One-Shot Classical-Quantum Capacity and Hypothesis Testing

Ligong Wang and Renato Renner

Phys. Rev. Lett. **108**, 200501 — Published 15 May 2012

DOI: 10.1103/PhysRevLett.108.200501

One-Shot Classical-Quantum Capacity and Hypothesis Testing

Ligong Wang*
Research Laboratory of Electronics, MIT, Cambridge, MA, USA

Renato Renner†
Institute for Theoretical Physics, ETH Zurich, Switzerland

The one-shot classical capacity of a quantum channel quantifies the amount of classical information that can be transmitted through a single use of the channel such that the error probability is below a certain threshold. In this work, we show that this capacity is well approximated by a relative-entropy-type measure defined via hypothesis testing. Combined with a quantum version of Stein's Lemma, our results give a conceptually simple proof of the well-known Holevo-Schumacher-Westmoreland Theorem for the capacity of memoryless channels. More generally, we obtain tight capacity formulas for arbitrary (not necessarily memoryless) channels.

PACS numbers: 89.70.-a,89.70.Kn,89.70.Cf

In Information Theory, a *channel* models a physical device that takes an input and generates an output. One may, for instance, think of a communication channel (such as an optical fiber) that connects a sender (who provides the input) with a receiver (who obtains an output, which may deviate from the input). Another example is a memory device, such as a hard drive, where the input consists of the data written into the device, and where the output is the (generally noisy) data that is retrieved from the device at a later point in time.

A central question studied in Information Theory is whether, and how, a channel can be used to transmit data reliably in spite of the channel noise. This is usually achieved by *coding*, where an *encoder* prepares the channel input by adding redundancy to the data to be transmitted, and where a *decoder* reconstructs the data from the noisy channel output.

Here we focus on the case of *classical-quantum channel coding*, where the data to be transmitted reliably are classical. No assumptions are made about the channel that is used to achieve this task, i.e., the inputs and outputs may be arbitrary quantum states. However, since the quantum-mechanical structure of the input space is irrelevant for the encoding of classical data, it can be represented by a (classical) set $\mathcal{X}$. For any input $x \in \mathcal{X}$, the channel produces an output, specified by a density operator ρ_x on a Hilbert space $\mathbb{B}$. For our purposes, it is therefore sufficient to characterize a channel by a mapping $x \mapsto \rho_x$ from a set $\mathcal{X}$ to a set of density operators.

Classical-quantum channel coding has been studied extensively in a scenario where the channel can be used arbitrarily many times. The *channel coding theorem for stationary memoryless classical-quantum channels*, established by Holevo [1] and Schumacher and Westmoreland [2], provides an explicit formula (see (10)) for the rate at which data can be transmitted under the assumption that each use of the channel is independent of the previous uses. More general channel coding theorems that do not rely on this independence assumption have been developed in later work by Hayashi and Nagaoka [3] and by Kretschmann and Werner [4]. These results are asymptotic, i.e., they refer to a limit where the number of channel uses tends to infinity while the probability of error is required to approach zero.

Here we consider a scenario where a given quantum channel is used only *once* and derive tight bounds on the number of classical bits that can be transmitted with a given average error probability ϵ, in the following referred to as the *ϵ-one-shot classical-quantum capacity*. This one-shot approach provides a high level of generality, as nothing needs to be assumed about the structure of the channel [18]. (Note that any situation in which a channel is used repeatedly can be equivalently described as one single use of a larger channel.) In particular, our bounds on the channel capacities imply the aforementioned Holevo-Schumacher-Westmoreland Theorem for the capacity of memoryless channels, as well as the generalizations by Hayashi and Nagaoka. On the other hand, our work generalizes similar one-shot results for classical channels [5–7]. Despite their generality, the bounds as well as their proofs are remarkably simple. We hope that our approach may therefore also be of pedagogical value.

Our derivation is based on the idea, already exploited in previous works (see, e.g., [3, 8–10]), of relating the problem of channel coding to hypothesis testing. Here, we use hypothesis testing directly to define a relative-entropy-type quantity, denoted $D_{\mathrm{H}}^{\epsilon}(\cdot\|\cdot)$ (see (1)). Our main result asserts that the one-shot channel capacity is well approximated by $D_{\mathrm{H}}^{\epsilon}(\cdot\|\cdot)$ (Theorem 1).

The remainder of this Letter is structured as follows. We briefly describe hypothesis testing and state a few properties of the quantity $D_{\mathrm{H}}^{\epsilon}(\cdot\|\cdot)$. We then state and prove our main result which provides upper and lower bounds on the ϵ-one-shot classical-quantum capacity in terms of $D_{\mathrm{H}}^{\epsilon}(\cdot\|\cdot)$. Finally, we show how the known asymptotic bounds (for arbitrarily many channel uses) can be obtained from Theorem 1.

Hypothesis Testing and $D_{\mathrm{H}}^{\epsilon}(\cdot\|\cdot)$.—Hypothesis testing is the task of distinguishing two possible states of a system, ρ and σ. A strategy for this task is specified by a Positive Operator Valued Measure (POVM) with two elements, Q and $I-Q$, corresponding to the two possible values for the guess. The probability that the strategy produces a *correct* guess on input ρ is given by $\mathrm{tr}[Q\rho]$, and the probability that it produces a *wrong* guess on input σ is $\mathrm{tr}[Q\sigma]$. We define the *hypothesis testing relative entropy* $D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma)$ as

$$D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma) \triangleq -\log_2 \inf_{\substack{Q:0\le Q\le I,\\ \mathrm{tr}[Q\rho]\ge 1-\epsilon}} \mathrm{tr}[Q\sigma]. \tag{1}$$

Note that $D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma)$ is a semidefinite program and can therefore be evaluated efficiently.

As its name suggests, $D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma)$ can be understood as a relative entropy. In particular, for $\epsilon = 0$, it is equal to *Rényi's relative entropy of order* 0, $D_0(\rho\|\sigma) = -\log_2 \mathrm{tr}[\rho^0\sigma]$, where ρ^0 denotes the projector onto the support of ρ. For $\epsilon > 0$, it corresponds to a "smoothed" variant of the relative Rényi entropy of order 0 used by Buscemi and Datta [11] for characterizing the quantum capacity of channels [19]. $D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma)$ has the following properties, all of which hold for all ρ, σ and $\epsilon \in [0,1)$:

1. *Positivity:*

$$D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma) \ge 0,$$

with equality if $\rho = \sigma$ and $\epsilon = 0$.

2. *Data Processing Inequality (DPI):* for any Completely Positive Map (CPM) $\mathcal{E}$,

$$D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma) \ge D_{\mathrm{H}}^{\epsilon}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)).$$

3. Let $D(\cdot\|\cdot)$ denote the usual quantum relative entropy, then

$$D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma) \le \big(D(\rho\|\sigma) + H_{\mathrm{b}}(\epsilon)\big)/(1-\epsilon), \tag{2}$$

where $H_{\mathrm{b}}(\cdot)$ is the binary entropy function.

Positivity follows immediately from the definition.

To prove the DPI, consider any POVM to distinguish $\mathcal{E}(\rho)$ from $\mathcal{E}(\sigma)$. We can then construct a new POVM to distinguish ρ from σ by preceding the given POVM with the CPM $\mathcal{E}$. This new POVM clearly gives the same error probabilities (in distinguishing ρ and σ) as the original POVM (in distinguishing $\mathcal{E}(\rho)$ and $\mathcal{E}(\sigma)$). The DPI then follows because an optimization over all possible strategies for distinguishing ρ and σ can only decrease the failure probability.

To prove (2), first see that it holds when $D(\rho\|\sigma)$ is replaced by $D(P_\rho\|P_\sigma)$, where P_ρ is the distribution of the outcomes of the optimal POVM performed on ρ, namely, it is $(1-\epsilon,\epsilon)$, and similarly for P_σ which is $(2^{-D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma)}, 1-2^{-D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma)})$. This can be shown by directly computing $D(P_\rho\|P_\sigma)$. Then (2) follows because $D(\cdot\|\cdot)$ satisfies the DPI so $D(\rho\|\sigma) \ge D(P_\rho\|P_\sigma)$.

A further connection between $D_{\mathrm{H}}^{\epsilon}(\cdot\|\cdot)$ and $D(\cdot\|\cdot)$ is the Quantum Stein's Lemma [8, 12], which we restate as follows.

Lemma 1 (Quantum Stein's Lemma)**.** *For any two states ρ and σ on a Hilbert space and for any $\epsilon \in (0,1)$,*

$$\lim_{n\to\infty} \frac{1}{n} D_{\mathrm{H}}^{\epsilon}(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma).$$

Statement and Proof of the Main Result.—Before stating our main result, we introduce some general terminology. The encoder is specified by a list of inputs, $\{x_i\}$, $i \in \{1,\dots,m\}$, called a *codebook* of size m. The decoder applies a corresponding *decoding POVM*, which acts on $\mathbb{B}$ and has m elements. A decoding error occurs if the output of the decoding POVM is not equal to the index i of the input x_i fed into the channel. An (m,ϵ)*-code* consists of a codebook of size m and a corresponding decoding POVM such that, when the message is chosen uniformly, the average probability of a decoding error is at most ϵ [20].

The main result of this Letter is the following theorem.

Theorem 1. *The ϵ-one-shot classical-quantum capacity of a channel $x \mapsto \rho_x$, i.e., the largest number R for which a $(2^R,\epsilon)$-code exists, satisfies*

$$\sup_{P_X} D_{\mathrm{H}}^{\epsilon}(\pi^{\mathbb{AB}}\|\pi^{\mathbb{A}}\otimes\pi^{\mathbb{B}}) \ge R \ge \sup_{P_X} D_{\mathrm{H}}^{\epsilon/2}(\pi^{\mathbb{AB}}\|\pi^{\mathbb{A}}\otimes\pi^{\mathbb{B}}) - \log_2\frac{1}{\epsilon} - 4\ , \tag{3}$$

where $\pi^{\mathbb{AB}}$ is the joint state of the input and output for an input chosen according to the distribution P_X, i.e.,

$$\pi^{\mathbb{AB}} \triangleq \sum_{x\in\mathcal{X}} P_X(x)|x\rangle\langle x|^{\mathbb{A}} \otimes \rho_x^{\mathbb{B}}\ ,$$

for any representation of the inputs x in terms of orthonormal vectors $|x\rangle^{\mathbb{A}}$ on a Hilbert space $\mathbb{A}$, and where $\pi^{\mathbb{A}}$ and $\pi^{\mathbb{B}}$ are the corresponding marginals.

The proof of Theorem 1 is divided into two parts, one for the first inequality (referred to as the *converse*) and the other for the second inequality (the *achievability*). We start with the converse which asserts that, if a $(2^R,\epsilon)$-code exists, then

$$R \le \sup_{P_X} D_{\mathrm{H}}^{\epsilon}(\pi^{\mathbb{AB}}\|\pi^{\mathbb{A}}\otimes\pi^{\mathbb{B}}). \tag{4}$$

Proof of Theorem 1—Converse Part. By definition, it is sufficient to prove (4) for a uniform distribution on the x's used in the codebook, so we can focus on states $\pi^{\mathbb{AB}}$ of the form

$$\pi^{\mathbb{AB}} = 2^{-R}\sum_{i=1}^{2^R} |x_i\rangle\langle x_i| \otimes \rho_{x_i}\ .$$

Note that the decoding POVM combined with the inverse of the encoding map (which is classical) can be viewed as a CPM. This CPM maps $\pi^{\mathbb{AB}}$ to the (classical) state $P_{MM'}$ denoting the joint distribution of the transmitted message M and the decoder's guess M'. Similarly, it maps $\pi^{\mathbb{A}}\otimes\pi^{\mathbb{B}}$ to $P_M\otimes P_{M'}$. Hence, it follows from the DPI for $D_{\mathrm{H}}^{\epsilon}(\rho\|\sigma)$ that

$$D_{\mathrm{H}}^{\epsilon}(P_{MM'}\|P_M\otimes P_{M'}) \leq D_{\mathrm{H}}^{\epsilon}(\pi^{\mathbb{AB}}\|\pi^{\mathbb{A}}\otimes\pi^{\mathbb{B}}) \ .$$

It thus remains to prove

$$R \leq D_{\mathrm{H}}^{\epsilon}(P_{MM'}\|P_M\otimes P_{M'}) \ . \tag{5}$$

For this, we consider a (possibly suboptimal) strategy to distinguish between $P_{MM'}$ and $P_M\otimes P_{M'}$. The strategy guesses $P_{MM'}$ if $M=M'$, and guesses $P_M\otimes P_{M'}$ otherwise. Using this distinguishing strategy, the probability of guessing $P_M\otimes P_{M'}$ on state $P_{MM'}$ is exactly the probability that $M\neq M'$ computed from $P_{MM'}$, namely, the average probability of a decoding error, and is thus not larger than ϵ by assumption. Furthermore, the probability of guessing $P_{MM'}$ on state $P_M\otimes P_{M'}$ is given by

$$\sum_{i=1}^{2^R} P_M(i)\cdot P_{M'}(i) = 2^{-R}\sum_{i=1}^{2^R} P_{M'}(i) = 2^{-R} \ .$$

This implies (5). □

We proceed with the achievability part of Theorem 1. We show a slightly stronger result which asserts that, for any $\epsilon>\epsilon'>0$ and $c>0$, there exists a $(2^R,\epsilon)$-code with

$$R \geq \sup_{P_X} D_{\mathrm{H}}^{\epsilon'}(\pi^{\mathbb{AB}}\|\pi^{\mathbb{A}}\otimes\pi^{\mathbb{B}}) - \log_2\frac{2+c+c^{-1}}{\epsilon-(1+c)\epsilon'} \ . \tag{6}$$

Choosing $c=1/3$ and $\epsilon'=\epsilon/2$, this bound implies the second inequality of (3).

The main technique we need for proving (6) is the following lemma by Hayashi and Nagaoka [3, Lemma 2]:

Lemma 2. *For any positive real c and any operators $0\leq S\leq I$ and $T\geq 0$, we have*

$$\begin{aligned} I-(S+T)^{-1/2}S(S+T)^{-1/2} \\ \leq (1+c)(I-S)+(2+c+c^{-1})T. \end{aligned}$$

Proof of Theorem 1—Achievability Part. Fix $\epsilon'\in(0,\epsilon)$, $c>0$, and P_X. We are going to show that there exists a $(2^R,\epsilon)$-code such that

$$\epsilon \leq (1+c)\epsilon' + (2+c+c^{-1})2^{R-D_{\mathrm{H}}^{\epsilon'}(\pi^{\mathbb{AB}}\|\pi^{\mathbb{A}}\otimes\pi^{\mathbb{B}})} \ ,$$

which immediately implies (6).

Let Q be an operator acting on $\mathbb{AB}$ such that $0\leq Q\leq I$ and $\mathrm{tr}\big[Q\pi^{\mathbb{AB}}\big]\geq 1-\epsilon'$. By definition, it suffices to prove that there exists a codebook and a decoding POVM with error probability

$$\epsilon \leq (1+c)\epsilon' + (2+c+c^{-1})2^R\mathrm{tr}\big[Q(\pi^{\mathbb{A}}\otimes\pi^{\mathbb{B}})\big] \ . \tag{7}$$

We generate a codebook by choosing its codewords x_j at random, each independently according to the distribution P_X. Furthermore, we define the corresponding decoding POVM by its elements,

$$E_i = \left(\sum_{j=1}^{2^R} A_{x_j}\right)^{-\frac{1}{2}} A_{x_i}\left(\sum_{j=1}^{2^R} A_{x_j}\right)^{-\frac{1}{2}} \ ,$$

where $A_x \triangleq \mathrm{tr}_{\mathbb{A}}\big[\big(|x\rangle\langle x|^{\mathbb{A}}\otimes I^{\mathbb{B}}\big)\,Q\big]$.

For a specific codebook $\{x_j\}$ and the transmitted codeword x_i, the probability of error is given by

$$\Pr(\mathrm{error}|x_i,\{x_j\}) = \mathrm{tr}[(I-E_i)\rho_{x_i}] \ .$$

We now use Lemma 2 with $S=A_{x_i}$ and $T=\sum_{j\neq i}A_{x_j}$ to bound this by

$$\begin{aligned}\Pr(\mathrm{error}|x_i,\{x_j\}) &\leq (1+c)\big(1-\mathrm{tr}[A_{x_i}\rho_{x_i}]\big) \\ &\quad + (2+c+c^{-1})\sum_{j\neq i}\mathrm{tr}[A_{x_j}\rho_{x_i}].\end{aligned}$$

Averaging over all codebooks, but keeping the transmitted codeword x_i fixed, we find

$$\begin{aligned}\Pr(\mathrm{error}|x_i) &\leq (1+c)\big(1-\mathrm{tr}[A_{x_i}\rho_{x_i}]\big) \\ &\quad + (2+c+c^{-1})(2^R-1)\mathrm{tr}\Big[\sum_{x'\in\mathcal{X}}P_X(x')A_{x'}\rho_{x_i}\Big].\end{aligned}$$

Averaging now in addition over the transmitted codeword x_i, we obtain the upper bound

$$\begin{aligned}\Pr(\mathrm{error}) &\leq (1+c)\big(1-\sum_x P_X(x)\mathrm{tr}[A_x\rho_x]\big) \\ &\quad + (2+c+c^{-1})2^R\mathrm{tr}\Big[\sum_{x'}P_X(x')A_{x'}\sum_x P_X(x)\rho_x\Big].\end{aligned} \tag{8}$$

Note that

$$\begin{aligned}\sum_x P_X(x)\mathrm{tr}[A_x\rho_x] &= \sum_x P_X(x)\mathrm{tr}\Big[Q|x\rangle\langle x|^{\mathbb{A}}\otimes\rho_x^{\mathbb{B}}\Big] \\ &= \mathrm{tr}\big[Q\pi^{\mathbb{AB}}\big] \geq 1-\epsilon'\end{aligned}$$

and

$$\begin{aligned}&\mathrm{tr}\Big[\sum_x P_X(x')A_{x'}\sum_x P_X(x)\rho_x\Big] \\ &\quad = \sum_{x'}P_X(x')\mathrm{tr}\Big[Q|x'\rangle\langle x'|\otimes\sum_x P_X(x)\rho_x\Big] \\ &\quad = \mathrm{tr}\big[Q(\pi^{\mathbb{A}}\otimes\pi^{\mathbb{B}})\big].\end{aligned}$$

Inserting these expressions into (8) we find that the upper bound (7) holds for the probability of error averaged over the class of codebooks we generated. Thus there must exist at least one codebook whose error probability ϵ satisfies (7). □

Asymptotic Analysis.—Theorem 1 applies to the transmission of a message in a single use of the channel. Obviously, a channel that can be used n times can always be modeled as one big single-use channel. We can thus retrieve the known expressions for the (usual) capacity of channels, i.e., the average number of bits that can be transmitted per channel use in the limit where the channel is used arbitrarily often and the error ϵ approaches 0. Most generally, a channel that can be used an arbitrary number of times is characterized by a sequence of mappings $x_n \mapsto \rho^n$, $n \in \{1, 2, \ldots\}$, where $x_n \in \mathcal{X}_n$ represents an input state over n channel-uses [21], and where ρ^n is a density operator on $\mathbb{B}^{\otimes n}$. Note that such a channel need not have any structure such as "causality" as defined in [4]. From Theorem 1 it immediately follows that the capacity of any channel is given by

$$C = \lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{n} \sup_{P_{X_n}} D_{\mathrm{H}}^{\epsilon}(\pi^{\mathbb{A}_n \otimes \mathbb{B}^{\otimes n}} \| \pi^{\mathbb{A}_n} \otimes \pi^{\mathbb{B}^{\otimes n}}), \qquad (9)$$

where $\mathbb{A}_n$ denotes the Hilbert space spanned by orthonormal states $|x_n\rangle$ for all $x_n \in \mathcal{X}_n$. This expression is equivalent to [3, Theorem 1] [22]. We can also derive similar results for the *optimistic capacity* and the *ϵ-capacity*, see [13].

Now consider a *memoryless* channel whose behavior in each use is independent of the previous uses. The capacity C of such a channel is given by the well-known Holevo-Schumacher-Westmoreland Theorem [1, 2]:

$$C = \lim_{k \to \infty} \frac{1}{k} \sup_{P_{X_k}} D(\pi^{\mathbb{A}_k \otimes \mathbb{B}^{\otimes k}} \| \pi^{\mathbb{A}_k} \otimes \pi^{\mathbb{B}^{\otimes k}}) \; . \qquad (10)$$

Note that $D(\pi^{\mathbb{A}_k \otimes \mathbb{B}^{\otimes k}} \| \pi^{\mathbb{A}_k} \otimes \pi^{\mathbb{B}^{\otimes k}})$ may equivalently be written as a mutual information $I(\mathbb{A}_k; \mathbb{B}^{\otimes k})$. This theorem can be proved easily using (9).

Proof of (10)*.* To show achievability, i.e., that C is lower-bounded by the right-hand side (RHS) of (10), we restrict the supremum in (9) to product distributions on k-use states, so the joint state $\pi^{\mathbb{A}_n \otimes \mathbb{B}^{\otimes n}}$ looks like $(\pi^{\mathbb{A}_k \otimes \mathbb{B}^{\otimes k}})^{\otimes (n/k)}$ [23]. We then let n tend to infinity and apply Lemma 1 to obtain that, for any k,

$$C \geq \frac{1}{k} \sup_{P_{X_k}} D(\pi^{\mathbb{A}_k \otimes \mathbb{B}^{\otimes k}} \| \pi^{\mathbb{A}_k} \otimes \pi^{\mathbb{B}^{\otimes k}}). \qquad (11)$$

This concludes the proof of the achievability part.

The converse, i.e., that C is upper-bounded by the RHS of (10), follows immediately from (9) and (2). □

To conclude, it may be interesting to compare Theorem 1 to other recently derived bounds on the one-shot capacity of classical-quantum channels [14, 15]. The bounds of [14] are different from ours in that they are not known to coincide asymptotically for arbitrary channels. In [15], it has been shown that the one-shot classical-quantum capacity R of a channel can be approximated (up to additive terms of the order $\log_2 1/\epsilon$) by

$$R \approx \max_{P_X} H_{\min}^{\epsilon}(A)_{\pi^{\mathbb{A}}} - H_{\max}^{\epsilon}(A|B)_{\pi^{\mathbb{AB}}} \; ,$$

where $H_{\min}^{\epsilon}$ and $H_{\max}^{\epsilon}$ denote the smooth min- and max-entropies, which have recently been shown to be the relevant quantities for characterizing a number of information-theoretic tasks (see, e.g., [16] for definitions and properties). Combined with our result, this suggests that there is a deeper and more general relation between hypothesis testing and smooth entropies (and, therefore, the associated operational quantities). Exploring this link is left as an open question for future work.

Acknowledgments.—LW acknowledges support from the US Air Force Office of Scientific Research (grant No. FA9550-11-1-0183) and the National Science Foundation (grant No. CCF-1017772). RR acknowledges support from the Swiss National Science Foundation (grant Nos. 200021-119868, 200020-135048, and the NCCR "QSIT") and the European Research Council (ERC) (grant No. 258932).

* Electronic address: `wlg@mit.edu`
† Electronic address: `renner@phys.ethz.ch`

[1] A. S. Holevo, IEEE Trans. Inform. Theory, **44**, 269 (1998).
[2] B. Schumacher and M. D. Westmoreland, Phys. Rev. A, **56**, 131 (1997).
[3] M. Hayashi and H. Nagaoka, IEEE Trans. Inform. Theory, **49**, 1753 (2003).
[4] D. Kretschmann and R. F. Werner, Phys. Rev. A, **72**, 062323 (2005).
[5] R. Renner, S. Wolf, and J. Wullschleger, in *Proc. IEEE Int. Symp. Inform. Theory* (Seattle, Washington, USA, 2006).
[6] L. Wang, R. Colbeck, and R. Renner, in *Proc. IEEE Int. Symp. Inform. Theory* (Seoul, Korea, 2009) pp. 1804–1808.
[7] Y. Polyanskiy, H. V. Poor, and S. Verdú, in *Proc. IEEE Int. Symp. Inform. Theory* (Toronto, Canada, 2008).
[8] T. Ogawa and H. Nagaoka, IEEE Trans. Inform. Theory, **46**, 2428 (2000).
[9] T. Ogawa and H. Nagaoka, in *Proc. IEEE Int. Symp. Inform. Theory* (Lausanne, Switzerland, 2002).
[10] M. Hayashi, Phys. Rev. A, **76**, 062301 (2007).
[11] F. Buscemi and N. Datta, IEEE Trans. Inform. Theory, **56**, 1447 (2010).
[12] F. Hiai and D. Petz, Comm. Math. Phys., **143**, 99 (1991).
[13] L. Wang, *Information-Theoretic Aspects of Optical Communications*, Ph.D. dissertation, volume 6 of the ETH Series in Information Theory and its Applications, ETH Zurich (2011), editor: Amos Lapidoth.
[14] M. Mosonyi and N. Datta, J. Math. Phys., **50**, 072104 (2009).
[15] J. Renes and R. Renner, IEEE Trans. Inform. Theory, **57**, 7377 (2011).

[16] M. Tomamichel, R. Colbeck, and R. Renner, IEEE Trans. Inform. Theory, **55**, 5840 (2009).
[17] N. Datta, IEEE Trans. Inform. Theory, **55**, 2816 (2009).
[18] In particular, in contrast to previous work, there is no need no define channels as sequences of mappings.
[19] It is also similar to a relative-entropy-type quantity used in [17], although the precise relation to this quantity is not known.
[20] It is well known that, in single-user scenarios, the (asymptotic) capacity does not depend on whether the average (over uniformly chosen messages) or the maximum probability of error is considered. In the one-shot case, one can construct a code that has maximum probability of error not larger than 2ϵ from a code with average probability of error ϵ, thereby sacrificing one bit.
[21] Here $\mathcal{X}_n$ cannot be replaced by $\mathcal{X}^{\times n}$, which only represents *product* states.
[22] While the expressions used in [3] look completely different, one can (non-operationally) prove their equivalence. See [13].
[23] As n tends to infinity, the problem that n might not be divisible by k becomes negligible.