h.s a sum of Holevo quantities

$$\chi\big(\frac{1}{M}; \mathcal{N}(\sum_{m=1}^{M} \frac{1}{M} \rho_{x_i(m)})\big)$$

which can be upper bounded by $N \max_{p_x,\rho} \chi(X; \mathcal{N}(\rho))$. Thus we conclude that if $R > C_{1,\infty} + \epsilon$

$$P_{error} \geq \frac{R - C_{1,\infty}}{R} - \frac{1}{N} h_2(P_{error})$$

Thus if $R > C_{1,\infty} + \epsilon$ for N large enough we have $P_{error} \geq \delta$.

Sketch of proof for the achievability. The details are more technical than in the classical case so we just give the main ideas. As in the classical case one uses a random code ensemble $\mathcal{C}$ in order to prove that $\mathbf{E}_{\mathcal{C}}[P_{error}(C)] \leq \delta$.

Since there must exist at least one code with error probability below the average, we conclude that there exist one coding scheme with $P_{error}(C) \leq \delta$. To construct the ensemble $\mathcal{C}$ first fix a probability distribution p_x, and corresponding 2×2 density matrices ρ_x, $x \in \{1, ..., D\}$. For each message $m \in \{1, ..., M\}$ randomly pick an input block

$$\rho_{x_1(m)} \otimes ... \otimes \rho_{x_N(m)}$$

with probability

$$p_{x_1(m)} \cdots p_{x_N(m)}$$

Once this has been done for all messages one has a code $C \in \mathcal{C}$. One communicates with this given code C. The ensemble $\mathcal{C}$ of all such possible codes obtained by repeating the random experiment is the code book or code ensemble $\mathcal{C}$ (note that the cardinality of this ensemble is D^N). The sender picks a message $m \in \{1, .., M\}$ and a code $C \in \mathcal{C}$ and sends the codeword through the channel. The receiver gets

$$\rho'_{x_1(m)} \otimes ... \otimes \rho'_{x_N(m)} = \sigma_m$$

In the classical case the decoding rule is based on jointly typical sequences of inputs and output typical. However as already stressed "X and Y do not exist at the same time" so here the situation will be more involved. Below we introduces two kind of typical subspaces of the Hilbert space.

Each ρ'_x has a spectral decomposition with eigenvalues λ_k and eigenprojectors P_k (here $k = 1, 2$). Thus the received word above has eigenvalues of the form

$$\lambda_{k_1}^{(m)} \lambda_{k_2}^{(m)} ... \lambda_{k_N}^{(m)}$$

with eigenprojectors

$$P_{k_1}^{(m)} \otimes ... \otimes P_{k_N}^{(m)}$$

We define the projector

$$P_{typ}^{(m)} = \sum_{typical\, e.v\, sequence} P_{k_1}^{(m)} \otimes ... \otimes P_{k_N}^{(m)}$$

where the sum is over typical sequences of eigenvalues, defined as those which have empirical entropy close to the average entropy of received words

$$\left\{ \lambda_{k_1}^{(m)}, ..., \lambda_{k_N}^{(m)} \mid \left| -\frac{1}{N} \sum_{i=1}^{N} \log_2 \lambda_{k_i}^{(m)} - \sum_x p_x S(\rho'_x) \right| \leq \epsilon \right\}$$

This projector is the one that projects on the space of typical output states (words) given an input message m. The law of large numbers implies that

$$\mathbf{E}_{\mathcal{C}}[Tr\sigma_m P_{typ}^m] \geq 1-\delta, \qquad \mathbf{E}_{\mathcal{C}}[TrP_{typ}^m] = 2^{N(\sum_x p_x S(\rho'_x)\pm\epsilon)} \tag{8.4}$$

Let us show here the first inequality: the trace is equal to

$$\begin{aligned} &\mathbf{E}_{\mathcal{C}} \sum_{typical\, e.v\, sequence} \lambda_{k_1}^{(m)}\lambda_{k_2}^{(m)}...\lambda_{k_N}^{(m)} = \mathbf{E}_{\mathcal{C},\lambda}[1_{typical\, e.v\, sequence}] \\ &= Prob_{\mathcal{C},\lambda}\Big[\big|-\frac{1}{N}\sum_{i=1}^{N}\log_2\lambda_{k_i}^{(m)} - \sum_x p_x S(\rho'_x)\big| \leq \epsilon\Big] \\ &\geq 1-\delta \end{aligned}$$

where the last inequality is the law of large numbers in the probability space $(\mathcal{C},\lambda)$. Consider now the mixture (of mixed states)

$$\sigma = \sum_{x_1(m)} p_{x_N(m)}p_{x_1(m)}...p_{x_N(m)}\rho_{x_1(m)}\otimes...\otimes\rho_{x_N(m)} = \otimes_{i=1}^N\big(\sum_x p_x\rho'_x\big)$$

The eigenvalues and eigenprojectors are of the form $\mu_{k_1}...\mu_{k_N}$ and $P_{k_1}\otimes...\otimes P_{k_N}$. The typical eigenvalue sequences are those in the ensemble

$$\Big\{\mu_1,..,\mu_N \mid \big|-\frac{1}{N}\sum_{i=1}^{N}\log_2\mu_{k_i} - S(\rho')\big| < \epsilon\Big\}$$

where $\rho' = \sum_x p_x\rho_x$. We set P_{typ} for the projector on the span of these typical eigenvalue sequences. As usual the law of large numbers implies

$$Tr\otimes_{i=1}^N\big(\sum_x p_x\rho'_x\big)P_{typ} \geq 1-\delta, \qquad TrP_{typ} = 2^{NS(\rho')\pm\epsilon)} \tag{8.5}$$

The decoding operation is a generalized measurement described by a "positive operator valued measure" (POVM). This notion is a convenient generalization of von Neumann's ordinary projective measurements. A POVM is a resolution of the identity in terms of positive operators. In fact it can be shown to be equivalent at the expense of purifying and making unitary transformations. Set

$$Q_m = \big(\sum_m P_{typ}P_{typ}^{(m)}P_{typ}\big)^{-1/2}P_{typ}P_{typ}^{(m)}P_{typ}\big(\sum_m P_{typ}P_{typ}^{(m)}P_{typ}\big)^{-1/2}$$

and

$$Q_0 = I - \sum_m Q_m$$

where the operators in the inverse square root are restricted over the complement of their kernel (so that the inverse exists). Let us check that $\{Q_0, Q_1, ..., Q_M\}$ is a POVM. Evidently they are selfadjoint and for all $m = 1, ..., M$ they are positive. Moreover

$$\sum_m Q_m = I_{kernel} < I$$

thus the extra Q_0 is also positive. Thus we have a POVM.

The decoding rule can now be stated. An apparatus described by a POVM works in the same way than in the usual measurement postulate:

$$\sigma_m \to Q_{m'}\sigma_m Q_{m'} \leftrightarrow \text{output}\, m' \,\text{with probability}\; Tr\sigma_m Q_{m'}$$

The probability of error is

$$P_{error} = \frac{1}{M}\sum_{m=1}^{M}\left(Tr\sigma_m Q_0 + \sum_{m'\neq m} Tr\sigma_m Q_{m'}\right)$$

The fist term in the parenthesis corresponds to a "decoder failure" and the second term corresponds to a "decoding error". As in the classical case one proves

$$\mathbf{E}_{\mathcal{C}}[P_{error}] \leq 4\delta + (M-1)2^{-N\left(S(\rho') - \sum_x p_x S(\rho'_x) - 2\epsilon\right)}$$

starting from (8.4) and (8.5). As is often the case for quantum calculations the estimates are a bit technical due to the non-commutative nature of the algebra. The reader is refered to pages 559 and 560 of [Nielsen and Chuang] or to the original papers of Holevo or Schumacher-Westmoreland for the details leading to this estimate.

When the rate $\frac{1}{N}\log_2 M < S(\rho') - \sum_x p_x S(\rho'_x) - 2\epsilon$ the average error probability can be made as small as we wish for N large enough. Thus this coding-decoding scheme achieves reliable transmission. the best rate attainable is found by taking a scheme where p_x and ρ_x maximize $\chi(X; \mathcal{N}(\rho))$. End of proof.

8.5 The capacities of quantum gaussian channels.