The Quantum Capacity of a Quantum Channel

Francesco Battistel

March 31, 2017
TopMath-IMPRS Spring School "Quantum Entropy and its Use"

Contents

1 Quantum Communication

Alice would like to send quantum information to Bob. More precisely, she has a quantum system A and she would like to prepare any state on A and send it to Bob in such a way that Bob gets it perfectly undamaged. We require in particular that also all the correlations with other systems are preserved. Unfortunately they have only a noisy quantum channel $\mathcal{N}^{A\to B}$ to communicate between them. Such channel will cause some errors over the state originally prepared by Alice, so Bob needs to apply some other channel $\mathcal{R}^{B\to\hat{B}}$, which we'll call the *recovery* or *decoder*, to correct such errors, where $\hat{B}$ is some preferred system over which he wants to extract the decoded state.

If ρ is the state prepared by Alice, we would like to require that $\mathcal{R}(\mathcal{N}(\rho)) = \rho$ for every density matrix on A. Unfortunately, it is known that the only channels which are invertible over the full $\mathcal{H}_A$ are the unitary channels, hence, in the general case, Alice can hope to convey reliably to Bob only the states supported in some subspace $\mathcal{C}_A \subseteq \mathcal{H}_A$, which we call the *code subspace*.

Recall that qubits (two dimensional quantum systems) are the smallest amount of quantum information, like bits for classical information. The number of qubits encoded in the code subspace is given by $\log_2 \dim \mathcal{C}_A := k$ (it's not necessarily an integer but it's not a problem) and we call them the *logical qubits*, while $\log_2 \dim \mathcal{H}_A$ are the *physical qubits*.

Let's emphasize here what it means for Alice to be able to send to Bob a qubit of information (i.e. a logical qubit): let's say that Alice is able to send reliably to Bob

only a single state $|\psi^A\rangle$; this state is a state of many physical qubits, but if Bob receives always only this one, then he doesn't get any kind of information (it's like sending always the bit 0 classically!). Moreover, Alice may just send classically the procedure to realize $|\psi^A\rangle$ in Bob's laboratory. The point is that, when she's able to send 2^k basis states (so k logical qubits), then she can also send any density matrix supported on the code subspace, which will have some purification on an auxiliary or *reference system* R. If Alice controls R, then in this sense her ability to send pure states on A to Bob is the same as sending entanglement through the channel (see remark 1). This is important because when Alice and Bob share some (maximally) entangled state, then they can perform many interesting protocols like quantum teleportation.

1.1 Operational definition of quantum capacity

Roughly speaking, the quantum capacity is related to the largest dimension of the code subspace, however, it turns out that this question is already difficult enough and a little bit restrictive, so we'll consider an asymptotic setting in which Alice takes n copies of A and tries to find the best code subspace in A^n (the decoder acts over the whole A^n at the same time: here it is where we may find some better code than in the one-shot setting, together with the fact that we allow for states entangled between such copies), and while doing this we send $n \to \infty$. We also allow for some small error probability ϵ in the recovery for finite n, provided that the error approaches 0 in the limit.

More formally, we define

Definition 1 (Rate)**.** Given a code subspace $\mathcal{C}_{A^n}$ with $k^{(n)} \coloneqq \log_2 \dim \mathcal{C}_{A^n}$, the rate for such a code is defined as $k^{(n)}/n$. Impropriately we may say that it's the number of logical qubits sent per channel use, but notice that at this stage we are not saying that we have a good code, with Bob being really able to receive such qubits undamaged.

Recall that the fidelity F of two density operators is defined as $F(\rho,\sigma) = ||\sqrt{\rho}\sqrt{\sigma}||_1^2 = (\mathrm{Tr}\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}})^2$ and it's a measure of how much different they are. We have $0 \leq F \leq 1$ and $F = 1 \iff \rho = \sigma$. If $\sigma = |\sigma\rangle\langle\sigma|$ is pure it reduces to $F(\rho, |\sigma\rangle) \coloneqq F(\rho, |\sigma\rangle\langle\sigma|) = \langle\sigma|\rho|\sigma\rangle$. It will be generally useful to introduce a reference system R, hold by Alice, which purifies ρ^A. We denote the purification by $|\psi^{RA}\rangle$.

Definition 2 (Achievable rate)**.** We say that Q is an achievable rate for a channel $\mathcal{N}$ if $\forall \epsilon, \delta > 0$ there exist a sequence of codes with increasing n and recoveries such that for n sufficiently large the rate is at least $Q - \delta$ and

$$F\Big(\mathcal{R}^{B^n \to \hat{B}^n} \circ \mathcal{N}^{A^n \to B^n}\left(|\psi^{R^n A^n}\rangle\langle\psi^{R^n A^n}|\right), |\psi^{R^n A^n}\rangle\Big) \geq 1 - \epsilon. \tag{1}$$

for every state supported in the code subspace and purified by the reference system.

Remark 1. We'll still continue to refer to it simply as the fidelity, but this is actually what is called the *entanglement fidelity* [5]. Closeness to 1 in such measure is known to imply not only that the state is recovered, but that also the entanglement with the purifying system is preserved. We can understand this with an example: the Bell pairs $|00\rangle + |11\rangle$ and $|00\rangle - |11\rangle$ have both single-qubit reduced density matrices equal to $\mathbb{1}/2$, so if we are able to recover such reduced state, it doesn't necessarily mean that we have recovered the correct purified one (it's even worse than that, because the purifying system may be a completely different one from the original!). Moreover, this tells us that the system R is not just a mathematical convenience, but it's crucial in our definition and in the understanding of the protocol.

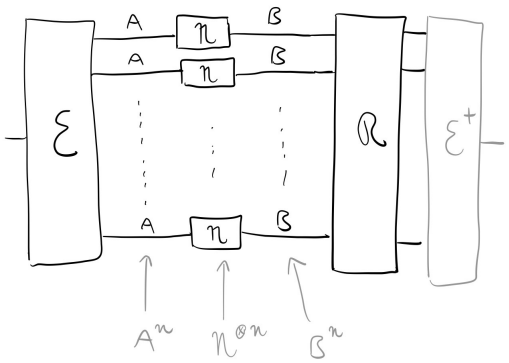

Figure 1: Alice encodes some state into n copies of A and uses n times the channel $\mathcal{N}$ to send them to Bob, who tries to perform a recovery on them. If recovery is successful he can apply $\mathcal{E}^\dagger$ to get the original state ($\mathcal{E}$ is invertible over the code subspace). We can just talk about (code) subspaces in A^n, generally spanned by entangled states, and associated recoveries. We remark that it's crucial that $\mathcal{R}$ acts collectively over B^n.

Definition 3 (Quantum capacity)**.** The quantum capacity $Q(\mathcal{N})$ of a channel $\mathcal{N}$ is defined as the supremum of all achievable rates.

We are assuming that Alice uses n times the channel $\mathcal{N}^{A\to B}$ (taken to be memoryless) to send A^n to B^n, sending one copy at a time, and we write $\mathcal{N}^{A^n\to B^n} = (\mathcal{N}^{A\to B})^{\otimes n}$ (see figure 1).

We may model the code subspace by saying that Alice has a system A' ($\dim \mathcal{H}_{A'} \leq \dim \mathcal{H}_A$) and she wants to send every state on it to Bob, so she uses an *encoder* $\mathcal{E}\colon \rho^{A'} \mapsto \rho^A \coloneqq V\rho^{A'}V^\dagger$, where V is an isometry $\mathcal{H}_{A'} \to \mathcal{H}_A$ ($V^\dagger V = \mathbb{1}$), in order to protect such information by making it non-locally redundant in the larger space. The code subspace corresponds to the image of $\mathcal{H}_{A'}$ under V. However, in the following we won't write down or draw the encoder any more, because the quantum capacity basically tells us what's the maximum dimension of the code subspace in an asymptotic limit, but it doesn't say anything about what's actually such subspace (so what's the encoder), nor about the decoder. Notice in particular that it depends only on the channel $\mathcal{N}$. In the following we'll derive a formula for the quantum capacity, but the proof will be non-constructive.

2 Coherent Information

We have already gotten rid of the encoder, and now we would like to get rid also of the recovery, in the sense that we want to find conditions guaranteeing the existence of such recovery, without the need to explicitly construct it. In this way we'll be able to simplify the proof a lot and in the meanwhile we'll explain the so-called decoupling principle, which underlies many important results in the study of quantum capacities and quantum error correction.

Recall that we have introduced a reference system R, hold by Alice, which purifies ρ^A, and we denote the purification by $|\psi^{RA}\rangle$. Consider the Stinespring dilations $U_{\mathcal{N}}^{A\to BE}$ of $\mathcal{N}^{A\to B}$ and $V_{\mathcal{R}}^{B\to \hat{B}E'}$ of $\mathcal{R}^{B\to\hat{B}}$.

Referring to figure 2, notice that the recovery works iff the final state $|\tilde{\psi}^{R\hat{B}EE'}\rangle$ has the form $|\psi^{R\hat{B}}\rangle \otimes |\chi^{EE'}\rangle$, where $|\chi^{EE'}\rangle$ is some residual junk and $|\psi^{R\hat{B}}\rangle$ is understood

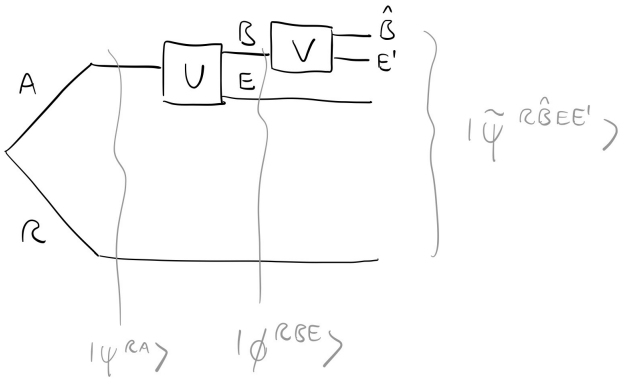

Figure 2: General setting consider here and in the following: we start from a purified code state $|\psi^{RA}\rangle$ which is transmitted across the noise channel and the recovery channel. The same picture holds if one substitutes each system with its n-th tensor product factor.

to be the same state as $|\psi^{RA}\rangle$, in the sense that, given some preferred choice of bases, the expansion of the state vector has the same coefficients in both cases.

Definition 4 (Coherent information)**.** In the above setting we define the coherent information as

$$I_c(R\rangle B)_\phi \coloneqq -H(R|B)_\phi \tag{2}$$

where ϕ is a short hand notation for $|\phi^{RBE}\rangle$.

Recalling that $|\phi^{RBE}\rangle$ is pure and recalling the definition of the conditional entropy, we see that

$$I_c(R\rangle B)_\phi = -\big(H(RB)_\phi - H(B)_\phi\big) = H(B)_\phi - H(E)_\phi \tag{3}$$

Notice that the quantities on the r.h.s. depend only on the marginals of $|\phi^{RBE}\rangle$ on B and E respectively, hence R is actually irrelevant for the computation of the coherent information. It can also be restated as $I_c(R\rangle B)_\phi = \frac{1}{2}(I(B;R)_\phi - I(E;R)_\phi)$, which says us that it's positive when the correlations of R with the system are stronger than with the environment (it's a quite quantum quantity because classically is always non-positive!).

Definition 5 (One-shot quantum capacity)**.**

$$Q_1(\mathcal{N}) \coloneqq \max_A I_c(R\rangle B)_\phi \tag{4}$$

where the maximum is taken over all density matrices on A.

The main result that we will discuss and prove in the following is:

Theorem 1 (Lloyd-Shor-Devetak or LSD Theorem)**.**

$$Q(\mathcal{N}) = \lim_n \frac{1}{n} \max_{A^n} I_c(R^n\rangle B^n)_{\phi^{R^n B^n E^n}} \tag{5}$$

$$= \lim_n \frac{1}{n} Q_1(\mathcal{N}^{\otimes n}) \tag{6}$$

Formulas where such a limit appear are called regularized. Before proving such a relation, we should say that such formula is useful only for some limited classes of channel, like degradable ones, whereas in general it's uncomputable. This happens because $Q_1(\mathcal{N})$ is not an additive quantity. Indeed, it's possible that $Q_1(\mathcal{N}^{\otimes 2}) > 2Q_1(\mathcal{N})$, which means that the knowledge of $Q_1(\mathcal{N})$ is not sufficient and we should compute $Q_1(\mathcal{N}^{\otimes n})$ for higher and higher n, which is basically hopeless. So in general we cannot get a single-letter formula for the asymptotic limit, contrarily to what happens in classical Shannon theory with the capacity of classical channels. Let's emphasize once more that we are talking about the asymptotic limit, so this doesn't tell us much about what's the best we can really achieve for any finite n, apart for the fact that we can't do better than this.

For the proof of theorem 1 we'll proceed as follows: we'll introduce the so-called decoupling principle, then we'll prove that the given formula is an upper bound on the quantum capacity, and finally we'll prove achievability for the coherent information, proving that random coding achieves approximate decoupling with high probability in the limit of large n.

We may first ask if such a limit ever exists. The answer is positive because $Q_1(\mathcal{N}^{\otimes n}) \leq n \log \dim \mathcal{H}_A$ (so the lim sup exists) and it can be proven that the lim inf is no smaller than the lim sup [1].

3 Decoupling Principle

We first need a monotonicity property for the coherent information, which will be crucial for the following results. Given channels $\mathcal{N}_1^{A\to B}$ and $\mathcal{N}_2^{B\to C}$, we know that the mutual information satisfies:

$$I(R;A) \geq I(R;B) \geq I(R;C) \tag{7}$$

from which it follows:

$$H(R) - H(R|A) \geq H(R) - H(R|B) \geq H(R) - H(R|C) \tag{8}$$

$$I_c(R\rangle A) \geq I_c(R\rangle B) \geq I_c(R\rangle C) \tag{9}$$

where the last inequality is called the *quantum data-processing inequality*. It states that a quantum channel acting on a system cannot increase the coherent information with the purifying reference system.

In the setting of figure 2 we define:

Definition 6 (Exact recoverability)**.** We say that Bob is able to exactly recover $|\psi^{RA}\rangle$ prepared by Alice iff the final state has the form

$$|\tilde{\psi}^{R\hat{B}EE'}\rangle = |\psi^{R\hat{B}}\rangle \otimes |\chi^{EE'}\rangle\,, \tag{10}$$

which is equivalent to the condition

$$F(\mathcal{R}^{B\to\hat{B}}(\phi^{RB}), |\psi^{RA}\rangle) = 1, \tag{11}$$

where $\phi^{RB} = \mathrm{Tr}_E\, |\phi^{RBE}\rangle\langle\phi^{RBE}|$ (in the following we use similar conventions to indicate reduced density matrices).

Definition 7 (Exact decoupling)**.** We say that there is exact decoupling between the environment and the reference system if:

$$\phi^{RE} = \phi^R \otimes \phi^E. \tag{12}$$

Theorem 2. *Exact decoupling* $\iff$ *exact recoverability.*

Proof. " $\Longleftarrow$ ": recalling that for pure states the von Neumann entropy is zero we get

$$I_c(R\rangle A)_\psi = -H(R|A) = -H(RA) + H(R) = H(R) \tag{13}$$

$$I_c(R\rangle \hat{B})_{\tilde{\psi}} = -H(R|\hat{B}) = -H(R\hat{B}) + H(R) = H(R) \tag{14}$$

(I omit the state over which these quantities are evaluated if it's clear. Remember that we do not do anything directly on system R, hence the marginal state is unchanged and so $H(R)$). We see that the coherent information of the initial and final states are the same, but then by the quantum data-processing inequality also in the intermediate state must have

$$H(R) = I_c(R\rangle B)_\phi \tag{15}$$

$$= H(B)_\phi - H(E)_\phi \tag{16}$$

$$= H(RE)_\phi - H(E)_\phi \tag{17}$$

$$\implies H(RE) = H(R) + H(E) \quad \implies \phi^{RE} = \phi^R \otimes \phi^E. \tag{18}$$

" $\Longrightarrow$ ": $|\psi^{RA}\rangle$ has some Schmidt decomposition $|\psi^{RA}\rangle = \sum \sqrt{\lambda_k}\, |k^R\rangle \otimes |k^A\rangle$. We diagonalize $\phi^R = \sum \lambda_k |k^R\rangle \langle k^R|$ (this form follows from the Schmidt decomposition and from the fact that the dynamics doesn't directly involve R) and $\phi^E = \sum \mu_l |l^E\rangle \langle l^E|$. Then if decoupling holds it means that (see also [4])

$$|\phi^{RBE}\rangle = \sum \sqrt{\lambda_k \mu_l}\, |k^R\rangle \otimes |\varphi^B_{kl}\rangle \otimes |l^E\rangle \tag{19}$$

where the $\{\varphi^B_{kl}\}$ are orthonormal. We construct a recovery as follows: we make a measurement with projectors $\Pi_l = \sum_k |\varphi^B_{kl}\rangle \langle \varphi^B_{kl}|$ and we apply an isometry V_l conditioned on the outcome such that $V_l |\varphi^B_{kl}\rangle = |k^{\hat{B}}\rangle$. Then the final state is

$$\Big(\sum_k \sqrt{\lambda_k}\, |k^R\rangle \otimes |k^{\hat{B}}\rangle\Big) \otimes |l^E\rangle = |\psi^{R\hat{B}}\rangle \otimes |l^E\rangle \tag{20}$$

where the last tensor product factor depends on the measurement outcome and it's just some residual junk. □

Definition 8 (Approximate recoverability)**.** We say that $|\psi^{RA}\rangle$ is ϵ-recoverable (or approximately recoverable generally speaking) if there exists a recovery such that

$$F(\mathcal{R}^{B\to\hat{B}}(\phi^{RB}), |\psi^{RA}\rangle) \geq 1 - \epsilon. \tag{21}$$

Definition 9 (Approximate decoupling)**.** We say that there's ϵ-decoupling between the environment and the reference if

$$||\phi^{RE} - \tilde{\phi}^R \otimes \tilde{\phi}^E||_1 \leq \epsilon, \tag{22}$$

where in general $\mathrm{Tr}_R\, \phi^{RE} \neq \tilde{\phi}^E$ and similarly tracing out E. Notice that when $\epsilon \to 0$ the only possibility left is $\phi^R \otimes \phi^E$, so $\tilde{\phi}^R \otimes \tilde{\phi}^E$ is not too much far from $\phi^R \otimes \phi^E$, because the trace norm is a distance.

Theorem 3. ϵ*-decoupling* $\implies$ ϵ*-recoverability.*

Remark 2. I'm not sure about what holds for the reverse implication. We don't need it anyway, while the above theorem will be crucial to prove achievability for the coherent information.

Proof. By Uhlmann's theorem, fixed $|\phi^{RBE}\rangle$ there exist a purification $|\tilde{\phi}^{RBE}\rangle$ of $\tilde{\phi}^R \otimes \tilde{\phi}^E$ such that:

$$F(\phi^{RE}, \tilde{\phi}^R \otimes \tilde{\phi}^E) = |\langle \phi^{RBE} | \tilde{\phi}^{RBE} \rangle|^2 \tag{23}$$

By theorem 2, $|\tilde{\phi}^{RBE}\rangle$ is exactly recoverable. Moreover, the fidelity is known to satisfy $F(\rho, \sigma) \geq (1 - ||\rho - \sigma||_1)^2$, the application of a quantum channel cannot decrease it and tracing out a subsystem can only make the states less distinguishable, hence

$$\begin{aligned}
F(\mathcal{R}^{B\to\hat{B}}(\phi^{RB}), |\psi^{RA}\rangle) &= F(\mathcal{R}^{B\to\hat{B}}(\phi^{RB}), \mathcal{R}^{B\to\hat{B}}(\tilde{\phi}^{RB})) && (24)\\
&\geq F(\phi^{RB}, \tilde{\phi}^{RB}) && (25)\\
&\geq |\langle \phi^{RBE} | \tilde{\phi}^{RBE} \rangle|^2 && (26)\\
&= F(\phi^{RE}, \tilde{\phi}^R \otimes \tilde{\phi}^E) && (27)\\
&\geq (1 - \frac{1}{2}||\tilde{\phi}^{RE} - \tilde{\phi}^R \otimes \tilde{\phi}^E||_1)^2 && (28)\\
&\geq 1 - \epsilon. && (29)
\end{aligned}$$

□

3.1 An upper bound on the quantum capacity

We now want to show that

$$Q(\mathcal{N}) \leq \lim_n \frac{1}{n} \max_{A^n} I_c(R^n\rangle B^n)_{\phi^{R^n B^n E^n}} \tag{30}$$

so we assume that $\bar{R}$ is an achievable rate and show that it's upper bounded by the above quantity. More specifically, we pick a sequence of codes and we assume that $\bar{R}$ is achievable for such sequence. We have said before that the ability of sending pure states through the channel is the same as the one of sending entanglement, in particular it's enough to be able to send maximal entanglement. We can alternatively think of it in the sense that once they share a maximally entangled state of the same dimension of the code subspace, then they can transmit reliably any state using quantum teleportation.

By definition of achievability for this choice of codes $\{\mathcal{C}_{A^n}\}$, $\forall \epsilon, \delta > 0$ for n sufficiently large the rate is at least $\bar{R} - \delta$ and Bob can recover with fidelity no smaller than $1 - \epsilon$. We pick a state $|\psi^{(n)}\rangle$ maximally entangled between R^n and a code subspace $\mathcal{C}_A^{(n)}$ of dimension $2^{n(\bar{R}-\delta)}$. Initially $I_c(R^n\rangle A^n)_{\psi^{(n)}} = H(R^n)_{\psi^{(n)}} = n(\bar{R} - \delta)$. Notice that, given such dimension for the code subspace, this is the maximum coherent information at the beginning and so also in the intermediate step by the quantum data-processing inequality.

We claim that if Bob is able to recover almost perfectly, then $I_c(R^n\rangle \hat{B}^n)_{\tilde{\psi}} = n(\bar{R}-\delta) - o(n)$, where linearity in n comes from the subadditivity of the von Neumann entropy. Suppose the contrary, i.e. that (the following can be made more rigorous using the quantum Fano inequality):

$$\begin{aligned}
H(R^n) - o(n) &\geq I_c(R^n\rangle \hat{B}^n)_{\tilde{\psi}} && (31)\\
&= H(\hat{B}^n) - H(R^n \hat{B}^n) && (32)\\
&\implies H(R^n \hat{B}^n) \geq H(\hat{B}^n) - H(R^n) + o(n) && (33)
\end{aligned}$$

If Bob is able to recover almost perfectly, then $\tilde{\psi}^{R^n \hat{B}^n}$ should be almost pure and hence $H(\hat{B}^n) \sim H(R^n)$, but then $H(R^n \hat{B}^n) \geq o(n)$, so that if the deviation is too large,

then there would be too much entanglement left between $R^n\hat{B}^n$ and the environment, hence Bob's recovery wouldn't be good enough, contrarily to our assumption.

By the quantum data processing inequality it then follows that:

$$I_c(R^n\rangle B^n)_\phi \geq I_c(R^n\rangle\hat{B}^n)_{\tilde{\psi}} = n(\bar{R}-\delta) - o(n) \tag{34}$$

$$\implies \bar{R} \leq \frac{1}{n} I_c(R^n\rangle B^n)_\phi + o(1) \tag{35}$$

where remember that the constant $o(1)$ can be taken arbitrarily small (we have also included δ inside it). The result holds for a generic sequence of codes, hence also for the best one. We can then maximize over all codes, which is equivalent to maximize over all possible input states on A^n, and finally taking the limit we get the upper bound (30).

4 Coherent information as an achievable rate

The purpose of this section is to prove that given some ρ^A with purification $|\psi^{RA}\rangle$, the coherent information $I_c(R\rangle B)_\phi$ is an achievable rate, i.e $Q(\mathcal{N}) \geq I_c(R\rangle B)_\phi$. Then it holds also when we replace $\mathcal{N}$ by $\mathcal{N}^{\otimes n}$, when we maximize over the input states and when we take the limit, but then we obtain exactly the upper bound (30) and theorem 1 is proved.

We won't go into the details of the proof but try to convey the most important ideas. The first one is that if we discard a random qubit from R, correlations with E will become weaker. The decoupling inequality shows that by discarding a sufficient amount of qubits the probability that we achieve approximate decoupling is close to 1. The second one is that in the limit of large n we can focus only on the typical subspaces of the code, so we discard qubits in these subspaces and we don't care too much about ensuring recoverability outside the typical subspaces. So then we apply theorem 3, which assures us that, given approximate decoupling for the codewords, there exists a sufficiently good recovery.

4.1 The decoupling inequality

Consider ϕ^{RE} and let's say that we want to discard a random qubit at a time from R and see what are then the correlations on average between what's left of R and the environment E.

In this case by discarding a random qubit, or a random quantum system with dimension $\dim R_1 \coloneqq |R_1|$, we mean that, chosen a random decomposition $R = R_1R_2$, we trace out R_1. A better way to model the random choice of R_1 is by picking a fixed decomposition $R = R_1R_2$ and by applying a random unitary W on R before tracing out R_1. The choice of a unitary is done according to the *Haar measure* over the group of $D \times D$ unitary matrices, which is the unique normalized left- and right-invariant measure over $U(D)$, i.e.

$$\mathbb{E}_W[\mathbb{1}] = 1 \tag{36}$$

$$\mathbb{E}_W[f(W)] = \mathbb{E}_W[f(W\tilde{W})] = \mathbb{E}_W[f(\tilde{W}W)] \tag{37}$$

for every fixed unitary $\tilde{W}$. Then, defined

$$\phi^{R_2E}(W) \coloneqq \mathrm{Tr}_{R_1}\Big((W\otimes\mathbb{1})\phi^{RE}(W^\dagger\otimes\mathbb{1})\Big) \tag{38}$$

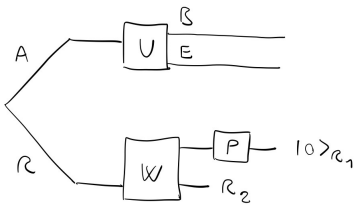

Figure 3: We apply a random unitary W over R and then we project some fixed subsystem R_1 onto some fixed state and we renormalise the outcome state.

we have a first version of the decoupling inequality [3]

$$\left(\mathbb{E}_W\left[\phi^{R_2E}(W) - \frac{\mathbb{1}}{|R_2|}\otimes\phi^E\right]\right)^2 \leq \frac{|R_2||E|}{|R_1|}\operatorname{Tr}\left[(\phi^{RE})^2\right] \tag{39}$$

Notice that ϕ^{RE} and ϕ^B have the same eigenvalues because $|\phi^{RBE}\rangle$ is pure, hence we may also write $\operatorname{Tr}\left[(\phi^{RE})^2\right] = \operatorname{Tr}\left[(\phi^B)^2\right]$.

It's interesting to see how it looks simply when there's no environment at all and the initial state is pure:

$$\left(\mathbb{E}_W\left[\phi^{R_2}(W) - \frac{\mathbb{1}}{|R_2|}\right]\right)^2 \leq \frac{|R_2|}{|R_1|}. \tag{40}$$

We see that if $|R_2|/|R_1| \ll 1$, then the reduced density matrix over what's left is almost completely mixed with high probability. Notice that $|R_i|$ is not the number of qubits but the dimension of the Hilbert space in which they live, so for example if we start with 1000 qubits and we trace out just a little bit more than half, let's say 525, we have $|R_2|/|R_1| = 2^{-50}$, which is already quite small!

We can devise a similar procedure in which instead of tracing out R_1 we project R_1 onto some fixed vector, let's call it $|0\rangle_{R_1}$ (so we can say that we are effectively projecting onto R_2), and we renormalise the state afterwards. Notice that tracing out doesn't really change the properties and the correlations in our system, but we just switch to a more local perspective, whereas projecting actually changes the state. The proof is basically the same and the final result is just modified by a factor of $|R_1|$ (see [3, 2]). In the following we'll use the decoupling inequality in this setting (see also figure 3), which reads:

Theorem 4 (Decoupling inequality)**.**

$$\left(\mathbb{E}_W\left[\phi^{R_2E}(W') - \frac{\mathbb{1}}{|R_2|}\otimes\phi^E\right]\right)^2 \leq |R_2||E|\operatorname{Tr}\left[(\phi^B)^2\right] \tag{41}$$

where $W' = PW$, P is the projector over R_2 and ϕ^{R_2E} is understood to be renormalized after this action.

4.2 Typical subspaces

Let's diagonalise $\rho^A = \sum p(x)\,|x\rangle\langle x|^A$, then $(\rho^A)^{\otimes n} = \sum p(\bar{x})\,|\bar{x}\rangle\langle\bar{x}|^{A^n}$, where $p(\bar{x}) = p(x^1)\,...\,p(x^n)$ and $|\bar{x}\rangle^{A^n} = |x^1\rangle\otimes\,...\,\otimes|x^n\rangle$. So, we prepare multiple copies of ρ^A and we would like to make sure that Bob can recover ρ^A from the output of the channel that he reads. In particular we want to show that $I_c(R\rangle B)_\phi$ for the input state ρ^A

is an achievable rate, i.e. we want to show the existence of a suitable sequence of codes. Notice that we refer in the following to the case in which we want to transmit a maximally entangled state between the reference R_2 and some code subspace (so the reduced density matrix over R_2 is the maximally mixed one $\mathbb{1}/|R_2|$), because as we have already said it's enough to ensure transmission of maximal entanglement.

Definition 10 (δ-typicality)**.** We say that $A_\delta \subseteq A^n$ is a *δ-typical subspace* if:

$$A_\delta \coloneqq \mathrm{span}\Big\{ |\bar{x}\rangle^{A^n} : \Big| -\frac{1}{n} p(\bar{x}) - H(\rho^A) \Big| \leq \delta \Big\} \tag{42}$$

We define also the *δ-typical projector* Π_δ^A projecting A^n onto A_δ.

We apply now the framework of the decoupling inequality projecting first $\{R^n, B^n, E^n\}$ onto δ-typical subspaces $\{R_\delta, B_\delta, E_\delta\}$, then averaging with respect to random unitaries supported only inside such subspaces and finally projecting out some fixed subsystem $R_1 \subseteq R_\delta$, $R_\delta = R_1 R_2$, with $|R_2| = 2^{nQ}$.

The proof uses the following lemma [2]:

Theorem 5. *Given $|\phi\rangle^{RBE}$, $\forall \delta > 0$ and sufficiently large n there exist δ-typical projectors $\Pi_\delta^{\{R,B,E\}}$ onto δ-typical subspaces $R_\delta \subseteq R^n$, $B_\delta \subseteq B^n$, $E_\delta \subseteq E^n$ such that*

$$|\phi_\delta^{R^n B^n E^n}\rangle \coloneqq (\Pi_\delta^R \otimes \Pi_\delta^B \otimes \Pi_\delta^E) \, |\phi^{R^n B^n E^n}\rangle \tag{43}$$

satisfies

$$|E_\delta| \leq 2^{nH(E)_\phi + n\delta} \tag{44}$$

$$\mathrm{Tr}\big[(\phi_\delta^{B_\delta})^2\big] \leq 2^{-nH(B)_\phi + n\delta} \tag{45}$$

$$||\phi^{R^n B^n E^n} - \phi_\delta^{R^n B^n E^n}||_1 \leq \epsilon \tag{46}$$

where $\epsilon = 2^{-nc\delta^2}$ for some c independent of n and δ.

So for n large the state is mostly supported inside δ-typical subspaces. Then proceeding as in the previous section one gets

$$\mathbb{E}_W \Big(||\phi_\delta^{R_2 E^n}(W') - \frac{\mathbb{1}}{|R_2|} \otimes \phi_\delta^{E^n}||_1 \Big) \leq \sqrt{|R_2||E_\delta| \, \mathrm{Tr}\big[(\phi_\delta^{B_\delta})^2\big]} \tag{47}$$

$$\leq \sqrt{2^{n(Q - H(B) + H(E) + 3\delta}} \tag{48}$$

$$\leq 2^{-n\delta} \leq \epsilon \tag{49}$$

if $0 \leq Q < H(B) - H(E) - 3\delta = I_c(R\rangle B)_\phi - 3\delta$. From the previous observation it finally follows that

$$\mathbb{E}_W \Big(||\phi^{R_2 E^n}(W') - \frac{\mathbb{1}}{|R_2|} \otimes \phi^{E^n}||_1 \Big) \leq 6\epsilon. \tag{50}$$

If the average value approaches decoupling, then there must exist a choice of codes approaching decoupling. Therefore, by theorem 3 there exist a good recovery for such a sequence and $I_c(R\rangle B)_\phi$ is an achievable rate.

5 Additivity Issues and Superactivation

We come now to the issue of additivity for the entropic quantities we have defined. There are two different questions that one can address [6]. The first one is to ask if quantities appearing in regularized formulas are additive: this would allow one to remove the regularization and the computation of capacities would be much easier. The second one refers to additivity of capacities of *different* channels. Notice that after regularization, capacities for parallel uses of the same channel *are* additive, i.e. $Q(\mathcal{N}^{\otimes m}) = mQ(\mathcal{N})$, whereas it turns out that it can happen that $Q(\mathcal{N} \otimes \mathcal{M}) > Q(\mathcal{N}) + Q(\mathcal{M})$.

Degradable Channels. We start showing that there exist classes of channels for which Q_1 is (sub)additive even for different channels in the same class, so in particular the regularization is superfluous.

Definition 11 (Degradable Channels)**.** Recall that given a channel $\mathcal{N}^{A\to B}$ with dilation $U_{\mathcal{N}}^{A\to BE}$, we can define the complementary channel $\widehat{\mathcal{N}}^{A\to E}$ by tracing out B instead of E. We say that $\mathcal{N}^{A\to B}$ is degradable if there exists a channel $\mathcal{T}^{B\to E}$ such that

$$\widehat{\mathcal{N}}^{A\to E} = \mathcal{T}^{B\to E} \circ \mathcal{N}^{A\to B}. \tag{51}$$

Intuitively it means that Eve gets less information about A than Bob, because of the monotonicity properties of the different distance measures for states under the action of channels, like the fact that in terms of fidelity two states can only be made less distinguishable by the action of a channel.

If Alice and Bob use two degradable channels $\mathcal{N}_1 \otimes \mathcal{N}_2$ in parallel, the key point is that there is a product channel $\mathcal{T}_1^{B_1\to E_1} \otimes \mathcal{T}_2^{B_2\to E_2}$ mapping B_1B_2 to E_1E_2. Then

$$I(B_1; B_2) \geq I(B_1; E_2) \geq I(E_1; E_2) \tag{52}$$

because we can use monotonicity of mutual information in two separate steps. With this one can show [3] that $Q_1(\mathcal{N}_1 \otimes \mathcal{N}_2) \leq Q_1(\mathcal{N}_1) + Q_1(\mathcal{N}_2)$, so that $Q(\mathcal{N}) = Q_1(\mathcal{N})$ for degradable channels.

Superactivation: superactivation refers to the surprising fact that not only it can happen that $Q(\mathcal{N} \otimes \mathcal{M}) > Q(\mathcal{N}) + Q(\mathcal{M})$, but that this can happen even if the two considered channels have both 0 capacity! It has been proven [7] that there exist an antidegradable channel and a Horodecki (PPT) channel (classes with 0 capacity) that combined together give precisely such a result.

This effect is quite important, because it tells us that in general the quantum capacity by itself is not a complete characterization of the ability of a channel to transmit quantum information, but it depends also on the other channels that are used in combination with it. This poses a big problem if one wants to compute capacities, because one has to consider combinations of different more and more channels, but from a practical point of view is a big opportunity: if we are smart enough in using entanglement between the inputs of different channels, we can transform apparently useless resources into something valuable.

References

[1] Howard Barnum, M. A. Nielsen, and Benjamin Schumacher. Information transmission through a noisy quantum channel. *Phys. Rev.*, A57:4153, June 1998.

[2] P. Hayden, M. Horodecki, A. Winter, and J. Yard. A decoupling approach to the quantum capacity. *eprint arXiv:quant-ph/0702005*, February 2007.

[3] John Preskill. Quantum Shannon Theory. *http://www.theory.caltech.edu/people/preskill/ph229/*, April 2016.

[4] B. Schumacher and M. D. Westmoreland. Approximate quantum error correction. *eprint arXiv:quant-ph/0112106*, December 2001.

[5] Benjamin Schumacher. Sending entanglement through noisy quantum channels. 54:2614–2628, October 1996.

[6] G. Smith. Quantum Channel Capacities. *ArXiv e-prints*, 110(4):040501, July 2010.

[7] G. Smith and J. Yard. Quantum Communication with Zero-Capacity Channels. *Science*, 321:1812, September 2008.