The coding theorem for a class of quantum channels with long-term memory

2007 J. Phys. A: Math. Theor. 40 8147

(http://iopscience.iop.org/1751-8121/40/28/S20)

IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **40** (2007) 8147–8164 doi:10.1088/1751-8113/40/28/S20

The coding theorem for a class of quantum channels with long-term memory

Nilanjana Datta[1] **and Tony C Dorlas**[2]

[1] Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB30WB, UK
[2] Dublin Institute for Advanced Studies, School of Theoretical Physics, 10 Burlington Road, Dublin 4, Ireland

E-mail: n.datta@statslab.cam.ac.uk and dorlas@stp.dias.ie

Received 24 October 2006, in final form 20 March 2007
Published 27 June 2007
Online at stacks.iop.org/JPhysA/40/8147

Abstract

In this paper, we consider the transmission of classical information through a class of quantum channels with long-term memory, which are convex combinations of memoryless channels. Hence, the memory of such channels can be considered to be given by a Markov chain which is aperiodic but not irreducible. We prove the coding theorem and weak converse for this class of channels. The main techniques that we employ are a quantum version of Feinstein's fundamental lemma (Feinstein A 1954 *IRE Trans. PGIT* **4** 2–22, Khinchin A I 1957 *Mathematical Foundations of Information Theory: II. On the Fundamental Theorems of Information Theory* (New York: Dover) chapter IV) and a generalization of Helstrom's theorem (Helstrom C W 1976 Quantum detection and estimation theory *Mathematics in Science and Engineering* vol 123 (London: Academic)).

PACS number: 03.67.−a

1. Introduction

The biggest hurdle in the path of efficient information transmission is the presence of noise, in both classical and quantum channels. This noise causes a distortion of the information sent through the channel. Error-correcting codes are used to overcome this problem. Instead of transmitting the original messages, they are encoded into codewords, which are then sent through the channel. Information transmission is said to be reliable if the probability of error, in decoding the output of the channel, vanishes asymptotically in the number of uses of the channel (see e.g. [4] and [16]). The aim is to achieve reliable transmission, whilst optimizing the rate, i.e. the ratio of the size of the message to its corresponding codeword. The optimal rate of reliable transmission is referred to as the capacity of the channel.

Shannon, in his noisy channel coding theorem [21], obtained an explicit expression for the channel capacity of discrete, memoryless[3], classical channels. The first rigorous proof of this fundamental theorem was provided by Feinstein [7]. He used a packing argument (see e.g. [13]) to find a lower bound to the maximal number of codewords that can be sent through the channel reliably, i.e. with an arbitrarily low probability of error. More precisely, he proved that for any given $\delta > 0$, and sufficiently large number, n, of uses of a memoryless classical channel, the lower bound to the maximal number, N_n, of codewords that can be transmitted through the channel reliably is given by

$$N_n \geqslant 2^{n(H(X:Y)-\delta)}.$$

Here $H(X:Y)$ is the mutual information of the random variables X and Y, corresponding to the input and the output of the channel, respectively. We refer to this result as *Feinstein's fundamental lemma*, following Khinchin [13]. It implies that for a real number $R < C$, where $C = \max H(X:Y)$ (the maximum being taken over all possible input distributions), $M_n \leqslant 2^{nR}$ classical messages can be transmitted through the channel reliably. In other words, any rate $R < C$ is *achievable*.

For real world communication channels, the assumption that noise is uncorrelated between successive uses of a channel cannot be justified. Hence, memory effects need to be taken into account. This leads us to the consideration of quantum channels with memory. The first model of such a channel was studied by Macchiavello and Palma [15]. They showed that the transmission of classical information through two successive uses of a quantum depolarizing channel, with Markovian-correlated noise, is enhanced by using inputs entangled over the two uses. An important class of quantum channels with memory consists of the so-called *forgetful channels*. The channel studied in [15] falls in this class. Roughly speaking, a forgetful channel is one for which the output after a large number of successive uses does not depend on the initial input state. Forgetful channels have been studied by Bowen and Mancini [3] and more recently by Kretschmann and Werner [14]. In [14], coding theorems for arbitrary forgetful channels were proved. The proof of the direct channel coding theorem for a class of quantum channels with Markovian-correlated noise, where the underlying Markov chain was aperiodic and irreducible, was sketched out in [5]. Recently, Bjelaković and Boche [2] have proved a coding theorem for causal ergodic classical-quantum channels with decaying input memory.

The capacities of channels with long-term memory (i.e. channels which are 'not forgetful') had remained an open problem to date. In this paper, we evaluate the classical capacity of a class of quantum channels with long-term memory. These channels are convex combinations of memoryless channels. For a channel Φ in this class, $\Phi^{(n)} : \mathcal{B}(\mathcal{H}^{\otimes n}) \to \mathcal{B}(\mathcal{K}^{\otimes n})$ and the action of $\Phi^{(n)}$ on any state $\rho^{(n)} \in \mathcal{B}(\mathcal{H}^{\otimes n})$ is given as follows:

$$\Phi^{(n)}(\rho^{(n)}) = \sum_{i=1}^{M} \gamma_i \Phi_i^{\otimes n}(\rho^{(n)}). \tag{1}$$

where $\Phi_i : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ $(i = 1, \ldots, M)$ are completely positive, trace-preserving (CPT) maps and $\gamma_i > 0$, $\sum_{i=1}^{M} \gamma_i = 1$. Here, $\mathcal{H}$ and $\mathcal{K}$ denote Hilbert spaces. On using the channel, an initial random choice is made as to which memoryless channel the successive input states are transmitted through. A classical version of such a channel was introduced by Jacobs [12] and studied further by Ahlswede [1], who obtained an expression for its capacity which is analogous to the one we obtain in theorem 3.1 of section 3.

The memory of the class of channels that we study can be considered to be given by a Markov chain which is aperiodic but not irreducible. This can be seen as follows. A quantum

[3] For such a channel, the noise affecting successive input states is assumed to be perfectly uncorrelated.

channel (of length n) with Markovian-correlated noise is a CPT map $\Phi^{(n)} : \mathcal{B}(\mathcal{H}^{\otimes n}) \to \mathcal{B}(\mathcal{K}^{\otimes n})$ defined as follows:

$$\Phi^{(n)}(\rho^{(n)}) = \sum_{i_1,\ldots,i_n} q_{i_n|i_{n-1}} \ldots q_{i_2|i_1}\gamma_{i_1}\big(\Phi_{i_1} \otimes \cdots \otimes \Phi_{i_n}\big)(\rho^{(n)}),$$

Here (i) $q_{j|i}$ denote the elements of the transition matrix of a discrete-time Markov chain with a finite state space I, (ii) $\{\gamma_i\}$ denotes an invariant distribution of the chain and (iii) for each $i \in I$, $\Phi_i : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ is a CPT map. Casting our channel (defined by (1)) in this form yields $q_{j|i} = \delta_{ij}$. Hence the transition matrix of the Markov chain, in this case, is the identity matrix. Hence, once a particular branch, $i = 1, \ldots, M$, has been chosen, the successive inputs are sent through this branch. The Markov chain is therefore aperiodic, but not irreducible. We prove the coding theorem and weak converse for this class of channels. The main techniques that we employ are a quantum version [5] of Feinstein's fundamental lemma [7, 13] and a generalization of Helstrom's theorem [9]. For a quantum memoryless channel, our method yields an alternative proof of the Holevo–Schumacher–Westmoreland (HSW) theorem [11, 20], similar in spirit to the proof in [22]. Our results can be extended to quantum channels with arbitrary Markovian-correlated noise. The proofs in this case are technically more involved and will be presented in a subsequent paper.

We start the main body of our paper with some preliminaries in section 2. Our main result is stated in section 3. For clarity of exposition, we follow this with a proof of the quantum analogue of Feinstein's fundamental lemma for memoryless channels in section 4. The proof of our main result, for a class of quantum channels with long-term memory, is given in section 5.

2. Preliminaries

Let $\mathcal{B}(\mathcal{H})$ denote the algebra of linear operators acting on a finite-dimensional Hilbert space $\mathcal{H}$. The von Neumann entropy of a state ρ, i.e. a positive operator of unit trace in $\mathcal{B}(\mathcal{H})$, is defined as $S(\rho) = -\mathrm{Tr}\,\rho \log \rho$, where the logarithm is taken to base 2. A quantum channel is given by a CPT map $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$, where $\mathcal{H}$ and $\mathcal{K}$ are the input and output Hilbert spaces of the channel, respectively. Let $\dim \mathcal{H} = d$ and $\dim \mathcal{K} = d'$. For any ensemble $\{p_j, \rho_j\}$ of states ρ_j and probability distributions $\{p_j\}$, the Holevo χ quantity is defined as

$$\chi(\{p_j, \rho_j\}) := S\left(\sum_j p_j\rho_j\right) - \sum_j p_j S(\rho_j). \tag{2}$$

The Holevo capacity of a memoryless quantum channel Φ is given by

$$\chi^*(\Phi) := \max_{\{p_j,\rho_j\}} \chi(\{p_j, \Phi(\rho_j)\}), \tag{3}$$

where the maximum is taken over all ensembles $\{p_j, \rho_j\}$ of possible input states $\rho_j \in \mathcal{B}(\mathcal{H})$ occurring with probabilities p_j. It is known that the maximum in (3) can be achieved by using an ensemble of pure states, and that it suffices to restrict the maximum to ensembles of at most d^2 pure states.

Let us consider the transmission of classical information through successive uses of a quantum channel Φ. Let n uses of the channel be denoted by $\Phi^{(n)}$. Suppose Alice has a set of messages, labelled by the elements of the set $\mathcal{M}_n = \{1, 2, \ldots, M_n\}$, which she would like to communicate to Bob, using the quantum channel Φ. To do this, she encodes each message into a quantum state of a physical system with Hilbert space $\mathcal{H}^{\otimes n}$, which she then sends to Bob through n uses of the quantum channel. In order to infer the message that Alice communicated

to him, Bob makes a measurement (described by POVM elements) on the state that he receives. The encoding and decoding operations, employed to achieve reliable transmission of information through the channel, together define a quantum error-correcting code (QECC). More precisely, a code $\mathcal{C}^{(n)}$ of size N_n is given by a sequence $\{\rho_i^{(n)}, E_i^{(n)}\}_{i=1}^{N_n}$ where each $\rho_i^{(n)}$ is a state in $\mathcal{B}(\mathcal{H}^{\otimes n})$ and each $E_i^{(n)}$ is a positive operator acting in $\mathcal{K}^{\otimes n}$, such that $\sum_{i=1}^{N_n} E_i^{(n)} \leqslant I_n$. Here I_n denotes the identity operator in $\mathcal{B}(\mathcal{K}^{\otimes n})$. Defining $E_0^{(n)} = I_n - \sum_{i=1}^{N_n} E_i^{(n)}$ yields a resolution of identity in $\mathcal{K}^{\otimes n}$. Hence, $\{E_i^{(n)}\}_{i=0}^{N_n}$ defines a POVM. An output $i \geqslant 1$ would lead to the inference that the state (or codeword) $\rho_i^{(n)}$ was transmitted through the channel $\Phi^{(n)}$, whereas the output 0 is interpreted as a failure of any inference. The average probability of error for the code $\mathcal{C}^{(n)}$ is given by

$$P_e(\mathcal{C}^{(n)}) := \frac{1}{N_n} \sum_{i=1}^{N_n} \left(1 - \mathrm{Tr}\left(\Phi^{(n)}\left(\rho_i^{(n)}\right) E_i^{(n)}\right)\right). \tag{4}$$

If there exists $N \in \mathbf{N}$ such that for all $n \geqslant N$, there exists a sequence of codes $\{\mathcal{C}^{(n)}\}_{n=1}^{\infty}$, of sizes $N_n \geqslant 2^{nR}$, for which $P_e(\mathcal{C}^{(n)}) \to 0$ as $n \to \infty$, then R is said to be an *achievable* rate.

The capacity of Φ is defined as

$$C(\Phi) := \sup R, \tag{5}$$

where R is an achievable rate. If the codewords $\rho_i^{(n)}$, $i = 1, 2, \ldots, N_n$, are restricted to product states in $\mathcal{B}(\mathcal{H}^{\otimes n})$, the capacity $C(\Phi)$ is referred to as the *product state capacity*.

3. Main result

In this paper, we study a class of channels with long-term memory defined by (1). As explained in the introduction, the memory of a channel in this class can be considered to be given by a Markov chain which is aperiodic but not irreducible [17].

Our main result is given by the following theorem.

Theorem 3.1. *The product state capacity of a channel Φ, with long-term memory, defined through (1), is given by*

$$C(\Phi) = \sup_{\{p_j, \rho_j\}} \left[\bigwedge_{i=1}^{M} \chi_i(\{p_j, \rho_j\})\right],$$

where $\chi_i(\{p_j, \rho_j\}) := \chi(\{p_j, \Phi_i(\rho_j)\})$. The supremum is taken over all finite ensembles of states $\rho_j \in \mathcal{B}(\mathcal{H})$ with probabilities p_j.

Here we use the standard notation $\bigwedge$ to denote the minimum.

The product state capacity can be generalized to give the classical capacity of the channel Φ in the usual manner, that is, by considering inputs which are product states over uses of blocks of n channels, but which may be entangled across different uses within the same block. The classical capacity $C_{\mathrm{classical}}(\Phi)$ is obtained in the limit $n \to \infty$ and is given by

$$C_{\mathrm{classical}}(\Phi) = \lim_{n\to\infty} \frac{1}{n} C(\Phi^{(n)}). \tag{6}$$

4. Analogue of Feinstein's fundamental lemma for a memoryless quantum channel

In this section, we prove an analogue of Feinstein's fundamental lemma [7] for a memoryless quantum channel Φ. This is given by theorem 4.1. It provides a lower bound to the maximal number of codewords that can be reliably sent through Φ.

The proof of our main result, theorem 3.1, employs a theorem which is a generalization of theorem 4.1.

Theorem 4.1. *Let $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a memoryless quantum channel. Given $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$ there exist at least $N_n \geqslant 2^{n(\chi^*(\Phi)-\epsilon)}$ product states $\tilde{\rho}_1^{(n)}, \ldots, \tilde{\rho}_{N_n}^{(n)} \in \mathcal{B}(\mathcal{H}^{\otimes n})$ and positive operators $E_1^{(n)}, \ldots, E_{N_n}^{(n)} \in \mathcal{B}(\mathcal{K}^{\otimes n})$ such that $\sum_{k=1}^{N_n} E_k^{(n)} \leqslant I_n$ and*

$$\mathrm{Tr}\big[\Phi^{\otimes n}\big(\tilde{\rho}_k^{(n)}\big)E_k^{(n)}\big] > 1 - \epsilon, \tag{7}$$

for each k.

Here $\chi^*(\Phi)$ is the Holevo capacity (3) of the memoryless quantum channel Φ.

Before giving the proof of theorem 4.1, let us briefly sketch the idea behind it. The proof employs the idea of construction of a maximal code. For a given $\epsilon > 0$, starting with an empty code, the proof gives a prescription for successively adding codewords $\rho_j^{(n)}$ and corresponding POVM elements $E_j^{(n)}$, $j = 1, 2, \ldots$, such that

$$\varepsilon_j^{(n)} := 1 - \mathrm{Tr}\big(E_j^{(n)}\Phi^{(n)}\big(\rho_j^{(n)}\big)\big) \leqslant \epsilon. \tag{8}$$

Note that $\varepsilon_j^{(n)}$ is the probability of error in inferring the jth codeword. This is done until no more codewords can be added without violating condition (8). The resulting code is maximal. Let the size of this code be N_n. The proof ensures that the number N_n is large and provides a lower bound for it in terms of the Holevo capacity $\chi^*(\Phi)$.

Proof. Let the maximum in (3) be attained for an ensemble $\{p_j, \rho_j\}_{j=1}^J$. Denote $\sigma_j = \Phi(\rho_j)$, $\bar{\sigma} = \sum_{j=1}^J p_j\Phi(\rho_j)$ and $\bar{\sigma}_n = \bar{\sigma}^{\otimes n}$. Since $\bar{\sigma}_n$ is a product state, its eigenvalues and eigenvectors can be labelled by sequences $\underline{k} = (k_1, \ldots, k_n) \in J^n$. □

Choose $\delta > 0$. We will relate δ to ϵ at a later stage. There exists $n_1 \in \mathbb{N}$ such that for $n \geqslant n_1$, there is a typical subspace $\overline{T}_\epsilon^{(n)}$ of $\mathcal{K}^{\otimes n}$, with projection $\bar{P}_n$ such that if $\bar{\sigma}_n$ has a spectral decomposition

$$\bar{\sigma}_n = \sum_{\underline{k}} \bar{\lambda}_{\underline{k}}^{(n)} \big|\psi_{\underline{k}}^{(n)}\big\rangle\big\langle\psi_{\underline{k}}^{(n)}\big|, \tag{9}$$

then

$$\left|\frac{1}{n}\log\bar{\lambda}_{\underline{k}}^{(n)} + S(\bar{\sigma})\right| < \frac{\epsilon}{3} \tag{10}$$

for all $\underline{k}$ such that $\big|\psi_{\underline{k}}^{(n)}\big\rangle \in \overline{T}_\epsilon^{(n)}$ and

$$\mathrm{Tr}(\bar{P}_n\bar{\sigma}_n) > 1 - \delta^2. \tag{11}$$

Further define

$$\bar{S} = \sum_{j=1}^J p_j S(\sigma_j). \tag{12}$$

Lemma 4.1. *Given a sequence* $\underline{j} = (j_1, \ldots, j_n) \in J^n$, *let* $P_{\underline{j}}^{(n)}$ *be the projection onto the subspace of* $\mathcal{K}^{\otimes n}$ *spanned by the eigenvectors of* $\sigma_{\underline{j}}^{(n)} = \sigma_{j_1} \otimes \ldots \otimes \sigma_{j_n}$ *with eigenvalues* $\lambda_{\underline{j},\underline{k}}^{(n)} = \prod_{i=1}^n \lambda_{j_i,k_i}$ *such that*

$$\left| \frac{1}{n} \log \lambda_{\underline{j},\underline{k}}^{(n)} + \bar{S} \right| < \frac{\epsilon}{3}. \tag{13}$$

For any $\delta > 0$ *there exists* $n_2 \in \mathbb{N}$ *such that for* $n \geqslant n_2$,

$$\mathbb{E}\big(\mathrm{Tr}\big(\sigma_{\underline{j}}^{(n)} P_{\underline{j}}^{(n)}\big)\big) > 1 - \delta^2, \tag{14}$$

where $\mathbb{E}$ *denotes the expectation with respect to the probability distribution* $\big\{p_{\underline{j}}^{(n)}\big\}$ *on the states* $\rho_{\underline{j}}^{(n)}$.

Proof. Define i.i.d. random variables $X_1, \ldots, X_n$ with distribution given by

$$\mathrm{Prob}(X_i = \lambda_{j,k}) = p_j \lambda_{j,k}, \tag{15}$$

where $\lambda_{j,k}$, $k = 1, 2, \ldots, d'$, are the eigenvalues of σ_j. By the *weak law of large numbers*,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \log X_i \to \mathbb{E}(\log X_i) &= \sum_{j=1}^J \sum_{k=1}^{d'} p_j \lambda_{j,k} \log \lambda_{j,k} \\ &= -\sum_{j=1}^J p_j S(\sigma_j) = -\bar{S}. \end{aligned} \tag{16}$$

It follows that there exists n_2 such that for $n \geqslant n_2$, the typical set $T_\epsilon^{(n)}$ of sequences of pairs $((j_1, k_1), \ldots, (j_n, k_n))$ such that

$$\left| \frac{1}{n} \sum_{i=1}^n \log \lambda_{j_i,k_i} + \bar{S} \right| < \frac{\epsilon}{3} \tag{17}$$

satisfies

$$\mathbb{P}\big(T_\epsilon^{(n)}\big) = \sum_{((j_1,k_1),\ldots,(j_n,k_n)) \in T_\epsilon^{(n)}} \prod_{i=1}^n p_{j_i} \lambda_{j_i,k_i} > 1 - \delta^2. \tag{18}$$

Obviously,

$$P_{\underline{j}}^{(n)} \geqslant \sum_{\substack{\underline{k}=(k_1,\ldots,k_n) \\ ((j_1,k_1),\ldots,(j_n,k_n)) \in T_\epsilon^{(n)}}} \big|\psi_{\underline{j},\underline{k}}^{(n)}\big\rangle\big\langle\psi_{\underline{j},\underline{k}}^{(n)}\big| \tag{19}$$

and

$$\mathbb{E}\big(\mathrm{Tr}\big(\sigma_{\underline{j}}^{(n)} P_{\underline{j}}^{(n)}\big)\big) \geqslant \mathbb{P}\big(T_\epsilon^{(n)}\big) > 1 - \delta^2. \tag{20}$$

□

Continuing the proof of the theorem, let N_n be the maximal number N for which there exist product states $\tilde{\rho}_1^{(n)}, \ldots, \tilde{\rho}_N^{(n)}$ on $\mathcal{H}^{\otimes n}$ and positive operators $E_1^{(n)}, \ldots, E_N^{(n)}$ on $\mathcal{K}^{\otimes n}$ such that

(i) $\sum_{k=1}^{N_n} E_k^{(n)} \leqslant \bar{P}_n$,
(ii) $\mathrm{Tr}\big[\tilde{\sigma}_k^{(n)} E_k^{(n)}\big] > 1 - \epsilon$ and
(iii) $\mathrm{Tr}\big[\bar{\sigma}_n E_k^{(n)}\big] \leqslant 2^{-n[S(\bar{\sigma}) - \bar{S} - \frac{2}{3}\epsilon]}$.

Here $\tilde{\sigma}_k^{(n)} = \Phi^{\otimes n}\big(\tilde{\rho}_k^{(n)}\big)$.

For any given $\underline{j} \in J^n$, define

$$V_{\underline{j}}^{(n)} = \left(\bar{P}_n - \sum_{k=1}^{N_n} E_k^{(n)}\right)^{1/2} \bar{P}_n P_{\underline{j}}^{(n)} \bar{P}_n \left(\bar{P}_n - \sum_{k=1}^{N_n} E_k^{(n)}\right)^{1/2}. \tag{21}$$

Clearly, $V_{\underline{j}}^{(n)} \leqslant \bar{P}_n - \sum_{k=1}^{N_n} E_k^{(n)}$, and we also have

Lemma 4.2.

$$\operatorname{Tr}\bigl(\bar{\sigma}_n V_{\underline{j}}^{(n)}\bigr) \leqslant 2^{-n[S(\bar{\sigma})-\bar{S}-\frac{2}{3}\epsilon]}. \tag{22}$$

Proof. Put $Q_n = \sum_{k=1}^{N_n} E_k$. Note that Q_n commutes with $\bar{P}_n$. Using the fact that $\bar{P}_n\bar{\sigma}_n\bar{P}_n \leqslant 2^{-n[S(\bar{\sigma})-\frac{1}{3}\epsilon]}$ by (10), we have

$$\begin{aligned}
\operatorname{Tr}\bigl(\bar{\sigma}_n V_{\underline{j}}^{(n)}\bigr) &= \operatorname{Tr}\bigl[\bar{\sigma}_n(\bar{P}_n - Q_n)^{1/2}\bar{P}_n P_{\underline{j}}^{(n)}\bar{P}_n(\bar{P}_n - Q_n)^{1/2}\bigr] \\
&= \operatorname{Tr}\bigl[\bar{P}_n\bar{\sigma}_n\bar{P}_n(\bar{P}_n - Q_n)^{1/2}P_{\underline{j}}^{(n)}(\bar{P}_n - Q_n)^{1/2}\bigr] \\
&\leqslant 2^{-n[S(\bar{\sigma})-\frac{1}{3}\epsilon]}\operatorname{Tr}\bigl[(\bar{P}_n - Q_n)^{1/2}P_{\underline{j}}^{(n)}(\bar{P}_n - Q_n)^{1/2}\bigr] \\
&\leqslant 2^{-n[S(\bar{\sigma})-\frac{1}{3}\epsilon]}\operatorname{Tr}\bigl(P_{\underline{j}}^{(n)}\bigr) \leqslant 2^{-n[S(\bar{\sigma})-\bar{S}-\frac{2}{3}\epsilon]},
\end{aligned} \tag{23}$$

where, in the last inequality, we used the standard upper bound on the dimension of the typical subspace: $\operatorname{Tr}\bigl(P_{\underline{j}}^{(n)}\bigr) \leqslant 2^{n[\bar{S}+\frac{1}{3}\epsilon]}$, which follows from lemma 4.1. □

Since N_n is maximal, it now follows that

$$\operatorname{Tr}\bigl(\sigma_{\underline{j}}^{(n)} V_{\underline{j}}^{(n)}\bigr) \leqslant 1 - \epsilon, \tag{24}$$

and hence

Corollary 4.1.

$$\mathbb{E}\bigl(\operatorname{Tr}\bigl[\sigma_{\underline{j}}^{(n)} V_{\underline{j}}^{(n)}\bigr]\bigr) \leqslant 1 - \epsilon. \tag{25}$$

Lemma 4.3. *For all $\eta > 0$, there exists $n_3 \in \mathbb{N}$ such that for all $n \geqslant n_3$,*

$$\mathbb{E}\bigl(\operatorname{Tr}\bigl[\sigma_{\underline{j}}^{(n)}\bar{P}_n P_{\underline{j}}^{(n)}\bar{P}_n\bigr]\bigr) > 1 - \eta. \tag{26}$$

Proof. We write

$$\begin{aligned}
\mathbb{E}\bigl(\operatorname{Tr}\bigl[\sigma_{\underline{j}}^{(n)}\bar{P}_n P_{\underline{j}}^{(n)}\bar{P}_n\bigr]\bigr) &= \mathbb{E}\bigl(\operatorname{Tr}\bigl[\sigma_{\underline{j}}^{(n)} P_{\underline{j}}^{(n)}\bigr]\bigr) - \mathbb{E}\bigl(\operatorname{Tr}\bigl[\sigma_{\underline{j}}^{(n)}(I_n - \bar{P}_n)P_{\underline{j}}^{(n)}\bigr]\bigr) \\
&\quad - \mathbb{E}\bigl(\operatorname{Tr}\bigl[\sigma_{\underline{j}}^{(n)}\bar{P}_n P_{\underline{j}}^{(n)}(I_n - \bar{P}_n)\bigr]\bigr).
\end{aligned} \tag{27}$$

By lemma 4.1, the first term is $> 1 - \delta^2$ provided $n \geqslant n_2$. The last two terms can be bounded using the Cauchy–Schwarz inequality as follows:

$$\begin{aligned}
\mathbb{E}\bigl(\operatorname{Tr}\bigl[\sigma_{\underline{j}}^{(n)}(I_n - \bar{P}_n)P_{\underline{j}}^{(n)}\bigr]\bigr) &= \mathbb{E}\bigl(\operatorname{Tr}\bigl[\bigl(\sigma_{\underline{j}}^{(n)}\bigr)^{1/2}(I_n - \bar{P}_n)P_{\underline{j}}^{(n)}\bigl(\sigma_{\underline{j}}^{(n)}\bigr)^{1/2}\bigr]\bigr) \\
&\leqslant \bigl\{\mathbb{E}\bigl(\operatorname{Tr}\bigl[(I_n - \bar{P}_n)\sigma_{\underline{j}}^{(n)}(I_n - \bar{P}_n)\bigr]\bigr)\bigr\}^{1/2}\bigl\{\mathbb{E}\bigl(\operatorname{Tr}\bigl[\bigl(\sigma_{\underline{j}}^{(n)}\bigr)^{1/2}P_{\underline{j}}^{(n)}\bigl(\sigma_{\underline{j}}^{(n)}\bigr)^{1/2}\bigr]\bigr)\bigr\}^{1/2} \\
&= \bigl\{\mathbb{E}\bigl(\operatorname{Tr}\bigl[\sigma_{\underline{j}}^{(n)}(I_n - \bar{P}_n)\bigr]\bigr)\bigr\}^{1/2}\bigl\{\mathbb{E}\bigl(\operatorname{Tr}\bigl[\sigma_{\underline{j}}^{(n)}P_{\underline{j}}^{(n)}\bigr]\bigr)\bigr\}^{1/2} \\
&\leqslant \bigl\{\mathbb{E}\bigl(\operatorname{Tr}\bigl[\sigma_{\underline{j}}^{(n)}(I_n - \bar{P}_n)\bigr]\bigr)\bigr\}^{1/2} \\
&= \bigl(\operatorname{Tr}\bigl[\bar{\sigma}_n(I_n - \bar{P}_n)\bigr]\bigr)^{1/2} \leqslant \delta
\end{aligned} \tag{28}$$

by (11) provided $n \geqslant n_1$. Analogously,

$$\mathbb{E}\big(\operatorname{Tr}\big[\sigma_{\underline{j}}^{(n)}\bar{P}_n P_{\underline{j}}^{(n)}(I_n-\bar{P}_n)\big]\big)\leqslant\delta. \tag{29}$$

Choosing $n_3 = n_1 \vee n_2$ and $\delta^2+2\delta<\eta$, the result follows. □

Lemma 4.4. *Assume $\eta<\frac{1}{3}\epsilon$. Then for $n\geqslant n_3$,*

$$\operatorname{Tr}\left[\bar{\sigma}_n\sum_{k=1}^{N}E_k^{(n)}\right]=\mathbb{E}\left(\operatorname{Tr}\left[\sigma_{\underline{j}}^{(n)}\sum_{k=1}^{N}E_k^{(n)}\right]\right)\geqslant\eta^2. \tag{30}$$

Proof. Define

$$Q_n'=\bar{P}_n-(\bar{P}_n-Q_n)^{1/2}. \tag{31}$$

By the above corollary,

$$\begin{aligned}1-\epsilon&\geqslant\mathbb{E}\big\{\operatorname{Tr}\big(\sigma_{\underline{j}}^{(n)}(\bar{P}_n-Q_n')P_{\underline{j}}^{(n)}(\bar{P}_n-Q_n')\big)\big\}\\&=\mathbb{E}\big\{\operatorname{Tr}\big(\sigma_{\underline{j}}^{(n)}\bar{P}_nP_{\underline{j}}^{(n)}\bar{P}_n\big)\big\}-\mathbb{E}\big\{\operatorname{Tr}\big(\sigma_{\underline{j}}^{(n)}Q_n'P_{\underline{j}}^{(n)}\bar{P}_n\big)+\operatorname{Tr}\big(\sigma_{\underline{j}}^{(n)}\bar{P}_nP_{\underline{j}}^{(n)}Q_n'\big)\big\}\\&\quad+\mathbb{E}\big\{\operatorname{Tr}\big(\sigma_{\underline{j}}^{(n)}Q_n'P_{\underline{j}}^{(n)}Q_n'\big)\big\}.\end{aligned} \tag{32}$$

Since the last term is positive, we have, by lemma 4.3,

$$\mathbb{E}\big\{\operatorname{Tr}\big(\sigma_{\underline{j}}^{(n)}Q_n'P_{\underline{j}}^{(n)}\bar{P}_n\big)+\operatorname{Tr}\big(\sigma_{\underline{j}}^{(n)}\bar{P}_nP_{\underline{j}}^{(n)}Q_n'\big)\big\}\geqslant\epsilon-\eta>2\eta. \tag{33}$$

On the other hand, using Cauchy–Schwarz for each term, we have

$$\begin{aligned}&\mathbb{E}\big\{\operatorname{Tr}\big(\sigma_{\underline{j}}^{(n)}Q_n'P_{\underline{j}}^{(n)}\bar{P}_n\big)+\operatorname{Tr}\big(\sigma_{\underline{j}}^{(n)}\bar{P}_nP_{\underline{j}}^{(n)}Q_n'\big)\big\}\\&\quad\leqslant2\big\{\mathbb{E}\big[\operatorname{Tr}\big(Q_n'\sigma_{\underline{j}}^{(n)}Q_n'\big)\big]\big\}^{1/2}\big\{\mathbb{E}\big[\operatorname{Tr}\big(\sigma_{\underline{j}}^{(n)}\bar{P}_nP_{\underline{j}}^{(n)}\bar{P}_n\big)\big]\big\}^{1/2}\\&\quad\leqslant2\big\{\mathbb{E}\big[\operatorname{Tr}\big(\sigma_{\underline{j}}^{(n)}Q_n'^2\big)\big]\big\}^{1/2}.\end{aligned} \tag{34}$$

Thus,

$$\mathbb{E}\big[\operatorname{Tr}\big(\sigma_{\underline{j}}^{(n)}Q_n'^2\big)\big]\geqslant\eta^2. \tag{35}$$

To complete the proof, we now claim that

$$Q_n\geqslant(Q_n')^2. \tag{36}$$

Indeed, on the domain of $\bar{P}_n$, (36) follows from the inequality $1-(1-x)^2\geqslant x^2$ for $0\leqslant x\leqslant 1$. □

To complete the proof of theorem 4.1, we now have by assumption,

$$\operatorname{Tr}\big[\bar{\sigma}_nE_k^{(n)}\big]\leqslant2^{-n[S(\bar{\sigma})-\bar{S}-\frac{2}{3}\epsilon]} \tag{37}$$

for all $k=1,\ldots,N_n$. On the other hand, choosing $\eta<\frac{1}{3}\epsilon$ and $\delta<\frac{1}{3}\eta$, we have by lemma 4.4,

$$\operatorname{Tr}\left[\bar{\sigma}_n\sum_{k=1}^{N_n}E_k^{(n)}\right]\geqslant\eta^2 \tag{38}$$

provided $n\geqslant n_3$. It follows that

$$N_n\geqslant\eta^2 2^{n[S(\bar{\sigma})-\bar{S}-\frac{2}{3}\epsilon]}\geqslant2^{n[S(\bar{\sigma})-\bar{S}-\epsilon]} \tag{39}$$

for $n\geqslant n_3$ and $n\geqslant-\frac{6}{\epsilon}\log\eta$.

5. A class of channels with long-term memory

We now consider the class of quantum channels with long-term memory, mentioned in the introduction:

$$\Phi^{(n)}(\rho^{(n)}) = \sum_{i=1}^{M} \gamma_i \Phi_i^{\otimes n}(\rho^{(n)}), \tag{40}$$

where $\Phi_i : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ $(i = 1, \ldots, M)$ are CPT maps and $\gamma_i > 0$, $\sum_{i=1}^{M} \gamma_i = 1$.

For an ensemble of states $\{p_j, \rho_j\}$ where $\rho_j \in \mathcal{B}(\mathcal{H})$, define

$$\widehat{\chi}(\{p_j, \rho_j\}) := \bigwedge_{i=1}^{M} \chi_i(\{p_j, \rho_j\}), \tag{41}$$

where $\chi_i(\{p_j, \rho_j\}) = \chi\left(\{p_j, \Phi_i(\rho_j)\}\right)$.

5.1. Proof of the direct part of theorem 3.1

To prove the direct part of theorem 3.1, i.e. the fact that a rate $R < C(\Phi)$ is achievable, we employ the quantum analogue of Feinstein's fundamental lemma for the class of channels defined by (40). This analogue is given by the following theorem, which we prove in section 5.1.1.

Theorem 5.1. *Given $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$ there exist at least $N_n \geqslant 2^{n(C(\Phi)-\epsilon)}$ product states $\rho_1^{(n)}, \ldots, \rho_{N_n}^{(n)} \in \mathcal{B}(\mathcal{H}^{\otimes n})$ and positive operators $E_1^{(n)}, \ldots, E_{N_n}^{(n)} \in \mathcal{B}(\mathcal{K}^{\otimes n})$ such that $\sum_{k=1}^{N_n} E_k^{(n)} \leqslant I_n$ and such that for each $k = 1, \ldots, N_n$,*

$$\mathrm{Tr}\left[\Phi^{(n)}\left(\rho_k^{(n)}\right) E_k^{(n)}\right] > 1 - \epsilon. \tag{42}$$

Here

$$C(\Phi) := \sup_{\{p_j, \rho_j\}} \left[\bigwedge_{i=1}^{M} \chi_i(\{p_j, \rho_j\})\right] = \sup_{\{p_j, \rho_j\}} \widehat{\chi}(\{p_j, \rho_j\}), \tag{43}$$

where the supremum is over all finite ensembles of states ρ_j with probabilities p_j.

The above theorem implies that a rate $R < C(\Phi)$ is achievable. This can be seen as follows. Given $R < C(\Phi)$, choose $\epsilon > 0$ such that $R < C(\Phi) - \epsilon$. Then, theorem 5.1 guarantees the existence of codes $\mathcal{C}^{(n)}$ of size

$$N_n \geqslant 2^{n(C(\Phi)-\epsilon)} \geqslant 2^{nR},$$

with codewords given by product states $\rho_j^{(n)}$, and POVM elements $E_j^{(n)}$, for which the probability of error, $\varepsilon_j^{(n)}$, can be made arbitrarily small, for each $j \in \{1, 2, \ldots, N_n\}$ and n large enough. Hence, the rate R is achievable.

5.1.1. Proof of theorem 5.1. Choose an ensemble $\{p_j, \rho_j\}_{j=1}^{J}$ such that

$$C(\Phi) < \widehat{\chi}(\{p_j, \rho_j\}) + \tfrac{1}{4}\epsilon. \tag{44}$$

Define $\sigma_{i,j} = \Phi_i(\rho_j)$, $\sigma_{i,\underline{j}}^{(n)} = \otimes_{r=1}^{n} \sigma_{i,j_r}$, $\bar{\sigma}_i = \sum_{j=1}^{J} p_j \Phi_i(\rho_j) = \Phi_i(\bar{\rho})$ and $\bar{\sigma}_i^{(n)} = \bar{\sigma}_i^{\otimes n}$. Let $\bar{P}_i^{(n)}$, $i = 1, \ldots, M$, be the orthogonal projections onto the typical subspaces for the states $\bar{\sigma}_i^{(n)}$ so that, as above,

$$\mathrm{Tr}\left(\bar{P}_i^{(n)} \bar{\sigma}_i^{(n)}\right) > 1 - \delta^2 \tag{45}$$

for n large enough, and

$$\bar{P}_i^{(n)}\bar{\sigma}_i^{(n)}\bar{P}_i^{(n)} \leqslant 2^{-n[S(\bar{\sigma}_i)-\frac{1}{4}\epsilon]}. \tag{46}$$

By lemma 4.1 there also exist typical subspaces with projections $P_{i,\underline{j}}^{(n)}$ for which

$$\mathbb{E}\big(\mathrm{Tr}\big(\sigma_{i,\underline{j}}^{(n)}P_{i,\underline{j}}^{(n)}\big)\big) > 1-\delta^2 \tag{47}$$

for n large enough.

To distinguish between the different memoryless branches, Φ_i, of the quantum channel Φ, we add a preamble to the input state encoding each message in the set $\mathcal{M}_n$. This is given by an m-fold tensor product of a suitable state (as described below). Let us first sketch the idea behind adding such a preamble. Helstrom [9] showed that two states σ_1 and σ_2, occurring with *a priori* probabilities γ_1 and γ_2 respectively, can be distinguished, with an asymptotically vanishing probability of error, if a suitable collective measurement is performed on the m-fold tensor products $\sigma_1^{\otimes m}$ and $\sigma_2^{\otimes m}$, for a large enough $m \in \mathbb{N}$. The optimal measurement is projection valued. The relevant projection operators, which we denote by Π^+ and Π^-, are the orthogonal projections onto the positive and negative eigenspaces of the difference operator $A_m = \gamma_1\sigma_1^{\otimes m} - \gamma_2\sigma_2^{\otimes m}$. Here we generalize this result to distinguish between the different branches Φ_i. If the preamble is given by a state $\omega^{\otimes m}$, then, by using Helstrom's result, we can construct a POVM which distinguishes between the output states $\sigma_i^{\otimes n} := (\Phi_i(\omega))^{\otimes n}$ corresponding to the different branches $\Phi_i, i = 1, 2, \ldots, M$. The outcome of this POVM measurement would in turn serve to determine which branch of the channel is being used for transmission.

Note that we may assume that all branches Φ_i are different. Indeed, otherwise we do not need to distinguish them and can introduce a compound probability for each set of identical branches. This assumption means that there exist states $\omega_{i,j}$ on $\mathcal{H}$ for each pair $1 \leqslant i < j \leqslant M$ such that $\Phi_i(\omega_{i,j}) \neq \Phi_j(\omega_{i,j})$. Introducing the fidelity of two states as in [16],

$$F(\sigma,\sigma') = \mathrm{Tr}\sqrt{\sigma^{1/2}\sigma'\sigma^{1/2}}, \tag{48}$$

we then have

$$F(\Phi_i(\omega_{i,j}),\Phi_j(\omega_{i,j})) \leqslant f < 1 \tag{49}$$

for all pairs (i, j). We now introduce, for any $m \in \mathbb{N}$ and $1 \leqslant i < j \leqslant M$, the difference operators

$$A_{i,j}^{(m)} = \gamma_i(\Phi_i(\omega_{i,j}))^{\otimes m} - \gamma_j(\Phi_j(\omega_{i,j}))^{\otimes m}. \tag{50}$$

Let $\Pi_{i,j}^{\pm}$ be the orthogonal projections onto the eigenspaces of $A_{i,j}^{(m)}$ corresponding to all non-negative, and all negative eigenvalues, respectively.

Lemma 5.1. *Suppose that for a given $\delta > 0$,*

$$|\,\mathrm{Tr}\big[|A_{i,j}^{(m)}|\big] - (\gamma_i+\gamma_j)| \leqslant \delta. \tag{51}$$

Then

$$\big|\,\mathrm{Tr}\big[\Pi_{i,j}^{+}\big(\Phi_i(\omega_{i,j})\big)^{\otimes m}\big] - 1\big| \leqslant \frac{\delta}{2\gamma_i} \tag{52}$$

and

$$\big|\,\mathrm{Tr}\big[\Pi_{i,j}^{-}\big(\Phi_j(\omega_{i,j})\big)^{\otimes m}\big] - 1\big| \leqslant \frac{\delta}{2\gamma_j}. \tag{53}$$

Proof. Write $A = A_{i,j}^{(m)}$ and $\Pi^{\pm} = \Pi_{i,j}^{\pm}$. First note that

$$\begin{aligned}\operatorname{Tr}[\Pi^{\pm}A] &= \tfrac{1}{2}\operatorname{Tr}[A \pm (\Pi^{+} - \Pi^{-})A] \\ &= \tfrac{1}{2}\left(\operatorname{Tr}[A] \pm \operatorname{Tr}[|A|]\right) \\ &= \tfrac{1}{2}(\gamma_i - \gamma_j) \pm \tfrac{1}{2}\operatorname{Tr}[|A|] \end{aligned} \tag{54}$$

so that we have by the assumption

$$|\operatorname{Tr}[\Pi^{+}A] - \gamma_i| \leqslant \tfrac{1}{2}\delta \tag{55}$$

and

$$|\operatorname{Tr}[\Pi^{-}A] + \gamma_j| \leqslant \tfrac{1}{2}\delta. \tag{56}$$

Now writing $\sigma_i = (\Phi_i(\omega_{i,j}))^{\otimes m}$ and $\sigma_j = (\Phi_j(\omega_{i,j}))^{\otimes m}$, we obviously have $\operatorname{Tr}[\Pi^{-}\sigma_i] \geqslant 0$, and on the other hand,

$$\gamma_i \operatorname{Tr}[\Pi^{-}\sigma_i] = \operatorname{Tr}[\Pi^{-}A] + \gamma_j \operatorname{Tr}[\Pi^{-}\sigma_j] \leqslant -\gamma_j + \tfrac{1}{2}\delta + \gamma_j = \tfrac{1}{2}\delta. \tag{57}$$

The first result thus follows from $\Pi^{+} + \Pi^{-} = I_m$ and $\operatorname{Tr}\sigma_i = 1$. Similarly,

$$\gamma_j \operatorname{Tr}[\Pi^{+}\sigma_j] = -\operatorname{Tr}[\Pi^{+}A] + \gamma_i \operatorname{Tr}[\Pi^{+}\sigma_i] \leqslant -\gamma_i + \tfrac{1}{2}\delta + \gamma_i = \tfrac{1}{2}\delta. \tag{58}$$

□

To compare the outputs of all the different branches of the channel, we define projections $\tilde{\Pi}_i$ on the tensor product space $\bigotimes_{1\leqslant i<j\leqslant M} \mathcal{K}^{\otimes m} = \mathcal{K}^{\otimes mL}$ with $L = \binom{M}{2}$ as follows:

$$\tilde{\Pi}_i = \bigotimes_{1\leqslant i_1<i_2\leqslant M} \Gamma_{i_1,i_2}^{(i)}, \quad \text{where } \Gamma_{i_1,i_2}^{(i)} = \begin{cases} I_m & \text{if } i_1 \neq i \text{ and } i_2 \neq i \\ \Pi_{i_1,i}^{-} & \text{if } i_2 = i \\ \Pi_{i,i_2}^{+} & \text{if } i_1 = i. \end{cases} \tag{59}$$

Note that it follows from the fact that $\Pi_{i,j}^{+}\Pi_{i,j}^{-} = 0$ that the projections $\tilde{\Pi}_i$ are also disjoint:

$$\tilde{\Pi}_i\tilde{\Pi}_j = 0 \qquad \text{for} \quad i \neq j. \tag{60}$$

Introducing the notation

$$\omega^{(mL)} = \bigotimes_{i_1<i_2} \omega_{i_1,i_2}^{\otimes m}, \tag{61}$$

we now have

Lemma 5.2. *For all $i = 1, \ldots, M$,*

$$\lim_{m\to\infty} \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] = 1. \tag{62}$$

Proof. Note that for all $i < j$,

$$F(\gamma_i\Phi_i(\omega_{i,j})^{\otimes m}, \gamma_j\Phi_j(\omega_{i,j})^{\otimes m}) = \sqrt{\gamma_i\gamma_j}F(\Phi_i(\omega_{i,j}), \Phi_i(\omega_{i,j}))^m < f^m. \tag{63}$$

Using the inequalities [16]

$$\operatorname{Tr}(A_1) + \operatorname{Tr}(A_2) - 2F(A_1, A_2) \leqslant \|A_1 - A_2\|_1 \leqslant \operatorname{Tr}(A_1) + \operatorname{Tr}(A_2) \tag{64}$$

for any two positive operators A_1 and A_2, we find that

$$\left|\operatorname{Tr}\left(\left|A_{i,j}^{(m)}\right|\right) - (\gamma_i + \gamma_j)\right| \leqslant 2f^m, \tag{65}$$

since

$$\operatorname{Tr}\left(\left|A_{i,j}^{(m)}\right|\right) = \|\gamma_i\Phi_i(\omega_{i,j})^{\otimes m} - \gamma_j\Phi_j(\omega_{i,j})^{\otimes m}\|_1. \tag{66}$$

Using lemma 5.1, we then have

$$\begin{aligned} 1 &\geqslant \mathrm{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}\left(\bigotimes_{i_1<i_2} \omega_{i_1,i_2}^{\otimes m}\right)\right] \\ &= \prod_{i_1<i} \mathrm{Tr}\left[\Pi_{i_1,i}^{-}\left(\Phi_i(\omega_{i_1,i})\right)^{\otimes m}\right] \prod_{i_2>i} \mathrm{Tr}\left[\Pi_{i,i_2}^{+}\left(\Phi_i(\omega_{i,i_2})\right)^{\otimes m}\right] \\ &\geqslant \left(1-\frac{f^m}{\gamma_i}\right)^{M-1}. \end{aligned} \tag{67}$$

□

We now fix m so large that

$$\mathrm{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] > 1-\delta \tag{68}$$

for all $i = 1, \ldots, M$. The product state $\omega^{(mL)}$, defined through (61), is used as a preamble to the input state encoding each message and serves to distinguish between the different branches, Φ_i, $i = 1, 2, \ldots, M$, of the channel. If $\rho_k^{(n)} \in \mathcal{B}(\mathcal{H}^{\otimes n})$ is a product state encoding the kth classical message in the set $\mathcal{M}_n$, then the kth codeword is given by the product state

$$\omega^{(mL)} \otimes \rho_k^{(n)}.$$

Continuing with the proof of theorem 5.1, let $N = \tilde{N}(n)$ be the maximal number of product states $\tilde{\rho}_1^{(n)}, \ldots, \tilde{\rho}_N^{(n)}$ on $\mathcal{H}^{\otimes n}$ (each of which is a tensor product of states in the maximizing ensemble $\{p_j, \rho_j\}_{j=1}^J$) for which there exist positive operators $E_1^{(n)}, \ldots, E_N^{(n)}$ on $\mathcal{K}^{\otimes mL} \otimes \mathcal{K}^{\otimes n}$ such that

(i) $E_k^{(n)} = \sum_{i=1}^M \tilde{\Pi}_i \otimes E_{k,i}^{(n)}$ and $\sum_{k=1}^N E_{k,i}^{(n)} \leqslant \bar{P}_i^{(n)}$,
(ii) $\sum_{i=1}^M \gamma_i \,\mathrm{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathrm{Tr}\left[\Phi_i^{\otimes n}\left(\tilde{\rho}_k^{(n)}\right) E_{k,i}^{(n)}\right] > 1-\epsilon$ and
(iii) $\sum_{i=1}^M \gamma_i \,\mathrm{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathrm{Tr}\left[\left(\Phi_i\left(\bar{\rho}\right)\right)^{\otimes n} E_{k,i}^{(n)}\right] \leqslant 2^{-n[C(\Phi)-\frac{3}{4}\epsilon]}$.

for $\bar{\rho} = \sum_{j=1}^J p_j \rho_j$. For each $i = 1, \ldots, M$ and $\underline{j} = (j_1, \ldots, j_n) \in J^n$, we define, as before

$$V_{i,\underline{j}}^{(n)} = \left(\bar{P}_i^{(n)} - \sum_{k=1}^N E_{k,i}^{(n)}\right)^{1/2} \bar{P}_i^{(n)} P_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)} \left(\bar{P}_i^{(n)} - \sum_{k=1}^N E_{k,i}^{(n)}\right)^{1/2}. \tag{69}$$

Clearly $V_{i,\underline{j}}^{(n)} \leqslant \bar{P}_i^{(n)} - \sum_{k=1}^N E_{k,i}^{(n)}$. Put

$$V_{\underline{j}}^{(n)} := \sum_{i=1}^M \tilde{\Pi}_i \otimes V_{i,\underline{j}}^{(n)}. \tag{70}$$

This is a candidate for an additional measurement operator, $E_{N+1}^{(n)}$, for Bob with a corresponding input state $\tilde{\rho}_{N+1}^{(n)} = \rho_{\underline{j}}^{(n)} = \rho_{j_1} \otimes \rho_{j_2} \ldots \otimes \rho_{j_n}$. Clearly, condition (i), given above, is satisfied and we also have

Lemma 5.3.

$$\sum_{i=1}^M \gamma_i \,\mathrm{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathrm{Tr}\left[\bar{\sigma}_i^{(n)} V_{i,\underline{j}}^{(n)}\right] \leqslant 2^{-n[C(\Phi)-\frac{3}{4}\epsilon]}, \tag{71}$$

where $\bar{\sigma}_i^{(n)} = (\Phi_i(\bar{\rho}))^{\otimes n}$.

Proof. By lemma 4.2, replacing $\frac{1}{3}\epsilon$ by $\frac{1}{4}\epsilon$ in the definition of the typical subspaces, we have

$$\operatorname{Tr}\left(\bar{\sigma}_i^{(n)} V_{i,\underline{j}}^{(n)}\right) \leqslant 2^{-n[S(\bar{\sigma}_i)-\bar{S}_i-\frac{1}{2}\epsilon]} = 2^{-n[\chi_i-\frac{1}{2}\epsilon]} \tag{72}$$

for n large enough. Then,

$$\begin{aligned}\sum_{i=1}^{M} \gamma_i \operatorname{Tr}[\tilde{\Pi}_i \Phi^{\otimes mL}(\omega^{(mL)})] \operatorname{Tr}\left[\bar{\sigma}_i^{(n)} V_{i,\underline{j}}^{(n)}\right] &\leqslant \sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\bar{\sigma}_i^{(n)} V_{i,\underline{j}}^{(n)}\right] \\ &\leqslant \sum_{i=1}^{M} \gamma_i 2^{-n[S(\bar{\sigma}_i)-\bar{S}_i-\frac{1}{2}\epsilon]} \\ &\leqslant 2^{-n[\widehat{\chi}(\Phi)-\frac{1}{2}\epsilon]}, \\ &\leqslant 2^{-n[C(\Phi)-\frac{3}{4}\epsilon]}, \end{aligned} \tag{73}$$

where we used the obvious fact that $\operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}\left(\omega_i^{(mL)}\right)\right] \leqslant 1$. □

By maximality of N it now follows that condition (ii) cannot hold, that is,

$$\sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \operatorname{Tr}\left[\Phi_i^{\otimes n}\left(\rho_{\underline{j}}^{(n)}\right) V_{i,\underline{j}}^{(n)}\right] \leqslant 1-\epsilon \tag{74}$$

for every $\underline{j}$, and this yields the following.

Corollary 5.1.

$$\sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left(\operatorname{Tr}\left[\Phi_i^{\otimes n}\left(\rho_{\underline{j}}^{(n)}\right) V_{i,\underline{j}}^{(n)}\right]\right) \leqslant 1-\epsilon. \tag{75}$$

We also need the following lemma.

Lemma 5.4. *For all $\eta' > \delta^2 + 3\delta$,*

$$\sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \operatorname{Tr}\left[\sigma_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)} P_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)}\right] > 1-\eta' \tag{76}$$

if n is large enough.

Proof. Using lemma 4.3 and (68), we have

$$\sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left(\operatorname{Tr}\left[\sigma_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)} P_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)}\right]\right) > (1-\delta)(1-\eta), \tag{77}$$

provided $\eta > \delta^2 + 2\delta$. Hence, the result follows. □

Lemma 5.5. *Assume $\eta' < \frac{1}{3}\epsilon$ and write*

$$Q_i^{(n)} = \sum_{k=1}^{N} E_{k,i}^{(n)}. \tag{78}$$

Then for n large enough,

$$\sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left(\operatorname{Tr}\left[\Phi_i^{(n)}\left(\rho_{\underline{j}}^{(n)}\right) Q_i^{(n)}\right]\right) \geqslant \eta'^2. \tag{79}$$

Proof. This is analogous to lemma 4.4. Define

$$Q_i^{(n)\prime} = \bar{P}_i^{(n)} - \left(\bar{P}_i^{(n)} - Q_i^{(n)}\right)^{1/2}. \tag{80}$$

By corollary 5.1,

$$\begin{aligned}
1-\epsilon \geqslant \sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left(\operatorname{Tr}\left[\Phi_i^{\otimes n}\left(\rho_{\underline{j}}^{(n)}\right) V_{i,\underline{j}}^{(n)}\right]\right) \\
= \sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left\{\operatorname{Tr}\left(\sigma_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)} P_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)}\right)\right\} \\
- \sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left\{\operatorname{Tr}\left(\sigma_{i,\underline{j}}^{(n)} Q_i^{(n)\prime} P_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)}\right)\right. \\
\left. + \operatorname{Tr}\left(\sigma_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)} P_{i,\underline{j}}^{(n)} Q_i^{(n)\prime}\right)\right\} \\
+ \sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left\{\operatorname{Tr}\left(\sigma_{i,\underline{j}}^{(n)} Q_i^{(n)\prime} P_{i,\underline{j}}^{(n)} Q_i^{(n)\prime}\right)\right\}.
\end{aligned} \tag{81}$$

Since the last term is positive, we have, by lemma 5.4,

$$\begin{aligned}
\sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left\{\operatorname{Tr}\left(\sigma_{i,\underline{j}}^{(n)} Q_i^{(n)\prime} P_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)}\right) + \operatorname{Tr}\left(\sigma_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)} P_{i,\underline{j}}^{(n)} Q_i^{(n)\prime}\right)\right\} \\
\geqslant \epsilon - \eta' > 2\eta'.
\end{aligned} \tag{82}$$

On the other hand, using the Cauchy–Schwarz inequality for each term, we have

$$\begin{aligned}
\sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left\{\operatorname{Tr}\left(\sigma_{i,\underline{j}}^{(n)} Q_i^{(n)\prime} P_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)}\right) + \operatorname{Tr}\left(\sigma_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)} P_{i,\underline{j}}^{(n)} Q_i^{(n)\prime}\right)\right\} \\
\leqslant 2\left\{\sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left[\operatorname{Tr}\left(\sigma_{i,\underline{j}}^{(n)} \left(Q_i^{(n)\prime}\right)^2\right)\right]\right\}^{1/2} \\
\times \left\{\sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left[\operatorname{Tr}\left(\sigma_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)} P_{i,\underline{j}}^{(n)} \bar{P}_i^{(n)}\right)\right]\right\}^{1/2} \\
\leqslant 2\left\{\sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left[\operatorname{Tr}\left(\sigma_{i,\underline{j}}^{(n)} \left(Q_i^{(n)\prime}\right)^2\right)\right]\right\}^{1/2}.
\end{aligned} \tag{83}$$

Thus,

$$\sum_{i=1}^{M} \gamma_i \operatorname{Tr}\left[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\right] \mathbb{E}\left[\operatorname{Tr}\left(\sigma_{i,\underline{j}}^{(n)} \left(Q_i^{(n)\prime}\right)^2\right)\right] \geqslant \eta'^2. \tag{84}$$

To complete the proof, we remark as before that

$$Q_i^{(n)} \geqslant \left(Q_i^{(n)\prime}\right)^2. \tag{85}$$

□

It now follows, as before, that for n large enough, $\tilde{N}(n) \geqslant (\eta')^2 2^{n[C(\Phi)-\frac{3}{4}\epsilon]}$. We take the following states as codewords:

$$\rho_k^{(mL+n)} = \omega^{(mL)} \otimes \tilde{\rho}_k^{(n)}. \tag{86}$$

For n sufficiently large, we then have

$$N_{n+mL} = \tilde{N}(n) \geqslant (\eta')^2 2^{n[C(\Phi)-\frac{3}{4}\epsilon]} \geqslant 2^{(mL+n)[C(\Phi)-\epsilon]}. \tag{87}$$

To complete the proof, we need to show that the set $E_k^{(n)}$ satisfies (42). But this follows immediately from condition (ii):

$$\begin{aligned}
\operatorname{Tr}\big[\Phi^{(mL+n)}\big(\rho_k^{(mL+n)}\big)E_k^{(n)}\big] &= \sum_{i=1}^{M} \gamma_i \operatorname{Tr}\big[\Phi_i^{\otimes(mL+n)}\big(\omega^{(mL)} \otimes \tilde{\rho}_k^{(n)}\big)E_k^{(n)}\big] \\
&= \sum_{i,j=1}^{M} \gamma_i \operatorname{Tr}\big[\tilde{\Pi}_j \Phi_i^{\otimes mL}(\omega^{(mL)})\big] \operatorname{Tr}\big[\Phi_i^{\otimes n}\big(\tilde{\rho}_k^{(n)}\big)E_{k,j}^{(n)}\big] \\
&\geqslant \sum_{i=1}^{M} \gamma_i \operatorname{Tr}\big[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})\big] \operatorname{Tr}\big[\Phi_i^{\otimes n}\big(\tilde{\rho}_k^{(n)}\big)E_{k,i}^{(n)}\big] > 1-\epsilon.
\end{aligned} \tag{88}$$

5.2. Proof of the converse of theorem 3.1

In this section, we prove that it is impossible for Alice to transmit classical messages reliably to Bob through the channel Φ defined in (40) at a rate $R > C(\Phi)$. This is the weak converse of theorem 3.1 in the sense that the probability of error does not tend to zero asymptotically as the length of the code increases, for any code with rate $R > C(\Phi)$. To prove the weak converse, suppose that Alice encodes messages labelled by $\alpha \in \mathcal{M}_n$ by product states $\rho_\alpha^{(n)} = \rho_{\alpha,1} \otimes \ldots \otimes \rho_{\alpha,n}$ in $\mathcal{B}(\mathcal{H}^{\otimes n})$. Let the corresponding outputs for the ith branch of the channel be denoted by $\sigma_{\alpha,i}^{(n)}$, i.e.

$$\sigma_{\alpha,i}^{(n)} = \Phi_i^{\otimes n}\big(\rho_\alpha^{(n)}\big) = \sigma_{\alpha,1}^i \otimes \ldots \otimes \sigma_{\alpha,n}^i, \quad \sigma_{\alpha,j}^i = \Phi_i(\rho_{\alpha,j}). \tag{89}$$

Further define

$$\bar{\sigma}_i^{(n)} = \frac{1}{|\mathcal{M}_n|} \sum_{\alpha \in \mathcal{M}_n} \sigma_{\alpha,i}^{(n)} \tag{90}$$

and

$$\bar{\sigma}_{i,j} = \frac{1}{|\mathcal{M}_n|} \sum_{\alpha \in \mathcal{M}_n} \sigma_{\alpha,j}^i. \tag{91}$$

Let Bob's POVM elements corresponding to the codewords $\rho_\alpha^{(n)}$ be denoted by $E_\alpha^{(n)}, \alpha = 1, \ldots, |\mathcal{M}_n|$. We may assume that Alice's messages are produced uniformly at random from the set $\mathcal{M}_n$. Then Bob's average probability of error is given by

$$\bar{p}_e^{(n)} := 1 - \frac{1}{|\mathcal{M}_n|} \sum_{\alpha \in \mathcal{M}_n} \operatorname{Tr}\big[\Phi^{(n)}\big(\rho_\alpha^{(n)}\big)E_\alpha^{(n)}\big]. \tag{92}$$

We also define the average error corresponding to the ith branch of the channel as

$$\bar{p}_{i,e}^{(n)} := 1 - \frac{1}{|\mathcal{M}_n|} \sum_{\alpha \in \mathcal{M}_n} \operatorname{Tr}\big[\Phi_i^{\otimes n}\big(\rho_\alpha^{(n)}\big)E_\alpha^{(n)}\big], \tag{93}$$

so that

$$\bar{p}_e^{(n)} = \sum_{i=1}^{M} \gamma_i \bar{p}_{i,e}^{(n)}. \tag{94}$$

Let $X^{(n)}$ be a random variable with a uniform distribution over the set $\mathcal{M}_n$, characterizing the classical message sent by Alice to Bob. Let $Y_i^{(n)}$ be the random variable corresponding to

Bob's inference of Alice's message, when the codeword is transmitted through the ith branch of the channel. It is defined by the conditional probabilities

$$\mathbb{P}\big[Y_i^{(n)} = \beta | X^{(n)} = \alpha\big] = \mathrm{Tr}\big[\Phi_i^{\otimes n}\big(\rho_\alpha^{(n)}\big)E_\beta^{(n)}\big]. \tag{95}$$

By Fano's inequality,

$$h\big(\bar{p}_{i,e}^{(n)}\big) + \bar{p}_{i,e}^{(n)} \log(|\mathcal{M}_n| - 1) \geqslant H\big(X^{(n)}\big|Y_i^{(n)}\big) = H(X^{(n)}) - H\big(X^{(n)} : Y_i^{(n)}\big). \tag{96}$$

Here $h(\cdot)$ denotes the binary entropy and $H(\cdot)$ denotes the Shannon entropy. Using the Holevo bound and the subadditivity of the von Neumann entropy, we have

$$\begin{aligned} H\big(X^{(n)} : Y_i^{(n)}\big) &\leqslant S\left(\frac{1}{|\mathcal{M}_n|}\sum_{\alpha\in\mathcal{M}_n}\Phi_i^{\otimes n}\big(\rho_\alpha^{(n)}\big)\right) - \frac{1}{|\mathcal{M}_n|}\sum_{\alpha\in\mathcal{M}_n} S\left(\Phi_i^{\otimes n}\big(\rho_\alpha^{(n)}\big)\right) \\ &= S\left(\frac{1}{|\mathcal{M}_n|}\sum_{\alpha\in\mathcal{M}_n}\sigma_{\alpha,i}^{(n)}\right) - \frac{1}{|\mathcal{M}_n|}\sum_{\alpha\in\mathcal{M}_n} S\big(\sigma_{\alpha,i}^{(n)}\big) \\ &\leqslant \sum_{j=1}^{n}\left[S(\bar{\sigma}_{i,j}) - \frac{1}{|\mathcal{M}_n|}\sum_{\alpha\in\mathcal{M}_n} S\left(\sigma_{\alpha,j}^{i}\right)\right] \\ &= \sum_{j=1}^{n}\chi_i\left(\left\{\frac{1}{|\mathcal{M}_n|}, \rho_{\alpha,j}\right\}_{\alpha\in\mathcal{M}_n}\right) \\ &= \sum_{j=1}^{n}\frac{1}{|\mathcal{M}_n|}\sum_{\alpha\in\mathcal{M}_n} S\left(\sigma_{\alpha,j}^{i}\big\|\bar{\sigma}_{i,j}\right) := \sum_{j=1}^{n} A_j. \end{aligned} \tag{97}$$

The expression A_j can be rewritten using Donald's identity

$$\sum_\alpha p_\alpha S(\omega_\alpha\|\rho) = \sum_\alpha p_\alpha S(\omega_\alpha\|\bar{\omega}) + S(\bar{\omega}\|\rho), \tag{98}$$

where $\bar{\omega} = \sum_\alpha p_\alpha\omega_\alpha$. We apply this with ρ replaced by

$$\bar{\sigma}_i = \frac{1}{n|\mathcal{M}_n|}\sum_{j=1}^{n}\sum_{\alpha\in\mathcal{M}_n}\sigma_{\alpha,j}^{i}, \tag{99}$$

ω_α replaced by $\sigma_{\alpha,j}^{i}$, p_α replaced by $1/|\mathcal{M}_n|$ and consequently $\bar{\omega}$ replaced by $\bar{\sigma}_{i,j}$. Hence,

$$\begin{aligned} \frac{1}{|\mathcal{M}_n|}\sum_{\alpha\in\mathcal{M}_n} S\left(\sigma_{\alpha,j}^{i}\big\|\bar{\sigma}_{i,j}\right) &\leqslant \frac{1}{|\mathcal{M}_n|}\sum_{\alpha\in\mathcal{M}_n} S\big(\sigma_{\alpha,j}^{i}\big\|\bar{\sigma}_i\big) - S(\bar{\sigma}_{i,j}\|\bar{\sigma}_i) \\ &\leqslant \frac{1}{|\mathcal{M}_n|}\sum_{\alpha\in\mathcal{M}_n} S\big(\sigma_{\alpha,j}^{i}\big\|\bar{\sigma}_i\big), \end{aligned} \tag{100}$$

where we have used the non-negativity of the von Neumann entropy. Inserting into (97), we now have

$$\frac{1}{n}H\big(X^{(n)} : Y_i^{(n)}\big) \leqslant \frac{1}{n|\mathcal{M}_n|}\sum_{j=1}^{n}\sum_{\alpha\in\mathcal{M}_n} S\big(\sigma_{\alpha,j}^{i}\big\|\bar{\sigma}_i\big) = \chi_i\left(\left\{\frac{1}{n|\mathcal{M}_n|}, \rho_{\alpha,j}\right\}_{(\alpha,j)}\right). \tag{101}$$

Fano's inequality (96) now yields

$$h\big(\bar{p}_{i,e}^{(n)}\big) + \bar{p}_{i,e}^{(n)}\log|\mathcal{M}_n| \geqslant \log|\mathcal{M}_n| - n\chi_i\left(\left\{\frac{1}{n|\mathcal{M}_n|}, \rho_{\alpha,j}\right\}_{(\alpha,j)}\right). \tag{102}$$

However, since

$$C(\Phi) \geqslant \bigwedge_{i=1}^{M} \chi_i \left(\left\{ \frac{1}{n|\mathcal{M}_n|}, \rho_{\alpha,j} \right\}_{(\alpha,j)} \right) \tag{103}$$

and $R = \frac{1}{n} \log |\mathcal{M}_n| > C(\Phi)$, there must be at least one branch i such that

$$\bar{p}_{i,e}^{(n)} \geqslant 1 - \frac{C(\Phi) + 1/n}{R} > 0. \tag{104}$$

We conclude from (94) and (104) that

$$\bar{p}_{e}^{(n)} \geqslant \left(1 - \frac{C(\Phi) + 1/n}{R} \right) \bigwedge_{i=1}^{M} \gamma_i. \tag{105}$$

Acknowledgments

The authors would like to thank Andreas Winter for a helpful suggestion. They are also grateful to Igor Bjelaković for carefully reading the paper and pointing out some typos. This work was supported by the European Commission through the Integrated Project FET/QIPC 'SCALA'.

References

[1] Ahlswede R 1968 The weak capacity of averaged channels *Z. Wahrscheinlichkeitstheor. Verwandte. Geb.* **11** 61–73
[2] Bjelaković I and Boche H 2006 Ergodic classical-quantum channels: structure and coding theorems *Preprint* quant-ph/0609229
[3] Bowen G and Mancini S 2004 Quantum channels with a finite memory *Phys. Rev.* A **69** 01236
[4] Cover T M and Thomas J A *Elements of Information Theory* (New York: Wiley)
[5] Datta N and Dorlas T 2006 A quantum version of Feinstein's lemma and its application to channel coding *Proc. Int. Symp. Inf. Th. ISIT 2006, Seattle* pp 441–5
[6] Datta N and Dorlas T C Classical capacity of quantum channels with general Markovian correlated noise, in preparation
[7] Feinstein A 1954 A new basic theorem of information theory *IRE Trans. PGIT* **4** 2–22
[8] Hayashi M and Nagaoka H 2003 General formulas for capacity of classical-quantum channels *IEEE Trans. Inf. Theory* **49** 1753–68
[9] Helstrom C W 1976 Quantum detection and estimation theory *Mathematics in Science and Engineering* vol 123 (London: Academic)
[10] Hiai F and Petz D 1991 The proper formula for the relative entropy and its asymptotics in quantum probability *Commun. Math. Phys.* **143** 257–81
[11] Holevo A S 1998 The capacity of a quantum channel with general signal states *IEEE Trans. Inf. Theory* **44** 269–73
[12] Jacobs K 1962 Almost periodic channels *Colloqium on Combinatorial Methods in Probabilistic Theory (Ahrhus)*
[13] Khinchin A I 1957 *Mathematical Foundations of Information Theory: II. On the Fundamental Theorems of Information Theory* (New York: Dover) chapter IV
[14] Kretschmann D and Werner R F 2005 Quantum channels with memory *Preprint* quant-ph/0502106
[15] Macchiavello C and Palma G M 2002 Entanglement–enhanced information transmission over a quantum channel with correlated noise *Phys. Rev.* A **65** 050301
[16] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[17] Norris J R 1997 *Markov Chains, Cambridge Series in Statistical and Probabilistic Mathematics* (Cambridge: Cambridge University Press)
[18] Ohya M and Petz D 1993 *Quantum Entropy and Its Use* (Berlin: Springer)

[19] Schumacher B 1995 Quantum coding *Phys. Rev.* A **51** 2738–47
[20] Schumacher B and Westmoreland M D 1997 Sending classical information via noisy quantum channels *Phys. Rev.* A **56** 131–8
[21] Shannon C E 1948 A mathematical theory of communication: I *Bell Syst. Tech. J.* **27** 379–423
Shannon C E 1948 A mathematical theory of communication: II *Bell Syst. Tech. J.* **27** 623–56
[22] Winter A 1999 Coding theorem and strong converse for quantum channels *IEEE Trans. Inf. Theory* **45** 2481–5