A random-coding based proof for the quantum coding theorem

Rochus Klesse*

Universität zu Köln, Institut für Theoretische Physik, Zülpicher Str. 77, D-50937 Köln, Germany

October 16, 2007

Abstract

We present a proof for the quantum channel coding theorem which relies on the fact that a randomly chosen code space typically is highly suitable for quantum error correction. In this sense, the proof is close to Shannon's original treatment of information transmission via a noisy classical channel.

1 Preliminaries

1.1 Quantum channel

In the theory of information transmission the information is ascribed to the configuration of a physical system, and the transmission is ascribed to the dynamical evolution of that configuration under the influence of an in general noisy environment. It is therefore customary to characterize an information carrying system solely by its configuration space, and to consider its intrinsic dynamics as part of the transmission.

In a quantum setting we identify a system Q with its Hilbert space, denoted by the same symbol Q. Its dimension $|Q|$ will be always assumed to be finite. The system's configuration is a quantum state described by a density operator ρ in $\mathcal{B}(Q)$, the set of bounded operators on Q.

The process of information transmission can be any dynamics of an open quantum system Q according to which an initial input state ρ evolves to a final output state ρ', defining in this way the operation of a quantum channel $\mathcal{N}$ [1]. Mathematically, $\mathcal{N}$ is a completely positive mapping of $\mathcal{B}(Q)$ onto itself, or, when we admit that the

*Email address: rk@thp.uni-koeln.de

[1] For an introduction into the theory of quantum information see e.g. [1, 2].

system may change to an other system Q' during the course of transmission, onto $\mathcal{B}(Q')$, the set of bounded operators on Q',

$$\begin{aligned} \mathcal{N} \,:\, \mathcal{B}(Q) &\to \mathcal{B}(Q') \\ \rho &\mapsto \rho' = \mathcal{N}(\rho)\,. \end{aligned}$$

According to Stinespring's theorem [3] the operation of the channel can be always understood as an isometric transformation followed by a restriction [4, 2]. That is, one always finds an ancilla system E with $|E| \geq 1$ and an isometric operation $V : Q \to Q'E$ such that for all states ρ

$$\mathcal{N}(\rho) = \mathrm{tr}_E V \rho V^\dagger\,,$$

where tr_E denotes the partial trace over E. In the following we refer to this construction as Stinespring representation. An elementary physical interpretation of it becomes obvious in the case $Q = Q'$. Here one can find a unitary operator U on QE and a state vector $|\varphi_E\rangle \in E$ such that $V|\psi\rangle = U|\psi\rangle \otimes |\varphi_E\rangle$ for all state vectors $|\psi\rangle \in Q$. Interpreting U as time evolution operator of the joint system QE, an initial state $\rho \otimes \varphi_E$, where $\varphi_E = |\varphi_E\rangle\langle\varphi_E|$, will evolve to final state $U\rho \otimes \varphi_E U^\dagger$. Its partial trace with respect to E yields indeed $\mathcal{N}(\rho)$ as the reduced density operator for Q,

$$\mathrm{tr}_E U\rho \otimes \varphi_E U^\dagger = \mathrm{tr}_E V\rho V^\dagger = \mathcal{N}(\rho)\,.$$

If we fix an orthonormal basis $|1\rangle, \ldots |N\rangle$ of E [2], the Stinespring representation can be rewritten more explicitly in an operator sum as

$$\mathcal{N}(\rho) = \sum_{k=1}^{N} A_k \rho {A_k}^\dagger\,,$$

where Kraus operators $A_1, \ldots, A_N : Q \to Q'$ are defined by $A|\psi\rangle := \langle k|V|\psi\rangle$ [4, 1, 2]. Because V is an isometry the Kraus operators satisfy the completeness relation $\sum_{k=1}^{N} {A_k}^\dagger A_k = \underline{1}_Q$.

Below, we will often have to refer to the number of Kraus operators of a channel $\mathcal{N}$ in a certain operator-sum representation, which, of course, equals the dimension $|E|$ of the ancilla E in the corresponding Stinespring representation. It is therefore convenient to define the length $|\mathcal{N}|$ of a channel $\mathcal{N}$ by the minimum number of Kraus operators in an operator-sum representation, or, equivalently, as the minimum dimension of an ancilla in a Stinespring representation needed to represent $\mathcal{N}$.

According to the above definition a quantum channel maps density operators to density operators, and therefore should be trace-preserving. As a matter of fact, it is sometimes advantageous to be less restrictive and to consider also trace-decreasing

[2] Since we assumed the dimensions $|Q|$ and $|Q'|$ to be finite also the ancilla E can be chosen to be of finite dimension $|E| = N$.

channels. Being still a completely positive mapping, an in general trace-decreasing channel $\mathcal{N} : \mathcal{B}(Q) \to \mathcal{B}(Q')$ has a Stinespring representation with an operator $V : Q \to Q'E$ satisfying $V^\dagger V \leq \underline{1}_Q$. As a consequence, corresponding Kraus operators $A_1, \ldots A_N$ of $\mathcal{N}$ may be incomplete, meaning that $\sum_{k=1}^N {A_k}^\dagger A \leq \underline{1}_Q$. Physically, a trace-decreasing channel describes a transmission that involves either some selective process or some leakage, as an effect of which a system does not necessarily reach its destination. This motivates us to denote $\mathrm{tr}\mathcal{N}(\rho)$ as the transmission probability of state ρ with respect to $\mathcal{N}$.

1.2 Fidelities

A frequently used quantity for measuring the distance of general quantum states is the fidelity [5, 6, 1]

$$F(\rho, \sigma) := \| \sqrt{\rho}\sqrt{\sigma} \|_{tr}^2 ,$$

where $\| \ldots \|_{tr}$ denotes the trace norm, $\| A \|_{tr} = \mathrm{tr}\sqrt{A^\dagger A}$. If one of the states is pure, say $\rho = \psi = |\psi\rangle\langle\psi|$, this reduces to

$$F(\psi, \sigma) = \langle\psi|\sigma|\psi\rangle .$$

Generally, $0 \leq F(\rho, \sigma) \leq 1$, and $F(\rho, \sigma) = 1$ if and only if $\rho = \sigma$. The fidelity of two states is related to their trace norm distance by [1]

$$1 - \| \rho - \sigma \|_{tr} \leq F(\rho, \sigma) \leq 1 - \frac{1}{4} \| \rho - \sigma \|_{tr}^2 .$$

Furthermore, the fidelity is monotonic under quantum operations in the sense that for any trace-preserving completely positive $\mathcal{E} : \mathcal{B}(Q) \to B(Q')$,

$$F(\rho, \sigma) \leq F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) .$$

A remarkably theorem by Uhlmann [5] states that the fidelity of ρ and σ can be also understood as the maximum transmission probability $|\langle\psi|\varphi\rangle|^2$ of purifications ψ and φ for ρ and σ, respectively. The fidelity $F(\rho, \sigma)$ thus tells us how close two pure states ψ and φ of a universe can be if they are known to reduce to states ρ and σ on a subsystem Q. More precisely, the theorem states that if ψ_{RQ} in RQ is a purification of ρ, and if σ can be also purified on RQ, then

$$F(\rho, \sigma) := \max_{\varphi_{RQ}} |\langle\psi_{RQ}|\varphi_{RQ}\rangle|^2 ,$$

where the maximum is taken over all purifications φ_{RQ} of σ in RQ [6].

To determine how well a state ρ is preserved under a channel $\mathcal{E} : \mathcal{B}(Q) \to \mathcal{B}(Q')$ we will generally use the entanglement fidelity [7]

$$F_e(\rho, \mathcal{E}) := \langle\psi_{RQ}|\mathcal{I}_R \otimes \mathcal{E}(\psi_{RQ})|\psi_{RQ}\rangle , \tag{1}$$

where ψ_{RQ} is any purification of ρ on Q extended by an ancilla system R, and $\mathcal{I}_R$ is the identity operation on R. In terms of Kraus operators $A_1, \ldots, A_{|\mathcal{E}|}$ of $\mathcal{E}$ the entanglement fidelity can be expressed as [7]

$$F_e(\rho, \mathcal{E}) = \sum_{k=1}^{|\mathcal{E}|} |\mathrm{tr}\, \rho A_k|^2 \,. \tag{2}$$

The entanglement fidelity of a state $\rho = \sum_i p_i \psi_i$ is known to be a lower bound of the averaged fidelities $F(\psi_i, \mathcal{E}(\psi_i))$ [1],

$$F_e(\rho, \mathcal{E}) \;\leq\; \sum_i p_i F(\psi_i, \mathcal{E}(\psi_i)) \,.$$

This relation becomes particularly useful if ρ is chosen to be the normalized projection π_C on a subspace C of Q, $\pi_C = \Pi_C/|C|$. Then the entanglement fidelity yields a lower bound of the average subspace fidelity,

$$F_e(\pi_C, \mathcal{E}) \;\leq\; \int_C d\psi \; F(\psi, \mathcal{E}(\psi)) =: F_{av}(C, \mathcal{E}) \,. \tag{3}$$

where the integral is taken with respect to the normalized, unitarily invariant measure on C. Actually, there also exists a strict relation between the two fidelities [8, 9],

$$F_{av}(C, \mathcal{E}) = \frac{|C|\, F_e(\pi_C, \mathcal{E}) + 1}{|C| + 1} \,.$$

We emphasize that with Eq. (1) also the entanglement fidelity with respect to a trace-decreasing channel $\mathcal{E}$ is defined. In this case representation (2) turns out to hold as well, leading to the following simple but nevertheless useful observation. Let a channel $\mathcal{E} : \mathcal{B}(Q) \to \mathcal{B}(Q')$ be defined by Kraus operators $A_1, \ldots, A_{|\mathcal{E}|}$. We call a second channel $\tilde{\mathcal{E}} : \mathcal{B}(Q) \to \mathcal{B}(Q')$ a reduction of $\mathcal{E}$ if it can be represented by a subset of the Kraus operators $A_1, \ldots, A_{|\mathcal{E}|}$, i.e.

$$\tilde{\mathcal{E}}(\rho) = \sum_{k \in \tilde{N}} A_k \rho {A_k}^\dagger \,, \quad \tilde{N} \subset \{1, \ldots, |\mathcal{E}|\} \,.$$

By Eq. (2) we notice that reducing a channel can never increase entanglement fidelity: for any reduction $\tilde{\mathcal{E}}$ of a channel $\mathcal{E}$

$$F(\rho, \tilde{\mathcal{E}}) \;\leq\; F(\rho, \mathcal{E}) \,. \tag{4}$$

2 Quantum coding theorem

2.1 Quantum capacity of a quantum channel

For the purpose of quantum-information transmission, Alice (sender) and Bob (receiver) may employ a quantum channel $\mathcal{N} : \mathcal{B}(Q) \to \mathcal{B}(Q')$ that conveys an input

quantum system Q from Alice to an in general different output system Q' received by Bob.

In the simplest case, Alice may prepare quantum information in form of some state ρ of Q, which after transmission via the channel becomes a state $\rho' = \mathcal{N}(\rho)$ of Q' received by Bob. In order to obtain Alice's originally sent state ρ, Bob may subject ρ' to suited physical manipulations, which eventually should result in a state ρ'' of Q close to ρ. Mathematically, this corresponds to the application of a trace-preserving, completely positive mapping $\mathcal{R} : \mathcal{B}(Q') \to \mathcal{B}(Q)$, which we denote as recovery operation in the following. Referring to Sec. 1.2, relation (3), the overall performance of this elementary transmission scheme can be conveniently assessed by the entanglement fidelity $F_e(\pi, \mathcal{R} \circ \mathcal{N})$ of the homogeneous density $\pi = \underline{1}_Q/|Q|$ of Q with respect to $\mathcal{R} \circ \mathcal{N}$, or, if we suppose that Bob has optimized the recovery operation $\mathcal{R}$, by the maximized entanglement fidelity

$$\max_R F_e(\pi, \mathcal{R} \circ \mathcal{N}) \,. \tag{5}$$

To improve the transmission scheme, Alice and Bob may agree upon using only states ρ whose supports lie in a certain linear subspace C of Q [3]. A subspace used for this purpose is called a (quantum) code. Its size k is defined as $k = \log_2 |C|$, meaning that a pure state in C carries k qubits of quantum information [12]. Corresponding to (5), an appropriate quantity for assessing the suitability of a code C for a channel $\mathcal{N}$ is the quantity

$$F_e(C, \mathcal{N}) := \max_{\mathcal{R}} F_e(\pi_C, \mathcal{R} \circ \mathcal{N}) \,,$$

where $\pi_C = \Pi_C/|C|$ is the normalized projection on C (again cf. Sec. 1.2). We refer to this quantity as the entanglement fidelity of the code C with respect to the channel $\mathcal{N}$.

The definition involves a non-trivial optimization of the recovery operation $\mathcal{R}$. At first sight, this makes the code entanglement fidelity rather difficult to determine and therefore may cast doubts on its usefulness. However, following Schumacher and Westmoreland [13] we will derive a useful explicit lower bound for $F_e(C, \mathcal{N})$ in Sec. (4).

In the elementary transmission scheme considered so far the quantum information is encoded in single quantum systems Q and transmitted in single uses ("shots") of the channel $\mathcal{N}$. Like in classical communication schemes the restriction to single-shot uses of the channel is very often far from being optimal. Since the work of Shannon [14] it is known that encoding and transmission of information in large blocks yields much better results.

[3] This can be advantageous when the interaction of system and environment does affect states in C significantly less than the average state, for instance, because C obeys certain symmetries of the system-environment interaction Hamiltonian. Moreover, the restriction to a suited subspace C may allow Bob to employ quantum error-correcting schemes in the recovery operation $\mathcal{R}$ [10, 11].

In an n-block transmission scheme, Alice uses n identical copies of the quantum system Q, in which she encodes quantum information as a state ρ with support in a chosen code $C_n \subset Q^n$. During the transmission each individual system Q is independently transformed by the channel $\mathcal{N}$, and Bob receives the state $\mathcal{N}^{\otimes n}(\rho)$, on which he applies a recovery operation $\mathcal{R}_n : \mathcal{B}(Q^n) \to \mathcal{B}(Q'^n)$. The crucial differences to a single-shot scheme are the usage of a code C_n and a recovery operation $\mathcal{R}_n$ which in general will not obey the tensor product structure, i.e. $C_n \neq C_1^{\otimes n}$ and $\mathcal{R}_n \neq \mathcal{R}_1^{\otimes n}$. The rate $R = \frac{1}{n}\log_2 |C_n|$ of an n-block code $C_n \subset Q^n$ denotes the average number of qubits encoded per system Q and sent per channel use.

In the end, we wish to know up to which rate the channel $\mathcal{N}$ can reliably transmit quantum information when an optimal block code C_n of arbitrarily large block number n is used. This rate defines the quantum capacity $Q(\mathcal{N})$ of the channel $\mathcal{N}$ [15, 16, 17] (for a recent review see e.g. [18]). A mathematically precise definition uses the notion of an achievable rate. A rate R is called achievable by the channel $\mathcal{N}$ if there is a sequence of codes $C_n \subset Q^n$, $n = 1, 2, \ldots$, such that

$$\lim_{n\to\infty} \sup \frac{\log_2 |C_n|}{n} \geq R\,, \quad \text{and} \quad \lim_{n\to\infty} F_e(C_n, \mathcal{N}^{\otimes n}) = 1\,.$$

The supremum of all achievable rates of a channel $\mathcal{N}$ is the quantum capacity $Q(\mathcal{N})$ of the channel $\mathcal{N}$.

2.2 Quantum coding theorem

Determining the quantum capacity of a channel $\mathcal{N}$ poses one of the central problems of quantum information theory. It is partially solved by the quantum coding theorem [15, 16, 17] which relates quantum capacity to coherent information [19], the quantum analogue to mutual information in classical information theory. The coherent information is defined for a state ρ with respect to a trace-preserving channel $\mathcal{N}$ as

$$I(\rho, \mathcal{N}) \;=\; S(\mathcal{N}(\rho)) - S_e(\rho, \mathcal{N})\,.$$

This is the von Neuman entropy of the channel output, $S(\mathcal{N}(\rho))$, minus the entropy exchange $S_e(\rho, \mathcal{N})$ between system and environment, which is given by

$$S_e(\rho, \mathcal{N}) = S(\mathcal{I}_R \otimes \mathcal{N}(\psi_{RQ}))\,,$$

where ψ_{RQ} is a purification of ρ, and $\mathcal{I}_R$ is the identity operation on the ancilla system R [7].

The quantum noisy coding theorem states that the quantum capacity $Q(\mathcal{N})$ of a channel $\mathcal{N}$ is the regularized coherent information $I_r(\mathcal{N})$ of $\mathcal{N}$,

$$Q(\mathcal{N}) = I_r(\mathcal{N}) := \lim_{n\to\infty} \frac{1}{n} \max_{\rho} I(\rho, \mathcal{N}^{\otimes n})\,.$$

The limiting procedure corresponds to the one in the definition of an achievable rate and thus contributes to the fact that generally optimal coding can be only asymptotically reached in the limit of block numbers $n \to \infty$. As a consequence of this limit the regularized coherent information and thus the quantum capacity of a channel is still difficult to determine.

The regularized coherent information has long been known an upper bound for $Q(\mathcal{N})$, which is the content of the converse coding theorem [16, 17]. The direct coding theorem, stating that $I_r(\mathcal{N})$ is actually attainable, has been strictly proven first by Devetak [20]. His proof utilizes a correspondence of classical private information and quantum information.

Sections 4, 5, 7, and 8 below represent the four stages of a different proof for the direct quantum coding theorem, of which an earlier version appeared in Ref. [21]. The working hypothesis underlying this proof is that randomly chosen block codes of sufficiently large block number typically allow for almost perfect quantum error correction. In this respect, the present proof as well as the one of Hayden *et al.* [22] and also the earlier approaches of Shor [23] and Lloyd [15] follow Shannon's original treatment [14] of the classical coding problem.

3 Outline of proof

In the first stage of the the proof (Sec. 4) we establish a lower bound for the code entanglement fidelity. It is essentially an earlier result of Schumacher and Westmoreland [13], of which has been also made good use of recently by Abeysinghe *et al.* [24] and Hayden *et al.* [22] in the same context. The bound can be explicitly determined in terms of Kraus operators of the channel $\mathcal{N}$, and its use will relieve us from the burden of optimizing a recovery operation $\mathcal{R}$ for a given code C and channel $\mathcal{N}$ in the course of proving the coding theorem. In deriving the lower bound the optimization of $\mathcal{R}$ is solved by means of Uhlmann's theorem.

In the next stage (Sec. 5) we investigate the error correcting ability of codes that are chosen at random from a unitarily invariant ensemble of codes with a given dimension K. Taking the average of the lower bound derived in Sec. 4 we will show the averaged code entanglement fidelity of a channel $\mathcal{N} : \mathcal{B}(Q) \to B(Q')$ to obey

$$[F_e(C,\mathcal{N})]_K \;\geq\; \mathrm{tr}\,\mathcal{N}(\pi) \;-\; \sqrt{K|\mathcal{N}|}\,\|\mathcal{N}(\pi)\|_F\,, \tag{6}$$

where $\pi = \underline{1}_Q/|Q|$, and $\|\ldots\|_F$ denotes the Frobenius norm or two norm.

In Sec. 6 we will illustrate the efficiency of random coding by means of the special case of a unital channel $\mathcal{U} : \mathcal{B}(Q) \to \mathcal{B}(Q')$, which by definition satisfies $\mathcal{U}(\pi) = \pi'$. In this case the lower bound (6) immediately proves the attainability of the quantum Hamming bound by random coding, and thus provides evidence for the validity of the above mentioned working hypothesis. Moreover, if we demand the channel $\mathcal{U}$ to be also uniform, as will be defined in Sec. 6, we can easily establish the coherent

information $I(\pi,\mathcal{U})$ to be a lower bound of the quantum capacity $Q(\mathcal{U})$,

$$Q(\mathcal{U}) \;\geq\; I(\pi,\mathcal{U})\,.$$

The third stage of the proof (Sec. 7) is merely the generalization of this relation to an arbitrary channel $\mathcal{N}:\mathcal{B}(Q)\to\mathcal{B}(Q')$. To this end we have to consider n-block transmission schemes. For large n it is possible to arrange for unitality and uniformity of $\mathcal{N}^{\otimes n}$ in an approximate sense by, as it will turn out, only minor modifications of $\mathcal{N}^{\otimes n}$. Approximate uniformity is achieved by reducing the operation $\mathcal{N}^{\otimes n}$ to an operation $\mathcal{N}_{\varepsilon,n}$ consisting only of typical Kraus operators. Furthermore, letting $\mathcal{N}_{\varepsilon,n}$ follow a projection on the typical subspace of $\mathcal{N}(\pi)$ in Q'^n establishes an approximatively uniform and unital channel $\tilde{\mathcal{N}}_{\varepsilon,n}$, which nevertheless is close to the original $\mathcal{N}^{\otimes n}$. In the end, this suffices to prove $Q(\mathcal{N})\geq I(\pi,\mathcal{N})$ for a general channel $\mathcal{N}$. A corollary is that for any subspace $V\subset Q$ with normalized projection $\pi_V=\Pi_V/|V|$

$$Q(\mathcal{N}) \;\geq\; I(\pi_V,\mathcal{N})\,.$$

Finally, in Sec. 8 we employ a lemma of Bennett, Shor, Smolin, and Thapliyal (BSST) [25] in order to deduce from the last relation

$$Q(\mathcal{N})\geq\frac{1}{m}I(\rho,\mathcal{N}^{\otimes m})$$

for an arbitrary integer m, and any density ρ of Q^n. This shows the regularized coherent information to be a lower bound of $Q(\mathcal{N})$ and thus concludes the proof of the direct coding theorem.

4 A lower bound for the code entanglement fidelity

Let a (possibly trace-decreasing) quantum channel $\mathcal{N}:\mathcal{B}(Q)\to\mathcal{B}(Q')$ have a Stinespring representation with an operator $V:Q\to Q'E$, and let $C\subset Q$ be a code whose normalized projection $\pi_C=\Pi_C/|C|$ may have a purification ψ_{RQ} on RQ, with R being an appropriate ancilla system. Following Schumacher and Westmoreland we will establish

$$F_e(C,\mathcal{N}) \;\geq\; p - p\,\|\,\rho'_{RE}-\rho_R\otimes\rho'_E\,\|_{tr}\,, \tag{7}$$

where $p=\mathrm{tr}\mathcal{N}(\pi_C)$, $\rho_R=\mathrm{tr}_Q\,\psi_{RQ}$, and the states ρ'_{RE} and ρ'_E are reduced density operators of the final normalized pure state

$$\psi'_{RQ'E}=\frac{1}{p}(\underline{1}_R\otimes V)\psi_{RQ}(\underline{1}_R\otimes V^\dagger)\,, \tag{8}$$

$$\rho'_{RE}=\mathrm{tr}_{Q'}\,\psi_{RQ'E}\,,\quad \rho'_E=\mathrm{tr}_{RQ'}\,\psi_{RQ'E}\,.$$

Furthermore, we will show show that the lower bound (7) can alternatively be formulated in terms of Kraus operators $A_1,\ldots,A_N$ of $\mathcal{N}$ as

$$F_e(C,\mathcal{N}) \;\geq\; p - \|\,D\,\|_{tr}\,, \tag{9}$$

where

$$D = |C| \sum_{ij=1}^{N} \Big(\pi_C A_i^\dagger A_j \pi_C \; - \; \mathrm{tr}(\pi_C A_i^\dagger A_j \pi_C)\, \pi_C \Big) \otimes |i\rangle\langle j| \,, \tag{10}$$

with $|1\rangle, \ldots, |N\rangle$ being orthonormal states of some ancilla system.

Proof of relation (7): We recall that the code entanglement fidelity involves a non-trivial optimization procedure of a recovery operation $\mathcal{R}$ (cf. Sec. 2.1). The idea is to hand over this job to Uhlmann's theorem. To this end we consider the pure state

$$\tilde{\psi} := \psi_{RQ} \otimes \psi'_{RQ'E}$$

of the joint system $RSQ'E$, where S denotes a copy of QR. Obviously, $\tilde{\psi}$ is a purification of the state $\rho_R \otimes \rho'_E$ with respect to the ancilla SQ'. Next, we extend $\psi'_{RQ'E}$ by the operation

$$\mathcal{E} : \mathcal{B}(Q') \to B(SQ') \,,\; \rho \mapsto \psi_S \otimes \rho \,,$$

where ψ_S is any fixed pure state of S, to a pure state

$$\psi' := \mathcal{I}_R \otimes \mathcal{E} \otimes \mathcal{I}_E(\psi'_{RQ'E})$$

of $RSQ'E$. ψ' is a purification of ρ'_{RE} with respect to SQ', since

$$\mathrm{tr}_{SQ'}\psi' = \mathrm{tr}_{Q'}\mathrm{tr}_S\psi' = \mathrm{tr}_{Q'}\psi'_{RQ'E} = \rho'_{RE} \,.$$

Now, let another purification φ of ρ'_{RE} in $RSQ'E$ maximize the transition amplitude to $\tilde{\psi}$,

$$|\langle\tilde{\psi}|\varphi\rangle|^2 = \max_{\chi \text{ purification of } \rho'_{RE}} |\langle\tilde{\psi}|\chi\rangle|^2 \,.$$

According to Uhlmann's theorem (cf. Sec. 1.2) we know that

$$|\langle\tilde{\psi}|\varphi\rangle|^2 = F(\rho_R \otimes \rho'_E, \rho'_{RE}) \,. \tag{11}$$

Then, an optimal recovery operation $\mathcal{R} : \mathcal{B}(Q') \to \mathcal{B}(Q)$ can be constructed my means of a unitary operation $U_{SQ'}$ on SQ' that rotates the actual (extended) final state ψ' to the maximizing state φ,

$$\varphi = (\underline{1}_R \otimes U_{SQ'} \otimes \underline{1}_E)\psi'(\underline{1}_R \otimes {U_{SQ'}}^\dagger \otimes \underline{1}_E) \,.$$

Keeping in mind that $S = QR$ we define

$$\mathcal{R}(\rho_{Q'}) := \mathrm{tr}_{RQ'} U_{SQ'} \mathcal{E}(\rho_{Q'}) {U_{SQ'}}^\dagger \,,$$

and realize that for the state $\rho_{RQ'} = \mathrm{tr}_E \psi'_{RQ'E}$

$$\begin{aligned} \mathcal{I}_R \otimes \mathcal{R}(\rho'_{RQ'}) &= \mathrm{tr}_{RQ'} \,(\underline{1}_R \otimes U_{SQ'}) \mathcal{I}_R \otimes \mathcal{E}(\rho'_{RQ'})(\underline{1}_R \otimes {U_{SQ'}}^\dagger) \\ &= \mathrm{tr}_{RQ'E} \,(\underline{1}_R \otimes U_{SQ'} \otimes \underline{1}_E)\psi'(\underline{1}_R \otimes {U_{SQ'}}^\dagger \otimes \underline{1}_E) \\ &= \mathrm{tr}_{RQ'E} \,\varphi \,, \end{aligned}$$

where here and in the following the partial trace over R refers to the second R appearing in the product Hilbert space $RSQ'E = RQRQ'E$. Since further

$$\psi_{RQ} = \mathrm{tr}_{RQ'E}\tilde{\psi}\,,$$

we conclude

$$\begin{aligned} F_e(\pi_C, \mathcal{R}\circ\mathcal{N}) &\geq p\,F(\psi_{RQ}, \mathcal{I}_R\otimes\mathcal{R}(\rho'_{RQ'})) \\ &= p\,F(\mathrm{tr}_{RQ'E}\,\tilde{\psi}, \mathrm{tr}_{RQ'E}\,\varphi) \\ &\geq p|\langle\tilde{\psi}|\varphi\rangle|^2\,, \end{aligned}$$

where the second inequality is due to the monotonicity of the fidelity under partial trace. With Eq. (11) and the general relation $F(\rho,\sigma) \geq 1 - \parallel \rho-\sigma \parallel_{tr}$ this proves relation (7).

Proof of relation (9): We choose a purification ψ_{RQ} of π_C with a state vector

$$|\psi\rangle_{RQ} = \frac{1}{\sqrt{K}}\sum_{l=1}^{K}|c_l^R\rangle|c_l^Q\rangle\,,$$

where $K = |C|$, and $|c_1^R\rangle, \ldots |c_K^R\rangle$ and $|c_1^Q\rangle, \ldots |c_K^Q\rangle$ denote orthonormal vectors that span R and C, respectively. Supposing that the orthonormal states $|1\rangle, \ldots, |N\rangle$ span the ancilla E and the Kraus operators $A_1, \ldots A_N$ are associated to V by $A_i|\psi\rangle_Q = \langle i|V|\psi_Q\rangle$, we immediately obtain from Eq. (8)

$$p\,\psi'_{RQ'E} = \frac{1}{K}\sum_{lm=1}^{K}\sum_{ij=1}^{N}|c_l^R\rangle\langle c_m^R|\otimes A_i|c_l^Q\rangle\langle c_m^Q|A_j{}^\dagger\otimes|i\rangle\langle j|\,.$$

Hence

$$\begin{aligned} p\,\rho'_{RE} &= \frac{1}{K}\sum_{lm=1}^{K}\sum_{ij=1}^{N}\langle c_m^Q|A_j{}^\dagger A_i|c_l^Q\rangle\,|c_l^R\rangle\langle c_m^R|\otimes|i\rangle\langle j| \\ p\,\rho_R\otimes\rho'_E &= \frac{1}{K^2}\sum_{m=1}^{K}|c_m^R\rangle\langle c_m^R|\otimes\sum_{l=1}^{K}\sum_{ij=1}^{N}\langle c_l^Q|A_j{}^\dagger A_i|c_l^Q\rangle\,|i\rangle\langle j|\,. \end{aligned}$$

The trace norm of $p(\rho'_{RE}-\rho_R\otimes\rho'_E)$ appearing in the lower bound (7) becomes more handy if we transform the operator difference by an isometry $\mathcal{J}:\mathcal{B}(RE)\to\mathcal{B}(QE)$,

$$\mathcal{J}:\sum_{lm,ij}\alpha_{lm,ij}|c_l^R\rangle\langle c_m^R|\otimes|i\rangle\langle j|\;\mapsto\;\sum_{lm,ij}\alpha^*_{lm,ij}|c_l^Q\rangle\langle c_m^Q|\otimes|i\rangle\langle j|\,.$$

$\mathcal{J}$ shifts from R to Q and then complex conjugates with respect to the basis $|c_l^Q\rangle\otimes|i\rangle$, which clearly leaves the trace norm invariant. A straightforward calculation then shows

$$D := p\mathcal{J}(\rho'_{RE}-\rho_R\otimes\rho'_E) \;=\; K\sum_{ij=1}^{N}\Big(\pi_C A_i{}^\dagger A_j\pi_C \;-\; \mathrm{tr}(\pi_C A_i{}^\dagger A_j\pi_C)\,\pi_C\Big)\otimes|i\rangle\langle j|\,,$$

as in Eq. (10), and further

$$F_e(\mathcal{C},\mathcal{N}) \;\geq\; p \;-p\,\|\,\rho'_{RE} - \rho_R \otimes \rho'_E\,\|_{tr} = \; p - \|\,\mathcal{J}(\rho'_{RE} - \rho_R \otimes \rho'_E)\,\|_{tr} = \; p - \|\,D\,\|_{tr} \;\; .$$

which is what we wanted to proof.

5 Random coding

Let the unitarily invariant code ensemble of all K-dimensional codes $C \subset Q$ be defined by the ensemble average

$$[A(C)]_K \;:=\; \int_{\underline{\mathrm{U}}(Q)} d\mu(U)\; A(UC_0) \tag{12}$$

of a code dependent variable $A(C)$. Here, C_0 is some fixed K-dimensional code space in Q, and μ is the normalized Haar measure on $\underline{\mathrm{U}}(Q)$, the group of all unitaries on Q. Below we will show that the ensemble averaged code entanglement fidelity of a (possibly trace-decreasing) channel $\mathcal{N} : \mathcal{B}(Q) \to \mathcal{B}(Q')$ obeys

$$[F_e(C,\mathcal{N})]_K \;\geq\; \mathrm{tr}\,\mathcal{N}(\pi) \;-\; \sqrt{K|\mathcal{N}|}\;\|\,\mathcal{N}(\pi)\,\|_F \;, \tag{13}$$

where $\pi = \underline{1}_Q/|Q|$ is the uniform density on Q.

We begin with the ensemble average of relation (9),

$$[F_e(C,\mathcal{N})]_K \;\geq\; [\mathrm{tr}\mathcal{N}(\pi_C)]_K - [\|\,D\,\|_{tr}]_K \;\; , \tag{14}$$

where, as always, $\pi_C = \Pi_C/|C|$, and

$$D = K\sum_{i,j=1}^{N} \Big(\pi_C A_i^\dagger A_j \pi_C \;-\; \mathrm{tr}(\pi_C A_i^\dagger A_j \pi_C)\,\pi_C\Big) \otimes |i\rangle\langle j| \;, \tag{15}$$

with $A_1, \ldots, A_N$ being $N = |\mathcal{N}|$ Kraus operators of a minimal operator-sum representation of $\mathcal{N}$. To average $\mathrm{tr}\mathcal{N}(\pi_C)$ we realize that $\rho \mapsto \mathrm{tr}\mathcal{N}(\rho)$ as a linear operation interchanges with the average. Since $[\pi_C]_K = \pi$ we thus obtain

$$[\mathrm{tr}\mathcal{N}(\pi_C)]_K = \mathrm{tr}\mathcal{N}([\pi_C]_K) = \mathrm{tr}\mathcal{N}(\pi)\;.$$

Directly averaging the trace norm of D turns out to be quite cumbersome. Therefore, we first estimate

$$[\|\,D\,\|_{tr}]_K^2 \;\leq\; KN\,[\|\,D\,\|_F]_K^2 \;\leq\; KN\left[\|\,D\,\|_F^2\right]_K \;\; ,$$

where $\|\,D\,\|_F = \;(\mathrm{tr}\;D^\dagger D)^{1/2}$ denotes the Frobenius norm (two-norm) of D. The first inequality follows from the general relation $\|\,A\,\|_{tr} \leq \sqrt{d}\;\|\,A\,\|_F$, where d is the rank of A, and the second inequality is Jensen's inequality. This leads us to

$$[F_e(C,\mathcal{N})]_K \;\geq\; \mathrm{tr}\mathcal{N}(\pi) \;-\; \sqrt{KN\;\left[\|\,D\,\|_F^2\right]_K}\;, \tag{16}$$

and it remains to determine the ensemble average of $\| D \|_F^2$. From the explicit representation Eq. (15) follows

$$\| D \|_F^2 = \text{tr } D^\dagger D \; = \; \sum_{ij=1}^{N} \text{tr}(\pi_C {W_{ij}}^\dagger \pi_C W_{ij}) - \frac{1}{K} |\text{tr } \pi_C W_{ij}|^2 \, ,$$

where operators W_{ij} are

$$W_{ij} = {A_i}^\dagger A_j \, .$$

It is useful to introduce a Hermitian form

$$b(V,W) \; := \; \Big[\text{tr}(\pi_C V^\dagger \pi_C W) \; - \; \frac{1}{K} \text{tr}(\pi_C V^\dagger) \, \text{tr}(\pi_C W) \Big]_K \, , \tag{17}$$

with which

$$\Big[\| D \|_F^2 \Big]_K \; = \; \sum_{ij=1}^{N} b(W_{ij}, W_{ij}) \, . \tag{18}$$

The point is that the unitary invariance of the ensemble average entails the unitary invariance of b, i.e., for any $U \in \mathrm{U}(Q)$

$$b(V,W) \; = \; b(UVU^\dagger, UWU^\dagger) \, .$$

which, in fact, already determines b to a large extend: According to Weyl's theory of group invariants [26, 27] $b(V,W)$ must be a linear combination of the only two fundamental unitarily invariant Hermitian forms $\text{tr } V^\dagger W$ and $\text{tr } V^\dagger \, \text{tr } W$,

$$b(V,W) \; = \; \alpha \, \text{tr } V^\dagger W \; + \beta \, \text{tr } V^\dagger \, \text{tr } W \, . \tag{19}$$

An elementary proof of this fact is outlined in Appendix A. To determine the coefficients α and β we consider two special choices of the operators V and W. For $V = W = 1\!\!1_Q$ Eqs. (17) and (19) yield

$$\alpha M + \beta M^2 = \frac{1}{K} \, , \tag{20}$$

where here and henceforth $M = |Q|$. Secondly, when we set V and W to a projection $\psi = |\psi\rangle\langle\psi|$ on Q we obtain from Eq. (17)

$$b(\psi,\psi) = \frac{K-1}{K} \Big[|\langle\psi|\pi_C|\psi\rangle|^2 \Big]_K \, .$$

Reverting to random matrix theory we find in Appendix (B) $\big[|\langle\psi|\pi_C|\psi\rangle|^2 \big]_K = (1 + 1/K)/(M^2 + M)$, and hence

$$b(\psi,\psi) = \frac{1 - K^{-2}}{M^2 + M} \, .$$

With $b(\psi,\psi) = \alpha + \beta$ from Eq. (19) this yields the second equation,

$$\alpha + \beta = \frac{1 - K^{-2}}{M^2 + M} \,. \tag{21}$$

Solving Eqs. (20) and (21) for α and β, and inserting the solution into (19) produces

$$b(V,W) = \frac{1 - K^{-2}}{M^2 - 1} \left(\mathrm{tr}\, V^\dagger W \;-\; \frac{1}{M} \mathrm{tr}\, V^\dagger \,\mathrm{tr}\, W \right) \,,$$

and, by Eq. (18),

$$\left[\| D \|_F^2 \right]_K = \frac{1 - K^{-2}}{M^2 - 1} \sum_{ij} \left(\mathrm{tr}\, {W_{ij}}^\dagger W_{ij} - \frac{1}{M} |\mathrm{tr}\, W_{ij}|^2 \right) \,. \tag{22}$$

In general, not much is given away when we use the upper bound for $\left[\| D \|_F^2 \right]_K$ that we obtain by using $(1 - 1/K^2)/(M^2 - 1) \leq 1/M^2$ and by omitting the negative terms $-|\mathrm{tr}\, W_{ij}|^2/M$ in the sum. Then

$$\left[\| D \|_F^2 \right]_K \;\leq\; \frac{1}{M^2} \sum_{ij} \mathrm{tr}\, {W_{ij}}^\dagger W_{ij} = \mathrm{tr}(\sum_j A_j \frac{\underline{1}_Q}{M} {A_j}^\dagger \sum_i A_i \frac{\underline{1}_Q}{M} {A_i}^\dagger) \,,$$

where we cyclically permuted operators under the trace to obtain the last equality. We realize that the argument of the trace is simply $\mathcal{N}(\pi)^2$ (with $\pi = \underline{1}_Q/M$). This yields the rather simple upper bound

$$\left[\| D \|_F^2 \right]_K \;\leq \| \mathcal{N}(\pi) \|_F^2 \,, \tag{23}$$

which finally proves the lower bound (13) by relation (16).

6 Unital and uniform channels

The efficiency of random coding can be easily demonstrated by relation (13) for the case of a unital channel $\mathcal{U} : \mathcal{B}(Q) \to \mathcal{B}(Q')$, which by definition maps the homogeneously distributed input state π to the homogeneously distributed output state π'. An example is a random unitary channel $\mathcal{U}_r : \mathcal{B}(Q) \to \mathcal{B}(Q)\,, \rho \mapsto \sum_i p_i U_i \rho {U_i}^\dagger$, where arbitrary unitary operators $U_1, \ldots, U_N$ are applied with probabilities $p_1, \ldots, p_N$ on the system Q.

Thus, for a unital channel $\| \mathcal{U}(\pi) \|_F = \| \pi' \|_F = |Q'|^{-1/2}$, which by relation (13) predicts the average entanglement fidelity of K-dimensional codes to obey

$$[F_e(C,\mathcal{U})]_K \;\geq\; 1 - \sqrt{\frac{K|\mathcal{U}|}{|Q'|}} \,.$$

This means that almost all codes of dimension K allow for almost perfect correction of the unital noise $\mathcal{U}$, provided that

$$K|\mathcal{U}| \ll |Q'| \, .$$

Recalling that $|\mathcal{U}|$ is the number of Kraus operators in an operator-sum representation of $\mathcal{U}$, this relation clearly shows the attainability of the quantum Hamming bound [28] by random coding. Formally, this is equivalent to the lower bound

$$Q(\mathcal{U}) \; \geq \; \log_2 |Q'| - \log_2 |\mathcal{U}| \tag{24}$$

of the quantum information capacity of $\mathcal{U}$. To see this, we consider the n-times replicated noise $\mathcal{U}^{\otimes n}$, and study the averaged entanglement fidelity of codes with dimension $K_n = \lfloor 2^{nR} \rfloor$ for some positive rate R. Since with $\mathcal{U}$ also $\mathcal{U}^{\otimes n}$ unital, and $|\mathcal{U}^{\otimes n}| = |\mathcal{U}|^n$, this time we arrive at

$$[F_e(C, \mathcal{U}^{\otimes n})]_{K_n} \; \geq \; 1 - \left(\frac{2^R \, |\mathcal{U}|}{|Q'|} \right)^{n/2} \, .$$

For $n \to \infty$ the right hand side converges to unity if $R < \log_2 |Q'| - \log_2 |\mathcal{U}|$. Hence, all rates below $\log_2 |Q'| - \log_2 |\mathcal{U}|$ are achievable, which by the definition of quantum capacity (cf. Sec. 2.1) shows relation (24).

Finally, let us assume that the channel $\mathcal{U}$ is also uniform, meaning that $\mathcal{U}$ has a minimal operator-sum representation with Kraus operators $A_1, \ldots, A_{|\mathcal{U}|}$ obeying $\mathrm{tr} A_i^\dagger A_j = 0$ for $i \neq j$ and $\frac{1}{|Q|} \mathrm{tr} A_i^\dagger A_i = const. = |\mathcal{U}|^{-1}$. The first condition is actually no restriction, because a non-diagonal representation can always be transformed to a diagonal one[4] . The second condition demands that errors E_i associated with Kraus operators A_i appear with equal probability $p_i = 1/|\mathcal{U}|$. We observe that by Schumacher's representation [7] the entropy exchange of π under a uniform U is simply given by

$$S_e(\pi, \mathcal{U}) = S(\underline{1}_{|\mathcal{U}|}/|\mathcal{U}|) = \log_2 |\mathcal{U}| \, .$$

Since $\mathcal{U}$ is unital we also have

$$S(\mathcal{U}(\pi)) = S(\pi') = \log_2 |Q'| \, .$$

Comparing these expressions with relation (24) and recalling the definition of coherent information (cf. Sec. 2.2) establishes the lower bound

$$Q(\mathcal{U}) \geq I(\pi, \mathcal{U}) \, .$$

In fact, the following section we will show this bound to hold for general channels.

[4] For arbitrary operation elements $B_1, \ldots, B_N$ of $\mathcal{N}$, $N = |\mathcal{N}|$, let an $N \times N$ matrix H be defined by $H_{ij} := \mathrm{tr} B_i^\dagger B_j$. Since $H = H^\dagger$, there is a unitary matrix U such that $UHU^\dagger$ is diagonal. Because of the unitary freedom in the operator-sum representation [1], the operators $A_m := \sum_j U^\dagger_{jm} B_j$ equivalently represent $\mathcal{N}$. It is readily verified that $\mathrm{tr} A_l^\dagger A_m = 0$ for $l \neq m$.

7 General channels

Starting again with relation (13) we will proof for a general channel $\mathcal{N} : \mathcal{B}(Q) \to \mathcal{B}(Q')$

$$Q(\mathcal{N}) \;\geq\; I(\pi, \mathcal{N})\,, \tag{25}$$

where $\pi = \underline{1}_Q/|Q|$, and, as a corollary,

$$Q(\mathcal{N}) \;\geq\; I(\pi_V, \mathcal{N})\,, \tag{26}$$

where π_V is the normalized projection $\pi_V = \Pi_V/|V|$ on any subspace $V \subset Q$.

The strategy of proving is to approximate $\mathcal{N}^{\otimes n}$ by an almost uniform and unital channel $\tilde{\mathcal{N}}_{\varepsilon,n}$, with which we then proceed as in the preceding section. We construct $\tilde{\mathcal{N}}_{\varepsilon,n}$ in two steps. The first step is to reduce $\mathcal{N}^{\otimes n}$ to its typical Kraus operators, as will be defined below. This yields an almost uniform operation $\mathcal{N}_{\varepsilon,n}$. In a second step, we let $\mathcal{N}_{\varepsilon,n}$ follow a projection on the typical subspace of $\mathcal{N}(\pi)$ in Q'^n, resulting in an operation $\tilde{\mathcal{N}}_{\varepsilon,n}$ with the desired properties.

We begin with briefly recalling definitions and basic properties of both typical sequences [29] and typical subspaces [12, 1].

7.1 Typical sequences

Let X_1, X_2, X_3, ... be independent random variables with an identical probability distribution $\mathcal{P}$ over an alphabet $\aleph$. Let $H(\mathcal{P}) = -\sum_{a\in\aleph} \mathcal{P}(a) \log_2 \mathcal{P}(a)$ denote the Shannon entropy of $\mathcal{P}$, let n be a positive integer, and let ε be some positive number. A sequence $\underline{a} = (a_1, a_2, \ldots, a_n) \in \aleph^n$ is defined to be ε-typical if its probability of appearance $p_{\underline{a}} = \mathcal{P}(a_1)\mathcal{P}(a_2)\ldots\mathcal{P}(a_n)$ satisfies

$$2^{-n(H(\mathcal{P})+\varepsilon)} \;\leq\; p_{\underline{a}} \;\leq\; 2^{-n(H(\mathcal{P})-\varepsilon)}\,.$$

Let $\aleph_{\varepsilon,n}$ denote the set of all ε-typical sequences of length n.

Below we will make use of the following two well-known facts:

(i) The number $|\aleph_{\varepsilon,n}|$ of all ε-typical sequences of length n is less than $2^{n(H(\mathcal{P})+\varepsilon)}$.

(ii) The probability $P_{\varepsilon,n} = \sum_{\underline{a}\in\aleph_{\varepsilon,n}} p_{\underline{a}}$ of a random sequence of length n being ε-typical exceeds $1 - 2e^{-n\psi(\varepsilon)}$, where $\psi(\varepsilon)$ is a positive number independent of n.

Proofs can be found in Appendix C.

7.2 Typical subspaces

Let ρ be some density operator of a quantum system Q, let n be a positive integer, and let ε be a positive number. An eigenvector $|\underline{v}\rangle$ of $\rho^{\otimes n}$ is called ε-typical if its eigenvalue $p_{\underline{v}}$ satisfies

$$2^{-n(S(\rho)+\varepsilon)} \;\leq\; p_{\underline{v}} \;\leq\; 2^{-n(S(\rho)-\varepsilon)}\,.$$

The ε-typical subspace $T_{\varepsilon,n}$ of ρ in $Q^{\otimes n}$ is defined as the span of all ε-typical eigenvectors of $\rho^{\otimes n}$. We denote the projection on $T_{\varepsilon,n}$ by $\Pi_{\varepsilon,n}$.

Notice that typical eigenvectors correspond to typical sequences when an orthonormal eigen-system $|v_1\rangle, \ldots, |v_{|Q|}\rangle$ of ρ is chosen as alphabet $\aleph$, a sequence of length n over $\aleph$ is identified with an eigenvector $|\underline{v}\rangle = |v_{j_1}\rangle|v_{j_2}\rangle \ldots |v_{j_n}\rangle$ of $\rho^{\otimes n}$, and the probability $\mathcal{P}(|\underline{v}\rangle)$ of an eigenvector $|\underline{v}\rangle$ of ρ is taken to be its eigenvalue. Then, the above stated properties of typical sequences translate to

(i') The dimension of $T_{\varepsilon,n}$ is less than $2^{n(S(\rho)+\varepsilon)}$.

(ii') The probability $P_{\varepsilon,n} = \mathrm{tr}\, \Pi_{\varepsilon,n}\rho^{\otimes n}$ of measuring an ε-typical eigenvalue of $\rho^{\otimes n}$ exceeds $1 - 2e^{-n\psi(\varepsilon)}$, where $\psi(\varepsilon)$ is a positive number independent of n.

7.3 Reduction of $\mathcal{N}^{\otimes n}$

Let a trace-preserving channel $\mathcal{N} : \mathcal{B}(Q) \to \mathcal{B}(Q')$ be given. $\mathcal{N}$ may be represented in a minimal operator sum with Kraus operators $A_1, \ldots, A_{|\mathcal{N}|}$, which without loss of generality we assume to be diagonal, i.e. $\mathrm{tr} A_j{}^\dagger A_i = 0$ for $i \neq j$ (cf. footnote 4). Accordingly, $\mathcal{N}^{\otimes n}$ can be represented by $|\mathcal{N}|^n$ Kraus operators $A_{j_1} \otimes A_{j_2} \otimes \ldots \otimes A_{j_n}$ where $j_\nu = 1, \ldots, |\mathcal{N}|$.

Now, letting an alphabet $\aleph$ be defined as the set of Kraus operators $A_1, \ldots, A_{|\mathcal{N}|}$ of $\mathcal{N}$, the Kraus operators of $\mathcal{N}^{\otimes n}$ can obviously be regarded as sequences over $\aleph$ of length n. In order to identify an ε-typical sequence of length n, and with it also an ε-typical Kraus operator of $\mathcal{N}^{\otimes n}$, we define a probability distribution $\mathcal{P}$ over $\aleph$ by

$$\mathcal{P}(A) = \frac{1}{|Q|}\mathrm{tr}\, A^\dagger A\,, \qquad A \in \aleph\,.$$

The normalization of $\mathcal{P}$ follows from the completeness relation $\sum_{A\in\aleph} A^\dagger A = \underline{1}_Q$, and, owing to the diagonality of the Kraus operators, the Shannon entropy $H(\mathcal{P})$ turns out to agree with the entropy exchange $S_e(\pi, \mathcal{N})$: Again by Schumacher's representation [7],

$$S_e(\pi,\mathcal{N}) = S\left(\sum_{i=1}^{|\mathcal{N}|} \frac{1}{|Q|}\mathrm{tr}(A_i{}^\dagger A_i)|i\rangle\langle i|\right) = S\left(\sum_{i=1}^{|\mathcal{N}|} \mathcal{P}(A_i)|i\rangle\langle i|\right) = H(\mathcal{P})\,.$$

Being in the possession of the probability distribution $\mathcal{P}$ over the set of Kraus operators $\aleph$, we can define the ε-typical channel $\mathcal{N}_{\varepsilon,n}$ of $\mathcal{N}^{\otimes n}$ to consist precisely of the operators $\underline{A}$ that are ε-typical with respect to $\mathcal{P}$,

$$\rho \mapsto \mathcal{N}_{\varepsilon,n}(\rho) := \sum_{\underline{A}\in\aleph_{\varepsilon,n}} \underline{A}\rho\underline{A}^\dagger\,.$$

As a direct consequence of properties (i) and (ii) of typical sequences one finds (cf. Appendix D)

$$\begin{aligned} |\mathcal{N}_{\varepsilon,n}| &\leq 2^{n(S_e(\pi,\mathcal{N})+\varepsilon)} , \\ \operatorname{tr} \mathcal{N}_{\varepsilon,n}(\pi_n) &\geq 1 - 2e^{n\psi_1(\varepsilon)} , \end{aligned}$$

where $\pi_n = \underline{1}_{Q^n}/|Q|^n$, and $\psi_1(\varepsilon)$ is a positive number independent of n. Furthermore, the relative weight $\frac{1}{|Q|^n}\mathrm{tr}\underline{A}^\dagger\underline{A}$ of an ε-typical operator $\underline{A} = A_{j_1} \otimes \ldots \otimes A_{j_n}$ is just the probability $p_{\underline{A}} = \mathcal{P}(A_{j_1}) \ldots \mathcal{P}(A_{j_n})$ and therefore obeys

$$2^{-n(S_e(\pi,\mathcal{N})+\varepsilon)} \leq p_{\underline{A}} \leq 2^{-n(S_e(\pi,\mathcal{N})-\varepsilon)} .$$

Hence, keeping only the ε-typical Kraus operators the original channel $\mathcal{N}^{\otimes n}$ reduces to a channel $\mathcal{N}_{\varepsilon,n}$ with Kraus operators $\underline{A} \in \aleph_{\varepsilon,n}$ of similar probability $p_{\underline{A}}$. In general, this strongly reduced the number of Kraus operators from $|\mathcal{N}|^n$ to $|\mathcal{N}_{\varepsilon,n}|$ and renders $\mathcal{N}_{\varepsilon,n}$ much closer to a uniform channel than the original channel $\mathcal{N}^{\otimes n}$. At the same time, the transmission probability of the homogeneously mixed state π_n deviates only by an exponentially small amount from unity.

In order to achieve also approximate unitality, we will further modify the channel by letting $\mathcal{N}_{\varepsilon,n}$ follow a projection $\mathcal{T}_{\varepsilon,n} : \rho \mapsto \Pi_{\varepsilon,n}\rho\,\Pi_{\varepsilon,n}$ on the ε-typical subspace $T_{\varepsilon,n} \subset Q^n$ of the density $\mathcal{N}(\pi)$. This defines the ε-reduced operation of $\mathcal{N}^{\otimes n}$ by

$$\tilde{\mathcal{N}}_{\varepsilon,n} := \mathcal{T}_{\varepsilon,n} \circ \mathcal{N}_{\varepsilon,n} ,$$

with the following properties shown in Appendix D:

$$|\tilde{\mathcal{N}}_{\varepsilon,n}| \leq 2^{n(S_e(\pi,\mathcal{N})+\varepsilon)} , \tag{27}$$

$$\operatorname{tr} \tilde{\mathcal{N}}_{\varepsilon,n}(\pi_n) \geq 1 - 4e^{-n\psi_3(\varepsilon)} , \tag{28}$$

$$\|\tilde{\mathcal{N}}_{\varepsilon,n}(\pi_n)\|_F^2 \leq 2^{-n(S(\mathcal{N}(\pi))-3\varepsilon)} , \tag{29}$$

where $\psi_3(\varepsilon)$ is a positive number independent of n. Now we are ready to proof relation (25):

7.4 $Q(\mathcal{N}) \geq I(\pi, \mathcal{N})$

We note that for any code $C \subset Q^{\otimes n}$

$$F_e(C, \mathcal{N}^{\otimes n}) \geq F_e(C, \mathcal{N}_{\varepsilon,n}) \geq F_e(C, \tilde{\mathcal{N}}_{\varepsilon,n}) . \tag{30}$$

The first inequality holds because $\mathcal{N}_{\varepsilon,n}$ is a reduction of $\mathcal{N}^{\otimes n}$ (cf. Sec. 1.2, relation (4)), and the second one follows from

$$\max_{\mathcal{R}} F(\pi_C, \mathcal{R} \circ \mathcal{N}_{\varepsilon,n}) \geq \max_{\mathcal{R}} F(\pi_C, \mathcal{R} \circ \mathcal{T}_{\varepsilon,n} \circ \mathcal{N}_{\varepsilon,n}) = \max_{\mathcal{R}} F(\pi_C, \mathcal{R} \circ \tilde{\mathcal{N}}_{\varepsilon,n}) .$$

Averaging relation (30) over the unitary ensemble of codes $C \subset Q^n$ of dimension

$$K_n = \lfloor 2^{nR} \rfloor ,$$

we immediately obtain with relation (13) and the bounds (27), (28), (29)

$$\begin{aligned} \left[F_e(C, \mathcal{N}^{\otimes n})\right]_{K_n} &\geq \mathrm{tr}\tilde{\mathcal{N}}_{\varepsilon,n}(\pi_n) - \sqrt{K_n |\tilde{\mathcal{N}}_{\varepsilon,n}|}\, \| \tilde{\mathcal{N}}_{\varepsilon,n}(\pi_n) \|_F \\ &\geq 1 - 4e^{-n\psi_3(\varepsilon)} - 2^{\frac{n}{2}(R + S_e(\pi,\mathcal{N}) - S(\mathcal{N}(\pi)) + 4\varepsilon)} . \end{aligned}$$

For all $\varepsilon > 0$, the right-hand side of inequality converges to unity in the limit $n \to \infty$ if the asymptotic rate R obeys

$$R + 4\varepsilon \;<\; S(\mathcal{N}(\pi)) - S_e(\pi, \mathcal{N}) \equiv I(\pi, \mathcal{N}) .$$

That is, all rates $R = \lim_{n\to\infty} \frac{1}{n} \log_2 K_n$ below $I(\pi, \mathcal{N})$ are achievable and therefore $I(\pi, \mathcal{N})$ is a lower bound of the capacity $Q(\mathcal{N})$.

Relation (26) follows as a corollary:

7.5 $Q(\mathcal{N}) \geq I(\pi_V, \mathcal{N})$

Let V be any subspace of the input Hilbert space Q of a channel $\mathcal{N} : B(Q) \to B(Q')$, and let $\pi_V = \Pi_V / |V|$ be the normalized projection on V. The restriction of $\mathcal{N}$ to densities with support in V,

$$\mathcal{N}_V : B(V) \to B(Q') \; , \; \rho \mapsto \mathcal{N}(\rho) ,$$

is a channel for which the result of the previous subsection obviously predicts $I(\pi_V, \mathcal{N}_V)$ an achievable rate. It is evident that then $I(\pi_V, \mathcal{N}) = I(\pi_V, \mathcal{N}_V)$ is also an achievable rate of the complete channel $\mathcal{N}$. Thus, for any subspace $V \subset Q$

$$Q(\mathcal{N}) \geq I(\pi_V, \mathcal{N}) .$$

8 $Q(\mathcal{N}) \geq I_r(\mathcal{N})$

Finally, we will show that with the BSST lemma the result of the last subsection implies the lower bound

$$Q(\mathcal{N}) \geq \frac{1}{m} I(\rho, \mathcal{N}^{\otimes m}) ,$$

where m is an arbitrary large integer, and ρ any density on Q^m. Clearly, this suffices to prove the regularized coherent information $I_r(\mathcal{N})$ (cf. Sec. 2.2) a lower bound of $Q(\mathcal{N})$.

The BSST lemma [25] states that for a channel $\mathcal{N}$ and an arbitrary state ρ on the input space of $\mathcal{N}$

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} S(\mathcal{N}^{\otimes n}(\pi_{\varepsilon,n})) \;=\; S(\mathcal{N}(\rho)) ,$$

where $\pi_{\varepsilon,n}$ is the normalized projection on the frequency-typical subspace $T^{(f)}_{\varepsilon,n}$ of ρ. As a corollary, one obtains an analogous relation for the coherent information,

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}I(\pi_{\varepsilon,n},\mathcal{N}^{\otimes n}) \;=\; I(\rho,\mathcal{N})\,.$$

$T^{(f)}_{\varepsilon,n}$ is similar to the ordinary typical subspace $T_{\varepsilon,n}$ which we have used above. The difference is that for $T^{(f)}_{\varepsilon,n}$ typicality of a sequence is defined via the relative frequency of symbols in this sequence, whereas for $T_{\varepsilon,n}$ it is defined by its total probability. For details we refer the reader to the work of Holevo [30], where an elegant proof of the BSST lemma is given.

Here, what matters is solely the fact that $\pi_{\varepsilon,n}$ is a homogeneously distributed subspace density of the kind that we used in the previous subsection. Thus we can make use of the bound $Q(\mathcal{E})\geq I(\pi_V,\mathcal{E})$ with, for instance, $\mathcal{E}=\mathcal{N}^{\otimes mn}$, and V being the frequency-typical subspace $T^{(f)}_{\varepsilon,n}\subset Q^{mn}$ of an arbitrary density ρ on Q^m. This means that for any $\varepsilon>0$ and any m,n

$$Q(\mathcal{N}^{\otimes mn})\geq I(\pi_{\varepsilon,n},\mathcal{N}^{\otimes mn})\,.$$

Using the trivial identity $Q(\mathcal{N}^{\otimes k})=kQ(\mathcal{N})$ we can therefore write

$$\begin{aligned}
Q(\mathcal{N}) &= \frac{1}{m}\lim_{n\to\infty}\frac{1}{n}Q(\mathcal{N}^{\otimes mn})\\
&\geq \frac{1}{m}\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}I(\pi_{\varepsilon,n},(\mathcal{N}^{\otimes m})^{\otimes n})\\
&= \frac{1}{m}I(\rho,\mathcal{N}^{\otimes m})\,,
\end{aligned}$$

where the last equation follows from the corollary.

I would like to thank Michal Horodecki and Milosz Michalski for inviting me to contribute to the present issue of OSID on the quantum coding theorem.

A Unitary invariant Hermitian form

Let H be a finite dimensional Hilbert space with an orthonormal basis $|1\rangle,\ldots,|N\rangle$, and let $b:\mathcal{B}(H)\times\mathcal{B}(H)\to\mathbf{C}$ be a unitary invariant Hermitian form. For $i,j\in\{1,\ldots,N\}$ let $E_{ij}:=|i\rangle\langle j|$. As a consequence of the unitary invariance one finds constants α,β and γ such that for $i,j,\in\{1,\ldots,N\}$, $i\neq j$

$$\begin{aligned}
b(E_{ij},E_{ij}) &= \alpha\,,\\
b(E_{ii},E_{jj}) &= \beta\,,\\
b(E_{ii},E_{ii}) &= \gamma\,,
\end{aligned}$$

and for all other combinations of indices $i, j, l, m \in \{1, \ldots, N\}$

$$b(E_{ij}, E_{lm}) = 0 \, .$$

This immediately leads to

$$b(V, W) = (\gamma - \alpha - \beta)\, b_1(V, W) \; + \; \beta \, \mathrm{tr} V^\dagger \mathrm{tr} W \; + \; \alpha \, \mathrm{tr} \, V^\dagger W \, ,$$

with

$$b_1(V, W) = \sum_{i=1}^{N} \langle i|V^\dagger|i\rangle \langle i|W|i\rangle \, .$$

Obviously, b_1 is not unitary invariant, from which we conclude $\gamma - \alpha - \beta = 0$ and thus

$$b(V, W) = \beta \, \mathrm{tr} V^\dagger \mathrm{tr} W \; + \; \alpha \, \mathrm{tr} \, V^\dagger W \, ,$$

which is what we wanted to prove.

B Average of $|\langle\psi|\pi_C|\psi\rangle|^2$

We show that independent of the normalized vector $|\psi\rangle \in Q$

$$[|\langle\psi|\pi_C|\psi\rangle|^2]_K = \frac{1 + K^{-1}}{M^2 + M} \tag{31}$$

(notations as in Sec. 5). By definition,

$$[|\langle\psi|\pi_C|\psi\rangle|^2]_K = \frac{1}{K^2} \int d\mu(U) \, |\langle\psi|U \, \Pi_0 U^\dagger|\psi\rangle|^2 \, ,$$

where the integral extends over $\mathrm{U}(Q)$ and Π_0 is the projection on an arbitrarily chosen linear subspace $C_0 \subset Q$ of dimension K. We extend $|\psi\rangle \equiv |\psi_1\rangle$ to an orthonormal basis $|\psi_1\rangle, \ldots, |\psi_M\rangle$ of Q, and chose

$$C_0 := \mathrm{span}\{|\psi_1\rangle, \ldots, |\psi_K\rangle\} \, .$$

Then

$$\int d\mu(U) \, |\langle\psi|U \, \Pi_0 U^\dagger|\psi\rangle|^2 = \sum_{i,j=1}^{K} \int d\mu(U) \, |U_{1i}|^2 |U_{1j}|^2 \, ,$$

where $U_{ij} = \langle\psi_i|U|\psi_j\rangle$. Making use of the unitary invariance of μ, this becomes

$$K \int d\mu(U) \, |U_{11}|^4 \; + \; (K^2 - K) \int d\mu(U) \, |U_{11}|^2 |U_{12}|^2 \, .$$

For the calculation of these integrals we refer to the work of Pereyra and Mello [31], in which, amongst others, the joint probability density for the elements $U_{11}, \ldots, U_{1k}$ of a random unitary matrix $U \in U_K$ has been determined to be

$$p(U_{11}, \ldots, U_{1k}) = c \left(1 - \sum_{a=1}^{k} |U_{1a}|^2\right)^{n-k-1} \Theta(1 - \sum_{a=1}^{k} |U_{1a}|^2) \,,$$

where c is a normalization constant, and $\Theta(x)$ denotes the standard unit step function. By a straightforward calculation, we obtain from this

$$\begin{aligned} \int d\mu(U)\, |U_{11}|^4 &= \frac{2}{M^2 + M} \,, \\ \int d\mu(U)\, |U_{11}|^2 |U_{12}|^2 &= \frac{1}{M^2 + M} \,, \end{aligned}$$

which immediately leads to Eq. (31).

C Typical Sequences

The first property follows from

$$1 = \sum_{\underline{a} \,\in\, \aleph^n} p_{\underline{a}} \;\geq \sum_{\underline{a} \,\in\, \aleph_{\varepsilon,n}} p_{\underline{a}} \;\geq\; |\aleph_{\varepsilon,n}| 2^{-n(H(\mathcal{P})+\varepsilon)} \,.$$

To prove the second property we first realize that by definition

$$\begin{aligned} P_{\varepsilon,n} &= \Pr(\text{ “}\underline{a} \in \aleph^n \text{ is } \varepsilon\text{-typical”}) \;=\; \Pr(\big|-\log_2(p_{\underline{a}}) - nH(\mathcal{P})\big| \;\leq\; n\varepsilon) \\ &= \Pr(\,|\sum_{l=1}^{n} (-\log_2 \mathcal{P}(a_l) - H(\mathcal{P}))\,| \;\leq\; n\varepsilon\,) \,. \end{aligned}$$

The negative logarithms of the probabilities $\mathcal{P}(a_l)$ can be understood as n independent random variables Y_l that assume values $-\log_2 \mathcal{P}(a)$ for all $a \in \aleph$ with probabilities $\mathcal{P}(a)$. Their mean is the Shannon entropy $H(\mathcal{P})$,

$$\mu = E(Y_1) = -\sum_{a \in \aleph} \mathcal{P}(a) \log_2 \mathcal{P}(a) = H(\mathcal{P}) \,.$$

This means that

$$1 - P_{\varepsilon,n} \;=\; \Pr(\,|\sum_{l=1}^{n} (Y_l - \mu)| \;\geq\; n\varepsilon\,)$$

is the probability of a large deviation $\propto n$. Since the variance σ and all higher moments of $Y_1 - \mu$ are finite we can employ a result from the theory of large deviations [32], according to which

$$\Pr(\,|\sum_{l=1}^{n} (Y_l - \mu)| \;\geq\; n\varepsilon\,) \;\leq\; 2e^{-n\psi(\varepsilon)} \,,$$

where $\psi(\varepsilon)$ is a positive number that is approximately $\varepsilon^2/2\sigma^2$.

D Properties of $\mathcal{N}_{\varepsilon,n}$ and $\tilde{\mathcal{N}}_{\varepsilon,n}$

We will show the following relations (definitions and notations as in Sec. 7.3):

$$\begin{aligned} |\mathcal{N}_{\varepsilon,n}| &\leq 2^{n(S_e(\pi,\mathcal{N})+\varepsilon)} && (32)\\ \mathrm{tr}\,\mathcal{N}_{\varepsilon,n}(\pi_n) &\geq 1-2e^{-n\psi_1(\varepsilon)} && (33)\\ |\tilde{\mathcal{N}}_{\varepsilon,n}| &\leq 2^{n(S_e(\pi,\mathcal{N})+\varepsilon)} && (34)\\ \mathrm{tr}\,\tilde{\mathcal{N}}_{\varepsilon,n}(\pi_n) &\geq 1-4e^{-n\psi_3(\varepsilon)} && (35)\\ \|\tilde{\mathcal{N}}_{\varepsilon,n}(\pi_n)\|_F^2 &\leq 2^{-n(S(\mathcal{N}(\pi))-3\varepsilon)}\,, && (36) \end{aligned}$$

where $\psi_1(\varepsilon)$ and $\psi_3(\varepsilon)$ are positive numbers independent of n.

The first relation follows from $|\mathcal{N}_{\varepsilon,n}| = |\aleph_{\varepsilon,n}| \leq 2^{n(H(\mathcal{P})+\varepsilon)}$ and $H(\mathcal{P}) = S_e(\pi,\mathcal{N})$. To prove relation (33) we note that for a Kraus operator $\underline{A} = A_{j_1}\otimes\ldots\otimes A_{j_n}$

$$\frac{1}{|Q|^n}\mathrm{tr}\;\underline{A}^\dagger\underline{A} = \frac{1}{|Q|}\mathrm{tr}\;{A_{j_1}}^\dagger A_{j_1}\ldots\frac{1}{|Q|}\mathrm{tr}\;{A_{j_n}}^\dagger A_{j_n} = \mathcal{P}(A_{j_1})\ldots\mathcal{P}(A_{j_n}) \equiv p_{\underline{A}}\,.$$

Making use of property (ii) of typical sequences this shows

$$\mathrm{tr}\,\mathcal{N}_{\varepsilon,n}(\pi_n) = \frac{1}{|Q|^n}\sum_{\underline{A}\in\aleph_{\varepsilon,n}}\mathrm{tr}\;\underline{A}^\dagger\underline{A} = \sum_{\underline{A}\in\aleph_{\varepsilon,n}} p_{\underline{A}} \geq 1-2e^{n\psi_1(\varepsilon)}\,,$$

where $\psi_1(\varepsilon)$ is a positive number independent of n. Relation (34) is evident by relation (32) and

$$\tilde{\mathcal{N}}_{\varepsilon,n}(\rho) = \Pi_{\varepsilon,n}\,\mathcal{N}_{\varepsilon,n}(\rho)\,\Pi_{\varepsilon,n} = \sum_{\underline{A}\in\aleph_{\varepsilon,n}}(\Pi_{\varepsilon,n}\underline{A})\rho(\Pi_{\varepsilon,n}\underline{A})^\dagger\,.$$

In order to show (35) it is convenient to introduce the complementary operation $\mathcal{M}_{\varepsilon,n}$ of $\mathcal{N}_{\varepsilon,n}$ by

$$\mathcal{N}^{\otimes n} = \mathcal{N}_{\varepsilon,n} + \mathcal{M}_{\varepsilon,n}\,,$$

i.e. $\mathcal{M}_{\varepsilon,n}$ consists of the ε-"untypical" Kraus operators of $\mathcal{N}^{\otimes n}$,

$$\mathcal{M}_{\varepsilon,n}(\rho) = \sum_{\underline{A}\in\aleph\setminus\aleph_{\varepsilon,n}}\underline{A}\,\rho\,\underline{A}^\dagger\,.$$

Then,

$$\begin{aligned} \mathrm{tr}\,\tilde{\mathcal{N}}_{\varepsilon,n}(\pi_n) &= \mathrm{tr}\,\Pi_{\varepsilon,n}(\mathcal{N}^{\otimes n}(\pi_n) - \mathcal{M}_{\varepsilon,n}(\pi_n))\\ &\geq \mathrm{tr}\Pi_{\varepsilon,n}\mathcal{N}^{\otimes n}(\pi_n) - \mathrm{tr}\,\mathcal{M}_{\varepsilon,n}(\pi_n)\,. \qquad (37) \end{aligned}$$

The inequality results from the fact that for two positive operators A, B always $\mathrm{tr}AB \geq 0$, and therefore (indices suppressed)

$$\mathrm{tr}\,\mathcal{M}(\rho) = \mathrm{tr}\,\Pi\mathcal{M}(\rho) + \mathrm{tr}\,(\underline{1}-\Pi)\mathcal{M}(\rho) \geq \mathrm{tr}\,\Pi\mathcal{M}(\rho)\,.$$

Taking into account that $\Pi_{\varepsilon,n}$ projects on the typical subspace $T_{\varepsilon,n}$ of $\mathcal{N}(\pi)$ and using property (ii') of typical subspaces, the first term in Eq. (37) can be bounded from below as

$$\mathrm{tr}\,\Pi_{\varepsilon,n}\mathcal{N}^{\otimes n}(\pi_n) = \mathrm{tr}\,\Pi_{\varepsilon,n}\mathcal{N}^{\otimes n}(\pi^{\otimes n}) = \mathrm{tr}\,\Pi_{\varepsilon,n}(\mathcal{N}(\pi))^{\otimes n} \;\geq\; 1 - 2e^{-n\psi_2(\varepsilon)}\,.$$

The second term in Eq. (37) obeys

$$\mathrm{tr}\,\mathcal{M}_{\varepsilon,n}(\pi_n) \;=\; \mathrm{tr}\,\mathcal{N}^{\otimes n}(\pi_n) \;-\; \mathrm{tr}\,\mathcal{N}_{\varepsilon,n}(\pi_n) \;\leq\; 2e^{-n\psi_1(\varepsilon)}\,,$$

by relation (33). We thus find

$$\mathrm{tr}\,\tilde{\mathcal{N}}_{\varepsilon,n}(\pi_n) \;\geq\; 1 - 2(e^{-n\psi_2(\varepsilon)} + e^{-n\psi_1(\varepsilon)}) \;\geq\; 1 - 4\,e^{-n\psi_3(\varepsilon)}\,,$$

when $\psi_3(\varepsilon) := \min\{\psi_1(\varepsilon), \psi_2(\varepsilon)\}$.

Finally, we address the Frobenius norm of $\tilde{\mathcal{N}}(\pi_n)$. For positive operators A, B

$$\|\,A+B\,\|_F^2 = \|\,A\,\|_F^2 + \|\,B\,\|_F^2 + 2\mathrm{tr}\,AB \;\geq \|\,A\,\|_F^2 + \|\,B\,\|_F^2\;.$$

This can be used to derive

$$\|\,\mathcal{T}_{\varepsilon,n}\circ\mathcal{N}^{\otimes n}(\pi_n)\,\|_F^2 = \|\,\mathcal{T}_{\varepsilon,n}\circ(\mathcal{N}_{\varepsilon,n}+\mathcal{M}_{\varepsilon,n})(\pi_n)\,\|_F^2 \geq \|\,\mathcal{T}_{\varepsilon,n}\circ\mathcal{N}_{\varepsilon,n}(\pi_n)\,\|_F^2\;.$$

Thus

$$\begin{aligned}
\|\,\tilde{\mathcal{N}}_{\varepsilon,n}(\pi_n)\,\|_F^2 &= \|\,\mathcal{T}_{\varepsilon,n}\circ\mathcal{N}_{\varepsilon,n}(\pi_n)\,\|_F^2 \\
&\leq \|\,\mathcal{T}_{\varepsilon,n}\circ\mathcal{N}^{\otimes n}(\pi_n)\,\|_F^2 \\
&= \|\,\Pi_{\varepsilon,n}\,(\mathcal{N}(\pi))^{\otimes n}\,\Pi_{\varepsilon,n}\,\|_F^2 \\
&= \sum_{|\underline{\mathrm{v}}\rangle\ \varepsilon\text{-typical eigenvector}} (p_{\underline{\mathrm{V}}})^2 \\
&\leq 2^{-n(S(\mathcal{N}(\pi))-3\varepsilon)}\,,
\end{aligned}$$

where we used $\dim T_{\varepsilon,n} \leq 2^{n(S(\mathcal{N}(\pi))+\varepsilon)}$ (property (i')) and $p_{\underline{\mathrm{V}}} \leq 2^{-n(S(\mathcal{N}(\pi))-\varepsilon)}$ to derive the last inequality.

References

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).

[2] M. Keyl, Phys. Rep. **369**, 431 (2002).

[3] W. F. Stinespring, Proc. Am. Math. Soc. **6**, 211 (1955).

[4] K. Kraus, *States, Effects, and Operations*, Lecture Notes in Physics Vol. 190 (Springer-Verlag, Berlin, Heidelberg, 1983).

[5] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).

[6] R. Jozsa, J. Mod. Opt., **41**, 2315 (1994).

[7] B. Schumacher, Phys. Rev. A **54**, 2614 (1996).

[8] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **60**, 1888 (1999).

[9] M. A. Nielsen, Phys. Lett. A **303**, 249 (2002).

[10] P. W. Shor, Phys. Rev. A **52**, R2493 (1995).

[11] A. M. Steane, Phys. Rev. Lett. **77**, 793 (1996).

[12] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).

[13] B. Schumacher and M. D. Westmoreland, Quantum Inf. Process. **1**, 5 (2002), quant-ph/0112106.

[14] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).

[15] S. Lloyd, Phys. Rev. A **55**, 1613 (1997).

[16] H. Barnum, M. A. Nielsen, and B. Schumacher, Phys. Rev. A **57**, 4153 (1998), quant-ph/9702049.

[17] H. Barnum, E. Knill, and M. A. Nielsen, IEEE Trans. Inf. Theory **46**, 1317 (2000), quanth-ph/9809010.

[18] D. Kretschmann and R. F. Werner, New J. Phys. **6**, 26 (2004).

[19] B. Schumacher and M. A. Nielsen, Phys. Rev. A **54**, 2629 (1996).

[20] I. Devetak, IEEE Trans. Inf. Theory **51**, 44 (2005), quant-ph/0304127.

[21] R. Klesse, Phys. Rev. A **75**, 062315 (2007).

[22] P. Hayden, M. Horodecki, J. Yard, and A. Winter, preprint arXiv:quant-ph/0702005v1 (2007)

[23] P. W. Shor, *The quantum channel capacity and coherent information*, Lecture Notes, MSRI Workshop on Quantum Computation, San Francisco, 2002 (unpublished); available at http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1

[24] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter, arXiv:quant-ph/0606225 (2006).

[25] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, IEEE Trans. Inf. Theory **48**, 2637 (2002), quant-ph/0106052.

[26] H. Weyl, *The Classical Groups* (Princeton University Press, New Jersey, 1946).

[27] R. Howe, in *Perspectives on Invariant Theory*, Schur Lectures, edited by I. Piatetski-Shapiro and S. Gelbart (Bar-Ilan University, Ramat-Gan, 1995).

[28] A. Ekert and C. Macchiavello, Phys. Rev. Lett. **77**, 2585 (1996).

[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley and Sons, New York, 1991).

[30] A. S. Holevo, J. Math. Phys. **43**, 4326 (2002).

[31] P. Pereyra and P. A. Mello, J. Phys. A **16**, 237 (1983).

[32] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes* (Oxford University Press, New York, 1992).