DIAS-06-02

A quantum version of the Feinstein Lemma and its application to channel coding

N. Datta
Statistical Laboratory
Centre for Mathematical Sciences
University of Cambridge
Wilberforce Road, Cambridge CB3 0WB
Email: n.datta@statslab.cam.ac.uk

T. C. Dorlas
Dublin Institute for Advanced Studies
School of Theoretical Physics
10 Burlington Road, Dublin 4, Ireland.
Email: dorlas@stp.dias.ie

January 19, 2006

Abstract

In this paper we develop a quantum version of Feinstein's Lemma and use it to give a new proof of the direct channel coding theorem for transmission of classical information through a quantum memoryless channel. Moreover, we extend the lemma to a class of quantum channels with memory and thus obtain a bound for the achievable rates in the case of product state inputs.

1 Introduction

The biggest hurdle in the path of efficient information transmission is the presence of noise in classical and quantum channels. This noise causes a distortion of the information sent through the channel. To overcome this problem, one uses error–correcting codes. Instead of transmitting the original messages, the latter are encoded into codewords which are then sent through the channel. The codewords necessarily have redundancies so that even if part of a codeword is distorted by the noise in the channel, the corresponding output of the channel can still be decoded to yield the original message with a low probability of error. The information transmission is said to be reliable if the probability of error in decoding the output of the channel vanishes asymptotically (see e.g. [3] and [11]).

Shannon, in his Noisy Channel Coding Theorem [15], showed that information can be reliably sent over a classical channel at all rates up to the channel capacity. The first rigorous proof of this fundamental theorem was provided by Feinstein [5]. He used a packing argument to find a upper bound to the maximal number of codewords that can be sent through the channel with a low probability of error. His argument is often referred to as Feinstein's Lemma.

In this paper we develop a quantum version of Feinstein's Lemma and use it to find an alternative proof of the direct Channel Coding Theorem for transmission of classical information through a quantum memoryless channel. For such a channel successive channel inputs are acted on independently.

The first proof of this theorem, which states that all rates up to the so–called Holevo capacity are achievable, was proved independently by Holevo [8] and Schumacher and Westmoreland [14]. Unlike our proof, they employed the random coding technique. Alternative proofs have been given by Winter [16], Ogawa [12], and Hayashi & Nagaoka [6].

The proof in [12] was based on the standpoint of quantum hypothesis testing and the quantum information spectrum, though it also employed an argument similar to Feinstein's lemma. In [6] the technique of quantum information spectrum was used and there were no structural assumptions imposed on the quantum channels.

Our version of the quantum Feinstein's lemma can be extended explicitly to a class of quantum channels with memory. This allows us to obtain a rigorous lower bound to the maximum achievable rate of transmission for this class of channels, for the case of product state inputs. The generalized quantum Feinstein lemma and the direct coding theorem for these channels with memory are given in Section 5. However, due to lack of space, the details of the proof have been omitted. They will be presented in [4].

The quantum Feinstein lemma for memoryless channels is stated and proved in Section 3, and the corresponding direct coding theorem is given in Section 4.

2 Preliminaries

Let $\mathcal{B}(\mathcal{H})$ denote the algebra of linear operators acting on a finite–dimensional Hilbert space $\mathcal{H}$, and $\mathcal{S}(\mathcal{H})$ denote the set of all postive operators of unit trace in $\mathcal{B}(\mathcal{H})$, i.e., states (or density matrices). The von Neumann entropy of a state ρ is defined as $S(\rho) = -\mathrm{Tr}\,\rho\log\rho$, where the logaritm is taken to base 2. A quantum channel is given by a completely positive trace–preserving (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. Let $\dim \mathcal{H} = d$ and $\dim \mathcal{K} = d'$. The Holevo

capacity of the channel is defined as follows:

$$\chi(\Phi) := \max_{\{p_j,\rho_j\}} \left\{ S\left(\sum_j p_j\, \Phi(\rho_j)\right) - \sum_j p_j\, S(\Phi(\rho_j)) \right\}, \tag{1}$$

where the maximum is taken over all ensembles $\{p_j, \rho_j\}$ of possible input states $\rho_j \in \mathcal{B}(\mathcal{H})$ and probability distributions $\{p_j\}$.

It can be shown that the maximum in (1) can be achieved by using an ensemble of pure states and that in the maximisation it suffices to consider ensembles of at most d^2 pure states.

3 The Quantum Feinstein Lemma

Theorem 1 *Let $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$, be a quantum channel and let $\chi(\Phi)$ be its Holevo capacity. Given $\epsilon > 0$, there exists $n_0 \in \mathbf{N}$ such that for all $n \geq n_0$ there exists $N \geq 2^{n(\chi(\Phi)-\epsilon)}$ and there exist product states $\tilde{\rho}_1^{(n)}, \ldots, \tilde{\rho}_N^{(n)} \in \mathcal{S}(\mathcal{H}^{\otimes n})$ and positive operators $E_1^{(n)}, \ldots, E_N^{(n)} \in \mathcal{B}(\mathcal{K}^{\otimes n})$ such that $\sum_{k=1}^N E_k^{(n)} \leq \mathbf{1}$ and*

$$Tr\, \Phi^{\otimes n}\left(\tilde{\rho}_k^{(n)}\right) E_k^{(n)} > 1 - \epsilon, \tag{2}$$

for each k.

Proof Let the maximum in (1) be attained for an ensemble $\{p_j, \rho_j\}_{j=1}^J$, where $J \leq d^2$. Denote $\sigma_j = \Phi(\rho_j)$, $\bar{\sigma} = \sum_{j=1}^J p_j \Phi(\rho_j)$ and $\bar{\sigma}_n = \bar{\sigma}^{\otimes n}$.

Choose $\delta > 0$. We will relate δ to ϵ at a later stage. The Typical Subspace Theorem (see e.g. [13] or [11]) ensures that there exists $n_1 \in \mathbf{N}$, such that for $n \geq n_1$, there is a typical subspace $\overline{\mathcal{T}}_{\delta,\epsilon}$ with projection P_n, such that if $\bar{\sigma}_n$ has a spectral decomposition

$$\bar{\sigma}_n = \sum_{\underline{k}} \bar{\lambda}_{\underline{k}}^{(n)} |\psi_{\underline{k}}^{(n)}\rangle\langle\psi_{\underline{k}}^{(n)}|, \tag{3}$$

then

$$\left| \frac{1}{n} \log \bar{\lambda}_{\underline{k}}^{(n)} + S(\bar{\sigma}) \right| < \frac{\epsilon}{3}, \tag{4}$$

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

$$\mathrm{Tr}\,(P_n \bar{\sigma}_n) > 1 - \delta^2. \tag{5}$$

Define

$$\bar{S} = \sum_{j=1}^{J} p_j\, S(\sigma_j). \tag{6}$$

We make use of the following lemma.

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

$$\left| \frac{1}{n} \log \lambda_{\underline{j},\underline{k}}^{(n)} + \bar{S} \right| < \frac{\epsilon}{3}. \tag{7}$$

Let $\delta > 0$. There exists $n_2 \in \mathbf{N}$ such that for $n \geq n_2$,

$$\mathbf{E}\left(Tr\, \sigma_{\underline{j}}^{(n)} P_{\underline{j}}^{(n)} \right) > 1 - \delta^2. \tag{8}$$

Proof *(of Lemma 1)* Define independent and identically distributed (i.i.d.) random variables $X_1, \ldots, X_n$ with distribution given by

$$\mathbf{P}(X_i = \lambda_{j,k}) = p_j\, \lambda_{j,k}, \tag{9}$$

where $\lambda_{j,k},\ k = 1, 2, \ldots, d'$ are the eigenvalues of σ_j. By the weak law of large numbers, we have the following convergence in probability

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} \log X_i \xrightarrow{\mathbf{P}} \mathbf{E}(\log X_i) &= \sum_{j=1}^{J} \sum_{k=1}^{d'} p_j\, \lambda_{j,k} \log \lambda_{j,k} \\ &= -\sum_{j=1}^{J} p_j\, S(\sigma_j) \\ &= -\bar{S}. \end{aligned} \tag{10}$$

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

$$\left| \frac{1}{n} \sum_{i=1}^{n} \log \lambda_{j_i,k_i} + \bar{S} \right| < \frac{\epsilon}{3}, \tag{11}$$

satisfies

$$\mathbf{P}\left(T_{\delta,\epsilon}^{(n)}\right) = \sum_{((j_1,k_1),\ldots,(j_n,k_n))\in T_{\delta,\epsilon}^{(n)}} \prod_{i=1}^{n} p_{j_i}\lambda_{j_i,k_i} > 1-\delta^2. \tag{12}$$

Obviously,

$$P_{\underline{j}}^{(n)} \geq \sum_{\underline{k}:(\underline{j},\underline{k})\in T_{\delta,\epsilon}^{(n)}} |\psi_{\underline{j},\underline{k}}^{(n)}\rangle\langle\psi_{\underline{j},\underline{k}}^{(n)}|, \tag{13}$$

and

$$\mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}P_{\underline{j}}^{(n)}\right) \geq \mathbf{P}\left(T_{\delta,\epsilon}^{(n)}\right) > 1-\delta^2. \tag{14}$$

QED

Continuing the proof of the theorem, let $N = N(n)$ be the *maximal* number for which there exist states $\tilde{\rho}_1^{(n)},\ldots,\tilde{\rho}_N^{(n)}$ on $\mathcal{H}^{\otimes n}$ of the tensor product form

$$\tilde{\rho}_k^{(n)} = \rho_{k_1}\otimes\rho_{k_2}\ldots\otimes\rho_{k_n}\ ,$$

and there exist positive operators $E_1^{(n)},\ldots,E_N^{(n)}$ on $\mathcal{K}^{\otimes n}$ such that, defining $\tilde{\sigma}_k^{(n)} = \Phi^{\otimes n}(\tilde{\rho}_k^{(n)})$, we have

(i) $\sum_{k=1}^{N} E_k^{(n)} \leq P_n$ and

(ii) $\mathrm{Tr}\,\tilde{\sigma}_k^{(n)}E_k^{(n)} > 1-2\epsilon$ for each k, and

(iii) $\mathrm{Tr}\,\bar{\sigma}_n E_k^{(n)} \leq 2^{-n[S(\bar{\sigma})-\bar{S}-\frac{2}{3}\epsilon]}$ for each k.

For any given $\underline{j}$ define

$$V_{\underline{j}}^{(n)} = \left(P_n - \sum_{k=1}^{N} E_k^{(n)}\right)^{1/2} P_n P_{\underline{j}}^{(n)} P_n \left(P_n - \sum_{k=1}^{N} E_k^{(n)}\right)^{1/2}. \tag{15}$$

Clearly, $V_{\underline{j}}^{(n)} \leq P_n - \sum_{k=1}^{N} E_k^{(n)}$, and we also have:

Lemma 2 *Define*

$$W_n = \{\underline{j} \mid Tr(\sigma_{\underline{j}}^{(n)}P_{\underline{j}}^{(n)}) > 1-\delta\}. \tag{16}$$

Then, for all $\underline{j}\in W_n$,

$$Tr(\bar{\sigma}_n V_{\underline{j}}^{(n)}) \leq 2^{-n[S(\bar{\sigma})-\bar{S}-\frac{2}{3}\epsilon]}. \tag{17}$$

Proof *(of Lemma 2)* Put $Q_n = \sum_{k=1}^{N(n)} E_k^{(n)}$. Note that Q_n commutes with P_n. Using the fact that $P_n \bar{\sigma}_n P_n \leq 2^{-n[S(\bar{\sigma})-\frac{1}{3}\epsilon]}$ by (4), we have

$$\begin{aligned}
\mathrm{Tr}\, \bar{\sigma}_n V_{\underline{j}}^{(n)} &= \mathrm{Tr}\, \bar{\sigma}_n (P_n - Q_n)^{1/2} P_n P_{\underline{j}}^{(n)} P_n (P_n - Q_n)^{1/2} \\
&= \mathrm{Tr}\, P_n \bar{\sigma}_n P_n (P_n - Q_n)^{1/2} P_{\underline{j}}^{(n)} (P_n - Q_n)^{1/2} \\
&\leq 2^{-n[S(\bar{\sigma}_n)-\frac{1}{3}\epsilon]} \mathrm{Tr} \Big[(P_n - Q_n)^{1/2} \\
&\qquad \times P_{\underline{j}}^{(n)} (P_n - Q_n)^{1/2} \Big] \\
&\leq 2^{-n[S(\bar{\sigma}_n)-\frac{1}{3}\epsilon]} \mathrm{Tr}\; P_{\underline{j}}^{(n)} \\
&\leq 2^{-n[S(\bar{\sigma}_n)-\bar{S}-\frac{2}{3}\epsilon]}, \qquad (18)
\end{aligned}$$

where, in the last inequality, we used the standard upper bound on the dimension of the typical subspace: $\mathrm{Tr}\, P_{\underline{j}}^{(n)} \leq 2^{n[\bar{S}+\frac{1}{3}\epsilon]}$, which follows from (7). QED

Since $N(n)$ is maximal it follows that for $\underline{j} \in W_n$,

$$\mathrm{Tr}\, \sigma_{\underline{j}}^{(n)} V_{\underline{j}}^{(n)} \leq 1 - 2\epsilon. \qquad (19)$$

We now show that the set W_n has high probability:

Lemma 3 *$\mu(W_n) > 1 - \delta$, where $\mu(W_n) := \sum_{\underline{j} \in W_n} p_{\underline{j}}^{(n)}$.*

Proof *(of Lemma 3)* If $\underline{j} \notin W_n$ then $\mathrm{Tr}\, \sigma_{\underline{j}}^{(n)} P_{\underline{j}}^{(n)} \leq 1 - \delta$. Hence

$$\sum_{\underline{j} \notin W_n} p_{\underline{j}}^{(n)} \mathrm{Tr}\, \sigma_{\underline{j}}^{(n)} (\mathbf{1} - P_{\underline{j}}^{(n)}) \geq \delta\, \mu(W_n^c), \qquad (20)$$

where $\mathbf{1}$ denotes the identity operator on $\mathcal{H}^{\otimes n}$. On the other hand,

$$\begin{aligned}
\sum_{\underline{j} \notin W_n} p_{\underline{j}}^{(n)} \mathrm{Tr}\, \sigma_{\underline{j}}^{(n)} (\mathbf{1} - P_{\underline{j}}^{(n)}) &\leq \mathbf{E} \left(\mathrm{Tr}\, \sigma_{\underline{j}}^{(n)} (\mathbf{1} - P_{\underline{j}}^{(n)}) \right) \\
&< \delta^2, \qquad (21)
\end{aligned}$$

by (8). Hence, $\mu(W_n^c) < \frac{\delta^2}{\delta} = \delta$. QED

Corollary 1 *Assume $\delta < \epsilon$. Then*

$$\mathbf{E}\left(Tr\,\sigma_{\underline{j}}^{(n)} V_{\underline{j}}^{(n)}\right) < 1 - \epsilon. \tag{22}$$

Proof *(of Corollary 1)* Using (19), we have

$$\begin{aligned}
&\mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)} V_{\underline{j}}^{(n)}\right) = \\
&= \sum_{\underline{j}\in W_n} p_{\underline{j}}^{(n)} \mathrm{Tr}\,\sigma_{\underline{j}}^{(n)} V_{\underline{j}}^{(n)} + \sum_{\underline{j}\in W_n^c} p_{\underline{j}}^{(n)} \mathrm{Tr}\,\sigma_{\underline{j}}^{(n)} V_{\underline{j}}^{(n)} \\
&\leq 1 - 2\epsilon + \mu(W_n^c) < 1 - \epsilon,
\end{aligned} \tag{23}$$

since $\delta < \epsilon$. QED

The bound given in the following lemma is essential for the proof of the lower bound (2).

Lemma 4 *For all $\eta > 0$, there exists $n_3 \in \mathbf{N}$ such that for all $n \geq n_3$,*

$$\mathbf{E}\left(Tr\,\sigma_{\underline{j}}^{(n)} P_n P_{\underline{j}}^{(n)} P_n\right) > 1 - \eta. \tag{24}$$

Proof *(of Lemma 4)* We write

$$\begin{aligned}
&\mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)} P_n P_{\underline{j}}^{(n)} P_n\right) = \\
&= \mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)} P_{\underline{j}}^{(n)}\right) - \mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)} (\mathbf{1} - P_n) P_{\underline{j}}^{(n)}\right) \\
&\quad - \mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)} P_n P_{\underline{j}}^{(n)} (\mathbf{1} - P_n)\right).
\end{aligned} \tag{25}$$

By *Lemma* 1, the first term is $> 1 - \delta^2$ provided $n \geq n_2$. The last two terms

can be bounded using Cauchy-Schwarz as follows:

$$\begin{aligned}
&\mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}(\mathbf{1}-P_n)P_{\underline{j}}^{(n)}\right) = \\
&= \mathbf{E}\left(\mathrm{Tr}\,\left(\sigma_{\underline{j}}^{(n)}\right)^{1/2}(\mathbf{1}-P_n)P_{\underline{j}}^{(n)}\left(\sigma_{\underline{j}}^{(n)}\right)^{1/2}\right) \\
&\leq \left\{\mathbf{E}\left(\mathrm{Tr}\,(\mathbf{1}-P_n)\sigma_{\underline{j}}^{(n)}(\mathbf{1}-P_n)\right)\right\}^{1/2} \\
&\qquad \times\left\{\mathbf{E}\left(\mathrm{Tr}\,\left(\sigma_{\underline{j}}^{(n)}\right)^{1/2}P_{\underline{j}}^{(n)}\left(\sigma_{\underline{j}}^{(n)}\right)^{1/2}\right)\right\}^{1/2} \\
&= \left\{\mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}(\mathbf{1}-P_n)\right)\right\}^{1/2}\left\{\mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}P_{\underline{j}}^{(n)}\right)\right\}^{1/2} \\
&\leq \left\{\mathbf{E}\left(\mathrm{Tr}\,\left[\sigma_{\underline{j}}^{(n)}(\mathbf{1}-P_n)\right]\right)\right\}^{1/2} \\
&= (\mathrm{Tr}\,\bar{\sigma}_n(\mathbf{1}-P_n))^{1/2} \leq \delta \qquad (26)
\end{aligned}$$

by (5) provided $n \geq n_1$. Similarly,

$$\begin{aligned}
&\mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}P_nP_{\underline{j}}^{(n)}(\mathbf{1}-P_n)\right) \\
&= \mathbf{E}\left(\mathrm{Tr}\,\left(\sigma_{\underline{j}}^{(n)}\right)^{1/2}P_nP_{\underline{j}}^{(n)}(\mathbf{1}-P_n)\left(\sigma_{\underline{j}}^{(n)}\right)^{1/2}\right) \\
&\leq \left\{\mathbf{E}\left(\mathrm{Tr}\,P_{\underline{j}}^{(n)}P_n\sigma_{\underline{j}}^{(n)}P_nP_{\underline{j}}^{(n)}\right)\right\}^{1/2} \\
&\qquad \times\left\{\mathbf{E}\left(\mathrm{Tr}\,\left(\sigma_{\underline{j}}^{(n)}\right)^{1/2}(\mathbf{1}-P_n)\left(\sigma_{\underline{j}}^{(n)}\right)^{1/2}\right)\right\}^{1/2} \\
&= \left\{\mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}P_nP_{\underline{j}}^{(n)}P_n\right)\right\}^{1/2} \\
&\qquad \times\left\{\mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}(\mathbf{1}-P_n)\right)\right\}^{1/2} \\
&\leq \left\{\mathbf{E}\left(\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}(\mathbf{1}-P_n)\right)\right\}^{1/2} \leq \delta. \qquad (27)
\end{aligned}$$

Choosing $n_3 = n_1 \vee n_2 \equiv \max\{n_1, n_2\}$, and $\delta^2 + 2\delta < \eta$ the result follows. QED

Lemma 5 *Assume $\eta < \frac{1}{3}\epsilon$ and $\delta < \epsilon$. Then for $n \geq n_3$,*

$$Tr\,\bar{\sigma}_n \sum_{k=1}^{N} E_k^{(n)} = \mathbf{E}\left(Tr\,\sigma_{\underline{j}}^{(n)} \sum_{k=1}^{N} E_k^{(n)}\right) \geq \eta^2. \qquad (28)$$

Proof *(of Lemma 5)* Define

$$Q'_n = P_n - (P_n - Q_n)^{1/2}, \tag{29}$$

where $Q_n = \sum_{k=1}^{N} E_k^{(n)}$. By Corollary 1,

$$\begin{aligned} 1-\epsilon &> \mathbf{E}\left\{\mathrm{Tr}\,\left(\sigma_{\underline{j}}^{(n)}(P_n - Q'_n)P_{\underline{j}}^{(n)}(P_n - Q'_n)\right)\right\} \\ &= \mathbf{E}\left\{\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}P_nP_{\underline{j}}^{(n)}P_n\right\} \\ &\quad -\mathbf{E}\left\{\mathrm{Tr}\,\left(\sigma_{\underline{j}}^{(n)}Q'_nP_{\underline{j}}^{(n)}P_n\right)\right\} \\ &\quad -\mathbf{E}\left\{\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}P_nP_{\underline{j}}^{(n)}Q'_n\right\} \\ &\quad +\mathbf{E}\left\{\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}Q'_nP_{\underline{j}}^{(n)}Q'_n\right\}. \end{aligned} \tag{30}$$

Since the last term is positive, we have, by *Lemma* 4,

$$\begin{aligned} \mathbf{E}\left\{\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}Q'_nP_{\underline{j}}^{(n)}P_n + \mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}P_nP_{\underline{j}}^{(n)}Q'_n\right\} &> \epsilon - \eta \\ &> 2\eta. \end{aligned} \tag{31}$$

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

$$\begin{aligned} &\mathbf{E}\left\{\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}Q'_nP_{\underline{j}}^{(n)}P_n + \mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}P_nP_{\underline{j}}^{(n)}Q'_n\right\} \leq \\ &\leq 2\left\{\mathbf{E}\left[\mathrm{Tr}\,Q'_n\sigma_{\underline{j}}^{(n)}Q'_n\right]\right\}^{1/2}\left\{\mathbf{E}\left[\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}P_nP_{\underline{j}}^{(n)}P_n\right]\right\}^{1/2} \\ &\leq 2\left\{\mathbf{E}\left[\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}Q'^2_n\right]\right\}^{1/2}. \end{aligned} \tag{32}$$

Thus,

$$\mathbf{E}\left[\mathrm{Tr}\,\sigma_{\underline{j}}^{(n)}Q'^2_n\right] \geq \eta^2. \tag{33}$$

To complete the proof of this lemma, we now claim that

$$Q_n \geq (Q'_n)^2. \tag{34}$$

Indeed, this follows on the domain of P_n from the inequality $1-(1-x)^2 \geq x^2$ for $0 \leq x \leq 1$. QED

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

$$\mathrm{Tr}\,\bar{\sigma}_nE_k^{(n)} \leq 2^{-n[S(\bar{\sigma})-\bar{S}-\frac{2}{3}\epsilon]} \tag{35}$$

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* 5,

$$\mathrm{Tr}\,\bar{\sigma}_n \sum_{k=1}^{N} E_k^{(n)} \geq \eta^2 \tag{36}$$

provided $n \geq n_3$. It follows from item (iii) in the definition of $N(n)$ (below eq. (14)) that

$$N(n) \geq \eta^2 2^{n[S(\bar{\sigma})-\bar{S}-\frac{2}{3}\epsilon]} \geq 2^{n[S(\bar{\sigma})-\bar{S}-\epsilon]} \tag{37}$$

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

4 The Direct Channel Coding Theorem

Theorem 1 can now be used to prove that the Holevo capacity, defined by (1), provides an lower bound to the maximum achievable rate of transmission of classical information through a quantum memoryless channel, when the inputs to multiple uses of the channel are restricted to product states.

Let the sender Alice have a set of classical messages labelled by n-letter words $x \in A^n$ from an alphabet A. We denote the probability distribution of the messages by μ_n. To transmit these messages to Bob through a quantum channel Φ in the form of product states, she encodes a message x by a codeword, which is a state $\tilde{\rho}_k^{(n)} \in \mathcal{S}(\mathcal{H}^{\otimes n})$. For this purpose, she makes use of an ensemble $\{p_k, \rho_k\}$ of quantum states, and sets $\tilde{\rho}_k^{(n)} = \rho_{k_1} \otimes \rho_{k_2} \ldots \otimes \rho_{k_n}$, with probability $p_{k_1} p_{k_2} \ldots p_{k_n}$. This state is transmitted through n uses of the channel, i.e., through $\Phi^{\otimes n}$. Bob receives the state $\tilde{\sigma}_k^{(n)} := \Phi^{\otimes n}(\tilde{\rho}_k^{(n)})$. To decode the label k of the message sent by Alice, Bob does a measurement on $\tilde{\sigma}_k^{(n)}$ described by positive operators (POVM elements) $E_1^{(n)}, \ldots, E_{N_n}^{(n)}$ (where $\sum_{j=1}^{N_n} E_j^{(n)} \leq \mathbf{I}_n$) and $E_0^{(n)} := \mathbf{I}_n - \sum_{j=1}^{N_n} E_j^{(n)}$. The POVM element $E_0^{(n)}$ corresponds to a failure in decoding. The (asymptotic) rate of information transmission in this scenario is given by

$$R := \lim_{n\to\infty} R_n = \lim_{n\to\infty} (\log N_n)/n, \tag{38}$$

where N_n is the number of code words. Assuming that all coded messages arise with uniform probability (note that this is roughly the case for a typical

set), the average probability of error is given by

$$p_e^{(n)} = \frac{1}{N_n} \sum_{k=1}^{N_n} \left[1 - \mathrm{Tr} \left(\Phi^{\otimes n} \left(\tilde{\rho}_k^{(n)} \right) E_k^{(n)} \right] \right].$$

In the following theorem, which is the direct part of the HSW theorem generalized to an ergodic source, we prove that for any rate $R < \chi(\Phi)$, where $\chi(\Phi)$ is the Holevo capacity defined through (1), the coding and decoding scheme given, respectively, by a code with product state codewords and a POVM, classical information can be transmitted reliably through the memoryless quantum channel.

Theorem 2 *Consider a memoryless quantum channel given by a completely positive trace-preserving map $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$, where $\mathcal{H}$ and $\mathcal{K}$ are finite-dimensional Hilbert spaces. Let $\chi(\Phi)$ be the Holevo capacity of the channel. If A is the alphabet of a classical ergodic source of information with probability distribution μ and Shannon entropy $H < \chi(\Phi)$, then there exists for any given $\epsilon > 0$, an $n_0 \in \mathbf{N}$ such that for all $n \geq n_0$ there exist a code map $\mathcal{C}_n : A^n \to \mathcal{S}(\mathcal{H}^{\otimes n})$ with image in the product states, and a random decoding $\mathcal{D}_n : \mathcal{S}(\mathcal{K}^{\otimes n}) \to A^n$ such that the average error probability given by*

$$p_e = \sum_{x \in A^n} \sum_{y \in A^n; y \neq x} \mu_n(x) \, \mathbf{P} \left[\mathcal{D}_n(\Phi^{\otimes n}(\mathcal{C}_n(x))) = y \right] < \epsilon, \tag{39}$$

where μ_n is the restriction of μ to A^n. Moreover, Given $\delta > 0$, the code can be chosen such that the rate of information transmission is exceeds $H - \delta$.

Proof By McMillan's theorem (see e.g. [3]), for $\epsilon > 0$ and n large enough, there exists a typical set $T_n \equiv T_\epsilon^{(n)}$ in A^n, such that for all $x \in T_n$, $\mu_n(x) > 2^{-n(H+\epsilon)}$ and $\mu_n(T_n) > 1 - \epsilon$. By the above Theorem 1, there exist product states $\rho_k^{(n)}$ and positive operators $E_k^{(n)}$ with $k = 1, \ldots, N$ and $N > 2^{n(\chi(\Phi)-\epsilon)}$ such that $\mathrm{Tr}\,[\Phi^{\otimes n}(\rho_k^{(n)}) E_k^{(n)}] > 1 - \epsilon$. Choose ϵ to be so small that $H + \epsilon < \chi(\Phi) - \epsilon$. Then we can define a one–to–one map $\mathcal{C}_n : T_0 \to \mathcal{S}(\mathcal{H}^{\otimes n})$ by $\mathcal{C}_n(x) = \rho_{k_x}^{(n)}$ for certain $k_x \in \{1, \ldots, N\}$. Choosing one k_0 in the complement of $I_n := \{k_x | \, x \in T_0\}$, we put $\mathcal{C}_n(x) = \rho_{k_0}^{(n)}$ if $x \notin T_0$. We define the decoding as follows: Given $\sigma^{(n)} \in \mathcal{S}(\mathcal{K}^{\otimes n})$ we define a probability measure on A^n by

$$\nu_n(\sigma^{(n)})(x) = \mathrm{Tr}\, [\sigma^{(n)} E_k^{(n)}(x)]. \tag{40}$$

To determine $\mathcal{D}_n(\sigma^{(n)})$ we sample A^n with this probability distribution. (If the result is k_0 we put it equal to a fixed $x_0 \notin T_0$). Clearly then,

$$\begin{aligned} p_e &= \sum_{x\in A^n} \mu_n(x)\, \mathbf{P}\left[\mathcal{D}_n(\Phi^{(n)}(\mathcal{C}_n(x))) \neq x\right] \\ &= \sum_{x\in T_0} \mu_n(x)(1 - \mathrm{Tr}\,[\sigma_{k_x}^{(n)} E_k^{(n)}(x)]) + \mu_n(T_0^c) < 2\epsilon, \end{aligned} \tag{41}$$

where $\sigma_{k_x}^{(n)} = \Phi^{(n)}(\mathcal{C}_n(x))$ and T_0^c denotes the complement of T_0. Obviously, the number of codewords is given by $|T_0| + 1$, which is bounded by $2^{n(H-\delta)}$. QED

5 A class of channels with memory

We consider a class of quantum channels with Markovian correlated noise. These were introduced in [10] and studied in more generality in [2] and [9]. Such a channel of length n is a CPT map $\Phi^{(n)} : \mathcal{B}(\mathcal{H}^{\otimes n}) \to \mathcal{B}(\mathcal{K}^{\otimes n})$ and needs the following ingredients for its definitions: (i) the transition matrix of a discrete–time Markov chain, with elements $q_{i|j}$, (ii) an initial (error) probability distribution $\{q_i\}$, which is the invariant distribution of the chain, and (iii) a finite set $\{V_j\}$, where V_j denotes a unitary Kraus operator for a single use of the channel, i.e., for $\Phi^{(1)} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$. The channel $\Phi^{(n)}$ is defined through its action on a state $\rho^{(n)} \in \mathcal{B}(\mathcal{H}^{\otimes n})$ as follows.

$$\begin{aligned} \Phi^{(n)}(\rho^{(n)}) &= \sum_{i_0,\ldots,i_n} q_{i_n|i_{n-1}} \cdots q_{i_1|i_0} q_{i_0} \\ &\quad (V_{i_n} \otimes .. \otimes V_{i_0}) \rho^{(n)} (V_{i_n}^* \otimes .. \otimes V_{i_0}^*). \end{aligned} \tag{42}$$

Extending *Theorem 1* to this class of channels yields the following theorem.

Theorem 3 *Let $\Phi^{(n)}$ denote a quantum memory channel with Markovian correlated noise, defined by (42). A quantity characterising it is*

$$\tilde{\chi} := \sup_{\{p_j,\rho_j\}} \left\{ S_M - \bar{S}(\{p_j, \rho_j\}) \right\}, \tag{43}$$

where S_M is defined as

$$S_M := \lim_{n\to\infty} \frac{1}{n} S\left(\Phi^{(n)}(\bar{\rho}^{\otimes n})\right) \tag{44}$$

with $\bar{\rho} = \sum_j p_j \rho_j$, and

$$\begin{aligned} \bar{S}(\{p_j, \rho_j\}) &= \lim_{n\to\infty} \frac{1}{n} \sum_{j_1,\ldots,j_n} p_{j_1} \ldots p_{j_n} \\ &\quad \times S\left(\Phi^{(n)}(\rho_{j_1} \otimes \cdots \otimes \rho_{j_n})\right). \end{aligned} \tag{45}$$

Given $\epsilon > 0$, there exists $n_0 \in \mathbf{N}$ such that for all $n \geq n_0$ there exists $N \geq 2^{n(\tilde{\chi}-\epsilon)}$ and there exist product states $\tilde{\rho}_1^{(n)}, \ldots, \tilde{\rho}_N^{(n)} \in \mathcal{S}(\mathcal{H}^{\otimes n})$ and positive operators $E_1^{(n)}, \ldots, E_N^{(n)} \in \mathcal{B}(\mathcal{K}^{\otimes n})$ such that $\sum_{k=1}^N E_k^{(n)} \leq \mathbf{1}$ and

$$Tr\,\Phi^{(n)}\left(\tilde{\rho}_k^{(n)}\right) E_k^{(n)} > 1 - \epsilon. \tag{46}$$

The proof of this theorem [4] is analogous to the proof of *Theorem 1* and relies on the existence of suitable typical subspaces. However, there are some important differences. Since the channel has memory, the outputs of the channel are not of a tensor product form, even when the inputs are chosen to be product states. They can, however, be proved to be ergodic. This allows us to use the theorem of Hiai and Petz [7], or more generally of Bjelakovic et al. [1], to define a typical subspace which is the analogue of the typical subspace $\overline{\mathcal{T}}_{\delta,\epsilon}$ used in the proof of *Theorem 1.* As a consequence, we have a direct coding theorem for classical information over a quantum channel with memory:

Theorem 4 *Consider a quantum channel with memory defined by completely positive maps $\Phi^{(n)}$ of the form (42). Let $\tilde{\chi}$ be the capacity of the channel defined by (43). If A is the alphabet of a classical ergodic source of information with probability distribution μ and Shannon entropy $H < \tilde{\chi}$, then there exists for any given $\epsilon > 0$, an $n_0 \in \mathbf{N}$ such that for all $n \geq n_0$ there exist a code map $\mathcal{C}_n : A^n \to \mathcal{S}(\mathcal{H}^{\otimes n})$, and a random decoding $\mathcal{D}_n : \mathcal{S}(\mathcal{K}^{\otimes n}) \to A^n$ such that the average error probability given by*

$$p_e = \sum_{x\in A^n} \sum_{y\in A^n; y\neq x} \mu_n(x) \mathbf{P}\left[\mathcal{D}_n(\Phi^{(n)}(\mathcal{C}_n(x))) = y\right] < \epsilon, \tag{47}$$

where μ_n is the restriction of μ to A^n. Moreover, if $\delta > 0$ then the code can be chosen such that the rate of information transmission $R > H - \delta$.

References

[1] I. Bjelaković, T. Krüger, R. Siegmund-Schultze & A. Szkoła, "The Shannon-McMillan theorem for ergodic quantum lattice systems," *Invent. math.* **155**, 203–222, 2004.

[2] G. Bowen and S. Mancini, "Quantum channels with a finite memory", *Phys. Rev. A* **69**, 01236, 2004.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory,* John Wiley & Sons, Inc.

[4] N. Datta and T. C. Dorlas, "Product state capacities of a class of channels with memory," *in preparation.*

[5] A. Feinstein, " A new basic theorem of information theory," *IRE Trans. PGIT,* **4**, pp. 2–22, 1954.

[6] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Trans. Inform. Theory* **49**, pp. 1753–1768, 2003.

[7] F. Hiai & D. Petz, "The proper formula for the relative entropy and its asymptotics in quantum probability". *Commun. Math. Phys.* **143**, 257–281, 1991.

[8] A. S. Holevo, "The capacity of a quantum channel with general signal states," *IEEE Trans. Info. Theory*, **44**, 269-273, 1998.

[9] D. Kretschmann and R. F. Werner, "Quantum channels with memory," *quant-ph/0502106.*

[10] C. Macchiavello and G. M. Palma, "Entanglement–enhanced information transmission over a quantum channel with correlated noise", *Phys. Rev. A* **65**, 050301, 2002.

[11] M.A. Nielson and I. L. Chuang, *Quantum Computation and Quantum Information,* Cambridge University Press, Cambridge, 2000.

[12] T. Ogawa and H. Nagaoka, "A New Proof of the Channel Coding Theorem via Hypothesis Testing in Quantum Information Theory," *Proc. 2002 IEEE International Symposium on Information Theory,* pp. 73, 2002.

[13] B. Schumacher, "Quantum Coding", *Phys. Rev. A* **51**, 2738-2747, 1995.

[14] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A* **56**, 131-138, 1997.

[15] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pt. I, pp. 379–423, 1948; pt. II, pp. 623–656, 1948.

[16] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Trans. Info. Theory*, **45**, 2481–2485, 1999.