

*Tema con variazioni:* quantum channel capacity

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2004 New J. Phys. 6 26

(<http://iopscience.iop.org/1367-2630/6/1/026>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 38.107.179.212

The article was downloaded on 14/02/2012 at 22:50

Please note that [terms and conditions apply](#).

## ***Tema con variazioni:* quantum channel capacity**

**Dennis Kretschmann and Reinhard F Werner**

Institut für Mathematische Physik, Technische Universität Braunschweig,  
Mendelssohnstr. 3, 38106 Braunschweig, Germany  
E-mail: [d.kretschmann@tu-bs.de](mailto:d.kretschmann@tu-bs.de)

*New Journal of Physics* **6** (2004) 26

Received 11 November 2003

Published 23 February 2004

Online at <http://www.njp.org/> (DOI: 10.1088/1367-2630/6/1/026)

**Abstract.** Channel capacity describes the size of the nearly ideal channels, which can be obtained from many uses of a given channel, using an optimal error correcting code. In this paper we collect and compare minor and major variations in the mathematically precise statements of this idea which have been put forward in the literature. We show that all the variations considered lead to equivalent capacity definitions. In particular, it makes no difference whether one requires mean or maximal errors to go to zero, and it makes no difference whether errors are required to vanish for any sequence of block sizes compatible with the rate, or only for one infinite sequence.

**Contents**

<b>1. Introduction</b>	<b>2</b>
1.1. Notations . . . . .	3
<b>2. Tema: quantum channel capacity</b>	<b>4</b>
2.1. Prima variazione: choice of units . . . . .	4
2.2. Seconda variazione: testing only one sequence . . . . .	4
2.3. Terza variazione: minimum fidelity . . . . .	5
2.4. Quarta variazione: average fidelity . . . . .	5
2.5. Quinta variazione: entanglement fidelity . . . . .	6
2.6. Sesta variazione: entropy rate . . . . .	6
2.7. Settima variazione: errors vanishing quickly or not at all . . . . .	7
2.8. Ottava variazione: isometric encodings and homomorphic decodings . . . . .	8
2.9. Nona variazione: coding with a little help from a classical friend . . . . .	8
2.10. Coda: the coding theorem . . . . .	8
<b>3. Elementary properties of channel capacity</b>	<b>9</b>
3.1. Basic inequalities . . . . .	9
3.2. Quantum capacity of noiseless channels . . . . .	10
3.3. Partial transposition bound . . . . .	11
<b>4. Alternative error criteria</b>	<b>12</b>
4.1. Preliminaries . . . . .	12
4.2. Four equivalent distance measures . . . . .	14
4.3. Average fidelity and channel fidelity . . . . .	15
4.4. Entanglement fidelity and entropy rate . . . . .	17
4.5. Entanglement generation capacity . . . . .	20
<b>5. Isometric encoding suffices</b>	<b>21</b>
<b>6. Classical side information</b>	<b>23</b>
6.1. Classical forward communication does not increase the channel capacity . . . . .	23
6.2. Average fidelity by forward communication . . . . .	25
<b>7. Testing a single sequence</b>	<b>26</b>
7.1. Subexponential sequences . . . . .	27
7.2. A counterexample . . . . .	28
7.3. Hashing helps . . . . .	29
<b>Acknowledgments</b>	<b>31</b>
<b>References</b>	<b>31</b>

**1. Introduction**

Quantum channel capacity is one of the key quantitative notions of the young field of quantum information theory. Whenever one asks ‘how much quantum information’ can be stored in a device, or sent on a transmission line, it is implicitly a question about the capacity of a channel. Like Shannon’s classical definition, the concept applies also to noisy channels, which do corrupt the signal. In this case one may apply an error correction scheme, and still use the channel like

an almost ideal one. Capacity expresses this quantitatively: it is the maximal number of ideal qubit (resp. bit) transmissions per use of the channel, taken in the limit of long messages and using error correction schemes asymptotically eliminating all errors.

Many of the terms in this informal definition can be, and in fact have been, formalized mathematically in different ways. As a result, there are many published definitions of quantum capacity in the literature. Some of these are immediately seen to be equivalent, but with other variants this is less obvious. Moreover, some of the differences seem to have gone unnoticed, creating the danger that some results would be unwittingly transferred between inequivalent concepts, creating a mixture of rigorous argument and folklore hard to unravel.

The purpose of the present paper is to show that, fortunately, all the major definitions are indeed equivalent. In order to make the presentation self-contained, we have also included abridged versions of arguments from the literature. Other points, however, e.g., concerning the question whether the rate has to be achieved on every sequence of increasing blocks, or just on an infinite, possibly sparse set of increasing block sizes, seem to be new. We have also made an effort to lay out the required tools carefully, so that they can be used in other applications.

All this does not help much to come closer to the proof of a coding theorem, i.e., to a rigorous formula for the capacity not requiring the solution of asymptotically large optimization problems. Major progress in this direction has recently been obtained by Shor [1, 2] and Devetak [3]. We hope that our work will contribute to an unambiguous interpretation of these results, as well.

The key chapter (section 2) of this paper begins with presenting the theme: a basic rigorous definition of quantum channel capacity. This is followed by nine logical variations on this theme, which like musical variations are not all of the same weight. In each variation a result is stated to the effect that a modified definition is equivalent to the basic one after all. All proofs, however, are left to the later sections. A coda at the end of the variations comments on the coding theorem and recent developments.

### 1.1. Notations

In order to state the basic definition of capacity and its variations, we have to introduce some notation. A *quantum channel* which transforms input systems described by a Hilbert space  $\mathcal{H}_1$  into output systems described by a (possibly different) Hilbert space  $\mathcal{H}_2$  is represented by a completely positive trace-preserving linear map  $T: \mathcal{B}_*(\mathcal{H}_1) \rightarrow \mathcal{B}_*(\mathcal{H}_2)$ , where by  $\mathcal{B}_*(\mathcal{H})$  we denote the space of trace class operators on  $\mathcal{H}$ . This map takes the input state to the output state, i.e., we work in the Schrödinger picture (see Kraus' textbook [4] for a detailed description of the concept of quantum operations).

The definition of channel capacity requires the comparison of the channel after correction with an ideal channel. As a measure of the distance between two channels we take the *norm of complete boundedness* (or *cb-norm*, for short) [5], denoted by  $\|\cdot\|_{cb}$ . For two channels  $T$  and  $S$ , the distance  $\frac{1}{2}\|T - S\|_{cb}$  can be defined as the largest difference between the overall probabilities in two statistical quantum experiments differing only by exchanging one use of  $S$  by one use of  $T$ . These experiments may involve entangling the systems on which the channels act with arbitrary further systems. Equivalently, we may set  $\|T\|_{cb} = \sup_n \|T \otimes \text{id}_n\|_\infty$ , where the norm is the norm of linear operators between the Banach spaces  $\mathcal{B}_*(\mathcal{H}_i)$ , and  $\text{id}_n$  denotes the identity map (ideal channel) on the  $n \times n$  matrices.

Among the properties which make the cb-norm well-suited for capacity estimates are multiplicativity,  $\|T_1 \otimes T_2\|_{cb} = \|T_1\|_{cb} \|T_2\|_{cb}$ , and unitality,  $\|T\|_{cb} = 1$  for any channel  $T$ . The equivalence with other error measures is discussed extensively below.

Note that throughout this work we use base two logarithms, and we write  $\text{ld } x := \log_2 x$ .

## 2. Tema: quantum channel capacity

**Definition 2.1.** A positive number  $R$  is called the achievable rate for the quantum channel  $T: \mathcal{B}_*(\mathcal{H}_1) \rightarrow \mathcal{B}_*(\mathcal{H}_2)$  with respect to the quantum channel  $S: \mathcal{B}_*(\mathcal{H}_3) \rightarrow \mathcal{B}_*(\mathcal{H}_4)$  iff for any pair of integer sequences  $(n_\nu)_{\nu \in \mathbb{N}}$  and  $(m_\nu)_{\nu \in \mathbb{N}}$  with  $\lim_{\nu \rightarrow \infty} n_\nu = \infty$  and  $\overline{\lim}_{\nu \rightarrow \infty} (m_\nu/n_\nu) \leq R$  we have

$$\lim_{\nu \rightarrow \infty} \Delta(n_\nu, m_\nu) = 0, \quad (1)$$

where we set

$$\Delta(n_\nu, m_\nu) := \inf_{D, E} \|DT^{\otimes n_\nu} E - S^{\otimes m_\nu}\|_{cb}, \quad (2)$$

the infimum taken over all encoding channels  $E$  and decoding channels  $D$  with suitable domain and range. The channel capacity  $Q(T, S)$  of  $T$  with respect to  $S$  is defined to be the supremum of all achievable rates. The quantum capacity is the special case  $Q(T) := Q(T, \text{id}_2)$ , with  $\text{id}_2$  being the ideal qubit channel.

This definition is a transcription of Claude E Shannon's definition of the capacity of a discrete memoryless channel in classical information theory, as presented originally in his famous 1948 paper [6] and now found in most standard textbooks on the subject (e.g., [7, 8]). To make the translation one only needs to express Shannon's maximal error probabilities in terms of norm estimates [9] and take an ideal one-bit channel rather than the one-qubit channel as the reference. This choice can also be made for quantum channels  $T$ , defining the capacity  $C(T)$  of a quantum channel for *classical information*. Much more is known about  $C(T)$  than about  $Q(T)$  [10, 11].

### 2.1. Prima variazione: choice of units

Formally, definition 2.1 assigns a special role to the ideal qubit channel  $\text{id}_2$ . Is this essential? What do we get if we take the ideal channel  $\text{id}_n$  on a Hilbert space of some dimension  $n > 2$  as reference?

We will show in section 3.2 that the choice  $n = 2$  only amounts to a choice of units, fixing the unit *bit*:

$$Q(T, \text{id}_n) = \frac{\text{ld } m}{\text{ld } n} Q(T, \text{id}_m). \quad (3)$$

### 2.2. Seconda variazione: testing only one sequence

At first sight definition 2.1 of channel capacity, as given above and widely used throughout the community [9], [12]–[15], seems a little impractical, since it involves checking an infinite number of pairs of sequences when testing a given rate  $R$ . Work would be substantially reduced if only

one such pair had to be tested. For the sake of discussion let us say that a rate  $R$  is *sporadically achievable* if, for *some* pair of sequences  $(n_\nu)_{\nu \in \mathbb{N}}$ ,  $(m_\nu)_{\nu \in \mathbb{N}}$ , with  $n_\nu \rightarrow \infty$  and vanishing errors, the rate  $R$  is achieved infinitely often:  $\overline{\lim}_{\nu \rightarrow \infty} (m_\nu/n_\nu) = R$ . For example, there might be a special coding scheme which utilizes some rare number theoretical properties of  $n$  and  $m$ . Many published definitions [16]–[19] would accept sporadically achievable rates as achievable in the capacity definition. Often the choice  $n_\nu = \nu$  is made [20]–[24]. While this sequence of block sizes can hardly be called ‘sporadic’, it is a logically similar variation to the sparse sequences, so we include it for convenience.

In section 7 we show that all sporadically achievable rates are, in fact, achievable. Hence there is no need to introduce a ‘sporadic capacity’. What we have to show is that coding schemes that work infinitely often can be extended to all block sizes. This is a non-trivial result, since we also show that by merely putting blocks together, and by perhaps not using some of the code bits such an extension is not possible.

### 2.3. Terza variazione: minimum fidelity

The cb-norm is by no means the only way to evaluate the distance between two channels. Another distance measure that has appeared particularly widely (e.g., in [17, 20, 21, 25]) is the minimal overlap between input and corresponding output states: the *minimum fidelity* of a quantum channel  $T : \mathcal{B}_*(\mathcal{H}) \rightarrow \mathcal{B}_*(\mathcal{H})$  is defined as

$$F(T) := \min\{\langle \psi | T(|\psi\rangle\langle \psi|) | \psi \rangle \mid \psi \in \mathcal{H}, \|\psi\| = 1\}. \quad (4)$$

When we want to particularly emphasize the Hilbert space  $\mathcal{H}$  on which the minimization is performed, we will write  $F(\mathcal{H}, T)$  instead.

Of course,  $0 \leq F(\mathcal{H}, T) \leq 1$ , and  $F(\mathcal{H}, T) = 1$  implies that  $T$  acts as the ideal channel on  $\mathcal{H}$ :  $T|_{\mathcal{H}} = \text{id}_{\mathcal{H}}$ . These features make the minimum fidelity a suitable distance measure. We might then call a positive number  $R$  an achievable rate for the channel  $T$  if there is a sequence  $(\mathcal{K}_n)_{n \in \mathbb{N}}$  of Hilbert spaces such that  $\overline{\lim}_{n \rightarrow \infty} (\text{ld dim}(\mathcal{K}_n)/n) = R$  and  $\lim_{n \rightarrow \infty} F(\mathcal{K}_n, D_n T^{\otimes n} E_n) = 1$  for suitable encodings  $(E_n)_{n \in \mathbb{N}}$  and decodings  $(D_n)_{n \in \mathbb{N}}$ .

In section 4.2 we show that the quantum channel capacity arising from this definition is the same.

### 2.4. Quarta variazione: average fidelity

Instead of requiring that the maximum error be small we might be less demanding, and just require an average error to vanish. In the previous section we would then have to replace the minimum fidelity  $F(T)$  by the *average fidelity*,

$$\overline{F}(T) := \int \langle \psi | T(|\psi\rangle\langle \psi|) | \psi \rangle d\psi, \quad (5)$$

where the integral is over the normalized unitarily invariant measure ‘ $d\psi$ ’ on the unit vectors in  $\mathcal{H}$ .

In sections 4.3 and 4.4 we show that this modification has no effect on the quantum channel capacity. An alternative proof is presented in section 6.2.

### 2.5. Quinta variazione: entanglement fidelity

*Entanglement fidelity* was introduced by Ben Schumacher in 1996 [26] and is closely related to minimum fidelity. It characterizes how well the entanglement between the input states and a reference system not undergoing the noise process is preserved: For a quantum channel  $T: \mathcal{B}_*(\mathcal{H}) \rightarrow \mathcal{B}_*(\mathcal{H})$  and a quantum state  $\varrho \in \mathcal{B}_*(\mathcal{H})$ , the *entanglement fidelity of  $\varrho$  with respect to  $T$* ,  $F_e(\varrho, T)$ , is given as

$$F_e(\varrho, T) := \langle \psi | T \otimes \text{id}_{\mathcal{B}_*(\mathcal{H})}(|\psi\rangle\langle\psi|) | \psi \rangle, \quad (6)$$

where  $\psi$  is a purification of  $\varrho$ . This quantity does not depend on the details of the purification process, as is made evident by the alternative expression [26]

$$F_e(\varrho, T) = \sum_i |\text{tr } \varrho t_i|^2, \quad (7)$$

where  $T(\sigma) = \sum_i t_i \sigma t_i^*$  is the Kraus decomposition of  $T$  [4]. Obviously,  $0 \leq F_e(\varrho, T) \leq 1$ . Moreover,  $F_e(\varrho, T) = 1$  implies that  $T$  is noiseless on the support of  $\varrho$ :  $T|_{\text{supp}(\varrho)} = \text{id}_{\text{supp}(\varrho)}$ .

We might then define achievable rates exactly as in section 2.3 above, replacing the condition  $\lim_{n \rightarrow \infty} F(\mathcal{K}_n, D_n T^{\otimes n} E_n) = 1$  by the requirement that

$$\lim_{n \rightarrow \infty} \inf_{\varrho \in \mathcal{B}_*(\mathcal{K}_n)} F_e(\varrho, D_n T^{\otimes n} E_n) = 1 \quad (8)$$

for suitable encodings  $(E_n)_{n \in \mathbb{N}}$  and decodings  $(D_n)_{n \in \mathbb{N}}$ .

The quantum capacity that stems from this definition of achievable rates is likewise equivalent, as shown in section 4.2.

In the previous definition, instead of minimizing over all density operators  $\varrho \in \mathcal{B}_*(\mathcal{H})$ , one can simply choose  $\varrho$  to be the maximally mixed state on  $\mathcal{H}$ ,  $\varrho := (1/d)1_{\mathcal{H}}$ , with the shorthand  $d := \dim \mathcal{H}$ . The resulting variant of entanglement fidelity we call *channel fidelity* [27] of the quantum channel  $T$ ,

$$F_c(T) := F_e\left(\frac{1}{d}1_{\mathcal{H}}, T\right) = \langle \Omega | (T \otimes \text{id}_{\mathcal{H}})(|\Omega\rangle\langle\Omega|) | \Omega \rangle, \quad (9)$$

where  $\Omega = d^{-1/2} \sum_{i=1}^d |i, i\rangle$  is a maximally entangled state on  $\mathcal{H} \otimes \mathcal{H}$ .

Channel fidelity is a very handy figure of merit, since it is a linear functional, does not involve a maximization process, and is completely equivalent to the error criteria discussed above. The details are spelled out in section 4.3.

A further variant arises when in the definition of channel fidelity instead of the maximally entangled state  $\Omega$  an arbitrary input state  $\Gamma \in \mathcal{H} \otimes \mathcal{H}$  is permitted, replacing the channel fidelity  $F_c(T)$  by the quantity  $F_c(\Gamma, T) := \langle \Omega | (T \otimes \text{id}_{\mathcal{H}})(|\Gamma\rangle\langle\Gamma|) | \Omega \rangle$ . This is the error quantity on which Devetak's *entanglement generating capacity* [3] is built on, and is also equivalent (section 4.5).

### 2.6. Sesta variazione: entropy rate

The original definition of quantum channel capacity in terms of entanglement fidelity involves a different concept of computing the rates [16, 18]. According to this definition, the capacity of

a quantum channel is the maximal *entropy rate* of a *quantum source* whose entanglement with the reference system is preserved by the noisy channel. A quantum source  $(\mathcal{K}_n, \varrho_n)_{n \in \mathbb{N}}$  consists of a pair of sequences of Hilbert spaces  $\mathcal{K}_n$  and corresponding density operators  $\varrho_n \in \mathcal{B}_*(\mathcal{K}_n)$ . It is meant to represent a stream of quantum particles produced by some physical process. Its *entropy rate* is defined as

$$R = \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} S(\varrho_n), \quad (10)$$

where  $S(\varrho) = -\text{tr}(\varrho \text{ld } \varrho)$  is the *von Neumann entropy*.

The quantum capacity as defined by Schumacher is then the supremum of all entropy rates for sources such that  $\lim_{n \rightarrow \infty} F_e(\varrho_n, D_n T^{\otimes n} E_n) = 1$  for suitable encodings  $(E_n)_{n \in \mathbb{N}}$  and decodings  $(D_n)_{n \in \mathbb{N}}$ .

It turns out that in order to make this definition equivalent to the others, some mild constraint on the sources is needed. In fact, we will show in section 4.4 that the supremum over *all* sources will be infinite for all channels with positive capacity. However, for a wide range of interesting sources equivalence does hold, namely (cf section 4.4),

- if  $\rho_n = 1_{\mathcal{K}_n} / \dim \mathcal{K}_n$ , which brings us back to the definition based on channel fidelity discussed in the previous section,
- if the source satisfies the so-called asymptotic equipartition property, which has recently been established for general stationary ergodic quantum sources [28]–[30], or
- if the dimension of the ambient space of the encodings grows at most exponentially (even at a rate much larger than the capacity).

### 2.7. Settima variazione: errors vanishing quickly or not at all

In the various definitions of achievable rates presented so far, instead of simply requiring the error quantity to approach zero in the large block limit, one could impose a certain minimum speed of convergence, e.g., linear, polynomial, exponential or super-exponential convergence, as a function of the number of channel invocations. We will show in section 7 that all these definitions coincide, as long as the speed of convergence is at the most exponential.

If we require the errors to vanish even faster or, in the extreme case, that  $\Delta(n_\nu, m_\nu) = 0$  for large enough  $\nu$ , as in the theory of error-correcting codes invented by Knill and Laflamme [31], equivalence no longer holds: if a channel has a small, but non-vanishing probability for depolarization, the same also holds for its tensor powers, and no such channel allows the perfect transmission of even one qubit. Hence the capacity based on exactly vanishing errors will be zero for such channels.

On the other hand, one might sometimes feel inclined to tolerate (small) finite errors in transmission: For some  $\varepsilon > 0$ , let  $Q_\varepsilon(T)$  denote the quantity defined exactly like the quantum channel capacity in definition 2.1, but requiring only  $\Delta(n_\nu, m_\nu) \leq \varepsilon$  for large  $\nu$  instead of  $\lim_{\nu \rightarrow \infty} \Delta(n_\nu, m_\nu) = 0$ . Obviously,  $Q_\varepsilon(T) \geq Q(T)$  for any quantum channel  $T$ . We even have  $\lim_{\varepsilon \rightarrow 0} Q_\varepsilon(T) = Q(T)$  (see section 7.3).

In the purely classical setting even more is known: if  $\varepsilon > 0$  is small enough, one cannot achieve bigger rates by allowing small errors, i.e.,  $C_\varepsilon(T) = C(T)$ . This is the so-called *strong converse* to Shannon's coding theorem. It is still unknown whether an analogous property holds for quantum channels.

### 2.8. *Ottava variazione: isometric encodings and homomorphic decodings*

Definition 2.1 of channel capacity involves an optimization over the set of all encoding and decoding maps. This set is very large, and it may thus seem favourable to restrict both encoding and decoding to smaller classes.

In [17] it has been shown that we may restrain our attention to *isometric* encodings, i.e., encodings of the form

$$E(\rho) = V\rho V^* \quad (11)$$

with isometric  $V$ , and still be left with the same capacity (see section 5 for details). Physically, this means that encoding can always be thought of as a unitary process augmented by an initial projection onto a subspace small enough to fit into the channel.

In the Knill and Laflamme [31] setting of perfect error correction, not only are encoding maps isometric, but in addition the decodings can be chosen to be of the (Heisenberg picture) form

$$D_*(X) = V(X \otimes 1)V^* + \text{tr}(\rho_0 X)(1 - VV^*) \quad (12)$$

with isometric  $V$  and an arbitrary reference state  $\rho_0$ . We call maps of this type *homomorphic*, because the first term is an algebraic homomorphism, and the second term only serves to render the whole channel unital.

Since the sufficiency of isometric encoding transfers from the perfect error correction setting to asymptotically perfect error correction, it may seem reasonable to conjecture that a similar result holds for homomorphic decodings. However, up to now no such result is known.

### 2.9. *Nona variazione: coding with a little help from a classical friend*

Here we consider a setup in which a quantum channel  $T$  is assisted by additional classical forward communication between the sender (Alice) and the receiver (Bob). Clearly, this allows Alice and Bob to collaborate in a more coordinated fashion: Alice may use the additional resource to transfer information about the encoding process, which Bob on his part may try to take advantage of in his choice of the decoding channel.

However, it is a straightforward consequence of the isometric encoding theorem that these new possibilities do not help to increase the channel capacity, even if the classical side channel is noiseless: we have  $Q(T \otimes \text{id}_c) = Q(T)$  [17, 25], where by  $\text{id}_c$  we denote an ideal channel of arbitrarily large dimensionality. That this is not a trivial statement is seen from the observation that classical feedback between successive channels uses may increase the capacity [18, 20, 25, 32].

The uselessness of classical forward communication may be extended to cover the so-called *separable* side channels, i.e., quantum channels with intermediate measurement and re-preparation processes. The details on both classes of side channels are spelled out in section 6.

### 2.10. *Coda: the coding theorem*

Computing channel capacities on the basis of the definitions given, even the simplified ones, is a tricky business. It involves optimization in systems of asymptotically many tensor factors. It

has therefore been a long-time challenge to find a quantum analogue of Shannon's noisy coding theorem [6], which would allow to compute the channel capacity as an optimization over a low dimensional space.

According to Shannon's famous theorem, the classical capacity is obtained by finding the supremum of the so-called *mutual information*, which itself is given in terms of the Shannon entropy. A quantum analogue of mutual information, *coherent information*, has been identified early. For the quantum channel  $T: \mathcal{B}_*(\mathcal{H}_1) \rightarrow \mathcal{B}_*(\mathcal{H}_2)$  and the density operator  $\varrho \in \mathcal{B}_*(\mathcal{H}_1)$ , it is defined as

$$I_c(\varrho, T) := S(T(\varrho)) - S(T \otimes \text{id}(|\psi\rangle\langle\psi|)), \quad (13)$$

where  $\psi \in \mathcal{H}_1 \otimes \mathcal{H}_1$  is a purification of the density operator  $\varrho$ , and  $S$ , as before, is the *von Neumann entropy*.

The regularized coherent information has long been known to be an upper bound on the quantum channel capacity [16]–[18], i.e.,

$$Q(T) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \max_{\varrho} I_c(\varrho, T^{\otimes n}). \quad (14)$$

Unlike the classical or quantum mutual information, coherent information is not additive; hence taking the limit  $n \rightarrow \infty$  in equation (14) is indeed required [21].

The first sketch of an argument to close the gap in equation (14) was given by Lloyd [33]. At a recent conference, Shor [1, 2] presented a coding scheme based on random coding to attain coherent information. His results have not been published yet. Shortly thereafter, Devetak [3] released another coding scheme based on a key generation protocol made 'coherent' [34, 35]. By the same techniques, Devetak and Winter [36, 37] very recently were able to prove the long-conjectured *hashing inequality* [25], which states that the regularized coherent information is an achievable entanglement distillation rate, and implies the channel capacity result by teleportation [22].

These achievements certainly mark a major step in the direction of a coding theorem, but do not satisfy all the properties desired of such a theorem. In particular, they still demand the solution of asymptotically large variational problems.

### 3. Elementary properties of channel capacity

#### 3.1. Basic inequalities

Before we enter the proof sections we need to review some basic properties of channel capacity, which will turn out to be helpful as we proceed, but are also interesting in their own right. All proofs are easy, and may be found in [14], albeit for noiseless reference channels only. The generalization is straightforward.

Running two channels,  $T_1$  and  $T_2$ , in succession, the capacity of the composite channel,  $T_1 \circ T_2$ , cannot be bigger than the capacity of the channel with the smallest bandwidth. This is known as the *bottleneck inequality*:

$$Q(T_1 \circ T_2, S) \leq \min\{Q(T_1, S), Q(T_2, S)\}. \quad (15)$$

Instead of running  $T_1$  and  $T_2$  in succession, we may also run them in parallel, which is represented mathematically by the tensor product  $T_1 \otimes T_2$ . In this case the capacity can be shown to be *super-additive*,

$$Q(T_1 \otimes T_2, S) \geq Q(T_1, S) + Q(T_2, S). \quad (16)$$

For the standard ideal channels we even have additivity. The same holds true if both  $S$  and one of the channels  $T_1, T_2$  are noiseless, the third channel being arbitrary. However, to decide whether additivity holds generally is one of the big open problems in the field.

Finally, the *two-step-coding inequality* tells us that by using an intermediate channel in the coding process we cannot increase the transmission rate:

$$Q(T_1, T_3) \geq Q(T_1, T_2)Q(T_2, T_3). \quad (17)$$

### 3.2. Quantum capacity of noiseless channels

There are special cases in which the quantum channel capacity can be evaluated relatively easily, the most relevant one being the noiseless channel  $\text{id}_n$ , where by the subscript  $n$  we denote the dimension of the underlying Hilbert space. In this case we have

$$Q(\text{id}_n, \text{id}_m) = \frac{\text{ld } n}{\text{ld } m}. \quad (18)$$

A proof follows below. Combining this with the two-step coding inequality (17), we see that for any quantum channel  $T$

$$Q(T, \text{id}_n) = \frac{\text{ld } m}{\text{ld } n} Q(T, \text{id}_m), \quad (19)$$

which shows that quantum channel capacities relative to noiseless quantum channels of different dimensionality only differ by a constant factor. Fixing the dimensionality of the reference channel then only corresponds to a choice of units. Conventionally the ideal qubit channel is chosen as a standard of reference, fixing the unit *bit*.

**Proof of equation 18.** This is an immediate consequence of estimates of the simulation error  $\Delta(\text{id}_n, \text{id}_m) = \inf_{D, E} \|D \text{id}_n E - \text{id}_m\|_{cb}$  between ideal channels. We have

$$\Delta(\text{id}_n, \text{id}_m) = 0, \quad \text{if } m \leq n; \quad (20)$$

$$\Delta(\text{id}_n, \text{id}_m) \geq 1 - \frac{n}{m}, \quad \text{if } m \geq n. \quad (21)$$

The first relation is shown by explicitly constructing  $E: \mathcal{B}_*(\mathbb{C}^m) \rightarrow \mathcal{B}_*(\mathbb{C}^n)$  and  $D: \mathcal{B}_*(\mathbb{C}^n) \rightarrow \mathcal{B}_*(\mathbb{C}^m)$  such that  $D \text{id}_n E = \text{id}_m$ . To this end we may consider  $\mathbb{C}^m \subset \mathbb{C}^n$  as a subspace with projection  $P_m$ . Then  $E$  is defined by extending each  $n \times n$  matrix by zeros for the additional  $(m - n)$  dimensions, and

$$D: \varrho \mapsto P_m \varrho P_m + \frac{\text{tr}(1_n - P_m) \varrho}{m} P_m, \quad (22)$$

where the second term serves to make  $D$  trace-preserving. Then, clearly,  $DE = \text{id}_m$  as claimed.

To prove the inequality (21), choose a maximal family of one-dimensional orthogonal projections  $\{P_\nu\}_{\nu=1,\dots,m} \subset \mathcal{B}_*(\mathbb{C}^m)$  such that  $\sum_{\nu=1}^m P_\nu = 1_m$ . Then for any decoding  $D: \mathcal{B}_*(\mathbb{C}^n) \rightarrow \mathcal{B}_*(\mathbb{C}^m)$ , the relation

$$\mathrm{tr} \varrho F_\nu := \mathrm{tr} D(\varrho) P_\nu \quad \forall \varrho \in \mathcal{B}_*(\mathbb{C}^n) \quad (23)$$

defines a set  $\{F_\nu\}_{\nu=1,\dots,m} \subset \mathcal{B}_*(\mathbb{C}^n)$  of positive operators satisfying  $\sum_{\nu=1}^m F_\nu = 1_n$ . For any encoding  $E: \mathcal{B}_*(\mathbb{C}^m) \rightarrow \mathcal{B}_*(\mathbb{C}^n)$  we thus have

$$\begin{aligned} n = \mathrm{tr} 1_n &= \mathrm{tr} \sum_{\nu=1}^m F_\nu \geq \sum_{\nu=1}^m \mathrm{tr} E(P_\nu) F_\nu = \sum_{\nu=1}^m \mathrm{tr} D(E(P_\nu)) P_\nu \\ &\geq \sum_{\nu=1}^m \mathrm{tr} P_\nu - \sum_{\nu=1}^m |\mathrm{tr}(P_\nu D(E(P_\nu)) - P_\nu)| \\ &\geq m - \sum_{\nu=1}^m \|D(E(P_\nu)) - P_\nu\|_\infty \\ &\geq m(1 - \|DE - \mathrm{id}_m\|_{cb}), \end{aligned} \quad (24)$$

where in the fourth step we have used equation (23). Equation (24) then immediately implies equation (21).

We now have to convert these estimates equations (20) and (21) into statements for achievable rates for  $S = \mathrm{id}_n$  and  $T = \mathrm{id}_m$ . Thus equations (20) and (21) apply with  $n$  replaced by  $n^\nu$  and  $m$  replaced by  $m^{\mu}$ . So let  $(n_\nu)_{\nu \in \mathbb{N}}$  and  $(m_\nu)_{\nu \in \mathbb{N}}$  be two integer sequences such that  $\lim_{\nu \rightarrow \infty} n_\nu = \infty$  and  $\lim_{\nu \rightarrow \infty} (m_\nu/n_\nu) < (\mathrm{ld} n)/(\mathrm{ld} m)$ . Then for all sufficiently large  $\nu$  we have  $n^\nu \geq m^{\mu}$ , and therefore  $\Delta(n_\nu, m_\nu) = 0$ , which implies that any  $R < (\mathrm{ld} n)/(\mathrm{ld} m)$  is achievable.

On the other hand, let  $R = ((\mathrm{ld} n)/(\mathrm{ld} m)) + \varepsilon$ , for some  $\varepsilon > 0$ , and choose diverging sequences such that  $\lim_{\nu \rightarrow \infty} (m_\nu/n_\nu) = R$ . Then  $n^\nu/m^{\mu} \leq m^{-\varepsilon n_\nu}$  infinitely often, and thus, by equation (21),  $\Delta(n_\nu, m_\nu)$  is close to 1 infinitely often. Hence the errors do not go to zero, and the rate  $R$  is not achievable. To summarize,  $Q(\mathrm{id}_n, \mathrm{id}_m) = (\mathrm{ld} n)/(\mathrm{ld} m)$  is the supremum of all achievable rates.  $\square$

By the same techniques, one may also show that the capacity of the ideal channel does not increase if the information to be transmitted is restricted to be classical.

### 3.3. Partial transposition bound

The upper bound on the capacity of ideal channels can also be obtained from a general upper bound on quantum capacities, which has the virtue of being easily calculated in many situations. It involves, on each system considered, the *transposition map*, which we denote by  $\Theta$ , defined as matrix transposition with respect to some fixed orthonormal basis. None of the quantities we consider will depend on this basis. As is well known, transposition is positive but not completely positive. Similarly, we have  $\|\Theta\|_\infty = 1$ , but generally  $\|\Theta\|_{cb} > 1$ . More precisely,  $\|\Theta\|_{cb} = d$ , when the system is described on a  $d$ -dimensional Hilbert space [5]. We claim that, for any channel  $T$  and small  $\varepsilon > 0$ ,

$$Q_\varepsilon(T) \leq \mathrm{ld} \|T\Theta\|_{cb} =: Q_\Theta(T), \quad (25)$$

where  $Q_\varepsilon (\geq Q(T))$  is the finite error capacity introduced in section 2.7. In particular, for the ideal channel this implies  $Q(\text{id}_d) \leq \text{ld}(d)$ .

The proof of equation (25) is quite simple [12]: suppose  $R$  is an achievable rate, and that  $m_\nu/n_\nu \rightarrow R \leq Q_\varepsilon(T)$ , and encoding  $E_\nu$  and decoding  $D_\nu$  are such that  $\Delta(n_\nu, m_\nu) = \|D_\nu T^{\otimes n_\nu} E_\nu - \text{id}_2^{\otimes m_\nu}\|_{cb} \rightarrow 0$ . Then we have

$$\begin{aligned} 2^{m_\nu} &= \|\text{id}_2^{\otimes m_\nu} \Theta\|_{cb} \leq \|(\text{id}_2^{\otimes m_\nu} - D_\nu T^{\otimes n_\nu} E_\nu) \Theta\|_{cb} + \|D_\nu T^{\otimes n_\nu} E_\nu \Theta\|_{cb} \\ &\leq \|\Theta_{2^{m_\nu}}\|_{cb} \|\text{id}_2^{\otimes m_\nu} - D_\nu T^{\otimes n_\nu} E_\nu\|_{cb} + \|D_\nu (T\Theta)^{\otimes n_\nu} \Theta E_\nu\|_{cb} \\ &\leq 2^{m_\nu} \Delta(n_\nu, m_\nu) + \|T\Theta\|_{cb}^{n_\nu}, \end{aligned} \quad (26)$$

where in the last step we have used that  $D_\nu$  and  $\Theta E_\nu \Theta$  are channels with  $\text{cb-norm} = 1$ , and that the  $\text{cb-norm}$  is exactly tensor multiplicative, so  $\|X^{\otimes n}\|_{cb} = \|X\|_{cb}^n$ . Hence, by taking the binary logarithm and dividing by  $n_\nu$ , we get

$$\frac{m_\nu}{n_\nu} + \frac{\text{ld}(1 - \Delta(n_\nu, m_\nu))}{n_\nu} \leq \text{ld}\|T\Theta\|_{cb}. \quad (27)$$

Then in the limit  $\nu \rightarrow \infty$  we find  $R \leq Q_\Theta(T)$  for any achievable rate  $R$ .

The upper bound  $Q_\Theta(T)$  computed in this way has some remarkable properties, which make it a capacity-like quantity in its own right. For example, it is exactly additive:

$$Q_\Theta(S \otimes T) = Q_\Theta(S) + Q_\Theta(T), \quad (28)$$

for any pair  $S, T$  of channels, and satisfies the bottleneck inequality  $Q_\Theta(ST) \leq \min\{Q_\Theta(S), Q_\Theta(T)\}$ . Moreover, it coincides with the quantum capacity on ideal channels:  $Q_\Theta(\text{id}_n) = Q(\text{id}_n) = \text{ld } n$ , and it vanishes whenever  $T\Theta$  is completely positive. In particular, if  $\text{id} \otimes T$  maps any entangled state to a state with positive partial transpose, we have  $Q_\Theta(T) = 0$ .

## 4. Alternative error criteria

In this section we will show that the various distance measures introduced in sections 2.3–2.6 are equivalent, as long as the reference channel is chosen to be noiseless, i.e.,  $S = \text{id}_d$  for some  $d < \infty$ . Hence the remarkable insensitivity of the quantum capacity to the choice of the error criterion holds only for the capacity  $Q(T)$  which is our main concern, but not for the more general  $Q(T, S)$  comparing two arbitrary channels. The reason for this difference is the analogous observation for distance measures on the state space: different ways of quantifying the distance between states become inequivalent in the limit of large dimensions, but all measures for the distance between a *pure* state and a general state essentially agree.

### 4.1. Preliminaries

The following lemma will serve as a starting point for showing the equivalence of fidelity, ordinary operator norm, and  $\text{cb-norm}$  criteria. By  $\|A\|_1 := \text{tr}\sqrt{(A^*A)}$  we will denote the *trace norm* of the operator  $A \in \mathcal{B}(\mathbb{C}^d)$ , by  $\|A\|_2 := \sqrt{\text{tr}(A^*A)}$  its *Hilbert–Schmidt norm*, and by  $\|A\|_\infty$  the ordinary operator norm. These norms are related by the following chain of inequalities:

$$\|A\|_\infty \leq \|A\|_2 \leq \|A\|_1 \quad (29)$$

(see chapter VI of [38] for a thorough discussion of these *Schatten classes*, these and other useful properties, and the relation to  $\mathcal{L}^p$ -spaces). Of course, all norms in a finite-dimensional space are equivalent, so there must also be a bound in the reverse direction. This is

$$\|A\|_1 \leq d\|A\|_\infty. \quad (30)$$

The crucial difference between these estimates is that the bound in equation (30) explicitly depends on the Hilbert space dimension  $d$ , which makes this inequality useless for applications in capacity theory, where dimensions grow exponentially. Our aim in this section is therefore to relate the various error measures with dimension-independent bounds (see proposition 4.3 below).

**Lemma 4.1.** *Let  $\varrho$  be a density operator and  $\psi$  be a unit vector in a Hilbert space  $\mathcal{H}$ . Then*

$$\|\varrho - |\psi\rangle\langle\psi|\|_1 \leq 2\sqrt{1 - \langle\psi|\varrho|\psi\rangle}, \quad (31)$$

with equality in equation (31) iff  $\varrho$  is pure or  $\psi$  is orthogonal to the support of  $\varrho$ .

**Proof.** Suppose first that  $\varrho = |\varphi\rangle\langle\varphi|$  is pure. Then we can compute the trace norm in the two-dimensional space spanned by  $\psi$  and  $\varphi$ . For the moment we will only use this property, i.e., we assume that  $\psi = (1, 0)$  is the first basis vector, and  $\varrho$  is an arbitrary  $(2 \times 2)$  density matrix. Then we may expand the traceless operator  $\varrho - |\psi\rangle\langle\psi|$  in terms of the Pauli matrices  $\{\sigma_i\}_{i=1,2,3}$ , as follows:

$$\varrho - |\psi\rangle\langle\psi| = (\varrho_{11} - 1)\sigma_3 + \text{Re}(\varrho_{12})\sigma_1 - \text{Im}(\varrho_{12})\sigma_2. \quad (32)$$

From this we find the eigenvalues of  $\varrho - |\psi\rangle\langle\psi|$  to equal  $\pm\sqrt{(\varrho_{11} - 1)^2 + |\varrho_{12}|^2}$ , and hence

$$\|\varrho - |\psi\rangle\langle\psi|\|_1 = 2\sqrt{(\varrho_{11} - 1)^2 + |\varrho_{12}|^2}. \quad (33)$$

Positivity of  $\varrho$  clearly requires  $\det \varrho \geq 0$ , implying  $|\varrho_{12}|^2 \leq \varrho_{11}\varrho_{22} = \varrho_{11}(1 - \varrho_{11})$ . Since  $\text{tr}(\varrho^2) = 1 + 2(\varrho_{11}(\varrho_{11} - 1) + |\varrho_{12}|^2)$ , equality holds if  $\varrho$  is pure. By inserting  $|\varrho_{12}|^2 = \varrho_{11}(1 - \varrho_{11})$  into equation (33) we see that for pure states we indeed have equality in equation (31).

We now drop the assumption that  $\varrho$  is pure and consider an arbitrary convex decomposition  $\varrho = \sum_i \lambda_i \varrho_i$  into pure states  $\varrho_i$ . Then because  $x \mapsto \sqrt{1 - x}$  is concave we obtain

$$\begin{aligned} \|\varrho - |\psi\rangle\langle\psi|\|_1 &\leq \sum_i \lambda_i \|\varrho_i - |\psi\rangle\langle\psi|\|_1 = 2 \sum_i \lambda_i \sqrt{1 - \langle\psi|\varrho_i|\psi\rangle} \\ &\leq 2 \sqrt{1 - \sum_i \lambda_i \langle\psi|\varrho_i|\psi\rangle} = 2\sqrt{1 - \langle\psi|\varrho|\psi\rangle}, \end{aligned} \quad (34)$$

where in the second step the result for pure states has been used. This establishes equation (31).

Now suppose that equality holds in equation (34). Then because the concavity of  $x \mapsto \sqrt{1 - x}$  is strict,  $\langle\psi|\varrho_i|\psi\rangle = \langle\psi|\varrho|\psi\rangle \forall i$ . But since the convex decomposition of  $\varrho$  is arbitrary, we may conclude that

$$|\langle\varphi|\psi\rangle|^2 = \langle\psi|\varrho|\psi\rangle \|\varphi\|^2 \quad (35)$$

for any vector  $\varphi$  in the support of  $\varrho$ . By polarization this implies  $\langle \varphi_1 | \psi \rangle \langle \psi | \varphi_2 \rangle = \langle \varphi_1 | \varphi_2 \rangle \langle \psi | \varrho | \psi \rangle$ , from which it follows that

$$S|\psi\rangle\langle\psi|S = \langle\psi|\varrho|\psi\rangle S, \quad (36)$$

where  $S$  denotes the projection operator on  $\text{supp}(\varrho)$ . Hence either the factor  $\langle\psi|\varrho|\psi\rangle$  vanishes, entailing  $S\psi = 0$ , or else  $S$  is a rank one operator, and thus  $\varrho$  is pure. This concludes the proof.  $\square$

From lemma 4.1 we may derive a fidelity-based expression for the deviation of a given channel from the ideal channel:

**Lemma 4.2.** *Let  $\mathcal{H}$  be a Hilbert space,  $\dim \mathcal{H} < \infty$ , and  $T : \mathcal{B}_*(\mathcal{H}) \rightarrow \mathcal{B}_*(\mathcal{H})$  be a channel. We then have*

$$\|T - \text{id}\|_\infty \leq 4 \sup_{\|\psi\|=1} \{\sqrt{1 - \langle\psi|T(|\psi\rangle\langle\psi|)|\psi\rangle}\}. \quad (37)$$

**Proof.** Note that the operator norm  $\|T - \text{id}\|_\infty$  equals the norm of the adjoint operator on the dual space, i.e.,

$$\|T - \text{id}\|_\infty = \sup_{\|\varrho\|_1 \leq 1} \|T_*\varrho - \varrho\|_1 \quad (38)$$

(cf chapter VI of [38] or section 2.4 of [39] for details). Any matrix  $\varrho$ , with  $\|\varrho\|_1 \leq 1$ , has a decomposition  $\varrho = \varrho_1 + i\varrho_2$  into Hermitian  $\varrho_i$  satisfying  $\|\varrho_i\|_1 \leq 1$ . Inserting this decomposition into equation (38) and using the triangle inequality, we find  $\|T - \text{id}\|_\infty \leq 2 \sup \|T_*\varrho - \varrho\|_1$ , where the supremum is now over all Hermitian matrices  $\varrho$  obeying  $\|\varrho\|_1 \leq 1$ .

By spectral decomposition, any Hermitian matrix  $\varrho$  can be given the form  $\varrho = \sum_i r_i \varrho_i$ , where the  $\varrho_i$  are rank one projectors and the coefficients  $r_i$  are real numbers satisfying  $\sum_i |r_i| = \|\varrho\|_1$ . Inserting this into the supremum, we see that  $\|T - \text{id}\|_\infty \leq 2 \sup \|T_*\varrho - \varrho\|_1$ , where optimization is now with respect to all one-dimensional projectors  $|\psi\rangle\langle\psi|$ . The inequality then directly follows from lemma 4.1.  $\square$

#### 4.2. Four equivalent distance measures

We now have in hand all the tools we need to prove that the distance measures presented in sections 2.3–2.6 do indeed coincide:

**Proposition 4.3.** *Let  $\mathcal{H}$  be a Hilbert space,  $\dim \mathcal{H} < \infty$ , and let  $T : \mathcal{B}_*(\mathcal{H}) \rightarrow \mathcal{B}_*(\mathcal{H})$  be a channel. Then*

$$\begin{aligned} 1 - \inf_{\varrho \in \mathcal{B}_*(\mathcal{H})} F_e(\varrho, T) &\leq 4\sqrt{1 - F(T)} \leq 4\sqrt{\|T - \text{id}\|_\infty} \leq 4\sqrt{\|T - \text{id}\|_{cb}} \\ &\leq 8 \left( 1 - \inf_{\varrho \in \mathcal{B}_*(\mathcal{H})} F_e(\varrho, T) \right)^{\frac{1}{4}}. \end{aligned} \quad (39)$$

These are the dimension-independent bounds we need: if a sequence of channels becomes close to ideal in the sense of any of the error measures appearing in this proposition, so it will be in terms of all the others. The equivalence of the basic capacity definition 2.1 based on the cb-norm and the definitions based on minimum fidelity and entanglement fidelity, as presented in sections 2.3 and 2.5, then directly follows.

It is crucial for proposition 4.3 that we are considering only the deviation of  $T$  from the ideal channel, so we can use lemma 4.1 for the distance between an output state and a pure state. Therefore, for the general capacity  $Q(T, S)$  the choice of the error quantity may remain important. General properties such as superadditivity (16), which are easy to see for the cb-norm criterion, might therefore fail for the simpler-looking operator norm  $\|T - S\|_\infty$ . This is the principal reason for choosing the cb-norm in the basic definition.

Mean fidelity, as used in section 2.4, and channel fidelity, as introduced in section 2.5, are conspicuously absent from proposition 4.3. Their role will be discussed in section 4.3.

The equivalence of Schumacher's original definition of channel capacity in terms of the entropy rate will then be treated in section 4.4.

**Proof of proposition 4.3.** Let  $\phi \in \mathcal{H} \otimes \mathcal{H}$  be a purification of  $\varrho \in \mathcal{B}_*(\mathcal{H})$ . We then have

$$1 - F_e(\varrho, T) = \langle \phi | (\text{id} - T) \otimes \text{id}(|\phi\rangle\langle\phi|) | \phi \rangle. \quad (40)$$

By Schmidt decomposition,  $\phi$  can be given a representation  $|\phi\rangle = \sum_j \lambda_j |j\rangle \otimes |j'\rangle$ , where  $\{|j\rangle\}_j$  and  $\{|j'\rangle\}_j$  are orthonormal systems in  $\mathcal{H}$ , and the so-called Schmidt coefficients  $\{\lambda_j\}_j$  are non-negative real numbers satisfying  $\sum_j \lambda_j^2 = 1$ . Inserting this representation into equation (40), we see that

$$\begin{aligned} 1 - F_e(\varrho, T) &= \sum_{j,k} \lambda_j^2 \lambda_k^2 \langle j | (\text{id} - T) (|j\rangle\langle k|) | k \rangle \\ &\leq \sum_{j,k} \lambda_j^2 \lambda_k^2 \|\text{id} - T\|_\infty \| |j\rangle\langle k| \|_\infty = \|\text{id} - T\|_\infty, \end{aligned} \quad (41)$$

where in the last step the normalization  $\sum_j \lambda_j^2 = 1$  has been applied.

The first inequality then immediately follows from lemma 4.2 and the definition of minimum fidelity, equation (4).

An application of the Schwarz inequality directly gives the second inequality:

$$\begin{aligned} 1 - \langle \psi | T(|\psi\rangle\langle\psi|) | \psi \rangle &= \langle \psi | (\text{id} - T) (|\psi\rangle\langle\psi|) | \psi \rangle \\ &\leq \|\text{id} - T\|_\infty \| |\psi\rangle\langle\psi| \|_\infty = \|T - \text{id}\|_\infty \end{aligned} \quad (42)$$

for all unit vectors  $\psi \in \mathcal{H}$ .

The third inequality is obvious from the definition of cb-norm, so we only need to prove the last step. Applying lemma 4.2 to the operator  $T \otimes \text{id}_n$  and then taking the supremum over  $n$  on both sides, we see that

$$\|T - \text{id}\|_{cb} \leq 4 \sqrt{1 - \inf_{n \in \mathbb{N}} F(T \otimes \text{id}_n)} = 4 \sqrt{1 - \inf_{\varrho \in \mathcal{B}_*(\mathcal{H})} F_e(\varrho, T)}, \quad (43)$$

concluding the proof.  $\square$

### 4.3. Average fidelity and channel fidelity

Average fidelity and channel fidelity have been shown [40, 41] to be directly related error criteria.

**Proposition 4.4.** Let  $\overline{F}(T)$  be the average fidelity and  $F_c(T)$  be the channel fidelity of a quantum channel  $T : \mathcal{B}_*(\mathcal{H}) \rightarrow \mathcal{B}_*(\mathcal{H})$ , as introduced in equations (5) and (9), respectively. We then have

$$\overline{F}(T) = \frac{dF_c(T) + 1}{d + 1}, \quad (44)$$

where  $d$  is the dimension of the underlying Hilbert space  $\mathcal{H}$ .

From proposition 4.4 we may conclude that both quantities coincide in the large-dimension limit  $d \rightarrow \infty$ . Consequently, average fidelity and channel fidelity are equivalent error criteria for capacity purposes.

However, neither appears in proposition 4.3. After giving a somewhat simplified proof of equation (44), we show by an explicit counterexample that this omission is not accidental. Since a coding for which the worst case fidelity goes to 1 also makes the average fidelity go to 1, the capacity defined with average fidelity might in principle be larger than the standard one. That these capacities nevertheless coincide will then follow from proposition 4.5 in section 4.4. A more direct proof for the equivalence of average fidelity, i.e., a proof not making use of proposition 4.4, is then presented in section 6.2.

**Proof of proposition 4.4.** Suppose that  $\{t_i\}_i$  is a set of Kraus operators for the quantum channel  $T$ , i.e.,  $T(\sigma) = \sum_i t_i \sigma t_i^* \forall \sigma \in \mathcal{B}_*(\mathcal{H})$ .

In the course of the proof we will repeatedly employ the so-called *flip* operator  $\mathbb{F} \in \mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{H})$ , defined by  $\mathbb{F}(\varphi \otimes \psi) := \psi \otimes \varphi$ . In a basis  $\{|n\rangle\}_{n=1, \dots, d}$  of  $\mathcal{H}$ ,  $d := \dim \mathcal{H}$ , this corresponds to the representation

$$\mathbb{F} = \sum_{n,m=1}^d |n, m\rangle\langle m, n|. \quad (45)$$

Working in this representation one easily verifies that for all operators  $A, B \in \mathcal{B}(\mathcal{H})$

$$\text{tr } \mathbb{F}(A \otimes B) = \text{tr } AB. \quad (46)$$

In terms of the Kraus operators  $\{t_i\}_i$  the average fidelity of  $T$  then reads

$$\begin{aligned} \overline{F}(T) &= \int \langle U\psi | T(|U\psi\rangle\langle U\psi|) |U\psi\rangle dU \\ &= \sum_i \text{tr} \int t_i^* U \varrho U^* t_i U \varrho U^* dU \\ &= \sum_i \text{tr} \mathbb{F}(t_i^* \otimes t_i) \int (U \otimes U)(\varrho \otimes \varrho)(U \otimes U)^* dU, \end{aligned} \quad (47)$$

where  $\varrho := |\psi\rangle\langle\psi|$  is an arbitrary pure reference state, integration is over all unitaries  $U \in \mathcal{B}(\mathcal{H})$  and in the last step we have applied equation (46). The second factor under the trace,

$$P(\varrho) := \int (U \otimes U)(\varrho \otimes \varrho)(U \otimes U)^* dU, \quad (48)$$

is obviously invariant under local unitary transformations, i.e.,  $[P(\varrho) | V \otimes V] = 0$  for all unitary operators  $V \in \mathcal{B}(\mathcal{H})$ . Such a state is usually called a *Werner state* [42], and it follows from the

theory of group representations that these states are spanned by the identity operator and the flip operator,  $P(\varrho) = \alpha 1 + \beta \mathbb{F}$  with complex coefficients  $\alpha, \beta$  (see [43] and chapter 3.1.2 of [14] for details). The coefficients can be easily obtained by tracing  $P(\varrho)$  with the identity and flip operator, respectively, and are both found to equal  $1/(d(d+1))$ . Inserting the expansion

$$P(\varrho) = \frac{1}{d(d+1)}(1 + \mathbb{F}) \quad (49)$$

into equation (47) and using equation (46) again, we see that

$$\begin{aligned} \bar{F}(T) &= \frac{1}{d(d+1)} \sum_i \text{tr} \mathbb{F}(t_i^* \otimes t_i)(1 + \mathbb{F}) \\ &= \frac{1}{d(d+1)} \left( \text{tr} \sum_i t_i^* t_i + \sum_i \text{tr} t_i^* \otimes t_i \right) \\ &= \frac{1}{d(d+1)} \left( d + \sum_i |\text{tr} t_i|^2 \right) = \frac{1}{d(d+1)} (d + d^2 F_c(T)) \\ &= \frac{1}{d+1} (1 + d F_c(T)), \end{aligned} \quad (50)$$

where in the second step we have used the normalization  $\sum_i t_i^* t_i = 1$ , and in the third step equation (7) has been applied for the state  $\varrho = 1/d$ .  $\square$

We proceed with the advertised

**Counterexample.** For  $\varrho \in \mathcal{B}_*(\mathbb{C}^d)$ , we set

$$T(\varrho) = P_+ \varrho P_+ + P_- \varrho P_-, \quad (51)$$

where  $P_+ := |\psi_1\rangle\langle\psi_1|$  is some one-dimensional projector and  $P_- := \text{id} - P_+$  its orthocomplement. Then by equation (7) we find

$$F_c(T) = \frac{1}{d^2} \sum_{i=\pm} |\text{tr} P_i|^2 = \frac{d^2 - 2d + 2}{d^2}, \quad (52)$$

and therefore  $\lim_{d \rightarrow \infty} \bar{F}(T) = \lim_{d \rightarrow \infty} F_c(T) = 1$ , the first equality by equation (44). However, using equation (38) we have

$$\|T - \text{id}\|_\infty = \sup_{\|\varrho\|_1 \leq 1} \|T_* \varrho - \varrho\|_1 \geq \|T_* \tilde{\varrho} - \tilde{\varrho}\|_1 \quad \forall \tilde{\varrho} \in \mathcal{B}_*(\mathbb{C}^d), \quad (53)$$

and by choosing  $\tilde{\varrho} = \frac{1}{2} |\psi_1 + \psi_2\rangle\langle\psi_1 + \psi_2|$  such that  $\psi_2 \perp \psi_1$ ,  $\|T - \text{id}\|_\infty$  can be easily shown to be non-zero, and independent of  $d$ . Hence there exists no bound of the form  $\|T - \text{id}\|_\infty \leq f(F_c(T))$  with a dimension independent function  $f$ , such that  $x \rightarrow 1$  implies  $f(x) \rightarrow 0$ .

#### 4.4. Entanglement fidelity and entropy rate

Let us briefly summarize what we have learned so far about the interrelation of the various distance measures introduced in section 2: from proposition 4.3 and the results of the previous section

we may infer the existence of two classes of equivalent error criteria, one of them containing average and channel fidelity, the other one cb-norm distance, operator norm, minimum fidelity and entanglement fidelity.

To show that both classes lead to the same quantum channel capacity, we will have to construct, from a given coding scheme with rate  $R$  and channel fidelity approaching one, a sequence of Hilbert spaces  $(\mathcal{K}_n)_{n \in \mathbb{N}}$  all pure states of which may be sent reliably with rate  $R$ . This is the essence of the following proposition, which closely follows the argument presented in section V of [17]. Although for this purpose we only need to consider channel fidelities, and thus the chaotic density operator, the statement is kept more general to apply to all density matrices, since this will immediately allow us to cope with Schumacher's definition of channel capacity in terms of entropy rates as well.

**Proposition 4.5.** *Let  $\mathcal{H}$  be a Hilbert space with  $d := \dim \mathcal{H} < \infty$ . Let  $T : \mathcal{B}_*(\mathcal{H}) \rightarrow \mathcal{B}_*(\mathcal{H})$  be a channel, and  $\varrho \in \mathcal{B}_*(\mathcal{H})$  a density operator. Then, for a suitable  $k$ -dimensional projection  $P_k \in \mathcal{B}(\mathcal{H})$ , and for the 'compressed channel'  $T_k : \mathcal{B}_*(P_k \mathcal{H}) \rightarrow \mathcal{B}_*(P_k \mathcal{H})$  given by*

$$T_k(\sigma) := P_k T(\sigma) P_k + \text{tr}((1 - P_k)T(\sigma)) \frac{1}{k} P_k, \quad (54)$$

the estimate

$$F(P_k \mathcal{H}, T_k) \geq 1 - \frac{1 - F_e(\varrho, T)}{1 - q^*}, \quad (55)$$

holds with both

$$q^* = k \|\rho\|_\infty \quad (56)$$

and

$$q^* = \frac{1 + \text{ld } d - S(\rho)}{\text{ld } d - \text{ld } k}, \quad (57)$$

where  $S$  again denotes the von Neumann entropy.

**Proof.** The idea of the proof is to recursively remove dimensions of low fidelity from the support of  $\varrho$  until we are left with a Hilbert space of given dimension  $k$  and a minimum pure state fidelity bounded from below in terms of  $F_e(\varrho, T)$ . To this end, define

$$f : \mathcal{H} \rightarrow \mathbb{R}, \quad |\psi\rangle \mapsto \langle \psi | T(|\psi\rangle\langle \psi|) | \psi \rangle. \quad (58)$$

Setting  $d := \dim \text{supp}(\varrho)$  and  $\varrho_0 := \varrho$ , we now recursively define a collection  $\{\varrho_i\}_{i=0, \dots, d}$  of positive operators, as follows:

$$\varrho_i := \varrho_{i-1} - q_i |\varphi_i\rangle\langle \varphi_i|, \quad (59)$$

where  $\varphi_i$  is the state vector in the support of  $\varrho_{i-1}$  that minimizes  $f$  and  $q_i$  is the largest positive number that leaves  $\varrho_i$  positive. Note that since  $\dim \mathcal{H}$  is finite,  $q_i$  can be chosen to be strictly positive. By construction,  $\text{supp}(\varrho_i) \subset \text{supp}(\varrho_{i-1})$ , and  $\text{rank}(\varrho_i) = \text{rank}(\varrho_{i-1}) - 1$ ; so our procedure removes dimensions from the support of  $\varrho$  one by one. It follows that

$$\varrho = \sum_{i=1}^d q_i |\varphi_i\rangle\langle \varphi_i|, \quad (60)$$

implying  $\sum_{i=1}^d q_i = \text{tr}(\varrho) = 1$ .

Using the convexity of entanglement fidelity in the density operator input, we see that

$$F_e(\varrho, T) = F_e\left(\sum_{i=1}^d q_i |\varphi_i\rangle\langle\varphi_i|, T\right) \leq \sum_{i=1}^d q_i f(\varphi_i) \leq f(\varphi_{d-k}) \sum_{i=1}^{d-k} q_i + \sum_{i=d-k+1}^d q_i, \quad (61)$$

where in the last step we have used that  $k \mapsto f(\varphi_k)$  is non-decreasing by construction. We now take the subspace  $P_k \mathcal{H}$  as the span of all vectors  $\{\varphi_i\}_{i=d-k+1, \dots, d}$ . Then, since  $\langle\psi|T_k(|\psi\rangle\langle\psi|)|\psi\rangle \geq \langle\psi|T(|\psi\rangle\langle\psi|)|\psi\rangle$  for  $\psi \in P_k \mathcal{H}$ , we have  $F(P_k \mathcal{H}, T_k) \geq f(\varphi_{d-k})$ . Introducing  $q^* = \sum_{i=d-k+1}^d q_i$ , and using  $\sum_{i=1}^{d-k} q_i = 1 - q^*$ , we immediately have the desired estimate.

Our remaining task is to give upper bounds on  $q^*$ , either in terms of the largest eigenvalue of  $\rho$  or its entropy. Note that from the sum representing  $\varrho$  in equation (60) we have

$$q_i \leq q_i + \sum_{\substack{j=1 \\ j \neq i}}^d q_j |\langle\varphi_j|\varphi_i\rangle|^2 = \langle\varphi_i|\varrho|\varphi_i\rangle \leq \|\varrho\|_\infty. \quad (62)$$

Therefore, each of the  $k$  terms in  $q^*$  is bounded by  $\|\varrho\|_\infty$ , and we get  $q^* \leq k\|\varrho\|_\infty$ , which gives the first estimate.

For the entropic estimate, note first that in the inequality

$$S\left(\sum_i q_i \sigma_i\right) \leq \sum_i q_i S(\sigma_i) - \sum_i q_i \text{ld } q_i, \quad (63)$$

which is valid for arbitrary convex combinations of states  $\sigma_i$  with weights  $q_i$  (cf chapter 11.3.6 of [44]), the case of pure states  $\sigma_i$  leaves just the entropy of the probability distribution  $q$ . On the other hand, it is obvious that among all probability distributions with given weight  $q^*$  for the last group of  $k$  indices, the one with the highest entropy is equidistribution, in each of the ranges  $1 \leq i \leq d - k$  and  $d - k + 1 \leq i \leq d$ . Evaluating the entropy of this distribution, and combining this with the previous estimate we find

$$\begin{aligned} S(\rho) &\leq H_2(q^*) + q^* \text{ld } k + (1 - q^*) \text{ld } (d - k) \\ &\leq 1 + \text{ld } d - q^* (\text{ld } d - \text{ld } k), \end{aligned} \quad (64)$$

where the first term denotes the binary Shannon entropy,

$$H_2(q^*) = -q^* \text{ld } q^* - (1 - q^*) \text{ld } (1 - q^*) \leq 1, \quad (65)$$

and we have also used  $\text{ld } (d - k) \leq \text{ld } d$ . Hence the result follows by writing this as an upper bound for  $q^*$ .  $\square$

Proposition 4.5 allows us to make the transition from average error criteria and entropy rates to maximal error criteria. So let us assume that a coding scheme  $(E_n, D_n)_{n \in \mathbb{N}}$  is given, together with a sequence  $(\rho_n)_{n \in \mathbb{N}}$  of source states, such that  $F_e(\rho_n, D_n T^{\otimes n} E_n) \rightarrow 1$ . Then the channel  $T_k$  will again be a corrected version of  $T^{\otimes n}$ , but we can now conclude that its worst case fidelity goes to one.

Let us first consider the case in which the source does not appear explicitly, i.e., in which we assume either the mean fidelity or, equivalently, the channel fidelity to go to one for a scheme with rate  $R$ . Since the channel fidelity is just the entanglement fidelity with respect to a maximally mixed  $\rho$ , we may apply proposition 4.5 with  $\rho_n = 1/\dim \mathcal{H}_n$  and  $\dim \mathcal{H}_n = 2^{\lfloor nR \rfloor}$ , where we denote the largest integer no larger than  $x$  conventionally by  $\lfloor x \rfloor$  (read: *floor of x*). We set  $k = \dim \mathcal{H}_n/2$ , which is to say that the modified coding scheme corrects *just one qubit*

less than the original one. Then  $k\|\rho\|_\infty = 1/2$ , and we immediately find that the minimum fidelity is at least  $1 - (1 - F_e)/2$ , and hence also goes to 1.

The second case of interest is that of a source satisfying the quantum asymptotic equipartition property (QAEP). That is to say, for large  $n$ , the Hilbert space  $\mathcal{H}_n$  can be decomposed into a subspace on which  $\rho_n$  essentially looks like a multiple of the identity, and a subspace of low probability: given any  $\varepsilon > 0$ , for large enough  $n$  essentially all the eigenvalues  $\lambda$  of a QAEP quantum source  $(\rho_n)_{n \in \mathbb{N}}$  with entropy rate  $R$  are concentrated in a so-called  $\varepsilon$ -typical subspace, i.e.,

$$2^{-n(R+\varepsilon)} \leq \lambda \leq 2^{-n(R-\varepsilon)}, \quad (66)$$

in the sense that the sum of the eigenvalues that do not satisfy equation (66) can be made arbitrarily small. We can then conclude that  $\|\rho_n\|_\infty \leq 2^{-n(R-\varepsilon)}$  for large  $n$ . Hence, if we choose  $k \approx 2^{n(R-2\varepsilon)}$ , we can guarantee that  $q^* \rightarrow 0$ , and once again the worst case fidelity has to go to 1. This case covers product sources and stationary ergodic sources [28]–[30], and many others of interest. The discussion of the equivalence between the minimum fidelity version and Schumacher's entanglement fidelity version of channel capacity in [17] is limited to this case.

Does the equivalence hold even without the equipartition property? We will give a counterexample below, which is, however, rather artificial from the point of view of typical coding situations: the dimension of the spaces  $\mathcal{H}_n$  grows superexponentially. This is indeed necessary. For if we have an upper bound  $\dim \mathcal{H}_n \leq \tau^n$  for some positive constant  $\tau$ ,  $S(\rho_n) \approx nR$ , and  $k \approx 2^{n(R-\varepsilon)}$ , we find that  $q^*$  in equation (57) goes to the constant  $(\text{ld } \tau - R)/(\text{ld } \tau - R + \varepsilon) < 1$ . Therefore, the maximal subspace fidelity in equation (55) in proposition 4.5 goes to one if the entanglement fidelity does.

**Counterexample.** Here we show the claim that the Schumacher capacity with unconstrained sources is infinite for all channels with positive quantum capacity. In fact, suppose that we are given a coding scheme with channel fidelity going to 1. Then we simply enlarge the Hilbert space  $\mathcal{H}_n$  by a direct summand  $\mathcal{K}_n$  of some large dimension, and let

$$\rho_n = (1 - \varepsilon_n) \frac{1_{\mathcal{H}_n}}{\dim \mathcal{H}_n} + \varepsilon_n \frac{1_{\mathcal{K}_n}}{\dim \mathcal{K}_n}. \quad (67)$$

The coding operations on  $\mathcal{K}_n$  can be completely depolarizing, for as long as  $\varepsilon_n \rightarrow 0$ , the entanglement fidelity of this source goes to 1, as required. On the other hand, the entropy of this source is

$$S(\rho_n) \geq \varepsilon_n (-\text{ld } \varepsilon_n + \text{ld } \dim \mathcal{K}_n). \quad (68)$$

Clearly, we can make  $S(\rho_n)/n$  diverge if only we let  $\dim \mathcal{K}_n$  go to  $\infty$  fast enough.

#### 4.5. Entanglement generation capacity

We now focus on Devetak's [3] *entanglement generation capacity*, as introduced in section 2.5, and verify that it is totally equivalent to the definitions discussed above. The proof is based on entanglement-assisted teleportation [45] and therefore involves classical forward communication from the encoding to the decoding apparatus. However, this additional resource is shown in section 6.1 not to affect the quantum channel capacity.

Due to the additional freedom of choosing an arbitrary pure input state  $\Gamma \in \mathcal{H} \otimes \mathcal{H}$  instead of the maximally entangled state  $\Omega$ , the entanglement generation capacity is certainly no smaller than the capacity based on channel fidelity, which was shown to be a valid figure of merit in the

previous section. So we only need to prove the converse. This is easily done with the help of teleportation: in the entanglement generation scenario, the sender and receiver end up sharing a state  $\sigma := (DTE \otimes \text{id}_{\mathcal{H}})(|\Gamma\rangle\langle\Gamma|)$  which has asymptotically perfect overlap with the maximally entangled state,

$$F := \langle\Omega|\sigma|\Omega\rangle \geq 1 - \varepsilon \quad (69)$$

for some (small)  $\varepsilon > 0$ . The output system can thus be readily interpreted as being in the maximally entangled state with probability  $F \approx 1$ , and hence can be used as a resource in the standard teleportation protocol [45] to transfer arbitrary quantum states from the sender to the receiver with fidelity no smaller than  $F$ , at the same rate  $R$ .

## 5. Isometric encoding suffices

In this section we will show that if there exists a coding scheme that achieves high fidelity transmission for a given source, there is another coding scheme with isometric encoding, as in equation (11), that also achieves high fidelity transmission. It then directly follows that in the definition of channel capacity we may restrict our attention to isometric encodings, as claimed in section 2.8. While this result is originally due to Barnum *et al* [17], here we give a slightly generalized version of Holevo's presentation (cf chapter 9 of [46]). All we need is the following

**Proposition 5.1.** *Let  $\mathcal{H}$ ,  $\mathcal{K}$  be Hilbert spaces with dimensions  $\eta := \dim \mathcal{H}$  and  $\kappa := \dim \mathcal{K}$ . Let  $\varrho \in \mathcal{B}_*(\mathcal{H})$  be a density operator, and  $E: \mathcal{B}_*(\mathcal{H}) \rightarrow \mathcal{B}_*(\mathcal{K})$  a completely positive map such that  $\text{tr}(E\varrho) = 1$ . Let  $T: \mathcal{B}_*(\mathcal{K}) \rightarrow \mathcal{B}_*(\mathcal{H})$  be a channel. We may then find a channel  $\tilde{E}: \mathcal{B}_*(\mathcal{H}) \rightarrow \mathcal{B}_*(\mathcal{K})$  such that*

$$F_e(\varrho, T\tilde{E}) \geq (F_e(\varrho, TE))^2, \quad (70)$$

where for  $\sigma \in \mathcal{B}_*(\mathcal{H})$  we have

$$\tilde{E}(\sigma) = \begin{cases} V\sigma V^* & \text{for } \eta \leq \kappa \\ W^*\sigma W + \frac{\text{tr}((1_{\mathcal{H}} - WW^*)\sigma)}{\kappa} 1_{\mathcal{K}} & \text{for } \eta > \kappa \end{cases} \quad (71)$$

with isometries  $V: \mathcal{H} \rightarrow \mathcal{K}$  and  $W: \mathcal{K} \rightarrow \mathcal{H}$ , respectively.

**Proof.** Let  $\{t_i\}_{i=1,\dots,\tau}$  and  $\{e_j\}_{j=1,\dots,\varepsilon}$  be sets of Kraus operators for the maps  $T$  and  $E$ , respectively. By equation (7) we have

$$F_e(\varrho, TE) = \sum_{i,j=1}^{\tau,\varepsilon} |\text{tr } t_i e_j \varrho|^2 = \sum_{i,j=1}^{\tau,\varepsilon} |X_{i,j}|^2, \quad (72)$$

where  $X_{i,j} := \text{tr } t_i e_j \varrho$ . If  $\tau \neq \varepsilon$ , add zero components so that  $X$  becomes an  $(m \times m)$  square matrix,  $m := \max\{\tau, \varepsilon\}$ .

By the singular value decomposition we may find unitary matrices  $A, B$  such that  $X = ADB$ , where  $D$  is diagonal with real non-negative entries. Since this decomposition simply corresponds to a change of the Kraus representation of  $T$  and  $E$ , we may assume without loss that  $X$  is diagonal already, and thus

$$F_e(\varrho, TE) = \sum_{i=1}^m X_{i,i}^2 = \sum_{i=1}^m (\text{tr } t_i e_i \varrho)^2. \quad (73)$$

Now for  $k = 1, \dots, m$ , we define  $\lambda_k := \text{tr } e_k \varrho e_k^*$ . Let  $\varrho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ ,  $p_j > 0$ ,  $\sum_j p_j = 1$ , be a diagonal representation of the density operator  $\varrho$ . Then  $\lambda_k = \sum_j p_j \|e_k \psi_j\|_\infty^2$  and  $\text{tr } t_k e_k \varrho = \sum_j p_j \langle\psi_j| t_k e_k |\psi_j\rangle$ . Thus,  $\lambda_k = 0$  implies that  $e_k \psi_j = 0 \forall j$ , and therefore  $\text{tr } t_k e_k \varrho = 0$ , so that these terms do not contribute to the sum in equation (73). We may therefore assume without loss that  $\lambda_k > 0 \forall k = 1, \dots, m$ . Moreover,

$$\sum_{k=1}^m \lambda_k = \sum_{k=1}^m \text{tr } e_k \varrho e_k^* = \text{tr } E(\varrho) = 1, \quad (74)$$

where in the last step we have used that  $E$  is trace-preserving on the state  $\varrho$ . Since

$$F_e(\varrho, TE) = \sum_{i=1}^m \lambda_i \frac{(\text{tr } t_i e_i \varrho)^2}{\lambda_i}, \quad (75)$$

we may find an index  $k$  such that

$$(\text{tr } t e \varrho)^2 = \frac{(\text{tr } t_k e_k \varrho)^2}{\lambda_k} \geq F_e(\varrho, TE), \quad (76)$$

where we have introduced the short-hand notation  $e := e_k / \sqrt{\lambda_k}$  and  $t := t_k$ , respectively. Applying the Schwarz inequality we see that

$$\begin{aligned} (\text{tr } t e \varrho)^2 &= |\text{tr}(t^* \sqrt{\varrho})^* e \sqrt{\varrho}|^2 \\ &\leq \text{tr}(t t^* \varrho) \text{tr}(e^* e \varrho) \\ &= \text{tr}(|t^*|^2 \varrho). \end{aligned} \quad (77)$$

Let us treat the case  $\eta \leq \kappa$  first: since  $t^* t \leq 1_{\mathcal{K}}$ , by working in the spectral representation one easily obtains  $t t^* \leq 1_{\mathcal{H}}$ , and  $|t^*|^2 \leq |t^*|$ , from which it follows that

$$\text{tr } |t^*|^2 \varrho \leq \text{tr } |t^*| \varrho = \text{tr } t V \varrho, \quad (78)$$

where by  $V: \mathcal{H} \rightarrow \mathcal{K}$  we denote the polar isometry of  $t^*$ , i.e.,  $t^* = V|t^*|$ . Existence of this isometry requires  $\eta \leq \kappa$ . Since

$$|\text{tr } t V \varrho|^2 \leq \sum_{i=1}^m |\text{tr } t_i V \varrho|^2 = F_e(\varrho, TV(\cdot)V^*), \quad (79)$$

tracing backwards our results leaves us with the following chain of inequalities:

$$\begin{aligned} F_e(\varrho, TV(\cdot)V^*) &\geq |\text{tr } t V \varrho|^2 && \text{by (79)} \\ &\geq (\text{tr } |t^*|^2 \varrho)^2 && \text{by (78)} \\ &\geq (\text{tr } t e \varrho)^4 && \text{by (77)} \\ &\geq (F_e(\varrho, TE))^2 && \text{by (76),} \end{aligned} \quad (80)$$

the result we set out to prove.

If  $\eta > \kappa$ , the proof proceeds very similarly: we denote the polar isometry of  $t$  by  $W: \mathcal{K} \rightarrow \mathcal{H}$ , i.e.,  $t = W|t|$ . From equation (77) we may then conclude that

$$(\text{tr } t e \varrho)^2 \leq \text{tr } t t^* \varrho = \text{tr } W |t|^2 W^* \varrho \leq \text{tr } W |t| W^* \varrho = \text{tr } t W^* \varrho, \quad (81)$$

where in the second to last step we have used that  $t^*t \leq 1_{\mathcal{K}}$ , and thus  $|t|^2 \leq |t|$ . Substituting  $W^*$  for  $V$ , we now mimic equations (79) and (80) to conclude that  $(F_e(\varrho, TE))^2 \leq F_e(\varrho, TW^*(\cdot)W)$ . The map  $W^*(\cdot)W$ , while being completely positive, is not necessarily trace-preserving. The desired result then follows by renormalization, as above in equations (22) and (54).  $\square$

In proposition 5.1 we have included cases in which the input space of the channel is strictly smaller than the input space of the encoding map,  $\kappa = \dim \mathcal{K} < \dim \mathcal{H} = \eta$ . Though we do not need to consider this situation in our settings, it may prove helpful in other applications to avoid cumbersome distinction of cases, and thus has been added for convenience.

To arrive at the statement that isometric encoding suffices, all we need to do then is to combine proposition 5.1 with the channel fidelity definition of quantum capacity, section 2.5, taking  $\varrho$  to be the chaotic density matrix,  $E$  to be the encoding channel, and thinking of  $T$  as the concatenation of quantum channel and decoding channel.

## 6. Classical side information

As claimed in section 2.9, it is a straightforward consequence of proposition 5.1 that classical forward communication has no effect on the quantum channel capacity [17, 25]. However, before entering the proof we need to make a few more comments on the assisted channel  $T \otimes \text{id}_{\Lambda}^c$ , where  $T: \mathcal{B}_*(\mathcal{H}_1) \rightarrow \mathcal{B}_*(\mathcal{H}_2)$  is an arbitrary quantum channel and  $\text{id}_{\Lambda}^c$  denotes the identity on a classical system with  $\Lambda \in \mathbb{N}$  states. Thus, in the limit of large dimensions, an  $n$ -fold tensor product of the channel  $T$  will be assisted by a classical system with a total of  $\Lambda^n$  states.

The results we are going to present in this section apply slightly more generally: instead of *a priori* fixing a side channel of given (if arbitrarily large) dimension, the encoder may choose the size of the side channel in the encoding process, which also covers the case of super-exponentially growing side channels. The capacity of a channel  $T$  assisted by this type of classical forward communication will be marked with a subscript,  $Q_{cf}(T)$ . This generalization will play a role in section 6.2.

Of course,  $Q_{cf}(T) \geq Q(T \otimes \text{id}_{\Lambda}^c) \geq Q(T)$ . It is the aim of section 6.1 to show that all these capacities are equal.

### 6.1. Classical forward communication does not increase the channel capacity

Obviously, for classically assisted channels the encoding is a channel with both a classical and a quantum output. Such channels are usually called *instruments* [47], and can be thought of as a collection of trace-non-increasing operators  $\{E_{\lambda}\}_{\lambda=1, \dots, \Lambda}$  summing up to a channel  $E = \sum_{\lambda=1}^{\Lambda} E_{\lambda}$ . The index  $\lambda \in \{1, \dots, \Lambda\}$  represents classical information that may be obtained in the encoding process and sent undisturbed to the decoder over a noiseless classical channel. Depending on the value of  $\lambda$ , one channel  $D_{\lambda_0}$  out of a collection of trace-preserving quantum channels  $\{D_{\lambda}\}_{\lambda=1, \dots, \Lambda}$  is used in the decoding process.

The definition of achievable rates and channel capacity now completely parallels the definition of the unassisted quantities. Here we focus on the channel fidelity version of channel capacity (cf section 2.5), since this is a definition for which proposition 5.1 is well suited. Of course, all other definitions of channel capacity can be extended to classically assisted capacities in the same spirit.

We may thus say that  $R$  is an achievable rate for the classically assisted quantum channel  $T$  iff there is a sequence of Hilbert spaces  $(\mathcal{K}_n)_{n \in \mathbb{N}}$  satisfying  $\overline{\lim}_{n \rightarrow \infty} (\text{Id dim } \mathcal{K}_n)/n = R$  and a sequence of encodings  $(E_{\lambda_n, n})_{\lambda_n=1, \dots, \Lambda_n, n \in \mathbb{N}}$  and decodings  $(D_{\lambda_n, n})_{\lambda_n=1, \dots, \Lambda_n, n \in \mathbb{N}}$  for some integer sequence  $(\Lambda_n)_{n \in \mathbb{N}}$  such that

$$\lim_{n \rightarrow \infty} \sum_{\lambda_n=1}^{\Lambda_n} F_c(D_{\lambda_n, n} T^{\otimes n} E_{\lambda_n, n}) = 1. \quad (82)$$

The quantum capacity  $Q_{cf}(T)$  of the channel  $T$  with classical forward communication is then defined as the supremum of all achievable rates.

Of course, the capacity of the channel  $T \otimes \text{id}_\Lambda^c$  for fixed  $\Lambda \in \mathbb{N}$  is obtained by setting  $\Lambda_n := \Lambda^n$  in the above definition.

**Theorem 6.1.** *Let  $T : \mathcal{B}_*(\mathcal{H}_1) \longrightarrow \mathcal{B}_*(\mathcal{H}_2)$  be a quantum channel and  $\Lambda \in \mathbb{N}$ . We then have*

$$Q_{cf}(T) = Q(T \otimes \text{id}_\Lambda^c) = Q(T). \quad (83)$$

Before giving the proof let us consider a seeming generalization of this theorem, which allows the side channel  $R$  to be any *separable channel*, i.e., a channel  $R = R_2 R_1$  operating by first collecting classical information, by a channel  $R_1$ , say and then recoding this into quantum information by another channel  $R_2$ . Equivalently,  $(\text{id} \otimes R)$  maps any input state to a separable state, so these channels are also called ‘entanglement breaking’ [48, 49]. Then, for every such  $R$  and any channel  $T$  we have the following

**Corollary 6.2.** *Separable side channels do not increase the quantum channel capacity, i.e., for any quantum channel  $T$  and any separable channel  $R = R_2 \circ R_1$  we have*

$$Q(T \otimes R) = Q(T). \quad (84)$$

**Proof of corollary 6.2.** For a separable channel  $R = R_1 \circ R_2 = R_2 \circ \text{id}_\Lambda^c \circ R_1$ , we have

$$\begin{aligned} Q(T) &\leq Q(T \otimes R) = Q((\text{id} \otimes R_2)(T \otimes \text{id}_\Lambda^c)(\text{id} \otimes R_1)) \\ &\leq Q(T \otimes \text{id}_\Lambda^c) = Q(T), \end{aligned} \quad (85)$$

where the first inequality follows by using codings ignoring the channel  $R$ , the second follows by the bottleneck inequality (15), and in the last step we have applied theorem 6.1. From this chain of inequalities we get  $Q(T) = Q(T \otimes R)$ , just as claimed.  $\square$

We now follow [17] in the

**Proof of theorem 6.1.** From our remarks it is clear that we only need to prove the inequality  $Q_{cf}(T) \leq Q(T)$ . Given a sequence of Hilbert spaces  $(\mathcal{K}_n)_{n \in \mathbb{N}}$  and suitable encodings  $(E_{\lambda_n, n})_{\lambda_n=1, \dots, \Lambda_n, n \in \mathbb{N}}$  and decodings  $(D_{\lambda_n, n})_{\lambda_n=1, \dots, \Lambda_n, n \in \mathbb{N}}$  such that the classically assisted channel  $T$  achieves the rate  $R$  with channel fidelity approaching one, we will show that the same rate can be achieved without using the side channel.

From the definition of channel capacity, for any  $\varepsilon > 0$  we may find  $n_\varepsilon \in \mathbb{N}$  such that

$$\sum_{\lambda_n=1}^{\Lambda_n} F_c(D_{\lambda_n, n} T^{\otimes n} E_{\lambda_n, n}) \geq 1 - \varepsilon \quad \forall n \geq n_\varepsilon. \quad (86)$$

For the remainder of the proof we will fix  $n \geq n_\varepsilon$  and drop the index to streamline the notation. Setting  $e_\lambda := \text{tr } E_\lambda 1/d$  with  $d := \dim \mathcal{K}_n$  and

$$\tilde{F}_c(D_\lambda TE_\lambda) := \frac{1}{e_\lambda} F_c(D_\lambda TE_\lambda), \quad (87)$$

we may then rewrite equation (86) as

$$\sum_{\lambda=1}^{\Lambda} e_\lambda \tilde{F}_c(D_\lambda TE_\lambda) \geq 1 - \varepsilon. \quad (88)$$

Since  $\sum_{\lambda=1}^{\Lambda} e_\lambda = 1$ , there is an index  $\mu$  such that  $\tilde{F}_c(D_\mu TE_\mu) \geq 1 - \varepsilon$ . However,

$$\tilde{F}_c(D_\mu TE_\mu) = F_c\left(D_\mu T \frac{E_\mu}{e_\mu}\right), \quad (89)$$

and we may therefore apply proposition 5.1 to conclude that there is a channel  $\tilde{E}$  such that

$$F_c(D_\mu T \tilde{E}) \geq (1 - \varepsilon)^2, \quad (90)$$

from which it follows that the same rate can be achieved without relying on the classical side channel.  $\square$

## 6.2. Average fidelity by forward communication

We already know that average fidelity may be considered a suitable error criterion for capacity purposes. The line of thought we followed to establish this fact proceeded via the equivalence of average fidelity and channel fidelity, proposition 4.4, and is thus ultimately based on proposition 4.5.

The results of the previous section regarding the uselessness of classical forward communication can be employed to give an alternative proof that average fidelity serves as a valid distance measure, making use only of the sufficiency of isometric encoding.

To this end, we will show that instead of evaluating the average fidelity  $\bar{F}(T)$  for a given channel  $T$ , we may just as well compute the minimum fidelity  $F(\tilde{T})$ , where the new channel  $\tilde{T}$  is simply the old channel  $T$  augmented by classical forward communication. However, by theorem 6.1 the quantum capacities of  $\tilde{T}$  and  $T$  coincide, and therefore average fidelity and minimum fidelity turn out to be equivalent error quantities.

While our concept of classically assisted channel capacity, as presented in section 6.1, only allows for discrete classical messages and therefore involves the calculation of finite sums, for the evaluation of average fidelity we will rather deal with integrals, corresponding to continuous classical messages. However, this extension poses no difficulties.

So suppose we have at hand a quantum channel  $T: \mathcal{B}_*(\mathcal{H}_1) \rightarrow \mathcal{B}_*(\mathcal{H}_2)$  together with encoding channel  $E: \mathcal{B}_*(\mathcal{H}_3) \rightarrow \mathcal{B}_*(\mathcal{H}_1)$  and decoding channel  $D: \mathcal{B}_*(\mathcal{H}_2) \rightarrow \mathcal{B}_*(\mathcal{H}_3)$ . By equation (5) the average fidelity of the concatenation  $D \circ T \circ E$  then reads

$$\begin{aligned} \bar{F}(\mathcal{H}_3, DTE) &= \int \langle U\psi | DTE(|U\psi\rangle\langle U\psi|) |U\psi\rangle dU \\ &= \text{tr} \left( \varrho \int U^* DTE(U\varrho U^*) U dU \right), \end{aligned} \quad (91)$$

where integration is over all unitaries  $U \in \mathcal{B}_*(\mathcal{H}_3)$ , and  $\varrho = |\psi\rangle\langle\psi| \in \mathcal{B}_*(\mathcal{H}_3)$  is an arbitrary reference state. We will now convert  $DTE$  into a channel with classical forward communication. Denoting by  $\mathcal{C}(X)$  the vector space of continuous functions on the set  $X$ , we may define  $\tilde{D}: \mathcal{B}_*(\mathcal{H}_2) \otimes \mathcal{C}(U(\mathcal{H}_3)) \rightarrow \mathcal{B}_*(\mathcal{H}_3)$  and  $\tilde{E}: \mathcal{B}_*(\mathcal{H}_3) \rightarrow \mathcal{B}_*(\mathcal{H}_1) \otimes \mathcal{C}(U(\mathcal{H}_3))$ , as follows:

$$\tilde{D}(\varrho \otimes f) := \int U^* D(\varrho) U f(U) dU, \quad \tilde{E}_U(\sigma) := E(U\sigma U^*), \quad (92)$$

where in the definition of  $\tilde{E}$  we have made use of the fact that  $\mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X)$  is isomorphic to the  $\mathcal{B}(\mathcal{H})$ -valued functions on  $X$ . We then see that for any state  $\varrho = |\psi\rangle\langle\psi| \in \mathcal{B}_*(\mathcal{H}_3)$

$$\begin{aligned} \langle\psi|\tilde{D}(T \otimes \text{id}_{\mathcal{C}(U(\mathcal{H}_3))})\tilde{E}(|\psi\rangle\langle\psi|)|\psi\rangle &= \text{tr } \varrho \tilde{D}(T \otimes \text{id}_{\mathcal{C}(U(\mathcal{H}_3))})\tilde{E}(\varrho) \\ &= \text{tr} \left( \varrho \int U^* DTE(U\varrho U^*) U dU \right). \end{aligned} \quad (93)$$

The average fidelity for the concatenation  $DTE$  thus equals the minimum pure state fidelity for the classically assisted channel  $\tilde{D}(T \otimes \text{id})\tilde{E}$ , which is what we wanted to show.

Note that for this proof to apply also in the setting of  $n$ -fold tensor products, as required by the definition of channel capacity, we may not restrict ourselves to exponentially growing side channels,  $(T \otimes \text{id}_{\Lambda})^{\otimes n}$ : this would correspond to an averaging over  $n$ -fold tensor products of unitary operators,  $U_1 \otimes U_2 \otimes \dots \otimes U_n$ . However, not all unitary operators on an  $n$ -fold tensor product are of this form.

## 7. Testing a single sequence

We will now prove the claim made in section 2.2: if a coding scheme construction works for a certain pair of integer sequences  $(n_\nu)_{\nu \in \mathbb{N}}$ ,  $(m_\nu)_{\nu \in \mathbb{N}}$  such that the rate  $R$  is achieved infinitely often, i.e.,  $\overline{\lim}_{\nu \rightarrow \infty} (m_\nu/n_\nu) = R$ , and the error tends to zero,  $\lim_{\nu \rightarrow \infty} \Delta(n_\nu, m_\nu) = 0$ , then coding works for all such pairs.

As mentioned in section 2.2, this requires extending a given coding scheme to more block sizes. Therefore this section will be organized by extension method: in section 7.1 we use only the method of *wasting resources*, i.e., either using the coded channels for fewer bits than allowed by the given coding scheme (i.e., decreasing  $m_\nu$ ) or requiring some additional channel uses (thus increasing  $n_\nu$ ) and simply not using them. This will allow the extension whenever we can find a subsequence along which, on the one hand, the desired rate is achieved and which, on the other hand, does not grow too fast.

A second method would be to use blocks from the given coding scheme and put them together as tensor products, to get to larger block sizes. We show in section 7.2 by an explicit example that this method, combined with the wasteful one, is not sufficient to extend a very sparse coding sequence to all large block sizes.

Finally, we show in section 7.3, based on the work in [24], how hashing codes can be used to achieve the desired extension in all cases.

Throughout, we will denote by  $(N_\mu)_{\mu \in \mathbb{N}}$ ,  $(M_\mu)_{\mu \in \mathbb{N}}$  the given coding sequences, and assume, without loss of generality, that the sporadic rate is attained, i.e.,  $R = \overline{\lim}_{\mu \rightarrow \infty} (M_\mu/N_\mu) = \lim_{\mu \rightarrow \infty} (M_\mu/N_\mu)$ . The sequences to which we seek to extend the scheme are denoted by  $(n_\nu)_{\nu \in \mathbb{N}}$  and  $(m_\nu)_{\nu \in \mathbb{N}}$ , as before.

### 7.1. Subexponential sequences

Obviously, good coding becomes easier the more parallel channels are available for transmission. Moreover, if a certain coding scheme works for some Hilbert space  $\mathcal{H}$ , it works at least as well for states supported on a lower dimensional Hilbert space  $\mathcal{H}'$ . Thus, the error quantity  $\Delta(n, m)$ , as introduced in definition 2.1, has the following monotonicity properties:

$$\Delta(n + 1, m) \leq \Delta(n, m) \leq \Delta(n, m + 1) \quad (94)$$

for all positive integers  $n, m$ . We call a diverging sequence  $(N_\mu)_{\mu \in \mathbb{N}}$  *subexponential* if

$$\lim_{\mu \rightarrow \infty} \frac{N_{\mu+1}}{N_\mu} = 1. \quad (95)$$

This covers, for example, all arithmetic sequences, and polynomially growing ones. For such sequences the desired result follows directly from the following lemma, which slightly generalizes lemma 3.2 of [24].

**Lemma 7.1.** *Suppose  $\Delta : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R}_+$  satisfies the monotonicity properties (94). Let  $(N_\mu)_{\mu \in \mathbb{N}}$ ,  $(M_\mu)_{\mu \in \mathbb{N}}$  be a pair of integer sequences such that  $(N_\mu)_{\mu \in \mathbb{N}}$  is subexponential, cf equation (95), and  $\lim_{\mu \rightarrow \infty} \Delta(N_\mu, M_\mu) = 0$ . Then for any pair of integer sequences  $(n_\nu)_{\nu \in \mathbb{N}}$ ,  $(m_\nu)_{\nu \in \mathbb{N}}$  such that  $\lim_{\nu \rightarrow \infty} n_\nu = \infty$  and*

$$\overline{\lim}_{\nu \rightarrow \infty} \frac{m_\nu}{n_\nu} < \underline{\lim}_{\mu \rightarrow \infty} \frac{M_\mu}{N_\mu},$$

we have  $\lim_{\nu \rightarrow \infty} \Delta(n_\nu, m_\nu) = 0$ .

**Proof.** If we have only the monotonicity property of  $\Delta$  to draw upon, the way to show that  $\Delta(n_\nu, m_\nu) \rightarrow 0$  is to find a suitable index  $\mu = \mu(\nu)$  for all sufficiently large  $\nu$  so that  $\Delta(n_\nu, m_\nu) \leq \Delta(N_{\mu(\nu)}, M_{\mu(\nu)})$ , for which we need

$$n_\nu \geq N_{\mu(\nu)} \quad \text{and} \quad m_\nu \leq M_{\mu(\nu)}. \quad (96)$$

The first inequality we will ensure by defining

$$\mu(\nu) = \min\{\alpha \mid N_\alpha \geq n_\nu\} - 1. \quad (97)$$

Then

$$N_{\mu(\nu)} \leq n_\nu \leq N_{\mu(\nu)+1}, \quad (98)$$

and  $\lim_{\nu} \mu(\nu) = \infty$ . Hence it remains to show that the second inequality in equation (96) holds for all sufficiently large  $\nu$ . We consider

$$\frac{m_\nu}{M_{\mu(\nu)}} = \frac{m_\nu}{n_\nu} \frac{n_\nu}{N_{\mu(\nu)+1}} \frac{N_{\mu(\nu)+1}}{N_{\mu(\nu)}} \frac{N_{\mu(\nu)}}{M_{\mu(\nu)}}. \quad (99)$$

In this product the second factor is  $\leq 1$  by equation (98), and the third converges to 1 because  $(N_\mu)_{\mu \in \mathbb{N}}$  is subexponential. Now pick  $R_-, R_+$  such that strict inequalities

$$\overline{\lim}_{\nu \rightarrow \infty} \frac{m_\nu}{n_\nu} < R_- < R_+ < \underline{\lim}_{\mu \rightarrow \infty} \frac{M_\mu}{N_\mu}, \quad (100)$$

hold. Then for all sufficiently large  $\nu$  the first factor in equation (99) is  $\leq R_-$ , and the last factor is  $\leq 1/R_+$ . Hence the product of the first and last factor in equation (99) is  $\leq R_-/R_+ < 1$ . Hence equation (96) holds for all sufficiently large  $\nu$ .  $\square$

This result covers most sequences  $(N_\mu)_{\mu \in \mathbb{N}}, (M_\mu)_{\mu \in \mathbb{N}}$  naturally arising for families of codes. In contrast to proposition 7.2 the result therefore remains useful even for the simulation of one noisy channel by another, i.e., for the definition of capacities  $Q(T, S)$  with non-ideal reference channel  $S$ .

### 7.2. A counterexample

From the way in which subexponentiality of  $(N_\mu)_{\mu \in \mathbb{N}}$  enters the proof of lemma 7.1, it is not clear whether this assumption is really necessary. In this section we will give an example showing that it cannot be omitted, implying that to establish full equivalence we do need the more sophisticated techniques presented in section 7.3. The example will also satisfy another natural constraint on the error function  $\Delta(n, m)$ , which reflects another elementary method of getting new coding schemes from old: we can always split the given number of channels into sub-blocks, and apply a known coding scheme to each block. The total error is then estimated as the sum of the errors of each block. Hence the error function  $\Delta(n, m)$  is *subadditive*:

$$\Delta(n_1 + n_2, m_1 + m_2) \leq \Delta(n_1, m_1) + \Delta(n_2, m_2). \quad (101)$$

Suppose we are given some codes for a possibly very sparse sequence of block sizes  $N_\mu$ , with  $M_\mu = N_\mu$  coded bits, and  $\varepsilon_\mu = \Delta(N_\mu, N_\mu) \rightarrow 0$ . Then the rate 1 is sporadically achievable. The error bound we get by the best combination of blocking and possibly wasting some resources is then

$$\Delta(n, m) := \inf \left\{ \sum_k \varepsilon_{\mu_k} \mid m \leq \sum_k N_{\mu_k} \leq n \right\}, \quad (102)$$

where the infimum is taken over all admissible sets  $\{N_{\mu_k}\}$ . This satisfies both monotonicity (94) and subadditivity (101). Our aim in constructing the counterexample is to choose  $(N_\mu)_{\mu \in \mathbb{N}}$  growing sufficiently rapidly, and  $(\varepsilon_\mu)_{\mu \in \mathbb{N}}$  decreasing sufficiently slowly, so that  $\Delta(n, m)$  can be bounded away from zero even though  $m/n$  gets small.

We assume that  $(N_\mu)_{\mu \in \mathbb{N}}$  is *superexponential* in the sense that  $N_{\mu+1}/N_\mu \rightarrow \infty$ . Of  $(\varepsilon_\mu)_{\mu \in \mathbb{N}}$  for the moment we only require that it decreases monotonically to zero. Then the infimum in equation (102) never contains sums arising by breaking up a block of size  $N_{\mu_k}$  into blocks of smaller sizes: in this way one would not only get more terms in the sum of  $\varepsilon$ s, but each term would be larger than  $\varepsilon_{\mu_k}$ . Therefore we can lower bound  $\Delta$  by considering only a decomposition into the largest available blocks. For  $N_\mu \leq m \leq n < N_{\mu+1}$  this means

$$\Delta(n, m) \geq \varepsilon_\mu \left\lfloor \frac{m}{N_\mu} \right\rfloor, \quad (103)$$

where  $\lfloor x \rfloor$  denotes the largest integer  $\leq x$ . Now we choose  $n_\mu = N_{\mu+1} - 1$ , and  $m_\mu$  close to the geometric mean:  $m_\mu \approx \sqrt{N_\mu N_{\mu+1}}$ . Then on the one hand  $m_\mu/n_\mu \approx \sqrt{N_\mu/N_{\mu+1}} \rightarrow 0$ , because  $(N_\mu)_{\mu \in \mathbb{N}}$  is superexponential. Hence this pair of sequences has rate zero. On the other hand, if we only let  $\varepsilon_\mu$  decrease slowly enough we can prevent  $\Delta(n_\mu, m_\mu)$  from going to zero. For example, with  $\varepsilon_\mu = \sqrt{N_\mu/N_{\mu+1}}/2$  we get  $\Delta(n_\mu, m_\mu) \geq 1/2$  asymptotically.

In summary, we have constructed a monotonic and subadditive function  $\Delta(n, m)$ , for which the rate 1 can be achieved sporadically, but for which the proper achievable rate is 0.

### 7.3. Hashing helps

We will now explain how hashing helps to establish full equivalence of the one-sequence and all-sequence definitions, showing that it is indeed sufficient to check only one pair of sequences when testing a given rate  $R$ . As we know from the previous subsection, this requires that if we have found a fairly good coding for some large block size, we must make better use of it than just repeating the blocks, and maybe not using some of the input bits.

This problem is in essence the same as that arising when we have a fairly good channel to begin with: just repeating it without further encoding will not make errors go to zero. Instead they will accumulate. But, on the other hand, a channel which is nearly ideal should also have nearly the capacity of an ideal channel, or else the whole idea of capacity would make no sense. In fact, in our paper, so far we have only shown one type of channel to have positive capacity, namely the ideal channels. As shown in section 3.2, in that case the problem of accumulating errors simply does not occur, and all coding is with  $\Delta(n, m) = 0$ .

So it would actually be conceivable that capacities are always zero, unless coding can be done without errors (see also section 2.7). However, it can be shown that *small errors can be corrected* with only a small loss in capacity. This problem is treated in a self-contained way in [24], and we refer to that paper for details and proofs. Here we only point out the statements needed in the present context, and sketch the main arguments.

The non-trivial family of codes needed for this argument are called hash codes. In [24] they are constructed as *random graph codes*, based on a scheme [50] which turns graphs into quantum error correcting codes of the Knill–Laflamme type. The verification that a certain number of errors is corrected by such a code amounts to showing that a certain system of linear equations is non-singular. Then the existence of codes with suitable parameters is shown by checking that this condition holds true in a generic random graph of suitable size. The random graphs are generated such that the probabilities for each edge are independent and equidistributed. This is quite different from Shannon’s idea of random coding, where the distribution depends on the noise in the channel and the input state. The argument based on graph codes works in any Hilbert space dimension  $d$  which is a prime number. It shows that if we want to encode  $m$  systems of dimension  $d$  into  $n$  systems of the same dimension, and

$$\left(\frac{m}{n} + \frac{4f}{n} - 1\right) \text{ld } d + H_2\left(\frac{2f}{n}\right) < 0 \quad (104)$$

then we can arrange for the code to correct arbitrary errors occurring on up to  $f$  of the  $n$  subsystems. Here

$$H_2(p) = -p \text{ld } (p) - (1 - p) \text{ld } (1 - p) \quad (105)$$

is again the binary entropy function. Moreover, the expression in equation (104) is an upper bound on the exponential rate at which the probability for a random graph code *not* to correct that many errors decreases. The crucial feature is that if  $f/n$  is small, i.e., we do not require many errors to be corrected, then we can get  $m$  close to  $n$ , i.e., the rate of the coding scheme is nearly that of the ideal channel.

The next step is to convert the correction of *rare* errors to that of arbitrary *small* errors. Here a straightforward norm estimate is

$$\|DT^{\otimes n}E - \text{id}_m\|_{cb} \leq \left(2^{H_2((f+1)/n)}\|T - \text{id}\|_{cb}^{\frac{f+1}{n}}\right)^n, \quad (106)$$

when  $E, D$  are a code correcting  $f$  out of  $n$  errors on  $m$  input systems, as above, and  $\text{id}_m$  denotes the ideal channel on  $m$   $d$ -level systems. Then as soon as the expression in parentheses is  $<1$ , in particular, if the channel  $T$  is close to ideal, we see that the errors go to zero exponentially in  $n$ .

We now apply these ideas to a given coding solution for some channel  $T$ , i.e., we assume that for the given channel we have some encoding of a  $d$ -level system through  $N$  parallel uses of the channel. The nominal rate of this coding scheme, expressed in the units ‘qubits per channel use’ is  $\text{ld } d/N$ . For large  $N$  we may as well assume that  $d$  is a prime number, because the gaps between consecutive primes go to zero [51]. Then we apply the above ideas to the encoded channel  $\tilde{T} = D(T^{\otimes N})E$ . The overall code will require  $nN$  channel uses, and the encoded systems are  $d^m$  dimensional. If we use  $n$  as the index of the resulting sequence of channel uses, the resulting sequence of block lengths grows linearly, hence is clearly subexponential, and has rate  $(m \text{ld } d)/(nN)$ . The errors go to zero exponentially, provided we can find an  $f$  satisfying equation (104) and such that the quantity in parentheses in equation (106) is strictly less than one. Combining all this gives the following estimate (theorem 8.2 in [24]):

**Proposition 7.2.** *Let  $T$  be a channel, not necessarily between systems of the same dimension. Let  $N, d \in \mathbb{N}$  with  $d$  a prime number, and suppose that there are channels  $E$  and  $D$  encoding and decoding a  $d$ -level system through  $N$  parallel uses of  $T$ , with error  $\Delta = \|DT^{\otimes N}E - \text{id}_d\|_{cb} < (1/2e)$ , with  $e = \exp 1$ . Then*

$$Q(T) \geq \frac{\text{ld } d}{N}(1 - 4e\Delta) - \frac{1}{N}H_2(2e\Delta). \quad (107)$$

Moreover,  $Q(T)$  is the least upper bound on all expressions of this form, and for coding rates below the bound the errors decrease exponentially.

Note that here a single successful coding scheme  $(E, D)$  guarantees at least a lower bound to the capacity. The most important aspect of this bound is once again that the precision  $\Delta$  required does not depend on the dimension  $d$ . Therefore, even if we know such codes only on an arbitrarily thinly spaced sequence of  $N$ 's, with vanishing errors along this thin subsequence, we can achieve all rates below the sporadic rate  $(\overline{\lim}_\mu \text{ld } d_\mu/N_\mu)$  by subexponential sequences as well, and hence for any sequence, as required by definition 2.1. Thus the sporadic capacity is equal to the capacity.

Note that proposition 7.2 also clarifies the questions brought up in section 2.7: indeed a requirement that errors should vanish exponentially fast can be met for any achievable rate strictly below the capacity. Analogous results have been presented very recently by Hamada [52], building on earlier work in [13, 19, 53].

Moreover, it is clear from proposition 7.2 that tolerating finite errors is possible: since we require the capacity  $Q_\varepsilon(T)$  to be achieved for arbitrarily large  $N$ , the second term in equation (107) also goes to zero, and we get the bound

$$Q_\varepsilon(T) \geq Q(T) \geq Q_\varepsilon(T)(1 - 4e\varepsilon). \quad (108)$$

Hence  $\lim_{\varepsilon \rightarrow 0} Q_\varepsilon(T) = Q(T)$ , as claimed.

## Acknowledgments

We thank A Winter for fruitful discussions and A S Holevo for letting us use his version of the isometric encoding theorem, as well as for his perceptive comments on an earlier version of the manuscript. Funding from Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged.

## References

- [1] Shor P W 2002 The quantum channel capacity and coherent information, *Lecture Notes, MSRI Workshop on Quantum Computation* (San Francisco, November 2002); available online at <http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1>
- [2] Shor P W 2003 Capacities of quantum channels and how to find them *Preprint* quant-ph/0304102
- [3] Devetak I 2003 The private classical information capacity and quantum information capacity of a quantum channel *Preprint* quant-ph/0304127
- [4] Kraus K 1983 *States, Effects, and Operations* (Berlin: Springer)
- [5] Paulsen V I 2002 *Completely Bounded Maps and Operator Algebras* (Cambridge: Cambridge University Press)
- [6] Shannon C E 1948 *Bell Syst. Tech. J.* **27** 379, 623, reprinted in: Shannon C E 1993 *Collected Papers* ed N J A Sloane and A D Wyner (Piscataway, NJ: IEEE Press); also available at <http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html>
- [7] McEliece R J 1977 *The Theory of Information and Coding* (Reading, MA: Addison-Wesley)
- [8] Ash R B 1990 *Information Theory* (New York: Dover)
- [9] Werner R F 2001 Quantum information theory—an invitation *Quantum Information* ed G Alber *et al* (Berlin: Springer) (*Preprint* quant-ph/0101061)
- [10] Bennett C H and Shor P W 1998 *IEEE Trans. Inf. Th.* **44** 2724
- [11] Shor P W 2000 *Geom. Func. Anal.* **GAF A 2000** (Special Vol) 816
- [12] Holevo A S and Werner R F 2001 Evaluating capacities of bosonic Gaussian channels *Phys. Rev. A* **63** 032312 (*Preprint* quant-ph/9912067)
- [13] Hamada M 2002 Lower bounds on the quantum capacity and highest error exponent of general memoryless channels *IEEE Trans. Inf. Th.* **48** 2547 (*Preprint* quant-ph/0112103)
- [14] Keyl M 2002 Fundamentals of quantum information theory *Phys. Rep.* **369** (5) (*Preprint* quant-ph/0202122)
- [15] Ogawa T and Nagaoka H 2002 A new proof of the channel coding theorem via hypothesis testing in quantum information theory *Proc. 2002 IEEE Int. Symp. on Inf. Th. (Piscataway, NJ)* (*Preprint* quant-ph/0208139)
- [16] Barnum H, Nielsen M A and Schumacher B 1998 Information transmission through a noisy quantum channel *Phys. Rev. A* **57** 4153 (*Preprint* quant-ph/9702049)
- [17] Barnum H, Knill E and Nielsen M A 2000 On quantum fidelities and channel capacities *IEEE Trans. Inf. Th.* **46** 1317 (*Preprint* quant-ph/9809010)
- [18] Barnum H, Smolin J A and Terhal B M 1998 Quantum capacity is properly defined without encodings *Phys. Rev. A* **58** 3496 (*Preprint* quant-ph/9711032)
- [19] Hamada M 2002 Information rates achievable with algebraic codes on quantum discrete memoryless channels *Preprint* quant-ph/0207113 *IEEE Trans. Inf. Th.* submitted
- [20] Bennett C H, DiVincenzo D P and Smolin J A 1997 Capacities of quantum erasure channels *Phys. Rev. Lett.* **78** 3217 (*Preprint* quant-ph/9701015)
- [21] DiVincenzo D P, Shor P W and Smolin J A 1998 Quantum channel capacities of very noisy channels *Phys. Rev. A* **57** 830 (*Preprint* quant-ph/9706061)

- [22] Horodecki M, Horodecki P and Horodecki R 2000 Unified approach to quantum capacities: towards quantum noisy coding *Phys. Rev. Lett.* **85** 433 (*Preprint* quant-ph/0003040)
- [23] Matsumoto R and Uyematsu T 2002 Lower bound for the quantum capacity of a discrete memoryless quantum channel *J. Math. Phys.* **43** 4391 (*Preprint* quant-ph/010515 v4)
- [24] Keyl M and Werner R F 2002 How to correct small quantum errors? *Coherent Evolution in Noisy Environments, Lecture Notes in Physics* ed A M Buchleitner and K Hornberger (Berlin: Springer) (*Preprint* quant-ph/0206086)
- [25] Bennett C H, DiVincenzo D P, Smolin J A and Wootters W K 1996 Mixed state entanglement and quantum error correction *Phys. Rev. A* **54** 3824 (*Preprint* quant-ph/9604024 v2)
- [26] Schumacher B 1996 Sending entanglement through noisy quantum channels *Phys. Rev. A* **54** 2614 (*Preprint* quant-ph/9604023)
- [27] Reimpell M and Werner R F 2003 Iterative optimization of quantum error correcting codes *Preprint* quant-ph/0307138
- [28] Bjelaković I, Krüger T, Siegmund-Schultze R and Szkoła A 2002 The Shannon-McMillan theorem for ergodic quantum lattice systems *Preprint* math.DS/0207121 v3
- [29] Bjelaković I, Krüger T, Siegmund-Schultze R and Szkoła A 2003 Chained typical subspaces—a quantum version of Breiman’s theorem *Preprint* quant-ph/0301177 v2
- [30] Bjelaković I and Siegmund-Schultze R 2003 An ergodic theorem for the quantum relative entropy *Preprint* quant-ph/0306094
- [31] Knill E and Laflamme R 1997 Theory of quantum error-correcting codes *Phys. Rev. A* **55** 900 (*Preprint* quant-ph/9604034)
- [32] Bowen G 2002 Quantum feedback channels *Preprint* quant-ph/0209076
- [33] Lloyd S 1997 Capacity of the noisy quantum channel *Phys. Rev. A* **55** 1613 (*Preprint* quant-ph/9604015)
- [34] Harrow A 2003 Coherent classical communication *Preprint* quant-ph/0307091
- [35] Devetak I, Harrow A W and Winter A 2003 A family of quantum protocols *Preprint* quant-ph/0308044
- [36] Devetak I and Winter A 2003 Distillation of secret key and entanglement from quantum states *Preprint* quant-ph/0306078
- [37] Devetak I and Winter A 2003 Relating quantum privacy and quantum coherence: an operational approach *Preprint* quant-ph/0307053
- [38] Reed M and Simon B 1980 *Methods of Modern Mathematical Physics I: Functional Analysis* (New York: Academic)
- [39] Bratteli O and Robinson D W 1979 *Operator Algebras and Quantum Statistical Mechanics I* (New York: Springer)
- [40] Horodecki M, Horodecki P and Horodecki R 1999 General teleportation channel, singlet fraction, and quasidistillation *Phys. Rev. A* **60** 1888 (*Preprint* quant-ph/9807091)
- [41] Nielsen M A 2002 A simple formula for the average gate fidelity of a quantum dynamical operation *Phys. Lett. A* **303** 249 (*Preprint* quant-ph/0205035 v2)
- [42] Werner R F 1989 *Phys. Rev. A* **40** 4277
- [43] Vollbrecht K G H and Werner R F 2001 Entanglement measures under symmetry *Phys. Rev. A* **64** 062307 (*Preprint* quant-ph/0010095)
- [44] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
- [45] Bennett C H, Brassard G, Crépeau C, Jozsa R, Peres A and Wootters W K 1993 *Phys. Rev. Lett.* **70** 1895
- [46] Holevo A S 2002 An introduction to quantum information theory: systems, channels, capacities (unpublished)
- [47] Davies E B 1976 *Quantum Theory of Open Systems* (London: Academic)
- [48] Horodecki M, Shor P W and Ruskai M B 2003 Entanglement breaking channels *Preprint* quant-ph/0302031
- [49] Ruskai M B 2003 Qubit entanglement breaking channels *Preprint* quant-ph/0302032

- [50] Schlingemann D M and Werner R F 2002 Quantum error-correcting codes associated with graphs *Phys. Rev. A* **65** 012308 (*Preprint* quant-ph/0012111)
- [51] Ingham A E 1937 *Quart. J. Math., Oxford Ser.* **8** 255
- [52] Hamada M 2003 Notes on the fidelity of symplectic quantum error-correcting codes *Preprint* quant-ph/0311003
- [53] Hamada M 2002 Exponential lower bound on the highest fidelity achievable by quantum error-correcting codes *Phys. Rev. A* **65** 052305-1-4 (*Preprint* quant-ph/0109114)