Achieving the Holevo bound via sequential measurements

Citation: Giovannetti, Vittorio, Seth Lloyd, and Lorenzo Maccone. "Achieving the Holevo Bound via Sequential Measurements." Physical Review A 85.1 (2012): n. pag. Web. 17 Feb. 2012. © 2012 American Physical Society

As Published: http://dx.doi.org/10.1103/PhysRevA.85.012302

Publisher: American Physical Society (APS)

Persistent URL: http://hdl.handle.net/1721.1/69151

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Achieving the Holevo bound via sequential measurements

Vittorio Giovannetti,[1] Seth Lloyd,[2] and Lorenzo Maccone[3]
[1]*NEST, Scuola Normale Superiore and Istituto Nanoscienze–CNR, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*
[2]*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3]*Dipartimento Fisica "A. Volta", INFN Sezione Pavia, Università di Pavia, via Bassi 6, I-27100 Pavia, Italy*
(Received 18 October 2011; published 3 January 2012)

We present a decoding procedure to transmit classical information in a quantum channel which, saturating asymptotically the Holevo bound, achieves the optimal rate of the communication line. In contrast to previous proposals, it is based on performing a sequence of projective YES-NO measurements which in N steps determines which codeword was sent by the sender (N being the number of the codewords). Our analysis shows that, as long as N is below the limit imposed by the Holevo bound, the error probability can be sent to zero asymptotically in the length of the codewords.

DOI: 10.1103/PhysRevA.85.012302 PACS number(s): 03.67.Hk, 89.70.Kn

I. INTRODUCTION

By constraining the amount of classical information which can be reliably encoded into a collection of quantum states [1], the Holevo bound sets a limit on the rates that can be achieved when transferring classical messages in a quantum communication channel. Even though, for a finite number of channel uses, the bound in general is not achievable, it is saturated [2,3] in the asymptotic limit of infinitely many channel uses. Consequently, via proper optimization and regularization [4], it provides the quantum analog of the Shannon capacity formula [5], i.e., the *classical capacity* of the quantum channel (e.g., see Refs. [6,7]).

Starting from the seminal works of Refs. [2,3], several alternative versions of the asymptotic attainability of the Holevo bound have been presented so far (e.g., see Refs. [7–12] and references therein). The original proof [2,3] was obtained by extending to the quantum regime the typical subspace-encoding argument of Shannon communication theory [5]. In this context an explicit detection scheme [sometime presented as the *pretty good measurement* (PGM) scheme [3,13]] was introduced that allows for exact message recovery in the asymptotic limit infinitely long codewords. More recently, Ogawa and Nagaoka [9,10], and Hayashi and Nagaoka [11] proved the asymptotic achievability of the bound by establishing a formal connection with the quantum-hypothesis-testing problem [14], and by generalizing a technique (the information-spectrum method) which was introduced by Verdú and Han [15] in the context of classical communication channels.

In this paper we analyze a decoding procedure for classical communication in a quantum channel which allows for an alternative proof of the asymptotic attainability of the Holevo bound. Here we give a formal proof, whereas in Ref. [16] we give a more intuitive take on the argument, based on a decoding measurement procedure that relies on a randomized search of the optimal match between the received message and the input codeword. The main advantage resides in the fact that, unlike the strategies which employ the PGM and its variants [13,17–27] the proposed scheme allows for a simple intuitive description and it appears to be more suited for practical implementations. As in Refs. [2,3] our approach is based on the notion of a typical subspace but it replaces the PGM scheme with a sequential decoding strategy in which, similarly to the quantum-hypothesis-testing approach of Refs. [9,10], the received quantum codeword undergoes a sequence of simple YES-NO projective measurements that try to determine which among all possible inputs might have originated it. It is worth mentioning that the possibility of adopting a sequential YES-NO projective decoding strategy to achieve the Holevo bound was implicitly anticipated in Ref. [8] by Winter. His approach, however, is completely different from the one given here: in particular, in his derivation Winter adopts a *greedy algorithm* which starting from a sufficiently small code allows one to expand it until it saturates the bound (in the construction each individual new codeword that is added to the code is decoded via a two-value detection strategy which can be implemented through proper von Neumann projections). In contrast, to prove that our strategy attains the bound, we invoke Shannon's averaging trick and show that the *average* error probability converges to zero in the asymptotic limit of long codewords (the average being performed over the codewords of a given code *and* over all the possible codes).

The paper is organized as follows: In Sec. II we set the problem, present the scheme in an informal, nontechnical way, and clarify an important aspect concerning the structure of our sequential detection scheme. The formal derivation of the procedure begins in the next section. Specifically, the notation and some basic definitions are presented in Sec. III. Next the sequential detection strategy is formalized in Sec. IV, and finally the main result is derived in Sec. V. Conclusions and perspectives are given in Sec. VI. The paper includes also some technical Appendixes.

II. INTUITIVE DESCRIPTION OF THE MODEL

The transmission of classical messages through a quantum channel can be decomposed into three logically distinct stages: the *encoding* stage in which the sender of the message (say, Alice) maps the classical information she wishes to communicate into the states of some quantum objects (the quantum-information carriers of the system); the *transmission* stage in which the carriers propagate along the communication line, reaching the receiver (say, Bob); and the *decoding* stage in which Bob performs some quantum measurement on the carriers in order to retrieve Alice's messages. For explanatory purposes we will restrict the analysis to the simplest scenario

where Alice is bound to use only unentangled signals and where the noise in the channel is memoryless.[1] Under this hypothesis the coding stage can be described as a process in which Alice encodes N classical messages into factorized states of n quantum carriers, producing a collection $\mathcal{C}$ of N quantum codewords of the form $\sigma_{\vec{j}} := \sigma_{j_1} \otimes \cdots \otimes \sigma_{j_n}$, where $j_1, \ldots, j_n$ are symbols extracted from a classical alphabet and where we use N different vectors $\vec{j}$. Due to the communication noise, these strings will be received as the factorized states $\rho_{\vec{j}} := \rho_{j_1} \otimes \cdots \otimes \rho_{j_n}$ (the output codewords of the system), where for each j we have

$$\rho_j = T(\sigma_j)\,, \tag{1}$$

T being the completely positive, trace-preserving channel [29] that defines the noise acting on each carrier. Finally, the decoding stage of the process can be characterized by assigning a specific positive-valued operator measure (POVM) [29] which Bob applies to $\rho_{\vec{j}}$ to get a (hopefully faithful) estimation $\vec{j}'$ of the value $\vec{j}$. Indicating with $\{X_{\vec{j}}, X_0 = \mathbb{1} - \sum_{\vec{j}\in\mathcal{C}} X_{\vec{j}}\}$ the elements which compose the selected POVM, the average error probability that Bob will mistake a given $\vec{j}$ sent by Alice for a different one can be expressed as (e.g., see Ref. [2]),

$$P_{\rm err} := \frac{1}{N}\sum_{\vec{j}\in\mathcal{C}} (1 - {\rm Tr}[X_{\vec{j}}\rho_{\vec{j}}]). \tag{2}$$

In the limit of infinitely long sequences $n \to \infty$, it is known [2,3,7–11] that $P_{\rm err}$ can be sent to zero under the condition that N scales as 2^{nR} with R being bounded by the optimized version of the Holevo information, i.e.,

$$R \leqslant \max_{\{p_j,\sigma_j\}} \chi(\{p_j,\rho_j\}), \tag{3}$$

where the maximization is performed over all possible choices of the inputs σ_j and over all possible probabilities p_j, and where for a given quantum output ensemble $\{p_j,\rho_j\}$ we have

$$\chi(\{p_j,\rho_j\}) := S\left(\sum_j p_j\rho_j\right) - \sum_j p_j S(\rho_j), \tag{4}$$

with $S(\cdot) := -{\rm Tr}[(\cdot)\log_2(\cdot)]$ being the von Neumann entropy [29]. The inequality in Eq. (3) is a direct consequence of the Holevo bound [1], and its right-hand side defines the so-called Holevo capacity of the channel T, i.e., the highest achievable rate of the communication line which guarantees asymptotically null zero error probability under the constraint of employing only unentangled codewords.[2] In Refs. [2,3] the achievability of the bound (3) was obtained by showing that from *any* output quantum ensemble $\{p_j,\rho_j\}$ it is possible to identify a set of $\sim 2^{n\chi(\{p_j,\rho_j\})}$ output codewords $\rho_{\vec{j}}$, and a decoding POVM for which the error probability of Eq. (2) goes to zero as n increases. Note that by proceeding this way one can forget about the initial mapping $\vec{j} \to \sigma_{\vec{j}}$ and work directly with the $\vec{j} \to \rho_{\vec{j}}$ mapping. This is an important simplification which typically is not sufficiently stressed (see, however, Ref. [11]). Within this framework, the proof [2,3] exploited the random coding trick by Shannon in which the POVM is shown to provide exponentially small error probability *on average*, when mediating over all possible groups of codewords associated with $\{p_j,\rho_j\}$.

The idea we present here follows the same typicality approach of Refs. [2,3] but assumes a different detection scheme. In particular, while in Refs. [2,3] the POVM produces all possible outcomes in a single step as shown schematically in the inset of Fig. 1, our scheme is sequential *and* constructed in terms of two-valued projections which allow Bob to test for each of the codewords. Specifically, in our approach Bob is supposed to perform a first YES-NO projective measure to verify whether or not the received signal corresponds to the first element of the list; see Fig. 1. If the answer is YES he stops and declares that the received message was the first one. If the answer is NO he takes the state which emerges from the measurement apparatus and performs a new YES-NO projective measure aimed to verify whether or not it corresponds to the second element of the list, and so on until he has checked for all possible outcomes. The difficulty resides in the fact that, due to the quantum nature of the codewords, at each step of the protocol the received message is partially modified by the measurement (a problem which will *not* occur in a purely classical communication scenario). This implies for instance that the state that is subject to the second measurement is not equal to what Bob received from the quantum channel. As a

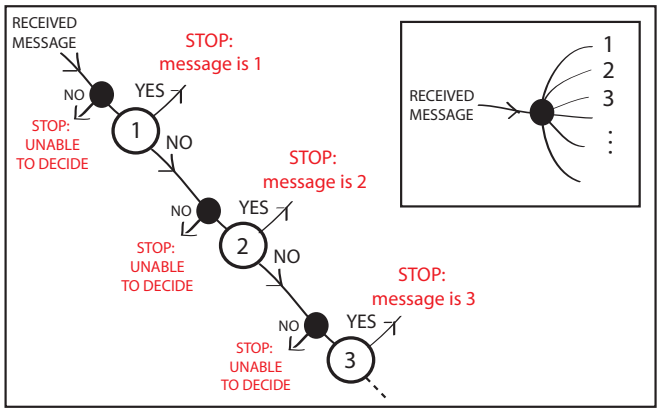

FIG. 1. (Color online) Flowchart representation of the detection scheme: the projections on the typical subspace $\mathcal{H}^{(n)}_{\rm typ}(\vec{j})$ of the codewords are represented by the open circles, while the projections on the typical subspace $\mathcal{H}^{(n)}_{\rm typ}$ of the average message of the source are represented by the black circles (they correspond to the *smoothing* steps of the protocol needed to compensate for the nonexact orthogonality of the $P_{\vec{j}}$—see text for details). The inset describes the standard PGM decoding scheme which produces all the possible outcomes in a single step.

[1] A similar formulation of the problem holds also when entangled signals are allowed: in this case, however, the σ_j defined in the text represents (possibly entangled) states of m-long *blocks* of carriers: for each possible choice of m, and for each possible coding and decoding strategy, one defines the error probability as in Eq. (2). The optimal transmission rate (i.e., the capacity of the channel) is also expressible as in the right-hand-side term of Eq. (3) via proper regularization over m (this is a consequence of the superadditivity of the Holevo information [4]). Finally the same construction can be applied also in the case of quantum communication channels with memory, e.g., see Ref. [28].

[2] See footnote 1.

consequence, to avoid divergence of the accumulated errors as the detection proceeds, the YES-NO projective measurements need to be carefully designed to have little impact on the received codewords. As will be clarified in the following section we tackle this problem by resorting on the notion of typical subspaces [30]: specifically our YES-NO projective measurements will be mild modifications of von Neumann projections on the typical subspaces of the codewords, in which their nonexact orthogonality is smoothed away by rescaling them through further projection on the typical subspace of the source average message (see Sec. IV).

A. Two-value projective measures vs two-valued measures

Before entering into the technical details of the derivation, it is worth emphasizing that the difficult part of our derivation consists exactly in showing that the decoding scheme is built upon two-value *projective* measures. Indeed, given a generic POVM of elements $\{Y_\ell\}_{\ell=1,\ldots,r}$ it is always possible to represent it as a sequence of properly concatenated *non-necessarily projective* two-valued measures. For instance, at the first step of the sequential measurement one probes the system with the generalized (non-necessarily projective) measurement of elements $\{Y_1{}^{1/2},(\mathbb{1}-Y_1)^{1/2}\}$, the first entry being associated with the outcome $\ell=1$ of the original POVM, and the second with the outcome $\ell\neq 1$. Accordingly the latter event occurs with probability $q_1 := \mathrm{Tr}[\rho(\mathbb{1}-Y_1)]$ while the input state ρ is mapped into $\rho_1 := (\mathbb{1}-Y_1)^{1/2}\rho(\mathbb{1}-Y_1)^{1/2}/q_1$. Checking ρ for the outcome $\ell=2$ is then equivalent to testing ρ_1 with the two-valued generalized measurement of elements $\{Y_2{}^{1/2},(\mathbb{1}-Y_2)^{1/2}\}$: in this case, at the level of ρ the first entry corresponds to the outcome $\ell=2$, while the second one provides the result $\ell\neq 1,2$. Proceeding along this line for all ℓ, we can finally decompose the POVM in terms of a sequence of concatenated two-valued generalized measurements which, at the level of the input state ρ, are expressed in terms of the operators $\{Y_1^{1/2},(\mathbb{1}-Y_1)^{1/2}\}$, $\{(\mathbb{1}-Y_1)^{-1/2}Y_2{}^{1/2},(\mathbb{1}-Y_1)^{-1/2}(\mathbb{1}-Y_1-Y_2)^{1/2}\}$, $\{(\mathbb{1}-Y_1-Y_2)^{-1/2}Y_3{}^{1/2},(\mathbb{1}-Y_1-Y_2)^{-1/2}\,(\mathbb{1}-Y_1-Y_2-Y_3)^{1/2}\}$, etc.

III. SOURCES, CODES, AND TYPICAL SUBSPACES

In this section we review some basic notions and introduce the definitions necessary to formalize our detection scheme.

An independent, identically distributed quantum source is defined by assigning the quantum ensemble $\mathcal{E}=\{p_j,\rho_j : j\in\mathcal{A}\}$ which specifies the density matrices $\rho_j\in\mathfrak{S}(\mathcal{H})$ emitted by the source as they emerge from the memoryless channel, as well as the probabilities p_j associated with those events (here j is the associated classical random variable which takes values on the domain $\mathcal{A}$). Since the channel is memoryless, when operated n consecutive times, it generates products states $\rho_{\vec{j}}\in\mathfrak{S}(\mathcal{H}^{\otimes n})$ of the form

$$\rho_{\vec{j}} := \rho_{j_1}\otimes\cdots\otimes\rho_{j_n}, \tag{5}$$

with probability

$$p_{\vec{j}} := p_{j_1}p_{j_2}\cdots p_{j_n} \tag{6}$$

[in these expressions $\vec{j}:=(j_1,\ldots,j_n)\in\mathcal{A}^n$]. In strict analogy to Shannon information theory, one defines an N-element code $\mathbf{C}$ as a collection of N states of the form (5), i.e.,

$$\mathbf{C} := \{\rho_{\vec{j}}\in\mathfrak{S}(\mathcal{H}^{\otimes n}) : \vec{j}\in\mathcal{C}\}, \tag{7}$$

with $\mathcal{C}$ being the subset of $\mathcal{A}^n$ which identifies the elements of $\mathbf{C}$ (i.e., the codewords of the code). The probability that the source will generate the code $\mathbf{C}$ can then be computed as the (joint) probability of emitting all the codewords that compose it, i.e.,

$$P(\mathbf{C}) := \prod_{\vec{j}\in\mathcal{C}} p_{\vec{j}} = \prod_{\vec{j}\in\mathcal{C}}\prod_{\ell=1}^{n} p_{j_\ell}. \tag{8}$$

A. Typical spaces

Consider $\rho=\sum_j p_j\rho_j\in\mathfrak{S}(\mathcal{H})$ the average density matrix associated with the ensemble $\mathcal{E}$, and let $\rho=\sum_\ell q_\ell|e_\ell\rangle\langle e_\ell|$ be its spectral decomposition (i.e., $|e_\ell\rangle$ are the orthonormal bases of $\mathcal{H}$ formed by the eigenvectors of ρ while q_ℓ are their eigenvalues). For fixed $\delta>0$, one defines [30] the typical subspace $\mathcal{H}_{\mathrm{typ}}^{(n)}$ of ρ as the subspace of $\mathcal{H}^{\otimes n}$ spanned by those vectors,

$$|e_{\vec{\ell}}\rangle := \left|e_{\ell_1}\right\rangle\otimes\cdots\otimes\left|e_{\ell_n}\right\rangle, \tag{9}$$

whose associated probabilities $q_{\vec{\ell}} := q_{\ell_1}q_{\ell_2}\cdots q_{\ell_n}$ satisfy the constraint

$$2^{-n[S(\rho)+\delta]}\leqslant q_{\vec{\ell}}\leqslant 2^{-n[S(\rho)-\delta]}, \tag{10}$$

where $S(\rho)=-\mathrm{Tr}[\rho\log_2\rho]$ is the von Neumann entropy of ρ (as in the classical case [31], the states $|e_{\vec{\ell}}\rangle$ defined above can be thought of as those which, on average, contain the symbol $|e_\ell\rangle$ almost nq_ℓ times). Identifying by $\mathcal{L}$ the set of those vectors $\vec{\ell}=(\ell_1,\ell_2,\ldots,\ell_n)$ which satisfy Eq. (10), the projector P on $\mathcal{H}_{\mathrm{typ}}^{(n)}$ can then be expressed as

$$P=\sum_{\vec{\ell}\in\mathcal{L}}|e_{\vec{\ell}}\rangle\langle e_{\vec{\ell}}|, \tag{11}$$

while the average state $\rho^{\otimes n}$ is clearly given by

$$\rho^{\otimes n}=\sum_{\vec{\ell}}q_{\vec{\ell}}\,|e_{\vec{\ell}}\rangle\langle e_{\vec{\ell}}|. \tag{12}$$

By construction, the two operators satisfy the inequalities

$$P\;2^{-n[S(\rho)+\delta]}\leqslant P\rho^{\otimes n}P\leqslant P\;2^{-n[S(\rho)-\delta]}. \tag{13}$$

Furthermore, it is known that the probability that $\mathcal{E}$ will emit a message which is not in $\mathcal{H}_{\mathrm{typ}}^{(n)}$ is exponentially depressed [30]. More precisely, for all $\epsilon>0$ it is possible to identify a sufficiently large n_0 such that for all $n\geqslant n_0$ we have

$$\mathrm{Tr}[\rho^{\otimes n}(\mathbb{1}-P)]<\epsilon. \tag{14}$$

Typical subsets can be defined also for each of the product states of Eq. (5) associated with each codeword at the output of the channel. In this case the definition is as follows [2]:

First, for each $j \in \mathcal{A}$ we define the spectral decomposition of the element ρ_j, i.e.,

$$\rho_j = \sum_k \lambda_k^j |e_k^j\rangle\langle e_k^j|, \tag{15}$$

where $|e_k^j\rangle$ are the eigenvectors of ρ_j and λ_k^j the corresponding eigenvalues (notice that while $\langle e_k^j|e_{k'}^j\rangle = \delta_{kk'}$ for all k, k', and j, in general the quantities $\langle e_k^j|e_{k'}^{j'}\rangle$ are *a priori* undefined). Now the spectral decomposition of the codeword $\rho_{\vec{j}}$ is provided by

$$\rho_{\vec{j}} = \sum_{\vec{k}} \lambda_{\vec{k}}^{(\vec{j})} |e_{\vec{k}}^{(\vec{j})}\rangle\langle e_{\vec{k}}^{(\vec{j})}|, \tag{16}$$

where for $\vec{k} := (k_1, \ldots, k_n)$ one has

$$\begin{aligned} |e_{\vec{k}}^{(\vec{j})}\rangle &:= |e_{k_1}^{j_1}\rangle \otimes |e_{k_2}^{j_2}\rangle \otimes \cdots \otimes |e_{k_n}^{j_n}\rangle, \\ \lambda_{\vec{k}}^{(\vec{j})} &:= \lambda_{k_1}^{j_1}\lambda_{k_2}^{j_2}\cdots\lambda_{k_n}^{j_n}. \end{aligned} \tag{17}$$

Notice that for fixed $\vec{j}$ the vectors $|e_{\vec{k}}^{(\vec{j})}\rangle$ are an orthonormal set of $\mathcal{H}^{\otimes n}$; notice also that in general such vectors have nothing to do with the vectors $|e_{\vec{\ell}}\rangle$ of Eq. (9).

Now the typical subspace $\mathcal{H}_{\rm typ}^{(n)}(\vec{j})$ of $\rho_{\vec{j}}$ is defined as the linear subspace of $\mathcal{H}^{\otimes n}$ spanned by the $|e_{\vec{k}}^{(\vec{j})}\rangle$ whose associated $\lambda_{\vec{k}}^{(\vec{j})}$ satisfy the inequality

$$2^{-n[S(\rho)-\chi(\mathcal{E})+\delta]} \leqslant \lambda_{\vec{k}}^{(\vec{j})} \leqslant 2^{-n[S(\rho)-\chi(\mathcal{E})-\delta]}, \tag{18}$$

with

$$\chi(\mathcal{E}) := S(\rho) - \sum_j p_j S(\rho_j) \tag{19}$$

being the Holevo information of the source $\mathcal{E}$. The projector on $\mathcal{H}_{\rm typ}^{(n)}(\vec{j})$ can then be written as

$$P_{\vec{j}} := \sum_{\vec{k}\in\mathcal{K}_{\vec{j}}} |e_{\vec{k}}^{(\vec{j})}\rangle\langle e_{\vec{k}}^{(\vec{j})}|, \tag{20}$$

where $\mathcal{K}_{\vec{j}}$ identify the set of the labels $\vec{k}$ which satisfy Eq. (18).

We notice that the bounds for the probabilities $\lambda_{\vec{k}}^{(\vec{j})}$ do not depend on the value of $\vec{j}$ which defines the selected codeword: they are only functions of the source $\mathcal{E}$ only [this of course does not imply that the subspace $\mathcal{H}_{\rm typ}^{(n)}(\vec{j})$ will not depend on $\vec{j}$]. It is also worth stressing that since the vectors $|e_{\vec{k}}^{(\vec{j})}\rangle$ in general are not orthogonal with respect to the label $\vec{j}$, there will be a certain overlap between the subspaces $\mathcal{H}_{\rm typ}^{(n)}(\vec{j})$. The reason why they are defined as detailed above stems from the fact that the probability that $\rho_{\vec{j}}$ will not be found in $\mathcal{H}_{\rm typ}^{(n)}(\vec{j})$ (averaged over all possible realization of $\rho_{\vec{j}}$) can be made arbitrarily small by increasing n, e.g., see Ref. [2]. More precisely, for fixed $\delta > 0$, one can show that for all $\epsilon > 0$ there exists n_0 such that for all $n > n_0$ integer one has

$$\sum_{\vec{j}} p_{\vec{j}} \,{\rm Tr}[\rho_{\vec{j}}(\mathbb{1} - P_{\vec{j}})] < \epsilon, \tag{21}$$

where $p_{\vec{j}}$ is the probability (6) that the source $\mathcal{E}$ has emitted the codeword $\rho_{\vec{j}}$.

B. Decoding and Shannon's averaging trick

The goal in the design of a decoding stage is to identify a POVM attached to the code $\mathbf{C}$ that yields a vanishing error probability in identifying the codewords as n increases. How can one prove that such a POVM exists? First of all let us recall that a POVM is a collection of positive operators $\{X_{\vec{j}}, X_0 = \mathbb{1} - \sum_{\vec{j}\in\mathcal{C}} X_{\vec{j}} : \vec{j} \in \mathcal{C}\}$. The probability of getting a certain outcome $\vec{j}'$ when measuring the codeword $\rho_{\vec{j}}$ is computed as the expectation value ${\rm Tr}[X_{\vec{j}'}\rho_{\vec{j}}]$ (the outcome associated with ${\rm Tr}[X_0\rho_{\vec{j}}]$ corresponds to the case in which the POVM is not able to identify any of the possible codewords). Then, the error probability (averaged over all possible codewords of $\mathbf{C}$) is given by the quantity

$$P_{\rm err}(\mathrm{C}) := \frac{1}{N}\sum_{\vec{j}\in\mathcal{C}} (1 - {\rm Tr}[X_{\vec{j}}\rho_{\vec{j}}]). \tag{22}$$

Proving that this quantity is asymptotically null will be in general quite complicated. However, the situation simplifies if one averages $P_{\rm err}(\mathrm{C})$ with all codewords $\mathbf{C}$ that the source $\mathcal{E}$ can generate, i.e.,

$$\langle P_{\rm err}\rangle := \sum_{\mathbf{C}} P(\mathbf{C})\, P_{\rm err}(\mathrm{C}), \tag{23}$$

$P(\mathbf{C})$ being the probability defined in Eq. (8). Proving that $\langle P_{\rm err}\rangle$ is nullified for $n \to \infty$ implies that at least one of the codes $\mathbf{C}$ generated by $\mathcal{C}$ allows for asymptotic null error probability with the selected POVM (indeed the result is even stronger as *almost all* those which are randomly generated by $\mathcal{C}$ will do the job). In Refs. [2,3] the achievability of the Holevo bound was proven by adopting the pretty good measurement detection scheme, i.e., the POVM of elements

$$X_{\vec{j}} = \left[\sum_{\vec{h}\in\mathcal{C}} P P_{\vec{h}} P\right]^{-\frac{1}{2}} P P_{\vec{j}} P \left[\sum_{\vec{h}\in\mathcal{C}} P P_{\vec{h}} P\right]^{-\frac{1}{2}}, \tag{24}$$

$$X_0 = \mathbb{1} - \sum_{\vec{j}\in\mathcal{C}} X_{\vec{j}}, \tag{25}$$

where P is the projector (11) on the typical subspace of the average state of the source, and for $\vec{j} \in \mathbf{C}$ the $P_{\vec{j}}$ are the projectors (20) associated with the codeword $\rho_{\vec{j}}$ [it is worth stressing that for generic choices of the set $\mathcal{C}$ the operators (24) do not represent orthogonal projectors]. With this choice one can verify that for given ϵ there exist n sufficiently large such that Eq. (23) yields the inequality [2]

$$\langle P_{\rm err}\rangle \leqslant 4\epsilon + (N-1)\, 2^{-n[\chi(\mathcal{E})-2\delta]}, \tag{26}$$

implying that as long as $N - 1$ is smaller than $2^{-n[\chi(\mathcal{E})-2\delta]}$ one can bound the (average) error probability close to zero.

IV. THE SEQUENTIAL DETECTION SCHEME

In this section we formalize our sequential detection scheme and compute its associated average error probability.

A. The scheme

As anticipated in the Introduction, the idea is to recover the value of the label $\vec{j}$ of the received codeword $\rho_{\vec{j}}$ by employing a sequence of concatenating YES-NO tests to determine in which of the typical subspaces $\{\mathcal{H}^{(n)}_{\rm typ}(\vec{j}')\}_{\vec{j}'\in\mathcal{C}}$ this state belongs. Of course, since Bob does not know *a priori* the true value of $\vec{j}$, he needs to check recursively for all the possibilities. The natural tools to perform these tests are the projectors (20); however, as mentioned earlier, we have to smooth such operations to account for the disturbance that the nonorthogonality among the $\mathcal{H}^{(n)}_{\rm typ}(\vec{j})$ might introduce in the process. For this purpose, each projective measurement on a given $\mathcal{H}^{(n)}_{\rm typ}(\vec{j})$ will be preceded by a smoothing stage in which Bob checks (via a von Neumann projective measurement on $\mathcal{H}^{(n)}_{\rm typ}$) whether or not the state is in the typical subspace of the average message.[3]

The resulting scheme is schematically sketched in Fig. 1 and consists in the following instructions:

(0) Bob fixes an ordering[4] of the codewords of $\mathcal{C}$ yielding the sequence $\vec{j}_1,\vec{j}_2,\vec{j}_3,\ldots,\vec{j}_N$ and introduces a discrete variable u that it is set equal to 1.

(1) Then Bob performs a smoothing transformation by checking (via a YES-NO projective measurement) whether or not the received state is in the typical subspace $\mathcal{H}^{(n)}_{\rm typ}$ of the average message. If he gets the result NO, he declares failure (the message cannot be decoded) and the protocol stops. Vice versa, if he gets YES the protocol proceeds with the operations that follow:

(2) Bob performs a YES-NO measurement that determines whether or not the received state is in the typical subspace of the uth element of the list (i.e., the one corresponding to codeword $\vec{j}_u$):[5]

(a) If the result of the measurement at step (2) is YES, the protocol stops and Bob declares that he has identified the received message as the uth element of the list (i.e., $\vec{j}_u$).

(b) Instead, if the answer of the measurement at step (2) is NO, Bob increments the value of the variable u by 1: If $u+1>N$ he declares failure (the message cannot be decoded) and the protocol stops; otherwise he goes back to point (1) of the instruction list and the protocol continues [i.e., Bob will perform a YES-NO measurement to check whether or not the state is in the typical subspace of the $(u+1)$th element of the list].

From the above scheme it should be clear that the protocol proceeds until Bob gets either a YES answer at the step (2) for some $u\in\{1,\ldots,N\}$ or a NO result at one of the smoothing stages (meaning that during the decoding procedure the received message has left the typical subspace $\mathcal{H}^{(n)}_{\rm typ}$ of the average message). In this way he is able to test all the N possibilities determining an estimate of the transmitted $\vec{j}$ or getting a null result (the message has not been identified, corresponding to an error in the communication).

B. The error probability

The statistical properties of our decoding procedure can be represented in terms of an effective POVM with $N+1$ elements $\{E_1,E_2,\ldots,E_N,E_0=\mathbb{1}-\sum_{u=1}^N E_u\}$, where the first N are associated with the YES results of the projections on the typical subspace of the codewords, while the last one accounts collectively for all the failure events which bring Bob to conclude that he is not able to decode the received message. More specifically, E_1 allows us to compute the joint probability that the transmitted state will pass successfully the first smoothing stage *and* give a YES result in the projective measurement on $\mathcal{H}^{(n)}_{typ}(\vec{j}_1)$. Accordingly it is described by the (positive) operator

$$E_1 := \bar{P}_{\vec{j}_1}, \tag{27}$$

where for any operator Θ the symbol $\bar{\Theta}$ stands for

$$\bar{\Theta} := P\Theta P, \tag{28}$$

P being the projector of Eq. (11). As explicitly shown in Appendix A 1, similar expressions hold also for the remaining elements of the POVM. In particular, indicating with $Q_{\vec{j}} := \mathbb{1}-P_{\vec{j}}$ the orthogonal complement of $P_{\vec{j}}$, we have

$$\begin{aligned} E_2 &:= \bar{Q}_{\vec{j}_1}\bar{P}_{\vec{j}_2}\bar{Q}_{\vec{j}_1},\\ E_3 &:= \bar{Q}_{\vec{j}_1}\bar{Q}_{\vec{j}_2}\bar{P}_{\vec{j}_3}\bar{Q}_{\vec{j}_2}\bar{Q}_{\vec{j}_1},\\ E_4 &:= \cdots, \end{aligned} \tag{29}$$

which, for all u, admit the following compact form:

$$E_u = M_u^\dagger M_u, \tag{30}$$

with

$$M_u := P_{\vec{j}_u}P\bar{Q}_{\vec{j}_{u-1}}\bar{Q}_{\vec{j}_{u-2}}\cdots\bar{Q}_{\vec{j}_1}. \tag{31}$$

The associated average error probability (23) can then be expressed as

$$\langle P_{\rm err}\rangle = \sum_{\vec{j}_1,\ldots,\vec{j}_N}\frac{p_{\vec{j}_1}\cdots p_{\vec{j}_N}}{N}\sum_{u=1}^N\left(1-{\rm Tr}\left[M_u\rho_{\vec{j}_u}M_u^\dagger\right]\right), \tag{32}$$

yielding the identity

$$\begin{aligned} 1-\langle P_{\rm err}\rangle &= \sum_{\ell=0}^{N-1}\sum_{\vec{j},\vec{j}_1,\ldots,\vec{j}_\ell}\frac{p_{\vec{j}}p_{\vec{j}_1}\cdots p_{\vec{j}_\ell}}{N}\\ &\quad\times{\rm Tr}\left[P_{\vec{j}}\bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}\rho_{\vec{j}}\bar{Q}_{\vec{j}_\ell}\cdots\bar{Q}_{\vec{j}_1}\right]\\ &= \sum_{\ell=0}^{N-1}\sum_{\vec{j},\vec{j}_1,\ldots,\vec{j}_\ell}\sum_{\vec{k}}\sum_{\vec{k}'\in\mathcal{K}_{\vec{j}}}\lambda_{\vec{k}}^{(\vec{j})}\frac{p_{\vec{j}}p_{\vec{j}_1}\cdots p_{\vec{j}_\ell}}{N}\\ &\quad\times\left|\langle e_{\vec{k}'}^{(\vec{j})}|\bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}|e_{\vec{k}}^{(\vec{j})}\rangle\right|^2, \end{aligned} \tag{33}$$

[3]The exact ordering between the smoothing steps and the projective measurements on a given $\mathcal{H}^{(n)}_{\rm typ}(\vec{j})$ is not relevant: indeed, due to the sequential character of the detection scheme, all projections will always be preceded and followed by a smoothing transformation.

[4]Fixing a prior ordering of the codebook is useful to formalize the protocol but it is not essential: indeed, in Ref. [16] we do not assume this and test randomly for the codewords.

[5]It is worth stressing that in Ref. [16] this test was implemented by performing a series of rank-1 projective measurements onto a basis of the subspace.

where the definitions of Eqs. (16) and (20) were employed to explicitly compute the trace.

V. BOUNDS ON THE ERROR PROBABILITY

In this section we derive an upper limit for the error probability (32) which will lead us to prove the achievability of the Holevo bound. The starting point of our analysis is to apply the Cauchy-Schwarz inequality to the right-hand side (RHS) of Eq. (33). In Appendix A 2 we prove that Eq. (33) can be written as

$$1-\langle P_{\rm err}\rangle \geqslant \frac{1}{N}\sum_{\ell=0}^{N-1}|{\rm Tr}[W_1\ \mathcal{Q}^\ell]|^2, \tag{34}$$

where for q integer we defined the operators

$$\begin{aligned} W_q &:= \textstyle\sum_{\vec{j}}\ p_{\vec{j}}\ P_{\vec{j}}\ \rho_{\vec{j}}^q\ P_{\vec{j}},\\ \mathcal{Q} &:= \textstyle\sum_{\vec{j}}\ p_{\vec{j}}\ \bar{\mathcal{Q}}_{\vec{j}_\ell} = \bar{\mathbb{1}} - \bar{W}_0. \end{aligned} \tag{35}$$

To proceed further it is important to notice that the quantity $\mathcal{Q}$ is always positive and smaller than $\mathbb{1}$, i.e.,

$$\mathbb{1}\geqslant \mathcal{Q}\geqslant 0. \tag{36}$$

Both properties simply follow from the identity

$$\mathcal{Q} = P\left(\mathbb{1}-\sum_{\vec{j}}p_{\vec{j}}P_{\vec{j}}\right)P = P\left[\sum_{\vec{j}}\ p_{\vec{j}}\ (\mathbb{1}-P_{\vec{j}})\right]P, \tag{37}$$

and from the fact that $\mathbb{1}\geqslant \mathbb{1}-P_{\vec{j}}\geqslant 0$. We also notice that

$$\mathbb{1}\geqslant W_1\geqslant W_0\times 2^{-n[S(\rho)-\chi(\mathcal{E})+\delta]}\geqslant 0, \tag{38}$$

where the last inequality is obtained by observing that the typical eigenvalues $\lambda_{\vec{k}}^{(\vec{j})}$ are lower bounded as in Eq. (18). From the above expressions we can conclude that the quantity in the summation that appears on the LHS of Eq. (34) is always smaller than 1 and that it is decreasing with ℓ. An explicit proof of this fact is as follows:

$$\begin{aligned} 0\leqslant {\rm Tr}[W_1\ \mathcal{Q}^\ell] &= {\rm Tr}\big[\sqrt{W_1}\ \mathcal{Q}^{\frac{\ell-1}{2}}\ \mathcal{Q}\ \mathcal{Q}^{\frac{\ell-1}{2}}\ \sqrt{W_1}\big]\\ &\leqslant {\rm Tr}\big[\sqrt{W_1}\ \mathcal{Q}^{\frac{\ell-1}{2}}\ \mathbb{1}\ \mathcal{Q}^{\frac{\ell-1}{2}}\ \sqrt{W_1}\big] = {\rm Tr}[W_1\ \mathcal{Q}^{\ell-1}], \end{aligned}$$

where we used the fact that the square root of a non-negative operator can be taken to be non-negative too (for a more detailed characterization of W_0 see Appendix A 3). A further simplification of the bound can be obtained by replacing the terms in the summation of Eq. (34) with the smallest addendum. This yields

$$1-\langle P_{\rm err}\rangle \geqslant |A|^2, \tag{39}$$

with

$$A := {\rm Tr}[W_1\mathcal{Q}^{N-1}] = \sum_{z=0}^{N-1}\binom{N-1}{z}(-1)^z f_z, \tag{40}$$

$$f_z := {\rm Tr}\big[W_1\ P\ \bar{W}_0^z\big], \tag{41}$$

where we used the fact that $\bar{\mathbb{1}}^2=\bar{\mathbb{1}}=P$ and we employed the definitions of Eq. (35). It turns out that the quantities f_z defined above are positive, smaller than 1, and decreasing in z. Indeed as shown in Appendix A 4 they satisfy the inequalities

$$0\leqslant f_z\leqslant f_0\ \ 2^{-nz[\chi(\mathcal{E})-2\delta]}\qquad \text{for all integer } z, \tag{42}$$

and, for each given ϵ, there exists a sufficiently large n_0 such that for $n\geqslant n_0$

$$1-\epsilon\leqslant f_0\leqslant 1. \tag{43}$$

Using these expressions, we can derive the following bound on A:

$$\begin{aligned} A &= f_0+\sum_{z=1}^{N-1}\binom{N-1}{z}(-1)^z f_z\\ &\geqslant f_0-\sum_{z=1}^{N-1}\binom{N-1}{z}f_z = 2f_0-\sum_{z=0}^{N-1}\binom{N-1}{z}f_z\\ &\geqslant 2f_0-f_0\sum_{z=0}^{N-1}\binom{N-1}{z}2^{-nz[\chi(\mathcal{E})-2\delta]}\\ &= f_0[2-(1+2^{-n[\chi(\mathcal{E})-2\delta]})^{N-1}], \end{aligned} \tag{44}$$

where in the first inequality we get a bound by taking all the terms of $k\geqslant 1$ with the negative sign, and the second is from (42). Now, on one hand if N is too large the quantity on the RHS side will become negative as we are taking the Nth power of a quantity which is larger than 1. On the other hand, if N is small then for large n the quantity in the square parentheses in the last equality will approach 1. This implies that there must be an optimal choice for N in order to have $[2-(1+2^{-n[\chi(\mathcal{E})-2\delta]})^{N-1}]$ approaching 1 for large n. To study such a threshold we rewrite Eq. (44) as

$$A\geqslant f_0[2-Y(x=2^{\chi(\mathcal{E})-2\delta},y=N,n)], \tag{45}$$

where we defined

$$Y(x,y,n) := (1+x^{-n})^{y^n-1}. \tag{46}$$

We notice that for $x,y\geqslant 1$ in the limit of $n\to\infty$ the quantity $\log_2[Y(x,y,n)]$ is an indeterminate form. Its behavior can be studied, for instance, using the de l'Hôpital formula, yielding

$$\lim_{n\to\infty}\log_2[Y(x,y,n)] = \frac{\log_2 x}{\log_2 y}\lim_{n\to\infty}\left(\frac{y}{x}\right)^n. \tag{47}$$

This shows that if $y<x$ the limit exists and it is zero, i.e., $\lim_{n\to\infty}Y(x,y,n)=1$. Vice versa, for $y>x$ the limit diverges, and thus $\lim_{n\to\infty}Y(x,y,n)=\infty$. Therefore, assuming $N=2^{nR}$, we can conclude that as long as

$$R<\chi(\mathcal{E})-2\delta \tag{48}$$

the quantity on the RHS of Eq. (45) approaches f_0 as n increases: this corresponds to having $y<x$ in the Y function, so that $Y\to 1$ for $n\to\infty$. Recalling then Eq. (43), we get

$$1-\langle P_{\rm err}\rangle \geqslant |A|^2 > f_0^2 > |1-\epsilon|^2 > 1-2\epsilon, \tag{49}$$

and thus

$$\langle P_{\rm err}\rangle < 2\epsilon. \tag{50}$$

On the contrary, if $R > \chi(\mathcal{E}) - 2\delta$, the lower bound on A becomes infinitely negative and hence useless to set a proper upper bound on $\langle P_{\rm err}\rangle$. This shows that by adopting the sequential detection strategy defined in Sec. IV it is possible to send $N = 2^{nR}$ messages with asymptotically vanishing error probability, for all rates R which satisfy the condition (48). ■

VI. CONCLUSIONS

Our analysis provides an explicit upper bound for the averaged error probability of our detection scheme (the average being performed over all codewords of a given code, and over all possible codes). Specifically, it shows that the error probability can be bounded close to zero for codes generated by sources $\mathcal{E}$ which have strictly fewer than $2^{n\chi(\mathcal{E})}$ elements. In other words, our detection scheme provides an alternative demonstration of the achievability of the Holevo bound [2]. In particular, an analogous procedure can be used to decode channels that transmit quantum information, to approach the coherent information limit [32–35]. This follows simply from the observation [34] that the transferring of quantum messages over the channel can be formally treated as the transferring of classical channels, imposing an extra constraint of privacy in the signaling process.

ACKNOWLEDGMENTS

V.G. is grateful to P. Hayden, A. S. Holevo, K. Matsumoto, J. Tyson, M. M. Wilde, and A. Winter for comments and discussions. V.G. acknowledges support from the FIRB-IDEAS project under Contract No. RBID08B3FM and the support of Institut Mittag-Leffler (Stockholm), where he was visiting while part of this work was done. S.L. was supported by the W. M. Keck Foundation, DARPA, NSF, NEC, ONR, and Intel. L.M. was supported by the W. M. Keck Foundation.

APPENDIX

This appendix is devoted to clarifying some technical aspects of the derivation.

1. Derivation of the POVM

Here we provide an explicit derivation of the POVM (30) associated with our iterative measurement procedure. It is useful to describe the whole process as a global unitary transformation that coherently transfers the information from the codewords to some external memory register.

Consider, for instance, the first step of the detection scheme where Bob tries to determine whether or not a given state $|\Psi\rangle \in \mathcal{H}^{\otimes n}$ corresponds to the first codeword $\rho_{\vec{j}_1}$ of his list. The corresponding measurement can be described as the following (two-step) unitary transformation:

$$\begin{aligned}|\Psi\rangle|00\rangle_{B_1} \rightarrow\ & P|\Psi\rangle|01\rangle_{B_1} + (\mathbb{1}-P)|\Psi\rangle|00\rangle_{B_1}\\ \rightarrow\ & P_{\vec{j}_1}P|\Psi\rangle|11\rangle_{B_1} + \big(\mathbb{1}-P_{\vec{j}_1}\big)P|\Psi\rangle|01\rangle_{B_1}\\ & + (\mathbb{1}-P)|\Psi\rangle|00\rangle_{B_1},\end{aligned} \tag{A1}$$

where B_1 represents a two-qubit memory register which stores the information extracted from the system. Specifically, the first qubit records with a "1" if the state $|\Psi\rangle$ belongs to the typical subspace $\mathcal{H}^{(n)}_{\rm typ}$ of the average state of the source (instead it will keep the value "0" if this is not the case). Similarly, the second qubit of B_1 records with a "1" if the projected component $P|\Psi\rangle$ is in the typical subspace $\mathcal{H}^{(n)}_{\rm typ}(\vec{j}_1)$ of $\rho_{\vec{j}_1}$. Accordingly the joint probability of success of finding $|\Psi\rangle$ in $\mathcal{H}^{(n)}_{\rm typ}$ and *then* in $\mathcal{H}^{(n)}_{\rm typ}(\vec{j}_1)$ is given by

$$\mathcal{P}_1(\Psi) = \langle\Psi|P P_{\vec{j}_1} P|\Psi\rangle, \tag{A2}$$

in agreement with the definition of E_1 given in Eq. (27). Vice versa, the joint probability of finding the state $|\Psi\rangle$ in in $\mathcal{H}^{(n)}_{\rm typ}$ and then not in $\mathcal{H}^{(n)}_{\rm typ}(\vec{j}_1)$ is given by $\langle\Psi|P(\mathbb{1}-P_{\vec{j}_1})P|\Psi\rangle$ and finally the joint probability of not finding $|\Psi\rangle$ in $\mathcal{H}^{(n)}_{\rm typ}$ is $\langle\Psi|\mathbb{1}-P|\Psi\rangle$. Let us now consider the second step of the protocol where Bob checks whether or not the message is in the typical subspace $\mathcal{H}^{(n)}_{\rm typ}(\vec{j}_2)$ of $\rho_{\vec{j}_2}$. It can be described as a unitary gate along the same lines of Eq. (A1) with $P_{\vec{j}_1}$ replaced by $P_{\vec{j}_2}$, and B_1 with a new two-qubit register B_2. Notice, however, that this gate only acts on that part of the global system which emerges from the first measurement with B_1 in $|01\rangle$. This implies the following global unitary transformation:

$$\begin{aligned}|\Psi\rangle|00\rangle_{B_1}|00\rangle_{B_2} \rightarrow\ & P_{\vec{j}_1}P|\Psi\rangle|11\rangle_{B_1}|00\rangle_{B_2}\\ & + \big[P_{\vec{j}_2}P\big(\mathbb{1}-P_{\vec{j}_1}\big)P|\Psi\rangle|01\rangle_{B_1}|11\rangle_{B_2}\\ & + \big(\mathbb{1}-P_{\vec{j}_2}\big)P\big(\mathbb{1}-P_{\vec{j}_1}\big)P|\Psi\rangle|01\rangle_{B_1}|01\rangle_{B_2}\\ & + (\mathbb{1}-P)\big(\mathbb{1}-P_{\vec{j}_1}\big)P|\Psi\rangle|01\rangle_{B_1}|00\rangle_{B_2}\big]\\ & + (\mathbb{1}-P)|\Psi\rangle|00\rangle_{B_1}|00\rangle_{B_2},\end{aligned} \tag{A3}$$

which shows that the joint probability of finding $|\Psi\rangle$ in $\mathcal{H}^{(n)}_{\rm typ}(\vec{j}_2)$ [after having found it in $\mathcal{H}^{(n)}_{\rm typ}$, not in $\mathcal{H}^{(n)}_{\rm typ}(\vec{j}_1)$, and again in $\mathcal{H}^{(n)}_{\rm typ}$] is

$$\mathcal{P}_2(\Psi) = \langle\Psi|P\big(\mathbb{1}-P_{\vec{j}_1}\big)P P_{\vec{j}_2}P\big(\mathbb{1}-P_{\vec{j}_1}\big)P|\Psi\rangle, \tag{A4}$$

in agreement with the definition of E_2 given in Eq. (29). Reiterating this procedure for all the remaining steps, one can then verify the validity of Eq. (30) for all $u \geqslant 2$. Moreover, it is clear [e.g., from Eqs. (A2) and (A4)] that it is a quite different POVM from the conventionally used pretty good measurement [2,3].

2. Derivation of Eq. (34)

The inequality (34) can be obtained via direct application of the Cauchy-Schwarz inequality to the RHS of Eq. (33). Specifically we notice that

$$\begin{aligned}&\sum_{\vec{k}}\sum_{\vec{k}'\in\mathcal{K}_{\vec{j}}} \lambda^{(\vec{j})}_{\vec{k}}\ \big|\langle e^{(\vec{j})}_{\vec{k}'}|\bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}|e^{(\vec{j})}_{\vec{k}}\rangle\big|^2\\ &\quad\geqslant \sum_{\vec{k}\in\mathcal{K}_{\vec{j}}} \lambda^{(\vec{j})}_{\vec{k}}\ \big|\langle e^{(\vec{j})}_{\vec{k}}|\bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}|e^{(\vec{j})}_{\vec{k}}\rangle\big|^2\\ &\quad= \sum_{\vec{k}\in\mathcal{K}_{\vec{j}}} \lambda^{(\vec{j})}_{\vec{k}}\big|\langle e^{(\vec{j})}_{\vec{k}}|\bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}|e^{(\vec{j})}_{\vec{k}}\rangle\big|^2 \sum_{\vec{k}}\lambda^{(\vec{j})}_{\vec{k}}\\ &\quad\geqslant \left|\sum_{\vec{k}\in\mathcal{K}_{\vec{j}}} \lambda^{(\vec{j})}_{\vec{k}}\langle e^{(\vec{j})}_{\vec{k}}|\bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}|e^{(\vec{j})}_{\vec{k}}\rangle\right|^2\\ &\quad= \big|{\rm Tr}\big[P_{\vec{j}}\rho_{\vec{j}}P_{\vec{j}}\bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}\big]\big|^2,\end{aligned} \tag{A5}$$

where the first inequality follows by dropping some positive terms (those with $\vec{k} \neq \vec{k}'$), the first identity simply exploits the fact that the $\lambda_{\vec{k}'}^{(\vec{j})}$ are normalized probabilities when summing over all $\vec{k}$, and the second inequality follows by applying the Cauchy-Schwarz inequality. Replacing this into Eq. (33) we can then write

$$1 - \langle P_{\rm err}\rangle \geqslant \sum_{\ell=0}^{N-1} \sum_{\vec{j},\vec{j}_1,\ldots,\vec{j}_\ell} \frac{p_{\vec{j}} p_{\vec{j}_1} \cdots p_{\vec{j}_\ell}}{N} \left|{\rm Tr}\left[P_{\vec{j}}\ \rho_{\vec{j}}\ P_{\vec{j}} \bar{Q}_{\vec{j}_1} \cdots \bar{Q}_{\vec{j}_\ell}\right]\right|^2. \tag{A6}$$

This can be further simplified by invoking again the Cauchy-Schwarz inequality, this time with respect to the summation over the $\vec{j}, \vec{j}_1, \ldots, \vec{j}_\ell$, i.e.,

$$\begin{aligned} &\sum_{\vec{j},\vec{j}_1,\ldots,\vec{j}_\ell} p_{\vec{j}} p_{\vec{j}_1} \cdots p_{\vec{j}_\ell} \left|{\rm Tr}\left[P_{\vec{j}} \rho_{\vec{j}} P_{\vec{j}} \bar{Q}_{\vec{j}_1} \cdots \bar{Q}_{\vec{j}_\ell}\right]\right|^2 \\ &\geqslant \left| \sum_{\vec{j},\vec{j}_1,\cdots,\vec{j}_\ell} p_{\vec{j}} p_{\vec{j}_1} \cdots p_{\vec{j}_\ell} {\rm Tr}\left[P_{\vec{j}} \rho_{\vec{j}} P_{\vec{j}} \bar{Q}_{\vec{j}_1} \cdots \bar{Q}_{\vec{j}_\ell}\right] \right|^2 \\ &= ({\rm Tr}[W_1\ \mathcal{Q}^\ell])^2, \end{aligned} \tag{A7}$$

with W_1 and $\mathcal{Q}$ defined as in Eq. (35).

3. Some useful identities

In this section we derive two inequalities which are not used in the main derivation but which allow us to better characterize the various operators that enter into our analysis. First of all we observe that

$$W_0 = \sum_{\vec{j}} p_{\vec{j}} P_{\vec{j}} \leqslant \rho^{\otimes n}\ 2^{n[S(\rho)-\chi(\mathcal{E})+\delta]}, \tag{A8}$$

which follows from the following chain of inequalities:

$$\begin{aligned} W_0 &= \sum_{\vec{j}} p_{\vec{j}} P_{\vec{j}} = \sum_{\vec{j}} p_{\vec{j}} \sum_{\vec{k}\in\mathcal{K}_{\vec{j}}} |e_{\vec{k}}^{(\vec{j})}\rangle\langle e_{\vec{k}}^{(\vec{j})}| \\ &\leqslant \sum_{\vec{j}} p_{\vec{j}} \sum_{\vec{k}\in\mathcal{K}_{\vec{j}}} |e_{\vec{k}}^{(\vec{j})}\rangle\langle e_{\vec{k}}^{(\vec{j})}| \lambda_{\vec{k}}^{(\vec{j})}\ 2^{n[S(\rho)-\chi(\mathcal{E})+\delta]} \\ &\leqslant \sum_{\vec{j}} p_{\vec{j}} \sum_{\vec{k}} |e_{\vec{k}}^{(\vec{j})}\rangle\langle e_{\vec{k}}^{(\vec{j})}| \lambda_{\vec{k}}^{(\vec{j})}\ 2^{n[S(\rho)-\chi(\mathcal{E})+\delta]} \\ &= \sum_{\vec{j}} p_{\vec{j}}\ \rho_{\vec{j}} 2^{n[S(\rho)-\chi(\mathcal{E})+\delta]} \\ &= \rho^{\otimes n} 2^{n[S(\rho)-\chi(\mathcal{E})+\delta]}, \end{aligned}$$

where we used Eq. (18). We can also prove the following identity:

$$\begin{aligned} \mathcal{Q} &= \sum_{\vec{j}} p_{\vec{j}} \bar{Q}_{\vec{j}_\ell} = P(\mathbb{1} - W_0)P \\ &\geqslant P(\mathbb{1} - \rho^{\otimes n} 2^{n[S(\rho)-\chi(\mathcal{E})+\delta]})P \\ &\geqslant P(1 - 2^{-n[\chi(\mathcal{E})-2\delta]}), \end{aligned} \tag{A9}$$

which follows by using Eq. (13). Notice that due to Eq. (A8) this also gives

$$P W_0 P \leqslant P\ 2^{-n[\chi(\mathcal{E})-2\delta]}. \tag{A10}$$

4. Characterization of the function f_z

We start by deriving the inequalities of Eq. (43) first. To do this, we observe that for all positive ϵ' we can write

$$\sum_{\vec{j}} p_{\vec{j}} {\rm Tr}[\rho_{\vec{j}}(\mathbb{1} - P_{\vec{j}})P] \leqslant \sum_{\vec{j}} p_{\vec{j}} {\rm Tr}[\rho_{\vec{j}}(\mathbb{1} - P_{\vec{j}})] < \epsilon',$$

where the first inequality follows by simply noticing that $\rho_{\vec{j}}(\mathbb{1} - P_{\vec{j}})$ is positive semidefinite (the two operators commute), while the last is just Eq. (21) which holds for sufficiently large n. Reorganizing the terms and using Eq. (14) this finally yields

$$\begin{aligned} f_0 = {\rm Tr}[W_1 P] &> \sum_{\vec{j}} p_{\vec{j}} {\rm Tr}[\rho_{\vec{j}} P] - \epsilon' \\ &= {\rm Tr}[\rho^{\otimes n} P] - \epsilon' > 1 - 2\epsilon', \end{aligned} \tag{A11}$$

which corresponds to the lefttmost inequality of Eq. (43) by setting $\epsilon = 2\epsilon'$. The rightmost inequality instead follows simply by observing that

$$f_0 = {\rm Tr}[W_1 P] \leqslant {\rm Tr}[W_1] = \sum_{\vec{j}} p_{\vec{j}} {\rm Tr}[P_{\vec{j}} \rho_{\vec{j}}] \leqslant 1. \tag{A12}$$

To prove the inequality (42) we finally notice that for $z \geqslant 1$ we can write

$$\begin{aligned} f_z &= {\rm Tr}\left[W_1 P \bar{W}_0^z\right] = {\rm Tr}\left[W_1 \bar{W}_0^z\right] \\ &= {\rm Tr}\left[\sqrt{W_1} \bar{W}_0^{\frac{z-1}{2}} \bar{W}_0 \bar{W}_0^{\frac{z-1}{2}} \sqrt{W_1}\right] \\ &\leqslant {\rm Tr}\left[\sqrt{W_1} \bar{W}_0^{\frac{z-1}{2}} P \bar{W}_0^{\frac{z-1}{2}} \sqrt{W_1}\right] 2^{-n[\chi(\mathcal{E})-2\delta]} \\ &\leqslant {\rm Tr}\left[\sqrt{W_1} \bar{W}_0^{\frac{z-1}{2}} \bar{W}_0^{\frac{z-1}{2}} \sqrt{W_1}\right] 2^{-n[\chi(\mathcal{E})-2\delta]} \\ &= {\rm Tr}\left[W_1 \bar{W}_0^{z-1}\right] 2^{-n[\chi(\mathcal{E})-2\delta]} = f_{z-1} 2^{-n[\chi(\mathcal{E})-2\delta]}, \end{aligned}$$

where we used the fact that the operators W_1 and $\bar{W}_0$ are non-negative. The expression (42) then follows by simply reiterating the above inequality z times.

[1] A. S. Holevo, Probl. Peredachi Inf. **9**, 3 (1973); Probl. Inf. Transm. (Engl. Transl.) **9**, 110 (1973).

[2] A. S. Holevo, IEEE Trans. Inf. Theory **44**, 269 (1998).

[3] B. Schumacher and M. D. Westmoreland, Phys. Rev. A **56**, 131 (1997); P. Hausladen, R. Jozsa, B. W. Schumacher, M. Westmoreland, and W. K. Wootters, *ibid.* **54**, 1869 (1996).

[4] M. B. Hastings, Nat. Phys. **5**, 255 (2008).

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

[6] C. H. Bennett and P. W. Shor, IEEE Trans. Inf. Theory **44**, 2724 (1998).

[7] A. S. Holevo, e-print arXiv:quant-ph/9809023 [see also Tamagawa University Research Review, no. 4] (1998).

[8] A. Winter, IEEE Trans. Inf. Theory **45**, 2481 (1999).
[9] T. Ogawa, Ph.D. dissertation, University of Electro-Communications, Tokyo, Japan, 2000; (in Japanese) T. Ogawa and H. Nagaoka, in *Proceedings of the 2002 IEEE International Symposium on Information Theory, Lausanne, Switzerland*, (IEEE, New, York, 2002), p. 73; T. Ogawa, IEEE Trans. Inf. Theory **45**, 2486 (1999).
[10] T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory **53**, 2261 (2007).
[11] M. Hayashi and H. Nagaoka, IEEE Trans. Inf. Theory **49**, 1753 (2003).
[12] M. Hayashi, Phys. Rev. A **76**, 062301 (2007); Commun. Math. Phys. **289**, 1087 (2009).
[13] P. Hausladen and W. K. Wooters, J. Mod. Opt. **41**, 2385 (1994).
[14] F. Hiai and D. Petz, Commun. Math. Phys. **143**, 99 (1991); T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory **46**, 2428 (2000).
[15] S. Verdú and T. S. Han, IEEE Trans. Inf. Theory **40**, 1147 (1994); T. S. Han, *Information-Spectrum Methods in Information Theory* (Springer, Berlin, 2002).
[16] S. Lloyd, V. Giovannetti, and L. Maccone, Phys. Rev. Lett. **106**, 250501 (2011).
[17] J. Tyson, J. Math. Phys. **50**, 032106 (2009); Phys. Rev. A **79**, 032343 (2009).
[18] C. Mochon, Phys. Rev. A **73**, 032328 (2006).
[19] V. P. Belavkin, Stochastics **1**, 315 (1975); P. Belavkin, Radio Eng. Electron. Phys. **20**, 39 (1975); V. P. Belavkin and V. Maslov, *in Mathematical Aspects of Computer Engineering*, edited by V. Maslov (MIR, Moscow, 1987).
[20] M. Ban, J. Opt. B **4**, 143 (2002).
[21] T. S. Usuda, I. Takumi, M. Hata, and O. Hirota, Phys. Lett. A **256**, 104 (1999).
[22] Y. C. Eldar and G. David Forney, IEEE Trans. Inf. Theory **47**, 858 (2001).
[23] H. Barnum and E. Knill, J. Math. Phys. **43**, 2097 (2002).
[24] A. Montanaro, Commun. Math. Phys. **273**, 619 (2007).
[25] M. Ježek, J. Řeháček, and J. Fiurášek, Phys. Rev. A **65**, 060301(R) (2002); Z. Hradil, J. Řeháček, J. Fiurášek, and M. Ježek, in *Quantum State Estimation*, Lecture Notes in Physics No. 649 (Springer, Berlin, 2004), p. 163.
[26] P. Hayden, D. Leung, and G. Smith, Phys. Rev. A **71**, 062339 (2005).
[27] A. S. Kholevo, Teor. Veroyatn. Ee Primen. **23**, 429 (1978); Theor. Probab. Appl. **23**, 411 (1978).
[28] D. Kretschmann and R. F. Werner, Phys. Rev. A **72**, 062323 (2005).
[29] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).
[30] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).
[31] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).
[32] S. Lloyd, Phys. Rev. A **55**, 1613 (1997).
[33] P. W. Shor [http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/]; MSRI Workshop on Quantum Information, Berkeley, 2002.
[34] I. Devetak, IEEE Trans. Inf. Theory **51**, 44 (2005).
[35] P. Hayden, P. W. Shor, and A. Winter, Open Syst. Inf. Dyn. **15**, 71 (2008).