Languages of Quantum Information Theory

Andreas Winter*

SFB 343, Fakultät für Mathematik

Universität Bielefeld, Postfach 100131, 33501 Bielefeld

(September 8, 1998)

Abstract

This note will introduce some notation and definitions for information theoretic quantities in the context of quantum systems, such as (conditional) entropy and (conditional) mutual information. We will employ the natural C*–algebra formalism, and it turns out that one has an allover dualism of language: we can define everything for (compatible) observables, but also for (compatible) C*–subalgebras. The two approaches are unified in the formalism of quantum operations, and they are connected by a very satisfying inequality, generalizing the well known Holevo bound. Then we turn to communication via (discrete memoryless) quantum channels: we formulate the Fano inequality, bound the capacity region of quantum multiway channels, and comment on the quantum broadcast channel.

1998 PACS number(s): 03.67.-a, 03.67.Hk, 89.70.+c

Typeset using REVTEX

*Electronic address: `winter@mathematik.uni-bielefeld.de`

I. INTRODUCTION

After the beginnings of quantum information theory in the sixties [1], and Holevo's now widely known investigations of the seventies [2–4] today there is again a tremendous interest in this field. This interest focuses on two areas which may be described, sightly abusing language introduced by Holevo twenty years ago [3], as *classical–quantum* problems on the one hand, and *quantum–quantum* problems on the other, and it mostly derives from the latter, as these include all problems of (quantum) information processing inside a quantum computer or memory. Whereas this area (which is charcterized by its attention to *entanglement*) poses many new and beautiful, and also very difficult problems, the present note is concerned wholly with the former area (though it is by now not altogether clear how to separate these two worlds, cf. e.g. opinion uttered by Adami and Cerf [5]). We take the view that classical–quantum problems are those in which classical information has to be stored in or sent trough some quantum system. Examples from recent work are the determination of the quantum channel capacity for fixed input states [6–8], quantum cryptographic protocols [9,10], and entanglement enhanced transmission (superdense coding) [11].
Our approach is somewhat reminiscent of "quantum probability" through its formulation in terms of C*–algebras and its emphasis on observable operators (which reflects our dwelling in the classical–quantum area), but we cannot respect the bounds of this field: we will use positive operator valued measures (instead of unbounded selfadjoint operators), and we will consider quantum operations, both quite uncommon in noncommutative probability. Finally it should be noted that we hardly present any new concepts or results — our contribution lies in introducing a reasonable and efficient calculus.

The outline of the paper is as follows: in section II we will basically recall the language of C*–algebras, completely positive maps, positive operator valued measures, and the notion of compatibility. In the following sections III and IV we will define various information theoretic quantities, first for observables, second for $*$–subalgebras. In section V we will unify these approaches using completely positive C*–algebra maps, and can give meaning to some hybrid expressions in section VI. The observable and subalgebra notions will be brought together in section VII where we prove an information inequality in generalization of the Holevo bound. Up to this point the work consists in the definition of concepts and information theoretic quantities, and proving some simple numerical relations (without providing any operational justifications which seem for the most part not easy to give). The last section VIII will discuss the application of these concepts to quantum channels, stating a Fano inequality, and determining a bound on the capacity region of the quantum multiway channel. We conclude by making some observations for the quantum broadcast channel.

About notation: finite sets will be denoted $\mathcal{A}, \mathcal{B}, \ldots$, the functions exp and log are always to basis 2.

II. MATHEMATICAL DESCRIPTION OF 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 common 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 [12], Dixmier [13], and standard references on basic mathematics of quantum mechanics: Davies [14], Kraus [15], and the more advanced [16] by Holevo.

A. Systems and their states

A C*–*algebra* with unit is a 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.
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. [12]) 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 $\mathcal{H}$. In this note 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 [12]).[1] This includes as extremal cases the algebras $\mathfrak{L}(\mathcal{H})$, and the commutative algebras $\mathbb{C}\mathcal{X}$ over a finite set $\mathcal{X}$. 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.
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

[1] It is certainly the case that most of the material presented may be generalized to infinite dimensional algebras (see e.g. Ohya/Petz [17]). We decided not to try for several reasons: one is that in information theory the interesting things already happen in the discrete and even finite domain, another (decisive) that the present author is only a stumbling beginner in the vast field of C*–algebras. At least it seems clear that the bulk of the things presented here carries over to algebras which are isomorphic to countable sums of full (bounded) operator algebras of separable Hilbert spaces: there we have trace, well behaved tensor products, and the Schatten decomposition (diagonalization) of density operators.

a convex set $\mathfrak{S}(\mathfrak{A})$ whose extreme points are called *pure* states, all others are *mixed.* 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 general this is only true for so–called *normal* states, which means that for an increasing sequence A_n converging in norm to A the values $\rho(A_n)$ converge to $\rho(A)$. 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 [18]. $\mathrm{Tr}\,(\rho A)$ then defines a real bilinear and nondegenerate 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 $\mathfrak{A}$.

B. 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.*[2] An observable X together with a state

[2]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.

ρ yields a probability measure P^X on Ω via

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

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.[3] 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 [19] shows that one comes into serious 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:

---

[3] 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 certainely is possible as demonstrated, but commutativity is not necessary.

the crux is that we cannot access it due to the forbidden noncompatibel observables (a good account of this aspect of quantum theory is in [20]).

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 $\mathfrak{A}_1\mathfrak{A}_2$.

C. Quantum 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 states to states, and being completely positive and trace preserving.[4] 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 [21]:

Theorem 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*

[4] In general this is only true if we restrict φ to be a *normal* map, see Davies [14].

$\mathcal{K}$, 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. [14].

D. Entropy and divergence

We will talk about information theory, so we need a concept of entropy: the von Neumann entropy $H(\rho) = -\mathrm{Tr}\,(\rho \log \rho)$ (introduced in [22][5]) of a state ρ (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 (first defined by Umegaki [24]), or simply divergence 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 [17], and to [25]. Two important facts we will use are

Theorem 2 (Klein inequality) *For positive operators ρ, σ (not necessary 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.

Proof. See [17]. □

Theorem 3 (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. Uhlmann [26], the situation we are in was already solved by Lindblad [27]. For a textbook account see [17]. □

---

[5] After Levitin [23] it was in fact used before by Landau 1929, and by Dirac 1930.

III. OBSERVABLE LANGUAGE

Fix a state on a C*–algebra, say ρ on $\mathfrak{A}$ and let X, Y, Z compatible observables on $\mathfrak{A}$. By the previous section II these are then random variables with a joint distribution, and one defines entropy $H(X)$, 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

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

(which is known from classical information theory).

IV. 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

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

(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 commutes obviously 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:

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

$$I(\mathfrak{X}_1 \wedge \mathfrak{X}_2) = H(\mathfrak{X}_1) + H(\mathfrak{X}_2) - H(\mathfrak{X}_1\mathfrak{X}_2)$$

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

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, 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 1 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 II.2 (we could also have seen this already from the definition by applying subadditivity of von Neumann entropy to the second last line, see theorem VII.1).

Lemma 1 *Let $\mu : \mathfrak{A} \to \mathfrak{B}$ a surjective $*$–algebra homomorphism. Then*

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

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

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 $q' \leq \mu(p)$. Then
$$1 = \mathrm{Tr}\,(q'\mu(p)) = \mathrm{Tr}\,(\mu_*(q')p) \leq \mathrm{Tr}\,(p) = 1$$
So we must have equality which implies $p \leq \mu_*(q')$, but both operators are states, so $p = \mu_*(q')$. Because μ_* is injective this means that there is only one pure state $q' \leq \mu(p)$, i.e. $\mu(p)$ is pure.

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

3. Let $A \in \mathfrak{A}$, $A \geq 0$. Then

$$\begin{aligned}\mathrm{Tr}\,(\mu_*(\mu(p))A) &= \mathrm{Tr}\,(\mu(p)\mu(A)) = \mathrm{Tr}\,(\mu(p)\mu(A)\mu(p)) \\ &= \mathrm{Tr}\,(\mu(pAp)) \leq \mathrm{Tr}\,(pAp) = \mathrm{Tr}\,(pA)\end{aligned}$$

   Thus $\mu_*(\mu(p)) \leq p$. If $\mu(p) \neq 0$ 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

$$\mu_*(\rho) = \mu_*(\mu(\mu_*(\rho))) = \sum_i \alpha_i \mu_*(\mu(p_i)) \leq \sum_i \alpha_i p_i = \mu_*(\rho)$$

   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 VII.1):

Consider the following commuative diagram:

$$\begin{tikzcd}
\mathfrak{Y} \arrow[r, "\varphi_1"] \arrow[d, equal] & \mathfrak{X}_1 \otimes \mathfrak{Y} \arrow[r, "\mu_1"] \arrow[d, "\varphi'_1"] & \mathfrak{X}_1\mathfrak{Y} \arrow[d, "\jmath_1"] & \\
\mathfrak{Y} \arrow[r, "\varphi"] & \mathfrak{X}_1 \otimes \mathfrak{X}_2 \otimes \mathfrak{Y} \arrow[r, "\mu"] & \mathfrak{X}_1\mathfrak{X}_2\mathfrak{Y} \arrow[r, "\jmath"] & \mathfrak{A} \\
\mathfrak{Y} \arrow[r, "\varphi_2"] \arrow[u, equal] & \mathfrak{X}_2 \otimes \mathfrak{Y} \arrow[r, "\mu_2"] \arrow[u, "\varphi'_2"] & \mathfrak{X}_2\mathfrak{Y} \arrow[u, "\jmath_2"] &
\end{tikzcd}$$

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 1 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 2 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 2 factors. In this case the restriction $\rho|_{\mathfrak{A}_i}$ is just a partial trace.

V. 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 [21]). That this is promising one can see from the observation that observables can be interpreted in a natural way as C*–algebra maps: $X : \Omega \rightarrow \mathfrak{A}$ corresponds by linear extension to $X : \mathfrak{B}(\Omega) \rightarrow \mathfrak{A}$, 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 $X_{\mathrm{tot}} : \mathfrak{B}(\Omega)\otimes\mathfrak{A} \rightarrow \mathfrak{A}$ mapping $j\otimes A \mapsto \sqrt{Y_j}A\sqrt{Y_j}$, its *interior* part $X_{\mathrm{int}} = X_{\mathrm{tot}} \circ \imath_{\mathfrak{A}} : \mathfrak{A} \rightarrow \mathfrak{A}$ with $A \mapsto \sum_j \sqrt{Y_j}A\sqrt{Y_j}$, and its *exterior* part $X_{\mathrm{ext}} = X_{\mathrm{tot}} \circ \imath_{\mathfrak{B}(\Omega)}$ which coincides with X.

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

$\mathfrak{A}$ mapping $X \otimes Y \mapsto \varphi(X)\psi(Y)$:

$$\begin{array}{ccc} \mathfrak{X} & \xrightarrow{\varphi} & \mathfrak{A} \\ {\scriptstyle\varphi_1}\downarrow & & \| \\ \mathfrak{X}\otimes\mathfrak{Y} & \xrightarrow{\exists!\varphi\psi} & \mathfrak{A} \\ {\scriptstyle\varphi_2}\uparrow & & \| \\ \mathfrak{Y} & \xrightarrow{\psi} & \mathfrak{A} \end{array}$$

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 [17]: with them the entropy of an operation is related to the mutual information of the operation as a channel (see section VIII). 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).

VI. PIDGIN

With the insight of the preceding section 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 also again there are some expressions which seem suspicious, like

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

But due to the inequality of theorem VII.10 in fact it behaves nicely.

VII. INEQUALITIES

A. Entropy

Theorem 1 *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:* $H(\mathfrak{A}_1\mathfrak{A}_2\mathfrak{A}_3) + H(\mathfrak{A}_2) \leq H(\mathfrak{A}_1\mathfrak{A}_2) + H(\mathfrak{A}_2\mathfrak{A}_3)$.

Proof. Subadditivity is a special case of strong subadditivity: $\mathfrak{A}_2 = \mathbb{C}$. The latter can be reduced to the familiar form (see e.g. Wehrl [25]) by the same type of argument as we used in section IV for the nonnegativity of conditional mutual information... □

Theorem 2 *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 IV.1) we may assume that we have a pure state ρ on $\mathfrak{X} \otimes \mathfrak{Y}$. 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 polar decomposition of ρ...). □

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 [17]). We will consider the less restrictive condition $\operatorname{Tr}\varphi(A) \leq \operatorname{Tr} A$, and for an observable X and subalgebra $\mathfrak{X}$ let us say they are *maximal in* $\mathfrak{A}$ if X and the inclusion map have this property (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 3 (Entropy increase) *Let* $\varphi : \mathfrak{Y} \to \mathfrak{X}$ *with* $\operatorname{Tr}\varphi(A) = \operatorname{Tr} A$, *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 3. 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

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

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 4 *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 5 *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. [28]): this means $\rho = \mathrm{Tr}_{\mathfrak{L}(\mathcal{L})}|\psi\rangle\langle\psi|$. Now by theorem 2 $H(\mathfrak{X}) = H(\mathfrak{Y}\mathfrak{Z})$, $H(\mathfrak{X}\mathfrak{Y}) = H(\mathfrak{Z})$, and the assertion follows from subadditivity theorem 1: $H(\mathfrak{Y}\mathfrak{Z}) \leq H(\mathfrak{Y}) + H(\mathfrak{Z})$. □

B. Information

The following inequality for mutual information is a straightforward generalization of the Holevo bound [2], see also next section VIII:

Theorem 6 *Let X, Y be compatible observables with values in compatible $*$–subalgebras $\mathfrak{X}, \mathfrak{Y}$, respectively. Then*

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

(Conditions of equality!).

Proof. Consider the diagram

$$\begin{array}{ccccccc}
\mathfrak{B}(\Omega_X) & \xrightarrow{X} & \mathfrak{X} & = & \mathfrak{X} & & \\
\downarrow & & \downarrow & & \downarrow{\scriptstyle\varphi} & & \\
\mathfrak{B}(\Omega_X) \otimes \mathfrak{B}(\Omega_Y) & \xrightarrow{X\otimes\mathrm{id}} & \mathfrak{X} \otimes \mathfrak{B}(\Omega_Y) & \xrightarrow{\mathrm{id}\otimes Y} & \mathfrak{X} \otimes \mathfrak{Y} & \xrightarrow{\mu} & \mathfrak{A} \\
\uparrow & & \uparrow & & \uparrow{\scriptstyle\varphi'} & & \\
\mathfrak{B}(\Omega_Y) & = & \mathfrak{B}(\Omega_Y) & \xrightarrow{Y} & \mathfrak{Y} & &
\end{array}$$

and apply the Lindblad–Uhlmann monotonicity theorem II.3 twice, with $\mu_*(\rho)$ and the maps $(\mathrm{id} \otimes Y)_*$ and $(X \otimes \mathrm{id})_*$, one after the other. $\square$

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

$$\begin{array}{ccccc}
\mathfrak{A}_1' & \xrightarrow{\varphi_1} & \mathfrak{A}_1 & \xrightarrow{\psi_1} & \mathfrak{A} \\
\downarrow & & \downarrow & & \| \\
\mathfrak{A}_1' \otimes \mathfrak{A}_2' & \xrightarrow{\varphi_1 \otimes \varphi_2} & \mathfrak{A}_1 \otimes \mathfrak{A}_2 & \xrightarrow{\psi = \psi_1 \psi_2} & \mathfrak{A} \\
\uparrow & & \uparrow & & \| \\
\mathfrak{A}_2' & \xrightarrow{\varphi_2} & \mathfrak{A}_2 & \xrightarrow{\psi_2} & \mathfrak{A}
\end{array}$$

Theorem 7 *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 13 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 1. $\square$

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 give $\mathfrak{X}_k$, 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 8 *Let $\mathfrak{X}_1,\dots,\mathfrak{X}_n$, $\mathfrak{Y}_1,\dots,\mathfrak{Y}_n$ C^*–algebras, $\mathfrak{X}_i = \mathbb{C}\mathcal{X}_i$, $\mathfrak{A} = \mathfrak{X}_1\otimes\cdots\otimes\mathfrak{X}_n\otimes\mathfrak{Y}_1\otimes\cdots\otimes\mathfrak{Y}_n$. and a probability distribution P on $\mathcal{X}_1\times\cdots\times\mathcal{X}_n$. 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 probability on $\mathcal{X}_1\times\cdots\times\mathcal{X}_n$ and W maps the $\mathcal{X}_i$ to states on $\mathfrak{Y}_i$):

$$I(\mathfrak{X}_1\cdots\mathfrak{X}_n\wedge\mathfrak{Y}_1\cdots\mathfrak{Y}_n) \leq \sum_k I(\mathfrak{X}_k\wedge\mathfrak{Y}_k)$$

Proof. We only have to check the conditional independence, which is left to the reader. $\square$

We note another simple estimate for the mutual information:

Theorem 9 *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 5. $\square$

C. Conditional entropy

Theorem 10 *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 IV.1

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

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 τ_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

$$\begin{aligned} H(\varphi|\psi) &= H(Q) - (H(Q\tau) - \sum_x Q(x)H(\tau_x)) \\ &= H(Q) - I(Q,\tau) \geq 0 \end{aligned}$$

(see section VIII, for the last line theorem VIII.1). □

Note 11 *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:*

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

From this it would follow that in this case $I(\mathfrak{X} \wedge \mathfrak{Y}) \leq \min\{H(\mathfrak{X}), H(\mathfrak{Y})\}$ (see theorem 9), which we now only get from the commutativity assumption.

Definition 12 *Call ρ* separable with respect to *the 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 13 (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 $H(\psi|\varphi) \leq H(\psi|\varphi \circ \varphi')$, 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. theorem 6. □

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

Theorem 14 (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 13 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 15 *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. □

VIII. QUANTUM CHANNELS

A. General remarks

We consider in the following only quantum channels with a priori fixed input states (i.e. *classical–quantum* channels after Holevo [3]). Formally such a system may be described by the collection $(W_x|x \in \mathcal{X})$ of states with W_x appearing at the receiving end when x is sent. This may also be described by its linear extension $W : \mathbb{C}\mathcal{X}_* \to \mathfrak{Y}_*$, a trace preserving quantum operation (this is the only occasion where we omit the subscript $*$ for a quantum map between state spaces).

Side remark: the most general quantum channel appears if we allow at the left any C*–algebra instead of the commutative one. In this case we are free to choose input states — in general from a continuum. Even more, we may (in block coding) use entangled states. For simplicity, and because of some unsolved problems in the more general case we decided here to stay with classical–quantum channels.

This idea of a channel as a process, after choosing a distribution P on $\mathcal{X}$ (i.e. a state on $\mathfrak{X} = \mathbb{C}\mathcal{X}$), which is an average input, leads to the notions of the average output $PW = W(P) =$

$\sum_{x\in\mathcal{X}} P(x)W_x$ and the mutual information $I(P,W) = H(PW) - \sum_{x\in\mathcal{X}} P(x)H(W_x)$. Whereas this is a physically perfectly reasonable model with its appropriate ideas, looking at classical information theory we see that there is also another way of thinking about channels: namely as stochastic two–end systems, one end of which is declared the sender, the other the receiver (even though formally the thing is symmetric), and their respective input and output distributions are marginals of some joint distribution (which reflects the dependence of the output on the input). To model this with quantum systems define the *channel state* $\gamma = \sum_x P(x)x \otimes W_x$ on $\mathfrak{X} \otimes \mathfrak{Y}$. Notice that we (abstractly, and somewhat unnaturally) divided the system into two: its past and its future, and γ describes the correlation between them. Obviously P and PW are obtained as marginals, by tracing over $\mathfrak{Y}$, $\mathfrak{X}$, respectively. In fact it is an easy exercise to verify that $I(P,W) = I(\mathfrak{X} \wedge \mathfrak{Y})$.
This second point of view (and its connection to the first, which was noticed before by Hall [29] in his investigation of what he calls *context mappings*) was the motive for the whole presentation in the preceding sections: to phrase the information and entropy concepts initially defined in the context of processing states via quantum operations in a "static" model that allows for the use of observables (i.e. random variables), and comparison with certain subalgebras.

B. Multiway channels

In the sequel we will also consider a more general channel: we call it the *(all–to–all) quantum multiway channel* with s senders and r receivers (or the *r–fold compound multiple access channel*), and it consists of s commutative C*–algebras $\mathfrak{X}_1, \dots, \mathfrak{X}_s$ (say $\mathfrak{X}_i = \mathbb{C}\mathcal{X}_i$, and let $\mathfrak{X} = \mathfrak{X}_1 \otimes \cdots \otimes \mathfrak{X}_s$), a quantum operation $W : \mathfrak{X} \to \mathfrak{Y}$, and r compatible $*$–subalgebras $\mathfrak{Y}_1, \dots, \mathfrak{Y}_r$ of $\mathfrak{Y}$. The idea here is that the $\mathfrak{X}_i$ are the senders, the $\mathfrak{Y}_j$ the receivers, and each sender wants to send the same message to every receiver, with small error probability. This we formalize in the notion of an $(n, \bar{\epsilon})$–code which consists of s mappings $f_i : \mathcal{M}_i \to \mathcal{X}_i^n$ with finite set $\mathcal{M}_i$, and r decoding observables Y_j, indexed by $\mathcal{M}_1' \times \cdots \times \mathcal{M}_s' \supset \mathcal{M}_1 \times \cdots \times \mathcal{M}_s$, with values in $\mathfrak{Y}_j^{\otimes n}$ (and so these are automatically compatible) with the r (average) error probabilities

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

all being at most $\bar{\epsilon}$. The *rate* of the code is the tuple $(R_1, \ldots, R_s)$ with $R_i = \dfrac{1}{n} \log |\mathcal{M}_i|$. The problem is then to determine the *capacity regions* $\mathbf{R}(\bar{\epsilon})$, i.e. the set of all achievable s–tuples with error probability $\bar{\epsilon}$ (where *achievable* means that for infinitely many n there exist $(n, \bar{\epsilon})$–codes with rate tuples converging to the given tuple), or more realistically $\mathbf{R} = \bigcap_{\bar{\epsilon} > 0} \mathbf{R}(\bar{\epsilon})$ (which is usually called *the* capacity region). Obviously this consists of two parts: first to exhibit the existence of codes with certain rate, second bounds on the rate for any code.

A little history: with classical communication the multiway channel was first considered by Shannon [30], and the exact determination of the capacity region was done by Ahlswede [31,32]. There are of course even more general multi–user communication models, most of which are unsolved: a good overview is in the paper [33] by El Gamal and Cover. Quantum channels for single sender and receiver were all around since the sixties, but the first formal definitions seem to have been given by Holevo [2,3]. The quantum multiway channel as defined here is a slightly smoothed presentation of the definition by Allahverdyan and Saakian [34] (where the channel is a general quantum map).

Before we can tackle this problem (of which we will solve in this paper only the second part, giving bounds) we have to collect a few facts.

The following is a corollary to the information inequality:

Theorem 1 (Holevo bound) *For any classical–quantum channel $W : \mathcal{X} \to \mathfrak{S}(\mathfrak{Y})$, any distribution P on $\mathcal{X}$, and any observable Y on $\mathfrak{Y}$*

$$I(P, W) \geq I(P, Y_* \circ W)$$

More generally, for any completely positive quantum operation $\varphi : \mathfrak{Z} \to \mathfrak{Y}$ one has $I(P, W) \geq I(P, \varphi_* \circ W)$. In particular $I(P, W) \leq I(P, \mathrm{id}) = H(P)$.

Proof. All ingredients are already known: we define a channel state $\gamma = \sum_x P(x)[x] \otimes [x]$ on $\mathbb{C}\mathcal{X} \otimes \mathbb{C}\mathcal{X}$ and observe that $I(P, \mathrm{id}) = I(\mathrm{id}_1 \wedge \mathrm{id}_2) = H(P)$. Now to apply the information bound let $\psi : \mathfrak{Y} \to \mathbb{C}\mathcal{X}$ such that $W = \psi_*$:

$$\begin{array}{ccccc} I(\mathrm{id}_1 \wedge \mathrm{id}_2) & \geq & I(\mathrm{id}_1 \wedge \psi) & \geq & I(\mathrm{id}_1 \wedge \psi \circ \varphi) \\ \| & & \| & & \| \\ H(P) & & I(P, W) & & I(P, \varphi_* \circ W) \end{array}$$

□

The formulation of the Holevo bound is of course in the manner of a data processing inequality, data processing in the sense of composition of two quantum operations. We can also formulate it in the language of observables, just like for classical correlated random variables:

For this consider the following state on $\mathfrak{X} \otimes \mathfrak{Y} \otimes \mathfrak{Z}$

$$\gamma = \sum_{x \in \mathcal{X}} P(x) x \otimes W_x \otimes \varphi_*(W_x)$$

which represents the correlation of the three stages of the system: preparation, reception, and detection of the signal (again note that this is artificial in the material sense). The data processing inequality now is in the familiar form $I(\mathfrak{X} \wedge \mathfrak{Z}) \leq I(\mathfrak{X} \wedge \mathfrak{Y})$. For proof check identity of the information terms with those in the Holevo bound.

We might want to try to imitate the well known classical proof for random variables: by obvious chain rules

$$\begin{aligned} I(\mathfrak{X} \wedge \mathfrak{Y}\mathfrak{Z}) &= I(\mathfrak{X} \wedge \mathfrak{Y}|\mathfrak{Z}) + I(\mathfrak{X} \wedge \mathfrak{Z}) \\ &= I(\mathfrak{X} \wedge \mathfrak{Z}|\mathfrak{Y}) + I(\mathfrak{X} \wedge \mathfrak{Y}) \end{aligned}$$

Since $I(\mathfrak{X} \wedge \mathfrak{Y}|\mathfrak{Z}) \geq 0$ the inequality will be proved if we could show that $I(\mathfrak{X} \wedge \mathfrak{Z}|\mathfrak{Y}) = 0$: but this is not true, as we will show immediately by example! Before we do that however let us discuss our definition of γ. Observe that it not even in the classical case reflects the dependence of $\mathfrak{Z}$ on $\mathfrak{Y}$ correctly: W_x is a sum of pure (deterministic) states, say $W_x = \sum_y W(y|x) V_y$ (classically of course this is unique, and V_y *is* just y), and φ_* invidually transforms these states. Thus a better choice whould be

$$\gamma = \sum_{x,y} P(x)W(y|x)x \otimes V_y \otimes \varphi_*(V_y)$$

Note that this does not change $I(\mathfrak{X} \wedge \mathfrak{Z})$ or $I(\mathfrak{X} \wedge \mathfrak{Y})$. On the other hand the decomposition of W_x is no longer unique in the quantum case. In our example however the W_x will be pure, so there is in fact no question of decomposition:

Example 2 Consider a binary channel, i.e. $\mathcal{X} = \{0,1\}$, $\mathfrak{Y} = \mathfrak{L}(\mathbb{C}^2)$. In $\mathbb{C}^2$ fix an orthonormal basis $|0\rangle, |1\rangle$ and let $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Let $W_0 = |0\rangle\langle 0|$, $W_1 = |+\rangle\langle +|$, and P the uniform distribution.

In the first scenario let $\varphi_* = \mathrm{id}$, so

$$\gamma = \frac{1}{2}[0] \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0| + \frac{1}{2}[1] \otimes |+\rangle\langle +| \otimes |+\rangle\langle +|$$

and a short calculation shows

$$\begin{aligned} H(\mathfrak{X}|\mathfrak{Y}) \quad &= 1 - h(\cos^2 \tfrac{\pi}{8}) \approx .399 \\ H(\mathfrak{X}|\mathfrak{Y}\mathfrak{Z}) &= 1 - h(\cos^2 \tfrac{\pi}{6}) \approx .189 \end{aligned}$$

The difference is easily explained: in the second quantity one has access to two clones of the original state W_x, so identifying x is better possible.

This principle of doubly using quantum information in a forbidden way still is possible even if we insist that φ should be a measurement: in the second scenario φ_* is the external operation of a von Neumann measurement in basis

$$|u\rangle = \cos\frac{\pi}{8}|0\rangle - \sin\frac{\pi}{8}|1\rangle, \quad |v\rangle = \sin\frac{\pi}{8}|0\rangle + \cos\frac{\pi}{8}|1\rangle$$

Thus (with $\alpha = \cos^2\frac{\pi}{8} = \frac{1+\sqrt{1/2}}{2}$)

$$\gamma = \frac{1}{2}[0] \otimes W_0 \otimes (\alpha[0] + (1-\alpha)[1]) + \frac{1}{2}[1] \otimes W_1 \otimes ((1-\alpha)[0] + \alpha[1])$$

and an easy calculation shows (with $\beta = \frac{1+\sqrt{1-2\alpha(1-\alpha)}}{2} = \frac{1+\sqrt{3/4}}{2}$)

$$\begin{aligned} H(\mathfrak{X}|\mathfrak{Y}) \quad &= 1 - h(\alpha) \approx .399 \\ H(\mathfrak{X}|\mathfrak{Y}\mathfrak{Z}) &= 1 - h(\beta) \approx .246 \end{aligned}$$

Again the reason for the failure is the same (which is unknown in the classical theory): in γ we consider states as coexistent which never can coexist, because the third stage evolves from the second by an operation (a measurement) which must needs disturb the system: we neglected this very fact in constructing γ, and we had to: otherwise we could not have incorporated both stages of the evolution, the one after W, and the one after φ_*.

After this digression we turn to an application of the Holevo bound: with the above notation we obtain the result reported in [34]

Theorem 3 (Upper capacity bounds) *The capacity region of the quantum multiway channel is contained in the closure of all nonnegative* $(R_1, \dots, R_s)$ *satisfying*

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

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

Proof. Assume an $(n, \bar{\epsilon})$–code $(f_1, \dots, f_s, Y_1, \dots, Y_r)$. Then the uniform distribution on the codewords induces a channel state γ on $(\mathfrak{X}_1 \cdots \mathfrak{X}_s \mathfrak{Y})^{\otimes n}$. 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 VII.14 (and corollary) we have

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

With

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

we conclude (with theorem VII.7 and corollary)

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

□

Note 4 *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 [31,32].*

Note 5 *The significance of the Holevo bound lies in that we can with it and the Fano inequality derive an upper bound on the capacity of a quantum channel. Holevo [7] and independently Schumacher and Westmoreland [8] recently showed that in the case* $\mathfrak{Y} = \mathfrak{L}(\mathcal{H})$ *this bound can be achieved. In [35] achievability in the case of the multiple access channel* ($r = 1$) *and for general* $\mathfrak{Y}$ *is demonstrated.*
We conjecture that also in the general case of $r > 1$ *the theorem gives already the right capacity region.*

C. Broadcast channels

To end this section let us think a bit about the quantum analog of the *broadcast channel* (see also recent work by Allahverdyan and Saakian [36]): suppose a sender wants to transmit messages from two sets to two receivers — over the same quantum channel (like a TV–station with several programs). Receiver 1 is interested in part 1 of the message, receiver 2 in part 2, both in a common part 0. A model of this situation is a map $W : \mathcal{X} \to \mathfrak{S}(\mathfrak{Y})$ for the channel, two $*$–subalgebras $\mathfrak{Y}_1, \mathfrak{Y}_2$ of $\mathfrak{Y}$ for the two receivers: the triple $(W, \mathfrak{Y}_1, \mathfrak{Y}_2)$ we call a broadcast channel. If these subalgebras are compatible we call the system *plug–and–play* (because then each receiver may choose any observable without interfering with the other. In the other case they may have to agree on compatible observables, or prescribe the order of access to the data). An n–block code for this channel is a triple (f, D_1, D_2) with a map $f : \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}^n$ and compatible observables D_i in $\mathfrak{Y}_i$, indexed by $\mathcal{M}'_0 \times \mathcal{M}'_i \supset \mathcal{M}_0 \times \mathcal{M}_i$ ($i = 1, 2$). The (maximum) error probability of the code is

$$e(f, D_1, D_2) = \max\{1 - \mathrm{Tr}\,(W_{f(m_0,m_1,m_2)} D_{1,m_0m_1} D_{2,m_0m_2}) | m_i \in \mathcal{M}_i,\ i = 0, 1, 2\}$$

(and analogously the average error probability $\bar{e}$). If it is at most ϵ we speak of a (n, ϵ)–code ($(n, \bar{\epsilon})$–code, respectively). The capacity of the code is, as expected, the triple $(R_1, R_0, R_2) =$

$(\frac{1}{n}\log|\mathcal{M}_1|, \frac{1}{n}\log|\mathcal{M}_0|, \frac{1}{n}\log|\mathcal{M}_2|)$, and the problem is to determine the capacity region. This is a problem exceedingly difficult, not even solved completely in the classical case ($\mathfrak{Y}$ commutative).

Thus we may consider a restricted situation, which has in the classical case a complete solution: the *degraded* broadcast channel: here the line to receiver 2 "factors" through 1, i.e. the degraded broadcast channel is a triple $(W, \varphi_*, \mathfrak{Y}_1)$ with W as above, and a quantum operation $\varphi_* : \mathfrak{Y}_* \to \mathfrak{Y}_{2*}$. Receiver 1 is the $*$–subalgebra $\mathfrak{Y}_1$ of $\mathfrak{Y}$, receiver 2 the algebra $\mathfrak{Y}_2$. This links with the previous explanation via the definitions $W_2 = \varphi_* \circ W$, $W_1 = \imath_* \circ W$ (for the inclusion $\imath : \mathfrak{Y}_1 \hookrightarrow \mathfrak{Y}$). This however gives not the correct picture because this model is manifestly not plug–and–play: The second receiver has to take what the first left to him. Formally: an n–block code $(f, \tilde{D}_1, D_2)$ now consists of f as before, and also the observable D_2 on $\mathfrak{Y}_2^{\otimes n}$, and a subtle modification of D_1 to the operation $\tilde{D}_1 = \psi_*$ for a quantum operation

$$
\begin{aligned}
\psi : \mathbb{C}\mathcal{M}_0' \otimes \mathbb{C}\mathcal{M}_1' \otimes \mathfrak{Y}^{\otimes n} &\to \mathfrak{Y}^{\otimes n} \\
m_0 \otimes m_1 \otimes A &\mapsto E_{m_0 m_1}^* A E_{m_0 m_1}
\end{aligned}
$$

with $E_{m_0 m_1} \in \mathfrak{Y}_1^{\otimes n}$. Obviously $\mathrm{Tr}_{\mathfrak{Y}^{\otimes n}} \circ \tilde{D}_1 = D_1$ for the observable D_1 indexed by $\mathcal{M}_0' \times \mathcal{M}_1'$ and consisting of the operators $D_{1,m_0 m_1} = E_{m_0 m_1} E_{m_0 m_1}^*$. With this we can formulate the error probability:

$$
e(f, \tilde{D}_1, D_2) = \max\{1 - \mathrm{Tr}\,(E_{m_0 m_1}^* W_{f(m_0,m_1,m_2)} E_{m_0 m_1} D_{2,m_0 m_2}) | m_i \in \mathcal{M}_i,\ i = 0,1,2\}
$$

(analogous for the average error probability). In direct analogy with the classical situation we present the following

Conjecture 6 *For $\mathfrak{Y}_1 = \mathfrak{Y}$ the rate region is the convex hull of the triples (R_1, R_0, R_2) with*

$$
\begin{aligned}
R_1 &\leq I(V(\cdot|u), W|Q) \\
R_0 + R_2 &\leq I(Q, \varphi_* \circ W \circ V) \\
R_1 + R_0 + R_2 &\leq I(QV, W)
\end{aligned}
$$

where Q is a distribution on a finite set $\mathcal{U}$, and V a classical channel from $\mathcal{U}$ to $\mathcal{X}$.

D. Open problems

Note 7 *Meaning of theorem VII.3 for coding theorems: The reason why for truely quantum channels one has strict inequality is that we cannot detect the W_x optimally in one common basis (for simplicity assume that we only employ von Neumann measurements). Assume we chose an eigenbasis of PW, then we "see" correctly the entropy $H(PW)$ of the output state, but for the letter states we introduce some additional entropy to their $H(W_x)$. Thus we get to low a mutual information because our measurements introduce noise. We want this noise increase to be small by choosing codewords appropriately, and then "approximating" with a von Neumann observable, all the codeword states nearly commute with. The problem here is to do this such that the von Neumann mutual information remains the same.*
Note that this is a different approach to coding than those used so far: there we directly construct codes approaching certain rate, using general observables. Here we would have a von Neumann observable approaching the Holevo bound, i.e. a classical channel for which we may construct codes by the known classical techniques.

Note 8 *For classical–quantum channels there does not appear to exist a reasonable notion of* transpose *channel. If however we see a channel as a quantum map from any one system to another, then given an input state one can define formally a transpose channel under certain circumstances, see [17]. This goes the opposite direction as the original channel, so in our case we get a measurement operation. It is to be explored whether this notion gives us new insight in the communication problem. In particular we may relate the classical–quantum channels with quantum–classical channels (i.e. fixed measurements, or if variable only product measurements). Maybe we can even prove that coding classical information with entangled states in quantum–quantum channels yields higher capacities...*

ACKNOWLEDGMENTS

Thanks to Peter Löber for discussions during the course of this work, especially for pointing out to me the importance of strong subadditivity. Thanks also to MJW Hall for drawing my attention to his work.

REFERENCES

[1] The reader should consult W. T. Grandy, Jr., "Resource letter ITP-1: Information Theory in Physics", Am. J. Phys. **65**,4(1997), 466–476, and A. S. Holevo, "Coding Theorems for Quantum Communication Channels", LANL eprint `quant-ph/9708064`. Also C. M. Caves, P. D. Drummond, "Quantum limits on bosinic communication rates", Rev. Mod. Phys. **66**,1(1994), 481–538 and H. P. Yuen, M. Ozawa, "Ultimate information carrying limit of quantum systems", Phys. Rev. Letters **70**,4(1993), 363–366 contain historical comments.

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

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

[4] A. S. Holevo, Problemy Peredachi Informatsii **15**,4(1979), 3–11 (english translation: "Capacity of a quantum communication channel", Probl. Inf. Transm. **15**,4(1979), 247–253)

[5] C. Adami, N. J. Cerf, "On the von Neumann capacity of noisy quantum channels", Phys. Rev. A **56**,5(1997), 3470–3483

[6] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, W. K. Wootters, "Classical information capacity of a quantum channel", Phys. Rev. A **54**,3(1997), 1869–1876

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

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

[9] C. H. Bennett, G. Brassard, "Quantum cryptography: Public key distribution and coin tossing", in: *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India*, New York 1984, 175–179

[10] D. Mayers, A. Yao, "Unconditional security in Quantum Cryptography", LANL eprint `quant-ph/9802025`

[11] C. H. Bennett, S. Wiesner, "Communication via One– and Two–Particle Operators on Einstein–Podolsky–Rosen States", Phys. Rev. Letters **69**,20(1992), 2881–2884

[12] W. Arveson, *An invitation to C^*–algebras*, New York, Heidelberg 1976

[13] J. Dixmier, *C^*–algebras*, Amsterdam, New York 1977

[14] E. B. Davies, *Quantum Theory of Open Systems*, London 1976

[15] K. Kraus, *States, Effects, and Operations*, Springer Lecture Notes in Physics 190, Berlin 1983

[16] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, Amsterdam 1982

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

[18] R. Schatten, *Norm Ideals of Completely Continuous Operators*, Berlin, Göttingen, Heidelberg 1960

[19] J. S. Bell, "On the Einstein Podolsky Rosen paradox", Physics **1**,3(1964), 195–200

[20] A. Peres, *Quantum Theory: Concepts and Methods*, Dordrecht 1995

[21] W. F. Stinespring, "Positive Functions on C*–Algebras", Proc. Amer. Math. Soc. **6**(1955), 211–216

[22] J. von Neumann, "Thermodynamik quantenmechanischer Gesamtheiten", Gött. Nachr., 273–291

[23] L. B. Levitin, "Conditional Entropy and Information in Quantum Systems", in: Proc.

ISIT 1998, 88

[24] H. Umegaki, "Conditional expectations in an operator algebra IV (entropy and information)", Kodai Math. Sem. Rep. **14**(1962), 59–85

[25] A. Wehrl, "General properties of entropy", Rev. Mod. Phys. **50**,2(1978), 221–260

[26] A. Uhlmann, "Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory", Commun. Math. Phys. **54**(1977), 21–32

[27] G. Lindblad, "Completely positive maps and entropy inequalities", Commun. Math. Phys. **40**(1975), 147–151, "Expectations and entropy inequalities for finite quantum systems", **39**(1974), 111–119, also: "Entropy, information, and quantum measurements", **33**(1973), 305–322

[28] B. Schumacher, "Sending entanglement through noisy quantum channels", Phys. Rev. A **54**,4(1996), 2614–2628

[29] M. J. W. Hall, "Quantum information and correlation bounds", Phys. Rev. A **55**,1(1997), 100–113, also contribution to *Quantum Communication, Computing and Measurement* (O. Hirota et. al. eds.), New York 1997, 53–61

[30] C. E. Shannon, "Two–way communication channels", in: *Proc. Fourth Berkeley Symposium Probability and Statistics* (ed. J. Neyman), Berkeley 1961, 611–644. Reprinted in: *Claude Elwood Shannon Collected Papers* (eds. N. J. A. Sloane, A. D. Wyner), New York 1993, 351–384

[31] R. Ahlswede, "Multi–way communication channels", in: *Second International Symposium on Information Theory*, Hungarian Academy of Sciences 1971, 23–52

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

[33] A. El Gamal, T. Cover, "Multiple User Information Theory", Proc. IEEE **68**(1980),

1466–1483

[34] A. E. Allahverdyan, D. B. Saakian, "Multi–access channels in quantum information theory", LANL eprint `quant-ph/9712034`

[35] A. Winter, "The Capacity of the Quantum Multiple Access Channel", LANL eprint `quant-ph/9807019`

[36] A. E. Allahverdyan, D. B. Saakian, "The broadcast quantum channel for classical information transmission", LANL eprint `quant-ph/9805067`