Chapter 8

Quantum channel capacities

This chapter is focused on *capacities* of quantum channels for transmitting information. The notion of a channel capacity has multiple, inequivalent formulations in the quantum setting. For example, one may consider the capacity with which classical or quantum information can be transmitted through a channel, and different resources may or may not be available to assist the information transmission—such as entanglement shared between a sender and receiver before the information transmission takes place.

Three fundamental theorems are presented, characterizing the capacities of quantum channels to transmit either classical or quantum information, both with and without the assistance of prior shared entanglement. When prior shared entanglement between the sender and receiver is not available, these characterizations have a somewhat undesirable property: they require a *regularization*—or an averaging over an increasingly large number of uses of a given channel—and fail to provide capacity formulas that are either explicit or efficiently computable for this reason. The apparent need for such regularizations is discussed in the last section of the chapter, along with the related phenomenon of *super-activation* of quantum capacity.

8.1 Classical information over quantum channels

The general scenario to be considered throughout this chapter involves two hypothetical individuals: a *sender* and a *receiver*. The sender wishes to transmit information, either classical or quantum, to the receiver, and is able to do this through multiple, independent uses of a given channel Φ. One aims

to design a scheme by which the sender prepares an input to these channel uses and the receiver processes their output in such a way that information is transmitted with a high degree of accuracy. As is standard in information theory, the chapter mainly deals with the asymptotic regime, making use of entropic notions to analyze rates of information transmission in the limit of an increasingly large number of independent channel uses.

The subject of the present section is the capacity of quantum channels to transmit *classical* information, including both the case in which the sender and receiver share prior entanglement and in which they do not. The first subsection below introduces notions and terminology concerning channel capacities that will be needed throughout the section, as well as in later parts of the chapter. The second subsection is devoted to a proof of the *Holevo–Schumacher–Westmoreland theorem*, which characterizes the capacity of a channel to transmit classical information without the use of prior shared entanglement. The final subsection proves the *entanglement-assisted capacity theorem*, which characterizes the capacity of a channel to transmit classical information with the assistance of prior shared entanglement.

8.1.1 Classical capacities of quantum channels

Five quantities that relate to the information-transmitting capabilities of channels are defined below. The first two quantities are fundamental with respect to the subject of quantum channel capacities: the *classical capacity* and the *entanglement-assisted classical capacity* of a quantum channel. The remaining three quantities are the *Holevo capacity*, the *entanglement-assisted Holevo capacity*, and the *coherent information*, all of which play important roles in the main results to be presented.

The classical capacity of a channel

Intuitively (and somewhat informally) speaking, the classical capacity of a channel describes the average number of classical bits of information that can be transmitted, with a high degree of accuracy, through each use of that channel. As is typical for information-theoretic notions, channel capacities are more formally defined in terms of asymptotic behaviors, where the limit of an increasing number of channel uses is considered.

When stating a precise mathematical definition of classical capacity, it is convenient to refer to the *emulation* of one channel by another, as well as to

the *approximation* of one channel by another, with the approximation being defined with respect to the completely bounded trace norm.

Definition 8.1. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{Z})$ be channels, for $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ being complex Euclidean spaces. It is said that the channel Φ *emulates* Ψ if there exist channels $\Xi_{\mathrm{E}} \in \mathrm{C}(\mathcal{Z}, \mathcal{X})$ and $\Xi_{\mathrm{D}} \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ such that

$$\Psi = \Xi_{\mathrm{D}} \Phi \, \Xi_{\mathrm{E}}. \tag{8.1}$$

When this relationship holds, the channel Ξ_{E} is called an *encoding channel* and Ξ_{D} is called a *decoding channel*.

Definition 8.2. Let $\mathcal{Z}$ be a complex Euclidean space, let $\Psi_0, \Psi_1 \in \mathrm{C}(\mathcal{Z})$ be channels, and let $\varepsilon > 0$ be a positive real number. The channel Ψ_0 is an ε-approximation to Ψ_1 (or, equivalently, Ψ_1 is an ε-approximation to Ψ_0) if and only if

$$\left\| \Psi_0 - \Psi_1 \right\|_1 < \varepsilon. \tag{8.2}$$

The definition of the classical capacity of a quantum channel, which makes use of the previous two definitions, is as follows.

Definition 8.3 (Classical capacity of a channel)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Let $\Gamma = \{0, 1\}$ denote the binary alphabet, let $\mathcal{Z} = \mathbb{C}^{\Gamma}$, and let $\Delta \in \mathrm{C}(\mathcal{Z})$ denote the completely dephasing channel defined with respect to the space $\mathcal{Z}$.

1. A value $\alpha \geq 0$ is an *achievable rate* for classical information transmission through Φ if and only if (i) $\alpha = 0$, or (ii) $\alpha > 0$ and the following holds for every choice of a positive real number $\varepsilon > 0$: for all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, the channel $\Phi^{\otimes n}$ emulates an ε-approximation to the channel $\Delta^{\otimes m}$.
2. The *classical capacity* of Φ, denoted $\mathrm{C}(\Phi)$, is the supremum value of all achievable rates for classical information transmission through Φ.

In the context of Definition 8.3, the completely dephasing channel Δ is to be viewed as an ideal channel for transmitting a single bit of classical information. When considering an emulation of the m-fold tensor product $\Delta^{\otimes m}$ of this ideal classical channel by the channel $\Phi^{\otimes n}$, no generality is lost in restricting one's attention to classical-to-quantum encoding channels Ξ_{E}

and quantum-to-classical decoding channels Ξ_{D}. That is, one may assume that the conditions

$$\Xi_{\mathrm{E}} = \Xi_{\mathrm{E}}\Delta^{\otimes m} \quad \text{and} \quad \Xi_{\mathrm{D}} = \Delta^{\otimes m}\Xi_{\mathrm{D}} \tag{8.3}$$

are met. This is so because

$$\begin{aligned} &\left|\!\left|\!\left| \left(\Delta^{\otimes m}\Xi_{\mathrm{D}}\right)\Phi^{\otimes n}\left(\Xi_{\mathrm{E}}\Delta^{\otimes m}\right) - \Delta^{\otimes m}\right|\!\right|\!\right|_1 \\ &\qquad = \left|\!\left|\!\left| \Delta^{\otimes m}\left(\Xi_{\mathrm{D}}\Phi^{\otimes n}\Xi_{\mathrm{E}} - \Delta^{\otimes m}\right)\Delta^{\otimes m}\right|\!\right|\!\right|_1 \\ &\qquad \leq \left|\!\left|\!\left| \Xi_{\mathrm{D}}\Phi^{\otimes n}\Xi_{\mathrm{E}} - \Delta^{\otimes m}\right|\!\right|\!\right|_1, \end{aligned} \tag{8.4}$$

which implies that replacing a given choice of Ξ_{E} and Ξ_{D} by $\Xi_{\mathrm{E}}\Delta^{\otimes m}$ and $\Delta^{\otimes m}\Xi_{\mathrm{D}}$ cannot decrease the quality of the emulation that is achieved.

In light of this observation, the implicit use of the completely bounded trace norm in Definition 8.3 may appear to be somewhat heavy-handed; an equivalent definition is obtained by requiring that $\Phi^{\otimes n}$ emulates any channel $\Psi \in \mathrm{C}(\mathcal{Z}^{\otimes m})$ satisfying

$$\left\| \left(\Delta^{\otimes m}\Psi\right)\left(E_{a_1\cdots a_m,a_1\cdots a_m}\right) - E_{a_1\cdots a_m,a_1\cdots a_m} \right\|_1 < \varepsilon, \tag{8.5}$$

which is equivalent to

$$\left\langle E_{a_1\cdots a_m,a_1\cdots a_m}, \Psi(E_{a_1\cdots a_m,a_1\cdots a_m}) \right\rangle > 1 - \frac{\varepsilon}{2}, \tag{8.6}$$

for all $a_1 \cdots a_m \in \Gamma^m$. An interpretation of this requirement is that every string $a_1 \cdots a_m \in \Gamma^m$ is transmitted by Ψ with a probability of error smaller than $\varepsilon/2$.

There is, on the other hand, a benefit to using the more general notion of channel approximation defined by the completely bounded trace norm in Definition 8.3, which is that it allows the quantum capacity to be defined in an analogous manner to the classical capacity—replacing the dephasing channel Δ by the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}$. (For the quantum capacity, which is discussed later in Section 8.2, the completely bounded trace norm provides a natural notion of channel approximation.)

The following proposition is, perhaps, self-evident, but it is nevertheless worth stating explicitly. The same argument used to prove it may be applied to other notions of capacity as well—there is nothing specific to the classical capacity that is required by the proof.

Proposition 8.4. *Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, and let k be a positive integer. It holds that $\mathrm{C}(\Phi^{\otimes k}) = k\,\mathrm{C}(\Phi)$.*

Proof. Assume first that $\alpha \geq 0$ is an achievable rate for classical information transmission through Φ. It follows immediately that αk is an achievable rate for information transmission through $\Phi^{\otimes k}$, and therefore

$$\mathrm{C}\big(\Phi^{\otimes k}\big) \geq k\,\mathrm{C}(\Phi). \tag{8.7}$$

Next, assume that α is an achievable rate for information transmission through $\Phi^{\otimes k}$. If it is the case that $\alpha = 0$, then α/k is trivially an achievable rate for classical information transmission through Φ, so one may focus on the case that $\alpha > 0$. For any choice of a positive integer $n \geq k$, the channel $\Phi^{\otimes n}$ evidently emulates every channel emulated by the channel

$$\Phi^{\otimes k\lfloor n/k \rfloor}. \tag{8.8}$$

For every choice of $\varepsilon > 0$, all but finitely many positive integers n, and all positive integers $m \leq \alpha\lfloor n/k \rfloor$, the channel $\Phi^{\otimes n}$ therefore emulates an ε-approximation to $\Delta^{\otimes m}$. For all $\delta \in (0, \alpha/k)$, it holds that

$$\alpha\lfloor n/k \rfloor \geq (\alpha/k - \delta)n \tag{8.9}$$

for all but finitely many positive integers n, implying that $\alpha/k - \delta$ is an achievable rate for classical information transmission through Φ.

Taking the supremum over all achievable rates, one finds that

$$\mathrm{C}(\Phi) \geq \frac{1}{k}\,\mathrm{C}\big(\Phi^{\otimes k}\big), \tag{8.10}$$

which completes the proof. □

The entanglement-assisted classical capacity of a channel

The entanglement-assisted classical capacity of a channel is defined in a similar way to the classical capacity, except that one assumes that the sender and receiver may share any entangled state of their choosing prior to the transmission of information through the channel. The ability of the sender and receiver to share entanglement, as compared with the situation in which they do not, can result in a significant increase in the classical capacity of a

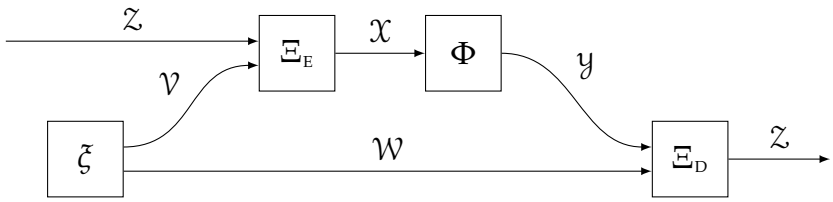

Figure 8.1: An illustration of the channel $\Psi(Z) = \big(\Xi_{\mathrm{D}}\big(\Phi\Xi_{\mathrm{E}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)\big)(Z \otimes \xi)$ referred to in Definition 8.5.

quantum channel. For instance, shared entanglement doubles the classical capacity of the identity channel through the use of dense coding (discussed in Section 6.3.1), and an arbitrary (constant-factor) increase is possible for other choices of channels.

A formal definition for the entanglement-assisted classical capacity of a channel requires only a minor change to the definition of the ordinary classical capacity. In particular, the definition of an emulation of one channel by another is modified to allow for the existence of a shared entangled state as follows.

Definition 8.5. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{Z})$ be channels, for $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ being complex Euclidean spaces. It is said that the channel Φ *emulates* Ψ *with the assistance of entanglement* if and only if there exist complex Euclidean spaces $\mathcal{V}$ and $\mathcal{W}$, a state $\xi \in \mathrm{D}(\mathcal{V} \otimes \mathcal{W})$, and channels $\Xi_{\mathrm{E}} \in \mathrm{C}(\mathcal{Z} \otimes \mathcal{V}, \mathcal{X})$ and $\Xi_{\mathrm{D}} \in \mathrm{C}(\mathcal{Y} \otimes \mathcal{W}, \mathcal{Z})$ such that

$$\Psi(Z) = \big(\Xi_{\mathrm{D}}\big(\Phi\Xi_{\mathrm{E}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)\big)(Z \otimes \xi) \tag{8.11}$$

for all $Z \in \mathrm{L}(\mathcal{Z})$. (See Figure 8.1 for an illustration of a channel Ψ satisfying this equation for all $Z \in \mathrm{L}(\mathcal{Z})$.) When this relationship holds, the channel Ξ_{E} is called an *encoding channel*, Ξ_{D} is called a *decoding channel*, and ξ is referred to as the *shared entangled state* that assists this emulation.

Aside from the modification represented by the previous definition, the entanglement-assisted classical capacity is defined in an analogous way to the ordinary classical capacity.

Definition 8.6 (Entanglement-assisted classical capacity of a channel)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Let $\Gamma = \{0, 1\}$ denote the binary alphabet, let $\mathcal{Z} = \mathbb{C}^{\Gamma}$, and let $\Delta \in \mathrm{C}(\mathcal{Z})$ denote the completely dephasing channel defined with respect to the space $\mathcal{Z}$.

1. A value $\alpha \geq 0$ is an *achievable rate* for entanglement-assisted classical information transmission through Φ if and only if (i) $\alpha = 0$, or (ii) $\alpha > 0$ and the following holds for every choice of a positive real number $\varepsilon > 0$: for all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, the channel $\Phi^{\otimes n}$ emulates an ε-approximation to the channel $\Delta^{\otimes m}$ with the assistance of entanglement.
2. The *entanglement-assisted classical capacity* of Φ, denoted $\mathrm{C}_{\mathrm{E}}(\Phi)$, is defined as the supremum over all achievable rates for entanglement-assisted classical information transmission through Φ.

Through the same argument used to prove Proposition 8.4, one has that the following simple proposition holds.

Proposition 8.7. *Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, and let k be a positive integer. It holds that $\mathrm{C}_{\mathrm{E}}(\Phi^{\otimes k}) = k\,\mathrm{C}_{\mathrm{E}}(\Phi)$.*

The Holevo capacity of a channel

Suppose that $\mathcal{X}$ is a complex Euclidean space, Σ is an alphabet, $p \in \mathcal{P}(\Sigma)$ is a probability vector, and $\{\rho_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X})$ is a collection of states. The Holevo information $\chi(\eta)$ of the ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ defined as

$$\eta(a) = p(a)\rho_a \tag{8.12}$$

for each $a \in \Sigma$ is given by

$$\chi(\eta) = \mathrm{H}\left(\sum_{a\in\Sigma} p(a)\rho_a\right) - \sum_{a\in\Sigma} p(a)\,\mathrm{H}(\rho_a). \tag{8.13}$$

Based on this quantity, one may define the *Holevo capacity* of a channel in the manner specified by Definition 8.8 below. This definition will make use of the following notation: for any ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ and any channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, one defines the ensemble $\Phi(\eta) : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ as

$$(\Phi(\eta))(a) = \Phi(\eta(a)) \tag{8.14}$$

for each $a \in \Sigma$. That is, $\Phi(\eta)$ is the ensemble obtained by evaluating Φ on the ensemble η in the most natural way.

Definition 8.8. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. The *Holevo capacity* of Φ is defined as

$$\chi(\Phi) = \sup_{\eta} \chi(\Phi(\eta)), \tag{8.15}$$

where the supremum is over all choices of an alphabet Σ and an ensemble of the form $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$.

Two restrictions may be placed on the supremum (8.15) in Definition 8.8 without decreasing the value that is defined for a given channel. The first restriction is that the supremum may be replaced by a maximum over all ensembles of the form $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$, for Σ being an alphabet of size

$$|\Sigma| = \dim(\mathcal{X})^2. \tag{8.16}$$

The second restriction is that one may restrict their attention to ensembles η for which $\mathrm{rank}(\eta(a)) \leq 1$ for each $a \in \Sigma$. The following proposition will be used in the proof that this is so.

Proposition 8.9. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, let Σ be an alphabet, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble. There exists an alphabet Γ and an ensemble $\tau : \Gamma \to \mathrm{Pos}(\mathcal{X})$ such that*

1. $\mathrm{rank}(\tau(b)) \leq 1$ *for each* $b \in \Gamma$, *and*
2. $\chi(\Phi(\eta)) \leq \chi(\Phi(\tau))$.

Proof. Assume that Δ is the alphabet for which $\mathcal{X} = \mathbb{C}^\Delta$, and let

$$\eta(a) = \sum_{c \in \Delta} \lambda_{a,c} x_{a,c} x_{a,c}^* \tag{8.17}$$

be a spectral decomposition of $\eta(a)$ for each $a \in \Sigma$. The requirements of the proposition hold for the ensemble $\tau : \Sigma \times \Delta \to \mathrm{Pos}(\mathcal{X})$ defined by

$$\tau(a, c) = \lambda_{a,c} x_{a,c} x_{a,c}^* \tag{8.18}$$

for each $(a, c) \in \Sigma \times \Delta$. It is evident that the first property holds, so it remains to verify the second.

Define $\mathcal{Z} = \mathbb{C}^\Sigma$ and $\mathcal{W} = \mathbb{C}^\Delta$, and consider three registers Y, Z, and W corresponding to the spaces $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$, respectively. For the density operator $\rho \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z} \otimes \mathcal{W})$ defined as

$$\rho = \sum_{(a,c) \in \Sigma \times \Delta} \lambda_{a,c} \Phi\big(x_{a,c} x_{a,c}^*\big) \otimes E_{a,a} \otimes E_{c,c}, \tag{8.19}$$

one has that the following two equalities hold:

$$\begin{aligned}\chi(\Phi(\tau)) &= \mathrm{D}\big(\rho[\mathsf{Y},\mathsf{Z},\mathsf{W}]\big\|\rho[\mathsf{Y}]\otimes\rho[\mathsf{Z},\mathsf{W}]\big),\\ \chi(\Phi(\eta)) &= \mathrm{D}\big(\rho[\mathsf{Y},\mathsf{Z}]\big\|\rho[\mathsf{Y}]\otimes\rho[\mathsf{Z}]\big).\end{aligned} \tag{8.20}$$

The inequality $\chi(\Phi(\eta)) \leq \chi(\Phi(\tau))$ follows from the monotonicity of the quantum relative entropy function under partial tracing (which represents a special case of Theorem 5.38). □

Theorem 8.10. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X},\mathcal{Y})$ be a channel, and let Σ be an alphabet having size $|\Sigma| = \dim(\mathcal{X})^2$. There exists an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ such that*

$$\chi(\Phi(\eta)) = \chi(\Phi). \tag{8.21}$$

One may assume, in addition, that $\mathrm{rank}(\eta(a)) \leq 1$ *for each $a \in \Sigma$.*

Proof. Consider an arbitrary ensemble of the form $\tau : \Gamma \to \mathrm{Pos}(\mathcal{X})$, for Γ being any alphabet, and let

$$\sigma = \sum_{a\in\Gamma} \tau(a) \tag{8.22}$$

denote the average state of the ensemble τ. Through Proposition 2.52, one finds that there must exist an alphabet Δ, a probability vector $p \in \mathcal{P}(\Delta)$, and a collection of ensembles $\{\tau_b : b \in \Delta\}$ taking the form $\tau_b : \Gamma \to \mathrm{Pos}(\mathcal{X})$, each satisfying the constraint

$$\sum_{a\in\Gamma} \tau_b(a) = \sigma \tag{8.23}$$

and possessing the property

$$\big|\{a \in \Gamma : \tau_b(a) \neq 0\}\big| \leq \dim(\mathcal{X})^2, \tag{8.24}$$

so that τ is given by the convex combination

$$\tau = \sum_{b\in\Delta} p(b)\tau_b. \tag{8.25}$$

By Proposition 5.51 it follows that

$$\chi(\Phi(\tau)) \leq \sum_{b\in\Delta} p(b)\chi(\Phi(\tau_b)), \tag{8.26}$$

and so there must exist at least one choice of a symbol $b \in \Delta$ for which $p(b) > 0$ and

$$\chi(\Phi(\tau)) \leq \chi(\Phi(\tau_b)). \tag{8.27}$$

Fix any such choice of $b \in \Delta$, and let

$$\Gamma_0 = \{a \in \Gamma \,:\, \tau_b(a) \neq 0\}. \tag{8.28}$$

For an arbitrarily chosen one-to-one mapping $f : \Gamma_0 \to \Sigma$, one obtains an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ such that

$$\chi(\Phi(\eta)) \geq \chi(\Phi(\tau)) \tag{8.29}$$

by setting $\eta(f(a)) = \tau_b(a)$ for every $a \in \Gamma_0$ and $\eta(c) = 0$ for $c \notin f(\Gamma_0)$.

Because the argument just presented holds for an arbitrary choice of an ensemble τ, it follows that

$$\chi(\Phi) = \sup_{\eta} \chi(\Phi(\eta)), \tag{8.30}$$

where the supremum is over all ensembles of the form $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$. As the set of all such ensembles is compact, there must exist an ensemble of the same form for which the equality (8.21) holds.

The additional restriction that $\mathrm{rank}(\eta(a)) \leq 1$ for each $a \in \Sigma$ may be assumed by first using Proposition 8.9 to replace a given ensemble τ by one satisfying the restriction $\mathrm{rank}(\tau(a)) \leq 1$ for each $a \in \Gamma$, and then proceeding with the argument above. This results in an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ with $\mathrm{rank}(\eta(a)) \leq 1$ for each $a \in \Sigma$, and such that (8.21) holds, which completes the proof. □

The entanglement-assisted Holevo capacity of a channel

Along similar lines to the entanglement-assisted classical capacity, which mirrors the definition of the classical capacity in the setting in which the sender and receiver initially share an entangled state of their choosing, one may define the entanglement-assisted Holevo capacity of a channel. The following definition is helpful when formalizing this notion.

Definition 8.11. Let Σ be an alphabet, let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ be an ensemble, and let

$$\rho = \sum_{a \in \Sigma} \eta(a) \tag{8.31}$$

denote the average state of η. It is said that η is *constant with respect to* $\mathcal{Y}$ if and only if there exists a probability vector $p \in \mathcal{P}(\Sigma)$ such that

$$\operatorname{Tr}_{\mathcal{X}}(\eta(a)) = p(a) \operatorname{Tr}_{\mathcal{X}}(\rho) \tag{8.32}$$

for each $a \in \Sigma$.

A simple operational characterization of ensembles constant with respect to a given complex Euclidean space is provided by the following proposition. In essence, it states that this sort of ensemble is one obtained by applying a randomly chosen channel to just one subsystem of a fixed bipartite state.

Proposition 8.12. *Let Σ be an alphabet, let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, and let $\eta : \Sigma \to \operatorname{Pos}(\mathcal{X} \otimes \mathcal{Y})$ be an ensemble. The following three statements are equivalent:*

1. *The ensemble η is constant with respect to $\mathcal{Y}$.*
2. *There exists a complex Euclidean space $\mathcal{Z}$, a state $\sigma \in \mathrm{D}(\mathcal{Z} \otimes \mathcal{Y})$, a probability vector $p \in \mathcal{P}(\Sigma)$, and a collection of channels $\{\Phi_a : a \in \Sigma\} \subseteq \mathrm{C}(\mathcal{Z}, \mathcal{X})$ such that*
$$\eta(a) = p(a) \big(\Phi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(\sigma) \tag{8.33}$$
for every $a \in \Sigma$.
3. *Statement 2 holds under the additional assumption that $\sigma = uu^*$ for some choice of a unit vector $u \in \mathcal{Z} \otimes \mathcal{Y}$.*

Proof. The fact that the second statement implies the first is immediate, and the third statement trivially implies the second. It therefore remains to prove that the first statement implies the third.

Assume that η is constant with respect to $\mathcal{Y}$, let ρ denote the average state of the ensemble η, as in Definition 8.11, and let

$$\xi = \operatorname{Tr}_{\mathcal{X}}(\rho). \tag{8.34}$$

Let $\mathcal{Z}$ be a complex Euclidean space having dimension $\dim(\mathcal{Z}) = \operatorname{rank}(\xi)$, and let $u \in \mathcal{Z} \otimes \mathcal{Y}$ be a unit vector that purifies ξ:

$$\operatorname{Tr}_{\mathcal{Z}}(uu^*) = \xi. \tag{8.35}$$

As η is constant with respect to $\mathcal{Y}$, it therefore holds that

$$p(a) \operatorname{Tr}_{\mathcal{Z}}(uu^*) = \operatorname{Tr}_{\mathcal{X}}(\eta(a)) \tag{8.36}$$

for each $a \in \Sigma$. By Proposition 2.29, one concludes that there must exist a channel $\Phi_a \in \mathrm{C}(\mathcal{Z}, \mathcal{X})$ such that

$$\eta(a) = p(a)\big(\Psi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)\big(uu^*\big) \tag{8.37}$$

for each $a \in \Sigma$. Setting $\sigma = uu^*$ completes the proof. □

Definition 8.13. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. The *entanglement-assisted Holevo capacity* of Φ is defined as

$$\chi_{\mathrm{E}}(\Phi) = \sup_{\eta} \chi\big(\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\eta)\big), \tag{8.38}$$

where the supremum is taken over all choices of an alphabet Σ, a complex Euclidean space $\mathcal{W}$, and an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{W})$ that is constant with respect to $\mathcal{W}$.

The relationship between the entanglement-assisted classical capacity and the entanglement-assisted Holevo capacity is discussed in Section 8.1.3 below. In this context, the fixed bipartite state whose existence is implied by Proposition 8.12, for a given ensemble that is constant with respect to $\mathcal{W}$, is representative of a (possibly entangled) state shared between a sender and receiver.

The coherent information

The final quantity, associated with a given channel, that is to be defined in the present subsection is the *coherent information*.

Definition 8.14. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel and let $\sigma \in \mathrm{D}(\mathcal{X})$ be a state, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. The *coherent information* of σ through Φ is defined as

$$\mathrm{I}_{\mathrm{C}}(\sigma\,;\Phi) = \mathrm{H}(\Phi(\sigma)) - \mathrm{H}\Big(\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)\Big(\mathrm{vec}\big(\sqrt{\sigma}\big)\,\mathrm{vec}\big(\sqrt{\sigma}\big)^*\Big)\Big). \tag{8.39}$$

The *maximum coherent information* of Φ is defined as

$$\mathrm{I}_{\mathrm{C}}(\Phi) = \max_{\sigma \in \mathrm{D}(\mathcal{X})} \mathrm{I}_{\mathrm{C}}(\sigma;\Phi). \tag{8.40}$$

In general terms, the coherent information of a state σ through a channel Φ quantifies the correlations that exist after Φ is applied to a purification

of σ. The definition implicitly takes this purification to be $\operatorname{vec}(\sqrt{\sigma})$ for the sake of simplicity and concreteness, but any other purification would result in the same quantity.

Consider the state

$$\rho = \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)\Big(\operatorname{vec}\big(\sqrt{\sigma}\big)\operatorname{vec}\big(\sqrt{\sigma}\big)^*\Big) \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{X}) \tag{8.41}$$

of a pair of registers (Y, X), corresponding to the spaces $\mathcal{Y}$ and $\mathcal{X}$, obtained as suggested above. One has that the coherent information $\mathrm{I_c}(\sigma\,;\,\Phi)$ of σ through Φ is equal to $\mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{Y}, \mathsf{X})$. The mutual information between Y and X is therefore given by

$$\mathrm{H}(\mathsf{Y} : \mathsf{X}) = \mathrm{I_c}(\sigma\,;\,\Phi) + \mathrm{H}(\sigma). \tag{8.42}$$

While it is not immediately clear that the coherent information is relevant to the notion of channel capacity, it will be proved later in the chapter that this quantity is fundamentally important with respect to both the entanglement-assisted classical capacity and the quantum capacity (to be defined later in Section 8.2).

The following proposition establishes an intuitive fact, which is that feeding the output of a channel into a second channel cannot increase the first channel's coherent information relative to a given state.

Proposition 8.15. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ be channels, and let $\sigma \in \mathrm{D}(\mathcal{X})$ be a state. It holds that*

$$\mathrm{I_c}(\sigma\,;\,\Psi\Phi) \leq \mathrm{I_c}(\sigma\,;\,\Phi). \tag{8.43}$$

Proof. Choose complex Euclidean spaces $\mathcal{W}$ and $\mathcal{V}$, along with isometries $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ and $B \in \mathrm{U}(\mathcal{Y}, \mathcal{Z} \otimes \mathcal{V})$, so that Stinespring representations of Φ and Ψ are obtained:

$$\Phi(X) = \operatorname{Tr}_{\mathcal{W}}(AXA^*) \quad \text{and} \quad \Psi(Y) = \operatorname{Tr}_{\mathcal{V}}(BYB^*) \tag{8.44}$$

for all $X \in \mathrm{L}(\mathcal{X})$ and $Y \in \mathrm{L}(\mathcal{Y})$. Define a unit vector $u \in \mathcal{Z} \otimes \mathcal{V} \otimes \mathcal{W} \otimes \mathcal{X}$ as

$$u = (B \otimes \mathbb{1}_{\mathcal{W}} \otimes \mathbb{1}_{\mathcal{X}})(A \otimes \mathbb{1}_{\mathcal{X}})\operatorname{vec}\big(\sqrt{\sigma}\big). \tag{8.45}$$

Now, consider four registers $\mathsf{Z}, \mathsf{V}, \mathsf{W}$, and X, corresponding to the spaces $\mathcal{Z}$, $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{X}$, respectively. Assuming the compound register $(\mathsf{Z}, \mathsf{V}, \mathsf{W}, \mathsf{X})$ is in the pure state uu^*, one has the following expressions:

$$\begin{aligned} \mathrm{I_c}(\sigma\,;\,\Phi) &= \mathrm{H}(\mathsf{Z}, \mathsf{V}) - \mathrm{H}(\mathsf{Z}, \mathsf{V}, \mathsf{X}), \\ \mathrm{I_c}(\sigma\,;\,\Psi\Phi) &= \mathrm{H}(\mathsf{Z}) - \mathrm{H}(\mathsf{Z}, \mathsf{X}). \end{aligned} \tag{8.46}$$

The proposition follows from the strong subadditivity of the von Neumann entropy (Theorem 5.39). □

It is convenient to refer to the notion of *complementary channels* in some of the proofs to be found in the present chapter that involve the coherent information. This notion is defined as follows.

Definition 8.16. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ be channels, for some choice of complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$. It is said that Φ and Ψ are *complementary* if and only if there exists an isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ for which it holds that

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}(AXA^*) \quad \text{and} \quad \Psi(X) = \mathrm{Tr}_{\mathcal{Y}}(AXA^*) \tag{8.47}$$

for every $X \in \mathrm{L}(\mathcal{X})$.

It is immediate from Corollary 2.27 that, for every channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, there must exist a complex Euclidean space $\mathcal{Z}$ and a channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ that is complementary to Φ; such a channel Ψ is obtained from any choice of a Stinespring representation of Φ.

Proposition 8.17. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, and let $\sigma \in \mathrm{D}(\mathcal{X})$ be a density operator. If $\mathcal{Z}$ is a complex Euclidean space and $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ is a channel that is complementary to Φ, then*

$$\mathrm{I_c}(\sigma\,;\Phi) = \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)). \tag{8.48}$$

Proof. Under the assumption that $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ is complementary to Φ, there must exist an isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ such that the equations (8.47) hold for every $X \in \mathrm{L}(\mathcal{X})$. Let X, Y, and Z be registers corresponding to the spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, define a unit vector $u \in \mathcal{Y} \otimes \mathcal{Z} \otimes \mathcal{X}$ as

$$u = \big(A \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)\,\mathrm{vec}\big(\sqrt{\sigma}\big), \tag{8.49}$$

and consider the compound register $(\mathsf{Y}, \mathsf{Z}, \mathsf{X})$. With respect to the pure state uu^* of this compound register, it holds that $\mathrm{H}(\mathsf{Z}) = \mathrm{H}(\mathsf{Y}, \mathsf{X})$, and therefore

$$\mathrm{H}\Big(\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)\Big(\mathrm{vec}\big(\sqrt{\sigma}\big)\,\mathrm{vec}\big(\sqrt{\sigma}\big)^*\Big)\Big) = \mathrm{H}(\Psi(\sigma)), \tag{8.50}$$

from which the proposition follows. □

8.1.2 The Holevo–Schumacher–Westmoreland theorem

This section states and proves the *Holevo–Schumacher–Westmoreland theorem*, which establishes that the classical capacity of a quantum channel is lower-bounded by its Holevo capacity, and that through a regularization of the Holevo capacity one obtains a characterization of the classical capacity.

The notion of a *classical-to-quantum product state channel code*, along with a few mathematical results that are useful for analyzing these codes, will be introduced prior to the statement and proof of the Holevo–Schumacher–Westmoreland theorem.

Classical-to-quantum product state channel codes

When studying the classical capacity of quantum channels, it is instructive to consider a related but somewhat more basic task of encoding classical information using fixed sets of quantum states. When this task is connected with the notion of the classical capacity of a given channel, a link must be made between that particular channel and the states that are used to encode classical information—but it is reasonable to begin by examining the task of encoding classical information into quantum states in isolation.

Throughout the discussion that follows, $\Gamma = \{0,1\}$ will denote the binary alphabet and

$$\{\sigma_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X}) \tag{8.51}$$

will denote a fixed collection of states, for $\mathcal{X}$ being a complex Euclidean space and Σ being an alphabet.[1] The situation to be considered is that binary strings, representing classical information, are to be encoded into tensor products of quantum states drawn from the collection (8.51) in such a way that each binary string can be recovered from its encoding with high probability.

In more precise terms, it is to be assumed that positive integers n and m have been selected, and that every binary string $b_1 \cdots b_m \in \Gamma^m$ of length m is to be *encoded* by a product state having the form

$$\sigma_{a_1} \otimes \cdots \otimes \sigma_{a_n} \in \mathrm{D}\big(\mathcal{X}^{\otimes n}\big), \tag{8.52}$$

[1] The entire discussion could be generalized to allow for arbitrary alphabets Γ in place of the binary alphabet. As there is little gain in doing this from the perspective of this book, the assumption that $\Gamma = \{0,1\}$ is made in the interest of simplicity.

for some choice of a string $a_1 \cdots a_n \in \Sigma^n$. That is, a function $f : \Gamma^m \to \Sigma^n$ is to be selected, and each string $b_1 \cdots b_m \in \Gamma^m$ is to be encoded by the state (8.52) for $a_1 \cdots a_n = f(b_1 \cdots b_m)$. When discussing this sort of code, it is convenient to make use of the shorthand notation

$$\sigma_{a_1 \cdots a_n} = \sigma_{a_1} \otimes \cdots \otimes \sigma_{a_n} \tag{8.53}$$

for each string $a_1 \cdots a_n \in \Sigma^n$, and with respect to this notation one has that

$$\sigma_{f(b_1 \cdots b_m)} \in \mathrm{D}\big(\mathcal{X}^{\otimes n}\big) \tag{8.54}$$

denotes the state that encodes the string $b_1 \cdots b_m \in \Gamma^m$.

From the encoding of a given binary string, one may hope to recover (or *decode*) this string by means of a measurement. Such a measurement takes the form $\mu : \Gamma^m \to \mathrm{Pos}\big(\mathcal{X}^{\otimes n}\big)$, and succeeds in successfully recovering a particular string $b_1 \cdots b_m$ from its encoding with probability

$$\big\langle \mu(b_1 \cdots b_m), \sigma_{f(b_1 \cdots b_m)} \big\rangle. \tag{8.55}$$

As a general guideline, one is typically interested in coding schemes for which the probability of a successful decoding is close to 1 and the ratio m/n, which represents the rate at which classical information is effectively transmitted, is as large as possible. The following definition summarizes these notions.

Definition 8.18. Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, let

$$\{\sigma_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X}) \tag{8.56}$$

be a collection of states, let $\Gamma = \{0, 1\}$ denote the binary alphabet, and let n and m be positive integers. A *classical-to-quantum product-state channel code* for the collection of states (8.56) is a pair (f, μ) consisting of a function and a measurement of the forms

$$f : \Gamma^m \to \Sigma^n \quad \text{and} \quad \mu : \Gamma^m \to \mathrm{Pos}\big(\mathcal{X}^{\otimes n}\big). \tag{8.57}$$

The *rate* of such a code is equal to the ratio m/n, and the code is said to have *error bounded by* δ if it holds that

$$\big\langle \mu(b_1 \cdots b_m), \sigma_{f(b_1 \cdots b_m)} \big\rangle > 1 - \delta \tag{8.58}$$

for every string $b_1 \cdots b_m \in \Gamma^m$.

Remark 8.19. The term *channel code* is used in this definition to distinguish this type of code from a *source code*, as discussed in Chapter 5. The two notions are, in some sense, complementary. A channel code represents the situation in which information is encoded into a state possessing a degree of noise or randomness, while a source code represents the situation in which information produced by a noisy or random source is encoded into any chosen state.

It is evident that some choices of sets $\{\sigma_a : a \in \Sigma\}$ are better suited to the construction of classical-to-quantum product state channel codes than others, assuming one wishes to maximize the rate and minimize the error probability for such a code. For the most part, the analysis that follows will be focused on the situation in which a set of states has been fixed, and one is interested in understanding the capabilities of this particular set, with respect to classical-to-quantum product state channel codes.

Typicality for ensembles of states

The notion of *typicality* is central to the proofs of multiple theorems to be presented in the current chapter, including a fundamental theorem on the existence of classical-to-quantum product-state channel codes possessing certain rates and error bounds.

A standard definition of typicality was introduced in Section 5.3.1—but it is an extension of this definition to ensembles of states that will be used in the context of channel coding. The following definition is a starting point for a discussion of this concept, providing a notion of typicality for joint probability distributions.

Definition 8.20. Let Σ and Γ be alphabets, let $p \in \mathcal{P}(\Sigma \times \Gamma)$ be a probability vector, and let $q \in \mathcal{P}(\Sigma)$ be the marginal probability vector defined as

$$q(a) = \sum_{b \in \Gamma} p(a, b) \tag{8.59}$$

for each $a \in \Sigma$. For each choice of a positive real number $\varepsilon > 0$, a positive integer n, and a string $a_1 \cdots a_n \in \Sigma^n$ satisfying $q(a_1) \cdots q(a_n) > 0$, a string $b_1 \cdots b_n \in \Gamma^n$ is said to be *ε-typical conditioned on* $a_1 \cdots a_n \in \Sigma^n$ if and only if

$$2^{-n(\mathrm{H}(p)-\mathrm{H}(q)+\varepsilon)} < \frac{p(a_1, b_1) \cdots p(a_n, b_n)}{q(a_1) \cdots q(a_n)} < 2^{-n(\mathrm{H}(p)-\mathrm{H}(q)-\varepsilon)}. \tag{8.60}$$

One writes $K_{a_1 \cdots a_n, \varepsilon}(p)$ to denote the set of all such strings $b_1 \cdots b_n \in \Gamma^n$.

When a joint probability vector $p \in \mathcal{P}(\Sigma \times \Gamma)$ is fixed, or can safely be taken as being implicit, the notation $K_{a_1 \cdots a_n, \varepsilon}$ may be used in place of $K_{a_1 \cdots a_n, \varepsilon}(p)$. It is also convenient to define $K_{a_1 \cdots a_n, \varepsilon}(p) = \varnothing$ for any string $a_1 \cdots a_n \in \Sigma^n$ for which $q(a_1) \cdots q(a_n) = 0$.

Intuitively speaking, if one were to select strings $a_1 \cdots a_n \in \Sigma^n$ and $b_1 \cdots b_n \in \Gamma^n$ by choosing the pairs $(a_1, b_1), \ldots, (a_n, b_n)$ independently at random according to a probability vector $p \in \mathcal{P}(\Sigma \times \Gamma)$, then it is reasonable to expect that $b_1 \cdots b_n$ will be contained in $K_{a_1 \cdots a_n, \varepsilon}(p)$, with the probability of this event becoming increasingly likely as n becomes large. This fact is established by the following proposition, based on the weak law of large numbers (Theorem 1.15)—the methodology is essentially the same as the analogous fact (Proposition 5.45) that was proved in regard to the standard definition of typicality discussed in Section 5.3.1.

Proposition 8.21. *Let Σ and Γ be alphabets, let $p \in \mathcal{P}(\Sigma \times \Gamma)$ be a probability vector, and let $q \in \mathcal{P}(\Sigma)$ be the marginal probability vector defined as*

$$q(a) = \sum_{b \in \Gamma} p(a, b) \tag{8.61}$$

for each $a \in \Sigma$. For every $\varepsilon > 0$ it holds that

$$\lim_{n \to \infty} \sum_{a_1 \cdots a_n \in \Sigma^n} \sum_{b_1 \cdots b_n \in K_{a_1 \cdots a_n, \varepsilon}} p(a_1, b_1) \cdots p(a_n, b_n) = 1. \tag{8.62}$$

Proof. Let n be a positive integer and let $X_1, \ldots, X_n$ be independent and identically distributed random variables, defined as

$$X_k(a, b) = -\log(p(a, b)) + \log(q(a)) \tag{8.63}$$

for each pair $(a, b) \in \Sigma \times \Gamma$, and distributed with respect to p. One has that the expected value $\alpha = \mathrm{E}(X_k)$ of each of these random variables is given by $\alpha = \mathrm{H}(p) - \mathrm{H}(q)$, and furthermore

$$\begin{aligned} &\Pr\left(\left| \frac{X_1 + \cdots + X_n}{n} - \alpha \right| < \varepsilon \right) \\ &\qquad = \sum_{a_1 \cdots a_n \in \Sigma^n} \sum_{b_1 \cdots b_n \in K_{a_1 \cdots a_n, \varepsilon}} p(a_1, b_1) \cdots p(a_n, b_n). \end{aligned} \tag{8.64}$$

The conclusion of the proposition therefore follows from the weak law of large numbers (Theorem 1.15). □

The next proposition places an upper bound on the expected size of the set $K_{a_1 \cdots a_n, \varepsilon}$. It is analogous to Proposition 5.46 for the standard definition of typicality.

Proposition 8.22. *Let Σ and Γ be alphabets, let $p \in \mathcal{P}(\Sigma \times \Gamma)$ be a probability vector, and let $q \in \mathcal{P}(\Sigma)$ be the marginal probability vector defined as*

$$q(a) = \sum_{b \in \Gamma} p(a, b) \tag{8.65}$$

for each $a \in \Sigma$. For every positive integer n and every positive real number $\varepsilon > 0$, it holds that

$$\sum_{a_1 \cdots a_n \in \Sigma^n} q(a_1) \cdots q(a_n) \left| K_{a_1 \cdots a_n, \varepsilon} \right| < 2^{n(\mathrm{H}(p) - \mathrm{H}(q) + \varepsilon)}. \tag{8.66}$$

Proof. For each string $a_1 \cdots a_n \in \Sigma^n$ satisfying $q(a_1) \cdots q(a_n) > 0$ and each string $b_1 \cdots b_n \in K_{a_1 \cdots a_n, \varepsilon}$, one has

$$2^{-n(\mathrm{H}(p) - \mathrm{H}(q) + \varepsilon)} < \frac{p(a_1, b_1) \cdots p(a_n, b_n)}{q(a_1) \cdots q(a_n)}, \tag{8.67}$$

and therefore

$$\begin{aligned} & 2^{-n(\mathrm{H}(p) - \mathrm{H}(q) + \varepsilon)} \sum_{a_1 \cdots a_n \in \Sigma^n} q(a_1) \cdots q(a_n) |K_{a_1 \cdots a_n, \varepsilon}| \\ & \quad < \sum_{a_1 \cdots a_n \in \Sigma^n} \sum_{b_1 \cdots b_n \in K_{a_1 \cdots a_n, \varepsilon}} p(a_1, b_1) \cdots p(a_n, b_n) \leq 1, \end{aligned} \tag{8.68}$$

from which the proposition follows. □

The notion of typicality for joint probability distributions established by Definition 8.20 may be extended to ensembles of quantum states in a fairly straightforward fashion, using spectral decompositions of the states in an ensemble.

Definition 8.23. Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble of states, and let Γ be an alphabet such that $|\Gamma| = \dim(\mathcal{X})$. By the spectral theorem (as stated by Corollary 1.4), one may write

$$\eta(a) = \sum_{b \in \Gamma} p(a, b) u_{a,b} u_{a,b}^* \tag{8.69}$$

for $p \in \mathcal{P}(\Sigma \times \Gamma)$ being a probability vector and $\{u_{a,b} \,:\, b \in \Gamma\}$ being an orthonormal basis of $\mathcal{X}$ for each $a \in \Sigma$. With respect to the ensemble η, and for each positive real number $\varepsilon > 0$, each positive integer n, and each string $a_1 \cdots a_n \in \Sigma^n$, the *projection onto the ε-typical subspace of $\mathcal{X}^{\otimes n}$ conditioned on* $a_1 \cdots a_n$ is defined as

$$\Lambda_{a_1 \cdots a_n, \varepsilon} = \sum_{b_1 \cdots b_n \in K_{a_1 \cdots a_n, \varepsilon}} u_{a_1,b_1} u_{a_1,b_1}^* \otimes \cdots \otimes u_{a_n,b_n} u_{a_n,b_n}^*. \tag{8.70}$$

Remark 8.24. For a fixed choice of a string $a_1 \cdots a_n \in \Sigma^n$, one has that the inclusion of each string $b_1 \cdots b_n$ in $K_{a_1 \cdots a_n, \varepsilon}$ is determined by the multiset of values $\{p(a_1, b_1), \ldots, p(a_n, b_n)\}$ alone. Thus, the same is true regarding the inclusion of each rank-one projection in the summation (8.70). It follows that the projection $\Lambda_{a_1 \cdots a_n, \varepsilon}$ specified by Definition 8.23 is uniquely defined by the ensemble η, and is independent of the particular choices of the spectral decompositions (8.69).

Facts analogous to the previous two propositions, holding for ensembles rather than joint probability distributions, follow directly.

Proposition 8.25. *Let Σ be an alphabet, let $\mathcal{X}$ be a complex Euclidean space, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble of states. For every $\varepsilon > 0$, it holds that*

$$\lim_{n \to \infty} \sum_{a_1 \cdots a_n \in \Sigma^n} \big\langle \Lambda_{a_1 \cdots a_n, \varepsilon}, \eta(a_1) \otimes \cdots \otimes \eta(a_n) \big\rangle = 1, \tag{8.71}$$

where, for each positive integer n, and each string $a_1 \cdots a_n \in \Sigma^n$, $\Lambda_{a_1 \cdots a_n, \varepsilon}$ is the projection onto the ε-typical subspace of $\mathcal{X}^{\otimes n}$ conditioned on $a_1 \cdots a_n$, with respect to the ensemble η. Moreover, one has

$$\sum_{a_1 \cdots a_n \in \Sigma^n} \mathrm{Tr}(\eta(a_1)) \cdots \mathrm{Tr}(\eta(a_n)) \; \mathrm{Tr}\big(\Lambda_{a_1 \cdots a_n, \varepsilon}\big) < 2^{n(\beta + \varepsilon)} \tag{8.72}$$

for

$$\beta = \sum_{\substack{a \in \Sigma \\ \eta(a) \neq 0}} \mathrm{Tr}(\eta(a)) \; \mathrm{H}\bigg(\frac{\eta(a)}{\mathrm{Tr}(\eta(a))} \bigg). \tag{8.73}$$

Proof. Assume that

$$\eta(a) = \sum_{b \in \Gamma} p(a,b) \, u_{a,b} u_{a,b}^* \tag{8.74}$$

is a spectral decomposition of $\eta(a)$ for each $a \in \Sigma$, and define $q \in \mathcal{P}(\Sigma)$ as $q(a) = \sum_{b\in\Gamma} p(a,b)$ for each $a \in \Sigma$ (which is equivalent to $q(a) = \mathrm{Tr}(\eta(a))$ for each $a \in \Sigma$). For each positive integer n, each positive real number $\varepsilon > 0$, and each string $a_1 \cdots a_n \in \Sigma^n$, one has

$$\begin{aligned}\big\langle \Lambda_{a_1\cdots a_n,\varepsilon}, \eta(a_1) \otimes \cdots \otimes \eta(a_n) \big\rangle \\ = \sum_{b_1\cdots b_n \in K_{a_1\cdots a_n,\varepsilon}} p(a_1,b_1)\cdots p(a_n,b_n),\end{aligned} \tag{8.75}$$

and moreover

$$\beta = \mathrm{H}(p) - \mathrm{H}(q) \quad \text{and} \quad \mathrm{Tr}\big(\Lambda_{a_1\cdots a_n,\varepsilon}\big) = \big|K_{a_1\cdots a_n,\varepsilon}\big|. \tag{8.76}$$

The proposition therefore follows from Propositions 8.21 and 8.22. □

A useful operator inequality

When analyzing the performance of classical-to-quantum product state channel codes, it is helpful to make use of the operator inequality to be stated as Lemma 8.28 below. The proof of this inequality will make use of the following fact regarding square roots of positive semidefinite operators.

Lemma 8.26 (Operator monotonicity of the square root)**.** *Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\sqrt{P} \leq \sqrt{P+Q}. \tag{8.77}$$

Proof. The block operator

$$\begin{pmatrix} P & \sqrt{P} \\ \sqrt{P} & \mathbb{1} \end{pmatrix} + \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} P+Q & \sqrt{P} \\ \sqrt{P} & \mathbb{1} \end{pmatrix} \tag{8.78}$$

is positive semidefinite. As $[P+Q, \mathbb{1}] = 0$ and $\sqrt{P}$ is Hermitian, it follows by Lemma 5.31 that

$$\sqrt{P} \leq \sqrt{P+Q}\sqrt{\mathbb{1}} = \sqrt{P+Q}, \tag{8.79}$$

as required. □

Remark 8.27. It is not difficult to prove Lemma 8.26 directly, without relying on Lemma 5.31, by using spectral properties of operators that were also employed in the proof of that lemma.

Lemma 8.28 (Hayashi–Nagaoka). *Let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators, for $\mathcal{X}$ being a complex Euclidean space, and assume $P \leq \mathbb{1}$. It holds that*

$$\mathbb{1} - \sqrt{(P+Q)^+}\, P \sqrt{(P+Q)^+} \leq 2(\mathbb{1} - P) + 4Q. \tag{8.80}$$

Proof. For every choice of operators $A, B \in \mathrm{L}(\mathcal{X})$, one has

$$0 \leq (A-B)(A-B)^* = AA^* + BB^* - (AB^* + BA^*), \tag{8.81}$$

and therefore $AB^* + BA^* \leq AA^* + BB^*$. Setting

$$A = X\sqrt{Q} \quad \text{and} \quad B = (\mathbb{1} - X)\sqrt{Q}, \tag{8.82}$$

for a given operator $X \in \mathrm{L}(\mathcal{X})$, yields

$$XQ(\mathbb{1} - X)^* + (\mathbb{1} - X)QX^* \leq XQX^* + (\mathbb{1} - X)Q(\mathbb{1} - X)^*, \tag{8.83}$$

and therefore

$$\begin{aligned} Q &= XQX^* + XQ(\mathbb{1} - X)^* + (\mathbb{1} - X)QX^* + (\mathbb{1} - X)Q(\mathbb{1} - X)^* \\ &\leq 2XQX^* + 2(\mathbb{1} - X)Q(\mathbb{1} - X)^*. \end{aligned} \tag{8.84}$$

For the specific choice $X = \sqrt{P+Q}$, one obtains

$$Q \leq 2\sqrt{P+Q}\, Q \sqrt{P+Q} + 2\Big(\mathbb{1} - \sqrt{P+Q}\Big)\, Q\, \Big(\mathbb{1} - \sqrt{P+Q}\Big), \tag{8.85}$$

and from the observation that $Q \leq P + Q$ it follows that

$$\begin{aligned} Q &\leq 2\sqrt{P+Q}\, Q \sqrt{P+Q} \\ &\qquad + 2\Big(\mathbb{1} - \sqrt{P+Q}\Big)\, (P+Q)\, \Big(\mathbb{1} - \sqrt{P+Q}\Big) \\ &= \sqrt{P+Q}\Big(2\mathbb{1} + 4Q - 4\sqrt{P+Q} + 2P\Big)\sqrt{P+Q}. \end{aligned} \tag{8.86}$$

Using the fact that $P \leq \mathbb{1}$ together with Lemma 8.26, one has

$$P \leq \sqrt{P} \leq \sqrt{P+Q}, \tag{8.87}$$

and therefore

$$Q \leq \sqrt{P+Q}\Big(2\mathbb{1} - 2P + 4Q\Big)\sqrt{P+Q}. \tag{8.88}$$

Conjugating both sides of this inequality by the Moore–Penrose pseudo-inverse of $\sqrt{P+Q}$ yields

$$\sqrt{(P+Q)^+}\, Q \sqrt{(P+Q)^+} \leq 2\Pi_{\mathrm{im}(P+Q)} - 2P + 4Q. \tag{8.89}$$

It follows that

$$\begin{aligned} &\mathbb{1} - \sqrt{(P+Q)^+}\, P \sqrt{(P+Q)^+} \\ &\qquad = \mathbb{1} - \Pi_{\mathrm{im}(P+Q)} + \sqrt{(P+Q)^+}\, Q \sqrt{(P+Q)^+} \\ &\qquad \leq \mathbb{1} + \Pi_{\mathrm{im}(P+Q)} - 2P + 4Q \\ &\qquad \leq 2(\mathbb{1} - P) + 4Q, \end{aligned} \tag{8.90}$$

as required. □

An existence proof for classical-to-quantum product state channel codes

Returning to the discussion of classical-to-quantum product-state channel codes, assume as before that an alphabet Σ, a complex Euclidean space $\mathcal{X}$, and a set of states

$$\{\sigma_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X}) \tag{8.91}$$

has been fixed, and let $\Gamma = \{0,1\}$ denote the binary alphabet. It is natural to ask, for any choice of a positive real number $\delta > 0$ and positive integers m and n, whether or not there exists a classical-to-quantum product-state channel code (f,μ) taking the form

$$f : \Gamma^m \to \Sigma^n \quad \text{and} \quad \mu : \Gamma^m \to \mathrm{Pos}(\mathcal{X}^{\otimes m}) \tag{8.92}$$

and having error bounded by δ.

In general, one may expect that making such a determination is not tractable from a computational point of view. It is possible, however, to prove the existence of reasonably good classical-to-quantum product-state channel codes through the probabilistic method: for suitable choices of n, m, and δ, a *random* choice of a function $f : \Gamma^m \to \Sigma^n$ and a well-chosen measurement $\mu : \Gamma^m \to \mathrm{Pos}(\mathcal{X}^{\otimes n})$ are considered, and a coding scheme having error bounded by δ is obtained with a nonzero probability. The next theorem gives a precise statement regarding the parameters n, m, and δ through which this methodology proves the existence of quantum-to-classical product-state channels codes.

Theorem 8.29. *Let Σ be an alphabet, let $\mathcal{X}$ be a complex Euclidean space, let*

$$\{\sigma_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X}) \tag{8.93}$$

be a collection of states, and let $\Gamma = \{0,1\}$ denote the binary alphabet. Also let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be the ensemble defined as

$$\eta(a) = p(a)\sigma_a \tag{8.94}$$

for each $a \in \Sigma$, let α be a positive real number satisfying $\alpha < \chi(\eta)$, and let $\delta > 0$ be a positive real number. For all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, there exists a function $f : \Gamma^m \to \Sigma^n$ and a measurement $\mu : \Gamma^m \to \mathrm{Pos}(\mathcal{X}^{\otimes n})$ such that

$$\big\langle \mu(b_1 \cdots b_m), \sigma_{f(b_1 \cdots b_m)} \big\rangle > 1 - \delta \tag{8.95}$$

for every $b_1 \cdots b_m \in \Gamma^m$.

Proof. It will first be assumed that n and m are arbitrary positive integers. As suggested previously, the proof makes use of the probabilistic method: a random function $g : \Gamma^{m+1} \to \Sigma^n$ is chosen from a particular probability distribution, a decoding measurement μ is defined for each possible choice of g, and the expected probability of a decoding error for the pair (g, μ) is analyzed. As is to be explained later in the proof, this analysis implies the existence of a channel coding scheme (f, μ), where $f : \Gamma^m \to \Sigma^n$ is derived from g, satisfying the requirements theorem for all but finitely many n and for $m \leq \alpha n$.

The particular distribution from which g is to be chosen is one in which each individual output symbol of g is selected independently according to the probability vector p. Equivalently, for a random selection of g according to the distribution being described, one has that

$$\Pr\big(g(b_1 \cdots b_{m+1}) = a_1 \cdots a_n\big) = p(a_1) \cdots p(a_n) \tag{8.96}$$

for every choice of $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$ and $a_1 \cdots a_n \in \Sigma^n$, and moreover the outputs of a randomly chosen g on distinct choices of the input string $b_1 \cdots b_{m+1}$ are uncorrelated.

The specification of the decoding measurement μ that is to be associated with a given g is not chosen randomly—a unique measurement is defined

for each g in a way that is dependent upon the ensemble η. First, let $\varepsilon > 0$ be a sufficiently small positive real number such that the inequality

$$\alpha < \chi(\eta) - 3\varepsilon \tag{8.97}$$

holds. For each string $a_1 \cdots a_n \in \Sigma^n$, let $\Lambda_{a_1 \cdots a_n}$ denote the projection onto the ε-typical subspace of $\mathcal{X}^{\otimes n}$ conditioned on $a_1 \cdots a_n$, with respect to the ensemble η, and let Π_n be the projection onto the ε-typical subspace of $\mathcal{X}^{\otimes n}$ with respect to the average state

$$\sigma = \sum_{a \in \Sigma} p(a) \sigma_a \tag{8.98}$$

of the ensemble η. (As ε has been fixed, the dependence of $\Lambda_{a_1 \cdots a_n}$ and Π_n on ε is not written explicitly, allowing for slightly less cluttered equations.) Next, for a given choice of $g : \Gamma^{m+1} \to \Sigma^n$, define

$$Q = \sum_{b_1 \cdots b_{m+1} \in \Gamma^{m+1}} \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n \,, \tag{8.99}$$

and, for each binary string $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$, define

$$Q_{b_1 \cdots b_{m+1}} = \sqrt{Q^+}\, \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n \sqrt{Q^+}. \tag{8.100}$$

One has that each operator $Q_{b_1 \cdots b_{m+1}}$ is positive semidefinite, and moreover

$$\sum_{b_1 \cdots b_{m+1} \in \Gamma^{m+1}} Q_{b_1 \cdots b_{m+1}} = \Pi_{\mathrm{im}(Q)}. \tag{8.101}$$

Finally, the measurement $\mu : \Gamma^{m+1} \to \mathrm{Pos}\big(\mathcal{X}^{\otimes n}\big)$ to be associated with g is defined as

$$\mu(b_1 \cdots b_{m+1}) = Q_{b_1 \cdots b_{m+1}} + \frac{1}{2^{m+1}} \big(\mathbb{1} - \Pi_{\mathrm{im}(Q)}\big) \tag{8.102}$$

for each $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$.

For each choice of g, the probability that the measurement μ associated with g errs in recovering a string $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$ from its encoding is equal to

$$\big\langle \mathbb{1} - \mu(b_1 \cdots b_{m+1}), \sigma_{g(b_1 \cdots b_{m+1})} \big\rangle. \tag{8.103}$$

The next phase of the proof establishes a lower-bound on the average error probability

$$\frac{1}{2^{m+1}} \sum_{b_1 \cdots b_{m+1} \in \Gamma^{m+1}} \big\langle \mathbb{1} - \mu(b_1 \cdots b_{m+1}), \sigma_{g(b_1 \cdots b_{m+1})} \big\rangle, \tag{8.104}$$

for a uniformly chosen string $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$. To bound this average probability of error, one may first observe that Lemma 8.28 implies that

$$\begin{aligned} &\mathbb{1} - Q_{b_1 \cdots b_{m+1}} \\ &\quad \leq 2\big(\mathbb{1} - \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n\big) + 4\big(Q - \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n\big) \end{aligned} \tag{8.105}$$

for each $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$. For a fixed choice of g, the probability of an error in recovering a given string $b_1 \cdots b_{m+1}$ is therefore upper-bounded by

$$\begin{aligned} &2\big\langle \mathbb{1} - \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n, \sigma_{g(b_1 \cdots b_{m+1})} \big\rangle \\ &\quad + 4\big\langle Q - \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n, \sigma_{g(b_1 \cdots b_{m+1})} \big\rangle. \end{aligned} \tag{8.106}$$

The expected value of this expression will be shown to be small, under the additional assumption that $m \leq \alpha n$, when $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$ is chosen uniformly and g is chosen according to the distribution described above.

The first term in the expression (8.106) will be considered first. To prove an upper bound on the expected value of this quantity, it is convenient to make use of the operator identity

$$ABA = AB + BA - B + (\mathbb{1} - A)B(\mathbb{1} - A). \tag{8.107}$$

In particular, for any choice of a string $a_1 \cdots a_n \in \Sigma^n$, this identity implies

$$\begin{aligned} &\big\langle \Pi_n \Lambda_{a_1 \cdots a_n} \Pi_n, \sigma_{a_1 \cdots a_n} \big\rangle \\ &\quad = \big\langle \Pi_n \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \big\rangle + \big\langle \Lambda_{a_1 \cdots a_n} \Pi_n, \sigma_{a_1 \cdots a_n} \big\rangle - \big\langle \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \big\rangle \\ &\qquad + \big\langle (\mathbb{1} - \Pi_n) \Lambda_{a_1 \cdots a_n} (\mathbb{1} - \Pi_n), \sigma_{a_1 \cdots a_n} \big\rangle \\ &\quad \geq \big\langle \Pi_n \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \big\rangle + \big\langle \Lambda_{a_1 \cdots a_n} \Pi_n, \sigma_{a_1 \cdots a_n} \big\rangle - \big\langle \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \big\rangle. \end{aligned} \tag{8.108}$$

As $\Lambda_{a_1 \cdots a_n}$ is a projection operator and commutes with $\sigma_{a_1 \cdots a_n}$, it follows that

$$\begin{aligned} &\big\langle \Pi_n \Lambda_{a_1 \cdots a_n} \Pi_n, \sigma_{a_1 \cdots a_n} \big\rangle \\ &\quad \geq \big\langle 2\Pi_n - \mathbb{1}, \Lambda_{a_1 \cdots a_n} \sigma_{a_1 \cdots a_n} \big\rangle \\ &\quad = \big\langle 2\Pi_n - \mathbb{1}, \sigma_{a_1 \cdots a_n} \big\rangle + \big\langle \mathbb{1} - 2\Pi_n, (\mathbb{1} - \Lambda_{a_1 \cdots a_n}) \sigma_{a_1 \cdots a_n} \big\rangle \\ &\quad \geq \big\langle 2\Pi_n - \mathbb{1}, \sigma_{a_1 \cdots a_n} \big\rangle - \big\langle \mathbb{1} - \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \big\rangle \\ &\quad = 2\big\langle \Pi_n, \sigma_{a_1 \cdots a_n} \big\rangle + \big\langle \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \big\rangle - 2. \end{aligned} \tag{8.109}$$

Averaging over all choices of $a_1 \cdots a_n \in \Sigma^n$, with each a_k being selected independently according to the probability vector p, one has that

$$\begin{aligned}&\sum_{a_1\cdots a_n\in\Sigma^n} p(a_1)\cdots p(a_n)\big\langle \Pi_n \Lambda_{a_1\cdots a_n}\Pi_n, \sigma_{a_1\cdots a_n}\big\rangle \\ &\quad\geq 2\big\langle \Pi_n, \sigma^{\otimes n}\big\rangle + \sum_{a_1\cdots a_n\in\Sigma^n} p(a_1)\cdots p(a_n)\big\langle \Lambda_{a_1\cdots a_n}, \sigma_{a_1\cdots a_n}\big\rangle - 2.\end{aligned} \tag{8.110}$$

The right-hand side of the expression (8.110) approaches 1 in the limit as n goes to infinity by Propositions 5.45 and 8.21, from which it follows that

$$\sum_{a_1\cdots a_n\in\Sigma^n} p(a_1)\cdots p(a_n)\big\langle \mathbb{1} - \Pi_n \Lambda_{a_1\cdots a_n}\Pi_n, \sigma_{a_1\cdots a_n}\big\rangle < \frac{\delta}{8} \tag{8.111}$$

for all but finitely many choices of a positive integer n. For any n for which the inequality (8.111) holds, and for a random selection of $g : \Gamma^{m+1} \to \Sigma^n$ as described above, it therefore holds that the expected value of the expression

$$2\big\langle \mathbb{1} - \Pi_n \Lambda_{g(b_1\cdots b_{m+1})}\Pi_n, \sigma_{g(b_1\cdots b_{m+1})}\big\rangle \tag{8.112}$$

is at most $\delta/4$ for an arbitrary choice of $b_1 \cdots b_{m+1}$, and therefore the same bound holds for a uniformly selected binary string $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$.

The second term in the expression (8.106) will be considered next. It may first be observed that

$$Q - \Pi_n \Lambda_{g(b_1\cdots b_{m+1})}\Pi_n = \sum_{\substack{c_1\cdots c_{m+1}\in\Gamma^{m+1}\\ c_1\cdots c_{m+1}\neq b_1\cdots b_{m+1}}} \Pi_n \Lambda_{g(c_1\cdots c_{m+1})}\Pi_n, \tag{8.113}$$

so that

$$\begin{aligned}&\big\langle Q - \Pi_n \Lambda_{g(b_1\cdots b_{m+1})}\Pi_n, \sigma_{g(b_1\cdots b_{m+1})}\big\rangle \\ &\quad= \sum_{\substack{c_1\cdots c_{m+1}\in\Gamma^{m+1}\\ c_1\cdots c_{m+1}\neq b_1\cdots b_{m+1}}} \big\langle \Pi_n \Lambda_{g(c_1\cdots c_{m+1})}\Pi_n, \sigma_{g(b_1\cdots b_{m+1})}\big\rangle.\end{aligned} \tag{8.114}$$

The value of the function g on each input string is chosen independently according to the probability vector $p^{\otimes n}$, so there is no correlation between $g(b_1 \cdots b_{m+1})$ and $g(c_1 \cdots c_{m+1})$ for $b_1 \cdots b_{m+1} \neq c_1 \cdots c_{m+1}$. It follows that the expected value of the above expression is given by

$$\big(2^{m+1} - 1\big) \sum_{a_1\cdots a_n\in\Sigma^n} p(a_1)\cdots p(a_n)\big\langle \Lambda_{a_1\cdots a_n}, \Pi_n \sigma^{\otimes n}\Pi_n\big\rangle. \tag{8.115}$$

By Proposition 8.25 it holds that

$$\sum_{a_1\cdots a_n\in\Sigma^n} p(a_1)\cdots p(a_n)\,\mathrm{Tr}\big(\Lambda_{a_1\cdots a_n}\big) \leq 2^{n(\beta+\varepsilon)} \tag{8.116}$$

for

$$\beta = \sum_{a\in\Sigma} p(a)\,\mathrm{H}(\sigma_a), \tag{8.117}$$

and by the definition of Π_n one has that

$$\lambda_1\big(\Pi_n\sigma^{\otimes n}\Pi_n\big) \leq 2^{-n(\mathrm{H}(\sigma)-\varepsilon)}. \tag{8.118}$$

It follows that

$$\begin{aligned}\big(2^{m+1}-1\big)\sum_{a_1\cdots a_n\in\Sigma^n} p(a_1)\cdots p(a_n)\big\langle\Lambda_{a_1\cdots a_n},\Pi_n\sigma^{\otimes n}\Pi_n\big\rangle \\ \leq 2^{m+1-n(\chi(\eta)-2\varepsilon)},\end{aligned} \tag{8.119}$$

so that the expected value of the second term in the expression (8.106) is upper-bounded by

$$2^{m-n(\chi(\eta)-2\varepsilon)+3}. \tag{8.120}$$

Now assume that $m \leq \alpha n$. For $g : \Gamma^{m+1} \to \Sigma^n$ chosen according to the distribution specified earlier and $b_1\cdots b_{m+1} \in \Gamma^{m+1}$ chosen uniformly, one has that the expected value of the error probability (8.104) is at most

$$\frac{\delta}{4} + 2^{\alpha n-n(\chi(\eta)-2\varepsilon)+3} \leq \frac{\delta}{4} + 2^{-\varepsilon n+3} \tag{8.121}$$

for all but finitely many choices of n. As

$$2^{-\varepsilon n} < \frac{\delta}{32} \tag{8.122}$$

for all sufficiently large n, it follows that the expected value of the error probability (8.104) is smaller than $\delta/2$ for all but finitely many choices of n. For all but finitely many choices of n, there must therefore exist at least one choice of a function $g : \Gamma^{m+1} \to \Sigma^m$ such that, for μ being the measurement associated with g, it holds that

$$\frac{1}{2^{m+1}}\sum_{b_1\cdots b_{m+1}\in\Gamma^{m+1}} \big\langle \mathbb{1} - \mu(b_1\cdots b_{m+1}), \sigma_{g(b_1\cdots b_{m+1})}\big\rangle < \frac{\delta}{2}. \tag{8.123}$$

Finally, for a given choice of n, $m \leq \alpha n$, g, and μ for which the bound (8.123) holds, consider the set

$$B = \Big\{ b_1 \cdots b_{m+1} \in \Gamma^{m+1} \,:\, \big\langle \mathbb{1} - \mu(b_1 \cdots b_{m+1}), \sigma_{g(b_1 \cdots b_{m+1})} \big\rangle \geq \delta \Big\} \tag{8.124}$$

of all strings whose encodings incur a decoding error with probability at least δ. It holds that

$$\frac{\delta |B|}{2^{m+1}} < \frac{\delta}{2}, \tag{8.125}$$

and therefore $|B| \leq 2^m$. By defining a function $f : \Gamma^m \to \Sigma^n$ as $f = gh$ for any one-to-one function $h : \Gamma^m \to \Gamma^{m+1} \backslash B$, one has that

$$\big\langle \mu(b_1 \cdots b_m), \sigma_{f(b_1 \cdots b_m)} \big\rangle > 1 - \delta \tag{8.126}$$

for every choice of $b_1 \cdots b_m \in \Gamma^m$, which completes the proof. □

Statement and proof of the Holevo–Schumacher–Westmoreland theorem

The Holevo–Schumacher–Westmoreland theorem will now be stated, and proved through the use of Theorem 8.29.

Theorem 8.30 (Holevo–Schumacher–Westmoreland theorem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. It holds that the classical capacity of Φ is equal to its regularized Holevo capacity:*

$$\mathrm{C}(\Phi) = \lim_{n \to \infty} \frac{\chi\big(\Phi^{\otimes n}\big)}{n}. \tag{8.127}$$

Proof. It will first be observed that the inequality

$$\mathrm{C}(\Phi) \geq \chi(\Phi) \tag{8.128}$$

follows from Theorem 8.29. Let Σ be any alphabet, let

$$\{\rho_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X}) \tag{8.129}$$

be a collection of states, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and define an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ as

$$\eta(a) = p(a)\rho_a \tag{8.130}$$

for each $a \in \Sigma$. Also set $\mathcal{Z} = \mathbb{C}^\Gamma$ for $\Gamma = \{0,1\}$ being the binary alphabet. As η is an arbitrarily chosen ensemble, the inequality (8.128) will follow from a demonstration that every positive real number less than $\chi(\Phi(\eta))$ is an achievable rate for classical information transmission through Φ.

Fix any choice of $\alpha > 0$ satisfying

$$\alpha < \chi(\Phi(\eta)), \tag{8.131}$$

and define $\sigma_a = \Phi(\rho_a)$ for each $a \in \Sigma$. By Theorem 8.29, the following statement holds: for every positive real number $\varepsilon > 0$, for all but finitely many choices of a positive integer n, and for all positive integers $m \leq \alpha n$, there exist a classical-to-quantum product state channel code (f, μ) for the collection

$$\{\sigma_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{Y}) \tag{8.132}$$

for which the probability of an error is strictly less than $\varepsilon/2$ on every binary string of length m.

With this fact in mind, for any fixed choice of positive integers n and m for which $m \leq \alpha n$, one may define encoding and decoding channels

$$\Xi_{\mathrm{E}} \in \mathrm{C}\big(\mathcal{Z}^{\otimes m}, \mathcal{X}^{\otimes n}\big) \quad \text{and} \quad \Xi_{\mathrm{D}} \in \mathrm{C}\big(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m}\big) \tag{8.133}$$

as follows:

$$\begin{aligned} \Xi_{\mathrm{E}}(Z) &= \sum_{b_1 \cdots b_m \in \Gamma^m} \big\langle E_{b_1 \cdots b_m, b_1 \cdots b_m}, Z \big\rangle \rho_{f(b_1 \cdots b_m)}, \\ \Xi_{\mathrm{D}}(Y) &= \sum_{b_1 \cdots b_m \in \Gamma^m} \big\langle \mu(b_1 \cdots b_m), Y \big\rangle E_{b_1 \cdots b_m, b_1 \cdots b_m}, \end{aligned} \tag{8.134}$$

for all $Z \in \mathrm{L}(\mathcal{Z}^{\otimes m})$ and $Y \in \mathrm{L}(\mathcal{Y}^{\otimes n})$. For n being sufficiently large, one has that

$$\big\langle E_{b_1 \cdots b_m, b_1 \cdots b_m}, \big(\Xi_{\mathrm{D}} \Phi^{\otimes n} \Xi_{\mathrm{E}}\big)(E_{b_1 \cdots b_m, b_1 \cdots b_m}) \big\rangle > 1 - \frac{\varepsilon}{2} \tag{8.135}$$

for every $b_1 \cdots b_m \in \Gamma^m$. As Ξ_{E} is a classical-to-quantum channel and Ξ_{D} is quantum-to-classical, it follows that $\Xi_{\mathrm{D}} \Phi^{\otimes n} \Xi_{\mathrm{E}}$ is a ε-approximation to the completely dephasing channel $\Delta^{\otimes m} \in \mathrm{C}(\mathcal{Z}^{\otimes m})$. It has been demonstrated that α is an achievable rate for classical information transmission through Φ.

Following the same reasoning, except replacing the channel Φ by the channel $\Phi^{\otimes n}$, one finds that

$$\chi(\Phi^{\otimes n}) \leq \mathrm{C}(\Phi^{\otimes n}) = n\,\mathrm{C}(\Phi) \tag{8.136}$$

for every positive integer n, and therefore

$$\mathrm{C}(\Phi) \geq \lim_{n\to\infty} \frac{\chi\big(\Phi^{\otimes n}\big)}{n}. \tag{8.137}$$

It remains to prove that the classical capacity of Φ is no larger than its regularized Holevo capacity. There is nothing to prove if $\mathrm{C}(\Phi) = 0$, so it will be assumed that $\mathrm{C}(\Phi) > 0$. Suppose that $\alpha > 0$ is an achievable rate for classical information transmission through Φ, and let $\varepsilon > 0$ be chosen arbitrarily. It must therefore hold, for all but finitely many positive integers n, and for all $m \leq \alpha n$, that $\Phi^{\otimes n}$ emulates an ε-approximation to the completely dephasing channel $\Delta^{\otimes m} \in \mathrm{C}(\mathcal{Z}^{\otimes m})$.

Let n be any positive integer for which this property holds and for which $\lfloor \alpha n \rfloor \geq 2$, and let $m = \lfloor \alpha n \rfloor$. The situation in which a sender generates a binary string of length m, uniformly at random, and transmits this string through the ε-approximation to $\Delta^{\otimes m}$ emulated by $\Phi^{\otimes n}$ will be considered. Let X and Z be classical registers both having state set Γ^m, where X is a register representing the randomly generated string selected by the sender and Z is a register representing the string obtained by the receiver when a copy of the string stored in X is transmitted through the ε-approximation to $\Delta^{\otimes m}$ emulated by $\Phi^{\otimes n}$. As $\Phi^{\otimes n}$ emulates an ε-approximation to $\Delta^{\otimes m}$, there must exist a collection of states

$$\big\{\rho_{b_1\cdots b_m} \,:\, b_1 \cdots b_m \in \Gamma^m\big\} \subseteq \mathrm{D}\big(\mathcal{X}^{\otimes n}\big) \tag{8.138}$$

along with a measurement $\mu : \Gamma^m \to \mathrm{Pos}\big(\mathcal{Y}^{\otimes n}\big)$ such that

$$\big\langle \mu(b_1 \cdots b_m), \Phi^{\otimes n}(\rho_{b_1\cdots b_m})\big\rangle \geq 1 - \frac{\varepsilon}{2} \tag{8.139}$$

for every binary string $b_1 \cdots b_m \in \Gamma^m$. With respect to the probability vector $p \in \mathcal{P}(\Gamma^m \times \Gamma^m)$ defined as

$$p(b_1 \cdots b_m, c_1 \cdots c_m) = \frac{1}{2^m}\big\langle \mu(c_1 \cdots c_m), \Phi^{\otimes n}(\rho_{b_1\cdots b_m})\big\rangle, \tag{8.140}$$

which represents the probabilistic state of (X, Z) suggested above, it follows from Holevo's theorem (Theorem 5.52) that

$$\mathrm{I}(\mathsf{X} : \mathsf{Z}) \leq \chi(\Phi^{\otimes n}(\eta)), \tag{8.141}$$

where $\eta : \Gamma^m \to \mathrm{Pos}(\mathcal{X}^{\otimes n})$ is the ensemble defined as

$$\eta(b_1 \cdots b_m) = \frac{1}{2^m} \rho_{b_1 \cdots b_m} \tag{8.142}$$

for each $b_1 \cdots b_m \in \Gamma^m$.

A lower-bound on the mutual information $\mathrm{I}(\mathsf{X} : \mathsf{Z})$ may be derived as follows. The distribution represented by the marginal probability vector $p[\mathsf{X}]$ is uniform, and therefore $\mathrm{H}(p[\mathsf{X}]) = m$. By (8.139), each entry of the probability vector $p[\mathsf{Z}]$ is lower-bounded by $(1 - \varepsilon/2)2^{-m}$, so the lower bound

$$\mathrm{H}(p[\mathsf{Z}]) \geq \Big(1 - \frac{\varepsilon}{2}\Big) m \tag{8.143}$$

follows by the concavity of the Shannon entropy function (Proposition 5.5). Finally, again making use of (8.139), one may conclude that

$$\begin{aligned} \mathrm{H}(p) &\leq -\Big(1 - \frac{\varepsilon}{2}\Big) \log\Big(\frac{1 - \varepsilon/2}{2^m}\Big) - \frac{\varepsilon}{2} \log\Big(\frac{\varepsilon/2}{2^{2m} - 2^m}\Big) \\ &< \Big(1 + \frac{\varepsilon}{2}\Big) m + \mathrm{H}\Big(1 - \frac{\varepsilon}{2}, \frac{\varepsilon}{2}\Big). \end{aligned} \tag{8.144}$$

It follows that

$$\begin{aligned} \chi\big(\Phi^{\otimes n}\big) \geq \mathrm{I}(\mathsf{X} : \mathsf{Z}) &= \mathrm{H}(p[\mathsf{X}]) + \mathrm{H}(p[\mathsf{Z}]) - \mathrm{H}(p) \\ &\geq (1 - \varepsilon) m - \mathrm{H}\Big(1 - \frac{\varepsilon}{2}, \frac{\varepsilon}{2}\Big). \end{aligned} \tag{8.145}$$

Now, given that $\varepsilon > 0$ was chosen arbitrarily, it follows that

$$\chi\big(\Phi^{\otimes n}\big) \geq m = \lfloor \alpha n \rfloor, \tag{8.146}$$

for all but finitely many positive integers n. One therefore has that

$$\alpha \leq \lim_{n \to \infty} \frac{\chi\big(\Phi^{\otimes n}\big)}{n}. \tag{8.147}$$

As $\mathrm{C}(\Phi)$ is equal to the supremum over all achievable rates α for classical information transmission through Φ, it follows that

$$\mathrm{C}(\Phi) \leq \lim_{n \to \infty} \frac{\chi\big(\Phi^{\otimes n}\big)}{n}, \tag{8.148}$$

as required. □

8.1.3 The entanglement-assisted classical capacity theorem

This section is focused on the *entanglement-assisted classical capacity theorem*, which characterizes the entanglement-assisted classical capacity of a given channel. This theorem stands out among the capacity theorems presented in the present chapter, as no regularization is required by the characterization it provides.

Holevo–Schumacher–Westmoreland with entanglement assistance

A preliminary step toward the proof of the entanglement-assisted classical capacity theorem is the observation that, when both the classical capacity and Holevo capacity are replaced by their entanglement-assisted forms, an analogous statement to the Holevo–Schumacher–Westmoreland theorem is true.

Theorem 8.31. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. The entanglement-assisted classical capacity of Φ equals the regularized entanglement-assisted Holevo capacity of Φ:*

$$\mathrm{C}_{\mathrm{E}}(\Phi) = \lim_{n\to\infty} \frac{\chi_{\mathrm{E}}\big(\Phi^{\otimes n}\big)}{n}. \tag{8.149}$$

Proof. The theorem is proved in essentially the same way as the Holevo–Schumacher–Westmoreland theorem (Theorem 8.30), with each step being modified to allow for the possibility of entanglement assistance.

In greater detail, let $\mathcal{W}$ be a complex Euclidean space and let η be an ensemble of the form

$$\eta : \Sigma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{W}) \tag{8.150}$$

that is constant with respect to $\mathcal{W}$. By Proposition 8.12, one may choose a complex Euclidean space $\mathcal{V}$, a state $\xi \in \mathrm{D}(\mathcal{V} \otimes \mathcal{W})$, a probability vector $p \in \mathcal{P}(\Sigma)$, and a collection of channels

$$\{\Psi_a \,:\, a \in \Sigma\} \subseteq \mathrm{C}(\mathcal{V}, \mathcal{X}) \tag{8.151}$$

such that

$$\eta(a) = p(a)\big(\Psi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\xi) \tag{8.152}$$

for every $a \in \Sigma$. It will be proved that every positive real number α with

$$\alpha < \chi\big(\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\eta)\big) \tag{8.153}$$

is an achievable rate for entanglement-assisted classical information transmission through Φ. Let

$$\sigma_a = \big(\Phi\Psi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\xi) \tag{8.154}$$

for each $a \in \Sigma$.

By Theorem 8.29 it follows, for every positive real number $\varepsilon > 0$, for all but finitely many choices of a positive integer n, and for all positive integers $m \leq \alpha n$, that there exist a classical-to-quantum product state channel code (f, μ) for the collection

$$\{\sigma_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{Y} \otimes \mathcal{W}) \tag{8.155}$$

having error bounded by $\varepsilon/2$. Assume hereafter that such a code, taking the form

$$f : \Gamma^m \to \Sigma^n \quad \text{and} \quad \mu : \Gamma^m \to \mathrm{Pos}\big((\mathcal{Y} \otimes \mathcal{W})^{\otimes n}\big) \tag{8.156}$$

has been selected.

It will be proved that the channel $\Phi^{\otimes n}$ emulates a ε-approximation to the completely dephasing channel $\Delta^{\otimes m} \in \mathrm{C}(\mathcal{Z}^{\otimes m})$ with the assistance of entanglement. The entangled state to be used to assist this emulation is

$$V\xi^{\otimes n}V^* \in \mathrm{D}\big(\mathcal{V}^{\otimes n} \otimes \mathcal{W}^{\otimes n}\big), \tag{8.157}$$

where $V \in \mathrm{U}\big((\mathcal{V} \otimes \mathcal{W})^{\otimes n}, \mathcal{V}^{\otimes n} \otimes \mathcal{W}^{\otimes n}\big)$ represents a permutation of tensor factors:

$$\begin{aligned} &V\big((v_1 \otimes w_1) \otimes \cdots \otimes (v_n \otimes w_n)\big) \\ &\quad = (v_1 \otimes \cdots \otimes v_n) \otimes (w_1 \otimes \cdots \otimes w_n) \end{aligned} \tag{8.158}$$

for every choice of vectors $v_1, \ldots, v_n \in \mathcal{V}$ and $w_1, \ldots, w_n \in \mathcal{W}$. Define an encoding channel $\Xi_{\mathrm{E}} \in \mathrm{C}\big(\mathcal{Z}^{\otimes m} \otimes \mathcal{V}^{\otimes n}, \mathcal{X}^{\otimes n}\big)$ as

$$\Xi_{\mathrm{E}} = \sum_{b_1 \cdots b_m \in \Gamma^m} \Theta_{b_1 \cdots b_m} \otimes \Psi_{f(b_1 \cdots b_m)}\,, \tag{8.159}$$

where one defines

$$\Psi_{a_1 \cdots a_n} = \Psi_{a_1} \otimes \cdots \otimes \Psi_{a_n} \tag{8.160}$$

for each $a_1 \cdots a_n \in \Sigma^n$, and where $\Theta_{b_1 \cdots b_m} \in \mathrm{CP}(\mathcal{Z}^{\otimes m}, \mathbb{C})$ is given by

$$\Theta_{b_1 \cdots b_m}(Z) = Z(b_1 \cdots b_m, b_1 \cdots b_m) \tag{8.161}$$

for every $Z \in \mathrm{L}(\mathcal{Z}^{\otimes m})$. Described in words, the encoding map Ξ_{E} takes as input a compound register $(\mathsf{Z}_1, \ldots, \mathsf{Z}_m, \mathsf{V}_1, \ldots, \mathsf{V}_n)$, measures $(\mathsf{Z}_1, \ldots, \mathsf{Z}_m)$ with respect to the standard basis measurement, and applies the channel $\Psi_{f(b_1 \cdots b_m)}$ to $(\mathsf{V}_1, \ldots, \mathsf{V}_n)$ for $b_1 \cdots b_m$ being the string obtained from the standard basis measurement on $(\mathsf{Z}_1, \ldots, \mathsf{Z}_m)$. Define a decoding channel $\Xi_{\mathrm{D}} \in \mathrm{C}\big(\mathcal{Y}^{\otimes n} \otimes \mathcal{W}^{\otimes n}, \mathcal{Z}^{\otimes m}\big)$ as

$$\Xi_{\mathrm{D}}(Y) = \sum_{b_1 \cdots b_m \in \Gamma^m} \big\langle W \mu(b_1 \cdots b_m) W^*, Y \big\rangle \, E_{b_1 \cdots b_m, b_1 \cdots b_m} \tag{8.162}$$

for all $Y \in \mathrm{L}(\mathcal{Y}^{\otimes n} \otimes \mathcal{W}^{\otimes n})$, where $W \in \mathrm{U}\big((\mathcal{Y} \otimes \mathcal{W})^{\otimes n}, \mathcal{Y}^{\otimes n} \otimes \mathcal{W}^{\otimes n}\big)$ is an isometry representing a permutation of tensor factors that is similar to V, but with $\mathcal{V}$ replaced by $\mathcal{Y}$:

$$\begin{aligned} &W((y_1 \otimes w_1) \otimes \cdots \otimes (y_n \otimes w_n)) \\ &\quad = (y_1 \otimes \cdots \otimes y_n) \otimes (w_1 \otimes \cdots \otimes w_n) \end{aligned} \tag{8.163}$$

for all choices of vectors $y_1, \ldots, y_n \in \mathcal{Y}$ and $w_1, \ldots, w_n \in \mathcal{W}$.

Now, let $\Psi \in \mathrm{C}(\mathcal{Z}^{\otimes m})$ denote the channel that has been emulated with the assistance of entanglement by the above construction:

$$\Psi(Z) = \big(\Xi_{\mathrm{D}}\big(\Phi^{\otimes n} \Xi_{\mathrm{E}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}^{\otimes n}\big)\big)\big(Z \otimes V \xi^{\otimes n} V^*\big) \tag{8.164}$$

for every $Z \in \mathrm{L}(\mathcal{Z}^{\otimes m})$. For every binary string $b_1 \cdots b_m \in \Gamma^m$, it holds that

$$\big(\Xi_{\mathrm{E}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}^{\otimes n}\big)\big(E_{b_1 \cdots b_m, b_1 \cdots b_m} \otimes V \xi^{\otimes n} V^*\big) = W \rho_{f(b_1 \cdots b_m)} W^*, \tag{8.165}$$

from which it follows that

$$\big\langle E_{b_1 \cdots b_m, b_1 \cdots b_m}, \Psi(E_{b_1 \cdots b_m, b_1 \cdots b_m}) \big\rangle \geq 1 - \frac{\varepsilon}{2}. \tag{8.166}$$

As $\Psi = \Delta^{\otimes m} \Psi \Delta^{\otimes m}$, it follows that Ψ is a ε-approximation to $\Delta^{\otimes m}$.

As $\varepsilon > 0$ has been chosen arbitrarily, and the analysis above may be considered for all but finitely many n and all $m \leq \alpha n$, one may conclude that α is an achievable rate for entanglement-assisted classical information transmission through Φ. Given that the alphabet Σ, the complex Euclidean space $\mathcal{W}$, and the ensemble η were chosen arbitrarily, subject to η being constant with respect to $\mathcal{W}$, it follows that

$$\chi_{\mathrm{E}}(\Phi) \leq \mathrm{C}_{\mathrm{E}}(\Phi). \tag{8.167}$$

Applying the same argument to the channel $\Phi^{\otimes n}$ in place of Φ, for any choice of a positive integer n, yields

$$\chi_{\mathrm{E}}(\Phi^{\otimes n}) \leq \mathrm{C}_{\mathrm{E}}(\Phi^{\otimes n}) = n\,\mathrm{C}_{\mathrm{E}}(\Phi), \tag{8.168}$$

and therefore

$$\mathrm{C}_{\mathrm{E}}(\Phi) \geq \lim_{n\to\infty} \frac{\chi_{\mathrm{E}}\big(\Phi^{\otimes n}\big)}{n}. \tag{8.169}$$

It remains to prove that the entanglement-assisted classical capacity of Φ cannot exceed its regularized entanglement-assisted Holevo capacity. As in the proof of Theorem 8.30, it may be assumed that $\mathrm{C}_{\mathrm{E}}(\Phi) > 0$, and it suffices to consider the situation in which a sender transmits a uniformly generated binary string of length m to a receiver.

Suppose $\alpha > 0$ is an achievable rate for entanglement-assisted classical information transmission through Φ, and let $\varepsilon > 0$ be chosen arbitrarily. It must therefore hold, for all but finitely many positive integers n, and for all $m \leq \alpha n$, that $\Phi^{\otimes n}$ emulates an ε-approximation to $\Delta^{\otimes m}$ with the assistance of entanglement. Let n be an arbitrarily chosen positive integer for which this property holds and for which $\lfloor \alpha n \rfloor \geq 2$, and let $m = \lfloor \alpha n \rfloor$.

By the assumption that $\Phi^{\otimes n}$ emulates an ε-approximation to $\Delta^{\otimes m}$ with the assistance of entanglement, one may conclude that there exists a choice of complex Euclidean spaces $\mathcal{V}$ and $\mathcal{W}$, a state $\xi \in \mathrm{D}(\mathcal{V} \otimes \mathcal{W})$, a collection of channels

$$\big\{\Psi_{b_1\cdots b_m} \,:\, b_1 \cdots b_m \in \Gamma^m\big\} \subseteq \mathrm{C}\big(\mathcal{V}, \mathcal{X}^{\otimes n}\big), \tag{8.170}$$

and a measurement $\mu : \Gamma^m \to \mathrm{Pos}(\mathcal{Y}^{\otimes n} \otimes \mathcal{W})$, such that

$$\Big\langle \mu(b_1 \cdots b_m), \big(\Phi^{\otimes n}\Psi_{b_1\cdots b_m} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\xi)\Big\rangle \geq 1 - \frac{\varepsilon}{2} \tag{8.171}$$

for every string $b_1 \cdots b_m \in \Gamma^m$.

Let X and Z be classical registers both having state set Γ^m, where X is a register representing the randomly generated string selected by the sender and Z is a register representing the string obtained by the receiver when a copy of the string stored in X is transmitted through the ε-approximation to $\Delta^{\otimes m}$ emulated by $\Phi^{\otimes n}$ with the assistance of entanglement. With respect to the probability vector $p \in \mathcal{P}(\Gamma^m \times \Gamma^m)$ defined as

$$\begin{aligned} &p(b_1 \cdots b_m, c_1 \cdots c_m) \\ &\quad = \frac{1}{2^m}\Big\langle \mu(c_1 \cdots c_m), \big(\Phi^{\otimes n}\Psi_{b_1\cdots b_m} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\xi)\Big\rangle, \end{aligned} \tag{8.172}$$

representing a probabilistic state of (X, Z), it follows from Holevo's theorem (Theorem 5.52) that

$$\mathrm{I}(\mathsf{X} : \mathsf{Z}) \leq \chi\big(\big(\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\eta)\big), \tag{8.173}$$

for $\eta : \Gamma^m \to \mathrm{Pos}(\mathcal{X}^{\otimes n} \otimes \mathcal{W})$ being the ensemble defined as

$$\eta(b_1 \cdots b_m) = \frac{1}{2^m}\big(\Psi_{b_1 \cdots b_m} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\xi). \tag{8.174}$$

The same lower-bound on the quantity $\mathrm{I}(\mathsf{X} : \mathsf{Z})$ derived in the proof of Theorem 8.30 holds in the present case, from which it follows that

$$\chi_{\mathrm{E}}\big(\Phi^{\otimes n}\big) \geq \mathrm{I}(\mathsf{X} : \mathsf{Z}) \geq (1-\varepsilon)m - \mathrm{H}\Big(1 - \frac{\varepsilon}{2}, \frac{\varepsilon}{2}\Big). \tag{8.175}$$

Given that $\varepsilon > 0$ was chosen arbitrarily, one has that

$$\chi_{\mathrm{E}}\big(\Phi^{\otimes n}\big) \geq m = \lfloor \alpha n \rfloor, \tag{8.176}$$

for all but finitely many positive integers n. Consequently,

$$\alpha \leq \lim_{n\to\infty} \frac{\chi_{\mathrm{E}}\big(\Phi^{\otimes n}\big)}{n}. \tag{8.177}$$

As $\mathrm{C}_{\mathrm{E}}(\Phi)$ is defined as the supremum over all achievable rates α for classical information transmission through Φ with the assistance of entanglement, it follows that

$$\mathrm{C}_{\mathrm{E}}(\Phi) \leq \lim_{n\to\infty} \frac{\chi_{\mathrm{E}}\big(\Phi^{\otimes n}\big)}{n}, \tag{8.178}$$

which completes the proof. □

Strongly typical strings and projections

The proof of the entanglement-assisted classical capacity theorem that is presented in this book will make use of a notion of typicality, known as *strong typicality*, that differs from the standard notion discussed previously in Section 5.3.1. True to its name, strong typicality is the more restrictive of the two notions; every strongly typical string will necessarily be a typical string, up to a simple change of parameters, while some typical strings are not strongly typical.

Similar to the standard notion of typicality, one may define an *ε-strongly typical subspace* with respect to a spectral decomposition of a given state. Unlike the standard typical subspace, the strongly typical subspace may not be uniquely determined by a given state—when the spectral decomposition is not unique, the strongly typical subspace may depend on the particular spectral decomposition with respect to which it was defined. Despite this apparent drawback, the notion of an ε-strongly typical subspace will prove to be an important concept in proving the entanglement-assisted classical capacity theorem.

The definition of strong-typicality to follow uses the following notation, for which it is to be assumed that Σ is an alphabet and n is a positive integer. For every string $a_1 \cdots a_n \in \Sigma^n$ and symbol $a \in \Sigma$, one writes

$$N(a \mid a_1 \cdots a_n) = \big|\{k \in \{1, \ldots, n\} \,:\, a_k = a\}\big|, \tag{8.179}$$

which is the number of times the symbol a occurs in the string $a_1 \cdots a_n$.

Definition 8.32. Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, and let $\varepsilon > 0$ be a positive real number. A string $a_1 \cdots a_n \in \Sigma^n$ is said to be *ε-strongly typical with respect to p* if and only if

$$\left| \frac{N(a \mid a_1 \cdots a_n)}{n} - p(a) \right| \leq p(a)\varepsilon \tag{8.180}$$

for every $a \in \Sigma$. The set of all ε-strongly typical strings of length n with respect to p is denoted $S_{n,\varepsilon}(p)$ (or by $S_{n,\varepsilon}$ when p is implicit and can safely be omitted).

The average behavior of a nonnegative real-valued function defined on the individual symbols of a strongly typical string may be analyzed using the following elementary proposition.

Proposition 8.33. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, let $\varepsilon > 0$ be a positive real number, let $a_1 \cdots a_n \in S_{n,\varepsilon}(p)$ be an ε-strongly typical string with respect to p, and let $\phi : \Sigma \to [0, \infty)$ be a nonnegative real-valued function. It holds that*

$$\left| \frac{\phi(a_1) + \cdots + \phi(a_n)}{n} - \sum_{a \in \Sigma} p(a)\phi(a) \right| \leq \varepsilon \sum_{a \in \Sigma} p(a)\phi(a). \tag{8.181}$$

Proof. The inequality (8.181) follows from the definition of strong typicality together with the triangle inequality:

$$\begin{aligned}
&\left| \frac{\phi(a_1) + \cdots + \phi(a_n)}{n} - \sum_{a \in \Sigma} p(a)\phi(a) \right| \\
&\qquad = \left| \sum_{a \in \Sigma} \left(\frac{N(a \mid a_1 \cdots a_n)\phi(a)}{n} - p(a)\phi(a) \right) \right| \\
&\qquad \leq \sum_{a \in \Sigma} \phi(a) \left| \frac{N(a \mid a_1 \cdots a_n)}{n} - p(a) \right| \leq \varepsilon \sum_{a \in \Sigma} p(a)\phi(a),
\end{aligned} \tag{8.182}$$

as required. □

As a corollary to Proposition 8.33, one has that every ε-strongly typical string, with respect to a given probability vector p, is necessarily δ-typical for every choice of $\delta > \varepsilon \, \mathrm{H}(p)$.

Corollary 8.34. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, let $\varepsilon > 0$ be a positive real number, and let $a_1 \cdots a_n \in S_{n,\varepsilon}(p)$ be an ε-strongly typical string with respect to p. It holds that*

$$2^{-n(1+\varepsilon)\,\mathrm{H}(p)} \leq p(a_1) \cdots p(a_n) \leq 2^{-n(1-\varepsilon)\,\mathrm{H}(p)}. \tag{8.183}$$

Proof. Define a function $\phi : \Sigma \to [0, \infty)$ as

$$\phi(a) = \begin{cases} -\log(p(a)) & \text{if } p(a) \neq 0 \\ 0 & \text{if } p(a) = 0. \end{cases} \tag{8.184}$$

With respect to this function, the implication provided by Proposition 8.33 is equivalent to (8.183). □

Strings that are obtained by independently selecting symbols at random according to a given probability vector are likely to be not only typical, but strongly typical, with the probability of strong typicality increasing with string length. The following lemma establishes a quantitative bound on this probability.

Lemma 8.35. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, and let $\varepsilon > 0$ be a positive real number. It holds that*

$$\sum_{a_1 \cdots a_n \in S_{n,\varepsilon}(p)} p(a_1) \cdots p(a_n) \geq 1 - \zeta_{n,\varepsilon}(p) \tag{8.185}$$

for

$$\zeta_{n,\varepsilon}(p) = 2 \sum_{a \in \operatorname{supp}(p)} \exp\bigl(-n\varepsilon^2 p(a)^2\bigr). \tag{8.186}$$

Proof. Suppose first that $a \in \Sigma$ is fixed, and consider the probability that a string $a_1 \cdots a_n \in \Sigma^n$, randomly selected according to the probability vector $p^{\otimes n}$, satisfies

$$\left| \frac{N(a \mid a_1 \cdots a_n)}{n} - p(a) \right| > p(a)\varepsilon. \tag{8.187}$$

To bound this probability, one may define $X_1, \ldots, X_n$ to be independent and identically distributed random variables, taking value 1 with probability $p(a)$ and value 0 otherwise, so that the probability of the event (8.187) is equal to

$$\Pr\left(\left| \frac{X_1 + \cdots + X_n}{n} - p(a) \right| > p(a)\varepsilon \right). \tag{8.188}$$

If it is the case that $p(a) > 0$, then Hoeffding's inequality (Theorem 1.16) implies that

$$\Pr\left(\left| \frac{X_1 + \cdots + X_n}{n} - p(a) \right| > p(a)\varepsilon \right) \leq 2 \exp\bigl(-n\varepsilon^2 p(a)^2\bigr), \tag{8.189}$$

while it holds that

$$\Pr\left(\left| \frac{X_1 + \cdots + X_n}{n} - p(a) \right| > p(a)\varepsilon \right) = 0 \tag{8.190}$$

in case $p(a) = 0$. The lemma follows from the union bound. □

The next proposition establishes upper and lower bounds on the number of strings in an ε-strongly typical set for a given length.

Proposition 8.36. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, and let $\varepsilon > 0$ be a positive real number. It holds that*

$$\bigl(1 - \zeta_{n,\varepsilon}(p)\bigr)\, 2^{n(1-\varepsilon)\,\mathrm{H}(p)} \leq \bigl|S_{n,\varepsilon}(p)\bigr| \leq 2^{n(1+\varepsilon)\,\mathrm{H}(p)}, \tag{8.191}$$

for $\zeta_{n,\varepsilon}(p)$ as defined in Lemma 8.35.

Proof. By Corollary 8.34, one has

$$p(a_1) \cdots p(a_n) \geq 2^{-n(1+\varepsilon)\,\mathrm{H}(p)} \tag{8.192}$$

for every string $a_1 \cdots a_n \in S_{n,\varepsilon}(p)$. Consequently,

$$1 \geq \sum_{a_1 \cdots a_n \in S_{n,\varepsilon}(p)} p(a_1) \cdots p(a_n) \geq \bigl|S_{n,\varepsilon}(p)\bigr| 2^{-n(1+\varepsilon)\,\mathrm{H}(p)}, \tag{8.193}$$

and therefore

$$\bigl|S_{n,\varepsilon}(p)\bigr| \leq 2^{n(1+\varepsilon)\,\mathrm{H}(p)}. \tag{8.194}$$

Along similar lines, one has

$$p(a_1) \cdots p(a_n) \leq 2^{-n(1-\varepsilon)\,\mathrm{H}(p)} \tag{8.195}$$

for every string $a_1 \cdots a_n \in S_{n,\varepsilon}(p)$. By Lemma 8.35, it follows that

$$1 - \zeta_{n,\varepsilon}(p) \leq \sum_{a_1 \cdots a_n \in S_{n,\varepsilon}(p)} p(a_1) \cdots p(a_n) \leq \bigl|S_{n,\varepsilon}(p)\bigr|\, 2^{-n(1-\varepsilon)\,\mathrm{H}(p)}, \tag{8.196}$$

and therefore

$$\bigl|S_{n,\varepsilon}(p)\bigr| \geq (1 - \zeta_{n,\varepsilon}(p))\, 2^{n(1-\varepsilon)\,\mathrm{H}(p)}, \tag{8.197}$$

as required. □

Finally, the ε-strongly typical subspaces associated with a given density operator are defined as follows.

Definition 8.37. Let $\mathcal{X}$ be a complex Euclidean space, let $\rho \in \mathrm{D}(\mathcal{X})$ be a density operator, let $\varepsilon > 0$ be a positive real number, and let n be a positive integer. Also let

$$\rho = \sum_{a \in \Sigma} p(a) x_a x_a^* \tag{8.198}$$

be a spectral decomposition of ρ, for Σ being an alphabet, $p \in \mathcal{P}(\Sigma)$ being a probability vector, and $\{x_a : a \in \Sigma\} \subset \mathcal{X}$ being an orthonormal set of vectors. The *projection operator onto the ε-strongly typical subspace of $\mathcal{X}^{\otimes n}$* with respect to the spectral decomposition (8.198) is defined as

$$\Lambda = \sum_{a_1 \cdots a_n \in S_{n,\varepsilon}(p)} x_{a_1} x_{a_1}^* \otimes \cdots \otimes x_{a_n} x_{a_n}^*. \tag{8.199}$$

With respect to the decomposition (8.198), the *ε-strongly typical subspace of $\mathcal{X}^{\otimes n}$* is defined as the image of Λ.

Example 8.38. Let $\Sigma = \{0,1\}$, let $\mathcal{X} = \mathbb{C}^\Sigma$, and let $\rho = \mathbb{1}/2 \in \mathrm{D}(\mathcal{X})$. With respect to the spectral decomposition

$$\rho = \frac{1}{2} e_0 e_0^* + \frac{1}{2} e_1 e_1^*, \tag{8.200}$$

for $n = 2$ and for any choice of $\varepsilon \in (0,1)$, one has that the corresponding projection operator onto the ε-strongly typical subspace is given by

$$\Lambda_0 = E_{0,0} \otimes E_{1,1} + E_{1,1} \otimes E_{0,0}. \tag{8.201}$$

Replacing the spectral decomposition by

$$\rho = \frac{1}{2} x_0 x_0^* + \frac{1}{2} x_1 x_1^*, \tag{8.202}$$

for

$$x_0 = \frac{e_0 + e_1}{\sqrt{2}} \quad \text{and} \quad x_1 = \frac{e_0 - e_1}{\sqrt{2}}, \tag{8.203}$$

one obtains the corresponding projection operator

$$\Lambda_1 = x_0 x_0^* \otimes x_1 x_1^* + x_1 x_1^* \otimes x_0 x_0^* \neq \Lambda_0. \tag{8.204}$$

Two lemmas on the output entropy of channels

The proof of the entanglement-assisted classical capacity theorem appearing at the end of the present section will make use of multiple lemmas. The two lemmas that follow concern the output entropy of channels. The first of these two lemmas will also be used in the next section of the chapter, when proving that the coherent information is a lower bound on the quantum capacity of a channel.

Lemma 8.39. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X},\mathcal{Y})$ be a channel, let $\rho \in \mathrm{D}(\mathcal{X})$ be a density operator, and let $\varepsilon > 0$ be a positive real number. Let*

$$\rho = \sum_{a \in \Sigma} p(a) x_a x_a^* \tag{8.205}$$

be a spectral decomposition of ρ, for Σ being an alphabet, $\{x_a : a \in \Sigma\} \subset \mathcal{X}$ being an orthonormal set, and $p \in \mathcal{P}(\Sigma)$ being a probability vector. For every positive integer n, let $\Lambda_{n,\varepsilon}$ denote the projection operator onto the ε-strongly typical subspace of $\mathcal{X}^{\otimes n}$ with respect to the decomposition (8.205)*, and let*

$$\omega_{n,\varepsilon} = \frac{\Lambda_{n,\varepsilon}}{\mathrm{Tr}(\Lambda_{n,\varepsilon})}. \tag{8.206}$$

It holds that

$$\left| \lim_{n\to\infty} \frac{\mathrm{H}\big(\Phi^{\otimes n}(\omega_{n,\varepsilon})\big)}{n} - \mathrm{H}(\Phi(\rho)) \right| \leq \big(2\,\mathrm{H}(\rho) + \mathrm{H}(\Phi(\rho))\big)\,\varepsilon. \tag{8.207}$$

Proof. It may be verified that the equation

$$\begin{aligned} &\mathrm{H}(\Phi(\rho)) - \frac{1}{n}\,\mathrm{H}\big(\Phi^{\otimes n}(\omega_{n,\varepsilon})\big) \\ &\qquad = \frac{1}{n}\,\mathrm{D}\big(\Phi^{\otimes n}(\omega_{n,\varepsilon}) \,\big\|\, \Phi^{\otimes n}\big(\rho^{\otimes n}\big)\big) \\ &\qquad\qquad + \frac{1}{n}\,\mathrm{Tr}\big(\big(\Phi^{\otimes n}(\omega_{n,\varepsilon}) - \Phi(\rho)^{\otimes n}\big)\log\big(\Phi(\rho)^{\otimes n}\big)\big) \end{aligned} \tag{8.208}$$

holds for every positive integer n. Bounds on the absolute values of the two terms on the right-hand side of this equation will be established separately.

The first term on the right-hand side of (8.208) is nonnegative, and an upper bound on it may be obtained from the monotonicity of the quantum relative entropy under the action of channels (Theorem 5.38). Specifically, by Corollary 8.34 and Proposition 8.36, one has

$$\begin{aligned} &\frac{1}{n}\,\mathrm{D}\big(\Phi^{\otimes n}(\omega_{n,\varepsilon}) \,\big\|\, \Phi^{\otimes n}(\rho^{\otimes n})\big) \leq \frac{1}{n}\,\mathrm{D}\big(\omega_{n,\varepsilon} \,\big\|\, \rho^{\otimes n}\big) \\ &= -\frac{1}{n}\log(|S_{n,\varepsilon}|) - \frac{1}{n|S_{n,\varepsilon}|} \sum_{a_1\cdots a_n \in S_{n,\varepsilon}} \log(p(a_1)\cdots p(a_n)) \\ &\qquad\qquad\qquad\qquad \leq 2\varepsilon\,\mathrm{H}(\rho) - \frac{\log(1-\zeta_{n,\varepsilon})}{n} \end{aligned} \tag{8.209}$$

for every positive integer n, where $S_{n,\varepsilon}$ denotes the set of ε-strongly typical strings of length n with respect to p and $\zeta_{n,\varepsilon}$ denotes the quantity defined in Lemma 8.35.

To bound the absolute value of second term on the right-hand side of (8.208), one may first define a function $\phi : \Sigma \to [0,\infty)$ as

$$\phi(a) = \begin{cases} -\,\mathrm{Tr}(\Phi(x_a x_a^*)\log(\Phi(\rho))) & \text{if } p(a) > 0 \\ 0 & \text{if } p(a) = 0 \end{cases} \tag{8.210}$$

for each $a \in \Sigma$. It is evident from its specification that $\phi(a)$ is nonnegative for each $a \in \Sigma$, and is finite by virtue of the fact that

$$\mathrm{im}(\Phi(x_a x_a^*)) \subseteq \mathrm{im}(\Phi(\rho)) \tag{8.211}$$

for each $a \in \operatorname{supp}(p)$. Using the identity

$$\log\bigl(P^{\otimes n}\bigr) = \sum_{k=1}^{n} \mathbb{1}^{\otimes(k-1)} \otimes \log(P) \otimes \mathbb{1}^{\otimes(n-k)}, \tag{8.212}$$

it may be verified that

$$\begin{aligned} &\operatorname{Tr}\bigl(\Phi^{\otimes n}(\omega_{n,\varepsilon}) \log\bigl(\Phi(\rho)^{\otimes n}\bigr)\bigr) \\ &\quad = -\frac{1}{|S_{n,\varepsilon}|} \sum_{a_1 \cdots a_n \in S_{n,\varepsilon}} (\phi(a_1) + \cdots + \phi(a_n)), \end{aligned} \tag{8.213}$$

and it is evident that

$$\mathrm{H}(\Phi(\rho)) = \sum_{a \in \Sigma} p(a)\phi(a). \tag{8.214}$$

It therefore holds that

$$\begin{aligned} &\left| \frac{1}{n} \operatorname{Tr}\bigl(\bigl(\Phi^{\otimes n}(\omega_{n,\varepsilon}) - \Phi(\rho)^{\otimes n}\bigr) \log\bigl(\Phi(\rho)^{\otimes n}\bigr)\bigr) \right| \\ &\leq \frac{1}{|S_{n,\varepsilon}|} \sum_{a_1 \cdots a_n \in S_{n,\varepsilon}} \left| \frac{\phi(a_1) + \cdots + \phi(a_n)}{n} - \sum_{a \in \Sigma} p(a)\phi(a) \right| \\ &\qquad\qquad \leq \varepsilon \sum_{a \in \Sigma} p(a)\phi(a) = \varepsilon\, \mathrm{H}(\Phi(\rho)). \end{aligned} \tag{8.215}$$

Combining the inequalities (8.209) and (8.215), one has

$$\begin{aligned} &\left| \frac{1}{n} \mathrm{H}\bigl(\Phi^{\otimes n}(\omega_{n,\varepsilon})\bigr) - \mathrm{H}(\Phi(\rho)) \right| \\ &\quad \leq 2\varepsilon\, \mathrm{H}(\rho) - \frac{\log(1 - \zeta_{n,\varepsilon})}{n} + \varepsilon\, \mathrm{H}(\Phi(\rho)), \end{aligned} \tag{8.216}$$

from which the lemma follows. □

Lemma 8.40. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. The function $f : \mathrm{D}(\mathcal{X}) \to \mathbb{R}$ defined by*

$$f(\rho) = \mathrm{H}(\rho) - \mathrm{H}(\Phi(\rho)) \tag{8.217}$$

is concave.

Proof. Let $\mathcal{Z}$ be any complex Euclidean space, and consider first the function $g : \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z}) \to \mathbb{R}$ defined as

$$g(\sigma) = \mathrm{H}(\sigma) - \mathrm{H}(\mathrm{Tr}_{\mathcal{Z}}(\sigma)) \tag{8.218}$$

for every $\sigma \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z})$. An alternative expression for g is

$$g(\sigma) = -\,\mathrm{D}(\sigma \,\|\, \mathrm{Tr}_{\mathcal{Z}}(\sigma) \otimes \mathbb{1}_{\mathcal{Z}}), \tag{8.219}$$

and the concavity of g therefore follows from the joint convexity of quantum relative entropy (Corollary 5.36).

For a suitable choice of a complex Euclidean space $\mathcal{Z}$, let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ be an isometry that yields a Stinespring representation of Φ:

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}(AXA^*) \tag{8.220}$$

for every $X \in \mathrm{L}(\mathcal{X})$. The function f is given by $f(\rho) = g(A\rho A^*)$ for every $\rho \in \mathrm{D}(\mathcal{X})$, and therefore the concavity of g implies that f is concave as well. □

An additivity lemma concerning the coherent information

The next lemma that will be used in the proof of the entanglement-assisted capacity theorem is the following lemma, which states that the quantity

$$\max_{\sigma \in \mathrm{D}(\mathcal{X})} \big(\mathrm{H}(\sigma) + \mathrm{I_c}(\sigma\,;\,\Phi)\big), \tag{8.221}$$

defined for each channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, is additive with respect to tensor products. (Indeed, this is the quantity the entanglement-assisted classical capacity theorem establishes is equal to the entanglement-assisted classical capacity.)

Lemma 8.41 (Adami–Cerf)**.** *Let $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$ be complex Euclidean spaces and let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be channels. It holds that*

$$\begin{aligned} &\max_{\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)} \big(\mathrm{H}(\sigma) + \mathrm{I_c}(\sigma; \Phi_0 \otimes \Phi_1)\big) \\ &\quad = \max_{\sigma_0 \in \mathrm{D}(\mathcal{X}_0)} \big(\mathrm{H}(\sigma_0) + \mathrm{I_c}(\sigma_0; \Phi_0)\big) + \max_{\sigma_1 \in \mathrm{D}(\mathcal{X}_1)} \big(\mathrm{H}(\sigma_1) + \mathrm{I_c}(\sigma_1; \Phi_1)\big). \end{aligned} \tag{8.222}$$

Proof. Choose isometries $A_0 \in \mathrm{U}(\mathcal{X}_0, \mathcal{Y}_0 \otimes \mathcal{Z}_0)$ and $A_1 \in \mathrm{U}(\mathcal{X}_1, \mathcal{Y}_1 \otimes \mathcal{Z}_1)$, for an appropriate choice of complex Euclidean spaces $\mathcal{Z}_0$ and $\mathcal{Z}_1$, so that Stinespring representations of Φ_0 and Φ_1 are obtained:

$$\Phi_0(X_0) = \mathrm{Tr}_{\mathcal{Z}_0}\big(A_0 X_0 A_0^*\big) \quad \text{and} \quad \Phi_1(X_1) = \mathrm{Tr}_{\mathcal{Z}_1}\big(A_1 X_1 A_1^*\big) \tag{8.223}$$

for all $X_0 \in \mathrm{L}(\mathcal{X}_0)$ and $X_1 \in \mathrm{L}(\mathcal{X}_1)$. The channels $\Psi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Z}_0)$ and $\Psi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Z}_1)$ defined as

$$\Psi_0(X_0) = \mathrm{Tr}_{\mathcal{Y}_0}\big(A_0 X_0 A_0^*\big) \quad \text{and} \quad \Psi_1(X_1) = \mathrm{Tr}_{\mathcal{Y}_1}\big(A_1 X_1 A_1^*\big) \tag{8.224}$$

for all $X_0 \in \mathrm{L}(\mathcal{X}_0)$ and $X_1 \in \mathrm{L}(\mathcal{X}_1)$ are therefore complementary to Φ_0 and Φ_1, respectively.

Now, consider registers X_0, X_1, Y_0, Y_1, Z_0, and Z_1 corresponding to the spaces $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, $\mathcal{Y}_1$, $\mathcal{Z}_0$, and $\mathcal{Z}_1$, respectively. Let $\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)$ be an arbitrary density operator. With respect to the state

$$(A_0 \otimes A_1)\sigma(A_0 \otimes A_1)^* \in \mathrm{D}(\mathcal{Y}_0 \otimes \mathcal{Z}_0 \otimes \mathcal{Y}_1 \otimes \mathcal{Z}_1) \tag{8.225}$$

of $(\mathsf{Y}_0, \mathsf{Z}_0, \mathsf{Y}_1, \mathsf{Z}_1)$, one has that

$$\begin{aligned} &\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \otimes \Phi_1) \\ &\quad = \mathrm{H}(\mathsf{Y}_0, \mathsf{Z}_0, \mathsf{Y}_1, \mathsf{Z}_1) + \mathrm{H}(\mathsf{Y}_0, \mathsf{Y}_1) - \mathrm{H}(\mathsf{Z}_0, \mathsf{Z}_1). \end{aligned} \tag{8.226}$$

For every state of $(\mathsf{Y}_0, \mathsf{Z}_0, \mathsf{Y}_1, \mathsf{Z}_1)$, including the state (8.225), it holds that

$$\begin{aligned} &\mathrm{H}(\mathsf{Y}_0, \mathsf{Z}_0, \mathsf{Y}_1, \mathsf{Z}_1) - \mathrm{H}(\mathsf{Z}_0, \mathsf{Z}_1) \\ &\quad \leq \mathrm{H}(\mathsf{Y}_0, \mathsf{Z}_0) - \mathrm{H}(\mathsf{Z}_0) + \mathrm{H}(\mathsf{Y}_1, \mathsf{Z}_1) - \mathrm{H}(\mathsf{Z}_1); \end{aligned} \tag{8.227}$$

two applications of the strong subadditivity of the von Neumann entropy (Theorem 5.39) yield this inequality. It follows from the subadditivity of the von Neumann entropy (Theorem 5.26) that

$$\mathrm{H}(\mathsf{Y}_0, \mathsf{Y}_1) \leq \mathrm{H}(\mathsf{Y}_0) + \mathrm{H}(\mathsf{Y}_1). \tag{8.228}$$

Consequently,

$$\begin{aligned} &\mathrm{H}(\mathsf{Y}_0, \mathsf{Z}_0, \mathsf{Y}_1, \mathsf{Z}_1) + \mathrm{H}(\mathsf{Y}_0, \mathsf{Y}_1) - \mathrm{H}(\mathsf{Z}_0, \mathsf{Z}_1) \\ &\qquad \leq \big(\mathrm{H}(\mathsf{Y}_0, \mathsf{Z}_0) + \mathrm{H}(\mathsf{Y}_0) - \mathrm{H}(\mathsf{Z}_0)\big) \\ &\qquad\quad + \big(\mathrm{H}(\mathsf{Y}_1, \mathsf{Z}_1) + \mathrm{H}(\mathsf{Y}_1) - \mathrm{H}(\mathsf{Z}_1)\big). \end{aligned} \tag{8.229}$$

For $\sigma_0 = \sigma[\mathsf{X}_0]$ and $\sigma_1 = \sigma[\mathsf{X}_1]$, one has the equations

$$\begin{aligned} \mathrm{H}(\mathsf{Y}_0, \mathsf{Z}_0) + \mathrm{H}(\mathsf{Y}_0) - \mathrm{H}(\mathsf{Z}_0) &= \mathrm{H}(\sigma_0) + \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0), \\ \mathrm{H}(\mathsf{Y}_1, \mathsf{Z}_1) + \mathrm{H}(\mathsf{Y}_1) - \mathrm{H}(\mathsf{Z}_1) &= \mathrm{H}(\sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1). \end{aligned} \tag{8.230}$$

It follows that

$$\begin{aligned} &\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \otimes \Phi_1) \\ &\quad \leq \big(\mathrm{H}(\sigma_0) + \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0)\big) + \big(\mathrm{H}(\sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1)\big). \end{aligned} \tag{8.231}$$

Maximizing over all $\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)$, one obtains the inequality

$$\begin{aligned} &\max_{\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \otimes \Phi_1)\big) \\ &\qquad \leq \max_{\sigma_0 \in \mathrm{D}(\mathcal{X}_0)} \big(\mathrm{H}(\sigma_0) + \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0)\big) \\ &\qquad\qquad + \max_{\sigma_1 \in \mathrm{D}(\mathcal{X}_1)} \big(\mathrm{H}(\sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1)\big). \end{aligned} \tag{8.232}$$

For the reverse inequality, it suffices to observe that

$$\begin{aligned} &\mathrm{H}(\sigma_0 \otimes \sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_0 \otimes \sigma_1; \Phi_0 \otimes \Phi_1) \\ &\quad = \mathrm{H}(\sigma_0) + \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0) + \mathrm{H}(\sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1) \end{aligned} \tag{8.233}$$

for every choice of $\sigma_0 \in \mathrm{D}(\mathcal{X}_0)$ and $\sigma_1 \in \mathrm{D}(\mathcal{X}_1)$, and therefore

$$\begin{aligned} &\max_{\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \otimes \Phi_1)\big) \\ &\quad \geq \max_{\sigma_0 \in \mathrm{D}(\mathcal{X}_0)} \big(\mathrm{H}(\sigma_0) + \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0)\big) + \max_{\sigma_1 \in \mathrm{D}(\mathcal{X}_1)} \big(\mathrm{H}(\sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1)\big), \end{aligned} \tag{8.234}$$

which completes the proof. □

A lower-bound on the Holevo capacity for flat states by dense coding

Next in the sequence of lemmas needed to prove the entanglement-assisted classical capacity theorem is the following lemma, which establishes a lower bound on the entanglement-assisted Holevo capacity of a given channel. Its proof may be viewed an application of dense coding (q.v. Section 6.3.1).

Lemma 8.42. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, let $\Pi \in \mathrm{Proj}(\mathcal{X})$ be a nonzero projection operator, and let $\omega = \Pi / \operatorname{Tr}(\Pi)$. It holds that*

$$\chi_{\mathrm{E}}(\Phi) \geq \mathrm{H}(\omega) + \mathrm{I}_{\mathrm{c}}(\omega; \Phi). \tag{8.235}$$

Proof. Let $m = \operatorname{rank}(\Pi)$, let $\mathcal{W} = \mathbb{C}^{\mathbb{Z}_m}$, let $V \in \mathrm{U}(\mathcal{W}, \mathcal{X})$ be any isometry satisfying $VV^* = \Pi$, and let

$$\tau = \frac{1}{m} \operatorname{vec}(V) \operatorname{vec}(V)^* \in \mathrm{D}(\mathcal{X} \otimes \mathcal{W}). \tag{8.236}$$

Recall the collection of discrete Weyl operators

$$\{W_{a,b} \,:\, a, b \in \mathbb{Z}_m\} \subset \mathrm{U}(\mathcal{W}), \tag{8.237}$$

as defined in Section 4.1.2 of Chapter 4, and define a collection of unitary channels

$$\big\{\Psi_{a,b} \,:\, a, b \in \mathbb{Z}_m\big\} \subseteq \mathrm{C}(\mathcal{W}) \tag{8.238}$$

in correspondence with these operators:

$$\Psi_{a,b}(Y) = W_{a,b} Y W_{a,b}^* \tag{8.239}$$

for each $Y \in \mathrm{L}(\mathcal{W})$. Finally, consider the ensemble

$$\eta : \mathbb{Z}_m \times \mathbb{Z}_m \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{W}) \tag{8.240}$$

defined as

$$\eta(a, b) = \frac{1}{m^2} \big(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Psi_{a,b}\big)(\tau), \tag{8.241}$$

for all $(a, b) \in \mathbb{Z}_m \times \mathbb{Z}_m$.

It holds that

$$\begin{aligned} &\mathrm{H}\left(\frac{1}{m^2} \sum_{a,b \in \mathbb{Z}_m} (\Phi \otimes \Psi_{a,b})(\tau)\right) \\ &\qquad = \mathrm{H}\left(\Phi(\omega) \otimes \frac{\mathbb{1}_{\mathcal{W}}}{m}\right) = \mathrm{H}\big(\Phi(\omega)\big) + \mathrm{H}(\omega) \end{aligned} \tag{8.242}$$

and

$$\begin{aligned} \frac{1}{m^2} \sum_{a,b \in \mathbb{Z}_n} \mathrm{H}\big((\Phi \otimes \Psi_{a,b})(\tau)\big) &= \mathrm{H}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\tau)\big) \\ &= \mathrm{H}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})\big(\operatorname{vec}\big(\sqrt{\omega}\big) \operatorname{vec}\big(\sqrt{\omega}\big)^*\big)\big), \end{aligned} \tag{8.243}$$

from which it follows that

$$\chi\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\eta)\big) = \mathrm{H}(\omega) + \mathrm{I}_{\mathrm{C}}(\omega; \Phi). \tag{8.244}$$

Moreover, η is constant with respect to $\mathcal{W}$, as is evident from the fact that

$$\mathrm{Tr}_{\mathcal{X}}(\eta(a,b)) = \frac{1}{m^3}\mathbb{1}_{\mathcal{W}} \tag{8.245}$$

for each choice of $(a,b) \in \mathbb{Z}_m \times \mathbb{Z}_m$. It therefore holds that

$$\chi_{\mathrm{E}}(\Phi) \geq \chi\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\eta)\big) = \mathrm{H}(\omega) + \mathrm{I}_{\mathrm{C}}(\omega;\Phi), \tag{8.246}$$

which completes the proof. □

An upper-bound on the Holevo capacity

The final lemma needed for the proof of the entanglement-assisted classical capacity theorem establishes an upper bound on the entanglement-assisted Holevo capacity of a channel.

Lemma 8.43. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X},\mathcal{Y})$ be a channel. Also let $\mathcal{W}$ be a complex Euclidean space, let Σ be an alphabet, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X}\otimes\mathcal{W})$ be an ensemble that is constant with respect to $\mathcal{W}$, and let*

$$\sigma = \sum_{a\in\Sigma} \mathrm{Tr}_{\mathcal{W}}(\eta(a)). \tag{8.247}$$

It holds that

$$\chi\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\eta)\big) \leq \mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{C}}(\sigma\,;\,\Phi). \tag{8.248}$$

Proof. Assume that $\mathcal{Z}$ is a complex Euclidean space and $A \in \mathrm{U}(\mathcal{X},\mathcal{Y}\otimes\mathcal{Z})$ is an isometry for which

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(AXA^*\big) \tag{8.249}$$

for all $X \in \mathrm{L}(\mathcal{X})$. The channel $\Psi \in \mathrm{C}(\mathcal{X},\mathcal{Z})$ defined by

$$\Psi(X) = \mathrm{Tr}_{\mathcal{Y}}\big(AXA^*\big) \tag{8.250}$$

for all $X \in \mathrm{L}(\mathcal{X})$ is therefore complementary to Φ. By Proposition 8.17, it follows that

$$\mathrm{I}_{\mathrm{C}}(\sigma\,;\,\Phi) = \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)). \tag{8.251}$$

It therefore suffices to prove that

$$\chi\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\eta)\big) \leq \mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)). \tag{8.252}$$

By the assumption that η is constant with respect to $\mathcal{W}$, Proposition 8.12 implies that there must exist a complex Euclidean space $\mathcal{V}$, a collection of channels

$$\{\Xi_a \,:\, a \in \Sigma\} \subseteq \mathrm{C}(\mathcal{V}, \mathcal{X}), \tag{8.253}$$

a unit vector $u \in \mathcal{V} \otimes \mathcal{W}$, and a probability vector $p \in \mathcal{P}(\Sigma)$ such that

$$\eta(a) = \big(\Xi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(uu^*) \tag{8.254}$$

for every $a \in \Sigma$. Assume hereafter that such a choice for these objects has been fixed, and define states $\tau \in \mathrm{D}(\mathcal{W})$ and $\xi \in \mathrm{D}(\mathcal{V})$ as

$$\tau = \mathrm{Tr}_{\mathcal{V}}(uu^*) \quad \text{and} \quad \xi = \mathrm{Tr}_{\mathcal{W}}(uu^*). \tag{8.255}$$

It may be noted that

$$\sigma = \sum_{a \in \Sigma} p(a) \Xi_a(\xi). \tag{8.256}$$

Let $\mathcal{U}$ be a complex Euclidean space such that $\dim(\mathcal{U}) = \dim(\mathcal{V} \otimes \mathcal{X})$, and select a collection of isometries

$$\{B_a \,:\, a \in \Sigma\} \subset \mathrm{U}(\mathcal{V}, \mathcal{X} \otimes \mathcal{U}) \tag{8.257}$$

satisfying

$$\Xi_a(V) = \mathrm{Tr}_{\mathcal{U}}(B_a V B_a^*) \tag{8.258}$$

for every $V \in \mathrm{L}(\mathcal{V})$.

Assume momentarily that $a \in \Sigma$ has been fixed, and define a unit vector

$$v_b = (A \otimes \mathbb{1}_{\mathcal{U}} \otimes \mathbb{1}_{\mathcal{W}})(B_a \otimes \mathbb{1}_{\mathcal{W}})u \in \mathcal{Y} \otimes \mathcal{Z} \otimes \mathcal{U} \otimes \mathcal{W}. \tag{8.259}$$

Let Y, Z, U, and W be registers having corresponding complex Euclidean spaces $\mathcal{Y}$, $\mathcal{Z}$, $\mathcal{U}$, and $\mathcal{W}$, and consider the situation in which the compound register $(\mathsf{Y}, \mathsf{Z}, \mathsf{U}, \mathsf{W})$ is in the pure state $v_b v_b^*$. The following equalities may be verified:

$$\begin{aligned}
\mathrm{H}(\mathsf{W}) &= \mathrm{H}(\tau), \\
\mathrm{H}(\mathsf{Y}, \mathsf{W}) &= \mathrm{H}\big(\big(\Phi \Xi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(uu^*)\big), \\
\mathrm{H}(\mathsf{U}, \mathsf{W}) &= \mathrm{H}(\mathsf{Y}, \mathsf{Z}) = \mathrm{H}\big(\Xi_a(\xi)\big), \\
\mathrm{H}(\mathsf{Y}, \mathsf{U}, \mathsf{W}) &= \mathrm{H}(\mathsf{Z}) = \mathrm{H}\big((\Psi \Xi_a)(\xi)\big).
\end{aligned} \tag{8.260}$$

By the strong subadditivity of the von Neumann entropy (Theorem 5.39), it holds that

$$\mathrm{H}(\mathsf{W}) - \mathrm{H}(\mathsf{Y}, \mathsf{W}) \leq \mathrm{H}(\mathsf{U}, \mathsf{W}) - \mathrm{H}(\mathsf{Y}, \mathsf{U}, \mathsf{W}), \tag{8.261}$$

and therefore

$$\mathrm{H}(\tau) - \mathrm{H}\big((\Phi\Xi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(uu^*)\big) \leq \mathrm{H}\big(\Xi_a(\xi)\big) - \mathrm{H}\big((\Psi\Xi_a)(\xi)\big). \tag{8.262}$$

Finally, in accordance with the probability vector p, one may average the two sides of (8.262) over all $a \in \Sigma$, obtaining

$$\begin{aligned} \mathrm{H}(\tau) - \sum_{a\in\Sigma} p(a)\,\mathrm{H}\big((\Phi\Xi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(uu^*)\big) \\ \leq \sum_{a\in\Sigma} p(a)\big(\mathrm{H}\big(\Xi_a(\xi)\big) - \mathrm{H}\big((\Psi\Xi_a)(\xi)\big)\big). \end{aligned} \tag{8.263}$$

Lemma 8.40 therefore implies that

$$\mathrm{H}(\tau) - \sum_{a\in\Sigma} p(a)\,\mathrm{H}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho_a)\big) \leq \mathrm{H}(\sigma) - \mathrm{H}(\Psi(\sigma)). \tag{8.264}$$

By the subadditivity of the von Neumann entropy (Proposition 5.10) one has

$$\mathrm{H}\left(\sum_{a\in\Sigma} p(a)(\Phi\Xi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(uu^*)\right) \leq \mathrm{H}(\Phi(\sigma)) + \mathrm{H}(\tau). \tag{8.265}$$

The inequality (8.252) follows from (8.264) and (8.265), which completes the proof. □

The entanglement-assisted classical capacity theorem

Finally, the entanglement-assisted classical capacity theorem will be stated, and proved through the use of the lemmas presented above.

Theorem 8.44 (Entanglement-assisted classical capacity theorem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. It holds that*

$$\mathrm{C}_{\mathrm{E}}(\Phi) = \max_{\sigma\in\mathrm{D}(\mathcal{X})} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{C}}(\sigma\,;\,\Phi)\big). \tag{8.266}$$

Proof. By Lemmas 8.41 and 8.43, one may conclude that

$$\begin{aligned} \chi_{\mathrm{E}}(\Phi^{\otimes n}) &\leq \max_{\sigma\in\mathrm{D}(\mathcal{X}^{\otimes n})} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{C}}\big(\sigma\,;\,\Phi^{\otimes n}\big)\big) \\ &= n \max_{\sigma\in\mathrm{D}(\mathcal{X})} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{C}}\big(\sigma\,;\,\Phi\big)\big) \end{aligned} \tag{8.267}$$

for every positive integer n. By Theorem 8.31, it therefore follows that

$$\mathrm{C}_{\mathrm{E}}(\Phi) = \lim_{n\to\infty} \frac{\chi_{\mathrm{E}}\big(\Phi^{\otimes n}\big)}{n} \leq \max_{\sigma\in\mathrm{D}(\mathcal{X})} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{C}}\big(\sigma\,;\,\Phi\big)\big). \tag{8.268}$$

For the reverse inequality, one may first choose a complex Euclidean space $\mathcal{Z}$ and an isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y}\otimes\mathcal{Z})$ such that

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(AXA^*\big) \tag{8.269}$$

for all $X \in \mathrm{L}(\mathcal{X})$. It holds that the channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$, defined by

$$\Psi(X) = \mathrm{Tr}_{\mathcal{Y}}\big(AXA^*\big) \tag{8.270}$$

for all $X \in \mathrm{L}(\mathcal{X})$, is complementary to Φ, so that Proposition 8.17 implies

$$\mathrm{I}_{\mathrm{C}}(\sigma\,;\,\Phi) = \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)) \tag{8.271}$$

for all $\sigma \in \mathrm{D}(\mathcal{X})$.

Let $\sigma \in \mathrm{D}(\mathcal{X})$ be any density operator, let $\delta > 0$ be chosen arbitrarily, and choose $\varepsilon > 0$ to be sufficiently small so that

$$(7\,\mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) + \mathrm{H}(\Psi(\sigma)))\varepsilon < \delta. \tag{8.272}$$

By Lemma 8.39, one may conclude that the inequality

$$\begin{aligned} \frac{1}{n}\Big(\mathrm{H}(\omega_{n,\varepsilon}) + \mathrm{H}(\Phi^{\otimes n}(\omega_{n,\varepsilon})) - \mathrm{H}(\Psi^{\otimes n}(\omega_{n,\varepsilon}))\Big) \\ \geq \mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)) - \delta \end{aligned} \tag{8.273}$$

holds for all but finitely many positive integers n, where

$$\omega_{n,\varepsilon} = \frac{\Lambda_{n,\varepsilon}}{\mathrm{Tr}(\Lambda_{n,\varepsilon})} \tag{8.274}$$

and $\Lambda_{n,\varepsilon}$ denotes the ε-strongly typical projection with respect to any fixed spectral decomposition of σ. By Lemma 8.42, it follows that

$$\frac{\chi_{\mathrm{E}}(\Phi^{\otimes n})}{n} \geq \mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)) - \delta \tag{8.275}$$

for all but finitely many positive integers n, and therefore

$$\mathrm{C}_{\mathrm{E}}(\Phi) = \lim_{n\to\infty} \frac{\chi_{\mathrm{E}}\big(\Phi^{\otimes n}\big)}{n} \geq \mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)) - \delta. \tag{8.276}$$

As this inequality holds for all $\delta > 0$, one has

$$\mathrm{C}_{\mathrm{E}}(\Phi) \geq \mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)) = \mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{C}}\big(\sigma\,;\,\Phi\big), \tag{8.277}$$

and maximizing over all $\sigma \in \mathrm{D}(\mathcal{X})$ completes the proof. □

8.2 Quantum information over quantum channels

This section is concerned with the capacity of quantum channels to transmit quantum information from a sender to a receiver. Along similar lines to the classical capacities considered in the previous section, one may consider the quantum capacity of a channel both when the sender and receiver share prior entanglement, used to assist the information transmission, and when they do not.

As it turns out, the capacity of a quantum channel to transmit quantum information with the assistance of entanglement is, in all cases, one-half of the entanglement-assisted classical capacity of the same channel. This fact is proved below through a combination of the teleportation and dense coding protocols discussed in Section 6.3.1. As the entanglement-assisted classical capacity has already been characterized by Theorem 8.44, a characterization of the capacity of a quantum channel to transmit quantum information with the assistance of entanglement follows directly. For this reason, the primary focus of the section is on an analysis of the capacity of quantum channels to transmit quantum information without the assistance of entanglement.

The first subsection below presents a definition of the quantum capacity of a channel, together with the closely related notion of a channel's capacity to generate shared entanglement. The second subsection presents a proof of the quantum capacity theorem, which characterizes the capacity of a given channel to transmit quantum information.

8.2.1 Definitions of quantum capacity and related notions

Definitions of the *quantum capacity* and the *entanglement-generation capacity* of a quantum channel are presented below, and it is proved that the two quantities coincide. The *entanglement-assisted quantum capacity* of a quantum channel is also defined, and its fairly straightforward relationship to the entanglement-assisted classical capacity of a channel is clarified.

The quantum capacity of a channel

Informally speaking, the quantum capacity of a channel is the number of qubits, on average, that can be accurately transmitted with each use of that channel. Like the capacities discussed in the previous section, the quantum capacity of a channel is defined in information-theoretic terms, referring to

a situation in which an asymptotically large number of channel uses, acting on a collection of possibly entangled registers, is made available.

The definition of quantum capacity that follows makes use of the same notions of an emulation of one channel by another (Definition 8.1) and of an ε-approximation of one channel by another (Definition 8.2) that were used in the previous section.

Definition 8.45 (Quantum capacity of a channel)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Also let $\Gamma = \{0, 1\}$ denote the binary alphabet, and let $\mathcal{Z} = \mathbb{C}^{\Gamma}$.

1. A value $\alpha \geq 0$ is an *achievable rate* for quantum information transmission through Φ if and only if (i) $\alpha = 0$, or (ii) $\alpha > 0$ and the following holds for every choice of a positive real number $\varepsilon > 0$: for all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, the channel $\Phi^{\otimes n}$ emulates an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}$.
2. The *quantum capacity* of Φ, which is denoted $\mathrm{Q}(\Phi)$, is defined as the supremum of all achievable rates for quantum information transmission through Φ.

Similar to the classical capacities considered in the previous section, the argument through which Proposition 8.4 was proved yields an analogous proposition for the quantum capacity.

Proposition 8.46. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, and let k be a positive integer. It holds that $\mathrm{Q}(\Phi^{\otimes k}) = k\,\mathrm{Q}(\Phi)$.*

The entanglement generation capacity of a channel

The *entanglement generation capacity* of a channel is defined in a similar way to the quantum capacity, except that the associated task is more narrowly focused—by means of multiple, independent uses of a channel, a sender and receiver aim to establish a state, shared between them, having high fidelity with a maximally entangled state.

Definition 8.47 (Entanglement generation capacity of a channel)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Also let $\Gamma = \{0, 1\}$ denote the binary alphabet, and let $\mathcal{Z} = \mathbb{C}^{\Gamma}$.

1. A value $\alpha \geq 0$ is an *achievable rate* for entanglement generation through Φ if and only if (i) $\alpha = 0$, or (ii) $\alpha > 0$ and the following holds for every choice of a positive real number $\varepsilon > 0$: for all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, there exists a state $\rho \in \mathrm{D}(\mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m})$ and a channel $\Xi \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$ such that

$$\mathrm{F}\Big(2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^*, \big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\rho)\Big) \geq 1 - \varepsilon. \tag{8.278}$$

2. The *entanglement generation capacity* of Φ, denoted $\mathrm{Q}_{\mathrm{EG}}(\Phi)$, is defined as the supremum of all achievable rates for entanglement generation through Φ.

Remark 8.48. For any choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, a unit vector $y \in \mathcal{Y}$, and a channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, the maximum value for the fidelity $\mathrm{F}(yy^*, \Psi(\rho))$ over $\rho \in \mathrm{D}(\mathcal{X})$ is achieved when ρ is a pure state. It follows from this observation that the quantity $\mathrm{Q}_{\mathrm{EG}}(\Phi)$ would not change if the states $\rho \in \mathrm{D}(\mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m})$ considered in the specification of achievable rates in Definition 8.47 are constrained to be pure states.

Equivalence of quantum capacity and entanglement generation capacity

The task associated with entanglement generation capacity would seem to be more specialized than the one associated with quantum capacity; it is apparent that the emulation of a close approximation to an identity channel allows a sender and receiver to generate a shared state having high fidelity with a maximally entangled state, but it is not clear that the reverse should be true. The relationship between entanglement generation and identity channel emulation provided by the following theorem allows one to prove that the reverse implication does indeed hold: the quantum capacity and entanglement generation capacity of any given channel always coincide.

Theorem 8.49. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $n = \dim(\mathcal{Y})$, and assume $\dim(\mathcal{Z}) \leq n/2$. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, let $u \in \mathcal{X} \otimes \mathcal{Y}$ be a unit vector, let $\varepsilon \geq 0$ be a nonnegative real number, and assume the inequality*

$$\mathrm{F}\left(\frac{1}{n} \operatorname{vec}(\mathbb{1}_{\mathcal{Y}}) \operatorname{vec}(\mathbb{1}_{\mathcal{Y}})^*, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(uu^*)\right) \geq 1 - \varepsilon \tag{8.279}$$

is satisfied. The channel Φ emulates a δ-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}$ for $\delta = \big(4\sqrt{2}\big)\varepsilon^{\frac{1}{4}}$.

Proof. Let $A \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ be the operator defined by the equation $\mathrm{vec}(A) = u$, and let

$$A = \sum_{k=1}^{r} \sqrt{p_k} x_k y_k^* \tag{8.280}$$

be a singular value decomposition of A, where $r \leq n$ is the rank of A, $\{x_1, \ldots, x_r\} \subset \mathcal{X}$ and $\{y_1, \ldots, y_r\} \subset \mathcal{Y}$ are orthonormal sets, and $(p_1, \ldots, p_r)$ is a probability vector. Define $W \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ as

$$W = \sum_{k=1}^{r} x_k y_k^*, \tag{8.281}$$

and define a unit vector $v \in \mathcal{X} \otimes \mathcal{Y}$ as

$$v = \frac{1}{\sqrt{r}} \mathrm{vec}(W). \tag{8.282}$$

By the monotonicity of the fidelity function under partial tracing, one has

$$\begin{aligned} \mathrm{F}(uu^*, vv^*) &= \frac{1}{\sqrt{r}} \sum_{k=1}^{r} \sqrt{p_k} \geq \frac{1}{\sqrt{n}} \sum_{k=1}^{r} \sqrt{p_k} = \mathrm{F}\left(\frac{1}{n} \mathbb{1}_{\mathcal{Y}}, \mathrm{Tr}_{\mathcal{X}}(uu^*)\right) \\ &\geq \mathrm{F}\left(\frac{1}{n} \mathrm{vec}(\mathbb{1}_{\mathcal{Y}}) \mathrm{vec}(\mathbb{1}_{\mathcal{Y}})^*, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(uu^*)\right) \geq 1 - \varepsilon. \end{aligned} \tag{8.283}$$

Consequently, by Theorems 3.30 and 3.32, one has

$$\begin{aligned} &\mathrm{F}\left(\frac{1}{n} \mathrm{vec}(\mathbb{1}_{\mathcal{Y}}) \mathrm{vec}(\mathbb{1}_{\mathcal{Y}})^*, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(vv^*)\right) + 1 \\ &\quad \geq \mathrm{F}\left(\frac{1}{n} \mathrm{vec}(\mathbb{1}_{\mathcal{Y}}) \mathrm{vec}(\mathbb{1}_{\mathcal{Y}})^*, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(uu^*)\right)^2 + \mathrm{F}\big(vv^*, uu^*\big)^2 \\ &\quad \geq 2(1-\varepsilon)^2, \end{aligned} \tag{8.284}$$

and therefore

$$\mathrm{F}\left(\frac{1}{n} \mathrm{vec}(\mathbb{1}_{\mathcal{Y}}) \mathrm{vec}(\mathbb{1}_{\mathcal{Y}})^*, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(vv^*)\right) \geq 1 - 4\varepsilon. \tag{8.285}$$

Next, define a projection operator $\Pi_r = W^* W \in \mathrm{Proj}(\mathcal{Y})$ and define $\mathcal{V}_r = \mathrm{im}(\Pi_r)$. For each choice of k beginning with r and decreasing to 1, choose $w_k \in \mathcal{V}_k$ to be a unit vector that minimizes the quantity

$$\alpha_k = \big\langle w_k w_k^*, \Phi\big(W w_k w_k^* W^*\big)\big\rangle, \tag{8.286}$$

and define

$$\mathcal{V}_{k-1} = \{z \in \mathcal{V}_k \,:\, \langle w_k, z\rangle = 0\}. \tag{8.287}$$

Observe that $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r$ and that $\{w_1, \ldots, w_k\}$ is an orthonormal basis for $\mathcal{V}_k$, for each $k \in \{1, \ldots, r\}$. In particular, it holds that

$$v = \frac{1}{\sqrt{r}}(W \otimes \mathbb{1}_{\mathcal{Y}}) \operatorname{vec}(\Pi_r) = \frac{1}{\sqrt{r}} \sum_{k=1}^{r} W w_k \otimes \overline{w_k}. \tag{8.288}$$

At this point, a calculation reveals that

$$\begin{aligned} &\mathrm{F}\biggl(\frac{1}{n}\operatorname{vec}(\mathbb{1}_{\mathcal{Y}})\operatorname{vec}(\mathbb{1}_{\mathcal{Y}})^*, \bigl(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\bigr)(vv^*)\biggr)^2 \\ &\qquad = \frac{1}{nr} \sum_{j,k \in \{1,\ldots,r\}} \bigl\langle w_j w_k^*, \Phi(W w_j w_k^* W^*)\bigr\rangle. \end{aligned} \tag{8.289}$$

By the complete positivity of Φ, one may conclude that

$$\begin{aligned} &\bigl|\bigl\langle w_j w_k^*, \Phi(W w_j w_k^* W^*)\bigr\rangle\bigr| \\ &\qquad \leq \sqrt{\bigl\langle w_j w_j^*, \Phi(W w_j w_j^* W^*)\bigr\rangle}\sqrt{\bigl\langle w_k w_k^*, \Phi(W w_k w_k^* W^*)\bigr\rangle} \\ &\qquad = \sqrt{\alpha_j \alpha_k}, \end{aligned} \tag{8.290}$$

for each choice of $j, k \in \{1, \ldots, r\}$. Therefore, by the triangle inequality one may conclude that

$$\mathrm{F}\biggl(\frac{1}{n}\operatorname{vec}(\mathbb{1}_{\mathcal{Y}})\operatorname{vec}(\mathbb{1}_{\mathcal{Y}})^*, \bigl(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\bigr)(vv^*)\biggr) \leq \frac{1}{\sqrt{nr}} \sum_{k=1}^{r} \sqrt{\alpha_k}. \tag{8.291}$$

Applying the Cauchy–Schwarz inequality, one obtains

$$\frac{1}{\sqrt{nr}} \sum_{k=1}^{r} \sqrt{\alpha_k} \leq \sqrt{\frac{1}{n}\sum_{k=1}^{r} \alpha_k}, \tag{8.292}$$

and therefore

$$\frac{1}{n}\sum_{k=1}^{r} \alpha_k \geq (1 - 4\varepsilon)^2 \geq 1 - 8\varepsilon. \tag{8.293}$$

Now choose $m \in \{0, \ldots, r\}$ to be the maximum value of m for which it holds that $\alpha_k \geq 1 - 16\varepsilon$ for all $k \leq m$. By (8.293) it necessarily holds that $m \geq n/2$. By the definition of the values $\alpha_1, \ldots, \alpha_r$, one may conclude that

$$\bigl\langle ww^*, \Phi(Www^*W^*)\bigr\rangle \geq 1 - 16\varepsilon \tag{8.294}$$

for every unit vector $w \in \mathcal{V}_m$. Finally, let $V \in \mathrm{U}(\mathcal{Z}, \mathcal{Y})$ be any isometry with the property that $\mathrm{im}(V) \subseteq \mathcal{V}_m$. Such an isometry exists by the assumption that $\dim(\mathcal{Z}) \leq n/2$ together with the fact that $n/2 \leq m = \dim(\mathcal{V}_m)$. Let $\Xi_{\mathrm{E}} \in \mathrm{C}(\mathcal{Z}, \mathcal{X})$ and $\Xi_{\mathrm{D}} \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ be channels of the form

$$\Xi_{\mathrm{E}}(Z) = WVZV^*W^* + \Psi_{\mathrm{E}}(Z) \quad \text{and} \quad \Xi_{\mathrm{D}}(Y) = V^*YV + \Psi_{\mathrm{D}}(Y), \tag{8.295}$$

for all $Z \in \mathrm{L}(\mathcal{Z})$ and $Y \in \mathrm{L}(\mathcal{Y})$, where $\Psi_{\mathrm{E}} \in \mathrm{CP}(\mathcal{Z}, \mathcal{X})$ and $\Psi_{\mathrm{D}} \in \mathrm{CP}(\mathcal{Y}, \mathcal{Z})$ are completely positive maps that cause Ξ_{E} and Ξ_{D} to be trace-preserving. For every unit vector $z \in \mathcal{Z}$ it holds that

$$\big\langle zz^*, (\Xi_{\mathrm{D}}\Phi\Xi_{\mathrm{E}})(zz^*)\big\rangle \geq \big\langle Vzz^*V^*, \Phi(WVzz^*V^*W^*)\big\rangle \geq 1 - 16\varepsilon, \tag{8.296}$$

and therefore one has

$$\big\| zz^* - (\Xi_{\mathrm{D}}\Phi\Xi_{\mathrm{E}})(zz^*) \big\|_1 \leq 8\sqrt{\varepsilon} \tag{8.297}$$

by one of the Fuchs–van de Graaf inequalities (Theorem 3.36). Applying Theorem 3.58, one therefore finds that

$$\big\| \Xi_{\mathrm{D}}\Phi\Xi_{\mathrm{E}} - \mathbb{1}_{\mathrm{L}(\mathcal{Z})} \big\|_1 \leq \big(4\sqrt{2}\big)\varepsilon^{\frac{1}{4}}, \tag{8.298}$$

which completes the proof. □

Theorem 8.50. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. The quantum capacity and entanglement generation capacity of Φ are equal:* $\mathrm{Q}(\Phi) = \mathrm{Q}_{\mathrm{E}}(\Phi)$.

Proof. It will first be proved that $\mathrm{Q}(\Phi) \leq \mathrm{Q}_{\mathrm{E}}(\Phi)$, which is straightforward. If the quantum capacity of Φ is zero, there is nothing to prove, so it will be assumed that $\mathrm{Q}(\Phi) > 0$. Let $\alpha > 0$ be an achievable rate for quantum information transmission through Φ, and let $\varepsilon > 0$ be chosen arbitrarily.

Setting $\Gamma = \{0, 1\}$ and $\mathcal{Z} = \mathbb{C}^{\Gamma}$, one therefore has that the channel $\Phi^{\otimes n}$ emulates an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}$ for all but finitely many positive integers n and for all positive integers $m \leq \alpha n$. That is, for all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, there must exist channels $\Xi_{\mathrm{E}} \in \mathrm{C}(\mathcal{Z}^{\otimes m}, \mathcal{X}^{\otimes n})$ and $\Xi_{\mathrm{D}} \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$ such that

$$\big\| \Xi_{\mathrm{D}}\Phi^{\otimes n}\Xi_{\mathrm{E}} - \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m} \big\|_1 < \varepsilon. \tag{8.299}$$

Supposing that n and m are positive integers for which such channels exist, one may consider the density operators

$$\tau = 2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^* \quad \text{and} \quad \rho = \big(\Xi_{\mathrm{E}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\tau), \tag{8.300}$$

along with the channel $\Xi = \Xi_{\mathrm{D}}$. One of the Fuchs–van de Graaf inequalities (Theorem 3.36) implies that

$$\begin{aligned} \mathrm{F}\Big(\tau, \big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\rho)\Big) &= \mathrm{F}\Big(\tau, \big(\Xi_{\mathrm{D}}\Phi^{\otimes n}\Xi_{\mathrm{E}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\tau)\Big) \\ &\geq 1 - \frac{1}{2}\big\| \big(\Xi_{\mathrm{D}}\Phi^{\otimes n}\Xi_{\mathrm{E}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\tau) - \tau \big\|_1 \geq 1 - \frac{\varepsilon}{2}. \end{aligned} \tag{8.301}$$

As this is so for all but finitely many positive integers n and all positive integers $m \leq \alpha n$, it is the case that α is an achievable rate for entanglement generation through Φ. Taking the supremum over all achievable rates α for quantum communication through Φ, one obtains $\mathrm{Q}(\Phi) \leq \mathrm{Q}_{\mathrm{E}}(\Phi)$.

It remains to prove that $\mathrm{Q}_{\mathrm{E}}(\Phi) \leq \mathrm{Q}(\Phi)$. As for the reverse inequality just proved, there is nothing to prove if $\mathrm{Q}_{\mathrm{E}}(\Phi) = 0$, so it will be assumed that $\mathrm{Q}_{\mathrm{E}}(\Phi) > 0$. Let $\alpha > 0$ be an achievable rate for entanglement generation through Φ and let $\beta \in (0, \alpha)$ be chosen arbitrarily. It will be proved that β is an achievable rate for quantum communication through Φ. The required relation $\mathrm{Q}_{\mathrm{E}}(\Phi) \leq \mathrm{Q}(\Phi)$ follows by taking the supremum over all achievable rates α for entanglement generation through Φ and over all $\beta \in (0, \alpha)$.

Let $\varepsilon > 0$ be chosen arbitrarily and let $\delta = \varepsilon^4/1024$, so that $\big(4\sqrt{2}\big)\delta^{\frac{1}{4}} = \varepsilon$. One has that, for all but finitely many positive integers n and all positive integers $m \leq \alpha n$, that there exists a state $\rho \in \mathrm{D}(\mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m})$ and a channel $\Xi \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$ such that

$$\mathrm{F}\big(2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^*, \big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\rho)\big) \geq 1 - \delta. \tag{8.302}$$

Fix n and $m \leq \alpha n$ to be positive integers for which this statement holds, and observe that the function

$$\begin{aligned} \rho \mapsto \mathrm{F}\big(2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^*, \big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\rho)\big)^2 \\ = \big\langle 2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^*, \big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\rho)\big\rangle \end{aligned} \tag{8.303}$$

must achieve its maximum value (over all density operators) on a pure state. Thus, there must exist a unit vector $u \in \mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m}$ so that the inequality (8.302) holds when $\rho = uu^*$. By Theorem 8.49, it follows that $\Phi^{\otimes n}$ emulates

an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes k}$ for every positive integer $k \leq m - 1$.

Under the assumption $n \geq 1/(\alpha - \beta)$, one has that $\beta n \leq \alpha n - 1$. Thus, for all but finitely many positive integers n and all positive integers $k \leq \beta n$, it holds that $\Phi^{\otimes n}$ emulates an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes k}$. As $\varepsilon > 0$ has been chosen arbitrarily, it follows that β is an achievable rate for quantum communication through Φ, which completes the proof. $\square$

The entanglement-assisted quantum capacity of a channel

The entanglement-assisted quantum capacity of a channel, which will be proved is equal to one-half of its entanglement-assisted classical capacity, may be formally defined as follows.

Definition 8.51 (Entanglement-assisited quantum capacity of a channel)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Also let $\Gamma = \{0, 1\}$ denote the binary alphabet, and let $\mathcal{Z} = \mathbb{C}^{\Gamma}$.

1. A value $\alpha \geq 0$ is an *achievable rate* for entanglement-assisted quantum information transmission through Φ if and only if (i) $\alpha = 0$, or (ii) $\alpha > 0$ and the following holds for every choice of a positive real number $\varepsilon > 0$: for all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, the channel $\Phi^{\otimes n}$ emulates an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}$ with the assistance of entanglement.
2. The *entanglement-assisted quantum capacity* of Φ, denoted $\mathrm{Q}_{\mathrm{E}}(\Phi)$, is the supremum of all achievable rates for entanglement-assisted quantum information transmission through Φ.

Proposition 8.52. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. It holds that*

$$\mathrm{Q}_{\mathrm{E}}(\Phi) = \frac{1}{2}\mathrm{C}_{\mathrm{E}}(\Phi). \tag{8.304}$$

Proof. Assume α is an achievable rate for entanglement-assisted classical communication through Φ. If $\alpha = 0$, then $\alpha/2$ is trivially an achievable rate for entanglement-assisted quantum information transmission through Φ. It will be observed that if $\alpha > 0$, then $\alpha/2 - \delta$ is an achievable rate for entanglement-assisted quantum communication through Φ for all real numbers $\delta \in (0, \alpha/2)$. Taking the supremum over all achievable rates α for

entanglement-assisted classical communication through Φ and the infimum over all $\delta \in (0, \alpha/2)$, one obtains

$$\mathrm{Q}_{\mathrm{E}}(\Phi) \geq \frac{1}{2}\mathrm{C}_{\mathrm{E}}(\Phi). \tag{8.305}$$

Suppose n and $m \leq \alpha n$ are positive integers and $\varepsilon > 0$ is a positive real number such that $\Phi^{\otimes n}$ emulates an ε-approximation to the channel $\Delta^{\otimes m}$, where $\Delta \in \mathrm{C}(\mathcal{Z})$ denotes the completely dephasing channel as usual. Let $k = \lfloor m/2 \rfloor$, and consider the maximally entangled state

$$\tau = 2^{-k}\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes k}\big)\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes k}\big)^{*}. \tag{8.306}$$

By tensoring τ with the state ξ used for the emulation of an ε-approximation to $\Delta^{\otimes m}$ by $\Phi^{\otimes n}$, one may define a new channel $\Psi \in \mathrm{C}(\mathcal{Z}^{\otimes k})$ through the use of the traditional teleportation protocol (q.v. Example 6.56 in Section 6.3.1), but where the classical communication channel required for teleportation is replaced by the ε-approximation to the channel $\Delta^{\otimes m}$ emulated by $\Phi^{\otimes n}$. It holds that Ψ is an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes k}$.

One therefore has that, for all $\varepsilon > 0$, for all but finitely many values of n, and for all $k \leq (\alpha n - 1)/2$, that $\Phi^{\otimes n}$ emulates an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes k}$ through the assistance of entanglement. For every $\delta \in (0, \alpha/2)$, it is therefore the case that $\alpha/2 - \delta$ is an achievable rate for entanglement-assisted quantum communication through Φ, as required.

Now assume α is an achievable rate for entanglement-assisted quantum communication through Φ. It will be proved that 2α is an achievable rate for entanglement-assisted classical communication through Φ. This statement is trivial in the case $\alpha = 0$, so it will be assumed that $\alpha > 0$. The proof is essentially the same as the reverse direction just considered, with dense coding replacing teleportation.

Suppose that n and $m \leq \alpha n$ are positive integers and $\varepsilon > 0$ is a positive real number such that $\Phi^{\otimes n}$ emulates an ε-approximation to $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}$. Using the maximally entangled state

$$\tau = 2^{-m}\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\big)\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\big)^{*}, \tag{8.307}$$

tensored with the state ξ used for the emulation of $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}$ by $\Phi^{\otimes n}$, one may define a new channel $\Psi \in \mathrm{C}(\mathcal{Z}^{\otimes 2m})$ through the traditional dense coding protocol (q.v. Example 6.61 in Section 6.3.1), where the quantum channel

required for dense coding is replaced by the ε-approximation to the channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}$ emulated by $\Phi^{\otimes n}$. It holds that Ψ is an ε-approximation to $\Delta^{\otimes 2m}$.

It therefore holds that, for all $\varepsilon > 0$, for all but finitely many values of n, and for all $m \leq \alpha n$, that $\Phi^{\otimes n}$ emulates an ε-approximation to the channel $\Delta^{\otimes 2m}$, which implies that 2α is an achievable rate for entanglement-assisted classical communication through Φ. The inequality

$$\mathrm{C}_{\mathrm{E}}(\Phi) \geq 2\,\mathrm{Q}_{\mathrm{E}}(\Phi) \tag{8.308}$$

is obtained when one takes the supremum over all achievable rates α for entanglement-assisted quantum communication through Φ.

The equality (8.304) therefore holds, which completes the proof. □

8.2.2 The quantum capacity theorem

The purpose of the present subsection is to state and prove the quantum capacity theorem, which yields an expression for the quantum capacity of a given channel. Similar to the Holevo–Schumacher–Westmoreland theorem (Theorem 8.30), the expression that is obtained from the quantum capacity theorem includes a regularization over an increasing number of uses of a given channel.

The subsections that follow include statements and proofs of various lemmas that will be used to prove the quantum capacity theorem, along with the statement and proof of the theorem itself.

A decoupling lemma

The first of several lemmas that will be used to prove the quantum capacity theorem concerns a phenomenon known as *decoupling*. Informally speaking, this is the phenomenon whereby the action of a sufficiently noisy channel on a randomly chosen subspace of its input space can be expected not only to destroy entanglement with an auxiliary system, but to destroy classical correlations as well. The lemma that follows proves a fact along these lines that is specialized to the task at hand.

Lemma 8.53. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{W}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $n = \dim(\mathcal{X})$ and $m = \dim(\mathcal{Z})$, and assume $m \leq n \leq \dim(\mathcal{Y} \otimes \mathcal{W})$. Assume moreover that $V \in \mathrm{U}(\mathcal{Z}, \mathcal{X})$ and $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ are isometries, and define density operators*

$\xi \in \mathrm{D}(\mathcal{W} \otimes \mathcal{X})$ *and* $\rho_U \in \mathrm{D}(\mathcal{W} \otimes \mathcal{Z})$*, for each unitary operator* $U \in \mathrm{U}(\mathcal{X})$*, as follows:*

$$\begin{aligned} \xi &= \frac{1}{n} \operatorname{Tr}_{\mathcal{Y}}\bigl(\operatorname{vec}(A)\operatorname{vec}(A)^*\bigr), \\ \rho_U &= \frac{1}{m} \operatorname{Tr}_{\mathcal{Y}}\bigl(\operatorname{vec}(AUV)\operatorname{vec}(AUV)^*\bigr). \end{aligned} \tag{8.309}$$

It holds that

$$\int \bigl\| \rho_U - \operatorname{Tr}_{\mathcal{Z}}(\rho_U) \otimes \omega \bigr\|_2^2 \, \mathrm{d}\eta(U) \leq \operatorname{Tr}(\xi^2)\,, \tag{8.310}$$

for η *denoting the Haar measure on* $\mathrm{U}(\mathcal{X})$ *and* $\omega \in \mathrm{D}(\mathcal{Z})$ *being the completely mixed state on* $\mathcal{Z}$*.*

Proof. Observe first that

$$\bigl\| \rho_U - \operatorname{Tr}_{\mathcal{Z}}(\rho_U) \otimes \omega \bigr\|_2^2 = \operatorname{Tr}\bigl(\rho_U^2\bigr) - \frac{1}{m} \operatorname{Tr}\Bigl(\bigl(\operatorname{Tr}_{\mathcal{Z}}(\rho_U)\bigr)^2 \Bigr). \tag{8.311}$$

The lemma requires a bound on the integral of the expression represented by (8.311) over all U, and toward this goal the two terms on the right-hand side of that equation will be integrated separately.

To integrate the first term on the right-hand side of (8.311), let Γ be the alphabet for which $\mathcal{Y} = \mathbb{C}^\Gamma$, define $B_a = \bigl(e_a^* \otimes \mathbb{1}_{\mathcal{W}}\bigr) A$ for each $a \in \Gamma$, and observe that

$$\rho_U = \frac{1}{m} \sum_{a \in \Gamma} \operatorname{vec}(B_a U V) \operatorname{vec}(B_a U V)^*. \tag{8.312}$$

It therefore holds that

$$\begin{aligned} \operatorname{Tr}\bigl(\rho_U^2\bigr) &= \frac{1}{m^2} \sum_{a,b \in \Gamma} \bigl| \operatorname{Tr}\bigl(V^* U^* B_a^* B_b U V\bigr) \bigr|^2 \\ &= \frac{1}{m^2} \sum_{a,b \in \Gamma} \operatorname{Tr}\bigl(V^* U^* B_a^* B_b U V \otimes V^* U^* B_b^* B_a U V\bigr) \\ &= \Bigl\langle U V V^* U^* \otimes U V V^* U^*, \frac{1}{m^2} \sum_{a,b \in \Gamma} B_a^* B_b \otimes B_b^* B_a \Bigr\rangle. \end{aligned} \tag{8.313}$$

Integrating over all $U \in \mathrm{U}(\mathcal{X})$ yields

$$\int \operatorname{Tr}\bigl(\rho_U^2\bigr) \, \mathrm{d}\eta(U) = \Bigl\langle \Xi\bigl(V V^* \otimes V V^*\bigr), \frac{1}{m^2} \sum_{a,b \in \Gamma} B_a^* B_b \otimes B_b^* B_a \Bigr\rangle, \tag{8.314}$$

for $\Xi \in \mathrm{C}(\mathcal{X} \otimes \mathcal{X})$ denoting the Werner twirling channel (q.v. Example 7.26 in the previous chapter). Making use of the expression

$$\Xi(X) = \frac{2}{n(n+1)} \langle \Pi_{\mathcal{X} \circledS \mathcal{X}}, X \rangle \, \Pi_{\mathcal{X} \circledS \mathcal{X}} + \frac{2}{n(n-1)} \langle \Pi_{\mathcal{X} \circledA \mathcal{X}}, X \rangle \, \Pi_{\mathcal{X} \circledA \mathcal{X}}, \tag{8.315}$$

which holds for every $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$, and observing the equations

$$\big\langle \Pi_{\mathcal{X} \circledS \mathcal{X}}, VV^* \otimes VV^* \big\rangle = \frac{m(m+1)}{2}, \tag{8.316}$$

$$\big\langle \Pi_{\mathcal{X} \circledA \mathcal{X}}, VV^* \otimes VV^* \big\rangle = \frac{m(m-1)}{2}, \tag{8.317}$$

it follows that

$$\begin{aligned} &\int \mathrm{Tr}\big(\rho_U^2\big) \, \mathrm{d}\eta(U) \\ &\quad = \Big\langle \frac{m(m+1)}{n(n+1)} \Pi_{\mathcal{X} \circledS \mathcal{X}} + \frac{m(m-1)}{n(n-1)} \Pi_{\mathcal{X} \circledA \mathcal{X}}, \frac{1}{m^2} \sum_{a,b \in \Gamma} B_a^* B_b \otimes B_b^* B_a \Big\rangle. \end{aligned} \tag{8.318}$$

A similar methodology can be used to integrate the second term on the right-hand side of (8.311). In particular, one has

$$\mathrm{Tr}_{\mathcal{Z}}(\rho_U) = \frac{1}{m} \sum_{a \in \Gamma} B_a U V V^* U^* B_a^*, \tag{8.319}$$

and therefore

$$\begin{aligned} &\mathrm{Tr}\Big(\big(\mathrm{Tr}_{\mathcal{Z}}(\rho_U)\big)^2 \Big) \\ &\quad = \frac{1}{m^2} \sum_{a,b \in \Gamma} \mathrm{Tr}\big(V^* U^* B_a^* B_b U V V^* U^* B_b^* B_a U V \big) \\ &\quad = \Big\langle W_{\mathcal{Z}}, \frac{1}{m^2} \sum_{a,b \in \Gamma} V^* U^* B_a^* B_b U V \otimes V^* U^* B_b^* B_a U V \Big\rangle \\ &\quad = \Big\langle (UV \otimes UV) W_{\mathcal{Z}} (UV \otimes UV)^*, \frac{1}{m^2} \sum_{a,b \in \Gamma} B_a^* B_b \otimes B_b^* B_a \Big\rangle, \end{aligned} \tag{8.320}$$

where $W_{\mathcal{Z}} \in \mathrm{U}(\mathcal{Z} \otimes \mathcal{Z})$ denotes the swap operator on $\mathcal{Z} \otimes \mathcal{Z}$, and the second equality has used the identity $\langle W_{\mathcal{Z}}, X \otimes Y \rangle = \mathrm{Tr}(XY)$. Integrating over all $U \in \mathrm{U}(\mathcal{X})$ yields

$$\begin{aligned} &\int \mathrm{Tr}\Big(\big(\mathrm{Tr}_{\mathcal{Z}}(\rho_U)\big)^2 \Big) \, \mathrm{d}\eta(U) \\ &\quad = \Big\langle \Xi\big((V \otimes V) W_{\mathcal{Z}} (V \otimes V)^* \big), \frac{1}{m^2} \sum_{a,b \in \Gamma} B_a^* B_b \otimes B_b^* B_a \Big\rangle. \end{aligned} \tag{8.321}$$

By making use of the equations

$$\begin{aligned}\left\langle \Pi_{\mathcal{X}\circledcirc\mathcal{X}}, (V\otimes V)W_{\mathcal{Z}}(V\otimes V)^*\right\rangle &= \frac{m(m+1)}{2},\\ \left\langle \Pi_{\mathcal{X}\circledcirc\mathcal{X}}, (V\otimes V)W_{\mathcal{Z}}(V\otimes V)^*\right\rangle &= -\frac{m(m-1)}{2},\end{aligned} \tag{8.322}$$

and performing a similar calculation to the one above, one finds that

$$\begin{aligned}&\int \operatorname{Tr}\Big(\big(\operatorname{Tr}_{\mathcal{Z}}(\rho_U)\big)^2\Big)\,\mathrm{d}\eta(U)\\ &\quad= \left\langle \frac{m(m+1)}{n(n+1)}\Pi_{\mathcal{X}\circledcirc\mathcal{X}} - \frac{m(m-1)}{n(n-1)}\Pi_{\mathcal{X}\circledcirc\mathcal{X}}, \frac{1}{m^2}\sum_{a,b\in\Gamma} B_a^*B_b\otimes B_b^*B_a\right\rangle.\end{aligned} \tag{8.323}$$

Combining (8.318) and (8.323), together with some algebra, it follows that

$$\begin{aligned}&\int \left\|\rho_U - \operatorname{Tr}_{\mathcal{Z}}(\rho_U)\otimes\omega\right\|_2^2\,\mathrm{d}\eta(U)\\ &\qquad= \frac{m^2-1}{n^2-1}\left\langle \mathbb{1}_{\mathcal{X}}\otimes\mathbb{1}_{\mathcal{X}} - \frac{1}{n}W_{\mathcal{X}}, \frac{1}{m^2}\sum_{a,b\in\Gamma} B_a^*B_b\otimes B_b^*B_a\right\rangle,\end{aligned} \tag{8.324}$$

where $W_{\mathcal{X}}$ denotes the swap operator on $\mathcal{X}\otimes\mathcal{X}$. By similar calculations to (8.313) and (8.320) above, but replacing U and V by $\mathbb{1}_{\mathcal{X}}$, it may be verified that

$$\operatorname{Tr}(\xi^2) = \frac{1}{n^2}\operatorname{Tr}\left(\sum_{a,b\in\Gamma} B_a^*B_b\otimes B_b^*B_a\right) \tag{8.325}$$

and

$$\operatorname{Tr}\Big(\big(\operatorname{Tr}_{\mathcal{X}}(\xi)\big)^2\Big) = \frac{1}{n^2}\left\langle W_{\mathcal{X}}, \sum_{a,b\in\Gamma} B_a^*B_b\otimes B_b^*B_a\right\rangle. \tag{8.326}$$

Consequently,

$$\begin{aligned}&\int \left\|\rho_U - \operatorname{Tr}_{\mathcal{Z}}(\rho_U)\otimes\omega\right\|_2^2\,\mathrm{d}\eta(U)\\ &\quad= \frac{1-m^{-2}}{1-n^{-2}}\left(\operatorname{Tr}(\xi^2) - \frac{1}{n}\operatorname{Tr}\Big(\big(\operatorname{Tr}_{\mathcal{X}}(\xi)\big)^2\Big)\right) \le \operatorname{Tr}(\xi^2),\end{aligned} \tag{8.327}$$

as required. □

A lower-bound on entanglement generation decoding fidelity

The next lemma is used, within the proof of the quantum capacity theorem, to infer the existence of a decoding channel for the task of entanglement generation. This inference is based on a calculation involving a Stinespring representation of the channel through which entanglement generation is to be considered.

Lemma 8.54. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{W}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $m = \dim(\mathcal{Z})$, and assume that $m \leq \dim(\mathcal{X}) \leq \dim(\mathcal{Y} \otimes \mathcal{W})$. Also let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ and $V \in \mathrm{U}(\mathcal{Z}, \mathcal{X})$ be isometries, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be the channel defined by*

$$\Phi(X) = \mathrm{Tr}_{\mathcal{W}}\big(AXA^*\big) \tag{8.328}$$

for all $X \in \mathrm{L}(\mathcal{X})$, let $\rho \in \mathrm{D}(\mathcal{W} \otimes \mathcal{Z})$ be the state defined as

$$\rho = \frac{1}{m}\,\mathrm{Tr}_{\mathcal{Y}}\big(\mathrm{vec}(AV)\,\mathrm{vec}(AV)^*\big), \tag{8.329}$$

and let $\omega \in \mathrm{D}(\mathcal{Z})$ denote the completely mixed state on $\mathcal{Z}$. There exists a channel $\Xi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ such that

$$\begin{aligned} \mathrm{F}\Bigg(\frac{1}{m}\,\mathrm{vec}(\mathbb{1}_{\mathcal{Z}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{Z}})^*,\ \frac{1}{m}(\Xi\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})\big(\mathrm{vec}(V)\,\mathrm{vec}(V)^*\big)\Bigg) \\ \geq \mathrm{F}\big(\rho, \mathrm{Tr}_{\mathcal{Z}}(\rho) \otimes \omega\big). \end{aligned} \tag{8.330}$$

Proof. Let $\mathcal{V}$ be a complex Euclidean space of sufficiently large dimension that the inequalities $\dim(\mathcal{V}) \geq \dim(\mathcal{W})$ and $\dim(\mathcal{V} \otimes \mathcal{Z}) \geq \dim(\mathcal{Y})$ hold, and let $B \in \mathrm{L}(\mathcal{W}, \mathcal{V})$ be an operator such that $\mathrm{Tr}_{\mathcal{V}}\big(\mathrm{vec}(B)\,\mathrm{vec}(B)^*\big) = \mathrm{Tr}_{\mathcal{Z}}(\rho)$. For the vector

$$u = \frac{1}{\sqrt{m}}\,\mathrm{vec}(B \otimes \mathbb{1}_{\mathcal{Z}}) \in (\mathcal{V} \otimes \mathcal{Z}) \otimes (\mathcal{W} \otimes \mathcal{Z}), \tag{8.331}$$

one has that $\mathrm{Tr}_{\mathcal{V}\otimes\mathcal{Z}}(uu^*) = \mathrm{Tr}_{\mathcal{Z}}(\rho) \otimes \omega$. It is evident that the vector

$$v = \frac{1}{\sqrt{m}}\,\mathrm{vec}(AV) \in \mathcal{Y} \otimes \mathcal{W} \otimes \mathcal{Z} \tag{8.332}$$

satisfies $\mathrm{Tr}_{\mathcal{Y}}(vv^*) = \rho$, so it follows by Uhlmann's theorem (Theorem 3.23) that there exists an isometry $W \in \mathrm{U}(\mathcal{Y}, \mathcal{V} \otimes \mathcal{Z})$ such that

$$\mathrm{F}\big(\rho, \mathrm{Tr}_{\mathcal{Z}}(\rho) \otimes \omega\big) = \big|\big\langle u, (W \otimes \mathbb{1}_{\mathcal{W}\otimes\mathcal{Z}})v\big\rangle\big|. \tag{8.333}$$

Now define a channel $\Xi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ as

$$\Xi(Y) = \mathrm{Tr}_{\mathcal{V}}(WYW^*) \tag{8.334}$$

for every $Y \in \mathrm{L}(\mathcal{Y})$. By the monotonicity of the fidelity under partial tracing (which is a special case of Theorem 3.30), one has

$$\begin{aligned}
&\mathrm{F}\big(\rho, \mathrm{Tr}_{\mathcal{Z}}(\rho) \otimes \omega\big) \\
&\quad = \mathrm{F}\big(uu^*, (W \otimes \mathbb{1}_{\mathcal{W}\otimes\mathcal{Z}})vv^*(W \otimes \mathbb{1}_{\mathcal{W}\otimes\mathcal{Z}})^*\big) \\
&\quad \leq \mathrm{F}\Bigg(\mathrm{Tr}_{\mathcal{V}}\big(\mathrm{Tr}_{\mathcal{W}}(uu^*)\big), \frac{1}{m}(\Xi\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})\big(\mathrm{vec}(V)\,\mathrm{vec}(V)^*\big)\Bigg) \\
&\quad = \mathrm{F}\Bigg(\frac{1}{m}\,\mathrm{vec}(\mathbb{1}_{\mathcal{Z}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{Z}})^*, \frac{1}{m}(\Xi\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\mathrm{vec}(V)\,\mathrm{vec}(V)^*)\Bigg).
\end{aligned} \tag{8.335}$$

The channel Ξ therefore satisfies the requirement of the lemma. □

Two additional lemmas needed for the quantum capacity theorem

The two lemmas that follow represent technical facts that will be utilized in the proof of the quantum capacity theorem. The first lemma concerns the approximation of one isometry by another isometry that meets certain spectral requirements, and the second lemma is a general fact regarding Haar measure.

Lemma 8.55. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{W}$ be complex Euclidean spaces, let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ be an isometry, let $\Lambda \in \mathrm{Proj}(\mathcal{Y})$ and $\Pi \in \mathrm{Proj}(\mathcal{W})$ be projection operators, and let $\varepsilon \in (0, 1/4)$ be a positive real number. Also let $n = \dim(\mathcal{X})$, and assume that the constraints*

$$\big\langle \Lambda \otimes \Pi, AA^* \big\rangle \geq (1 - \varepsilon)n \tag{8.336}$$

and

$$2\,\mathrm{rank}(\Pi) \leq \dim(\mathcal{W}) \tag{8.337}$$

are satisfied. There exists an isometry $B \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ such that

1. $\|A - B\|_2 < 3\varepsilon^{1/4}\sqrt{n}$,
2. $\mathrm{Tr}_{\mathcal{W}}(BB^*) \leq 4\Lambda\,\mathrm{Tr}_{\mathcal{W}}(AA^*)\Lambda$, *and*
3. $\mathrm{rank}(\mathrm{Tr}_{\mathcal{Y}}(BB^*)) \leq 2\,\mathrm{rank}(\Pi)$.

Proof. By means of the singular value theorem, one may write

$$(\Lambda \otimes \Pi)A = \sum_{k=1}^{n} s_k u_k x_k^* \tag{8.338}$$

for an orthonormal basis $\{x_1, \ldots, x_n\}$ of $\mathcal{X}$, an orthonormal set $\{u_1, \ldots, u_n\}$ of vectors in $\mathcal{Y} \otimes \mathcal{W}$, and a collection $\{s_1, \ldots, s_n\} \subset [0, 1]$ of nonnegative real numbers. It holds that

$$\sum_{k=1}^{n} s_k^2 = \langle \Lambda \otimes \Pi, AA^* \rangle \geq (1 - \varepsilon)n. \tag{8.339}$$

Define $\Gamma \subseteq \{1, \ldots, n\}$ as

$$\Gamma = \Big\{ k \in \{1, \ldots, n\} \,:\, s_k^2 \geq 1 - \sqrt{\varepsilon} \Big\}, \tag{8.340}$$

and observe the inequalities

$$\begin{aligned} (1 - \varepsilon)n &\leq \sum_{k=1}^{n} s_k^2 \\ &\leq |\Gamma| + (n - |\Gamma|)\big(1 - \sqrt{\varepsilon}\big) = \big(1 - \sqrt{\varepsilon}\big)n + \sqrt{\varepsilon}|\Gamma|, \end{aligned} \tag{8.341}$$

from which it follows that

$$|\Gamma| \geq \big(1 - \sqrt{\varepsilon}\big)n > \frac{n}{2}. \tag{8.342}$$

Let $f : \{1, \ldots, n\} \backslash \Gamma \to \Gamma$ be any one-to-one function, and let $W \in \mathrm{U}(\mathcal{W})$ be any unitary operator satisfying $\Pi W \Pi = 0$; the existence of such an operator W follows from the assumption $2\,\mathrm{rank}(\Pi) \leq \dim(\mathcal{W})$. Finally, define an isometry $B \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ as follows:

$$B = \sum_{k \in \Gamma} u_k x_k^* + \sum_{k \in \{1, \ldots, n\} \backslash \Gamma} (\mathbb{1}_{\mathcal{Y}} \otimes W) u_{f(k)} x_k^*. \tag{8.343}$$

It remains to prove that B has the properties required by the statement of the lemma.

First, it will be verified that B is indeed an isometry. The set $\{u_k \,:\, k \in \Gamma\}$ is evidently orthonormal, as is the set

$$\{(\mathbb{1}_{\mathcal{Y}} \otimes W) u_{f(k)} \,:\, k \in \{1, \ldots, n\} \backslash \Gamma\}. \tag{8.344}$$

Moreover, for every $k \in \Gamma$, it must hold that $s_k > 0$, and therefore

$$u_k \in \operatorname{im}\big((\Lambda \otimes \Pi)A\big) \subseteq \operatorname{im}\big(\Lambda \otimes \Pi\big), \tag{8.345}$$

which implies that $u_k = (\mathbb{1}_{\mathcal{Y}} \otimes \Pi)u_k$. For an arbitrary choice of $j, k \in \Gamma$, one therefore has

$$\begin{aligned} \big\langle u_j, (\mathbb{1}_{\mathcal{Y}} \otimes W)u_k \big\rangle &= \big\langle (\mathbb{1}_{\mathcal{Y}} \otimes \Pi)u_j, (\mathbb{1}_{\mathcal{Y}} \otimes W\Pi)u_k \big\rangle \\ &= \big\langle u_j, (\mathbb{1}_{\mathcal{Y}} \otimes \Pi W \Pi)u_k \big\rangle = 0, \end{aligned} \tag{8.346}$$

which implies that the set

$$\{u_k \,:\, k \in \Gamma\} \cup \big\{(\mathbb{1}_{\mathcal{Y}} \otimes W)u_{f(k)} \,:\, k \in \{1, \ldots, n\} \backslash \Gamma\big\} \tag{8.347}$$

is an orthonormal set. This implies that B is an isometry.

Next, observe that

$$\| A - B \|_2 \leq \| A - (\Lambda \otimes \Pi)A \|_2 + \| (\Lambda \otimes \Pi)A - B \|_2. \tag{8.348}$$

The first term in this expression is bounded as

$$\| A - (\Lambda \otimes \Pi)A \|_2 = \sqrt{\big\langle \mathbb{1} - \Lambda \otimes \Pi, AA^* \big\rangle} \leq \sqrt{\varepsilon n}. \tag{8.349}$$

For the second term, it holds that

$$\begin{aligned} \big\| (\Lambda \otimes \Pi)A - B \big\|_2^2 &= \sum_{k \in \Gamma} (s_k - 1)^2 + \sum_{k \in \{1, \ldots, n\} \backslash \Gamma} \big(s_k^2 + 1\big) \\ &= n + \sum_{k=1}^{n} s_k^2 - 2 \sum_{k \in \Gamma} s_k. \end{aligned} \tag{8.350}$$

Consequently, one finds that

$$\big\| (\Lambda \otimes \Pi)A - B \big\|_2^2 \leq 2n - 2\big(1 - \sqrt{\varepsilon}\big)^{\frac{3}{2}} n < 4n\sqrt{\varepsilon}. \tag{8.351}$$

It follows that

$$\| A - B \|_2 < 3\varepsilon^{1/4}\sqrt{n}, \tag{8.352}$$

so that the first requirement on B is fulfilled.

The second requirement on B may be verified as follows:

$$\begin{aligned} \operatorname{Tr}_{\mathcal{W}}\big(BB^*\big) &\leq 2 \sum_{k \in \Gamma} \operatorname{Tr}_{\mathcal{W}}\big(u_k u_k^*\big) \\ &\leq \frac{2}{1 - \sqrt{\varepsilon}} \operatorname{Tr}_{\mathcal{W}}\big((\Lambda \otimes \Pi)AA^*(\Lambda \otimes \Pi)\big) \leq 4\Lambda \operatorname{Tr}_{\mathcal{W}}(AA^*)\Lambda. \end{aligned} \tag{8.353}$$

Finally, to verify that the third requirement on B is satisfied, one may again use the observation that $(\mathbb{1} \otimes \Pi)u_k = u_k$, which implies that

$$\operatorname{im}(\operatorname{Tr}_{\mathcal{Y}}(u_k u_k^*)) \subseteq \operatorname{im}(\Pi), \tag{8.354}$$

for each $k \in \Gamma$. As

$$\operatorname{Tr}_{\mathcal{Y}}(BB^*) = \sum_{k \in \Gamma} \operatorname{Tr}_{\mathcal{Y}}(u_k u_k^*) + \sum_{k \in \{1,\ldots,n\} \setminus \Gamma} W(\operatorname{Tr}_{\mathcal{Y}}(u_{f(k)} u_{f(k)}^*))W^*, \tag{8.355}$$

it follows that

$$\operatorname{im}\big(\operatorname{Tr}_{\mathcal{Y}}(BB^*)\big) \subseteq \operatorname{im}(\Pi) + \operatorname{im}(W\Pi) \tag{8.356}$$

and therefore

$$\operatorname{rank}\big(\operatorname{Tr}_{\mathcal{Y}}(BB^*)\big) \leq 2 \operatorname{rank}(\Pi), \tag{8.357}$$

as required. □

Lemma 8.56. *Let $\mathcal{X}$, $\mathcal{W}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $m = \dim(\mathcal{Z})$ and $n = \dim(\mathcal{X})$, and assume $m \leq n$. For every choice of an isometry $V \in \mathrm{U}(\mathcal{Z}, \mathcal{X})$ and every operator $Z \in \mathrm{L}(\mathcal{W} \otimes \mathcal{X})$, it holds that*

$$\int \big\| \big(\mathbb{1}_{\mathcal{W}} \otimes V^* U^*\big) Z \big(\mathbb{1}_{\mathcal{W}} \otimes UV\big) \big\|_1 \, \mathrm{d}\eta(U) \leq \frac{m}{n} \|Z\|_1 \tag{8.358}$$

for η denoting the Haar measure on $\mathrm{U}(\mathcal{X})$.

Proof. Let

$$\big\{W_1, \ldots, W_{n^2}\big\} \subset \mathrm{U}(\mathcal{X}) \tag{8.359}$$

be any collection of unitary operators for which it holds that the completely depolarizing channel $\Omega \in \mathrm{C}(\mathcal{X})$ is given by

$$\Omega(X) = \frac{1}{n^2} \sum_{k=1}^{n^2} W_k X W_k^* \tag{8.360}$$

for all $X \in \mathrm{L}(\mathcal{X})$. (Such a collection may, for instance, be derived from the discrete Weyl operators defined in Section 4.1.2.) Define $\mathcal{Y} = \mathbb{C}^{n^2}$, and define a channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Z} \otimes \mathcal{Y})$ as

$$\Phi(X) = \frac{1}{nm} \sum_{k=1}^{n^2} V^* W_k^* X W_k V \otimes E_{k,k} \tag{8.361}$$

for every $X \in \mathrm{L}(\mathcal{X})$. The fact that Φ is a channel follows from Corollary 2.27 together with the calculation

$$\frac{1}{nm}\sum_{k=1}^{n^2} W_k V V^* W_k^* = \frac{n}{m}\Omega(VV^*) = \mathbb{1}_{\mathcal{X}}. \tag{8.362}$$

Next, by the right unitary invariance of the Haar measure, it holds that

$$\begin{aligned}&\int \big\| \big(\mathbb{1}_{\mathcal{W}} \otimes V^* U^*\big) Z \big(\mathbb{1}_{\mathcal{W}} \otimes UV\big) \big\|_1 \,\mathrm{d}\eta(U) \\ &\quad = \int \big\| \big(\mathbb{1}_{\mathcal{W}} \otimes V^* W_k^* U^*\big) Z \big(\mathbb{1}_{\mathcal{W}} \otimes U W_k V\big) \big\|_1 \,\mathrm{d}\eta(U)\end{aligned} \tag{8.363}$$

for every choice of $k \in \{1, \ldots, n^2\}$, and therefore

$$\begin{aligned}&\int \big\| \big(\mathbb{1}_{\mathcal{W}} \otimes UV\big)^* Z \big(\mathbb{1}_{\mathcal{W}} \otimes UV\big) \big\|_1 \,\mathrm{d}\eta(U) \\ &\quad = \frac{1}{n^2}\sum_{k=1}^{n^2} \int \Big\| \big(\mathbb{1}_{\mathcal{W}} \otimes U W_k V\big)^* Z \big(\mathbb{1}_{\mathcal{W}} \otimes U W_k V\big) \Big\|_1 \,\mathrm{d}\eta(U) \\ &\quad = \frac{1}{n^2} \int \Bigg\| \sum_{k=1}^{n^2} \big(\mathbb{1}_{\mathcal{W}} \otimes U W_k V\big)^* Z \big(\mathbb{1}_{\mathcal{W}} \otimes U W_k V\big) \otimes E_{k,k} \Bigg\|_1 \,\mathrm{d}\eta(U) \\ &\quad = \frac{m}{n} \int \big\| \big(\mathbb{1}_{\mathrm{L}(\mathcal{W})} \otimes \Phi\big)\big((\mathbb{1}_{\mathcal{W}} \otimes U) Z (\mathbb{1}_{\mathcal{W}} \otimes U^*)\big) \big\|_1 \,\mathrm{d}\eta(U).\end{aligned} \tag{8.364}$$

As the trace norm is non-increasing under the action of channels, as well as unitarily invariant, it follows that

$$\begin{aligned}&\big\| \big(\mathbb{1}_{\mathcal{W}} \otimes UV\big)^* Z \big(\mathbb{1}_{\mathcal{W}} \otimes UV\big) \big\|_1 \,\mathrm{d}\eta(U) \\ &\quad \leq \frac{m}{n} \int \big\| (\mathbb{1}_{\mathcal{W}} \otimes U) Z (\mathbb{1}_{\mathcal{W}} \otimes U^*) \big\|_1 \,\mathrm{d}\eta(U) = \frac{m}{n} \| Z \|_1,\end{aligned} \tag{8.365}$$

which completes the proof. □

The quantum capacity theorem

As the following theorem establishes, the entanglement-generation capacity of a given channel is always at least as large as the coherent information of the completely mixed state through that channel. This fact, which will be generalized to arbitrary states in place of the completely mixed state in a corollary to the theorem, lies at the heart of the proof of the quantum capacity theorem.

Theorem 8.57. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. The entanglement generation capacity of Φ is lower-bounded by the coherent information of the completely mixed state $\omega \in \mathrm{D}(\mathcal{X})$ through Φ:*

$$\mathrm{I_C}(\omega\,;\,\Phi) \leq \mathrm{Q_{EG}}(\Phi). \tag{8.366}$$

Proof. Let $\mathcal{W}$ be a complex Euclidean space such that

$$\dim(\mathcal{W}) = 2\dim(\mathcal{X} \otimes \mathcal{Y}), \tag{8.367}$$

and let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ be an isometry for which

$$\Phi(X) = \mathrm{Tr}_{\mathcal{W}}\big(AXA^*\big) \tag{8.368}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Define a channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{W})$ as

$$\Psi(X) = \mathrm{Tr}_{\mathcal{Y}}\big(AXA^*\big) \tag{8.369}$$

for all $X \in \mathrm{L}(\mathcal{X})$, so that Ψ is complementary to Φ. It therefore holds that

$$\mathrm{I_C}(\omega\,;\,\Phi) = \mathrm{H}(\Phi(\omega)) - \mathrm{H}(\Psi(\omega)). \tag{8.370}$$

The theorem is vacuous in the case that $\mathrm{I_C}(\omega\,;\,\Phi) \leq 0$, so hereafter it will be assumed that $\mathrm{I_C}(\omega\,;\,\Phi)$ is positive. To prove the theorem, it suffices to demonstrate that every positive real number smaller than $\mathrm{I_C}(\omega\,;\,\Phi)$ is an achievable rate for entanglement generation through Φ. Toward this goal, assume that an arbitrary positive real number α satisfying $\alpha < \mathrm{I_C}(\omega\,;\,\Phi)$ has been fixed, and that $\varepsilon > 0$ is a positive real number chosen to be sufficiently small so that the inequality

$$\alpha < \mathrm{I_C}(\omega\,;\,\Phi) - 2\varepsilon(\mathrm{H}(\Phi(\omega)) + \mathrm{H}(\Psi(\omega))) \tag{8.371}$$

is satisfied. The remainder of the proof is devoted to proving that α is an achievable rate for entanglement generation through Φ.

Consider an arbitrary positive integer $n \geq 1/\alpha$, and let m be any positive integer such that $m \leq \alpha n$. Also let $\Gamma = \{0, 1\}$ denote the binary alphabet, and let $\mathcal{Z} = \mathbb{C}^{\Gamma}$. The task in which a state having high fidelity with the maximally entangled state

$$2^{-m}\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\big)\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\big)^* \tag{8.372}$$

is established between a sender and receiver through the channel $\Phi^{\otimes n}$ is to be considered. For any choice of an isometry $W \in \mathrm{U}(\mathcal{Z}^{\otimes m}, \mathcal{X}^{\otimes n})$ and a channel $\Xi \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$, the state

$$2^{-m}\big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)\big(\mathrm{vec}(W)\,\mathrm{vec}(W)^*\big) \tag{8.373}$$

may be established through the channel $\Phi^{\otimes n}$, so one may aim to prove that there exists a choice of Ξ and W for which the fidelity between the states (8.372) and (8.373) is high.

It is helpful at this point to let $A_n \in \mathrm{U}(\mathcal{X}^{\otimes n}, \mathcal{Y}^{\otimes n} \otimes \mathcal{W}^{\otimes n})$ be the isometry defined by the equation

$$\begin{aligned}(y_1 \otimes \cdots \otimes y_n \otimes w_1 \otimes \cdots \otimes w_n)^* A_n (x_1 \otimes \cdots \otimes x_n) \\ = \prod_{k=1}^{n} (y_k \otimes w_k)^* A x_k\end{aligned} \tag{8.374}$$

holding for every choice of vectors $x_1, \ldots, x_n \in \mathcal{X}$, $y_1, \ldots, y_n \in \mathcal{Y}$, and $w_1, \ldots, w_n \in \mathcal{W}$. In effect, A_n is equivalent to $A^{\otimes n}$, except that the tensor factors in its output space have been permuted, so that the output space becomes $\mathcal{Y}^{\otimes n} \otimes \mathcal{W}^{\otimes n}$ rather than $(\mathcal{Y} \otimes \mathcal{W})^{\otimes n}$. It may be noted that

$$\Phi^{\otimes n}(X) = \mathrm{Tr}_{\mathcal{W}^{\otimes n}}\big(A_n X A_n^*\big) \quad \text{and} \quad \Psi^{\otimes n}(X) = \mathrm{Tr}_{\mathcal{Y}^{\otimes n}}\big(A_n X A_n^*\big) \tag{8.375}$$

for every $X \in \mathrm{L}(\mathcal{X}^{\otimes n})$. Under the assumption that the decoding channel $\Xi \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$ has been selected optimally, Lemma 8.54 implies that the fidelity between the states (8.372) and (8.373) is lower-bounded by

$$\mathrm{F}\big(\rho\,,\, \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\rho) \otimes \omega_{\mathcal{Z}}^{\otimes m}\big) \tag{8.376}$$

for $\rho \in \mathrm{D}(\mathcal{W}^{\otimes n} \otimes \mathcal{Z}^{\otimes m})$ being defined as

$$\rho = 2^{-m}\,\mathrm{Tr}_{\mathcal{Y}^{\otimes n}}\big(\mathrm{vec}(A_n W)\,\mathrm{vec}(A_n W)^*\big) \tag{8.377}$$

and for $\omega_{\mathcal{Z}} \in \mathrm{D}(\mathcal{Z})$ denoting the completely mixed state on $\mathcal{Z}$.

The probabilistic method will be employed to prove the existence of an isometry W for which the expression (8.376) is close to 1, provided that n is sufficiently large. In particular, one may fix $V \in \mathrm{U}(\mathcal{Z}^{\otimes m}, \mathcal{X}^{\otimes n})$ to be an arbitrary isometry, and let $W = UV$ for U chosen at random with respect to the Haar measure on $\mathrm{U}(\mathcal{X}^{\otimes n})$. The analysis that follows demonstrates that, for an operator W chosen in this way, one expects the quantity (8.376) to be

close to 1, for sufficiently large n, which proves the existence of a choice of W for which this is true.

Let $k = \dim(\mathcal{X})$ and define $\xi \in \mathrm{D}(\mathcal{W}^{\otimes n} \otimes \mathcal{X}^{\otimes m})$ as

$$\xi = \frac{1}{k^n} \operatorname{Tr}_{\mathcal{Y}^{\otimes n}}\bigl(\operatorname{vec}(A_n)\operatorname{vec}(A_n)^*\bigr). \tag{8.378}$$

Also define $\rho_U \in \mathrm{D}(\mathcal{W}^{\otimes n} \otimes \mathcal{Z}^{\otimes m})$ as

$$\rho_U = \frac{1}{2^m} \operatorname{Tr}_{\mathcal{Y}^{\otimes n}}\bigl(\operatorname{vec}(A_n UV)\operatorname{vec}(A_n UV)^*\bigr), \tag{8.379}$$

for each unitary operator $U \in \mathrm{U}(\mathcal{X}^{\otimes n})$, and observe that

$$\rho_U = \frac{k^n}{2^m}\bigl(\mathbb{1}_{\mathcal{W}}^{\otimes n} \otimes V^{\mathsf{T}}U^{\mathsf{T}}\bigr)\,\xi\,\bigl(\mathbb{1}_{\mathcal{W}}^{\otimes n} \otimes V^{\mathsf{T}}U^{\mathsf{T}}\bigr)^*. \tag{8.380}$$

For the isometry $W = UV$, the fidelity between the states (8.372) and (8.373) is lower-bounded by

$$\mathrm{F}\bigl(\rho_U\,,\, \operatorname{Tr}_{\mathcal{Z}^{\otimes m}}(\rho_U) \otimes \omega_{\mathcal{Z}}^{\otimes m}\bigr), \tag{8.381}$$

for a suitable choice of the decoding channel Ξ.

Let $\Lambda_{n,\varepsilon} \in \operatorname{Proj}(\mathcal{Y}^{\otimes n})$ and $\Pi_{n,\varepsilon} \in \operatorname{Proj}(\mathcal{W}^{\otimes n})$ be the projection operators onto the ε-strongly typical subspaces of $\mathcal{Y}^{\otimes n}$ and $\mathcal{W}^{\otimes n}$, with respect to any fixed choice of spectral decompositions of the operators $\Phi(\omega)$ and $\Psi(\omega)$, respectively. By Lemma 8.35, one has the inequalities

$$\begin{aligned}
&\frac{1}{k^n}\bigl\langle \Lambda_{n,\varepsilon} \otimes \mathbb{1}_{\mathcal{W}}^{\otimes n} \otimes \mathbb{1}_{\mathcal{X}}^{\otimes n}, \operatorname{vec}(A_n)\operatorname{vec}(A_n)^*\bigr\rangle \\
&\qquad\qquad = \bigl\langle \Lambda_{n,\varepsilon}, (\Phi(\omega))^{\otimes n}\bigr\rangle > 1 - \zeta_{n,\varepsilon}, \\
&\frac{1}{k^n}\bigl\langle \mathbb{1}_{\mathcal{Y}}^{\otimes n} \otimes \Pi_{n,\varepsilon} \otimes \mathbb{1}_{\mathcal{X}}^{\otimes n}, \operatorname{vec}(A_n)\operatorname{vec}(A_n)^*\bigr\rangle \\
&\qquad\qquad = \bigl\langle \Pi_{n,\varepsilon}, (\Psi(\omega))^{\otimes n}\bigr\rangle > 1 - \zeta_{n,\varepsilon}
\end{aligned} \tag{8.382}$$

for

$$\zeta_{n,\varepsilon} = K\exp(-\delta n\varepsilon^2), \tag{8.383}$$

where $K \geq 1$ and $\delta > 0$ are positive real numbers that are independent of n and ε. It follows that

$$\frac{1}{k^n}\bigl\langle \Lambda_{n,\varepsilon} \otimes \Pi_{n,\varepsilon} \otimes \mathbb{1}_{\mathcal{X}}^{\otimes n}, \operatorname{vec}(A_n)\operatorname{vec}(A_n)^*\bigr\rangle > 1 - 2\zeta_{n,\varepsilon}, \tag{8.384}$$

which is equivalent to

$$\left\langle \Lambda_{n,\varepsilon} \otimes \Pi_{n,\varepsilon}, A_n A_n^* \right\rangle \geq (1 - 2\zeta_{n,\varepsilon})\, k^n. \tag{8.385}$$

If n is sufficiently large so that $\zeta_{n,\varepsilon} < 1/4$, it follows by Lemma 8.55 that there exists an isometry $B_n \in \mathrm{U}(\mathcal{X}^{\otimes n}, \mathcal{Y}^{\otimes n} \otimes \mathcal{W}^{\otimes n})$ satisfying the conditions

$$\begin{aligned} \left\| A_n - B_n \right\|_2 &\leq 3\, \zeta_{n,\varepsilon}^{1/4}\, k^{n/2}, \\ \mathrm{Tr}_{\mathcal{W}^{\otimes n}}\left(B_n B_n^*\right) &\leq 4 \Lambda_{n,\varepsilon}\, \mathrm{Tr}_{\mathcal{W}^{\otimes n}}\left(A_n A_n^*\right) \Lambda_{n,\varepsilon}, \end{aligned} \tag{8.386}$$

and

$$\mathrm{rank}\left(\mathrm{Tr}_{\mathcal{Y}^{\otimes n}}\left(B_n B_n^*\right)\right) \leq 2\, \mathrm{rank}(\Pi_{n,\varepsilon}). \tag{8.387}$$

By Proposition 8.36, the third condition implies that

$$\mathrm{rank}\left(\mathrm{Tr}_{\mathcal{Y}^{\otimes n}}\left(B_n B_n^*\right)\right) \leq 2^{n(1+\varepsilon)\, \mathrm{H}(\Psi(\omega))+1}. \tag{8.388}$$

Using the second condition, together with Corollary 8.34, one obtains

$$\begin{aligned} &\mathrm{Tr}\left(\left(\frac{1}{k^n}\, \mathrm{Tr}_{\mathcal{W}^{\otimes n}}\left(B_n B_n^*\right)\right)^2\right) \\ &\qquad \leq \mathrm{Tr}\left(\left(\frac{4}{k^n} \Lambda_{n,\varepsilon}\, \mathrm{Tr}_{\mathcal{W}^{\otimes n}}\left(A_n A_n^*\right) \Lambda_{n,\varepsilon}\right)^2\right) \\ &\qquad = 16\, \mathrm{Tr}\left(\left(\Lambda_{n,\varepsilon} \Phi(\omega)^{\otimes n} \Lambda_{n,\varepsilon}\right)^2\right) \\ &\qquad \leq 2^{-n(1-\varepsilon)\, \mathrm{H}(\Phi(\omega))+4}. \end{aligned} \tag{8.389}$$

Finally, define

$$\sigma = \frac{1}{k^n}\, \mathrm{Tr}_{\mathcal{Y}^{\otimes n}}\left(\mathrm{vec}(B_n)\, \mathrm{vec}(B_n)^*\right), \tag{8.390}$$

and also define

$$\begin{aligned} \tau_U &= \frac{1}{2^m}\, \mathrm{Tr}_{\mathcal{Y}^{\otimes n}}\left(\mathrm{vec}(B_n U V)\, \mathrm{vec}(B_n U V)^*\right) \\ &= \frac{k^n}{2^m} \left(\mathbb{1}_{\mathcal{W}}^{\otimes n} \otimes V^{\mathsf{T}} U^{\mathsf{T}}\right) \sigma \left(\mathbb{1}_{\mathcal{W}}^{\otimes n} \otimes V^{\mathsf{T}} U^{\mathsf{T}}\right)^* \end{aligned} \tag{8.391}$$

for each $U \in \mathrm{U}(\mathcal{X}^{\otimes n})$. It holds that

$$\begin{aligned}
&\Big\| \rho_U - \mathrm{Tr}_{\mathcal{Z}}(\rho_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \Big\|_1 \\
&\qquad \leq \big\| \rho_U - \tau_U \big\|_1 + \Big\| \tau_U - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \Big\|_1 \\
&\qquad\quad + \Big\| \big(\mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\rho_U) \big) \otimes \omega_{\mathcal{Z}}^{\otimes m} \Big\|_1 \\
&\qquad \leq \Big\| \tau_U - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \Big\|_1 + 2 \big\| \rho_U - \tau_U \big\|_1,
\end{aligned} \tag{8.392}$$

and it remains to consider the average value of the two terms in the final expression of this inequality. When considering the average value of first term in the final expression of (8.392), it may be noted that

$$\begin{aligned}
\mathrm{rank}\Big(\tau_U - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \Big) &\leq \mathrm{rank}\Big(\mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \Big) \\
&\leq 2^m \ \mathrm{rank}\big(\mathrm{Tr}_{\mathcal{Y}^{\otimes n}}\big(B_n B_n^* \big) \big) \leq 2^{n(1+\varepsilon)\,\mathrm{H}(\Psi(\omega)) + m + 1}
\end{aligned} \tag{8.393}$$

and

$$\mathrm{Tr}(\sigma^2) = \mathrm{Tr}\left(\left(\frac{1}{k^n} \mathrm{Tr}_{\mathcal{W}^{\otimes n}}\big(B_n B_n^* \big) \right)^2 \right) \leq 2^{-n(1-\varepsilon)\,\mathrm{H}(\Phi(\omega)) + 4}. \tag{8.394}$$

Making use of Lemma 8.53, it therefore follows that

$$\begin{aligned}
&\int \big\| \tau_U - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \big\|_1^2 \, \mathrm{d}\eta(U) \\
&\qquad \leq 2^{n(1+\varepsilon)\,\mathrm{H}(\Psi(\omega)) + m + 1} \int \big\| \tau_U - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \big\|_2^2 \, \mathrm{d}\eta(U) \\
&\qquad \leq 2^{n((1+\varepsilon)\,\mathrm{H}(\Psi(\omega)) - (1-\varepsilon)\,\mathrm{H}(\Phi(\omega))) + m + 5} \\
&\qquad = 2^{-n(\mathrm{I_c}(\omega;\Phi) - 2\varepsilon(\mathrm{H}(\Phi(\omega)) + \mathrm{H}(\Psi(\omega)))) + m + 5}.
\end{aligned} \tag{8.395}$$

By the assumption (8.371), along with the assumption that $m \leq \alpha n$, one has that this quantity approaches 0 in the limit as n approaches infinity. It therefore holds (by Jensen's inequality) that the quantity

$$\int \big\| \tau_U - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \big\|_1 \, \mathrm{d}\eta(U) \tag{8.396}$$

also approaches 0 in the limit as n approaches infinity. The average value of the second term in the final expression of (8.392) may be upper-bounded as

$$\begin{aligned}
&\int \|\rho_U - \tau_U\|_1 \,\mathrm{d}\eta(U) \\
&= \frac{k^n}{2^m} \int \Big\| \big(\mathbb{1}_{\mathcal{Y}^{\otimes n}} \otimes V^{\mathsf{T}} U^{\mathsf{T}}\big)(\xi - \sigma)\big(\mathbb{1}_{\mathcal{Y}^{\otimes n}} \otimes V^{\mathsf{T}} U^{\mathsf{T}}\big)^* \Big\|_1 \,\mathrm{d}\eta(U) \\
&\leq \|\xi - \sigma\|_1 \leq \frac{1}{k^n} \big\| \operatorname{vec}(A_n)\operatorname{vec}(A_n)^* - \operatorname{vec}(B_n)\operatorname{vec}(B_n)^* \big\|_1 \\
&\qquad\qquad \leq \frac{2}{k^{n/2}} \big\| A_n - B_n \big\|_2 \leq 6\,\zeta_{n,\varepsilon}^{1/4}
\end{aligned} \tag{8.397}$$

by Lemma 8.56. Once again, this quantity approaches 0 in the limit as n approaches infinity. It follows that the entanglement generation capacity of Φ is at least α, which completes the proof. □

Corollary 8.58. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, and let $\sigma \in \mathrm{D}(\mathcal{X})$ be a density operator. The quantum capacity of Φ is lower-bounded by the coherent information of σ through Φ:*

$$\mathrm{I_C}(\sigma; \Phi) \leq \mathrm{Q}(\Phi). \tag{8.398}$$

Proof. Observe first that it is a consequence of Theorem 8.57 that

$$\mathrm{I_C}(\omega_{\mathcal{V}}; \Phi) \leq \mathrm{Q}(\Phi) \tag{8.399}$$

for every nontrivial subspace $\mathcal{V} \subseteq \mathcal{X}$, where

$$\omega_{\mathcal{V}} = \frac{\Pi_{\mathcal{V}}}{\dim(\mathcal{V})} \tag{8.400}$$

is the state that is maximally mixed over the subspace $\mathcal{V}$. To verify that this is so, let $\mathcal{Z}$ be any complex Euclidean space with $\dim(\mathcal{Z}) = \dim(\mathcal{V})$, let $V \in \mathrm{U}(\mathcal{Z}, \mathcal{X})$ be an isometry such that $VV^* = \Pi_{\mathcal{V}}$, and define a channel $\Xi \in \mathrm{C}(\mathcal{Z}, \mathcal{Y})$ as

$$\Xi(Z) = \Phi(VZV^*) \tag{8.401}$$

for all $Z \in \mathrm{L}(\mathcal{Z})$. It is evident that $\mathrm{Q}(\Xi) \leq \mathrm{Q}(\Phi)$; the channel Φ emulates Ξ, so for every positive integer n it holds that $\Phi^{\otimes n}$ emulates every channel that can be emulated by $\Xi^{\otimes n}$. It follows that

$$\begin{aligned}
\mathrm{Q}(\Phi) \geq \mathrm{Q}(\Xi) = \mathrm{Q_{EG}}(\Xi) &\geq \mathrm{I_C}(\omega_{\mathcal{Z}}; \Xi) \\
&= \mathrm{I_C}(V\omega_{\mathcal{Z}}V^*; \Phi) = \mathrm{I_C}(\omega_{\mathcal{V}}; \Phi),
\end{aligned} \tag{8.402}$$

as claimed.

Now, let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ be an isometry such that

$$\Phi(X) = \mathrm{Tr}_{\mathcal{W}}(AXA^*) \tag{8.403}$$

for all $X \in \mathrm{L}(\mathcal{X})$, for a suitable choice of a complex Euclidean space $\mathcal{W}$, and define a channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{W})$ as

$$\Psi(X) = \mathrm{Tr}_{\mathcal{Y}}(AXA^*) \tag{8.404}$$

for all $X \in \mathrm{L}(\mathcal{X})$. It therefore holds that Ψ is complementary to Φ, so that

$$\mathrm{I_c}(\sigma; \Phi) = \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)). \tag{8.405}$$

Let

$$\sigma = \sum_{a \in \Sigma} p(a) x_a x_a^* \tag{8.406}$$

be a spectral decomposition of σ, and let

$$\omega_{n,\varepsilon} = \frac{\Lambda_{n,\varepsilon}}{\mathrm{Tr}(\Lambda_{n,\varepsilon})} \in \mathrm{D}(\mathcal{X}^{\otimes n}) \tag{8.407}$$

for each positive integer n and each positive real number $\varepsilon > 0$, for $\Lambda_{n,\varepsilon}$ denoting the projection onto the ε-strongly typical subspace of $\mathcal{X}^{\otimes n}$, with respect to the spectral decomposition (8.406).

Next, let $\varepsilon > 0$ be a positive real number, to be chosen arbitrarily. By Lemma 8.39, it holds that

$$\left| \lim_{n \to \infty} \frac{\mathrm{H}\big(\Phi^{\otimes n}(\omega_{n,\varepsilon})\big)}{n} - \mathrm{H}(\Phi(\sigma)) \right| \leq (2\,\mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)))\varepsilon, \tag{8.408}$$

and therefore there must exist a positive integer n_0 such that, for all $n \geq n_0$, one has

$$\left| \frac{1}{n} \mathrm{H}\big(\Phi^{\otimes n}(\omega_{n,\varepsilon})\big) - \mathrm{H}(\Phi(\sigma)) \right| \leq (2\,\mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) + 1)\varepsilon. \tag{8.409}$$

Along similar lines, there must exist a positive integer n_1 such that, for all $n \geq n_1$, one has

$$\left| \frac{1}{n} \mathrm{H}\big(\Psi^{\otimes n}(\omega_{n,\varepsilon})\big) - \mathrm{H}(\Psi(\sigma)) \right| \leq (2\,\mathrm{H}(\sigma) + \mathrm{H}(\Psi(\sigma)) + 1)\varepsilon. \tag{8.410}$$

It follows that there must exist a positive integer n such that

$$\begin{aligned}\left|\frac{1}{n}\mathrm{I_C}\big(\omega_{n,\varepsilon};\Phi^{\otimes n}\big)-\mathrm{I_C}(\sigma;\Phi)\right| \\ \leq (4\,\mathrm{H}(\sigma)+\mathrm{H}(\Phi(\sigma))+\mathrm{H}(\Psi(\sigma))+2)\varepsilon.\end{aligned} \tag{8.411}$$

By the argument presented at the beginning of the proof, it holds that

$$\frac{\mathrm{I_C}\big(\omega_{n,\varepsilon};\Phi^{\otimes n}\big)}{n} \leq \frac{\mathrm{Q}(\Phi^{\otimes n})}{n} = \mathrm{Q}(\Phi), \tag{8.412}$$

and therefore

$$\mathrm{Q}(\Phi) \geq \mathrm{I_C}\big(\sigma;\Phi\big) - (4\,\mathrm{H}(\sigma)+\mathrm{H}(\Phi(\sigma))+\mathrm{H}(\Psi(\sigma))+2)\varepsilon. \tag{8.413}$$

As ε has been chosen to be an arbitrary positive real number, it follows that

$$\mathrm{Q}(\Phi) \geq \mathrm{I_C}\big(\sigma;\Phi\big), \tag{8.414}$$

which completes the proof. □

Finally, the quantum capacity theorem may be stated and proved.

Theorem 8.59 (Quantum capacity theorem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X},\mathcal{Y})$ be a channel. It holds that*

$$\mathrm{Q}(\Phi) = \lim_{n\to\infty} \frac{\mathrm{I_C}(\Phi^{\otimes n})}{n} \tag{8.415}$$

Proof. For every positive integer n and every density operator $\sigma \in \mathrm{D}(\mathcal{X}^{\otimes n})$, one has

$$\mathrm{I_C}(\sigma\,;\,\Phi^{\otimes n}) \leq \mathrm{Q}(\Phi^{\otimes n}) = n\,\mathrm{Q}(\Phi) \tag{8.416}$$

by Corollary 8.58. It therefore holds that

$$\mathrm{Q}(\Phi) \geq \lim_{n\to\infty} \frac{\mathrm{I_C}(\Phi^{\otimes n})}{n}. \tag{8.417}$$

Now suppose that α is an achievable rate for entanglement generation through Φ. It will be proved that

$$\lim_{n\to\infty} \frac{\mathrm{I_C}(\Phi^{\otimes n})}{n} \geq \alpha. \tag{8.418}$$

If $\alpha = 0$, then this inequality holds trivially, so it will be assumed hereafter that $\alpha > 0$.

Let $\varepsilon \in (0, 1/2)$ be chosen arbitrarily, and let $\Gamma = \{0, 1\}$ and $\mathcal{Z} = \mathbb{C}^{\Gamma}$. As α is an achievable rate for entanglement generation through Φ, it holds that for all but finitely many positive integers n and all positive integers $m \leq \alpha n$ that there must exist a unit vector $u \in \mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m}$ and a channel $\Xi \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$ such that

$$\mathrm{F}\Big(2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^*, \big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(uu^*)\Big) > 1 - \varepsilon, \tag{8.419}$$

and therefore

$$\Big\| 2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^* - \big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(uu^*) \Big\|_1 < 2\sqrt{2\varepsilon} \tag{8.420}$$

by one of the Fuchs–van de Graaf inequalities (Theorem 3.36). For any unit vector $u \in \mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m}$ for which the inequality (8.420) holds, one concludes from the Fannes–Audenaert inequality (Theorem 5.28) that

$$\mathrm{H}\big(\big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(uu^*)\big) \leq 2\delta m + \mathrm{H}(\delta, 1-\delta) \tag{8.421}$$

and

$$m - \mathrm{H}\big(\Xi\Phi^{\otimes n}(\rho)\big) \leq \delta m + \mathrm{H}(\delta, 1-\delta), \tag{8.422}$$

for

$$\rho = \operatorname{Tr}_{\mathcal{Z}^{\otimes m}}(uu^*) \quad \text{and} \quad \delta = \sqrt{2\varepsilon}. \tag{8.423}$$

Together, these inequalities imply that

$$\mathrm{I}_{\mathrm{C}}\big(\rho; \Xi\Phi^{\otimes n}\big) \geq (1 - 3\delta)m - 2\,\mathrm{H}(\delta, 1-\delta), \tag{8.424}$$

and therefore

$$\mathrm{I}_{\mathrm{C}}\big(\rho; \Phi^{\otimes n}\big) \geq (1 - 3\delta)m - 2\,\mathrm{H}(\delta, 1-\delta) \geq (1 - 3\delta)m - 2 \tag{8.425}$$

by Proposition 8.15. Choosing $m = \lfloor \alpha n \rfloor \geq \alpha n - 1$, it follows that

$$\frac{\mathrm{I}_{\mathrm{C}}\big(\rho; \Phi^{\otimes n}\big)}{n} \geq (1 - 3\delta)\alpha - \frac{3}{n} \tag{8.426}$$

so that

$$\lim_{n \to \infty} \frac{\mathrm{I}_{\mathrm{C}}(\Phi^{\otimes n})}{n} \geq (1 - 3\delta)\alpha. \tag{8.427}$$

Because δ may be taken to be arbitrarily small by a suitable choice of ε, the inequality (8.418) follows, which completes the proof. $\square$

8.3 Non-additivity and super-activation

Expressions for the classical and quantum capacities of a quantum channel are given by regularizations of the Holevo capacity and maximum coherent information,

$$\mathrm{C}(\Phi) = \lim_{n\to\infty} \frac{\chi(\Phi^{\otimes n})}{n} \quad \text{and} \quad \mathrm{Q}(\Psi) = \lim_{n\to\infty} \frac{\mathrm{I_c}(\Psi^{\otimes n})}{n}, \tag{8.428}$$

as has been established by the Holevo–Schumacher–Westmoreland theorem and quantum capacity theorem (Theorems 8.30 and 8.59). Non-regularized analogues of these formulas do not hold, in general. In particular, the strict inequalities

$$\chi(\Phi \otimes \Phi) > 2\chi(\Phi) \quad \text{and} \quad \mathrm{I_c}(\Psi \otimes \Psi) > 2\,\mathrm{I_c}(\Psi) \tag{8.429}$$

hold for a suitable choice of channels Φ and Ψ, as is demonstrated in the subsections that follow. These examples reveal that the Holevo capacity does not coincide directly with the classical capacity, and likewise for the maximum coherent information and quantum capacity.

With respect to the Holevo capacity, the fact that a strict inequality may hold for some channels Φ in (8.429) will be demonstrated in Section 8.3.1, through the use of Theorem 7.53 from the previous chapter. The existence of such channels is far from obvious, and no explicit examples are known at the time of this book's writing—it is only the existence of such channels that is known. The now falsified conjecture that the equality

$$\chi(\Phi_0 \otimes \Phi_1) = \chi(\Phi_0) + \chi(\Phi_1) \tag{8.430}$$

should hold for all choices of channels Φ_0 and Φ_1 was known for some time as the *additivity conjecture*.

In contrast, it is not difficult to find an example of a channel Ψ for which a strict inequality in (8.429) holds. There are, in fact, very striking examples of channels that go beyond the demonstration of non-additivity of maximum coherent information. In particular, one may find channels Ψ_0 and Ψ_1 such that both Ψ_0 and Ψ_1 have zero quantum capacity, and therefore

$$\mathrm{I_c}(\Psi_0) = \mathrm{I_c}(\Psi_1) = 0, \tag{8.431}$$

but for which

$$\mathrm{I_c}(\Psi_0 \otimes \Psi_1) > 0, \tag{8.432}$$

and therefore $\Psi_0 \otimes \Psi_1$ has nonzero quantum capacity. This phenomenon is known as *super-activation*, and is discussed in Section 8.3.2. From such a choice of channels Ψ_0 and Ψ_1, the construction of a channel Ψ for which the strict inequality (8.429) holds is possible.

8.3.1 Non-additivity of the Holevo capacity

The fact that there exists a channel Φ for which

$$\chi(\Phi \otimes \Phi) > 2\chi(\Phi) \tag{8.433}$$

is demonstrated below. The proof makes use of Theorem 7.53, together with two basic ideas: one concerns the *direct sum* of two channels, and the other is a construction that relates the minimum output entropy of a given channel to the Holevo capacity of a channel constructed from the one given.

Direct sums of channels and their minimum output entropy

The direct sum of two maps is defined as follows. (One may also consider direct sums of more than two maps, but it is sufficient for the needs of the present section to consider the case of just two maps.)

Definition 8.60. Let $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$ be complex Euclidean spaces and let $\Phi_0 \in \mathrm{T}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{T}(\mathcal{X}_1, \mathcal{Y}_1)$ be maps. The direct sum of Φ_0 and Φ_1 is the map

$$\Phi_0 \oplus \Phi_1 \in \mathrm{T}(\mathcal{X}_0 \oplus \mathcal{X}_1, \mathcal{Y}_0 \oplus \mathcal{Y}_1) \tag{8.434}$$

defined as

$$(\Phi_0 \oplus \Phi_1) \begin{pmatrix} X_0 & \cdot \\ \cdot & X_1 \end{pmatrix} = \begin{pmatrix} \Phi_0(X_0) & 0 \\ 0 & \Phi_1(X_1) \end{pmatrix} \tag{8.435}$$

for every $X_0 \in \mathrm{L}(\mathcal{X}_0)$ and $X_1 \in \mathrm{L}(\mathcal{X}_1)$. The dots in (8.435) indicate arbitrary operators in $\mathrm{L}(\mathcal{X}_1, \mathcal{X}_0)$ and $\mathrm{L}(\mathcal{X}_0, \mathcal{X}_1)$ that have no influence on the output of the map $\Phi_0 \oplus \Phi_1$.

The direct sum of two channels is also a channel, as is established by the following straightforward proposition.

Proposition 8.61. *Let $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$ be complex Euclidean spaces and let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be channels. The direct sum of Φ_0 and Φ_1 is a channel: $\Phi_0 \oplus \Phi_1 \in \mathrm{C}(\mathcal{X}_0 \oplus \mathcal{X}_1, \mathcal{Y}_0 \oplus \mathcal{Y}_1)$.*

Proof. It is immediate from the definition of the direct sum of Φ_0 and Φ_1 that $\Phi_0 \oplus \Phi_1$ is trace-preserving, so it suffices to prove that $\Phi_0 \oplus \Phi_1$ is completely positive. As Φ_0 and Φ_1 are completely positive, Kraus representations

$$\Phi_0(X_0) = \sum_{a\in\Sigma} A_a X_0 A_a^* \quad \text{and} \quad \Phi_1(X_1) = \sum_{b\in\Gamma} B_b X_1 B_b^* \tag{8.436}$$

of these maps must exist. Through a direct computation, one may verify that

$$\begin{aligned}(\Phi_0 \oplus \Phi_1)(X) = &\sum_{a\in\Sigma} \begin{pmatrix} A_a & 0 \\ 0 & 0 \end{pmatrix} X \begin{pmatrix} A_a & 0 \\ 0 & 0 \end{pmatrix}^* \\ &+ \sum_{b\in\Gamma} \begin{pmatrix} 0 & 0 \\ 0 & B_b \end{pmatrix} X \begin{pmatrix} 0 & 0 \\ 0 & B_b \end{pmatrix}^*\end{aligned} \tag{8.437}$$

for all $X \in \mathrm{L}(\mathcal{X}_0 \oplus \mathcal{X}_1)$. It follows that $\Phi_0 \oplus \Phi_1$ is completely positive, as required. □

By Theorem 7.53, there exist channels Φ_0 and Φ_1 such that

$$\mathrm{H}_{\min}(\Phi_0 \otimes \Phi_1) < \mathrm{H}_{\min}(\Phi_0) + \mathrm{H}_{\min}(\Phi_1). \tag{8.438}$$

It is possible to obtain, from this fact, an example of a single channel Φ such that

$$\mathrm{H}_{\min}(\Phi \otimes \Phi) < 2\,\mathrm{H}_{\min}(\Phi). \tag{8.439}$$

The following corollary (to Theorem 7.53) establishes that this is so.

Corollary 8.62. *There exists a channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, for some choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, such that*

$$\mathrm{H}_{\min}(\Phi \otimes \Phi) < 2\,\mathrm{H}_{\min}(\Phi). \tag{8.440}$$

Proof. By Theorem 7.53, there exist complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$ and channels $\Phi_0, \Phi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ such that

$$\mathrm{H}_{\min}(\Phi_0 \otimes \Phi_1) < \mathrm{H}_{\min}(\Phi_0) + \mathrm{H}_{\min}(\Phi_1). \tag{8.441}$$

Assume that such a choice of channels has been fixed for the remainder of the proof. Let $\sigma_0, \sigma_1 \in \mathrm{D}(\mathcal{X})$ be density operators satisfying

$$\mathrm{H}(\Phi_0(\sigma_0)) = \mathrm{H}_{\min}(\Phi_0) \quad \text{and} \quad \mathrm{H}(\Phi_1(\sigma_1)) = \mathrm{H}_{\min}(\Phi_1). \tag{8.442}$$

Define channels $\Psi_0, \Psi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Y})$ as

$$\Psi_0(X) = \Phi_0(X) \otimes \Phi_1(\sigma_1) \quad \text{and} \quad \Psi_1(X) = \Phi_0(\sigma_0) \otimes \Phi_1(X) \tag{8.443}$$

for all $X \in \mathrm{L}(\mathcal{X})$, and define

$$\Phi = \Psi_0 \oplus \Psi_1 \in \mathrm{C}(\mathcal{X} \oplus \mathcal{X}, (\mathcal{Y} \otimes \mathcal{Y}) \oplus (\mathcal{Y} \otimes \mathcal{Y})). \tag{8.444}$$

It remains to verify that $\mathrm{H}_{\min}(\Phi \otimes \Phi) < 2\,\mathrm{H}_{\min}(\Phi)$.

For any state $\rho \in \mathrm{D}(\mathcal{X} \oplus \mathcal{X})$, one may write

$$\rho = \begin{pmatrix} \lambda\rho_0 & X \\ X^* & (1-\lambda)\rho_1 \end{pmatrix} \tag{8.445}$$

for some choice of $\lambda \in [0,1]$, $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$, and $X \in \mathrm{L}(\mathcal{X})$. Evaluating Φ on ρ yields

$$\Phi(\rho) = \begin{pmatrix} \lambda\Phi_0(\rho_0) \otimes \Phi_1(\sigma_1) & 0 \\ 0 & (1-\lambda)\Phi_0(\sigma_0) \otimes \Phi_1(\rho_1) \end{pmatrix}, \tag{8.446}$$

so that

$$\begin{aligned} \mathrm{H}(\Phi(\rho)) &= \lambda\big(\mathrm{H}(\Phi_0(\rho_0)) + \mathrm{H}(\Phi_1(\sigma_1))\big) \\ &\quad + (1-\lambda)\big(\mathrm{H}(\Phi_0(\sigma_0)) + \mathrm{H}(\Phi_1(\rho_1))\big) + \mathrm{H}(\lambda, 1-\lambda). \end{aligned} \tag{8.447}$$

One concludes that

$$\mathrm{H}_{\min}(\Phi) = \mathrm{H}_{\min}(\Phi_0) + \mathrm{H}_{\min}(\Phi_1). \tag{8.448}$$

Finally, define an isometry $V \in \mathrm{U}(\mathcal{X} \otimes \mathcal{X}, (\mathcal{X} \oplus \mathcal{X}) \otimes (\mathcal{X} \oplus \mathcal{X}))$ by the equation

$$V(x_0 \otimes x_1) = (x_0 \oplus 0) \otimes (0 \oplus x_1) \tag{8.449}$$

holding for all $x_0, x_1 \in \mathcal{X}$. A calculation reveals that

$$\begin{aligned} &\mathrm{H}\big((\Phi \otimes \Phi)(V\xi V^*)\big) \\ &\quad = \mathrm{H}\big((\Phi_0 \otimes \Phi_1)(\xi)\big) + \mathrm{H}(\Phi_0(\sigma_0)) + \mathrm{H}(\Phi_1(\sigma_1)) \end{aligned} \tag{8.450}$$

for every density operator $\xi \in \mathrm{D}(\mathcal{X} \otimes \mathcal{X})$. In particular, for any choice of $\xi \in \mathrm{D}(\mathcal{X} \otimes \mathcal{X})$ satisfying

$$\mathrm{H}((\Phi_0 \otimes \Phi_1)(\xi)) = \mathrm{H}_{\min}(\Phi_0 \otimes \Phi_1), \tag{8.451}$$

one has

$$\begin{aligned} &\mathrm{H}\big((\Phi \otimes \Phi)(V\xi V^*)\big) \\ &\quad = \mathrm{H}_{\min}(\Phi_0 \otimes \Phi_1) + \mathrm{H}_{\min}(\Phi_0) + \mathrm{H}_{\min}(\Phi_1), \end{aligned} \tag{8.452}$$

and therefore $\mathrm{H}_{\min}(\Phi \otimes \Phi) < 2\,\mathrm{H}_{\min}(\Phi)$ as claimed. $\square$

From low minimum output entropy to high Holevo capacity

The construction to be described below allows one to conclude that there exists a channel Ψ for which the Holevo capacity is super-additive, meaning that

$$\chi(\Psi \otimes \Psi) > 2\chi(\Psi), \tag{8.453}$$

by means of Corollary 8.62.

Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are complex Euclidean spaces and $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is an arbitrary channel. Suppose further that Σ is an alphabet and

$$\{U_a \,:\, a \in \Sigma\} \subset \mathrm{U}(\mathcal{Y}) \tag{8.454}$$

is a collection of unitary operators with the property that the completely depolarizing channel $\Omega \in \mathrm{C}(\mathcal{Y})$ is given by

$$\Omega(Y) = \frac{1}{|\Sigma|} \sum_{a \in \Sigma} U_a Y U_a^* \tag{8.455}$$

for all $Y \in \mathrm{L}(\mathcal{Y})$. (Such a collection may, for instance, be derived from the discrete Weyl operators defined in Section 4.1.2.) Let $\mathcal{Z} = \mathbb{C}^\Sigma$ and define a new channel $\Psi \in \mathrm{C}(\mathcal{Z} \otimes \mathcal{X}, \mathcal{Y})$ by the equation

$$\Psi(E_{a,b} \otimes X) = \begin{cases} U_a \Phi(X) U_a^* & \text{if } a = b \\ 0 & \text{otherwise} \end{cases} \tag{8.456}$$

holding for all $a, b \in \Sigma$ and $X \in \mathrm{L}(\mathcal{X})$.

In more intuitive terms, the action of the channel Ψ may be described as follows. A pair of registers (Z, X) is taken as input, and a measurement of the register Z with respect to the standard basis of $\mathcal{Z}$ is made, yielding a symbol $a \in \Sigma$. The channel Φ is applied to X, resulting in a register Y, and the unitary channel described by U_a is applied to Y. The measurement outcome a is discarded and Y is taken to be the output of the channel.

As the following proposition shows, the Holevo capacity of the channel Ψ constructed in this way is determined by the minimum output entropy of the channel Φ.

Proposition 8.63. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Also let Σ be an alphabet, let*

$$\{U_a \,:\, a \in \Sigma\} \subset \mathrm{U}(\mathcal{Y}) \tag{8.457}$$

be a collection of unitary operators for which the equation (8.455) *holds for all* $Y \in \mathrm{L}(\mathcal{Y})$*, let* $\mathcal{Z} = \mathbb{C}^{\Sigma}$*, and let* $\Psi \in \mathrm{C}(\mathcal{Z} \otimes \mathcal{X}, \mathcal{Y})$ *be a channel defined by the equation* (8.456) *holding for all* $a, b \in \Sigma$ *and* $X \in \mathrm{L}(\mathcal{X})$*. It holds that*

$$\chi(\Psi) = \log(\dim(\mathcal{Y})) - \mathrm{H}_{\min}(\Phi). \tag{8.458}$$

Proof. Consider first the ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Z} \otimes \mathcal{X})$ defined as

$$\eta(a) = \frac{1}{|\Sigma|} E_{a,a} \otimes \rho \tag{8.459}$$

for all $a \in \Sigma$, where $\rho \in \mathrm{D}(\mathcal{X})$ is any state for which

$$\mathrm{H}_{\min}(\Phi) = \mathrm{H}(\Phi(\rho)). \tag{8.460}$$

One has

$$\begin{aligned} \chi(\Psi(\eta)) &= \mathrm{H}\left(\frac{1}{|\Sigma|} \sum_{a \in \Sigma} U_a \Phi(\rho) U_a^* \right) - \frac{1}{|\Sigma|} \sum_{a \in \Sigma} \mathrm{H}\big(U_a \Phi(\rho) U_a^*\big) \\ &= \mathrm{H}\big(\Omega(\rho)\big) - \mathrm{H}(\Phi(\rho)) \\ &= \log(\dim(\mathcal{Y})) - \mathrm{H}_{\min}(\Phi). \end{aligned} \tag{8.461}$$

It therefore holds that

$$\chi(\Psi) \geq \log(\dim(\mathcal{Y})) - \mathrm{H}_{\min}(\Phi). \tag{8.462}$$

Next, consider an arbitrary state $\sigma \in \mathrm{D}(\mathcal{Z} \otimes \mathcal{X})$. For $\Delta \in \mathrm{C}(\mathcal{Z})$ denoting the completely dephasing channel, one may write

$$\big(\Delta \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(\sigma) = \sum_{a \in \Sigma} q(a) E_{a,a} \otimes \xi_a, \tag{8.463}$$

for some choice of a probability vector $q \in \mathcal{P}(\Sigma)$ and a collection of states

$$\{\xi_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X}). \tag{8.464}$$

It holds that

$$\Psi(\sigma) = \sum_{a \in \Sigma} q(a) U_a \Phi(\xi_a) U_a^* \tag{8.465}$$

and therefore

$$\mathrm{H}(\Psi(\sigma)) \geq \sum_{a \in \Sigma} q(a)\, \mathrm{H}\big(\Phi(\xi_a)\big) \geq \mathrm{H}_{\min}(\Phi) \tag{8.466}$$

by the concavity of the von Neumann entropy function (Theorem 5.25).

Finally, consider an arbitrary ensemble $\eta : \Gamma \to \mathrm{Pos}(\mathcal{Z} \otimes \mathcal{X})$, written as

$$\eta(b) = p(b)\sigma_b \tag{8.467}$$

for each $b \in \Gamma$, for $p \in \mathcal{P}(\Gamma)$ being a probability vector and

$$\{\sigma_b \,:\, b \in \Gamma\} \subseteq \mathrm{D}(\mathcal{Z} \otimes \mathcal{X}) \tag{8.468}$$

being a collection of states. It holds that

$$\begin{aligned} \chi(\Psi(\eta)) &= \mathrm{H}\left(\sum_{b\in\Gamma} p(b)\Psi(\sigma_b)\right) - \sum_{b\in\Gamma} p(b)\,\mathrm{H}(\Psi(\sigma_b)) \\ &\leq \log(\dim(\mathcal{Y})) - \mathrm{H}_{\min}(\Phi). \end{aligned} \tag{8.469}$$

The ensemble η was chosen arbitrarily, and therefore

$$\chi(\Psi) \leq \log(\dim(\mathcal{Y})) - \mathrm{H}_{\min}(\Phi), \tag{8.470}$$

which completes the proof. □

Theorem 8.64. *There exists a channel $\Psi \in \mathrm{C}(\mathcal{W}, \mathcal{Y})$, for some choice of complex Euclidean spaces $\mathcal{W}$ and $\mathcal{Y}$, such that*

$$\chi(\Psi \otimes \Psi) > 2\chi(\Psi). \tag{8.471}$$

Proof. By Corollary 8.62 there exist complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$ and a channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ for which the inequality

$$\mathrm{H}_{\min}(\Phi \otimes \Phi) < 2\,\mathrm{H}_{\min}(\Phi) \tag{8.472}$$

holds. Let Σ be an alphabet and let

$$\{U_a \,:\, a \in \Sigma\} \subset \mathrm{U}(\mathcal{Y}) \tag{8.473}$$

be a collection of unitary operators for which

$$\Omega(Y) = \frac{1}{|\Sigma|} \sum_{a\in\Sigma} U_a Y U_a^* \tag{8.474}$$

for all $Y \in \mathrm{L}(\mathcal{Y})$. Also let $\mathcal{Z} = \mathbb{C}^\Sigma$ and let $\Psi \in \mathrm{C}(\mathcal{Z} \otimes \mathcal{X}, \mathcal{Y})$ be a channel defined by the equation (8.456) above for all $a, b \in \Sigma$ and $X \in \mathrm{L}(\mathcal{X})$.

Up to a permutation of the tensor factors of its input space, $\Psi \otimes \Psi$ is equivalent to the channel $\Xi \in \mathrm{C}((\mathcal{Z} \otimes \mathcal{Z}) \otimes (\mathcal{X} \otimes \mathcal{X}), \mathcal{Y} \otimes \mathcal{Y})$ that would be obtained from the channel $\Phi \otimes \Phi$ through a similar construction, using the collection of unitary operators

$$\{U_a \otimes U_b \,:\, (a,b) \in \Sigma \times \Sigma\} \subset \mathrm{U}(\mathcal{Y} \otimes \mathcal{Y}). \tag{8.475}$$

It therefore follows from Proposition 8.63 that

$$\chi(\Psi) = \log(\dim(\mathcal{Y})) - \mathrm{H}_{\min}(\Phi) \tag{8.476}$$

while

$$\chi(\Psi \otimes \Psi) = \log(\dim(\mathcal{Y} \otimes \mathcal{Y})) - \mathrm{H}_{\min}(\Phi \otimes \Phi) > 2\chi(\Psi). \tag{8.477}$$

Taking $\mathcal{W} = \mathcal{Z} \otimes \mathcal{X}$, the theorem is therefore proved. □

One consequence of this theorem is that an analogous statement to the Holevo–Schumacher–Westmoreland theorem (Theorem 8.30), but without a regularization, does not hold in general; because

$$\mathrm{C}(\Phi) \geq \frac{\chi(\Phi \otimes \Phi)}{2}, \tag{8.478}$$

it is the case that $\mathrm{C}(\Phi) > \chi(\Phi)$ for some choices of a channel Φ.

8.3.2 Super-activation of quantum channel capacity

The purpose of the present subsection is to demonstrate the phenomenon of *super-activation*, in which the tensor product of two zero-capacity channels have positive quantum capacity. As a byproduct, one obtains an example of a channel Ψ satisfying

$$\mathrm{I_c}(\Psi \otimes \Psi) > 2\,\mathrm{I_c}(\Psi). \tag{8.479}$$

Two classes of zero-capacity channels

It is possible to prove that certain classes of channels have zero quantum capacity. Channels whose Choi operators are PPT and self-complementary channels fall into this category. The following proposition establishes that channels whose Choi operators are PPT must have zero capacity.

Proposition 8.65. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel such that $J(\Phi) \in \mathrm{PPT}(\mathcal{Y} : \mathcal{X})$. It holds that $\mathrm{Q}(\Phi) = 0$.*

Proof. The first step of the proof is to establish that, for every choice of a complex Euclidean space $\mathcal{W}$ and a density operator $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{W})$, one has

$$\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\rho) \in \mathrm{PPT}(\mathcal{Y} : \mathcal{W}). \tag{8.480}$$

Toward this goal, observe that, for any choice of a complex Euclidean space $\mathcal{W}$ and a positive semidefinite operator $P \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{W})$, there must exist a completely positive map $\Psi_P \in \mathrm{CP}(\mathcal{X}, \mathcal{W})$ satisfying

$$P = \big(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Psi_P\big)\big(\mathrm{vec}(\mathbb{1}_{\mathcal{X}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^*\big). \tag{8.481}$$

(The map Ψ_P is uniquely defined by this requirement—one may obtain its Choi representation by swapping the tensor factors of P.) It follows that, for any complex Euclidean space $\mathcal{W}$ and any density operator $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{W})$, one must have

$$\begin{aligned} &\big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\rho) \\ &\quad = \big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho\big)\big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(J(\Phi)) \in \mathrm{Pos}(\mathcal{Y} : \mathcal{W}) \end{aligned} \tag{8.482}$$

by virtue of the fact that Ψ_ρ is completely positive and $J(\Phi) \in \mathrm{PPT}(\mathcal{Y} : \mathcal{X})$, which establishes (8.480).

Now, it follows from the assumption $J(\Phi) \in \mathrm{PPT}(\mathcal{Y} : \mathcal{X})$ that

$$J\big(\Phi^{\otimes n}\big) \in \mathrm{PPT}\big(\mathcal{Y}^{\otimes n} : \mathcal{X}^{\otimes n}\big) \tag{8.483}$$

for every positive integer n. For every choice of positive integers n and m, for $\mathcal{Z} = \mathbb{C}^\Gamma$ for $\Gamma = \{0, 1\}$, and for any channel $\Xi \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$, it therefore holds that

$$\big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\rho) \in \mathrm{PPT}\big(\mathcal{Z}^{\otimes m} : \mathcal{Z}^{\otimes m}\big) \tag{8.484}$$

for every density operator $\rho \in \mathrm{D}(\mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m})$. By Proposition 6.47, one therefore has

$$\mathrm{F}\Big(2^{-m}\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\big)\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\big)^*,\ \big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\rho)\Big) \leq 2^{-m/2}. \tag{8.485}$$

For every choice of a positive real number $\alpha > 0$, it must therefore be the case that α fails to be an achievable rate for entanglement generation though Φ. Consequently, Φ has zero capacity for entanglement generation, which implies $\mathrm{Q}(\Phi) = 0$ by Theorem 8.50. $\square$

The second category of channels mentioned above having zero quantum capacity are *self-complementary* channels. These are channels $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ such that there exists an isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Y})$ such that

$$\Phi(X) = \big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \mathrm{Tr}\big)(AXA^*) = \big(\mathrm{Tr} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(AXA^*) \tag{8.486}$$

for every $X \in \mathrm{L}(\mathcal{X})$.

It follows from Proposition 8.17 that the coherent information of every state $\sigma \in \mathrm{D}(\mathcal{X})$ through a self-complementary channel Φ must be zero:

$$\mathrm{I_c}(\sigma; \Phi) = \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Phi(\sigma)) = 0. \tag{8.487}$$

As every tensor power of a self-complementary channel is necessarily self-complementary, the quantum capacity theorem (Theorem 8.59) implies that self-complementary channels have zero quantum capacity. The following proposition states a more general variant of this observation.

Proposition 8.66. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ be complementary channels, and suppose that there exists a channel $\Xi \in \mathrm{C}(\mathcal{Z}, \mathcal{Y})$ such that $\Phi = \Xi\Psi$. It holds that Φ has zero quantum capacity: $\mathrm{Q}(\Phi) = 0$.*

Proof. For every positive integer n, one has

$$\mathrm{I_c}(\sigma; \Phi^{\otimes n}) = \mathrm{I_c}(\sigma; \Xi^{\otimes n}\Psi^{\otimes n}) \leq \mathrm{I_c}(\sigma; \Psi^{\otimes n}) \tag{8.488}$$

by Proposition 8.15. Because Ψ is complementary to Φ, it holds that $\Psi^{\otimes n}$ is complementary to $\Phi^{\otimes n}$ for every positive integer n, and therefore

$$\begin{aligned} \mathrm{I_c}(\sigma; \Phi^{\otimes n}) &= \mathrm{H}\big(\Phi^{\otimes n}(\sigma)\big) - \mathrm{H}\big(\Psi^{\otimes n}(\sigma)\big) \\ &= -\mathrm{I_c}(\sigma; \Psi^{\otimes n}) \leq -\mathrm{I_c}(\sigma; \Phi^{\otimes n}), \end{aligned} \tag{8.489}$$

which implies $\mathrm{I_c}(\sigma; \Phi^{\otimes n}) \leq 0$. By Theorem 8.59, it therefore holds that $\mathrm{Q}(\Phi) = 0$, which completes the proof. □

Remark 8.67. Channels of the form $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ for which there exists a channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ complementary to Φ, as well as a channel $\Xi \in \mathrm{C}(\mathcal{Z}, \mathcal{Y})$ for which $\Phi = \Xi\Psi$, are known as *anti-degradable channels*.

50% erasure channels

A *50%-erasure channel* is a simple type of self-complementary channel that plays a special role in the example of super-activation to be presented below. For any choice of a complex Euclidean space $\mathcal{X}$, the 50%-erasure channel defined with respect to $\mathcal{X}$ is the channel $\Xi \in \mathrm{C}(\mathcal{X}, \mathbb{C} \oplus \mathcal{X})$ defined as

$$\Xi(X) = \frac{1}{2}\begin{pmatrix} \mathrm{Tr}(X) & 0 \\ 0 & X \end{pmatrix} \tag{8.490}$$

for each $X \in \mathrm{L}(\mathcal{X})$.

Intuitively speaking, a 50%-erasure channel acts as the identity channel with probability $1/2$, and otherwise its input is lost (or erased). Under the assumption that $\mathcal{X} = \mathbb{C}^{\Sigma}$, for Σ being a given alphabet, one may associate the complex Euclidean space $\mathbb{C} \oplus \mathcal{X}$ with $\mathbb{C}^{\{\#\} \cup \Sigma}$, for # being a special *blank symbol* that is not contained in Σ. With this interpretation, the event that the input is erased may be associated with the blank symbol # being produced, so that

$$\Xi(X) = \frac{1}{2}X + \frac{1}{2}\,\mathrm{Tr}(X)E_{\#,\#} \tag{8.491}$$

for every $X \in \mathrm{L}(\mathcal{X})$. It should be noted that there is no ambiguity about whether an erasure has occurred for this channel. The situation is analogous to one in which a letter is sent through a postal system; and with probability $1/2$, the letter is received, and otherwise an empty envelope is received (as opposed to the letter being lost without the receiver's knowledge).

For every choice of $\mathcal{X}$, the 50%-erasure channel $\Xi \in \mathrm{C}(\mathcal{X}, \mathbb{C} \oplus \mathcal{X})$ is self-complementary: using the association of $\mathbb{C} \oplus \mathcal{X}$ with $\mathbb{C}^{\{\#\} \cup \Sigma}$, for Σ being a given alphabet that does not contain the blank symbol #, one has

$$\Xi(X) = \big(\mathrm{Tr} \otimes \mathbb{1}\big)\big(AXA^*\big) = \big(\mathbb{1} \otimes \mathrm{Tr}\big)\big(AXA^*\big) \tag{8.492}$$

for $A \in \mathrm{U}(\mathcal{X}, (\mathbb{C} \oplus \mathcal{X}) \otimes (\mathbb{C} \oplus \mathcal{X}))$ being the isometry defined as

$$A = \frac{1}{\sqrt{2}} \sum_{a \in \Sigma} (e_a \otimes e_\# + e_\# \otimes e_a) e_a^*. \tag{8.493}$$

It follows that $\mathrm{Q}(\Xi) = 0$.

A theorem of Smith and Yard

The following theorem allows one to prove lower bounds on the maximum coherent information of a channel tensored with a 50%-erasure channel on a sufficiently large space. For a suitable choice of a zero-capacity channel tensored with a 50%-erasure channel, the theorem leads to a demonstration of the super-activation phenomenon.

Theorem 8.68 (Smith–Yard)**.** *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ be an isometry, and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ be complementary channels defined as*

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}(AXA^*) \quad \textit{and} \quad \Psi(X) = \mathrm{Tr}_{\mathcal{Y}}(AXA^*) \tag{8.494}$$

for every $X \in \mathrm{L}(\mathcal{X})$. Also let Σ be an alphabet, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble of states, and let $\mathcal{W}$ be a complex Euclidean space satisfying

$$\dim(\mathcal{W}) \geq \sum_{a \in \Sigma} \mathrm{rank}(\eta(a)). \tag{8.495}$$

There exists a density operator $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{W})$ such that

$$\mathrm{I_c}(\rho; \Phi \otimes \Xi) = \frac{1}{2}\chi(\Phi(\eta)) - \frac{1}{2}\chi(\Psi(\eta)), \tag{8.496}$$

for $\Xi \in \mathrm{C}(\mathcal{W}, \mathbb{C} \oplus \mathcal{W})$ denoting the 50%-erasure channel on $\mathcal{W}$.

Proof. By the assumption

$$\dim(\mathcal{W}) \geq \sum_{a \in \Sigma} \mathrm{rank}(\eta(a)), \tag{8.497}$$

one may choose a collection of vectors $\{u_a \,:\, a \in \Sigma\} \subset \mathcal{X} \otimes \mathcal{W}$ for which it holds that

$$\mathrm{Tr}_{\mathcal{W}}(u_a u_a^*) = \eta(a) \tag{8.498}$$

for each $a \in \Sigma$, and for which it holds that

$$\left\{\mathrm{Tr}_{\mathcal{X}}(u_a u_a^*) \,:\, a \in \Sigma\right\} \tag{8.499}$$

is an orthogonal set of operators. Let $\mathcal{V} = \mathbb{C}^{\Sigma}$, define a unit vector

$$u = \sum_{a \in \Sigma} e_a \otimes u_a \in \mathcal{V} \otimes \mathcal{X} \otimes \mathcal{W}, \tag{8.500}$$

and let $\rho = \mathrm{Tr}_{\mathcal{V}}(uu^*)$. One may observe that, by virtue of the fact that (8.499) is an orthogonal set, it holds that

$$\mathrm{Tr}_{\mathcal{W}}(uu^*) = \sum_{a\in\Sigma} E_{a,a} \otimes \eta(a). \tag{8.501}$$

For the unit vector $v \in \mathcal{V} \otimes \mathcal{Y} \otimes \mathcal{Z} \otimes \mathcal{W}$ defined as $v = (\mathbb{1}_{\mathcal{V}} \otimes A \otimes \mathbb{1}_{\mathcal{W}})u$, it therefore holds that

$$\mathrm{Tr}_{\mathcal{W}}(vv^*) = \sum_{a\in\Sigma} E_{a,a} \otimes A\eta(a)A^*. \tag{8.502}$$

The 50%-erasure channel Ξ has the property that

$$\mathrm{H}\big((\Phi \otimes \Xi)(\rho)\big) = \frac{1}{2}\mathrm{H}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho)\big) + \frac{1}{2}\mathrm{H}\big(\Phi(\mathrm{Tr}_{\mathcal{W}}(\rho))\big) + 1, \tag{8.503}$$

and likewise for the channel Ψ in place of Φ. As Ψ is complementary to Φ and Ξ is self-complementary, it follows that

$$\begin{aligned}
\mathrm{I_C}(\rho; \Phi \otimes \Xi) &= \mathrm{H}\big((\Phi \otimes \Xi)(\rho)\big) - \mathrm{H}\big((\Psi \otimes \Xi)(\rho)\big) \\
&= \frac{1}{2}\big(\mathrm{H}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho)\big) - \mathrm{H}\big((\Psi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho)\big)\big) \\
&\quad + \frac{1}{2}\big(\mathrm{H}\big(\Phi(\mathrm{Tr}_{\mathcal{W}}(\rho))\big) - \mathrm{H}\big(\Psi(\mathrm{Tr}_{\mathcal{W}}(\rho))\big)\big).
\end{aligned} \tag{8.504}$$

Now, let V, Y, Z, and W be registers corresponding to the spaces $\mathcal{V}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$, respectively, and consider the situation in which the compound register $(\mathsf{V}, \mathsf{Y}, \mathsf{Z}, \mathsf{W})$ is in the pure state vv^*. It holds that

$$\begin{aligned}
\mathrm{H}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho)\big) &= \mathrm{H}(\mathsf{Y}, \mathsf{W}) = \mathrm{H}(\mathsf{V}, \mathsf{Z}), \\
\mathrm{H}\big((\Psi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho)\big) &= \mathrm{H}(\mathsf{Z}, \mathsf{W}) = \mathrm{H}(\mathsf{V}, \mathsf{Y}), \\
\mathrm{H}\big(\Phi(\mathrm{Tr}_{\mathcal{W}}(\rho))\big) &= \mathrm{H}(\mathsf{Y}), \\
\mathrm{H}\big(\Psi(\mathrm{Tr}_{\mathcal{W}}(\rho))\big) &= \mathrm{H}(\mathsf{Z}),
\end{aligned} \tag{8.505}$$

and therefore

$$\mathrm{I_C}(\rho; \Phi \otimes \Xi) = \frac{1}{2}\mathrm{I}(\mathsf{V} : \mathsf{Y}) - \frac{1}{2}\mathrm{I}(\mathsf{V} : \mathsf{Z}) = \frac{1}{2}\chi(\Phi(\eta)) - \frac{1}{2}\chi(\Psi(\eta)), \tag{8.506}$$

as required. □

$$A_1 = \begin{pmatrix} 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 \\ \gamma & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha \\ -\gamma & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$A_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad A_5 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\beta \end{pmatrix}, \quad A_6 = \begin{pmatrix} 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \end{pmatrix},$$

Figure 8.2: Kraus operators $A_1, \ldots, A_6 \in \mathrm{L}(\mathbb{C}^4)$ for the channel Φ.

An explicit example of super-activation

An example of the super-activation phenomenon, based on Theorem 8.68, will now be described. The first step is to define a zero-capacity channel Φ as follows. Let

$$\alpha = \sqrt{\sqrt{2} - 1}, \quad \beta = \sqrt{1 - \frac{1}{\sqrt{2}}}, \quad \text{and} \quad \gamma = \sqrt{\frac{1}{\sqrt{2}} - \frac{1}{2}}, \tag{8.507}$$

define $A_1, \ldots, A_6 \in \mathrm{L}(\mathbb{C}^4)$ as in Figure 8.2, and define $\Phi \in \mathrm{C}(\mathbb{C}^4)$ as

$$\Phi(X) = \sum_{k=1}^{6} A_k X A_k^* \tag{8.508}$$

for every $X \in \mathrm{L}(\mathbb{C}^4)$.

The fact that Φ is a zero-capacity channel follows from the fact that the Choi representation of Φ is a PPT operator. One way to verify this claim is to check that

$$\big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^4)}\big)(J(\Phi)) = J(\Theta) \tag{8.509}$$

for $\Theta \in \mathrm{C}(\mathbb{C}^4)$ being the channel defined as

$$\Theta(X) = \sum_{k=1}^{6} B_k X B_k^* \tag{8.510}$$

for every $X \in \mathrm{L}(\mathbb{C}^4)$, where $B_1, \ldots, B_6 \in \mathrm{L}(\mathbb{C}^4)$ are as specified in Figure 8.3. It therefore follows from Proposition 8.65 that Φ has zero quantum capacity.

$$B_1 = \begin{pmatrix} 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 \\ \gamma & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha \\ \gamma & 0 & 0 & 0 \\ 0 & -\gamma & 0 & 0 \end{pmatrix}, \quad B_3 = \begin{pmatrix} \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$B_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & -\beta \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_5 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta \end{pmatrix}, \quad B_6 = \begin{pmatrix} 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \end{pmatrix}.$$

Figure 8.3: Kraus operators $B_1, \ldots, B_6 \in \mathrm{L}(\mathbb{C}^4)$ for the channel Θ.

A channel complementary to Φ is given by $\Psi \in \mathrm{C}(\mathbb{C}^4, \mathbb{C}^6)$ defined as

$$\Psi(X) = \sum_{k=1}^{4} C_k X C_k^* \tag{8.511}$$

for every $X \in \mathrm{L}(\mathbb{C}^4)$, where

$$\begin{aligned} C_1 &= \begin{pmatrix} 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 \\ \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha \\ 0 & 0 & 0 & 0 \\ \beta & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\ C_3 &= \begin{pmatrix} \gamma & 0 & 0 & 0 \\ -\gamma & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & \beta \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad C_4 = \begin{pmatrix} 0 & \gamma & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\beta \\ 0 & 0 & \beta & 0 \end{pmatrix}. \end{aligned} \tag{8.512}$$

Finally, define $\Sigma = \{0, 1\}$, define density operators

$$\sigma_0 = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \sigma_1 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{8.513}$$

and define an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathbb{C}^4)$ as $\eta(0) = \sigma_0/2$ and $\eta(1) = \sigma_1/2$. It holds that

$$\Phi(\sigma_0) = \begin{pmatrix} \frac{2-\sqrt{2}}{2} & 0 & 0 & 0 \\ 0 & \frac{2-\sqrt{2}}{2} & 0 & 0 \\ 0 & 0 & \frac{\sqrt{2}-1}{2} & 0 \\ 0 & 0 & 0 & \frac{\sqrt{2}-1}{2} \end{pmatrix} \tag{8.514}$$

and

$$\Phi(\sigma_1) = \begin{pmatrix} \frac{\sqrt{2}-1}{2} & 0 & 0 & 0 \\ 0 & \frac{\sqrt{2}-1}{2} & 0 & 0 \\ 0 & 0 & \frac{2-\sqrt{2}}{2} & 0 \\ 0 & 0 & 0 & \frac{2-\sqrt{2}}{2} \end{pmatrix}, \tag{8.515}$$

while

$$\Psi(\sigma_0) = \Psi(\sigma_1) = \begin{pmatrix} \frac{\sqrt{2}-1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\sqrt{2}-1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{2-\sqrt{2}}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2-\sqrt{2}}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{2-\sqrt{2}}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2-\sqrt{2}}{4} \end{pmatrix}. \tag{8.516}$$

One therefore has that

$$\chi(\Phi(\eta)) = \mathrm{H}\left(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}\right) - \mathrm{H}\left(\tfrac{2-\sqrt{2}}{2}, \tfrac{2-\sqrt{2}}{2}, \tfrac{\sqrt{2}-1}{2}, \tfrac{\sqrt{2}-1}{2}\right) > \tfrac{1}{50}, \tag{8.517}$$

while

$$\chi(\Psi(\eta)) = 0. \tag{8.518}$$

By Theorem 8.68, there must exist a density operator $\rho \in \mathrm{D}(\mathbb{C}^4 \otimes \mathbb{C}^4)$ such that

$$\mathrm{I_c}(\rho; \Phi \otimes \Xi) > \frac{1}{100}, \tag{8.519}$$

for $\Xi \in \mathrm{C}(\mathbb{C}^4, \mathbb{C} \oplus \mathbb{C}^4)$ being a 50%-erasure channel. One therefore has that $\mathrm{Q}(\Phi) = \mathrm{Q}(\Xi) = 0$, while $\mathrm{Q}(\Phi \otimes \Xi) > 0$.

The need for a regularization in the quantum capacity theorem

The super-activation example described above illustrates that the maximum coherent information is not additive; one has

$$\mathrm{I_c}(\Phi \otimes \Xi) > \mathrm{I_c}(\Phi) + \mathrm{I_c}(\Xi) \tag{8.520}$$

for the channels Φ and Ξ specified in that example. As these channels are different, it does not follow immediately that a strict inequality of the form

$$\mathrm{I_c}(\Psi^{\otimes n}) > n\,\mathrm{I_c}(\Psi) \tag{8.521}$$

holds for any choice of a channel Ψ and a positive integer n. It is possible, however, to conclude that such an inequality does hold (for $n = 2$) using a direct sum construction along similar lines to the one used in the context of the Holevo capacity and minimum output entropy. The following three propositions that concern direct sums of channels will be used to reach this conclusion.

Proposition 8.69. *Let $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, $\mathcal{Y}_1$, $\mathcal{Z}_0$, and $\mathcal{Z}_1$ be complex Euclidean spaces, and let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$, $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$, $\Psi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Z}_0)$, and $\Psi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Z}_1)$ be channels such that Ψ_0 is complementary to Φ_0 and Ψ_1 is complementary to Φ_1. The channel $\Psi_0 \oplus \Psi_1$ is complementary to $\Phi_0 \oplus \Phi_1$.*

Proof. Let $A_0 \in \mathrm{U}(\mathcal{X}_0, \mathcal{Y}_0 \otimes \mathcal{Z}_0)$ and $A_1 \in \mathrm{U}(\mathcal{X}_1, \mathcal{Y}_1 \otimes \mathcal{Z}_1)$ be isometries such that the following equations hold for all $X_0 \in \mathrm{L}(\mathcal{X}_0)$ and $X_1 \in \mathrm{L}(\mathcal{X}_1)$:

$$\begin{aligned} \Phi_0(X_0) &= \mathrm{Tr}_{\mathcal{Z}_0}\bigl(A_0 X_0 A_0^*\bigr), \qquad & \Psi_0(X_0) &= \mathrm{Tr}_{\mathcal{Y}_0}\bigl(A_0 X_0 A_0^*\bigr), \\ \Phi_1(X_1) &= \mathrm{Tr}_{\mathcal{Z}_1}\bigl(A_1 X_1 A_1^*\bigr), \qquad & \Psi_1(X_1) &= \mathrm{Tr}_{\mathcal{Y}_1}\bigl(A_1 X_1 A_1^*\bigr). \end{aligned} \tag{8.522}$$

Let $W \in \mathrm{U}\bigl((\mathcal{Y}_0 \otimes \mathcal{Z}_0) \oplus (\mathcal{Y}_1 \otimes \mathcal{Z}_1), (\mathcal{Y}_0 \oplus \mathcal{Y}_1) \otimes (\mathcal{Z}_0 \oplus \mathcal{Z}_1)\bigr)$ be the isometry defined by the equation

$$\begin{aligned} &W\bigl((y_0 \otimes z_0) \oplus (y_1 \otimes z_1)\bigr) \\ &\quad = (y_0 \oplus 0) \otimes (z_0 \oplus 0) + (0 \oplus y_1) \otimes (0 \oplus z_1) \end{aligned} \tag{8.523}$$

for every $y_0 \in \mathcal{Y}_0$, $y_1 \in \mathcal{Y}_1$, $z_0 \in \mathcal{Z}_0$, and $z_1 \in \mathcal{Z}_1$. The equations

$$\begin{aligned} (\Phi_0 \oplus \Phi_1)(X) &= \mathrm{Tr}_{\mathcal{Z}_0 \oplus \mathcal{Z}_1}\left(W \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix} X \begin{pmatrix} A_0^* & 0 \\ 0 & A_1^* \end{pmatrix} W^* \right) \\ (\Psi_0 \oplus \Psi_1)(X) &= \mathrm{Tr}_{\mathcal{Y}_0 \oplus \mathcal{Y}_1}\left(W \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix} X \begin{pmatrix} A_0^* & 0 \\ 0 & A_1^* \end{pmatrix} W^* \right) \end{aligned} \tag{8.524}$$

hold for all $X \in \mathrm{L}(\mathcal{X}_0 \oplus \mathcal{X}_1)$, which implies that $\Psi_0 \oplus \Psi_1$ is complementary to $\Phi_0 \oplus \Phi_1$, as required. $\square$

Proposition 8.70. *Let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be channels, for $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$ being complex Euclidean spaces, and let $\sigma \in \mathrm{D}(\mathcal{X}_0 \oplus \mathcal{X}_1)$ be an arbitrary density operator, written as*

$$\sigma = \begin{pmatrix} \lambda\sigma_0 & X \\ X^* & (1-\lambda)\sigma_1 \end{pmatrix} \tag{8.525}$$

for $\lambda \in [0,1]$, $\sigma_0 \in \mathrm{D}(\mathcal{X}_0)$, $\sigma_1 \in \mathrm{D}(\mathcal{X}_1)$, and $X \in \mathrm{L}(\mathcal{X}_1, \mathcal{X}_0)$. It holds that

$$\mathrm{I_c}(\sigma; \Phi_0 \oplus \Phi_1) = \lambda\, \mathrm{I_c}(\sigma_0; \Phi_0) + (1-\lambda)\, \mathrm{I_c}(\sigma_1; \Phi_1). \tag{8.526}$$

Proof. Observe first that

$$\begin{aligned} \mathrm{H}((\Phi_0 \oplus \Phi_1)(\sigma)) &= \mathrm{H}\begin{pmatrix} \lambda\Phi_0(\sigma_0) & 0 \\ 0 & (1-\lambda)\Phi_1(\sigma_1) \end{pmatrix} \\ &= \lambda\, \mathrm{H}(\Phi_0(\sigma_0)) + (1-\lambda)\, \mathrm{H}(\Phi_1(\sigma_1)) + \mathrm{H}(\lambda, 1-\lambda). \end{aligned} \tag{8.527}$$

Assuming that $\mathcal{Z}_0$ and $\mathcal{Z}_1$ are complex Euclidean spaces and $\Psi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Z}_0)$ and $\Psi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Z}_1)$ are channels complementary to Φ_0 and Φ_1, respectively, one has that

$$\begin{aligned} &\mathrm{H}((\Psi_0 \oplus \Psi_1)(\sigma)) \\ &\quad = \lambda\, \mathrm{H}(\Psi_0(\sigma_0)) + (1-\lambda)\, \mathrm{H}(\Psi_1(\sigma_1)) + \mathrm{H}(\lambda, 1-\lambda) \end{aligned} \tag{8.528}$$

by a similar calculation to (8.527). As $\Psi_0 \oplus \Psi_1$ is complementary to $\Phi_0 \oplus \Phi_1$, as established in Proposition 8.69, it follows that

$$\begin{aligned} \mathrm{I_c}(\sigma; \Phi_0 \oplus \Phi_1) &= \mathrm{H}((\Phi_0 \oplus \Phi_1)(\sigma)) - \mathrm{H}((\Psi_0 \oplus \Psi_1)(\sigma)) \\ &= \lambda\big(\mathrm{H}(\Phi_0(\sigma_0)) - \mathrm{H}(\Psi_0(\sigma_0))\big) \\ &\qquad + (1-\lambda)\big(\mathrm{H}(\Phi_1(\sigma_1)) - \mathrm{H}(\Psi_1(\sigma_1))\big) \\ &= \lambda\, \mathrm{I_c}(\sigma_0; \Phi_0) + (1-\lambda)\, \mathrm{I_c}(\sigma_1; \Phi_1) \end{aligned} \tag{8.529}$$

as required. $\square$

Proposition 8.71. *Let $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$ be complex Euclidean spaces and let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be channels. It holds that*

$$\mathrm{I_c}((\Phi_0 \oplus \Phi_1) \otimes (\Phi_0 \oplus \Phi_1)) \geq \mathrm{I_c}(\Phi_0 \otimes \Phi_1). \tag{8.530}$$

Proof. Define an isometry $W \in \mathrm{U}\big(\mathcal{X}_0 \otimes \mathcal{X}_1, (\mathcal{X}_0 \oplus \mathcal{X}_1) \otimes (\mathcal{X}_0 \oplus \mathcal{X}_1)\big)$ by the equation

$$W(x_0 \otimes x_1) = (x_0 \oplus 0) \otimes (0 \oplus x_1) \tag{8.531}$$

holding for all $x_0 \in \mathcal{X}_0$ and $x_1 \in \mathcal{X}_1$, and along similar lines, define an isometry $V \in \mathrm{U}\big(\mathcal{Y}_0 \otimes \mathcal{Y}_1, (\mathcal{Y}_0 \oplus \mathcal{Y}_1) \otimes (\mathcal{Y}_0 \oplus \mathcal{Y}_1)\big)$ by the equation

$$V(y_0 \otimes y_1) = (y_0 \oplus 0) \otimes (0 \oplus y_1) \tag{8.532}$$

for all $y_0 \in \mathcal{Y}_0$ and $y_1 \in \mathcal{Y}_1$. One has that

$$\begin{aligned}
&\big((\Phi_0 \oplus \Phi_1) \otimes (\Phi_0 \oplus \Phi_1)\big)(W(X_0 \otimes X_1)W^*) \\
&\qquad = \begin{pmatrix} \Phi_0(X_0) & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & \Phi_1(X_1) \end{pmatrix} \\
&\qquad = V(\Phi_0(X_0) \otimes \Phi_1(X_1))V^*
\end{aligned} \tag{8.533}$$

for all $X_0 \in \mathrm{L}(\mathcal{X}_0)$ and $X_1 \in \mathrm{L}(\mathcal{X}_1)$. For every choice of a density operator $\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)$, it follows that

$$\mathrm{I_c}(W\sigma W^*; (\Phi_0 \oplus \Phi_1) \otimes (\Phi_0 \oplus \Phi_1)) = \mathrm{I_c}(\sigma; \Phi_0 \otimes \Phi_1), \tag{8.534}$$

which implies the proposition. □

Finally, consider the channel $\Psi = \Phi \oplus \Xi$, for Φ and Ξ as in the example of super-activation described above. By Proposition 8.70, one may conclude that $\mathrm{I_c}(\Phi \oplus \Xi) = 0$, while Proposition 8.71 implies

$$\mathrm{I_c}\big((\Phi \oplus \Xi) \otimes (\Phi \oplus \Xi)\big) \geq \mathrm{I_c}(\Phi \otimes \Xi) > 0. \tag{8.535}$$

It therefore holds that the channel $\Psi = \Phi \oplus \Xi$ satisfies the strict inequality (8.521) for $n = 2$.

As a consequence of this fact, one has that the quantum capacity and maximum coherent information differ for some channels. In this sense, the regularization in the quantum capacity theorem (Theorem 8.59) is similar to the one in the Holevo–Schumacher–Westmoreland theorem (Theorem 8.30) in that it cannot generally be removed.

8.4 Exercises

8.1. Let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be channels, for an arbitrary choice of complex Euclidean spaces $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$.

(a) Prove that

$$\mathrm{I_c}(\Phi_0 \oplus \Phi_1) = \max\bigl\{\mathrm{I_c}(\Phi_0), \mathrm{I_c}(\Phi_1)\bigr\}. \tag{8.536}$$

(b) Prove that

$$\chi(\Phi_0 \oplus \Phi_1) = \max_{\lambda \in [0,1]} \Bigl(\lambda\chi(\Phi_0) + (1-\lambda)\chi(\Phi_1) + \mathrm{H}(\lambda, 1-\lambda)\Bigr). \tag{8.537}$$

8.2. Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{Z}, \mathcal{W})$ be channels, and assume that Φ is an entanglement breaking channel (q.v. Exercise 6.1). Prove that the following identities hold:

(a) $\mathrm{H_{min}}(\Phi \otimes \Psi) = \mathrm{H_{min}}(\Phi) + \mathrm{H_{min}}(\Psi)$.

(b) $\chi(\Phi \otimes \Psi) = \chi(\Phi) + \chi(\Psi)$.

(c) $\mathrm{I_c}(\Phi \otimes \Psi) = \mathrm{I_c}(\Psi)$.

The inequality established by a correct answer to Exercise 5.6 is useful for solving this exercise.

8.3. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. It is said that Φ is *degradable* if and only if there exists a complex Euclidean space $\mathcal{Z}$ and a channel $\Psi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ for which it holds that $\Psi\Phi$ is complementary to Φ.

(a) Prove that, for any choice of a degradable channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, states $\sigma_0, \sigma_1 \in \mathrm{D}(\mathcal{X})$, and a real number $\lambda \in [0,1]$, the following inequality holds:

$$\mathrm{I_c}\bigl(\lambda\sigma_0 + (1-\lambda)\sigma_1; \Phi\bigr) \geq \lambda\,\mathrm{I_c}(\sigma_0; \Phi) + (1-\lambda)\,\mathrm{I_c}(\sigma_1; \Phi). \tag{8.538}$$

(Equivalently, the function $\sigma \mapsto \mathrm{I_c}(\sigma; \Phi)$ defined on $\mathrm{D}(\mathcal{X})$ is concave.)

(b) Let X_0, X_1, Y_0, and Y_1 be registers, let $\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)$ be an arbitrary state of the pair $(\mathsf{X}_0, \mathsf{X}_1)$, and let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be degradable channels. Prove that

$$\mathrm{I_c}(\sigma; \Phi_0 \otimes \Phi_1) = \mathrm{I_c}(\sigma[\mathsf{X}_0]; \Phi_0) + \mathrm{I_c}(\sigma[\mathsf{X}_1]; \Phi_1). \tag{8.539}$$

8.4. Let $\mathcal{X}$ be a complex Euclidean space, let $\lambda \in [0,1]$, and define a channel $\Xi \in \mathrm{C}(\mathcal{X}, \mathbb{C} \oplus \mathcal{X})$ as

$$\Xi(X) = \begin{pmatrix} \lambda \operatorname{Tr}(X) & 0 \\ 0 & (1-\lambda)X \end{pmatrix} \tag{8.540}$$

for all $X \in \mathrm{L}(\mathcal{X})$.

(a) Give a closed-form expression for the coherent information $\mathrm{I}_{\mathrm{C}}(\sigma;\Xi)$ of an arbitrary state $\sigma \in \mathrm{D}(\mathcal{X})$ through Ξ.

(b) Give a closed-form expression for the entanglement-assisted classical capacity $\mathrm{C}_{\mathrm{E}}(\Xi)$ of Ξ.

(c) Give a closed-form expression for the quantum capacity $\mathrm{Q}(\Xi)$ of Ξ.

The results of Exercise 8.3 may be helpful when solving this exercise. The closed-form expressions should be functions of λ and $n = \dim(\mathcal{X})$ alone.

8.5. Let n be a positive integer, let $\mathcal{X} = \mathbb{C}^{\mathbb{Z}_n}$, and let

$$\{W_{a,b} \,:\, a, b \in \mathbb{Z}_n\} \tag{8.541}$$

denote the set of discrete Weyl operators acting on $\mathcal{X}$ (q.v. Section 4.1.2 of Chapter 4). Also let $p \in \mathcal{P}(\mathbb{Z}_n)$ be a probability vector, and define a channel $\Phi \in \mathrm{C}(\mathcal{X})$ as

$$\Phi(X) = \sum_{a \in \mathbb{Z}_n} p(a) W_{0,a} X W_{0,a}^* \tag{8.542}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Prove that

$$\mathrm{I}_{\mathrm{C}}(\Phi) = \log(n) - \mathrm{H}(p). \tag{8.543}$$

Like the previous problem, the results of Exercise 8.3 may be helpful when solving this exercise.

8.6. For every positive integer n and every real number $\varepsilon \in [0,1]$, define a channel $\Phi_{n,\varepsilon} \in \mathrm{C}(\mathbb{C}^n)$ as

$$\Phi_{n,\varepsilon} = \varepsilon \mathbb{1}_n + (1-\varepsilon)\Omega_n, \tag{8.544}$$

where $\mathbb{1}_n \in \mathrm{C}(\mathbb{C}^n)$ and $\Omega_n \in \mathrm{C}(\mathbb{C}^n)$ denote the identity and completely depolarizing channels defined with respect to the space $\mathbb{C}^n$. Prove that, for every choice of a positive real number K, there exists a choice of n and ε for which

$$\mathrm{C}_{\mathrm{E}}(\Phi_{n,\varepsilon}) \geq K\,\mathrm{C}(\Phi_{n,\varepsilon}) > 0. \tag{8.545}$$

8.5 Bibliographic remarks

The study of quantum channel capacities is, perhaps obviously, motivated in large part by Shannon's channel coding theorem [191], and the goal of obtaining analogous statements for quantum channels. It was soon realized, however, that there would not be a single capacity of a quantum channel, but rather several inequivalent but nevertheless fundamentally interesting capacities. The 1998 survey of Bennett and Shor [38] provides a summary of what was known about channel capacities at a relatively early point in their study.

Holevo [111] and Schumacher and Westmoreland [188] independently proved the Holevo–Schumacher–Westmoreland theorem (Theorem 8.30), in both cases building on Hausladen, Jozsa, Schumacher, Westmoreland, and Wootters [92]. The definition of what is now called the Holevo capacity (or the *Holevo information* of a channel) originates with the work of Holevo and Schumacher and Westmoreland. Lemma 8.28 was proved by Hayashi and Nagaoka [94], who used it in the analysis of generalizations of the Holevo–Schumacher–Westmoreland theorem.

The entanglement-assisted classical capacity theorem (Theorem 8.44) was proved by Bennett, Shor, Smolin, and Thapliyal [39]. The proof of this theorem presented in this chapter is due to Holevo [112]. Lemma 8.41 is due to Adami and Cerf [2].

Tasks involving quantum information transmission through quantum channels, along with fundamental definitions connected with such tasks, were investigated in several papers, including papers of Schumacher [186], Schumacher and Nielsen [187], Adami and Cerf [2], and Barnum, Nielsen, and Schumacher [22]. The entanglement generation capacity of a channel was defined explicitly by Devetak [58], and Theorems 8.49 and 8.50 follow from results proved by Barnum, Knill, and Nielsen [21].

The coherent information of a state through a channel was defined by Schumacher and Nielsen [187]. Lloyd [150] recognized the fundamental connection between the maximum coherent information of a channel and its quantum capacity, and provided a heuristic argument in support of the quantum capacity theorem (Theorem 8.59). The first rigorous proof of the quantum capacity theorem to be published was due to Devetak [58]. Shor reported a different proof of this theorem prior to Devetak's proof, although it was not published. A proof appearing in a subsequent paper of Hayden,

Shor, and Winter [98] reportedly resembles Shor's original proof.

The proof of the quantum capacity theorem presented in this chapter is due to Hayden, M. Horodecki, Winter, and Yard [95], incorporating some simplifying ideas due to Klesse [132], who independently proved the same theorem based on similar techniques. The phenomenon of decoupling (as represented by Lemma 8.53) provides a key step in this proof; this basic technique was used by Devetak [58], and was identified more explicitly by M. Horodecki, Oppenheim, and Winter [118] and Abeyesinghe, Devetak, Hayden, and Winter [1]. The PhD thesis of Dupuis [62] may be consulted for further information on this technique.

Shor [192] proved that the non-additivity of Holevo capacity follows from the non-additivity of minimum output entropy. In the same paper, Shor also proved the converse implication, which naturally had greater relevance prior to Hastings proof that the minimum output entropy is non-additive, along with the equivalence of these two non-additivity statements with two other statements concerning the entanglement of formation. The direct sum construction of channels and its implications to the additivity of channel capacities was investigated by Fukuda and Wolf [76].

DiVincenzo, Shor, and Smolin [61] proved that the coherent information is non-additive in 1998. Various properties of quantum erasure channels were established by Bennett, DiVincenzo, and Smolin [35]. Theorem 8.68, along with the realization that it gives an example of the super-activation phenomenon, is due to Smith and Yard [194]. The channel Φ described in the chapter giving rise to an example of super-activation, which appears in Smith and Yard's paper as well, was identified by K. Horodecki, Pankowski, M. Horodecki, and P. Horodecki [115], as it relates to a different capacity known as the *private capacity* of a channel.