A Strong Converse for Classical Channel Coding Using Entangled Inputs

Robert König* and Stephanie Wehner†

Institute for Quantum Information, Caltech, Pasadena California 91125, USA

(Received 30 April 2009; published 14 August 2009)

A fully general strong converse for channel coding states that when the rate of sending classical information exceeds the capacity of a quantum channel, the probability of correctly decoding goes to zero exponentially in the number of channel uses, even when we allow code states which are entangled across several uses of the channel. Such a statement was previously only known for classical channels and the quantum identity channel. By relating the problem to the additivity of minimum output entropies, we show that a strong converse holds for a large class of channels, including all unital qubit channels, the d-dimensional depolarizing channel and the Werner-Holevo channel. This further justifies the interpretation of the classical capacity as a sharp threshold for information transmission.

DOI: 10.1103/PhysRevLett.103.070504 PACS numbers: 03.67.Hk

A fundamental problem in quantum information theory is the transmission of classical information over (noisy) quantum channels. As a simple example, suppose we send M classical bits using a qubit identity channel n times. Clearly [1], this can be done reliably if $M \leq n$, but if the number of classical bits exceeds the number of qubits sent ($M > n$) we are no longer able to recover the encoded information with perfect accuracy [2]. This situation is analogous to the problem of information transmission over a noisy classical channel. Here, there exists a constant C, called the classical capacity, which determines the maximal number of classical bits that can be sent reliably per channel use: by using the channel n times, we can reliably transmit M bits if and only if the rate $R = \frac{M}{n}$ satisfies $R \leq C$ in the asymptotic limit. This is known as the coding theorem due to Shannon [4]. For example, for the binary bit flip channel, which flips an input bit with probability p, this constant is given by $C = 1 - h(p)$, where h is the binary entropy function. The unifying concept for both scenarios is that of the classical capacity C. For the qubit identity channel Holevo's seminal result [3] shows that the classical capacity is equal to 1.

In fact, for both the qubit identity channel and any classical channel, the classical capacity C imposes a sharp bound on our ability to recover classical information sent over the channel: On the one hand if $R \leq C$, then it is possible to send nR classical bits by using the channel n times in such a way that the probability P_{succ} of successful decoding goes to 1 exponentially as $n \to \infty$. This is also referred to as the achievability of the capacity. On the other hand, if $R > C$, then for any encoding and decoding scheme, P_{succ} is exponentially small in the difference $n(R - C)$. This is referred to as the strong converse of the coding theorem for these channels.

For classical noisy channels, the strong converse was established by Wolfowitz [5]. For the qubit identity channel $\text{id}_2 \equiv \text{id}_{\mathcal{B}(\mathbb{C}^2)}$, the argument is rather simple: Suppose we encode a uniformly distributed nR-bit string $X \in \{0,1\}^{nR}$ using a family of 2^{nR} states $\{\rho_x\}_{x=1}^{2^{nR}}$ on $(\mathbb{C}^2)^{\otimes n}$ (i.e., of n qubits). Then, for any decoding positive operator valued measure (POVM) $\{E_x\}_{x=1}^{2^{nR}}$ on $(\mathbb{C}^2)^{\otimes n}$, the average success probability of correctly decoding is bounded by

$$P_{\text{succ}}^{\text{id}_2}(n, R) = \frac{1}{2^{nR}} \sum_x \text{tr}(E_x \rho_x) \leq \frac{1}{2^{nR}} \sum_x \text{tr}(E_x) = 2^{-n(R-1)}.$$

Here, we used the operator inequality $\rho_x \leq \mathbb{I}_{(\mathbb{C}^2)^{\otimes n}}$ for every x, and the fact that the operator elements of a POVM sum to the identity. Because of the strong converse property, we can regard the capacity C as an exact measure of the information-carrying power of any classical channel and the quantum identity channel.

Unfortunately, this appealing operational interpretation of the classical capacity C is not quite as complete for general quantum channels. While the achievability of the capacity has been established in [6,7] (building on [8]), only a weak converse has been shown without assumptions [3]. It merely states that for rates $R > C$ above the capacity, the success probability is bounded away from 1. This is in contrast to a strong converse, which shows that this probability goes to zero exponentially, in the limit as n goes to infinity.

Here, we are interested in the validity of the strong converse property for a general quantum channel. Establishing such a converse is more difficult than for classical channels for the same reason it is difficult to compute the classical capacity of a quantum channel: We have to take into account the possibility that entanglement over several uses of the channel may help to increase the probability of successful decoding. Indeed, a recent breakthrough result by Hastings [9] shows that using entangled states can be advantageous. Formally, this is expressed by the product-state capacity C_Φ^{prod}: This is defined in the same way as the capacity, but with the restriction that the input states to the channel $\Phi^{\otimes n}$ have to be of tensor product form.

Hasting's result shows that there are channels Φ with $C_\Phi^{\text{prod}} < C_\Phi$.

In light of the advantage of entanglement for coding, it is natural to ask whether entanglement may invalidate the strong converse property: In particular, we study whether allowing arbitrary (entangled) input states affects the exponential decay of the success probability. Previous studies of the region $R > C_\Phi$ were restricted to the case where the inputs are not entangled across different uses of the channel [10,11], and are thus conceptually similar to the study of the achievability of the product-state capacity C_Φ^{prod} instead of the more general C_Φ.

Main result.—In this Letter, we prove a strong converse for a large number of quantum channels Φ. In particular, our result applies to (i) the qudit depolarizing channel

$$\Delta_r(\rho) = r\rho + (1-r)\frac{\mathbb{I}}{d}, \tag{1}$$

replacing any input state with the fully mixed state with probability $(1-r)$ for $-1/(d^2-1) \leq r \leq 1$, (ii) any unital qubit channel [12], and, more generally, (iii) any channel which has additive minimum output α entropy $S_\alpha^{\min}$ for $\alpha \geq 1$ (close to 1) as defined below [13], and the following covariance property: there is a pair of unitary representations of some group G on the input space $\mathcal{H}_{\text{in}}$ and the output space $\mathcal{H}_{\text{out}}$, respectively, such that

$$g\Phi(\rho)g^\dagger = \Phi(g\rho g^\dagger) \quad \text{for all } g \in G,$$

where the representation on $\mathcal{H}_{\text{out}}$ is irreducible. An example of such a channel is the Werner-Holevo channel [14].

More formally, we are concerned with (noisy) quantum channels, i.e., completely positive trace-preserving maps (CPTPM) Φ: $\mathcal{B}(\mathcal{H}_{\text{in}}) \to \mathcal{B}(\mathcal{H}_{\text{out}})$. Throughout, we restrict our attention to finite-dimensional Hilbert spaces $\mathcal{H}_{\text{in}}$ and $\mathcal{H}_{\text{out}}$. A code of rate R for Φ specifies (for every n) a family $\{\rho_x\}_{x=1}^{2^{nR}}$ of states on $\mathcal{H}_{\text{in}}^{\otimes n}$, where ρ_x is the quantum code word associated with the classical message $x \in \{1, \ldots, 2^{nR}\}$. A corresponding decoder is a POVM $\{E_x\}_{x=1}^{2^{nR}}$ on $\mathcal{H}_{\text{out}}^{\otimes n}$. We are interested in the average success probability of decoding correctly, that is, the quantity

$$P_{\text{succ}}^{\Phi}(n, R) = \frac{1}{2^{nR}} \sum_{x=1}^{2^{nR}} \operatorname{tr}(E_x \Phi^{\otimes n}(\rho_x)). \tag{2}$$

In this terminology, we show the following:

Theorem.—Let Φ be a CPTPM described by (i)–(iii), and let C_Φ be its classical capacity. There exists a constant $\gamma > 0$ such that the following holds: For any code of rate R, and any corresponding decoder, the success probability $P_{\text{succ}}^{\Phi}(n, R)$ is upper bounded by $2^{-\gamma \cdot n(R-C_\Phi)}$ (for sufficiently large n).

Thus the success probability decays exponentially when coding at rates above the capacity.

Background.—Before giving a short overview of our proof, let us briefly recall how the study of the achievability of rates below the capacity can be subdivided into three major components: one begins by setting up a connection between the operational problem of coding and an entropic quantity. More precisely, one can show that there exists codes such that the success probability has a behavior of the form

$$P_{\text{succ}}^{\Phi}(n, R) = 1 - e^{-n\delta(\bar{\chi}^*(\Phi)-R)}, \tag{3}$$

with $\delta > 0$ for rates R smaller than

$$\bar{\chi}^*(\Phi) := \lim_{n\to\infty} \frac{1}{n} \chi^*(\Phi^{\otimes n}). \tag{4}$$

This quantity is the regularized version of the Holevo quantity of the channel Φ, i.e.,

$$\chi^*(\Phi) := \max_{\{p_x, \rho_x\}_x} \chi(\{p_x, \Phi(\rho_x)\}_x), \tag{5}$$

which in turn is defined in terms of the Holevo quantity of an ensemble $\{p_x, \sigma_x\}_x$, given by

$$\chi(\{p_x, \sigma_x\}_x) := S\Big(\sum_x p_x \sigma_x\Big) - \sum_x p_x S(\sigma_x). \tag{6}$$

This is the first step in the study of the coding problem. It reduces the operational problem of coding to the study of the quantity (4). In particular, (3) tells us that we can code with exponentially small error at any rate $R < \bar{\chi}^*(\Phi)$.

The second component is to study general properties of the quantity $\bar{\chi}^*(\Phi)$. The computation of this value is drastically simplified in cases where the Holevo quantity is additive, that is,

$$\chi^*(\Phi^{\otimes n-1} \otimes \Phi) = \chi^*(\Phi^{\otimes n-1}) + \chi^*(\Phi) \tag{7}$$

for all $n > 1$, since this implies $\bar{\chi}^*(\Phi) = \chi^*(\Phi)$. Note that part of this statement, the so-called subadditivity

$$\chi^*(\Phi^{\otimes n-1} \otimes \Phi) \geq \chi^*(\Phi^{\otimes n-1}) + \chi^*(\Phi),$$

is trivial, as it corresponds to restricting to product states. Showing whether or not (7) holds for a given channel Φ is called an additivity problem. It has several equivalent formulations: for example, the quantity $\chi^*(\Lambda)$, for any CPTPM Λ, can be re-expressed in terms of the relative entropy D as

$$\chi^*(\Lambda) = \min_{\sigma} \max_{\rho} D(\Lambda(\rho) \,\|\, \Lambda(\sigma)) \tag{8}$$

as shown in [15]. The physical significance of the additivity property (7) stems from the fact that (4) is a formula for the capacity C_Φ, while (5) is equal to the product-state capacity C_Φ^{prod} [16]. Additivity of χ^* for a channel Φ therefore implies that there is no advantage in using entangled states for coding in the asymptotic limit.

Finally, one needs to investigate the additivity problem [cf. (7)], which is poorly understood in general. King [17] has shown additivity of χ^* for the depolarizing channel (1). His proof uses the fact that for any covariant channel Φ, the Holevo quantity is related to the minimum output entropy [18]

$$S^{\min}(\Phi) := \min_{\rho} S(\Phi(\rho)) \tag{9}$$

by

$$\chi^*(\Phi) = \log d_{\text{out}} - S^{\min}(\Phi), \tag{10}$$

where d_{out} is the dimension of the output space $\mathcal{H}_{\text{out}}$. King then establishes the additivity of $S^{\min}$ for the depolarizing channel Δ_r by showing that the related minimum α-Rényi entropies $S_\alpha^{\min}$ (defined below) are additive for Δ_r. This implies additivity of χ^*, and leads to an explicit formula for the capacity C_{Δ_r}.

Proof outline.—Our approach to coding at rates above the capacity has the same overall structure as the study of the achievability explained above. The strong converse theorem is obtained by (a) relating the decoding probability to entropic quantities, (b) rephrasing the resulting additivity problems and finally (c) showing that the channels (i)–(iii) satisfy these additivity properties.

The relevant quantities in our case turn out to be the following Rényi-entropic versions of the above quantities. For $\alpha \geq 1$, we use [19]

$$\begin{aligned} S_\alpha(\rho) &:= \frac{1}{1-\alpha} \log \operatorname{tr}(\rho^\alpha) \\ D_\alpha(\rho \parallel \sigma) &:= \frac{1}{\alpha - 1} \log \operatorname{tr}(\rho^\alpha \sigma^{1-\alpha}) \\ \chi_\alpha(\{p_x, \sigma_x\}_x) &:= \frac{\alpha}{\alpha - 1} \log \operatorname{tr}\Big(\sum_x p_x \sigma_x^\alpha\Big)^{1/\alpha}. \end{aligned} \tag{11}$$

We also need the corresponding derived quantities $\chi_\alpha^*(\Phi)$, $\bar{\chi}_\alpha^*(\Phi)$, and $S_\alpha^{\min}(\Phi)$ defined as in (5), (4), and (9), respectively.

We now give a sketch of the proof, following the three steps (a)–(c) outlined above (details can be found in [20]). First, we relate our operational problem to the regularized quantity $\bar{\chi}_\alpha^*(\Phi)$ by showing that for any code of rate R, we have

$$P_{\text{succ}}^{\Phi}(n, R) \lesssim 2^{-n[1-(1/\alpha)](R - \bar{\chi}_\alpha^*(\Phi))} \quad \text{for all } \alpha \geq 1 \tag{12}$$

for sufficiently large n. The proof employs techniques used by Ogawa and Nagaoka [10]. This is the analog of (3). It shows that for any rate $R > \bar{\chi}_\alpha^*(\Phi)$, the success probability decays exponentially with n.

Clearly, the quantity $\bar{\chi}_\alpha^*(\Phi)$ again has a particularly simple form if χ_α^* is additive as in (7). To study additivity of the quantity χ_α^*, the second step of our proof is to derive the following analog of (8), essentially following the steps of Schumacher and Westmoreland [15]

$$\begin{aligned} \min_{\sigma_{\text{out}}} \max_{\rho} D_\alpha(\Lambda(\rho) \parallel \sigma_{\text{out}}) &\leq \chi_\alpha^*(\Lambda) \\ &\leq \min_{\sigma_{\text{in}}} \max_{\rho} D_\alpha(\Lambda(\rho) \parallel \Lambda(\sigma_{\text{in}})). \end{aligned} \tag{13}$$

As before, additivity of the quantity χ_α^* is intimately connected to the classical capacity C_Φ: As shown in [10], for every $\varepsilon > 0$, we have $\chi_\alpha^*(\Phi) < C_\Phi + \varepsilon$ for all $\alpha \geq 1$ in some neighborhood of 1. In particular, with (12), this shows that additivity of χ_α^* for all α in the vicinity of 1 implies a strong converse, that is, an exponential decay of the success probability for any rates $R > C_\Phi$. Since it is known [10,11] that coding with product states at rates above the capacity leads to the same exponential behavior, we can conclude that entanglement provides no operational advantage.

Finally, we show additivity of χ_α^* for the special class of channels Φ satisfying our assumptions (i)–(iii). For these channels, the covariance properties imply that both the lower and upper bound in (13) coincide and are attained when σ_{in} and σ_{out} are completely mixed. By definition, this means that these channels satisfy the Rényi-entropic version

$$\chi_\alpha^*(\Phi) = \log d_{\text{out}} - S_\alpha^{\min}(\Phi) \tag{14}$$

of (10). Additivity of χ_α^* is shown by combining (13) with (14), as follows. For $\sigma_{\text{in}} = \mathbb{I}/d$ equal to the fully mixed state, we get

$$\begin{aligned} \chi_\alpha^*(\Phi^{\otimes n}) &\leq \max_{\rho} D_\alpha[\Phi^{\otimes n}(\rho) \parallel \Phi^{\otimes n}((\mathbb{I}/d_{\text{in}})^{\otimes n})] \\ &= \log d_{\text{out}}^n - S_\alpha^{\min}(\Phi^{\otimes n}) = n \log d_{\text{out}} - n S_\alpha^{\min}(\Phi). \end{aligned} \tag{15}$$

In the last step, we used the additivity of the minimum output α-entropy $S_\alpha^{\min}$ for the channels of interest for $\alpha \geq 1$ close to 1 (cf. [21] for qubit unital channels, [17] for the depolarizing channel, and [22–24] for the Werner-Holevo channel). By the subadditivity property of the quantity χ_α^*, we know that $n\chi_\alpha^*(\Phi) \leq \chi_\alpha^*(\Phi^{\otimes n})$. Combining this with (14) and (15) proves additivity, that is, $\bar{\chi}_\alpha^*(\Phi) = \chi_\alpha^*(\Phi) = \log d_{\text{out}} - S_{\min}(\Phi)$. This concludes the proof of our main result.

Conclusions.—In summary, we have shown that for a large class of practically relevant quantum channels, the probability of reliably transmitting nR classical bits by n uses of the channel has an asymptotic behavior of the form $2^{-\gamma n(R-C)}$ for some constant $\gamma > 0$ when coding at rates R above the classical capacity C. Such a statement was previously only known for classical channels and the identity channel. Our result has direct practical applications to quantum cryptography, especially in the so-called noisy-quantum-storage model [25,26], where the adversary is restricted to using low-capacity channels. For these applications, some knowledge about the optimal constant γ will be useful. Our work provides bounds on this value, about which little is known even in the classical case.

On a more fundamental level, our result implies that for the quantum channels considered, using entanglement provides no advantage in all rate regimes. These channels therefore behave just as classical channels with respect to the transmission of classical information. Establishing strong converses for a wider class of channels is of fundamental importance, as this is the natural counterpart of the achievability statement of the capacity. Of particular interest in this context are channels whose Holevo quantity is nonadditive [9]. While we do not explicitly use this fact, the Holevo quantity is additive for the channels considered in this Letter.

Showing that the success probability of decoding has an exponential behavior both below and above the capacity confirms our interpretation of the classical capacity as the single relevant measure of the usefulness of a quantum channel for classical communication.

We acknowledge support by NSF Grants No. PHY-04056720 and No. PHY-0803371.

*rkoenig@caltech.edu
†wehner@caltech.edu

[1] By coding into orthogonal states.

[2] This statement can be seen as a special case of Holevo's channel coding theorem [3].

[3] A. S. Holevo, Prob. Peredachi Inf. **9**, 3 (1973) [Probl. Inf. Transm. **9**, 177 (1973)].

[4] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).

[5] J. Wolfowitz, *Coding Theorems of Information Theory* (Springer, New York, 1964).

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

[7] B. Schumacher and M. Westmoreland, Phys. Rev. A **56**, 131 (1997).

[8] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. Wootters, Phys. Rev. A **54**, 1869 (1996).

[9] M. B. Hastings, Nature Phys. **5**, 255 (2009).

[10] T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory **45**, 2486 (1999).

[11] A. Winter, IEEE Trans. Inf. Theory **45**, 2481 (1999).

[12] A unital channel maps the completely mixed state on $\mathcal{H}_{\text{in}}$ to the completely mixed state on $\mathcal{H}_{\text{out}}$.

[13] The additivity property $S_\alpha^{\min}(\Phi^{\otimes n}) = n \cdot S_\alpha^{\min}(\Phi)$ is equivalent to the multiplicativity of the maximum output α-norm defined as $\nu_\alpha(\Phi) = \max_\rho \|\Phi(\rho)\|_\alpha$, where $\|A\|_\alpha = (\text{tr}|A|^\alpha)^{1/\alpha}$.

[14] R. F. Werner and A. S. Holevo, J. Math. Phys. (N.Y.) **43**, 4353 (2002).

[15] B. Schumacher and M. D. Westmoreland, Phys. Rev. A **63**, 022308 (2001).

[16] In addition to (3), Fano's inequality gives a (weak) converse involving the quantity (4), which shows that $C_\Phi = \bar{\chi}^*(\Phi)$.

[17] C. King, IEEE Trans. Inf. Theory **49**, 221 (2003).

[18] A. S. Holevo, arXiv:quant-ph/0212025.

[19] Note that the expression $\sigma^{1-\alpha}$ in $D_\alpha(\rho \parallel \sigma)$ requires to augment definition (11): We will only invert σ on its support supp(σ), and set $D_\alpha(\rho \parallel \sigma) = \infty$ if supp$(\sigma) \not\subset$ supp(ρ).

[20] See EPAPS Document No. E-PRLTAO-103-061934 for the proof sketched in the main paper worked out in detail. For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.

[21] C. King, J. Math. Phys. (N.Y.) **43**, 4641 (2002).

[22] N. Datta, A. S. Holevo, and Y. M. Suhov, arXiv:quant-ph/0403072.

[23] R. Alicki and M. Fannes, Open Syst. Inf. Dyn. **11**, 1230 (2004).

[24] K. Matsumoto and F. Yura, J. Phys. A **37**, L167 (2004).

[25] S. Wehner, C. Schaffner, and B. M. Terhal, Phys. Rev. Lett. **100**, 220502 (2008).

[26] C. Schaffner, B. Terhal, and S. Wehner, arXiv:0807.1333.