Vol. 10 (1976) REPORTS ON MATHEMATICAL PHYSICS No. 1

QUANTUM INFORMATION THEORY

ROMAN S. INGARDEN

Institute of Physics, Nicholas Copernicus University, 87-100 Toruń, Poland

(*Received December 13, 1975*)

A conceptual analysis of the classical information theory of Shannon (1948) shows that this theory cannot be directly generalized to the usual quantum case. The reason is that in the usual quantum mechanics of closed systems there is no general concept of joint and conditional probability. Using, however, the generalized quantum mechanics of open systems (A. Kossakowski 1972) and the generalized concept of observable ("semiobservable", E. B. Davies and J. T. Lewis 1970) it is possible to construct a quantum information theory being then a straightforward generalization of Shannon's theory.

1. Introduction

Information theory, as it is understood in this paper and as it is usually understood by mathematicians and engineers following the pioneer paper of Shannon [57], is not only a theory of the entropy concept itself (in this aspect information theory is most interesting for physicists), but also a theory of transmission and coding of information, i.e., a theory of information sources and channels. In the case of classical (i.e., non-quantum) systems both parts of the theory are closely connected, this connection being actually accomplished in probability theory forming a theoretical background of information theory. The clue concepts are those of joint and conditional probability which enable to formulate the definition of information sources and channels and then of the concept of channel capacity which is the most important for Shannon's coding theorems. In the case of quantum systems the probability theory has to be essentially generalized, cf. [4], [18], [63], [64], [65], and the concepts of joint and conditional probability appear to be very specialized, i.e., cannot be defined for the general case, except in principle for the trivial case, i.e., for the commuting observables; for details cf. [63]. This circumstance caused that when it was necessary to consider communication problems for the needs of quantum electronics and optics, this had to be done by other means, not those of Shannon's information theory, e.g., by the so-called signal detection theory, cf. [24], [25], by Fourier analysis methods, cf. [13], [14], or by combination of these methods with some other heuristic physical arguments, cf. [21], [61], [36]. Finally, however, it has been realized that, because of the lack of the quantitative concept of information in these methods, they cannot fully answer the practical needs. The existing solutions of information problems were only partial or special, cf. [58], [59], [42], [44], and it was felt that some rather radical generalization of the theoretical frame of quantum probability theory is needed. Of course, to do this we have actually to generalize the quantum mechanics itself. Recently it has been realized that the conventional quantum mechanics, although probabilistic "from the beginning", is not a general quantum probabilistic theory, i.e., it describes dynamical, but not stochastical processes in a closed (isolated) system. The only stochastic element in the theory is connected with considering the outside measuring instrument, macroscopical in character, which introduces an unexpected disturbance of the system called a "reduction of the wave packet". This reduction can be predicted only statistically as has been recognized by Max Born in his statistical interpretation of quantum mechanics. When a result of the measurement is given, the quantum description is interrupted and we have to begin it anew with a new wave function (quantum state). Thus quantum mechanics describes only a physical process (time evolution of the state) between two consecutive measurements, it cannot cross the measuring process itself to the next ones. Therefore, we cannot speak about a quantum stochastic process (consisting of a sequence of random measurements or random events) in the frame of the usual quantum mechanics. The only extrapolation of this theory known in literature was the so-called "principle of repeatability" of von Neumann [67] which says that an immediate repetition of a measurement gives the same result. It was felt, however, that this principle is too special and it was frequently discussed and criticized, cf. [43], [19], [12].

The foundations of a new "stochastic" quantum mechanics of open systems was laid from two independent sides. On the one hand, in 1961 E. C. G. Sudarshan, P. Mathews and J. Rau published a paper [60] in which they formulated the first general idea of "stochastic quantum dynamics", rather in connection with elementary particles physics. Independently, in 1972 A. Kossakowski formulated a more specific theory of quantum "dynamical semigroups" or non-Hamiltonian quantum statistical mechanics of open system [39], cf. also [32], based on the theory of Banach spaces and motivated rather by the laser theory. On the other hand, since the famous experiments of Hanbury Brown and Twiss of 1954 [23] a new field of "photon statistics" appeared and quickly developed, cf. [7], [37], intensified by the invention of lasers and all "quantum electronics" and "quantum optics", as well as by Glauber's idea of "coherent states" [16]. E. B. Davies [6] rightly recognized that for describing such processes as photon counts, etc., a new generalized theory of quantum stochastic processes is needed and he formulated such a theory, cf. also [8], [7], where a generalized concept of observable, as well as that of "instrument", and generalized theory of repeated measurements have been given.

The fact that Davies' concepts and theorems are fundamental for the construction of quantum information theory has been realized by A. S. Holevo [27], [28]. He has shown how to solve with these tools some quantum information problems, and explained also the connection of Davies' theory with Glauber's method of coherent states. Holevo did not formulate, however, any general scheme of quantum information theory as a generalization of Shannon's theory. The purpose of the present paper is to give such a general scheme.

2. Basic definitions and notation

Let us associate, as usually, with a quantum system a complex separable Hilbert space $\mathscr{H}$. Elements of $\mathscr{H}$ will be denoted by $x, y, z, \ldots$, and the scalar product in $\mathscr{H}$ by $(\cdot, \cdot)$. Let $L^1(\mathscr{H})$ be the real Banach space of self-adjoint trace class operators on $\mathscr{H}$. Elements of $L^1(\mathscr{H})$ will be denoted by $\varrho, \sigma, \tau, \ldots$, and the (trace) norm by

$$\|\varrho\|_1 := \sup_{\{x_n\}\{y_n\}} \sum_{n=1}^{\infty} |(x_n, \varrho y_n)|, \tag{2.1}$$

where $\{x_n\}$ and $\{y_n\}$ are any complete orthonormal bases in $\mathscr{H}$, cf. [15], [39]. The *positive cone* in $L^1(\mathscr{H})$, i.e. the set of all positive definite operators from $L^1(\mathscr{H})$, will be denoted by $L^1_+(\mathscr{H})$.

We remark that

$$\text{for all } \varrho \in L^1_+(\mathscr{H}): \quad \|\varrho\|_1 = \operatorname{Tr}\varrho, \tag{2.2}$$

$$\begin{aligned} \text{for all } \varrho \in L^1(\mathscr{H}): \quad \|\varrho\|_1 &= \operatorname{Tr}|\varrho| := \operatorname{Tr}(\varrho^*\varrho)^{1/2} \\ &= \operatorname{Tr}(\varrho^2)^{1/2}. \end{aligned} \tag{2.3}$$

The Banach space $L^1(\mathscr{H})$ is the least linear space in which the convex set

$$\Delta(\mathscr{H}) := \{\varrho \in L^1_+(\mathscr{H}) \colon \|\varrho\|_1 = 1\} \tag{2.4}$$

of all *states* (mixed states, normal states, density operators) is embedded.

The set of all real-valued continuous linear functionals on $L^1(\mathscr{H})$, $\langle A, \varrho\rangle$, $\varrho \in L^1(\mathscr{H})$, is called the *dual space* to $L^1(\mathscr{H})$ and will be denoted by $L^\infty(\mathscr{H}) := L^1(\mathscr{H})^*$. Accordingly, we introduce the *dual cone* $L^\infty_+(\mathscr{H})$ in $L^\infty(\mathscr{H})$ as the set of all elements $A \in L^\infty(\mathscr{H})$ such that $\langle A, \varrho\rangle \geqslant 0$ for all $\varrho \in L^1_+(\mathscr{H})$. Since it is known [10] that any real continuous linear functional on $L^1(\mathscr{H})$ can be presented in the form

$$\langle A, \varrho\rangle = \operatorname{Tr}(A\varrho), \qquad \varrho \in L^1(\mathscr{H}),\ A \in \boldsymbol{B}(\mathscr{H}),\ A = A^*, \tag{2.5}$$

where $\boldsymbol{B}(\mathscr{H})$ is the Banach space of all bounded linear operators on $\mathscr{H}$, we see that $L^\infty_+(\mathscr{H})$ consists of all positive definite bounded linear operators on $\mathscr{H}$, i.e., is a positive cone in $L^\infty(\mathscr{H})$. We remark that, if $\mathscr{H}$ is infinite-dimensional, $L^1(\mathscr{H})$ is *not reflexive*, i.e., $L^1(\mathscr{H})^{**} \neq L^1(\mathscr{H})$, we can only write $L^1(\mathscr{H}) \subset L^1(\mathscr{H})^{**} = L^\infty(\mathscr{H})^*$ by the so-called canonical embedding.

On the other hand, we may generalize (2.4) to

$$\tilde{\Delta}(\mathscr{H}) := \{\varrho \in L^1_+(\mathscr{H}) \colon 0 < \|\varrho\|_1 \leqslant 1\} \tag{2.6}$$

called the set of *incomplete* or *partial states* (in particular, those with $\|\varrho\|_1 = \operatorname{Tr}\varrho < 1$); cf. [51], p. 460.

The dual space $L^\infty(\mathscr{H})$ is only a proper part of the set $S(\mathscr{H})$ of all *observables* on $\mathscr{H}$ or of *self-adjoint*, not necessarily bounded linear *operators* on $\mathscr{H}$ defined by their spectral representations

$$A = \int_{-\infty}^{+\infty} \lambda a(d\lambda), \tag{2.7}$$

where $a(\cdot)$ is a spectral measure on the real axis $\boldsymbol{R}$. We recall that a *spectral measure* on $\boldsymbol{R}$ can be defined as a function a: $\mathscr{B}(\boldsymbol{R}) \to \boldsymbol{B}(\mathscr{H})$ (where $\mathscr{B}(\boldsymbol{R})$ is the Baire algebra[1] on $\boldsymbol{R}$) such that

(1) $a(E) \geqslant 0 \quad$ for all $E \in \mathscr{B}(\boldsymbol{R}) \quad$ (*positivity*), (2.8)

(2) $a\big(\bigcup_{n=1}^{\infty} E_n\big) = \sum_{n=1}^{\infty} a(E_n) \quad$ for all $\{E_n\}_{n=1}^{\infty} \subset \mathscr{B}(\boldsymbol{R})$ such that

$E_n \cap E_m = \emptyset \quad$ for all $n \neq m \quad$ (σ-*additivity*), (2.9)

(3) $a(\boldsymbol{R}) = I :=$ identity operator on $\mathscr{H}$ (*normalization*), (2.10)

(4) $\big(a(E)\big)^2 = a(E) \quad$ for all $E \in \mathscr{B}(\boldsymbol{R}) \quad$ (*projectivity*), (2.11)

The dual space $L^\infty(\mathscr{H})$ is a Banach space with the (operator) norm

$$||A||_\infty := \sup_{||\varrho||_1 = 1} |\langle A, \varrho\rangle| = \sup_{\substack{x\in\mathscr{H} \\ x\neq 0}} \frac{||Ax||}{||x||}, \tag{2.12}$$

where $||x||^2 := (x, x)$ for all $x \in \mathscr{H}$.

If T is a linear operator on $L^1(\mathscr{H})$ (called a *superoperator* on $\mathscr{H}$), we define its norm by

$$||T||_1 := \sup_{||\varrho||_1 = 1} ||T\varrho||. \tag{2.13}$$

We denote the set of all linear operators on and into $L^1(\mathscr{H})$ by $K(\mathscr{H})$, and the set of all bounded linear operators on $L^1(\mathscr{H})$, which is a Banach space with the norm (2.13), by $\boldsymbol{L}^1(\mathscr{H})$. To every $T \in \boldsymbol{L}^1(\mathscr{H})$ there corresponds one and only one linear operator T^* on $L^\infty(\mathscr{H})$ called the *dual superoperator* to T and defined by the identity

$$\langle A, T\varrho\rangle = \langle T^*A, \varrho\rangle \quad \text{for all } A \in L^\infty(\mathscr{H}) \text{ and all } \varrho \in L^1(\mathscr{H}). \tag{2.14}$$

We have [71]

$$||T^*||_\infty := \sup_{||A||_\infty = 1} ||T^*A||_\infty = ||T||_1 . \tag{2.15}$$

All dual superoperators T^* to any $T \in \boldsymbol{L}^1(\mathscr{H})$ form the *second dual space* $\boldsymbol{L}^\infty(\mathscr{H})$ being a Banach space with the norm (2.15).

---

[1] The Baire algebra, and not the (natural) Borel algebra as usual, is taken here only for mathematical convenience. The *Baire algebra* on $\boldsymbol{R}$ is the least σ-algebra containing every compact set which is the intersection of a countable number of open subsets of $\boldsymbol{R}$. The Baire algebra on the compactified real axis $\bar{\boldsymbol{R}}$ coincides with the Borel algebra on $\bar{\boldsymbol{R}}$; cf. [71], p. 18.

A superoperator $T \in K(\mathscr{H})$ mapping $L^1_+(\mathscr{H})$ into itself and such that

$$\|T\varrho\|_1 = \|\varrho\|_1 \quad \text{for all } \varrho \in L^1_+(\mathscr{H}) \tag{2.16}$$

is said to be a *positive linear endomorphism* of $L^1(\mathscr{H})$ or a *positive superoperator* on $\mathscr{H}$, cf. [39]. A positive linear endomorphism T of $L^1(\mathscr{H})$ maps $\Delta(\mathscr{H})$ into itself, is a *contracting operator*, i.e.,

$$\|T\varrho\|_1 \leqslant \|\varrho\|_1 \quad \text{for all } \varrho \in L^1(\mathscr{H}), \tag{2.17}$$

and satisfies the relation

$$\mathrm{Tr}(T\varrho) = \mathrm{Tr}\,\varrho \quad \text{for all } \varrho \in L^1(\mathscr{H}), \tag{2.18}$$

cf. [39]. Iff a superoperator $T \in K(\mathscr{H})$ maps $L^1_+(\mathscr{H})$ into itself and satisfies only (2.17), we shall call it an *incompletely (partially) positive linear endomorphism* of $L^1(\mathscr{H})$ (*incompletely positive superoperator on* $\mathscr{H}$). It maps $\tilde{\Delta}(\mathscr{H})$ into itself.

3. Quantum information

Let us consider an arbitrary state $\varrho \in \Delta(\mathscr{H})$. Since ϱ as an operator is completely continuous (compact), positive definite and with unit trace, it has a purely point spectrum $\{p_n\}_{n=1}^{\infty}$ such that

$$p_n \geqslant 0 \quad \text{for all } n \in N = \{1, 2, \ldots\}, \qquad \sum_{n=1}^{\infty} p_n = 1, \tag{3.1}$$

cf. [15]. Denoting the eigenvectors of ϱ by x_n, $(x_n, x_m) = \delta_{nm}$ (Kronecker's delta), we may write

$$\varrho x = \sum_{n=1}^{\infty} p_n(x, x_n)x_n \quad \text{for all } x \in \mathscr{H} \tag{3.2}$$

or

$$\varrho = \sum_{n=1}^{\infty} p_n P_n, \quad \text{where} \quad P_n := (\,\cdot\,, x_n)x_n,\ n \in N. \tag{3.3}$$

P_n are projectors on one-dimensional subspaces of $\mathscr{H}$,

$$P_n^2 = P_n, \quad P_n \geqslant 0, \quad \mathrm{Tr}\,P_n = 1 \quad (n \in N), \tag{3.4}$$

and are called *pure states* (represented also by vectors x_n).

Von Neumann, [66], [67], V.2, introduced the concept of *entropy* (*information*) of a quantum (mixed) state ϱ (3.3) as

$$H(\varrho) := -\mathrm{Tr}(\varrho \ln \varrho) = -\sum_{n=1}^{\infty} p_n \ln p_n. \tag{3.5}$$

Since all terms in the sum (3.5) are non-negative (we put $0 \cdot \ln 0 := 0$), $0 \leqslant H(\varrho) < \infty$ when the sum is convergent or $H(\varrho) = +\infty$ when it is divergent. We call a state ϱ with a finite $H(\varrho)$ *thermodynamically regular*, one with infinite $H(\varrho)$ *thermodynamically irregular*. We see that all pure states have 0 entropy,

$$H(P_n) = 0 \quad \text{for all } n \in \boldsymbol{N}. \tag{3.6}$$

Conversely, if entropy $H(\varrho) = 0$, ϱ is a pure state. Each pure state is, of course, thermodynamically regular. If $\mathscr{H}$ is finite-dimensional, every state is thermodynamically regular.

The existence of thermodynamically irregular states for general $\mathscr{H}$ is a rather unpleasant feature, the more that they form a set of the second category in $\Delta(\mathscr{H})$, while the set of regular states is *meager* or *of the first category*, i.e., is a union of a countable family of nowhere dense sets; a set is *nowhere dense* in a topological space X iff the interior of its closure is void, and it is *of the second category* iff it is not meager in X, cf. [68], [69]. This fact is the reason why $H(\varrho)\colon \Delta(\mathscr{H}) \to \bar{\boldsymbol{R}}$ is discontinuous in ϱ with respect to the trace norm (2.1) and is unbounded in every open ball $\{\varrho \in \Delta(\mathscr{H})\colon \operatorname{Tr}|\varrho - \varrho'| < \varepsilon,\ \varepsilon > 0\}$; cf. [68], [69]. $H(\varrho)$ is, however, weakly lower semi-continuous in ϱ, [68], [69], and this is, perhaps, sufficient for its physical interpretation.

The entropy $H(\varrho)$ is a measure for the degree of "mixedness" of the state ϱ, it also measures the information necessary to select an eigenstate of ϱ. But usually we do not measure directly ϱ, but some physical observable $A \in S(\mathscr{H})$ in a given state ϱ, cf. (2.7). Then the probability of finding the result of measurement in a subset $E \in \mathscr{B}(\boldsymbol{R})$ is

$$p(A, \varrho, E) := \operatorname{Tr}\big(\varrho a(E)\big). \tag{3.7}$$

Because of (2.8)–(2.10) we have, as should be,

$$p(A, \varrho, E) \geqslant 0 \quad \text{for all } E \in \mathscr{B}(\boldsymbol{R}), \quad p(A, \varrho, \boldsymbol{R}) = 1, \tag{3.8}$$

and $p(A, \varrho, E)$ is σ-additive in $\mathscr{B}(\boldsymbol{R})$. If A is *thermodynamically regular* [31], i.e., has a purely point spectrum

$$A = \int_{-\infty}^{\infty} \lambda a(d\lambda) = \sum_{n=1}^{\infty} \lambda_n P_n, \quad a(E) = \sum_{\lambda_n \in E} P_n \quad \text{for all } E \in \mathscr{B}(\boldsymbol{R}), \tag{3.9}$$

and there exists a real constant c such that

$$\sum_{n=1}^{\infty} \exp(c\lambda_n) < \infty, \tag{3.10}$$

then we can define the *entropy* (*information*) *of observable A in state ϱ* or *A-information in state ϱ*

$$\begin{aligned} H_A(\varrho) = H(A, \varrho) &:= -\sum_{n=1}^{\infty} p(A, \varrho, \lambda_n) \ln p(A, \varrho, \lambda_n) \\ &= -\sum_{n=1}^{\infty} \operatorname{Tr}(\varrho P_n) \ln \operatorname{Tr}(\varrho P_n), \end{aligned} \tag{3.11}$$

cf. [33] (where only pure states were considered). Thus there exists a non-empty class $\Delta_A(\mathscr{H}) \subset \Delta(\mathscr{H})$ that

$$H(A, \varrho) < \infty \quad \text{for all } A \in S_{\text{th}}(\mathscr{H}) \text{ and all } \varrho \in \Delta_A(\mathscr{H}), \tag{3.12}$$

where $S_{\text{th}}(\mathscr{H}) \subset S(\mathscr{H})$ denotes the set of all thermodynamically regular observables. In the following we shall assume that all the states considered for measurement of an observable $A \in S_{\text{th}}(\mathscr{H})$ belong to $\Delta_A(\mathscr{H})$.[2]

An example of a thermodynamically regular observable is the energy of a linear harmonic oscillator of frequency ω and mass $m = 1$ with the spectrum

$$\lambda_n = (n+\tfrac{1}{2})\hbar\omega \quad (n = 0, 1, 2, \ldots). \tag{3.13}$$

If we take for ϱ the *Gibbs ensemble* with temperature T, i.e., we put in (3.3) for x_n the eigenvectors belonging to λ_n (3.13) and

$$p_n = (1-e^{-\alpha})e^{-\alpha n} \quad (n = 0, 1, 2, \ldots), \tag{3.14}$$

where α is the dimensionless constant

$$\alpha := \frac{\hbar\omega}{kT} > 0 \tag{3.15}$$

(as usually, $\hbar$ denotes Planck's constant and k Boltzmann's constant), we easily obtain

$$p(A, \varrho, \lambda_n) = \operatorname{Tr}(P_n \varrho) = p_n \quad (n = 0, 1, 2, \ldots). \tag{3.16}$$

Then

$$H(A, \varrho) = H(\varrho) = -\ln(1-e^{-\alpha}) + \alpha e^{-\alpha}(1-e^{-\alpha})^{-1}. \tag{3.17}$$

The first term of (3.17) is identical with the maximum information when mean energy of the oscillator is fixed in the class of pure states as calculated in [33], equation (95), while the second term is the addition to this expression when we maximize in the class of all mixed states. Denoting

$$H_1(\alpha) := -\ln(1-e^{-\alpha}), \qquad H_2(\alpha) := \alpha e^{-\alpha}(1-e^{-\alpha})^{-1} \tag{3.18}$$

and their sum by $H(\alpha)$, we have

$$\begin{aligned} &\lim_{\alpha \downarrow 0} H(\alpha) = \lim_{\alpha \downarrow 0} H_1(\alpha) = \infty, \qquad \lim_{\alpha \downarrow 0} H_2(\alpha) = 1, \\ &\lim_{\alpha \uparrow \infty} H(\alpha) = \lim_{\alpha \uparrow \infty} H_1(\alpha) = \lim_{\alpha \uparrow \infty} H_2(\alpha) = 0, \end{aligned} \tag{3.19}$$

[2] *Note added in the proof*: In the following paper of the author (this issue, p. 131) another more elegant approach is proposed in connection with the generalization of the Kolmogorov–Sinai entropy to the quantum case by E. Sąsiada (this issue, p. 129).

while

$$\lim_{\alpha \downarrow 0} \frac{H_2(\alpha)}{H_1(\alpha)} = 0, \qquad \lim_{\alpha \uparrow \infty} \frac{H_2(\alpha)}{H_1(\alpha)} = \infty . \tag{3.20}$$

Both functions, $H_1(\alpha)$ and $H_2(\alpha)$, are positive and strongly monotonically decreasing, when α increases from 0 to ∞, and they are equal, $H_1(\alpha) = H_2(\alpha)$ for

$$\alpha = \alpha_0 \approx 0.7; \tag{3.21}$$

cf. Fig. 1. Since $H_1(\alpha)$ prevails for $\alpha \ll 1$ (much smaller than 1) or $\hbar\omega \ll kT$, i.e., in the classical region, H_1 can be called the *classical part* of H. Analogously, since H_2 prevails for $\alpha \gg 1$ ($\hbar\omega \gg kT$) or in the quantum region, H_2 can be called the *quantum part* of H.

Generalizing (3.17) we remark that if $A = f(\varrho)$ and hence for $x \in D_A$ (the domain of A)

$$[A, \varrho]x := (A\varrho - \varrho A)x = 0, \tag{3.22}$$

we can choose for x_n in (3.3) as a development of ϱ the eigenvectors of A and therefore

$$H(A, \varrho) = H(\varrho). \tag{3.23}$$

If A and ϱ do not commute, $H(A, \varrho)$ and $H(\varrho)$ can be different,

$$H(A, \varrho) \geqslant H(\varrho) \tag{3.23a}$$

(O. Klein's theorem).

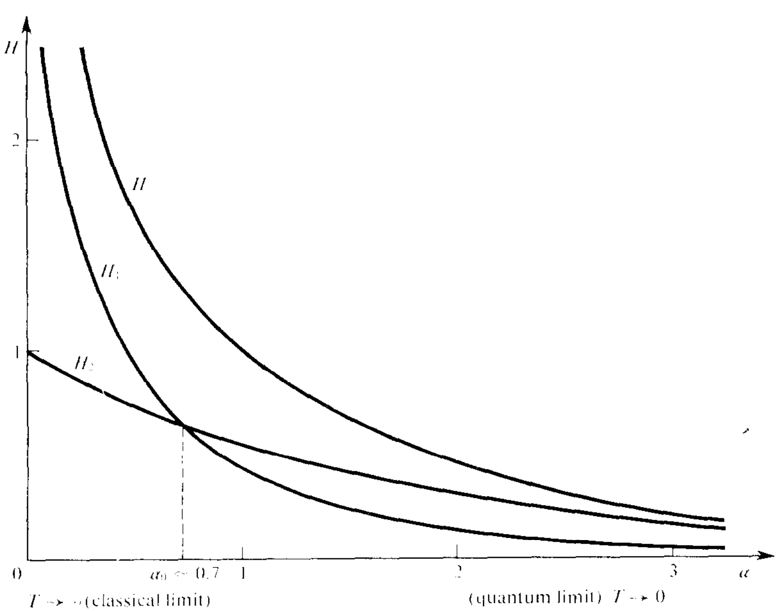

Fig. 1. Classical part H_1 and quantum part H_2 of information H of energy of a linear harmonic oscillator and the Gibbs ensemble as functions of inverse temperature

There are generalizations of the quantum entropy concept, cf. [55], [56], [22], [54], corresponding roughly to the relative and continuous entropies in the classical case, as well as axiomatic approaches to this concept, cf. [31], [49]. For the purposes of the present paper, however, we do not need these constructions.

4. Classical information channels

For the convenience of the reader and easier understanding we briefly recapitulate here the essential points of the classical Shannon information theory of communication channels with discrete messages [57], [33], [11], [70], [3], [26]. There is also a generalization of this theory for continuous messages by Kolmogorov [38] and Dobrushin [9], but for the sake of simplicity we do not go here into this generalization (its quantum counterpart would require the general concept of quantum entropy mentioned at the end of the previous section).

We use Kolmogorov's scheme of classical probability space $(\Omega, \mathscr{B}(\Omega), p)$, where Ω is a set of elementary events, $\mathscr{B}(\Omega)$ a σ-algebra of subsets of Ω (algebra of events, usually the Baire or Borel algebra on Ω), and p a probability measure on $\mathscr{B}(\Omega)$. Let $\boldsymbol{Z}$ be the set of all integers $\{\ldots, -1, 0, 1, 2, \ldots\}$. We consider the usual block scheme of an information transmitting arrangement, cf. Fig. 2, as composed of an input source, a channel, an output source and a noise source (the latter representing the surroundings giving undesired information). The channel and noise are treated together by means of the so-called *double source input-output* and by taking the supremum over the input, as will be explained below.

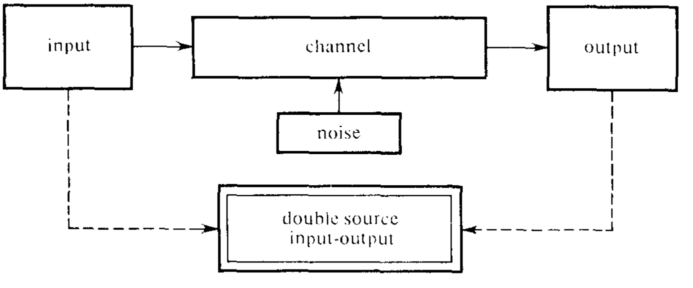

Fig. 2. The block scheme of an information transmitting arrangement

We assume that each unit of time[3] (second, microsecond, etc., depending on the arrangement), $t \in \boldsymbol{Z}$, in the input and output sources some objects (called *symbols* or *letters*), taken from a given finite set called *alphabet*, appear with some given probabilities. We have the following scheme:

[3] Time is only one of possible interpretations of the parameter t. In optical instruments where optical images are transmitted, t may be considered as a spacial coordinate (or coordinates), cf. [20]. Thus $\boldsymbol{Z}$ may be generalized to the many-dimensional array $\boldsymbol{Z}^n = \boldsymbol{Z} \times \ldots \times \boldsymbol{Z}$ (n times).

Concept	Input	Output
Alphabet	$A = \{x_1, \dots, x_a\}$	$B = \{y_1, \dots, y_b\}$
Set of sequences of $n \in N$ letters (*n-words*)	$\Omega_A^{(n)} = A^n = \underbrace{A \times \dots \times A}_{n}$	$\Omega_B^{(n)} = B^n = \underbrace{B \times \dots \times B}_{n}$
Set of doubly infinite sequences of *Z-words*	$\Omega_A = A^Z$	$\Omega_B = B^Z$
Probability measure of n-words	$p_A^{(n)}$ on $\mathscr{B}(\Omega_A^n)$	$p_B^{(n)}$ on $\mathscr{B}(\Omega_B^{(n)})$
Probability measure on Z-words	p_A on $\mathscr{B}(\Omega_A)$	p_B on $\mathscr{B}(\Omega_B)$
Information of n-words	$H_n(A) = -\sum_{i\in\Omega_A^{(n)}} p_A^{(n)}(i)\ln p_A^{(n)}(i)$	$H_n(B)$ of $p_B^{(n)}$ analogously
Information of Z-words per one letter	$H(A) = \lim_{n\to\infty} \frac{1}{n} H_n(A)$	$H(B) = \lim_{n\to\infty} \frac{1}{n} H_n(B)$

We see that the difficult problem how to calculate information of p_A and p_B (the latter are actually continuous probability measures since Ω_A and Ω_B are not denumerable) is here avoided by taking the limits of $n^{-1}H_n(A)$ and $n^{-1}H_n(B)$ for $n \to \infty$. These limits exist when $p_A^{(n)}$ and $p_B^{(n)}$ are *stationary*, i.e., are invariant with respect to the shift in t, as we here assume. We remark that if all letters in each n-word $(a_1, \dots, a_n) \in A^n$ are *stochastically independent*, i.e.,

$$p_A^{(n)}(a_1, \dots, a_n) = p_A^{(1)}(a_1) \dots p_A^{(1)}(a_n) \quad \text{for all } n \in N, \tag{4.1}$$

we obviously obtain

$$H_n(A) = nH_1(A) \quad \text{and} \quad H(A) = H_1(A), \tag{4.2}$$

and analogously for B. Thus the problem of calculating information $H(A)$ of a source A is not trivial only if there are statistical correlations between letters in words and $p_A^{(n)}$ for $n > 1$ is not determined by $p_A^{(m)}$'s with $m < n$ (of course, converselly, all $p_A^{(m)}$'s are determined by $p_A^{(n)}$ with $n > m$ as the so-called marginal probability measures). Thus we have non-trivial information of a source if there are non-trivial (not factorizing) *joint probabilities* $p_A^{(n)}$ of letters in an n-word for $n > 1$.

The *double source* is a source with the alphabet $C = A \times B$ of pairs (a, b), $a \in A$, $b \in B$. We define as above the corresponding $\Omega_C^{(n)}$, Ω_C, $p_C^{(n)}$, p_C, $H_n(C)$ and $H(C) = H(A \times B)$. Also here if the sources are *stochastically independent*, i.e.,

$$p_C^{(n)}(a, b) = p_A^{(n)}(a)p_B^{(n)}(b) \quad \text{for all } a \in \Omega_A^{(n)}\ b \in \Omega_B^{(n)} \text{ and all } n \in N, \tag{4.3}$$

we have the additivity of information

$$H(A \times B) = H(A) + H(B). \tag{4.4}$$

In the general case, however, when correlations between A and B occur, we have only the *subadditivity*

$$H(A \times B) \leqslant H(A)+H(B). \tag{4.5}$$

Therefore,

$$R(A, B) := H(A)+H(B)-H(A \times B) \geqslant 0. \tag{4.6}$$

Since $R(A, B)$ vanishes when there is no informational connection between A and B, $R(A, B)$ is a good measure of the transition of information from A to B (or conversely) called *transinformation* (*synentropy*) or *information rate* per letter or per unit of time used, e.g., per second, etc. Expression (4.6) is symmetric in A and B, but when we introduce conditional information, cf. [26],

$$H(A \mid B) = \text{information of } A \text{ under condition } B := H(A \times B)-H(B), \tag{4.7}$$

$$H(B \mid A) = \text{information of } B \text{ under condition } A := H(A \times B)-H(A), \tag{4.8}$$

we obtain the asymmetrical expressions for information rate

$$R(A, B) = H(A)-H(A \mid B) = H(B)-H(B \mid A). \tag{4.9}$$

The second terms on the right-hand sides of (4.9) can be interpreted as a measure of the *noise* in the channel, $H(A \mid B)$ for direction $A \to B$ and $H(B \mid A)$ for direction $B \to A$. They vanish when the channel is noiseless (no loss of information, *translation*), and are equal to $H(A)$ or $H(B)$, respectively, when the noise is total (the stochastic independence of sources A and B) and no information is transmitted through the channel.

We obtain a real asymmetry and characterization of A as the input of information if we introduce Shannon's definition of *channel capacity* which is a property of the channel itself and not also of the source A (B being now considered as depending on A)

$$\mathscr{C} := \sup_{H(A)} R(A, B) = \sup_{p_A} R(A, B) \geqslant 0. \tag{4.10}$$

The channel capacity is a key concept of the celebrated Shannon coding theorems, cf. [35], [70], saying roughly that if $H(A) < \mathscr{C}$, we can choose such a *code*, i.e., a noiseless channel translating the messages (words) expressed in alphabet A into messages in some alphabet A' such that the given channel works then practically also as a noiseless one, i.e., translates with any prescribed accuracy (arbitrarily small probability of distortion) the messages in alphabet A' into messages in some alphabet B' from which they can be again translated into alphabet B (decoding), cf. Fig. 3. (In this approximate formulation we omitted the conditions that the input source should be ergodic and the channel non-anticipating and with finite memory, cf. [35].) Thus the negative influence of noise can be practically eliminated by an appropriate choice of code if $H(A) < \mathscr{C}$.

For calculating $\mathscr{C}$ in a special example of practical importance[4] Shannon assumed in [57] that the input message (signal) and the noise are stochastically independent, that the signal is contained in a frequency band $W = \omega/2\pi$ and a time interval τ (then by

[4] Actually this example has continuous messages, but they can be approximated by discrete ones.

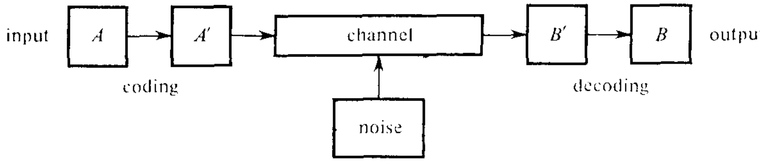

Fig. 3. Scheme of Shannon's coding theorems

the so-called *sampling theorem* it is defined by $2W\tau$ numbers), and that the probability measures for signal and noise are of the Gaussian type with powers S and N, respectively. Then

$$\mathscr{C} = W \ln\left(1 + \frac{S}{N}\right), \tag{4.11}$$

i.e., the channel capacity is determined by the frequency band width W and the signal-to-noise ratio S/N (for other heuristic derivation of this relation cf. [50]).

Recapitulating this simplified summary of classical information theory we can state that it is critically dependent on the concept of joint probability since $p_A^{(n)}$, $p_B^{(n)}$ ($n > 1$, $n \in N$), $p^n_{A\times B}$ ($n \in N$) define $H(A)$, $H(B)$, and $H(A \times B)$. Alternatively, *conditional probabilities*

$$p^{(n)}(a\,|\,b) := \frac{p^{(n)}((a, b))}{p_B^{(n)}(b)}, \quad p_B^{(n)}(b) > 0, \quad a \in \mathscr{B}(\Omega_A^{(n)}), \quad b \in \mathscr{B}(\Omega_B^{(n)}) \tag{4.12}$$

may be used instead of $p^{(n)}_{A\times B}$ since $H(A\,|\,B)$ can be expressed by $p^{(n)}(a\,|\,b)$ (4.12), cf. [26] and Section 7 below. But, as we see, joint probabilities are needed for $H(A)$ and $H(B)$ and, according to (4.12), joint probabilities determine conditional probabilities.

5. Why quantum information theory?

We have seen in Section 3 that for $\alpha \gg 1$ or $\hbar\omega \gg kT$ information of the energy of an oscillator in thermal equilibrium has a quantum part which prevails definitely over the classical part. Now we are interested how the information capacity of a quantum channel, e.g., electromagnetic one, should behave for $\alpha \to \infty$ or $T \to 0$. If we assume for the moment that the Shannon formula (4.11) is correct also in the quantum domain, we may insert into it for comparison the classical and quantum expressions for noise N as function of temperature, and then go over to $T = 0$. In the classical case we have the well-known Nyquist formula for noise [48], cf. [5], Chap. 11,

$$N = KkT \quad (K > 0), \tag{5.1}$$

where K is some constant connected with the friction in the system. In the quantum case to get some estimation we can either generalize (5.1) by putting in place of kT the Planck expression for the mean energy of a photon in the black body electromagnetic radiation, cf. [5] 11.3,

$$N = K\hbar\omega(e^{\alpha} - 1)^{-1}, \tag{5.2}$$

or use for N simply the fluctuation (standard deviation) of photon energy in the black body radiation, cf. [40], p. 427, [52],

$$N = \hbar\omega e^{-\alpha/2}(e^{\alpha}-1)^{-1}. \tag{5.3}$$

In both cases we obtain essentially the same limit as in the classical case (5.1), i.e.,

$$\lim_{T\to 0} \mathscr{C} = \infty \tag{5.4}$$

if we assume that W and S in (4.11) are independent of T.[5] This result seems to be incorrect in the quantum limit. Indeed, independent calculations for some special quantum

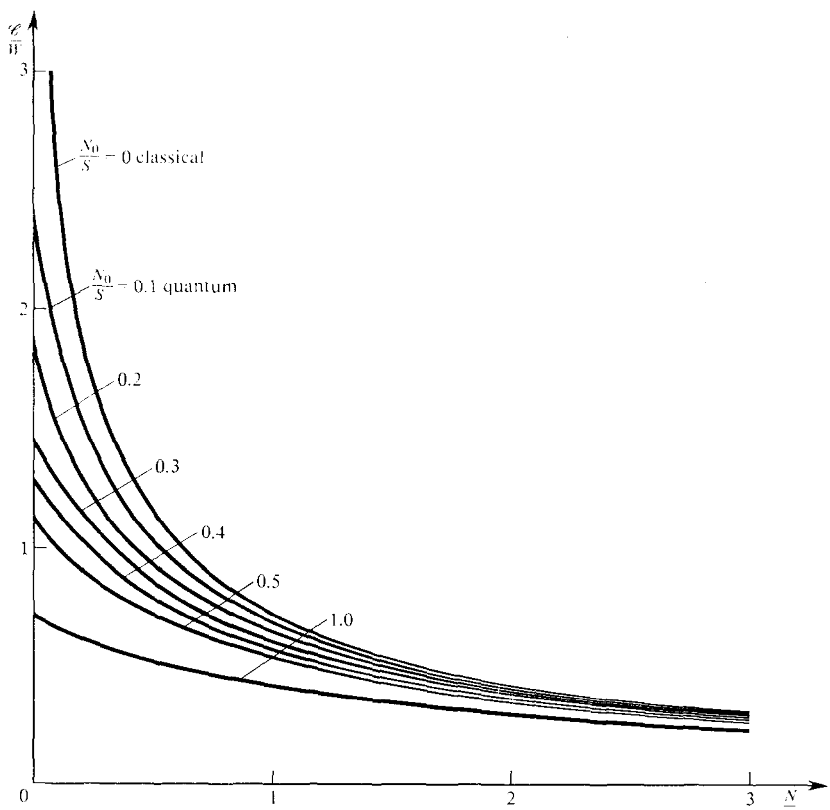

Fig. 4. Relative channel capacity $\mathscr{C}/W$ as function of the "thermal" relative noise N/S and the "zero quantum" relative noise N_0/S according to formula (5.5)

[5] We remark that if we take for our system not the whole electromagnetic field in thermal equilibrium, but only one oscillator in thermal state as in Section 3, we also obtain the same result. Then, namely, although the mean energy gives for $T \to 0$ a finite limit $\frac{1}{2}\hbar\omega$ ("zero energy"), the energy fluctuation vanishes at $T = 0$ as $a^{-\alpha}$ for $\alpha \to \infty$.

models [21], [27] show that in the quantum case the Shannon formula (4.11) should be rather modified to the form

$$\mathscr{C}_{qu} = W \ln \left(1 + \frac{S}{N+N_0} \right), \tag{5.5}$$

where N_0 is a positive constant, as if a "zero temperature fluctuation" or "zero quantum noise". Then we would have

$$\lim_{T \to 0} \mathscr{C}_{qu} = W \ln \left(1 + \frac{S}{N_0} \right) < \infty, \tag{5.6}$$

i.e., the capacity is bounded from above as we can expect (for $T \to \infty$, $\mathscr{C} \to 0$ strictly monotonously, cf. Fig. 4).

We see from the above short discussion of the present situation that a new quantum information theory is really desirable. The old theory cannot be improved only by inserting into it some quantum formulae, as we have tried to do. The question has, namely, a much deeper origin, i.e., the existence of a different probability theory for the "non-commuting" case.

To complete our answer to the question in the title of the section we shall indicate where the "quantum land" is situated. For radiation systems in thermal equilibrium the "quantum border" is determined by the condition $\alpha = 1$ or $\hbar\omega = kT$ which gives the critical circular frequency

$$\omega = \omega_0 = \frac{kT}{\hbar}, \tag{5.7}$$

or the critical wave length

$$\lambda_0 = \frac{2\pi c}{\omega_0} = \frac{2\pi c\hbar}{kT} \tag{5.8}$$

(c is the velocity of light), cf. Table 1.

TABLE 1

Temperature T(K)	Example	Critical circular frequency $\omega_0(\mathrm{s}^{-1})$	Critical wave length λ_0(m)	Spectral region
3	liquid helium	$4 \cdot 10^{11}$	$5 \cdot 10^{-3}$	millimeter microwaves
300	room temperature	$4 \cdot 10^{13}$	$5 \cdot 10^{-5}$	infrared light
3000	thermal electric lamp	$4 \cdot 10^{14}$	$5 \cdot 10^{-6}$	near infrared light
30000	sun surface	$4 \cdot 10^{15}$	$5 \cdot 10^{-7}$	visible light
$3 \cdot 10^6$	sun centre and atomic bomb	$4 \cdot 10^{17}$	$5 \cdot 10^{-9}$	soft X-rays

We see that for all practically important temperatures of equilibrium radiation the radio waves and the long microwaves belong to the classical region, while the short microwaves (in low temperatures) and infrared and visible light (in normal and high temperatures) belong to the quantum region. Thus the recently progressing shift of communication devices from the radio frequencies to the light frequencies (lasers) is a passage across the border of the quantum domain, this Alice's wonderland of physics, as well as mathematics. The passage is only accelerated by the fact that lassers are far from the Gibbs thermal equilibrium (their "generalized equilibria" are usually determined by the negative absolute temperature [21] or by the so-called higher order temperatures [29], [2]). Indeed, many facts of experiment and theory show that in the field of "quantum electronics" or "quantum optics" the quantum and at least semi-quantum laws begin to govern.[6] Thus the quantum information theory is not only a scientifically interesting subject, but it is a practical need.

6. Semiobservables, instruments, and expectations

As we have said in the introduction, for closed quantum systems no general concepts of joint and conditional probability exist and, therefore, no information theory (to be precise, no theory of transport of information, no theory of communication) can be constructed, at least on Shannon's lines. This is, however, natural. A closed system cannot communicate with the surroundings, in particular, a technical device intended for communication cannot be controlled, or "pumped" longer from outside and cannot give any "message" to the outside world. In other words, no closed system can be an "input" or an "output". To be more rigorous, closed quantum systems can contact with the outside world at most twice: during their "preparation" and their "measurement" (in the theory both acts are idealized as momentary, of no duration). Such single acts are not sufficient for an information channel, which has to work constantly, in discontinuous or continuous way.

The other point which should be emphasized is that the contact of a communication device with the outside world has to be of a stochastic nature. Indeed, the point is that the message is not fixed in a deterministic way. If it were fixed, it would be no message at all since it would be known beforehand. Thus, not because of any technical reasons, but as a matter of principle, the quantum theory of communication should be a stochastic theory of open quantum systems.

As mentioned, such a theory has been formulated by Davies and Lewis [8], and to some extent and independently by Kossakowski [39], cf. also [30], [32]. In our presenta-

[6] There are people (as W. E. Lamb) who say that lasers are classical instruments, but such a saying has only a limited meaning of stating that lasers are macroscopic systems and that the classical laws in some sense are prevailing in their theory. But it cannot be denied that in many more subtle effects (as quantum noise, spiking, spectral characteristics, etc.), as well in a more fundamental theory, the quantum laws cannot be neglected. Actually, we have to use there quantum mechanics of open systems, cf. [29], which is a semi-macroscopic theory, and not the usual quantum mechanics (of closed systems).

tion we shall combine both methods in this sense that to the more abstract formulation of Davies and Lewis we shall give a concrete representation as proposed by Kossakowski and recapitulated above in Section 2. Then we shall generalize the obtained theory to include finite memory effects (more general than in the Kossakowski theory).

First of all we shall generalize the concept of quantum observable as a self-adjoint operator (2.7), cf. [8]. A self-adjoint operator is uniquely determined by a spectral measure defined by properties (2.8)–(2.11). But from the point of view of the theory of probability it is only essential that $p(A, \varrho, E)$ as defined by (3.7) is a probability measure in the classical Kolmogorov sense. We easily see that for the latter only the axioms (1)–(3), (2.8) to (2.10), are necessary and sufficient, since they guarantee that p is positive (non-negative), σ-additive and normalized. Thus in a generalized theory axiom (4) (projectivity) may be considered as superfluous. If we omit it, however, we lose the 1-1-correspondence between an operator A and its spectral measure $a(E)$ in (2.7): now to a given A there exist many operator measures $a(E)$ such that (2.7) holds. Since only the measure $a(E)$ is directly connected with probability p, we have to resign from operators as representing observables and go over to bounded operator valued measures $a(E)$ satisfying only (2.8), (2.9), and (2.10) and called *operator measures* or *semispectral measures*, cf. [45]. Since in general we need many-dimensional spaces of elementary events, Ω, it is more convenient to formulate our definition in a somewhat more general form:

DEFINITION 1. Let $\mathscr{H}$ be the Hilbert space of a physical system, Ω a compact[7] topological space called *value space* and $\mathscr{B}(\Omega)$ the Baire σ-algebra of subsets of Ω. A *semiobservable* on $\mathscr{H}$ and Ω or *semispectral measure* on $\mathscr{B}(\Omega)$ into $\boldsymbol{B}(\mathscr{H})$ is a function a: $\mathscr{B}(\Omega) \to \boldsymbol{B}(\mathscr{H})$ such that

(1) $a(E) \geqslant 0 \quad$ for all $E \in \mathscr{B}(\Omega) \quad$ (*positivity*), (6.1)

(2) $a(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} a(E_n) \quad$ for all mutually disjoint E_n (the series is understood in the sense of weak operator topology) $\quad$ (σ-*additivity*), (6.2)

(3) $a(\Omega) = I \quad$ (*normalization*). (6.3)

That a semiobservable has a good physical interpretation for open quantum systems can be inferred from the following theorem due to E. M. Alfsen (private communication):

THEOREM 1. *Let a and b be spectral and semispectral measures, respectively, on $\mathscr{B}(\bar{\boldsymbol{R}})$ into $\boldsymbol{B}(\mathscr{H})$, and let the mean values of a and b coincide in any state $\varrho \in \Delta(\mathscr{H})$, i.e.*,

$$\langle a, \varrho \rangle := \mathrm{Tr}\Big(\varrho \int_{\bar{\boldsymbol{R}}} \lambda a(d\lambda)\Big) = \mathrm{Tr}\Big(\varrho \int_{\boldsymbol{R}} \lambda b(d\lambda)\Big) =: \langle b, \varrho \rangle =: \bar{\lambda}. \tag{6.4}$$

---

[7] Compactification of the value space is very convenient mathematically and has no serious physical consequences. E.g., we may compactify $\boldsymbol{R}$ to $\bar{\boldsymbol{R}}$ not only by one point in infinity, but also by *two* points, $\boldsymbol{R}^2$ not only by one point, but also by a circle or line in infinity, etc., accordingly to the topological property of the physical situation.

Then for all ϱ

$$ {}_{\varrho}\sigma_a^2 := \mathrm{Tr}\Big(\varrho \int_{\boldsymbol{R}} (\lambda-\bar{\lambda})^2 a(d\lambda)\Big) \leqslant \mathrm{Tr}\Big(\varrho \int_{\boldsymbol{R}} (\lambda-\bar{\lambda})^2 b(d\lambda)\Big) =: {}_{\varrho}\sigma_b^2. \tag{6.5} $$

In other words, the spectral measures on $\boldsymbol{R}$ are distinguished among all semispectral measures giving the same mean values by the smallest possible fluctuations (standard deviations). This situation corresponds rightly to our concept of physical isolation of a closed system: such a system is under the smallest possible disturbances from outside. If we allow outside disturbances of different strength, we go over from one observable A or a (2.7) to an infinite set of semiobservables b which give the same mean values but greater (not smaller) fluctuations. The preservation of the mean values is a generalization of the Langevin idea of a stochastic force having the zero mean value but a positive fluctuation.

THEOREM 2 (M. A. Naimark, cf. [27], p. 33). *Let* $a(E)$ *be a semispectral measure on* $\mathscr{B}(\Omega)$ *into* $\boldsymbol{B}(\mathscr{H})$. *Then there exists a Hilbert space* $\mathscr{H}_0$ *such that* $\mathscr{H}$ *is its subspace, i.e.,*

$$ \mathscr{H} = P\mathscr{H}_0 \tag{6.6} $$

(*where* P *is some projector operator on* $\mathscr{H}_0$), *and a spectral*[8] *measure* $a_0(E)$ *on* $\mathscr{B}(\Omega)$ *into* $\boldsymbol{B}(\mathscr{H}_0)$ *such that*

$$ a(E) = Pa_0(E)P \quad \textit{for all } E \in \mathscr{B}(\Omega). \tag{6.7} $$

Conversely, if $a_0(E)$ *is a spectral measure on* $\mathscr{B}(\Omega)$ *into* $\boldsymbol{B}(\mathscr{H}_0)$, *and* $\mathscr{H}$ *is a subspace of* $\mathscr{H}_0$ *given by* (6.6), *then* $a(E)$ *defined by* (6.7) *is a semispectral measure on* $\mathscr{B}(\Omega)$ *into* $\boldsymbol{B}(\mathscr{H})$.

Theorem 2 allows us to show that if we build a semiobservable by means of Glauber's coherent states [16] in Dirac's notation

$$ a(E) := \frac{1}{\pi}\int_E |z\rangle\langle z|\, d^2z \qquad \text{for all } E \in \mathscr{B}(\boldsymbol{C}), \tag{6.8} $$

where $\boldsymbol{C}$ is a compactificated complex plane, $a(E)$ can be considered as a projection (6.7) of a spectral measure on $\mathscr{B}(\boldsymbol{C})$ composed of two commuting observables on a bigger Hilbert space, cf. [27], also [34].

In the case where we measure first an observable A with a pure point spectrum (3.9) and then an arbitrary observable

$$ B = \int_{-\infty}^{\infty} \lambda b(d\lambda) \tag{6.9} $$

[8] Here by spectral measure we mean a somewhat more general concept than defined by (2.8)–(2.11), namely, instead of $\boldsymbol{R}$ we take any compact topological space Ω. E.g., if $\Omega = \boldsymbol{R}^n$, a spectral measure can be built by taking the product of the spectral measures of n commuting observables on $\mathscr{H}$.

Commutativity of the observables is necessary and sufficient condition that the product be a spectral measure.

(where $b(E)$ is a spectral measure), semiobservables can be used for construction of conditional and joint probability measures, even if A and B do not commute, cf. [8], [62], [46], [47]. If A is measured on a system in state ϱ and an eigenvalue λ_n is obtained, then according to the usual interpretation, cf. [8], [67], p. 214, [43], [19], [12], the state is changed

$$\varrho \mapsto \varrho_n := \frac{P_n \varrho P_n}{\mathrm{Tr}(P_n \varrho P_n)}. \tag{6.10}$$

When A is measured, but we do not know or are not interested which eigenvalue is obtained (no selection) we get the transformation

$$\varrho \mapsto \varrho_A := \sum_{n=1}^{\infty} P_n \varrho P_n =: T_A \varrho, \tag{6.11}$$

where T_A is a positive linear endomorphism of $L^1(\mathscr{H})$. T_A causes that after the measurement in the domain of A

$$[A, \varrho_A] = 0, \tag{6.12}$$

even when before measurement $[A, \varrho] \neq 0$. Similarly, we can define the conditioning of the observable B by the measurement of A

$$b_A(E) := \sum_{n=1}^{\infty} P_n b(E) P_n = T_A b(E) \quad \text{for all } E \in \mathscr{B}(\boldsymbol{R}). \tag{6.13}$$

Since

$$\mathrm{Tr}\big(\varrho b_A(E)\big) = \mathrm{Tr}\big(\varrho_A b(E)\big) \quad \text{for all } \varrho \in \Lambda(\mathscr{H}) \text{ and all } E \in \mathscr{B}(\boldsymbol{R}), \tag{6.14}$$

the function

$$p(b_A, \varrho, E) := \mathrm{Tr}\big(\varrho b_A(E)\big) \tag{6.15}$$

is a probability measure on $\mathscr{B}(\boldsymbol{R})$ and $b_A(E)$ is a semiobservable on $\mathscr{B}(\boldsymbol{R})$ into $\boldsymbol{B}(\mathscr{H})$. Only when $[A, B] = 0$, $b_A(E) = b(E)$ is a spectral measure, and the measurement of B is not influenced (*conditioned*) by the measurement of A. We may also define [8] the semiobservable

$$c(F \times E) := \sum_{\lambda_n \in E} P_n b(F) P_n \quad \text{for all } E, F \in \mathscr{B}(\boldsymbol{R}) \tag{6.16}$$

which can be shown to extend uniquely to $c(G)$ for all $G \in \mathscr{B}(\boldsymbol{R}^2)$ and may be called a *joint semiobservable* of a and b (more precisely, of a followed by b or b following a), while

$$p(c, \varrho, G) := \mathrm{Tr}\big(\varrho c(G)\big) \quad \text{for all } G \in \mathscr{B}(\boldsymbol{R}^2) \tag{6.17}$$

is a *joint probability measure* for measuring b following a. We have the marginal measures

$$\begin{aligned} c(\boldsymbol{R} \times E) &= a(E) && \text{for all } E \in \mathscr{B}(\boldsymbol{R}), \\ c(F \times \boldsymbol{R}) &= b_A(F) && \text{for all } F \in \mathscr{B}(\boldsymbol{R}). \end{aligned} \tag{6.18}$$

We remark that the presented procedure cannot be repeated, however, for the third measurement even for B with purely point spectrum, expect for the trivial case of commuting A and B. The procedure cannot also be generalized for A with continuous spectrum. The deeper reason of these difficulties is rooted in the fact that, in general, the product of two spectral (semispectral) measures is not even a semispectral one (except for the commuting observables). To surmount these difficulties Davies and Lewis [8] defined the concept of "instrument". In our more concrete setting this definition can be formulated as follows:

DEFINITION 2. An *instrument* is a superoperator-valued probability measure, i.e., a function $\mathscr{E}\colon \mathscr{B}(\Omega) \to K(\mathscr{H})$ such that

(1) $\mathscr{E}(E) \geqslant 0$ (is incompletely positive) for all $E \in \mathscr{B}(\Omega)$ (*positivity*), (6.19)

(2) $\mathscr{E}\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mathscr{E}(E_n)$ for mutually disjoint $E_n \in \mathscr{B}(\Omega)$ (where the series is understood in the sense of strong operator topology) (*σ-additivity*), (6.20)

(3) $\mathrm{Tr}\big(\mathscr{E}(\Omega)\varrho\big) = \mathrm{Tr}\,\varrho$ for all $\varrho \in L^1_+(\mathscr{H})$ (*normalization*). (6.21)

An instrument (semiobservable) is said to be *discrete* iff its value space Ω is discrete.

THEOREM 3 [8]. *To every instrument $\mathscr{E}$ on $\mathscr{B}(\Omega)$ into $K(\mathscr{H})$ there exists a unique semiobservable a on $\mathscr{B}(\Omega)$ into $\boldsymbol{B}(\mathscr{H})$, called the* associated semiobservable *with $\mathscr{E}$, such that*

$$\mathrm{Tr}\big(\mathscr{E}(E)\varrho\big) = \mathrm{Tr}\big(\varrho a(E)\big) \quad \textit{for all } \varrho \in L^1(\mathscr{H}) \textit{ and all } E \in \mathscr{B}(\Omega). \tag{6.22}$$

Every instrument a on $\mathscr{B}(\Omega)$ into $\boldsymbol{B}(\mathscr{H})$ is determined in such a way by at least one instrument.

THEOREM 4 [28]. *Let $\mathscr{E}_1$ and $\mathscr{E}_2$ be instruments on $\mathscr{B}(\Omega_1)$ and $\mathscr{B}(\Omega_2)$, respectively, into $K(\mathscr{H})$. Then there exists one and only one instrument $\mathscr{E}_{12} =: \mathscr{E}_1 \circ \mathscr{E}_2$ on $\mathscr{B}(\Omega_1 \times \Omega_2)$ into $K(\mathscr{H})$, called the composition of $\mathscr{E}_1$ following $\mathscr{E}_2$, such that*

$$\mathscr{E}_{12}(E \times F) = \mathscr{E}_1(E)\mathscr{E}_2(F) \quad \textit{for all } E \in \mathscr{B}(\Omega_1) \textit{ and all } F \in \mathscr{B}(\Omega_2). \tag{6.23}$$

Generalizing the above concepts (6.13) and (6.16) and following [8], we introduce

DEFINITION 3. Let $\mathscr{E}_1$ and $\mathscr{E}_2$ be instruments on $\mathscr{B}(\Omega_1)$ and $\mathscr{B}(\Omega_2)$, respectively, into $K(\mathscr{H})$, and a_1 and a_2 the associated semiobservables with $\mathscr{E}_1$ and $\mathscr{E}_2$. Then the semiobservable on $\mathscr{B}(\Omega_2)$ into $\boldsymbol{B}(\mathscr{H})$

$$a_{21}(F) := \mathscr{E}_1(\Omega_1)^* a_2(F) \quad \text{for all } F \in \mathscr{B}(\Omega_2) \tag{6.24}$$

is said to be the semiobservable a_2 *conditioned* by the measurement of a_1 with the instrument $\mathscr{E}_1$. The semiobservable $c_{21}(G)$ on $\mathscr{B}(\Omega_2 \times \Omega_1)$ into $\boldsymbol{B}(\mathscr{H})$ associated with the composition $\mathscr{E}_2 \circ \mathscr{E}_1$ is said to be the *joint semiobservable* of a_2 measured by instrument

$\mathscr{E}_2$ following a_1 measured by instrument $\mathscr{E}_1$ (and for the reverse order of the measurements $c_{12}(G)$ associated with $\mathscr{E}_1 \circ \mathscr{E}_2$). Iff

$$\mathscr{E}_1 \circ \mathscr{E}_2 = \mathscr{E}_2 \circ \mathscr{E}_1 \quad \text{(and, therefore, } c_{12} = c_{21}\text{)}, \tag{6.25}$$

the instruments (joint semiobservables) are said to be *compatible*. (In general, instruments are not compatible.)

The probability measure

$$\mathrm{Tr}\big((\mathscr{E}_2 \circ \mathscr{E}_1)(G)\varrho\big) \quad \text{for all } G \in \mathscr{B}(\Omega_2 \times \Omega_1) \text{ and all } \varrho \in \Delta(\mathscr{H}) \tag{6.26}$$

is said to be the *joint probability measure* for measurement by instrument $\mathscr{E}_2$ following measurement by instrument $\mathscr{E}_1$ in state ϱ.

THEOREM 5 [8]. *The joint semiobservable $c_{21}(G)$ on $\mathscr{B}(\Omega_2 \times \Omega_1)$ into $\boldsymbol{B}(\mathscr{H})$ has the marginal semiobservables*

$$\begin{aligned} c_{21}(\Omega_2 \times E) &= a_1(E) \quad \textit{for all } E \in \mathscr{B}(\Omega_1), \\ c_{21}(F \times \Omega_1) &= a_{21}(F) \quad \textit{for all } F \in \mathscr{B}(\Omega_2). \end{aligned} \tag{6.27}$$

Each instrument $\mathscr{E}$ is, of course, compatible with itself, but it is not necessarily repeatable in the sense of the following definition, cf. [8]:

DEFINITION 4. An instrument $\mathscr{E}$ on $\mathscr{B}(\Omega)$ into $K(\mathscr{H})$ is said to be *repeatable* iff

$$\mathscr{E}(E)\mathscr{E}(F)\varrho = \mathscr{E}(E \cap F)\varrho \quad \text{for all } E, F \in \mathscr{B}(\Omega) \text{ and all } \varrho \in L^1(\mathscr{H}). \tag{6.28}$$

It is said to be *weakly repeatable* iff

$$\mathrm{Tr}\big(\mathscr{E}(E)\mathscr{E}(F)\varrho\big) = \mathrm{Tr}\big(\mathscr{E}(E \cap F)\varrho\big) \quad \text{for all } E, F \in \mathscr{B}(\Omega) \text{ and all } \varrho \in L^1(\mathscr{H}), \tag{6.29}$$

and *strongly repeatable* iff Ω is discrete ($\Omega = \{\lambda_n\}_{n=1}^{\infty}$, $\mathscr{E}(\{\lambda_n\}) =: \mathscr{E}_n$, $n \in N$) and

(1) $\mathscr{E}_n \mathscr{E}_m \varrho = \delta_{nm} \mathscr{E}_m \varrho$ for all $x, y \in \Omega$ and all $\varrho \in L^1(\mathscr{H})$ (*von Neumann repeatability hypothesis*), (6.30)

(2) if $\mathrm{Tr}(\mathscr{E}_n \varrho) = \mathrm{Tr}\varrho$ for all $\varrho \in L^1(\mathscr{H})$, then $\mathscr{E}_n \varrho = \varrho$ (*minimum disturbance principle*), (6.31)

(3) if $\mathrm{Tr}(A\mathscr{E}_n \varrho) = 0$ for all $A \in L_+^{\infty}(\mathscr{H})$, all $n \in N$ and all $\varrho \in L_+^1(\mathscr{H})$, then $A = 0$ (*non-degeneracy condition*). (6.32)

As Davies and Lewis said, [8], p. 247, the existence of repeatable instruments is doubtful, although there are instruments which are repeatable in some approximate sense, cf. [8], [7]. As regards strongly repeatable instruments, we have their connection with the transformation T_A (6.11) of an observable A with a purely point spectrum (3.9) by means of

THEOREM 6 [8]. *Let Ω be a discrete value space and $A = \sum_{n=1}^{\infty} \lambda_n P_n$ any observable on $\mathscr{H}$ with the spectrum $\Omega = \{\lambda_n\}_{n=1}^{\infty}$. Then the formula*

$$\mathscr{E}(E)\varrho = \sum_{\lambda_n \in E} P_n \varrho P_n \tag{6.33}$$

sets up a 1-1-*correspondence between the discrete spectral measures* $a(E) = \sum_{\lambda_n \in E} P_n$ *on* $\mathscr{B}(\Omega)$ *and strongly repeatable instruments on* $L^1(\mathscr{H})$. *In particular, a strongly repeatable instrument is uniquely determined by its associate semiobservable which is a spectral measure (observable).*

Thus the class of instruments for which the von Neumann repeatability hypothesis [67] is valid is uniquely determined by all observables with purely point spectra. Therefore, if we immediately repeat the measurement of an observable whose spectrum is not purely point, we do not obtain, in general, the same result. We get some statistics of results which is just fixed by a considered instrument $\mathscr{E}$ on $\mathscr{B}(\Omega)$ into $K(\mathscr{H})$ and a state $\varrho \in \Delta(\mathscr{H})$ by means of the probability measure

$$p(\mathscr{E}, \varrho, E) := \operatorname{Tr}(\mathscr{E}(E)\varrho) \quad \text{for all } E \in \mathscr{B}(\mathscr{H}). \tag{6.34}$$

The dual superoperator $\mathscr{E}^*(E)$ to an instrument $\mathscr{E}(E)$ on $\mathscr{B}(\Omega)$ into $K(\mathscr{H})$ is called by Davies an *expectation* [7] because it is a generalization of the concept of conditional expectation in an operator algebra introduced by Umegaki [62], [46]. We remark that the whole theory developed by Davies and Lewis [8], [7] and Holevo [27], [28] is formulated either in more abstract terms for the general case or in terms of Von Neumann algebras for the quantum case. As we said our theory is formulated a little less abstract and is therefore more elementary. In our setting we can also define an expectation independently of an instrument by

DEFINITION 5. An *expectation* on $\mathscr{B}(\Omega)$ into $K^*(\mathscr{H})$ (= the set of all linear operators on and into $L^\infty(\mathscr{H})$) is a function $\mathscr{E}^*\colon \mathscr{B}(\Omega) \to K^*(\mathscr{H})$ such that

(1) $A \geqslant 0$ implies $\mathscr{E}^*(E)A \geqslant 0$ for all $E \in \mathscr{B}(\Omega)$ and all $A \in L^\infty(\mathscr{H})$ (*positivity*), (6.35)

(2) $\mathscr{E}^*\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mathscr{E}^*(E_n)$ for all mutually disjoint $E_n \in \mathscr{B}(\mathscr{H})$, the series being understood in the weak operator topology (*σ-additivity*), (6.36)

(3) $\mathscr{E}^*(\Omega)I = I$ (*normalization*). (6.37)

Thus an expectation is a dual superoperator-valued probability measure, another example of generalized probability measures. Of course, we have

$$\langle A, \mathscr{E}(E)\varrho\rangle = \operatorname{Tr}(A\mathscr{E}(E)\varrho) = \operatorname{Tr}(\varrho\mathscr{E}^*(E)A) = \langle \mathscr{E}^*(E)A, \varrho\rangle \quad \text{for all } E \in \mathscr{B}(\Omega) \tag{6.38}$$

and all $A \in L^\infty(\mathscr{H})$ and $\varrho \in L^1(\mathscr{H})$.

7. Quantum information channels

After the mathematical preparation we are finally able to formulate a quantum information theory in analogy to the classical one presented in a short summary in Section 4.

For this purpose we have to give first a schematic description of the physical procedures and processes occurring in a quantum information channel, e.g., in an arrangement for laser communication, cf. [52], [53].

In the classical case in each unit of time we register some object called *letter* in the input and in the output. This letter may be considered as a value of a stochastic variable, and the process as its measurement. The quantum analogue will be a measurement of an observable (semiobservable), and a measured eigenvalue (value) corresponds to a letter.

The letters in the input source can be controlled by means of some modulation arrangement (modulator) which prepares states of the system to get such or other letter (eigenvalue) one after another in a sequence, of course, with some probabilities which define the information of the communicated message. In the case of laser communication we have the following examples of modulators: a *Pockels cell* (changes the intensity of light), a *Kerr cell* (changes the polarization of light), a *dye laser cell* (changes the frequency of light), etc., for details cf. [52], [53]. Practically the most important are two sorts of modulation: the *amplitude* or *intensity modulation* (AM, effected, e.g., by a Pockels cell) and the *frequency* or *energy modulation* (FM, effected, e.g., by a dye laser cell), the second being more modern and technically more perfect (it is much less sensitive to the noise caused by absorption, scatterring, density fluctuations, etc.).

The next question is how to define the quantum system or the Hilbert space on which an observable (semiobservable) acts. It is clear that this is the quantum system of the physical carrier of information. In case of an AM laser sender, this is a light oscillator (mode) of a given frequency whose quantum energy levels correspond to numbers of photons emitted, and the same by the receiver. Thus we have two Hilbert spaces: the input $\mathscr{H}_{\mathrm{in}}$ and the output $\mathscr{H}_{\mathrm{out}}$ of the two oscillators (this may be, of course, generalized for many modes), and the third one $\mathscr{H} = \mathscr{H}_{\mathrm{in}} \otimes \mathscr{H}_{\mathrm{out}}$ which corresponds to the double source and describes the interaction of both systems through the channel. For description of the higher correlations of measurements we do not use the higher tensor products of these Hilbert spaces, but the joint instruments and probability measures. By an FM laser system the situation is reverse: the number n of photons is fixed, but their frequency is changing, so we can conveniently use the n-photon quantum mechanical description in the momentum space representation, cf. [1], [41]. So we obtain a picture of particles (photons) moving from the input to the output through the channel (the *running wave approach* in contradistinction to the previous *standing excitation approach* by AM). Thus we have only one Hilbert space $\mathscr{H}$ on which we make measurements at input, at the channel (without selection), and at output. To obtain the "double source" we have only to apply the joint instrument concept and the joint probability measure. But to define the higher correlation of letters we have to take the higher tensor products of our $\mathscr{H}$. The same method can be applied when the modulation consists in changes of polarization of photons (e.g., by a Kerr cell), but then actually a finite-dimensional Hilbert space is sufficient.

In the present paper we shall consider only the running wave approach, not specifying whether it is for the frequency or the polarization modulation. The other approach can easily be reconstructed by the reader.

As always in physics, we can use for description of time processes two pictures, the Heisenberg or Schrödinger ones, which are mutually dual (cf. Sections 2 and 6 above). Here we prefer to use the Schrödinger picture, the other can be also easily reconstructed.

We shall only consider a discrete quantum channel working in an impulse way, i.e., such in which both the input and output messages have the character of short impulses and are registered in the form of discrete numbers, λ_n^{in} and $\lambda_n^{\text{out}}(n \in \boldsymbol{N})$, respectively. The sets $\Omega_{\text{in}} := \{\lambda_n^{\text{in}}\}_{n=1}^{\infty}$, $\Omega_{\text{out}} := \{\lambda_n^{\text{out}}\}_{n=1}^{\infty}$ correspond to the alphabets A and B discussed in Section 4 and serve to define the input and output instruments

$$\mathscr{E}_{\text{in}}: \mathscr{B}(\Omega_{\text{in}}) \to K(\mathscr{H}), \quad \mathscr{E}_{\text{out}}: \mathscr{B}(\Omega_{\text{out}}) \to K(\mathscr{H}). \tag{7.1}$$

We assume as in Section 4 that time t is also discrete, i.e., that the registration (measurement) goes on in discrete instants. At each moment of time $t \in \boldsymbol{Z}$ we have a different physical system with its Hilbert space $\mathscr{H}_t$ (a beam of photons, electrons, ions, phonons, etc.) which serves as a medium (carrier) of communication and goes from the input to the output. Each system has the same (or similar) physical constitution (Hilbert space), but is in a different state ϱ_t which is influenced by the modulation arrangement and the outside noise, is sent from the sender and received by the receiver. For simplicity we assume that the time scales in input and output are shifted by the retardation defined by the distance between two end stations and the velocity of the carrier in the channel.

Under the given assumptions we can apply Theorem 6 and say that to the instruments $\mathscr{E}_{\text{in}}$ and $\mathscr{E}_{\text{out}}$ there correspond bijectively the associated observables (spectral measures)

$$a_{\text{in}}: \mathscr{B}(\Omega_{\text{in}}) \to B(\mathscr{H}), \quad a_{\text{out}}: \mathscr{B}(\Omega_{\text{out}}) \to B(\mathscr{H}). \tag{7.2}$$

Denoting $a_n^{\text{in}} := a_{\text{in}}(\{\lambda_n^{\text{in}}\}) = P_n^{\text{in}}$, $a_n^{\text{out}} := a_{\text{out}}(\{\lambda_n^{\text{out}}\}) = P_n^{\text{out}}$ we can introduce probabilities

$$p_{nt}^{\text{in}} := p(\mathscr{E}_{\text{in}}, \varrho_t^{\text{in}}, \lambda_n^{\text{in}}) = p(a_{\text{in}}, \varrho_t^{\text{in}}, \lambda_n^{\text{in}}) = \operatorname{Tr}(a_n^{\text{in}}\varrho_t^{\text{in}}), \tag{7.3}$$

$$p_{nt}^{\text{out}} := p(\mathscr{E}_{\text{out}}, \varrho_t^{\text{out}}, \lambda_n^{\text{out}}) = p(a_{\text{out}}, \varrho_t^{\text{out}}, \lambda_n^{\text{out}}) = \operatorname{Tr}(a_n^{\text{out}}\varrho_t^{\text{out}}) \tag{7.4}$$

for all $n \in \boldsymbol{N}$ and all $t \in \boldsymbol{Z}$. For stationary channels $p_{nt}^{\text{in}}, p_{nt}^{\text{out}}$ are independent of t and we may write p_n^{in} and p_n^{out} or p_n^{in} (1) and p_n^{in} (1) to indicate that these probabilities correspond to 1-words (1-letter words). To be able to calculate information we assume that instruments $\mathscr{E}_{\text{in}}$, $\mathscr{E}_{\text{out}}$ or observables a_{in}, a_{out} (we may use also self-adjoint operators A_{in} and A_{out} on $\mathscr{H}$) are thermodynamically regular (*informationally regular*), i.e., their spectra $\{\lambda_n^{\text{in}}\}_{n=1}^{\infty}$, $\{\lambda_n^{\text{out}}\}_{n=1}^{\infty}$ fulfil the condition (3.10). Then according to (3.11) we obtain the information of 1-words in input and output ($\varrho_n^{\text{in}} \in \Delta_{A_{\text{in}}}(H)$; $\varrho_n^{\text{out}} \in \Delta_{A_{\text{out}}}(H)$; cf. (3.12))

$$H_1^{\text{in}} := -\sum_{n=1}^{\infty} p_n^{\text{in}}(1)\ln p_n^{\text{in}}(1), \quad H_1^{\text{out}} := -\sum_{n=1}^{\infty} p_n^{\text{out}}(1)\ln p_n^{\text{out}}(1), \tag{7.5}$$

being finite non-negative numbers. We remark once more that for each $t \in \boldsymbol{Z}$ we make a measurement on a different physical system (set of photons, electrons, etc., emitted at time t), therefore ϱ_t^{in} (ϱ_t^{out}) work in seperate Hilbert spaces for each t and we have no problem of repeated measurement.

For k-words ($k > 1$) we build the Hilbert space

$$\mathscr{H}^k := \underbrace{\mathscr{H} \otimes \ldots \otimes \mathscr{H}}_{k \text{ times}} = \mathscr{H}_t \otimes \ldots \otimes \mathscr{H}_{t+k}, \tag{7.6}$$

i.e., the kth tensor power of $\mathscr{H}$. Then we have states $\varrho(k) \in \Delta(\mathscr{H}^k)$ which are not necessarily the tensor products of the states on $\mathscr{H}$ (the latter occurs only when the letters are stochastically independent). Because the systems (subsystems) $\mathscr{H}_{t+i}$ ($0 \leqslant i \leqslant k$) are distinguished by time which is considered as known, we have no problem of mixing them (no necessity of symmetrizing or antisymmetrizing). Then we can also apply to the ith subsystem operators $a_{n_i}^{\mathrm{in}}$ (multiplied tensorially by the respective unity operators with respect to all other subsystems). Of course, since the respective observables commute, as in the classical case, we have no problem with the joint probability. We obtain

$$\Big(\mathscr{E}^{(i)}_{\substack{\mathrm{in}\\ \mathrm{out}}}(n_i) := \mathscr{E}^{(i)}_{\substack{\mathrm{in}\\ \mathrm{out}}}(\{\overset{\mathrm{in}}{\lambda}{}^{\mathrm{out}}_{n_i}\})\Big),$$

$$\overset{\mathrm{in}}{p}{}^{\mathrm{out}}_k(n_1, \ldots, n_k) := \mathrm{Tr}\big(\mathscr{E}^{(1)}_{\substack{\mathrm{in}\\ \mathrm{out}}}(n_1) \ldots \mathscr{E}^{(k)}_{\substack{\mathrm{in}\\ \mathrm{out}}}(n_k)\overset{\mathrm{in}}{\varrho}{}^{\mathrm{out}}(k)\big)$$

$$= \mathrm{Tr}\big(\overset{\mathrm{out}}{a}{}^{\mathrm{in}}_{n_1} \ldots \overset{\mathrm{out}}{a}{}^{\mathrm{in}}_{n_k} \overset{\mathrm{out}}{\varrho}{}^{\mathrm{in}}(k)\big) = \mathrm{Tr}\big(\overset{\mathrm{in}}{P}{}^{\mathrm{out}}_{n_1} \ldots \overset{\mathrm{in}}{P}{}^{\mathrm{out}}_{n_k} \overset{\mathrm{in}}{\varrho}{}^{\mathrm{out}}(k)\big) \tag{7.7}$$

and

$$\overset{\mathrm{in}}{H}{}^{\mathrm{out}}_k := -\sum_{n_1, \ldots, n_k = 1}^{\infty} \overset{\mathrm{in}}{p}{}^{\mathrm{out}}_k(n_1, \ldots, n) \ln \overset{\mathrm{in}}{p}{}^{\mathrm{out}}_k(n_1, \ldots, n_k). \tag{7.8}$$

Then we define as in the classical case

$$H^{\mathrm{in}} := \lim_{k \to \infty} \frac{1}{k} H_k^{\mathrm{in}}, \qquad H^{\mathrm{out}} := \lim_{k \to \infty} \frac{1}{k} H_k^{\mathrm{out}}. \tag{7.9}$$

Since probabilities (7.7) have the usual classical sense, the convergence of information in (7.10) is secured for stationary quantum sources by the same theorem as in the classical case. The correlations of letters in k-words are, of course, given by the states $\varrho(k)$ which describe the statistics of modulation of information in the quantum communication apparatus.

Before we go over to the double source, let us discuss the connection between $\varrho^{\mathrm{in}}(k)$ and $\varrho^{\mathrm{out}}(k)$. We obtain, according to our model ($k \in N$),

$$\varrho^{\mathrm{out}}(k) = \mathscr{E}_{\mathrm{ch}}(k) \circ \mathscr{E}^{(1)}_{\mathrm{in}}(\Omega_{\mathrm{in}}) \circ \ldots \circ \mathscr{E}^{(k)}_{\mathrm{in}}(\Omega_{\mathrm{in}})\varrho^{\mathrm{in}}(k), \tag{7.10}$$

where $\mathscr{E}_{\mathrm{ch}}(k)$ is a positive linear endomorphism of $L^1(\mathscr{H}^k)$ describing the influence of the noise in the channel on the statistics of k-words. $\mathscr{E}_{\mathrm{ch}}(k)$ is as if an instrument without "selection", i.e., for $E = \Omega$, cf. (6.21).

Now the statistics of the double source "input-output", cf. Fig. 2, can be described by the following definitions, analogous to those in Section 4 and corresponding to our present quantum model:

$$p_k(n_1, \ldots, n_k;\ m_1, \ldots, m_k)$$
$$:= \operatorname{Tr}\big(\mathscr{E}^{(1)}_{\text{out}}(m_1) \circ \ldots \circ \mathscr{E}^{(k)}_{\text{out}}(m_k) \circ \mathscr{E}_{\text{ch}}(k) \circ \mathscr{E}^{(1)}_{\text{in}}(n_1) \circ \ldots \circ \mathscr{E}^{(k)}_{\text{in}}(n_k)\varrho^{\text{in}}(k)\big), \tag{7.11}$$

$$H^{\text{in-out}}_k := -\sum_{\substack{n_1, \ldots, n_k = 1\\ m_1, \ldots, m_k = 1}}^{\infty} p_k(n_1, \ldots, n_k;\ m_1, \ldots, m_k) \ln p_k(n_1, \ldots, n_k;\ m_1, \ldots, m_k), \tag{7.12}$$

$$H^{\text{in-out}} = \lim_{k\to\infty} \frac{1}{k} H^{\text{in-out}}_k. \tag{7.13}$$

We assume that the noise in the channel described by $\{\mathscr{E}_{\text{ch}}(k)\}_{k=1}^{\infty}$ is such that $H^{\text{in-out}}$ in (7.13) is finite (in general, $H^{\text{in-out}}$ can be non-negatively finite or positively infinite). We say that channel with such a noise is *thermodynamically regular.*

We can easily check from (7.7), (7.10) and (7.11) by using (6.21) and (2.17) that

$$p^{\text{in}}_k(n_1, \ldots, n_k) = \sum_{m_1, \ldots, m_k = 1}^{\infty} p_k(n_1, \ldots, n_k;\ m_1, \ldots, m_k), \tag{7.14}$$

$$p^{\text{out}}_k(m_1, \ldots, m_k) = \sum_{n_1, \ldots, n_k = 1}^{\infty} p_k(n_1, \ldots, n_k;\ m_1, \ldots, m_k), \tag{7.15}$$

i.e., that p^{in}_k and p^{out}_k are marginal probability measures of p_k, as it should be by the construction of the double source. Of course, the other conditions for $p^{\substack{\text{in}\\ \text{out}}}_k$, cf. [3], pp. 5–6 (the stationary case)

$$\begin{gathered} p^{\substack{\text{in}\\ \text{out}}}_k(n_1, \ldots, n_k) \geqslant 0, \\ p^{\substack{\text{in}\\ \text{out}}}_k(n_1, \ldots, n_k) = \sum_{n=1}^{\infty} p^{\substack{\text{in}\\ \text{out}}}_{k+1}(n_1, \ldots, n_k, n) = \sum_{n=1}^{\infty} p^{\substack{\text{in}\\ \text{out}}}_{k+1}(n, n_1, \ldots, n_k), \sum_{n=1}^{\infty} p^{\substack{\text{in}\\ \text{out}}}_1(n) = 1 \end{gathered} \tag{7.16}$$

are also satisfied, as well as the analogous ones for p_k.

Now we can define as in Section 4:

$$R^{\text{in-out}} := H^{\text{in}} + H^{\text{out}} - H^{\text{in-out}}, \tag{7.17}$$

$$H(\text{in}\,|\,\text{out}) := H^{\text{in-out}} - H^{\text{out}}, \tag{7.18}$$

$$H(\text{out}\,|\,\text{in}) := H^{\text{in-out}} - H^{\text{in}}, \tag{7.19}$$

so

$$R^{\text{in-out}} = H^{\text{in}} - H(\text{in}\,|\,\text{out}) = H^{\text{out}} - H(\text{out}\,|\,\text{in}). \tag{7.20}$$

Introducing the conditional probabilities, cf. (4.12),

$$p_k^{\text{in out}}(n_1, \dots, n_k | m_1, \dots, m_k) := \frac{p_k(n_1, \dots, n_k; m_1, \dots, m_k)}{p_k^{\text{out}}(m_1, \dots, m_k)}, \tag{7.21}$$

$$p_k^{\text{out in}}(m_1, \dots, m_k | n_1, \dots, n_k) := \frac{p_k(n_1, \dots, n_k; m_1, \dots, m_k)}{p_k^{\text{in}}(n_1, \dots, n_k)}, \tag{7.22}$$

where we assumed $p_k^{\substack{\text{in}\\\text{out}}} > 0$, we can easily obtain in the well-known way, cf. [26],

$$H_k(\text{in}|\text{out}) := - \sum_{m_1, \dots, m_k = 1}^{\infty} p^{\text{out}}(m_1, \dots, m_k) \sum_{n_1, \dots, n_k = 1}^{\infty} p^{\text{in out}}(n_1, \dots, n_k | m_1, \dots, m_k) \times \\ \times \ln p_k^{\text{in out}}(n_1, \dots, n_k | m_1, \dots, m_k). \tag{7.23}$$

$$H_k(\text{out}|\text{in}) := - \sum_{n_1, \dots, n_k = 1}^{\infty} p^{\text{in}}(n_1, \dots, n_k) \sum_{m_1, \dots, m_k = 1}^{\infty} p^{\text{out in}}(m_1, \dots, m_k | n_1, \dots, n_k) \times \\ \times \ln p_k^{\text{out in}}(m_1, \dots, m_k | n_1, \dots, n_k), \tag{7.24}$$

$$H(\text{in}|\text{out}) = \lim_{k \to \infty} \frac{1}{k} H_k(\text{in}|\text{out}), \qquad H(\text{out}|\text{in}) = \lim_{k \to \infty} \frac{1}{k} H_k(\text{out}|\text{in}). \tag{7.25}$$

Now we can write as in (4.10) for the channel capacity

$$\mathscr{C} := \sup_{H^{\text{in}}} R^{\text{in-out}} = \sup_{p^{\text{in}}} R^{\text{in-out}}. \tag{7.26}$$

We see that in such a way the quantum information theory has been as if reduced to the classical one. We have, however, to remember that under the classical "level" there exists the quantum one (as is the case of the quantum probability theory). In other words, $p_k^{\substack{\text{in}\\\text{out}}}$ and p_k are not arbitrary, but those given by (7.7) and (7.11). Thus, from this point of view, the quantum theory is a special case of the classical one. Therefore, the coding theory (the Shannon coding theorems) require a special treatment and in principle we have two distinct coding theories, classical and quantum. We plan to consider the quantum coding problem in a separate paper in future.

8. Examples

For the lack of space we cannot consider here any real application of our theory (this may be also the subject of future papers), except for checking the theory in simplest limiting cases.

The most trivial case is that of the *ideal channel* without noise and without non-trivial translation, i.e., such in which the input and output are identical and one is connected directly with the other. Then we have $\mathscr{E}_{\text{ch}}(k) = \boldsymbol{I}_k$ (= identity operator in $L^1(\mathscr{H}^k)$) and $\mathscr{E}_{\text{out}}^{(i)} = \mathscr{E}_{\text{in}}^{(i)}$, and using Theorem 6 and (6.30) we obtain

$$p_k(n_1, \dots, n_k; m_1, \dots, m_k) = \delta_{n_1 m_1} \dots \delta_{n_k m_k} p_k^{\text{in}}(n_1, \dots, n_k), \tag{8.1}$$

or

$$p_k^{\mathrm{out}|\mathrm{in}}(m_1, \ldots, m_k | n_1, \ldots, n_k) = \delta_{n_1 m_1} \ldots \delta_{n_k m_k} \tag{8.2}$$

for all $k \in \boldsymbol{N}$ and all $n_i, m_i \in \boldsymbol{N}$ ($i = 1, \ldots, k$). We have from (7.24)

$$H_k(\mathrm{out}|\mathrm{in}) = 0 \quad \text{for all } k \in \boldsymbol{N}, \quad H(\mathrm{out}|\mathrm{in}) = 0. \tag{8.3}$$

From (7.20) we obtain

$$R^{\text{in-out}} = H^{\mathrm{out}} = H^{\mathrm{in}} \tag{8.4}$$

since due to (8.1) and (7.15)

$$p_k^{\mathrm{out}}(n_1, \ldots, n_k) = p_k^{\mathrm{in}}(n_1, \ldots, n_k) \quad \text{for all } k, n_i \in \boldsymbol{N}. \tag{8.5}$$

Hence from (7.25)

$$\mathscr{C} = \infty \tag{8.6}$$

and we have no limit of transmission of information.

As a second example we shall consider the case in which the input is classical and the output is quantum. For simplicity we additionally assume that there are no correlations between letters (the so-called *Bernoulli shift*, cf. [3]). Then we have

$$p_1^{\mathrm{in}|\mathrm{out}}(m|n) = \mathrm{Tr}(P_m \varrho_n), \tag{8.7}$$

where the observable A in the output is given by (3.9), the value of the classical stochastic variable in the input is indexed by $n \in \boldsymbol{N}$, and ϱ_n denotes the output state occurring under the condition that in the input we have the nth event. We obtain

$$\begin{aligned} J_A(p) := R^{\text{in-out}} &= H_1^{\mathrm{out}} - H_1(\mathrm{out}|\mathrm{in}) = H_A\big(\varrho^{\mathrm{out}}(1)\big) - \sum_{n=1}^{\infty} p_1(n) H_A(\varrho_n) \\ &= H_A\Big(\sum_{n=1}^{\infty} p_1(n)\varrho_n\Big) - \sum_{n=1}^{\infty} p_1(n) H_A(\varrho_n), \end{aligned} \tag{8.8}$$

where $p = \{p_n\}_{n=1}^{\infty}$ and we used the first notation in (3.11). This case exactly corresponds to the case discussed by Holevo in his excellent papers [27], [28], except for the additional assumption in [27], [28] that the Hilbert space $\mathscr{H}$ is finite dimensional ($N < \infty$). By this assumption Holevo showed [35] that

$$J_A(p) \leqslant H\Big(\sum_{n=1}^{N} p_1(n)\varrho_n\Big) - \sum_{n=1}^{N} p_1(n) H(\varrho_n), \tag{8.9}$$

where H denotes von Neumann information (3.5) and where the equality sign occurs if and only if all ϱ_n mutually commute. (In the same paper also some stronger estimation of $J_A(p)$ from above has been given.) These estimations are, of course, also estimations for the channel capacity $\mathscr{C}$ (in the Bernoulli case).

Now we can mention the important extrapolation of such a theory similar to that discussed by Shannon and leading to his celebrated formula (4.11) used very much by engineers. Namely, Holevo [27] discussed instead of discrete observable A a semi-observable (6.8) based on Glauber's coherent states. Then n is replaced by a complex number $\alpha \in C$, and the Shannon Gaussian signal and noise distributions by the Gaussian distributions

in input (classical):

$$p(\alpha) = (\pi S)^{-1}\exp\left(-\frac{|\alpha|^2}{S}\right) \qquad (S > 0), \tag{8.10}$$

in output (quantum):

$$\varrho(\alpha) = (\pi N)^{-1}\int_C \exp\left(-\frac{|z-\alpha|^2}{N}\right)|z\times z|\,d^2z, \qquad (N > 0). \tag{8.11}$$

By these assumptions one obtains, [27] p. 40,

$$J_a(p) = \ln\left(1+\frac{S}{N+1}\right) \tag{8.12}$$

which is a result of the type (5.5).

Thus we see that our theory gives qualitatively reasonable results. For quantitative checking and practical applications the theory requires further investigations and mathematical development. The aim of the present paper was only to give a general formulation of the quantum information theory of the Shannon type.

REFERENCES

[1] Akhiezer, A. I., and V. B. Berestetski: *Quantum electrodynamics*, 3rd ed., Nauka, Moscow, 1969 (in Russian).
[2] Antoniewicz, R.: Reports Math. Phys. 7(1975), 289.
[3] Billingsley, P.: *Ergodic theory and information*, Wiley, New York, 1965.
[4] Birkhoff, G., and J. von Neumann: Ann. Math. **37** (1936), 823.
[5] Brillouin, L., *Science and information theory*, 2nd ed., Academic Press, New York, 1962.
[6] Davies, E. B., Commun. Math. Phys. **15** (1969), 277.
[7] —: J. Funct. Anal. **6** (1970), 318.
[8] — and J. T. Lewis: Commun. Math. Phys. **17** (1970), 239.
[9] Dobruschin, R. I.: *Arbeiten zur Informationstheorie IV*, DVW, Berlin, 1963; Russ. orig.: Usp. Mat. Nauk **14** (90) (1959), 3.
[10] Dunford, N., and J. T. Schwartz: *Linear operators*, vol. 2, Interscience, New York, 1963.
[11] Feinstein, A.: *Foundation of information theory*, McGraw-Hill, New York, 1958.
[12] Furry, W. H.: *Lectures in theoretical physics*, vol. 8, Univ. of Colorado Press, Boulder, 1966, 1.
[13] Gabor, D.: Phil. Mag. **41** (1950), 1161.
[14] —: Arch. Elektr. Übertrag. **7** (1953), 95.
[15] Gelfand, I. M., and N. Ya. Vilenkin: *Generalized functions*, vol. 4, Fizmatgiz, Moscow, 1961 (in Russian).
[16] Glauber, R. J.: Phys. Rev. **131** (1963), 109.
[17] —: *Fundamental problems in statistical physics*, ed. E.G.D. Cohen, North-Holland, Amsterdam 1968.
[18] Gleason, A. M.: J. Math. Mech. **6** (1957), 885.
[19] Goldberger, M. L., and K. M. Watson: Phys. Rev. **134** (1964), B 919.

[20] Goodman, J. W.: *Introduction to Fourier Optics*, McGraw-Hill, New York, 1968.
[21] Gordon, J. P.: Proc. IRE **50** (1962), 1898.
[22] Gudder, S., and J. P. Marchand: J. Math. Phys. **13** (1972), 799.
[23] Hanbury Brown, R., and R. Q. Twiss: Phil. Mag. **45** (1954), 663.
[24] Helstrom, C. W., *Statistical theory of signal detection*, Pergamon Press, Oxford, 1960.
[25] —, J. W. S. Lin and J. P. Gordon: Proc. IEEE **58** (1970), 1578.
[26] Henze, E.: *Einführung in die Informationstheorie*, 2 Aufl., DVW, Berlin, 1969.
[27] Holevo, A. S.: Probl. Pered. Inform. **9** (2) (1973), 31 (in Russian).
[28] —: *ibid.* **9** (3) (1973), 3 (in Russian).
[29] Ingarden, R. S.: Acta Phys. Polon. **A43** (1973), 3.
[30] —: *ibid.* **A43** (1973), 15.
[31] —, and A. Kossakowski: Bull. Acad. Polon. Sci., Ser. Math. Phys. **16** (1968), 61.
[32] —, —: Ann. Phys. (New York) **89** (1975), 451.
[33] —, and K. Urbanik: Acta Phys. Polon. **21** (1962), 281.
[34] Jadczyk, A. Z.: Bull. Acad. Polon. Sci. Ser. Mat. Phys. **22** (1974), 963.
[35] Khinchin, A. I.: *Mathematical foundations of information theory*, Dover, New York, 1957, Russ. orig.: Usp. Fiz. Nauk **8** (3) (1953) 3; **11** (1) (1956), 17.
[36] Kennedy, R. S.: Proc. IEEE **58** (1970), 1951.
[37] Klauder, J. R., and E. C. G. Sudarshan: *Fundamentals of quantum optics*, Benjamin, New York, 1968.
[38] Kolmogorov, A. N.: *Arbeiten zur Informationstheorie I*, 3 Aufl., DVW, Berlin, 1967, 91, Russ. orig.: *Theory of information transmission*, Akad. N. SSSR, Moscow, 1956.
[39] Kossakowski, A.: Reports Math. Phys. **3** (1972), 247.
[40] Landau, L., and Lifshitz, E.: *Statisticalphysics*, Nauka, Moscow, 1964 (in Russian).
[41] Landau, L., and R. Peierls: Z. Phys. **62** (1930), 188.
[42] Lebedev, D. S., and L. B. Levitin: Inform. Control **9** (1966), 1.
[43] Lüders, G.: Ann. Phys. (Leipzig) **8** (1951), 322.
[44] Mityugov, V. V: Probl. Pered. Inform. **2** (3) (1967), 86 (in Russian).
[45] Mlak, W.: *Introduction to the theory of Hilbert spaces*, PWN, Warsaw, 1970 (in Polish).
[46] Nakamura, M., and T. Turumaru: Tohôku Math. J. **6** (1954), 182.
[47] —, and H. Umegaki, Math. Japon. **7** (1962), 151.
[48] Nyquist, N.: Phys. Rev. **32** (1928), 110.
[49] Ochs, W.: Reports Math. Phys. **8** (1975), 109.
[50] Raisbeck, G.: *Information theory. An introduction for scientists and engineers*, MIT Press, Cambridge, 1963.
[51] Rényi, A.: *Wahrscheinlichkeitsrechnung mit einem Anhang über Informationstheorie*, DVW, Berlin, 1962.
[52] Ross, M.: *Laser receivers*, Wiley, New York, 1966.
[53] —: *Laser applications*, ed. M. Ross, vol. 1, Academic Press, New York, 1971.
[54] Ruskai, M. B.: Ann. Inst. H. Poincaré **A19** (1973), 387.
[55] Segal, I. E.: Ann. Math. **57** (1953), 401.
[56] —: J. Math. Mech. **9** (1960), 623.
[57] Shannon, C. E.: Bell Syst. Techn. J. **27** (1948), 379, 623.
[58] Stratonovich, R. L.: Zh. Eksp. Teor. Fiz. **28** (1955), 547 (in Russian).
[59] —: Probl. Inform. **2** (2) (1966), 45 (in Russian).
[60] Sudarshan, E. C. G., P. M. Mathews and J. Rau: Phys. Rev. **121** (1961), 920.
[61] Takahashi, H.: *Advances in communication systems theory and applications*, ed. A. V. Balakrishnan vol. 1, Academic Press, New York, 1965.
[62] Umegaki, H.: Tohôku Math. J. **6** (1954), 177.

[63] Urbanik, K.: Studia Math. **21** (1961), 117.
[64] Varadarajan, V. S.: Commun. Pure Appl. Math. **15** (1962), 189, corr. **18** (1965).
[65] —: *Geometry of quantum theory*, vol. 1, Van Nostrand, Princeton, 1968.
[66] von Neumann, J. Göttingen Nach. (1927), 273, *Collected Works*, ed. A. H. Taub, vol. 1, Pergamon Press, Oxford, 1961, 236.
[67] —: *Mathematical foundations of quantum mechanics*, Princeton Univ. Press, Princeton, 1955, Germ. orig.: *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin, 1932.
[68] Wehrl, A.: IUPAP Conf. Stat. Phys. Budapest 1975, Abstracts, Akad. Kiado, Budapest, 1975, 39.
[69] —: *Two theorems about entropy and convergence of density matrices*, Preprint Univ. Vienna, 1975, Commun. Math. Phys., to appear.
[70] Wolfowitz, J.: *Coding theorems of information theory*, 2nd ed., Springer, Berlin, 1969.
[71] Yosida, K.: *Functional analysis*, Springer, Berlin, 1965.