Sequential Projective Measurements for Channel Decoding

Seth Lloyd,[1] Vittorio Giovannetti,[2] and Lorenzo Maccone[3]

[1]*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2]*NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, piazza dei Cavalieri 7, I-56126 Pisa, Italy*
[3]*Dipartimento Fisica "A. Volta," INFN Sezione Pavia, Università di Pavia, via Bassi 6, I-27100 Pavia, Italy*

(Received 30 November 2010; published 20 June 2011)

We study the transmission of classical information in quantum channels. We present a decoding procedure that is very simple but still achieves the channel capacity. It is used to give an alternative straightforward proof that the classical capacity is given by the regularized Holevo bound. This procedure uses only projective measurements and is based on successive "yes-no" tests only.

DOI: 10.1103/PhysRevLett.106.250501 PACS numbers: 03.67.Hk, 03.67.Ac, 89.70.Kn

According to quantum information theory, to transfer classical signals we must encode them into the states of quantum information carriers, transmit these through the (possibly noisy) communication channel, and then decode the information at the channel output [1]. Frequently, even if no entanglement between successive information carriers is employed in the encoding or is generated by the channel, a joint measurement procedure is necessary (e.g., see [2]) to achieve the capacity of the communication line, i.e., the maximum transmission rate per channel use [1]. This is clear from the original proofs [3,4] that the classical channel capacity is provided by the regularization of the Holevo bound [5]: These proofs employ a decoding procedure based on detection schemes (the "pretty good measurement" or its variants [6–17]). Alternative decoding schemes were also derived in Ref. [18] by using a iterative scheme which, given any good small code, allows one to increase the number of transmitted messages up to the size set by the bound and in Refs. [19–22] with an application of quantum hypothesis testing (which was introduced in this context in Refs. [19,23] for the quantum and classical setting, respectively). Here we present a simple decoding procedure which uses only dichotomic projective measurements acting on the channel output, which is nonetheless able to achieve the channel capacity for transmission of classical information through a quantum channel. Our procedure sidesteps most of the technicalities associated with similar prior proofs.

The main idea is that even if the possible alphabet states (i.e., the states of a single information carrier) are not orthogonal at the output of the channel, the code words composed of a long sequence of alphabet states approach orthogonality asymptotically, as the number of letters in each code word goes to infinity. Thus, one can sequentially test whether each code word is at the output of the channel. When one gets the answer "yes," the probability of error is small (as the other code words have little overlap with the tested one). When one gets the answer "no," the state has been ruined very little and can be still employed to further test for the other code words. To reduce the accumulation of errors during a long sequence of tests that yield no answers, every time a no is obtained, we have to project the state back to the space that contains the typical output of the channel. In summary, the procedure is (i) test whether the channel output is the first code word; (ii) if yes, we are done; if no, then project the system into the typical subspace and abort with an error if the projection fails; (iii) repeat the above procedure for all the other code words until we get a yes (or abort with an error if we test all of them without getting yes); (iv) in the end, we identified the code word that was sent or we had to abort.

After reviewing some basic notions on typicality, we will prove that the above procedure succeeds in achieving the classical capacity of the channel by focusing on an implementation where yes-no projective measurements are employed to test *randomly* for *each single base vector* of the typical subspaces. An alternative proof referring to this same procedure is presented in Ref. [24] by using a decoding strategy where instead one discriminates directly among the various typical subspaces of the code words through a *deterministic* (not random) sequence of yes-no projective measurements which do not discriminate among the basis vectors of each subspace.

Definitions and review.—For notational simplicity we will consider code words composed of unentangled states. For general channels, entangled code words must be used to achieve capacity [25], but the extension of our theory to this case is straightforward (replacing the Holevo bound with its regularized version).

Consider a quantum channel that is fed with a letter j from a classical alphabet with probability p_j. The letter j is encoded into a state of the information carriers which is evolved by the channel into an output $\rho_j = \sum_k p_{k|j} |k\rangle_j\langle k|$, where ${}_j\langle k'|k\rangle_j = \delta_{k'k}$. Hence, the average output is

$$\rho = \sum_j p_j \rho_j = \sum_{j,k} p_j p_{k|j} |k\rangle_j\langle k| = \sum_k p_k |k\rangle\langle k|, \quad (1)$$

where $|k\rangle_j$ and $|k\rangle$ are the eigenvectors of the jth output-alphabet density matrix and of the average output, respectively. The subtleties of quantum channel decoding arise

because the ρ_j typically commute neither with each other nor with ρ. The Holevo-Schumacher-Westmoreland theorem [3,4] implies that we can send classical information reliably down the channel at a rate (bits per channel use) given by the Holevo quantity [5]

$$\chi \equiv S(\rho) - \sum_j p_j S(\rho_j), \tag{2}$$

where $S(\cdot) \equiv -\mathrm{Tr}[(\cdot)\log_2(\cdot)]$ is the von Neumann entropy. This rate can be asymptotically attained in the multichannel uses scenario as $\lim_{n\to\infty}(\log_2 N_n)/n$, where a set $\mathcal{C}_n$ of N_n code words $\vec{j} = (j_1, \ldots, j_n)$ formed by long sequences of the letters j are used to reliably transfer N_n distinct classical messages. Similarly to the Shannon random-coding theory [26], the code words $\vec{j} \in \mathcal{C}_n$ can be chosen at random among the *typical sequences* generated by the probability p_j, in which each letter j of the alphabet occurs approximately $p_j n$ times. As mentioned above, the Holevo-Schumacher-Westmoreland theorem uses the pretty good measurement to decode the code words of $\mathcal{C}_n$ at the output of the channel. We will now show that a sequence of binary projective measurements suffices [27].

Sequential measurements for channel decoding.—The channel output state $\rho_{\vec{j}} \equiv \rho_{j_1} \otimes \cdots \otimes \rho_{j_n}$ associated to a generic typical sequence $\vec{j} = (j_1, \ldots, j_n)$ possesses a *typical subspace* $\mathcal{H}_{\vec{j}}$ spanned by the vectors $|k_1\rangle_{j_1} \ldots |k_n\rangle_{j_n} \equiv |\vec{k}\rangle_{\vec{j}}$, where $|k\rangle_j$ occurs approximately $p_j p_{k|j} n = p_{jk} n$ times; e.g., see Ref. [3]. The subspace $\mathcal{H}_{\vec{j}}$ has dimensions $\sim 2^{n\sum_j p_j S(\rho_j)}$ independent of the input $\vec{j} \in \mathcal{C}_n$. Moreover, a typical output subspace $\mathcal{H}$ and a projector P onto it exist such that, for any $\epsilon > 0$ and sufficiently large n,

$$\mathrm{Tr}\,\bar{\rho} > 1 - \epsilon, \tag{3}$$

where $\bar{\rho} \equiv P\rho \otimes \cdots \otimes \rho P$ is the projection of the n-output average density matrix onto $\mathcal{H}$. Notice that $\mathcal{H}$ and the $\mathcal{H}_{\vec{j}}$'s in general differ. Typicality for $\mathcal{H}$ implies that, for $\delta > 0$ and sufficiently large n, the eigenvalues λ_i of $\bar{\rho}$ and the dimension of $\mathcal{H}$ are bounded as [3,4]

$$\lambda_i \le 2^{-n[S(\rho)-\delta]}, \tag{4}$$

$$\#\ \text{nonzero eigenvalues} = \dim(\mathcal{H}) \le 2^{n[S(\rho)+\delta]}. \tag{5}$$

Define then the operator

$$\tilde{\rho} = P\Big(\sum_{\vec{j},\vec{k}\in typ} p_{\vec{j}} p_{\vec{k}|\vec{j}} |\vec{k}\rangle_{\vec{j}}\langle\vec{k}|\Big)P \le \bar{\rho}, \tag{6}$$

where the inequality follows because the summation is restricted to the $\vec{j}$'s that are typical sequences of the classical source and to the states $|\vec{k}\rangle_{\vec{j}}$ which span the typical subspace of the $\vec{j}$th output. [Without these limitations, the inequality would be replaced by an equality.] Then, the maximum eigenvalue of $\tilde{\rho}$ is no greater than that of $\bar{\rho}$, while the number of nonzero eigenvalues of $\tilde{\rho}$ cannot be greater than those of $\bar{\rho}$; i.e., Eqs. (3)–(5) apply also to $\tilde{\rho}$.

Now we come to our main result. To distinguish between the N_n distinct code words of $\mathcal{C}_n$, we perform sequential von Neumann measurements corresponding to projections onto the possible outputs $|\vec{k}\rangle_{\vec{j}}$ to find the channel input (as shown in Ref. [24], these can also be replaced by joint projectors on the spaces $\mathcal{H}_{\vec{j}}$). In between these measurements, we perform von Neumann measurements that project onto the typical output subspace $\mathcal{H}$.

We will show that, as long as the rate at which we send information down the channel is bounded above by the Holevo quantity (2), these measurements identify the proper input to the channel with probability one in the limit that the number of uses of the channel goes to infinity. That is, we send information down the channel at a rate R smaller than χ, so that there are $N_n \simeq 2^{nR}$ possible randomly selected code words $\vec{j}$ that could be sent down over n uses. Each code word gives rise to $\sim 2^{n\sum_j p_j S(\rho_j)}$ possible typical outputs $|\vec{k}\rangle_{\vec{j}}$. As always with Shannon-like random-coding arguments [26], our set of possible outputs occupy only a fraction $2^{-n(\chi-R)}$ of the full output space. This sparseness of the actual outputs in the full space is the key to obtaining asymptotic zero error probability: All our error probabilities will scale as $2^{-n(\chi-R)}$.

The code word sent down the channel is some typical sequence $\vec{j}$, which yields some typical output $|\vec{k}\rangle_{\vec{j}}$ with probability $p_{\vec{k}|\vec{j}}$. We begin with a von Neumann measurement corresponding to projectors P and $\mathbb{1} - P$ to check whether the output lies in the typical subspace $\mathcal{H}$. From Eq. (3) we can conclude that for any $\epsilon > 0$, for sufficiently large n, this measurement yields the result yes with probability larger than $1 - \epsilon$. We follow this with a binary projective measurement with projectors

$$P_{\vec{k}_1|\vec{j}_1} \equiv |\vec{k}_1\rangle_{\vec{j}_1}\langle\vec{k}_1|, \qquad \mathbb{1} - P_{\vec{k}_1|\vec{j}_1}, \tag{7}$$

to check whether the input was $\vec{j}_1$ and the output was $\vec{k}_1$. If this measurement yields the result yes, we conclude that the input was indeed $\vec{j}_1$. Usually, however, this measurement yields the result no. In this case, we perform another measurement to check for typicality and move on to a second trial output state, e.g., $|\vec{k}_2\rangle_{\vec{j}_1}$. If this measurement yields the result yes, we conclude that the input was $\vec{j}_1$. Usually, of course, the measurement yields the result no, and so we project again and move on to a third trial output state, $|\vec{k}_3\rangle_{\vec{j}_1}$, etc. Having exhausted the $O(2^{n\sum_k p_k S(\rho_k)})$ typical output states from the code word $\vec{j}_1$, we turn to the typical output states from the input $\vec{j}_2$, then $\vec{j}_3$, and so on, moving through the $N_n \simeq 2^{nR}$ code words until we eventually find a match. The maximum number of measurements that must be performed is hence

$$M \simeq 2^{nR} 2^{n\sum_k p_k S(\rho_k)}. \quad (8)$$

The probability amplitude that, after m trials without finding the correct state, we find it at the $m+1$th trial can then be expressed as

$$\mathcal{A}_m(\text{yes}) = {}_{\vec{j}}\langle \vec{k}|P(\mathbb{1}-P_{\ell_m})P\ldots P(\mathbb{1}-P_{\ell_1})P|\vec{k}\rangle_{\vec{j}}, \quad (9)$$

where for $q=1,\ldots,m$ the operators P_{ℓ_q} represent the first m elements $P_{\vec{k}_r|\vec{j}_s}$ that compose the decoding sequence of projectors. The error probability $P_{\text{err}}(\vec{j},\vec{k})$ of mistaking the vector $|\vec{k}\rangle_{\vec{j}}$ can then be bounded by considering the worst case scenario in which the code word sent is the last one tested in the sequence. Since this is the worst that can happen, $|\mathcal{A}_M(\text{yes})|$ with $m=M$ is the smallest possible, so that $P_{\text{err}}(\vec{j},\vec{k}) \le 1-|\mathcal{A}_M(\text{yes})|^2$. Recall that the input code words $\vec{j}$ are randomly selected from the set of typical input sequences, and $\vec{k}$'s are typical output sequences. Then, the average error probability for a randomly selected set of input code words can bounded as $\langle P_{\text{err}}\rangle \le 1-\langle|\mathcal{A}_M(\text{yes})|^2\rangle \le 1-|\langle\mathcal{A}_M(\text{yes})\rangle|^2$. Here $\langle\cdots\rangle$ represents the average over all possible code words of a given selected code book $\mathcal{C}_n$ *and* the averaging over all possible code books of code words. The Cauchy-Schwarz inequality $\langle|\mathcal{A}_M(\text{yes})|^2\rangle \ge |\langle\mathcal{A}_M(\text{yes})\rangle|^2$ was employed. The last term can be evaluated as

$$\begin{aligned}\langle\mathcal{A}_m(\text{yes})\rangle &= \text{Tr}\Big[P\Big(\mathbb{1}-\sum_{\ell_m}\pi_{\ell_m}P_{\ell_m}\Big)P\ldots P\Big(\mathbb{1}-\sum_{\ell_1}\pi_{\ell_1}P_{\ell_1}\Big)P\tilde{\rho}\Big]\\ &= \text{Tr}[(P-\tilde{\rho})^m\tilde{\rho}] = \sum_{k=0}^{m}\binom{m}{k}(-1)^k\text{Tr}[\tilde{\rho}^{k+1}], \quad (10)\end{aligned}$$

where π_ℓ stands for the probability $p_{\vec{j}}p_{\vec{k}|\vec{j}}$ and where we used (6) and (7) to write $\tilde{\rho}=\sum_\ell \pi_\ell P P_\ell P$. To prove the optimality of our decoding, it is hence sufficient to show that $\langle\mathcal{A}_m(\text{yes})\rangle \sim 1$ even when the number m of measurements is equal to its maximum possible value M of Eq. (8). Consider then Eqs. (4) and (5), which imply

$$\text{Tr}\,\tilde{\rho}^j \le \sum_{i=0}^{\dim(\mathcal{H})}\lambda_i^j \le 2^{n[S(1-j)+\delta(1+j)]}. \quad (11)$$

Use this and Eq. (3) to rewrite Eq. (10) as

$$\begin{aligned}\langle\mathcal{A}_m(\text{yes})\rangle &\ge \text{Tr}\tilde{\rho} + \sum_{k=1}^{m}\binom{m}{k}(-1)^k\text{Tr}[\tilde{\rho}^{k+1}]\\ &\ge 1-\epsilon-\sum_{k=1}^{m}\binom{m}{k}2^{n[-kS(\rho)+\delta(k+2)]}\\ &= 1-\epsilon-\gamma, \quad (12)\end{aligned}$$

where $\gamma \equiv 2^{2n\delta}[(1+\zeta_n)^m-1]$, with $\zeta_n = 2^{n[-S(\rho)+\delta]}$. If $S(\rho)>\delta$, for large n we can write

$$(1+\zeta_n)^m - 1 \simeq e^{m\zeta_n}-1 \simeq m\zeta_n. \quad (13)$$

Hence, γ is asymptotically negligible as long as $2^{2n\delta}m\zeta_n$ is vanishing for $n\to\infty$. This yields the constraint

$$m \le 2^{n[S(\rho)-\delta]} \quad \text{for all } m. \quad (14)$$

In particular, it must hold for M, the largest value of m given in (8). By imposing this, the decoding procedure yields a vanishing error probability if the rate R satisfies

$$R < \chi - \delta, \quad (15)$$

as required by the Holevo bound [5].

In summary, we have shown that under the condition (15) the average amplitude $\langle\mathcal{A}_m(\text{yes})\rangle$ of identifying the correct code word is asymptotically close to 1 even in the worst case in which we had to check over *all* the other code words $m=M$. This implies that the average probability of error in identifying the codew ord asymptotically vanishes. In other words, the procedure works even when the measurements are chosen so that the code word sent is the last one tested in the sequence of tests. Note that the same results presented here can be obtained also by starting from the direct calculation of the error probability [24] (instead of by using the probability amplitude).

We conclude by noting that from Eq. (9) one sees that the probabilities associated with the various outcomes can be described in terms of a positive operator-valued measure $\{E_\ell\}$ as

$$\begin{aligned}E_1 &= PP_1P; \qquad E_2 = P(\mathbb{1}-P_1)PP_2P(\mathbb{1}-P_1)P;\\ E_\ell &= P(\mathbb{1}-P_1)P(\mathbb{1}-P_2)P\ldots P(\mathbb{1}-P_{\ell-1})P\\ &\quad\times P_\ell\ldots(\mathbb{1}-P_1)P;\\ E_0 &= \mathbb{1}-\sum_{\ell=1}^{M}E_\ell, \quad (16)\end{aligned}$$

where P_ℓ is defined as in (7) and E_0 is the "abort" result. We gave a simple realization of this positive operator-valued measure by using sequential yes-no projections, but different realizations may be possible. It is an alternative to the conventional pretty good measurement. The operators P_ℓ in this positive operator-valued measure are simply projections onto separable pure states or on their orthogonal complement, and P projects into the typical output subspace (with which the states involved have asymptotically complete overlap). Such a sequence of projective measurements shows that the output state departs at most infinitesimally from its original (nonentangled) form throughout the entire decoding procedure. This clarifies that the role of entanglement in the decoding is analogous to [28] increasing the distinguishability of a multipartite set of states that are not orthogonal when considered by separate parties. Note that also the pretty good measurement becomes projective when employed to discriminate among a sufficiently small set of states [29,30].

Conclusions.—Using projective measurements acting on the channel output in a sequential fashion, we gave a new proof that it is possible to attain the Holevo capacity when a noisy quantum channel is used to transmit classical information. Such measurements provide an alternative to the usual pretty good measurements for channel decoding and can be used in many of the same situations. In particular, an analogous procedure can be used to decode channels that transmit quantum information, to approach the coherent information limit [31–33]. This follows simply from the observation [33] that the transfer of quantum messages over the channel can be formally treated as a transfer of classical messages imposing an extra constraint of privacy in the signaling.

We acknowledge C. Fuchs, P. Hayden, A. S. Holevo, K. Matsumoto, S. Tan, J. Tyson, M. M. Wilde, and A. Winter for comments and discussions. V. G. was supported by the FIRB-IDEAS project, RBID08B3FM, and by Institut Mittag-Leffler. S. L. was supported by the WM Keck Foundation, DARPA, NSF, NEC, ONR, and Intel. L. M. was supported by the EU through the FP7 STREP project COQUIT.

[1] C. H. Bennett and P. W. Shor, IEEE Trans. Inf. Theory **44**, 2724 (1998).
[2] C. A. Fuchs, Phys. Rev. Lett. **79**, 1162 (1997).
[3] A. S. Holevo, IEEE Trans. Inf. Theory **44**, 269 (1998).
[4] B. Schumacher and M. D. Westmoreland, Phys. Rev. A **56**, 131 (1997); P. Hausladen, R. Jozsa, B. Schumacher, M. D. Westmoreland, and W. K. Wootters, Phys. Rev. A **54**, 1869 (1996).
[5] A. S. Holevo, Prob. Peredachi Inf. **9**, 3 (1973) [Probl. Inf. Transm. **9**, 177 (1973)].
[6] J. Tyson, J. Math. Phys. (N.Y.) **50**, 032106 (2009); Phys. Rev. A **79**, 032343 (2009).
[7] C. Mochon, Phys. Rev. A **73**, 032328 (2006).
[8] V. P. Belavkin, Stochastics **1**, 315 (1975); P. Belavkin, Radio Eng. Electron. Phys. **20**, 1177 (1975); V. P. Belavkin and V. Maslov, in *Mathematical Aspects of Computer Engineering*, edited by V. Maslov (MIR, Moscow, 1987).
[9] M. Ban, J. Opt. B **4**, 143 (2002).
[10] T. S. Usuda, I. Takumi, M. Hata, and O. Hirota, Phys. Lett. A **256**, 104 (1999).
[11] Y. C. Eldar and G. David Forney, IEEE Trans. Inf. Theory **47**, 858 (2001).
[12] P. Hausladen and W. K. Wootters, J. Mod. Opt. **41**, 2385 (1994).
[13] H. Barnum and E. Knill, J. Math. Phys. (N.Y.) **43**, 2097 (2002).
[14] A. Montanaro, Commun. Math. Phys. **273**, 619 (2007).
[15] M. Jëzek, J. Řeháček, and J. Fiurášek, Phys. Rev. A **65**, 060301(R) (2002); Z. Hradil, J. Řeháček, J. Fiurášek, and M. Jëzek, Lect. Notes Phys. **649**, 163 (2004).
[16] P. Hayden, D. Leung, and G. Smith, Phys. Rev. A **71**, 062339 (2005).
[17] A. S. Kholevo, Teor. Veroyatn. Primen. **23**, 429 (1978) [Theory Probab. Appl. **23**, 411 (1979)].
[18] A. Winter, IEEE Trans. Inf. Theory **45**, 2481 (1999).
[19] T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory **46**, 2428 (2000).
[20] T. Ogawa and H. Nagaoka, in *Proceedings of the 2002 IEEE International Symposium on Information Theory, Lausanne, Switzerland, 2002* (IEEE, New York, 2002), p. 73; IEEE Trans. Inf. Theory **45**, 2486 (1999).
[21] M. Hayashi and H. Nagaoka, IEEE Trans. Inf. Theory **49**, 1753 (2003); M. Hayashi, Phys. Rev. A **76**, 062301 (2007); Commun. Math. Phys. **289**, 1087 (2009).
[22] L. Wang and R. Renner, arXiv:1007.5456v1.
[23] 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).
[24] V. Giovannetti, S. Lloyd, and L. Maccone, arXiv:1012.0386v1.
[25] M. B. Hastings, Nature Phys. **5**, 255 (2009).
[26] T. M. Cover, and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[27] Binary projective measurements have Kraus operators Π and $\mathbb{1} - \Pi$ (Π being a projector): The outputs yes and no correspond to Π and $\mathbb{1} - \Pi$, respectively. The probability of each outcome is $p = \mathrm{Tr}[\rho M]$, with $M = \Pi$ or $M = \mathbb{1} - \Pi$, and the postmeasurement state is $M\rho M/p$.
[28] C. H. Bennett *et al.*, Phys. Rev. A **59**, 1070 (1999).
[29] C. Fuchs (private communication).
[30] R. Jozsa, D. Robb, and W. K. Wootters, Phys. Rev. A **49**, 668 (1994).
[31] S. Lloyd, Phys. Rev. A **55**, 1613 (1997).
[32] P. W. Shor, in *Proceedings of the MSRI Workshop on Quantum Information, Berkeley, 2002* (Mathematical Sciences Research Institute, Berkeley, CA, 2002), http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1.
[33] I. Devetak, IEEE Trans. Inf. Theory **51**, 44 (2005).