RIGOROUS RESULTS IN NON-EQUILIBRIUM STATISTICAL MECHANICS*†

Gérard G. Emch‡
Joint Institute for Laboratory Astrophysics
University of Colorado
Boulder, Colorado

Introduction
A mathematical formulation of some physical postulates will first be proposed together with a precise statement of the limitations inherent in the mathematical model considered.

The second lecture will then be devoted to the derivation and to some discussion of the evolution equations pertinent to the model.

Finally, some equilibrium conditions will be analyzed in connection with ergodic theory.

FIRST LECTURE
The Model
The object of this first lecture is not to present some kind of 'axiomatic' non-equilibrium statistical mechanics: It will soon become apparent that some situations may exist which escape from the frame of the present discussion. The intention is rather to develop a useful model, the characterization of which could be made in terms appealing also to a mathematically inclined mind. The viewpoint adopted here represents in fact a compromise between the aspiration to the greatest generality and the need to be able to go further than the axioms themselves.

It was recognized almost from the beginning of the theory that the essence of Statistical Mechanics can be sketched as follows: Given a 'physical system' \( \bigcup \), it may happen that, when one performs on it the measurement of only a restricted set \( M \) of observables, one is led to a non-mechanical description \( [M] \) of \( \bigcup \). The elements A

*Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1965.
†This research was supported by the Advanced Research Projects Agency (PROJECT DEFENDER), monitored by the U. S. Army Research Office, Durham, under Contract DA-31-124-ARO-D-139.
‡Present address: Institute of Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland.
of $M$ will be referred to as macroscopic observables. The name 'actual' observables will also be attributed to them for reasons which will be clear in the next paragraph. The description $[M]$ will be referred to as the macroscopic description of $\sum$. The phenomenological treatment of the information obtained through the macroscopic measurements is the aim of Thermodynamics. The purpose of (non-equilibrium) Statistical Mechanics is then to restore the validity of Mechanics modulo some further assumptions of statistical character.

To achieve this, one first assumes:

A 1a: The macroscopic description $[M]$ of $\sum$ can be embedded in a microscopic description $[m]$ of $\sum$ in terms of a set $m$ which contains $M$ as a proper subset.

The elements $a$ of $m$ will be referred to as microscopic or 'ideal' observables. The idea is that it could be possible in principle to perform the measurements of these observables $a$, but that in fact one restricts ones attention to the subset $M$ of $m$. By this it is meant that one actually performs only the measurements associated with the elements $A$ of $M$. Hence the names 'ideal' vs. 'actual' observables.

The above assumption can be completed by:

A 1b: Each state $\Phi$ defined on $M$ can be extended to at least one state $\varphi$ defined on $m$.

The lack of structure of the above scheme is patent. To be able to go further, one is forced to impose further restrictions. The first two of these will be mainly of mathematical convenience. Postulates of the kind of $A2a$ and $A2b$ below were suggested to the physicists by Segal.1 They were later on strongly advocated by Haag2 and his school as a convenient tool for the formulation of axiomatic field theory. These two postulates will be given here in a rather dry mathematical form. It should be noted, however, that they can be justified from considerations of more immediate physical significance. Since a detailed discussion of this step would lead to a great digression from the main topic of these lectures, the best we can do here is to mention that a starting point for this justification can be found in Reference 3 and several references quoted therein, and for a form closer related to the formulation below, in Reference 4.

A 2a: To both $M$ and $m$ one can associate $\mathbb{B}^*$-algebras (respectively $\mathcal{U}_M$ and $\mathcal{U}_m$) in such a way that the building elements of $M$ (respectively $m$) are associated to self-adjoint elements of $\mathcal{U}_M$ (respectively $\mathcal{U}_m$). One moreover assumes that $\mathcal{U}_M$ is a subalgebra of $\mathcal{U}_m$.

A 2b: To each state either defined on $M$ or on $m$ one can associate a positive linear bounded functional respectively defined on $\mathcal{U}_M$ or $\mathcal{U}_m$.

To make these assumptions at least intelligible, it would perhaps be useful to quickly recall that:
A $B^*$-algebra $\mathbb{B}$ is a vector space defined on the complex numbers (the Mathematician speaks of the field $\mathbb{C}$) and equipped with the following attributes:

(i) To each pair of elements $A$ and $B$ of $\mathbb{B}$, one can associate an element $AB$ of $\mathbb{B}$ called the product of $A$ and $B$, and defined in such a way that $\mathbb{B}$ becomes an algebra on $\mathbb{C}$.

(ii) To each element $A$ of $\mathbb{B}$ one associates a real number, denoted by $\|A\|$, in such a way that $\|\ldots\|$ becomes a norm on $\mathbb{B}$.

(iii) Each element $A$ of $\mathbb{B}$ possesses an adjoint $A^*$ in $\mathbb{B}$. This definition is admittedly sketchy; more can be found in the classical textbooks by Naimark and Rickart. A useful prototype of a $B^*$-algebra is $\mathcal{B}(\mathcal{H})$ (formed by the bounded linear operators acting on a Hilbert space $\mathcal{H}$). This algebra and some other related algebras will be extensively used in the following.

As far as the assumption A2b is concerned, it is good to remember that

A positive bounded linear functional $\Phi$ on a $B^*$-algebra $\mathbb{B}$ is defined as a mapping of $\mathbb{B}$ into $\mathbb{C}$ which satisfies the following properties:

(i) $\Phi$ is linear

(ii) $|\Phi(A)| \leq \|\Phi\| \|A\|$ $\forall A \in \mathbb{B}$

(iii) $\Phi(A^*) \in \mathbb{R}^+$ $\forall A \in \mathbb{B}$

The set of all $\Phi$ satisfying the properties (i) and (ii) above will be denoted by $\mathbb{B}^*$. The subset of $\mathbb{B}^*$ formed by the elements of $\mathbb{B}^*$, which moreover satisfy (iii), will be denoted by $(\mathbb{B}^*)^+$.

That a physical state must at least satisfy the requirements of assumption A2b should have become clear—if it was not before—to this audience from the quite explicit lectures presently given here by Professor Furry.

As far as axiomatic is concerned, one can still point out that A2a and A2b together make A1b redundant.

The following postulate will now introduce the fact that the microscopic description is mechanistic:

**A3:** The time evolution of the system $\sum_{m}$ as described through $\mathfrak{B}_m$ is given by a continuous representation of the additive group $\mathfrak{R}$ in the group of the automorphisms of $\mathfrak{B}_m$. In plain words this means that for every instant $t$ there exists a mapping of $\mathfrak{B}_m$ onto itself which preserves the structure of $\mathfrak{B}_m$ and is continuous in $t$. This mapping will be referred to as $\mathfrak{B}^{-t}$ (needless to say, the minus sign is completely irrelevant and is introduced here for notational convenience only).

Up to now the scheme defined by assumptions A1a to A3 is quite general. In particular, it covers the following situations:
(a) Both descriptions \([M]\) and \([m]\) are in terms of classical theory. This corresponds to the usual Classical Statistical Mechanics.

(b) \([M]\) is still classical, whereas \([m]\) is a quantum mechanical description. This is then nothing but the usual Quantum Statistical Mechanics.

Many other, more or less unorthodox, situations are also covered by the above axioms. As a curiosity, one could for instance remark that this would be the case of a theory attempting a classical interpretation of Quantum Mechanics (in that case \([M]\) would be quantal and \([m]\) classical). We, however, want to be able to disregard such very general situations in order to concentrate on our main problem which is the exhibition of a mathematical model of some use for the study of Non-equilibrium Statistical Mechanics. To do so, further restrictions will be imposed on the structure emerging from the three types of axioms proposed above.

**A 4a:** The system \(\sum\), as described by \(\mathbb{H}_m\), is a 'simple' (or irreducible) quantum system.

By a 'simple' quantum system, we will understand that there exists a Hilbert space \(\mathbb{H}\) which can be used to represent \(\mathbb{H}_m\) as \(\mathbb{H}(\mathcal{S})\), where \(\mathbb{H}(\mathcal{S})\) is the set of all bounded linear operators defined on \(\mathcal{S}\). In the language used in Reference 3 this is to say that the propositions system associated with \([m]\) is irreducible. *Physically, this postulate means that we will concentrate on the simplest form of Quantum Statistical Mechanics to the detriment of Classical Statistical Mechanics. The latter will not, however, completely escape from the considerations presented later on and it should become clear in the second of these lectures that slight changes would also allow us to include Classical Statistical Mechanics in the present scheme. Taken literally—as it should be!—the above assumption A 4a is even more drastic: it explicitly excludes the cases where the microscopic description exhibits superselection rules. A realistic theory should obviously allow us to treat such situations, if only because of the many-body structure of most of the problems of Statistical Mechanics. It is, however, the hope of the present discussion that one can describe some of the major aspects of statistical mechanics (such as irreversibility) even within the restricted frame imposed by the above assumptions. The following axiom is postulated in the same spirit: were it to the expense of some not too drastic lost in generality, we are willing to make some simplifying assumptions which could lead to a more concise discussion of a precisely formulated model.

---

*As a mathematical consequence \(^8\) of this, we have now

\[
A^t = U^{-t} A = U^{-t} A U^t.
\]
A 4b: \([\mathcal{M}]\) is a classical description.

We say that a description is classical when the set of all the observables used to formulate it is a compatible set. This means that the dispersion, obtained when an observable of this description is measured, is not affected by the simultaneous measurement of any other observable of the same description. Mathematically the assumption A4b is equivalent to the following statement (Remember that as a result of the preceding axioms \(\mathcal{M} - \mathcal{M}_m = \mathcal{A}(\emptyset)\)):

\[ AB = BA \text{ for all } A \text{ and } B \text{ in } \mathcal{M}_m, \]

i.e.,

\[ \mathcal{M}_m \subseteq \mathcal{M}. \]

Using the language of Reference 3, this is equivalent to saying that the propositions-system associated with \([\mathcal{M}]\) is a Boolean lattice. For an attempt to treat a situation where A4b is not satisfied, see Reference 16. We shall not, however, enter here in the discussion of such a generalization.

The next restriction is

A 4c: Each self-adjoint element of \(\mathcal{M}_m\) has a discrete spectrum.

The physical justification for this could be that actual measurements are always made with a limited precision only, and can therefore be described by a discrete set of numbers. (Because of the confusion which persists in the current literature on this point, it is perhaps not useless to remark again here that the set of all possible values of an observable is not the same thing as the spectrum of this observable: the former is much larger than the latter.) It should then be noted that nothing like A4c is assumed on \(\mathcal{M}_m\). In particular, the microscopic energy, defined as the generator of the time translations in \(\mathcal{A}\), can still have a continuous spectrum.

For the purpose of these talks, it will be important to remark on the following mathematical consequence of the above assumptions: \(\mathcal{M}_m\) generates a partition \(\{E_\Delta\}\) of \(\mathcal{A}\) into mutually orthogonal subspaces which are defined as the maximal common eigensubspaces of all the self-adjoint elements of \(\mathcal{M}_m\). One can therefore write each of these elements in the form

\[ A = \sum_\Delta A(\Delta)E_\Delta. \]

Physically the \(E_\Delta\) will be interpreted as the quantum analog of the classical macrocells. Macroscopic measurements will then determine the relative population of the macrocells. However, nothing
can be inferred, from the macroscopic measurements at a given time, about the distribution inside the macrocells at that given instant. This remark is made here to give an intuitive meaning to the term used. Mathematically, it is not quite precise but will be improved upon in a while.

Before we go into this, I wish to come back to the microscopic description and make some comments on the definition of the microscopic states. In the first two lectures and in a good deal of the third we shall not admit as physical states all the elements of \( (W^*_{m})^+ \) but restrict our attention to those which, moreover, satisfy the following postulate:

**A 5:** The physical microscopic states are normal.

This postulate should not be confused with the assertion that the states are normalized (or at least normalizable). That physical states have to be normalized is of common use in the whole of physics and this condition deserves no other comments here than those presented by Professor Furry. 7) The assumption A 5 is of a more subtle nature, both from the mathematical point of view and from the physical consistency of the theory. Explicitly stated, a positive bounded linear functional \( \Phi \) defined on a \( B^* \)-algebra \( \mathcal{W} \) is said to be normal if

\[
\Phi \left( \sum_i p_i \right) = \sum_i \Phi(p_i)
\]

for any set \( \{ p_i \} \) of non-intersecting projectors of \( \mathcal{W} \). This is nothing but the ‘or’-law of the theory of probabilities. The mathematical convenience aspect of A 5 enters now in the following. The theory of probabilities requires the above additive law for any sequence of mutually exclusive events, including infinite sequences, whereas physically one can only require that for finite sequences. In fact, the above relation follows from linearity for finite sequences! Therefore, the content of A 5 is that this law is also required for infinite sequences. This requirement is obviously an idealization which goes outside of the range of physical testability. The importance of A 5 lies then entirely in the fact that, in case of infinite-dimensional \( \mathcal{W} \), it is a necessary condition to prove that one can associate to each physical state \( \Phi \) an element \( W \in \mathcal{B}(\mathcal{W}) \) satisfying

1. \( W^* = W \quad \text{Tr} W = 1 \quad W \in \mathcal{B}(\mathcal{W})^+ \)
2. \( \Phi(A) = \text{Tr} W A \).

These states are usually referred to as density operators or density matrices. In these lectures we shall use this standard terminology or simply refer to them as states. When the moment comes where we shall be forced to consider more general states, we shall refer to states which do not satisfy A 5 as 'generalized states.'

The mathematical structure determined by the axioms A 1 to A 5 is now rich enough to enable us to go further than a somewhat...
sterile axiomatic. A single further postulate will be needed later on, but even without it we can already make some remarks about our model.

We first want to analyze the structure of the set $C$ of all the states (remember that $A\mathcal{S}$ is assumed). We remark that $C$ is a convex set, the extremal points of which are the pure states (the latter being recognized as those states for which $W^2 = W$). We can, moreover, define on $C \times C$ a positive definite functional, linear in the sense of convex sets:

$$(W_1, W_2) = \text{Tr} \, W_1 W_2.$$ 

This looks so much like the germ of a scalar product that it is strongly tempting to try to embed $C$ in a Hilbert space, a Banach space, a $B^*$-algebra or even better, if possible, in a Hilbert algebra. The reason for this temptation is obviously that we intend to use as much Functional Analysis as will be useful.

The starting point for this is to see that $C$ can be embedded, without any mathematical or physical restrictions, in the set $\mathfrak{B}$ defined as

$$\mathfrak{B} = \{A | A \in \mathfrak{B}(\mathfrak{B}); \text{Tr} A^* A < \infty\}$$

equipped with the sesquilinear form

$$(A, B) = \text{Tr} \, A^* B.$$ 

The next question is: What can the Mathematician teach us about $\mathfrak{B}$? First of all he recognizes in $\mathfrak{B}$ the set of all Hilbert-Schmidt operators acting on $\mathfrak{B}$. He then remembers that $\mathfrak{B}$ has a name in the mathematical literature: it is the Hilbert-Schmidt class. As physicists, however, we prefer to use the more intuitive name of Liouville space attached to $\mathfrak{B}$, for reasons which will become evident later on. For the moment, however, let us still take advantage of the presence of the Mathematician. We learn from him that the Hilbert-Schmidt class is a complete Hilbert space so that we can work on $\mathfrak{B}$ with tools which are of common use among Physicists. Moreover, $\mathfrak{B}$ is also a Hilbert*-algebra. These pieces of information will suffice us for the moment. We shall, however, need some more knowledge about $\mathfrak{B}$ and its mathematical surroundings in the third lecture.

As Physicists, our familiarity with $\mathfrak{B}$ will be nothing but improved if we emphasize that:

(i) The scalar product in $\mathfrak{B}$ has a well-defined physical meaning: for any self-adjoint element $A$ of $\mathfrak{B}$ and any density operator $W$, $(A, W)$ is the expectation value of
the observable $A$ on the state $W$. To some extent, then, the scalar product in $\mathfrak{g}$ has even a more immediate meaning than the usual scalar product in $\mathfrak{g}$.

(ii) The norm in $\mathfrak{g}$ is defined as

$$\|A\|_{\mathfrak{g}} = (\text{Tr} \ A^* A)^{\frac{1}{2}}$$

and should not be confused with the ordinary norm of the same element considered as an operator acting on $\mathfrak{g}$. When calculated for states, this new norm is some kind of a measure of the degree of mixture of $W$. It is therefore tempting to use this norm to define a substitute for the entropy:

$$S' = -k \ln \left\{ \frac{\|W\|_{\mathfrak{g}}^2}{\mathfrak{g}} \right\}.$$ 

The Physicist will convince himself that for many purposes this could be quite a convenient substitute for the usual entropy, this being especially true for isolated systems.

(iii)

$$\frac{d}{dt} \ u^\dagger W = -iL u^\dagger W$$

i.e., the operator $L$ acting on $\mathfrak{g}$ is defined as the infinitesimal generator of $\{u^\dagger\}$. Formally

$$u^\dagger = e^{-iLt}.$$ 

$L$ is related to the infinitesimal generator of the time evolution $\{U^t\}$ in $\mathfrak{g}$ by the formal relation

$$L = [H , \cdot ]$$

where $H$ is the microscopic energy. $L$ is then the exact quantum analog of the classical Liouville operator defined as

$$L = [H , \cdot ]$$

where now $[ , ]$ denotes the Poisson bracket. Hence the name of Liouville space for the space on which the operator $L$ acts.

The next thing to do is to say something about the relation between the microscopic and the macroscopic descriptions. Quite generally we shall say that

**D.3**: Two microscopic states are macroscopically equivalent if their respective restrictions to $\mathfrak{m}$ are identical.
In symbols:

\[ \phi \sim \phi' \iff \phi(A) = \phi'(A) \quad \text{for all } A \in M. \]

In our model this amounts to saying that

\[ W \sim W' \iff p(\Delta) = \text{Tr} WE_\Delta = \text{Tr} W'E_\Delta = p'(\Delta) \quad \text{for all } \Delta. \]

The above definition gives precisely what the Mathematician calls an equivalence relation. It therefore induces a partition of \( \mathcal{E} \) into equivalence classes, each of which is characterized by the set \( \{p(\Delta)\} \). To avoid possible misunderstanding we shall change from the terminology used in References 9, 10, 11 and 12 and call each of these equivalence classes a macroscopic state.\(^{13}\) This concept of macroscopic equivalence classes of microscopic states is of central importance in a study of statistical mechanics (and is perhaps even more useful there than it is in the quantum theory of measurement\(^{7,14}\)).

It should be clear that the very definition of the \( p(\Delta) \) implies that an evolution equation for them would be all that is needed to describe the macroscopic evolution of the system of interest. Such an equation is usually referred to as a master equation. The second lecture of this series will be devoted to the establishment of the connections between the microscopic dynamics and the master equation.

What we did up to now was to describe a microscopically simple (remember that this adjective has a precise mathematical and physical meaning!) quantum system. We then introduced the fact that we are not interested in the complete information attached to the evolution of this system, but rather to the partial information relevant to the measurement of a restricted set of compatible observables. This is then nothing but a particular problem in quantum mechanics. The statistical elements of the theory enter only when we must decide what kind of object we will choose as initial state on the basis of the partial information obtained through a macroscopic measurement at time \( t=0 \). In order to be able to make this choice on reasonable grounds, we will need a further (and last!) postulate for our model.

A.6: Given any macroscopic state \( \{p(\Delta)\} \), there always exists at least one microscopic state \( W \) which reproduces exactly the microscopic information contained in the macroscopic state considered.

In order to give a mathematical transcription of this postulate, we define a macroscopically trivial symmetry as an automorphism \( U \) of \( \mathcal{H}_M \) which leaves \( \mathcal{H}_M \) element-wise invariant. Such an automorphism can be implemented by a unitary operator \( U \in \mathcal{U}_M \),\(^{15}\), and conversely any unitary operator of \( \mathcal{U}_M \) implements a macroscopically trivial symmetry. The image \( UW \) of any state \( W \) through a macroscopically
trivial symmetry is macroscopically equivalent to \( W \) itself. Conversely, for any two macroscopically equivalent microscopic states there exists a macroscopically trivial symmetry which allows passage from one to the other. Two such states, however, contain different microscopic informations since they lead to different expectation values for at least some microscopic observables. Therefore, a necessary and sufficient condition for a state \( W \) to give the same microscopic information as that contained in its equivalence class only is that it should be left invariant by all macroscopically trivial symmetries. As a consequence of this requirement, \( W \) has to be uniform in each macrocell. It is then possible to satisfy A6 only when all the \( E_\Delta \) are finite-dimensional. This condition is not only necessary, it is also sufficient. Therefore, an alternative form of A6 is:

\[
N_\Delta = \dim E_\Delta = \| E_\Delta \|_B^2
\]

for all \( \Delta \).

An explicit form of \( W \) can then be given:

\[
W = \sum_\Delta p(\Delta) W_\Delta
\]

where

\[
W_\Delta = \frac{E_\Delta}{N_\Delta}.
\]

\( W \) is characteristic of its equivalence class, and unique. It will be referred to as the maximal representative of its equivalence class. It is worth noting that the entropy \( S(W) \) reaches its maximum inside an equivalence class for \( W \) and has then the value \( k \cdot N_\Delta \).

Our last remark will be the following. Given any state \( W \) the maximal representative of its equivalence class can be obtained by applying to \( W \) an operator \( \mathcal{A} \) defined on \( \mathcal{A} \) by:

\[
\mathcal{A} \cdot A = \sum_\Delta (E_\Delta, A) W_\Delta
\]

\[
= \sum_\Delta (W_\Delta, A) E_\Delta
\]

From the point of view of principles, it is important to realize that \( \mathcal{A} \) is introduced here as an operation on states and not on observables. \( \mathcal{A} \) satisfies the following properties:

(i) \( \mathcal{A} \) is linear and bounded

(ii) \( \mathcal{A}^* = \mathcal{A} \)

(iii) \( \mathcal{A}^* = \mathcal{A} \)
i.e., $\mathcal{P}$ is a projector on $\mathcal{B}$. It will be referred to as the coarse-graining projector and its range in $\mathcal{B}$ will be called the macroscopic subspace of $\mathcal{B}$ since only what happens in $\mathcal{P}\mathcal{B}$ is relevant for the macroscopic description. An orthonormal basis in $\mathcal{P}\mathcal{B}$ is provided by the set of vectors:

$$\Delta = \frac{\mathcal{E}_\Delta}{\|\mathcal{E}_\Delta\|},$$

and one can prove the following important relation:

$$\mathcal{D}\mathcal{L}\mathcal{E} = 0.$$

**SECOND LECTURE**

**Master Equations**

We want now to discuss the possible evolution equations which are relevant for the model proposed in the first lecture. The main features of this model can be summarized as follows: We developed a quantum Liouville space formalism for non-equilibrium statistical mechanics. In this formalism all the physical states (no matter whether pure or mixtures) and all the observables which are necessary for a statistical mechanical description are represented as vectors in the same Hilbert space: the Liouville space. Moreover, the scalar product $(A, W)$ of an observable $A$ and a state $W$ is the expectation value $\text{Tr} AW$ of this observable for this state. This is quite similar to the situation encountered in Classical (Statistical) Mechanics. The similarity even goes further and, for instance, the time evolution in the Liouville space is generated by an operator $L$ which is the exact quantum analog of the classical Liouville operator. We also introduced the fact that we are willing to deal with the partial information which can be obtained from a restricted set of observables: the macroscopic observables. These were (in our model) of the form:

$$A = \sum_{\Delta} A(\Delta) \mathcal{E}_\Delta$$

where $\{\mathcal{E}_\Delta\}$ is a fixed partition of the usual Hilbert space $\mathcal{B}$ in orthogonal projectors. These $\mathcal{E}_\Delta$ were interpreted as macrocells and, when considered as vectors in $\mathcal{B}$, they span a subspace $\mathcal{P}\mathcal{B}$ of $\mathcal{B}$ which we called the macroscopic subspace of $\mathcal{B}$. The projector $\mathcal{P}$ on this subspace was called the coarse-graining projector. We finally made the observation that, given an initial state $W^0$, every statement about the macroscopic evolution of the system could be made in terms of the occupation probabilities $p(\Delta)$ defined as
\[ p^f(\Delta) = (E_\Delta, W^f) = (E_\Delta', u^{t'} W^f_\Delta). \]

It should also be remembered that by 'prediction about the macroscopic evolution' we only mean the possibility to calculate the evolution in time of the expectation values of the macroscopic observables.

Since the \( p^f(\Delta) \) are interpreted as probabilities, the most naive approach would be to discuss their evolution in terms of Stochastic Theory. In this spirit the most simple question would be to see whether the \( p^f(\Delta) \) could obey a Markovian evolution equation. (The complete answer to this question is not yet to be considered as known.) We shall first see that if the coarse-graining operation only is taken into account, then the Markovian assumption is inconsistent with the 'rules of the game', i.e., with the assumption that the microscopic evolution is governed by a microscopic quantum mechanical law.

Since Professor Uhlenbeck\(^{17}\) is lecturing on the theory of Markov processes just now, here in Boulder, it is rather inapposite on my part to present this theory again in front of this audience. Let me however do so, for the sake of self-completeness of this lecture. I shall in any case be very brief.

Let \( \Sigma \) be the system of interest and suppose that the outcome of an experiment on \( \Sigma \) can be described with the help of a discrete set \( \{ \Delta \} \) of 'possible states'. The evolution of \( \Sigma \) as described by \( \{ \Delta \} \) is by definition a Markovian process if the conditional probabilities satisfy:

\[
p_n^{(\Delta_n^{t_n}, \Delta_{n-1}^{t_{n-1}}, \ldots, \Delta_1^{t_1})} = p_n^{(\Delta_n^{t_n}, \Delta_{n-1}^{t_{n-1}}, \ldots, \Delta_1^{t_1})}
\]

for any sequence \( t_n \geq t_{n-1} \geq \ldots \geq t_1 \). If, moreover,

\[
p_n^{(\Delta_n^{t_n}, \Delta_{n-1}^{t_{n-1}})} = p_n^{(\Delta_n^{t_n}, \Delta_{n-1}^{t_{n-1}})}
\]

whenever \( t_n - t_{n-1} = t_n' - t_{n-1}' \), then the Markovian process is said to be homogeneous.

In the case of the present model, we have

\[
p_n^{(\Delta')^{(\Delta)}} = (E_\Delta', u^{t'} W_\Delta)
\]

and the evolution is therefore homogeneous in time without any regard to its possible Markovian character.

As an immediate consequence of the definition, one has for any Markovian process:
STATISTICAL MECHANICS

\[ p(t''|t) (\Delta''|\Delta) = \sum_{\Delta'} p(t''|t')(\Delta''|\Delta') p(t'|t)(\Delta'|\Delta) \]

for any \( t' \in (t,t'') \), which reduces for an homogeneous process to

\[ p_{t_1+t_2} (\Delta''|\Delta) = \sum_{\Delta'} p_{t_2} (\Delta''|\Delta') p_{t_1} (\Delta'|\Delta) \]

valid for any positive \( t_1 \) and \( t_2 \). In our model the assumption that the macroscopic evolution is Markovian is equivalent to the requirement (semi-group property) that

\[ \mathcal{T} u_{t_1+t_2} \mathcal{T} = \mathcal{T} u_{t_2} \mathcal{T} u_{t_1} \mathcal{T} \]

from which we conclude that

\[ \frac{d}{dt} \mathcal{T} u^t \mathcal{T} = -i \mathcal{D} \mathcal{T} u^t \mathcal{T} \]

We saw, however, at the end of the preceding lecture that

\[ \mathcal{D} \mathcal{T} \mathcal{T} = 0 \]

from which we now conclude that the macroscopic evolution cannot be Markovian without being stationary.

The conclusion of the above elementary analysis is that the coarse-graining of a simple quantum mechanical system as described in the first lecture can never by itself allow a macroscopic description in terms of a Markovian process. This result could also have been derived in a more sophisticated way from the main embedding theorem of the Appendix of Reference 18.

We, however, know that there exists in nature some systems whose macroscopic description can be made, at least within some approximation, in terms of Markovian equations and whose description should be possible, at a microscopic level, in terms of ordinary quantum mechanics. This suggests that in order to allow, at least in principle, the very existence of such situations, we must introduce some further ingredient in our model. The idea which naturally occurs is the following: Since we have introduced coarse-graining in order to take into account the limitation inherent to any real physical experiment, we should also include in our formalism some kind of time-smoothing which would reflect the fact that actual measurements have a finite duration. In the same sense that the coarse-graining is not only linked to the limited precision of the experiment,
but rather to the limited class of measurements that the experimenter is willing to consider, time-smoothing could also be justified by the fact that the same experimentalist is interested in phenomena, the time-scale of which is large enough to justify a time-averaging over a certain microscopic time. Such a time could be, for instance, of the order of the collision time if one is interested in the approach of equilibrium of a gas. The fact that the same evolution equations are not valid for submicroscopic times and for macroscopic times provides a further a priori justification for an attempt to introduce time-smoothing and to discuss its possible influence on the considerations presented above.

So much now for the heuristic justification of the next step. Let us therefore become more formal again. We want to introduce the fact that we are not interested (even for macroscopic observables) in

\[ \langle A \rangle_W(t) = \langle A, W \rangle \]

itself, but rather in

\[ \langle A \rangle_W(t) = \frac{1}{\tau} \int_0^{\tau} dt' \langle A \rangle_W(t') \]

which can be written for macroscopic observables as

\[ \langle A \rangle_W(t) = \sum_{\Delta} A(\Delta) \overline{p}(\Delta) \]

where

\[ \overline{p}(\Delta) = \frac{1}{\tau} \int_0^{\tau} dt' p(t') \]

and where \( \tau \) is the non-vanishing microscopic time over which one wishes to average.

Similarly, one defines

\[ \overline{P}(\Delta' | \Delta) = \frac{1}{\tau} \int_0^{\tau} dt' P(t' | \Delta) \]

so that

\[ \overline{p}(\Delta') = \sum_{\Delta} \overline{P}(\Delta' | \Delta) p(\Delta) \]

\( \overline{P} \) can also be written explicitly as

\[ \overline{P}(\Delta' | \Delta) = \left( E_{\Delta'}, W(\tau) U W_{\Delta} \right) \]
where
\[ \mathcal{W}(\tau) = \frac{1}{\tau} \int_{0}^{\tau} dt \ u^t. \]

\( \mathcal{W}(\tau) \) is a well-defined operator to which we shall refer as the time-smoothing operator. It should be noted at this point that \( \mathcal{W}(\tau) \) obviously commutes with \( u^t \) but not with \( \mathcal{T} \) in general. As a consequence

\[ \mathcal{F} \mathcal{W}(\tau) \mathcal{W} = \mathcal{W}(\tau) \mathcal{F} \mathcal{W} \]

is not true in general. The scruples of the physicist will, however, be removed if he remarks that (because of linearity)

\[ \left( \mathcal{T} \frac{1}{\tau} \int_{t}^{t+\tau} dt \right) \mathcal{W}^t' = \left( \frac{1}{\tau} \int_{t}^{t+\tau} dt \right) \mathcal{W}^t', \]

the physical meaning of which is obvious.

We can now reproduce for the time-smoothed quantities the analysis presented previously for the unsmoothed ones and get under the Markovian assumption:

\[ \frac{d}{dt} \mathcal{T} u^t \mathcal{W}(\tau) \mathcal{W} = -i \mathcal{L} \mathcal{W}(\tau) \mathcal{W}(\tau) u^t \mathcal{W}. \]

The main difference between this expression and the one obtained before is that whereas \( \mathcal{L} \mathcal{L} \mathcal{W} \) was identically zero, this is not the case in general for \( \mathcal{L} \mathcal{L} \mathcal{W}(\tau) \mathcal{W} \). It is then easy to verify that the time-smoothed \( \overline{p^t}(\Delta) \) satisfy the following equation

\[ \frac{d}{dt} \overline{p^t}(\Delta) = \sum_{\Delta'} K(\Delta', \Delta) \frac{1}{N_{\Delta}} \overline{p^t}(\Delta) \]

which can be rewritten in the 'gain-loss' form as

\[ \frac{d}{dt} \overline{p^t}(\Delta) = \sum_{\Delta'} K(\Delta', \Delta) \left( \frac{1}{N_{\Delta}} \overline{p^t}(\Delta) - \frac{1}{N_{\Delta'}} \overline{p^t}(\Delta') \right) \]

where

\[ K(\Delta', \Delta) = -i \langle E_{\Delta'}, L \mathcal{W}(\tau) E_{\Delta} \rangle. \]
Remarks:

(i) \( K(\Delta', \Delta) \) are all real.

(ii) \[ \sum_{\Delta} K(\Delta', \Delta) = 0 = \sum_{\Delta'} K(\Delta', \Delta). \]

This property was already used in the last step above to derive the gain-loss form of the equations.

(iii) \[ \lim_{\tau \to 0} \mathbb{M}(\tau) = 1 \]

and therefore

\[ \lim_{\tau \to 0} K(\Delta', \Delta) = 0. \]

This is in agreement with our previous result when coarse-graining only was used and time-smoothing was not yet introduced. Therefore we see that the compatibility between a mechanical microscopic description and a Markovian macroscopic description of the model depends crucially on the introduction of the time-smoothing operation.

(iv) \( K(\Delta', \Delta) \) depends explicitly on \( \tau \) through \( \mathbb{M}(\tau) \). This probably has to be interpreted in the light of the heuristic comments made when we introduced the concept of time-smoothing. Some attempts have been made to dissolve this dependence into an evaluation of the transition matrix \( K \) in terms of a perturbation technique. This, however, was not quite convincing and doubts even remain as to whether this should be done. We therefore shall content ourselves here with the mention of the existence of this problem without committing ourselves further. In any case, we would like to postpone the discussion of this point until after the derivation of the so-called master equation in its generalized form.

(v) We now want to consider the case where \( \tau \) is very small so that an expansion of \( K \) in first order in \( \tau \) would make sense. We then get

\[ K^{(1)}(\Delta', \Delta) = -\frac{1}{2}(E_{\Delta'}, L^2 E_{\Delta}) \tau. \]

In its gain-loss form the master equation obtained above only depends on the non-diagonal elements of the transition matrix \( K \) which, in the approximation considered here, reduces to
\[ k^{(1)}(\Delta', \Delta) = -\text{Tr}(E_{\Delta'}^V E_{\Delta}^V) \tau \]

where

\[ V = H - \mathcal{D} H. \]

In this approximation our equation resembles very closely that derived originally by Pauli.\(^{20}\) This deserves then some comments. First we would like to recall that our equation was derived under two basic assumptions:

(a) The quantities of interest are time-smoothed expectations values of macroscopic observables.
(b) The macroscopic evolution can be described as a Markovian process.

These two assumptions constitute, in fact, the appropriate generalization of Pauli's original random phase assumption repeated at all times. Therefore, our equation appears as the generalization to all order in \( V \) of the Pauli's equation. It was derived under similar physical assumptions but there was no need to use any perturbation technique.

Our second comment will be that the 'interaction \( V \)' which is responsible for the approach to equilibrium (see below) introduces itself quite naturally as the difference between the microscopic Hamiltonian \( H \) and the macroscopic energy \( \mathcal{D} H \) without reference to any more or less artificial perturbation calculation.

In the above remarks we mentioned that the first approximation in \( \tau \) of our equation implies a macroscopic approach to equilibrium. This is well-known from the Pauli equation and, in fact, was the main reason for the success of this equation. We would like to show explicitly how this appears in our model. We first remember that the macroscopic evolution was entirely described by the macroscopic state \( \{ \bar{\rho}^{(\Delta)} \} \) which can be represented in the macroscopic subspace \( \mathcal{B}^\Delta \) by the vector

\[ \bar{W}(t) = \sum_{\Delta} \bar{\rho}^{(\Delta)}(t) W_\Delta. \]

From the equation

\[ \frac{d}{dt} \bar{\rho}^{(\Delta')} = \sum_{\Delta} k^{(1)}(\Delta', \Delta) \frac{1}{N_\Delta} \bar{p}^{(\Delta)} \]

and
we conclude that
\[
\frac{d}{dt} \overline{W}(t) = -\Lambda \overline{W}(t),
\]
where
\[
\Lambda = -\frac{1}{2} D L^2 \xi \tau.
\]
One can formally integrate the above equation and find
\[
\overline{W}(t) = e^{-\Lambda t} \overline{W}(0).
\]
The essential point is now that the operator $\Lambda$ which maps $S_0$ into itself is a positive self-adjoint operator and can therefore be written as
\[
\Lambda = \int_0^\infty \lambda dP_\lambda
\]
so that
\[
\overline{W}(t) = \int_0^\infty e^{-\lambda t} dP_\lambda \overline{W}(0).
\]
Consequently, the expectation value of any macroscopic observable can be written as
\[
\overline{\langle A, W(t) \rangle} = \int_0^\infty e^{-\lambda t} d\langle A, P_\lambda W(0) \rangle.
\]
Should then $\langle A, P_\lambda W(0) \rangle$ be equivalent to a Lebesgue measure, we would have
\[
\overline{\langle A \rangle}_W(t) = \int_0^\infty e^{-\lambda t} A_W(\lambda) d\lambda
\]
where $A_W(\lambda)$ can be interpreted as the relaxation spectrum of the observable $A$ on the initial state $\overline{W}(0)$. The fact that $\Lambda$ is positive implies that one has for the time-dependent time-smoothed expectation values of macroscopic observables a superposition of non-increasing exponential in time. This is in general sufficient to insure a macroscopic approach to equilibrium. One moreover remarks
that the expectation values and their relaxation spectrum are linked by a Laplace–Stieltjes transformation. This can provide a test for the validity of a Pauli-type master equation when one knows experimentally the time-dependence of the expectation values.\(^{34}\)

The conclusion of the above discussion is that it is in principle not contradictory to assume that a macroscopically Markovian description is compatible with a microscopic quantum mechanical description. This is, however, only possible if one introduces both coarse-graining and time-smoothing. We moreover saw that in some approximation we obtained a macroscopic approach to equilibrium.

It is now tempting to compare the above results with some others which can be found in the literature on kinetic equations. The most recent contribution I know of is the result derived by Chappell and Swenson\(^{33}\) in connection with the statistical theory of plasmas. These authors found an expression for the transition rate \(W_{ab}\) between the eigenstates of an unperturbed Hamiltonian \(H_0\). They next try to interpret their result as the approximation to some order in the interaction \(V\) of the following expression derived from an analogy with Scattering Theory:

\[
W_{ab} = \lim_{\eta \to 0} \eta \int_{-\infty}^{0} dt \, e^{\eta t} 2 \text{Im} \left\{ \langle a | V U^t | b \rangle \langle a | \overline{U^t b} \rangle \right\}.
\]

To make contact with our formalism let us use the following notation:

\[
E_a = \langle a | a \rangle, \quad E_b = \langle b | b \rangle
\]

so that

\[
2 \text{Im} \left\{ \langle a | V U^t | b \rangle \langle a | \overline{U^t b} \rangle \right\}
\]

can be rewritten as

\[
- i \langle L U^{-t} E_a, E_b \rangle
\]

and therefore

\[
W_{ab} = - i \lim_{\eta \to 0} \eta \int_{-\infty}^{0} dt \, e^{\eta t} \langle L U^{-t} E_a, E_b \rangle
\]

\[
= - i \langle L \Re E_a, E_b \rangle
\]

where
$$\mathfrak{r} = \lim_{\eta \to \infty} \eta \int_0^\infty \exp{-\eta t} \ u^t \ dt.$$ 

One remarks that

(i) \(\mathfrak{r} \) exists!

(ii) \(\mathfrak{r} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau u^t \ dt\)

\[= \lim_{\tau \to \infty} \mathfrak{m}(\tau).\]

The formal analogy with our result is, to say the least, surprising! There is, however, a quite severe objection to the present result. Due to a theorem of von Neumann we know what \(\mathfrak{r} \) is. It is a projector defined on \(\mathfrak{r} \), namely the projector on the eigenspace of \(L\) corresponding to the eigenvalue zero. Consequently, \(L\mathfrak{r}\) vanishes and so does \(W_{ab}\). The physicist's reaction to this would almost certainly be that if the result obtained by Chappell and Swenson has some meaning, then one has

(i) to read \(u^t\) not as the complete evolution operator, but as an operator corresponding to some 'reduced hamiltonian';

(ii) to verify that the approximation to the relevant order in \(V\) is not drastically changed as a result of the above substitution.

If one wants to take this kind of argument seriously, one should at least be able to specify what kind of 'reduced hamiltonian' has to be considered. Such a question cannot be answered rigorously without looking more closely to the dynamics, leaving for a moment any conjecture on the possible Markovian character of the macroscopic evolution. To do so, we shall derive on the basis of the dynamics only an evolution equation for the time-smoothed occupation probabilities \(\mathfrak{p}^t(\Delta)\). Since the time evolution preserves any partition in energy shells, one can study the time evolution in each energy shell separately and assume without loss of generality that the hamiltonian is bounded. This assumption implies that the Liouville operator \(L\) is also bounded. We can then write for any complex number \(z\) outside the spectrum of \(L\) (and in particular for all \(z\) such that \(|z| \geq \|L\|\))

\[(z + iL) R(z) = I = R(z) (z + iL)\]
\[ R(z) = \int_0^\infty dt \ e^{-zt} \ u^t. \]

This immediately suggests the use of Laplace-transform techniques. We recall that we have to evaluate:

\[ \frac{d}{dt} p^t(\Delta) = \frac{d}{dt} (\mathbb{E}_{\Delta} \ \mathbb{W}(\tau) \ u^t W). \]

To be able to calculate these quantities, it is sufficient to have an expression for

\[ \frac{d}{dt} \mathbb{W}(\tau). \]

(To abbreviate the notation, we shall temporarily omit the \( \tau \)-dependence of \( \mathbb{W}(\tau) \).)

The Laplace-transform of the above expression is

\[ \mathcal{L} \{zR(z) - I\} \mathbb{W}. \]

From the first resolvent identity one gets

\[ \mathcal{L} \{zR(z) - I\} \mathbb{W} = -i \mathcal{L} LR(z) \mathbb{W} \]

and because of the identity \( \mathcal{L} L \mathcal{L} = 0 \)

\[ \mathcal{L} \{zR(z) - I\} \mathbb{W} = -i \mathcal{L} L (I - \mathcal{L}) R(z) \mathbb{W} \]

From the first resolvent identity again one can evaluate \((I - \mathcal{L}) R(z)\)

as follows:

\[ [z + i(I - \mathcal{L})L] (I - \mathcal{L}) R(z) = (I - \mathcal{L}) - i(I - \mathcal{L}) L \mathcal{L} R(z) \]

and again because of \( \mathcal{L} L \mathcal{L} = 0 \), this leads to

\[ [z + i(I - \mathcal{L})L] (I - \mathcal{L}) R(z) = (I - \mathcal{L}) - iL \mathcal{L} R(z). \]

For the \( z \) chosen from the beginning, the inverse of

\[ [z + i(I - \mathcal{L})L] \]

exists. Let us call it \( S(z) \). We can then write
\((I - \mathcal{D})R(z) = S(z)(I - \mathcal{D}) - iS(z) L \mathcal{D} R(z)\).

Substituting back in (*) one gets:
\[
\mathcal{D} \{ z R(z) - I \} \mathcal{D} = -i \mathcal{D} L S(z)(I - \mathcal{D}) \mathcal{D} - \mathcal{D} L S(z) L \mathcal{D} R(z) \mathcal{D}.
\]

We can now inverse Laplace-transform this equation to obtain
\[
\frac{d}{dt} \mathcal{D} U^t \mathcal{D} = -i \mathcal{D} L \mathcal{D}^t (I - \mathcal{D}) \mathcal{D} - \int_0^t dt' \mathcal{D} L \mathcal{D}^t \mathcal{D} U^{t-t'} \mathcal{D}.
\]

As a side remark, we see that it does not make the slightest difference if one replaces in the above expression
\[
\mathcal{D}^t = e^{-i(I-\mathcal{D})L} t
\]

by
\[
U^t_{(I-\mathcal{D})} = e^{-i(I-\mathcal{D})L(I-\mathcal{D}) t}.
\]

We can now take the matrix elements of this operator equation between the vectors \(E_{\Delta'}\) and \(W^0\). If we, moreover, remember that the set \(\{\Delta = E_{\Delta'}/(N_{\Delta'})^{1/2}\}\) is an orthonormal basis in \(\mathcal{D}\) we get
\[
\frac{d}{dt'} \vec{p}^t(\Delta') = -i \langle E_{\Delta'}, L U^t_{(I-\mathcal{D})} (I-\mathcal{D}) \mathcal{D} W^0 \rangle
\]
\[
+ \int_0^{t'} dt \sum_{\Delta} K^t(\Delta', \Delta) \left\{ \left( \frac{1}{N_{\Delta}} \right) \vec{p}^{t-t'}(\Delta) - \left( \frac{1}{N_{\Delta'}} \right) \vec{p}^{t-t'}(\Delta') \right\}
\]

where
\[
K^t(\Delta', \Delta) = \langle E_{\Delta'}, L E_{\Delta} \rangle.
\]

The equation for the time-smoothed macroscopic evolution we just arrived at is quite general and it deserves some comments. First of all, it should be remarked that no statistical assumption has been introduced up to now. Therefore, our equation can be considered as equivalent to the von Neumann equation. It therefore contains too much information. This surplus of information is contained in the first term of the right-hand side of our equation. In fact, we must be aware that we do not know precisely what \(W^0\) is because of the partial information provided by the initial macroscopic information.
are therefore obliged to make some assumption on the initial state. This assumption could be expressed as an average over all the possible initial situations compatible with our initial macroscopic knowledge. This, however, is not a precise statement as long as one does not make precise what kind of a priori probability we decide to choose in order to perform this average. We therefore prefer to say that we will choose as initial state the microscopic density matrix which corresponds to the most chaotic situation compatible with our initial knowledge. We saw in the first lecture that this statement can be given a precise meaning and that it is satisfied if \( W^0 = \mathcal{D} W^0 \), i.e., if we choose for \( W^0 \) the maximal representative of its equivalence class. In the absence of time-smoothing, this leads to the vanishing of the first term in the right-hand side of the equation. (Incidentally, we would like to mention that in the special case where all the macrocells are one-dimensional ('fine-grained' situation), this assumption corresponds exactly to one's saying that the initial state is diagonal in the uniquely defined representation where all the macroscopic observables are diagonal). In the presence of time-smoothing we feel justified in completing the condition \( W^0 = \mathcal{D} W^0 \) by the condition \( W^0 = \mathcal{M} W^0 \). This will then lead, also in the presence of time-smoothing, to the vanishing of the first term of the right-hand side of our equation. Under these supplementary assumptions of statistical character, we obtain the following equation:

\[
\frac{d}{dt} \tilde{p}^t(\Delta') = + \int_0^{t'} dt \, K^t(\Delta', \Delta) \left\{ \frac{1}{N_\Delta} \tilde{p}^{t-t}(\Delta) - \frac{1}{N_{\Delta'}} \tilde{p}^{t-t}(\Delta') \right\}
\]

which we shall refer to as the generalized master equation (GEM).

Because of the integration over time, the GEM contains all the so-called 'memory effects' and is therefore quite different in principle from a Markovian equation. We would, however, point out that it is quite possible that for practical purpose the GEM could be reduced to a seemingly Markovian equation in some approximation or for a certain time-scale. Our first remark in this direction would be that the kernel \( K^t \) appearing in the GEM does not depend on \( \mathcal{M}(\tau) \). We therefore conclude that under an appropriate assumption on the initial states the time-smoothed occupation probabilities for the macrocells satisfy the same GEM as the corresponding non-time-smoothed quantities. Up to now everything we said about the GEM (especially its derivation) was rigorous. Our next comment will now be of a more heuristic character. Since in the GEM the occupation probabilities can be time-smoothed 'at will', it is reasonable to assume that they evolve with a characteristic time of macroscopic magnitude. If now the kernel \( K^t \) is interpreted as a collision operator, several authors have been tempted to assume that its main variation takes place in a
time interval where the (time-smoothed) occupation probabilities are almost constant and then 'practically vanishes' outside of a time interval of the order of the microscopic collision time. These assumptions are meant to justify—at least from the physicist's point of view—that in the 'long time limit' (i.e., for macroscopic times) the GEM can reduce to

$$\frac{d}{dt} \mathcal{P}(\Delta') = \sum_\Delta K(\Delta', \Delta) \left\{ \frac{1}{N_\Delta} \dot{\mathcal{P}}^t(\Delta) - \frac{1}{N_{\Delta'}} \dot{\mathcal{P}}^t(\Delta') \right\}$$

where

$$K(\Delta', \Delta) = \int_0^\infty dt \, K^t(\Delta', \Delta).$$

This is often referred to as the 'long time limit' or as the 'Markovian approximation' of the GEM. Whether this use of the term 'Markovian' is appropriate is still open to speculations. Most probably it is not correct as our previous remarks should have shown. We would therefore prefer the term of 'Paulianization' to characterize this procedure which in fact is an attempt to justify the Pauli equation from purely dynamical considerations. I do not want to discuss here the conditions which the system should fulfill in order that this approximation could be valid. One can in fact exhibit physical system for which it is not true that the evolution can in its long-time limit be approximated by a Pauli-like equation.\textsuperscript{34} Suppose, however, for a moment that there exist at least some systems for which the above assumptions make sense. The natural question is then to ask what is the form of $K(\Delta', \Delta)$. From the theory of the Laplace-transform we have

$$K(\Delta', \Delta) = \lim_{\eta \to 0} \mathcal{K}_\eta(\Delta', \Delta)$$

where $\mathcal{K}_\eta(\Delta', \Delta)$ is the Laplace-transform of $K^t(\Delta', \Delta)$. We already know this quantity from the derivation of the GEM:

$$\mathcal{K}_\eta(\Delta', \Delta) = -(E_{\Delta'}, \text{LS}(z)L E_{\Delta}).$$

It is now interesting to note that

$$\mathcal{D} \text{LS}(z)L \mathcal{D} = -iz \mathcal{D} \text{LS}(z) \mathcal{D}$$

so that

$$K(\Delta', \Delta) = \lim_{\eta \to 0} \mathcal{K}_\eta(\Delta', \Delta) = \int_0^\infty dt \, e^{-\eta t} (E_{\Delta'}, \text{LS}(z)L E_{\Delta}).$$
The Paulianization of the GEM leads therefore to a result analogous to that of Chappell and Swenson referred to previously. There is, however, an important difference: the exact Liouville operator which occurs in the result of these authors and which makes it identically zero has been replaced here by a reduced evolution operator, namely, \((I - D)L\), which occurs in \(\mathbb{H}^1\). We would also like to point out that the approximation of this result to the appropriate order in \(V\) reproduces the starting point of the investigation of Chappell and Swenson and can therefore be considered as its correct generalization to all order as well as an alternative justification of the approximative result they started from.

These comments about the Paulianization of the GEM are admittedly of rather heuristic nature as long as one has not linked, in a one-to-one way, the assumptions underlying this procedure with the properties of the interaction. We, however, think that it was not without interest to indicate here how the rigorous considerations possible on our model can be taken as a starting point for the discussion of some more real physical situation.

THIRD LECTURE

Ergodic Theory

The purpose of this last lecture is to point out some of the problems relevant for the discussion of the actual approach to macroscopic equilibrium. Let me begin with a perhaps quite subjective assertion: I think that the difficulties I am about to mention are not at all characteristic of the model under consideration. I rather suspect that the reason why they do not appear in most of the modern papers where they should have been felt is mainly that they are indeed hidden behind a cloud of semi-heuristic arguments, the lack of mathematical precision of which is quite often too patent. I realize that this statement is a rather strong one. I shall, however, immediately smooth its apparent severity by the two following remarks:

(i) We shall mainly point out the difficulties.

(ii) It would be erroneous to conclude from the above remarks that most of the results of modern non-equilibrium have to be rejected by the Physicist on the mere basis of their lack of mathematical rigor. The actual aim of this lecture is rather to make it clear that in spite of all the material accumulated in this field (and probably because of it) still much remains to be done in the direction of a better understanding of the foundations of Non-Equilibrium Statistical Mechanics.

The first thing to do when one wishes to discuss the actual approach to macroscopic equilibrium is to define precisely what is meant by this expression and in particular to describe what kind of
properties one would expect from the equilibrium state(s). In this
respect ergodic theory can be of some real use. The reason is that
if the system under consideration ever approaches (in any sense) an
equilibrium state, then this state has to be equal to the average, over
an extremely long time, of the evolution of the initial state. One
then has to discuss

$$\lim_{\tau \to \infty} \mathfrak{M}(\tau) = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau dt \; \mathfrak{U}^t.$$ 

In the formalism used in these lectures the evolution operator \( \mathfrak{U}^t \) is a
unitary operator acting on a complex Hilbert space: the Liouville
space. We can, therefore use von Neumann's mathematical theorem\(^23\) to assert:

(i) The above limit exists in the strong topology in \( \mathfrak{U} \).

(ii) It is equal to the projector on the invariant subspace of

\( \{ \mathfrak{U}^t \} \) which is also the eigensubspace of \( L \) attached to

the eigenvalue zero.

Let me remind you that we have already met this operator in the
second lecture under the name of \( \mathfrak{R} \). Von Neumann's 'ergodic'
theorem allows us in particular to assert that for any \( A \) and \( W \) both
in \( \mathfrak{U} \) the time-averaged expectation values:

$$\langle A \rangle_{W}(t) = \frac{1}{\tau} \int_t^{t+\tau} dt \; \langle A, \mathfrak{U}^t W \rangle$$

approach a limit when \( \tau \) goes to infinity and that this limit is given by

$$\lim_{\tau \to \infty} \langle A \rangle_{W}(t) = \langle A, \mathfrak{R} \rangle_{W^0}.$$ 

Since the first part of this third lecture is essentially based on Reference 12, we shall only mention here the results without proofs. Those
who are interested in the formal proofs are referred to Reference 12.

Lemma 1: \( \mathfrak{R} \) is a closed subalgebra of \( \mathfrak{U} \).

Lemma 2: \( \mathfrak{R} \) is either empty or it contains at least one
finite-dimensional projector acting on \( \mathfrak{U} \).

The interest of these two lemmas is that they lead to a short proof of
the following two theorems which relate \( \mathfrak{R} \) to the microscopic hamil-
tonian \( H \) and allow one then to calculate the effect of \( \mathfrak{R} \) in terms of
the spectral projectors of \( H \).

Theorem 1: A necessary and sufficient condition for \( \mathfrak{R} \) to be
different from 0 is that the discrete part of the
spectrum of \( H \) be non-empty.
Theorem 2: Let \( \{P_i\} \) be the set of all projectors on the proper eigensubspaces of \( H \). Then for all \( A \) in \( \mathfrak{g} \) one has

\[
\mathfrak{g} A = \sum_i P_i A P_i.
\]

This second theorem can now be used to prove the following:

Theorem 3: A necessary and sufficient condition for the ergodic average \( \mathfrak{g} W \) of any state \( W \) to be still a state (in the sense of the first lecture) is that the spectrum of \( H \) is discrete.

(You will immediately see that if the spectrum of \( H \) is not totally discrete, then it is the normalization which goes wrong.) In order to emphasize the physical meaning of this theorem, let us rephrase it as follows:

(i) If the spectrum of \( H \) is discrete, the ergodic average of any state can be replaced by an ensemble average.

(ii) If the spectrum of \( H \) is not totally discrete, there will be states whose ergodic average cannot be replaced by a (properly normalized) ensemble average.

(iii) If the spectrum of \( H \) is totally continuous, no state is such that its ergodic average can be replaced by an ensemble average.

For the purpose of Statistical Mechanics it is of interest to note that the above theorem is not affected if one requires the condition only for state of the form

\[
W = \mathfrak{g} W.
\]

Let us for a moment consider only the case where the spectrum of the hamiltonian is discrete. The case of a totally continuous spectrum will be discussed at the end of this lecture.

Up to now, our remarks were almost only dealing with microscopic arguments. The next step is then to discuss the relations between the microscopic and macroscopic descriptions.

Theorem 4: A necessary and sufficient condition for the ergodic average of any state of the form \( W = \mathfrak{g} W \) to be still of this form is that the coarse-graining projector commute with the ergodic projector.

The physical meaning of the first term of this theorem should be obvious from the first lecture. We shall therefore assume in the following that:

\[ A \mathfrak{g} : \mathfrak{g} \text{ and } \mathfrak{g} \text{ commute.} \]
The above conditions all deal with the existence of at least one equilibrium state for each initial state. We want now to assume some kind of uniqueness.

A 8: Within each energy shell the ergodic average of all the states of the form \( W' = \mathcal{S} W \) is the same state.

Remarks:

(i) One can verify that this assumption is tenable only when the dimension of the energy shell is finite. (Compare with the similar consequence of A 6.)

(ii) The above conditions on the existence and uniqueness of the equilibrium state already determine that state: it is the microcanonical distribution.

(iii) Clearly enough an assumption such as A 8 implies a link between microscopic and macroscopic description. This link can be made explicit through the following theorem:

Theorem 5: A necessary and sufficient condition for A 8 is

\[
\frac{P_i W' \Delta P_i}{N_i} = \frac{P_i}{N_i}
\]

for all \( i \) and \( \Delta \) in the same energy shell \( J \) of dimension \( N_i \).

We are now equipped with two conditions:

(i) \( \mathcal{S} \) and \( \mathcal{S} \) commute

(ii) \( P_i W' \Delta P_i = P_i / N_i \)

which are characterized by two advantages:

(a) they are formally simple enough to be of some use;

(b) their justification in terms of observable requirements seems to be good enough.

As in all kinds of axiomatic attempts, we are now faced with the problem of the connection of our 'axioms' with other plausible physical requirements. This is the common problem of checking the independence of a certain set of axioms. In this respect we have a collection of small theorems, some of which will be reported below.

Theorem 6: The commutation of \( \mathcal{S} \) and \( \mathcal{S} \) implies that all states of the same macroscopic equivalence class have macroscopically equivalent time-averages.

Theorem 7: The reciprocal theorem is also true.

Theorem 8: To assume that

(i) \( \mathcal{S} W = \mathcal{S} W \) for all states \( W \), and

(ii) \( \mathcal{S} W \) does not depend on \( W \) in any given energy shell

is equivalent to assume that

(iii) for all \( W \), \( \mathcal{S} W \) is macroscopically equivalent to the microcanonical distribution.
Theorem 9: Our conditions are equivalent with the requirement that there exists essentially no other macroscopic constant of the motion than the macroscopic energy $\mathcal{E}$. (This suggests a close connection with the classical metric indecomposability; see for instance the remarkable treatise by Janel, 27). We should remark, however, that our present condition is more physical in the sense that it refers only to macroscopic observables.)

Theorem 10: Our conditions imply the validity of the translation in our language of the condition of the 'interconnection of states'.

(For a formulation of this condition in the usual formalism as well as for the exploitation of its consequences, see References 24, 25, and 26.)

The interest of this theorem is the following. Up to now we only considered conditions for the existence and uniqueness (the latter in the restricted sense pointed out above) of ergodic averages. The reason why so much attention was devoted to this point was, as mentioned before, that if the system under consideration ever actually approaches an equilibrium, then the equilibrium values must be equal to the ergodic average. We just wanted to emphasize this point again since it is not the usual justification of ergodic theory. We, however, did not say anything on the occurrence of an actual approach to equilibrium. The remark suggested by the last theorem mentioned before is the following: Our ergodic conditions imply the interconnection of states which in its turn was a decisive condition for the 'dissipative' behaviour of a wide class of physical systems, namely the systems for which the infinite-volume limit exhibits the famous diagonal singularities of van Hove. This shows that our ergodic conditions could be of some use in the definition of dissipative systems. We do not want to go further in this direction here but rather we will present some critical comments (some of which are certainly well-known) on the problem of the actual approach to equilibrium of quantum systems.

We saw that the discreteness of the spectrum of the microscopic Hamiltonian $H$ is needed to assure the existence as states of the ergodic averages of all states (or even the restricted class of those which satisfy $W = \mathcal{E} W$). One can prove, however (see, for instance, Reference 28), that the discreteness of $H$ has another consequence, namely, it implies the occurrence of the quantum analog of the Poincaré cycles. This phenomena is known under the name of recurrences and its mathematical expression is that all states $W^t$ are almost-periodic functions of the time in the norm of $\mathcal{E}$, i.e.,

Given any state $W^0$ and any $\epsilon > 0$, there always exists a $T_\epsilon < \infty$ such that
\[ \| W^t - W^0 \|_B < \epsilon \]

will occur in the interval \((0, T_\epsilon)\) whether or not \(W^t\) has
left this neighborhood of \(W^0\) before.
In plain words, this means that given any neighborhood of any state
\(W^0\), \(U^t\) is such that \(U^t W^0\) will either never go out of this neighbor-
hood, or reenter it again and again, this statement being translatable
in time. Consequently, the limit for infinite time of \(W^t\) can never
exist under this circumstance.

In order still to be able to assert something about the actual
approach to equilibrium, two ways have been proposed:

(i) either one agrees on another definition of the approach to
    equilibrium (some 'approach in the mean') so that one can
    formally disregard the occurrence of recurrence by the use
    of probabilistic arguments;

(ii) or, one takes the infinite-volume limit in such a way that
    the spectrum of the microscopic hamiltonian becomes
    truly continuous. Doing so, one makes the recurrence
    theorem unapplicable.

It is clear that both of these tricks involve some unsatisfac-
tory manipulations from the point of view of principles. The first line
of approach has, in fact, been in the center of ergodic theory from the
very beginning. For a general survey, see again Reference 27. The
modern attempts in this direction can also be found in Reference 29
and, in particular, in the works of the Italian School.\(^{30}\) The second
way to avoid the recurrence difficulties has now reached a quite wide-
spread consensus among Statistical Physicists working on non-equili-
rium phenomena. Aside from any question about the physical
meaning of this approximation, one can also raise some mathematical
objections against this procedure. We want to point out here one of
these objections which is connected with our previous remarks about
ergodic averages.

We saw that the continuity of the spectrum of \(H\) implies that \(\mathcal{A}\)
reduces to zero. This means that for every \(A\) in \(\mathcal{B}\) one has

\[ \lim_{T \to \infty} \frac{1}{T} \int_0^T dt (A, W^t) = 0. \]

For infinite system, however, there are certainly observables which
are not in \(\mathcal{A}\). (Naturally all the states on the system defined as in
the first lecture still belong to \(\mathcal{B}\).) The above equality will not then
necessarily hold true for at least some of these 'outside' observables.
Take, for instance, any observable \(A\) which commutes with the hamil-
tonian. Then the expectation value
is still defined by its expression as a trace and is obviously constant. Therefore, if it is not zero initially its average value in the ergodic sense will not be zero either. We are consequently about to feel that something has become wrong in our formalism. The only thing which could be wrong in this respect is, I think, our definition of states.

We saw in the first lecture that we have good physical reasons to assume that states are positive bounded linear functionals on \( \mathcal{B}(\mathcal{H}) \). I, however, pointed out that this is not sufficient, when \( \mathcal{H} \) is infinite-dimensional, to prove that states may be represented in general by the usual density matrix. To prove this representation theorem, one needs, moreover, to assume that the states are normal. We also saw that this condition is not physical but rather of mathematical convenience. This is not a defect of the proof: there exist positive linear functionals on \( \mathcal{B}(\mathcal{H}) \) which are not representable as density matrices. Here we are obliged to enter a little deeper into the mathematical structure of the so-called simple quantum systems whose description can be made through the set of all linear bounded operators acting on a single Hilbert space \( \mathcal{H} \) of infinite dimension.

First of all, let me remind you what we are concerned with. For a while, still let us assume that A5 is valid. Let then \( W^t \) be the evolution in time of any state \( W^0 \) and that

\[
W^t = U^t W^0 U^{-t}
\]

where \( U^t \) is a continuous one-parameter unitary group acting on \( \mathcal{H} \). Let \( A \) be any bounded observable (the fact that we restrict our attention to bounded self-adjoint operators as observables is well-known to be no restriction since the case of unbounded observables can be reduced to the present case by the use of the spectral decomposition theorem). We now want to know whether

(i) \[
\lim_{T \to \infty} \frac{1}{T} \int_0^T dt \langle A, W^t \rangle \quad \text{exists}
\]

(here also \( \langle A, W^t \rangle \) is defined as \( \text{Tr} A W^t \), but now without reference to \( \mathcal{H} \));

(ii) there exists a state \( \bar{W} \) (by which we still mean a density operator) which could be used to write the above limit (when it exists) as \( \langle A, \bar{W} \rangle \).

To answer this problem, let me remind you that the density operators all belong to the Trace-class \( \mathfrak{T} \). (For all the definitions and properties used below, cf. Reference 6.) \( \mathfrak{T} \) is not a Hilbert algebra itself but is, however, a Banach algebra, the dual of which is \( \mathcal{B}(\mathcal{H}) \) itself. Our question can then be formulated as follows: Can one
prove that the group \( \{ U^t \} \) of automorphisms of \( \mathcal{B}(\mathcal{A}) \) generated by \( \{ U^t \} \) and restricted to \( \mathcal{I} \) is weakly ergodic. Our question is now put in a language that the Mathematician is willing to understand. He even knows the answer\(^{31} \) which is: \( \{ U^t \} \) restricted to \( \mathcal{I} \) is not weakly ergodic, in general. The conclusion of all of this for the Physicist is that in order to represent ergodic averages in case of continuous hamiltonians, he is in general obliged to go outside the class of all the density operators. How is this possible?

Let us denote by \( \mathcal{B}^* \) the set of all bounded linear functionals on \( \mathcal{B}(\mathcal{A}) \). This is a Banach space, known as the dual of \( \mathcal{B}(\mathcal{A}) \). We know something more about \( \mathcal{B}^* \): It can be split into two complementary subspaces, the respective characterizations of which can be easily formulated:\(^{32} \)

\[
\mathcal{B}^* = \mathcal{B}_\# + \mathcal{U}^\perp
\]

where

\( \mathcal{B}_\# \) is the set of all normal bounded linear functionals on \( \mathcal{B}(\mathcal{A}) \). The set of all positive normalized elements of \( \mathcal{B}_\# \) is then precisely the set of all density matrices

and

\( \mathcal{U}^\perp \) is the set of all linear functionals which vanish on the class \( \mathcal{U} \) of all compact operators (these operators are also referred to as completely continuous. This latter terminology is, however, misleading for the careless reader since these operators have a completely discrete spectrum!)

Since we introduced many algebras in the last part of this lecture, it is perhaps not useless to recall that

\[
\mathcal{B}(\mathcal{A}) \supset \mathcal{U} \supset \mathcal{I} \supset \mathcal{E}.
\]

Now it is obvious that there are many observables which are outside of \( \mathcal{U} \), in particular, all of those which have some continuous part in their spectrum. It is then quite reasonable to assume that they are physical states which vanish when restricted to observables belonging to \( \mathcal{U} \) but which are not identically zero when considered as defined on the whole of \( \mathcal{B}(\mathcal{A}) \). Our assertion is then that for systems with continuous hamiltonians the ergodic states, and consequently the equilibrium states, must, when they exist, belong to that class, and therefore are more general than the usual density matrices.

The conclusion of this last remark is then the following alternative:
either one studies better the phenomena of recurrences in the
case of discrete hamiltonians (and this can be done on par-
ticular examples without any use of probabilistic arguments);
or, if one uses continuous hamiltonians to avoid formally the
problem of the recurrences, one must be prepared to be
obliged to enlarge the class of the physically admissible
states outside of the class of the density matrices.

A detailed analysis of this dilemma and of some of its consequences
will be published elsewhere.35)

Acknowledgements

During the time when I was occupied with the problems dis-
cussed in these lectures I had the opportunity to have many interesting
contacts with various persons at the following Institutions:
Institut de Physique théorique de l'université de Genève,
 Battelle Memorial Institute, Geneva, Switzerland.
 CERN, Geneva, Switzerland.
 Istituto di Fisica dell'Università, Pavia, Italy,
 Palmer Physical Laboratory, Princeton University,
 Princeton, New Jersey.
 Joint Institute for Laboratory Astrophysics at the University of
 Colorado, Boulder, Colorado.
 It is a pleasure for me to acknowledge here quite especially the
 friendly disputes I had in the later stage of the preparation of
 these lectures with Dr. R. J. Swenson.

References
 No. 12; G. W. Mackey, Mathematical Foundations of Quan-
tum Mechanics, Benjamin 1963.
2. R. Haag, in Les problèmes mathématiques de la théories des
 champs, CNRS 1959.
3. G. Emch, Theoretical Physics Institute, Boulder, Vol. VIIa
 (1964).
4. M. Guenin, Axiomatic Foundations of Quantum Theories,
 Preprint, Princeton 1965. For an actual field-theoretic non-
 relativistic illustration of these ideas, see G. Emch and
 M. Guenin, Gauge Invariant Formulation of the BCS model,
6. C. E. Rickart, Introduction to Banach Algebras, van Nostrand
 1960.
7. W. H. Furry, Special Topics in Quantum Theory, this volume.
22. The literature on these types of equations is so extensive that a list of references would certainly be incomplete. For some of the most relevant of them, see R. Zwanzig, Physica 30 (1964) and also some references quoted in References 9 and 10.
A quite interesting derivation was proposed recently by A. Strahl, Zeit. fur Physik 185 (1965), 178. This derivation is, however, not quite free of gaps but the only delicate passage can be made safe by the use of the techniques expounded in these lectures.
30. A. Loinger, in the above lecture notes, and the reference quoted by this author.