Achieving the Holevo bound via sequential measurements

V. Giovannetti
NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

S. Lloyd and L. Maccone
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

We present a new decoding procedure to transmit classical information in a quantum channel which, saturating asymptotically the Holevo bound, achieves the optimal rate of the communication line. Differently from 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.

arXiv:1012.0386v1 [quant-ph] 2 Dec 2010

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 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 Ref. [2, 3] several alternative versions of the asymptotic attainability of the Holevo bound have been presented so far (e.g. see Refs. [7–11] and references therein). The original proof [2, 3] was obtained 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, 12]) was introduced that allows for exact message recovery in the asymptotic limit infinitely long codewords. More recently, Ogawa and Nagaoka [9], and Hayashi and Nagaoka [10] proved the asymptotic achievability of the bound by establishing a formal connection with quantum hypothesis testing problem [13], and by generalizing a technique (the information-spectrum method) which was introduced by Verdú and Han [14] in the context of classical communication channel.

In this paper we analyze a new decoding procedure for classical communication in a quantum channel. Here we give a formal proof using conventional methods, whereas in [15] we give a more intuitive take on the argument. Our decoding procedure allows for a new proof of the asymptotic attainability of the Holevo bound. As in Refs. [2, 3] it is based on the notion of typical subspace but it replaces the PGM scheme with a sequential decoding strategy in which, similarly to the quantum hypothesis testing approach of Ref. [9], the received quantum codeword undergoes to a sequence of simple YES/NO projective measurements which try to determine which among all possible inputs my have originated it. To prove that this strategy attains the bound we compute its associated *average* error probability and show that it 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 main advantage of our scheme resides on the fact that, differently from PGM and its variants [12, 16–26], it allows for a simple intuitive description, it clarifies the role of entanglement in the decoding procedure [15], its analysis avoids some technicalities, and it appears to be more suited for practical implementations.

The paper is organized as follows: in Sec. II we set the problem and present the scheme in an informal, non technical way. 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 new 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 in three logically distinct stages: the *encoding* stage in which the sender of the message (say, Alice) maps the classical information she wish 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 $j_1, \cdots, j_n$ 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 [28] 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 Measurement (POVM) [28] 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 message, can now be expressed as, e.g. see Ref. [2],

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

In the limit infinitely long sequences $n \to \infty$, it is known [2, 3, 7–10] that P_{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

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

with $S(\cdot) := -\mathrm{Tr}[(\cdot)\log_2(\cdot)]$ being the von Neumann entropy [28]. 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 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 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. [10]). Within this framework, the proof [2, 3] exploited the random coding trick by Shannon in which the POVM is shown to provide exponential small error probability *in 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. Namely, Bob performs a sequence of measurements to test for each of the codewords. Specifically, he performs a first YES/NO 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 measure aimed to verify whether or not it corresponds to the second elements 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 consequence, to avoid that the accumulated errors diverge as the detection proceeds, the YES/NO measurements needs 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 [29]: specifically our YES/NO measurements will be mild modifications of von Neumann projections on the typical subspaces of the codewords, in which their non exact orthogonality is smoothed away by rescaling them through further projection on the typical subspace of the source average message (see Sec. IV for details).

[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-longs *blocks* of carriers: for each possible choice of m, and for each possible coding/decoding strategy one define the error probability as in Eq. (2). The optimal transmission rate (i.e. the capacity of the channel) is also expressible as in the rhs term of Eq. (3) via proper regularization over m (this is a consequence of the super-additivity 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. [27].

[2] See footnote 1.

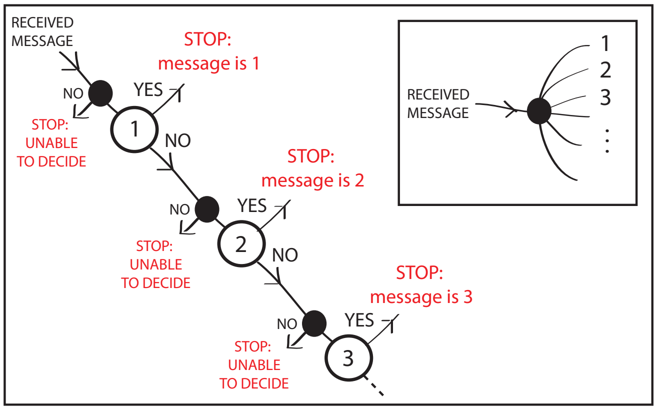

Figure 1: Flowchart representation of the detection scheme: the projections on the typical subspace $\mathcal{H}_{typ}^{(n)}(\vec{j})$ of the codewords are represented by the open circles, while the projections on the typical subspace $\mathcal{H}_{typ}^{(n)}$ of the average message of the source are represented by the black circles. The inset describes the standard PGM decoding scheme which produces all the possible outcomes in a single step.

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 $\vec{j} := (j_1, \cdots, j_n) \in \mathcal{A}^n$). In strict analogy to Shannon information theory, one defines a 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 $\rho = \sum_\ell q_\ell |e_\ell\rangle\langle e_\ell|$ its spectral decomposition (i.e. $|e_\ell\rangle$ are the orthonormal basis of $\mathcal{H}$ formed by the eigenvectors of ρ while q_ℓ are their eigenvalues). For fixed $\delta > 0$, one defines [29] the typical subspace $\mathcal{H}_{typ}^{(n)}$ of ρ as the subspace of $\mathcal{H}^{\otimes n}$ spanned by those vectors

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

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 [30], the states $|e_{\vec{\ell}}\rangle$ defined above can be thought as those which, in average, contain the symbol $|e_\ell\rangle$ almost nq_ℓ times). Identifying with $\mathcal{L}$ the set of those vectors $\vec{\ell} = (\ell_1, \ell_2, \cdots, \ell_n)$ which satisfies Eq. (10), the projector P on $\mathcal{H}_{typ}^{(n)}$ 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 $\mathcal{H}_{typ}^{(n)}$ is exponentially depressed [29]. More precisely, for all $\epsilon > 0$ it is possible to identify a sufficiently large n_0 such 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 to 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 $\vec{k} := (k_1, \cdots, k_n)$ 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 $\mathcal{H}_{typ}^{(n)}(\vec{j})$ 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 $\mathcal{H}_{typ}^{(n)}(\vec{j})$ 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 function of the source $\mathcal{E}$ only (this of course does not imply that the subspace $\mathcal{H}_{typ}^{(n)}(\vec{j})$ 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 $\mathcal{H}_{typ}^{(n)}(\vec{j})$. 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 $\mathcal{H}_{typ}^{(n)}(\vec{j})$ (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,

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

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 as n increases in identifying the codewords. How can one prove that such a POVM exists? First of all let us remind 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 $\mathrm{Tr}[X_{\vec{j}'}\rho_{\vec{j}}]$ (the outcome associated with $\mathrm{Tr}[X_0\rho_{\vec{j}}]$ 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

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

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

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

$P(\mathbf{C})$ being the probability defined in Eq. (8). Proving that $\langle P_{err}\rangle$ nullifies 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 adopting the pretty good measurement detection scheme, i.e. the POVM of elements

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

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

where P is the projector (11) on the typical subspace of the average state of the source, for $\vec{j} \in \mathbf{C}$ the $P_{\vec{j}}$ are the projectors (20) associated with the codeword $\rho_{\vec{j}}$. With this choice one can verify that, for given ϵ there exist n sufficiently large such that Eq. (23) yields the inequality [2]

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

This implies 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 detection scheme.

As anticipated in the introduction, the idea is to determine the value of the label $\vec{j}$ associated with the received codeword $\rho_{\vec{j}}$, by checking whether or not such state pertains to the typical subspace of the element $\vec{j}$ of the selected code $\mathcal{C}$.

Specifically we proceed as follows

- first we fix an ordering of the codewords of $\mathcal{C}$ yielding the sequence $\vec{j}_1, \vec{j}_2, \vec{j}_3, \cdots, \vec{j}_N$ with $\vec{j}_u \in \mathcal{C}$ for all $u = 1, \cdots, N$ (this is not really relevant but it is useful to formalize the protocol);
- then Bob performs a YES/NO measurement that determines whether or not the received state is the typical subspace of the first codeword $\vec{j}_1$[3];
- if the answer is YES the protocol stops and Bob declares to have identified the received message as the first of the list (i.e. $\vec{j}_1$);
- if the answer is NO Bob, performs a YES/NO measurement to check whether or not the state is in the typical sub of $\vec{j}_2$;
- the protocol goes on, testing similarly for all N possibilities. In the end we will either determine an estimate of the transmitted $\vec{j}$ or we will get a null result (the messages has not been identified, corresponding to an error in the communication).

We now better specify the YES/NO measurements. Indeed, as mentioned earlier, we have to "smooth" them to account for the disturbance they might introduce in the process. For this purpose, each of such measurements will consist in two steps in which first we check (via a von Neumann projective measurement) whether or not the incoming state is in the typical subspace $\mathcal{H}_{typ}^{(n)}$ of the average message. Then we apply a von Neumann projective measurement on the typical subspace $\mathcal{H}_{typ}^{(n)}(\vec{j}_i)$ of the i-th codeword of Bob's list (see Fig. 1). Hence, the POVM elements are defined as follows. The first element E_1 tests if the transmitted state is in $\mathcal{H}_{typ}^{(n)}(\vec{j}_1)$, so 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). Similarly the remaining elements can be expressed as follows

$$\begin{aligned} E_2 &:= (\mathbb{1} - \bar{P}_{\vec{j}_1}) \bar{P}_{\vec{j}_2} (\mathbb{1} - \bar{P}_{\vec{j}_1}) , \qquad (29) \\ E_3 &:= (\mathbb{1} - \bar{P}_{\vec{j}_1})(\mathbb{1} - \bar{P}_{\vec{j}_2}) \bar{P}_{\vec{j}_3} (\mathbb{1} - \bar{P}_{\vec{j}_2})(\mathbb{1} - \bar{P}_{\vec{j}_1}) , \\ E_4 &:= \cdots , \end{aligned}$$

[3] It is worth stressing that in Ref. [15] this test was implemented by performing a series of rank-one projective measurements on to a basis of the subspace.

(see Appendix A for an explicit derivation). A compact expression can be derived by writing

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

where

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

with $Q_{\vec{j}}$ being the orthogonal complement of $P_{\vec{j}}$, i.e.

$$Q_{\vec{j}} := \mathbb{1} - P_{\vec{j}} . \tag{32}$$

With such definitions the associated average error probability (23) can then be expressed as,

$$\begin{aligned} \langle P_{err} \rangle &= \sum_{\vec{j}_1, \cdots, \vec{j}_N} \frac{p_{\vec{j}_1} \cdots p_{\vec{j}_N}}{N} \sum_{u=1}^{N} (1 - \mathrm{Tr}[M_u \rho_{\vec{j}_u} M_u^\dagger]) \\ &= 1 - \frac{1}{N} \sum_{\vec{j}} p_{\vec{j}} \sum_{\ell=0}^{N-1} \mathrm{Tr}[P_{\vec{j}} \ \Phi^\ell(\rho_{\vec{j}})] , \end{aligned} \tag{33}$$

where we used the fact that the summations over the various $\vec{j}_i$ are independent. In writing the above expression we introduced the following super-operator

$$\Phi(\Theta) := \sum_{\vec{j}} p_{\vec{j}} \, \bar{Q}_{\vec{j}} \, \Theta \, \bar{Q}_{\vec{j}} , \tag{34}$$

which is completely positive and trace decreasing, and we use the notation Φ^ℓ to indicate the ℓ-fold concatenation of super-operators, e.g. $\Phi^2(\cdot) = \Phi(\Phi(\cdot))$. It is worth noticing that the possibility of expressing $\langle P_{err} \rangle$ in term of a single super-operator follows directly from the average we have performed over all possible codes $\mathbf{C}$. For future reference we find it useful to cast Eq. (33) in a slightly different form by exploiting the the definitions of Eqs. (16) and (20). More precisely, we write

$$\begin{aligned} 1 - \langle P_{err} \rangle &= \sum_{\ell=0}^{N-1} \sum_{\vec{j}, \vec{j}_1, \cdots, \vec{j}_\ell} \frac{p_{\vec{j}} p_{\vec{j}_1} \cdots p_{\vec{j}_\ell}}{N} \\ &\quad \times \mathrm{Tr}[P_{\vec{j}} \bar{Q}_{\vec{j}_1} \cdots \bar{Q}_{\vec{j}_\ell} \rho_{\vec{j}} \bar{Q}_{\vec{j}_\ell} \cdots \bar{Q}_{\vec{j}_1}] \\ &= \sum_{\ell=0}^{N-1} \sum_{\vec{j}, \vec{j}_1, \cdots, \vec{j}_\ell} \sum_{\vec{k}} \sum_{\vec{k}' \in \mathcal{K}_{\vec{j}}} \lambda_{\vec{k}}^{(\vec{j})} \frac{p_{\vec{j}} p_{\vec{j}_1} \cdots p_{\vec{j}_\ell}}{N} \\ &\quad \times \left| \langle e_{\vec{k}'}^{(\vec{j})} | \bar{Q}_{\vec{j}_1} \cdots \bar{Q}_{\vec{j}_\ell} | e_{\vec{k}}^{(\vec{j})} \rangle \right|^2 . \end{aligned} \tag{35}$$

V. BOUNDS ON THE ERROR PROBABILITY

In this section we derive an upper limit for the error probability (33) which will lead us to the prove the achievability of the Holevo bound.

Specifically, we notice that

$$\begin{aligned}
&\sum_{\vec{k}} \sum_{\vec{k}'\in\mathcal{K}_{\vec{j}}} \lambda_{\vec{k}}^{(\vec{j})} \left|\langle e_{\vec{k}'}^{(\vec{j})}|\bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}|e_{\vec{k}}^{(\vec{j})}\rangle\right|^2 \\
&\geqslant \sum_{\vec{k}\in\mathcal{K}_{\vec{j}}} \lambda_{\vec{k}}^{(\vec{j})} \left|\langle e_{\vec{k}}^{(\vec{j})}|\bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}|e_{\vec{k}}^{(\vec{j})}\rangle\right|^2 \\
&= \sum_{\vec{k}\in\mathcal{K}_{\vec{j}}} \lambda_{\vec{k}}^{(\vec{j})} \left|\langle e_{\vec{k}}^{(\vec{j})}|\bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}|e_{\vec{k}}^{(\vec{j})}\rangle\right|^2 \sum_{\vec{k}} \lambda_{\vec{k}}^{(\vec{j})} \\
&\geqslant \Big| \sum_{\vec{k}\in\mathcal{K}_{\vec{j}}} \lambda_{\vec{k}}^{(\vec{j})} \langle e_{\vec{k}}^{(\vec{j})}|\bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}|e_{\vec{k}}^{(\vec{j})}\rangle\Big|^2 \\
&= \left|\mathrm{Tr}[P_{\vec{j}}\ \rho_{\vec{j}}\ P_{\vec{j}}\ \ \bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}]\right|^2 , \qquad (36)
\end{aligned}$$

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. (35) we can write

$$\begin{aligned}
&1-\langle P_{err}\rangle \geqslant \qquad (37)\\
&\sum_{\ell=0}^{N-1}\sum_{\vec{j},\vec{j}_1,\cdots,\vec{j}_\ell} \frac{p_{\vec{j}}p_{\vec{j}_1}\cdots p_{\vec{j}_\ell}}{N} \left|\mathrm{Tr}[P_{\vec{j}}\ \rho_{\vec{j}}\ P_{\vec{j}}\ \ \bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}]\right|^2 .
\end{aligned}$$

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

$$\begin{aligned}
&\sum_{\vec{j},\vec{j}_1,\cdots,\vec{j}_\ell} p_{\vec{j}}p_{\vec{j}_1}\cdots p_{\vec{j}_\ell} \left|\mathrm{Tr}[P_{\vec{j}}\rho_{\vec{j}}P_{\vec{j}}\ \ \bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}]\right|^2 \\
&\geqslant \Big|\sum_{\vec{j},\vec{j}_1,\cdots,\vec{j}_\ell} p_{\vec{j}}p_{\vec{j}_1}\cdots p_{\vec{j}_\ell}\ \mathrm{Tr}[P_{\vec{j}}\rho_{\vec{j}}P_{\vec{j}}\ \ \bar{Q}_{\vec{j}_1}\cdots\bar{Q}_{\vec{j}_\ell}]\Big|^2 \\
&\qquad\qquad = \left(\mathrm{Tr}[W_1\ \mathcal{Q}^\ell]\right)^2 , \qquad (38)
\end{aligned}$$

where for q integer we defined

$$W_q := \sum_{\vec{j}} p_{\vec{j}}\ P_{\vec{j}}\ \rho_{\vec{j}}^q\ P_{\vec{j}} , \qquad (39)$$

$$\mathcal{Q} := \sum_{\vec{j}} p_{\vec{j}}\ \bar{Q}_{\vec{j}_\ell} = \bar{\mathbb{1}} - \bar{W}_0 , \qquad (40)$$

(notice that W_0 *is not* $\rho^{\otimes n}$, e.g. see Eq. (B1)). Therefore one gets

$$1-\langle P_{err}\rangle \geqslant \frac{1}{N}\sum_{\ell=0}^{N-1}\left|\mathrm{Tr}[W_1\ \mathcal{Q}^\ell]\right|^2 . \qquad (41)$$

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

$$\mathbb{1} \geqslant \mathcal{Q} \geqslant 0 . \qquad (42)$$

Both properties simply follow from the identity

$$\mathcal{Q} = P(\mathbb{1} - \sum_{\vec{j}} p_{\vec{j}}P_{\vec{j}})P = P\big[\sum_{\vec{j}} p_{\vec{j}}\ (\mathbb{1}-P_{\vec{j}})\big]P , \qquad (43)$$

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

$$\mathbb{1} \geqslant W_1 \geqslant W_0\ \ 2^{-n(S(\rho)-\chi(\mathcal{E})+\delta)} \geqslant 0 , \qquad (44)$$

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. (41) is always smaller than one and that it is decreasing with ℓ. An explicit proof of this fact is as follows

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

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 B). A further simplification of the bound can be obtained by replacing the terms in the summation of Eq. (41) with the smallest addendum. This yields

$$1-\langle P_{err}\rangle \geqslant |A|^2 , \qquad (45)$$

where, using the fact that $\bar{\mathbb{1}}^2 = \bar{\mathbb{1}} = P$, we defined

$$A := \mathrm{Tr}[W_1\mathcal{Q}^{N-1}] = \sum_{z=0}^{N-1}\binom{N-1}{z}(-1)^z\ f_z, \qquad (46)$$

$$f_z := \mathrm{Tr}[W_1\ P\ \bar{W}_0^z] . \qquad (47)$$

It turns out that the quantities f_z defined above are positive, smaller than one, and decreasing in z. Indeed as shown in the Appendix C they satisfy the inequalities

$$0\leqslant f_z \leqslant f_0\ \ 2^{-nz(\chi(\mathcal{E})-2\delta)} \qquad \text{for all integer } z, \qquad (48)$$

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

$$1-\epsilon \leqslant\ f_0\ \leqslant 1 . \qquad (49)$$

Using these expressions, we can derive the following bound on A, i.e.

$$\begin{aligned}
A &= f_0 + \sum_{z=1}^{N-1}\binom{N-1}{z}(-1)^z f_z \\
&\geqslant f_0 - \sum_{z=1}^{N-1}\binom{N-1}{z} f_z = 2f_0 - \sum_{z=0}^{N-1}\binom{N-1}{z} f_z \\
&\geqslant 2f_0 - f_0\sum_{z=0}^{N-1}\binom{N-1}{z} 2^{-nz(\chi(\mathcal{E})-2\delta)} \\
&= f_0\,[2-(1+2^{-n(\chi(\mathcal{E})-2\delta)})^{N-1}] , \qquad (50)
\end{aligned}$$

where in the first inequality we get a bound by taking all the terms of $k \geqslant 1$ with the negative sign, the second from (48). Now, on one hand if N is too large the quantity on the rhs side will become negative as we are taking the N 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 parenthesis 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 one for large n. To study such threshold we rewrite Eq. (50) as

$$A \geqslant f_0[2 - Y(x = 2^{\chi(\mathcal{E})-2\delta}, y = N, n)] \;, \tag{51}$$

where we defined

$$Y(x,y,n) := (1+x^{-n})^{y^n-1} \;. \tag{52}$$

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

$$\lim_{n\to\infty} \log[Y(x,y,n)] = \frac{\log x}{\log y} \lim_{n\to\infty} \left(\frac{y}{x}\right)^n \;. \tag{53}$$

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

$$R < \chi(\mathcal{E}) - 2\delta \;, \tag{54}$$

the quantity on the rhs of Eq. (51) approaches f_0 as n increases (this corresponds to having $y < x$ in the Y function). Reminding then Eq. (49) we get

$$1 - \langle P_{err}\rangle \geqslant |A|^2 > f_0^2 > |1-\epsilon|^2 > 1 - 2\epsilon \;, \tag{55}$$

and thus

$$\langle P_{err}\rangle < 2\epsilon \;. \tag{56}$$

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 $\langle P_{err}\rangle$.

To summarize, we have shown that adopting the sequential detection strategy defined in Sec. IV we can conclude that it is possible to send $N = 2^{nR}$ messages with asymptotically vanishing error probability, for all rates R which satisfy the condition (54). ■

VI. CONCLUSIONS

To summarize: the above analysis provides an explicit upper bound for the averaged error probability of the new 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 bound close to zero for codes generated by sources $\mathcal{E}$ which have strictly less than $2^{n\chi(\mathcal{E})}$ elements. In other words, our new detection scheme provides an alternative demonstration of the achievability of the Holevo bound [2].

An interesting open question is to extend the technique presented here to a decoding procedure that can achieve the quantum capacity of a channel [31–34].

Acknowledgments

VG is grateful to P. Hayden, A. S. Holevo, K. Matsumoto, J. Tyson and A. Winter for comments and discussions.

VG acknowledges support from the FIRB-IDEAS project under the contract RBID08B3FM and support of Institut Mittag-Leffler (Stockholm), where he was visiting while part of this work was done. SL was supported by the WM Keck Foundation, DARPA, NSF, and NEC. LM was supported by the WM Keck Foundation.

Appendix A: 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

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

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 $\mathcal{H}_{typ}^{(n)}$ 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 $\mathcal{H}_{typ}^{(n)}(\vec{j}_1)$ of $\rho_{\vec{j}_1}$. Accordingly the joint probability of success of finding $|\Psi\rangle$ in $\mathcal{H}_{typ}^{(n)}$ and *then* in $\mathcal{H}_{typ}^{(n)}(\vec{j}_1)$ is given by

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

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 $\mathcal{H}_{typ}^{(n)}$ and *then not* in $\mathcal{H}_{typ}^{(n)}(\vec{j}_1)$ 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 in $\mathcal{H}_{typ}^{(n)}$ is $\langle\Psi|\mathbb{1}-P|\Psi\rangle$. Let us now consider the second step of the protocol where Bob checks wether or not the message is in the typical subspace $\mathcal{H}_{typ}^{(n)}(\vec{j}_2)$ 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,

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

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

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

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 Eq. (A2) and (A4)) that it is a quite different POVM from the conventionally used pretty good measurement [2, 3].

Appendix B: Some useful identities

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

$$W_0 = \sum_{\vec{j}} p_{\vec{j}} P_{\vec{j}} \leqslant \rho^{\otimes n}\ 2^{n(S(\rho)-\chi(\mathcal{E})+\delta)}\ , \tag{B1}$$

which follows by 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

$$\begin{aligned}\mathcal{Q} &= \sum_{\vec{j}} p_{\vec{j}}\ \bar{Q}_{\vec{j}_\ell} = P(\mathbb{1}-W_0)P\\ &\geqslant P(\mathbb{1}-\rho^{\otimes n}\ 2^{n(S(\rho)-\chi(\mathcal{E})+\delta)})P\\ &\geqslant P\ (1-2^{-n(\chi(\mathcal{E})-2\delta)})\ ,\end{aligned} \tag{B2}$$

which follows by using Eq. (13). Notice that due to Eq. (B1) this also gives

$$PW_0P \leqslant P\ 2^{-n(\chi(\mathcal{E})-2\delta)}\ . \tag{B3}$$

Appendix C: Characterization of the function f_z

We start deriving the inequalities of Eq. (49) first. To do we observe that for all ϵ' positive we can write

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

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

$$\begin{aligned}f_0 = \mathrm{Tr}[W_1\ P]\ &>\ \sum_{\vec{j}} p_{\vec{j}}\ \mathrm{Tr}[\rho_{\vec{j}}P] - \epsilon'\\ &= \mathrm{Tr}[\rho^{\otimes n}P] - \epsilon' > 1-2\epsilon'\ ,\end{aligned} \tag{C1}$$

which corresponds to the lefttmost inequality of Eq. (49) by setting $\epsilon = 2\epsilon'$. The rightmost inequality instead follows simply by observing that

$$f_0 = \mathrm{Tr}[W_1\ P] \leqslant \mathrm{Tr}[W_1] = \sum_{\vec{j}} p_{\vec{j}} \mathrm{Tr}[P_{\vec{j}}\rho_{\vec{j}}] \leqslant 1\ . \tag{C2}$$

To prove the inequality (48) we finally notice that for $z \geqslant 1$ we can write

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

where we used the fact that the operators operators W_1, $\bar{W}_0$ are non negative. The expression (48) then follows by simply reiterating the above inequality z times.

[1] A. S. Holevo, "Some estimates of the information transmitted by quantum communication channel," *Probl. Pered. Inform.* **9**, 3 (1973) [Probl. Inf. Transm. **9**, 110 (1973)].

[2] A. S. Holevo, "The Capacity of the Quantum Channel with General Signal States," *IEEE Trans. Inform. Theory* **44**, 269-273, (1998).

[3] B.W. Schumacher and M. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A* **56**, 131-138 (1997); P. Hausladen, R. Jozsa, B.W. Schumacher, M. Westmoreland, and W.K. Wootters, "Classical information capacity of a quantum channel," *Phys. Rev. A* **54**, 1869-1876 (1996).

[4] M. B. Hastings, "Superadditivity of communication capacity using entangled inputs," *Nature Phys.* **5**, 255-257 (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, "Quantum Information Theory", IEEE Trans. Inf. Theory **44**, 2724-2742 (1998).

[7] A. S. Holevo, "Coding Theorems for Quantum Channels", arXiv:quant-ph/9809023 [see also Tamagawa University Research Review, no. 4, 1998].

[8] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Trans. Inform. Theory,* vol. **45**, 2481-2485 (1999).

[9] T. Ogawa, *A study on the asymptotic property of the hypothesis testing and the channel coding in quantum mechanical systems*, Ph.D. dissertation (in Japanese), Univ. Electro-Communications, Tokyo, Japan, 2000; T. Ogawa and H. Nagaoka, "A new proof of the channel coding theorem via hypothesis testing in quantum information theory", in *Proc. 2002 IEEE Int. Symp. Information Theory, Lausanne, Switzerland*, June/July 2002, p. 73; "Strong converse to the quantum channel coding theorem," *IEEE Trans.Info.Theor.* **45** 2486-2489, (1999).

[10] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Trans. Inform. Theory* **49**, 1753-1768 (2003).

[11] M. Hayashi, "Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding", *Phys. Rev. A* **76**, 062301 (2007); M. Hayashi, "Universal coding for classical-quantum channel", *Commun. Math. Phys.* **289**, 1087-1098 (2009).

[12] P. Hausladen and W. K. Wooters, "A 'Pretty Good' Measurement for Distinguishing Quantum States", *J. Mod. Opt.* **41**, 2385-2390 (1994).

[13] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," *Commun. Math. Phys.* **143**, 99-114 (1991); T. Ogawa and H. Nagaoka, "Strong converse and Stein's lemma in quantum hypothesis testing," *IEEE Trans. Inf. Theory* **46**, 2428-2433 (2000).

[14] S. Verdú and T. S. Han, "A general formula for channel capacity", *IEEE Trans. Inform. Theory,* **40**, 1147-1157 (1994); T. S. Han, *Information-Spectrum Methods in Information Theory* (Springer, Berlin, 2002).

[15] S. Lloyd, V. Giovannetti, and L. Maccone, unpublished (2010).

[16] J. Tyson, "Two-sided estimates of minimum-error distinguishability of mixed quantum states via generalized Holevo-Curlander bounds", *J. Math. Phys.* **50**, 032106 (2009); "Error rates of Belavkin weighted quantum measurements and a converse to Holevo's asymptotic optimality theorem", *Phys. Rev. A* **79**, 032343 (2009).

[17] C. Mocohon, "Family of generalized 'pretty good' measurements and the minimal-error pure-state discrimination problems for which they are optimal," *Phys. Rev. A* **73**, 032328 (2006).

[18] V. P. Belavkin, "Optimal multiple quantum statistical hypothesis testing," *Stochastics* **1**, 315-345 (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).

[19] M. Ban, "Optimal signal detection in entanglement-assisted quantum communication systems" *J. Opt. B: Quantum Semiclassical Opt.* **4**, 143 (2002).

[20] T. S. Usuda, I. Takumi, M. Hata, and O. Hirota, "Minimum error detection of classical linear code sending through a quantum channel," *Phys. Lett. A* **256**, 104-108 (1999).

[21] Y. C. Eldar and G. David Forney, "On quantum detection and the square-root measurement", *IEEE Trans. Inf. Theory* **47**, 858-872 (2001).

[22] H. Barnum and E. Knill, "Reversing quantum dynamics with near-optimal quantum and classical fidelity," *J. Math. Phys.* **43**, 2097-3017 (2002).

[23] A. Montanaro, "On the Distinguishability of Random Quantum States," *Commun. Math. Phys.* **273**, 619-636 (2007).

[24] M. Jĕzek, J. Řeháček, and J. Fiurášek, "Finding optimal strategies for minimum-error quantum-state discrimination," *Phys. Rev. A* **65**, 060301(R) (2002); Z. Hradil, J. Řeháček, J. Fiurášek, and M. Jĕzek, "Maximum-Likelihood Methods in Quantum Mechanics," *Lect. Notes Phys.* **649**, 163-172 (2004).

[25] P. Hayden, D. Leung, and G. Smith, "Multiparty data hiding of quantum information," *Phys. Rev. A* **71**, 062339 (2005).

[26] A. S. Kholevo, Teor. Veroyatn. Ee Primen. **23**, 429 (1978); [Theor. Probab. Appl. **23**, 411 (1978)].

[27] D. Kretschmann and R. F. Werner, "Quantum channels with memory,", *Phys. Rev. A* **72**, 062323 (2005).

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

[29] B. Schumacher, "Quantum coding," *Phys. Rev. A* **51**, 2738-2747 (1994).

[30] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, **27**, 379-423 (1948); *ib.* **27** 623-656 (1948).

[31] S. Lloyd, "Capacity of the noisy quantum channel", *Phys. Rev. A* **55**, 1613-1622 (1997).

[32] P. W. Shor, "The quantum channel capacity and coherent information", unpublished lecture notes. Online at `http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/`; MSRI Workshop on Quantum Information, Berkeley, 2002.

[33] I. Devetak, "The Private Classical Capacity and Quantum Capacity of a Quantum Channel", IEEE Trans. Inf. Theory **51**, 44-55 (2005).

[34] P. Hayden, P. W. Shor, A. Winter, "Random quantum codes from Gaussian ensembles and an uncertainty relation", Open Syst. Inf. Dyn. **15**, 71-89 (2008).