J. S. BARAS

NONCOMMUTATIVE PROBABILITY MODELS IN QUANTUM COMMUNICATION AND MULTI-AGENT STOCHASTIC CONTROL

Estratto da: *Ricerche di Automatica*
Vol. 10, No. 2, December 1979

ODERISI GUBBIO 1981

RICERCHE DI AUTOMATICA
Vol. 10, No. 2, December 1979
Printed in Italy

NONCOMMUTATIVE PROBABILITY MODELS IN QUANTUM COMMUNICATION AND MULTI-AGENT STOCHASTIC CONTROL

J. S. BARAS

Abstract. In this paper we present a survey of basic results in quantum communication theory that utilize heavily so called noncommutative probability models. We indicate by means of examples how these mathematical methods can lead to actual computations in specific examples. We offer a careful review of these non-commutative models and we observe the similarities with desired features in the representation of the statistics in multi-agent stochastic control problems. We then give suggestions for interpreting some of the basic constructs of quantum models in the language of multi-agent stochastic control systems.

1. Introduction.

In stochastic control problems the main objective is to achieve satisfactory performance of a system operating in an uncertain environment. Typically, performance is measured by the expected value of a performance criterion (or cost function). In classical stochastic control there is one controller (control station, control agent) and one performance measure. The controller employs sensors to perform measurements on the system, collects and stores the resulting data (observations) which are subsequently processed in actuators to produce the decisions (inputs) that optimize the performance criterion. This process is customarily identified as employing feedback (more precisely information feedback)

Manuscript received June, 15, 1979, and in revised form February 15, 1980.
The Author is with the Electrical Engineering Department, University of Maryland, College Park, Md. 20742, U. S. A.
Research supported partially by the National Science Foundation under Grant ENG 75-20900, and by the Army Research Office contract ARO DAAG 29-77-C-0042 while visiting the Information Systems Laboratory, Stanford University.

from the system. The mathematical formulation of this classical stochastic control problem is thought to be well understood to date but explicit solutions are known only for few special cases. For example, in [1] results on existence of optimal strategies are presented for the case where decisions (or controls) affect only the performance measure (or criterion) and not the trajectory of a stochastic dynamical system, while the information available for use in feedback is arbitrary but prescribed in advance. In [2], via the celebrated theorem of Girsanov existence results are obtained for the case where decisions affect the performance measure and the « state » satisfies an Itô type stochastic differential equation. The case where the information available to the controller is explicitly generated by some sort of noisy observation of the « state » (i. e. the case of so called partially observable stochastic systems) is treated in [3-5]. Different approaches to this problem can be found in the books [6-9], at various levels of generality and mathematical sophistication. An interesting approach to obtain explicit solutions to some stochastic control problems appears in Beneš [10]. In classical stochastic control there is no concern for the interaction between information (and its transmission) and control. Furthermore, there is no interaction between measurement process and system dynamics and typically selection of measurements (or observations) is not part of the problem. If we can obtain an explicit solution there is not much difficulty in implementing the optimal controller which does not depend explicitly on the observation model (in the sense that the latter is fixed). Available information is modelled by σ-algebras, the system's « state » by a vector valued stochastic process, but is not in general well defined and understood [14].

Serious complications arise, however, when one considers the control of a stochastic system by various controllers with different available information and possibly different criteria. Such problems go under the categories of non-classical information patterns, stochastic control of large systems, decentralized or hierarchical control, etc. The recent special issue [16] and in particular the review article [17] contain a wealth of information about the current status of such problems and the major unresolved issues. Thus seemingly simple problems lead to major departures from classical stochastic control results (c. f. Witsenhausen's well known and often quoted counterexample [11]). It is fair to state that despite many worthwhile and enlightening contributions by several people the major questions still remain unanswered. In our opinion there are two fundamental problems, whose resolution is widely recognized as key to further progress:

(*a*) The interaction between information and control. Under this heading are such problems as communication between controllers via « signaling strategies », « information neighborhoods » for controllers, cost of information versus cost of control ([16-19], and the references therein). Despite the pioneering work of Witsenhausen [11-15], who developed a number of important formulations and results about the separation of the use of information (i. e. estimation) and control, it is the author's opinion that there does not exist to date an agreed upon and satisfactory formulation of the joint « optimization » problem in information flow and control. However, recent results by Whittle and Rudge [20-21] and by Ho, Kastner and Wong [22-23] suggest that, through a combination of methods from information theory and control, progress can be achieved in certain cases, although basically these results give up the « real-time » character of controls in stochastic control problems.

(*b*) The concept of « state » for such a system is not well understood, although Witsenhausen [13-14] gives results towards a resolution of this problem. The problem stems primarily from the fact that in multi-agent stochastic problems there need not be a preassigned total time order of actions. In fact, action times may depend on observations and controls by the same or other agents. This problem is also related to the availability (theoretically, despite the obvious computational complexity) of a dynamic programming algorithm for solution.

In the present paper it is our aim to present a third major problem that may appear in stochastic control problems with many agents, which has not been emphasized to date. This is centered around the possible interactions between measurements by different agents and between system dynamics and measurements. We shall see that these concepts are related to some of the difficulties encountered to date and are akin to *very strong interaction between information and control.* This is typically the case where one cannot prove existence of an optimal control law (or strategy or design) [19]. We point out that similar problems appear in communication problems with quantum mechanical signal and noise models. We offer a brief review of some of the major results in this area which is known as « quantum detection and estimation». Pursuing further the similarities between these two different problems we suggest certain new formulations for the problems of interaction between information and control, and system dynamics and measurement inspired by the methodologies used in the quantum communication theory. It is seen that a « non-commutative » probability theory (in the sense the term is used in the axiomatic foundation of quantum mechanics [54-58]) may be necessary for some of these problems. Our primary objective is to stimulate further investigation of the conceptual similarities unraveled here between quantum physics and stochastic systems control, by researchers in these different fields. In particular, we believe that system theorists would benefit by close study of the vast mathematical arsenal that physicists have so successfully exploited in their problem areas.

2. Multi-agent stochastic control problems.

2.1. Heuristics.

We are interested in decentralized control problems where one wishes to specify an optimal design, in Witsenhausen's terminology [12] [13] but with the additional characteristic that there is some flexibility over the information pattern. By that we mean that there are available several alternatives for the information available to each controller, on which decisions have to be based. What can be said abstractly about the joint selection of information and control pattern? Obviously, we are not interested in an exhaustive search between all possible information patterns. We are also interested in problems where there is no strict preassigned order of action times to be followed by the various controllers (or agents). Both phenomena occur quite naturally in systems with very large number of agents. Thus in economic systems each agent (which can be an individual or an organization) has ample choice of sets of data on which to base decisions (consider, for example, the various economic indicators or statistical data reduction results available to the public and the government). Furthermore it does not appear that there exists a strict preassigned order of action times in economic systems. In systems with such properties, we encounter a new kind of difficulty. Mainly, since control actions by one controller affect the measurements (or observations) of another there may very well exist situations where efforts by two agents (by choice of information and control) to obtain as accurate as possible values for two critical (for their actions) variables will be in conflict, resulting in the impossibility of such *simultaneous* accuracy. This certainly requires very strong information-control interaction. Finally, we are interested in systems where the agents can anticipate certain control actions by other agents when they are informed of the type, *and not the results,* of measurements performed by these other agents. Such so called *anticipatory* systems have been studied by others [25]. Often it is possible to give a statistical description of such a system's reaction to measurement. Roughly, the system's « state » changes due to the measurement performed. It is quite interesting to note that such properties were used by Wigner [25] [31, p. 187] to produce a well known (but rather easily resolved) paradox in quantum physics. Such phenomena can be seen again in economic systems where, for example, measurement of income levels for taxation may have adverse effects on productivity. In another setting the knowledge by a driver that a traffic detector exists nearby, may cause changes in his velocity in order to, for example, catch the green or yellow at an intersection. Some immediate questions of interest are: i) What are the implications of such phenomena on the probabilistic models used in stochastic control? ii) Are there any assumptions that will permit a reasonable definition of a « state »? iii) How can we formulate optimization problems for such control systems? iv) Can we solve any?

It is not our objective to provide complete answers to any of the above questions, we do not believe this is feasible. Instead, we will promote the thesis that for several classes of such systems, certain models and methodologies from quantum physics can be successfully employed. We shall also make an effort to discover and isolate assumptions (or properties) that permit the use of such models.

2.2. The need for a non-commutative structure.

The heuristic discussion of the previous section suggests that a careful examination of the information available for decisions and a precise description of its place in the mathematical formulation are necessary for further understanding of these problems. Witsenhausen in [11] [13] proposed a model for doing this. According to that model the system's dynamics are determined by the realizations of the noise variables and all control variables. The performance measure can also be expressed as a function of these same variables (by solving the system's equations). Finally for each decision, the data available for that decision, are functions of these same variables, and define a σ-field in an appropriate space. The complete specification of the control problem consists, according to [13], of the specification of the performance measure and these σ-fields.

Following [11], [13] let $(\Omega, \mathcal{B}, P)$ be the probability space for the intrinsic random variables of the system (called actions of nature in [13]) and suppose we have a finite set A of agents acting on the system. In this formulation a controller acting at two different times will be considered as two different agents. Let $(U_a, \mathcal{F}_a)$ be the measurable space in which agent $a \in A$ selects his control action u_a. Considering product sets and product σ-fields, to a subset B of agents we associate the set

$$H_B = \Omega \times \prod_{a \in B} U_a \tag{2.1}$$

and the σ-field

$$\mathcal{F}_B = \mathcal{B} \times \prod_{a \in B} \mathcal{F}_a. \tag{2.2}$$

All σ-fields (2.2) are considered as subfields of $\mathcal{F}_A$ using the natural projection of H_A onto H_B. The information available to agent a is characterized by a subfield $\mathcal{J}_a$ of $\mathcal{F}_A$. The possible control laws for agent a are the functions $\gamma_a: H_A \to U_a$ which are measurable from $\mathcal{J}_a$ to $\mathcal{F}_a$, they form the set Γ_a. Then the control for the subset of agents $B \subset A$ can be chosen from $\Gamma_B = \prod_{a \in B} \Gamma_a$ and the whole design in Γ_A. Witsenhausen then goes on to characterize various types of *information patterns* (i. e. the collection of $\mathcal{J}_a$), such as causal, classical, quasi classical, without self-information etc.

Our first point is that in a well-posed multi-agent stochastic control problem, the agents will make inferences about system variables based on their own information fields $\mathcal{J}_a$ (appropriately pulled back in $\mathcal{B}$). That is the agents will compute conditional expectations (and/or probabilities) either implicitly or explicitly. To us one important difference between classical and nonclassical information patterns is that these operations *commute* in a classical pattern and *do not commute* in a nonclassical pattern. Indeed according to [13] an information pattern is *classical* if it is *sequential* (i. e. there is an ordering $(a_1, a_2, \dots, a_n)$ of A such that $\mathcal{J}_{a_k} \subset \mathcal{F}_{\{a_1, \dots, a_{k-1}\}}$ for $1 \leq k \leq n$) and $\mathcal{J}_{a_1} \subset \mathcal{F}_{\phi}$, $\mathcal{J}_{a_{k-1}} \subset \mathcal{J}_{a_k}$ for $k=2, \dots, n$. Then the commutativity of conditional expectations is just a consequence of the smoothing property of conditional expectations [26]. The statement for the nonclassical patterns is also obvious. The non-commutative modifier of our title refers to the corresponding property of conditional expectations. We adopt the point of view that the σ-fields $\mathcal{J}_a$ forming the information pattern are generated (or correspond to) by measurements (or observations) performed on the system. The natural question is then: Does there exist a different model than the one described above, which can describe statistically the events (observations) associated with a multi-agent stochastic control problem, including an intrinsically non-commutative conditional expectation operation? We shall see in section 4 that there is an affirmative answer, and that a prime example of such probability models is the von-Neumann model of quantum mechanics [27].

Second, we firmly believe that one of the deficiencies in the current development of decentralized stochastic control is that the information pattern is assumed fixed and given apriori. Little effort has been directed towards « optimal » selection of information pattern. The articles that address such problems (see references in [12] and [17]), use a formulation that represents the choice of information pattern as an optimization problem over a finite set of parameters. The usual method of attacking such problems is to solve a parametrized family of stochastic control problems, then select the parameters which result in better performance, and thus choose the corresponding information pattern. This approach has been followed for example in problems of optimizing sensor locations in distributed systems, problems with considerable communication costs between agents, optimal selection of parameters in the output equations (corresponding to sensor optimization). It is important to realize that in the design of a distributed control scheme the choice of information pattern is at least as important a task as the choice of control actions. However a proper formulation of the joint optimization problem escapes us todate. One reason is that all formulations of multi-agent stochastic control problems are unbalanced, in the sense that controls dominate while the information pattern is carried at best as a set of parameters. Consider how difficult would be to pose such a problem in Witsenhausen's model [13]. From practical examples it is well known, that a wisely chosen information pattern can reduce considerably the control optimization task. The non-commutative probability model of quantum mechanics developed by von-Neumann [27] allows a better formulation of the measurement selection problem. Properly extended to current axiomatic models of quantum physics, it contains several suggestions in formulating the information pattern selection problem. We discuss these ideas in section 4. There are however differences between the quantum mechanical logic (or propositional calculus) and a proper probabilistic model for a multi-agent stochastic control problem. The most important one is that the non-commutative model was created in quantum mechanics to characterize the *passive* interaction between a measurement process and a system.

In stochastic control however we have in addition the *active* operation of controls which we try to use to influence the system's behavior in addition to changes in the system due to measurements.

3. **Quantum communication.**

3.1. Classical versus « quantum » noise and signals.

In classical communication theory, the signal (message) to be transmitted is used to « modulate » a carrier signal which typically is an electromagnetic wave at radio frequency (low frequency end of spectrum used for telecommunication), which (modulated carrier) is then propagated through a communication channel to the receiver. The function of the receiver is then to resolve the distortion introduced into the signal due to propagation and reception, in a way that minimizes a given distortion measure, and recover the transmitted signal. At radio frequencies we have thermal noise as the primary source of distortion in the received signal. This may be intrinsic thermal noise or thermal noise processed by some linear or nonlinear electronic device. A fundamental assumption in classical communication systems design is that the optimal receiver does not depend on the channel model and the noise model. This is primarily due to the suppression of any nonreversible interaction between measuring apparatus and the carrier field, and is a consequence of the noise (random contamination) model assumed.

In communication at optical frequencies such as laser communication systems, however, a completely different noise model is necessary in order to explain observed experimental data. So theory [29] [30] predicts that as the frequency increases thermal noise power decreases, but a new type of noise appears, the so called « quantum noise ». The most common manifestations of quantum noise are in the spontaneous emission by a laser amplifier and in the detection of light by a photodetector. Both noises are ultimately determined by the measuring process of quantized radiation and, therefore, are quantum phenomena. It is, therefore, necessary to bring quantum physics into the description of the receiver with the result being that the measurement outcomes are random (no matter what measurement we perform) with statistics dependent on the measurement performed. It is, as a consequence, no longer true that the receiver can be constructed independently of the signal and noise model.

Glauber's article [28] contains a very enlightening discussion of the differences between noise and signal models in the radio (low) and light (high) frequencies end of the telecommunications spectrum. An additional reference is [29]. Briefly, there are two main differences: *a*) at the low frequency end of the electromagnetic spectrum we have large quantum densities of nonenergetic quanta while at high frequencies we have very small quantum densities of highly energetic quanta; *b*) at low frequencies we have almost complete spatial and temporal control of the waveforms we generate, while at high (optical) frequencies we loose this possibility rather quickly (for example, we can have spatial control of the field but the amplitudes tend to fluctuate uncontrollably). Both *a*) and *b*) are manifestations of the uncertainty principle of quantum mechanics [30] [31] which can also be thought of as measurement-control interactions or as state-measurement interactions (in the language of stochastic control).

Clearly, the same remarks are valid for the transmitter (or source), since measurement-control interaction manifests itself in the modulation of a carrier signal at optical frequencies.

In conclusion, at high (optical) frequencies the interaction between control variables affecting measured variables of an optical field requires the introduction of quantum physics laws for proper treatment. We shall see that this is a result of the different, nonclassical structure of the fundamental propositional logic which has to be created to explain these interactions. In this section we present an outline of the major results obtained in the simplest of communication problems, that of signal detection, when the classical Hilbert space model of quantum physics is used. While doing this we emphasize and interpret the differences from the classical theory.

Comprehensive references on Quantum Detection are the recent monographs by Helstrom [32] on physics-communications aspects and by A. S. Holevo [33] on mathematical aspects. In the author's opinion, the best exposition of some of the difficult optimization problems involved for the Hilbert space model of quantum physics can be found in the thesis of S. Young [34]. Holevo's monograph treats a more general model however (operator algebra model).

3.2. Classical *M*-ary detection.

In the classical formulation of detection theory (Bayesian hypothesis testing) [32] a certain « system » is observed and we obtain n numbers

$v_1, \ldots, v_n$ on the basis of which we have to decide about the « state » of the system. The system may be in one of M states and we call hypothesis H_j, $j=1, \ldots, M$ the proposition « The system is in state j ». The observed data vector $v=\{v_1, v_2, \ldots, v_m\}^T$ is random with probability density $p_j(v)$ depending on the hypothesis. (Note the « system » may be absolutely fictitious). We also assume we know the prior probabilities ζ_i of hypothesis H_i being correct. We then let C_{ij} be the cost incurred by choosing hypothesis H_i when H_j is true. A *decision strategy* is a rule which assigns a hypothesis H_i, to be chosen, for every value of observed data v. A *randomized decision strategy* consists of M functions $\pi_i(v)$ such that

$$0 \leq \pi_i(v) \leq 1, \quad \sum_{i=1}^{M} \pi_i(v) = 1, \tag{3.1}$$

where

$$\pi_i(v) = Pr\{H_i \text{ is chosen when observed data vector is } v\}.$$

Then the « risk function » for hypothesis H_i is

$$W_i(v) = \sum_{j=1}^{M} \zeta_j C_{ij} p_j(v) \tag{3.2}$$

and the average Bayes cost for a strategy is

$$J = \int_{\mathbf{R}^n} \sum_{i=1}^{M} W_i(v) \pi_i(v) \, dv. \tag{3.3}$$

The solution to this classical problem (which is just a convex optimization problem) implies that the optimal decision strategy is

$$\begin{gathered} \pi_j^*(v) = 1, \quad \pi_i^*(v) \equiv 0 \quad i \neq j, \\ \text{at all } v \in \mathbf{R}^n \text{ such that } W_j(v) < W_i(v) \quad i \neq j. \end{gathered} \tag{3.4}$$

That is a *pure strategy*, not a randomized one. The effect of including randomized strategies is in this case to convexify the set of admissible strategies, but since the extreme points [36] are pure strategies, it is done here only for mathematical convenience and has no further significance. As a consequence there is no difficulty in implementing this classical optimal decision strategy. If we let

$$Y(v) = \min_j W_j(v) \tag{3.5}$$

the above solution is equivalent to

$$[W_i(v)-Y(v)]\,\pi_i(v)=0$$
$$W_i(v)\geq Y(v) \tag{3.6}$$

for all v, i. Furthermore,

$$Y(v)=\sum_{i=1}^{M} W_i(v)\,\pi_i(v) \tag{3.7}$$

and

$$J_{\min}=\int_{\mathbf{R}^n} Y(v)\,dv. \tag{3.8}$$

3.3. Quantum States and Measurements.

In the quantum formulation of the same detection problem, « system » states and observations have to be described according to the laws of quantum physics. Now in the Hilbert space model of quantum mechanics, we associate with a quantum system [30-32] a complex Hilbert space $\mathcal{H}$. Let us denote by $\mathcal{L}(\mathcal{H})$ ($\mathcal{L}_s(\mathcal{H})$) the space of all bounded (and selfadjoint) operators on $\mathcal{H}$ [30]; by $\mathcal{T}(\mathcal{H})$ ($\mathcal{T}_s(\mathcal{H})$) the space of all trace-class (and selfadjoint) operators on $\mathcal{H}$; by $\mathcal{L}_s^+(\mathcal{H})$, $\mathcal{T}_s^+(\mathcal{H})$ the nonnegative operators in $\mathcal{L}_s(\mathcal{H})$, $\mathcal{T}_s(\mathcal{H})$ [37]. The basic notion in quantum mechanics is that of an observable which is thought of as corresponding to some physical measurable variable (i. e. energy, momentum, location). Observables actually form the starting point in the mathematical development of quantum models in certain studies (see for example Segal [38]). In the Hilbert space model the observables are represented as elements of $\mathcal{L}_s(\mathcal{H})$. Since an observable A is selfadjoint it has a corresponding projection valued measure (PVM) E_A via the spectral theorem. Recall that an operator $B\in\mathcal{L}(\mathcal{H})$ is trace-class if it is compact and

$$\sum_{i=1}^{\infty} \lambda_i(A^*A)^{1/2}<\infty \tag{3.9}$$

where * denotes adjoint and $\lambda_i(B)$ the ith eigenvalue of the operator B. The state of a quantum system is represented by an operator $\rho\in\mathcal{T}_s^+(\mathcal{H})$ with $Tr[\rho]=1$, the so called *density operator*. Here

$$Tr: \mathcal{T}(\mathcal{H}) \rightarrow \mathbf{C}$$
$$Tr[A] = \sum_n \langle A\phi_n, \phi_n \rangle \tag{3.10}$$

and is a linear functional; in (3.10) ϕ_n is any orthonormal sequence and $\langle \cdot, \cdot \rangle$ the inner product in $\mathcal{H}$. Now since ρ is compact, selfadjoint and nonnegative it has a sequence of nonnegative real eigenvalues ν_i. Furthermore since ρ is trace-class

$$0 \leq \nu_i \leq 1, \quad \sum_i \nu_i = 1 \tag{3.11}$$

and we have

$$\rho = \sum_i \nu_i \langle \cdot, \psi_i \rangle \psi_i, \tag{3.12}$$

where ψ_i are the normalized eigenvectors of ρ. In classical quantum mechanics the unit vectors of $\mathcal{H}$ are the so called pure states. The fundamental postulate being that if we measure the variable v, which is represented by the observable $V \in \mathcal{L}_s(\mathcal{H})$, while the system is in pure state ψ, we do not get exact values but instead several values with certain probabilities. The distribution function giving the statistics of the measurement outcome when the system is in pure state ψ is

$$F_v(\xi) = Tr[E_V(-\infty, \xi] \langle \cdot, \psi \rangle \psi] \tag{3.13}$$

where E_V is the PVM associated with V. *The model thus, via the simple rule* (3.13), *avoids detailed description of the numerous possibilities of microscopic state transitions in the measurement-system interaction, and instead provides statistical information which allows computations of macroscopic variables* (*such as moments*) *that can be recorded by the instruments.* Often, and in particular in communication problems we do not know precisely the state of the system, we rather have some prior probabilities about the system being in various states. This is exactly the interpretation of the real scalars ν_i in (3.11) (3.12). As a consequence ρ in (3.12) is also called a « mixture state » [30, 31].

The modelling of measurements via projection valued measures (see (3.13)) is inadequate for communication problems. This was demonstrated by Holevo [39] and independently by Davies [40]. Davies considered repeated measurements on a quantum system and examined the difficulties arising from the notorious « reduction of the wave packet » formula. Conventionally it is assumed [37] that if a measurement of the observable A with discrete spectrum $\{\lambda_i\}$ (assumed for simplicity here) is performed on a quantum system at state ρ, the state immediately after the measurement is

$$\rho_i = \frac{P_i \rho P_i}{Tr\,[\rho P_i]} \tag{3.14}$$

where P_i is the spectral projection corresponding to λ_i. Suppose the first measurement is immediately followed by a second, represented by B with eigenvalues μ_j and spectral projections Q_j. Now the probability of obtaining λ_i at the first measurement and μ_j at the second is

$$\left.\begin{aligned} p_{ij} &= Tr\,[\rho P_i Q_j P_i] \\ \sum_{i,j} p_{ij} &= 1. \end{aligned}\right\} \tag{3.15}$$

Now an important consequence of quantum physics, is that there are incompatible observables; that is there are variables which cannot be measured simultaneously with any desired accuracy. In the Hilbert space model this is equivalent to noncommutativity of A and B. If then A and B are not compatible what is an appropriate model for their joint statistics? Well (3.15) suggests defining

$$R: \begin{array}{c}\text{subsets of}\\ \mathbf{Z}\times\mathbf{Z}\end{array} \rightarrow \mathcal{L}(\mathcal{H}) \tag{3.16}$$

$$R(E) = \Sigma\,\{P_i Q_j P_i;\ (i,j)\in E\}$$

(where $\mathbf{Z}$ is the set of integers) which has the properties: $R(E)\geq 0$, $R(\phi)=0$, $R(\mathbf{Z}\times\mathbf{Z})=1$, $R(\coprod_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} R(E_n)$ in the strong operator topology ($\coprod$ denotes disjoint union). That is R is a probability operator measure *POM* [37, 41]. Furthermore

$$Pr\left\{\begin{array}{l}\text{repeated measurement outcome}\\ \text{is in } E,\text{ given state } \rho\end{array}\right\} = Tr\,[\rho R(E)]. \tag{3.17}$$

So it is possible to define a joint distribution for two non-commuting discrete observables! In general this *POM* depends on which of the two measurements is made first. Furthermore the state-measurement interaction

$$\rho \rightarrow \frac{Q_j P_i \rho P_i Q_j}{tr\,[Q_j P_i \rho P_i Q_j]} \tag{3.18}$$

can be completely characterized by the map

$$S_E(\rho)=\Sigma\{Q_j P_i \rho P_i Q_j:\ (i, j)\in E\}, \tag{3.19}$$

which defines (modulo an obvious normalization) the change of state induced by the composite measurement conditioned upon the outcome lying in the set E. Clearly S_E is a positive linear map and it contains all the information about the measurement *as well as its interaction with states.* Such maps were called *operations* by Haag and Kastler [42]. We refer to the excellent treatment by Davies [37] for further developments along these lines. We would like only to emphasize that *the notions of operation and POM were introduced in order to handle interactions between noncompatible measurements and between states and measurements, in quantum systems.*

This seems to be an appropriate place for introducing the formal mathematical definition of a probability operator measure. A *POM* with values in the measurable space $(U, \mathcal{B})$ [41] (recall measurable space = a set U with a σ-algebra of subsets of $U, \mathcal{B}$) is a map $M: \mathcal{B} \rightarrow \mathcal{L}(\mathcal{H})$ such that

$$\left.\begin{array}{l}\text{(i)}\quad M(B)\geq 0, \quad \forall B\in\mathcal{B}\\ \text{(ii)}\quad \text{if } \{B_i\} \text{ is a partition of } U \text{ then}\\ \sum_i M(B_i)=I \text{ (in the strong operator topology on } \mathcal{H}).\end{array}\right\} \tag{3.20}$$

That is a *POM* is a *positive operator valued* measure [41, p. 6] such that $M(U)=I$. It is also called a *generalized resolution of the identity* [44, p. 121]. It is worth noting that if M is an *orthogonal resolution of the identity,* i. e. if in addition $B\cap C=\phi$, for $B, C\in\mathcal{B}$, implies $M(B)\,M(C)=0$, then M is necessarily a spectral measure (or *PVM*) [41, p. 12], i. e. $M(B)$ is a projection $\forall B\in\mathcal{B}$. Holevo [43, p. 341] termed such *POM*'s *measurements* for quantum systems. *PVM*'s are termed *simple* measurements. As in (3.17) a measurement represented by the *POM* M, when the quantum system is in state ρ, produces a random variable v whose statistics are described by the probability measure

$$\mu(B)=Tr\,[\rho\, M(B)] \tag{3.21}$$

$\forall B\in\mathcal{B}$.

Davies [40], [45] and Benioff [46-48] have shown that the joint statistics of the sequence of outcomes produced by *any* sequential measuring procedure (involving simple measurements, or observables) on *any* quantum system can always be described by a *POM* via a formula like (3.21). Note that this result is an extension of the example above involving two simple measurements. In [45] Davies and Lewis showed that von Neumann's repeatability hypothesis is too restrictive in describing statistics of repeated measurements and its abandonment leads to the *POM* model. This representation is valid even if the observables to be measured at a later stage of this process are chosen on the basis of the outcomes of earlier measurements [47], and even if the quantum system evolves in time, between measurement times, according to the laws of quantum mechanics (i. e. Schrödinger's equation or its consequences). Thus the notion of a *POM* encompasses the statistical description of the outcomes of every conceivable measurement process. However, whether or not there exists an instrument which realizes these statistics and how to construct it are questions we know *very little* about.

3.4. Quantum M-ary detection.

Holevo arrived at this generalization of a quantum measurement by considering the quantum formulation of the M-ary detection problem [49]. In analogy with the classical detection problem a certain quantum « system » is observed and we obtain the outcome v. The system may be in one of M states which are represented by density operators ρ_i in $\mathcal{T}_s^+(\mathcal{H})$ corresponding to the hypotheses $H_1, \ldots, H_M$. In view of the interpretation of ν_i in (3.11) the operators ρ_i are the analogues of the densities p_i in the classical problem. Now if the outcomes v are represented in the quantum model by a *POM* M, then $p_i(v)\,dv = Tr\,[\rho_i M(dv)]$.

Then for a randomized decision strategy (3.1) the average Bayes cost from (3.3) is

$$
\begin{aligned}
J &= \int_{\mathbf{R}^n} \sum_{i=1}^{M} \sum_{j=1}^{M} \zeta_j C_{ij}\, Tr\,[\rho_j M(dv)]\, \pi_i(v) \\
&= Tr \sum_{i=1}^{M} W_i \Pi_i
\end{aligned}
\tag{3.22}
$$

where $W_i \in \mathcal{T}_s(\mathcal{H})$ are the « risk operators » [32]

$$
W_i = \sum_{j=1}^{M} \zeta_j C_{ij} \rho_j, \quad i = 1, \ldots, M \tag{3.23}
$$

and Π_i are the « detection operators » [32]

$$\Pi_i = \int_{\mathbf{R}^n} \pi_i(v)\, M(dv). \tag{3.24}$$

Here it is easy to see that Π_i, $i=1, \dots, M$, are self-adjoint, non-negative and

$$\sum_{i=1}^{M} \Pi_i = I. \tag{3.25}$$

That is the set Π_i, $i=1, \dots, M$ forms a *POM* on $\mathcal{H}$. Note that since p_i depends on M, for the quantum problem it is necessary to specify not only the scheme for processing the observed data (i. e. the functions $\pi_i(\cdot)$) but also the measurement to be performed. It is in this sense that the receiver depends on the noise and signal model. It is also important to emphasize that the detection operators form a *POM* and not a *PVM*. For, an instrument that simply guesses which hypothesis might be true, selecting an arbitrary one with probability $1/M$ without even interacting with the system corresponds to a measurement with detection operators

$$\Pi_i = \frac{1}{M}\, I, \quad i=1, \dots, M$$

which certainly are not projections.

Thus mathematically the solution of the quantum M-ary detection problem is to select a *POM* Π_i, $i=1, \dots, M$ to minimize the Bayes cost (3.22). This is a linear programming problem with convex restraint set. The duality theory is delicate and for the details we refer to [34]. The solution is described by (see also [50], [35]):

THEOREM 3.4.1. *There exists a solution to the problem*

$$\min Tr \sum_{i=1}^{M} W_i \Pi_i$$

over all M-component POM's, where $W_i \in \mathcal{T}_s(\mathcal{H})$. *A necessary and sufficient condition for the POM* Π_i^*, $i=1, \dots, M$, *to be optimal is that*

(i) $\sum_{j=1}^{M} W_j \Pi_j \leq W_i$, $i=1, \dots, M$

or

$$\text{(ii)} \quad \sum_{j=1}^{M} \Pi_j W_j \leq W_i, \ i=1, \dots, M.$$

Furthermore, under any of the above conditions the operator

$$Y = \sum_{j=1}^{M} W_j \Pi_j = \sum_{j=1}^{M} \Pi_j W_j \tag{3.26}$$

is self-adjoint and is the unique solution of the dual problem.

It is easy to see that in view of (3.26) (i) or (ii) are equivalent to Y (in (3.26)) being self-adjoint and

$$W_i \geq Y, \quad i=1, \dots, M. \tag{3.27}$$

Then (3.26) (3.27) imply

$$(W_i - Y)\, \Pi_i = \Pi_i\, (W_i - Y) = 0, \quad i=1, 2, \dots, M. \tag{3.28}$$

Furthermore,

$$J_{\min} = Tr\, Y. \tag{3.29}$$

We observe that (3.26)-(3.29) are very similar to the equations describing the solution to the classical problem (3.6)-(3.8). There are, however, several fundamental differences, which we now discuss.

First, as we mentioned above the design of an optimal receiver consists of the *simultaneously optimal* choice of a measurement process *and* a data processing scheme.

Second, the consideration of measurements (i. e. *POM*'s) instead of just simple measurements (i. e. *PVM*'s) is not done just to convexify the problem. *The extreme points of the optimization problem considered here, are not PVM's.* To see this consider the simple example [43] of linearly polarized photons, with polarization angle equal to $\theta_k = 2\pi(k-1)/M$, $k=1, 2, \dots, M$ with equal probabilities $\zeta_k = 1/M$, and take the simple cost assignment $C_{ij} = 1 - \delta_{ij}$. Let $\mathcal{H}$ be two dimensional complex space $\mathbf{C}^2$, E_i, $i=1, \dots, M$ the projections on M directions of polarization. Hypothesis H_k, corresponds to the density operator

$$\rho_k = \frac{1}{2} \begin{bmatrix} 1 & \exp(-i\,\theta_k) \\ \exp(i\,\theta_k) & 1 \end{bmatrix} = E_k. \tag{3.30}$$

Then from (3.23) $W_i = 0.5\, I - M^{-1} \rho_i$ and it is easy to see that the operators

$$\Pi_i = 2M^{-1} E_i \tag{3.31}$$

satisfy the optimality condition of Theorem 3.4.1. The minimum average Bayes cost (which here equals the average probability of error) is (3.29) $J_{\min} = 1 - 2M^{-1}$. Pure guessing gives $1 - M^{-1}$. But more importantly [49] the minimum over all *PVM* is higher than the above value. For example, for $M = 3$ the minimum average Bayes cost over all PVM is $(2 - \sqrt{3}/2)\,|\,3 > 1 - 2/3 = 1/3$. It is clear that *PVM*'s correspond to the pure strategies of the classical problem. Indeed, the decision maker measures the observable corresponding to the *PVM*, he observes one of M possible values, he chooses then the corresponding hypothesis. *So we have the important difference that the solution to the M-ary Bayes Decision problem with known apriori statistics does not lead to a pure strategy, in the quantum model.* Holevo showed in [49] that *POM*'s constitute an appropriate generalization of the randomized decision strategies in the classical theory. To see this, observe that if the detection operators Π_k commute, they have a common spectral measure [44], E. So there exist non-negative functions f_i such that

$$\Pi_j = \int f_j(v)\, E(dv).$$

Since $\sum_{j=1}^{M} \Pi_j = I$, $\sum_{j=1}^{M} f_j(v) = 1$ for each $v \in U$, that is $f_j(v)$ can be considered as probabilities. Then such a *POM* is equivalent to measuring the observable corresponding to E and deciding hypothesis H_i is true with probability $f_i(v)$, when the measurement outcome is v.

Third, while there is no implementation problem in the classical M-ary detection, solving for the optimal *POM* in the quantum detection does not guarantee an implementation for the optimal decision rule as well. It is only for very few cases that the solution to this implementation problem is known [32].

Helstrom [32] examines several examples of detection problems and gives details about their solution. One can also find in [32] solutions to the implementation problem in specific cases. Recently Yuen and Shapiro [69-71] have analyzed the other end of the communication problem (i. e. the source) and found that transmission in certain states called two-photon coherent states can reduce considerably the quantum noise and therefore induce improved detection at the receiver. The so called quantum channel has also been investigated from the point of view of information theory and in particular its capacity has been analyzed by Holevo [72] [74-76] and Ingarden [73].

4. **Non-commutative probability models.**

4.1. Logic of Quantum Mechanics.

In the previous section we gave a brief description of a rather well known model for quantum physics, although some of its aspects (such as the generalization from *PVM* to *POM*) are not widely known. These extensions are primarily due to proper formulations of communication problems. Since the initial proposal of von Neumann's model of quantum physics a subfield has been created with its main objective being the deductive derivation of the initial model, or other models proposed later, starting from few phenomenological axioms. Such work is of paramount importance and has led to various improvements of the initial model which closely reflect physical reality. We note that similar studies *have not* been undertaken in the area of multi-agent stochastic systems. The primary benefit from such works is that the various physical assumptions, engineering intuition, etc., are built into the algebra of the model and this process mechanizes the subsequent derivations. The realization theory of deterministic dynamical systems in its current algebraic form [51-53], with its manyfold uses in various control and systems problems, should be a motivating example. However, a non-classical stochastic realization (or representation) theory is needed for multi-agent stochastic systems.

In quantum mechanics such an approach was originated by Birkoff and von Neumann. The starting point was the structure of *propositions,* that is yes-no measurements [31]. Due to its similarity to a logical system the set of propositions is called *quantum logic.* This set can be easily given the structure of a lattice and the basic question addressed by physicists was: is there any set of phenomenological axioms that can allow one to identify the quantum logic with the set of orthogonal projections on a complex Hilbert space $\mathcal{H}$? Details of such theories can be found in [31]. There is one major disadvantage in this school of thought, however, as pointed out by Pool [54]. It tacitly assumes the structure of quantum logic is sufficient in itself to determine the mathematical formalism which should be employed in the quantum theory. This is not true, however. It is a fact that quantum mechanics is used not so much to reproduce the logical properties of simple yes-no experiments (and sequences of those) but rather to compute transition probabilities, cross sections, etc. Therefore, the probabilistic aspects must be unified with the logical aspects. The interested reader is referred to [54-58] for details in the developments of the axiomatic foundation. Briefly, (we follow [58]) there are in the model three basic elements:

i) the set $\mathcal{E}$, of simple events (or effects), which can be ascertained by some sort of physical measurement (or sequence of measurements),

ii) the set $\mathcal{S}$ of states (or ensembles) of a physical system,

iii) the probability function P: $\mathcal{E}\times\mathcal{S}\to[0,1]$, with $P(p,a)$ (for $p\in\mathcal{E}, a\in\mathcal{S}$) giving the probability of occurrence of the event p in the state a.

The triple $(\mathcal{E},\mathcal{S},P)$ is called by Pool an *event-state-structure* and it represents a specific class of physical systems. Phenomenological interpretations of the general aspects of the quantum logic approach are discussed in [31], [54-58]. A set of physically motivated axioms are imposed on $(\mathcal{E},\mathcal{S},P)$ [58]. These axioms guarantee the existence of the certain and impossible events; imply an antisymmetric *relation of implication,* and thus a partial order on $\mathcal{E}$ (denoted $p\leq q$ for elements p, $q\in\mathcal{E}$); imply existence of the negation p' of an event $p\in\mathcal{E}$. Two events p, q are mutually exclusive if $p\leq q'$. A set « dual » to $\mathcal{S}$, $\hat{\mathcal{S}}$ is constructed as the set of functions

$$\begin{aligned}&\mu_a\colon \mathcal{E}\to[0,1]\\&\mu_a(p)=P(p,a),\quad a\in\mathcal{S},\quad p\in\mathcal{E}\end{aligned}\tag{4.1}$$

Finally an axiom asserts the existence of the least upper bound (for the partial order $\leq$) of countable sets of pairwise mutually exclusive events and the law of additivity of probabilities for mutually exclusive events. A *poset* is a set together with a partial order. So $(\mathcal{E},\leq)$ is a poset. An *orthocomplementation* on a poset $\mathcal{X}$ with a *least* (denoted by 0) and *greatest* (denoted by 1) elements is a mapping $'$: $\mathcal{X}\to\mathcal{X}$ such that: (i) $(x')'=x$, $x\in\mathcal{X}$, (ii) $x, y\in\mathcal{X}$ and $x\leq y$, then $y'\leq x'$, (iii) if $x\in\mathcal{X}$ then the greatest lower bound $x\wedge x'$, and the least upper bound $x\vee x'$ of x and x' exist and equal 0 and 1, respectively. The relation $\perp$ of *orthogonality* is defined via $x\perp y$ if $x\leq y'$. An *orthoposet* $(\mathcal{X},\leq,')$ is a poset $(\mathcal{X}, \leq)$ together with an orthocomplementation such that if $x, y \in \mathcal{X}$ and $x \perp y$ then $x \vee y$ exists. An orthoposet is a *σ-orthoposet* if $x_1, x_2, \ldots \in \mathcal{X}$ and $x_i \perp x_j$, $i \neq j$ then $\bigvee_i x_i$ exists. An orthoposet is *orthomodular* if $x, y \in \mathcal{X}$, $x \leq y$ imply $y = x \vee (x' \wedge y)$. If $(\mathcal{X}, \leq, ')$ is a σ-orthoposet then a probability measure on $\mathcal{X}$ is a function μ: $\mathcal{X} \to [0, 1]$ such that: (*a*) $\mu(0)=0$, $\mu(1)=1$, (*b*) if $x_1, x_2, \ldots \in \mathcal{X}$, $x_i \perp x_j$ for $i \neq j$ then $\mu(\bigvee_i x_i) = \sum_i \mu(x_i)$. Let $\mathcal{U}$ be the set of probability measures on $\mathcal{X}$. Then $\mathcal{U}$ is: (*a*) *order-determining* if $x, y \in \mathcal{U}$ and $\mu(x) \leq \mu(y)$ for all $\mu \in \mathcal{U}$ then $x \leq y$; (*b*) *strongly-order-determining* if $x, y \in \mathcal{X}$, $\{\mu \in \mathcal{U}: \mu(x)=1\} \subset \subset \{\mu \in \mathcal{U}: \mu(y)=1\}$ then $x \leq y$; (*c*) *separating* if $x, y \in \mathcal{X}$ and $\mu(x) = \mu(y)$ for all $\mu \in \mathcal{U}$ then $x = y$; (*d*) *σ-convex* if $\mu_1, \mu_2, \ldots \in \mathcal{U}$, $\lambda_1, \lambda_2, \ldots \in [0, 1]$ with $\sum_i \lambda_i = 1$ then $\exists\ \mu \in \mathcal{U}$ such that $\mu(x) = \sum_i \lambda_i \mu_i(x)$ for all $x \in \mathcal{X}$.

Then one has the following results [58] as a consequence of the axioms imposed on an event-state structure $(\mathcal{E}, \mathcal{S}, P)$.

THEOREM 4.1.1: *If* $(\mathcal{E}, \mathcal{S}, P)$ *is an event-state structure then*

a) $(\mathcal{E}, \leq, ')$ *is an orthomodular σ-orthoposet,*

b) $\hat{\mathcal{S}}$ *is a strongly-order-determining, σ-convex set of probability measures on* $(\mathcal{E}, \leq, ')$,

c) $a \to \mu_a$ *is a bijection of* $\mathcal{S}$ *onto* $\hat{\mathcal{S}}$.

THEOREM 4.1.2: *If*

a) $(\mathcal{X}, \lesssim, \perp)$ *is an orthomodular σ-orthoposet,*

b) $\mathcal{U}$ *is a σ-convex, strongly-order-determining set of probability measures on* $\mathcal{X}$, *and*

c) P: $\mathcal{X} \times \mathcal{U} \to [0, 1]$ *is defined by*

$$P(x, m) = m(x),\ x \in \mathcal{X}, m \in \mathcal{U}$$

then $(\mathcal{X}, \mathcal{U}, P)$ *is an event-state structure. Moreover, for* $x, y \in \mathcal{X}$, $x \lesssim y$ *iff* $x \leq y$, *for* $x \in \mathcal{X}$, $x^{\perp} = x'$ *and* $\mathcal{U} = \hat{\mathcal{U}}$.

It is a consequence of these theorems that an event-state structure may be viewed either as a triple $(\mathcal{E}, \mathcal{S}, P)$ satisfying certain axioms or as a pair $(\mathcal{E}, \hat{\mathcal{S}})$ where $\hat{\mathcal{S}}$ is a σ-convex, strongly-order-determining set of probability measures on an orthomodular σ-orthoposet.

The reason for this discussion is to indicate that such an event-state structure is a reasonable model for the probabilistic structure of a multi-agent stochastic control system. In fact, it is not possible to find well founded objections in postulating that the set of simple events in a multi-agent stochastic control system is an orthomodular σ-orthoposet. Note, as we shall see shortly (see example 2 below), that Witsenhausen's formulation [13] in terms of product σ-fields clearly satisfies the defining axioms. We find no objection in adopting the axioms for the set of states $\mathcal{S}$ either; and thus that the set of states has a structure like $\hat{\mathcal{S}}$ above. Actually, there may be several options for $\hat{\mathcal{S}}$ [11], [13], [14].

EXAMPLE 1: Let $\mathcal{P}(\mathcal{H})$ be the set of all orthogonal projections on a separable complex Hilbert space $\mathcal{H}$, and let $\leq$ be the usual order of projections: $Q_1 \leq Q_2$ iff $Q_1 Q_2 = Q_1$. Let $Q' = I - Q$ be the orthogonal complement of Q. Then $(\mathcal{P}(\mathcal{H}), \leq, ')$ is an orthomodular σ-orthoposet. Let $\mathcal{S} = \{\rho \in \mathcal{T}_s^+(\mathcal{H}), Tr[\rho] = 1\}$ and $\hat{\mathcal{S}} = \{\mu_\rho(\cdot), \rho \in \mathcal{S};\ \mu_\rho(Q) = Tr[\rho Q]\}$. Then $\hat{\mathcal{S}}, \mathcal{P}(\mathcal{H})$ satisfy the conditions of Theorem 4.1.2 and, therefore, they represent an event - state structure with the probability function being $P(Q, \rho) = Tr[\rho Q]$, which is of course von Neumann's Hilbert space model used in 3.3. It is instructive to study the physical arguments that lead to additional axioms [31] [54-58], which allow one to identify the general event-state structure discussed above with that of von Neumann.

EXAMPLE 2: Consider the event-state structure $(\mathcal{E}, \hat{\mathcal{S}})$ where $\mathcal{E}$ is a σ-algebra of subsets of a set X and $\hat{\mathcal{S}}$ is a σ-convex, strongly-order-determining set of probability measures on $\mathcal{E}$. This is the classical Kolmogorov model of probability theory with several probability measures.

4.2. NONCOMPATIBILITY AND OPERATIONS.

We know from the brief exposition in 3.3 that there is one, at least, fundamental difference between the event-state structures described in examples 1 and 2 of 4.1. Namely, example 1 allows for events which are not compatible (i. e. not simultaneously observable), while this cannot happen in example 2. We recall from our discussion in 2.2 that this is required in a model for a multi-agent stochastic system. Formally, to describe compatibility, one introduces a relation C (compatibility) on $\mathcal{E}$ via:

$$\text{for } p, q \in \mathcal{E},\ pCq \text{ iff } \exists \text{ a Boolean sublogic } \mathcal{B} \subset \mathcal{E} \text{ such that } p, q \in \mathcal{B} \tag{4.2}$$

This relation C may be defined as follows: for $p, q \in \mathcal{E}$, pCq means $\exists$ $p_0, q_0, r \in \mathcal{E}$ such that: (i) $p_0 \perp q_0$, (ii) $p_0 \perp r$, and $p = p_0 \vee r$, (iii) $q_0 \perp r$ and $q = q_0 \vee r$. Now, clearly if $p, q \in \mathcal{E}$, and pCq then $p \wedge q$ exists in $\mathcal{E}$. Then one deficiency of the event-state model of 4.1 is that the following question cannot be answered satisfactorily: if $p, q \in \mathcal{E}$, and $p \not{C} q$ (p, q non-compatible) then does $p \wedge q$ exist in $\mathcal{E}$? Physically, this translates to: if observation (measurement) procedures for two incompatible (i. e. non-simultaneously observable) events are given, then how does one describe the observation (measurement) procedure for the « and » (or conjunction) of these two events. A similar problem appeared in section 3.3 where it was circumvented by introducing the notion of an operation in the von Neumann model (3.19).

This question is linked with the concept of «conditional probability» in an event-state structure. In the Kolmogorov model the concept of conditional probability and the associated concept of Radon-Nikodym derivative are of paramount significance for the analysis of classical stochastic control systems [26] [6-9]. In the classical model conditional probability is expressed as a mathematical object defined constructively in terms of the primitive entities of the theory. How can this be done in the generality of our discussion here? The hint comes from the interpretation of the operation introduced in (3.19) in order to handle joint statistics of repeated non-compatible measurements on the same system. That is operations are a form of conditioning. Indeed, the transformations (3.14), (3.18) are widely employed to represent the concept of conditional probability in von Neumann's model. This approach leads to difficulties, however, [37]. These are circumvented by introducing *event-state-operation structures*, (Pool [58]). We follow the exposition in [58]. An event-state-operation structure is a 4-tuple $(\mathcal{E}, \mathcal{S}, P, T)$ where $(\mathcal{E}, \mathcal{S}, P)$ is an event-state structure and T is a mapping from $\mathcal{E}$ into $\Sigma = \{\text{set of all maps from } \mathcal{S} \text{ into } \mathcal{S}\}$, which satisfies certain axioms. If $p \in \mathcal{E}$, T_p is the *operation corresponding to the event p*. The first axiom defines the domain of T_p as $\mathcal{D}_p = \{a \in \mathcal{S};\ P(p, a) \neq 0\}$. Since for $p \in \mathcal{E}$ and $a \in \mathcal{D}_p$, $T_p a$ is interpreted as the state conditioned on the event p and the state a, the first axiom is just for consistency. The elements of the set $S_T = \{T_{p_1} \circ T_{p_2} \circ \ldots \circ T_{p_n};\ p_1, \ldots, p_n \in \mathcal{E}\}$ are called *operations* ($\circ$ is composition, order of application is from right to left). Other (necessary) consistency axioms are: (ii) if $P(p, a) = 1$, then $T_p a = a$, (iii) if $a \in \mathcal{D}_p$, $P(p, T_p a) = 1$. There is an axiom guaranteeing consistency in « reversal » of experimental procedure. Now $(S_T, \circ)$ is a sub-semigroup of Σ and the latter axiom provides this semigroup with an *involution* denoted by $*$; i. e. a map*: $S_T \to S_T$ which maps the operation $x = T_{p_1} \circ \dots \circ T_{p_n}$ into $x^* = T_{p_n} \circ \dots \circ T_{p_1}$, such that: (*a*) for $x \in S_T$, $(x^*)^* = x$; (*b*) for $x, y \in S_T$, $(x \circ y)^* = y^* \circ x^*$. That is $(S_T, \circ, ^*)$ becomes an *involution semigroup.* Note that if 0 and 1 denote the least and greatest elements of $\mathcal{E}$, then T_1 is the identity of S_T, while T_0 (has empty domain) is a zero of S_T. The meaning of the latter becomes clear by observing that for any $p \in \mathcal{E}$, $T_p \circ T_0 = T_0 \circ T_p = T_0$. Furthermore, we need an experimental procedure to determine whether a state belongs to the domain $\mathcal{D}_x$ of an operation. There is an axiom which guarantees the existence of an event $q_x \in \mathcal{E}$, such that q_x occurs with certainty in the state a if and only if a does not belong to $\mathcal{D}_x$.

For each $p \in \mathcal{E}$, T_p is in the set of *projections* of S_T

$$\mathcal{P}(S_T) = \{e \in S_T = e \circ e = e^* = e\}. \tag{4.3}$$

$\mathcal{P}(S_T)$ is a poset where $e \leq f$ means $e \circ f = e$. Then $p \in \mathcal{E} \mapsto T_p \in \mathcal{P}(S_T)$ is an order preserving map. Also consider the map$^\sim$: $S_T \to \mathcal{P}(S_T)$ which maps $x \in S_T$ into the element $T_{q_x} \in \mathcal{P}(S_T)$, where q_x is the event which helps us test if a state is not in $\mathcal{D}_x$. Clearly $\tilde{T}_0 = T_1$ and if $\tilde{x} = T_1$ then $x = T_0$.

Now a *Baer*-semigroup* $(S, \circ, ^*, ^\sim)$ is an involution semigroup $(S, \circ, ^*)$ with a zero and a mapping$^\sim$: $S \to \mathcal{P}(S)$ such that if $x \in S$, then $\{y \in S\colon x \circ y = 0\} = \{y \in S\colon y = \tilde{x} \circ z$, for some $z \in S\}$ (i. e. an involution semigroup where the annihilator of each element is a principal left (right) ideal generated by a self-adjoint idempotent). The closed projections of a Baer*-semigroup are the elements of $\tilde{\mathcal{P}}(S) = \{e \in \mathcal{P}(S) = (\tilde{\tilde{e}}) = e\}$.

Then we have,

THEOREM 4.2.1: *If $(\mathcal{E}, \mathcal{S}, P, T)$ is an event-state-operation structure, then $(S_T, \circ, ^*, ^\sim)$ as constructed above is a Baer*-semigroup. Moreover, the mapping $p \in \mathcal{E} \mapsto T_p \in \mathcal{P}(S_T)$ is an isomorphism of the orthomodular orthoposet $(\mathcal{E}, \leq, ')$ onto the orthomodular orthoposet $(\tilde{\mathcal{P}}(S_T), \leq, ^\sim)$.*

Let us see how these constructions look in the two examples.

EXAMPLE 1: The event-state structure $(\mathcal{P}(\mathcal{H}), \mathcal{S}, P)$ of von Neumann's model admits an operation map T. If a is the state with density operator $\rho_a \in \mathcal{S}$ and $P(Q, a) = Tr\,[\rho_a Q] \neq 0$ then $T_Q\, a$ is the state a' with density operator $\rho_{a'} = \dfrac{Q\,\rho_a\,Q}{Tr\,[\rho_a Q]}$. If $x \in S_T$ and $x = T_{Q_1} \circ \dots \circ T_{Q_n}$ where

$Q_1, \ldots, Q_n \in \mathcal{P}(\mathcal{H})$ then q_x is the projection Q on the null space of $Q_1 Q_2 \ldots Q_n$.

EXAMPLE 2: The event-state structure $(\mathcal{E}, \hat{\mathcal{S}})$ of the Kolmogorov model admits an operation map. For $p \in \mathcal{E}$, T_p is defined via:

$$\mathcal{D}_p = \{\mu \in \hat{\mathcal{S}}, \mu(p) \neq 0\}$$

and

$$(T_p \mu)(q) = \frac{\mu(p \wedge q)}{\mu(p)}, \quad q \in \mathcal{E}, \tag{4.4}$$

which is the usual conditional probability.

If one views the event-state structure of the previous section as a *passive* picture (in the sense that it considers only the probability of occurrence of events), then the introduction of the concept of operation provides an *active* picture. Now it is easily seen from Theorem 4.2.1 and the construction above that for $p, q \in \mathcal{E}$,

$$p \leq q \text{ iff } T_p \circ T_q = T_p. \tag{4.5}$$

The natural question (from which we started in this section) is then: Can the greatest lower bound $p \wedge q$, for $p, q \in \mathcal{E}$ be interpreted via the composition of the Baer*-semigroup? The answer is in [58]:

THEOREM 4.2.2: *If $(\mathcal{E}, \mathcal{S}, P, T)$ is an event-state-operation structure, then $(\mathcal{E}, \leq, ')$ is an ortholattice. Moreover, if $p, q \in \mathcal{E}$ then*

$$T_{p \wedge q} = \widetilde{(T_{p'} \circ T_q)} \circ T_q.$$

Furthermore,

THEOREM 4.2.3: *Assumptions as above. Then for $p, q \in \mathcal{E}$ the following are equivalent:*

(*a*) pCq

(*b*) $T_p \circ T_q = T_q \circ T_p$.

If pCq then $T_{p \wedge q} = T_p \circ T_q$.

So in the setting of Baer*-semigroups we can associate the compatible events with commutativity of the corresponding operations. Note [37] that operations and observables are quite different kinds of entities. It is accepted that the various constructs of Baer*-semigroups have nice physical interpretations in this model of quantum physics.

Returning to multi-agent stochastic control systems we note that classical systems are characterized by a *commutative Baer*-semigroup* structure, while nonclassical information patterns correspond to *non-commutative semigroups.* Some natural questions that need to be answered are: how specific structural properties of the system appear in the structure of the semigroup? In particular, how does the information pattern appear in the semigroup? Another important point is that an operation associates to every event a map of the state set into itself, which can be interpreted as due to a control law. That is we can think of an operation as a model for the combined operation of obtaining a measurement and applying a control law by an agent. This is the active interpretation of an operation which we believe is significant for stochastic control. The passive can also be used to model the system's interaction to measurements.

4.3. Further Developments on Models for Measurements.

Let us consider here the particular Baer*-semigroup that pertains in the quantum model used in section 3 to formulate the M-ary detection problem. We saw in 4.2 that this model admits an operation map. It turns out that it is more convenient to work with the *unnormalized* version of operations [37, p. 17]. So in the Hilbert space model an operation is a positive linear map T: $\mathcal{T}_s(\mathcal{H}) \to \mathcal{T}_s(\mathcal{H})$ which also satisfies

$$0 \leq Tr\,[T(\rho)] \leq Tr\,[\rho] \tag{4.6}$$

for all $\rho \in \mathcal{T}_s^+(\mathcal{H})$. We emphasize again the phenomenological interpretation of an operation: An operation describes the change of state associated with a measurement which passes only a proportion of the ensemble tested. The probability of transmission of a state ρ by an operation S is taken to be $Tr\,[S(\rho)]$, while the output state conditional upon transmission is taken to be

$$\rho_{\text{out}} = \frac{S(\rho)}{Tr\,S(\rho)}\,. \tag{4.7}$$

Associated with the operation S is its *effect*, defined as the unique operator A for which

$$Tr\,[S(\rho)] = Tr\,[\rho A] \tag{4.8}$$

for all $\rho \in \mathcal{T}_s(\mathcal{H})$. The interpretation of A is that it determines the probability of transmission but *not* the form of transmitted state. It is now seen that this unnormalized version of operations leads to a slightly different model for the propositional calculus which is actually more satisfactory. The set of all effects $\mathcal{EF}$ consists of all bounded operators A on $\mathcal{H}$ such that $0 \leq A \leq 1$. $\mathcal{EF}$ is a poset and has a least and greatest element. Furthermore, it has an ortho-complementation given by the map $A' = 1 - A$. It is easily seen that the set of orthogonal projections $\mathcal{P}(\mathcal{H})$ is the set of extreme points of $\mathcal{EF}$. Note, however, that $\mathcal{EF}$ is not a lattice, but on the other hand it is a convex set in $\mathcal{L}(\mathcal{H})$. Furthermore, $\mathcal{P}(\mathcal{H})$ is dense in $\mathcal{EF}$ for the weak operator topology. This circle of ideas emphasizes $\mathcal{T}_s(\mathcal{H})$ as the state (or ensemble) set for the quantum model, i. e. consider the normalization $Tr[\rho]=1$ of secondary importance. Since we know that the density operator of a quantum system is the analog of the probability density of a stochastic system it is seen that the above argumentation is akin to considering the unnormalized probability density in classical formulations of filtering and control problems [26]. This often turns some of the crucial equations to linear ones! See, for example, the unnormalized conditional density equation of classical nonlinear filtering of diffusion processes [26].

These ideas can be extended to the general Baer*-semigroup setting, and this was done by Ludwig [55, 56] and Dähn [57]. Starting from an event-state structure $(\mathcal{E}, \mathcal{S}, P)$ one embeds $\mathcal{S}$ into the real vector space $\mathcal{V}$ of functions on $\mathcal{E}$ defined by $\chi(p) = \sum_{i=1}^{n} c_i P(p, a_i)$, $a_i \in \mathcal{S}$, c_i real numbers, n arbitrary. Under this embedding the state a goes into the function $P(\cdot, a)$. By letting

$$||\chi|| = \sup_{p \in \mathcal{E}} |\chi(p)| \tag{4.9}$$

for $\chi \in \mathcal{V}$, $\mathcal{V}$ becomes a normed linear space and we consider its completion which we also write as $\mathcal{V}$. Then $\mathcal{V}$ is a real Banach space and we let $\mathcal{V}^*$ be its dual. By considering $P(p, \chi) = \chi(p)$, P can be defined on the whole of $\mathcal{V}$, and for p fixed this defines a linear functional on $\mathcal{V}$, allowing us to identify $\mathcal{E}$ with a subset of $\mathcal{V}^*$. Furthermore, one can introduce a partial order in $\mathcal{V}$ by a cone $\mathcal{V}^+$ such that: (i) $\mathcal{V}^+$ is closed in $\mathcal{V}$; (ii) if $x, y \in \mathcal{V}^+$ then $||x|| + ||y|| = ||x+y||$; (iii) given $x \in \mathcal{V}$ and $\varepsilon > 0$ then there exist $x_1, x_2 \in \mathcal{V}^+$ such that $x = x_1 - x_2$ and $||x_1|| + ||x_2|| < ||x|| + \varepsilon$. $\mathcal{V}$ is usually called the state space by Davies [37]. In $\mathcal{V}$ the norm is linear on $\mathcal{V}^+$ and, therefore, can be uniquely extended to a positive linear functional τ: $\mathcal{V} \to \mathbf{R}$, with $|\tau(x)| \leq ||x||$ for all $x \in \mathcal{V}$, and $\tau(x) = \|x\|$ for all $x \in \mathcal{V}^+$. The states $\mathcal{S}$ are indentified as the elements of $\{x \in \mathcal{V}^+ : \tau(x) = 1\}$ and form a convex set. The set of effects $\mathcal{E}\mathcal{F}$ is in this general setting identified with

$$\mathcal{E}\mathcal{F} = \{\phi \in \mathcal{V}^*;\ 0 \leq \phi \leq \tau\}. \tag{4.10}$$

$\mathcal{E}\mathcal{F}$ is convex, weak* compact, partially ordered and has the orthocomplementation

$$\phi^\perp \triangleq \tau - \phi.$$

The events $\mathcal{E}$ are identified as the extreme points of $\mathcal{E}\mathcal{F}$.

There are several advantages of doing this transformation from an event-state structure to a state space structure: (*a*) we embed a nonlinear structure into a linear richer structure, (*b*) the relationship with classical probability theory becomes more apparent, (*c*) the theory fits in nicely with the use of C^*-algebras [59] in quantum statistical mechanics and quantum field theory. The two examples we have been using now appear as,

EXAMPLE 1: $\mathcal{V} = \mathcal{T}_s(\mathcal{H})$ with trace norm.

EXAMPLE 2: $\mathcal{V}$ is the space of all real finite signed measures on the Baire σ-field $\mathcal{B}$ of a locally compact Hausdorff space X. Let $\mathcal{V}^+$ be the cone of nonnegative measures and let

$$\|\mu\| = \sup\{|\mu(E) - \mu(X - E)|;\ E \in \mathcal{B}\}.$$

We turn now in a discussion of more complex (but more realistic) models of the measurement process in quantum systems, utilizing the framework of a state space and in particular the state space of example 1 above. It is important to realize that the concepts and constructions to be introduced can actually be worked out for the general setting of an abstract state space (or a Baer*-semigroup). This is of particular importance to us, since we do not have at the moment axioms for multi-agent stochastic control systems that will permit a concrete representation of these constructs (be it an event-state structure, or a Baer*-semigroup, or an abstract state space).

Thinking of the transformation performed on a state by an operation as one corresponding to a simple yes-no measurement, and recalling the discussion in section 3.3 (of equations (3.16)-(3.18)) it is easily seen that for a general continuous measurement we need to consider *operation-valued measures OVM*. This was first done by Davies and Lewis [45] in order to incorporate state-measurement interaction in the measurement process model. They introduced the concept of an *instrument* $\mathcal{IM}$ on a measurable space $(U, \mathcal{B})$, which is a map $\mathcal{IM}: \mathcal{B} \to \mathcal{L}^+(\mathcal{V}) = \{$the set of positive linear maps on $\mathcal{V} = \mathcal{T}_s(\mathcal{H})\}$, such that: (i) $\mathcal{IM}(B) \geq \mathcal{IM}(\phi)$ for all $B \in \mathcal{B}$; (ii) if $\{B_i\}$ is a countable collection of disjoint sets in $\mathcal{B}$ then $\mathcal{IM}(\coprod_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mathcal{IM}(B_n)$ (in the strong operator topology); (iii) $Tr[\mathcal{IM}(U)\rho] = Tr[\rho]$, for all $\rho \in \mathcal{V} = \mathcal{T}_s(\mathcal{H})$.

The idea behind this is to capture the concept of a measurement which accepts a state, measures some physical variable, and gives an output state conditional on the value observed. We would like to emphasize that it is a notion stronger than an observable and a *POM*. In fact, if $\mathcal{IM}$ is an instrument on $(U, \mathcal{B})$ then there is a unique *POM*, M on $(U, \mathcal{B})$ such that

$$Tr[\rho M(B)] = Tr[\rho\, \mathcal{IM}(B)] \tag{4.11}$$

for all $B \in \mathcal{B}$ and $\rho \in \mathcal{T}_s(\mathcal{H})$. The *POM* thus associated to the instrument is called the *measurement performed by the instrument*. Considering two instruments $\mathcal{IM}_1$ on U_1 and $\mathcal{IM}_2$ on U_2 with values in a state space $\mathcal{V}$, we can then define their composition $\mathcal{IM}_{12}$ as an instrument on $U_1 \times U_2$, which represents the measurement of first $\mathcal{IM}_1$ and then $\mathcal{IM}_2$ [37]. The composition is uniquely determined from the equation

$$\mathcal{IM}_{12}(B_1 \times B_2)\rho = \mathcal{IM}_2(B_2)\, \mathcal{IM}_1(B_1)\rho \tag{4.12}$$

for all $\rho \in \mathcal{V}$ and all B_1, B_2. This then leads naturally to a family of instruments, parametrized by time if we wish to describe repeated measurements (observations) from the same system. In a series of papers [60-62] Davies introduced such families when the measurement outcomes form a marked point process [64] and he termed them *quantum stochastic processes*. We introduced a generalization in [63] to allow for outcomes with continuous sample paths. Let U be a complete separable metric space, $\mathcal{B}$ the Borel σ-algebra on $U, \mathcal{Y}_t$ the set of all measurable functions from $[0, t]$ into U and $\mathcal{F}_t$ a σ-algebra on $\mathcal{Y}_t$. A *quantum stochastic process with outcomes adapted to* $\mathcal{F}_t$ is a family of instruments $\mathcal{IM}_t$ on $\mathcal{Y}_t$ such that:

(i) $\lim_{t \to 0} \mathcal{IM}_t(\mathcal{Y}_t)\rho = \rho$ for all $\rho \in \mathcal{T}_s(\mathcal{H})$,

(ii) $\mathcal{IM}_s(B)\, \mathcal{IM}_t(A)\rho = \mathcal{IM}_{t+s}(c(A \times B))\rho$ for $A \in \mathcal{F}_t, B \in \mathcal{F}_s$, where c maps $\mathcal{Y}_t \times \mathcal{Y}_s$ onto $\mathcal{Y}_{t+s}$ via concatenation of sample paths. Note that (ii) is the appropriate analog of the Chapman-Kolmogorov equation of classical probability. The physical interpretation is clear: $\mathcal{J}\mathcal{M}_t(A)\rho$ is the new state given the initial state was ρ and that the sample path of the outcome process was in $A \in \mathcal{F}_t$. The one parameter family $T_t = \mathcal{J}\mathcal{M}_t(\mathcal{Y}_t)$ forms a semigroup of operators on $\mathcal{T}_s(\mathcal{H})$ describing the evolution of the state as perturbed by measurement. If we let z denote the empty sample path then $S_t = \mathcal{J}\mathcal{M}_t(\{z\})$ is also a semigroup on $\mathcal{T}_s(\mathcal{H})$ describing the evolution of the state unperturbed from measurements. For certain processes Davies was able to characterize the differential version of the effects of measurement.

We close this section with an important result of Naimark [44] which has significant implications to questions of implementation of *POM*, instruments, etc. A natural question is how are *POM*'s interpreted from phenomenological measurement theory? A nice exposition of the underlying ideas can be found in Davies [37, p. 38]. Briefly let us consider a *POM*, M on the product measurable space $(U, \mathcal{B}) = (U_1, \mathcal{B}_1) \times (U_2, \mathcal{B}_2)$. If one could measure together two quantities which take values in U_1, U_2 then their composite measurement should be describable by such a *POM* on $(U, \mathcal{B})$. Then we can define the « marginal measurements » M_1 on U_1, M_2 on U_2 via $M_1(B) = M(B \times \mathcal{B}_2)$, and similarly for M_2. It is a well known fact that if both M_1, M_2 are projection valued (and, therefore, directly interpretable by measurement phenomenology) then M is projection valued and M_1, M_2, commute. There is a way to interpret M as « approximate » joint measurement of two noncompatible observables. An example of this kind is given in Davies [37 pp. 39-41] where a *POM* corresponding to approximate simultaneous joint measurement of position and momentum is constructed. A different interpretation is provided by Naimark's theorem which asserts that given a *POM* M on $(U, \mathcal{B})$ with values in $\mathcal{L}(\mathcal{H})$, there exist a pure state ρ_e on a Hilbert space $\mathcal{H}_e$ and a *PVM* E_M on $\mathcal{H} \times \mathcal{H}_e$ such that $Tr[\rho M(A)] = Tr[(\rho \times \rho_e) E(A)]$, for any ρ and any $A \in \mathcal{B}$. The triple $\{\mathcal{H}_e, \rho_e, E_M\}$ is called a realization of M and the physical interpretation is that M is statistically equivalent to the simultaneous measurement of compatible observables on an augmented system (the original augmented by the auxiliary system $\rho_e, \mathcal{H}_e$). Examples of such constructions appear in quantum communication problems [32], [33], [65-67].

4.4. Analogies with multi-agent stochastic control.

Having analyzed noncommutative probability models of various degrees of complexity in quantum physics we want to explicitly indicate here the properties of these models that we believe make them attractive candidates for use in multi-agent stochastic control.

We start with the set of simple propositions that can be verified by the agents in such a control problem. As indicated earlier this set, let us call it $\mathcal{E}$, has a natural partial order, that of implication $\leq$. It is also natural to assume the existence of the certain and impossible events and the negation p' of an event p, which defines an orthocomplementation on $\mathcal{E}$.

To postulate that $(\mathcal{E}, \leq, ')$ is an orthoposet is, therefore, in complete agreement with fundamentals of multi-agent control systems. The property of orthomodularity is a necessary consequence of the meaning we attach to implication, and the modifier σ in orthoposet can be justified on technical and intuitive grounds. We think of « states » or ensembles of such a system as measures which assign probabilities on these simple propositions. The notion of σ-convexity has an obvious interpretation in terms of « mixture » states or prior probabilities about states. We think of states here as all possible configurations of the multi-agent system, we do not necessarily assume a memory interpretation (as is done for example in [13] [14]). As a result our causality notion is « loose ». There is a reason for this in view of our remarks on anticipatory agents in section 2.1. To assume that the set of states is strongly order determining is clearly natural. The only objection is whether or not it can be verified, since to find all states for which an elementary proposition is true may not be feasible.

We thus have,

POSTULATE 4.4.1: The data bases in a multi-agent stochastic control system can be used to construct an event-state structure $(\mathcal{E}, \mathcal{S}, P)$ or $(\mathcal{E}, \hat{\mathcal{S}})$ as in section 4.1. Moreover, this representation is faithful in the sense it can generate the data on which it was based.

We note that the assumption that $\mathcal{E}$ can be made isomorphic to a σ-algebra on a set X implies that the operations $(\wedge)$ and (or) $(\vee)$ implied by the partial order $\leq$ are distributive and this seems to violate basic properties of multi-agent control systems, such as different information, noncompatibility of data analysis by two different agents, etc. *That is the logic of a multi-agent stochastic system cannot be Boolean.*

To introduce the operation structure we assume that there are events that are incompatible. Now this singles out a particular class of multi-agent control systems. Namely, those where there is strong information-control interaction between at least a pair of agents. This leads naturally to operations as conditioning on a simple event. The only axioms on the set of operations, introduced in 4.2, that need discussion are those that imply the involution and the map$^{\sim}$. Recall that involution is identical to consistency in « reversal » of experimental procedure. This may not be true in certain stochastic systems. Also the property of existence of perfect test events (the q_x in 4.2) may not hold. Nevertheless, there are many multi-agent stochastic systems having both properties. We restrict our attention to those. Then we have,

POSTULATE 4.4.2: It is possible to represent inferences made by agents via an operation map on $(\mathcal{E}, \mathcal{S}, P)$ and thus obtain an event-state-operation structure characterizing the statistics of the multi-agent control system. We can then according to theorem 4.2.1 represent these statistics and conditioning via a Baer*-semigroup.

This latter identification we find very intriguing because it algebrizes the notion of information pattern. For example, agents using common data bases will be represented by commuting conditioning operations. It is tempting to conjecture that there are classes of multi-agent stochastic systems where the information pattern can be completely specified by describing the structure of the Baer*-semigroup of the system. By this we mean the identification of commutative subsemigroups, ideal structure, etc. It will be interesting to investigate the statistical meaning of these algebraic constructs.

These algebraic constructions, we believe can help us understand how these classes of multi-agent stochastic systems operate. When we want to consider optimization, however, we need an analytic framework (in particular, norms) and if possible convexity. This is provided in the models discussed in 4.3 via embedding the event-state structure into a pair of Banach spaces connected by a duality which is induced by the probabilistic laws of the system. In this general framework there are analogs of constructs in quantum mechanics such as *POM*, instruments, etc. We have aluded to their stochastic control interpretation before. Clearly, the results of Davies and Benioff on repeated measurements being represented by a *POM*, can be interpreted as providing a useful model for analyzing observation patterns in certain multi-agent stochastic control problems. Optimal selection of a sequence of instruments can be interpreted as a joint optimization of information pattern and control.

Conceptually, the similarities between quantum physics and multi-agent stochastic control are striking. This is not a surprise. After all quantum mechanics provides meaningful statistical interpretation of macroscopic variables by avoiding microscopic considerations; exactly the objective in stochastic control systems with a large number of interacting controllers. We have pointed out differences. It remains to be seen if such formalisms can be effectively used in solving such problems. We hope the results in quantum detection (sec. 3) and quantum filtering (sec. 5) will convince the reader that there are feasible and challenging optimization problems in this setting.

5. **Quantum estimation and filtering.**

5.1. QUANTUM ESTIMATION.

As was discussed in 3, the main objective of detection and estimation theory with quantum mechanical models is the development of performance bounds for communication devices at the optical end of the spectrum used for communication. Despite its complicated appearance, one can extend some of the results of classical detection and estimation theory to the non-commutative probability model. In 3 we gave a summary of the major results in quantum detection theory. Here we would like to review estimation and filtering problems which are clearly more important for stochastic control.

The mathematical foundations of quantum estimation theory have been developed in [33] [34] [43], although earlier results by Helstrom on maximum likelihood and minimum variance estimators existed (see [32] and the references therein, in particular Ch. VIII). The case of continuous measurements require delicate mathematics, such as integration theory of operator valued functions with *POM*'s. Thus suppose the state of the quantum field depends on a parameter θ, $\rho(\theta)$. For similar reasons as in the detection problem we do not only have to specify the data processing scheme used to compute the estimator but also the measurement to be performed. Let M be the *POM* representing the measurement performed which gives the outcome v in the measurable space $(U, \mathcal{B})$. The estimator is a measurable function $\hat{\theta}(\ )$ of the data, and is also to be chosen. Let $(\Theta, \mathcal{F})$ be the measurable space where $\theta, \hat{\theta}$ take values. Again we are compelled to consider *randomized estimation strategies,* that is we have to specify the conditional distribution

$$\pi_{\hat{\theta}\mid v}(\vartheta; \xi) = Pr\{\hat{\theta} \leq \vartheta, \text{ given } v = \xi\}. \tag{5.1}$$

The statistics of the measurement outcomes (see 3.21) are now characterized by the conditional probability density

$$p_{v\mid\theta}(\xi;\vartheta)\,d\xi=Tr\,[\rho\,(\vartheta)\,M\,(d\xi)]. \tag{5.2}$$

Letting $C\,(\hat{\theta},\theta)$ be the cost function according to which we choose the estimator $\hat{\theta}$, and recalling the Bayesian formulation of the M-ary detection in 3.4, we easily see that the average cost for the randomized estimation strategy (5.1) is

$$J=\int_{\Theta}\int_{U}\int_{\Theta}C\,(\hat{\vartheta},\vartheta)\,\pi_{\hat{\theta}\mid v}\,(d\hat{\vartheta},\xi)\;p_{v\mid\theta}(\xi,\vartheta)\,d\xi\,\pi\,(d\vartheta) \tag{5.3}$$

where $\pi\,(\cdot)$ is the prior distribution function of θ. In view of (5.2), (5.3) becomes

$$J=\int_{\Theta}\int_{U}\int_{\Theta}C\,(\hat{\vartheta},\vartheta)\,\pi_{\hat{\theta}\mid v}(d\hat{\vartheta},\xi)\,Tr\,\rho\,(\vartheta)\,M\,(d\xi)\,\pi\,(d\vartheta). \tag{5.4}$$

The next step requires a rigorous theory of integration of operator valued functions with *POM*'s and a Fubini-type theorem in order to exchange the order of integration and trace operation. Such theories were developed first by Holevo [43] [33] and later in a more complete form by Young [34]. This allows the introduction of the operator valued *risk function*

$$W\,(\hat{\vartheta})=\int_{\Theta}C\,(\hat{\vartheta},\vartheta)\,\rho\,(\vartheta)\,\pi\,(d\vartheta). \tag{5.5}$$

Note that this is the continuous analog of (3.23) and with the appropriate integration theory it can be shown that for each $\hat{\vartheta}$, $W\,(\hat{\vartheta})\in\mathcal{T}_s\,(\mathcal{H})$. Next for every set $F\in\mathcal{F}$ define

$$\Pi\,(F)=\int_{U}\mu_{\hat{\theta}\mid v}(F,\xi)\,M\,(d\xi) \tag{5.6}$$

where $\mu_{\hat{\theta}|v}$ is the probability measure corresponding to $\pi_{\hat{\theta}|v}$. It is straightforward to verify that Π is a *POM* on $(\Theta,\mathcal{F})$. Using now (5.5) and (5.6) we can rewrite the average cost for this strategy as

$$J=Tr\left[\int_{\Theta}W\,(\hat{\vartheta})\,\Pi\,(d\hat{\vartheta})\right], \tag{5.7}$$

which is easily seen to be the continuous analog of (3.22) as expected. We shall use Holevo's terminology and notation. So the integral in (5.7) is called the *trace integral* and is denoted by

$$J = \langle W, \Pi \rangle_{\Theta}. \tag{5.8}$$

It is important to realize that the problem of optimally estimating θ now becomes

$$\min Tr\left[\int_{\Theta} W(\hat{\vartheta})\, \Pi(d\hat{\vartheta}) \right] \tag{5.9}$$

over all *POM* on $(\Theta, \mathcal{F})$.
That is the selection of an optimal measurement M on the data space and of an optimal randomized strategy are both subsumed in the selection of a *POM* Π. This may create implementation (or realization) problems, once the optimal Π is found. Now (5.9) is a convex optimization problem which has a very interesting duality theory (see [33] [34] and [68] for details). Existence of an optimal *POM* is not difficult to establish under mild conditions on $C(\cdot, \cdot)$ and is typically done by proving that in an appropriate topology the set of *POM*'s is compact and the cost functional lower semicontinuous [33] [68]. To formulate the duality properly one considers the set $\mathcal{M}(\mathcal{F}, \mathcal{L}_s(\mathcal{H}))$ of countably additive operator valued measures on the measurable space $(\Theta, \mathcal{F})$ (i. e. $M \in \mathcal{M}(\mathcal{F}, \mathcal{L}_s(\mathcal{H}))$ implies $M(F) \in \mathcal{L}_s(\mathcal{H})$ for every $F \in \mathcal{F}$, and $M(\bigsqcup_n F_n) = \sum_n M(F_n)$ for a disjoint sequence of subsets F_n). Then (5.9) which is the primal problem can be stated as: given the operator valued function $W(\cdot)\colon \to \mathcal{T}_s(\mathcal{H})$ solve the optimization problem

$$P = \inf \left\{ Tr \int_{\Theta} W(\vartheta)\, M(d\vartheta);\ M \in \mathcal{M}(\mathcal{F}, \mathcal{L}_s(\mathcal{H})),\right.$$
$$\left. M(\Theta) = I,\ M(F) \geq 0 \text{ for every } F \in \mathcal{F} \right\} \tag{5.10}$$

Then the dual problem is to solve for

$$D = \sup \{ Tr\, Y\colon Y \in \mathcal{T}_s(\mathcal{H}),\ Y \leq W(\vartheta), \quad \forall \vartheta \in \Theta \}. \tag{5.11}$$

The optimal *POM* is characterized by ([33] [34] [68])

THEOREM 5.1.1: *Necessary and sufficient conditions for* M_* *to be the optimal solution to* (5.10) (*or equivalently* (5.9)) *are*

$$\text{(i)} \quad \int_\Theta W(\vartheta)\, M_*(d\vartheta) \leq W(\theta) \text{ for every } \theta \in \Theta$$

or

$$\text{(ii)} \quad \int_\Theta M_*(d\vartheta)\, W(\vartheta) \leq W(\theta) \text{ for every } \theta \in \Theta.$$

Furthermore under any of the above conditions it follows that

$$Y = \int_\Theta W(\vartheta)\, M_*(d\vartheta) = \int_\Theta M_*(d\vartheta)\, W(\vartheta)$$

is selfadjoint, is the unique solution of the dual (5.11) *and*

$$J_{\min} = P = D = Tr\,[Y]. \tag{5.12}$$

This result has been obtained in various equivalent forms in [33] [34] [68]; here we presented the result as in [34]. Note that this theorem is the continuous analog of theorem 3.4.1 of quantum detection. As a consequence our remarks there are valid here as well. A plethora of specializations of these conditions can be found in [33]. Their importance lies primarily in computing solutions for specific applications problems.

Two special cases of this result deserve attention. First consider the case of a quadratic cost function

$$C(\hat{\theta}, \theta) = \|\hat{\theta} - \theta\|^2, \tag{5.13}$$

which corresponds to minimum error variance estimators. Then assuming all densities (or distributions) appearing in (5.3) have finite second moments we get existence of an optimal *POM*. For simplicity assume $\Theta = \mathbf{R}^n$ and $\mathcal{F} = \mathcal{B}^n =$ the Borel σ-algebra on $\mathbf{R}^n$. The following operators are well defined under our assumptions (the integrals are Bochner-integrals [43])

$$\int_{\mathbf{R}^n} \rho(\vartheta)\, \pi(d\vartheta) = \eta \in \mathcal{T}_s^{+}(\mathcal{H})$$

$$\int_{\mathbf{R}^n} \vartheta\, \rho(\vartheta)\, \pi(d\vartheta) = \delta \in \mathcal{T}_s^n(\mathcal{H}) = \mathcal{T}_s(\mathcal{H}) \times \overset{n}{\dots} \times \mathcal{T}_s(\mathcal{H}) \tag{5.14}$$

$$\int_{\mathbf{R}^n} \vartheta^T \vartheta\, \rho(\vartheta)\, \pi(d\vartheta) = \lambda \in \mathcal{T}_s^+(\mathcal{H}).$$

Then in (5.5)

$$W(u) = \eta\, u^T u - 2u^T \delta + \lambda;\ u \in \mathbf{R}^n. \tag{5.15}$$

We say that the optimal measurement is a *minimum error variance measurement* for the family $\{\rho(\theta)\}$. We then have

COROLLARY 5.1.2: *The POM $\hat{M}$ is minimum error variance for the family $\{\rho(\theta)\}$ iff*

$$\eta\left(u^T u - \int_{\mathbf{R}^n} u^T u\, \hat{M}(du)\right) - 2\,\delta^T\left(u - \int_{\mathbf{R}^n} u\, \hat{M}(du)\right) \geq 0 \tag{5.16}$$

for all $u \in \mathbf{R}^n$ ($\delta^T u$ stands for the operator $\sum_{i=1}^{n} u_i\, \delta_i$).

We can now introduce the *operator moments* of the *POM M*, following Holevo [74]

$$U_{i_1, \dots, i_k} = \int_{\mathbf{R}^n} u_{i_1} \dots u_{i_k} M(du). \tag{5.17}$$

So U_i are the *first operator moments* and U_{ij} the *second operator moments*. Observe that in the case of a simple measurement M is a *PVM* and it is *uniquely* defined by its first operator moment. For M a *POM* $U_{ij} \neq U_i U_j$. Then (5.16) yields

$$\eta \sum_{i=1}^{n} (u_i^2 - U_{ii}) - 2 \sum_{i=1}^{n} \delta_i (u_i - U_i) \geq 0 \tag{5.18}$$

for all $u \in \mathbf{R}^n$. It can be seen that this solution defines an appropriate extension of the conditional expectation which solves the corresponding classical problem [43, section 8]. If we cast the classical problem using the state space structure of 4.3, it turns out that the operators δ_i commute and thus the analysis of Holevo [43] applies. In a simpler case suppose $n=1$. Then (5.17) implies that the solution to the minimum variance estimator of $\theta \in \mathbf{R}$ is provided by the measurement of the observable $\hat{\Theta}$ which satisfies the Lyapunov equation

$$n\,\hat{\Theta}+\hat{\Theta}\,n=2\delta. \tag{5.19}$$

In the classical case all operators commute and the solution can be written (degenerate cases apart) as $\hat{\Theta}=\delta\, n^{-1}=n^{-1}\,\delta$ which if we think of $\rho(\theta)$ as a « density » is in some sense like conditional expectation. Cases where $\rho(\theta)$ commute for all values of θ are thus termed « classical ». As an example of (5.18) consider the estimation problem corresponding to the detection problem discussed in 3.4 (eq. (3.30)-(3.31)).

The second case corresponds to the maximum likelihood estimator for which $\pi(d\theta)$ is uniform, $\frac{1}{\delta}\,d\theta$ say, and $C(\hat{\theta},\theta)=-\delta(\hat{\theta}-\theta)$.

Then from (5.5)

$$W(\hat{\vartheta})=-\frac{1}{\delta}\,\rho(\hat{\vartheta}) \tag{5.20}$$

and therefore the optimal *POM* maximizes

$$J=\frac{1}{\delta}\,Tr\left[\int_{\Theta}\rho(\vartheta)\,M(d\vartheta)\right]. \tag{5.21}$$

We say that the optimal *POM* is a *maximum likelihood measurement* for the family $\{\rho(\theta)\}$ (a term due to Holevo [43]).

COROLLARY 5.1.3: *The POM M_L is maximum likelihood for the family $\rho(\theta)$ iff*

(i) $\rho(\theta)\leq\int_{\Theta}\rho(\vartheta)\,M_L(d\vartheta), \quad \forall\theta\in\Theta$

or

(ii) $\rho(\theta)\leq\int_{\Theta}M_L(d\vartheta)\,\rho(\vartheta), \quad \forall\theta\in\Theta.$

As an example consider again the analog of the example treated in

3.4. Now

$$\rho(\theta)=\frac{1}{2}\begin{bmatrix} 1 & \exp(-i\,\theta) \\ \exp(i\,\theta) & 1 \end{bmatrix},\ 0\leq\theta<2\pi.$$

Then consider the *POM*

$$M_L(d\theta)=\rho(\theta)\,\frac{d\theta}{\pi}\,.$$

Now

$$\int_{\Theta}\rho(\vartheta)\,M_L(d\vartheta)=\int_0^{2\pi}\rho(\theta)\,\frac{d\theta}{\pi}=1,$$

and since

$$1-\rho(\theta)=\frac{1}{2}\begin{bmatrix} 1 & -\exp(-i\,\theta) \\ -\exp(i\,\theta) & 1 \end{bmatrix}\geq 0$$

we have optimality. The conditional density of the measurement outcome is

$$p(\hat{\theta}\mid\theta)=Tr\,[\rho(\theta)\,\rho(\hat{\theta})]\frac{d\hat{\theta}}{\pi}=\frac{1}{\pi}\cos^2\left(\frac{\hat{\theta}-\theta}{2}\right).$$

The book by Helstrom [32] and Holevo's paper [43] contain several worked out examples and special cases of the general theoretical results presented here.

5.2. Quantum Filtering.

The extension of classical filtering theory to the noncommutative model presents several difficulties. First, there are state dynamics and, therefore, the state (density operator) depends explicitly on time. Second, the parameter is now the sample path of the signal process x_t that modulates the quantum field. Third, there is state- measurement interaction which has to be properly modelled. If we suppress all these interactions and use Davies's and Benioff's results [40, 45] [46-48] discussed in 3.3 to represent repeated measurements by a *POM* on a large σ-field, we abandon in a way our hopes for recursive solutions and the implementation of the filter is not taken into consideration. A more promising approach is to study a hierarchy of problems with increasing complexity with the hope to discover cases with feasible solutions. We have followed this approach in a series of papers [63, 65-68].

In [65] the linear filtering problem for a scalar signal process in discrete time was analyzed. This is the simplest case. Measurements were modelled by observables and the solution was obtained via an elegant application of the projection theorem. However, there were several important assumptions made: (*a*) the density operator depends at each time only on the signal value at the same time (*b*) time is discrete and (*c*) the measurement outcomes at different times are independent conditioned on the signal sequence. These assumptions prevent the inclusion of the state evolution and state-measurement interaction in the model. However, there are cases where these assumptions are satisfied in practical systems. Furthermore, the bounds obtained by these methods are conservative. In [66], the same problem was analyzed for vector processes. The formulation and solution are far more complex than in the scalar case. We briefly describe the main ideas here.

We consider a quantum field (for example, a laser) which is modulated in some fashion by a vector discrete time stochastic process x_t, $t=0, 1, \ldots$ in $\mathbf{R}^n$. The field's state is described by a density operator $\rho(x_t)$. Note the lack of memory in $\rho(\cdot)$. Furthermore, we assume that the outcomes v_t (classical vector random variables) are independent when conditioned upon the sequence $\{x_t\}$. At times $t=0, 1, \ldots$ measurements are performed on the field, represented by *POM*'s, M_t, $t=0, 1, \ldots$ and giving outcomes v_t, $t=0, 1, \ldots$ in $\mathbf{R}^n$. We combine the outcomes linearly to compute the estimator

$$\hat{x}_t = \sum_{i=0}^{t} C_i(t)\, v_i \tag{5.22}$$

where $C_i(t)$, $t=0, 1, \ldots$ are $n \times n$ matrices. The problem is then to optimize *jointly* over the selection of measurements (*POM*'s M_t) and processing scheme (matrices $C_i(t)$) to obtain the minimum error variance linear filtered estimate of x_t, i. e. minimize

$$J = E\{\|x_t - \hat{x}_t\|^2\} \tag{5.23}$$

This problem is similar in nature with the quadratic cost estimation problem treated in (5.1). At time t we have to choose the new measurement and the processing scheme, having observed previous measurements. We can express (5.23) as a trace integral [66] of an operator valued function with M_t

$$J = Tr\left[\int_{\mathbf{R}^n} \mathcal{F}(u, C(t))\, M_t(du)\right] \tag{5.24}$$

similarly as in (5.7), where $\mathcal{F}$ is quadratic in u and linear in the matrices $C_i(t)$. The interested reader will find the details in [66]. Existence to this joint optimization problem is settled in [68], where also the necessary and sufficient conditions for optimality are proven employing the duality discussed in relation to Theorem 5.1.1. The main result is:

THEOREM 5.2.1: *Necessary and sufficient conditions for* $\hat{C}_0(t), \ldots$ $\ldots, \hat{C}_{t-1}(t)$, *and* $\hat{M}_t$ *to be optimal processing coefficient matrices and measurement at time t are*

$$\text{(i)} \quad \sum_{j=0}^{t} E\{v_i v_j^T\}\, \hat{C}_j^T(t) = E\{v_i x_k^T\}, \quad i=0, 1, \ldots, t$$

where $\hat{C}_t(t) = I_n$ *and* $v_t = \hat{v}_t$,

$$\text{(ii)} \quad \mathcal{F}(u, \hat{C}(t)) \geq \int_{\mathbf{R}^n} \mathcal{F}(u, \hat{C}(t))\, \hat{M}_t(du), \text{ for all } u \in \mathbf{R}^n.$$

Note that (ii) is identical to the result of Theorem 5.5.1, and that (i) are just the normal equations for the classical random variables of the measurement outcomes. It is important to compare the solution of this vector problem to that of the scalar one. The main point is that in the vector case the above equations characterize only the first and second operator moments of the optimal measurements. In the scalar case, this is enough to uniquely determine the measurements. Here we have more flexibility, and so we can look for other properties such as implementation. An example of such an approach is [74], which contains results on estimation problems with Gaussian field states and canonical measurements. More recently these were related to two-photon coherent states and photon counting measurements [77].

In [66] we pursued further structural properties of the filter. In particular, the question we analyzed was: how different is the minimum error variance *filtered* measurement at t (see 5.1 for definition) from the minimum error variance measurement at t? The latter measurements produce the minimum error variance estimator for x_t without postprocessing (i. e. where $C_i(t) = 0$, $i = 0, \ldots, t-1$ in (5.22)). This is a natural question to analyze since due to our assumptions we only utilize correlation between $\{x_t\}$. When $\{x_t\}$ is an uncorrelated sequence the filter measurements can be chosen independently at each time t. We discovered in [66] that when $\{x_t\}$ is a Gaussian sequence, the optimal measurement outcomes are jointly Gaussian with $\{x_t\}$ and a certain linearity condition is satisfied, then the filter separates. That is, we showed that in this case, the optimal quantum measurements are chosen separately from the optimal (classical) linear postprocessing of the measurement outcomes. In such cases, one wishes to establish that the same measurement is performed for all t and then there is hope for a recursive solution, if the classical problem has one. Such an example was given in [66] for a monochromatic laser, which is carrying a 2 dimensional vector signal process $x_t = [x_{1t}, x_{2t}]^T$ as its in phase and quadrature amplitudes, and is received along with thermal noise in a single-mode cavity. We assumed x_t is generated by a linear model

$$x_{t+1} = \phi_t x_t + W_t.$$

The optimal linear filter reduced to a very simple form (see details in [66]). Namely, the same measurement (called optical heterodyning) was performed on the received field and then the measurement outcomes were processed by a Kalman-filter. Unfortunately, we do not have many examples as successful to report in this area.

More recently in [63] [67] [78], we have, using the results of Davies on quantum stochastic processes [60-62], formulated the general continuous quantum filtering problem incorporating state evolution and state-measurement interaction. Davies has also used quantum stochastic processes to model a quantum communication system [79]. The main ideas are as follows. We are given the state representation of a quantum system $\rho(t, x^t)$ which at time t depends on the sample path x^t of the signal up to time t. This for example can be obtained by modelling the modulation via a stochastic Liouville operator equation

$$\frac{\partial \rho(t)}{\partial t} = -i\,[H(x(t)), \rho(t)]; \tag{5.25}$$

see the references for explicit examples. Incidentally, it is interesting to note that often equations like (5.25) arising in quantum optics, when explicit representations for ρ are assumed (i. e. as integral operators) lead to stochastic partial differential equations of the same type as the unnormalized conditional density equation of nonlinear (classical) filtering

[67]. We are given a class C_M of quantum measurement processes (cf. sec. 4.3) (representing measurements on the system) and a class C_p of processing schemes (i. e. functionals on the sample paths of the outcomes of measurements). We let y_t denote the outcome process for a quantum stochastic process, $\hat{f}_t$ the processing scheme. Then the filtered estimate is

$$\hat{x}_t = \hat{f}_t (y^t). \tag{5.26}$$

The problem is to select a quantum stochastic process from C_M and a scheme from C_p to minimize the error variance $E\{\|x_t - \hat{x}_t\|^2\}$. First, one establishes that there exists a quantum stochastic process in C_M parametrized by the sample paths x^t of the signal process on $\mathcal{Y}^t, \mathcal{I}\mathcal{N}_t(x^t, E)\rho(0)$ such that $\rho(t) = \mathcal{I}\mathcal{N}_t(x^t, \mathcal{Y}^t)\rho(0)$ (see also 4.3). Then we have to compute the differential description of the dynamics of $\rho(\cdot)$, where $\mathcal{I}\mathcal{N}_t$ will show its influence. Once this is done, further progress can be achieved since the probability measure

$$\mu_t(x^t, E) = Tr\,[\mathcal{I}\mathcal{N}_t(x^t, E)\,\rho(0)] \tag{5.27}$$

characterizes the statistics of the classical measurement process y_t given the classical signal process x_t. Explicit examples appear in [63], [67], [78], [79].

6. Conclusions.

We have outlined here a number of noncommutative probability models and their basic constructs that we believe are common in modelling quantum systems and certain multi-agent stochastic control systems. We have provided no proofs, only plausibility arguments. More remains to be done in evaluating the ideas presented here. A concrete suggestion is to try to identify the basic constructs used in quantum communication (rigorously) in the language of stochastic control. Recently, this was done for classical communication systems with promising results. It would be interesting to know if the Naimark construction has an interpretation in stochastic control. What additional postulates will permit concrete representations of the algebraic objects introduced here? Is there a classification theory? If the present paper stimulated some thoughts along these lines we would feel that it fulfilled its purpose.

REFERENCES

[1] V. E. BENEŠ, *Existence of Optimal Strategies Based on Specified Information, for a Class of Stochastic Decision Problems*, SIAM J. Control, Vol. 8, No. 2, 1970, pp. 179-188.

[2] V. E. BENEŠ, *Existence of Optimal Stochastic Control Laws*, SIAM J. Control, Vol. 9, No. 3, 1971, pp. 446-472.

[3] M. H. A. DAVIS, P. P. VARAIYA, *Dynamic Programming Conditions for Partially Observable Stochastic Systems*, SIAM J. Control, Vol. 11, 1973, pp. 226-261.

[4] W. H. FLEMING, *Optimal Control of Partially Observable Diffusions*, SIAM J. Control, Vol. 6, 1968, pp. 194-214.

[5] W. H. FLEMING, *Optimal Continuous - Parameter Stochastic Control*, SIAM Rev. 11, 1969, pp. 470-509.

[6] H. KUSHNER, *Stochastic Stability and Control*, New York: Academic Press, 1967.

[7] C. STRIEBEL, *Optimal Control of Discrete Time Stochastic Systems*, Lecture Notes in Economics and Math. Systems, Vol. 110, Springer-Verlag, 1975.

[8] W. FLEMING, R. RISHEL, *Deterministic and Stochastic Optimal Control*, Springer-Verlag, 1975.

[9] D. BERTSEKAS, S. SHREVE, *Stochastic Optimal Control, The Discrete Time Case*, Academic Press, 1978.

[10] V. E. BENEŠ, *Composition and Invariance Methods for Solving Some Stochastic Control Problems*, Adv. Appl. Prob., 7, 1975, pp. 299-329.

[11] H. S. WITSENHAUSEN, *On Information Structures, Feedback and Causality*, SIAM J. Control, Vol. 9, No. 2, pp. 149-160, 1971.

[12] H. S. WITSENHAUSEN, *Separation of Estimation and Control for Discrete Time Systems*, Proc. of the IEEE, Vol. 59, No. 11, pp. 1557-1566, 1971.

[13] H. S. WITSENHAUSEN, *The Intrinsic Model for Discrete Stochastic Control: Some Open Problems*, Lecture Notes in Economics and Mathem. Systems, Vol. 107, pp. 322-335, Springer-Verlag, 1975.

[14] H. S. WITSENHAUSEN, *Some Remarks on the Concept of State*, Directions in Large Scale Systems, (Y. C. Ho and S. K. Mitter, Edts.), New York: Plenum Press, pp. 69-75, 1976.

[15] H. S. WITSENHAUSEN, *A Standard Form for Sequential Stochastic Control*, J. Math. System Theory, Vol. 7, No. 1, pp. 5-11, 1973.

[16] Special Issue on « *Large-Scale Systems and Decentralized Control* », IEEE Trans. on Aut. Control, Vol. 23, No. 2, 1978.

[17] N. R. SANDELL, JR., P. P. VARAIYA, M. ATHANS, M. G. SAFONOV, *Survey of Decentralized Control Methods for Large Scale Systems,* IEEE Trans. on Aut. Control, Vol. 23, No. 2, 1978, pp. 108-128.

[18] J. M. BISMUT, *Analyse Convexe et Probabilités,* Ph. D. Dissertation, Faculté des Sciences de Paris.

[19] J. M. BISMUT, *An Example of Interaction Between Information and Control: The Transparency of a Game,* IEEE Trans. Aut. Control, Vol. AC-18, 1973, pp. 518-522.

[20] P. WHITTLE, J. F. RUDGE, *Simultaneous Communication and Control,* Adv. Appl. Prob. 8, 1976, pp. 365-384.

[21] J. F. RUDGE, *Multi-Agent Linear-Quadratic-Gaussian Systems with « Poor » Observations,* SIAM J. Control and Opt., Vol. 16, No. 5, 1978, pp. 832-846.

[22] Y. C. HO, M. P. KASTNER, E. WONG, *Teams, Signaling and Information Theory,* IEEE Trans. on Aut. Control, Vol. AC-23, No. 2, 1978, pp. 305-311.

[23] Y. C. HO, M. P. KASTNER, *Market Signaling: An Example of A Two-Person Decision Problem with Dynamic Information Structure,* IEEE Trans. Aut. Control, Vol. AC-23, No. 2, 1978, pp. 350-361.

[24] R. ROSEN, *On Anticipatory Systems I,* J. Social Biol. Struct., 1978, 1, pp. 155-162, *II,* Ibid, 1978, 1, pp. 163-180.

[25] E. WIGNER, *The Scientist Speculates,* p. 284; I. J. Good, ed. London: Heinemann, 1962.

[26] E. WONG, *Stochastic Processes in Information and Dynamical Systems,* McGraw-Hill, 1971.

[27] J. VON NEUMANN, *Mathematical Foundations of Quantum Mechanics,* transl. by R. T. Beyer, Princeton: Princeton Univ. Press, 1955.

[28] R. J. GLAUBER, *Quantum Theory of Coherence,* Quantum Optics S. M. Kay and A. Maitland Edts., Academic Press, 1970, pp. 53-125.

[29] D. MARCUSE, *Engineering Quantum Electrodynamics,* Harcourt, Brace & World, 1970.

[30] E. PRUGOVECKI, *Quantum Mechanics in Hilbert Space,* Academic Press, 1971.

[31] J. M. JAUCH, *Foundations of Quantum Mechanics,* Addison-Wesley, 1968.

[32] C. W. HELSTROM, *Quantum Detection and Estimation Theory,* Academic Press, 1976.

[33] A. S. HOLEVO, *Investigations in the General Theory of Statistical Decisions,* Proceedings of the Steklov Math. Inst., 124, Nauka, Moscow 1976, (Engl. Transl. 1978, issue 3).

[34] S. YOUNG, *Convexity and Duality in Optimization Theory,* Ph. D. Dissertation, Math. Dept., MIT September 1977.

[35] H. YUEN, R. KENNEDY, M. LAX, *Optimum Testing of Multiple Hypotheses in Quantum Detection Theory,* IEEE Trans. on Inform. Th., IT-21, 1975, pp. 125-134.

[36] I. EKELAND, R. TEMAM, *Convex Analysis and Variational Problems,* American Elsevier, 1976.

[37] E. B. DAVIES, *Quantum Theory of Open Systems,* Academic Press, 1976.

[38] I. E. SEGAL, *Mathematical Problems of Relativistic Physics,* Amer. Math. Soc., Providence, R. I., 1963.

[39] A. S. HOLEVO, *An Analogue of the Theory of Statistical Decisions in Non-Commutative Probability Theory,* Trans. Moscow Math. Soc. 26, 1972. pp. 133-149.

[40] E. B. DAVIES, *On the Repeated Measurement of Continuous Observables in Quantum Mechanics,* J. Funct. Anal. 6, 1970, pp. 318-346.

[41] S. K. BERBERIAN, *Notes on Spectral Theory,* New York, D. Van Nostrand, 1966.

[42] R. HAAG, D. KASTLER, *An Algebraic Approach to Quantum Field Theory,* J. Math. Phys. 5, 1964, pp. 848-861.

[43] A. S. HOLEVO, *Statistical Decision Theory for Quantum Systems,* J. of Multiv. Analysis 3, 1973, pp. 337-394.

[44] N. I. AKHIEZER, I. M. GLAZMAN, *Theory of Operators in Hilbert Space,* Vol. II, New York: Frederick Ungar, 1963.

[45] E. B. DAVIES, J. T. LEWIS, *An Operational Approach to Quantum Probability,* Commun. Math. Phys. 17, 1970, pp. 239-260.

[46] P. A. BENIOFF, *Operator Valued Measures in Quantum Mechanics: Finite and Infinite Processes,* J. Math. Phys. 13, 1972, pp. 231-242.

[47] P. A. BENIOFF, *Decision Procedures in Quantum Mechanics,* J. Math. Phys. 13, 1972, pp. 908-915.

[48] P. A. BENIOFF, *Procedures in Quantum Mechanics Without von Neumann's Postulate,* J. Math. Phys. 13, 1972, pp. 1347-1355.

[49] A. S. HOLEVO, *The Theory of Statistical Decisions on an Operator Algebra,* Soviet Math. Pokl., Vol. 15 (1974) No. 5, pp. 1276-1280.

[50] A. S. HOLEVO, *Remarks on Optimal Quantum Measurements,* Problemy Peredachi Informatsii, Vol. 10, No. 4, 1974, pp. 51-55 (English translation by Plenum Publ. Comp.).

[51] R. E. KALMAN, P. L. FALB, M. A. ARBIB, *Topics in Mathematical System Theory,* New York: McGraw-Hill, 1969.

[52] G. MARCHESINI, S. K. MITTER (Edts.), *Mathematical System Theory,* Lecture Notes in Economics and Mathematical Systems, Vol. 131, Springer-Verlag, 1976.

[53] Special Issue on « *Recent Trends in System Theory,* Proc. IEEE, Vol. 64, No. 1, Jan. 1976.

[54] J. C. T. POOL, *Simultaneous Observability and the Logic of Quantum Mechanics,* Ph. D. Dissertation, State University of Iowa, Dept. of Physics, Report SUI-63-17, 1963.

[55] G. LUDWIG, *Attempt of an Axiomatic Foundation of Quantum Mechanics and More General Theories II*, Commun. Math. Phys. 4, pp. 331-348, 1967.

[56] G. LUDWIG, *Attempt of an Axiomatic Foundation of Quantum Mechanics and More General Theories III*, Commun. Math. Phys. 9, pp. 1-12, 1968.

[57] G. DÄHN, *An Attempt at an Axiomatic Formulation of Quantum Mechanics and More General Theories IV*, Commun. Math. Phys. 9, pp. 192-211, 1968.

[58] J. T. POOL, *Baer* - Semigroups and the Logic of Quantum Mechanics*, Commun. Math. Phys. 9, pp. 118-141, 1968.

[59] J. DIXMIER, *Les C* - Algèbres et Leurs Representations*, 2nd. edition, Gauthier-Villars, Paris, 1969.

[60] E. B. DAVIES, *Quantum Stochastic Processes*, Commun. Math. Phys. 45, pp. 277-304, 1969.

[61] E. B. DAVIES, *Quantum Stochastic Processes - II*, Commun. Math. Phys. 49, pp. 83-105, 1970.

[62] E. B. DAVIES, *Quantum Stochastic Processes - III*, Commun. Math. Phys. 22, pp. 51-70, 1971.

[63] J. S. BARAS, *Continuous Quantum Filtering*, Proc. 15th Allerton Conference on Communication, Control and Computing, pp. 68-77, 1977.

[64] D. SNYDER, *Random Point Processes*, John Wiley, 1975.

[65] J. S. BARAS, R. O. HARGER, Y. H. PARK, *Quantum-Mechanical Linear Filtering of Random Signal Sequences*, IEEE Tran. on Info. Th., Vol. IT-22, No. 1, 1976, pp. 59-64.

[66] J. S. BARAS, R. O. HARGER, *Quantum Mechanical Linear Filtering of Vector Signal Processes*, IEEE Tran. on Info. Th., Vol. IT-23, No. 6, 1977, pp. 683-693.

[67] J. S. BARAS, *Stochastic Partial Differential Equations in Optical Communication Problems*, in Recent Developments in Variable Structure Systems, R. R. Mohler and A. Ruberti (Edts.), Lecture Notes in Economics and Mathematical Systems, Vol. 162, Springer-Verlag, 1978, pp. 21-43.

[68] J. S. BARAS, *An Optimization Problem from Linear Filtering with Quantum Measurements*, to appear J. Multiv. Analysis.

[69] H. P. YUEN, *Two-photon Coherent States of the Radiation Field*, Physical Review A, Vol. 13, No. 6, 1976, pp. 2226-2243..

[70] H. P. YUEN, J. H. SHAPIRO, *Optical Communication with Two-photon Coherent States - Part I: Quantum State Propagation and Quantum Noise Reduction*, IEEE Trans. on Info. Th., Vol. IT-24, pp. 657-668, 1978.

[71] J. H. SHAPIRO, H. P. YUEN, J. A. MACHADO MATA, *Optical Communication with Two-photon Coherent States - Part II: Photoemissive Detection and Structured Receiver Performance*, IEEE Trans. on Info. Th., Vol. IT-25, No. 2, pp. 179-192, 1979.

[72] A. S. HOLEVO, *Problems in the Mathematical Theory of Quantum Communication Channels*, Reports on Math. Physics, Vol. 12 (1977), pp. 273-278.

[73] R. INGARDEN, Reprint N. 317, Institute of Physics, Toruń, 1975.

[74] A. S. HOLEVO, *Some Statistical Problems for Quantum Gaussian States,* IEEE Trans. Info. Th., Vol. IT-21, pp. 533-543, 1975.

[75] A. S. HOLEVO, *Informational Aspects of Quantum Measurements,* Probl. Peredach. Inform. 9 (2), pp. 31-42, 1973.

[76] A. S. HOLEVO, *Some Estimates for the Quantity of Information Transmitted by a Quantum Communication Channel,* Probl. Peredach. Inform. 9 (3), pp. 3-11, 1973.

[77] H. P. YUEN, J. SHAPIRO, *Optical Communication with Two-photon Coherent States - Part III: Quantum Measurements Realizable with Photoemissive Detectors,* Preprint.

[78] J. S. BARAS, *Quantum Filtering with Counting Measurements,* Proc. 1979 IEEE Int. Symp. on Infor. Theory, Grignano, Italy, 1979.

[79] E. B. DAVIES, *Quantum Communication Systems,* IEEE Trans. on Info. Th., 1977, pp. 530-534.

Gubbio - Tipografia Oderisi Editrice - 1981