# 1 Classical communication

In this first part we will review the basic concept of a quantum channel (i.e. a dynamical map on a quantum system) in its most general form, and recall several of its properties. Then we will start with the topic of sending classical information via a quantum channel, which in later parts we will refine to quantum and private communication.

## 1.1 Quantum channels

**Definition 1.1** *A quantum channel, aka quantum map, aka physical map, between two quantum systems with Hilbert spaces $A$ and $B$ is a linear, completely positive and trace preserving (cptp) map $\mathcal{N} : \mathcal{L}(A) \to \mathcal{L}(B)$, often written a bit lax as $\mathcal{N} : A \to B$. The set of all channels from $A$ to $B$ is denoted as* $\mathrm{CPTP}(A \to B)$.

Recall that a linear map $\mathcal{N}$ is called trace preserving if $\operatorname{Tr} \mathcal{N}(\rho) = \operatorname{Tr} \rho$ for all matrices $\rho$, and positive (or more precisely, positivity-preserving) if $\rho \geq 0$ implies $\mathcal{N}(\rho) \geq 0$ for all $\rho$. It is completely positive if $\mathrm{id}_C \otimes \mathcal{N}$ is positive for all auxiliary systems $C$.

**Example 1.2** *We discuss several important special cases of this definition, where cptp-ness can be checked by hand.*

1. *The identity map* $\mathrm{id}_A : A \to A$ *(aka ideal channel) is cptp.*
2. *More generally, for an isometry $V : A \to B$ between Hilbert spaces, the map $\mathcal{V}(\rho) = V\rho V^\dagger$ is cptp.*
3. *For $A = B \otimes E$, the partial trace map* $\operatorname{Tr}_E : \mathcal{L}(A) \to \mathcal{L}(B)$ *is cptp.*
4. *Compositions and tensor products of cptp maps are cptp: if $\mathcal{N} : A \to B$ and $\mathcal{M} : B \to C$ are cptp, then so is $\mathcal{M} \circ \mathcal{N} : A \to C$, and if $\mathcal{N}_i : A_i \to B_i$ are cptp for $i = 1, 2$, then so is $\mathcal{N}_1 \otimes \mathcal{N}_2 : A_1 \otimes A_2 \to B_1 \otimes B_2$.*
5. *Convex combinations of cptp maps are cptp: if $\mathcal{N}_i : A \to B$ are cptp for $i \in \mathcal{I}$, and $p_i \geq 0$, $\sum_{i\in\mathcal{I}} p_i = 1$ is a probability vector, then $\sum_{i\in\mathcal{I}} p_i \mathcal{N}_i$ is cptp.*

The next lemma, providing a number of exhaustive characterisations of cptp maps, shows that the above collection of examples is already essential complete.

**Lemma 1.3** *For a linear map $\mathcal{N} : \mathcal{L}(A) \to \mathcal{L}(B)$, the following are equivalent:*

1. *$\mathcal{N}$ is cptp;*
2. *$\mathcal{N}$ is trace preserving and* $\mathrm{id}_{A'} \otimes \mathcal{N}$ *is positive for a Hilbert space $A' \simeq A$;*
3. *$J(\mathcal{N})^{A'B} := (\mathrm{id}_{A'} \otimes \mathcal{N})\Phi^{A'A}$ (called* Choi-Jamiołkowski matrix *or simply* Choi matrix*) is positive semidefinite, $J(\mathcal{N}) \geq 0$, and satisfies $\operatorname{Tr}_B J(\mathcal{N}) = \frac{1}{|A|}\mathbb{1}$;*
4. *there exists an isometry $V : A \to B \otimes E$ (called* Stinespring isometry *or* Stinespring dilation of $\mathcal{N}$*), with a suitable Hilbert space $E$, such that $\mathcal{N}(\rho) = \operatorname{Tr}_E V\rho V^\dagger$;*
5. *there exist operators $K_\alpha : A \to B$ (called* Kraus operators*) such that $\sum_\alpha K_\alpha^\dagger K_\alpha = \mathbb{1}_A$ and $\mathcal{N}(\rho) = \sum_\alpha K_\alpha \rho K_\alpha^\dagger$.*

**Proof.** *1* $\Rightarrow$ *2* is trivial.

*2* $\Rightarrow$ *3* is trivial since $\Phi^{A'A}$ is a state; note that in particular $\operatorname{Tr}_B J(\mathcal{N}) = \operatorname{Tr}_B(\mathrm{id}_{A'} \otimes \mathcal{N})\Phi^{A'A} = \operatorname{Tr}_A \Phi^{A'A} = \frac{1}{|A|}\mathbb{1}$ since $\mathcal{N}$ is trace preserving, i.e. $\operatorname{Tr}_B \circ \mathcal{N} = \operatorname{Tr}_A$.

*3* $\Rightarrow$ *4*. We show first that $\mathcal{N}(\rho) = |A| \operatorname{Tr}_{A'} \left( \left(\rho^\top \otimes \mathbb{1}_B\right) J(\mathcal{N})^{A'B} \right)$. Indeed, for the maximally entangled state $\Phi^{A'A}$, it holds that

$$\left(\rho^\top \otimes \mathbb{1}_A\right) |\Phi\rangle^{A'A} = \left(\mathbb{1}_{A'} \otimes \rho\right) |\Phi\rangle^{A'A},$$

hence $\rho = |A| \operatorname{Tr}_{A'} \left( \left(\rho^\top \otimes \mathbb{1}_A\right) \Phi^{A'A} \right)$. Thus,

$$\begin{aligned} \mathcal{N}(\rho) &= |A| \mathcal{N} \left( \operatorname{Tr}_{A'} \left( \left(\rho^\top \otimes \mathbb{1}_A\right) \Phi^{A'A} \right) \right) \\ &= |A| \operatorname{Tr}_{A'} (\mathrm{id}_{A'} \otimes \mathcal{N}) \left( \left(\rho^\top \otimes \mathbb{1}_A\right) \Phi^{A'A} \right) \\ &= |A| \operatorname{Tr}_{A'} \left( \left(\rho^\top \otimes \mathbb{1}_A\right) (\mathrm{id}_{A'} \otimes \mathcal{N}) \Phi^{A'A} \right) \\ &= |A| \operatorname{Tr}_{A'} \left( \left(\rho^\top \otimes \mathbb{1}_B\right) J(\mathcal{N})^{A'B} \right). \end{aligned} \tag{1}$$

Now choose a purification $\psi^{A'BE}$ of $J(\mathcal{N})^{A'B}$, with a suitable system $E$, which always exists. Indeed, since $\psi^{A'} = \frac{1}{|A|}\mathbb{1}_{A'} = J(\mathcal{N})^{A'} = \Phi^{A'}$, there exists an isometry $V : A \to B \otimes E$ such that $|\psi\rangle^{A'BE} = (\mathbb{1}_{A'} \otimes V^{A \to BE})|\Phi\rangle^{A'A}$. Define the isometric channel $\mathcal{V}(\rho) = V\rho V^\dagger$, and observe that $\psi^{A'BE} = (\mathrm{id}_{A'} \otimes \mathcal{V})\Phi^{A'A} = J(\mathcal{V})$. Thus we can continue the last line of the previous chain of identities in Eq. (1),

$$\begin{aligned} \mathcal{N}(\rho) &= |A| \operatorname{Tr}_{A'} \left( \left(\rho^\top \otimes \mathbb{1}_B\right) \operatorname{Tr}_E J(\mathcal{V})^{A'BE} \right) \\ &= \operatorname{Tr}_E \left[ |A| \operatorname{Tr}_{A'} \left( \left(\rho^\top \otimes \mathbb{1}_{BE}\right) J(\mathcal{V})^{A'BE} \right) \right] \\ &= \operatorname{Tr}_E \mathcal{V}(\rho) = \operatorname{Tr}_E V\rho V^\dagger, \end{aligned}$$

the last step by Eq. (1) for $\mathcal{V}$ in place of $\mathcal{N}$.

*4* $\Rightarrow$ *5*. Choose an orthonormal basis $\{|\alpha\rangle\}$ of $E$, and write $V = \sum_\alpha K_\alpha \otimes |\alpha\rangle$ with suitable operators $K_\alpha : A \to B$. Then,

$$\begin{aligned} \mathcal{N}(\rho) &= \operatorname{Tr}_E V\rho V^\dagger \\ &= \operatorname{Tr}_E \left( \sum_{\alpha,\beta} K_\alpha \rho K_\beta^\dagger \otimes |\alpha\rangle\langle\beta|^E \right) = \sum_\alpha K_\alpha \rho K_\alpha^\dagger. \end{aligned}$$

. Furthermore, since $V$ is an isometry,

$$\begin{aligned} \mathbb{1}_A &= V^\dagger V \\ &= \sum_{\alpha,\beta} K_\alpha^\dagger K_\beta \langle\alpha|\beta\rangle = \sum_\alpha K_\alpha^\dagger K_\alpha. \end{aligned}$$

.

*5 ⇒ 1.* It is quite evident that all maps $\rho \mapsto K\rho K^\dagger$ are positive and indeed completely positive, and so is the sum of such maps. Furthermore,

$$\operatorname{Tr}\mathcal{N}(\rho) = \sum_\alpha \operatorname{Tr} K_\alpha \rho K_\alpha^\dagger = \sum_\alpha \operatorname{Tr} K_\alpha^\dagger K_\alpha \rho = \operatorname{Tr}\left(\sum_\alpha K_\alpha^\dagger K_\alpha\right)\rho = \operatorname{Tr}\rho,$$

so $\mathcal{N}$ is cptp. □

**Example 1.4** *Here we present a few concrete channels that are of importance in theory and practice.*

1. *Identity channel* $\mathrm{id}_A : A \to A$.
2. *Constant channel (aka state preparation channel)* $\mathcal{P}_\sigma : A \to B$, *with a state* $\sigma$ *on* $B$, *acting as* $\mathcal{P}_\sigma(\rho) = \sigma \operatorname{Tr}\rho$.
3. *Qubit depolarizing channel* $\mathcal{D}_p(\rho) = (1-p)\rho + \frac{p}{3}(X\rho X^\dagger + Y\rho Y^\dagger + Z\rho Z^\dagger)$ *for* $0 \le p \le 1$, *where* $X$, $Y$ *and* $Z$ *are the Pauli unitaries:*

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

4. *Qubit dephasing channel* $\mathcal{Z}_p(\rho) = (1-p)\rho + pZ\rho Z^\dagger$ *for* $0 \le p \le 1$.
5. *Qubit Pauli error channels generalise the previous two examples: for a probability vector* $\vec{p} = (p_0, p_x, p_y, p_z)$, *let* $\mathcal{N}_{\vec{p}}(\rho) = p_0\rho + p_x X\rho X^\dagger + p_y Y\rho Y^\dagger + p_z Z\rho Z^\dagger$.
6. *Qubit amplitude damping channel* $\mathcal{A}_\delta(\rho) = A_0\rho A_0^\dagger + A_1\rho A_1^\dagger$, *where*

$$A_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\delta} \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & \sqrt{\delta} \\ 0 & 0 \end{bmatrix}.$$

7. *Erasure channel for an arbitrary input system* $A$: *let* $B := A \oplus \mathbb{C}|\perp\rangle$ *be an extension of* $A$ *by a state vector* $|\perp\rangle$ *orthogonal to all of* $A$, *and define* $\mathcal{E}_q(\rho) = (1-q)\rho + q|\perp\rangle\langle\perp|$.

There are two fundamental operations in quantum mechanics that stand out, one being state preparation, which is in principle governed by the constant channels $\mathcal{P}_\sigma$, and measurements, given by POVMs $(M_y : y \in \mathcal{Y})$. The former can be generalised to a a class of quantum channels with classical inputs, the latter modelled as quantum channels with classical output.

**Definition 1.5** *A* classical-quantum channel (cq-channel) *is a cptp map* $\mathcal{N} : X \to B$ *of the form*

$$\mathcal{N}(\xi) = \sum_x \langle x|\xi|x\rangle \rho_x,$$

*for an orthonormal basis* $\{|x\rangle\}$ *of* $X$ *and states* $\rho_x \in \mathcal{S}(B)$. *This channel allows for the preparation of* $\rho_x^B$ *at the output, by choosing input* $|x\rangle\langle x|$, *and the mixtures of such states by choosing mixtures or superpositions of these basis states.*

*A* quantum-classical channel (qc-channel) *is a cptp map $\mathcal{M} : A \to Y$ of the form*

$$\mathcal{N}(\alpha) = \sum_y |y\rangle\langle y| \operatorname{Tr} \alpha M_y,$$

*for an orthonormal basis $\{|y\rangle\}$ of $Y$ and a POVM $(M_y : y \in \mathcal{Y})$ on $A$. This channel implements the latter measurement and records the outcome in one of suitable orthogonal states of $Y$.*

Lecture 1 (12/12/2022)

Channels as we defined them, map states to states. Thanks to the Born rule, which for a state $\rho$ and a measurement POVM $(M_i : i \in \mathcal{I})$ states that

$$\Pr(i|\rho) = \operatorname{Tr} \rho M_i,$$

we can give an equivalent description in terms of a map taking observables to observables, the so-called adjoint cp map:

**Definition 1.6** *For a linear map $\mathcal{T} : \mathcal{L}(A) \to \mathcal{L}(B)$, define $\mathcal{T}^* : \mathcal{L}(B) \to \mathcal{L}(A)$ as the unique linear map with the property that for all $\xi \in \mathcal{L}(A)$ and $Y \in \mathcal{L}(B)$*

$$\operatorname{Tr} \mathcal{T}(\xi)^\dagger Y = \operatorname{Tr} \xi^\dagger \mathcal{T}^*(Y).$$

The definition is such that the adjoint of the adjoint is the original map, $\mathcal{T}^{**} = \mathcal{T}$.

A cptp map $\mathcal{N}$ is in particular Hermitian-preserving, hence $\mathcal{N}(\xi)^\dagger = \mathcal{N}(\xi^\dagger)$, so we can omit the dagger in the above defining formula, and indeed may restrict to Hermitian $\xi$ and $Y$. It can easily be seen that the adjoint map $\mathcal{N}^*$ of a cptp map $\mathcal{N}$ is completely positive and unit preserving (cpup), i.e. $\mathcal{N}^*(\mathbb{1}_B) = \mathbb{1}_A$.

**Example 1.7** *We record the adjoint maps for some of the channels and normal forms discussed above.*

1. *The ideal channel is its own adjoint* $\mathrm{id}_A^* = \mathrm{id}_A$.
2. *For the cp map $\mathcal{K}(\rho) = K\rho K^\dagger$, with $K \in \mathcal{L}(A \to B)$, the cyclicity of the trace implies $\mathcal{K}^*(Y) = K^\dagger Y K$. Imporant special cases: $K$ unitary or isometry.*
3. *By summing the previous example, a channel $\mathcal{N}(\rho) = \sum_\alpha K_\alpha \rho K_\alpha^\dagger$ in Kraus form, has adjoint map $\mathcal{N}^*(Y) = \sum_\alpha K_\alpha^\dagger Y K_\alpha$.*
4. *For the partial trace map $\operatorname{Tr}_E : \mathcal{L}(B \otimes E) \to \mathcal{L}(B)$, $\operatorname{Tr}_E^*(Y) = Y \otimes \mathbb{1}_E$.*
5. *For two channels $\mathcal{N}$ and $\mathcal{M}$, $(\mathcal{M} \circ \mathcal{N})^* = \mathcal{N}^* \circ \mathcal{M}^*$, and for two channels $\mathcal{N}_1$ and $\mathcal{N}_2$, $(\mathcal{N}_1 \otimes \mathcal{N}_2)^* = \mathcal{N}_1^* \otimes \mathcal{N}_2^*$.*
6. *By combining items 2, 4 and 5, for a channel in Stinespring form, $\mathcal{N}(\rho) = \operatorname{Tr}_E V\rho V^\dagger$, the adjoint map has the form $\mathcal{N}^*(Y) = V^\dagger(Y \otimes \mathbb{1}_E)V$.*
7. *The adjoint map of a constant channel $\mathcal{P}_\sigma$ is $\mathcal{P}_\sigma^*(Y) = (\operatorname{Tr} \sigma Y)\mathbb{1}$.*

Lecture 1 (13/12/2021)

Lecture 1 (11/12/2021)

## 1.2 One-shot communication, plain and entanglement-assisted

A channel, as modelled by a cptp map $\mathcal{N} : A \to B$, represents a causal influence from $A$ to $B$. As such, it is natural that it can be used to communicate by choosing different input states depending on the message.

To formalise this, we make the following definition.

**Definition 1.8** *A* code *for $\mathcal{N} : A \to B$ is a pair $(E, D)$, consisting of a map $E : [K] \to \mathcal{S}(A)$ from the set $[K] = \{1, \dots, K\}$ of messages (any set of $K$ elements would do) into the states on $A$, and a measurement (POVM) $D = (D_m : m \in [K])$ on $B$. We call $E$ the encoder and $D$ the decoder of the message.*

*It has two crucial performance criteria: we call $R = \log K$ the (one-shot) rate of the code, and the (average) error probability is*

$$P_e = 1 - \frac{1}{K} \sum_{m=1}^{K} \operatorname{Tr} \mathcal{N}\big(E(m)\big) D_m.$$

*The largest rate $R$ such that a code with error $\leq \epsilon$ exists, is called the $\epsilon$-one-shot-capacity of $\mathcal{N}$, and denoted $C_\epsilon(\mathcal{N})$.*

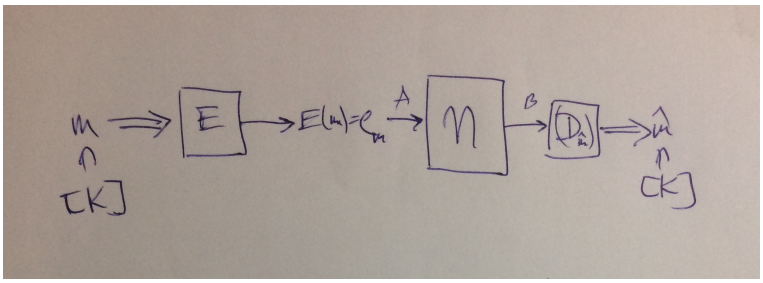


Figure 1: Diagramme of classical communication via a noisy quantum channel, including encoding and decoding of a uniformly distributed message.

The communication diagramme in Fig. 1 elucidates this definition. Indeed, we can introduce a uniformly distributed random variable $M \in [K]$, making the encoding $E(M) = \rho_M$ a random state, and so is the channel output $\mathcal{N}(\rho_M)$; finally, Born's rule defines the conditional probability of the decoded message $\widehat{M}$, a random variable jointly distributed with $M$:

$$\Pr\left\{M = m, \widehat{M} = m'\right\} = \frac{1}{K} \operatorname{Tr} \mathcal{N}\big(E(m)\big) D_{m'}. \tag{2}$$

**Remark 1.9** *The one-shot capacity is monotonic non-increasing under pre- and post-processings of the channel and under decrease of $\epsilon$. Indeed, for $\mathcal{N} : A \to B$ and two channel $\mathcal{A} : A_0 \to A$, $\mathcal{B} : B \to B_0$, as well as $0 \leq \epsilon \leq \epsilon' \leq 1$,*

$$C_{\epsilon'}(\mathcal{N}) \geq C_\epsilon(\mathcal{N}) \geq C_\epsilon(\mathcal{N} \circ \mathcal{A}) \geq C_\epsilon(\mathcal{B} \circ \mathcal{N} \circ \mathcal{A}).$$

*This can be seen by converting a code for $\mathcal{B} \circ \mathcal{N} \circ \mathcal{A}$ with error probability $\leq \epsilon$ into one for $\mathcal{N}$ with the same error probability and same rate: Indeed, given the encoder map $E(m) = \rho_m$, we can*

*define $E'(m) = \mathcal{A}(\rho_m)$, and given the decoding POVM $(D_m : m \in [M])$, let $D'_m = \mathcal{B}^*(D_m)$. Then,*

$$\operatorname{Tr}\mathcal{N}\big(E'(m)\big)D'_m = \operatorname{Tr}\mathcal{N}\big(\mathcal{A}(\rho_m)\big)\mathcal{B}^*(D_m) = \operatorname{Tr}\mathcal{B}\Big(\mathcal{N}\big(\mathcal{A}(\rho_m)\big)\Big)D_m,$$

*and averaging over messages we are done.*

**Example 1.10** *It may help to consider some simple examples.*

1. *For the ideal channel $\mathrm{id}_A$, we can choose $K = |A|$ orthonormal basis states $\rho_m = |m\rangle\langle m|$, and let $D_m = |m\rangle\langle m|$ be the projectors of the corresponding basis measurement, showing that $C_\epsilon(\mathrm{id}_A) \geq C_0(\mathrm{id}_A) \geq \log|A|$. On the other hand, assume a code with $K$ messages and error $\leq \epsilon$. By definition,*

$$1 - \epsilon \leq \frac{1}{K}\sum_{m=1}^{K} \operatorname{Tr}\rho_m D_m \leq \frac{1}{K}\sum_{m=1}^{K} \operatorname{Tr} D_m = \frac{1}{K}\operatorname{Tr}\mathbb{1}_A = \frac{|A|}{K},$$

   *implying $K \leq \frac{|A|}{1-\epsilon}$. Thus, $C_\epsilon(\mathrm{id}_A) \leq \log|A| - \log(1-\epsilon)$.*

2. *For a constant channel $\mathcal{P}_\sigma$, intuitively we cannot transmit anything, since the output does not depend on the input. Indeed, the error probability evaluates to*

$$P_e = 1 - \frac{1}{K}\sum_{m=1}^{K} \operatorname{Tr}\sigma D_m = 1 - \frac{1}{K}.$$

   *This shows $C_\epsilon(\mathcal{P}_\sigma) \leq \log\left\lfloor\frac{1}{1-\epsilon}\right\rfloor$, which is actually an equality.*

Pre-shared quantum entanglement between sender and receiver can enhance the communication, by reducing the error probability or increasing the rate. Recall dense coding, which we describe for an ideal qubit channel $\mathrm{id}_2$: Alice and Bob are connected by the ideal qubit channel $\mathrm{id}_2$ and share a maximally entangled state $\Phi^{T_A T_B} = |\Phi\rangle\langle\Phi|^{T_A T_B}$, where $T_A$ is a qubit held by Alice ("Alice's share") and $T_B$ a qubit held by Bob ("Bob's share"); $|\Phi\rangle^{T_A T_B} = \frac{1}{\sqrt{2}}\left(|0\rangle^{T_A}|0\rangle^{T_B} + |1\rangle^{T_A}|1\rangle^{T_B}\right)$. Alice encodes one of four messages by applying the Pauli unitaries $\mathbb{1}$, $X$, $Y$, or $Z$ to her qubit $T_A$, and sending it across the channel $\mathrm{id}_2$. This gives Bob a two-qubit pure state, which depending on the message has state vector

$$|\Phi^+\rangle^{BT_B} = |\Phi\rangle,\ |\Psi^+\rangle^{BT_B} = (X\otimes\mathbb{1})|\Phi\rangle,\ |\Psi^-\rangle^{BT_B} = (Y\otimes\mathbb{1})|\Phi\rangle,\ |\Phi^-\rangle^{BT_B} = (Z\otimes\mathbb{1})|\Phi\rangle,$$

i.e. one of an orthonormal basis of the Hilbert space (known as the Bell basis). Thus, Bob can distinguish the states perfectly by a suitable orthogonal projective measurement (known as the Bell measurement).

For a general noisy channel $\mathcal{N}$, we can consider how well Alice and Bob can communicate via the channel, exploiting arbitrary pre-shared entanglement.

**Definition 1.11** *An* entanglement-assisted code *for $\mathcal{N} : A \to B$ is a triple $(\omega, E, D)$, consisting of a state $\omega$ on $T_A \otimes T_B$, a map $E : [K] \to \mathrm{CPTP}(T_A \to A)$ from the set $[K] = \{1, \ldots, K\}$ of messages (any set of $K$ elements would do) into the channels mapping $T_A$ to $A$, and a measurement (POVM) $D = (D_m : m \in [K])$ on $B \otimes T_B$. We call $E$ the encoder and $D$ the decoder of the message; $E(m) = \mathcal{E}_m$ is called the modulation operation for message $m$.*

*As before, we call $R = \log K$ the (one-shot) rate of the code, and the (average) error probability is*

$$P_e = 1 - \frac{1}{K}\sum_{m=1}^{K} \operatorname{Tr}\big((\mathcal{N}\circ\mathcal{E}_m \otimes \mathrm{id}_{T_B})\omega\big) D_m.$$

*The largest rate $R$ such that an entanglement-assisted code with error $\leq \epsilon$ exists, is called the entanglement-assisted $\epsilon$-one-shot-capacity of $\mathcal{N}$, and denoted $C_\epsilon^{(ea)}(\mathcal{N})$, see Fig. 2.*

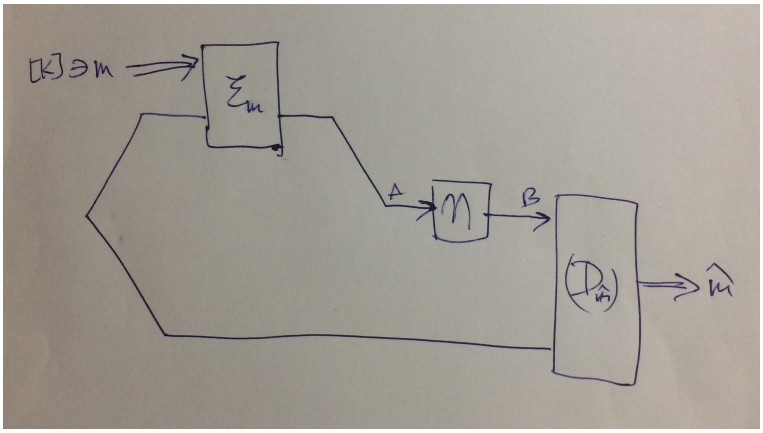


Figure 2: Diagramme of classical communication via a noisy quantum channel, assisted by entanglement, including encoding and decoding of a uniformly distributed message.

**Remark 1.12** *The discussion of dense coding shows that $C_\epsilon^{(ea)}(\mathrm{id}_2) \geq C_0^{(ea)}(\mathrm{id}_2) \geq 2$.*

**Remark 1.13** *Just as the plain one-shot capacity, the entanglement-assisted one is monotonic non-increasing under pre- and post-processings of the channel and under decrease of $\epsilon$. Indeed, for $\mathcal{N} : A \to B$ and two channel $\mathcal{A} : A_0 \to A$, $\mathcal{B} : B \to B_0$, as well as $0 \leq \epsilon \leq \epsilon' \leq 1$,*

$$C_{\epsilon'}^{(ea)}(\mathcal{N}) \geq C_\epsilon^{(ea)}(\mathcal{N}) \geq C_\epsilon^{(ea)}(\mathcal{N}\circ\mathcal{A}) \geq C_\epsilon^{(ea)}(\mathcal{B}\circ\mathcal{N}\circ\mathcal{A}).$$

*The proof is the same as before (Remark 1.9).*

**Remark 1.14** *As a matter of fact, the notion of entanglement-assisted code (Definition 1.11) generalises that of a plain code (Definition 1.8), $C_\epsilon^{(ea)}(\mathcal{N}) \geq C_\epsilon(\mathcal{N})$.*

*Indeed, any plain code with rate $R$ and $P_e \leq \epsilon$ can be used to define an entanglement assisted code with the same rate and error as follows. For a given encoder $E(m) = \rho_m$ and decoder $(D_m)$ of a plain code, define for the entanglement assisted code 1) any preshared state $\omega^{T_A T_B}$, 2) the encoder $\mathcal{E}_m = \mathcal{P}_{\rho_m}$ to be the constant channel so that $(\mathcal{E}_m \otimes \mathrm{id}_{T_B})\,\omega^{T_A T_B} = \rho_m \otimes \omega^{T_B}$, and 3) the decoder $\widetilde{D}_{m'} = D_{m'} \otimes \mathbb{1}_{T_B}$ which just ignores the $T_B$ space. It remains to check that this entanglement assisted code has the same error as the original plain code.*

$$\begin{aligned}\operatorname{Tr}\big((\mathcal{N}\circ\mathcal{E}_m \otimes \mathrm{id}_{T_B})\omega^{T_A T_B}\big)\widetilde{D}_{m'} &= \operatorname{Tr}\big(\mathcal{N}(\rho_m)\otimes\omega^{T_B}\big)\big(D_{m'}\otimes\mathbb{1}_{T_B}\big)\\ &= \big(\operatorname{Tr}\mathcal{N}(\rho_m)D_{m'}\big)\big(\operatorname{Tr}\omega^{T_B}\big)\\ &= \operatorname{Tr}\mathcal{N}(\rho_m)D_{m'},\end{aligned}$$

*which means the constructed entanglement assisted code has the same error as the plain code.*

Lecture 2 (15/12/2021)

Lecture 2 (14/12/2022)

## 1.3 Coding theorems via hypothesis testing

While in simple examples as the ones discussed above, it may be easily possible to characterise $C_\epsilon(\mathcal{N})$ or $C_\epsilon^{(ea)}(\mathcal{N})$ exactly or approximately, and to exhibit explicit good codes, general channels require a different approach. On the one hand, this includes *random coding*, pioneered by Shannon in his 1948 paper. On the other hand, we have to work with decoders that are potentially suboptimal, but at the same time easier to analyze. We begin with the definition of an important non-Shannon information quantity.

**Definition 1.15** *For two states $\rho$ and $\sigma$ on the same system, and $0 \leq \epsilon \leq 1$, define the* hypothesis testing relative entropy

$$D_h^\epsilon(\rho\|\sigma) := -\log \min \operatorname{Tr} \sigma M \text{ s.t. } 0 \leq M \leq \mathbb{1} \text{ and } \operatorname{Tr} \rho M \geq 1-\epsilon.$$

The hypothesis testing relative entropy encodes the tradeoff between the error probabilities of first and second kind in the hypothesis test between states $\rho$ and $\sigma$. While this seems elementary enough, and the logarithm may appear as no more than a decoration, the quantity has a number of good properties that make it behave like a relative entropy $D(\rho\|\sigma) = \operatorname{Tr} \rho(\log \rho - \log \sigma)$.

**Lemma 1.16** *The hypothesis testing relative entropy has the following properties.*

1. *For any states $\rho$, $\sigma$,*
$$D_h^\epsilon(\rho\|\sigma) \geq D_h^\epsilon(\rho\|\rho) = -\log(1-\epsilon) \geq 0.$$
2. *For any states $\rho$, $\sigma$ on $A$, and any quantum channel $\mathcal{N} : A \to B$,*
$$D_h^\epsilon(\rho\|\sigma) \geq D_h^\epsilon\big(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)\big).$$
3. *For the maximally mixed state $\tau_A = \frac{1}{|A|}\mathbb{1}_A$ and a pure state $\rho = |\psi\rangle\langle\psi|$, we have $D_h^\epsilon(\rho\|\tau) = \log|A| - \log(1-\epsilon)$.*

**Proof.** *1.* The first inequality follows from item 2, applied with the constant channel $\mathcal{P}_\rho$. The equality follows because in the definition we demand $\operatorname{Tr} \rho M \leq 1-\epsilon$, hence $D_h^\epsilon(\rho\|\rho) \geq -\log(1-\epsilon)$; this value is attained with $M = (1-\epsilon)\mathbb{1}$.

*2.* Let $M$ be the optimiser of $D_h^\epsilon\big(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)\big)$, in particular $1-\epsilon \leq \operatorname{Tr} \mathcal{N}(\rho)M = \operatorname{Tr} \rho\mathcal{N}^*(M)$. Thus, $M' = \mathcal{N}^*(M)$ is feasible for the problem $D_h^\epsilon(\rho\|\sigma)$. Furthermore, $2^{-D_h^\epsilon\left(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)\right)} = \operatorname{Tr} \mathcal{N}(\sigma)M = \operatorname{Tr} \sigma\mathcal{N}^*(M) \geq 2^{-D_h^\epsilon(\rho\|\sigma)}$.

*3.* A feasible $M$ for the problem $D_h^\epsilon(\rho\|\tau)$ must have $1-\epsilon \leq \operatorname{Tr} \rho M \leq \operatorname{Tr} M$, thus $\operatorname{Tr} \tau M \geq \frac{1-\epsilon}{|A|}$. Equality is attained for $M = (1-\epsilon)\rho$. □

Lecture 2 (13/12/2021)

**Theorem 1.17 (Wang/Renner, arXiv:1007.5456)** *For any channel $\mathcal{N} : A \to B$, and any $0 < \epsilon < 1$, we have for $\epsilon' < \epsilon$*

$$C_\epsilon(\mathcal{N}) \geq \sup_{\{P(x),\rho_x\}} \log \left\lfloor 2^{D_h^{\epsilon'}(\omega^{XB}\|\omega^X\otimes\omega^B) - \log\frac{4\epsilon}{(\epsilon-\epsilon')^2}} \right\rfloor$$

*where $\rho_x \in \mathcal{S}(A)$ are states, $P(x)$ is a probability function, and*

$$\omega^{XB} = \sum_x P(x)|x\rangle\langle x|^X \otimes \mathcal{N}(\rho_x).$$

We will prove this achievability result in this section. In a certain sense the theorem is best possible, up to additive constants, because of the following converse:

**Theorem 1.18** *For any channel $\mathcal{N} : A \to B$,*

$$C_\epsilon(\mathcal{N}) \leq \sup_{\{P(x),\rho_x\}} D_h^\epsilon(\omega^{XB} \| \omega^X \otimes \omega^B).$$

We will prove the converse in the next section. To prove the achievability result, Theorem 1.17, we need a few auxiliary concepts and lemmas.

**Definition 1.19** *For a family of operators $S_i \geq 0$, $i \in \mathcal{I}$, the* square-root measurement *is defined as the POVM $(M_i : i \in \mathcal{I})$, with*

$$M_i = \left(\sum_{i\in\mathcal{I}} S_i\right)^{-\frac{1}{2}} S_i \left(\sum_{i\in\mathcal{I}} S_i\right)^{-\frac{1}{2}}.$$

**Lemma 1.20 (Hayashi/Nagaoka, arXiv:quant-ph/0206186)** *For operators $0 \leq S \leq \mathbb{1}$ and $T \geq 0$, and any $c > 0$, it holds*

$$\mathbb{1} - (S+T)^{-\frac{1}{2}} S (S+T)^{-\frac{1}{2}} \leq (1+c)(\mathbb{1} - S) + \left(2 + c + c^{-1}\right) T.$$

**Proof.** We have $(A - cB)^\dagger (A - cB) \geq 0$, for arbitrary operators $A$ and $B$, and any real number $c > 0$. Thus, by rearranging the terms in this inequality,

$$A^\dagger B + B^\dagger A \leq c^{-1} A^\dagger A + c B^\dagger B. \tag{3}$$

We will apply this to $A = \sqrt{T}$ and $B = \sqrt{T}\left((S+T)^{-\frac{1}{2}} - \mathbb{1}\right)$. Namely, we have

$$\begin{aligned}
\mathbb{1} - (S+T)^{-\frac{1}{2}} S (S+T)^{-\frac{1}{2}} &= (S+T)^{-\frac{1}{2}} T (S+T)^{-\frac{1}{2}} \\
&= T + T\left((S+T)^{-\frac{1}{2}} - \mathbb{1}\right) + \left((S+T)^{-\frac{1}{2}} - \mathbb{1}\right) T \\
&\qquad + \left((S+T)^{-\frac{1}{2}} - \mathbb{1}\right) T \left((S+T)^{-\frac{1}{2}} - \mathbb{1}\right) \\
&= A^\dagger A + B^\dagger B + A^\dagger B + B^\dagger A \\
&\leq \left(1 + c^{-1}\right) T + (1+c) \left((S+T)^{-\frac{1}{2}} - \mathbb{1}\right) T \left((S+T)^{-\frac{1}{2}} - \mathbb{1}\right) \\
&\leq \left(1 + c^{-1}\right) T + (1+c) \left((S+T)^{-\frac{1}{2}} - \mathbb{1}\right) (S+T) \left((S+T)^{-\frac{1}{2}} - \mathbb{1}\right) \\
&= \left(1 + c^{-1}\right) T + (1+c) \left(1 + S + T - 2(S+T)^{\frac{1}{2}}\right) \\
&\leq \left(1 + c^{-1}\right) T + (1+c) \left(1 + S + T - 2S\right) \\
&= (1+c)(\mathbb{1} - S) + \left(2 + c + c^{-1}\right) T,
\end{aligned}$$

where in the penultimate line we have used $S + T \geq S$, hence $(S+T)^{\frac{1}{2}} \geq S^{\frac{1}{2}}$ by the next Lemma 1.21, as well as $S^{\frac{1}{2}} \geq S$ since $S$ is a POVM element. □

**Lemma 1.21** *The inverse function is operator anti-monotone on positive definite operators: if $0 < A \leq B$, it follows that $A^{-1} \geq B^{-1}$.*

*The square root function is operator monotone on positive semidefinite operators: for operators $0 \leq A \leq B$, it follows that $\sqrt{A} \leq \sqrt{B}$.*

**Proof.** To start with the first part, $A \leq B$ implies

$$\mathbb{1} \leq A^{-\frac{1}{2}} B A^{-\frac{1}{2}} = A^{-\frac{1}{2}} B^{\frac{1}{2}} B^{\frac{1}{2}} A^{-\frac{1}{2}}.$$

This in turn yields

$$\mathbb{1} \leq B^{\frac{1}{2}} A^{-1} B^{\frac{1}{2}},$$

thus the claim $B^{-1} \leq A^{-1}$.

For the second part, first assume that $B > 0$. We start similarly

$$\mathbb{1} \geq B^{-\frac{1}{2}} A^{\frac{1}{2}} A^{\frac{1}{2}} B^{-\frac{1}{2}} =: Y^\dagger Y.$$

This means that all singular values of $Y = A^{\frac{1}{2}} B^{-\frac{1}{2}}$ are $\leq 1$, in particular its eigenvalues are bounded by 1. But $Y$ is similar to $Z = B^{-\frac{1}{4}} A^{\frac{1}{2}} B^{-\frac{1}{4}}$, which has the same eigenvalues and is Hermitian, so $Z \leq \mathbb{1}$, which can be rearranged to say $A^{\frac{1}{2}} \leq B^{\frac{1}{2}}$. For the general case, set $B' = B + \epsilon\mathbb{1}$, $\epsilon > 0$ so that $A^{\frac{1}{2}} \leq B'^{\frac{1}{2}}$, so that $A^{\frac{1}{2}} \leq B^{\frac{1}{2}}$ follows by taking the limit $\epsilon \to 0$. □

Lecture 3 (19/12/2022)

**Proof of Theorem 1.17.** Let us fix $P(x)$ and $\rho_x$, furthermore $\epsilon' < \epsilon$ and $c > 0$. Then, it is enough to show that a code with rate $R$ exists that has error

$$P_e \leq (1+c)\epsilon' + \left(2 + c + c^{-1}\right) 2^{R - D_h^{\epsilon'}(\omega^{XB} \| \omega^X \otimes \omega^B)} \leq \epsilon,$$

because we can choose $c = \frac{\epsilon - \epsilon'}{\epsilon + \epsilon'}$. The second inequality is satisfied by choosing

$$\begin{aligned} R = \log K &= \log \left\lfloor 2^{D_h^{\epsilon'}(\omega^{XB} \| \omega^X \otimes \omega^B) - \log \frac{2+c+c^{-1}}{\epsilon - (1+c)\epsilon'}} \right\rfloor \\ &= \log \left\lfloor 2^{D_h^{\epsilon'}(\omega^{XB} \| \omega^X \otimes \omega^B) - \log \frac{4\epsilon}{(\epsilon - \epsilon')^2}} \right\rfloor. \end{aligned}$$

Indeed, choose an optimal POVM element $S$ according to the definition of the hypothesis testing relative entropy $D_h^{\epsilon'}(\omega^{XB} \| \omega^X \otimes \omega^B)$, which without loss of generality has the form $S = \sum_x |x\rangle\langle x|^X \otimes S_x$ with POVM elements $0 \leq S_x \leq \mathbb{1}$ on $B$. In particular,

$$\operatorname{Tr} \omega^{XB} S = \sum_x P(x) \operatorname{Tr} \omega_x S_x \geq 1 - \epsilon', \tag{4}$$

$$\operatorname{Tr}(\omega^X \otimes \omega^B) S = \sum_x P(x) \operatorname{Tr} \omega^B S_x = 2^{-D_h^{\epsilon'}(\omega^{XB} \| \omega^X \otimes \omega^B)}, \tag{5}$$

with $\omega_x = \mathcal{N}(\rho_x)$ and $\omega^B = \sum_x P(x) \mathcal{N}(\rho_x)$.

We define the encoder by picking $K$ elements $x_1, \ldots, x_K$ independently at random according to the distribution $P$, i.e. we let $E(m) := \rho_{x_m}$. The decoder will be the square-root measurement of the operators $S_{x_m}$,

$$D_m := \left( \sum_{m'=1}^K S_{x_{m'}} \right)^{-\frac{1}{2}} S_{x_m} \left( \sum_{m'=1}^K S_{x_{m'}} \right)^{-\frac{1}{2}}.$$

By Lemma 1.20, applied to $S = S_{x_m}$ and $S + T = \sum_{m=1}^K S_{x_m} =: R$, we have

$$\mathbb{1} - D_m \leq (1+c)(\mathbb{1} - S_{x_m}) + \left(2 + c + c^{-1}\right)(R - S_{x_m}). \tag{6}$$

This means that we can upper bound the probability of error of our code as follows:

$$\begin{aligned} P_e &= \frac{1}{K}\sum_{m=1}^{K} \operatorname{Tr}\omega_{x_m}(\mathbb{1}-D_m) \\ &\leq \frac{1+c}{K}\sum_{m=1}^{K}\operatorname{Tr}\omega_{x_m}(\mathbb{1}-S_{x_m}) + \frac{2+c+c^{-1}}{K}\sum_{m=1}^{K}\operatorname{Tr}\omega_{x_m}(R-S_{x_m}). \end{aligned}$$

Now, we take the average over codes (i.e. the independent choices of $x_m$ according to the distribution $P$), yielding

$$\begin{aligned} \mathbb{E}P_e &\leq (1+c)\sum_x P(x)\operatorname{Tr}\omega_x(\mathbb{1}-S_x) + \left(2+c+c^{-1}\right)(K-1)\sum_{x,x'}P(x)P(x')\operatorname{Tr}\omega_x S_{x'} \\ &\leq (1+c)\left(1-\operatorname{Tr}\omega^{XB}S\right) + K\left(2+c+c^{-1}\right)\operatorname{Tr}(\omega^X\otimes\omega^B)S \\ &\leq (1+c)\epsilon' + K\left(2+c+c^{-1}\right)2^{-D_h^{\epsilon'}(\omega^{XB}\|\omega^X\otimes\omega^B)}, \end{aligned}$$

which is what we wanted to show. □

Lecture 3 (20/12/2021)
Lecture 3 (18/12/2023)

To analyse the entanglement-assisted capacity $C_\epsilon^{(ea)}(\mathcal{N})$ of a channel $\mathcal{N}$, we will use the hypothesis testing relative entropy and the Hayashi/Nagaoka Lemma 1.20 once more, but perhaps surprisingly, we abandon the random coding strategy in favour of an explicit code.

**Lemma 1.22 (Position-based coding, Anshu/Jain/Warsi, 1702.01940)** *Let Alice and Bob share unlimited copies of a state $\rho^{A\widetilde{B}}$, i.e. $\omega^{A^n\widetilde{B}^n} = \rho^{\otimes n}$ for arbitrary integers $n$. Then, if $\log K \leq D_h^\eta\left((\mathcal{N}\otimes\mathrm{id}_{\widetilde{B}})\rho^{A\widetilde{B}}\middle\|\mathcal{N}(\rho^A)^B\otimes\rho^{\widetilde{B}}\right) - \log\frac{1}{\delta}$, there exists an entanglement-assisted code for $\mathcal{N}$ of rate $\log K$, with error $\leq \epsilon = (1+c)\eta + \left(2+c+c^{-1}\right)\delta$, for any $c>0$. The protocol actually uses $n=K$ copies of $\rho$.*

**Proof.** Alice and Bob initially share $n=K$ copies of $\rho^{A\widetilde{B}}$, i.e.

$$\omega^{A^K\widetilde{B}^K} = \rho^{A_1\widetilde{B}_1}\otimes\rho^{A_2\widetilde{B}_2}\otimes\cdots\otimes\rho^{A_K\widetilde{B}_K}.$$

Alice's modulation for message $m\in[K]$ consists in tracing out all $A_i$, $i\neq m$ and using the unitary isomorphism $\iota_m : A_m\to A$ to map to the channel input:

$$\mathcal{E}_m = \iota_m\circ\bigotimes_{i\neq m}\operatorname{Tr}_{A_i}.$$

Bob, on receiving the channel output, has the following state on $\widetilde{B}^K B$:

$$\Theta_m^{\widetilde{B}^K B} = (\mathcal{N}^{A_m\to B}\otimes\mathrm{id}_{\widetilde{B}_m})\rho^{A_m\widetilde{B}_m}\otimes\bigotimes_{i\neq m}\rho^{\widetilde{B}_i}$$

Motivated by the observation that the two-party marginal states of $\Theta$ are

$$\Theta_m^{B\widetilde{B}_{m'}} = \begin{cases} (\mathcal{N}^{A_m\to B}\otimes\mathrm{id}_{\widetilde{B}_m})\rho^{A_m\widetilde{B}_m} & \text{if } m'=m, \\ \mathcal{N}(\rho^A)^B\otimes\rho^{\widetilde{B}_{m'}} & \text{if } m'\neq m, \end{cases} \tag{7}$$

we let $S$ be the optimal POVM element for $D_h^\eta := D_h^\eta\big((\mathcal{N}\otimes \mathrm{id}_{\widetilde{B}})\rho^{A\widetilde{B}}\big\|\mathcal{N}(\rho^A)^B\otimes\rho^{\widetilde{B}}\big)$, so that

$$\begin{aligned}\mathrm{Tr}\big((\mathcal{N}\otimes \mathrm{id}_{\widetilde{B}})\rho^{A\widetilde{B}}\big)S &\geq 1-\eta,\\ \mathrm{Tr}\big(\mathcal{N}(\rho^A)^B\otimes\rho^{\widetilde{B}}\big)S &= 2^{-D_h^\eta}.\end{aligned} \tag{8}$$

Defining

$$S_m := S^{B\widetilde{B}_m}\otimes\bigotimes_{i\neq m}\mathbb{1}_{\widetilde{B}_i},$$

Eqs. (7) and (8) translate into

$$\mathrm{Tr}\,\Theta_m S_{m'}\begin{cases}\geq 1-\eta & \text{if } m'=m,\\ = 2^{-D_h^\eta} & \text{if } m'\neq m.\end{cases} \tag{9}$$

Now we define the decoding POVM once again by the square-root measurement, i.e.

$$D_m := \left(\sum_{m'=1}^K S_{m'}\right)^{-\frac{1}{2}} S_m \left(\sum_{m'=1}^K S_{m'}\right)^{-\frac{1}{2}},$$

we can bound the error probability of this code using the Hayashi-Nagaoka Lemma 1.20:

$$\begin{aligned}P_e &= \frac{1}{K}\sum_{m=1}^K \mathrm{Tr}\,\Theta_m(\mathbb{1}-D_m)\\ &\leq \frac{1}{K}\sum_{m=1}^K (1+c)\,\mathrm{Tr}\,\Theta_m(\mathbb{1}-S_m) + \frac{1}{K}\sum_{m=1}^K\sum_{m'\neq m}\left(2+c+c^{-1}\right)\mathrm{Tr}\,\Theta_m S_{m'}\\ &\leq (1+c)\eta + K\left(2+c+c^{-1}\right)2^{-D_h^\eta}.\end{aligned}$$

By assumption, $\log K \leq D_h^\eta - \log\frac{1}{\delta}$, so the error probability is bounded by $\epsilon := (1+c)\eta + \left(1+c+c^{-1}\right)\delta$. □

We make two applications of this, one with an optimal state, and the second with a separable state of cq-form.

**Theorem 1.23** *For a quantum channel $\mathcal{N}: A\to B$, let $0\leq\eta<\epsilon\leq 1$. Then we have*

$$\begin{aligned}&\sup_{\rho^{A\widetilde{B}}}\log\left\lfloor 2^{D_h^\eta\left((\mathcal{N}\otimes\mathrm{id}_{\widetilde{B}})\rho^{A\widetilde{B}}\middle\|\mathcal{N}(\rho^A)^B\otimes\rho^{\widetilde{B}}\right)-\log\frac{4\epsilon}{(\epsilon-\eta)^2}}\right\rfloor\\ &\qquad\leq C_\epsilon^{(ea)}(\mathcal{N}) \leq \sup_{\rho^{A\widetilde{B}}} D_h^\epsilon\big((\mathcal{N}\otimes\mathrm{id}_{\widetilde{B}})\rho^{A\widetilde{B}}\big\|\mathcal{N}(\rho^A)^B\otimes\rho^{\widetilde{B}}\big).\end{aligned}$$

*In both suprema, without loss of generality $\widetilde{B}\simeq A$ and the state $\rho^{A\widetilde{B}}$ is pure.*

We prove the lower bound (achievability) here, and the upper bound in the next section.

**Proof of the purity claim and the lower bound (achievability).** We begin with the purity claim. Indeed, consider an arbitrary state $\rho^{A\widetilde{B}}$ with arbitrary system $\widetilde{B}$. As is well-known, it can be purified to a pure state $|\psi\rangle\langle\psi|^{A\widetilde{B}\widetilde{C}}$ such that

$$\rho^{A\widetilde{B}} = \mathrm{Tr}_{\widetilde{C}} |\psi\rangle\langle\psi|^{A\widetilde{B}\widetilde{C}}.$$

The partial trace is a cptp map, so

$$D_h^\epsilon\big((\mathcal{N}\otimes \mathrm{id}_{\widetilde{B}})\rho^{A\widetilde{B}}\big\|\mathcal{N}(\rho^A)^B\otimes\rho^{\widetilde{B}}\big) \le D_h^\epsilon\big((\mathcal{N}\otimes \mathrm{id}_{\widetilde{B}\widetilde{C}})\psi^{A\widetilde{B}\widetilde{C}}\big\|\mathcal{N}(\psi^A)^B\otimes\psi^{\widetilde{B}\widetilde{C}}\big).$$

But up to an isometry on $\widetilde{B}\widetilde{C}$, the pure state $|\psi\rangle^{A\widetilde{B}\widetilde{C}}$ is equivalent to a state where the second system is $\simeq A$; however, this isometry leaves the latter hypothesis testing relative entropy invariant.

Now, the lower bound on $C_\epsilon^{(ea)}(\mathcal{N})$ follows from applying Lemma 1.22 with $\rho^{A\widetilde{B}}$, $c = \frac{\epsilon-\eta}{\epsilon+\eta}$ and $\delta = \frac{(\epsilon-\eta)^2}{4\epsilon}$ (cf. Theorem 1.17). □

Lecture 4 (21/12/2022)

By employing a separable state rather than a pure entangled one, we can recover the Wang/Renner coding theorem for the plain one-shot capacity.

**Proof of Theorem 1.17.** Consider the state

$$\rho^{AX} = \sum_x P(x)\rho_x^A \otimes |x\rangle\langle x|^X$$

shared between Alice (A) and Bob (X). Then, Lemma 1.22 yields a code, assisted by many copies of the shared state $\rho^{AX}$, of error $\le \epsilon$ and rate

$$R = \log\left\lfloor 2^{D_h^\eta\left((\mathcal{N}\otimes\mathrm{id}_X)\rho^{AX}\middle\|\mathcal{N}(\rho^A)^B\otimes\rho^X\right) - \log\frac{4\epsilon}{(\epsilon-\eta)^2}}\right\rfloor.$$

Since $\left(\rho^{AX}\right)^{\otimes K} = \rho^{A^K X^K}$ is a probabilistic mixture of states $\rho_{x^K}^{A^K}\otimes|x^K\rangle\langle x^K|^{X^K}$, for strings $x^K = x_1\ldots x_K$, there exists one such string which has the property that if we run the code with that initial state, it has at most the same error $\epsilon$. But now Remark 1.14 kicks in, showing that the code thus described is really a plain code for $\mathcal{N}$. □

## 1.4 Converse bounds and a metaconverse from generalised divergences

Here we first prove the upper bounds for the plain capacity $C_\epsilon(\mathcal{N})$ and the entanglement assisted capacity $C_\epsilon^{(ea)}(\mathcal{N})$ stated in the previous section. We will see that the proofs for these upper bounds have a similar form and with the introduction of a generic notion of "generalised divergences" the results can be generalised into a "metaconverse".

**Proof of Theorem 1.18.** Let $\mathcal{X} = [K]$ for an optimal code of error $\le \epsilon$, i.e. $\log K = C_\epsilon(\mathcal{N})$. Let the code input states be $\rho_m$ and form the state

$$\gamma^{XB} := \frac{1}{K}\sum_m |m\rangle\langle m|^X \otimes \mathcal{N}(\rho_m).$$

It will be enough to show that $\log K \le D_h^\epsilon(\gamma^{XB}\|\gamma^X\otimes\gamma^B)$. Note that $\gamma^X$ is the maximally mixed state, due to the uniform distribution on $\mathcal{X}$.

Now, in a first step, we can apply $\mathrm{id}_X \otimes \mathcal{D}$ to both arguments in this hypothesis testing relative entropy, where $\mathcal{D}$ is the qc-channel of the code's decoding POVM $(D_{m'})$. This maps $\gamma^{XB}$ to $\kappa^{XX'} = \sum_{mm'} \left(\frac{1}{K} \operatorname{Tr} \mathcal{N}(\rho_m) D_{m'}\right) |m\rangle\langle m|^X \otimes |m'\rangle\langle m'|^{X'}$, and $\gamma^X \otimes \gamma^B$ to $\kappa^X \otimes \kappa^{X'}$; by Lemma 1.16, point 2, we thus have

$$D_h^\epsilon(\gamma^{XB} \| \gamma^X \otimes \gamma^B) \geq D_h^\epsilon(\kappa^{XX'} \| \kappa^X \otimes \kappa^{X'}).$$

Secondly, we claim that $D_h^\epsilon(\kappa^{XX'} \| \kappa^X \otimes \kappa^{X'}) \geq \log K$. Indeed, consider $M = \sum_m |m\rangle\langle m|^X \otimes |m\rangle\langle m|^{X'}$, which is a POVM element that satisfies

$$\operatorname{Tr} \kappa^{XX'} M = \sum_m \frac{1}{K} \operatorname{Tr} \mathcal{N}(\rho_m) D_m \geq 1 - \epsilon$$

by assumption, so it is eligible in the definition of $D_h^\epsilon$. Thus,

$$D_h^\epsilon(\kappa^{XX'} \| \kappa^X \otimes \kappa^{X'}) \geq -\log \operatorname{Tr}(\kappa^X \otimes \kappa^{X'}) M = -\log \left( \sum_m \frac{1}{K} \langle m | \kappa^{X'} | m \rangle \right) = \log K,$$

concluding the proof. □

**Proof of the upper bound of Theorem 1.23.** The upper bound of Theorem 1.23 is similar to Theorem 1.18, and due to Matthews and Wehner (arXiv:1210.4722). The idea is that we compare the performance of the actual code over the channel $\mathcal{N}$ with that over a constant channel $\mathcal{P}_\sigma$, via a suitable hypothesis test. For that purpose, define

$$\alpha^{XAT_B} := \frac{1}{K} \sum_{m=1}^K |m\rangle\langle m|^X \otimes (\mathcal{E}_m \otimes \mathrm{id}_{T_B}) \omega^{T_A T_B},$$

and observe

$$\gamma^{XBT_B} = (\mathcal{N} \otimes \mathrm{id}_{XT_B})\alpha, \quad \alpha^X \otimes \sigma^B \otimes \omega^{T_B} = (\mathcal{P}_\sigma \otimes \mathrm{id}_{XT_B})\alpha = \sigma^B \otimes \alpha^{XT_B}.$$

Then we have the following chain of inequalities:

$$\begin{aligned} D_h^\epsilon\big((\mathcal{N} \otimes \mathrm{id}_{XT_B})\alpha \big\| (\mathcal{P}_\sigma \otimes \mathrm{id}_{XT_B})\alpha\big) &\geq D_h^\epsilon\big((\mathcal{D} \otimes \mathrm{id}_X)\gamma \big\| (\mathcal{D} \otimes \mathrm{id}_X)(\alpha^X \otimes \sigma^B \otimes \omega^{T_B})\big) \\ &= D_h^\epsilon(\kappa^{XX'} \| \kappa^X \otimes \kappa^{X'}) \\ &\geq \log K, \end{aligned}$$

where the first line is due to the monotonicity of $D_h^\epsilon$ under cptp maps, in this case the decoding qc-channel; the second line uses the notation of the proof of Theorem 1.18, where also the third line is argued.

On the other hand, we can purify $\alpha^{XAT_B}$ to a state $\psi^{AA'}$, where $A' = XT_BC$ includes another system $C$, and then we get

$$\begin{aligned} D_h^\epsilon\big((\mathcal{N} \otimes \mathrm{id}_{XT_B})\alpha \big\| (\mathcal{P}_\sigma \otimes \mathrm{id}_{XT_B})\alpha\big) &\leq D_h^\epsilon\big((\mathcal{N} \otimes \mathrm{id}_{A'})\psi \big\| (\mathcal{P}_\sigma \otimes \mathrm{id}_{A'})\psi\big) \\ &= D_h^\epsilon\big((\mathcal{N} \otimes \mathrm{id}_{A'})\psi \big\| \sigma^B \otimes \psi^{A'}\big), \end{aligned}$$

once more by the monotonicity of the hypothesis testing relative entropy under cptp maps, here a partial trace. Applying this with $\sigma = \mathcal{N}(\alpha^A)$ concludes the proof. □

Let us review how the converse bounds above came about, focusing on the plain capacity $C_\epsilon(\mathcal{N})$. Indeed, while the direct coding theorems used properties of the hypothesis testing relative entropy, the converses relied on more generic, axiomatic properties of "generalised divergences".

**Definition 1.24** *A function $\mathbb{D}(\rho\|\sigma)$ on pairs of states and semidefinite operators from the same system is called a* (generalised) divergence *if*

1. *$\mathbb{D}$ takes values in $\mathbb{R} \cup \{\infty\}$*
2. *$\mathbb{D}(\rho\|\sigma) \geq d_0$ for all states $\rho$ and $\sigma$, with a system-independent constant $d_0$;*
3. *$\mathbb{D}(\rho\|S) \geq \mathbb{D}\big(\mathcal{N}(\rho)\|\mathcal{N}(S)\big)$ for all states $\rho$ and semidefinite operators $S$ and all quantum channels $\mathcal{N}$.*

Examples include the quantum relative entropy (aka Umegaki relative entropy)

$$D(\rho\|\sigma) = \operatorname{Tr} \rho(\log \rho - \log \sigma),$$

the hypothesis testing relative entropy $D_h^\epsilon(\rho\|\sigma)$, and a number of other quantities, which will be discussed in more detail in the next chapter. We record two, however, without proof, which we want to exploit here.

**Definition 1.25** *For $\alpha \in (0;1) \cup (1;\infty)$, define the* Rényi relative entropy *as*

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \operatorname{Tr} \rho^\alpha \sigma^{1-\alpha},$$

*which is a generalised divergence for $\alpha \in (0;1) \cup (1;2)$.*

*Define the* sandwiched Rényi relative entropy *as*

$$\widetilde{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \operatorname{Tr} \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha,$$

*which is a generalised divergence for $\alpha \in \left(\frac{1}{2};1\right) \cup (1;\infty)$.*

**Lemma 1.26 (Metaconverse)** *For a channel $\mathcal{N} : A \to B$, $0 < \epsilon < 1 - \frac{1}{K}$ and any generalised divergence $\mathbb{D}$, if there exists a code with error $\leq \epsilon$ for $K$ messages, then*

$$\begin{aligned}\mathbb{D}\left((1-\epsilon,\epsilon)\middle\|\left(\frac{1}{K}, 1-\frac{1}{K}\right)\right) &\leq \sup_{\{P(x),\rho_x\}} \inf_{\sigma^B} \mathbb{D}(\omega^{XB}\|\omega^X \otimes \sigma^B) \\ &\leq \sup_{\{P(x),\rho_x\}} \mathbb{D}(\omega^{XB}\|\omega^X \otimes \omega^B),\end{aligned}$$

*where*

$$\omega^{XB} = \sum_x P(x)|x\rangle\langle x|^X \otimes \mathcal{N}(\rho_x),$$

*and $\sigma^B$ is an arbitrary state on $B$.*

**Proof.** As in the proof of Theorem 1.18, we use the code itself to come up with a candidate for $\omega$. Indeed, take a code of $K$ messages and error $\leq \epsilon$ and define

$$\gamma^{XB} := \frac{1}{K} \sum_m |m\rangle\langle m|^X \otimes \mathcal{N}(\rho_m).$$

We compare this state to the analogous one, when instead of the channel $\mathcal{N}$ the constant channel $\mathcal{P}_\sigma$ is applied:

$$\frac{1}{K}\sum_m |m\rangle\langle m|^X \otimes \mathcal{P}_\sigma(\rho_m) = \gamma^X \otimes \sigma^B,$$

and consider how each is transformed under the decoding operation of the code.

That is, in the first step we apply $\mathrm{id}_X \otimes \mathcal{D}$ to both these states as arguments in the generalised divergence, where $\mathcal{D}$ is the qc-channel of the code's decoding POVM $(D_{m'})$. This maps $\gamma^{XB}$ to $\kappa^{XX'} = \sum_{mm'} \frac{1}{K} \operatorname{Tr} \mathcal{N}(\rho_m) D_{m'} |m\rangle\langle m|^X \otimes |m'\rangle\langle m'|^{X'}$, and $\gamma^X \otimes \sigma^B$ to $\kappa^X \otimes \lambda^{X'}$, and by definition of $D$, we have

$$\mathbb{D}(\gamma^{XB} \| \gamma^X \otimes \sigma^B) \geq \mathbb{D}(\kappa^{XX'} \| \kappa^X \otimes \lambda^{X'}).$$

Secondly, we consider the binary POVM with elements $M = \sum_m |m\rangle\langle m|^X \otimes |m\rangle\langle m|^{X'}$ and $\mathbb{1} - M$. Recall

$$\operatorname{Tr} \kappa^{XX'} M = \sum_m \frac{1}{K} \operatorname{Tr} \mathcal{N}(\rho_m) D_m = P_e \geq 1 - \epsilon,$$

by assumption on the code, whereas

$$\operatorname{Tr}(\kappa^X \otimes \lambda^{X'}) M = \sum_m \frac{1}{K} \langle m | \lambda^{X'} | m \rangle = \frac{1}{K}.$$

We thus obtain

$$\begin{aligned}
\mathbb{D}(\kappa^{XX'} \| \kappa^X \otimes \lambda^{X'}) &\geq \mathbb{D}\left( (1 - P_e, P_e) \middle\| \left( \frac{1}{K}, 1 - \frac{1}{K} \right) \right) \\
&\geq \mathbb{D}\left( (1 - \epsilon, \epsilon) \middle\| \left( \frac{1}{K}, 1 - \frac{1}{K} \right) \right),
\end{aligned}$$

where the second inequality follows from $0 \leq P_e \leq \epsilon < 1 - \frac{1}{K}$ and the fact that under this condition there exists a binary stochastic map $R : \{0,1\} \to \{0,1\}$ having $\left(\frac{1}{K}, 1 - \frac{1}{K}\right)$ as a fixed point and such that $T(1 - P_e, P_e) = (1 - \epsilon, \epsilon)$. □

**Corollary 1.27** *For a channel $\mathcal{N} : A \to B$, $0 < \epsilon < 1$ and $\alpha > 1$,*

$$C_\epsilon(\mathcal{N}) \leq \sup_{\{P(x), \rho_x\}} \inf_{\sigma^B} \widetilde{D}_\alpha(\omega^{XB} \| \omega^X \otimes \sigma^B) - \frac{\alpha}{\alpha - 1} \log(1 - \epsilon).$$

**Proof.** We simply use the Metaconverse Lemma 1.26 with the sandwiched Rényi relative entropy. Observe that the left hand side of the central relation can be evaluated and lower bounded as follows:

$$\begin{aligned}
\widetilde{D}_\alpha\left( (1 - \epsilon, \epsilon) \middle\| \left( \frac{1}{K}, 1 - \frac{1}{K} \right) \right) &= \frac{1}{\alpha - 1} \log \left( (1 - \epsilon)^\alpha \frac{1}{K^{1-\alpha}} + \epsilon^\alpha \left( 1 - \frac{1}{K} \right)^{1-\alpha} \right) \\
&\geq \frac{1}{\alpha - 1} \log (1 - \epsilon)^\alpha \frac{1}{K^{1-\alpha}} \\
&= \log K + \frac{\alpha}{\alpha - 1} \log(1 - \epsilon),
\end{aligned}$$

which implies the claim. □

Lecture 4 (20/12/2023)

In a similar way, we can prove the following upper bound on the entanglement-assisted capacity:

**Lemma 1.28 (Metaconverse entanglement-assisted)** *For a channel $\mathcal{N} : A \to B$, $0 < \epsilon < 1 - \frac{1}{K}$ and any generalised divergence $\mathbb{D}$, if there exists an entanglement-assisted code with error $\leq \epsilon$ for $K$ messages, then*

$$\begin{aligned}\mathbb{D}\left((1-\epsilon,\epsilon)\middle\|\left(\frac{1}{K},1-\frac{1}{K}\right)\right) &\leq \sup_{\psi^{AA'}}\inf_{\sigma^B}\mathbb{D}((\mathrm{id}_A\otimes\mathcal{N})\psi\|\psi^A\otimes\sigma^B)\\ &\leq \sup_{\psi^{AA'}}\mathbb{D}((\mathrm{id}_A\otimes\mathcal{N})\psi\|\psi^A\otimes\mathcal{N}(\psi^{A'})^B),\end{aligned}$$

*where $\sigma^B$ is an arbitrary state on $B$.*

**Proof.** Like Lemma 1.26, this is an abstract version of our previous argument in terms of $D_h^\epsilon$. Namely, consider an entanglement-assisted code for $\mathcal{N}$ with $K$ messages and error $\epsilon$, and construct the state

Lecture 4 (22/12/2021)

$$\alpha^{XAT_B} := \frac{1}{K}\sum_{m=1}^{K}|m\rangle\langle m|^X\otimes(\mathcal{E}_m\otimes\mathrm{id})\omega^{T_AT_B}.$$

Observe now

$$\begin{aligned}\gamma^{XBT_B} &= (\mathcal{N}\otimes\mathrm{id}_{XT_B})\alpha^{XAT_B} = \frac{1}{K}\sum_{m=1}^{K}|m\rangle\langle m|^X\otimes(\mathcal{N}\circ\mathcal{E}_m\otimes\mathrm{id})\omega^{T_AT_B},\\ \alpha^X\otimes\sigma^B\otimes\omega^{T_B} &= (\mathcal{P}_\sigma\otimes\mathrm{id}_{XT_B})\alpha^{XAT_B},\end{aligned}$$

with any constant channel $\mathcal{P}_\sigma$.

We have then, using the decoding measurement (qc-channel) and the test $M$ as before,

$$\begin{aligned}\mathbb{D}((\mathcal{N}\otimes\mathrm{id}_{XT_B})\alpha\|(\mathcal{P}_\sigma\otimes\mathrm{id}_{XT_B})\alpha) &= \mathbb{D}(\gamma^{XBT_B}\|\gamma^X\otimes\sigma^B\otimes\omega^{T_B})\\ &\geq \mathbb{D}(\kappa^{XX'}\|\kappa^X\otimes\lambda^{X'})\\ &\geq \mathbb{D}\left((1-P_e,P_e)\middle\|\left(\frac{1}{K},1-\frac{1}{K}\right)\right)\\ &\geq \mathbb{D}\left((1-\epsilon,\epsilon)\middle\|\left(\frac{1}{K},1-\frac{1}{K}\right)\right).\end{aligned}\tag{10}$$

We can now purify $\alpha^{XAT_B}$ to $\psi^{AA'}$, with $A' = XT_BC$, which can only increase the leftmost generalised divergence. As this holds for any $\sigma^B$, we get our upper bound as claimed. □

**Corollary 1.29** *For a channel $\mathcal{N} : A \to B$, $0 < \epsilon < 1$ and $\alpha > 1$,*

$$C_\epsilon^{(ea)}(\mathcal{N}) \leq \sup_{\psi^{AA'}}\inf_{\sigma^B}\widetilde{D}_\alpha((\mathrm{id}_A\otimes\mathcal{N})\psi\|\psi^A\otimes\sigma^B) - \frac{\alpha}{\alpha-1}\log(1-\epsilon).$$

## 1.5 Application to memoryless channels: Holevo capacity & strong converse

To show that the bounds of Theorem 1.17 and Corollary 1.27 are actually good, and can be evaluated, we consider a cq-channel $\mathcal{N} : X \to B$ mapping $x \equiv |x\rangle\langle x|$ to $\rho_x$ in the i.i.d. regime. In other words, we really look at $\mathcal{N}^{\otimes n}$ for $n \gg 1$. We shall evaluate the asymptotic capacity $\frac{1}{n} C_\epsilon(\mathcal{N}^{\otimes n})$ when $n \to \infty$.

Lecture 5 (9/1/2023)

Recall that for a quantum state $\rho$, its von Neumann entropy is

$$H(\rho) = -\operatorname{Tr} \rho \log \rho = \min -\operatorname{Tr} \rho \log \sigma \text{ s.t. } \sigma \geq 0, \operatorname{Tr} \sigma = 1,$$

and that for a bipartite state $\rho^{AB}$, the quantum mutual information is

$$I(A:B)_\rho = D(\rho^{AB} \| \rho^A \otimes \rho^B) = H(\rho^A) + H(\rho^B) - H(\rho^{AB})$$

**Theorem 1.30 (Holevo-Schumacher-Westmoreland)** *For every $0 < \epsilon < 1$, the cq-channel $\mathcal{N} : X \to B$, mapping $|x\rangle\langle x|$ to $\mathcal{N}(|x\rangle\langle x|) =: \omega_x^B$ has asymptotic capacity*

$$\lim_{n\to\infty} \frac{1}{n} C_\epsilon(\mathcal{N}^{\otimes n}) = \max_{P(x)} I(X:B)_\omega =: \chi(\mathcal{N}),$$

*where as before*

$$\omega^{XB} = \sum_x P(x) |x\rangle\langle x|^X \otimes \mathcal{N}(|x\rangle\langle x|)^B.$$

*and*

$$I(X:B)_\omega = H\left(\sum_x P(x) \mathcal{N}(|x\rangle\langle x|)\right) - \sum_x P(x) H\big(\mathcal{N}(|x\rangle\langle x|)\big)$$

*is the Holevo information.*

**Proof.** We prove the equality by separately showing $\geq$ for the $\liminf$ and $\leq$ for the $\limsup$.

"$\geq$": In Theorem 1.30, we have proved that for a probability distribution $P$ on $\mathcal{X}$, we have $C_\epsilon(\mathcal{N}) \geq D_h^{\epsilon/2}\left(\omega^{XB} \| \omega^X \otimes \omega^B\right) - \log \frac{16}{\epsilon}$. Thus, likewise, for the channel $\mathcal{N}^{\otimes n}$ and the probability distribution $P^{\otimes n}$, the associated cq-state is $\left(\omega^{XB}\right)^{\otimes n}$, and we have

$$C_\epsilon\left(\mathcal{N}^{\otimes n}\right) \geq D_h^{\epsilon/2}\left((\omega^{XB})^{\otimes n} \middle\| (\omega^X \otimes \omega^B)^{\otimes n}\right) - \log \frac{16}{\epsilon}.$$

Now, we invoke a result from the next section, the so-called *asymptotic equipartition property (AEP)* of the hypothesis testing relative entropy (Theorem 2.9 in the next chapter), which states that

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} D_h^{\epsilon/2}\left((\omega^{XB})^{\otimes n} \middle\| (\omega^X \otimes \omega^B)^{\otimes n}\right) &= D\left(\omega^{XB} \| \omega^X \otimes \omega^B\right) \\ &= \operatorname{Tr} \omega^{XB}\big(\log \omega^{XB} - \log(\omega^X \otimes \omega^B)\big) \\ &= I(X:B)_\omega.\end{aligned}$$

This proves the direct part.

"$\leq$": We use Corollary 1.27 for $\mathcal{N}^{\otimes n}$, which adapted to cq-channels reads

$$C_\epsilon(\mathcal{N}^{\otimes n}) \leq \sup_{P^{(n)}} \inf_{\sigma^{B^n}} \widetilde{D}_\alpha\left(\omega^{X^nB^n} \middle\| \omega^{X^n} \otimes \sigma^{B^n}\right) - \frac{\alpha}{\alpha-1}\log(1-\epsilon),$$

where $P^{(n)}$ is a distribution on $\mathcal{X}^n$, and $\omega^{X^nB^n} = \sum_{x^n} P(x^n)|x^n\rangle\langle x^n|^{X^n} \otimes \omega_{x^n}^{B^n}$, using the notations $|x^n\rangle\langle x^n| = |x_1\rangle\langle x_1| \otimes \cdots \otimes |x_n\rangle\langle x_n|$ and $\omega_{x^n}^{B^n} = \omega_{x_1}^{B_1} \otimes \cdots \otimes \omega_{x_n}^{B_n}$. Furthermore, $\sigma^{B^n}$ is an arbitrary state, but we shall restrict the minimisation to tensor product states $\sigma^{B^n} = \sigma_1^{B_1} \otimes \cdots \otimes \sigma_n^{B_n}$, which only can make the right hand side larger.

Now, by plugging the two states into the expression for the sandwiched Rényi relative entropy, we have

$$\widetilde{D}_\alpha\left(\omega^{X^nB^n} \middle\| \omega^{X^n} \otimes \sigma^{B^n}\right) = \frac{1}{\alpha-1}\log\sum_{x^n} P^{(n)}(x^n)\widetilde{Q}_\alpha(\omega_{x^n}\|\sigma^{B^n}),$$

with

$$\widetilde{Q}_\alpha(\rho\|\sigma) := \operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha.$$

As both $\omega_{x^n}$ and $\sigma^{B^n}$ are product states, we have

$$\widetilde{Q}_\alpha(\omega_{x^n}\|\sigma^{B^n}) = \prod_{i=1}^n \widetilde{Q}_\alpha(\omega_{x_i}\|\sigma_i),$$

implying

$$\widetilde{D}_\alpha\left(\omega^{X^nB^n} \middle\| \omega^{X^n} \otimes \sigma^{B^n}\right) \leq \sum_{i=1}^n \max_{x_i} \widetilde{D}_\alpha(\omega_{x_i}\|\sigma_i).$$

All taken together, we thus obtain

$$C_\epsilon(\mathcal{N}^{\otimes n}) \leq n\left[\inf_{\sigma^B}\max_P \widetilde{D}_\alpha\left(\omega^{XB}\middle\|\omega^X\otimes\sigma^B\right)\right] - \frac{\alpha}{\alpha-1}\log(1-\epsilon).$$

It remains to prove that the expression inside the square bracket converges to $\chi(\mathcal{N})$ for $\alpha \to 1$. Indeed, as $\widetilde{D}_\alpha \to D$ (the Umegaki relative entropy) for $\alpha \to 1$, – see the next chapter –, this is equivalent to

$$\begin{aligned}\chi(N) &= \max_P D(\omega^{XB}\|\omega^X\otimes\omega^B)\\ &= \max_P \min_{\sigma^B} D(\omega^{XB}\|\omega^X\otimes\sigma^B)\\ &= \min_{\sigma^B}\max_P D(\omega^{XB}\|\omega^X\otimes\sigma^B),\end{aligned}$$

where the second equality is a consequence of

$$\begin{aligned}D(\omega^{XB}\|\omega^X\otimes\sigma^B) &= \operatorname{Tr}\omega^{XB}\log\omega^{XB} - \operatorname{Tr}\omega^{XB}(\log\omega^X\otimes\mathbb{1} + \mathbb{1}\otimes\log\sigma^B)\\ &= \sum_x P(x)\log P(x) - \sum_x P(x)H(\omega_x)\\ &\quad - \sum_x P(x)\log P(x) + D\left(\omega^B\|\sigma^B\right) + H(\omega^B)\\ &= D(\omega^{XB}\|\omega^X\otimes\omega^B) + D\left(\omega^B\|\sigma^B\right),\end{aligned}$$

and the third equality can be proved by a minimax argument. For a full proof, see Khathi and Wilde, Chapter 7. □

**Theorem 1.31 (Bennett-Shor-Smolin-Thapliyal)** *For any channel $\mathcal{N} : A \to B$ and $0 < \epsilon < 1$,*

$$\lim_{n\to\infty} \frac{1}{n} C_\epsilon^{(ea)}(\mathcal{N}^{\otimes n}) = \max_{\psi^{AA'}} I(A:B)_\omega,$$

*where $|\psi\rangle \in A \otimes A'$ varies over pure states and $\omega^{AB} = (\mathrm{id}_A \otimes \mathcal{N})\psi$.*