Coding Theorems of Quantum Information Theory

Andreas Winter

Dissertation zur Erlangung des Doktorgrades,
vorgelegt der Fakultät für Mathematik,
Universität Bielefeld

13. April 1999

Abstract

Coding theorems and (strong) converses for memoryless quantum communication channels and quantum sources are proved: for the quantum source the coding theorem is reviewed, and the strong converse proven. For classical information transmission via quantum channels we give a new proof of the coding theorem, and prove the strong converse, even under the extended model of nonstationary channels. As a by–product we obtain a new proof of the famous Holevo bound. Then multi–user systems are investigated, and the capacity region for the quantum multiple access channel is determined. The last chapter contains a preliminary discussion of some models of compression of correlated quantum sources, and a proposal for a program to obtain operational meaning for quantum conditional entropy. An appendix features the introduction of a notation and calculus of entropy in quantum systems.

Acknowledgements

This work grew out of an attempt to learn about quantum mechanics and information theory at the same time. My innocence in the one field and almost–ignorance in the other enabled me to proceed in both in the kind of naïve way that is necessary for success.

Thanks are due to Prof. Rudolf Ahlswede, who became my Doktorvater out of his desire to explore quantum information theory, and thus I was his student: I am indebted to him for his teaching and constant support. To Prof. Martin Aigner (FU Berlin) who, after being my Diplom advisor, obtained for me the opportunity to study with Prof. Ahlswede. Also to Prof. Volker Strassen (Uni Konstanz) whom I regard as a teacher, though he never was in a formal sense, for widening my mathematical interests, and for educating me in combinatorics and information theory, without which preparation I would have been lost at Bielefeld. Not least I want to thank my collegue Peter Löber for many discussions during the last one and a half years, and sparing me the ill fate of isolation by working in close neighbourhood: mathematically (as one also pursuing problems of quantum communication), and physically (by sharing the bureau room). Finally, the justification of the present work's pages was done by hand based on explanations by Prof. Eberhard König (FU Berlin), which hereby I acknowledge.

Gutachter: Prof. Dr. R. Ahlswede und Prof. Dr. F. Götze. Datum der mündlichen Prüfung: 1. Juli 1999.
Gedruckt auf alterungsbeständigem Papier ∞ ISO 9706.

Contents

Introduction

In the present thesis problems of information in quantum systems are discussed, mainly in the context of coding problems of various kind. Thus we follow a line of research initiated by SHANNON (1948), where informational–operational meaning was lent to terms like entropy, information, capacity, building on models of a stochastic nature. This is where quantum theory enters, which is generally understood to be a stochastic theory (starting with BORN (1926), now in any modern textbook, e.g. PERES (1995)). A stochastic theory however of a novel type: it was soon understood that the statistical predictions of quantum theory cannot be described in ordinary ("classical") stochastic theories (EINSTEIN et al. (1935), BELL (1964)), and this is formally mirrored in the necessity to introduce a "noncommutative probability".

These observations led physicists during the 1960s to speculate about the role of quantum probabilism in information theory: cf. the works of GORDON (1964), LEVITIN (1969), and FORNEY (1963). HOLEVO (1973) however is to be credited with founding an appropriate mathematical theory (after a first step by STRATONOVICH (1966)) and proving the justly named HOLEVO bound on quantum channel capacities. This work was extended subsequently by HOLEVO (1979). Apart from this and formulating the definite model (HOLEVO (1977), relying on earlier clarifying work on quantum stochastics by LUDWIG (1954), HOLEVO, and DAVIES & LEWIS) efforts concentrated on the analysis of specific restrictive or highly symmetric situations.

Then progress in foundations ceased, until the stormy revival and extension of the subject in 1994, which year saw two important contributions: the quantum algorithm of SHOR (1994) for factoring integers, proving the power of quantum information processing, and by SCHUMACHER (1995) the successful interpretation of VON NEUMANN entropy as asymptotic source coding rate for quantum information (at the same time establishing quantum information at all as a quantity, distinguished from what is now called "classical information". The reader should be aware however that it was known from the early days of quantum theory on that operationally quantum states are "more" than the knowlegde we can acquire about them. A true expression of this qualitative distinction is the *no–cloning theorem* of WOOTTERS & ZUREK (1982), stating that quantum states cannot be duplicated, i.e. "copied", whereas classical data obviously can).

Both works continue to exert a tremendous influence on the new thinking about quantum information theory. After that soon the coding theorem complementing the HOLEVO bound was proved (HAUSLADEN et al. (1997), HOLEVO (1998a), SCHUMACHER & WEST-

MORELAND (1997)), and today we face a variety of classical, quantum, or mixed information models, some of which at least we understand.

The present work opens and closes with quantum information: beginning with a review of SCHUMACHER's quantum source coding, to which we contribute the strong converse, ending with some speculations about multiple quantum source coding. In between we deal with transmission of classical information via quantum channels. Here our achievements include new proofs of the channel coding theorem (which is new for nonstationary channels), and the completely new strong converse (independently OGAWA & NAGAOKA (1998) have proved the strong converse for stationary and finite alphabet channels by a different method), estimates on the reliability function, and — as a by–product a new proof of the HOLEVO bound. In the third chapter we determine the capacity region of quantum multiple access channels, using our results on multiple quantum source coding with side information from the fourth chapter, where also a number of simple estimates on the rate region and some examples are discussed. Among the positive results of this part are the weak subadditivity for the so–called *coherent information* (while the ordinary subadditivity one would expect fails), and the determination of the rate region for multiple classical source coding with quantum side information at the decoder. Thus we completely skirt all questions of channel coding of quantum information and noise protection of quantum information by quantum error correcting codes, these issues only entering implicitely in the discussion of multiple quantum source coding. Also we choose to stay with discrete (i.e. finite, or, in the quantum case, finite–dimensional) and memoryless systems: this is not an essential restriction for our results, but allows to work consistently with techniques of a combinatorial flavor and to skip technicalities (such as finite variance conditions etc.) which, at the present state of techniques, could not have been avoided. The restriction is further justified by the ignorance on many questions even in this somewhat narrow setting. An appendix contains the necessary elements of quantum probabilistic theory and a calculus of entropy and information in quantum systems. It will be referred to for any concept of that field needed in the main text.

Parts of this work have been pre–published in the author's work: appendix A is distilled from its (very inadequate) predecessor WINTER (1998c), chapter I is from WINTER (1999c), and the results of chapters II and III were reported by WINTER (1998b), WINTER (1999a), WINTER (1999b), and WINTER (1998a).

I have tried to give due credit (or else a reasonable reference) to any result of some importance, especially in the main text. If there is no credit it is implicit that I am the inventor. However this does not apply to a number of propositions of less weight, especially in the appendix, which I found on my own but which I regard as "folklore", and thus never tried to trace them back to an original inventor.

Chapter I

Quantum Source Coding

In this chapter quantum information and the notion of its compression are introduced. To prove the corresponding coding theorem and strong converse basic techniques are developed: a relation between fidelity and trace norm distance, different notions of *typical subspace*, and an estimate on general η–*shadows*. Finally we comment on the relation to classical source coding.

Models of quantum data compression

Fix the complex Hilbert space $\mathcal{H}$, $d = \dim \mathcal{H} < \infty$. $\mathfrak{L}(\mathcal{H})$ denotes the algebra of (bounded) linear operators of $\mathcal{H}$, $\mathfrak{L}(\mathcal{H})_*$ its predual under the trace pairing.[1]

A *(discrete memoryless) quantum source* (q–DMS) is a pair (P, P) with a finite set $\mathsf{P} \subset \mathfrak{L}(\mathcal{H})_*$ of pure states on $\mathfrak{L}(\mathcal{H})$ and a p.d. P on P. The *average state* of the source is $P\mathsf{P} = \sum_{\pi \in \mathsf{P}} P(\pi)\pi$.

An n–*block code* for the q–DMS (P, P) is a pair $(\varepsilon_*, \delta_*)$ where $\varepsilon_* : \mathsf{P}^n \to \mathfrak{L}(\mathcal{K})_*$ maps P^n into the states on $\mathfrak{L}(\mathcal{K})$ (with some Hilbert space $\mathcal{K}$), and $\delta_* : \mathfrak{L}(\mathcal{K})_* \to \mathfrak{L}(\mathcal{H})_*^{\otimes n}$ is trace preserving and completely positive (i.e. it is a physical state transformation, see appendix A).

We say that $(\varepsilon_*, \delta_*)$ is *quantum encoding* if ε_* is the restriction to P^n of a trace preserving and completely positive map $\varepsilon_* : \mathfrak{L}(\mathcal{H})_*^{\otimes n} \to \mathfrak{L}(\mathcal{K})_*$. If there is no condition on ε_* we say that $(\varepsilon_*, \delta_*)$ is *arbitrary encoding.*

For an n–block code $(\varepsilon_*, \delta_*)$ define

1. the *(average) fidelity*

$$\bar{F} = \bar{F}(\varepsilon_*, \delta_*) = \sum_{\pi^n \in \mathsf{P}^n} P^n(\pi^n) \cdot \mathrm{Tr}\,((\delta_* \varepsilon_* \pi^n)\pi^n),$$

[1]For these notions (algebras, states, operations, trace pairing, trace norm, etc.) see appendix A, section *Quantum systems.*

2. the *(average) distortion*

$$\bar{D} = \bar{D}(\varepsilon_*, \delta_*) = \sum_{\pi^n \in \mathsf{P}^n} P^n(\pi^n) \cdot \frac{1}{2} \|\delta_* \varepsilon_* \pi^n - \pi^n\|_1 \ ,$$

3. the *entanglement fidelity* (see SCHUMACHER (1996))

$$F_e = F_e(\varepsilon_*, \delta_*) = \mathrm{Tr}\ \left(((\delta_* \varepsilon_* \otimes \mathrm{id}) \Psi_{P\mathsf{P}}^{\otimes n}) \Psi_{P\mathsf{P}}^{\otimes n}\right) ,$$

   where $\Psi_{P\mathsf{P}}$ is a purification of $P\mathsf{P}$, i.e. it is a pure state on an extended system (by tensor product with some space $\mathcal{H}_0$), and $P\mathsf{P} = \Psi_{P\mathsf{P}}|_{\mathfrak{L}(\mathcal{H})}$ (cf. SCHUMACHER (1996) who proves that F_e does not depend on the purification chosen). Note that this makes sense only if $(\varepsilon_*, \delta_*)$ is quantum encoding.

Observe that generally $\rho^n = \rho_1 \otimes \cdots \otimes \rho_n$ denotes a product state of n factors, while $\rho^{\otimes n} = \rho \otimes \cdots \otimes \rho$ is the n–fold tensor power of ρ.

Theorem I.1

$$\bar{D}^2 \leq 1 - \bar{F} \leq \bar{D} \ \textit{and}\ 1 - \bar{F} \leq 1 - F_e \ .$$

Proof. For the last inequality see SCHUMACHER (1996). The first double inequality follows from lemma I.3 below by linearity, and by convexity of the square function. □

A digression on fidelity First note that both $D(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1$ and $1 - F(\rho, \sigma) = 1 - \mathrm{Tr}\,(\rho\sigma)$ obey a triangle inequality:

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

and

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

Lemma I.2 (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 I.3 (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 the observable $(\rho, \mathbb{1} - \rho)$ and consider the quantum operation

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

Then (with monotonicity of $\|\cdot\|_1$ under quantum operations, see appendix A, section *Quantum systems*)

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

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

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

for a state π supported in $\mathbb{1} - \rho$. $\square$

Observe that the inequalities of this lemma still hold if only $\sum_j q_j \leq 1$. To close our digression we want to note two useful lemmata concerning "good" measurements:

Lemma I.4 (Tender operator) *Let ρ be a state, and X a positive operator with $X \leq \mathbb{1}$ and $1 - \mathrm{Tr}\,(\rho X) \leq \lambda \leq 1$. Then*

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

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 I.3, $1 - x^2 \leq 2(1-x)$, and $X \leq Y$. $\square$

Lemma I.5 (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 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[2]

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

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:

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

using triangle inequality and lemma I.4. The second part (which actually implies the first) is similar. □

Remark I.6 *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.*

Let us return to the source coding schemes: The n–block code $(\varepsilon_*, \delta_*)$ is called an $(n, \lambda)_{\bar{F}}$*–code* if $1 - \bar{F}(\varepsilon_*, \delta_*) \leq \lambda$. Similarly an $(n, \lambda)_{F_e}$–code is defined. The n–block code $(\varepsilon_*, \delta_*)$ is called an $(n, \lambda)_{\bar{D}}$*–code* if $\bar{D}(\varepsilon_*, \delta_*) \leq \lambda$.

[2] See also appendix A, section *Common tongue*, for D_{int} and D_{tot}.

The *rate* of an n–block code $(\varepsilon_*, \delta_*)$ is defined as $R(\varepsilon_*, \delta_*) = \frac{1}{n} \log \dim \mathcal{K}$.[3]

From the previous theorem it is clear that the most restrictive model is where we have to find an $(n, \lambda)_{F_e}$–code with quantum encoding, whereas the most powerful model is where we have to find an $(n, \lambda)_{\bar{F}}$–code with arbitrary encoding (equivalently we may use $\bar{D}$). Now we define for a q–DMS (P, P)

1. the λ–(quantum,F_e)–*rate* as
$$R_{q,F_e}(\lambda) = \limsup_{n\to\infty} \min\{R(\varepsilon_*, \delta_*) : (\varepsilon_*, \delta_*) \text{ an } (n, \lambda)_{F_e}\text{–code with qu. encoding}\},$$
2. the λ–(quantum,$\bar{F}$)–*rate* as
$$R_{q,\bar{F}}(\lambda) = \limsup_{n\to\infty} \min\{R(\varepsilon_*, \delta_*) : (\varepsilon_*, \delta_*) \text{ an } (n, \lambda)_{\bar{F}}\text{–code with qu. encoding}\},$$
3. the λ–(arbitrary,$\bar{F}$)–*rate* as
$$R_{a,\bar{F}}(\lambda) = \limsup_{n\to\infty} \min\{R(\varepsilon_*, \delta_*) : (\varepsilon_*, \delta_*) \text{ an } (n, \lambda)_{\bar{F}}\text{–code with arb. encoding}\}.$$

Despite our lot of definitions the situation turns out to be quite simple:

Theorem I.7 *For all $\lambda \in (0, 1)$ the three λ–rates of the q–DMS (P, P) are equal to the* VON NEUMANN *entropy of the ensemble (P, P):*
$$R_{q,F_e}(\lambda) = R_{q,\bar{F}}(\lambda) = R_{a,\bar{F}}(\lambda) = H(P\mathsf{P}),$$
where $H(\rho) = -\mathrm{Tr}\,(\rho \log \rho)$ (see appendix A, section Entropy and divergence*).*

Proof. Between the first two members of the chain we have "$\geq$" by theorem I.1, between the second and the third "$\geq$" is obvious. $R_{q,F_e}(\lambda) \leq H(P\mathsf{P})$ follows from the coding theorem I.16. Finally $R_{a,\bar{F}}(\lambda) \geq H(P\mathsf{P})$ follows from the strong converse theorem I.19. □

Typical subspaces and shadows

Let P a p.d. on the set $\mathcal{X}$, with $|\mathcal{X}| = a < \infty$. Define for $\alpha > 0$ the set
$$\mathcal{T}^n_{V,P,\alpha} = \{x^n \in \mathcal{X}^n : \forall x \in \mathcal{X}\ |N(x|x^n) - nP(x)| \leq \alpha\sqrt{P(x)(1 - P(x))}\sqrt{n}\}$$
of *variance–typical sequences* with constant α (in the sense of WOLFOWITZ (1964)), where $N(x|x^n) = |\{i : x_i = x\}|$. For a sequence x^n the empirical distribution P_{x^n} on $\mathcal{X}$ (i.e. $P_{x^n}(x) = \frac{1}{n}N(x|x^n)$) is called *type* of x^n.

It is easily seen that there are at most $(n+1)^a$ types; this kind of reasoning is generally called *type counting*.

[3] Here and in the sequel log is always understood to base 2, as well as exp. The unit of this rate is usually called *qubit* (short for quantum bit: the states of a two–level quantum system $\mathfrak{L}(\mathbb{C}^2)$).

Lemma I.8 (cf. Wolfowitz (1964)) *For every p.d. P on $\mathcal{X}$ and $\alpha > 0$*

$$P^{\otimes n}(\mathcal{T}^n_{V,P,\alpha}) \geq 1 - \frac{a}{\alpha^2}$$

$$|\mathcal{T}^n_{V,P,\alpha}| \leq \exp\left(nH(P) + Kd\alpha\sqrt{n}\right).$$

Proof. $\mathcal{T}^n_{V,P,\alpha}$ is the intersection of a events, namely for each $x \in \mathcal{X}$ that the mean of the independent Bernoulli variables X_i with value 1 iff $x_i = x$ has a deviation from its expectation $P(x)$ at most $\alpha\sqrt{P(x)(1-P(x))}/\sqrt{n}$. By Chebyshev's inequality each of these has probability at least $1 - 1/\alpha^2$.

The cardinality estimate is like in the proof of the following lemma I.9. $\square$

Now construct *variance–typical projectors* $\Pi^n_{V,\rho,\alpha}$ using typical sequences: for a diagonalization $\rho = \sum_j q_j\pi_j$ let $s_j = \sqrt{q_j(1-q_j)}$ and

$$\mathcal{T}^n_{V,\rho,\alpha} = \{(j_1, \dots, j_n) : \forall j\ |N(j|j^n) - nq_j| \leq s_j\alpha\sqrt{n}\},$$

and define

$$\Pi^n_{V,\rho,\alpha} = \sum_{(j_1,\dots,j_n)\in\mathcal{T}^n_{V,\rho,\alpha}} \pi_{j_1} \otimes \cdots \otimes \pi_{j_n}\ .$$

For a state ρ define $\mu(\rho)$ as the minimal nonzero eigenvalue of $\sqrt{\rho(\mathbb{1}-\rho)}$ and $N(\rho) = \dim\operatorname{supp}\sqrt{\rho(\mathbb{1}-\rho)}$, finally $K = 2\frac{\log e}{e}$. Then one has

Lemma I.9 *For every state ρ and $n > 0$*

$$\operatorname{Tr}(\rho^{\otimes n}\Pi^n_{V,\rho,\alpha}) \geq 1 - \frac{d}{\alpha^2}$$
$$\operatorname{Tr}(\rho^{\otimes n}\Pi^n_{V,\rho,\alpha}) \geq 1 - 2N(\rho)e^{-2\mu(\rho)^2\alpha^2}\ ,$$

and with $\Pi^n = \Pi^n_{V,\rho,\alpha}$

$$\Pi^n \exp\left(-nH(\rho) - Kd\alpha\sqrt{n}\right) \leq \Pi^n\rho^{\otimes n}\Pi^n \leq \Pi^n \exp\left(-nH(\rho) + Kd\alpha\sqrt{n}\right)$$
$$\operatorname{Tr}\Pi^n_{V,\rho,\alpha} \leq \exp\left(nH(\rho) + Kd\alpha\sqrt{n}\right).$$

*Every η–*shadow *B of $\rho^{\otimes n}$ (this means $0 \leq B \leq \mathbb{1}$ and $\operatorname{Tr}(\rho^{\otimes n}B) \geq \eta$) satifies*

$$\operatorname{Tr} B \geq \left(\eta - 2N(\rho)e^{-2\mu(\rho)^2\alpha^2}\right)\exp\left(nH(\rho) - Kd\alpha\sqrt{n}\right).$$

Proof. The first estimate is the Chebyshev inequality, as before: the trace is the probability of a set of variance–typical sequences of eigenvectors of the ρ_i' in the product of the measures given by the eigenvalue lists. Similarly the second estimate is the well known inequality of Hoeffding (1963). The third estimate is the key: to prove it let $\pi^n = \pi_{j_1} \otimes \cdots \otimes \pi_{j_n}$ one of the eigenprojections of $\rho^{\otimes n}$ which contributes to $\Pi^n_{V,\rho,\alpha}$. Then

$$\operatorname{Tr}(\rho^{\otimes n}\pi^n) = q_{j_1}\cdots q_{j_n} = \prod_j q_j^{N(j|j^n)}\ .$$

Taking logs and using the defining relation for the $N(j|j^n)$ we find

$$\begin{aligned}\left|\sum_j -N(j|j^n)\log q_j - nH(\rho)\right| &\leq \sum_j -\log q_j |N(j|j^n) - nq_j| \\ &\leq \sum_j -\alpha\sqrt{n}\sqrt{q_j}\log q_j \\ &= 2\alpha\sqrt{n}\sum_j -\sqrt{q_j}\log\sqrt{q_j} \\ &\leq 2d\frac{\log e}{e}\alpha\sqrt{n}\ .\end{aligned}$$

The rest follows from the following lemma I.10. □

Lemma I.10 (Shadow bound) *Let* $0 \leq \Lambda \leq \mathbb{1}$ *and* ρ *a state such that for some* $\lambda, \mu_1, \mu_2 > 0$

$$\mathrm{Tr}\,(\rho\Lambda) \geq 1-\lambda \text{ and } \mu_1\Lambda \leq \sqrt{\Lambda}\rho\sqrt{\Lambda} \leq \mu_2\Lambda.$$

Then $(1-\lambda)\mu_2^{-1} \leq \mathrm{Tr}\,\Lambda \leq \mu_1^{-1}$ *and for* $0 \leq B \leq \mathbb{1}$ *with* $\mathrm{Tr}\,(\rho B) \geq \eta$ *one has* $\mathrm{Tr}\,B \geq \left(\eta - \sqrt{8\lambda}\right)\mu_2^{-1}$. *If* ρ *and* Λ *commute this can be improved to* $\mathrm{Tr}\,B \geq (\eta-\lambda)\,\mu_2^{-1}$.

Proof. The bounds on $\mathrm{Tr}\,\Lambda$ follow by taking traces in the inequalities in $\sqrt{\Lambda}\rho\sqrt{\Lambda}$ and using $1-\lambda \leq \mathrm{Tr}\,(\rho\Lambda) \leq 1$. For the η–shadow B observe

$$\begin{aligned}\mu_2\mathrm{Tr}\,B &\geq \mathrm{Tr}\,(\mu_2\Lambda B) \geq \mathrm{Tr}\,\left(\sqrt{\Lambda}\rho\sqrt{\Lambda}B\right) \\ &= \mathrm{Tr}\,(\rho B) - \mathrm{Tr}\,\left(\left(\rho - \sqrt{\Lambda}\rho\sqrt{\Lambda}\right)B\right) \geq \eta - \left\|\rho - \sqrt{\Lambda}\rho\sqrt{\Lambda}\right\|_1\ .\end{aligned}$$

If ρ and Λ commute the trace norm can obviously be estimated by λ, else we have to invoke the tender operator lemma I.4 to bound it by $\sqrt{8\lambda}$. □

For the benefit of discussions in later chapters let us mention here two other notions of typical projector:

Entropy typical projectors Let $\rho_1, \ldots, \rho_n$ 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

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

Define the *entropy–typical projector*[4] *of* ρ^n *with constant* δ as

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

Then we have the following

[4]This is essentially what SCHUMACHER (1995) calls *typical subspace*.

Lemma I.11 *There is a constant K depending only on d (in fact one may choose $K \leq \max\{(\log 3)^2, (\log d)^2\}$) such that for arbitrary states $\rho_1, \dots \rho_n$*

$$\operatorname{Tr}\left(\rho^n \Pi^n_{H,\rho^n,\delta}\right) \geq 1 - \frac{K}{\delta^2} \ .$$

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. □

Concerning its size we have

Lemma I.12 *For the entropy–typical projector*

$$\left(1 - \frac{K}{\delta^2}\right) \exp\left(\sum_{i=1}^n H(\rho_i) - \delta\sqrt{n}\right) \leq \operatorname{Tr} \Pi^n_{H,\rho^n,\delta} \leq \exp\left(\sum_{i=1}^n H(\rho_i) + \delta\sqrt{n}\right) .$$

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 $\Pi^n = \Pi^n_{H,\rho^n,\delta}$

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

Now the lemma follows by the shadow bound lemma I.10. □

Constant typical projectors Let ρ a state with diagonalization $\rho = \sum_j q_j \pi_j$, and $\delta > 0$, then define

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

and the *constant–typical projector*

$$\begin{aligned}\Pi^n_{C,\rho,\delta} &= \sum_{j^n \in \mathcal{T}^n_{C,\rho^n,\delta}} \pi_{j_1} \otimes \cdots \otimes \pi_{j_n} \\ &= \sum_{j^n \text{ with } \|\sum_{i=1}^n \pi_{j_i} - n\rho\|_\infty \leq \delta\sqrt{n}} \pi_{j_1} \otimes \cdots \otimes \pi_{j_n} \ .\end{aligned}$$

Then one has

Lemma I.13 (Weak law) *Let $\tilde{\rho}, \rho_1, \dots, \rho_n$ states of a system and $\delta, \epsilon > 0$ such that*

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

Then

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

Proof. Consider the diagonalization $\tilde{\rho} = \sum_j q_j \pi_j$, and the conditional expectation 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

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

Thus for $j^n = (j_1, \dots, j_n)$ with

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

we have by triangle inequality

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

So we can estimate

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

the last line by d uses of CHEBYSHEV's inequality, as in the proof of lemma I.9. □

Concerning the size of this projector we have

Lemma I.14 *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 type close to the p.d. given by the eigenvalue list 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 A.4) the statement follows. $\square$

The constant typical projectors will be used as shadows of whole *sets* (namely of states which satisfy the "average" condition of the weak law lemma I.13).

SCHUMACHER's quantum coding

Let $\alpha > 0$. The SCHUMACHER *scheme with constant* α for the q–DMS (P, P) is the following family of n–block codes $(\varepsilon_*, \delta_*)$ with quantum encoding: define $\Pi^n = \Pi^n_{V,P\mathsf{P},\alpha}$ and the Hilbert space $\mathcal{K} = \mathrm{im}\Pi^n$, and

$$\begin{aligned} \varepsilon_* : \mathfrak{L}(\mathcal{H})_*^{\otimes n} &\longrightarrow \mathfrak{L}(\mathcal{K})_* \\ \sigma &\longmapsto \Pi^n \sigma \Pi^n + \frac{1 - \mathrm{Tr}\,(\sigma\Pi^n)}{\dim \mathcal{K}} \mathbb{1} \\ \delta_* : \mathfrak{L}(\mathcal{K})_* &\longrightarrow \mathfrak{L}(\mathcal{H})_*^{\otimes n} \\ \sigma &\longmapsto \sigma\ . \end{aligned}$$

Remark I.15 *Essentially the above scheme was first defined by* SCHUMACHER *(1995), with a slightly different definition of* Π^n. *The great contribution of* SCHUMACHER *(1995) was to notice the possibility and importance of having a* typical subspace*, and the following theorem is just a variation of the original argument. Subsequently there appeared minor modifications and refinements (*JOZSA & SCHUMACHER *(1994) and* JOZSA et al. *(1998)), but all rely on one or another notion of typical subspace of* $\mathcal{H}^{\otimes n}$.

Theorem I.16 *The* SCHUMACHER *scheme has rate*

$$R(\varepsilon_*, \delta_*) \leq H(P\mathsf{P}) + \frac{Kd\alpha}{\sqrt{n}}$$

and entanglement fidelity

$$F_e(\varepsilon_*, \delta_*) \geq 1 - 4N(P\mathsf{P})e^{-2\mu(P\mathsf{P})^2\alpha^2}.$$

Proof. The rate estimate is immediate from lemma I.9. For the fidelity consider a purification of $P\mathsf{P} = \sum_j q_j |\varphi_j\rangle\langle\varphi_j|$, e.g. the projector of $|\psi\rangle = \sum_j \sqrt{q_j}|\varphi_j\rangle \otimes |\varphi_j\rangle$ on $\mathfrak{L}(\mathcal{H}^{\otimes 2})$. With that the fidelity estimate follows easily from the shadow lemma I.9. $\square$

By slightly changing the definition of the subspace used we arrive at the JHHH–*scheme* of JOZSA et al. (1998): just take for Π^n the projector

$$\Pi^n_{H(\cdot)\leq R} = \mathrm{l.\,c.\,supp}_{H(\nu)\leq R} \Pi^n_{V,\nu,0}$$

(the least common support) with some rate $R \geq 0$. Then in JOZSA et al. (1998) it is proved that this gives universally good compression of all sources (P, P) with $H(P\mathsf{P}) < R$. For one thing (see JOZSA et al. (1998))

$$\mathrm{Tr}\,\Pi^n_{H(\cdot)\leq R} \leq (n+1)^{d^2+d} \exp(nR),$$

and for the fidelity one has

Theorem I.17 *Let* $(\varepsilon_*, \delta_*)$ *the* JHHH*–scheme with rate* R *and block length* n*. Then for every q–DMS* (P, P)

$$F_e(\varepsilon_*, \delta_*) \geq 1 - 2(n+1)^d \exp\left(-n \cdot \min_{H(\nu)\geq R} D(\nu\|P\mathsf{P})\right).$$

Proof. First note that by direct calculation for ν codiagonal with a state ρ we have

$$\Pi^n_{V,\nu,0}\rho^{\otimes n}\Pi^n_{V,\nu,0} = \Pi^n_{V,\nu,0} \exp\left(-nD(\nu\|\rho) - nH(\nu)\right)$$

(see lemma II.12). Fix a diagonalization $P\mathsf{P} = \sum_j q_j\pi_j$ and observe

$$\Pi^n_{H(\cdot)\leq R} \geq \sum_{\nu\in\mathbb{C}[\pi_1,\dots,\pi_d],H(\nu)\leq R} \Pi^n_{V,\nu,0}\ .$$

Using the simple facts that $\Pi_{V,\nu,0} \neq 0$ only if $\nu \in \frac{1}{n}\mathbb{N}[\pi_j|j]$, and $\mathrm{Tr}\,\Pi^n_{V,\nu,0} \leq \exp(nH(\nu))$, we find as in the previous theorem

$$\begin{aligned}1 - \bar{F}(\varepsilon_*, \delta_*) &\leq 2 \sum_{\nu\in\frac{1}{n}\mathbb{N}[\pi_j|j],H(\nu)>R} \exp(-nD(\nu\|P\mathsf{P}))\\ &\leq 2(n+1)^d \exp\left(-n \cdot \min_{H(\nu)\geq R} D(\nu\|P\mathsf{P})\right),\end{aligned}$$

where the last estimate is by type counting: there are at most $(n+1)^d$ different ν diagonal in the basis $\{\pi_j|j\}$ and $\Pi^n_{V,\nu,0} \neq 0$. □

Strong converse

The first proofs by SCHUMACHER (1995) and JOZSA & SCHUMACHER (1994) for the optimality of the SCHUMACHER scheme where valid only under the additional assumption that δ_* is of the form $\delta_*(\sigma) = U\sigma U^*$ for a unitary embedding U of $\mathcal{K}$ into $\mathcal{H}^{\otimes n}$. Also they achieved the bound $H(P\mathsf{P})$ only in the limit of $\lambda \to 0$ (so they proved a *weak converse*). The proof of BARNUM et al. (1996) removed the restriction on δ_*, but still yields only a weak converse. Also it works with some surprising and difficult fidelity estimates, involving even mixed state fidelity, see JOZSA (1994) (We may note that they seem to be related to our inequalities of theorem I.1). We should also mention the work of ALLAHVERDYAN & SAAKIAN (1997a) where a weak converse was proved for quantum encodings and using entanglement fidelity (compare our theorem IV.7 with $s=1$). The criticism of the authors on BARNUM et al. (1996) however is unjustified: by the above discussion (proof of theorem I.7) their result is weaker than that of BARNUM et al. (1996). Then HORODECKI (1998) noticed that considering $\bar{D}$ instead of $\bar{F}$ drastically simplifies the proof. His argument is as follows:

Assume that we are given a code $(\varepsilon_*, \delta_*)$ with arbitrary encoding in the states on a k–dimensional Hilbert space and

$$\bar{D} = \frac{1}{2}\sum_{\pi^n\in\mathsf{P}^n} P^n(\pi^n)\|\pi^n - \delta_*\varepsilon_*\pi^n\|_1 \leq \lambda \leq \frac{1}{4}\ .$$

So by MARKOV's inequality there is a subset $\mathcal{C}\subset\mathsf{P}^n$ with $P^n(\mathcal{C})\geq 1-2\sqrt{\lambda}$ and

$$\forall\pi^n\in\mathcal{C}\quad \|\pi^n - \delta_*\varepsilon_*\pi^n\|_1 \leq \sqrt{\lambda}\ .$$

Now form the state $\sigma = \sum_{\pi^n\in\mathsf{P}^n} P^n(\pi^n)\varepsilon_*\pi^n$, then by UHLMANN's monotonicity of the quantum I–divergence (theorem A.5)

$$\forall\pi^n\in\mathsf{P}^n\quad D(\varepsilon_*\pi^n\|\sigma)\geq D(\delta_*\varepsilon_*\pi^n\|\delta_*\sigma).$$

Averaging we obtain

$$\sum_{\pi^n\in\mathsf{P}^n} P^n(\pi^n)D(\varepsilon_*\pi^n\|\sigma)\geq \sum_{\pi^n\in\mathsf{P}^n} P^n(\pi^n)D(\delta_*\varepsilon_*\pi^n\|\delta_*\sigma).$$

Now it is straightforward to calculate the l.h.s. of this to $H(\sigma)-\sum_{\pi^n\in\mathsf{P}^n} P^n(\pi^n)H(\varepsilon_*\pi^n)$, whereas the r.h.s. evaluates similarly to $H(\delta_*\sigma)-\sum_{\pi^n\in\mathsf{P}^n} P^n(\pi^n)H(\delta_*\varepsilon_*\pi^n)$. Since $2\lambda\leq 1/2$ and $\sqrt{\lambda}\leq 1/2$ we can use a continuity property of H (see lemma A.4):

$\|\delta_*\sigma - (P\mathsf{P})^{\otimes n}\|_1\leq 2\lambda$ implies

$$|H(\delta_*\sigma) - nH(P\mathsf{P})|\leq -2\lambda\log\frac{2\lambda}{d^n}\ ,$$

and (for $\pi^n\in\mathcal{C}$) $\|\delta_*\varepsilon_*\pi^n - \pi^n\|_1\leq\sqrt{\lambda}$ implies

$$|H(\delta_*\varepsilon_*\pi^n) - H(\pi^n)|\leq -\sqrt{\lambda}\log\frac{\sqrt{\lambda}}{d^n}\ .$$

Combining we get the chain of inequalities

$$\begin{aligned}
\log k &\geq H(\sigma)\\
&\geq H(\sigma) - \sum_{\pi^n\in\mathsf{P}^n} P^n(\pi^n)H(\varepsilon_*\pi^n)\\
&\geq H(\delta_*\sigma) - \sum_{\pi^n\in\mathsf{P}^n} P^n(\pi^n)H(\delta_*\varepsilon_*\pi^n)\\
&\geq nH(P\mathsf{P}) - \sum_{\pi^n\in\mathsf{P}^n} P^n(\pi^n)H(\pi^n) - 2\sqrt{\lambda}\log d^n + 2\lambda\log\frac{2\lambda}{d^n} + \sqrt{\lambda}\log\frac{\sqrt{\lambda}}{d^n}\\
&= nH(P\mathsf{P}) - n(2\lambda+3\sqrt{\lambda})\log d + 2\lambda\log 2\lambda + \sqrt{\lambda}\log\sqrt{\lambda}\ .
\end{aligned}$$

Thus we proved

Theorem I.18 (Weak converse) *For every q–DMS* (P, P)

$$\liminf_{\lambda\to 0} R_{a,\bar F}(\lambda) = \liminf_{\lambda\to 0} R_{a,\bar D}(\lambda) \geq H(P\mathsf{P}).$$

□

But in fact much more is true:

Theorem I.19 (Strong converse) *Let* (P, P) *a q–DMS and* $(\varepsilon_*, \delta_*)$ *an* $(n, \lambda)_{\bar F}$*–code with arbitrary encoding, and* $\alpha > 0$*. Then*

$$\dim \mathcal{K} \geq \left(1 - \lambda - 4\sqrt{N(P\mathsf{P})}e^{-\mu(P\mathsf{P})^2\alpha^2}\right) \cdot \exp\left(nH(P\mathsf{P}) - Kd\alpha\sqrt{n}\right).$$

Proof. Let $B = \delta_*(\mathbb{1}_\mathcal{K})$ and $\Pi^n = \Pi^n_{V,P\mathsf{P},\alpha}$. Since $\varepsilon_*\pi^n \leq \mathbb{1}_\mathcal{K}$ for every $\pi^n \in \mathsf{P}^n$ it is clear that $\delta_*\varepsilon_*\pi^n \leq B$. Thus

$$\begin{aligned}
\mathrm{Tr}\,(B \cdot \Pi^n\pi^n\Pi^n) &\geq \mathrm{Tr}\,((\delta_*\varepsilon_*\pi^n)\Pi^n\pi^n\Pi^n) \\
&= \mathrm{Tr}\,((\delta_*\varepsilon_*\pi^n)\pi^n) - \mathrm{Tr}\,(\delta_*\varepsilon_*\pi^n(\pi^n - \Pi^n\pi^n\Pi^n)) \\
&\geq \mathrm{Tr}\,((\delta_*\varepsilon_*\pi^n)\pi^n) - \|\pi^n - \Pi^n\pi^n\Pi^n\|_1 \\
&\geq \mathrm{Tr}\,((\delta_*\varepsilon_*\pi^n)\pi^n) - \sqrt{8(1 - \mathrm{Tr}\,\pi^n\Pi^n)}
\end{aligned}$$

(the last estimate by lemma I.4). Averaging over $P^{\otimes n}$ we find, with the shadow lemma I.9 and concavity of the square root:

$$\begin{aligned}
\mathrm{Tr}\,\left(\Pi^n(P\mathsf{P})^{\otimes n}\Pi^n B\right) &\geq \bar F - \sqrt{8\,(1 - \mathrm{Tr}\,(P\mathsf{P})^{\otimes n}\Pi^n)} \\
&\geq 1 - \lambda - 4\sqrt{N(P\mathsf{P})}e^{-\mu(P\mathsf{P})^2\alpha^2}\ .
\end{aligned}$$

Since by lemma I.9

$$\Pi^n(P\mathsf{P})^{\otimes n}\Pi^n \leq \Pi^n \exp\left(-nH(P\mathsf{P}) + Kd\alpha\sqrt{n}\right)$$

we conclude

$$\mathrm{Tr}\,B \geq \mathrm{Tr}\,(B\Pi^n) \geq \left(1 - \lambda - 4\sqrt{N(P\mathsf{P})}e^{-\mu(P\mathsf{P})^2\alpha^2}\right) \cdot \exp\left(nH(P\mathsf{P}) - Kd\alpha\sqrt{n}\right),$$

and with $\dim\mathcal{K} = \mathrm{Tr}\,\mathbb{1}_\mathcal{K} = \mathrm{Tr}\,B$ the proof is complete. □

Corollary I.20 *Let* $E_n = o(n)$ *and* $\lambda_n \leq 1 - e^{-E_n}$*. Then for every sequence* $(\varepsilon_{n*}, \delta_{n*})$ *of* $(n, \lambda_n)_{\bar F}$*–codes with arbitrary encoding for the q–DMS* (P, P)

$$\liminf_{n\to\infty} R(\varepsilon_{n*}, \delta_{n*}) \geq H(P\mathsf{P}).$$

□

Remark I.21 *The proof of the above theorem is remarkable in that it employs a positive operator which is not necessarily bounded by* $\mathbb{1}$ *(this is why we could not directly apply the shadow). Even though it has consequently no interpretation as a physical measurement (maybe it has one as a* quantity*), it can be analyzed to give information about the coding scheme.*

Relation to classical source coding

Consider a slight variation of our initial model: P is now a set of pure states on a finite dimensional C*–algebra $\mathfrak{A}$ (which is a direct sum of full matrix algebras $\mathfrak{L}(\mathcal{H})$), and consider only $\bar{F}$ as a fidelity measure. A major (and extremal) example is a classical source, i.e. $\mathfrak{A} = \mathbb{C}\mathcal{X}$ is commutative, with a finite set $\mathcal{X}$, and w.l.o.g. $\mathsf{P} = \mathcal{X}$ (all possible pure states). The general case may be seen as an interpolation between this and the quantum case $\mathfrak{A} = \mathfrak{L}(\mathcal{H})$.

Observe that since $P\mathsf{P} \in \mathfrak{A}_*$ we find the typical projectors Π^n in $\mathfrak{A}^{\otimes n}$ (note that for $\mathfrak{A} = \mathbb{C}\mathcal{X}$ such a projector is given just by a set of typical sequences from $\mathcal{X}^n$). This means that the SCHUMACHER and JHHH–schemes make sense by just replacing $\mathfrak{L}(\mathcal{H})$ in the definitions by $\mathfrak{A}$, without changing the fidelity values (note again that for $\mathfrak{A} = \mathbb{C}\mathcal{X}$ the average fidelity is just the classical success probability). The strong converse need not be modified at all as $\mathfrak{L}(\mathcal{H})$ is already the most "spacious" algebra imaginable. Thus we arrive (with obvious definitions) at

Theorem I.22 *For all $\lambda \in (0,1)$ the arbitary and quantum encoding rates of the discrete memoryless source (P, P) on the C*–algebra $\mathfrak{A}$ are equal to the* VON NEUMANN *entropy of the ensemble (P, P):*

$$R_{q,\bar{F}}(\lambda) = R_{a,\bar{F}}(\lambda) = H(P\mathsf{P}).$$

□

Open questions

Dimension Why stay with finite dimensional spaces? In fact there is no obstruction to defining sensibly a SCHUMACHER scheme, indeed the original paper of SCHUMACHER (1995) had no dimension restriction, instead (implicitly) requiring bounded variance of the information density, i.e. in the present setting the condition $\mathrm{Tr}\,(\rho(\log\rho)^2) < \infty$. Then the typical projector of choice is the entropy typical one, and in fact the reader may as an exercise translate the coding theorem and our strong converse to this situation.

Memory It appears that no one has formalized the concept of coding a "quantum MARKOV chain".

Lossless coding It might be worthwhile to try and to convert the techniques of HUFFMAN coding, and especially of arithmetic coding of the source to quantum sources. See BRAUNSTEIN et al. (1998) for a discussion.

Rate distortion theory Develop further a rate distortion theory: the start to this was made by BENDJABALLAH et al. (1998), and a short note of BARNUM (1998).

Refined resource analysis A not yet investigated (and perhaps most interesting) problem is, how much "quantum" one actually needs to compress the source (P, P): whereas $\dim \mathcal{K}$ is shown by theorem I.22 to be a good resource measure, it is oblivious to the difference between an orthogonal ensemble (for whose coding a commutative algebra, i.e. a classical system, suffices), and a highly non–orthogonal one (which presumably needs all the quantum resources, i.e. possibilities of superpositions, of k degrees of freedom). As a measure of this "quantum" resource I propose the following:

A coding scheme is a pair $(\varepsilon_*, \delta_*)$ with

$$\begin{aligned} \varepsilon_* :& \mathsf{P}^n \longrightarrow \mathfrak{K}_* \quad \text{a mapping,} \\ \delta_* :& \mathfrak{K}_* \longrightarrow \mathfrak{A}_*^{\otimes n} \quad \text{a quantum operation,} \end{aligned}$$

where $\mathfrak{K}$ is a finite dimensional C*–algebra. Quantum and arbitrary encoding schemes are as before. Observe that $\operatorname{Tr} \mathbb{1}_{\mathfrak{K}}$ takes now the place of the previous $\dim \mathcal{K}$. Define the, say, *rate of superposition* as

$$r(\varepsilon_*, \delta_*) = \frac{1}{n} \left(\log \dim_{\mathbb{C}} \mathfrak{K} - \log \operatorname{Tr} \mathbb{1}_{\mathfrak{K}} \right).$$

Observe that $0 \leq r(\varepsilon_*, \delta_*) \leq \frac{1}{n} \log \operatorname{Tr} \mathbb{1}_{\mathfrak{K}}$, with $r(\varepsilon_*, \delta_*) = 0$ iff $\mathfrak{K}$ is commutative.

Now define for $\lambda \in (0,1)$, $R \geq 0$ the λ–rates of superposition with arbitrary and quantum encoding:

$$\begin{aligned} r_{a,\bar{F}}(\lambda, R) &= \limsup_{n\to\infty} \min\{r(\varepsilon_*, \delta_*) : \ (\varepsilon_*, \delta_*) \text{ an } (n,\lambda)_{\bar{F}}\text{–code (arb. enc.)}, R(\varepsilon_*, \delta_*) \leq R\}, \\ r_{q,\bar{F}}(\lambda, R) &= \limsup_{n\to\infty} \min\{r(\varepsilon_*, \delta_*) : \ (\varepsilon_*, \delta_*) \text{ an } (n,\lambda)_{\bar{F}}\text{–code (qu. enc.)}, R(\varepsilon_*, \delta_*) \leq R\}. \end{aligned}$$

It is obvious that $r_{a,\bar{F}}$ and $r_{q,\bar{F}}$ are nonincreasing functions of R, and that both are upper bounded by $H(P\mathsf{P})$. The problem is now to analyze $r_{a,\bar{F}}$ and $r_{q,\bar{F}}$ depending on λ and R.

- It is clear that $r_{a,\bar{F}}(\lambda, R) = 0$ if R is large enough ($R = H(P)$ suffices). It would be interesting to determine the exact threshold, the value at $R = H(P\mathsf{P})$ and the behavior between these points. In any case, I conjecture that $r_{a,\bar{F}}(\lambda, R)$ does not depend on $\lambda \in (0,1)$.

- I conjecture further that $r_{q,\bar{F}}$ depends neither on $\lambda \in (0,1)$ nor on $R > H(P\mathsf{P})$. If this is true $r_{q,\bar{F}}$ is an interesting ensemble property of (P, P).

Chapter II

Quantum Channel Coding

In this chapter we introduce the notion of a quantum channel. From the beginning we focus on the *product state capacity for transmission of classical information*, and prove coding theorem and strong converse, even for nonstationary channels. In the finite stationary case we can sharpen our rate estimates and derive some bounds for the reliability function. As a corollary to our strong converse we obtain another proof of the HOLEVO bound.

Quantum channels and codes

The following definition is after HOLEVO (1977): 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 $\mathfrak{L}(\mathcal{H})$, where $d = \dim \mathcal{H}$ is assumed to be finite.

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.

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 $\mathfrak{S}(\mathfrak{A}_1) \times \cdots \times \mathfrak{S}(\mathfrak{A}_n)$, 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

$$e(f, D) = \max\{1 - \mathrm{Tr}\,(\varphi_*^{\otimes n}(f(m))D_m) : m \in \mathcal{M}\}.$$

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 II.1 *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 II.2 is still true in the more general model: maybe higher capacities can be achieved there, see the discussion of* SCHUMACHER & WESTMORELAND *(1997).*

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. The main result of the present chapter (to be proved in the following sections) is

Theorem II.2 *Let $(\mathfrak{W}_1, \mathfrak{W}_2, \dots)$ a nonstationary q–DMC, and*

$$C(\mathfrak{W}_\mathrm{i}) = \sup_{P\ p.d.\ on\ \mathfrak{W}_i} I(P; \mathfrak{W}_i)$$

(with $I(P; \mathfrak{W}) = H(P\mathfrak{W}) - H(\mathfrak{W}|P)$, see remark A.15). Then for every $\lambda \in (0, 1)$

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

Proof. Combine the coding theorem II.4 and the strong converse theorem II.7. □

This theorem justifies the name *capacity* (of the channel $\mathfrak{W}$) for the quantity $C(\mathfrak{W})$, even in the strong sense of Wolfowitz (1964). Observe that this theorem is a quantum generalization of a theorem by Ahlswede (1968).

Remark II.3 *It should be clear that the same (including proofs) applies if the output system $\mathfrak{L}(\mathcal{H})$ is replaced by a $*$–subalgebra $\mathfrak{A}$.*

Maximal code construction

Theorem II.4 (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*

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

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

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

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 I.4 to obtain

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

Thus

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

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 I.12 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

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

the last inequality again by lemma I.12, and we are done. $\square$

Remark II.5 *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.*

Remark II.6 *The above coding theorem — for stationary channels and with slightly weaker bounds — was first proved by* HOLEVO *(1998a) (and independently by* SCHUMACHER & WESTMORELAND *(1997)), building on ideas of* HAUSLADEN et al. *(1997) for the* pure state channel*.*

Strong converse

Theorem II.7 (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 $(\mathfrak{W}_i)_i$*

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

Before proving this we need to follow a short technical digression:

Approximation of channels We have continuum many states on $\mathfrak{L}(\mathcal{H})$ to deal with, 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 II.8 *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 the digression 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.

Proof of theorem II.7. 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 I.13 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 lemmata I.14 and A.4 we have for good γ

$$\mathrm{Tr}\,\Pi_\gamma \leq \exp\left(|I_\gamma|(H(\tilde{\sigma}_\gamma)+\epsilon)\right) \leq \exp\left(\sum_{i\in I_\gamma} H(\tilde{\rho}_{i\gamma}) + 2|I_\gamma|\epsilon\right).$$

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 I.4 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 I.12 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 arrive at

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

and thus

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

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 II.8), $\delta = n^{1/3}$, and let n large enough. $\square$

Remark II.9 *The* weak converse *is already a consequence of the information bound of* Holevo *(1973), see theorem A.16, together with subadditivity of quantum mutual information (corollary A.18) and the classical* Fano *inequality (see theorem A.24).*

Refined analysis for stationary channels

From this point on we restrict ourselves to the finite and stationary case.

Let $W : \mathcal{X} \to \mathfrak{S}(\mathfrak{L}(\mathcal{H}))$ a finite q–DMC, mapping $x \in \mathcal{X}$ to the state W_x, with a set $\mathcal{X}$, say of cardinality $|\mathcal{X}| = a < \infty$, for a fixed complex Hilbert space $\mathcal{H}$ of finite dimension d (i.e., in slight variation to the previous sections, we label the set of channel states by $\mathcal{X}$). We will have occasion to consider other channels, say V, implicitely all with the same $\mathcal{X}$. Note that we drop here the subscript $*$ for state maps, to be closer to the notation in use in the literature.

For an n–block code (f, D) for W we will here interpret f as a mapping from the finite set $\mathcal{M}$ into $\mathcal{X}^n$. The (maximum) error probability of the code then reads as

$$e(f, D) = \max\{1 - \mathrm{Tr}\,(W_{f(m)}D_m) : m \in \mathcal{M}\}.$$

(For $f(m) = x^n \in \mathcal{X}^n$ we adopt the convention $W_{f(m)} = W_{x^n} = W_{x_1} \otimes \cdots \otimes W_{x_n}$). The *rate* of an n–block code is defined as $\frac{1}{n}\log|\mathcal{M}|$. Recall that $N(n, \lambda)$ is the maximal size (i.e. $|\mathcal{M}|$) of an (n, λ)–code, and define

$$e_{\min}(n, R) = \min\{e(f, D) : (f, D) \text{ is } n\text{–block code, } |\mathcal{M}| \geq \exp(nR)\}.$$

Finally for states ρ and ν, and another channel V and p.d. P on $\mathcal{X}$ let

$$\begin{aligned}
D(\nu\|\rho) &= \mathrm{Tr}\,\big(\nu(\log\nu - \log\rho)\big) \\
D(V\|W|P) &= \sum_{x\in\mathcal{X}} P(x)D(V_x\|W_x),
\end{aligned}$$

the (conditional) quantum I–divergence, see appendix A, section *Entropy and divergence.*

The rewards of our restriction are stronger estimates on $N(n,\lambda)$, and — more interestingly — upper and lower bounds on $e_{\min}(n,R)$, which lead to nontrivial lower and upper bounds on the reliability function of the channel. This extends results of BURNASHEV & HOLEVO (1997) from pure state to general channels, and thus gives (partial) answers to two problems posed by HOLEVO (1998b).

Some more typicalities We begin with an extension of lemma I.9: define the *conditional variance–typical projectors* $\Pi^n_{V,W,\delta}(x^n)$ with $x^n \in \mathcal{X}^n$ to be

$$\Pi^n_{V,W,\delta}(x^n) = \bigotimes_{x\in\mathcal{X}} \Pi^{I_x}_{V,W_x,\delta} ,$$

where $I_x = \{i \in [n] : x_i = x\}$.

Lemma II.10 *For every $x^n \in \mathcal{X}^n$ of type P, and with $\Pi^n = \Pi^n_{V,W,\delta}(x^n)$*

$$\operatorname{Tr} W_{x^n}\Pi^n \geq 1 - \frac{ad}{\delta^2}$$

$$\Pi^n \exp\left(-nH(W|P) - Kd\sqrt{a}\delta\sqrt{n}\right) \leq \Pi^n W_{x^n} \Pi^n \leq \Pi^n \exp\left(-nH(W|P) + Kd\sqrt{a}\delta\sqrt{n}\right)$$

$$\begin{aligned}\operatorname{Tr}\Pi^n_{V,W,\delta}(x^n) &\leq \exp\left(nH(W|P) + Kd\sqrt{a}\delta\sqrt{n}\right)\\ \operatorname{Tr}\Pi^n_{V,W,\delta}(x^n) &\geq \left(1 - \frac{ad}{\delta^2}\right)\exp\left(nH(W|P) - Kd\sqrt{a}\delta\sqrt{n}\right).\end{aligned}$$

Every η–shadow B of W_{x^n} satifies

$$\operatorname{Tr} B \geq \left(\eta - \frac{ad}{\delta^2}\right)\exp\left(nH(W|P) - Kd\sqrt{a}\delta\sqrt{n}\right).$$

Proof. The first inequality is just a times the estimate from lemma I.9. The estimate for $\Pi^n_{V,W,\delta}(x^n) W_{x^n} \Pi^n_{V,W,\delta}(x^n)$ follows from piecing together the estimates for the $\Pi^{I_x}_{V,W_x,\delta}$ in the same lemma (using $\sum_{x\in\mathcal{X}}\sqrt{P(x)} \leq \sqrt{a}$). The rest follows from the shadow bound lemma I.10. □

From this we get the following

Lemma II.11 *Let $\delta > 0$ and $x^n \in \mathcal{X}^n$ of type P. Then*

$$\operatorname{Tr}\left(W_{x^n}\Pi^n_{V,PW,\delta\sqrt{a}}\right) \geq 1 - \frac{ad}{\delta^2} .$$

Proof. Diagonalize $PW = \sum_j q_j \pi_j$, and let $\kappa_* : \mathfrak{L}(\mathcal{H})_* \to \mathfrak{L}(\mathcal{H})_*$ the conditional expectation be defined by $\kappa_*(\sigma) = \sum_j \pi_j \sigma \pi_j$. We claim that

$$\Pi^n_{V,PW,\delta\sqrt{a}} \geq \Pi^n_{V,\kappa_* W,\delta}(x^n).$$

Indeed let $\pi_{j_1} \otimes \cdots \otimes \pi_{j_n}$ one of the product states constituting $\bigotimes_{x\in\mathcal{X}} \Pi^{I_x}_{V,\kappa_*(W_x),\delta}$, i.e. with $\kappa_*(W_x) = \sum_j q_{j|x}\pi_j$

$$\forall x \in \mathcal{X}\ \forall j \ \left|N(j|j^{I_x}) - q_{j|x}|I_x|\right| \leq \delta\sqrt{|I_x|}\sqrt{q_{j|x}(1-q_{j|x})}.$$

Hence (with $|I_x| = P(x)n$)

$$\begin{aligned}
|N(j|j^n) - q_j n| &\leq \sum_{x\in\mathcal{X}} \left|N(j|j^{I_x}) - q_{j|x}|I_x|\right| \\
&\leq \sum_{x\in\mathcal{X}} \delta\sqrt{n}\sqrt{P(x)}\sqrt{q_{j|x}(1-q_{j|x})} \\
&\leq \delta\sqrt{a}\sqrt{n}\sqrt{\sum_{x\in\mathcal{X}} P(x) q_{j|x}(1-q_{j|x})} \\
&\leq \delta\sqrt{a}\sqrt{n}\sqrt{q_j(1-q_j)},
\end{aligned}$$

the last inequality by concavity of the map $x \mapsto x(1-x)$, and $q_j = \sum_{x\in\mathcal{X}} P(x) q_{j|x}$.

Thus we can estimate

$$\begin{aligned}
\mathrm{Tr}\,(W_{x^n}\Pi^n_{V,PW,\delta\sqrt{a}}) &= \mathrm{Tr}\,\left((\kappa_*^{\otimes n} W_{x^n})\Pi^n_{V,PW,\delta\sqrt{a}}\right) \\
&\geq \mathrm{Tr}\,\left((\kappa_*^{\otimes n} W_{x^n})\Pi^n_{V,\kappa_* W,\delta}(x^n)\right) \\
&\geq 1 - \frac{ad}{\delta^2}\ ,
\end{aligned}$$

the last line by lemma II.10. □

Of particular interest are the variance–typical projectors with $\delta = 0$, i.e. the $\Pi^n_\rho = \Pi^n_{V,\rho,0}$ and $\Pi^n_W(x^n) = \Pi^n_{V,W,0}(x^n)$, which we call *exact types*.

For the following fix diagonalizations $\rho = \sum_j q_j \pi_j$ and $W_x = \sum_j q_{j|x}\pi_{xj}$. The commutative algebras $\mathbb{C}[\pi_j|j]$ and $\mathbb{C}[\pi_{xj}|j]$ (which are maximal commutative subalgebras of the *commutants* $\mathbb{C}[\rho]'$ and $\mathbb{C}[W_x]'$) will be important below.

Lemma II.12 *For $\nu \in \mathbb{C}[\rho]'$ we have*

$$\Pi^n_\nu \rho^{\otimes n}\Pi^n_\nu = \Pi^n_\nu \exp\left(-nD(\nu\|\rho) - nH(\nu)\right)$$

$$(n+1)^{-d}\exp(nH(\nu)) \leq \mathrm{Tr}\,\Pi^n_\nu \leq \exp(nH(\nu)).$$

For $V_x \in \mathbb{C}[W_x]'$ and $x^n \in \mathcal{X}^n$ of type P

$$\Pi^n_V(x^n) W_{x^n} \Pi^n_V(x^n) = \Pi^n_V(x^n)\exp\left(-nD(V\|W|P) - nH(V|P)\right)$$

$$(n+1)^{-ad}\exp(nH(V|P)) \leq \mathrm{Tr}\,\Pi^n_V(x^n) \leq \exp(nH(V|P)).$$

Proof. The first equation is straightforward. To estimate $\operatorname{Tr}\Pi^n_\nu$ let $\rho=\nu$ and note that

$$(n+1)^{-d} \leq \operatorname{Tr}\left(\nu^{\otimes n}\Pi^n_\nu\right) \leq 1.$$

There the upper bound is trivial, whereas the lower bound is by type counting, i.e. observing that in the decomposition $\mathbb{1}=\sum_{\hat\nu\in\mathbb{C}[\pi_j|j]}\Pi^n_{\hat\nu}$ there appear at most $(n+1)^d$ nonzero terms, and the fact that for such $\hat\nu$ the quantity $\operatorname{Tr}\left(\nu^{\otimes n}\Pi^n_{\hat\nu}\right)$ is maximized with $\hat\nu=\nu$ (Compare Csiszár & Körner (1981), lemma 1.2.3). The second part of the lemma follows from the first by collecting positions of equal letters in x^n. □

Corollary II.13 *If $\nu\in\mathbb{C}[\rho]'$ and $\Pi^n_\nu\neq 0$ then*

$$(n+1)^{-d}\exp(-nD(\nu\|\rho)) \leq \operatorname{Tr}\left(\rho^{\otimes n}\Pi^n_\nu\right) \leq \exp(-nD(\nu\|\rho)).$$

□

Define for a state ρ, channel W, $x^n\in\mathcal{X}^n$ of type P, and a real number L:

$$\begin{aligned}
\Pi^n_{\rho,H(\cdot)\leq L} &= \sum_{\nu\in\mathbb{C}[\pi_j|j],H(\nu)\leq L}\Pi^n_\nu\\
\Pi^n_{\rho,H(\cdot)\geq L} &= \sum_{\nu\in\mathbb{C}[\pi_j|j],H(\nu)\geq L}\Pi^n_\nu\\
\Pi^n_{W,H(\cdot|P)\leq L}(x^n) &= \sum_{V_x\in\mathbb{C}[\pi_{xj}|j],H(V|P)\leq L}\Pi^n_V(x^n)\\
\Pi^n_{W,H(\cdot|P)\geq L}(x^n) &= \sum_{V_x\in\mathbb{C}[\pi_{xj}|j],H(V|P)\geq L}\Pi^n_V(x^n).
\end{aligned}$$

Lemma II.14 *For ρ, W, $x^n\in\mathcal{X}^n$ of type P, and L as above*

$$\operatorname{Tr}\left(\Pi^n_{W,H(\cdot|P)\leq L}(x^n)\right) \leq (n+1)^{ad}\exp(nL)$$

$$\operatorname{Tr}\left(W_{x^n}\Pi^n_{W,H(\cdot|P)\leq L}(x^n)\right) \geq 1-(n+1)^{ad}\exp\left(-n\cdot\inf_{H(V|P)>L}D(V\|W|P)\right)$$

$$\operatorname{Tr}\left(\rho^{\otimes n}\Pi^n_{\rho,H(\cdot)\geq L}\right) \geq 1-(n+1)^{d}\exp\left(-n\cdot\min_{H(\nu)\leq L}D(\nu\|\rho)\right).$$

Proof. The inequalities all follow from lemma II.12 and corollary II.13 together with type counting. □

Lemma II.15 *For ρ and L as above*

$$\begin{aligned}
\Pi^n_{\rho,H(\cdot)\geq L}\rho^{\otimes n}\Pi^n_{\rho,H(\cdot)\geq L} &\leq \Pi^n_{\rho,H(\cdot)\geq L}\exp\left(-n\cdot\min_{H(\nu)\geq L}(H(\nu)+D(\nu\|\rho))\right)\\
&= \Pi^n_{\rho,H(\cdot)\geq L}\exp\left(-nL-n\cdot\min_{H(\nu)=L}D(\nu\|\rho)\right)\\
&\leq \Pi^n_{\rho,H(\cdot)\geq L}\exp\left(-nL-n\cdot\min_{H(\nu)\leq L}D(\nu\|\rho)\right).
\end{aligned}$$

For an η–shadow B of $\rho^{\otimes n}$

$$\operatorname{Tr} B \geq \left(\eta - (n+1)^d \exp(-n \cdot \min_{H(\nu)\leq L} D(\nu\|\rho))\right) \cdot \exp\left(nL + n \cdot \min_{H(\nu)\leq L} D(\nu\|\rho)\right).$$

Proof. The first estimate is directly from lemma II.12. To see that the required min is assumed at the boundary of the (convex) region where $H(\nu) \geq L$ observe that the minimized quantity is linear in ν.

For the η–shadow B: note that by lemma II.14 with $\Pi^n = \Pi^n_{\rho,H(\cdot)\geq L}$

$$\operatorname{Tr}\left(\rho^{\otimes n}\Pi^n B \Pi^n\right) \geq \eta - (n+1)^d \exp\left(-n \cdot \min_{H(\nu)\leq L} D(\nu\|\rho)\right)$$

and the rest follows by the estimate on $\Pi^n \rho^{\otimes n} \Pi^n$. □

Code bounds up to $\mathbf{O}(\sqrt{\mathbf{n}})$ terms Our first result is a variation of theorem II.4:

Theorem II.16 *For every $\lambda \in (0,1)$ there is a constant $K(a,d,\lambda)$ such that for every q–DMC W*

$$N(n,\lambda) \geq \exp\left(nC(W) - K(a,d,\lambda)\sqrt{n}\right).$$

Proof. Let P a p.d. on $\mathcal{X}$ with $C(W) = H(PW) - H(W|P)$. Let (f,D) a maximal (n,λ)–code with the property

$$\forall m \in \mathcal{M} \qquad f(m) \in \mathcal{T}^n_{V,P,\sqrt{2a}},\ \operatorname{Tr} D_m \leq \operatorname{Tr} \Pi^n_{V,W,\delta}(f(m)),$$

with $\delta = \sqrt{\frac{2ad}{\lambda}}$. In particular (by lemma II.10)

$$\operatorname{Tr} D_m \leq \exp\left(nH(W|P) + (Kd\sqrt{a}\delta + Ka\sqrt{2a}\log d)\sqrt{n}\right).$$

Let $B = \sum_{m\in\mathcal{M}} D_m$, we claim that for all $x^n \in \mathcal{T}^n_{V,P,\sqrt{2a}}$

$$\operatorname{Tr}(W_{x^n} B) \geq \eta = \min\{1-\lambda, \lambda^2/32\}.$$

This is clear if x^n is a code word, and true else, for otherwise we could extend our code with the word x^n and decoding operator

$$D' = \sqrt{\mathbb{1} - B}\,\Pi^n_{V,W,\delta}(x^n)\sqrt{\mathbb{1} - B}\ .$$

This is exactly as in the proof of theorem II.4. Thus we arrive at

$$\operatorname{Tr}\left((PW)^{\otimes n} B\right) \geq \eta/2$$

which by lemma I.9 implies the estimate

$$\operatorname{Tr} B \geq \left(\frac{\eta}{2} - \frac{d}{\delta_0^2}\right) \exp\left(nH(PW) - Kd\delta_0\sqrt{n}\right).$$

Choosing $\delta_0 = \sqrt{\frac{4d}{\eta}}$ the proof is complete. □

The next theorem improves upon our previous converse, theorem II.7:

Theorem II.17 *For every $\lambda \in (0,1)$ there is a constant $K(a,d,\lambda)$ such that for every q–DMC W and every (n,λ)–code (f,D)*

$$|\mathcal{M}| \leq (n+1)^a \exp\left(nC(W) + K(a,d,\lambda)\sqrt{n}\right).$$

Proof. We will prove even more: under the additional assumption that all code words are of the same type P (such codes are called *constant composition*) one has

$$|\mathcal{M}| \leq \exp\left(nI(P;W) + K(a,d,\lambda)\sqrt{n}\right)$$

(from which the theorem clearly follows). To see this modify the decoder as follows: let

$$D'_m = \Pi^n_{V,PW,\delta} D_m \Pi^n_{V,PW,\delta}$$

with $\delta = \frac{\sqrt{32ad}}{1-\lambda}$. Then (f,D') is an $(n, \frac{1+\lambda}{2})$–code:

$$\begin{aligned}
\mathrm{Tr}\,(W_{f(m)} D'_m) &= \mathrm{Tr}\,(W_{f(m)} \Pi^n_{V,PW,\delta} D_m \Pi^n_{V,PW,\delta}) \\
&= \mathrm{Tr}\,(W_{f(m)} D_m) - \mathrm{Tr}\,\left((W_{f(m)} - \Pi^n_{V,PW,\delta} W_{f(m)} \Pi^n_{V,PW,\delta}) D_m\right) \\
&\geq 1 - \lambda - \frac{1-\lambda}{2}
\end{aligned}$$

(the last line by lemma II.11 and the tender operator lemma I.4). Now from lemma II.10

$$\begin{aligned}
\mathrm{Tr}\,D'_m &\geq \left(\frac{1-\lambda}{2} - \frac{ad}{\delta^2}\right) \exp\left(nH(W|P) - Kd\sqrt{a}\delta\sqrt{n}\right) \\
&\geq \frac{1-\lambda}{4} \exp\left(nH(W|P) - Kd\sqrt{a}\delta\sqrt{n}\right).
\end{aligned}$$

On the other hand $\sum_{m\in\mathcal{M}} D'_m \leq \Pi^n_{V,PW,\delta}$, hence by lemma I.9

$$\sum_{m\in\mathcal{M}} \mathrm{Tr}\,D'_m \leq \exp\left(nH(PW) + Kd\delta\sqrt{n}\right)$$

and we are done. □

Reliability function For the finite q–DMC W with capacity $C(W)$ the *reliability function* $E(R)$ is defined by

$$E(R) = \liminf_{n\to\infty,\ \delta\to 0} -\frac{1}{n} \log e_{\min}(n, R-\delta).$$

From the previous section we see that $E(R) = 0$ for $R > C(W)$. On the other hand define the *greedy* bound

$$E_{\mathrm{g}}(R,P) = \max\{\min\{\mu_{\mathrm{i}}(L,P), \frac{1}{2}\mu_{\mathrm{c}}(L',P)\} : R \leq L' - L\},$$

with the *individual exponent* (which may be $+\infty$)

$$\mu_{\mathrm{i}}(L,P)=\inf\{D(V\|W|P):H(V|P)>L\},$$

and the *collective exponent* (which is finite)

$$\mu_{\mathrm{c}}(L',P)=\min\{D(\rho\|PW):H(\rho)\leq L'\}.$$

Then we have

Theorem II.18 *For $n>0$, a type P, and $R<I(P,W)$ there exist constant composition n–block codes (f,D) of type P with*

$$|\mathcal{M}|\geq(n+1)^{d-ad}\exp(nR)$$

and error probability

$$e(f,D)\leq 8(n+1)^{ad}\exp(-nE_g(R,P))$$

if $n\geq n_0(a,d,P)$.

Proof. Let L,L' a pair of numbers with $R\leq L'-L$ and

$$E_{\mathrm{g}}(R,P)=\min\{\mu_{\mathrm{i}}(L,P),\frac{1}{2}\mu_{\mathrm{c}}(L',P)\}.$$

It is easily seen that we may assume $\mu_{\mathrm{i}}(L,P)\geq\frac{1}{2}\mu_{\mathrm{c}}(L',P)$. Also that in this case $L'<H(PW)$ and $L\geq H(W|P)$, in particular $E_{\mathrm{g}}(R,P)>0$.

Define $\lambda=8(n+1)^{ad}\exp(-nE_{\mathrm{g}}(R,P))$ and assume n to be large enough such that $\eta=\frac{\lambda^2}{32}\leq 1-\lambda$. Let (f,D) a maximal (n,λ)–code with the additional requirement

$$\forall m\in\mathcal{M}\qquad \mathrm{Tr}\,D_m\leq(n+1)^{ad}\exp(nL).$$

We claim that with $B=\sum_{m\in\mathcal{M}}D_m$

$$\forall x^n\text{ of type }P\qquad \mathrm{Tr}\,(W_{x^n}B)\geq\eta.$$

For else we could extend our code by an exceptional x^n and corresponding decoding operator

$$D'=\sqrt{\mathbb{1}-B}\,\Pi^n_{W,H(\cdot|P)\leq L}(x^n)\sqrt{\mathbb{1}-B}\;.$$

The argument is as in the proof of theorem II.4: observe that $\Pi^n_{W,H(\cdot|P)\leq L}(x^n)$, and hence D', satisfies the trace requirement, and

$$\mathrm{Tr}\,\left(W_{x^n}\Pi^n_{W,H(\cdot|P)\leq L}(x^n)\right)\geq 1-(n+1)^{ad}\exp(-n\mu_{\mathrm{i}}(L,P)).$$

Consequently

$$\mathrm{Tr}\,\left((PW)^{\otimes n}B\right)\geq\eta(n+1)^{-a}$$

and by lemma II.15

$$\begin{aligned}\mathrm{Tr}\,B&\geq\left(\eta(n+1)^{-a}-(n+1)^d\exp(-n\mu_{\mathrm{c}}(L',P))\right)\cdot\exp(nL'+n\mu_{\mathrm{c}}(L',P))\\&\geq(n+1)^d\exp(nL'),\end{aligned}$$

from which the estimate on $|\mathcal{M}|$ follows immediately. $\square$

Corollary II.19 *For $0 \leq R \leq C(W)$*

$$E(R) \geq E_g(R) = \max_{P\ p.d.:\ R \leq I(P;W)} E_g(R,P).$$

$\square$

Conversely, defining the *sphere packing* bound

$$E_{\rm sp}(R,P) = \min_{V \text{ channel: } I(P;V) \leq R} D(V\|W|P)$$

we have

Theorem II.20 *For $R \geq 0$ and $n > 0$ let (f,D) a constant composition n–block code (of type P) with*

$$|\mathcal{M}| \geq \exp(n(R+\delta)).$$

Then for the error probability

$$e(f,D) \geq \frac{1}{2}\exp(-nE_{sp}(R,P)(1+\delta))$$

if $n \geq n_0(a,d,\delta)$.

Proof. We can directly apply the original idea of Haroutunian (1968): consider a channel $V : \mathcal{X} \to \mathfrak{S}(\mathfrak{L}(\mathcal{H}))$ with $I(P;V) \leq R$. From the proof of the strong converse theorem II.17 we see that $e(f,D) \geq 1 - \frac{\delta}{2}$ if n is large enough (we assume $\delta < 1$). I.e. for some message $m \in \mathcal{M}$ and $S_m = \mathbb{1} - D_m$

$$\mathrm{Tr}\,(V_{f(m)}S_m) \geq 1 - \frac{\delta}{2}$$

Now generally for two states ρ, σ and complementary positive operators S, D (i.e. $S+D = \mathbb{1}$) one has

$$\mathrm{Tr}\,(\rho S)\log\frac{\mathrm{Tr}\,(\rho S)}{\mathrm{Tr}\,(\sigma S)} + \mathrm{Tr}\,(\rho D)\log\frac{\mathrm{Tr}\,(\rho D)}{\mathrm{Tr}\,(\sigma D)} \leq D(\rho\|\sigma).$$

This follows immediately from the monotonicity of quantum I–divergence, theorem A.5, applied to the completely positive, trace preserving map

$$\begin{aligned} \mathfrak{L}(\mathcal{H})_* &\longrightarrow \mathbb{C}^2 \\ \alpha &\longmapsto \mathrm{Tr}\,(\alpha S)e_1 + \mathrm{Tr}\,(\alpha D)e_2\ . \end{aligned}$$

From this we get by elementary operations

$$\mathrm{Tr}\,(\sigma S) \geq \exp\left(-\frac{D(\rho\|\sigma) + h(\mathrm{Tr}\,(\rho S))}{\mathrm{Tr}\,(\rho S)}\right).$$

Applying this to $\rho = V_{f(m)}$, $\sigma = W_{f(m)}$ and $S = S_m$, $D = D_m$ we find

$$\begin{aligned}\mathrm{Tr}\,(W_{f(m)}S_m) &\geq \exp\left(-\frac{nD(V\|W|P) + h\left(1-\frac{\delta}{2}\right)}{1-\frac{\delta}{2}}\right) \\ &\geq \frac{1}{2}\exp(-nD(V\|W|P)(1+\delta))\end{aligned}$$

if only δ is small enough (which is no real restriction). Now we choose V such that $D(V\|W|P)$ is minimal. □

Corollary II.21 *For $0 \leq R \leq C(W)$ (with the possible exception of the leftmost finite value of E_{sp})*

$$E(R) \leq E_{sp}(R) = \max_{P\ p.d.} E_{sp}(R, P).$$

Proof. To apply the theorem we have just to note the continuity of E_{sp} in R, which follows from its convexity. □

Remark II.22 *The proof obviously also works for infinite input alphabet, if only we have a strong converse which indeed we have, by the previous section.*

Remark II.23 *The reader may wish to apply the techniques of the previous proofs to show that $e(f, D)$ tends to 1 exponentially for rates above the capacity. The results however yield nothing of interest beyond the analysis of* Ogawa & Nagaoka *(1998).*

Holevo bound

An interesting application of our converse theorem II.17 is in a new, and completely elementary, proof of the Holevo bound (theorem A.16):

For a q–DMC $W : \mathcal{X} \to \mathfrak{S}(\mathfrak{L}(\mathcal{H}))$, a p.d. P on $\mathcal{X}$ and D an observable on $\mathfrak{L}(\mathcal{H})$, say indexed by $\mathcal{Y}$, the composition $D_* \circ W : \mathcal{X} \to \mathcal{Y}$ is a classical channel.

Holevo (1973) considered $C_1 = \max_{P,D} I(P; D_* \circ W)$ (the capacity if one is restricted to tensor product observables!) and proved $C_1 \leq C(W)$. For us this is now clear, since all codes for the classical channel $D_* \circ W$ (whose maximal rates are asymptotically just C_1) can be interpreted as special channel codes for W.

But we can show even a little more, namely Holevo's original *information bound* $I(P; D_* \circ W) \leq I(P; W)$ (from which the capacity estimate clearly follows).

Proof. Assume the opposite, $I(P; D_* \circ W) > I(P; W)$. Then by the well known classical coding theorem (Shannon (1948) — alternatively theorem II.4 which by remark II.3 generalizes the classical case) there is to every $\delta > 0$ an infinite sequence of $(n, 1/2)$–codes with codewords chosen from $\mathcal{T}^n_{V,P,\sqrt{2a}}$ for the channel $D_* \circ W$ with rates exceeding $I(P; D_* \circ W) - \delta$. Restricting to a single type of codewords we find constant composition codes (of type P_n) with rate exceeding $I(P; D_* \circ W) - 2\delta$ (if n is large enough).

As already explained these are special channel codes for W, so by theorem II.17 (proof) their rates are bounded by $I(P_n; W) + \delta$ (again, n large enough), hence

$$I(P; D_* \circ W) - 2\delta \leq I(P_n; W) + \delta.$$

Collecting inequalities we find

$$I(P; W) < I(P; D_* \circ W) \leq I(P_n; W) + 3\delta.$$

But since $P_n \to P$ by assumption and by the continuity of I in P (see lemma A.4), since furthermore δ is arbitrarily small, we end up with

$$I(P; W) < I(P; D_* \circ W) \leq I(P; W),$$

a contradiction. □

Open questions

We left open a number of problems:

Entangled input Is it possible to exceed the rate $C^{(1)} = C(\varphi_*) = \max_P I(P; \varphi_*)$ by using block codes where not only product states but arbitrary (entangled) states are allowed as "codewords"? We conjecture that the "ultimate" classical information capacity of φ_*,

$$\tilde{C} = \limsup_{n\to\infty} \frac{1}{n} \max_P I(P; \varphi_*^{\otimes n})$$

equals $C^{(1)}$ (compare SCHUMACHER & WESTMORELAND (1997)).

Computations Closely related is the issue of constructing a feasible algorithm to numerically compute the quantity $C^{(1)}$, maybe by an adaption of ARIMOTO's algorithm for computing the capacity of a classical channel (cf. ideas of NAGAOKA (1998)). This could be used for experimental tests of whether $C^{(n)} = \frac{1}{n} \max_P I(P; \varphi_*^{\otimes n})$ exceeds $C^{(1)}$.

Abstract approach In the proofs so far we relied heavily on the product structure of the n–fold channel. For reasons of better understanding of the foundations, as well as for having a unified framework for proof, it is desireable to have "abstract" coding theorems and converses at ones disposal. What this means is that time structure (blocks, in our case even products) is not used: after all the n–fold use of a channel is just a channel with larger alphabet. This is e.g. how FANO's inequality is used in weak converses. For something closer to our present setting compare WOLFOWITZ (1964), chapter 7.

- Prove an abstract coding theorem in this spirit!
- Prove the abstract converse, by exhibiting a usable "packing lemma", as is known in the classical theory.

Blowing up Prove a blowing up lemma as in the classical theory (commutative $\mathfrak{A}$), due to Ahlswede et al. (1976)! I suggest the following definition:

Let $\mathfrak{A} = \mathfrak{L}(\mathcal{H})$ a C*–algebra with $q = \dim_{\mathbb{C}} \mathfrak{A}$, and $\Pi \in \mathfrak{A}^{\otimes n}$ a projector. Define the *blow–up* of Π as

$$\Gamma\Pi = \mathrm{l.\,c.\,supp}\{A_{(i)}\Pi A_{(i)}^* :\ 1 \leq i \leq n,\ A \in \mathfrak{A},\ A^*A \leq \mathbb{1}\}$$

where $A_{(i)} = \mathbb{1}^{\otimes(i-1)} \otimes A \otimes \mathbb{1}^{\otimes(n-i)}$. The l^{th} *blow–up* of Π is $\Gamma^l\Pi$, defined as

$$\Gamma^l\Pi = \mathrm{l.\,c.\,supp}\{A_{(I)}\Pi A_{(I)}^* :\ I \subset [n],\ |I| = l,\ A \in \mathfrak{A}^{\otimes l},\ A^*A \leq \mathbb{1}\}$$

where $A_{(I)} = \mathbb{1}^{\otimes([n]\setminus I)} \otimes A$ (in the right order).

In loose words: $\Gamma^l\Pi$ is the least common support of all images of Π under all quantum operations confined to l positions (factors in the tensor product).

Lemma II.24 *The blowing up operation has the following properties:*

1. *$\Gamma^l\Pi$ is a projector.*
2. *Γ^l is the l–fold iteration of Γ.*
3. *For $0 \leq l \leq l'$ one has $\Pi \leq \Gamma^l\Pi \leq \Gamma^{l'}\Pi$.*
4. *$\mathrm{Tr}\,\Gamma^l\Pi \leq (qn)^l \cdot \mathrm{Tr}\,\Pi$.*

Proof. Points (1) and (3) are obvious. For (2) and (4) write $\Pi = \sum_{\pi\in\mathsf{P}} \pi$ for a set P of (necessarily orthogonal) minimal idempotents. Clearly $\mathrm{Tr}\,\Pi = |\mathsf{P}|$. Then

$$\Gamma^l\Pi = \mathrm{l.\,c.\,supp}\{A_{(I)}\pi A_{(I)}^* :\ \pi \in \mathsf{P},\ I \subset [n],\ |I| = l,\ A \in \mathfrak{A}^{\otimes l},\ A^*A \leq \mathbb{1}\}$$

and the supporting subspace[2] of this is

$$\sum_{\pi=|\psi\rangle\langle\psi|\in\mathsf{P}} \mathrm{span}\{A_{(I)}|\psi\rangle :\ I \subset [n],\ |I| = l,\ A \in \mathfrak{A}^{\otimes l},\ A^*A \leq \mathbb{1}\}.$$

But $\mathfrak{A}$ has a linear basis $(A_1, \ldots, A_q)$ which produces by tensor products a basis of length q^l of $\mathfrak{A}^{\otimes l}$. This shows (2), and since there are at most n^l many $I \subset [n]$ of cardinality l we get (4). □

Conjecture II.25 *Let W a fixed q–DMC, m_W the smallest non–zero eigenvalue of the W_x, $x^n \in \mathcal{X}^n$, and B a projector. Then*

$$\mathrm{Tr}\,(W_{x^n}\Gamma^l B) \geq \Phi\left(\Phi^{-1}(\mathrm{Tr}\,(W_{x^n}B)) + a\frac{l-1}{\sqrt{n}}\right),$$

with $a = c\dfrac{m_W}{\sqrt{-\ln m_W}}$, where $c > 0$ is a universal constant and $\Phi : \mathbb{R} \to [0,1]$ is the Gaussian distribution function: $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt$.

Among the possible applications would be the transition from weak to strong converses (after Ahlswede & Dueck, cf. Csiszár & Körner (1981), chapter 2.1).

[2]In $\mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_m$, which we think of $\mathfrak{A} = \bigoplus_{i=1}^m \mathfrak{L}(\mathcal{H}_i)$ to live on!

Reliability function We proved the sphere packing bound and a lower bound on the reliability function which at least shows its positivity for rates below the capacity. For the pure state channel this is matched by random coding and expurgated lower bounds of Burnashev & Holevo (1997). Unfortunately in this case our sphere packing bound is trivial!

We leave as open problems: the proof of a random coding lower bound in the general case (which should enable us to *determine* the reliability function above a critical rate), and (at least in the pure state case) to find a suitable modification of the sphere packing bound (as the present formulation does not take into account possible noncommutativity).

Chapter III

Quantum Multiple Access Channels

The multiway channel with s senders and r receivers in classical information theory was already studied by SHANNON (1961). AHLSWEDE (1971) and AHLSWEDE (1974a) first determined its capacity region. For a good overview on multiuser communication theory one should consult EL GAMAL & COVER (1980). In the present chapter we will define the corresponding quantum channel (after recent work by ALLAHVERDYAN & SAAKIAN (1997b)), extending the results of the previous chapter: we bound the capacity region (in the limit of vanishing error probability), and — for the multiple access channel, i.e. one receiver — we are able to prove the corresponding coding theorem.

Quantum multiway channels and capacity region

This is the simplest situation of multi–user communication in general: consider s independent senders, sender i using an alphabet $\mathcal{X}_i$, say with an a priori probability distribution P_i. We describe this by the quantum state $\sigma_i = \sum_{x_i\in\mathcal{X}_i} P_i(x_i)x_i$ on the commuative C*–algebra $\mathfrak{X}_i = \mathbb{C}\mathcal{X}_i$ generated by the x_i which are mutually orthogonal idempotents (to distinguish these as generators of this algebra we will sometimes write $[x_i]$). The channel is then a map

$$W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_s \to \mathfrak{S}(\mathfrak{Y})$$

with a (finite dimensional) C*–algebra $\mathfrak{Y}$, which connects the input $(x_1, \ldots, x_s)$ with the output $W_{x_1\ldots x_s}$. By linear extension we may view W as a completely positive, trace preserving map from $\mathfrak{X}_{1*} \otimes \cdots \otimes \mathfrak{X}_{s*}$ to $\mathfrak{Y}_*$. The receivers are modelled by compatible $*$–subalgebras $\mathfrak{Y}_j$ (see appendix A, section *Quantum systems*).

If all the $W_{x_1\ldots x_s}$ commute with each other (hence have a common diagonalization) the channel is called *classical*.

For fixed a priori distributions we have the *channel state*

$$\gamma = \sum_{\forall i\ x_i\in\mathcal{X}_i} P_1(x_1)\cdots P_s(x_s)[x_1]\otimes\cdots\otimes[x_s]\otimes W_{x_1\ldots x_s}$$

on $\mathfrak{X}_1 \otimes \cdots \otimes \mathfrak{X}_s \otimes \mathfrak{Y}$.

For a subset $J \subset [s]$ denote $P_J = \bigotimes_{i\in J} P_i$, i.e. $P_J(x_i|i \in J) = \prod_{i\in J} P_i(x_i)$, and $\mathcal{X}(J) = \prod_{i\in J} \mathcal{X}_i$ (similarly $\mathfrak{X}(J) = \bigotimes_{i\in J} \mathfrak{X}_i$).

Further define a *reduced channel* $P_{J^c}W : \mathcal{X}(J) \to \mathfrak{S}(\mathfrak{Y})$ by

$$P_{J^c}W : (x_i|i \in J) \longmapsto \sum_{\forall i\in J^c:\ x_i\in\mathcal{X}_i} P_{J^c}(x_i|i \in J^c)W_{x_1\dots x_s}\ .$$

Note that

$$\operatorname{Tr}_{\mathfrak{X}(J^c)}\gamma = \sum_{\forall i\in J:\ x_i\in\mathcal{X}_i} P_J(x_i|i \in J)[x_i|i \in J] \otimes (P_{J^c}W)_{(x_i|i\in J)}\ .$$

An n–*block code* is a collection $(f_1, \dots , f_s, D_1, \dots , D_r)$ of maps $f_i : \mathcal{M}_i \to \mathcal{X}_i^n$ and decoding observables $D_j \subset \mathfrak{Y}_j^{\otimes n}$, indexed by $\mathcal{M}'_1 \times \cdots \times \mathcal{M}'_s \supset \mathcal{M}_1 \times \cdots \times \mathcal{M}_s$. There are r (average) *error probabilities* of the code, the probability that the receiver j guesses wrongly any one of the sent words, taken over the uniform distribution on the codebooks:

$$\bar{e}_j(f_1, \dots , f_s, D_j) = 1 - \frac{1}{|\mathcal{M}_1| \cdots |\mathcal{M}_s|} \sum_{\forall i: m_i\in\mathcal{M}_i} \operatorname{Tr}\left(W^{\otimes n}(f(m_1), \dots , f(m_s))D_{j,m_1\dots m_s}\right).$$

We call $(f_1, \dots , f_s, D_1, \dots , D_r)$ an $(n, \bar{\lambda})$–code if all $\bar{e}_j(f_1, \dots , f_s, D_j)$ are at most $\bar{\lambda}$.

The *rates* of the code are the $R_i = \frac{1}{n}\log|\mathcal{M}_i|$. A tuple $(R_1, \dots , R_s)$ is said to be *achievable*, if for any $\bar{\lambda}, \delta > 0$ there exists for any large enough n an $(n, \bar{\lambda})$–code with i–th rate at least $R_i - \delta$. The set of all achievable tuples (which is clearly closed, and convex by the *time sharing principle*, cf. Csiszár & Körner (1981), lemma 2.2.2) is called the *capacity region* of the channel.

Outer bounds

In the case $r = 1$, $s = 2$ the following theorem was already stated by Allahverdyan & Saakian (1997b), who also gave hints on the proof.

Theorem III.1 (Outer bounds) *The capacity region of the quantum multiway channel is contained in the closure of all nonnegative $(R_1, \dots , R_s)$ satisfying*

$$\forall J \subset [s], j \in [r] \qquad R(J) = \sum_{i\in J} R_i \le \sum_u q_u I_{\gamma_u}\left(\mathfrak{X}(J) \wedge \mathfrak{Y}_j|\mathfrak{X}(J^c)\right)$$

for some channel states γ_u (belonging to appropriate input distributions) and $q_u \ge 0$, $\sum_u q_u = 1$.

Proof. Consider any $(n, \bar{\lambda})$–code $(f_1, \dots , f_s, D_1, \dots , D_r)$ with rate tuple $(R_1, \dots , R_s)$. Then the uniform distribution on the codewords induces a channel state γ on the block $(\mathfrak{X}_1 \cdots \mathfrak{X}_s\mathfrak{Y})^{\otimes n}$. Its restriction to the u–th copy in this tensor power will be denoted γ_u. Let $j \in [r]$, $J \subset [s]$. By Fano inequality in the form of corollary A.25 we have

$$H(\mathfrak{X}^{\otimes n}(J)|\mathfrak{Y}_j^{\otimes n}\mathfrak{X}^{\otimes n}(J^c)) \le 1 + \bar{\lambda} \cdot nR(J).$$

With

$$H(\mathfrak{X}^{\otimes n}(J)|\mathfrak{Y}_j^{\otimes n}\mathfrak{X}^{\otimes n}(J^c)) = H(\mathfrak{X}^{\otimes n}(J)) - I(\mathfrak{X}^{\otimes n}(J) \wedge \mathfrak{Y}_j^{\otimes n}\mathfrak{X}^{\otimes n}(J^c))$$
$$= nR(J) - I(\mathfrak{X}^{\otimes n}(J) \wedge \mathfrak{Y}_j^{\otimes n}\mathfrak{X}^{\otimes n}(J^c))$$

we conclude (with subadditivity of mutual information, corollary A.18) that

$$(1-\bar{\lambda})R(J) \leq \frac{1}{n} + \frac{1}{n} I_\gamma(\mathfrak{X}^{\otimes n}(J) \wedge \mathfrak{Y}_j^{\otimes n}\mathfrak{X}^{\otimes n}(J^c))$$
$$\leq \frac{1}{n} + \frac{1}{n}\sum_{u=1}^{n} I_{\gamma_u}(\mathfrak{X}(J) \wedge \mathfrak{Y}_j\mathfrak{X}(J^c)).$$

□

Remark III.2 *In the case of classical channels the region described in the theorem is the exact capacity region (i.e. all the rates there are achievable), as was first proved by* AHLSWEDE *(1971) and* AHLSWEDE *(1974a).*

Remark III.3 *The numeric computation of the above regions is not yet possible from the given description: we need a bound on the number of different single–letter channel states one has to consider in the convex combinations. For the multiple access channel ($r = 1$) this is easy: by* CARATHEODORY*'s theorem s will suffice. For general r there are also classical bounds, which carry over unchanged to the quantum case (since the quantum mutual information has properties similar to those of classical mutual information): $r(2^s - 1)$ always suffice, as was observed by* AHLSWEDE *(1974b).*

Coding theorem for multiple access channels

With the notation as before for a quantum multiway channel W with one receiver we have

Theorem III.4 *An s–tuple $(R_1, \dots, R_s)$ is achievable (i.e. there is an infinite sequence of $(n, \bar{\lambda}_n)$–codes with $\bar{\lambda}_n \to 0$ and rate tuple tending to $(R_1, \dots, R_s)$), if and only if it is in the convex hull of the pairs satifying (for some input distributions which induce a channel state γ)*

$$\forall J \subset [s] \qquad R(J) = \sum_{i \in J} R_i \leq I_\gamma\left(\mathfrak{X}(J) \wedge \mathfrak{Y}|\mathfrak{X}(J^c)\right).$$

We shall prove this only in the case $s = 2$, the reader should have no difficulty to see the extension to larger numbers. In this case the conditions reduce to

$$R_1 + R_2 \leq I(\mathfrak{Y} \wedge \mathfrak{X}_1\mathfrak{X}_2),$$

$$R_1 \leq I(\mathfrak{Y} \wedge \mathfrak{X}_1|\mathfrak{X}_2), \qquad R_2 \leq I(\mathfrak{Y} \wedge \mathfrak{X}_2|\mathfrak{X}_1).$$

That these are necessary is of course theorem III.1. For proof of the achievability it is (by the time sharing principle) sufficient to consider an extreme point of the region described by the above inequalities for a particular channel state. It is easily seen that w.l.o.g. $R_1 = I(\mathfrak{X}_1 \wedge \mathfrak{Y})$, $R_2 = I(\mathfrak{X}_2 \wedge \mathfrak{Y}\mathfrak{X}_1)$. That this point is achievable follows immediately from theorem IV.14 and the following theorem, applied with $\bar{R}_1 = I(\mathfrak{X}_1 \wedge \mathfrak{Y}) + \delta$ and $\bar{R}_2 = I(\mathfrak{X}_2 \wedge \mathfrak{X}_1\mathfrak{Y}) + \delta$.

Theorem III.5 *(Cf.* Csiszár & Körner *(1981), proof of theorem 3.2.3) Let $\bar{\lambda}, \delta > 0$, W a quantum multiple access channel with two senders, and P_i probability distributions on the sender alphabets $\mathcal{X}_i$. Define the c^2h^1–source (see chapter IV, section* Correlated quantum sources*) $(\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{Y}, \mathcal{X}_1 \times \mathcal{X}_2 \times \mathsf{P}, P)$ on $\mathfrak{X}_1 \otimes \mathfrak{X}_2 \otimes \mathfrak{Y}$ by $P(x_1 \otimes x_2 \otimes \pi) = P_1(x_1)P_2(x_2)q_{\pi|x_1x_2}$, where P is a set of pure states on $\mathfrak{Y}$ and the $q_{\pi|x_1x_2} \geq 0$ are such that $W_{x_1x_2} = \sum_{\pi\in\mathsf{P}} q_{\pi|x_1x_2}\pi$ (e.g. diagonalize all $W_{x_1x_2}$ and take P to be the set of all eigenstates occuring.*

Then from any $(n, \bar{\lambda})$–coding scheme $(g_1, g_2, D^{(0)})$ with quantum side information at the decoder for this source, with rates $\bar{R}_1, \bar{R}_2$, one can construct an $(n, 4\bar{\lambda})$–code (f_1, f_2, D) for W with rates $R_i \geq H(P_i) - \bar{R}_i - \delta$, provided $n \geq n_0(|\mathcal{X}_1|, |\mathcal{X}_2|, \delta)$.

Proof. Let $g_1 : \mathcal{X}_1^n \to \mathcal{M}_1$ and $g_2 : \mathcal{X}_2^n \to \mathcal{M}_2$ the encodings, $D^{(0)}$ the observable on $\mathbb{C}\mathcal{M}_1 \otimes \mathbb{C}\mathcal{M}_2 \otimes \mathfrak{Y}$ indexed by $\mathcal{X}_1^n \times \mathcal{X}_2^n$. Observe that it is of the form

$$D^{(0)}_{x_1^n x_2^n} = \sum_{m_1\in\mathcal{M}_1, m_2\in\mathcal{M}_2} m_1 \otimes m_2 \otimes D'_{m_1m_2,x_1^nx_2^n}\ .$$

Define for every $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$

$$\mathcal{A}_{m_1} = g_1^{-1}\{m_1\}, \qquad \mathcal{B}_{m_2} = g_2^{-1}\{m_2\}.$$

Assume that the $\mathcal{A}_{m_1}$, $\mathcal{B}_{m_2}$ consist of words of single type (otherwise one modifies the coding by also encoding the type of the sequences, increasing the rate negligibly, in the asymptotics).

Construct now codes $(f_1^{(m_1m_2)}, f_2^{(m_1m_2)}, D^{(m_1m_2)})$ for W as follows:
$f_1^{(m_1m_2)} = \mathrm{id}_{\mathcal{A}_{m_1}}$, $f_2^{(m_1m_2)} = \mathrm{id}_{\mathcal{B}_{m_2}}$ and $D^{(m_1m_2)}$ an observable on $\mathfrak{Y}$ indexed by $\mathcal{A}_{m_1} \times \mathcal{B}_{m_2}$ with $D^{(m_1m_2)}_{x_1^nx_2^n} \geq D'_{m_1m_2,x_1^nx_2^n}$.

As in Csiszár & Körner (1981), pp.272 we can see that for the error probabilities

$$\sum_{m_1\in\mathcal{M}_1}\sum_{m_2\in\mathcal{M}_2} P_1^n(\mathcal{A}_{m_1})P_2^n(\mathcal{B}_{m_2})\bar{e}(f_1^{(m_1m_2)}, f_2^{(m_1m_2)}, D^{(m_1m_2)}) \leq \bar{e}(g_1, g_2, D^{(0)})$$

and again copying from Csiszár & Körner (1981) we find that there is one of them having $\bar{e}(f_1^{(m_1m_2)}, f_2^{(m_1m_2)}, D^{(m_1m_2)}) \leq 4\bar{\lambda}$ and rates $R_i \geq H(P_i) - \bar{R}_i - \delta$, if n is large enough. $\square$

Open questions

Random coding The major drawback of the above method of proof is that it allows no direct code construction for every point in the capacity region, as does the proof of Ahlswede (1974a) (we needed to invoke the time sharing principle). It seems that this approach is no longer possible if there are two or more receivers present. The above outer bounds however we conjecture to be the correct ones (by formal analogy with the classical case). A proof of the corresponding coding theorem would be highly desireable, possibly by a cleverly adapted random coding argument (see the proofs of the quantum channel coding theorem by Holevo (1998a) and Schumacher & Westmoreland (1997)). It should be clear that such a proof is far more natural than the one we presented here. For a proof of the quantum multiple access channel coding theorem which does not rely on code partitions and reduction to a source coding problem but instead uses iterated "slicing" of the rate with random code selection, see Winter (1998a).

Chapter IV

Quantum Multiple Source Coding

Having investigated in chapter I the problem of quantum source coding we now turn to the problem of (independent) source coding of possibly dependent sources. In the first section we will introduce the mathematical model, and venture then to analyze this model as far as possible (which, as it will turn out, is not very much): we will restrict ourselves mostly to double sources, proving some general bounds and presenting characteristic examples. Then we study the particular case that only one of the sources is quantum, the others being classical. We are thus led to consider the problem of coding with side information, which for this kind of source we can in part solve. In general however there is to be distinguished between multiple source coding and coding with side information.

Correlated quantum sources

A *multiple (s–fold) quantum source* is a tuple $(\mathfrak{A}_1, \ldots, \mathfrak{A}_s, \mathsf{P}, P)$ of C*–algebras $\mathfrak{A}_i$ (with us: finite dimensional), a finite set P of pure states on $\mathfrak{A} = \mathfrak{A}_1 \otimes \cdots \otimes \mathfrak{A}_s$ and a p.d. P on P.

The *average state* of the source is the state $P\mathsf{P}$ on $\mathfrak{A}$, its marginal restricted to $\mathfrak{A}^{\otimes I} = \bigotimes_{i\in I} \mathfrak{A}_i$ is denoted $P\mathsf{P}|_I$.

We call the source *classically correlated* if all the states $\pi \in \mathsf{P}$ are product states with respect to $\mathfrak{A}_1, \ldots, \mathfrak{A}_s$: $\pi = \pi_1 \otimes \cdots \otimes \pi_s$, $\pi_i \in \mathfrak{S}(\mathfrak{A}_i)$. In this case we obtain for each $J \subset [n]$ a multiple source $((\mathfrak{A}_j | j \in J), \mathsf{P}|_J, P)$ by restricting the $\pi \in \mathsf{P}$ to $\mathfrak{A}^{\otimes J}$, i.e. replacing π by $\pi|_J$. Always in this situation we assume w.l.o.g. $\mathsf{P} = \mathsf{P}_1 \times \cdots \times \mathsf{P}_s$

If in particular k of the $\mathfrak{A}_i$ are classical (i.e. commutative), l are fully quantum (i.e. full matrix algebras) and the remaining m are arbitrary ("hybrid"), we speak of a $c^k q^l h^m$–*source.*

An *n–block coding scheme with quantum encoding* for a multiple quantum source $(\mathfrak{A}_1, \ldots, \mathfrak{A}_s, \mathsf{P}, P)$ is a tuple $(\varepsilon_{1*}, \ldots, \varepsilon_{s*}, \delta_*)$ with quantum operations

$$\begin{aligned} \varepsilon_{i*} :& \mathfrak{A}_{i*}^{\otimes n} \longrightarrow \mathfrak{L}(\mathcal{K}_i)_* \\ \delta_* :& \mathfrak{L}(\mathcal{K}_1 \otimes \cdots \otimes \mathcal{K}_s)_* \longrightarrow \mathfrak{A}_{1*}^{\otimes n} \otimes \cdots \otimes \mathfrak{A}_{s*}^{\otimes n}. \end{aligned}$$

An *n–block coding scheme with arbitrary encoding* for a classically correlated (!) multiple quantum source $(\mathfrak{A}_1, \ldots, \mathfrak{A}_s, \mathsf{P}, P)$ is a tuple $(\varepsilon_{1*}, \ldots, \varepsilon_{s*}, \delta_*)$ with

$$\begin{aligned} \varepsilon_{i*} :& \mathsf{P}_i^n \longrightarrow \mathfrak{S}(\mathfrak{L}(\mathcal{K}_i)) \quad \text{mappings and} \\ \delta_* :& \mathfrak{L}(\mathcal{K}_1 \otimes \cdots \otimes \mathcal{K}_s)_* \longrightarrow \mathfrak{A}_{1*}^{\otimes n} \otimes \cdots \otimes \mathfrak{A}_{s*}^{\otimes n} \quad \text{a quantum operation.} \end{aligned}$$

Writing $\varepsilon_* = \varepsilon_{1*} \otimes \cdots \otimes \varepsilon_{s*}$ we define the *average fidelity* and *average distortion* of the scheme $(\varepsilon_{1*}, \ldots, \varepsilon_{s*}, \delta_*)$ as expected:

$$\begin{aligned} \bar{F}(\varepsilon_{1*}, \ldots, \varepsilon_{s*}, \delta_*) &= \sum_{\pi^n \in \mathsf{P}^n} P^n(\pi^n) \cdot \mathrm{Tr}\left((\delta_* \varepsilon_* \pi^n) \pi^n\right), \\ \bar{D}(\varepsilon_{1*}, \ldots, \varepsilon_{s*}, \delta_*) &= \sum_{\pi^n \in \mathsf{P}^n} P^n(\pi^n) \cdot \frac{1}{2} \|\delta_* \varepsilon_* \pi^n - \pi^n\|_1 \ . \end{aligned}$$

If all $\mathfrak{A}_i$ are fully quantum, say $\mathfrak{A}_i = \mathfrak{L}(\mathcal{H}_i)$, we can define the *entanglement fidelity* by

$$F_e(\varepsilon_{1*}, \ldots, \varepsilon_{s*}, \delta_*) = \mathrm{Tr}\ \left(((\delta_* \varepsilon_* \otimes \mathrm{id}) \Psi_{P\mathsf{P}}^{\otimes n}) \Psi_{P\mathsf{P}}^{\otimes n}\right).$$

Quite obviously theorem I.1 for these quality measures is still valid. It should be clear also what we mean by $(n, \lambda)_{\bar{F}}$–, $(n, \lambda)_{\bar{D}}$–, and $(n, \lambda)_{F_e}$–coding schemes.

The *rate tuple* $(R_1, \ldots, R_s)$ of the coding scheme is defined by $R_i = \frac{1}{n} \log \dim \mathcal{K}_i$. A tuple $(R_1, \ldots, R_s)$ is called (quantum, $\bar{F}$)–achievable if there is a sequence of $(n, \lambda_n)_{\bar{F}}$–coding schemes with rate tuples converging to $(R_1, \ldots, R_s)$ and $\lambda \to 0$ as $n \to \infty$. The set $\mathbf{R}_{\mathrm{q},\bar{F}}$ of all (quantum, $\bar{F}$)–achievable rate tuples is called (quantum, $\bar{F}$)–*rate region.*

Analogously (arbitrary, $\bar{F}$)–, the same with $\bar{D}$, and (quantum, F_e)–achievability are defined, with their respective rate regions $\mathbf{R}_{\mathrm{a},\bar{F}}$, $\mathbf{R}_{\mathrm{q},\bar{D}}$, $\mathbf{R}_{\mathrm{a},\bar{D}}$ and $\mathbf{R}_{\mathrm{q},F_e}$.

It is clear from the definition that the rate regions are closed, convex (by the time sharing principle) and right upper closed (increasing some of the R_i does not leave the rate region). Also we have the following quite obvious inclusions:

$$\begin{array}{ccccc} \mathbf{R}_{\mathrm{q},F_e} & \subset & \mathbf{R}_{\mathrm{q},\bar{F}} & \subset & \mathbf{R}_{\mathrm{a},\bar{F}} \\ & & \| & & \| \\ & & \mathbf{R}_{\mathrm{q},\bar{D}} & & \mathbf{R}_{\mathrm{a},\bar{D}} \end{array}$$

Note that the different rate regions depend on the *ensemble* (P, P), only $\mathbf{R}_{\mathrm{q},F_e}$ is obvious to depend only on the average state $P\mathsf{P}$. For the others we will present evidence that they do in fact depend on further properties of (P, P) besides $P\mathsf{P}$.

Some general bounds Consider first a double source, quantum encoding with average fidelity:

Theorem IV.1 *Let* $(\mathfrak{A}_1, \mathfrak{A}_2, \mathsf{P}, P)$ *a double quantum source and* (R_1, R_2) *a* (*quantum,* $\bar{D}$)*–achievable pair. Then with the average state* $P\mathsf{P}$ *on* $\mathfrak{A} = \mathfrak{A}_1 \otimes \mathfrak{A}_2$

$$R_1 + R_2 \geq H(\mathfrak{A}_1\mathfrak{A}_2), \;\; R_1 \geq H(\mathfrak{A}_1|\mathfrak{A}_2), \;\; R_2 \geq H(\mathfrak{A}_2|\mathfrak{A}_1).$$

Proof. The first inequality follows from the converse to source coding, in the generalized form of theorem I.22. For the second, consider an $(n, \lambda)_{\bar{D}}$–coding scheme $(\varepsilon_{1*}, \varepsilon_{2*}, \delta_*)$ with quantum encoding which has rate pair $(R_1 + \epsilon, R_2 + \epsilon)$. Modify the coding scheme as follows (for n large enough):

$\mathfrak{A}_1$ encodes just as before, but $\mathfrak{A}_2$ uses SCHUMACHER's data compression to encode his part in $H(\mathfrak{A}_2) + \epsilon$ qubits per symbol and with $\bar{D} \leq \frac{1-\lambda}{2}$. The decoder first "unpacks" the signal from $\mathfrak{A}_2$ and then applies $\mathfrak{A}_2$'s previous encoding ε_{2*}. After that she applies her previous decoding δ_*. Let us estimate the average trace norm distortion of the new scheme: by the non–increasing of $\|\cdot\|_1$ under quantum operations and triangle inequality it is at most $\frac{1+\lambda}{2}$. Thus from theorem I.22 it follows that $R_1 + H(\mathfrak{A}_2) + 2\epsilon \geq H(\mathfrak{A})$, and since ϵ is arbitrarily small we get the second inequality. The third one is exactly symmetrical. □

Example IV.2 (Cloned wheel) Consider the c^0q^2–source $(\mathfrak{A}_1, \mathfrak{A}_2, \mathsf{P}, P)$ given by $\mathfrak{A}_1 = \mathfrak{A}_2 = \mathfrak{L}(\mathbb{C}^2)$, and P is equidistributed on

$$\mathsf{P} = \{|00\rangle\langle 00|, \; |11\rangle\langle 11|, \; |{++}\rangle\langle{++}|, \; |{--}\rangle\langle{--}|\},$$

where $\{|0\rangle, |1\rangle\}$ is an orthonormal basis of $\mathbb{C}^2$, and $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. So the average state of the source is

$$P\mathsf{P} = \frac{1}{4}\left(|00\rangle\langle 00| + |11\rangle\langle 11| + |{++}\rangle\langle{++}| + |{--}\rangle\langle{--}|\right)$$

and clearly the marginals are

$$P\mathsf{P}|_{\mathfrak{A}_1} = P\mathsf{P}|_{\mathfrak{A}_2} = \frac{1}{2}\mathbb{1}.$$

Since each of the sent pairs is clearly invariant under exchange of $\mathfrak{A}_1$ and $\mathfrak{A}_2$ we see that so is $P\mathsf{P}$, i.e. $P\mathsf{P}$ is supported on the three–dimensional *symmetrical subspace* $\mathrm{Sym}_2(\mathbb{C}^2)$ of $\mathbb{C}^2 \otimes \mathbb{C}^2$. In fact, an orthonormal basis of $\mathrm{Sym}_2(\mathbb{C}^2)$ is given by the *triplet* BELL *states*

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right)$$
$$|\Phi^-\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle - |11\rangle\right)$$
$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle + |10\rangle\right)$$

and it is readily checked that

$$P\mathsf{P} = \frac{1}{2}|\Phi^+\rangle\langle\Phi^+| + \frac{1}{4}|\Phi^-\rangle\langle\Phi^-| + \frac{1}{4}|\Psi^+\rangle\langle\Psi^+|.$$

Thus $H(P\mathsf{P}) = 3/2$ and it is clear from the previous theorem IV.1 that with quantum encoding one gets $R_1 + R_2 \geq 3/2$, $R_1, R_2 \geq 1/2$:

$$\mathbf{R}_{\mathrm{q},\bar{F}} \subset \{(R_1, R_2): \ R_1, R_2 \geq 1/2, R_1 + R_2 \geq 3/2\}.$$

This might appear strange: naïvely, in the coding $\mathfrak{A}_2$ (say) is unnecessary, since its state is identical to that of $\mathfrak{A}_1$ (which would mean that the uncertainty of the state of $\mathfrak{A}_2$ given that of $\mathfrak{A}_1$ is zero). So let's try the following coding scheme: $\mathfrak{A}_2$ transmits nothing, whereas $\mathfrak{A}_1$ transmits his state π faithfully using one qubit. But the task of the decoder is to reconstruct the total state, i.e. $\pi \otimes \pi$, which is clearly impossible by the no–cloning theorem. So we see that there is indeed a sense in the above inequalities.

However, in the model with *arbitrary encoding*, the first encoder *can* replace his state π by $\pi \otimes \pi$ and code it into (asymptotically) 3/2 qubits per symbol using SCHUMACHER's quantum coding. Hence

$$\mathbf{R}_{\mathrm{a},\bar{F}} = \{(R_1, R_2): \ R_1, R_2 \geq 0, R_1 + R_2 \geq 3/2\},$$

and thus we learn:

$$\boxed{\text{In general } \mathbf{R}_{\mathrm{a},\bar{F}} \text{ and } \mathbf{R}_{\mathrm{q},\bar{F}} \text{ are different.}}$$

Remark IV.3 *In the proof of theorem IV.1 a coding theorem (*SCHUMACHER*'s) was used. Thus, to prove lower bounds for more than two sources, we need some coding theorem for correlated quantum sources.*

Interestingly we *can* prove directly lower bounds on the resources needed for schemes with quantum encoding having high entanglement fidelity. We employ for this the following concepts from SCHUMACHER (1996), SCHUMACHER & NIELSEN (1996):

For a quantum operation $\varphi_* : \mathfrak{L}(\mathcal{H})_* = \mathfrak{A}_* \to \mathfrak{A}_*$ and a state ρ on $\mathfrak{A}$ choose a purification Ψ_ρ of ρ on the extended system $\mathfrak{A} \otimes \mathfrak{R}$ (for *reference* system). The *entropy exchange*[1] is defined as

$$S_e(\rho; \varphi_*) = H\left((\varphi_* \otimes \mathrm{id}_{\mathfrak{R}*})\Psi_\rho\right)$$

and SCHUMACHER (1996) shows that it does not depend on the purification chosen. It can be seen as a measure for the quantum information exchange between system and environment.

Thus it is natural to define the *coherent information* (after SCHUMACHER & NIELSEN (1996)) as

$$I_e(\rho; \varphi_*) = H(\varphi_*\rho) - S_e(\rho; \varphi_*).$$

From BARNUM et al. (1998) we take the following lemma, which is a direct consequence of the *quantum* FANO *inequality* from SCHUMACHER (1996).

[1]We adopt the name S_e for this following SCHUMACHER (1996) and general physical fashion.

Lemma IV.4 *Let φ_*, ψ_* quantum operations on the system $\mathfrak{A}$, ρ a state on $\mathfrak{A}$ and denote $d^2 = \dim_{\mathbb{C}} \mathfrak{A}$. Then*

$$H(\rho) \leq I_e(\rho; \varphi_*) + 2 + 4(1 - F_e(\psi_* \circ \varphi_*)) \log d.$$

$\square$

We are now ready to prove

Lemma IV.5 (Weak subadditivity of coherent information) *Let ρ a state on $\mathfrak{A}_1 \otimes \mathfrak{A}_2$ with marginals ρ_1, ρ_2, and $\varphi_{1*}, \varphi_{2*}$ quantum operations on $\mathfrak{A}_1, \mathfrak{A}_2$, respectively. Then*

$$I_e(\rho; \varphi_{1*} \otimes \varphi_{2*}) \leq I_e(\rho_1; \varphi_{1*}) + H(\rho_2).$$

Proof. Introducing environment systems $\mathfrak{E}_1$, $\mathfrak{E}_2$, pure "null" states τ_1 on $\mathfrak{E}_1$, τ_2 on $\mathfrak{E}_2$ and unitaries on the underlying Hilbert space of $\mathfrak{A}_1 \otimes \mathfrak{E}_1$, $\mathfrak{A}_2 \otimes \mathfrak{E}_2$, respectively, such that

$$\begin{aligned} \varphi_{1*}(\sigma) &= \operatorname{Tr}_{\mathfrak{E}_1} (U_1(\sigma \otimes \tau_1) U_1^*)) \\ \varphi_{2*}(\sigma) &= \operatorname{Tr}_{\mathfrak{E}_2} (U_2(\sigma \otimes \tau_2) U_2^*)) \end{aligned}$$

(which is possible by STINESPRING's theorem A.1). Now what we have to prove (with $\mathfrak{R} = \mathfrak{R}_1 \otimes \mathfrak{R}_2$) is

$$H((\varphi_{1*} \otimes \varphi_{2*})\rho) - H((\varphi_{1*} \otimes \varphi_{2*} \otimes \mathrm{id}_{\mathfrak{R}*})\Psi_\rho) \leq H(\varphi_* \rho_1) - H((\varphi_{1*} \otimes \mathrm{id}_{\mathfrak{A}_2 *} \otimes \mathrm{id}_{\mathfrak{R}*})\Psi_\rho) + H(\rho_2).$$

Defining operations

$$\begin{aligned} E_{1*} &= (U_1 \cdot \sqcup \cdot U_1^*) \otimes \mathrm{id}_{\mathfrak{E}_2 *} \otimes \mathrm{id}_{\mathfrak{A}_2 *} \otimes \mathrm{id}_{\mathfrak{R}*} \\ E_{2*} &= \mathrm{id}_{\mathfrak{E}_1 *} \otimes \mathrm{id}_{\mathfrak{A}_1 *} \otimes (U_2 \cdot \sqcup \cdot U_2^*) \otimes \mathrm{id}_{\mathfrak{R}*} \end{aligned}$$

on $\mathfrak{E}_1 \otimes \mathfrak{A}_1 \otimes \mathfrak{E}_2 \otimes \mathfrak{A}_2 \otimes \mathfrak{R}$, and the state $\sigma = \tau_1 \otimes \tau_2 \otimes \Psi_\rho$ we can write this as

$$H_{E_{1*}E_{2*}\sigma}(\mathfrak{A}_1\mathfrak{A}_2) + H_{E_{1*}\sigma}(\mathfrak{A}_1\mathfrak{A}_2\mathfrak{R}) \leq H_{E_{1*}E_{2*}\sigma}(\mathfrak{A}_1\mathfrak{A}_2\mathfrak{R}) + H_{E_{1*}\sigma}(\mathfrak{A}_1) + H_\sigma(\mathfrak{A}_2).$$

Notice that all the states here are pure! Thus by theorem A.12

$$\begin{aligned} H_{E_{1*}E_{2*}\sigma}(\mathfrak{A}_1\mathfrak{A}_2) &= H_{E_{1*}E_{2*}\sigma}(\mathfrak{E}_1\mathfrak{E}_2\mathfrak{R}) \\ H_{E_{1*}E_{2*}\sigma}(\mathfrak{A}_1\mathfrak{A}_2\mathfrak{R}) &= H_{E_{1*}E_{2*}\sigma}(\mathfrak{E}_1\mathfrak{E}_2) \\ H_{E_{1*}\sigma}(\mathfrak{A}_1\mathfrak{A}_2\mathfrak{R}) &= H_{E_{1*}\sigma}(\mathfrak{E}_1\mathfrak{E}_2) \\ &= H_{E_{1*}\sigma}(\mathfrak{E}_1) = H_{E_{1*}E_{2*}\sigma}(\mathfrak{E}_1) \end{aligned}$$

(the last step since $E_{1*}\sigma|_{\mathfrak{E}_2}$ is pure and E_{2*} acts trivially on $\mathfrak{E}_{1*}$), and our inequality transforms to

$$H_{E_{1*}E_{2*}\sigma}(\mathfrak{E}_1\mathfrak{E}_2\mathfrak{R}) + H_{E_{1*}E_{2*}\sigma}(\mathfrak{E}_1) \leq H_{E_{1*}E_{2*}\sigma}(\mathfrak{E}_1\mathfrak{E}_2) + H_{E_{1*}\sigma}(\mathfrak{A}_1) + H_{E_{1*}\sigma}(\mathfrak{A}_2).$$

Here with strong subadditivity of entropy (theorem A.9) the left hand side can be estimated by

$$H_{E_{1*}E_{2*}\sigma}(\mathfrak{E}_1\mathfrak{E}_2) + H_{E_{1*}E_{2*}\sigma}(\mathfrak{E}_1\mathfrak{R})$$

and we are done if we can prove that

$$H_{E_{1*}E_{2*}\sigma}(\mathfrak{E}_1\mathfrak{R}) \leq H_{E_{1*}\sigma}(\mathfrak{A}_1) + H_{E_{1*}\sigma}(\mathfrak{A}_2).$$

But $E_{1*}\sigma|_{\mathfrak{A}_1\mathfrak{A}_2\mathfrak{E}_1\mathfrak{R}}$ is pure, so again by theorem A.12

$$H_{E_{1*}\sigma}(\mathfrak{A}_1) = H_{E_{1*}\sigma}(\mathfrak{A}_2\mathfrak{E}_1\mathfrak{R}).$$

And since E_{2*} acts trivially on $\mathfrak{E}_{1*}\mathfrak{R}_*$ we have

$$H_{E_{1*}E_{2*}\sigma}(\mathfrak{E}_1\mathfrak{R}) = H_{E_{1*}\sigma}(\mathfrak{E}_1\mathfrak{R})$$

which renders our last inequality equivalent to

$$H_{E_{1*}\sigma}(\mathfrak{E}_1\mathfrak{R}) - H_{E_{1*}\sigma}(\mathfrak{A}_2) \leq H_{E_{1*}\sigma}(\mathfrak{A}_2\mathfrak{E}_1\mathfrak{R}),$$

and this is the triangle inequality, theorem A.13. □

Remark IV.6 *Subadditivity*

$$I_e(\rho; \varphi_{1*} \otimes \varphi_{2*}) \leq I_e(\rho_1; \varphi_{1*}) + I_e(\rho_2; \varphi_{2*})$$

which is by $I_e(\rho_2; \varphi_{2*}) \leq H(\rho_2)$ stronger than our lemma, and which one would expect of an information, actually fails*: see* Barnum et al. *(1998).*

Theorem IV.7 *Let $(\mathfrak{A}_1, \dots, \mathfrak{A}_s, \mathsf{P}, P)$ a multiple quantum source with $\mathfrak{A}_i = \mathfrak{L}(\mathcal{H}_i)$ and $(R_1, \dots, R_s)$ a (quantum, F_e)–achievable tuple. Then*

$$\forall I \subset [s] \qquad \sum_{i \in I} R_i \geq H(\mathfrak{A}(I)|\mathfrak{A}(I^c)) = H(P\mathsf{P}) - H(P\mathsf{P}|_{I^c}).$$

Proof. Let an $(n, \lambda)_{F_e}$–coding scheme $(\varepsilon_{1*}, \dots, \varepsilon_{s*}, \delta_*)$ with rate tuple $(R_1, \dots, R_s)$ be given. Denote $d = \sum_{i=1}^s \dim \mathcal{H}_i$.

We may think of ε_{i*} as acting on $\mathfrak{A}_{i*}^{\otimes n}$ by embedding the coding space $\mathfrak{L}(\mathcal{K}_i)_*$. Thus we can apply for $I \subset [s]$ lemma IV.4 to $\varphi_* = \varphi_{1*} \otimes \varphi_{2*}$ (with $\varphi_{1*} = \bigotimes_{i \in I} \varepsilon_{i*}$, $\varphi_{2*} = \bigotimes_{i \in I^c} \varepsilon_{i*}$) and $\psi_* = \delta_*$, and obtain

$$\begin{aligned} nH(P\mathsf{P}) &\leq I_e((P\mathsf{P})^{\otimes n}; \varphi_{1*} \otimes \varphi_{2*}) + 2 + 4n\lambda \log d \\ &\leq I_e((P\mathsf{P}|_I)^{\otimes n}; \varphi_{1*}) + nH(P\mathsf{P}|_{I^c}) + 2 + 4n\lambda \log d \end{aligned}$$

(using weak subadditivity of the coherent information). Since trivially

$$I_e((P\mathsf{P}|_I)^{\otimes n}; \varphi_{1*}) \leq n \sum_{i \in I} R_i$$

we get the theorem in the limit of $n \to \infty$ and $\lambda \to 0$. □

The following example shows that our nice theorem IV.1 is too weak, at least for non-classically correlated sources. At the same time it shows that also theorem IV.7 is too weak.

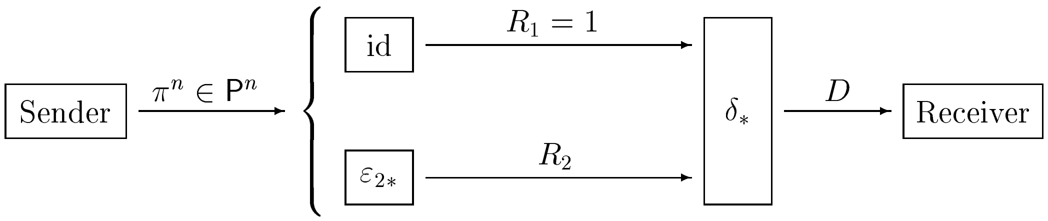
Example IV.8 (EPR source) Consider the source $(\mathfrak{A}_1, \mathfrak{A}_2, \mathsf{P}, P)$ with $\mathfrak{A}_1 = \mathfrak{A}_2 = \mathfrak{L}(\mathbb{C}^2)$, $\mathsf{P} = \{|\Phi^+\rangle\langle\Phi^+|,\ |\Phi^-\rangle\langle\Phi^-|\}$ (two of the Bell states) and P equidistributed on P. Clearly

$$P\mathsf{P} = \frac{1}{2}|\Phi^+\rangle\langle\Phi^+| + \frac{1}{2}|\Phi^-\rangle\langle\Phi^-| = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11|$$

and both marginals equal $\frac{1}{2}\mathbb{1}$. Theorems IV.1 and IV.7 both give only the lower bound $R_1 + R_2 \geq 1$. But we will prove that in fact

$$\mathbf{R}_{\mathrm{q},F_e} \subset \mathbf{R}_{\mathrm{q},\bar{F}} \subset \{(R_1, R_2) :\ R_1, R_2 \geq 1/2\}.$$

To see this let an $(n, \lambda)_{\bar{F}}$–coding scheme $(\varepsilon_{1*}, \varepsilon_{2*}, \delta_*)$ be given with rate pair $(1, R_2)$, the first encoder being the identity. Now imagine that two people want to use this scheme to transmit information: the sender encodes 0–1–sequences as sequences of $|\Phi^+\rangle\langle\Phi^+|$ and $|\Phi^-\rangle\langle\Phi^-|$, giving the according shares of these entangles states to the two encoders. The receiver measures the decoded states in (the tensor power of) the basis $\{|\Phi^+\rangle, |\Phi^-\rangle\}$, call the corresponding observable D. The transmission rate of this system clearly is 1, with average error probability bounded by λ:



Allowing that the sender cooperates with the encoder ε_{2*}, and the receiver with the decoder δ_*, can only increase the transmission rate. We may describe the new situation in a different, equivalent way: the two encoders get the n^{th} power of the maximally entangled state $|\Phi^+\rangle\langle\Phi^+|$, while the second encoder, before performing his ε_{2*}, does the *message encoding (!)* for the sender. This is done with the help of the *phase flip operator*

$$\beta : \begin{cases} |0\rangle & \longmapsto \ \ |0\rangle \\ |1\rangle & \longmapsto -|1\rangle \end{cases}$$

on $\mathbb{C}^2$, as it is readily checked that $(\mathrm{id}\otimes\beta)|\Phi^+\rangle = |\Phi^-\rangle$. But here the first encoder becomes superfluous: thus we can assume that initially sender and receiver share n maximally entangled pairs $|\Phi^+\rangle\langle\Phi^+|$, and the second encoder (viz., the sender!) transmits nR_2 qubits to the receiver. This is exactly the situation of *superdense coding*, invented by Bennett & Wiesner (1992): and it is well known that the maximal transmission rate in this situation is $2R_2$, forcing $R_2 \geq 1/2$ in the limit of $n \to \infty$, $\lambda \to 0$. Of course symmetrically for R_1.

We can note the two lessons we learned:

$\boxed{\text{Theorem IV.1 is too weak.}}$

$\boxed{\text{Theorem IV.7 is too weak.}}$

The last example shows the difference between average and entanglement fidelity:

Example IV.9 (Cloned cross) Consider the source $(\mathfrak{A}_1, \mathfrak{A}_2, \mathsf{P}, P)$ with $\mathfrak{A}_1 = \mathfrak{A}_2 = \mathfrak{L}(\mathbb{C}^2)$ and P equidistributed on $\mathsf{P} = \{|00\rangle\langle 00|,\ |11\rangle\langle 11|\}$. Clearly

$$P\mathsf{P} = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11|$$

with both marginals equal to $\frac{1}{2}\mathbb{1}$. A natural purification of this source would be by the GHZ–*state* $\frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$, invented by Greenberger et al. (1990) to extend Bell's theorem to multi–party entanglement.

Since the average state is the same as in the EPR source we have

$$\mathbf{R}_{\mathrm{q},F_e} \subset \{(R_1, R_2) :\ R_1, R_2 \geq 1/2\}.$$

On the other hand it is obvious that

$$\mathbf{R}_{\mathrm{q},\bar{F}} = \{(R_1, R_2) :\ R_1, R_2 \geq 0,\ R_1 + R_2 \geq 1\}.$$

It is clear from theorem IV.1 that $R_1 + R_2 \geq 1$ is necessary (even with arbitrary encoding). And also one sees easily that $R_1 = 1,\ R_2 = 0$ is (quantum, $\bar{F}$)–achievable: $\mathfrak{A}_2$ sends nothing, whereas $\mathfrak{A}_1$ transmits his qubit faithfully, the decoder has just to copy it to obtain the initial joint state (this is only possible because the two alternative states sent by $\mathfrak{A}_1$ are orthogonal!).

Again collecting our lessons:

$\boxed{\mathbf{R}_{\mathrm{q},\bar{F}} \text{ depends not just on } P\mathsf{P}.}$

$\boxed{\text{In general } \mathbf{R}_{\mathrm{q},\bar{F}} \text{ and } \mathbf{R}_{\mathrm{q},F_e} \text{ are different.}}$

Concluding this section we may state that the pleasing situation of chapter I has completely dissolved: all three rate concepts differ, and (except for entanglement fidelity) the rate region depends not only on the average state.

Classical source with quantum side information

In this and the following section we will turn to the study of a restricted kind of multiple source, namely $c^s h^1$–sources, and we will be able to complement the above bewildering picture by some positive results (coding theorems).

Theorem IV.10 (Code partition) *(Cf.* CSISZÁR & KÖRNER *(1981), proof of theorem 3.1.2) Let* $W : \mathcal{X} \to \mathfrak{S}(\mathfrak{Y})$ *a q–DMC,* P *a probability distribution on* $\mathcal{X}$*,* $\lambda, \delta, \eta > 0$*. Then for* $n \geq n_0(|\mathcal{X}|, \dim\mathcal{H}, \lambda, \delta, \eta)$ *there exist* $m - 1 \leq \exp(n(H(P) - I(P;W) + 3\delta))$ *many* (n, λ)*–codes with pairwise disjoint "large" codebooks* $\mathcal{C}_i$*:*

$$|\mathcal{C}_i| \geq \exp(n(I(P;W) - 2\delta)),$$

such that $P^n(\mathcal{X}^n \setminus \bigcup_{i=1}^{m-1} \mathcal{C}_i) < \eta$*.*

Proof. Choose $\alpha > 0$ such that $P^n(\mathcal{T}^n_{V,P,\alpha}) \geq 1 - \eta/2$ and n large enough such that for every $\mathcal{A} \subset \mathcal{T}^n_{V,P,\alpha}$ with $P^n(\mathcal{A}) \geq \eta/2$ there is a (n, λ)–code with codebook $\mathcal{C} \subset \mathcal{A}$ and $|\mathcal{C}| \geq \exp(n(I(P;W) - 2\delta))$ (by the coding theorem II.4). Now choose such a codebook $\mathcal{C}_1 \subset \mathcal{A}_1 = \mathcal{T}^n_{V,P,\alpha}$ and inductively $\mathcal{C}_i \subset \mathcal{A}_i = \mathcal{A}_{i-1} \setminus \mathcal{C}_{i-1}$ until $P^n(\mathcal{A}_i) < \eta/2$, say for $i = m$. Obviously the codebooks are disjoint, and the rest has weight less than η. It remains to estimate m:

$$(m-1) \cdot \exp(n(I(P;W) - 2\delta)) \leq \sum_{i=1}^{m-1} |\mathcal{C}_i| \leq \left|\mathcal{T}^n_{V,P,\alpha}\right| ,$$

and since by lemma I.8 $|\mathcal{T}^n_{V,P,\alpha}| \leq \exp(n(H(P) + \delta))$ for large enough n we get the statement. □

Consider the problem to encode the classical part of a c^1h^1–source $(\mathfrak{X} = \mathbb{C}\mathcal{X}, \mathfrak{Y}, \mathcal{X} \times \mathsf{P}, P)$, using the quantum source as side information at the decoder:

An *n–block coding scheme with quantum side information at the decoder* is a pair (f, D), with a mapping $f : \mathcal{X}^n \longrightarrow \mathcal{M}$ and an observable D on $\mathbb{C}\mathcal{M} \otimes \mathfrak{Y}$ indexed by $\mathcal{X}$. Its error probability (averaged over P) is

$$\bar{e}(f, D) = 1 - \sum_{x^n \in \mathcal{X}^n, \pi^n \in \mathsf{P}^n} P^n(x^n, \pi^n) \mathrm{Tr}\left((f(x^n) \otimes \pi^n) D_{x^n}\right).$$

The proof of the following theorem goes back to an idea of AHLSWEDE (1974b):

Theorem IV.11 (Rate slicing) *(Cf.* CSISZÁR & KÖRNER *(1981), theorem 3.1.2) For every* $\bar{\lambda}, \delta > 0$ *and* c^1h^1*–source* $(\mathfrak{X} = \mathbb{C}\mathcal{X}, \mathfrak{Y}, \mathcal{X} \times \mathsf{P}, P)$ *there exists an n–block code* (f, D) *with quantum side information at the decoder such that*

$$\frac{1}{n} \log |\mathcal{M}| \leq H(\mathfrak{X}|\mathfrak{Y}) + 3\delta, \ \textit{and}\ \bar{e}(f, D) \leq \bar{\lambda}$$

whenever $n \geq n_0(|\mathcal{X}|, \dim\mathcal{H}, \bar{\epsilon}, \delta)$*. Furthermore, the observable may be modified to the operation* $D'_* = \mathrm{Tr}_{\mathbb{C}\mathcal{M}} \circ D_{tot*}$ *from* $(\mathbb{C}\mathcal{M})_* \otimes \mathfrak{Y}_*$ *to* $(\mathbb{C}\mathcal{X}^n)_* \otimes \mathfrak{Y}_*$ *which satisfies*

$$\sum_{x^n \in \mathcal{X}^n, \pi^n \in \mathsf{P}^n} P^n(x^n, \pi^n) \left\| x^n \otimes \pi^n - D'_*(f(x^n) \otimes \pi^n) \right\|_1 \leq \sqrt{8\bar{\lambda}} + \bar{\lambda}.$$

Proof. Define the q–DMC $W : \mathcal{X} \to \mathfrak{Y}_*$ by

$$W_x = \frac{1}{P_{\mathcal{X}}(x)} \sum_{\pi \in \mathsf{P}} P(x, \pi)\pi$$

(with the marginal distribution $P_{\mathcal{X}}$ of P on $\mathcal{X}$). Choose $\eta \leq \bar{\lambda}$ in theorem IV.10, and n accordingly large such that codes (g_i, D_i), $i \in [m-1]$ like in that theorem exist. Assume that their message sets coincide with their codebooks and that g_i is the identity.

Define now

$$f(x^n) = \begin{cases} i & \text{if } x^n \in \mathcal{C}_i \ , \\ m & \text{else.} \end{cases}$$

The decoder reads $i = f(x^n)$ and if $i \neq m$ uses D_i to recover x^n from the side information: formally, D consists of the operators $[i] \otimes D_{ic}$ for $i \in [m-1]$, $c \in \mathcal{C}'_i$, and $[m] \otimes \mathbb{1}$. That this has the desired properties is easily checked. Now for the second part: observe that

$$D'_* : [j] \otimes \rho \longmapsto \begin{cases} \sum_{c \in \mathcal{C}_j} [c] \otimes \sqrt{D_{jc}}\rho\sqrt{D_{jc}} & \text{if } j < m, \\ [m] \otimes \rho & \text{if } j = m. \end{cases}$$

By the tender measurement lemma I.5 and note I.6 the assertion follows. $\square$

Remark IV.12 *The decoder either says "don't know" (with probability at most $\bar{\lambda}$ over the source distribution P^n), or decodes correctly with maximal error probability $\bar{\lambda}$.*

Corollary IV.13 *For the c^1h^1–source $(\mathfrak{X} = \mathbb{C}\mathcal{X}, \mathfrak{Y}, \mathcal{X} \times \mathsf{P}, P)$ the pair $(H(\mathfrak{X}|\mathfrak{Y}), H(\mathfrak{Y}))$ is $(quantum, \bar{F})$–achievable.*

Proof. Combine theorem IV.11 with Schumacher's quantum coding. $\square$

Consider now the $c^s h^1$–source

$$((\mathfrak{X}_i = \mathbb{C}\mathcal{X}_i | i \in [s]), \mathfrak{Y}, \mathcal{X}_1 \times \cdots \times \mathcal{X}_s \times \mathsf{P}, P).$$

An *n–block coding scheme with quantum side information at the decoder* for this is a $(s+1)$–tuple $(f_1, \dots, f_s, D)$ of mappings $f_i : \mathcal{X}_i^n \to \mathcal{M}_i$ and an observable D on $\mathfrak{X}_1 \otimes \cdots \otimes \mathfrak{X}_s \otimes \mathfrak{Y}$, indexed by $\mathcal{X}_1^n \times \cdots \times \mathcal{X}_s^n$. Its error probability (averaged over P) is

$$\bar{e}(f_1, \dots, f_s, D) = 1 - \sum_{x_i^n \in \mathcal{X}_i^n, \rho \in \mathsf{P}^n} P(x_1^n, \dots, x_s^n, \rho) \mathrm{Tr}\left((f_1(x_1^n) \otimes \cdots \otimes f_s(x_s^n) \otimes \rho) D_{x_1^n \dots x_s^n}\right).$$

Theorem IV.14 *With the notation above and $\bar{\lambda}, \delta > 0$ there exists an n–block coding scheme with quantum side information at the decoder with*

$$\forall J \subset [s] \quad \frac{1}{n} \sum_{j \in J} \log |\mathcal{M}_j| \leq H(\mathfrak{X}(J)|\mathfrak{X}(J^c)\mathfrak{Y}) + |J| \cdot 3\delta$$

and error probability at most $\bar{\lambda}$, whenever $n \geq n_0(|\mathcal{X}_i|, \dim \mathcal{H}, \bar{\lambda}, \delta)$.

Moreover for the operation $D'_* = \mathrm{Tr}_{\mathbb{C}(\mathcal{M}_1 \times \cdots \times \mathcal{M}_s)} \circ D_{tot*}$,

$$D'_* : \mathbb{C}(\mathcal{M}_1 \times \cdots \times \mathcal{M}_s)_* \otimes \mathfrak{Y}_*^{\otimes n} \to \mathbb{C}(\mathcal{X}_1^n \times \cdots \times \mathcal{X}_s^n)_* \otimes \mathfrak{Y}_*^{\otimes n},$$

it holds that

$$\sum_{x^n{}_i \in \mathcal{X}_i^n, \rho \in \mathsf{P}^n} P(x_1^n, \ldots, x_s^n, \rho) \| [x_1^n \ldots x_s^n] \otimes \rho - D'_*([f_1(x_1^n) \ldots f_s(x_s^n)] \otimes \rho) \|_1 \leq \bar{\lambda}.$$

Proof. Only the second statement is to be proved. We use induction on s, the number of sources: $s = 1$ is clear by direct application of the rate slicing theorem IV.11. For $s > 1$ it is sufficient (by the time sharing principle) to consider only extreme points of the region: thus w.l.o.g.

$$\begin{aligned}
\frac{1}{n} \log |\mathcal{M}_1| &\leq H(\mathfrak{X}_1 | \mathfrak{Y}) + 3\delta \\
\frac{1}{n} \log |\mathcal{M}_2| &\leq H(\mathfrak{X}_2 | \mathfrak{X}_1 \mathfrak{Y}) + 3\delta \\
&\ldots \\
\frac{1}{n} \log |\mathcal{M}_s| &\leq H(\mathfrak{X}_s | \mathfrak{X}_1 \cdots \mathfrak{X}_{s-1} \mathfrak{Y}) + 3\delta.
\end{aligned}$$

The proof that these are indeed the extreme points is in the section *Extreme points of rate regions* below.

Now by induction we have an $(n, \bar{\lambda}/2)$–coding scheme for the source

$$((\mathfrak{X}_i = \mathbb{C}\mathcal{X}_i | i \in [s-1]), \mathfrak{X}_s \otimes \mathfrak{Y}, \mathcal{X}_1 \times \cdots \times \mathcal{X}_{s-1} \times (\mathcal{X}_s \times \mathsf{P}), P),$$

call its decoding operation D'_{1*}. By rate slicing we also have an $(n, \bar{\lambda}/2)$–coding scheme for the source $(\mathfrak{X}_s, \mathfrak{Y}, \mathcal{X}_s \times \mathsf{P}, P)$ with side information at the decoder, call its decoding operation D'_{2*}. Then the concatenation $D'_* = D'_{1*} \circ (\mathrm{id} \otimes D'_{2*})$ of the two processes obviously has the desired error properties, and it is readily checked that it has the stated form. By tracing out $\mathfrak{Y}^{\otimes n}$ we recover the observable D. □

Remark IV.15 *The theorem shows that not only we can use quantum side information "just like" classical information to improve compression but also that we can do so with almost not disturbing the quantum information.*

Corollary IV.16 *For the above source all tuples* $(R_1, \ldots, R_s, H(\mathfrak{Y}))$ *satisfying*

$$\forall J \subset [s] \quad \sum_{j \in J} R_j \geq H(\mathfrak{X}(J) | \mathfrak{X}(J^c) \mathfrak{Y})$$

are (quantum, $\bar{F}$*)–achievable.*

Proof. Combine theorem IV.14 with SCHUMACHER's quantum coding. □

We close this section with a converse to these coding theorems:

Theorem IV.17 *Still with the above source all* $(quantum, \bar{F})$*–achievable rate tuples of the form* $(R_1, \dots, R_s, H(\mathfrak{Y}))$ *satisfy*

$$\forall J \subset [s] \quad \sum_{j\in J} R_j \geq H(\mathfrak{X}(J)|\mathfrak{X}(J^c)\mathfrak{Y}).$$

Proof. Otherwise we could by theorem III.5 construct an infinite sequence of transmission n–block codes for the quantum multiple access channel

$$\begin{aligned} W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_s &\longrightarrow \mathfrak{S}(\mathfrak{Y}) \\ (x_1, \dots, x_s) &\longmapsto \frac{1}{P_{\mathcal{X}_1\times\cdots\times\mathcal{X}_s}(x_1 \dots x_s)} \sum_{\pi\in\mathsf{P}} P(x_1 \dots x_s, \pi) \end{aligned}$$

which violate the outer bounds of theorem III.1. □

Quantum source with classical side information

The simplest instance of the problem considered in the previous section is the case of the c^1q^1–source. There we solved the problem of compressing the classical source with the quantum information as side information at the decoder, which gave us one extreme point of the rate region of the multiple source coding problem. It is natural, therefore, to consider the complementary problem of compressing the quantum source, using the classical information as side information, preferably only at the decoder: this would give us another extreme point, presumably completing the determination of the rate region of the c^1q^1–source (if the bounds of theorem IV.1 are already the correct ones).

An *n–block quantum source coding scheme with side information at the decoder* for the c^1q^1–source $(\mathfrak{X} = \mathbb{C}\mathcal{X}, \mathfrak{Y} = \mathfrak{L}(\mathcal{H}), \mathcal{X} \times \mathsf{P}, P)$ is a pair $(\varepsilon_*, \delta_*)$ with a mapping

$$\varepsilon_* : \mathsf{P}^n \longrightarrow \mathfrak{S}(\mathfrak{L}(\mathcal{K}))$$

and a family of quantum operations

$$\delta_* : \mathcal{X}^n \times \mathfrak{L}(\mathcal{K})_* \longrightarrow \mathfrak{Y}_*^{\otimes n}.$$

Quantum and arbitrary encoding are as before, also rate, and the average fidelity is

$$\bar{F} = \bar{F}(\varepsilon_*, \delta_*) = \sum_{(x^n,\pi^n)\in\mathcal{X}^n\times\mathsf{P}^n} P^n(x^n, \pi^n)\cdot \mathrm{Tr}\left((\delta_*(x^n)\varepsilon_*\pi^n)\pi^n\right)$$

(average distortion $\bar{D}(\varepsilon_*, \delta_*)$ similarly).

The limiting rates $R_{q,\bar{F}}(\lambda)$ and $R_{a,\bar{F}}(\lambda)$ are defined obviously. What can we say about them? From theorem IV.1 we get at least

$$\liminf_{\lambda\to 0} R_{q,\bar{F}}(\lambda) \geq H(\mathfrak{Y}|\mathfrak{X}) = \sum_{x\in\mathcal{X}} P_{\mathcal{X}}(x) H\left(\sum_{\pi\in\mathsf{P}} \frac{P(x,\pi)}{P_{\mathcal{X}}(x)}\pi\right).$$

In fact even $R_{a,\bar{F}}(\lambda) \geq H(\mathfrak{Y}|\mathfrak{X})$ for $\lambda \in (0,1)$: otherwise we could (with compressing $\mathfrak{X}$ classically, e.g. by ignoring all non–typical sequences) compress the total source $\mathfrak{X}\mathfrak{Y}$ with asymptotically at most $R_{a,\bar{F}}(\lambda) + H(\mathfrak{X}) < H(\mathfrak{Y}|\mathfrak{X}) + H(\mathfrak{X}) = H(\mathfrak{X}\mathfrak{Y})$ qubits per symbol, contradicting theorem I.22.

At present we do not know if one can approach this bound. But let us make an experiment! Assume that also the encoder has the side information, i.e. now

$$\varepsilon_* : \mathcal{X}^n \times \mathsf{P}^n \longrightarrow \mathfrak{S}(\mathfrak{Y}).$$

Since we are interested only in average performance it suffices that the scheme works well for typical $x^n \in \mathcal{X}^n$, say $x^n \in \mathcal{T}^n_{V,P_\mathcal{X},\alpha}$. To encode this the encoder has just to collect the positions of equal $x \in \mathcal{X}$ and do SCHUMACHER quantum coding on blocks of length $nP_\mathcal{X}(x) \pm \alpha\sqrt{P_\mathcal{X}(x)(1-P_\mathcal{X}(x))}\sqrt{n}$. This scheme — with side information both at the encoder and the decoder — obviously achieves the rate $H(\mathfrak{Y}|\mathfrak{X})$ asymptotically with arbitary high fidelity.

The c^0q^2–source: coding vs. side information

With the c^1q^1–source the idea to consider extreme points in a certain convex region proved useful, and in connection with this the idea to encode only part of the source while using the rest as side information at the decoder.

Whereas this paradigm is of undoubted worth in the classical theory, where we took it from (and which gave us some insights already for quantum communication problems, not just in the two previous sections but also in chapter III), in general one must be cautious with it: *using* quantum information often means using it *up*. As an illustration consider once more the cloned wheel example IV.2:

Obviously we can encode $\mathfrak{A}_1$ with rate zero, with side information from $\mathfrak{A}_2$ at the decoder, because the state π on $\mathfrak{A}_2$ is a faithful copy of the lost state π on $\mathfrak{A}_1$. This is of course in contrast to theorem IV.1, and we can note our last lesson:

Coding independent sources is not reducible to coding with side information.

Extreme points of rate regions

Here we prove the claim in the proof of theorem IV.14 that every extremal point of the region of all $(R_1, \dots, R_s)$ which satisfy for all $J \subset [s]$

$$R(J) = \sum_{i \in J} R_i \geq H(\mathfrak{X}(J)|\mathfrak{Y}\mathfrak{X}(J^c)) \qquad (J)$$

is of the form

$$R_{\pi(i)} = H(\mathfrak{X}_{\pi(i)}|\mathfrak{Y}\mathfrak{X}_{\pi(1)} \cdots \mathfrak{X}_{\pi(i-1)})$$

for a permutation π of the set $[s]$, and that these points all belong to the above region.

Assume that we have an extremal point: it follows that s of the inequalities (J) are met with equality. Choose one, say K:

$$R(K) = H(\mathfrak{X}(K)|\mathfrak{Y}\mathfrak{X}(K^c)).$$

We claim that we can find the remaining inequalities (J) met with equality among the $J \subset K$ or $J \supset K$. This follows from the following

Lemma IV.18 *From $R(K) = H(\mathfrak{X}(K)|\mathfrak{Y}\mathfrak{X}(K^c))$ the validity of (J) for all J follows from the validity for those which contain K or are contained in K.*

Proof. First consider $J \supset K$: there we have

$$R(J \setminus K) \geq H(\mathfrak{X}(J)|\mathfrak{Y}\mathfrak{X}(J^c)) - H(\mathfrak{X}(K)|\mathfrak{Y}\mathfrak{X}(K^c)).$$

Thus for arbitrary J, setting $J_1 = J \cap K$, $J_2 = J \cap K^c$, one obtains

$$\begin{aligned} R(J) &\geq H(\mathfrak{X}(J_1)|\mathfrak{Y}\mathfrak{X}(J_1^c)) + H(\mathfrak{X}(J_2 \cup K)|\mathfrak{Y}\mathfrak{X}(J_2^c \cap K^c)) - H(\mathfrak{X}(K)|\mathfrak{Y}\mathfrak{X}(K^c)) \\ &= H(\mathfrak{Y}\mathfrak{X}_1 \cdots \mathfrak{X}_s) - H(\mathfrak{Y}\mathfrak{X}(J_1^c)) - H(\mathfrak{Y}\mathfrak{X}(J_2^c \cap K^c)) + H(\mathfrak{Y}\mathfrak{X}(K^c)) \\ &\geq H(\mathfrak{Y}\mathfrak{X}_1 \cdots \mathfrak{X}_s) - H(\mathfrak{X}(J_1^c \cap J_2^c)) \\ &= H(\mathfrak{X}(J)|\mathfrak{Y}\mathfrak{X}(J^c)) \end{aligned}$$

by strong subadditivity (theorem A.9), applied to $\mathfrak{A}_1 = \mathfrak{X}(J_2)$, $\mathfrak{A}_2 = \mathfrak{Y}\mathfrak{X}(K^c \setminus J_2)$ and $\mathfrak{A}_3 = \mathfrak{X}(K \setminus J_1)$. □

If K is not a singleton there must be equalities below K, if $K \neq [s]$ there must be some above: otherwise it is easily seen that we are not in an extremal point. So by induction we arrive at a chain $\emptyset \neq K_1 \subset K_2 \subset \ldots \subset K_s = [s]$ of equalities, w.l.o.g. $K_i = \{s, s-1, \ldots, s+1-i\}$, which produces

$$R_i = H(\mathfrak{X}_i|\mathfrak{Y}\mathfrak{X}_1 \cdots \mathfrak{X}_{i-1}).$$

To see that this is indeed a point of the region apply again the lemma, iteratively.

Open questions

The reader will have noticed that the past chapter consisted mainly of open questions, skilfully disguised as half theorems, examples and suggestions. For convenience we collect here some of the more important problems:

Examples Clarify example IV.2: are there coding schemes with $R_1 = 1, R_2 = 1/2$?

Clarify the examples IV.8 and IV.9: can one improve the bounds? or can one actually construct coding schemes with $R_1 = R_2 = 1/2$, or at least $R_1 = 1, R_2 = 1/2$ for one or both of them?

The c^1q^1–source Solve the c^1q^1–source completely! From the above it is enough to consider quantum source coding with classical side information, since we conjecture that theorem IV.1 gives (at least in this case) already the right bounds.

More complicated sources Solve the c^0q^2–source: it seems that it is easier if we insist on no entanglement, but this might be a deception.

Consider entanglement fidelity This seems to be the only right choice if dealing with arbitrary kinds of correlation. Also it simplifies things a bit: namely at least the result will depend only on the average state of the source.

Techniques The only technique for code construction was the "code partition" trick. This is not satisfactory, as it destroys artificially the symmetry of the situation; also we have to resort to a channel coding theorem.

A promising direct approach that may be converted to work for the quantum problems is the *hypergraph coloring paradigm* (see AHLSWEDE (1979) and AHLSWEDE (1980)). Such a program would involve to elaborate further on techniques describable maybe by the term *noncommutative combinatorics*.

Guiding ideas? One of the initial motivations of the work in this chapter was the idea that the classical SLEPIAN–WOLF theorem is one possibility to give operational meaning to conditional entropies. As such a thing is completely lacking in quantum information theory, and on the other hand only formal definitions of quantum conditional entropy exist (derived from analogies, say with classical quantities), without consistent operational meaning, one sees that solving the above coding theorems would clarify this point dramatically.

It is interesting to note that already at this stage we can foresee (from the lessons we learned from our examples) that there must be necessarily *several* natural notions of conditional entropy.

Also we observe that the theory around SCHUMACHER's coherent information fails to give the right answers even to simple problems. I suspect that this comes from the fact that this theory builds on *pair entanglement*, whereas our situations involved *multi–party entanglement*.

Appendix A

Quantum Probability and Information

In this appendix the basic mathematical machinery of quantum probability with special attention to information theory is collected. Alongside we introduce a calculus of entropy and information quantities in quantum systems. Whenever possible we refer to the literature instead of giving full proofs.

Quantum systems

In classical probability theory one has generally two ways of seeing things: either through distributions (and the relation of their images, mostly marginals), or through random variables (with a joint distribution). Both ways have their merits (though random variables are considered more elegant), but basically they are equivalent, in particular none lacks anything without the other. Things are different in quantum probability, and we will take the following view: the analog of a distribution is a density operator on some complex Hilbert space, whereas the analog of random variables are *observables*, defined below. With density operators alone we can study physical processes transforming them, but every experiment involves some observable. Studying observables one usually fixes the underlying density operator (as the statistics of the experiments depend on the latter), but this falls short of not appropriately reflecting our manipulating quantum states, or having several alternative states.

For the following we refer to textbooks on C*–algebras like Arveson (1976), and standard references on basic mathematics of quantum mechanics: Davies (1976), Kraus (1983), and the more advanced Holevo (1982).

A C*–*algebra* with unit is a complex Banach space $\mathfrak{A}$ which is also a $\mathbb{C}$–algebra with unit $\mathbb{1}$ and a $\mathbb{C}$–antilinear involution $*$, such that

$$\|AB\| \leq \|A\|\|B\|, \qquad \|A^*\|^2 = \|A\|^2 = \|AA^*\|$$

These algebras will be the mathematical models for quantum systems, and subsystems are simply $*$–subalgebras (which are always assumed to be closed).

The set $\mathfrak{A}^+$ of $A \in \mathfrak{A}$ that can be written as $A = BB^*$ is called the *positive cone* of $\mathfrak{A}$ which is norm closed, and induces a partial order $\leq$. By the famous GELFAND–NAIMARK–SEGAL representation theorem (see e.g. ARVESON (1976)) every C*–algebra is isomorphic to a closed $*$–subalgebra of some $\mathfrak{L}(\mathcal{H})$, the algebra of bounded linear operators on the Hilbert space $\mathcal{H}$. With us all C*–algebras will be of finite dimension. It is known that those algebras are isomorphic to a direct sum of $\mathfrak{L}(\mathcal{H}_i)$ (see e.g. ARVESON (1976)). This includes as extremal cases the algebras $\mathfrak{L}(\mathcal{H})$, and the commutative algebras $\mathbb{C}\mathcal{X}$ over a finite set $\mathcal{X}$, with the generators $x \in \mathcal{X}$ as idempotents. In particular we have on every such algebra a well defined and unique *trace functional*, denoted Tr, that assigns trace one to all minimal positive idempotents.

States A *state* on a C*–algebra $\mathfrak{A}$ is a positive $\mathbb{C}$–linear functional ρ with $\rho(\mathbb{1}) = 1$. Positivity here means that its values on the positive cone are nonnegative. Clearly the states form a convex set $\mathfrak{S}(\mathfrak{A})$ whose extreme points are called *pure* states, all others are *mixed*. For $\mathfrak{A} = \mathfrak{L}(\mathcal{H})$ the pure states are exactly the one–dimensional projectors, i.e. using DIRAC's *bra–ket*–notation, the $|\psi\rangle\langle\psi|$ with unit vector $|\psi\rangle \in \mathcal{H}$.

One can easily see that every state ρ can be represented uniquely in the form $\rho(X) = \mathrm{Tr}\,(\hat{\rho}X)$ for a positive, selfadjoint element $\hat{\rho}$ of $\mathfrak{A}$ with trace one (such elements are called *density operators*). In the sequel we will therefore make no distinction between ρ and its density operator $\hat{\rho}$. The set of operators with finite trace will be denoted $\mathfrak{A}_*$, the *trace class* in $\mathfrak{A}$ which contains the states and is a two–sided ideal in $\mathfrak{A}$, the SCHATTEN–ideal (in our — finite dimensional — case this is of course just $\mathfrak{A}$). Then $\mathrm{Tr}\,(\rho A)$ defines a real bilinear and positive definite pairing of $\mathfrak{A}_{*s}$ and $\mathfrak{A}_s$, the selfadjoint parts of $\mathfrak{A}_*$ and $\mathfrak{A}$, which makes $\mathfrak{A}_s$ the dual of $\mathfrak{A}_{*s}$. Notice that in this sense pure states are equivalently described as minimal selfadjoint idempotents of $\mathfrak{A}$.

Observables Let $\mathcal{F}$ be a σ–algebra on some set Ω, $\mathfrak{X}$ a C*–algebra. A map $X : \mathcal{F} \longrightarrow \mathfrak{X}$ is called a *positive operator valued measure* (POVM), or an *observable*, with values in $\mathfrak{X}$ (or on $\mathfrak{X}$), if:

1. $X(\emptyset) = 0,\ X(\Omega) = \mathbb{1}$.
2. $E \subset F$ implies $X(E) \leq X(F)$.
3. If $(E_n)_n$ is a countable family of pairwise disjoint sets in $\mathcal{F}$ then $X(\bigcup_n E_n) = \sum_n X(E_n)$ (in general the convergence is to be understood in the weak topology: for every state its value at the left equals the limit value at the right hand side).

If the values of the observable are all projection operators and Ω is the real line one speaks of a *spectral measure* or a VON NEUMANN *observable*.[1] An observable X together with a

[1]Strictly speaking this term only applies to the expectation of the measure (in general an unbounded operator), but this in turn by the spectral theorem determines the measure.

state ρ yields a probability measure P^X on Ω via the formula

$$P^X(E) = \mathrm{Tr}\,(\rho X(E)).$$

In this way we may view X as a random variable with values in $\mathfrak{X}$, its distribution we denote P_X (note that P_X may not be isomorphic to P^X: if X takes the same value on disjoint events, which means that X introduces randomness by itself).

Two observables X, Y are said to be *compatible*, if they have values in the same algebra and $XY = YX$ elementwise, i.e. for all $E \in \mathcal{F}_X$, $F \in \mathcal{F}_Y$: $X(E)Y(F) = Y(F)X(E)$ (Note that it is possible for an observable not to be compatible with itself). By the way, the term *compatible* may be defined in obvious manner for arbitrary sets or collections of operators, in which meaning we will use it in the sequel. If X, Y are compatible we may define their *joint observable* $XY : \mathcal{F}_X \times \mathcal{F}_Y \longrightarrow \mathfrak{X}$ mapping $E \times F$ to $X(E)Y(F)$ (this defines the product mapping uniquely just as in the classical case of product measures). In fact we can analogously define the joint observable for any collection of pairwise compatible observables.[2] As the random variable of a product XY we will take $X \times Y$, rather than XY itself, with values in $\mathfrak{X} \times \mathfrak{X}$ (because the same product operator may be generated in two different ways which we want to distinguish). To indicate this difference we will sometimes write $X \cdot Y$ for the product.

Note that here we can see the reason why we cannot just consider all observables as random variables (and forget about the state): they will not have a joint distribution, at first of course only by our definition. But BELL's theorem (BELL (1964)) shows that one comes into trouble if one tries to allow a joint distribution for noncompatible observables. Conversely we see why we cannot do without observables, even though ρ contains all possible information: the crux is that we cannot access it due to the forbidden noncompatibel observables (a good account of this aspect of quantum theory is by PERES (1995)).

From now on all observables will be *countable*, i.e. w.l.o.g. are they defined on a countable Ω with σ–algebra 2^Ω. This means that we may view an observable X as a resolution of $\mathbb{1}$ into a countable sum $\mathbb{1} = \sum_{j\in\Omega} X_j$ of positive operators X_j.

If $\mathfrak{A}_1, \mathfrak{A}_2$ are subalgebras of $\mathfrak{A}$, they are compatible if they commute elementwise (again note, that a subalgebra need not not be compatible with itself: in fact it is iff it is commutative). In this case the closed subalgebra generated (in fact: spanned) by the products A_1A_2, $A_i \in \mathfrak{A}_i$ is denoted $\mathfrak{A}_1\mathfrak{A}_2$.

Operations Now we describe the transformations between quantum systems: a $\mathbb{C}$–linear map $\varphi : \mathfrak{A}_2 \to \mathfrak{A}_1$ is called a *quantum operation* if it is completely positive (i.e. positive, so that positive elements have positive images, and also the $\varphi \otimes \mathrm{id}_n$ are positive, where id_n is the identity on the algebra of $n \times n$–matrices), and unit preserving. These maps are in 1–1 correspondence with their (pre–)adjoints φ_* by the trace form, mapping

[2] Observe however that in general a joint observable might exist for non–compatible (i.e. non–commuting) observables. The operational meaning of this is that there is a common refinement of the involved observables. If they commute then this certainly is possible as demonstrated, but commutativity is not necessary.

states to states, and being completely positive and trace preserving.[3] Since here we restrict ourselves to finite dimensional algebras the adjoint map simply goes from $\mathfrak{A}_1$ to $\mathfrak{A}_2$, but to keep things well separated (which they actually are in the infinite case) we write the adjoint as $\varphi_* : \mathfrak{A}_{1*} \to \mathfrak{A}_{2*}$, the dual map (in fact we consider this as the primary object and the operator maps as their adjoint, which is the reason for writing subscript $*$). Notice that φ_* is sometimes considered as restricted to $\varphi_* : \mathfrak{S}(\mathfrak{A}_1) \to \mathfrak{S}(\mathfrak{A}_2)$. A characterization of quantum operations is by the Stinespring dilation theorem (Stinespring (1955)):

Theorem A.1 (Dilation) *Let $\varphi : \mathfrak{A} \to \mathfrak{L}(\mathcal{H})$ a linear map of C^*–algebras. Then φ is completely positive if and only if there exist a representation $\alpha : \mathfrak{A} \to \mathfrak{L}(\mathcal{K})$, with Hilbert space $\mathcal{K}$, and a bounded linear map $V : \mathcal{H} \to \mathcal{K}$ such that*

$$\forall A \in \mathfrak{A} \qquad \varphi(A) = V^*\alpha(A)V.$$

For proof see e.g. Davies (1976). A commonly used corollary of this is

Corollary A.2 *(cf.* Kraus *(1983)) Let $\varphi : \mathfrak{L}(\mathcal{H}_2) \to \mathfrak{L}(\mathcal{H}_1)$ a linear map of C^*–algebras. Then φ is completely positive and unit preserving if and only if there exist linear maps $B_i : \mathcal{H}_1 \to \mathcal{H}_2$ with $\sum_i B_i^* B_i = \mathbb{1}_{\mathcal{H}_1}$ and*

$$\forall A \in \mathfrak{L}(\mathcal{H}_2) \qquad \varphi(A) = \sum_i B_i^* A B_i \ .$$

Norms and norm inequalities For the C*–algebra $\mathfrak{L}(\mathcal{H})$ of linear operators on the complex Hilbert space $\mathcal{H}$ (of dimension d) the norm is the supremum norm $\|\cdot\|_\infty$, i.e. $\|A\|_\infty$ is the largest absolute value of an eigenvalue of A. The other important norm we use is the *trace norm* $\|\cdot\|_1$: $\|\alpha\|_1$ is the sum of the absolute values of all eigenvalues of α. Note the important formula

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

which explains the name "trace norm". Its proof is by the polar decomposition of α, see e.g. Arveson (1976). Also it implies immediately that $\|\cdot\|_1$ is nonincreasing under quantum operations. Obviously

$$\|\alpha\|_\infty \leq \|\alpha\|_1 \leq d\|\alpha\|_\infty \ .$$

If α is self–adjoint we have the unique decomposition (via diagonalization) $\alpha = \alpha_+ - \alpha_-$ into the positive and negative part of α, where $\alpha_+, \alpha_- \geq 0$ and $\alpha_+\alpha_- = 0$. Then note

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

and

$$\mathrm{Tr}\,\alpha_+ = \sup\{\mathrm{Tr}\,(\alpha B) : 0 \leq B \leq \mathbb{1}\}.$$

It should be clear that all the above suprema are in fact maxima.

Finally note that these observations still hold for any direct sum of $\mathfrak{L}(\mathcal{H}_i)$, d being replaced by the sum of the $\dim \mathcal{H}_i$.

[3] In general this is only true if we restrict φ to be a *normal* map, cf. Davies (1976).

Entropy and divergence

The VON NEUMANN *entropy* of a state ρ (introduced by VON NEUMANN (1927)[4]) is defined as $H(\rho) = -\mathrm{Tr}\,(\rho \log \rho)$, which reduces to the usual SHANNON entropy for a commutative algebra because then a state is nothing but a probability distribution. For states ρ, σ also introduce the *I–divergence*, or simply *divergence* (first defined by UMEGAKI (1962)) as $D(\rho\|\sigma) = \mathrm{Tr}\,(\rho(\log \rho - \log \sigma))$ with the convention that this is ∞ if $\mathrm{supp}\,\rho \not\leq \mathrm{supp}\,\sigma$ ($\mathrm{supp}\,\rho$ being the *support* of ρ, the minimal selfadjoint idempotent p with $p\rho p = \rho$). For properties of these quantities we will often refer to OHYA & PETZ (1993), and to WEHRL (1978). Three important facts we will use are

Theorem A.3 (KLEIN inequality) *For positive operators ρ, σ (not necessarily states)*

$$D(\rho\|\sigma) \geq \frac{1}{2}\mathrm{Tr}\,(\rho - \sigma)^2 + \mathrm{Tr}\,(\rho - \sigma).$$

In particular for states the divergence is nonnegative, and zero if and only if they are equal.

Proof. See OHYA & PETZ (1993). □

Lemma A.4 (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 OHYA & PETZ (1993). □

Theorem A.5 (Monotonicity) *Let ρ, σ be states on a C^*–algebra $\mathfrak{A}$, and φ_* a trace preserving, completely positive linear map from states on $\mathfrak{A}$ to states on $\mathfrak{B}$. Then*

$$D(\varphi_*\rho\|\varphi_*\sigma) \leq D(\rho\|\sigma).$$

Proof. See UHLMANN (1977); the situation we are in was already solved by LINDBLAD (1975). For a textbook account see OHYA & PETZ (1993). □

Observable language

This and the following two sections will introduce language (or formalism) to talk about entropy and information in the context of quantum systems in a transparent fashion.

Fix a state on a C*–algebra, say ρ on $\mathfrak{A}$ and let X, Y, Z compatible observables on $\mathfrak{A}$. These are then random variables with a joint distribution, and one defines entropy

[4] It was in fact introduced independently in the same year by LANDAU and WEYL.

$H(X)$, conditional entropy $H(X|Y)$, mutual information $I(X \wedge Y)$, and conditional mutual information $I(X \wedge Y|Z)$ for these observables as the respective quantities for them interpreted as random variables. Note however that these depend on the underlying state ρ. In case of need we will thus add the state as an index, like $H_\rho(X) = H(X)$, etc.

As things are there is not much to say about that part of the theory. We only note some useful formulas:

$$H(X|Y) = \sum_j \mathrm{Tr}\,(\rho Y_j) H_{\rho_j}(X), \quad \text{with } \rho_j = \frac{1}{\mathrm{Tr}\,(\rho Y_j)} \sqrt{Y_j}\rho\sqrt{Y_j}$$

(which is an easy calculation using the compatibility of X and Y), and

$$\begin{aligned} I(X \wedge Y) &= H(X) + H(Y) - H(XY) \\ &= D(P^{XY} \| P^X \otimes P^Y) = D(P_{X \cdot Y} \| P_X \otimes P_Y) \end{aligned}$$

(whose analogue is known from classical information theory).

Subalgebra language

Let $\mathfrak{X}, \mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{Y}$ compatible $*$–subalgebras of the C*–algebra $\mathfrak{A}$, and ρ a fixed state on $\mathfrak{A}$.

First consider the inclusion map $\imath : \mathfrak{X} \hookrightarrow \mathfrak{A}$ (which is certainly completely positive) and its adjoint $\imath_* : \mathfrak{A}_* \to \mathfrak{X}_*$. Define

$$H(\mathfrak{X}) = H(\imath_* \rho)$$

(where at the right hand appears the von Neumann entropy). For example for $\mathfrak{X} = \mathfrak{A}$ we obtain just the von Neumann entropy of ρ. For the trivial subalgebra $\mathbb{C} = \mathbb{C}\mathbb{1}$ (which obviously commutes with every subalgebra) we obtain, as expected, $H(\mathbb{C}) = 0$. The general philosophy behind this definition is that $H(\mathfrak{X})$ is the von Neumann entropy of the global state *viewed through (or restricted to) the subsystem* $\mathfrak{X}$. To reflect this in the notation we define $\rho|_{\mathfrak{X}} = \imath_* \rho$.

Now conditional entropy, mutual information, and conditional mutual information are defined by reducing them to entropy quantities:

$$H(\mathfrak{X}|\mathfrak{Y}) = H(\mathfrak{X}\mathfrak{Y}) - H(\mathfrak{Y})$$

$$\begin{aligned} I(\mathfrak{X}_1 \wedge \mathfrak{X}_2) &= H(\mathfrak{X}_1) + H(\mathfrak{X}_2) - H(\mathfrak{X}_1\mathfrak{X}_2) \\ &= H(\mathfrak{X}_2) - H(\mathfrak{X}_2|\mathfrak{X}_1) \end{aligned}$$

$$\begin{aligned} I(\mathfrak{X}_1 \wedge \mathfrak{X}_2|\mathfrak{Y}) &= H(\mathfrak{X}_1|\mathfrak{Y}) + H(\mathfrak{X}_2|\mathfrak{Y}) - H(\mathfrak{X}_1\mathfrak{X}_2|\mathfrak{Y}) \\ &= H(\mathfrak{X}_1\mathfrak{Y}) + H(\mathfrak{X}_2\mathfrak{Y}) - H(\mathfrak{X}_1\mathfrak{X}_2\mathfrak{Y}) - H(\mathfrak{Y}). \end{aligned}$$

It is not at all clear a priori that these definitions are all well behaved: while it is obvious from the definition that the entropy is always nonnegative, this is not true for the conditional entropy (as was observed by several authors before): if $\mathfrak{A} = \mathfrak{X} \otimes \mathfrak{Y}$ and ρ is a pure entangled state then $H(\mathfrak{X}|\mathfrak{Y}) = -H(\mathfrak{Y}) < 0$. This might raise pessimism whether the other two quantities also are (at least sometimes) pathological. This they are not (at least not in this way), as will be shown in a moment:

We have the following commutative diagram of inclusions, and the natural multiplication map μ (which is in fact a $*$–algebra homomorphism, and thus completely positive!):

$$\begin{tikzcd}
\mathfrak{X}_1 \arrow[r, equal] \arrow[d, "\varphi_1"] & \mathfrak{X}_1 \arrow[r, equal] \arrow[d, "\imath_1"] & \mathfrak{X}_1 \arrow[d, "\jmath_1"] \\
\mathfrak{X}_1 \otimes \mathfrak{X}_2 \arrow[r, "\mu"] & \mathfrak{X}_1\mathfrak{X}_2 \arrow[r, "\jmath"] & \mathfrak{A} \\
\mathfrak{X}_2 \arrow[r, equal] \arrow[u, "\varphi_2"] & \mathfrak{X}_2 \arrow[r, equal] \arrow[u, "\imath_2"] & \mathfrak{X}_2 \arrow[u, "\jmath_2"]
\end{tikzcd}$$

And hence the corresponding commutative diagram of adjoint maps (note that φ_{1*} and φ_{2*} are just partial traces). With this we find

$$\begin{aligned}
I(\mathfrak{X}_1 \wedge \mathfrak{X}_2) &= H(\mathfrak{X}_1) + H(\mathfrak{X}_2) - H(\mathfrak{X}_1\mathfrak{X}_2) \\
&= H(\jmath_{1*}\rho) + H(\jmath_{2*}\rho) - H(\jmath_*\rho) \\
&= H(\varphi_{1*}\mu_*\jmath_*\rho) + H(\varphi_{2*}\mu_*\jmath_*\rho) - H(\mu_*\jmath_*\rho) \\
&= D(\mu_*\jmath_*\rho \| \varphi_{1*}\mu_*\jmath_*\rho \otimes \varphi_{2*}\mu_*\jmath_*\rho)
\end{aligned}$$

by definition, then by commutativity of the diagram and the fact that μ_* preserves eigenvalues of density operators (because μ is a surjective $*$–homomorphism, see lemma A.6 below), the last by direct calculation on the tensor product (just as for the classical formula). From the last line we see that the mutual information is nonnegative because the divergence is, by theorem A.3 (we could also have seen this already from the definition by applying subadditivity of von Neumann entropy to the second last line, see theorem A.9).

Lemma A.6 *Let $\mu : \mathfrak{A} \to \mathfrak{B}$ a surjective $*$–algebra homomorphism. Then*

1. *For all pure states $p \in \mathfrak{S}(\mathfrak{A})$: $\mu(p)$ is pure or 0.*

2. *For all $A \in \mathfrak{A}$, $A \geq 0$: $\operatorname{Tr} A \geq \operatorname{Tr} \mu(A)$.*

3. *For pure $p \in \mathfrak{S}(\mathfrak{A})$, $q \in \mathfrak{S}(\mathfrak{B})$:*

$$\mu_*(\mu(p)) = p \text{ or } \mu(p) = 0, \;\; \mu(\mu_*(\mu(p))) = \mu(p), \;\; \mu(\mu_*(q)) = q.$$

4. *For $\rho \in \mathfrak{S}(\mathfrak{B})$, $\mu_*(\rho) = \sum_i \alpha_i p_i$ diagonalization with the $\alpha_i > 0$, then $\rho = \sum_i \alpha_i \mu(p_i)$ is a diagonalization.*

5. *Conversely every diagonalization of a state on $\mathfrak{B}$ is by μ_* translated into a diagonalization of its μ_*–image.*

Proof.

1. We have only to show that $\mu(p)$ is minimal if it is not 0: let q' any pure state with $q' \le \mu(p)$. Then
$$1 = \mathrm{Tr}\,(q'\mu(p)) = \mathrm{Tr}\,(\mu_*(q')p) \le \mathrm{Tr}\,(p) = 1.$$
So we must have equality which implies $p \le \mu_*(q')$, but both operators are states, so $p = \mu_*(q')$. Because μ_* is injective this means that there is only one pure state $q' \le \mu(p)$, i.e. $\mu(p)$ is pure.

2. We may write $A = \sum_i a_i p_i$ with pure states p_i and $a_i \ge 0$. Then $\mu(A) = \sum_i a_i \mu(p_i)$ and since pure states have trace 1 the assertion follows from (1).

3. Let $A \in \mathfrak{A}$, $A \ge 0$. Then
$$\begin{aligned}\mathrm{Tr}\,(\mu_*(\mu(p))A) &= \mathrm{Tr}\,(\mu(p)\mu(A)) = \mathrm{Tr}\,(\mu(p)\mu(A)\mu(p)) \\ &= \mathrm{Tr}\,(\mu(pAp)) \le \mathrm{Tr}\,(pAp) = \mathrm{Tr}\,(pA).\end{aligned}$$
Thus $\mu_*(\mu(p)) \le p$. If $\mu(p) \ne 0$ it is a pure state, hence $\mu_*(\mu(p))$ a state which forces $\mu_*(\mu(p)) = p$. This proves the left formula, the middle follows immediately, and for the right observe that we may choose a pure pre–image p of q (in fact that will be $\mu_*(q)$, as one can see from (4)).

4. $\sum_i \alpha_i \mu(p_i)$ is certainly the diagonalization of some positive operator since the $\mu(p_i)$ which are not 0 are by the homomorphism property and by (1) pairwise orthogonal pure states. Now observe $\mu(\mu_*(\rho)) = \sum_i \alpha_i \mu(p_i)$ and
$$\mu_*(\rho) = \mu_*(\mu(\mu_*(\rho))) = \sum_i \alpha_i \mu_*(\mu(p_i)) \le \sum_i \alpha_i p_i = \mu_*(\rho),$$
hence equality, i.e. all $\mu(p_i)$ are pure. From
$$\mu_*(\rho) = \sum_i \alpha_i \mu_*(\mu(p_i)) = \mu_*(\sum_i \alpha_i \mu(p_i))$$
and injectivity of μ_* the assertion follows.

5. This is a direct consequence of (3) and (4).

□

For the conditional mutual information we have to do somewhat more (yet from the definition we see that its positivity will have something to do with the strong subadditivity of von Neumann entropy, see theorem A.9):

Consider the following commuative diagram:

$$\begin{array}{ccccccc}
\mathfrak{Y} & \xrightarrow{\varphi_1} & \mathfrak{X}_1 \otimes \mathfrak{Y} & \xrightarrow{\mu_1} & \mathfrak{X}_1\mathfrak{Y} & & \\
\| & & \downarrow{\scriptstyle \varphi_1'} & & \downarrow{\scriptstyle \jmath_1} & & \\
\mathfrak{Y} & \xrightarrow{\varphi} & \mathfrak{X}_1 \otimes \mathfrak{X}_2 \otimes \mathfrak{Y} & \xrightarrow{\mu} & \mathfrak{X}_1\mathfrak{X}_2\mathfrak{Y} & \xrightarrow{\jmath} & \mathfrak{A} \\
\| & & \uparrow{\scriptstyle \varphi_2'} & & \uparrow{\scriptstyle \jmath_2} & & \\
\mathfrak{Y} & \xrightarrow{\varphi_2} & \mathfrak{X}_2 \otimes \mathfrak{Y} & \xrightarrow{\mu_2} & \mathfrak{X}_2\mathfrak{Y} & &
\end{array}$$

All maps there are completely positive, μ, μ_1, μ_2 being $*$–homomorphisms. Thus the adjoints of the various φ's are partial traces and with $\sigma = \mu_* \jmath_* \rho$: $H(\mathfrak{X}_1\mathfrak{X}_2\mathfrak{Y}) = H(\sigma)$, $H(\mathfrak{X}_1\mathfrak{Y}) = H(\mathrm{Tr}_{\mathfrak{X}_2}\sigma)$, $H(\mathfrak{X}_2\mathfrak{Y}) = H(\mathrm{Tr}_{\mathfrak{X}_1}\sigma)$, $H(\mathfrak{Y}) = H(\mathrm{Tr}_{\mathfrak{X}_1\otimes\mathfrak{X}_2}\sigma)$ (where we have made use of lemma A.6 several times), and we can indeed apply strong subadditivity.

Finally let us remark the nice formulas

$$H(\mathfrak{X}) = H(\mathfrak{X}|\mathbb{C}), \qquad I(\mathfrak{X}_1 \wedge \mathfrak{X}_2) = I(\mathfrak{X}_1 \wedge \mathfrak{X}_2|\mathbb{C}).$$

Example A.7 A very important special case of the definitions of this and the preceding section occurs for tensor products of Hilbert spaces $\mathfrak{L}(\mathcal{H}_1 \otimes \mathcal{H}_2) = \mathfrak{L}(\mathcal{H}_1) \otimes \mathfrak{L}(\mathcal{H}_2)$, or more generally tensor products of C*–algebras: $\mathfrak{A} = \mathfrak{A}_1 \otimes \mathfrak{A}_2$. $\mathfrak{A}_1, \mathfrak{A}_2$ are $*$–subalgebras of $\mathfrak{A}$ in the natural way, and are obviously compatible. The same then holds for observables $A_i \subset \mathfrak{A}_i$, and similarly for more than two factors. In this case the restriction $\rho|_{\mathfrak{A}_i}$ is just a partial trace.

Remark A.8 *It should be clear that we introduced (having H) conditional entropy and mutual information by* formal analogy *to the classical quantities. We cannot claim to have an operational meaning of them in general — the theorems in the main text must be seen as exceptions to this rule.*

We are in this respect in accordance with Levitin *(1998) who went even further by rejecting the very name "conditional entropy" for $H(\cdot|\cdot)$, and proposed to return to the name "correlation entropy" given by* Stratonovich *to the quantity $I(\cdot \wedge \cdot)$, on the grounds of a strictly operational reasoning (which is only open to the one criticism that* Levitin *always sticks with* classical *information, never acknowledging the unprecedented properties of* quantum *information).*

Common tongue

The languages of the two preceding sections may be phrased in a unified formalism (the "common tongue") using completely positive C*–algebra maps (in particular those from or to commutative algebras, inclusion maps, and $*$–algebra homomorphisms, cf. Stinespring (1955)).

That this is promising one can see from the observation that observables can be interpreted in a natural way as C*–algebra maps: $X : \Omega \to \mathfrak{A}$ corresponds by linear extension to $X : \mathfrak{B}(\Omega) \to \mathfrak{A}$, where $\mathfrak{B}(\Omega) = \mathfrak{B}(\Omega, \mathcal{F})$ is the algebra of bounded measurable functions on Ω. We follow the convention that in this algebra $j \in \Omega$ shall denote the function that is 1 on j and 0 elsewhere, so $X(j) = X_j$, and obviously $X_*(\rho)$ equals the distribution P^X on Ω induced by X with ρ.

Let us also introduce some notation for the observable X: the *total* observable operation $X_{\text{tot}} : \mathfrak{B}(\Omega) \otimes \mathfrak{A} \to \mathfrak{A}$ mapping $j \otimes A \mapsto \sqrt{Y_j} A \sqrt{Y_j}$, its *interior* part $X_{\text{int}} = X_{\text{tot}} \circ \imath_{\mathfrak{A}} : \mathfrak{A} \to \mathfrak{A}$ with $A \mapsto \sum_j \sqrt{Y_j} A \sqrt{Y_j}$, and its *exterior* part $X_{\text{ext}} = X_{\text{tot}} \circ \imath_{\mathfrak{B}(\Omega)}$ which coincides with X.

Consider compatible quantum operations $\varphi : \mathfrak{X} \to \mathfrak{A}$, $\psi : \mathfrak{Y} \to \mathfrak{A}$, etc. ($\varphi, \psi$ are compatible if their images commute elementwise). In this case their *product* is the operation $\varphi\psi : \mathfrak{X} \otimes \mathfrak{Y} \to \mathfrak{A}$ mapping $X \otimes Y \mapsto \varphi(X)\psi(Y)$:

$$
\begin{tikzcd}
\mathfrak{X} \arrow[r, "\varphi"] \arrow[d, "\varphi_1"'] & \mathfrak{A} \arrow[d, equal] \\
\mathfrak{X} \otimes \mathfrak{Y} \arrow[r, "\exists!\varphi\psi"] & \mathfrak{A} \\
\mathfrak{Y} \arrow[r, "\psi"] \arrow[u, "\varphi_2"] & \mathfrak{A} \arrow[u, equal]
\end{tikzcd}
$$

Note that this generalizes the product of observables, as well as the product map μ of subalgebras.

Now simply define $H(\varphi) = H(\varphi_*\rho)$, and again the conditional entropy and the informations are defined by reduction to entropy, e.g. $H(\varphi|\psi) = H(\varphi\psi) - H(\psi)$, or $I(\varphi \wedge \psi) = H(\varphi) + H(\psi) - H(\varphi\psi)$.

For the mutual information observe that (see previous diagram)

$$
\begin{aligned}
I(\varphi \wedge \psi) &= D((\varphi\psi)_*\rho \| \varphi_*\rho \otimes \psi_*\rho) \\
&= D(\sigma \| \mathrm{Tr}_{\mathfrak{Y}}\sigma \otimes \mathrm{Tr}_{\mathfrak{X}}\sigma), \qquad \text{with } \sigma = (\varphi\psi)_*\rho.
\end{aligned}
$$

Note the difference to Ohya & Petz (1993): with them the entropy of an operation is related to the mutual information of the operation as a channel. With us the entropy of an operation is the entropy of a state "viewed through" this operation (as was the idea with the entropy of a subsystem, and obviously also with the entropy of an observable).

With these insights we may now form hybrid expressions involving observables and $*$–subalgebras at the same time: let $\imath : \mathfrak{X} \hookrightarrow \mathfrak{A}$, $\jmath : \mathfrak{Y} \hookrightarrow \mathfrak{A}$ $*$–subalgebra inclusions, and X, Y observables on $\mathfrak{A}$, all four compatible. Then we have

$$H(\mathfrak{X}|Y) = H(\imath Y) - H(Y)$$

$$I(\mathfrak{X} \wedge Y) = H(\imath) + H(Y) - H(\imath Y),$$

and lots of others. From the previous section we know that the information quantities are nonnegative, but also the entropy conditional on an observable, from the formula

$$H(\mathfrak{X}|Y) = \sum_j \mathrm{Tr}\,(\rho Y_j) H_{\rho_j}(\mathfrak{X}), \quad \text{with } \rho_j = \frac{1}{\mathrm{Tr}\,(\rho Y_j)} \sqrt{Y_j} \rho \sqrt{Y_j}\ .$$

But again there are some expressions which seem suspicious, like

$$H(X|\mathfrak{Y}) = H(X\mathfrak{Y}) - H(\mathfrak{Y}).$$

However, due to the inequality of theorem A.20 in fact it behaves nicely.

Inequalities

Entropy Let us first note the basic

Theorem A.9 *For compatible $*$–subalgebras $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3$ one has:*

1. *Subadditivity:* $H(\mathfrak{A}_1\mathfrak{A}_2) \leq H(\mathfrak{A}_1) + H(\mathfrak{A}_2)$.
2. *Strong subadditivity:* $H(\mathfrak{A}_1\mathfrak{A}_2\mathfrak{A}_3) + H(\mathfrak{A}_2) \leq H(\mathfrak{A}_1\mathfrak{A}_2) + H(\mathfrak{A}_2\mathfrak{A}_3)$.

   (In our language this is equivalent to the more natural form

   $H(\mathfrak{A}_1\mathfrak{A}_3|\mathfrak{A}_2) \leq H(\mathfrak{A}_1|\mathfrak{A}_2) + H(\mathfrak{A}_3|\mathfrak{A}_2)$*).*

Proof. Subadditivity is a special case of strong subadditivity: $\mathfrak{A}_2 = \mathbb{C}$. The latter can be reduced to the familiar form, proved first by LIEB & RUSKAI (see the references in UHLMANN (1977)), by the same type of argument as we used in the section *Subalgebra language* for the nonnegativity of conditional mutual information. □

Another kind of inequality may serve as an operational justification of the definition of VON NEUMANN entropy. Call a quantum operation $\varphi : \mathfrak{A}_1 \to \mathfrak{A}_2$ *doubly stochastic* if it preserves the trace, i.e. for all $A \in \mathfrak{A}_1$: $\mathrm{Tr}\,\varphi(A) = \mathrm{Tr}\,A$ (see OHYA & PETZ (1993)). We will consider the less restrictive condition $\mathrm{Tr}\,\varphi(A) \leq \mathrm{Tr}\,A$, and for an observable X, a $*$–subalgebra $\mathfrak{X}$ let us say it is *maximal in* $\mathfrak{A}$ if X, the inclusion map has this property, respectively (obviously for the $*$–subalgebra this implies doubly stochastic). Main examples are: an observable whose atoms are minimal in the target algebra, i.e. have only trivial decompositions into positive operators, and a maximal commutative $*$–subalgebra.

Theorem A.10 (Entropy increase) *Let $\varphi : \mathfrak{Y} \to \mathfrak{X}$ with $\mathrm{Tr}\,\varphi(A) \leq \mathrm{Tr}\,A$, and $\psi : \mathfrak{X} \to \mathfrak{A}$ quantum operations. Then $H(\psi \circ \varphi) \geq H(\psi)$. (Notice that in the physical sense the operation φ_* is applied* after *ψ_*).*

Before we prove this let us note two important case of equality: Let $\rho = \sum_i \lambda_i p_i$ with mutually orthogonal pure states p_i, $\lambda_i \geq 0$, $\sum_i p_i = \mathbb{1}$. Then equality holds for the $*$–subalgebra generated by the p_i (in fact for any $*$–subalgebra which contains them), and for the observable that corresponds to the p_i's resolution of $\mathbb{1}$.
Proof of theorem A.10. Let $\sigma = \psi_* \rho$, we have to prove $H(\varphi_* \sigma) \geq H(\sigma)$. From the previous discussion we see that we may assume $\mathfrak{Y}$ to be commutative, without changing the trace relation. Let $\sigma = \sum_i \alpha_i p_i$ a diagonalization with pure states p_i on $\mathfrak{X}$, and q_j the family of minimal idempotents of $\mathfrak{Y}$ (which by commutativity are othogonal). Then we have decompositions $\varphi_* p_i = \sum_j \beta_{ij} q_j$, hence

$$\varphi_* \sigma = \sum_i \alpha_i \varphi_* p_i = \sum_j \left(\sum_i \alpha_i \beta_{ij} \right) q_j \ .$$

Now observe that for all j

$$\sum_i \beta_{ij} = \mathrm{Tr} \left(q_j \sum_i \varphi_* p_i \right) = \mathrm{Tr} \left((\varphi q_j) \sum_i p_i \right) = \mathrm{Tr}\,(\varphi q_j) \leq \mathrm{Tr}\,(q_j) = 1,$$

and the result follows from the formulas $H(\sigma) = H(\alpha_i | i)$, $H(\varphi_* \sigma) = H(\sum_i \beta_{ij} \alpha_i | j)$. □

Let us formulate the special cases of maximal observables and maximal $*$–subalgebras as a corollary:

Corollary A.11 *Let X an observable maximal in $\mathfrak{X}$, then $H(X) \geq H(\mathfrak{X})$. Let $\mathfrak{X}'$ a $*$–subalgebra maximal in $\mathfrak{X}$, then $H(\mathfrak{X}') \geq H(\mathfrak{X})$.* □

An application of this is in the proof of

Theorem A.12 *Let $\mathfrak{X}, \mathfrak{Y}$ compatible, $\rho|_{\mathfrak{X}\mathfrak{Y}}$ pure. Then $H(\mathfrak{X}) = H(\mathfrak{Y})$.*

Proof. By retracting the state ρ to $\mathfrak{X} \otimes \mathfrak{Y}$ by the multiplication map $\mu : \mathfrak{X} \otimes \mathfrak{Y} \to \mathfrak{X}\mathfrak{Y}$ (see lemma A.6) and embedding $\mathfrak{X}$ and $\mathfrak{Y}$ into full matrix algebras (see the proof of the next theorem) we may assume that we have a pure state ρ on $\mathfrak{L}(\mathcal{H}_1) \otimes \mathfrak{L}(\mathcal{H}_2)$ (entropies do not change as the $*$–subalgebras are maximal). Then the assertion of the theorem is $H(\mathrm{Tr}_{\mathfrak{X}} \rho) = H(\mathrm{Tr}_{\mathfrak{Y}} \rho)$ which is well known (proof via the SCHMIDT decomposition of $|\psi\rangle$, where $\rho = |\psi\rangle\langle\psi|$: cf. PERES (1995)). □

Theorem A.13 *Let $\mathfrak{X}, \mathfrak{Y}$ compatible, ρ any state. Then $|H(\mathfrak{X}) - H(\mathfrak{Y})| \leq H(\mathfrak{X}\mathfrak{Y})$.*

Proof. Like in the previous theorem we may assume that ρ is a state on $\mathfrak{X} \otimes \mathfrak{Y}$, and by symmetry we have to prove that

$$H(\mathfrak{X}) - H(\mathfrak{Y}) \leq H(\mathfrak{X}\mathfrak{Y}).$$

If we think of $\mathfrak{X}$ and $\mathfrak{Y}$ as sums of full operator algebras, say $\mathfrak{X} = \bigoplus_i \mathfrak{L}(\mathcal{H}_i)$, $\mathfrak{Y} = \bigoplus_j \mathfrak{L}(\mathcal{K}_j)$, then embedding them into $\mathfrak{L}(\bigoplus_i \mathcal{H}_i)$, $\mathfrak{L}(\bigoplus_j \mathcal{K}_j)$, respectively, does not change the entropies involved (because the $*$–subalgebras are maximal). Thus we may assume that $\mathfrak{X} = \mathfrak{L}(\mathcal{H})$, $\mathfrak{Y} = \mathfrak{L}(\mathcal{K})$. Now consider a *purification* $|\psi\rangle$ of ρ on the Hilbert space $\mathcal{H} \otimes \mathcal{K} \otimes \mathcal{L}$ (see e.g. SCHUMACHER (1996)): this means $\rho = \operatorname{Tr}_{\mathfrak{L}(\mathcal{L})} |\psi\rangle\langle\psi|$. Now by theorem A.12 $H(\mathfrak{X}) = H(\mathfrak{Y}\mathfrak{Z})$, $H(\mathfrak{X}\mathfrak{Y}) = H(\mathfrak{Z})$, and the assertion follows from the subadditivity theorem A.9: $H(\mathfrak{Y}\mathfrak{Z}) \leq H(\mathfrak{Y}) + H(\mathfrak{Z})$. □

Information The following inequality for mutual information is a straightforward generalization of the HOLEVO bound (HOLEVO (1973), see theorem A.16 below):

Theorem A.14 *Let X, Y be compatible observables with values in the compatible $*$–subalgebras $\mathfrak{X}, \mathfrak{Y}$, respectively. Then*

$$I(X \wedge Y) \leq I(\mathfrak{X} \wedge Y) \leq I(\mathfrak{X} \wedge \mathfrak{Y}).$$

Proof. Consider the diagram

$$\begin{tikzcd}
\mathfrak{B}(\Omega_X) \arrow[r, "X"] \arrow[d] & \mathfrak{X} \arrow[r, equal] \arrow[d] & \mathfrak{X} \arrow[d, "\varphi"'] & \\
\mathfrak{B}(\Omega_X) \otimes \mathfrak{B}(\Omega_Y) \arrow[r, "X\otimes\mathrm{id}"] & \mathfrak{X} \otimes \mathfrak{B}(\Omega_Y) \arrow[r, "\mathrm{id}\otimes Y"] & \mathfrak{X} \otimes \mathfrak{Y} \arrow[r, "\mu"] & \mathfrak{A} \\
\mathfrak{B}(\Omega_Y) \arrow[r, equal] \arrow[u] & \mathfrak{B}(\Omega_Y) \arrow[r, "Y"] \arrow[u] & \mathfrak{Y} \arrow[u, "\varphi'"] &
\end{tikzcd}$$

and apply the Lindblad–Uhlmann monotonicity theorem A.5 twice, with $\mu_*(\rho)$ and the maps $(\mathrm{id} \otimes Y)_*$ and $(X \otimes \mathrm{id})_*$, one after the other. □

This can be greatly extended: for example if $\mathfrak{X} \subset \mathfrak{X}'$, $\mathfrak{Y} \subset \mathfrak{Y}'$, then

$$I(\mathfrak{X} \wedge \mathfrak{Y}) \leq I(\mathfrak{X}' \wedge \mathfrak{Y}').$$

The most general form is

$$I(\psi_1 \circ \varphi_1 \wedge \psi_2 \circ \varphi_2) \leq I(\psi_1 \wedge \psi_2)$$

in the diagram

$$\begin{tikzcd}
\mathfrak{A}_1' \arrow[r, "\varphi_1"] \arrow[d] & \mathfrak{A}_1 \arrow[r, "\psi_1"] \arrow[d] & \mathfrak{A} \arrow[d, equal] \\
\mathfrak{A}_1' \otimes \mathfrak{A}_2' \arrow[r, "\varphi_1\otimes\varphi_2"] & \mathfrak{A}_1 \otimes \mathfrak{A}_2 \arrow[r, "\psi=\psi_1\psi_2"] & \mathfrak{A} \\
\mathfrak{A}_2' \arrow[r, "\varphi_2"] \arrow[u] & \mathfrak{A}_2 \arrow[r, "\psi_2"] \arrow[u] & \mathfrak{A} \arrow[u, equal]
\end{tikzcd}$$

Remark A.15 *It is worth noting that the above formulation of the information bound has the nice form of a data processing inequality. To dwell on this point a little more, and at the same time link our discussion with the traditional view and the language employed in the chapters I and II of the main text let us define for a (measureable) map* $\varphi_* : \mathcal{X} \to \mathfrak{S}(\mathfrak{Y})$ *(which we identify with its linear extension to* $\mathbb{C}\mathcal{X}$ *and regard as a* quantum channel*, see chapter II) and a p.d.* P *on* $\mathcal{X}$

$$I(P; \varphi_*) = I_\gamma(\mathbb{C}\mathcal{X} \wedge \mathfrak{Y})$$

with the channel state $\gamma = \sum_{x\in\mathcal{X}} P(x)[x] \otimes \varphi_*(x)$. *It is easily verified that*

$$I(P;\varphi_*) = H(P\varphi_*) - H(\varphi_*|P) \quad \textit{where} \begin{cases} P\varphi_* = \mathrm{Tr}_{\mathbb{C}\mathcal{X}}\gamma = \sum_{x\in\mathcal{X}} P(x)\varphi_*(x), \\ H(\varphi_*|P) = \sum_{x\in\mathcal{X}} P(x)H(\varphi_*(x)). \end{cases}$$

Now with a quantum operation $\psi_* : \mathfrak{Y}_* \to \mathfrak{Z}_*$ *the data processing inequality specializes to*

$$I(P; \psi_* \circ \varphi_*) \leq I(P; \varphi_*).$$

In particular if $\mathfrak{Z}$ is commutative, i.e. the operation, now denoted D_*, is a measurement, we recover the

Theorem A.16 (Holevo bound) $I(P; D_* \circ \varphi_*) \leq I(P; \varphi_*)$. □

In chapter II an elementary proof of this inequality is presented.

Theorem A.17 *Let* $\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{Y}_1, \mathfrak{Y}_2$ *compatible* $*$*–subalgebras of* $\mathfrak{A}$*,* ρ *a state on* $\mathfrak{A}$*. Then*

$$I(\mathfrak{X}_1\mathfrak{X}_2 \wedge \mathfrak{Y}_1\mathfrak{Y}_2) \leq I(\mathfrak{X}_1 \wedge \mathfrak{Y}_1) + I(\mathfrak{X}_2 \wedge \mathfrak{Y}_2)$$

if $I(\mathfrak{Y}_1 \wedge \mathfrak{X}_2\mathfrak{Y}_2|\mathfrak{X}_1) = 0$ *and* $I(\mathfrak{Y}_2 \wedge \mathfrak{X}_1\mathfrak{Y}_1|\mathfrak{X}_2) = 0$ *(i.e.* $\mathfrak{Y}_k$ *is* independent *from the other* $*$*–subalgebras conditional on* $\mathfrak{X}_k$*).*

Proof. First observe that the conditional independence mentioned, $I(\mathfrak{Y}_1 \wedge \mathfrak{X}_2\mathfrak{Y}_2|\mathfrak{X}_1) = 0$, is equivalent to $H(\mathfrak{Y}_1|\mathfrak{X}_1\mathfrak{X}_2\mathfrak{Y}_2) = H(\mathfrak{Y}_1|\mathfrak{X}_1)$. By theorem A.23 we then have also $H(\mathfrak{Y}_1|\mathfrak{X}_1\mathfrak{X}_2) = H(\mathfrak{Y}_1|\mathfrak{X}_1)$. Now observe (with the obvious chain rule)

$$\begin{aligned} H(\mathfrak{Y}_1\mathfrak{Y}_2|\mathfrak{X}_1\mathfrak{X}_2) &= H(\mathfrak{Y}_1|\mathfrak{X}_1\mathfrak{X}_2\mathfrak{Y}_2) + H(\mathfrak{Y}_2|\mathfrak{X}_1\mathfrak{X}_2) \\ &= H(\mathfrak{Y}_1|\mathfrak{X}_1) + H(\mathfrak{Y}_2|\mathfrak{X}_2) \end{aligned}$$

and hence

$$\begin{aligned} I(\mathfrak{X}_1\mathfrak{X}_2 \wedge \mathfrak{Y}_1\mathfrak{Y}_2) &= H(\mathfrak{Y}_1\mathfrak{Y}_2) - H(\mathfrak{Y}_1\mathfrak{Y}_2|\mathfrak{X}_1\mathfrak{X}_2) \\ &\leq H(\mathfrak{Y}_1) + H(\mathfrak{Y}_2) - H(\mathfrak{Y}_1|\mathfrak{X}_1) - H(\mathfrak{Y}_2|\mathfrak{X}_2) \\ &= I(\mathfrak{X}_1 \wedge \mathfrak{Y}_1) + I(\mathfrak{X}_2 \wedge \mathfrak{Y}_2) \end{aligned}$$

where we have used the subadditivity of von Neumann entropy, theorem A.9. □

The same obviously applies if we have n $*$–subalgebras $\mathfrak{X}_k$, and n $\mathfrak{Y}_k$, all compatible, and if $\mathfrak{Y}_k$ is independent from the others given $\mathfrak{X}_k$, i.e. for all k

$$H(\mathfrak{Y}_k|\mathfrak{X}_1 \cdots \mathfrak{X}_n\mathfrak{Y}_1 \cdots \widehat{\mathfrak{Y}_k} \cdots \mathfrak{Y}_n) = H(\mathfrak{Y}_k|\mathfrak{X}_k).$$

Corollary A.18 *Let $\mathfrak{X}_1, \dots, \mathfrak{X}_n$, $\mathfrak{Y}_1, \dots, \mathfrak{Y}_n$ C^*–algebras, $\mathfrak{X}_i = \mathbb{C}\mathcal{X}_i$ commutative, and $\mathfrak{A} = \mathfrak{X}_1 \otimes \cdots \otimes \mathfrak{X}_n \otimes \mathfrak{Y}_1 \otimes \cdots \otimes \mathfrak{Y}_n$. Then with the state*

$$\gamma = \sum_{x_i \in \mathcal{X}_i} P(x_1, \dots, x_n)[x_1] \otimes \cdots \otimes [x_n] \otimes W_{x_1} \otimes \cdots \otimes W_{x_n}$$

on $\mathfrak{A}$ (where P is a p.d. on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ and W maps the $\mathcal{X}_i$ to states on $\mathfrak{Y}_i$):

$$I(\mathfrak{X}_1 \cdots \mathfrak{X}_n \wedge \mathfrak{Y}_1 \cdots \mathfrak{Y}_n) \leq \sum_{k=1}^{n} I(\mathfrak{X}_k \wedge \mathfrak{Y}_k).$$

Proof. We only have to check the conditional independence, which is left to the reader. □

We note another estimate for the mutual information:

Theorem A.19 *For compatible $*$–subalgebras $\mathfrak{X}, \mathfrak{Y}$: $I(\mathfrak{X} \wedge \mathfrak{Y}) \leq 2\min\{H(\mathfrak{X}), H(\mathfrak{Y})\}$.*

Proof. Put together the formula $I(\mathfrak{X} \wedge \mathfrak{Y}) = H(\mathfrak{X}) - H(\mathfrak{X}|\mathfrak{Y})$ and the simple estimate $H(\mathfrak{X}|\mathfrak{Y}) \geq -H(\mathfrak{X})$ from theorem A.13. □

Conditional entropy We start with a simple positivity condition:

Theorem A.20 *Let $\varphi : \mathfrak{X} \to \mathfrak{A}$, $\psi : \mathfrak{Y} \to \mathfrak{A}$ compatible quantum operations with $\mathfrak{X}$ or $\mathfrak{Y}$ commutative. Then $H(\varphi|\psi) \geq 0$.*

Proof. Let $\sigma = (\varphi\psi)_*\rho$, then by definition and lemma A.6

$$H(\varphi|\psi) = H(\sigma) - H(\mathrm{Tr}_{\mathfrak{Y}}\sigma).$$

First case: $\mathfrak{X}$ is commutative, so we can write $\sigma = \sum_x Q(x)[x] \otimes \tau_*(x)$ with a distribution Q on $\mathcal{X}$, and states $\tau_*(x)$ on $\mathfrak{Y}$. Obviously $H(\sigma) = H(Q) + \sum_x Q(x)H(\tau_*(x))$, and $\mathrm{Tr}_{\mathfrak{Y}}\sigma = \sum_x Q(x)[x] = Q$, and hence $H(\varphi|\psi) = \sum_x Q(x)H(\tau_*(x)) \geq 0$.

Second case: $\mathfrak{Y}$ is commutative, so we can write $\sigma = \sum_x Q(x)[x]\tau_*(x) \otimes [x]$, like in the first case. $H(\sigma)$ is calculated as before, but now $\mathrm{Tr}_{\mathfrak{Y}}\sigma = \sum_x Q(x)\tau_*(x) = Q\tau_*$, and

$$\begin{aligned} H(\varphi|\psi) &= H(Q) - \left(H(Q\tau_*) - \sum_x Q(x)H(\tau_*(x)) \right) \\ &= H(Q) - I(Q; \tau_*) \geq 0, \end{aligned}$$

the last step by an application of the Holevo bound, theorem A.16. □

Remark A.21 *From the proof we see that the commutativity of $\mathfrak{X}$ or $\mathfrak{Y}$ enters in the representation of σ as a particular separable state with respect to the $*$–subalgebras $\mathfrak{X}$, $\mathfrak{Y}$ (see definition below), namely with one party admitting common diagonalization of her states. We formulate as a conjecture the more general:*

$$H(\mathfrak{X}|\mathfrak{Y}) \geq 0 \text{ if } \rho \text{ is separable with respect to } \mathfrak{X} \text{ and } \mathfrak{Y}.$$

From this it would follow that in this case $I(\mathfrak{X} \wedge \mathfrak{Y}) \leq \min\{H(\mathfrak{X}), H(\mathfrak{Y})\}$ (compare theorem A.19), which we now only get from the commutativity assumption.

Definition A.22 *Call* ρ separable with respect to *compatible* $*$*–subalgebras* $\mathfrak{X}_1, \dots, \mathfrak{X}_m$ *of* $\mathfrak{A}$*, if, for the natural multiplication map* $\mu : \mathfrak{X}_1 \otimes \cdots \otimes \mathfrak{X}_m \to \mathfrak{A}$*,* $\mu_* \rho$ *is a separable state on* $\mathfrak{X}_1 \otimes \cdots \otimes \mathfrak{X}_m$*, i.e. a convex combination of product states* $\sigma_1 \otimes \cdots \otimes \sigma_m$*,* $\sigma_i \in \mathfrak{S}(\mathfrak{X}_i)$*. If* $\mu_* \rho$ *is a product state, we call also* ρ *a* product state with respect to $\mathfrak{X}_1, \dots, \mathfrak{X}_m$.

Theorem A.23 (Knowledge decreases uncertainty) *Let* $\varphi : \mathfrak{X} \to \mathfrak{A}$*,* $\psi : \mathfrak{Y} \to \mathfrak{A}$ *compatible quantum operations, and* $\varphi' : \mathfrak{X}' \to \mathfrak{X}$ *any quantum operation.*
Then $H(\psi|\varphi) \leq H(\psi|\varphi \circ \varphi')$*, and in particular* $H(\psi|\varphi) \leq H(\psi)$*.*

Proof. The inequality is obviously equivalent to $I(\psi \wedge \varphi) \geq I(\psi \wedge \varphi \circ \varphi')$, i.e. to theorem A.14. □

Defining $h(x) = -x \log x - (1-x)\log(1-x)$ for $x \in [0,1]$ we have the famous

Theorem A.24 (Fano inequality) *Let* ρ *a state on* $\mathfrak{A}$*, and* $\mathfrak{Y}$ *be a* $*$*–subalgebra of* $\mathfrak{A}$*, compatible with the observable* X *(indexed by* $\mathcal{X}$*). Then for any observable* Y *with values in* $\mathfrak{Y}$ *the probability that "*$X \neq Y$*", i.e.* $P_e = 1 - \sum_j \mathrm{Tr}\,(\rho X_j Y_j)$*, satisfies*

$$H(X|\mathfrak{Y}) \leq h(P_e) + P_e \log(|\mathcal{X}| - 1).$$

Proof. By the previous theorem A.23 it suffices to prove the inequality with $H(X|Y)$ instead of $H(X|\mathfrak{Y})$. But then we have the classical Fano inequality: the uncertainty on X given Y may be estimated by the uncertainty of the event that they are equal plus the uncertainty on the value of X if they are not. □

Corollary A.25 *Let* $\mathfrak{X}$ *a commutative* $*$*–subalgebra compatible with* $\mathfrak{Y}$*, and* X *the — uniquely determined — maximal observable on* $\mathfrak{X}$*,* P_e *as in the theorem, then*

$$H(\mathfrak{X}|\mathfrak{Y}) \leq h(P_e) + P_e \log(\mathrm{Tr}\,\mathrm{supp}\,(\rho|_{\mathfrak{X}}) - 1).$$

Proof. First observe that $H(\mathfrak{X}|\mathfrak{Y}) = H(X|\mathfrak{Y})$. To apply the theorem we only have to restrict the range of X to those values that are actually assumed. □

Some philosophical remarks may be in order: quantum theory stipulates the channel as a process, an asymmetric notion, and this brings about the formula $I(P; \varphi_*) = H(P\varphi_*) - H(\varphi_*|P)$: input, average and conditional output entropy. In classical information theory however we like to see things more symmetric, namely the channel as a stochastic two–end system, with some underlying joint distribution. Following this idea produces our channel states γ, and a symmetric "information" expression $I(\mathfrak{X} \wedge \mathfrak{Y})$. Even though there are questions in quantum information where these two pictures can be brought to relation, for example in the above results (a connection that was noticed before by Hall (1997) in his investigation of what he calls *context mappings*), they are not reducible to each other: the "dynamic" picture is asymmetric (there may not even exist a backward channel producing the same channel state), whereas the "static" picture is obviously symmetric. Even worse, for a joint state it is not obvious that a channel and input distribution generating it exist at all. And if it exists, there is no uniqueness in its choice. On the other hand, modelling a situation of quantum evolutions statically may produce unphysical effects, see the example from Winter (1998c), VIII.B.2, pp.24: the channel state incorporates parts of a system which can never be simultaneously accessible.

References

AHLSWEDE, R. (1968), Beiträge zur Shannonschen Informationstheorie im Falle nichtstationärer Kanäle. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete*, 10, 1–42.

AHLSWEDE, R. (1971), Multi–way communication channels. In *Second International Symposium on Information Theory*, 23–52, Hungarian Academy of Sciences.

AHLSWEDE, R. (1974a), The capacity region of a channel with two senders and two receivers. *Ann. Prob.*, 2(5), 805–814.

AHLSWEDE, R. (1974b), Paper presented at 7^{th} *Hawaii International Conference on System Sciences*, Jan. 1974. Published in AHLSWEDE, R. & KÖRNER, J. (1975), Source coding with side information at the decoder and a converse for degraded broadcast channels. *IEEE Trans. Inf. Theory*, 21, 629–637.

AHLSWEDE, R. (1979), Coloring Hypergraphs: A New Approach to Multi–user Source Coding — I. *J. Combinatorics, Information & System Sciences*, 4(1), 76–115.

AHLSWEDE, R. (1980), Coloring Hypergraphs: A New Approach to Multi–user Source Coding — II. *J. Combinatorics, Information & System Sciences*, 5(3), 220–268.

AHLSWEDE, R., GACŚ, P. & KÖRNER, J. (1976), Bounds on Conditional Probabilities with Applications to Multi–User Communication. *Z. Wahrscheinlichkeitstheorie und verw. Geb.*, 34, 157–177.

ALLAHVERDYAN, A. E. & SAAKIAN, D. B. (1997a), Converse coding theorems for quantum source and noisy channels. LANL eprint `quant-ph/9702034`, `http://xxx.lanl.gov/`.

ALLAHVERDYAN, A. E. & SAAKIAN, D. B. (1997b), Multi–access channels in quantum information theory. LANL eprint `quant-ph/9712034`, `http://xxx.lanl.gov/`.

ARVESON, W. (1976), *An Invitation to C^*–Algebras*. Springer, New York.

BARNUM, H. (1998), Quantum Rate–Distortion Coding. LANL eprint `quant-ph/9806065`, `http://xxx.lanl.gov/`.

Barnum, H., Fuchs, C. A., Jozsa, R. & Schumacher, B. (1996), General Fidelity Limit for Quantum Channels. *Phys. Rev. A*, 54, 4707–4711.

Barnum, H., Nielsen, M. A. & Schumacher, B. (1998), Information transmission through a noisy quantum channel. *Phys. Rev. A*, 57(6), 4153–4175.

Bell, J. (1964), On the Einstein Podolsky Rosen paradox. *Physics*, 1(3), 195–200.

Bendjaballah, C., Leroy, J. M. & Vourdas, A. (1998), Rate Distortion and Detection in Quantum Communication. *IEEE Trans. Inf. Theory*, 44(4), 1658–1665.

Bennett, C. H. & Wiesner, S. J. (1992), Communication via one– and two–particle operators on Einstein–Podolsky–Rosen states. *Phys. Rev. Letters*, 69(20), 2881–2884.

Born, M. (1926), Zur Quantenmechanik der Stoßvorgänge. *Zeitschrift für Physik*, 37, 863–867.

Braunstein, S. L., Fuchs, C. A., Gottesman, D. & Lo, H.-K. (1998), A quantum analog of Huffman coding. LANL eprint `quant-ph/9805080`, `http://xxx.lanl.gov/`, presented at the IEEE International Symposium on Information Theory, Boston, 1998.

Burnashev, M. V. & Holevo, A. (1997), On Reliability Function of Quantum Communication Channel. LANL eprint `quant-ph/9703013`, `http://xxx.lanl.gov/`.

Csiszár, I. & Körner, J. (1981), *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, New York.

Davies, E. B. (1976), *Quantum Theory of Open Systems*. Academic Press, London.

Einstein, A., Podolsky, B. & Rosen, N. (1935), Can Quantum–Mechanical Description of Physical Reality be Considered Complete? *Physical Review*, 47, 777–780.

El Gamal, A. & Cover, T. (1980), Multiple User Information Theory. *Proc. IEEE*, 68, 1466–1483.

Forney, jr., G. D. (1963), *(unpublished)*. Master's thesis, MIT, Boston.

Gordon, J. P. (1964), Noise at optical frequencies: information theory. In *Proc. Int. School Phys. "Enrico Fermi"* (P. A. Miles, ed.), 156–181, Academic Press, New York.

Greenberger, D. M., Horne, M. A., Shimony, A. & Zeilinger, A. (1990), Bell's theorem without inequalities. *Am. J. Phys.*, 58(12), 1131–1143.

Hall, M. J. W. (1997), Quantum information and correlation bounds. *Phys. Rev. A*, 55(1), 100–113.

HAROUTUNIAN, A. E. (1968), Estimates on the error exponent for the semicontinuous memoryless quantum channel. *Probl. Peredachi Inform.*, 4(4), 37–48, in Russian.

HAUSLADEN, P., JOZSA, R., SCHUMACHER, B., WESTMORELAND, M. & WOOTTERS, W. K. (1997), Classical information capacity of a quantum channel. *Phys. Rev. A*, 54(3), 1869–1876.

HOEFFDING, W. (1963), Probability Inequalities for Sums of Bounded Random Variables. *J. Amer. Statist. Assoc.*, 58, 13–30.

HOLEVO, A. S. (1973), Bounds for the quantity of information transmitted by a quantum channel. *Probl. Inf. Transm.*, 9(3), 177–183.

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

HOLEVO, A. S. (1979), Capacity of a Quantum Communication Channel. *Probl. Inf. Transm.*, 15(4), 247–253.

HOLEVO, A. S. (1982), *Probabilistic and Statistical Aspects of Quantum Theory*. North–Holland, Amsterdam.

HOLEVO, A. S. (1998a), The Capacity of the Quantum Channel with General Signal States. *IEEE Trans. Inf. Theory*, 44(1), 269–273.

HOLEVO, A. S. (1998b), Coding Theorems for Quantum Channels. Research Review 4, Tamagawa University, extended version as LANL eprint `quant-ph/9809023`.

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

JOZSA, R. (1994), Fidelity for mixed states. *J. Mod. Optics*, 41(12), 2315–2323.

JOZSA, R., HORODECKI, M., HORODECKI, P. & HORODECKI, R. (1998), Universal Quantum Information Compression. *Phys. Rev. Letters*, 81, 1714–1717.

JOZSA, R. & SCHUMACHER, B. (1994), A new proof of the quantum noiseless coding theorem. *J. Mod. Optics*, 41(12), 2343–2349.

KRAUS, K. (1983), *States, Effects and Operations*. Number 190 in Lecture Notes in Physics, Springer, Berlin.

LEVITIN, L. B. (1969), On quantum measure of information. In *Proc. IV All–Union Conference on Information Transmission and Coding Theory*, 111–115, Tashkent.

LEVITIN, L. B. (1998), Conditional Entropy and Information in Quantum Systems. In *Proc. IEEE International Symposium on Information Theory, Boston*, 88.

LINDBLAD, G. (1975), Completely positive maps and entropy inequalities. *Comm. Math. Phys.*, 40, 147–151.

LUDWIG, G. (1954), *Die Grundlagen der Quantenmechanik*. Springer, Berlin.

NAGAOKA, H. (1998), Algorithms of Arimoto–Blahut Type for Computing Quantum Channel Capacity. In *Proc. IEEE International Symposium on Information Theory, Boston*.

OGAWA, T. & NAGAOKA, H. (1998), Strong Converse to the Quantum Channel Coding Theorem. LANL eprint `quant-ph/9808063`, `http://xxx.lanl.gov/`.

OHYA, M. & PETZ, D. (1993), *Quantum Entropy and Its Use*. Springer, Berlin.

PERES, A. (1995), *Quantum Theory: Concepts and Methods*. Kluwer, Dordrecht.

SCHUMACHER, B. (1995), Quantum Coding. *Phys. Rev. A*, 51(4), 2738–2747.

SCHUMACHER, B. (1996), Sending entanglement through noisy quantum channels. *Phys. Rev. A*, 54(4), 2614–2628.

SCHUMACHER, B. & NIELSEN, M. A. (1996), Quantum data processing and error correction. *Phys. Rev. A*, 54(4), 2629–2635.

SCHUMACHER, B. & WESTMORELAND, M. (1997), Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56(1), 131–138.

SHANNON, C. E. (1948), A Mathematical Theory of Communication. *Bell System Thechnical Journal*, 27, 379–423.

SHANNON, C. E. (1961), Two–way communication channels. In *Proc. Fourth Berkeley Symposium Probability and Statistics* (J. Neyman, ed.), 611–644, Berkeley.

SHOR, P. W. (1994), Algorithms for Quantum Computation: Discrete Logarithm and Factoring. In *Proc. 35th Ann. Symp. Foundations of Computer Science* (S. Goldwasser, ed.), 124–134, IEEE Press, Santa Fe, full paper: Polynomial–Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer, *SIAM J. Computing*, 26(5), 1484–1509 (1997).

STINESPRING, W. F. (1955), Positive Functions on C*–Algebras. *Proc. Amer. Math. Soc.*, 6, 211–216.

STRATONOVICH, R. L. (1966), Information transmission rate in certain quantum communication channels. *Probl. Peredachi Inform.*, 2(2), 45–57, in Russian.

UHLMANN, A. (1977), Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory. *Comm. Math. Phys.*, 54, 21–32.

Umegaki, H. (1962), Conditional expectations in an operator algebra, IV (entropy and information). *Kodai Math. Sem. Rep.*, 14, 59–85.

von Neumann, J. (1927), Thermodynamik quantenmechanischer Gesamtheiten. *Nachr. der Gesellschaft der Wiss. Gött.*, 273–291.

Wehrl, A. (1978), General properties of entropy. *Rev. Mod. Phys.*, 50(2), 221–260.

Winter, A. (1998a), The Capacity of the Quantum Multiple Access Channel. LANL eprint `quant-ph/9807019`, `http://xxx.lanl.gov/`.

Winter, A. (1998b), Coding Theorem and Strong Converse for Quantum Channels. Preprint 98–074, Sonderforschungsbereich 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld, to appear in *IEEE Trans. Inf. Theory*, Nov. 1999.

Winter, A. (1998c), Languages of Quantum Information Theory. Preprint E98–009, Sonderforschungsbereich 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld, also as LANL eprint `quant-ph/9807008`.

Winter, A. (1999a), Coding Theorem and Strong Converse for Nonstationary Quantum Channels. Preprint 99–033, Sonderforschungsbereich 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld.

Winter, A. (1999b), Rate and Error Estimates for Quantum Channels. Preprint 99–032, Sonderforschungsbereich 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld.

Winter, A. (1999c), Schumacher's Quantum Coding Revisited. Preprint 99–034, Sonderforschungsbereich 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld.

Wolfowitz, J. (1964), *Coding Theorems of Information Theory*. Springer, Berlin, second edition.

Wootters, W. K. & Zurek, W. H. (1982), A single quantum cannot be cloned. *Nature*, 299, 802–803.