Coding Theorem and Strong Converse for Nonstationary Quantum Channels

Andreas Winter

The author is with SFB 343, Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany. Electronic address: `winter@mathematik.uni-bielefeld.de`

Abstract

We study the nonstationary discrete memoryless quantum channel with arbitrary (product state) input, and prove a coding theorem and the strong converse for classical information transmission. The method is that of maximal codes and of packing decoding operators.

Keywords

Quantum channel, nonstationary, classical capacity, coding, strong converse.

I. Introduction

Twentyfive years after his bound on the classical information capacity of quantum channels [6] Holevo succeeded to prove the corresponding coding theorem [10], making the previous bound the weak converse (this was independently done by Schumacher and Westmoreland [16]). For a history of the problem see [9], for a modern account of the Holevo bound as a weak converse see [18].

The present work will extend these results in several ways: first, we will consider nonstationary discrete memoryless quantum channels, second, we will use a different method for proving the coding theorem, and third, we will prove the strong converse in this situation (in the finite and stationary case Ogawa, Nagaoka [13] proved the strong converse by a different approach). The method is an adaption to quantum channels of the ideas of maximal coding, and of set packing. In a subsequent work [20] the method will be demonstrated to yield good bounds on maximal code size for finite stationary channels, as well as upper and lower bounds for the reliability function.

Note that the present paper widens the scope and deepens the results of its predecessor [19], while there are minor differences in notation.

II. Prerequisites and notational conventions

We will use the definitions and notation of [18], in particular finite sets will be $\mathcal{A}$, $\mathcal{B}$, $\mathcal{M}$, $\mathcal{X}$, ..., quantum states (density operators) ρ, σ, ..., probability distributions P, Q, ..., and classical–quantum operations V, W, ... (also stochastic matrices), whereas general quantum operations (trace preserving, and completely positive maps) of state spaces of C*–algebras $\mathfrak{X}$, $\mathfrak{Y}$, ... are denoted as (pre-)adjoint maps φ_*, ψ_*, ... Our algebras will be of finite dimension, and apart from commutative ones we will confine ourselves to the C*–algebra $\mathfrak{L}(\mathcal{H})$, the algebra of (bounded) linear operators of the complex Hilbert space $\mathcal{H}$, though everything works equally well for the general case. In the sequel $d = \dim \mathcal{H}$.

The exponential function exp is always understood to basis 2, as well as the logarithm log. The function η on $[0, 1]$ is as usual defined by $\eta(x) = -x \log x$. The same symbol H denotes the Shannon and the von Neumann entropy (as Shannon's is the commutative case of von Neumann's).

We will use two different norms on operators: the supremum norm $\|\cdot\|_\infty$ which is the largest absolute value of an eigenvalue, and (for states etc.) the trace norm $\|\cdot\|_1$ which is the sum of the absolute values of the eigenvalues. It is known (and not difficult to prove, using the polar decomposition of α, see Arveson [2]) that

$$\|\alpha\|_1 = \sup_{\|B\|_\infty \leq 1} |\mathrm{Tr}\,(\alpha B)|$$

If α is selfadjoint we may write it as the difference $\alpha_+ - \alpha_-$ of its positive and its negative part. Observe that then

$$\|\alpha\|_1 = \mathrm{Tr}\,(\alpha_+) + \mathrm{Tr}\,(\alpha_-) = \sup_{-\mathbb{1} \leq B \leq \mathbb{1}} \mathrm{Tr}\,(\alpha B)$$

Further we note

Lemma 1 (Continuity) Let ρ, σ states with $\|\rho - \sigma\|_1 \leq \theta \leq \frac{1}{2}$. Then

$$|H(\rho) - H(\sigma)| \leq -\theta \log \frac{\theta}{d} = d\eta\left(\frac{\theta}{d}\right)$$

Proof: See [14]. ∎

III. Quantum channels and codes

The following definition is after [7]: a (discrete memoryless) *quantum channel* (q–DMC) is a completely positive trace preserving mapping φ_* from the states on a C*–algebra $\mathfrak{A}$ into the states on $\mathfrak{L}(\mathcal{H})$.

A *nonstationary q–DMC* is a sequence $(\varphi_{n*})_{n\in\mathbb{N}}$ of q–DMCs, with a global Hilbert space $\mathcal{H}$. This extends the concept of q–DMCs which are contained as *constant* sequences.

From [18] we recall: for a probability distribution P on $\mathfrak{S}(\mathfrak{A})$ let $P\varphi_*$ the average state of the channel φ_*. $H(\varphi_*|P) = \int H(\varphi_*(\rho))dP(\rho)$ is the conditional von Neumann entropy, the mutual information between a distribution and the channel is $I(P, \varphi_*) = H(P\varphi_*) - H(\varphi_*|P)$.

An *n–block code* for a nonstationary quantum channel $(\varphi_{n*})_n$ is a pair (f, D), where f is a mapping from a finite set $\mathcal{M}$ into $\mathfrak{S}(\mathfrak{A})^n$, and D is an observable on $\mathfrak{L}(\mathcal{H})^{\otimes n}$ indexed by $\mathcal{M}' \supset \mathcal{M}$, i.e. a partition of $\mathbb{1}$ into positive operators D_m, $m \in \mathcal{M}'$. The (maximum) error probability of the code is defined as

$$e(f, D) = \max\{1 - \mathrm{Tr}\,(\varphi_*(f(m))D_m) | m \in \mathcal{M}\}$$

We call (f, D) an (n, λ)*–code*, if $e(f, D) \leq \lambda$. The *rate* of an n–block code is defined as $\frac{1}{n}\log|\mathcal{M}|$. Finally define $N(n, \lambda)$ as the maximal size (i.e. $|\mathcal{M}|$) of an (n, λ)–code.

Remark 2: Observe that we did not allow all joint states of the system $\mathfrak{A}_1 \otimes \cdots \otimes \mathfrak{A}_n$ as code words, but only *product states.* This is the restriction under which the current theory was done. It is unknown if the following theorem 3 is still true in the more general model: maybe higher capacities can be achieved there, see the discussion in [16]. With our restriction we may without harm identify a channel mapping φ_* with its image $\mathfrak{W}_\varphi = \varphi_*(\mathfrak{S}(\mathfrak{A}))$ in the set of states on $\mathfrak{L}(\mathcal{H})$ (for then the image of an input state under φ_*^n is a

product state on $\mathfrak{L}(\mathcal{H})^{\otimes n}$).
Generalizing, a nonstationary quantum channel is now a sequence $(\mathfrak{W}_n)_n$ of arbitrary (measureable) subsets of states on a fixed $\mathfrak{L}(\mathcal{H})$. In this spirit we reformulate the definition of an n–block code as a pair (f, D) with a mapping $f : \mathcal{M} \to \mathfrak{W}_1 \times \cdots \times \mathfrak{W}_n$[1] and D as before.
Our main result is

Theorem 3: Let $(\mathfrak{W}_1, \mathfrak{W}_2, \dots)$ a nonstationary q–DMC, and

$$C(\mathfrak{W}_i) = \sup_{P \text{ p.d. on } \mathfrak{W}_i} I(P, \mathfrak{W}_i)$$

Then for every $\lambda \in (0, 1)$

$$\left| \frac{1}{n} \log N(n, \lambda) - \frac{1}{n} \sum_{i=1}^{n} C(\mathfrak{W}_i) \right| \to 0 \text{ as } n \to \infty$$

Proof: Combine the coding theorem VI.11 and the strong converse theorem VIII.14. ∎

This theorem justifies the name *capacity* (of the channel $\mathfrak{W}$) for the quantity $C(\mathfrak{W})$, even in the strong sense of Wolfowitz [21]. Observe that this theorem is a quantum generalization of a theorem by Ahlswede [1].

IV. Shadows

Let $\rho_1, \dots, \rho_n$ states on $\mathfrak{L}(\mathcal{H})$, with diagonalizations $\rho_i = \sum_j q_{j|i} \pi_{ij}$ with one–dimensional projectors π_{ij}. Let $\delta > 0$, and define

$$\mathcal{T}^n_{H,\rho^n,\delta} = \{(j_1, \dots, j_n) : \left| \sum_{i=1}^{n} - \log q_{ij_i} - \sum_{i=1}^{n} H(\rho_i) \right| \le \delta\sqrt{n}\}$$

Define the *entropy–typical projector*[2] *of* ρ^n *with constant* δ as

$$\Pi^n_{H,\rho^n,\delta} = \sum_{(j_1,\dots,j_n)\in\mathcal{T}^n_{H,\rho^n,\delta}} \pi_{1j_1} \otimes \cdots \otimes \pi_{nj_n}$$

Then we have the following

Lemma 4: There is a constant K only depending on d (in fact $K \le \max\{(\log 3)^2, (\log d)^2\}$) such that for arbitrary states $\rho_1, \dots \rho_n$

$$\mathrm{Tr}\,(\rho^n \Pi^n_{H,\rho^n,\delta}) \ge 1 - \frac{K}{\delta^2}$$

Proof: This is just Chebyshev's inequality applied to the random variables $X_i = -\log q_{\cdot|i}$ for the diagonalizations $\rho_i = \sum_j q_{j|i} \pi_{ij}$. Observe that K may be any bound for the variance of the X_i. ∎

[1] Where we identify $(\rho_1, \dots, \rho_n)$ with $\rho^n = \rho_1 \otimes \cdots \otimes \rho_n$.

[2] This is essentially what Schumacher [15] calls *typical subspace.* We add the predicate "entropy" to distinguish it from another typicality concept introduced below.

Motivated by this we define an η–*shadow* of ρ^n to be a positive operator $B \leq \mathbb{1}$ with $\operatorname{Tr}(\rho^n B) \geq \eta$. More generally an operator B as before is called an η–shadow of a set $\mathcal{A}$ of states if it is an η–shadow of every element of $\mathcal{A}$. Concerning the size of shadows we have

Lemma 5: For the entropy–typical projector

$$\left(1-\frac{K}{\delta^2}\right)\exp\left(\sum_{i=1}^n H(\rho_i)-\delta\sqrt{n}\right) \leq$$

$$\leq \operatorname{Tr}\Pi^n_{H,\rho^n,\delta} \leq \exp\left(\sum_{i=1}^n H(\rho_i)+\delta\sqrt{n}\right)$$

Conversely, if B is an η–shadow of ρ^n then

$$\operatorname{Tr} B \geq \left(\eta-\frac{K}{\delta^2}\right)\exp\left(\sum_{i=1}^n H(\rho_i)-\delta\sqrt{n}\right)$$

Proof: Observe that by definition of $\Pi^n = \Pi^n_{H,\rho^n,\delta}$

$$\Pi^n \exp\left(-\sum_{i=1}^n H(\rho_i)-\delta\sqrt{n}\right) \leq \Pi^n\rho^n\Pi^n$$

and

$$\Pi^n\rho^n\Pi^n \leq \Pi^n \exp\left(-\sum_{i=1}^n H(\rho_i)+\delta\sqrt{n}\right)$$

Taking traces yields (with lemma 4) the estimates for the typical projector.
Now let B an η–shadow. Then with the above

$$\operatorname{Tr}(\Pi^n\rho^n\Pi^n B) \leq \exp\left(-\sum_{i=1}^n H(\rho_i)+\delta\sqrt{n}\right)\operatorname{Tr} B$$

On the other hand

$$\begin{aligned}\operatorname{Tr}(\Pi^n\rho^n\Pi^n B) &= \operatorname{Tr}(\rho^n\Pi^n B\Pi^n)\\ &= \operatorname{Tr}(\rho^n B)-\operatorname{Tr}\big((\rho^n-\Pi^n\rho^n\Pi^n)B\big)\\ &\geq \eta-\frac{K}{\delta^2}\end{aligned}$$

■

We introduced the Π^n_H for the reason to *have* a good shadow for any product state, and to estimate *any* shadow for that state. There is another concept of typicality which is specially suited for stationary channels: let ρ a state with diagonalization $\rho = \sum_j q_j\pi_j$, and $\delta > 0$, then define

$$\mathcal{T}^n_{C,\rho,\delta} = \{(j_1,\ldots,j_n) : \forall j\ \left|N(j|j^n)-nq_j\right| \leq \delta\sqrt{n}\}$$

with $N(j|j^n) = |\{i : j_i = j\}|$, and the *constant–typical projector*

$$\Pi^n_{C,\rho,\delta} = \sum_{(j_1,\dots,j_n)\in\mathcal{T}^n_{C,\rho^n,\delta}} \pi_{j_1}\otimes\cdots\otimes\pi_{j_n}$$
$$= \sum_{(j_1,\dots,j_n)\text{ with }\|\sum_{i=1}^n \pi_{j_i}-n\rho\|_\infty\leq\delta\sqrt{n}} \pi_{j_1}\otimes\cdots\otimes\pi_{j_n}$$

Then one has

Lemma 6 (Weak law) Let $\tilde{\rho}, \rho_1, \dots, \rho_n$ states and $\delta, \epsilon > 0$ such that $\|\frac{1}{n}\sum_{i=1}^n \rho_i - \tilde{\rho}\|_\infty \leq \epsilon$. Then

$$\mathrm{Tr}\left(\rho^n \Pi^n_{C,\tilde{\rho},\delta+\epsilon\sqrt{n}}\right) \geq 1 - \frac{1}{\delta^2}$$

Proof: Consider the diagonalization $\tilde{\rho} = \sum_j q_j \pi_j$, and the *corset* map

$$\kappa_* : \sigma \longmapsto \sum_j \pi_j \sigma \pi_j$$

Defining $\rho_i' = \kappa_*(\rho_i)$ we claim that

$$\Pi^n_{C,\frac{1}{n}\sum_{i=1}^n \rho_i',\delta} \leq \Pi^n_{C,\tilde{\rho},\delta+\epsilon\sqrt{n}}$$

Indeed observe that we have

$$\|\frac{1}{n}\sum_{i=1}^n \rho_i' - \tilde{\rho}\|_\infty \leq \|\frac{1}{n}\sum_{i=1}^n \rho_i - \tilde{\rho}\|_\infty \leq \epsilon$$

Thus for $\mathbf{j} = (j_1, \dots, j_n)$ with

$$\|\sum_{i=1}^n \pi_{j_i} - \sum_{i=1}^n \rho_i'\|_\infty \leq \delta\sqrt{n}$$

we have by triangle inequality

$$\|\sum_{i=1}^n \pi_{j_i} - n\tilde{\rho}\|_\infty \leq (\delta + \epsilon\sqrt{n})\sqrt{n}$$

So we can estimate

$$\begin{aligned}\mathrm{Tr}\left(\rho^n \Pi^n_{C,\tilde{\rho},\delta+\epsilon\sqrt{n}}\right) &\geq \mathrm{Tr}\left(\rho^n \Pi^n_{C,\frac{1}{n}\sum_{i=1}^n \rho_i',\delta}\right)\\ &= \mathrm{Tr}\left(\rho'^n \Pi^n_{C,\frac{1}{n}\sum_{i=1}^n \rho_i',\delta}\right)\\ &\geq 1 - \frac{1}{\delta^2}\end{aligned}$$

the last line by Chebyshev's inequality: the last trace is the probability of a set of constant–typical sequences of eigenvectors of the ρ_i' in the product of the measures given by the eigenvalue lists. ■

Concerning the size of this projector we have

Lemma 7: For every state ρ and $0 < \delta \leq \frac{1}{2d}\sqrt{n}$

$$\mathrm{Tr}\,\Pi^n_{C,\rho,\delta} \leq (n+1)^d \exp\left(nH(\rho) + nd\eta\left(\frac{\delta}{\sqrt{n}}\right)\right)$$

Proof: The whole question reduces obviously to counting sequences of eigenvectors of ρ with empirical distribution (also called *type*) close to that given by the eigenvalues of ρ. Each sequence of type P has $P^{\otimes n}$–probability $\exp(-nH(P))$. Thus there are at most $\exp(nH(P))$ of these. Since there at most $(n+1)^d$ many types, and by the continuity of entropy (lemma II.1) the statement follows. ■

The constant typical projectors will be used as shadows of sets (namely of states which satisfy the "average" condition of the weak law).

V. Decoding

Lemma 8 (Tender operator) Let ρ be a state, and X a positive operator with $X \leq \mathbb{1}$ and $1 - \mathrm{Tr}\,(\rho X) \leq \lambda < 1$. Then

$$\|\rho - \sqrt{X}\rho\sqrt{X}\|_1 \leq \sqrt{8\lambda}$$

Proof: Let $Y = \sqrt{X}$ and write $\rho = \sum_k p_k \pi_k$ with one–dimensional projectors π_k and weights $p_k \geq 0$. Now

$$\begin{aligned}
\|\rho - Y\rho Y\|_1^2 &\leq \left(\sum_k p_k \|\pi_k - Y\pi_k Y\|_1\right)^2 \\
&\leq \sum_k p_k \|\pi_k - Y\pi_k Y\|_1^2 \\
&\leq 4\sum_k p_k(1 - \mathrm{Tr}\,(\pi_k Y \pi_k Y)) \\
&\leq 8\sum_k p_k(1 - \mathrm{Tr}\,(\pi_k Y)) \\
&= 8(1 - \mathrm{Tr}\,(\rho Y)) \\
&\leq 8(1 - \mathrm{Tr}\,(\rho X)) \leq 8\lambda
\end{aligned}$$

by triangle inequality, convexity of $x \mapsto x^2$, lemma A.2, $1 - x^2 \leq 2(1-x)$, and $X \leq Y$. ■

Lemma 9 (Tender measurement) Let ρ_a $(a \in \mathcal{A})$ a family of states on $\mathcal{H}$, and D an observable indexed by $\mathcal{B}$. Let $\varphi : \mathcal{A} \longrightarrow \mathcal{B}$ a map and $\lambda > 0$ such that

$$\forall a \in \mathcal{A} \qquad 1 - \mathrm{Tr}\,(\rho_a D_{\varphi(a)}) \leq \lambda$$

(I.e. the observable identifies $\varphi(a)$ from ρ_a with maximal error probability λ). Then the canonically corresponding (trace preserving) quantum operation

$$D_{\mathrm{int}*} : \rho \longmapsto \sum_{b \in \mathcal{B}} \sqrt{D_b}\rho\sqrt{D_b}$$

disturbes the states ρ_a only a little:

$$\forall a \in \mathcal{A} \qquad \|\rho_a - D_{\mathrm{int}*}\rho_a\|_1 \leq \sqrt{8\lambda} + \lambda$$

Furthermore the total observable operation

$$D_{\mathrm{tot}*} : \rho \longmapsto \sum_{b \in \mathcal{B}} [b] \otimes \sqrt{D_b}\rho\sqrt{D_b}$$

satisfies

$$\forall a \in \mathcal{A} \qquad \|[\varphi(a)] \otimes \rho_a - D_{\mathrm{tot}*}\rho_a\|_1 \leq \sqrt{8\lambda} + \lambda$$

Proof: An easy calculation:

$$\begin{aligned}
\|\rho_a - D_{\mathrm{int}*}\rho_a\|_1 &\leq \|\rho_a - \sqrt{D_{\varphi(a)}}\rho_a\sqrt{D_{\varphi(a)}}\|_1 + \\
&\qquad + \sum_{b \neq \varphi(a)} \|\sqrt{D_b}\rho_a\sqrt{D_b}\|_1 \\
&= \|\rho_a - \sqrt{D_{\varphi(a)}}\rho_a\sqrt{D_{\varphi(a)}}\|_1 + \\
&\qquad + \sum_{b \neq \varphi(a)} \mathrm{Tr}\,(\rho_a D_b) \\
&\leq \sqrt{8(1 - \mathrm{Tr}\,(\rho_a D_{\varphi(a)}))} + \\
&\qquad + 1 - \mathrm{Tr}\,(\rho_a D_{\varphi(a)}) \\
&\leq \sqrt{8\lambda} + \lambda
\end{aligned}$$

using triangle inequality and lemma 8. The second part (which actually implies the first) is similar. ■

Remark 10: If we modify the statement of the lemma to that the average error in identifying $\varphi(a)$ from ρ_a should be at most $\bar{\lambda}$ (relative a distribution on $\mathcal{A}$), then also the distortion bound of the lemma holds — on average: see appendix A.

VI. Maximal codes

Theorem 11 (Maximal codes) For $0 < \tau, \lambda < 1$ there is a constant K' and $\delta > 0$ such that for every nonstationary q–DMC $(\mathfrak{W}_i)_i$, distributions P_i on $\mathfrak{W}_i$ and $\mathcal{A} \subset \mathfrak{W}^n$ with $P^n(\mathcal{A}) \geq \tau$ there exists an (n, λ)–code (f, D) with the properties

$$\forall m \in \mathcal{M}\ f(m) \in \mathcal{A} \text{ and } \mathrm{Tr}\, D_m \leq \mathrm{Tr}\,\Pi^n_{H, f(m), \delta}$$

$$\begin{aligned}
\log|\mathcal{M}| &\geq H(P^n\mathfrak{W}^n) - H(\mathfrak{W}^n|P^n) - K'\sqrt{n} \\
&= \sum_{i=1}^{n} (H(P_i\mathfrak{W}_i) - H(\mathfrak{W}_i|P_i)) - K'\sqrt{n}
\end{aligned}$$

Proof: On every $\mathfrak{W}_i$ the entropy H is a random variable with expectation $H(\mathfrak{W}_i|P_i)$ and variance bounded by $(\log d)^2$. Define $\delta = \max\{\sqrt{2/\lambda}, \sqrt{2/\tau}\log d\}$, then by Chebyshev's inequality the set

$$\mathcal{A}' = \{\rho^n \in \mathcal{A} : \left| H(\rho^n) - \sum_{i=1}^{n} H(\mathfrak{W}_i|P_i) \right| \leq \delta\sqrt{n}\}$$

has probability $P^n(\mathcal{A}') \geq \tau/2$. Now let (f, D) a maximal (n, λ)–code with

$$\forall m \in \mathcal{M} \quad f(m) \in \mathcal{A}' \text{ and } \operatorname{Tr} D_m \leq \operatorname{Tr} \Pi^n_{H,f(m),\delta}$$

Define $B = \sum_{m\in\mathcal{M}} D_m$. We claim that with $\eta = \min\{1-\lambda, \lambda^2/32\}$

$$\forall \rho^n \in \mathcal{A}' \quad \operatorname{Tr}(\rho^n B) \geq \eta$$

This is clear for codewords, and true for the other states because otherwise we could extend our code by the codeword ρ^n with corresponding observable operator

$$D = \sqrt{\mathbb{1} - B}\,\Pi^n_{H,\rho^n,\delta}\sqrt{\mathbb{1} - B}$$

which clearly satisfies the trace bound (note that $B + D \leq \mathbb{1}$): to see this apply lemma V.8 to obtain

$$\|\rho^n - \sqrt{\mathbb{1} - B}\rho^n\sqrt{\mathbb{1} - B}\|_1 \leq \sqrt{8\eta} \leq \frac{\lambda}{2}$$

Thus

$$\begin{aligned}
\operatorname{Tr}(\rho^n\sqrt{\mathbb{1} - B}\,\Pi^n_{H,\rho^n,\delta}\sqrt{\mathbb{1} - B}) &= \\
&= \operatorname{Tr}\left(\rho^n \Pi^n_{H,\rho^n,\delta}\right) - \\
&\quad - \operatorname{Tr}\left((\rho^n - \sqrt{\mathbb{1} - B}\rho^n\sqrt{\mathbb{1} - B})\Pi^n_{H,\rho^n,\delta}\right) \\
&\geq (1 - \frac{\lambda}{2}) - \frac{\lambda}{2} = 1 - \lambda
\end{aligned}$$

So B is an η–shadow of $\mathcal{A}'$, and consequently

$$\operatorname{Tr}(P^n\mathfrak{W}^n B) \geq \eta\tau/2$$

By lemma IV.5 there is K with

$$\operatorname{Tr} B \geq \exp\left(\sum_{i=1}^{n} H(P_i\mathfrak{W}_i) - K\sqrt{n}\right)$$

On the other hand

$$\begin{aligned}
\operatorname{Tr} B = \sum_{m\in\mathcal{M}} \operatorname{Tr} D_m &\leq \sum_{m\in\mathcal{M}} \operatorname{Tr} \Pi^n_{H,f(m),\delta} \\
&\leq |\mathcal{M}| \exp\left(\sum_{i=1}^{n} H(\mathfrak{W}_i|P_i) + 2\delta\sqrt{n}\right)
\end{aligned}$$

the last inequality again by lemma IV.5, and we are done. ■

Remark 12: We can strengthen the theorem to that all the D_m are projectors. The proof goes through unchanged but for the construction of the code extension: there we take the support of the above D. The trace estimate holds because the trace of a projector is the dimension of its range.

VII. Approximation of channels

We have continuum many states on $\mathfrak{L}(\mathcal{H})$, and even more channels, so we introduce a simple approximation scheme: a partition $\mathfrak{Z}$ of $\mathfrak{S}(\mathfrak{L}(\mathcal{H}))$ into t sections $\mathfrak{Z}_1, \ldots, \mathfrak{Z}_t$ each having $\|\cdot\|_1$–diameter at most $\theta > 0$ is called *θ–fine*. The relation of the parameters t and θ is:

Lemma 13: For any $\theta > 0$ there is a θ–fine partition of $\mathfrak{S}(\mathfrak{L}(\mathcal{H}))$ into $t \leq C\theta^{-d^2}$ sections, with a constant C depending only on d.

Proof: The set of states is $\|\cdot\|_1$–isometric to the set of positive $d \times d$–matrices with trace one. This is obviously a compact set of real dimension $d^2 - 1$. It is contained in the set of all selfadjoint matrices with the real and imaginary parts of all its entries in the interval $[-1, 1]$ which is geometrically a d^2–dimensional cube. Now obviously we may decompose this cube into $(2\sqrt{2}d^3)^{d^2}\theta^{-d^2}$ cubes of edge length $\theta/(d^3\sqrt{2})$. We claim that for two states ρ, ρ' in the same small cube $\|\rho - \rho'\|_1 \leq \theta$. But this follows from the fact that a matrix with all entries absolutely bounded by ϵ has all its eigenvalues bounded by $d^2\epsilon$, which is straightforward (and rather crude). ■

We close this section with two definitions: the *$\mathfrak{Z}$–type* of a state ρ^n is the empirical distribution on sections in which $\mathfrak{Z}_j$ has weight proportional to the number of $\rho_i \in \mathfrak{Z}_j$. The *$\mathfrak{Z}$–class* of a channel $\mathfrak{W}_i$ is the set of sections $\mathfrak{Z}_j$ which have nonempty intersection with $\mathfrak{W}_i$.

Obviously the number of $\mathfrak{Z}$–types is bounded by $(n+1)^t$, the number of $\mathfrak{Z}$–classes is bounded by 2^t.

VIII. Strong converse

Theorem 14 (Strong Converse) For every $\lambda \in (0, 1)$ and $\epsilon > 0$ there is $n_0 = n_0(\lambda, \epsilon)$ such that for every $n \geq n_0$ and every nonstationary q–DMC $(\mathfrak{W}_i)_i$

$$\log N(n, \lambda) \leq \sum_{i=1}^{n} C(\mathfrak{W}_i) + n\epsilon$$

Proof: Let (f, D) an (n, λ)–code. Consider a θ–fine partition $\mathfrak{Z}$ of $\mathfrak{S}(\mathfrak{L}(\mathcal{H}))$ into t sections and choose representatives $\sigma_j \in \mathfrak{Z}_j$. For every ($\mathfrak{Z}$–)class γ let I_γ the set of indices $i \in [n]$ with $\mathfrak{W}_i$ of class γ. Consider the ($\mathfrak{Z}$–)types of the restrictions $f(m)^{I_\gamma}$ of the codewords to the positions I_γ. For each γ with $I_\gamma \neq \emptyset$ there is a type P_γ occuring in a fraction of at least $(|I_\gamma| + 1)^{-t}$ of the codewords. Successively choosing subcodes we arrive at a code $\mathcal{M}'$ with at least $|\mathcal{M}| \cdot (n+1)^{-t2^t}$ codewords and $f(m)^{I_\gamma}$ of type P_γ for all $m \in \mathcal{M}'$, whenever $I_\gamma \neq \emptyset$.

For each i, $i \in I_\gamma$ choose states $\tilde\rho_{ij} \in \mathfrak{W}_i \cap \mathfrak{Z}_j$ and define a distribution P_i on $\mathfrak{W}_i$ by $P_i(\tilde\rho_{ij}) = P_\gamma(j)$. Finally let $\tilde\rho_{i\gamma} = P_i\mathfrak{W}_i = \sum_j P_\gamma(j)\tilde\rho_{ij}$ and $\tilde\sigma_\gamma = \sum_j P_\gamma(j)\tilde\sigma_j$.
For classes γ with $|I_\gamma| \geq n2^{-2t}$ (which we call *good*) define (with some $\delta > 0$)

$$\Pi_\gamma = \Pi^{I_\gamma}_{C,\tilde\sigma_\gamma,\delta+\theta\sqrt{|I_\gamma|}} \quad \text{in } \mathfrak{L}(\mathcal{H})^{\otimes I_\gamma}$$

For bad γ define $\Pi_\gamma = \mathbb{1}$ in $\mathfrak{L}(\mathcal{H})^{\otimes I_\gamma}$. Then by the weak law lemma IV.6 for *every* γ

$$\forall m \in \mathcal{M}' \quad \mathrm{Tr}\left(f(m)^{I_\gamma}\Pi_\gamma\right) \geq 1 - \frac{1}{\delta^2}$$

and thus defining $\Pi_0 = \bigotimes_\gamma \Pi_\gamma$ we obtain

$$\forall m \in \mathcal{M}' \quad \mathrm{Tr}\left(f(m)\Pi_0\right) \geq 1 - \frac{2^t}{\delta^2}$$

Now assume that $n2^{-2t}$ is large enough and θ is small enough so that according to lemmas IV.7 and II.1 we have for good γ

$$\begin{aligned}\mathrm{Tr}\,\Pi_\gamma &\leq \exp\left(|I_\gamma|(H(\tilde\sigma_\gamma)+\epsilon)\right)\\ &\leq \exp\left(\sum_{i\in I_\gamma} H(\tilde\rho_{i\gamma}) + 2|I_\gamma|\epsilon\right)\end{aligned}$$

Hence we get (collecting the contributions of good and bad classes)

$$\mathrm{Tr}\,\Pi_0 \leq \exp\left(\sum_{i=1}^n H(\tilde\rho_{i\gamma}) + 2n\epsilon + n2^{-t}\log d\right)$$

Now consider the code (f', D') with $f' = f|_{\mathcal{M}'}$ and $D'_m = \Pi_0 D_m \Pi_0$ for $m \in \mathcal{M}'$. By the above considerations and lemma V.8 it is an $(n, \lambda + \sqrt{8}2^{t/2}\delta^{-1})$–code. Assuming $\sqrt{8}2^{t/2}\delta^{-1} \leq \frac{1-\lambda}{2}$, by lemma IV.5 we get

$$\mathrm{Tr}\,D'_m \geq \exp\left(\sum_{i=1}^n H(\mathfrak{W}_i|P_i) - n\epsilon\right)$$

if n is large enough. So we get

$$\begin{aligned}\mathrm{Tr}\,\Pi_0 &\geq \sum_{m\in\mathcal{M}'} D'_m\\ &\geq |\mathcal{M}'|\exp\left(\sum_{i=1}^n H(\mathfrak{W}_i|P_i) - n\epsilon\right)\end{aligned}$$

and thus

$$
\begin{aligned}
|\mathcal{M}| &\leq (n+1)^{t2^t} \exp\Bigg(\sum_{i=1}^{n} (H(P_i\mathfrak{W}_i) - H(\mathfrak{W}_i|P_i)) + \\
&\qquad\qquad + 3n\epsilon + n2^{-t}\log d \Bigg) \\
&\leq \exp\left(\sum_{i=1}^{n} (H(P_i\mathfrak{W}_i) - H(\mathfrak{W}_i|P_i)) + 5n\epsilon \right) \\
&\leq \exp\left(\sum_{i=1}^{n} C(\mathfrak{W}_i) + 5n\epsilon \right)
\end{aligned}
$$

if we can adjust our parameters accordingly: choose for example $t = \lceil \frac{1}{3}\log n \rceil$ with $\theta \leq \left(\frac{3C}{\log n}\right)^{d^{-2}}$ (which is possible by lemma VII.13), $\delta = n^{1/3}$, and let n large enough. ■

IX. Conclusion

The exclusive use of constant–typical projectors yields even more insights for the stationary case (for simplicity assume a finite alphabet, i.e. $|\mathfrak{W}| < \infty$): coding theorem and strong converse under linear constraints (of a continuous cost function, see Holevo [8] for definitions and capacity formula), simply because satisfying the linear constraint is a property invariant under permutations of the word's letters, i.e. a property of whole *types* of input sequences, not just individual sequences (see Csiszár, Körner [4]).

Also, it may be interesting to note that the proof of the converse did not use the Holevo bound [6] (which usually is derived from Uhlmann's monotonocity of the quantum I–divergence [17], which in turn is a consequence of a deep concavity result, the Wigner–Yanase–Dyson conjecture, first proved by Lieb [12]). Thus in turn we get a new proof of the Holevo bound.

Finally some general remarks: note that our proof of the coding theorem follows the original idea of Feinstein [5], the proof of the converse is after Wolfowitz [21]. In fact we exploited a close analogy between the classical and the quantum model (e.g. distributions become states, sets become positive operators upper bounded by $\mathbb{1}$, sets of typical sequences become typical projectors, etc.). In particular we used an analog of set intersection for positive operators (in both proofs it is crucial for the construction of the respective code modification). Now set intersection has some nice properties: it is the infimum in the partial order by inclusion, it is commutative, and distributive (with unions). Birkhoff and von Neumann [3] conceived an analog for projection operators on a Hilbert space (with the aim to construct a logic of propositions about a quantum system), presupposing that it has to be the order infimum. It turns out that beside good properties (commutativity, the "intersection" of two operators is upper bounded by both) it has also bad: it is not distributive. Our decision was for an "intersection" that is not commutative, and has one–sided order and distributivity properties: for $0 \leq A, B \leq \mathbb{1}$

$$0 \leq \sqrt{A}B\sqrt{A} \leq A, \quad \mathrm{Tr}\,(\sqrt{A}B\sqrt{A}) \leq \mathrm{Tr}\,B$$

$$\sqrt{A}(B+C)\sqrt{A} = \sqrt{A}B\sqrt{A} + \sqrt{A}C\sqrt{A}$$

But generically

$$\sqrt{A}B\sqrt{A} \neq \sqrt{B}A\sqrt{B} \text{ and } \sqrt{A}B\sqrt{A} \not\leq B$$

And (most importantly) we have the tender operator lemma V.8. The success of the construction indicates its possible relevance to propositional calculi for quantum mechanics.

Acknowledgements

Thanks to Peter Löber and to Prof. Rudolf Ahlswede for discussions on various aspects of the present work.

Appendix

A. Distortion and fidelity

This appendix compares the distortion measure D coming from the trace norm $\|\cdot\|_1$

$$D(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1$$

on the states of a quantum system with the (pure state) fidelity F: $F(\rho, \sigma) = \mathrm{Tr}\,(\sigma\rho)$ for ρ pure. First note that $\|\cdot\|_1$ is indeed a distortion measure:

$$\|\rho_1 \otimes \rho_2 - \sigma_1 \otimes \sigma_2\|_1 \leq \|\rho_1 - \sigma_1\|_1 + \|\rho_2 - \sigma_2\|_1$$

In contrast, the fidelity has a multiplicative law under tensor products, hence

$$1 - F(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) \leq 1 - F(\rho_1, \sigma_1) + 1 - F(\rho_2, \sigma_2)$$

so $1 - F$ is also a distortion measure.

Lemma 1 (Pure state) Let $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\phi\rangle\langle\phi|$ pure states. Then

$$1 - F(\rho, \sigma) = D(\rho, \sigma)^2$$

Proof: W.l.o.g. we may assume $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ and $|\phi\rangle = \alpha|0\rangle - \beta|1\rangle$ ($|\alpha|^2 + |\beta|^2 = 1$). A straightforward calculation shows $F = (|\alpha|^2 - |\beta|^2)^2$, and $D = 2|\alpha\beta|$. Now

$$\begin{aligned} 1 - F &= 1 - (|\alpha|^2 - |\beta|^2)^2 \\ &= (1 + |\alpha|^2 - |\beta|^2)(1 - |\alpha|^2 + |\beta|^2) \\ &= 4|\alpha\beta|^2 = D^2 \end{aligned}$$

■

Lemma 2 (Mixed state) Let σ an arbitrary mixed state (and ρ pure as above). Then

$$D \geq 1 - F \geq D^2$$

Proof: Write $\sigma = \sum_j q_j \pi_j$ with pure states π_j. Then

$$\begin{aligned} 1 - F(\rho, \sigma) &= \sum_j q_j \left(1 - F(\rho, \pi_j)\right) = \sum_j q_j D(\rho, \pi_j)^2 \\ &\geq \left(\sum_j q_j D(\rho, \pi_j)\right)^2 \geq D(\rho, \sigma)^2 \end{aligned}$$

Conversely: extend ρ to an observable $(\rho, \mathbb{1} - \rho)$ and consider the trace preserving quantum operation

$$\kappa_* : \sigma \longmapsto \rho\sigma\rho + (\mathbb{1} - \rho)\sigma(\mathbb{1} - \rho)$$

Then (with monotonocity of $\|\cdot\|_1$ under quantum operations)

$$2D = \|\rho - \sigma\|_1 \geq \|\kappa_*\rho - \kappa_*\sigma\|_1 = \|\rho - \kappa_*\sigma\|_1$$

(since $\rho = \kappa_*\rho$). Hence with $F = \mathrm{Tr}\,(\sigma\rho)$

$$\begin{aligned} 2D &\geq \left\|(1 - F)\rho - \mathrm{Tr}\,(\sigma(\mathbb{1} - \rho))(\mathbb{1} - \rho)\right\|_1 \\ &= (1 - F) + (1 - F) = 2(1 - F) \end{aligned}$$

■

Observe that the inequalities of this lemma still hold if only $\sum_j q_j \leq 1$.

Lemma 3 (Ensembles) Let ρ_i pure states, σ_i arbitrary states, and p_i probabilities. Then the average fidelity $\bar{F} = \sum_i p_i F(\rho_i, \sigma_i)$ and the average distortion $\bar{D} = \sum_i p_i D(\rho_i, \sigma_i)$ satisfy

$$\bar{D} \geq 1 - \bar{F} \geq \bar{D}^2$$

Proof: From lemma 2 the inequality $\bar{D} \geq 1 - \bar{F}$ is immediate, and also the inequality $1 - \bar{F} \geq \bar{D}^2$ (with the convexity of the square function). ■

Observe that in the main text we only need lemma 2. But we believe that this presentation has its own value: for example it shows that the fidelity criterion in quantum coding [15] may be replaced by a $\|\cdot\|_1$–criterion, which leads to a short proof of the converse, see [11].

References

[1] R. Ahlswede, "Beiträge zur Shannonschen Informationstheorie im Falle nichtstationärer Kanäle" (in German), Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **10** (1968), 1-42

[2] W. Arveson, *An Invitation to C^*–Algebras*, New York, Heidelberg, Berlin 1976

[3] G. Birkhoff, J. von Neumann, "The Logic of Quantum Mechanics", Ann. Math. **37**,4 (1936), 823–843

[4] I. Csiszár, J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, New York 1981

[5] A. Feinstein, "A new basic theorem of information theory", IRE Trans. Inf. Theory **4** (Sept. 1954), 2–22

[6] A. S. Holevo, Problemy Peredachi Informatsii **9**,3 (1973), 3–11 (English translation: "Bounds for the quantity of information transmitted by a quantum channel", Probl. Inf. Transm. **9**,3 (1973), 177–183)

[7] A. S. Holevo, "Problems in the mathematical theory of quantum communication channels", Rep. Math. Phys. **12**,2 (1977), 273–278

[8] A. S. Holevo, "On Quantum Communication Channels with Constrained Inputs", LANL eprint `quant-ph/9705054`

[9] A. S. Holevo, "Coding Theorems for Quantum Channels", Tamagawa University Research Review no.4, 1998 (an extended version as LANL eprint `quant-ph/9809023`)

[10] A. S. Holevo, "The Capacity of the Quantum Channel with General Signal States", IEEE Trans. Inf. Theory **44**,1 (1998), 269–273 (also available as LANL eprint `quant-ph/9611023`)

[11] M. Horodecki, "Limits for compression of quantum information carried by ensembles of mixed states", Phys. Rev. A **57** (1998), 3364–3369

[12] E. H. Lieb, "Convex Trace Functions and the Wigner–Yanase–Dyson Conjecture", Advan. Math. **11** (1973), 267–288

[13] T. Ogawa, H. Nagaoka, "Strong Converse to the Quantum Channel Coding Theorem", LANL eprint `quant-ph/9808063`

[14] M. Ohya, D. Petz, *Quantum Entropy and its Use*, Berlin, Heidelberg, New York 1993

[15] B. Schumacher, "Quantum Coding", Phys. Rev. A **51**,4 (1995), 2738–2747

[16] B. Schumacher, M. Westmoreland, "Sending classical information via noisy quantum channels", Phys. Rev. A **56**,1 (1997), 131–138

[17] A. Uhlmann,"Relative Entropy and the Wigner–Yanase–Dyson–Lieb Concavity in an Interpolation Theory", Comm. Math. Phys. **54** (1977), 21–32

[18] A. Winter, "Languages of Quantum Information Theory", Preprint Ergänzungsreihe 98–009 of SFB 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld, 1998 (also LANL eprint `quant-ph/9807008`)

[19] A. Winter, "Coding Theorem and Strong Converse for Quantum Channels", Submitted to IEEE Trans. Inf. Theory. Preprint 98–074 of SFB 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld, 1998

[20] A. Winter, "Rate and Error Estimates for Quantum Channels", to appear

[21] J. Wolfowitz, *Coding Theorems of Information Theory*, Berlin 1964