ISSN 0032-9460, Problems of Information Transmission, 2017, Vol. 53, No. 3, pp. 222–228. © Pleiades Publishing, Inc., 2017.
Original Russian Text © M. Dalai, 2017, published in Problemy Peredachi Informatsii, 2017, Vol. 53, No. 3, pp. 23–29.

INFORMATION THEORY

A Note on Random Coding Bounds for Classical-Quantum Channels[1]

M. Dalai

Department of Information Engineering, University of Brescia, Brescia, Italy
e-mail: marco.dalai@unibs.it

Received March 14, 2017; in final form, April 10, 2017

Abstract—A modified derivation of achievability results in classical-quantum channel coding theory is proposed, which has, in our opinion, two main benefits over previously used methods: it allows to (i) follow in a simple and clear way how binary hypothesis testing relates to channel coding achievability results, and (ii) derive in a unified way all previously known random coding achievability bounds on error exponents for classical and classical-quantum channels.

DOI: 10.1134/S0032946017030036

1. INTRODUCTION

Random coding has been used as a primary tool in achievability results since the early days of information theory [1–3]. Extensions of classical results allowed to derive the capacity [4–7] and achievable error exponents [8–10] for classical-quantum channels. In particular, the achievability proof for the capacity of a general mixed-state channel is derived in [6], and an achievable error exponent for pure-state channels is derived in [8], which formally coincides with the one for classical channels [3] and matches a converse bound derived in [11, 12]. In [9] Holevo conjectured that a formal extension of the classical bound also holds for general mixed-state channels, but no proof up to now has been obtained. The best known achievable exponent for mixed-state channels is Hayashi's [10].

The above-mentioned results were derived by combining the idea of random coding, i.e., the study of the average performance of random codes, with different choices of measurements used at the receiver. The measurements used in [4–6] extend the typicality method [13] and do not provide satisfactory bounds on error exponents at rates R smaller than the capacity. The "pretty good measurement" used in [8] allows one to prove achievability of Gallager's exponent for pure-state channels and, in a modified form [9], for classical channels, but it does not seem to be amenable to generalizations to arbitrary mixed-state channels. Hayashi proves achievability of a positive exponent in [10] for arbitrary mixed state channels using the method introduced in [7], which reduces the achievability proof in channel coding theory to an achievability proof in binary hypothesis testing (see [14] for recent results on the connection between multiple and binary hypothesis testing in the quantum setting). In particular, the binary hypothesis used in [7] is between a given codeword and a fixed (i.e., code-independent) *expected state*. This procedure resembles Shannon's first approach to error exponents [15]; it has the advantage of applying to general mixed-state channels, but, on the other hand, it does not recover the optimal Gallager's exponent for classical channels.

In this paper we present an additional possible choice for the measurement and the associated error analysis. It has the following benefits: (i) it shows in a clear way how the channel coding problem can be reduced to binary hypothesis testing even for the most powerful error exponent

[1] Partially supported by the Italian Ministry of Education under grant PRIN 2015 D72F1600079000.

achievability results, and (ii) to the best of the author's knowledge, it is the only "unified proof" which gives at the same time all the above-mentioned achievability results for capacity and error exponents [3,6,8,10]. The main innovation with respect to the proofs in [7,10] relies on the fact that we use a binary hypothesis test between a codeword and an *empirical average* (*code-dependent*) *tilted codeword*. This allows us to recover at the same time the results of [10] for general channels and the correct exponent for classical channels.

We close with the observation that, since Holevo's conjecture on the achievability of Gallager's exponent [9] has not been either proved or disproved yet, we believe that any new derivation of achievability results in classical-quantum channel coding is worth being investigated.

In our analysis, we will need the following results from the literature (or slight variations thereof).

Lemma 1 ([7, Lemma 2] with $c = 1$). *For operators $0 \le S \le I$ and $T \ge 0$, we have*

$$I - (S+T)^{-1/2} S (S+T)^{-1/2} \le 2(I-S) + 4T. \tag{1}$$

Given a self-adjoint operator A with spectral decomposition $A = \sum_i \lambda_i E_i$, where the λ_i are the operator eigenvalues and the E_i are the orthogonal projectors on the associated eigenspaces, we set by definition

$$\{A \ge 0\} = \sum_{i:\, \lambda_i \ge 0} E_i, \qquad \{A < 0\} = \sum_{i:\, \lambda_i < 0} E_i. \tag{2}$$

Then we have the following lemma.

Lemma 2 (variation of Lemma 1 in [10]). *For any two positive semidefinite operators A and B and real $s \in [0,1]$, setting $t = \max(s, 1-s)$, we have*

$$\operatorname{Tr} A^s B^{1-s} \ge \operatorname{Tr}[\{A^t - B^t \ge 0\} B] + \operatorname{Tr}[\{A^t - B^t < 0\} A]. \tag{3}$$

Proof. We start with Lemma 1 in [10], which states that for $0 \le s \le 1/2$

$$\operatorname{Tr} A^s B^{1-s} \ge \operatorname{Tr}[\{A^{1-s} - B^{1-s} \ge 0\} B] + \operatorname{Tr}[\{A^{1-s} - B^{1-s} < 0\} A]. \tag{4}$$

By inspection of the proof we see that the following similar inequality also holds for $0 \le s \le 1/2$:

$$\operatorname{Tr} A^s B^{1-s} \ge \operatorname{Tr}[\{A^{1-s} - B^{1-s} > 0\} B] + \operatorname{Tr}[\{A^{1-s} - B^{1-s} \le 0\} A],$$

which we can rearrange as

$$\operatorname{Tr} A^s B^{1-s} \ge \operatorname{Tr}[\{B^{1-s} - A^{1-s} \ge 0\} A] + \operatorname{Tr}[\{B^{1-s} - A^{1-s} < 0\} B].$$

Swapping the roles of A and B, we get for $0 \le s \le 1/2$

$$\operatorname{Tr} A^{1-s} B^s \ge \operatorname{Tr}[\{A^{1-s} - B^{1-s} \ge 0\} B] + \operatorname{Tr}[\{A^{1-s} - B^{1-s} < 0\} A],$$

which can also be stated as

$$\operatorname{Tr} A^s B^{1-s} \ge \operatorname{Tr}[\{A^s - B^s \ge 0\} B] + \operatorname{Tr}[\{A^s - B^s < 0\} A]$$

for $1/2 \le s \le 1$. This, together with (4), proves (3).

2. PROPOSED MEASUREMENT AND ERROR ANALYSIS

We derive our analysis in a "one-shot" setting, i.e., derive an *upper bound* on the probability of error in a classical-quantum communication by means of M given quantum states, each associated to one of M different messages, when a properly chosen positive operator-valued measurement (POVM) is used at the receiver to decide between the possible messages. We then use this to compute an upper bound on the the average probability of error over an ensemble of random codes with i.i.d. signals, later also considering the case where signals are randomly generated as tensor products of random i.i.d. density operators from some fixed finite set.

Let $W_1, \dots, W_M$ be finite-dimensional density operators representing the signals associated to M different messages. A decoder is defined in terms of a POVM (see [16,17]), i.e., a set of M positive semidefinite hermitian operators $\{Y_i\}_{i=1,\dots,M}$ such that $\sum\limits_i Y_i \le I$. Consider a variation of the POVM used in [7] defined by

$$Y_i = \Big(\sum_j \pi_j\Big)^{-1/2} \pi_i \Big(\sum_j \pi_j\Big)^{-1/2}, \tag{5}$$

where π_i is the following parametrized projector:

$$\pi_i = \left\{ W_i^t - \Big(\sum_{j\ne i} W_j^r\Big)^{\frac{t}{r}} > 0 \right\}, \qquad t = \max(r\rho, 1-r\rho), \quad 0 \le r, \rho \le 1. \tag{6}$$

This is another possible way of extending the classical approach to the quantum case, which does not seem to have been considered up to now. In particular, note that the operator $\Big(\sum\limits_{j\ne i} W_j^r\Big)^{\frac{1}{r}}$, which depends on the code at hand, plays a similar role as the fixed *code-independent* operator $W_{P^{(n)}}^{(n)}$ used in [7, Lemma 3]. Thus, we depart from [7, Lemma 3] in that we consider binary hypothesis tests between codewords and a *code-dependent* "average" state (with some tilting).

Using Lemma 1, we have

$$I - Y_i \le 2(I - \pi_i) + 4\sum_{j\ne i} \pi_j. \tag{7}$$

Hence, the average probability of error of the code using the described POVM can be bounded as

$$P_e = \frac{1}{M}\sum_i \mathrm{Tr}[W_i(I - Y_i)] \tag{8}$$

$$\le 2\frac{1}{M}\sum_i \mathrm{Tr}\left[W_i\left\{W_i^t - \Big(\sum_{j\ne i} W_j^r\Big)^{\frac{t}{r}} \le 0\right\}\right] \tag{9}$$

$$+ 4\frac{1}{M}\sum_i\sum_{j\ne i} \mathrm{Tr}\left[W_i\left\{W_j^t - \Big(\sum_{k\ne j} W_k^r\Big)^{\frac{t}{r}} > 0\right\}\right]. \tag{10}$$

Let us first consider the term in (9). Using Lemma 2, remembering that by definition $t = \max(r\rho, 1-r\rho)$, the ith term in the sum can be bounded as

$$\mathrm{Tr}\left[W_i\left\{W_i^t - \Big(\sum_{j\ne i} W_j^r\Big)^{\frac{t}{r}} \le 0\right\}\right] \le \mathrm{Tr}\left[W_i^{1-r\rho}\Big(\sum_{j\ne i} W_j^r\Big)^{\rho}\right]. \tag{11}$$

The second term of the error probability, i.e., (10), can be rewritten by interchanging the order of summation as

$$4\frac{1}{M}\sum_j \mathrm{Tr}\left[\Big(\sum_{i\ne j} W_i\Big)\left\{W_j^t - \Big(\sum_{k\ne j} W_k^r\Big)^{\frac{t}{r}} > 0\right\}\right]$$

or, renaming indices in the sums,

$$4\frac{1}{M}\sum_i \operatorname{Tr}\left[\left(\sum_{j\neq i} W_j\right)\left\{W_i^t - \left(\sum_{j\neq i} W_j^r\right)^{\frac{t}{r}} > 0\right\}\right]. \tag{12}$$

A possible attempt to bound the ith term in this sum could be as follows. For classical channels[2] we can use the bound

$$\left(\sum_{j\neq i} W_j\right)^r \leq \sum_{j\neq i} W_j^r, \tag{13}$$

valid for all $0 \leq r \leq 1$, to conclude that

$$\sum_{j\neq i} W_j \leq \left(\sum_{j\neq i} W_j^r\right)^{1/r}. \tag{14}$$

Alternatively, for general channels the above inequality works obviously for $r = 1$. Then we can use Lemma 2 to obtain for the ith term the bound

$$\begin{aligned}\operatorname{Tr}\left[\left(\sum_{j\neq i} W_j\right)\left\{W_i^t - \left(\sum_{j\neq i} W_j^r\right)^{\frac{t}{r}} > 0\right\}\right] &\leq \operatorname{Tr}\left[\left(\sum_{j\neq i} W_j^r\right)^{1/r}\left\{W_i^t - \left(\sum_{j\neq i} W_j^r\right)^{\frac{t}{r}} > 0\right\}\right] \\ &\leq \operatorname{Tr}\left[\left(\sum_{j\neq i} W_j^r\right)^{\rho} W_i^{1-r\rho}\right],\end{aligned}$$

which is thus valid for all r for classical channels and for $r = 1$ for all channels. This has the same form as in equation (11), and hence we deduce

$$P_e \leq 6\frac{1}{M}\sum_i \operatorname{Tr}\left[W_i^{1-r\rho}\left(\sum_{j\neq i} W_j^r\right)^{\rho}\right]. \tag{15}$$

This is our "one-shot" upper bound on the probability of error for a given specific code. Taking the expected value over an ensemble of codes with i.i.d. selection of the states W_i, we can upper bound the expected value of the probability of error as

$$\begin{aligned}\mathbf{E}[P_e] &\leq 6\frac{1}{M}\sum_i \mathbf{E}\left[\operatorname{Tr}\left[W_i^{1-r\rho}\left(\sum_{j\neq i} W_j^r\right)^{\rho}\right]\right] = 6\operatorname{Tr}\left[\mathbf{E}[W_1^{1-r\rho}]\,\mathbf{E}\left[\left(\sum_{j\neq 1} W_j^r\right)^{\rho}\right]\right] \\ &\overset{(a)}{\leq} 6\operatorname{Tr}\left[\mathbf{E}[W_1^{1-r\rho}]\left(\mathbf{E}\left[\sum_{j\neq 1} W_j^r\right]\right)^{\rho}\right] = 6(M-1)^{\rho}\operatorname{Tr}\left[\mathbf{E}[W^{1-r\rho}]\,\mathbf{E}[W^r]^{\rho}\right],\end{aligned} \tag{16}$$

which again is valid for any r for classical channels and with $r = 1$ for any channel. Here, W is a generic random signal of which signals $W_1, W_2, \ldots$ are i.i.d. extractions. In inequality (a) above we have used the operator concavity of the map $A \mapsto A^{\rho}$, $0 \leq \rho \leq 1$.

[2] We point out that for classical channels the whole double summation in (10) is not really necessary. Since all the W_i are diagonal in the same basis, the π_i can be written as $\pi_i = \left\{W_i > \left(\sum_{j\neq i} W_i^r\right)^{1/r}\right\}$ and are all pairwise orthogonal (in particular, a common eigenvector ψ of the W_i is in the range of π_i only if the eigenvalue of W_i associated to ψ is larger than that of any other of the states W_j, $j \neq i$). We could then define $Y_i = \pi_i$, adding an extra operator $Y_e = I - \sum_i \pi_i$ which accounts for extra errors, avoiding the use of the Hayashi–Nagaoka lemma. This boils down to one of the possible presentations of the classical bound.

For classical channels we can now set as usual $r = 1/(1+\rho)$ to obtain—apart from coefficients—the standard random coding bound

$$\mathbf{E}[P_e] \le 6(M-1)^\rho \operatorname{Tr} \mathbf{E}[W^{1/(1+\rho)}]^{1+\rho}. \tag{17}$$

For general channels, instead, the bound in (16) only holds with $r = 1$, which gives

$$\mathbf{E}[P_e] \le 6(M-1)^\rho \operatorname{Tr}\big[\mathbf{E}[W^{1-\rho}]\, \mathbf{E}[W]^\rho\big], \tag{18}$$

which equals Hayashi's exponent [10]. For pure-state channels this expression clearly equals (17) and hence gives again the correct random coding exponent, which can be written in simplified form as

$$\mathbf{E}[P_e] \le 6(M-1)^\rho \operatorname{Tr} \mathbf{E}[W]^{1+\rho}. \tag{19}$$

We observe that while for general channels the bound in (18) is not the one conjectured by Holevo, it is good enough to prove achievability of the capacity. This follows already from [7,10], but we recast its derivation here for completeness. In fact, assuming as usual that the randomization is such that the W are tensor products of independent identically distributed states, i.e., $W = S_1 \otimes S_2 \ldots \otimes S_n$ with the S_i distributed i.i.d. as some random state S with values in a fixed finite set, we find

$$\mathbf{E}[P_e] \le 6e^{nR\rho} \operatorname{Tr}\big[\mathbf{E}[S^{1-\rho}]\, \mathbf{E}[S]^\rho\big]^n. \tag{20}$$

Hence, the probability of error vanishes exponentially with n for all rates R such that

$$R \le -\frac{\log \operatorname{Tr}\big[\mathbf{E}[S^{1-\rho}]\, \mathbf{E}[S]^\rho\big]}{\rho}. \tag{21}$$

As $\rho \to 0$, the bound on R can be computed as usual using L'Hôpital's rule as

$$R < -\frac{\operatorname{Tr}\big[-\mathbf{E}[S^{1-\rho}\log S]\, \mathbf{E}[S]^\rho\big] + \operatorname{Tr}\big[\mathbf{E}[S^{1-\rho}]\, \mathbf{E}[S]^\rho \log \mathbf{E}[S]\big]}{\operatorname{Tr}\big[\mathbf{E}[S^{1-\rho}]\, \mathbf{E}[S]^\rho\big]}, \tag{22}$$

which as $\rho \to 0$ gives

$$R < \mathbf{E}[\operatorname{Tr}(S \log S)] - \operatorname{Tr}[\mathbf{E}[S] \log \mathbf{E}[S]], \tag{23}$$

and hence achievability of all rates below the capacity.

Thus, the proposed (parametrized) POVM achieves the reliability function of classical and pure-state channels as well as Hayashi's exponent and hence the capacity of any channel. A question which remains open is whether the POVM is provably not good enough to achieve the conjectured random coding exponent (17) for general channels or if it is the analysis which is not tight enough.

3. A CLOSING COMMENT

A personal impression of the author is that if Holevo's conjecture holds, then there should be a hope of proving it by reducing the estimation of the probability of error of a code to a binary hypothesis test between a state W_i and (a scaled version of) the state $\mathbf{E}[W^{1/(1+\rho)}]^{1+\rho}$ (the bound is then precisely the scale parameter). The above procedure has perhaps the advantage of showing the hypothesis testing road fairly clearly for classical and for pure-state channels. But still, there is a huge problem for general channels, since (14) cannot be applied. It is interesting to observe that, using (14) in our derivation for classical channels, we essentially use the same property used by Holevo in [9] for the same setting, but—at least to this author—the way that this requirement emerges looks different.

A final observation seems in order. The first proof of the optimal achievable error exponent in random coding for general classical channels was derived by Fano [2]. His proof, while being longer and less elegant than Gallager's one, has the benefit of deriving the right exponent while using a starting point very similar to Shannon's [15, Theorem 1, last equation on p. 9] and particularly fit to the binary hypothesis testing formulation, which is instead lost in Gallager's shorter proof. A detailed and clear description of those proofs and of their relations and applications can be found in [18]. The measurement proposed here for classical-quantum channels is in a sense an attempt to modify the method in [7, 10], which is similar to Shannon's [15], in a way similar to what was done by Fano. Still, the procedure does not go through in the goal of proving Holevo's conjecture. In this author's opinion, the key point relies on a tool which Fano uses and which does not seem to have a counterpart in the quantum theory yet. Fano bounds the probability that one random variable is less than another by applying large deviation results to their difference. When we try to isolate the simplest possible scenario for such a problem, it seems that the analog quantum tool needed (at least to start with) is a way to bound

$$\operatorname{Tr} A^{\otimes n}\{B^{\otimes n} - C^{\otimes n} > 0\} \tag{24}$$

as a function of n in terms of A, B, and C. The Chernoff bound corresponds to the particular case $A = C$. In particular, we would need a bound on

$$\operatorname{Tr} A\{B - C > 0\} \tag{25}$$

which is multiplicative under tensor products. Assuming that a "tilting" trick could be used in this case as in binary hypothesis testing [19], one could perhaps hope to succeed by studying the quantity

$$\operatorname{Tr} A\{B^t - C^t > 0\}. \tag{26}$$

Note that this is actually the problem that we have in (12), and the reason why we would like to use (14) is precisely because we have to make $A = C$ to get back to the Chernoff bound.

REFERENCES

1. Shannon, C.E., A Mathematical Theory of Communication, *Bell Syst. Tech. J.*, 1948, vol. 27, no. 3, pp. 379–423; no. 4, pp. 623–656.
2. Fano, R.M., *Transmission of Information: A Statistical Theory of Communication*, New York: Wiley, 1961.
3. Gallager, R.G., A Simple Derivation of the Coding Theorem and Some Applications, *IEEE Trans. Inform. Theory*, 1965, vol. 11, no. 1, pp. 3–18.
4. Hausladen, P., Jozsa, R., Schumacher, B., Westmoreland, M., and Wootters, W.K., Classical Information Capacity of a Quantum Channel, *Phys. Rev. A*, 1996, vol. 54, no. 3, pp. 1869–1876.
5. Schumacher, B. and Westmoreland, M.D., Sending Classical Information via Noisy Quantum Channels, *Phys. Rev. A*, 1997, vol. 56, no. 1, pp. 131–138.
6. Holevo, A.S., The Capacity of the Quantum Channel with General Signal States, *IEEE Trans. Inform. Theory*, 1998, vol. 44, no. 1, pp. 269–273.
7. Hayashi, M. and Nagaoka, H., General Formulas for Capacity of Classical-Quantum Channels, *IEEE Trans. Inform. Theory*, 2003, vol. 49, no. 7, pp. 1753–1768.
8. Burnashev, M.V. and Holevo, A.S., On the Reliability Function for a Quantum Communication Channel, *Probl. Peredachi Inf.*, 1998, vol. 34, no. 2, pp. 3–15 [*Probl. Inf. Trans.* (Engl. Transl.), 1998, vol. 34, no. 2, pp. 97–107].

9. Holevo, A.S., Reliability Function of General Classical-Quantum Channel, *IEEE Trans. Inform. Theory*, 2000, vol. 46, no. 6, pp. 2256–2261.

10. Hayashi, M., Error Exponent in Asymmetric Quantum Hypothesis Testing and Its Application to Classical-Quantum Channel Coding, *Phys. Rev. A*, 2007, vol. 76, no. 6, pp. 062301.

11. Dalai, M., Sphere Packing Bound for Quantum Channels, in *Proc. 2012 IEEE Int. Sympos. on Information Theory (ISIT'2012), Cambridge, MA, USA, July 1–6, 2012*, pp. 160–164.

12. Dalai, M., Lower Bounds on the Probability of Error for Classical and Classical-Quantum Channels, *IEEE Trans. Inform. Theory*, 2013, vol. 59, no. 12, pp. 8027–8056.

13. Cover, T.M. and Thomas, J.A., *Elements of Information Theory*, New York: Wiley, 1991.

14. Vazquez-Vilar, G., Multiple Quantum Hypothesis Testing Expressions and Classical-Quantum Channel Converse Bounds, in *Proc. 2016 IEEE Int. Sympos. on Information Theory (ISIT'2016), Barcelona, Spain, July 10–15, 2016*, pp. 2854–2857.

15. Shannon, C.E., Certain Results in Coding Theory for Noisy Channels, *Inform. Control*, 1957, vol. 1, pp. 6–25.

16. Holevo, A.S., Coding Theorems for Quantum Channels, `arXiv:quant-ph/9809023v1`, 1998.

17. Wilde, M.M., *Quantum Information Theory*, Cambridge, UK: Cambridge Univ. Press, 2013.

18. Shamai, S. and Sason, I., Variations on the Gallager Bounds, Connections, and Applications, *IEEE Trans. Inform. Theory*, 2002, vol. 48. no. 12, pp. 3029–3051.

19. Audenaert, K.M.R., Calsamiglia, J., Muñoz-Tapia, R., Bagan, E., Masanes, L., Acin, A., and Verstraete, F., Discriminating States: The Quantum Chernoff Bound, *Phys. Rev. Lett.*, 2007, vol. 98, no. 16, pp. 160501.