Approximate quantum error correction, random codes, and quantum channel capacity

Rochus Klesse*

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

(Received 15 January 2007; revised manuscript received 24 April 2007; published 13 June 2007)

We work out a theory of approximate quantum error correction that allows us to derive a general lower bound for the entanglement fidelity of a quantum code. The lower bound is given in terms of Kraus operators of the quantum noise. This result is then used to analyze the average error correcting performance of codes that are randomly drawn from unitarily invariant code ensembles. Our results confirm that random codes of sufficiently large block size are highly suitable for quantum error correction. Moreover, employing a lemma of Bennett, Shor, Smolin, and Thapliyal, we prove that random coding attains information rates of the regularized coherent information.

DOI: 10.1103/PhysRevA.75.062315 PACS number(s): 03.67.Hk, 03.67.Pp

I. INTRODUCTION

Physical processing, transmission, and storage of quantum information unavoidably suffers from decohering interactions with the environment. The insight that the resulting errors can, in principle, be corrected has been a major breakthrough in the field of quantum information theory [1,2]. A theory of quantum error correction (QEC) rapidly evolved [3–5] and eventually led to the concept of quantum fault tolerance [6], which, in fact, put large-scale quantum computation back in the realms of possibility. Quantum error correction stands in close relation to the information capacity of a noisy quantum channel and the quantum coding theorem [7–9].

In this paper we elaborate a theory of approximate QEC. We obtain a general and easily computable lower bound for the entanglement fidelity of a noisy channel $\mathcal{N}$ that is attainable when the information is encoded in a given error correcting code. The bound is expressed in terms of Kraus operators of $\mathcal{N}$, and the projection on the code space (Sec. III).

We employ this theory to analyze the average error correcting performance of codes that are chosen at random from certain code ensembles. For the unitarily invariant ensemble of all K-dimensional code spaces we find a surprisingly simple lower bound for the averaged code entanglement fidelity. Its deviation from unity is determined by $\sqrt{KN}\|\mathcal{N}(\pi_Q)\|_F$, where N is the number of Kraus operators in an operator-sum representation of the noise $\mathcal{N}$ under consideration, π_Q is the homogeneously distributed state of the system Q on which $\mathcal{N}$ is operating, and $\|A\|_F$ denotes the Frobenius norm $\sqrt{\mathrm{tr}A^\dagger A}$ of an operator A. We derive this result by reverting to random matrix theory. For the special case of unital noise the lower bound immediately reveals that randomly chosen codes attain with high probability the quantum Hamming bound [3] (Sec. IV).

Our next issue is the extension of the foregoing considerations to the case of noise operations that do not conserve the trace. We find it useful to understand them as the result of a selective process and therefore define fidelities and coherent information in this situation slightly different from the standard definitions in literature (see, e.g., [10]) (Sec. V).

One motivation why we extend our theory to trace-decreasing operations becomes apparent in the last section. Here we show that with the aid of a recent lemma of Bennett, Shor, Smolin, and Thapliyal (BSST) [11,12] our results allow a relatively simple proof of the direct coding theorem. Our proof follows ideas of Shor [13] and Lloyd [7] by showing that QEC based on random code spaces attains rates of the regularized coherent information. The proof is therefore quite different from Devetak's one [14], which is based on a correspondence of classical private information and quantum information (Sec. VI).

After having clarified some conventions and notations, we will start in Sec. II with a brief introduction to QEC and quantum channel capacity. The remaining sections are organized as laid out above.

A. Conventions and notations

We denote a general mixed state by ρ, a general pure state by ψ, and add subscripts to indicate the system. For instance, ψ_{QR} means a pure state of the joint system of Q and R.

We will use the trace norm $\|A\|_{\mathrm{tr}}=\mathrm{tr}\sqrt{A^\dagger A}$, and the Frobenius (Hilbert-Schmidt) norm $\|A\|_F=\sqrt{\mathrm{tr}A^\dagger A}$ for a linear operator A. The two-state fidelity is here defined as $F(\rho_1,\rho_2)=\|\sqrt{\rho_1}\sqrt{\rho_2}\|_{\mathrm{tr}}^2$.

II. QUANTUM ERROR CORRECTION AND QUANTUM CAPACITY

Throughout the paper, we consider a quantum system Q that is supposed to store or transmit quantum information. We denote the Hilbert space of Q by H_Q and its finite dimension by M. In addition to a possible internal unitary dynamics, Q is subjected to external noise during storage or transmission. Let the effect of both be described by a completely positive, trace-preserving mapping $\mathcal{N}$ that maps an initial density operator ρ to a final density operator $\rho'=\mathcal{N}(\rho)$ [15,16]. We call $\mathcal{N}$ either a noise operation or, synonymously, a noisy channel. $\mathcal{N}$ can be always represented in an operator sum

$$\mathcal{N}(\rho)=\sum_{i=1}^{N} A_i\rho A_i^{\dagger},$$

where the (nonunique) Kraus operators $A_1,\ldots,A_N$ are linear operators on H_Q. They satisfy the completeness relation $\Sigma_i A_i^{\dagger}A_i=\mathbf{1}_Q$.

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

A. Quantum error correction

In general, a QEC scheme for the noise $\mathcal{N}$ on Q is based on a quantum error correcting code C, which, by definition, is a certain linear subspace C of H_Q. Let K be the dimension of C, and let P be the projection on C. We call a state ρ a state in C or a code state (of C) if the support of ρ is a subset of C. If the code C is suitably chosen, one may find a recovery operation $\mathcal{R}$ that exactly recovers all code state from corruption by $\mathcal{N}$, i.e., for all code states ρ of C, $\mathcal{R}\circ\mathcal{N}(\rho)=\rho$.

Finding an optimal code C for the correction of some given noise $\mathcal{N}$ is a difficult task. The code C should be of course as large as possible, but at the same time the encoding in C must also be sufficiently redundant such that errors caused by $\mathcal{N}$ can still be identified and corrected. In practice, the code may also satisfy additional technical constraints. Somewhat simpler than this problem but nevertheless instructive is the following related one: Given the noise operation $\mathcal{N}$, what can be gained by the use of a certain quantum code C? Here, theory does provide definite answers in the form of necessary and sufficient conditions for the feasibility of quantum error correction.

First, there are quite elementary necessary and sufficient conditions for exact QEC [3–5,16].

Exact recovery of all code states is possible if and only if for all i, j the operators $PA_j^\dagger A_i P$ are proportional to P,

$$PA_j^\dagger A_i P = \frac{1}{K}(\mathrm{tr} PA_j^\dagger A_i P)P. \tag{1}$$

For explicitly given Kraus operators A_i it is usually no problem to check these conditions. If they are satisfied, it is also possible to explicitly construct the Kraus operators for the recovery operation $\mathcal{R}$. Things become more complicated when the conditions are violated. In this case, it can become quite difficult to foresee whether the violation is serious, and therefore error correction virtually impossible, or whether the violation is harmless and code states are still essentially correctable up to some small deviations. An early approach to this problem has been given in [17].

An alternative condition for QEC can be formulated in terms of coherent information [7,18]. The coherent information $I(\rho,\mathcal{N})$ of a state ρ with respect to the noise $\mathcal{N}$ is defined by

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

where $S(\varrho)=-\mathrm{tr}\varrho \log_2 \varrho$ is the von Neumann entropy, ψ_{RQ} is a purification of ρ, and $\mathcal{I}_R$ is the identity operation on the ancilla system R. The last term, $S(\mathcal{I}_R\otimes\mathcal{N}(\psi_{RQ}))$, is the entropy exchange $S_e(\rho,\mathcal{N})$ of ρ with respect to $\mathcal{N}$ [19]. The coherent information obeys an important inequality [18]: For any two operations $\mathcal{E}_1$ and $\mathcal{E}_2$

$$S(\rho) \geqslant I(\rho,\mathcal{E}_1) \geqslant I(\rho,\mathcal{E}_2 \circ \mathcal{E}_1). \tag{2}$$

Moreover, equality in the first inequality holds if and only if the action of $\mathcal{E}_1$ on ρ can be completely reversed, meaning that there exists an $\mathcal{R}$ such that $\mathcal{I}_R\otimes(\mathcal{R}\circ\mathcal{E}_1)(\psi_{RQ})=\psi_{RQ}$, for any purification ψ_{RQ} of ρ. This leads to the following necessary and sufficient condition for error correction [18].

Exact recovery of all code states is possible if and only if for a state ρ_C with $\mathrm{supp}(\rho_C)=C$

$$S(\rho_C) = I(\rho_C,\mathcal{N}). \tag{3}$$

Schumacher and Westmoreland [20] have shown that this condition is robust against small perturbations, i.e., if it is only approximately satisfied, then errors can still be approximately corrected. Their central result is a lower bound for the entanglement fidelity [19] of an arbitrary state ρ under the noise $\mathcal{N}$ and a subsequent recovery operation $\mathcal{R}$. It is proven that for given ρ and $\mathcal{N}$ there exists an $\mathcal{R}$ such that

$$F_e(\rho,\mathcal{R}\circ\mathcal{N}) \geqslant 1 - 2\sqrt{S(\rho) - I(\rho,\mathcal{N})}. \tag{4}$$

To elaborate on this, let us discuss entanglement fidelity and its relevance for our purposes.

B. Entanglement fidelity

The entanglement fidelity $F_e(\rho,\mathcal{E})$ of a state ρ under an operation $\mathcal{E}$ on Q is defined by

$$F_e(\rho,\mathcal{E}) \coloneqq \langle\psi_{RQ}|\mathcal{I}_R \otimes \mathcal{E}(\psi_{RQ})|\psi_{RQ}\rangle,$$

where ψ_{RQ} is any purification of ρ. That this is independent of the chosen purification can be seen from the representation in terms of Kraus operators of $\mathcal{N}$, $F_e(\rho,\mathcal{E})=\Sigma_{i=1}^{N}|\mathrm{tr}\rho A_i|^2$ [19]. Especially interesting is the entanglement fidelity of the homogeneously distributed code state $\pi_C=P/K$. The reason is that $F_e(\pi_C,\mathcal{E})$ is a lower bound of the code-averaged channel fidelity $F_{\mathrm{av}}(C,\mathcal{E})$ (Appendix A 1; cf. [21,22]). Moreover, it can be shown that when $F_e(\pi_C,\mathcal{E})$ is close to unity, C must have a large subcode $\tilde{C}\subset C$ with a similar high minimum fidelity $F_{\min}(\tilde{C},\mathcal{E})$ (Appendix A 2). The entanglement fidelity $F_e(\pi_C,\mathcal{E})$ is therefore a convenient figure of merit that characterizes the distortion of states in C under the operation $\mathcal{E}$.

In order to capture the suitability of a code C for QEC without referring to a certain recovery operation we introduce

$$F_e(C,\mathcal{N}) \coloneqq \max_{\mathcal{R}} F_e(\pi_C,\mathcal{R}\circ\mathcal{N}), \tag{5}$$

the entanglement fidelity of the code C under noise $\mathcal{N}$. By relation (4) it is then clear that

$$F_e(C,\mathcal{N}) \geqslant 1 - 2\sqrt{S(\pi_C) - I(\pi_C,\mathcal{N})}. \tag{6}$$

This shows that for small $S(\pi_C)-I(\pi_C,\mathcal{N})\ll 1$ the code entanglement fidelity is close to unity and thus approximate QEC is possible.

Building on ideas of Schumacher and Westmoreland's proof of relation (4), here we will derive an alternative lower bound for the code entanglement fidelity $F_e(C,\mathcal{N})$ that is explicitly given in terms of the Kraus operators of $\mathcal{N}$ [cf. relation (9) in Sec. III]. However, before we start, let us briefly point out that the code entanglement fidelity (5) can also be used to conveniently define quantum capacity of a noisy channel.

C. Quantum capacity of a noisy channel

We consider the following scheme of information transmission from Alice (sender) to Bob (receiver) by means of the channel $\mathcal{N}$ [8]: Alice is allowed to encode quantum information in blocks of n identical copies of Q, with the block size n and the encoding operation $\mathcal{E}_n$ at her disposal. Sending the block to Bob, each individual system Q is independently disturbed by the noise operation $\mathcal{N}$, i.e., the whole block Q^n is subjected to $\mathcal{N}^{\otimes n}$. Bob is allowed to perform any decoding operation $\mathcal{R}_n$ in order to restore the message which Alice originally sent. The maximum amount of quantum information, measured in units of qubits, that can be reliably transmitted per channel use in such a scheme defines the quantum capacity $Q(\mathcal{N})$ of the noisy channel $\mathcal{N}$ [8].

Precise mathematical definitions of the quantum capacity can be given in many ways [23]. Here we use one that fits in the present context of approximate QEC and the code entanglement fidelity.

It has been shown that restricting the encoding operation $\mathcal{E}_n$ to isometric embeddings into $H_Q^{\otimes n}$ has no effect on the capacity [9]. $\mathcal{E}_n$ is thus sufficiently described by the subspace C_n of $H_Q^{\otimes n}$ whose code states represent the encoded information. Viewing C_n as an error correcting code, Bob is able to reconstruct Alice's message within a precision that is given by the code entanglement fidelity $F_e(C_n,\mathcal{N}^{\otimes n})$. We follow the standard definitions and call R an achievable rate of $\mathcal{N}$ if there is a sequence of code spaces $C_n \subset H_Q^{\otimes n}$, $n=1,2,\ldots$, such that

$$\limsup_{n\to\infty} \frac{\log_2 \dim C_n}{n} = R, \quad \text{and} \lim_{n\to\infty} F_e(C_n,\mathcal{N}^{\otimes n}) = 1. \tag{7}$$

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

The quantum coding theorem for noisy channels [7–9] states that the quantum capacity $Q(\mathcal{N})$ of a channel $\mathcal{N}$ equals the regularized coherent information

$$I_r(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} \max_{\rho} I(\rho,\mathcal{N}^{\otimes n}). \tag{8}$$

$I_r(\mathcal{N})$ has long been known an upper bound for $Q(\mathcal{N})$, which is the content of the converse coding theorem [8,9]. The direct coding theorem, stating that $I_r(\mathcal{N})$ is actually attainable, has been strictly proven only recently by Devetak [14]. His proof utilizes the correspondence of private classical information and quantum information. More direct proofs in the spirit of Shannon's ideas on random coding [24] have been earlier outlined by Shor [13] and Lloyd [7]. In the last section, we will employ our theory to provide a strict proof along these lines.

III. LOWER BOUND FOR THE CODE ENTANGLEMENT FIDELITY

In this section we derive a lower bound for the code entanglement fidelity $F_e(C,\mathcal{N})$ in terms of Kraus operators $A_1,\ldots,A_N$ of $\mathcal{N}$ and the projection P on the K-dimensional code C. We will show that

$$F_e(C,\mathcal{N}) \geqslant 1 - \|D\|_{\mathrm{tr}}, \tag{9}$$

where

$$D = \frac{1}{K}\sum_{ij=1}^{N} \left(PA_i^{\dagger}A_jP - \frac{1}{K}(\mathrm{tr}PA_i^{\dagger}A_jP)P \right) \otimes |i\rangle\langle j|, \tag{10}$$

is an operator on $C \otimes H_E$, with H_E being an ancilla Hilbert-space spanned by orthonormal vectors $|1\rangle,\ldots,|N\rangle$.

The coefficients of D precisely correspond to the conditions (1) for exact error correction. If these are fulfilled the operator D vanishes and inequality (9) also predicts perfect error correction. In this sense, the lower bound (9) can be considered as a generalization of the elementary conditions (1) to the case of approximate QEC. It is worth mentioning that the lower bound does not depend on the chosen set of Kraus operators $A_1,\ldots,A_N$ for $\mathcal{N}$. Equivalent sets are related by a unitary transformation [16] which in Eq. (10) amounts merely to a unitary basis change, and therefore leaves $\|D\|_{\mathrm{tr}}$ invariant.

To prove relation (9) we describe $\mathcal{N}$ as a unitary U_{QE} on Q and an environment E, followed by a partial trace over E [15,16]. That is, for a general state ρ_Q

$$\mathcal{N}(\rho_Q) = \mathrm{tr}_E U_{QE}\rho_Q \otimes \psi_E U_{QE}^{\dagger},$$

where ψ_E is some fixed initial state of E. Further, let ψ_{RQ} be a purification of ρ_Q, let $\rho_R = \mathrm{tr}_Q \psi_{RQ}$, and let a pure state ψ'_{RQE} on RQE be defined by

$$\psi'_{RQE} = (\mathbf{1}_R \otimes U_{QE})\psi_{RQ} \otimes \psi_E(\mathbf{1}_R \otimes U_{QE}^{\dagger}).$$

ψ'_{RQE} purifies its partial states

$$\rho'_Q = \mathrm{tr}_{RE}\psi'_{RQE}, \quad \rho'_E = \mathrm{tr}_{RQ}\psi'_{RQE},$$

$$\rho'_{RE} = \mathrm{tr}_Q\psi'_{RQE}, \quad \rho'_{RQ} = \mathrm{tr}_E\psi'_{RQE}. \tag{11}$$

Following ideas that has been utilized in [20,25] we show that there exists a recovery operation $\mathcal{R}$ on Q such that

$$F_e(\rho_Q,\mathcal{R}\circ\mathcal{N}) \geqslant 1 - \|\rho'_{RE} - \rho_R \otimes \rho'_E\|_{\mathrm{tr}}. \tag{12}$$

The idea is to find in the vicinity of the actual final state ψ'_{RQE} (or an extension ψ' of it) a state $\tilde{\psi}$ from which ψ_{RQ} can be perfectly recovered by an operation $\mathcal{R}$ on Q. The distance between ψ' and $\tilde{\psi}$ will then determine a lower bound for the entanglement fidelity $F_e(\rho_Q,\mathcal{R}\circ\mathcal{N})$.

To this end, we consider the product state $\rho_R \otimes \rho'_E$ with its obvious purification

$$\tilde{\psi} \coloneqq \psi_{RQ} \otimes \psi'_{RQE}$$

on the joint system $RQSE$, where S denotes a copy of RQ. We extend ψ'_{RQE} to a pure state ψ' on $RQSE$ by some pure state ψ_S of S (i.e., tracing out S or RQE yields ψ'_{RQE} or ψ_S, respectively). According to Uhlmann's theorem [29,30,16] there is a unitary U_{QS} on QS such that

$$|\langle\tilde{\psi}|U_{QS}\psi'\rangle|^2 = F(\rho_R \otimes \rho'_E,\rho'_{RE}). \tag{13}$$

Then, for a recovery operation $\mathcal{R}$ on Q defined by

$$\mathcal{R}(\rho_Q) \coloneqq \mathrm{tr}_S U_{QS}\rho_Q \otimes \psi_S U_{QS}^{\dagger}$$

we find

$$\mathcal{I}_R \otimes \mathcal{R}(\rho'_{RQ}) = \mathrm{tr}_{SE} U_{QS} \psi' U_{QS}{}^{\dagger},$$

which by the monotonicity of the fidelity under partial trace [16] and $\psi_{RQ}=\mathrm{tr}_{SE}\tilde{\psi}$ yields

$$F_e(\rho_Q,\mathcal{R}\circ\mathcal{N}) \equiv F(\psi_{RQ},\mathcal{I}_R\otimes\mathcal{R}(\rho'_{RQ})) \geq |\langle\tilde{\psi}|U_{QS}\psi'\rangle|^2.$$

With Eq. (13) and the general relation $F(\rho,\sigma)\geq 1-\|\rho-\sigma\|_{\mathrm{tr}}$ [16] this proves relation (12).

Now, we become more specific and chose for given Kraus operators $A_1,\ldots,A_N$ of $\mathcal{N}$ its representing unitary U_{QE} such that

$$U_{QE}|\psi_Q\rangle|1\rangle = \sum_{i=1}^{N} A_i|\psi_Q\rangle|i\rangle, \qquad (14)$$

where $|1\rangle\equiv|\psi_E\rangle,|2\rangle,\ldots,|N\rangle$ are orthonormal vectors in H_E. Further, let $\rho_Q=\pi_C\equiv P/K$ with purification

$$|\psi_{RQ}\rangle = \frac{1}{\sqrt{K}}\sum_{l=1}^{K}|c_l^R\rangle|c_l^Q\rangle, \qquad (15)$$

where the orthonormal vectors $|c_1^R\rangle,\ldots|c_K^R\rangle$ and $|c_1^Q\rangle,\ldots|c_K^Q\rangle$ span H_R and C, respectively. For this setting, we obtain

$$\rho'_{RE} = \frac{1}{K}\sum_{ij=1}^{N}\sum_{l,m=1}^{K}\mathrm{tr}_Q(A_i|c_l^Q\rangle\langle c_m^Q|A_j{}^{\dagger})|c_l^R\rangle\langle c_m^R|\otimes|i\rangle\langle j|, \qquad (16)$$

$$\rho_R\otimes\rho'_E = \sum_{ij=1}^{N}\mathrm{tr}_Q(A_i\pi_C A_j{}^{\dagger})\rho_R\otimes|i\rangle\langle j|. \qquad (17)$$

Things become more convenient if we isometrically map both states with an isometry defined by

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

to

$$X := I(\rho'_{RE}) = \frac{1}{K}\sum_{ij=1}^{N}PA_i{}^{\dagger}A_jP\otimes|i\rangle\langle j|,$$

$$Y := I(\rho_R\otimes\rho'_E) = \frac{1}{K}\sum_{ij=1}^{N}\frac{1}{K}\mathrm{tr}(PA_i{}^{\dagger}A_jP)P\otimes|i\rangle\langle j|.$$

Hence, $\|\rho'_{RE}-\rho_R\otimes\rho'_E\|_{\mathrm{tr}}=\|X-Y\|_{\mathrm{tr}}$, which with relation (12) leads us to

$$F_e(\pi_C,\mathcal{R}\circ\mathcal{N}) \geq 1-\|X-Y\|_{\mathrm{tr}}.$$

Since the left-hand side is a lower bound of the code entanglement fidelity $F_e(C,\mathcal{N})$, and $X-Y=D$, this finally proves relation (9).

IV. RANDOM QUANTUM CODES

Random codes play an important role in classical as well as in quantum information theory. In this section we will analyze the average error correcting performance of random codes by means of the lower bound (9) for the entanglement fidelity of the codes. We consider the same setting as before: a quantum information storing system Q with M-dimensional Hilbert space H_Q that is exposed to noise $\mathcal{N}$ with a set of Kraus operators $A_1,\ldots A_N$.

A. Ensemble averaged code fidelity

Let E_K be an ensemble of K-dimensional codes in H_Q with an ensemble average $[A]$ defined for code dependent variables $A=A(C)$. We are interested in the ensemble averaged code entanglement fidelity $[F_e(C,\mathcal{N})]$. By inequality (9),

$$[F_e(C,\mathcal{N})] \geq 1-[\|D\|_{\mathrm{tr}}], \qquad (18)$$

where D is the code dependent operator Eq. (10).

In many cases, averaging the trace norm of D would be quite a difficult undertaking. We therefore prefer to estimate $[\|D\|_{\mathrm{tr}}]$ by the more convenient average of the squared Frobenius norm, $[\|D\|_F^2]=[\mathrm{tr}D^{\dagger}D]$: Trace norm and Frobenius norm of D with domain $C\otimes H_E$ of dimension $d=KN$ satisfy

$$\|D\|_{\mathrm{tr}} \leq \sqrt{d}\|D\|_F.$$

We remark that this inequality is a good estimate only if the eigenvalues of D are of similar magnitude. Using this estimate and employing Jensen's inequality [26] we obtain

$$[\|D\|_{\mathrm{tr}}] \leq \sqrt{d}[\|D\|_F] = \sqrt{d}[\sqrt{\|D\|_F^2}] \leq \sqrt{d[\|D\|_F^2]}, \qquad (19)$$

and so

$$[F_e(C,\mathcal{N})] \geq 1-\sqrt{KN[\|D\|_F^2]}. \qquad (20)$$

In the next subsection we will evaluate this lower bound for unitarily invariant code ensembles.

B. Unitarily invariant code ensembles

Let U_K be the unitarily invariant code ensemble that consists of all K-dimensional codes in H_Q, furnished with the unitarily invariant ensemble average

$$[A(C)]_{U_K} := \int_{\mathbf{U}(H_Q)} d\mu(U)A(UC_0),$$

where $C_0\subset H_Q$ is some fixed code space of dimension K, and μ is the (normalized) Haar measure on $\mathbf{U}(H_Q)$, the group of all unitaries on H_Q. Later on we will also consider an analogously defined ensemble $U_K(V)$ that consists of K-dimensional codes in some subspace V of H_Q.

Our task is to calculate $[\|D\|_F^2]_{U_K}$. By the explicit representation Eq. (10) of operator D we immediately find

$$\|D\|_F^2 = \mathrm{tr}D^{\dagger}D = \frac{1}{K^2}\sum_{ij=1}^{N}\mathrm{tr}(PW_{ij}{}^{\dagger}PW_{ij}) - \frac{1}{K}|\mathrm{tr}PW_{ij}|^2,$$

where the operators W_{ij} are

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

The ensemble average of $\|D\|_F^2$ can be conveniently calculated if we introduce a Hermitian form

$$b(V,W) := \left[\mathrm{tr}(PV^\dagger PW) - \frac{1}{K}\mathrm{tr}(PV^\dagger)\mathrm{tr}(PW) \right]_{U_K}, \qquad (21)$$

such that

$$[\|D\|_F^2]_{U_K} = \frac{1}{K^2}\sum_{ij} b(W_{ij}, W_{ij}). \qquad (22)$$

We recall that P is the projection on the K-dimensional code space that is chosen with unitarily invariant probability from the ensemble U_K. By Eq. (21) it is therefore clear that $b(V,W)$ is a unitarian invariant on H_Q, i.e., for any $U \in \mathbf{U}(H_Q)$

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

This places us in a position to utilize the general theory of group invariants by Weyl [27,28]: In the present situation it means that $b(V,W)$ must be a linear combination of the two fundamental unitarily invariant Hermitian forms $\mathrm{tr}V^\dagger W$ and $\mathrm{tr}V^\dagger \mathrm{tr}W$,

$$b(V,W) = \alpha \mathrm{tr}V^\dagger W + \beta \mathrm{tr}V^\dagger \mathrm{tr}W. \qquad (23)$$

To determine the coefficients α and β we derive two linear independent equations by equating Eqs. (21) and (23) for two special choices of the operators V and W. For $V=W=\mathbf{1}_{H_Q}$ we obtain as a first equation,

$$\alpha M + \beta M^2 = 0. \qquad (24)$$

Next, we set $V=W=P_1$, where P_1 is the projection on a one-dimensional space spanned by some unit vector $|\psi\rangle \in H_Q$. From Eq. (21) we immediately find

$$b(P_1, P_1) = \left(1 - \frac{1}{K}\right)[|\langle\psi|P|\psi\rangle|^2]_{U_K}.$$

Reverting to results from random matrix theory, we obtain in Appendix B $[|\langle\psi|P|\psi\rangle|^2]_{U_K} = (K^2+K)/(M^2+M)$ (which for large K and M is close to the naive estimate $[|\langle\psi|P|\psi\rangle|^2]_{U_K} \approx [\langle\psi|P|\psi\rangle]_{U_K}^2 = K^2/M^2$). Thus,

$$b(P_1, P_1) = \frac{K^2-1}{M^2+M}.$$

With $b(P_1,P_1) = \alpha + \beta$ from Eq. (23) this yields the second equation,

$$\alpha + \beta = \frac{K^2-1}{M^2+M}. \qquad (25)$$

Solving Eqs. (24) and (25) for α and β, and inserting the solution into Eq. (23) produces

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

and, by Eq. (22),

$$[\|D\|_F^2]_{U_K} = \frac{1-1/K^2}{M^2-1}\sum_{ij}\left(\mathrm{tr}W_{ij}{}^\dagger W_{ij} - \frac{1}{M}|\mathrm{tr}W_{ij}|^2\right). \qquad (26)$$

In general, not much is given away if instead of this exact result we use an upper bound for $[\|D\|_F^2]_{U_K}$ that we obtain by using $(1-1/K^2)/(M^2-1) \leqslant 1/M^2$ and by omitting the negative terms $-|\mathrm{tr}W_{ij}|^2/M$ in the sum. Then

$$[\|D\|_F^2]_{U_K} \leqslant \frac{1}{M^2}\sum_{ij}\mathrm{tr}W_{ij}{}^\dagger W_{ij} = \mathrm{tr}\left(\sum_j A_j \frac{1}{M} A_j{}^\dagger \sum_i A_i \frac{\mathbf{1}}{M} A_i{}^\dagger\right),$$

where we cyclically permuted operators under the trace to obtain the last equality. We realize that the argument of the trace is simply $\mathcal{N}(\pi_Q)^2$, with $\pi_Q = \mathbf{1}_Q/M$ being the homogeneously distributed density operator on H_Q. This yields the rather simple upper bound

$$[\|D\|_F^2]_{U_K} \leqslant \|\mathcal{N}(\pi_Q)\|_F^2. \qquad (27)$$

By relation (20) this means

$$[F_e(C,\mathcal{N})]_{U_K} \geqslant 1 - \sqrt{KN}\|\mathcal{N}(\pi_Q)\|_F. \qquad (28)$$

Before discussing this result let us generalize it to the unitarily invariant ensemble $U_K(V)$ of K-dimensional codes in a subspace $V \subset H_Q$ ($\dim V \geqslant K$). Here the average is given by

$$[A(C)]_{U_K(V)} := \int_{\mathbf{U}(V)} d\mu_V(U) A(UC_0),$$

where μ_V is the normalized Haar measure on the group $\mathbf{U}(V)$ of unitaries on the subspace V. Up to the fact that now the role of H_Q is taken over by the linear space V nothing has changed compared to the situation before. Hence, the derivation given above for the ensemble U_K applies to the ensemble $U_K(V)$ as well, showing that

$$[\|D\|_F^2]_{U_K(V)} \leqslant \|\mathcal{N}(\pi_V)\|_F^2, \qquad (29)$$

and consequently,

$$[F_e(C,\mathcal{N})]_{U_K(V)} \geqslant 1 - \sqrt{KN}\|\mathcal{N}(\pi_V)\|_F, \qquad (30)$$

where $\pi_V = \Pi_V / \dim V$.

C. Discussion

It is instructive to discuss the just obtained lower bounds for the case of unital noise, which by definition leaves the homogeneously distributed state π_Q invariant, $\mathcal{N}(\pi_Q) = \pi_Q$. A unital operation is, for instance, the process where arbitrary unitary operations $U_1, \ldots, U_N$ are applied to the system Q with probabilities $p_1, \ldots, p_N$. For unital noise $\|\mathcal{N}(\pi_Q)\|_F^2 = \|\pi_Q\|_F^2 = \mathrm{tr}(\pi_Q^2) = 1/M$. Hence, by the lower bound (28),

$$[F_e(C,\mathcal{N})]_{U_K} \geqslant 1 - \sqrt{\frac{KN}{M}}.$$

This means that on almost all codes C of the ensemble U_K the unital noise $\mathcal{N}$ can be almost perfectly corrected, provided that

$$KN \ll M.$$

Recalling that K is the code dimension, N is the number of Kraus operators in an operator-sum representation of $\mathcal{N}$, and M is the dimension of H_Q, we recover that randomly chosen codes attain the quantum Hamming bound [3].

The requirement $K \ll M/N$ suggests that $\log_2 M - \log_2 N$ is a lower bound of the capacity $Q(\mathcal{N})$, what we will now formally derive. To this end, we consider the n-fold replicated noise $\mathcal{N}^{\otimes n}$, and study the averaged entanglement fidelity of the code ensemble U_{K_n}, where we chose the code dimension to be $K_n = \lfloor 2^{nR} \rfloor$ for some positive R. $\mathcal{N}^{\otimes n}$ operates on states in $H_Q^{\otimes n}$ and has N^n operation elements. With $\mathcal{N}$ also $\mathcal{N}^{\otimes n}$ is unital, thus $\|\mathcal{N}^{\otimes n}(\rho_{Q_n})\|_F^2 = M^{-n}$, and by Eq. (28)

$$[F_e(C,\mathcal{N}^{\otimes n})]_{U_{K_n}} \geq 1 - \left(\frac{2^R N}{M}\right)^{n/2}.$$

In the limit $n \to \infty$ the right-hand side converges to unity if $R < \log_2 M - \log_2 N$. Since $\lim_{n\to\infty}(1/n)\log_2 K_n = R$ this implies that all rates below $\log_2 M - \log_2 N$ are achievable and so, by the definition of quantum capacity in Sec. II C,

$$Q(\mathcal{N}) \geq \log_2 M - \log_2 N.$$

We note that since $\mathcal{N}$ is unital $\log_2 M = S(\pi_Q) = S(\mathcal{N}(\pi_Q))$. Now, if we could identify the second term, $\log_2 N$, with the entropy exchange $S_e(\pi_Q, \mathcal{N})$ we would obtain that the lower bound $\log_2 M - \log_2 N$ is just the coherent information $I(\pi_Q, \mathcal{N})$, in accordance to the capacity formula. However, this is the case only for a special kind of unital operation. $\mathcal{N}$ must have a Kraus representation with operation elements $A_1, \ldots, A_N$ such that $\mathrm{tr} A_j^\dagger A_i = 0$ for $i \neq j$, and $(1/M)\mathrm{tr} A_i^\dagger A_i = \mathrm{const} = 1/N$. Then by Schumacher's relation indeed

$$S_e(\pi_Q, \mathcal{N}) = S(\{\mathrm{tr} A_i \pi_Q A_j^\dagger\}_{i,j=1,\ldots N}) = S(\mathbf{1}_N/N) = \log_2 N.$$

The first condition is actually no restriction, since a nondiagonal representation $B_1, \ldots, B_N$ with $\mathrm{tr} B_i^\dagger B_j \neq 0$ can always be unitarily transformed to a diagonal one (cf. footnote 1). The second condition demands that, roughly speaking, different kinds of errors appear with equal probability. In the end, this ensures that by the estimation $\|D\|_{\mathrm{tr}} \leq \sqrt{KN}\|D\|_F$ not much is lost and therefore the lower bound (28) is good.

To recapitulate, for unital noise $\mathcal{N}$ the lower bounds for the ensemble averaged code fidelities immediately make evident that the quantum Hamming bound is attainable by random codes. Moreover, if the noise $\mathcal{N}$ satisfies the condition of equally probable errors as specified above we can establish

$$Q_{\mathcal{N}} \geq I(\pi_Q, \mathcal{N}). \tag{31}$$

V. ERROR CORRECTION IN SELECTIVE NOISE

The hitherto presented analysis is restricted to trace-preserving noise operations. Here we will extend the considerations of the preceding sections to the case of trace-decreasing noise, which we find to be convenient in later use. First, we define channel fidelity and entanglement fidelity for a trace-decreasing channel. Within this definitions we will then generalize the lower bound (9) and the result (30) on the ensemble averaged code fidelity.

A. Fidelities for trace-decreasing channels

For a (possibly) trace-decreasing operation $\mathcal{N}$ on a system Q we define the channel fidelity with respect to a state ρ_Q as

$$F_{\mathrm{ch}}(\rho_Q, \mathcal{N}) \coloneqq \mathrm{tr}\mathcal{N}(\rho_Q) F\left(\rho_Q, \frac{\mathcal{N}(\rho_Q)}{\mathrm{tr}\mathcal{N}(\rho_Q)}\right), \tag{32}$$

where $F(\rho, \sigma)$ is the usual two-state fidelity. The definition deviates from the standard one by a factor $\mathrm{tr}\mathcal{N}(\rho_Q)$. This makes sense, when one interprets a trace-decreasing $\mathcal{N}$ as a selective operation that selects individual elements of the initial ensemble ρ_Q with probability $\mathrm{tr}\mathcal{N}(\rho_Q)$ [15]. Consequently, in order that $F_{\mathrm{ch}}(\rho_Q, \mathcal{N})$ is close to unity not only the selected final state $\mathcal{N}(\rho_Q)/\mathrm{tr}\mathcal{N}(\rho_Q)$ must be close to ρ_Q, but also the selection probability must be close to unity.

We define the entanglement fidelity of $\mathcal{N}$ with respect to ρ_Q as

$$F_e(\rho_Q, \mathcal{N}) \coloneqq F_{\mathrm{ch}}(\psi_{RQ}, \mathcal{I}_R \otimes \mathcal{N}) = \langle \psi_{RQ} | (\mathcal{I}_R \otimes \mathcal{N})(\psi_{RQ}) | \psi_{RQ} \rangle, \tag{33}$$

where ψ_{RQ} purifies ρ_Q. Note that if $\mathcal{N}$ is trace-decreasing also its extension $\mathcal{I}_R \otimes \mathcal{N}$ is trace-decreasing, in which case F_{ch} means the just defined fidelity (32). Repeating the arguments of Schumacher [19], it is not difficult to see that also the entanglement fidelity of a trace-decreasing $\mathcal{N}$ can be expressed by its Kraus operators $A_1, \ldots, A_N$ of $\mathcal{N}$ by the usual formula

$$F_e(\rho_Q, \mathcal{N}) = \sum_{i=1}^{N} |\mathrm{tr}\rho_Q A_i|^2. \tag{34}$$

A simple but important consequence of this relation is the following: Let for a subset $\tilde{N} \subset \{1, \ldots, N\}$ a quantum operation $\tilde{\mathcal{N}}$ be defined by

$$\tilde{\mathcal{N}}(\rho_Q) \coloneqq \sum_{i \in \tilde{N}} A_i \rho_Q A_i^\dagger,$$

which we will call a reduction of the operation $\mathcal{N}$. Then by Eq. (34),

$$F_e(\rho, \mathcal{N}) \geq F_e(\rho, \tilde{\mathcal{N}}).$$

Further, since for any operation $\mathcal{R}$ on Q clearly $\mathcal{R} \circ \tilde{\mathcal{N}}$ is a reduction of $\mathcal{R} \circ \mathcal{N}$, we conclude that for any code C

$$F_e(C, \mathcal{N}) \geq F_e(C, \tilde{\mathcal{N}}), \tag{35}$$

where the code entanglement fidelity for a trace-decreasing $\mathcal{N}$ is defined as for trace-preserving noise by $F_e(C, \mathcal{N}) \coloneqq \max_{\mathcal{R}} F_e(\pi_C, \mathcal{R} \circ \mathcal{N})$.

B. Lower bound for code entanglement fidelity

Let $\mathcal{N}$ be a noise operation on Q that can be represented by Kraus operators $A_1, \ldots, A_N$. The entanglement fidelity of a K-dimensional code C satisfies

$$F_e(C, \mathcal{N}) \geq \mathrm{tr}\mathcal{N}(\pi_C) - \|D\|_{\mathrm{tr}}, \tag{36}$$

where $\pi_C = P/K$ is the homogeneously distributed code state, and the operator D is defined exactly as in Eq. (10).

This relation generalizes the lower bound (9) to the case of a trace-decreasing operation $\mathcal{N}$. Its proof given in Appendix C is almost identical to the one of Eq. (9) in Sec. III.

C. Unitarily invariant code ensembles

We consider the ensemble $U_K(V)$ of all K-dimensional codes in a subspace V of H_Q which we introduced in Sec. IV B. According to the lower bound (36), the averaged code entanglement fidelity under a (possibly trace-decreasing) noise $\mathcal{N}$ with Kraus operators $A_1, \ldots, A_N$ satisfies

$$[F_e(C,\mathcal{N})]_{U_K(V)} \geqslant [\mathrm{tr}\mathcal{N}(\pi_C)]_{U_K(V)} - [\|D\|_{\mathrm{tr}}]_{U_K(V)},$$

where D is given by Eq. (10). As shown in Sec. IV,

$$[\|D\|_{\mathrm{tr}}]_{U_K(V)} \leqslant \sqrt{KN[\|D\|_F^2]_{U_K(V)}} \leqslant \sqrt{KN}\|\mathcal{N}(\pi_V)\|_F,$$

where $\pi_V = \Pi_V/\dim V$ [cf. Eqs. (19) and (29)]. Furthermore, we will show below that

$$[\mathrm{tr}\mathcal{N}(\pi_C)]_{U_K(V)} = \mathrm{tr}\mathcal{N}(\pi_V), \tag{37}$$

and thus obtain

$$[F_e(C,\mathcal{N})]_{U_K(V)} \geqslant \mathrm{tr}\mathcal{N}(\pi_V) - \sqrt{KN}\|\mathcal{N}(\pi_V)\|_F. \tag{38}$$

We show Eq. (37) by again referring to unitarian invariants: Let a linear form a on the set of all linear operators on H_Q be defined by

$$W \mapsto a(W) \coloneqq \frac{1}{K}[\mathrm{tr}PWP]_{U_K(V)},$$

where, as always, $P = \Pi_C$. Since the codes C are subspaces of V it is clear that $a(W) = a(\Pi_V W \Pi_V)$. Further, the unitarian invariance of the code ensemble entails $a(W) = a(UWU^\dagger)$ for all unitary transformations U on H_Q with $U(V) = V$. It follows that a must be proportional to the fundamental invariant linear form on V, $W \mapsto \mathrm{tr}(\Pi_V W)$. From $a(\Pi_V) = 1$ we can then deduce that $a(W) = \mathrm{tr}(\pi_V W)$. To conclude the proof of Eq. (37) note that

$$\begin{aligned}[\mathrm{tr}\mathcal{N}(\pi_C)]_{U_K(V)} &= \sum_{i=1}^{N} \frac{1}{K}[\mathrm{tr}A_i P A_i^\dagger]_{U_K(V)} \\ &= \sum_{i=1}^{N} a(A_i^\dagger A_i) = \mathrm{tr}\sum_{i=1}^{N} A_i \pi_V A_i^\dagger = \mathrm{tr}\mathcal{N}(\pi_V).\end{aligned}$$

VI. LOWER BOUNDS FOR THE QUANTUM CAPACITY

In this section we will prove that the quantum capacity $Q(\mathcal{N})$ of a general trace-preserving channel $\mathcal{N}$ satisfies

$$Q(\mathcal{N}) \geqslant I(\pi_V, \mathcal{N}), \tag{39}$$

where π_V is the homogeneously distributed density on an arbitrary subspace V of the system's Hilbert space H_Q. We will then use the lemma of BSST in order to establish the regularized coherent information $I_r(\mathcal{N})$ [cf. Eq. (8)] as a lower bound of $Q(\mathcal{N})$.

We first prove inequality (39) for the case $V = H_Q$ or $\pi_V = \pi_Q$. A strategy of proof becomes evident when we look back at Sec. IV C, where we showed $Q(\mathcal{N}) \geqslant I(\pi_Q, \mathcal{N})$ under the conditions of (i) equally probable errors and (ii) unitality: $\mathcal{N}(\pi_Q) = \pi_Q$.

For general noise $\mathcal{N}$ these two requirements are certainly not fulfilled, not even approximately. However, since our concern is the channel capacity of $\mathcal{N}$ we are free to consider the n-times replicated channel $\mathcal{N}^{\otimes n}$. For large n it is possible to arrange for the conditions (i) and (ii) in an approximate sense by, as it will turn out, only minor modifications of the operation $\mathcal{N}^{\otimes n}$. Following Shor [13], we (a) reduce the operation $\mathcal{N}^{\otimes n}$ to an operation $\mathcal{N}_n$ that consists only of the *typical* Kraus operators of $\mathcal{N}_n$ (cf. Sec. VI A 2). Thereafter we (b) project on the typical subspace T_n of $\mathcal{N}(\pi_Q)$ in $H_Q^{\otimes n}$ (cf. Sec. VI A 3). The purpose of reduction (a) is to approximately establishes a situation of equally probable errors (i). The second step allows to restrict the output Hilbert space of $\mathcal{N}_n$ to the typical subspace T_n, on which the density $\mathcal{N}_n(\pi_{Q_n})$ is approximately homogeneously distributed. This establishes a situation similar to (ii). After having proven Eq. (39) for $V = H_Q$ in Sec. VI B 1, we will argue in Sec. VI B 2 that its generalization is trivially obtained by restricting the original input Hilbert space H_Q of $\mathcal{N}$ to a subspace $V \subset H_Q$. Finally, in Sec. VI B 3 we use the lemma of BSST in order to show that $Q(\mathcal{N}) \geqslant I_r(\mathcal{N})$.

A. Reduction of the noise $\mathcal{N}^{\otimes n}$

Both, typical Kraus operators and typical subspaces are defined on the basis of typical sequences (see, e.g. [16]). We briefly recall their definition and state two basic facts that are important for our purposes.

1. Typical sequences

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables that assume values $A_1, \ldots, A_N$ with probabilities $p_1, p_2, \ldots, p_N$. We denote the probability distribution by $\mathcal{A}$. Its Shannon entropy is $H(\mathcal{A}) = -\Sigma_{i=1}^N p_i \log_2 p_i$. Let ε be some positive number. A sequence $\mathbf{A} = A_{j_1}, A_{j_2}, \ldots, A_{j_n}$ is defined to be ε typical if its probability of appearance $p_{\mathbf{A}} = p_{j_1} p_{j_2} \cdots p_{j_n}$ satisfies

$$2^{-n[H(\mathcal{A})+\varepsilon]} \leqslant p_{\mathbf{A}} \leqslant 2^{-n[H(\mathcal{A})-\varepsilon]}.$$

Below we will make use of the following two facts: (1) The number of all ε typical sequences $N_{\varepsilon,n}$ is less than $2^{n(H(\mathcal{A})+\varepsilon)}$ and (2) the probability $P_{\varepsilon,n}$ that a random sequence of length n is ε typical satisfies $1 - P_{\varepsilon,n} \leqslant 2e^{-n\psi(\varepsilon)}$, where $\psi(\varepsilon)$ is a positive number independent of n. Proofs can be found in Appendix D.

2. Restriction to typical Kraus operators

Let a trace-preserving noise $\mathcal{N}$ on Q be represented by Kraus operators $A_1, \ldots, A_N$. Without loss of generality we can assume that the A_i are diagonal in the sense that

$\mathrm{tr} A_j^\dagger A_i = 0$ for $i \neq j$.[1] We define the probability p_i of the Kraus operator A_i as

$$p_i := \frac{1}{M} \mathrm{tr} A_i^\dagger A_i, \tag{40}$$

and we denote the corresponding probability distribution by $\mathcal{A}_\mathcal{N}$. The definition makes sense, because the p_i are positive and, as a consequence of the trace preservation of $\mathcal{N}$, sum up to unity.

The n-times replicated noise $\mathcal{N}^{\otimes n}$ can be represented by N^n Kraus operators

$$A_{j_1} \otimes A_{j_2} \otimes \ldots \otimes A_{j_n} \equiv A_{\mathbf{j}},$$

where $j_\nu = 1, \ldots, N$ and $\mathbf{j} = (j_1, j_2, \ldots, j_n)$. By the diagonality of the operators A_i of $\mathcal{N}$ also the operators $A_{\mathbf{j}}$ of $\mathcal{N}^{\otimes n}$ are diagonal, and the probability $p_{\mathbf{j}}$ of the element $A_{\mathbf{j}}$ appears to be the product of the probabilities p_{j_ν} of its constituent elements A_{j_ν},

$$p_{\mathbf{j}} = \frac{1}{M^n} \mathrm{tr} A_{\mathbf{j}}^\dagger A_{\mathbf{j}} = \frac{1}{M^n} \mathrm{tr}(A_{j_1}^\dagger A_{j_1}) \ldots \mathrm{tr}(A_{j_n}^\dagger A_{j_n}) = p_{j_1} \ldots p_{j_n}.$$

In other words, the Kraus operators $A_{\mathbf{j}}$ of $\mathcal{N}^{\otimes n}$ are sequences of length n in which symbols A_i of an alphabet $A_1, \ldots, A_N$ appear according to the distribution $\mathcal{A}_\mathcal{N}$. Hence we are in the domain of classical random sequences and can employ the notions of Sec. VI A 1 to define the ε-typical operation $\mathcal{N}_{\varepsilon,n}$ of $\mathcal{N}^{\otimes n}$ by

$$\rho \mapsto \mathcal{N}_{\varepsilon,n}(\rho) := \sum_{A_{\mathbf{j}} \varepsilon\text{-typical}} A_{\mathbf{j}} \rho A_{\mathbf{j}}^\dagger,$$

i.e., $\mathcal{N}_{\varepsilon,n}$ consists only of the ε-typical Kraus operators of $\mathcal{N}$. In general, this strongly reduces the number of Kraus operators from N^n to

$$N_{\varepsilon,n} \leq 2^{n[H(\mathcal{A}_\mathcal{N})+\varepsilon]}$$

[cf. Sec. VI A 1, property (1)]. It is time to remark that $H(\mathcal{A}_\mathcal{N})$ is nothing other than the entropy exchange $S_e(\pi_Q, \mathcal{N})$, such that the last relation becomes

$$N_{\varepsilon,n} \leq 2^{n[S_e(\pi_Q,\mathcal{N})+\varepsilon]}. \tag{41}$$

To see this, we notice that $H(\mathcal{A}_\mathcal{N})$ equals the von Neumann entropy of an N-dimensional diagonal density matrix W with elements $W_{ii} = (1/M) \mathrm{tr} A_i^\dagger A_i$. Since we are working in a diagonal operator-sum representation, this actually means that $W_{ij} = (1/M) \mathrm{tr} A_j^\dagger A_i = \mathrm{tr} A_i \pi_Q A_j^\dagger$, where $\pi_Q = \mathbf{1}_Q / M$. By Schumacher's representation of the entropy exchange we thus realize that $H(\mathcal{A}_\mathcal{N}) = S_e(\pi_Q, \mathcal{N})$.

Despite its strongly reduced number of Kraus operators, in average the operation $N_{\varepsilon,n}$ does not much reduce the trace when n becomes large. This can be seen by the selection probability $\mathrm{tr}\mathcal{N}_{\varepsilon,n}(\pi_{Q_n})$ of the homogeneously distributed state $\pi_{Q_n} = \mathbf{1}_{Q_n} / M^n$. A lower bound can be derived by observing that

$$\mathrm{tr}\mathcal{N}_{\varepsilon,n}(\pi_{Q_n}) = \frac{1}{M^n} \sum_{A_{\mathbf{j}}\ \varepsilon\text{-typical}} \mathrm{tr} A_{\mathbf{j}} A_{\mathbf{j}}^\dagger = \sum_{A_{\mathbf{j}}\ \varepsilon\text{-typical}} p_{\mathbf{j}}$$

is the probability that an operation element $A_{\mathbf{j}}$ of $\mathcal{N}^{\otimes n}$ is ε-typical. Thus, by Sec. VI A 1, property (2),

$$\mathrm{tr}\mathcal{N}_{\varepsilon,n}(\pi_{Q_n}) \geq 1 - 2e^{-n\psi_1(\varepsilon)}, \tag{42}$$

where $\psi_1(\varepsilon)$ is a positive number independent of n.

3. Projection on typical subspace

We will further reduce the operation $\mathcal{N}^{\otimes n}$ by letting $\mathcal{N}_{\varepsilon,n}$ follow a projection on the ε-typical subspace $T_{\varepsilon,n} \subset H_Q^{\otimes n}$ of the density $\mathcal{N}(\pi_Q)$. The benefit of this procedure is that the so obtained operation $\tilde{\mathcal{N}}_{\varepsilon,n}$ maps π_{Q_n} to an almost homogeneously distributed state on $T_{\varepsilon,n}$, and thus establishes a situation similar to (ii) in Sec. V.

The ε-typical subspace $T_{\varepsilon,n} \subset H_Q^{\otimes n}$ of $\sigma \equiv \mathcal{N}(\pi_Q)$ is spanned by the ε-typical eigenvectors of $\sigma^{\otimes n}$ [16]. These are precisely the eigenvectors $v_{\mathbf{l}}$ with eigenvalues $p_{\mathbf{l}}$ satisfying

$$2^{-n\{S(\mathcal{N}(\pi_Q))+\varepsilon\}} \leq p_{\mathbf{l}} \leq 2^{-n\{S(\mathcal{N}(\pi_Q))-\varepsilon\}}.$$

The dimension of $T_{\varepsilon,n}$ obeys

$$\dim T_{\varepsilon,n} \leq 2^{n\{S(\mathcal{N}(\pi_Q))+\varepsilon\}}. \tag{43}$$

If n is large, almost the entire weight of $\sigma^{\otimes n}$ lies in the ε-typical subspace: Let $\Pi_{\varepsilon,n}$ be the projection on $T_{\varepsilon,n}$, then

$$\mathrm{tr}\Pi_{\varepsilon,n}\sigma^{\otimes n} = \sum_{\mathbf{l}:|v_{\mathbf{l}}\rangle\ \varepsilon\text{-typical}} p_{\mathbf{l}},$$

which in the notions of Sec. VI A 1 is the probability that an eigenvalue $|v_{\mathbf{l}}\rangle = |v_{l_1}\rangle |v_{l_2}\rangle \ldots |v_{l_M}\rangle$ is ε-typical. Thus, by the second property in Sec. VI A 1,

$$\mathrm{tr}\Pi_{\varepsilon,n}\sigma^{\otimes n} \geq 1 - 2e^{-n\psi_2(\varepsilon)}, \tag{44}$$

where $\psi_2(\varepsilon)$ is a positive number independent of n.

We define the ε-reduced operation of $\mathcal{N}^{\otimes n}$ by

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

where the operation $\mathcal{P}_{\varepsilon,n}$ describes the projective measurement on $T_{\varepsilon,n}$,

$$\mathcal{P}_{\varepsilon,n} : \rho \mapsto \Pi_{\varepsilon,n} \rho \Pi_{\varepsilon,n},$$

and $\mathcal{N}_{\varepsilon,n}$ is the ε-typical operation of $\mathcal{N}^{\otimes n}$ as defined in the previous subsection.

4. Properties of the ε-reduced operation $\tilde{\mathcal{N}}_{\varepsilon,n}$

The ε-reduced operation $\tilde{\mathcal{N}}_{\varepsilon,n}$ can be represented by Kraus operators of the form $\Pi_{\varepsilon,n} A_{\mathbf{j}}$, where $A_{\mathbf{j}}$ is an ε-typical operation element of $\mathcal{N}^{\otimes n}$. Their total number $\tilde{N}_{\varepsilon,n}$ is therefore bounded by

[1] For arbitrary operation elements $B_1, \ldots, B_N$ of $\mathcal{N}$ 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 [16], the operators $A_m := \sum_j U_{jm}^\dagger B_j$ equivalently represent $\mathcal{N}$. It is readily verified that $\mathrm{tr} A_l^\dagger A_m = 0$ for $l \neq m$.

$$\widetilde{N}_{\varepsilon,n} = N_{\varepsilon,n} \leqslant 2^{n[S_e(\pi_Q,\mathcal{N})+\varepsilon]}.$$

Besides the number of Kraus operators, the two other crucial figures are $\mathrm{tr}\widetilde{\mathcal{N}}_{\varepsilon,n}(\pi_{Q_n})$ and $\|\widetilde{\mathcal{N}}_{\varepsilon,n}(\pi_{Q_n})\|_F^2$ [cf. relation (38)]. In Appendix E we derive the followings bounds:

$$\mathrm{tr}\widetilde{\mathcal{N}}_{\varepsilon,n}(\pi_{Q_n}) \geqslant 1 - 4e^{-n\psi_3(\varepsilon)},$$

$$\|\widetilde{\mathcal{N}}_{\varepsilon,n}(\pi_{Q_n})\|_F^2 \leqslant 2^{-n\{S(\mathcal{N}(\pi_Q))-3\varepsilon\}},$$

where $\psi_3(\varepsilon)$ is a positive number independent of n. Finally, we note that for any code $C \subset H_Q^{\otimes n}$

$$F_e(C,\mathcal{N}^{\otimes n}) \geqslant F_e(C,\mathcal{N}_{\varepsilon,n}) \geqslant F_e(C,\widetilde{\mathcal{N}}_{\varepsilon,n}).$$

The first inequality holds because $\mathcal{N}_{\varepsilon,n}$ is a reduction of $\mathcal{N}^{\otimes n}$ and the second one is explained by the fact that $\widetilde{\mathcal{N}}_{\varepsilon,n}$ results from postprocessing of $\mathcal{N}_{\varepsilon,n}$ by $P_{\varepsilon,n}$, which cannot increase the code entanglement fidelity [cf. Eq. (5)].

B. Lower bounds for $Q(\mathcal{N})$

Lower bounds of the quantum capacity $Q(\mathcal{N})$ are given by the achievable rates of $\mathcal{N}$. Finding out whether a rate R is achievable or not requires to investigate the code entanglement fidelities $F_e(C_n,\mathcal{N}^{\otimes n})$ for suitable codes $C_n \subset H_Q^{\otimes n}$ (cf. Sec. II C). Our working hypothesis is that no special care has to be taken in choosing C_n. Rather, we suppose that randomly chosen codes in general do provide high achievable rates and therefore will study the averaged entanglement fidelity of the code ensembles introduced in Sec. IV B.

1. $Q(\mathcal{N}) \geqslant I(\pi_Q,\mathcal{N})$

We begin with the average code fidelity $[F_e(C_n,\mathcal{N}^{\otimes n})]_{U_K}$ of the unitarily invariant ensemble U_{K_n}. As in Sec. IV C, we chose the code dimension to be

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

meaning that $R = \lim_{n\to\infty}(1/n)\log_2 K_n$ is the asymptotic rate. By relation (38) and the results of the previous subsection we immediately find

$$[F_e(C,\mathcal{N}^{\otimes n})]_{U_{K_n}} \geqslant [F_e(C,\widetilde{\mathcal{N}}_{\varepsilon,n})]_{U_{K_n}} \geqslant 1 - \alpha_n - \beta_n \quad (45)$$

with coefficients

$$\alpha_n = 1 - \mathrm{tr}\widetilde{\mathcal{N}}_{\varepsilon,n}(\pi_{Q_n}) \;\leqslant\; 4e^{-n\psi_3(\varepsilon)},$$

$$\beta_n = \sqrt{K_n \widetilde{N}_{\varepsilon,n}}\|\widetilde{\mathcal{N}}_{\varepsilon,n}(\pi_{Q_n})\|_F \;\leqslant\; 2^{(n/2)\{R+S_e(\pi_Q,\mathcal{N})-S(\mathcal{N}(\pi_Q))+4\varepsilon\}}.$$

Clearly, for all $\varepsilon > 0$, the right-hand side of inequality (45) converges to unity in the limit $n \to \infty$ if the asymptotic rate R obeys

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

That is, all rates R below $I(\pi_Q,\mathcal{N})$ are achievable and therefore $I(\pi_Q,\mathcal{N})$ is a lower bound of the capacity $Q(\mathcal{N})$.

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

Let V be an arbitrary linear subspace of the system's Hilbert space H_Q, and let $\pi_V = \Pi_V/\dim V$. In short, the coherent information $I(\pi_V,\mathcal{N})$ can be established as a lower bound of $Q(\mathcal{N})$ in exactly the same way as before $I(\pi_Q,\mathcal{N})$ if we consider instead of $\mathcal{N}$ the operation $\mathcal{L}$ that is defined as the restriction of $\mathcal{N}$ to states ρ_V on a reduced input Hilbert space $V \subset H_Q$. For the sake of completeness, we briefly repeat the arguments.

This starts with reducing $\mathcal{L}^{\otimes n}$ to an ε-typical $\mathcal{L}_{\varepsilon,n}$ as described in Sec. VI A 2: The reduced input Hilbert space V of $\mathcal{L}$ entails that now the probability p_i of a Kraus operator A_i has to be defined as

$$p_i = \frac{1}{L}\mathrm{tr}\Pi_V A_i^\dagger A_i \Pi_V, \quad (46)$$

where $L = \dim V$, and Π_V is the projection on V. Here it is assumed that the operators $A_1,\ldots,A_N$ are diagonal with respect to V, i.e., $\mathrm{tr}\Pi_V A_i^\dagger A_j \Pi_V = 0$ for $i \neq j$. Accordingly, the probability of a $A_{\mathbf{j}} = A_{j_1} \otimes \ldots \otimes A_{j_n}$ is

$$p_{\mathbf{j}} = \frac{1}{L^n}\mathrm{tr}\Pi_V^{\otimes n} A_{\mathbf{j}}^\dagger A_{\mathbf{j}} \Pi_V^{\otimes n} = p_{j_1} \ldots p_{j_n}.$$

As before, $\mathcal{L}_{\varepsilon,n}$ is defined to consist only of the ε-typical $A_{\mathbf{j}}$. Its number $L_{\varepsilon,n}$ is bounded by $2^{n(H+\varepsilon)}$, with H being the Shannon entropy of the normalized probability distribution (46). Therefore, H coincides with the von Neumann entropy of a diagonal density matrix W with entries

$$W_{ij} = \frac{1}{L}\mathrm{tr}\Pi_V A_i^\dagger A_j \Pi_V = \mathrm{tr}A_j \frac{\Pi_V}{L} A_i^\dagger.$$

By Schumacher's representation of the entropy exchange we obtain $H = S_e(\pi_V,\mathcal{N})$, where $\pi_V = \Pi_V/L$.

The next step is to further reduce $\mathcal{L}_{\varepsilon,n}$ to an operation $\widetilde{\mathcal{L}}_{\varepsilon,n}$ by projecting the output of $\mathcal{L}_{\varepsilon,n}$ on the typical subspace $T_{\varepsilon,n} \subset H_Q^{\otimes n}$ of the density $\mathcal{L}(\pi_V) = \mathcal{N}(\pi_V)$. This follows precisely Sec. VI A 3 with $\sigma = \mathcal{N}(\pi_Q)$ replaced by $\sigma = \mathcal{N}(\pi_V)$. The resulting $\widetilde{\mathcal{L}}_{\varepsilon,n}$ is characterized by (cf. Sec. VI A 4)

$$\widetilde{L}_{\varepsilon,n} \leqslant 2^{n[S_e(\pi_V,\mathcal{N})+\varepsilon]},$$

$$\mathrm{tr}\widetilde{\mathcal{L}}_{\varepsilon,n} \;\geqslant\; 1 - 4e^{-n\psi_3(\varepsilon)},$$

$$\|\widetilde{\mathcal{L}}_{\varepsilon,n}\|_F^2 \;\leqslant\; 2^{-n\{S(\mathcal{N}(\pi_V))-3\varepsilon\}},$$

$$F_e(C,\mathcal{L}^{\otimes n}) \;\geqslant\; F_e(C,\widetilde{\mathcal{L}}_{\varepsilon,n}),$$

where $\widetilde{L}_{\varepsilon,n}$ is the number of Kraus operators that is needed to represent $\widetilde{\mathcal{L}}_{\varepsilon,n}$. Thus, by inequality (38),

$$[F_e(C,\mathcal{L}^{\otimes n})]_{U_{K_n}(V^{\otimes n})} \geqslant 1 - \alpha_n - \beta_n,$$

where the coefficients α_n and β_n are as in the previous subsection, but with π_Q replaced by π_V. Since further $[F_e(C,\mathcal{N}^{\otimes n})]_{U_{K_n}(V^{\otimes n})} = [F_e(C,\mathcal{L}^{\otimes n})]_{U_{K_n}(V^{\otimes n})}$ we can thus conclude that all rates R below

$$S(\mathcal{N}(\pi_V)) - S_e(\pi_V,\mathcal{N}) \equiv I(\pi_V,\mathcal{N})$$

are achievable by $\mathcal{N}$, meaning that $Q(\mathcal{N}) \geqslant I(\pi_V,\mathcal{N})$.

3. $Q(\mathcal{N}) \geq I_r(\mathcal{N})$

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 $H_Q^{\otimes m}$. Clearly, this suffices to prove the regularized coherent information $I_r(\mathcal{N})$ (cf. Sec. II C) a lower bound of $Q(\mathcal{N})$.

The BSST lemma [11] 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 homogeneously distributed state on the frequency-typical subspace $T_{\varepsilon,n}^{(f)}$ of ρ. As a corollary, one obtains an analogous relation for the coherent information,

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

$T_{\varepsilon,n}^{(f)}$ is similar to the ordinary typical subspace $T_{\varepsilon,n}$ which we have used above. The difference is that for $T_{\varepsilon,n}^{(f)}$ 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 [12], 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 $T_{\varepsilon,n}^{(f)} \subset H_Q^{\otimes mn}$ of an arbitrary density ρ on $H_Q^{\otimes m}$. 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.

VII. CONCLUDING REMARKS

We expect that the lower bound (9) for the code entanglement fidelity is also useful for directly evaluating the error correcting capability of a particular code for a particular noise operation. In this case, there is no need to estimate the trace norm of the operator D by its Frobenius norm. The only reason why we used this in general rather poor estimate here is that it enabled us to perform the ensemble average.

The above proof of the direct coding theorem shows that a randomly chosen code of sufficiently large block-size is typically a good quantum error correcting code. Studying the properties of unitarily invariant code ensembles might be therefore always a good thing to do when general aspects of QEC are of concern.

Note added. We would like to mention the recent eprint of Hayden *et al.* [31], in which a similar proof of the direct coding theorem has been independently obtained.

ACKNOWLEDGMENTS

I am grateful to M. R. Zirnbauer for pointing out the use of group invariants. I would also like to thank the referee for suggesting an improved lower bound for the code entanglement fidelity.

APPENDIX A: FIDELITY RELATIONS

1. $F_e(\pi_C, \mathcal{E}) \leq F_{\text{av}}(C, \mathcal{E})$

The average fidelity of the code C with respect to noise $\mathcal{E}$ is defined as

$$F_{\text{av}}(C, \mathcal{E}) = \int_{\mathbf{U}(C)} d\mu_C(U) F_{\text{ch}}(U\psi_0 U^\dagger, \mathcal{E}),$$

where $F_{\text{ch}}(\rho, \mathcal{E}) = F(\rho, \mathcal{E}(\rho))$, ψ_0 is an arbitrary pure state in C, and μ_C is the normalized Haar measure on the group $\mathbf{U}(C)$ of unitaries on the code space C. For a complete ensemble $\psi_1, \ldots, \psi_K$ of orthogonal pure states in C, $K = \dim C$, we find

$$\begin{aligned} F_{\text{av}}(C, \mathcal{E}) &= \int_{\mathbf{U}(C)} d\mu_C(U) \frac{1}{K} \sum_{i=1}^{K} F_{\text{ch}}(U\psi_i U^\dagger, \mathcal{E}) \\ &\geq \int_{\mathbf{U}(C)} d\mu_C(U) F_e\left(\frac{1}{K} \sum_{i=1}^{K} U\psi_i U^\dagger, \mathcal{E}\right) = F_e(\pi_C, \mathcal{E}). \end{aligned}$$

The inequality follows from the general relation [16]

$$\sum_i p_i F_{\text{ch}}(\rho_i, \mathcal{E}) \geq F_e\left(\sum_i p_i \rho_i, \mathcal{E}\right). \tag{A1}$$

2. Subcodes with high minimum fidelity

Let C be a code of dimension K with entanglement fidelity

$$F_e(\pi_C, \mathcal{E}) = 1 - \varepsilon.$$

We will show that there is a subcode $\tilde{C}$ of C of dimension $\tilde{K} = \lfloor K/2 \rfloor$ with minimum fidelity

$$F_{\min}(\tilde{C}, \mathcal{E}) := \min_{|\psi\rangle \in C} F_{\text{ch}}(\psi, \mathcal{E}) \geq 1 - 2\varepsilon.$$

To this end, we recursively define a sequence of subspaces $C_0 \supset C_1 \supset \ldots \supset C_{K-1}$, and a corresponding sequence of code vectors $|\psi_0\rangle, |\psi_1\rangle, \ldots, |\psi_{K-1}\rangle$ as follows:

$$i=0: \quad C_0 \coloneqq C$$

$$|\psi_0\rangle \text{ is the vector of minimal fidelity in } C_0,$$

$$i>0: \quad C_i \coloneqq C_{i-1} \cap |\psi_{i-1}\rangle^{\perp}$$

$$|\psi_i\rangle \text{ is the vector of minimal fidelity in } C_i.$$

By construction, $\dim C_i = K-i$, and $F_{\min}(C_i, \mathcal{E}) = F(\psi_i, \mathcal{E}(\psi_i)) \equiv F_i$. It is also clear that the minimum vectors $|\psi_0\rangle, |\psi_1\rangle, \ldots |\psi_{K-1}\rangle$ form an orthonormal basis of C. Hence $\pi_C = (1/K)\Sigma_{i=0}^{K-1}\psi_i$, and, by relation (A1),

$$1-\varepsilon \leqslant \frac{1}{K}\sum_{i=0}^{K} F_i.$$

For any $0<t<K$ we therefore obtain

$$1-\varepsilon \leqslant \frac{1}{K}\sum_{i=0}^{K-1-t} F_i + \frac{1}{K}\sum_{i=K-t}^{K-1} F_i \leqslant \frac{K-t}{K} + \frac{t}{K}F_{K-t},$$

where the last inequality follows from $1 \geqslant F_0 \geqslant F_1 \geqslant \cdots \geqslant F_{K-1} \geqslant 0$ and

$$1-\varepsilon \leqslant \frac{K-t}{K} + \frac{t}{K}F_{K-t}$$

is equivalent to

$$1-\frac{K}{t}\varepsilon \leqslant F_{K-t},$$

meaning that subspace C_{K-t} of dimension t has minimum fidelity larger than $1-\varepsilon K/t$. Setting $t = \lfloor K/2 \rfloor$ completes the proof.

APPENDIX B: AVERAGE OF $|\langle\psi|P|\psi\rangle|^2$

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

$$[|\langle\psi|P|\psi\rangle|^2]_{U_K} = \frac{K^2+K}{M^2+M} \tag{B1}$$

(notations as in Sec. IV B). By definition,

$$[|\langle\psi|P|\psi\rangle|^2]_{U_K} = \int d\mu(U)|\langle\psi|UP_0U^{\dagger}|\psi\rangle|^2,$$

where the integral extends over $\mathbf{U}(H_Q)$ and P_0 is the projection on an arbitrarily chosen linear subspace $C_0 \subset H_Q$ of dimension K. We extend $|\psi\rangle \equiv |\psi_1\rangle$ to an orthonormal basis $|\psi_1\rangle, \ldots, |\psi_M\rangle$ of H_Q, and chose

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

Then

$$[|\langle\psi|P|\psi\rangle|^2]_{U_K} = \sum_{i,j=1}^{K}\int d\mu(U)|U_{1i}|^2|U_{1j}|^2,$$

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 [34], 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

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

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

$$\int d\mu(U)|U_{11}|^4 = \frac{2}{M^2+M},$$

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

which immediately leads to Eq. (B1).

APPENDIX C: LOWER BOUND FOR CODE ENTANGLEMENT FIDELITY

Without loss of generality we can describe a possibly trace-decreasing $\mathcal{N}$ as a unitary operation U_{QE} on QE which is followed by a projective measurement on E that may reduce the trace. That is, for a general state ρ_Q

$$\mathcal{N}(\rho_Q) = \mathrm{tr}_E(\mathbf{1}_Q \otimes P_W)U_{QE}\rho_Q \otimes \psi_E U_{QE}{}^{\dagger},$$

where ψ_E is a fixed initial pure state of E, and P_W projects on some subspace of H_E. Let again ψ_{RQ} be a purification of ρ_Q, $\rho_R = \mathrm{tr}_Q\psi_{RQ}$, and let a normalized pure state ψ'_{RQE} on RQE be defined by its state vector

$$|\psi'_{RQE}\rangle = \frac{1}{\sqrt{p}}(\mathbf{1}_{RQ} \otimes P_W)(\mathbf{1}_R \otimes U_{QE})|\psi_{RQ}\rangle \otimes |\psi_E\rangle,$$

where $p = \mathrm{tr}\mathcal{N}(\rho_Q)$. The state ψ'_{RQE} is purification of its properly normalized partial states ρ'_Q, ρ'_E, ρ'_{RQ}, and ρ'_{RE}. Note that $\mathcal{N}(\rho_Q) = p\rho'_{RE}$.

Precisely as in Sec. III it follows that there exists a recovery operation $\mathcal{R}$ on Q satisfying

$$F(\psi_{RQ}, \mathcal{I}_R \otimes \mathcal{R}(\rho'_{RQ})) \geqslant 1 - \|\rho'_{RE} - \rho_R \otimes \rho'_E\|_{\mathrm{tr}}.$$

By definition (33) of entanglement fidelity for trace-decreasing operations this immediately leads to

$$F_e(\rho_Q, \mathcal{R} \circ \mathcal{N}) \geqslant p - \|p\rho'_{RE} - p\rho_R \otimes \rho'_E\|_{\mathrm{tr}},$$

which generalizes relation (12).

Continuing in a similar manner as before in Sec. III, we consider $\rho_Q = \pi_C$ with the purification (15), and chose the unitary U_{QE} with projection P_W such that

$$(\mathbf{1}_Q \otimes P_W)U_{QE}|\psi_Q\rangle|1\rangle = \sum_{i=1}^{N} A_i|\psi_Q\rangle|i\rangle, \tag{C1}$$

where $|1\rangle \equiv |\psi_E\rangle, |2\rangle, \ldots, |N\rangle$ are again orthonormal vectors in H_E. Then, it is readily verified that

$$p\rho'_{RE} = \frac{1}{K}\sum_{ij=1}^{N}\sum_{l,m=1}^{K} \mathrm{tr}_Q(A_i|c_l^Q\rangle\langle c_m^Q|A_j{}^{\dagger})|c_l^R\rangle\langle c_m^R| \otimes |i\rangle\langle j|,$$

$$p\rho_R \otimes \rho'_E = \sum_{ij=1}^{N} \mathrm{tr}_Q(A_i\pi_C A_j{}^{\dagger})\rho_R \otimes |i\rangle\langle j|,$$

where $p = \mathrm{tr}\mathcal{N}(\pi_C)$, which precisely correspond to expressions (16) and (17). As in Sec. III we conclude that

$$F_e(\pi_C, \mathcal{R}\circ\mathcal{N}) \geqslant p - \|D\|_{\mathrm{tr}},$$

showing that $\mathrm{tr}\mathcal{N}(\pi_C) - \|D\|_{\mathrm{tr}}$ is indeed a lower bound of $F_e(C,\mathcal{N})$.

APPENDIX D: TYPICAL SEQUENCES

The first property follows from

$$1 = \sum_{\mathbf{A}} p_{\mathbf{A}} \geqslant \sum_{\mathbf{A}\ \varepsilon\text{-typical}} p_{\mathbf{A}} \geqslant N_{\varepsilon,n}2^{-n[H(\mathcal{A})+\varepsilon]}.$$

To prove the second property we first realize that by definition

$$\begin{aligned}P_{\varepsilon,n} &= \mathrm{Prob}(\text{“}A_{j_1},\ldots,A_{j_n}\text{ is }\varepsilon\text{-typical”})\\ &= \mathrm{Prob}[|-\log_2(p_{j_1}\cdots p_{j_n}) - nH(\mathcal{A})| \leqslant n\varepsilon]\\ &= \mathrm{Prob}\left\{\left|\sum_{l=1}^{n}[-\log_2 p_{j_l} - H(\mathcal{A})]\right| \leqslant n\varepsilon\right\}.\end{aligned}$$

The negative logarithms of the probabilities p_{j_l} can be understood as n independent random variables Y_l that assume values $-\log_2 p_1,\ldots,-\log_2 p_N$ with probabilities $p_1,\ldots,p_N$. Their mean is the Shannon entropy $H(\mathcal{A})$,

$$\mu = E(Y_1) = -\sum_{i=1}^{N} p_i \log_2 p_i = H(\mathcal{A}).$$

This means that

$$1 - P_{\varepsilon,n} = \mathrm{Prob}\left[\left|\sum_{l=1}^{n}(Y_l - \mu)\right| \geqslant n\varepsilon\right]$$

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 [33], according to which

$$\mathrm{Prob}\left[\left|\sum_{l=1}^{n}(Y_l - \mu)\right| \geqslant n\varepsilon\right] \leqslant 2e^{-n\psi(\varepsilon)},$$

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

APPENDIX E: BOUNDS FOR $\mathrm{tr}\tilde{N}_{\varepsilon,n}(\pi_{Q_N})$ AND $\|\tilde{N}_{\varepsilon,n}(\pi_{Q_n})\|_F^2$

It is convenient to introduce the complementary operation $\mathcal{M}_{\varepsilon,n}$ of $\mathcal{N}_{\varepsilon,n}$ by

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

The operation elements of $\mathcal{M}_{\varepsilon,n}$ are exactly the ε-“untypical” operation elements of $\mathcal{N}^{\otimes n}$. Then,

$$\begin{aligned}\mathrm{tr}\tilde{\mathcal{N}}_{\varepsilon,n}(\pi_{Q_n}) &= \mathrm{tr}\Pi_{\varepsilon,n}[\mathcal{N}^{\otimes n}(\pi_{Q_n}) - \mathcal{M}_{\varepsilon,n}(\pi_{Q_n})]\\ &\geqslant \mathrm{tr}\Pi_{\varepsilon,n}\mathcal{N}^{\otimes n}(\pi_{Q_n}) - \mathrm{tr}\mathcal{M}_{\varepsilon,n}(\pi_{Q_n}).\end{aligned} \tag{E1}$$

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

$$\mathrm{tr}\mathcal{M}(\rho) = \mathrm{tr}\Pi\mathcal{M}(\rho) + \mathrm{tr}(\mathbf{1}-\Pi)\mathcal{M}(\rho) \geqslant \mathrm{tr}\Pi\mathcal{M}(\rho).$$

The first term in Eq. (E1) can be bounded from below as

$$\begin{aligned}\mathrm{tr}\Pi_{\varepsilon,n}\mathcal{N}^{\otimes n}(\pi_{Q_n}) &= \mathrm{tr}\Pi_{\varepsilon,n}\mathcal{N}^{\otimes n}(\pi_Q^{\otimes n})\\ &= \mathrm{tr}\Pi_{\varepsilon,n}\mathcal{N}(\pi_Q)^{\otimes n} \geqslant 1 - 2e^{-n\psi_2(\varepsilon)},\end{aligned}$$

where we used inequality (44). The second term in Eq. (E1) obeys

$$\mathrm{tr}\mathcal{M}_{\varepsilon,n}(\pi_{Q_n}) = \mathrm{tr}\mathcal{N}^{\otimes n}(\pi_{Q_n}) - \mathrm{tr}\mathcal{N}_{\varepsilon,n}(\pi_{Q_n}) \leqslant 2e^{-n\psi_1(\varepsilon)},$$

by inequality (42). We thus find

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

when $\psi_3(\varepsilon) \coloneqq \min\{\psi_1(\varepsilon),\psi_2(\varepsilon)\}$. For large n the homogeneously distributed state π_{Q_n} is almost certainly selected by the reduced operation $\tilde{\mathcal{N}}_{\varepsilon,n}$.

Now, let us address the Frobenius norm of $\tilde{\mathcal{N}}(\pi_{Q_n})$. For positive operators A,B

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

This can be used to derive

$$\begin{aligned}\|\mathcal{P}_{\varepsilon,n}\circ\mathcal{N}^{\otimes n}(\pi_{Q_n})\|_F^2 &= \|\mathcal{P}_{\varepsilon,n}\circ(\mathcal{N}_{\varepsilon,n} + \mathcal{M}_{\varepsilon,n})(\pi_{Q_n})\|_F^2\\ &\geqslant \|\mathcal{P}_{\varepsilon,n}\circ\mathcal{N}_{\varepsilon,n}(\pi_{Q_n})\|_F^2.\end{aligned}$$

Thus

$$\begin{aligned}\|\tilde{\mathcal{N}}_{\varepsilon,n}(\pi_{Q_n})\|_F^2 = \|\mathcal{P}_{\varepsilon,n}\circ\mathcal{N}_{\varepsilon,n}(\pi_{Q_n})\|_F^2 &\leqslant \|\mathcal{P}_{\varepsilon,n}\circ\mathcal{N}^{\otimes n}(\pi_{Q_n})\|_F^2\\ &= \|\Pi_{\varepsilon,n}\mathcal{N}(\pi_Q)^{\otimes n}\Pi_{\varepsilon,n}\|_F^2\\ &= \sum_{\mathbf{l}:|v_{\mathbf{l}}\rangle\ \varepsilon-\text{typical}} (p_{\mathbf{l}})^2 \leqslant 2^{-n\{S(\mathcal{N}(\pi_Q))-3\varepsilon\}},\end{aligned}$$

*where we used Eq. (43) and $p_{\mathbf{l}} \leqslant 2^{-n\{S(\mathcal{N}(\pi_Q))-\varepsilon\}}$ to derive the last inequality.

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

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

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

[4] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[5] E. Knill and R. Laflamme, Phys. Rev. A **55**, 900 (1997).

[6] For a review see, e. g., J. Preskill, *Quantum Information and Computation* (World Scientific, Singapore, 1998); arXiv:quant-ph/9712048.

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

[8] H. Barnum, M. A. Nielsen, and B. Schumacher, Phys. Rev. A **57**, 4153 (1998).

[9] H. Barnum, E. Knill, and M. A. Nielsen, IEEE Trans. Inf. Theory **46**, 1317 (2000).

[10] M. A. Nielsen, C. M. Caves, B. Schumacher, and H. Barnum, Proc. R. Soc. London, Ser. A **454**, 277 (1998).

[11] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, IEEE Trans. Inf. Theory **48**, 2637 (2002).

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

[13] 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

[14] I. Devetak, IEEE Trans. Inf. Theory **51**, 44 (2005).

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

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

[17] D. W. Leung, M. A. Nielsen, I. L. Chuang, and Y. Yamamoto, Phys. Rev. A **56**, 2567 (1997).

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

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

[20] B. Schumacher and M. D. Westmoreland, Quantum Inf. Process. **1**, 5 (2002).

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

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

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

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

[25] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter, e-print arXiv:quant-ph/0606225.

[26] If h is a concave function, and X a random variable in the domain of h, then by Jensen's inequality $[h(X)] < h([X])$.

[27] H. Weyl, *The Classical Groups* (Princeton University Press, Princeton, NJ, 1946).

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

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

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

[31] P. Hayden, M. Horodecki, J. Yard, and A. Winter, e-print arXiv:quant-ph/0702005v1.

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

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