Symmetry of Physical Laws
Part I. Symmetry in Space-Time and Balance Theorems

SATOSI WATANABE
U. S. Naval Postgraduate School, Monterey, California

In default of the theorem of "detailed balance": $P_{ij}=P_{ji}$, with regard to elementary transition probabilities, several "balance" theorems are introduced and proved on the basis of symmetry of physical laws in space-time. (1) First theorem of "averaged balance" (Sec. 5): We can establish $\bar{P}_{ij}=\bar{P}_{ji}$ by averaging over quantities of "minus class." Table V (Sec. 3) gives a list of "minus" quantities. (2) The so-called "detailed balance of collisions" in classical physics is a special case of Theorem I. (3) Heitler-Coester's theorem of "semidetailed balance" is also a special case of Theorem I. (4) Second theorem of "averaged balance" (Sec. 5): We can establish $\bar{P}_{ij}=\bar{P}_{ji}$ by averaging over quantities with $\rho_R=-1$. The quantities with $\rho_R=-1$ are listed in Table II (Sec. 2). (5) Theorem of "cyclic balance" (Sec. 7): In classical physics, a chain of transitions $i\rightarrow j\rightarrow k\rightarrow\cdots\rightarrow i$ repeats itself cyclically. (6) Theorem of "long-range balance" (Sec. 7): The time average of transition probability from i to j is equal to the time average of transition probability from j to i. Theorems I, II, and III, are direct consequences of inversibility (covariance for space-and-time inversion). Theorem IV is a consequence of reversibility (covariance for time reversal). Theorems V and VI are connected with ergodicity of Markoff's chains. This ergodicity is proved by the condition of bilateral normalization of transition probabilities: $\sum_j P_{ij}=1$, $\sum_i P_{ij}=1$. This bilateral normalization in turn can be derived from either reversibility or inversibility. The limits of validity of all these balance theorems in actual applications are carefully examined in the text.

1. INTRODUCTION

It has previously come to general attention that the principle of detailed balance by no means represents a universal rule in quantum physics.[1] Thus, thanks to Pauli's timely remark, the demonstration of the H-theorem which does not utilize the assumption of detailed balance has acquired a new importance.[2]

The transition probability P_{ij} from state S_i to state S_j has to obey, because of the very nature of probability,

$$\sum_j P_{ij}=1, \quad P_{ij}\geq 0. \tag{1.1}$$

However, the inverse normalization,

$$\sum_i P_{ij}=1, \tag{1.2}$$

is not self-evident.

Husimi[3] and Stückelberg[2] demonstrated that conditions (1.1) and (1.2) are sufficient for derivation of the H-theorem.* A simplified version of this proof was given by Pauli.[2]

Obviously, the detailed balance condition,

$$P_{ij}=P_{ji}, \tag{1.3}$$

allows one to deduce Eq. (1.2) from Eq. (1.1), but this is too restrictive a condition.

It is known that the bilateral normalization, Eqs. (1.1) and (1.2), can be derived from the unitarity of transition matrix (S-matrix) in quantum physics. But, in this paper, the bilateral normalization is considered in connection with the symmetry of physical laws in space-time. It will be pointed out that either reversibility (covariance for time-reversal) or inversibility (covariance for space-and-time inversion) is sufficient to deduce the bilateral normalization, without making use of the unitarity of transition matrices or of detailed balance. Indeed, reversibility or inversibility has a very clear physical meaning and may be considered as a more basic physical principle than the unitarity of transition matrices, which is specifically a quantum-mechanical situation. In the quantum theory of elementary processes there never appear transition matrices which do not obey reversibility or inversibility; therefore, reversibility or inversibility can be considered to be a sufficiently general rationalization of the bilateral normalization.

The principle of detailed balance, Eq. (1.3), is sometimes resorted to in problems other than the H-theorem. It is therefore worthwhile investigating its limits of validity. From Boltzmann's classical work, it is clear that this principle is intimately related to inversibility. Our investigation will show that if the physical system has inversibility, the theorem of detailed balance can be re-established in a broader sense with the help of hypotheses of elementary disorder (or simply chaos hypotheses) with regard to the physical quantities of what will be called the "minus class." To the minus class belong regular tensors and first kind pseudo-

[1] J. Hamilton and H. W. Peng, Proc. Roy. Irish Acad. **A49**, 197 (1944); W. Heitler, *Quantum Theory of Radiation* (Oxford University Press, London, 1944), second edition, p. 252.

[2] E. C. G. Stückelberg, Helv. Phys. Acta, **25**, 577 (1952).

[3] Kodi Husimi, *Theory of Probability and Statistics* (Kawade Shobo, Tokyo, 1942, in Japanese), p. 277.

* *Note added in proof*.—See also K. Yosida, Proc. Acad. (Tokyo) **19**, 43 (1940). Dr. Brockway McMillan was kind enough to point out that the mathematical theorem underlying Husimi-Stückelberg's proof has been known to mathematicians for a long time and that von Neumann also used the same theorem in his book: Johann von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Verlag Julius Springer, Berlin, Germany, 1932), in particular, p. 207.

tensors of odd ranks and second and third kind pseudotensors of even ranks. This classification of tensors will be discussed in detail in Sec. 2.[4]

This result clearly explains that the principle of detailed balance in classical physics is bound to utilize the chaos hypothesis with regard to the positions of molecules, which are minus variables. Heitler-Coester's so-called theorem of semidetailed balance[5] is also a variant of the above-mentioned general rule. Rather inexactly expressed, this theorem of semidetailed balance means that we can re-establish the detailed balance by averaging over the spin directions of particles. Since spin is a minus variable, it is natural that a chaos hypothesis is necessitated with regard to this variable. Although this theorem is particularly convenient for considerations in the perturbation theory, its domain of validity should not be overestimated. In order to apply this theorem, we have to describe the particles only by plane waves and to describe the electromagnetic field, not by the field strengths, but by its sources. Otherwise, we need further averaging or chaos hypothesis regarding other variables of minus class. On the other hand, it is also not a general rule that each time the particles have an "internal" freedom, averaging or chaos hypothesis is needed regarding this freedom.

Boltzmann already noticed that even if the detailed balance does not hold, i.e., if the system in state S_i does not return to S_i after a double transition, $S_i \rightarrow S_j \rightarrow S_i$, there will be a chain of transitions, $S_i \rightarrow S_j \rightarrow S_k \rightarrow \cdots \rightarrow S_i$, by which the system will come back to the original state. This "cyclic balance," or "closed cycle of corresponding collisions" as Tolman[6] calls it, can be considered as a generalization of detailed balance. Although lacking proof, Stückelberg[2] pointed out that the mechanism of his H-theorem is connected with cyclic balance.[7] It is obvious that cyclic balance is a manifestation of the ergodic nature of physical phenomena.

In the last section of this paper, we shall give a simplified version of the ergodic theorem, using only the hypothesis of bilateral normalization of transition probabilities. This will provide a general (though schematized) basis for the theorem of cyclic balance, without referring to collision processes.

This simplified ergodic theorem cannot be directly applied to the actual physical problems, because of various simplifying conditions which will be explained at an appropriate place. For instance, the number of states is assumed to be finite, which is not permissible for the applications in classical physics. We also ignore the important notion of "macroscopic cell" on the energy shell. But, on the other hand, our ergodic theorem has the advantages not only of being very simple and mathematically rigorous, but also of exhibiting all the essential assumptions necessary for the deduction of the ergodic nature of transition probabilities. It will be shown that the simplifying conditions being admitted, the bilateral normalization is the necessary and sufficient condition for the "ergodicity." The term ergodicity will be defined in accordance with the general ergodic theorem in physics. If reversibility or inversibility is taken as the foundation of the bilateral normalization, we can attribute ergodicity to the reversibility or inversibility of the physical laws. As in other versions of the H-theorem and the ergodic theorem in quantum physics, here also, the noncommutability of the Hamiltonian with the operators defining the states plays an essential role.

It is intended in Part II to examine reversibility, reflectibility (symmetry in space) and inversibility of quantum field theory,[8] and to discuss their bearings on the interaction types and other allied problems. In this paper, these symmetry properties are formally defined and, assumed to exist. The classification of physical quantities into four "kinds" is explained in a fashion which may seem unduly elaborate. But this will prove to be instrumental not only for the discussion of the principle of semidetailed balance but also for the discussion of reversibility, reflectibility, and inversibility in general.

2. "KINDS" AND "CLASSES" OF TENSORS

In this section, we shall give the mathematical definitions of the four "kinds" and two "classes" of tensors. In the next section, we shall first introduce a formal method to determine the kinds of tensors representing various physical quantities, and then clarify the physical implication of this determination.

We consider the entire group of congruent transformations of coordinates,

$$x'^{\mu}=a^{\mu}{}_{\cdot\nu}x^{\nu}, \quad (u,\ \nu=1,\ 2,\ 3,\ 0) \tag{2.1}$$

which leave invariant

$$x_{\mu}x^{\mu}=g_{\mu\nu}x^{\mu}x^{\nu}=(x^1)^2+(x^2)^2+(x^3)^2-(x^0)^2. \tag{2.2}$$

It is essential in the investigations involving inversions to use only the real coordinates, lest the connectivity of the Minkowski space may be altered.

The tensors $t^{\mu\nu\cdots}$ of the "regular" kind are defined by the transformation rule,

$$t'^{\mu\nu\cdots}=a^{\mu}{}_{\cdot\kappa}a^{\nu}{}_{\cdot\lambda}\cdots t^{\kappa\lambda\cdots}, \text{ (regular)} \tag{2.3}$$

[4] S. Watanabe, Sci. Pap. Inst. Phys. Chem. Res. (Tokyo) **39**, 157 (1941); S. Watanabe, Phys. Rev. **84**, 1008 (1951).

[5] W. Heitler, reference 1; W. Heitler, lecture notes, École d'Été de Physique Théorique, 1952; F. Coester, Phys. Rev. **84**, 1259 (1951).

[6] R. C. Tolman, *The Principles of Statistical Mechanics* (Oxford University Press, London, 1938), p. 114.

[7] The author is indebted to Professor W. Pauli, who in a private communication, emphasized the importance of cyclic balance.

[8] As far as reversibility is concerned, the subject is fairly fully covered in the second paper quoted under reference 4. See also the earlier works on this subject: E. P. Wigner, Göttinger Nachr. 546 (1932); S. Watanabe, *Le Deuxième Théorème de la Thermodynamique et la Mécanique Ondulatoire* (Hermann et Cie, Paris, 1935); S. Watanabe, Sci. Pap. Inst. Phys. Chem. Research (Tokyo) **31**, 109 (1937).

TABLE I. The kind to which a product of two tensors belongs.

	Reg.	1	2	3
Reg.	Reg.	1	2	3
1	1	Reg.	3	2
2	2	3	Reg.	1
3	3	2	1	Reg

which we write for simplicity as

$$t'=Tt. \qquad \text{(regular)} \quad (2.4)$$

The pseudotensors $t^{\mu\nu\cdots}$ of the first, second, and third kinds are defined, respectively, by[4]

$$t'=\sigma Tt, \qquad \text{(1st kind)} \quad (2.5)$$

$$t'=\sigma_t Tt, \qquad \text{(2nd kind)} \quad (2.6)$$

and

$$t'=\sigma_s Tt, \qquad \text{(3rd kind)} \quad (2.7)$$

where

$$\sigma=\sigma_t\sigma_s=\frac{\partial(x'^1,x'^2,x'^3,x'^0)}{\partial(x^1,x^2,x^3,x^0)}, \qquad (2.8)$$

$$\sigma_t=\sigma_s\sigma=\frac{\partial x'^0}{\partial x^0}\bigg/\left|\frac{\partial x'^0}{\partial x^0}\right|, \qquad (2.9)$$

$$\sigma_s=\sigma\sigma_t=\frac{\partial(x'^1,x'^2,x'^3)}{\partial(x^1,x^2,x^3)}\bigg/\left|\frac{\partial(x'^1,x'^2,x'^3)}{\partial(x^1,x^2,x^3)}\right|. \qquad (2.10)$$

The usual definition of ordinary tensors includes regular and second kinds, and that of pseudotensors includes first and third kinds, since only the "orthochronous" transformations are considered in their definition.

It is obvious from the foregoing definition that the kind to which a product (with or without contraction) of two tensors belongs is determined by the following rules: (a) The product of two tensors of the same kind is a regular tensor. (b) The product of a regular tensor and a pseudotensor of a given kind belongs to the last-named kind. (c) The product of two pseudotensors of different kinds is a pseudotensor whose kind is different from either one of the two factors (see Table I). These rules are symmetrical regarding three tensors involved in the operation of multiplication.

The antisymmetric tensor[9] $r^{\mu\nu\cdots}$ of the rank n which is "complementary" to an antisymmetric tensor[9] $t^{\mu\nu\cdots}$ of the rank $(4-n)$ is defined by

$$r=(1/24)\epsilon_{\mu\nu\kappa\lambda}t^{\mu\nu\kappa\lambda}, \qquad (2.11)$$

$$r_\mu=(1/6)\epsilon_{\mu\nu\kappa\lambda}t^{\nu\kappa\lambda}, \qquad (2.12)$$

$$r_{\mu\nu}=(1/2)\epsilon_{\mu\nu\kappa\lambda}t^{\kappa\lambda}, \qquad (2.13)$$

$$r_{\mu\nu\kappa}=\epsilon_{\mu\nu\kappa\lambda}t^{\lambda}, \qquad (2.14)$$

$$r_{\mu\nu\kappa\lambda}=\epsilon_{\mu\nu\kappa\lambda}t, \qquad (2.15)$$

where the tensor ϵ is completely antisymmetric, and its component $\epsilon_{\mu\nu\kappa\lambda}(=-\epsilon^{\mu\nu\kappa\lambda})$ is $+1$ or -1 in *any* coordinate system, accordingly as (μ,ν,κ,λ) is an even or odd permutation of (1,2,3,0). To satisfy this definition, ϵ must be a pseudotensor of the fourth rank of the first kind.

This complementary relation is reciprocal with regard to r and t.[10] The kinds of r, t, and ϵ (1st kind) are related by the product rule. Thus, the second line of Table I will give the kind of r as dependent on the kind of t. The scalar defined by Eq. (2.11), as complementary to ϵ itself, is a regular scalar and has the value -1 in any coordinate system. This $\epsilon_{\mu\nu\kappa\lambda}$ should not be confused with a regular tensor $\eta_{\mu\nu\kappa\lambda}$ which is $+1$ or -1 accordingly as (μ,ν,κ,λ) is an even or odd permutation of (1,2,3,0) in a particular coordinate system. Such a tensor changes the signs of its components by a transformation in which $\sigma=-1$. The complementary scalar to η is then a pseudoscalar of the first kind.

If a physical quantity is to be expressed as a tensor component, we have to determine (a) the rank of the tensor, (b) the component which represents it, and (c) the kind of the tensor. We assume in this work that (a) and (b) are already determined by the Lorentz transformations in the narrow sense, except for the ambiguity due to the possible complementary representation.

To determine the kind of a tensor, it is sufficient to examine its behavior for time-reversal (hereinafter reversion),

$$x^1,\ x^2,\ x^3 \to x^1,\ x^2,\ x^3 \qquad x^0\to -x^0, \qquad (2.16)$$

and its behavior for total space-reflection (hereinafter mirage),

$$x^1,\ x^2,\ x^3 \to -x^1,\ -x^2,\ -x^3 \qquad x^0\to x^0. \qquad (2.17)$$

For reversion or mirage, a component Q of any tensor will retain or change its sign, but its absolute value remains unchanged. We write for reversion

$$Q'=\rho_R Q, \quad \rho_R=+1 \quad \text{or} \quad -1, \qquad (2.18)$$

and for mirage

$$Q'=\rho_M Q, \quad \rho_M=+1 \quad \text{or} \quad -1. \qquad (2.19)$$

The four possible combinations of the values of ρ_R and ρ_M will lead the classification into four kinds.

TABLE II. The sign of ρ_R for various components as dependent on the kind.

Rank	Component	Reg.	1st	2nd	3rd
Scalar		+	−	−	+
Vector	Space	+	−	−	+
	Time	−	+	+	−
Tensor, 2nd rank	Space-space				
	Time-time	+	−	−	+
	Space-time	−	+	+	−

[9] The modifier "antisymmetric" applies only to ranks higher than one.

[10] A little caution must be taken regarding the sign. For instance, $r^{12}=t^{30}$ but $t^{12}=-r^{30}$, also $r=t^{1230}$ but $t=-r^{1230}$ according to the above definition.

According to the definition of the four kinds of tensors, ρ_R and ρ_M are directly determined by the kind of the tensor and by the nature of the component in consideration. Tables II and III list ρ_R and ρ_M for various components up to the second rank. In the designation of the nature of components in these tables, "space" means $\mu=1, 2, 3$ and "time" means $\mu=0$.

For the combination of reversion and mirage (hereinafter total inversion or inversion),

$$x^1, x^2, x^3, x^0 \to -x^1, -x^2, -x^3, -x^0, \tag{2.20}$$

we have

$$Q'=\rho_I Q, \tag{2.21}$$

with

$$\rho_I=\rho_R\rho_M. \tag{2.22}$$

For total inversion, the transformation matrix T of Eqs. (2.4)–(2.7) becomes simply

$$T=+I \quad \text{or} \quad -I, \tag{2.23}$$

as the rank of the tensor is even or odd. In Eq. (2.23), I means the identity matrix. Here the σ's are

$$\sigma=+1, \quad \sigma_t=-1, \quad \sigma_s=-1. \tag{2.24}$$

Definitions (2.4)–(2.7) show that the coefficient ρ_I is then simply the product of T in Eq. (2.23) and one of the σ's in Eq. (2.24). Thus we obtain a simple rule for ρ_I which depends only on the rank and the kind of the tensor: ρ_I is positive for regular tensors and first kind pseudotensors of even ranks and for second and third kind of pseudotensors of odd ranks. For regular tensors and first kind pseudotensors of odd ranks and for second and third kind pseudotensors of even ranks, ρ_I is negative. All the quantities of the former group will form the "plus class," and all the quantities of the latter group the "minus class." See Table IV.

3. DETERMINATION OF THE KINDS OF PHYSICAL QUANTITIES

It is a basic assumption of this entire work that any physical quantity can be represented as a component of a tensor of a certain kind. We are now going to introduce a set of formal prescriptions by which the kinds of various physical quantities can be determined according to the definition of each quantity. Admittedly, "definitions" of physical quantities and "physical laws" involving those quantities are hardly separable in many cases. As a result, one may raise an objection to the "proof" of reversibility, etc., to the effect that the kinds of the physical quantities are determined in such a way that the reversibility, etc. may hold automatically. The point is, however, that the same physical quantities appear in various physical laws, and that it is meaningful to verify that there is no internal contradictions among these laws. We shall use as elementary a definition as possible to determine each physical quantity's kind. The basic rules serving this purpose are as follows:

(a) The attributes of elementary particles, i.e., rest-mass, electric charge, mesic charge, and magnitude of spin, are regular scalars.
(b) The proper-time differential ds is a pseudoscalar of the second kind.

TABLE III. The sign of ρ_M for various components as dependent on the kind.

Rank	Component	Reg.	1st	2nd	3rd
Scalar		+	−	+	−
Vector	Space	−	+	−	+
	Time	+	−	+	−
Tensor, 2nd rank	Space-space				
	Time-time	+	−	+	−
	Space-time	−	+	−	+

TABLE IV. The class of a tensor determined by its rank and kind.

Rank	Reg.	1st	2nd	3rd
Even	+	+	−	−
Odd	−	−	+	+

Rule (b) means that the sign of ds is determined by the sign of the time differential dt. In addition to these rules, we notice that the transformations (2.16) and (2.17) do not change the sign of the operation

$$\int\!\!\int\!\!\int_{-\infty}^{\infty} dx^1dx^2dx^3.$$

This means that, as far as the signs of ρ_R and ρ_M are concerned, a physical quantity and its density behave in the same manner.

To begin with, we note that four-velocity dx^μ/ds is a second kind pseudovector, since dx^μ is a regular vector while ds is a second kind pseudoscalar. By the regular invariance of intrinsic mass, the momentum of a particle mdx^μ/ds then becomes a second kind pseudovector. This warrants the positive-definite definition of energy, since $\rho_R=\rho_M=1$ for the time component of a second kind vector. Force md^2x^μ/ds^2 must be a regular vector, for dx^μ is a regular vector and ds stands here squared. Since the total charge of an elementary particle is a regular scalar the current-density vector should be a second kind pseudovector, the only kind of vector whose time components (charge-density in this case) has $\rho_R=\rho_M=1$. The orbital angular momentum, being the product of a position-vector (regular) and a momentum-vector (second kind), should be represented by the space-space components of a second kind pseudotensor. From Tables II and III, we see that the space-space components of a second kind pseudotensor have the same values of ρ_R and ρ_M as the space components of a first kind pseudovector. This suggests that the spin-density of a particle, if expressed as a vector, should belong to the first kind. The magnetic moment density can be pictured as the product of charge (regular) and angular momentum (second kind); hence, it must be repre-

TABLE V. Classification of various quantities into kinds and classes.

Rank	Kind	Class	Quantities
Scalar	Reg.	+	Intrinsic mass, electric charge, mesic charge, magnitude of spin, Lagrangian
	1st	+	Magnetic pole strength
Vector	Reg.	−	Position-time, force
	1st	−	Spin density
	2nd	+	Four-velocity, momentum-energy, current-charge density, electromagnetic potentials, linear polarization of photon
Tensor	Reg.	+	Energy-momentum density
	2nd	−	Angular momentum (space-space), electric field (space-time), magnetic field (space-space), electric moment density (space-time), magnetic moment density (space-space), circular polarization of photon (space-space)
	3rd	−	All the quantities listed under 2nd class pseudotensor become 3rd class pseudotensors by the interchange of space-space components and space-time components

sented as the space-space components of a second kind pseudotensor. If it is represented as space-time components of a tensor, this tensor must belong to the third kind in virtue of the theorem of complementary tensors. By the definition of the electric field as the force on a charge, the ρ_R and ρ_M of the electric field must be the same as those of the space components of a regular vector which is force. From Tables II and III we see that the electric field, if represented as the space-time components of a tensor, must belong to the second kind. If we know, from the Lorentz transformation in the narrower sense, that the electric and magnetic fields build a tensor, the magnetic field should then be represented as space-space components of a second kind tensor. An alternative representation of the electromagnetic field is, in virtue of the theorem of complementary tensors, such that the electric and the magnetic fields are respectively represented as space-space components and space-time components of a third kind pseudotensor. If the magnetic pole strength is defined as the ratio of the force to the magnetic field, it must behave like a first kind scalar, since we have $\rho_R=1$ and $\rho_M=-1$ for force and $\rho_R=-1$ and $\rho_M=1$ for magnetic field.

The kind to which energy-momentum density tensor belongs can be determined by the requirement that its space-time components and time-time component should behave like the momentum-energy vector which is a second kind vector. This classifies the energy-momentum density tensor as a regular tensor. From the relation between the energy-momentum density tensor and the Lagrangian density, it follows that the latter has to behave like the diagonal elements of the former. In other words it is a regular scalar.

The ranks and kinds of various physical quantities being thus determined, their classes immediately follow from the rule which is tabulated in Table IV. All these results together with some results which can easily be inferred are listed in Table V.

The determination of the kinds to which various "internal" variables (i.e., other than position-time and energy-momentum) of a spinor field belong, requires a further discussion in quantum field theory, which will be given in Part II. Only some of the results will be given here. The ρ_R of these quantities have already been determined in a previous paper[11] while the ρ_M of these quantities can be shown to be the same as in the c-number theory as determined by the transformation properties of spinors. Table VI lists the classification of these quantities, assuming that the two spinors appearing in each expression belong to the same transformation rule for reversion, and mirage.[11]

The kind of the pi-meson field is the same as the kind of its source, since the differentiation operator, if involved, is a regular vector. If the spinors representing the nucleons before and after the emission or absorption of a pi meson are of the same kind, then the above table (under q-number theory) will immediately give the kind of the pi-meson field.[12] An inspection of Table VI will tell that a combination of scalar and vector types of interaction and a combination of pseudovector and pseudotensor types of interaction are not allowed. This "exclusion rule" of combination arises not from mirage but from reversion due to the change of ρ_R in q-number theory.[13] This change of ρ_R is exactly what is required to give to these quantities their respective physical meanings.[11] (See for instance that spin $i\psi^\dagger\gamma_5\gamma_\mu\psi$ becomes, as it should, a first kind vector.)

We now proceed to introduce the notions of "reversed state," "miraged state," and "inverted state." The determination of kinds of the physical quantities given above is based essentially on a comparison of the two descriptions of the same physical phenomenon referring to two different coordinate systems related to

TABLE VI. The kinds and classes of various tensorial quantities built with two spinors of the same kind. The results for two spinors of different kinds can easily be inferred from this table (see reference 11).

	c-number theory Kind	c-number theory Class	q-number theory Kind	q-number theory Class
$\psi^\dagger\psi$	2nd	−	reg.	+
$i\psi^\dagger\gamma_\mu\psi$	2nd	+	2nd	+
$i\psi^\dagger\gamma_\mu\gamma_\nu\psi$	2nd	−	2nd	−
$i\psi^\dagger\gamma_5\psi$	3rd	−	1st	+
$i\psi^\dagger\gamma_5\gamma_\mu\psi$	3rd	+	1st	−
$\psi^\dagger\gamma_5\gamma_\mu\gamma_\nu\psi$	3rd	−	3rd	−

[11] The second paper quoted under reference 4. These topics will be discussed in detail in Part II, Secs. 9 and 10.

[12] This is true for neutral pi mesons. The situation is more complicated for charged pi mesons. See Part II.

[13] Thus, this is a direct consequence of the results of the second paper quoted under footnote 4. See also G. Lüders, Z. Physik **133**, 325 (1952). For a more general discussion, see Sec. 10, Part II of the present paper. We shall see that the above "exclusion rules" hold also in cases where the two spinors involved belong to different transformation laws.

each other by Eq. (2.16) or Eq. (2.17). The physical insight into the meanings of "kinds" can be obtained more easily by an alternative interpretation of the transformation (2.16) or (2.17), namely by considering two phenomena connected by this transformation described by the *same* coordinate system.

Two phenomena are said to be reversed phenomena of each other if, by suitably choosing the coordinate origin, all the space coordinates involved in one phenomenon at any instant $t=x^0$ become the same as those involved in the other phenomenon at $-t$. It is hereby understood that the corresponding coordinates refer to the same physical entities, e.g., the particles of the same attributes. The two states of the physical system, one referring to a phenomenon at t, the other referring to its reversed phenomenon at $-t$, are said to be reversed states of each other.

From this definition it follows that in two mutually reversed states, the same particles have the same positions but the opposite velocities. Thus the current in the reversed state should have the opposite sign, resulting in the opposite sign of magnetic field, etc. The rest of the argument then follows the same pattern as in the preceding determination of kinds of tensors. We can confirm, in this manner, that the invariance or change of the signs of the physical quantities in the reversed state is exactly the same as ρ_R determined in the foregoing. The reversed state S_R of a state S can now be redefined as a state in which all the physical quantities with $\rho_R=1$ have the same values as in S and all the quantities with $\rho_R=-1$ have the same absolute values but with the opposite signs.

Two phenomena are said to be miraged phenomena of each other if, by a suitable choice of the coordinate origin, all the space coordinates involved in one phenomenon at t are the negative of all the coordinates involved in the other phenomenon at the same instant t; these coordinates are supposed to refer to the same physical entities. The states of the physical system in such two phenomena at the same instant are said to be the miraged states of each other.

Comparing the consequences of this definition of the miraged states with the preceding determination of the kinds of tensors, we can redefine the miraged state S_M of a state S as a state in which all the physical quantities with $\rho_M=1$ have the same values as in S and all the physical quantities with $\rho_M=-1$ have the same absolute values but with the opposite signs.

Two phenomena are said to be totally inverted phenomena of each other if, with a suitable choice of the space-time origin, all the space-coordinates involved in one phenomenon at t are the negative of the corresponding space-coordinates involved in the other phenomenon at $-t$. The two states compared here are totally inverted states of each other. The totally inverted state S_I of a state S can be defined as a state in which all the physical quantities with $\rho_I=1$ have the same values as in S and all the physical quantities with $\rho_I=-1$ have the same absolute values but with the opposite signs.

4. REVERSIBILITY, REFLECTIBILITY AND INVERSIBILITY

Every closed system of physical laws must include a time-dependent law from which it is possible to deduce predictive statements. Such a theoretical system should be capable of answering questions of the following type: What is the probability $P(S\rightarrow S'; t)$ of finding a physical system in state S' at the end of a period of time t if the system was found in state S at the beginning of this period? Such a probability will be simply called transition probability from S to S'.

If the description of the system by states S and S' is maximal, i.e., as detailed as allowed in principle, the prediction may be called a microscopic or dynamical prediction, while in other cases it is only statistical.[14] If the transition probability refers to a "statistical" prediction, we shall use the symbol W instead of P. We shall deal only with the dynamical probability P in this section. In classical physics, P is either 1 or 0, while in quantum physics we have only

$$0\leq P\leq 1.$$

In classical physics, a state is maximally defined if the values of all the independent physical quantities are furnished. In quantum physics, a maximally defined state is a "pure state," or a quantum state, in contrast to a "mixture" (Gibbs ensemble or density matrix). Such a pure state may be considered as an eigenstate of a set of mutually commuting operators, representing a group of physical quantities, although in some cases these operators may be quite complicated.

Covariance for reversion, or reversibility, means that a process and its reversed process have the same probability, i.e., the transition probability from S to S' during t is equal to the transition probability from the reversed state S_R' of S' to the reversed state S_R of S during time t. Symbolically:

$$P(S\rightarrow S'; t)=P(S_R'\rightarrow S_R; t). \tag{4.1}$$

If a state S is characterized by the values Q of physical quantities, S_R is characterized by $\rho_R Q$. We write for brevity

$$S=\{Q\},\quad S_R=\{\rho_R Q\},\quad S'=\{Q'\},\quad S_R'=\{\rho_R Q'\}. \tag{4.2}$$

Covariance for mirage, or reflectibility, then means, in a similar symbolism,

$$P(S\rightarrow S'; t)=P(S_M\rightarrow S_M'; t) \tag{4.3}$$

[14] According to this usage of words, the ordinary transition probability in quantum physics from one quantum state to another should be qualified as microscopical or dynamical and not statistical. The "statistical" transition probability in quantum physics then refers to a transition of a system known to be in a Hilbert subspace to another Hilbert subspace, where the dimensions of subspaces are more than one.

with

$$S=\{Q\},\quad S_M=\{\rho_M Q\},\quad S'=\{Q'\},\quad S_M'=\{\rho_M Q'\}. \tag{4.4}$$

Finally, covariance for total inversion, or inversibility, means

$$P(S\rightarrow S';t)=P(S_I'\rightarrow S_I;t) \tag{4.5}$$

with

$$S=\{Q\},\quad S_I=\{\rho_I Q\},\quad S'=\{Q'\},\quad S_I'=\{\rho_I Q'\}. \tag{4.6}$$

There is a simple theorem which follows directly from this definition, because of (2.22).

Theorem: If a physical system obeying a certain set of physical laws has any two of the three kinds of covariance, reversibility, reflectibility, and inversibility, then it also has the third one.

For instance, suppose that a system enjoys reversibility and inversibility. First, by reversibility, we have

$$P(\{Q\}\rightarrow\{Q'\};t)=P(\{\rho_R Q'\}\rightarrow\{\rho_R Q\};t), \tag{4.7}$$

and second, by inversibility,

$$P(\{\rho_R Q'\}\rightarrow\{\rho_R Q\};t)=P(\{\rho_I\rho_R Q\}\rightarrow\{\rho_I\rho_R Q'\};t). \tag{4.8}$$

Combining these two we obtain Eq. (4.3), for

$$\rho_M=\rho_I\rho_R. \tag{4.9}$$

It should be noted, however, that it is quite possible that a physical system possesses only one of the three kinds of covariance. Physical systems which obey reflectibility but not reversibility (hence not inversibility) are familiar to us. It may also very well happen that a physical system does not obey Eqs. (4.1) and (4.3) with the right signs of ρ_R and ρ_M but does obey them with wrong signs of ρ_R and ρ_M for some of the quantities, leading however to the right signs of $\rho_I=\rho_R\rho_M$. This statement is true for any permutation of ρ_R, ρ_M, and ρ_I.

In classical physics, all the physical laws are written in terms of tensorial components. Therefore, if the physical quantities appearing in each equation belong to the same kind, reversibility and reflectibility (hence also inversibility) are automatically guaranteed, since the existence of a solution representing a process will imply the existence of another solution representing the reversed or miraged process. As this situation is well known, we shall limit ourselves to some remarks of general nature.

The mechanical laws are covariant for both reversion and mirage as far as the force md^2x^μ/ds^2 is equated to a regular vector. The space components should then have $\rho_R=+1$, $\rho_M=-1$. The frictional force, e.g., $-k(\mathbf{v}/|\mathbf{v}|)\mathbf{v}^2$ (Newtonian), has $\rho_M=-1$ but $\rho_R=-1$; thus, it satisfies reflectibility but not reversibility. However, the Lorentz force $e(\mathbf{E}+[\mathbf{v}\times\mathbf{H}])$ has the right signs. The Maxwellian equations are covariant for both reversion and mirage, since the electromagnetic field tensor and its source, the current vector, belong both to the second kind. (See Table V.) In contrast to this, Ohm's law, $\mathbf{I}=\sigma\mathbf{E}$ is not covariant for reversion; the left side has $\rho_R=-1$, $\rho_M=-1$ while the right side has $\rho_R=+1$, $\rho_M=-1$ (see Tables II, III, and V).

In quantum physics, the physical laws are not written in tensorial expressions. Therefore we have to examine whether we can construct the whole theory in such a way that the expectation values (including eigenvalues) of all the relevant physical quantities behave for reversion and mirage as their respective kinds will dictate. The kinds of the physical quantities which have classical analogs can easily be determined by classical physics. The purely quantum mechanical quantities such as spin, magnetic moment, etc., can also be determined by their relations to known classical quantities. For instance, from the conservation law of total angular momentum, we have to assume the same ρ_R and ρ_M for spin as for orbital angular momentum. This is what has been done in Sec. 3. The question as to whether quantum physics in its entirety can be formulated in a covariant way for reversion and mirage will be studied in Part II.

In the remainder of this paper it is assumed that the physical laws governing the physical systems under consideration obey reversibility, reflectibility, and inversibility. This assumption may be considered to be warranted as far as atomistic laws are concerned. In particular, it is understood that these covariance properties exist independently of how the states S and S' are defined, in so far as they are maximally defined.

5. DETAILED BALANCE, SEMIDETAILED BALANCE AND AVERAGED BALANCE

The theorem of detailed balance, literally taken, would mean

$$P(S\rightarrow S';t)=P(S'\rightarrow S;t). \tag{5.1}$$

This type of theorem holds only in the first-order perturbation theorem in quantum theory and, of course, is not of a general validity.

The so-called theorem of detailed balance in classical physics by no means claims Eq. (5.1), which is a dynamical or microscopic law, but it does represent a statistical law in which the state of a system is characterized by a distribution function in velocities (or momenta). It may be written, to exploit the distinction between P and W, as

$$W(S\rightarrow S';t)=W(S'\rightarrow S;t), \tag{5.2}$$

where S is defined only by a certain distribution in velocities. This theorem is based essentially on the interesting fact that, as far as the linear momentum and energy of particles are concerned, a state and its totally inverted state are identical, since $\rho_I=1$ for all the four components of momentum-energy vectors which are pseudovectors of the second kind. (See Table IV.)

This indicates that the basic fact underlying the theorem of detailed balance is the theorem of total inversibility. This theorem which can be written as

$$P(S\rightarrow S';t)=P(S_I'\rightarrow S_I;t), \tag{5.3}$$

means that the (dynamical) transition probability from state S to state S' is equal to the (dynamical) transition probability from the inverted state S_I' of S' to the inverted state S_I of S, whereby the inverted state is constructed from the original state by keeping the values of all the quantities of the "plus" class and changing the signs of all the quantities of the "minus" class (see Table V). It should be noted that since the position is a minus quantity, comparison must be made between a quantity at $\mathbf{x}$ in S and the same quantity at $-\mathbf{x}$ in S_I. For instance the electric field (a minus quantity) at $\mathbf{x}$ in S must be equal in magnitude and opposite to the electric field at $-\mathbf{x}$ in S_I. This is in agreement with the situation created by the source point of this electric field placed at $\mathbf{x}$ in S and the same source point placed at $-\mathbf{x}$ in S_I.

Let us now disregard the electromagnetic field strengths and the notion "force." Then, if we consider only particles without spin, the only difference between a state and its inverted state lies in position coordinates and angular momenta (see Table V). This means that coordinates should be miraged; the change in sign of angular momentum ensues automatically. This mirage of coordinates ($\mathbf{x}\rightarrow-\mathbf{x}$) involves not only the mirage of positions of particles but also the mirage of shapes of particles and mirage of the boundary.

It is now clear that, if the molecules are spherical (or points) and spinless and the boundary is symmetrical with regard to mirage, then the classical theorem of detailed balance in Eq. (5.2) can be deduced, without discussion of collision processes, simply by the assumption that the distribution function is independent of position. This shows that the classical theorem of detailed balance is based on the "chaos" hypothesis in regard to the positions of molecules. (Chaos in angular momenta is a result of the chaos in positions.)

If the molecules are not spherical and/or the boundary is not mirage-invariant, the chaos hypothesis regarding positions does not guarantee Eq. (5.2), since we still have miraged molecules and the miraged boundary in the inverted state. If the particles have at least one plane of symmetry, the miraged shapes of the particles can be reached by some rotations. If the boundary has at least one plane of symmetry, the miraged boundary can be considered as a rotated position of the same boundary. In this case, we can further introduce a chaos hypothesis in regard to the orientations of the molecules. This will secure a type of relation in Eq. (5.2), but the right hand side will still refer to the rotated position of the boundary. Only if the boundary has a symmetry with regard to mirage, can we have the classical theorem of detailed balance on assumption of two kinds of chaos, one regarding positions, the other regarding the orientations of molecules. This situation explains why the usual illustrations of the breakdown of the theorem of detailed balance concerns either nonspherical molecules[15] or some irregular boundary.[16] The first categories of cases of breakdown can be remedied by the assumption of chaos with regard to the orientations of molecules if they have at least one plane of symmetry. It is possible that this hypothesis should be sufficient even if the molecules do not have a plane of symmetry, but in such a case we shall have to discuss the collision process in more detail. In case of a boundary which does not have a mirage-invariant shape we had better resort to a long-range balance, or cyclic balance, on which we shall touch in the last section.

The above derivation of the classical theorem of detailed balance suggests an immediate generalization. We can always establish a statistical balance of type in Eq. (5.2), by assuming as many chaos hypotheses as we need quantities of minus class to describe the state.[17] This general rule will be called the "principle of averaged balance."

In classical physics as well as in quantum physics, we can limit the number of necessary minus quantities to a certain degree by suitably choosing the employed variables. For instance, we can use the electromagnetic potentials (plus quantities) instead of electromagnetic field strengths (minus quantities); we can even describe the electromagnetic field by its source alone. In more elementary examples, "force" can be replaced by "potential." [18]

In quantum physics, there is further a simplifying situation due to the existence of noncommuting quantities. If we characterize particles by their linear momenta (plus), we must disregard their positions (minus) and angular momenta (minus). Hence, for the quantum mechanics of spinless particles, we have the detailed balance in the strictest sense (5.1) if we adopt the plane wave description; the total Hamiltonian is assumed to satisfy inversibility.

In case of the particles with spin, we can only derive the statistical balance of Eq. (5.2) by one chaos hypothesis regarding spin. This can be done by describing the electromagnetic field by its sources and describing these charged particles by linear momenta. From Table V, we see that the remaining physical quantities of the minus class are only spin and electromagnetic moments. Therefore, averaging over spin directions will yield the averaged balance of Eq. (5.2). This is what Heitler[5] calls the principle of semidetailed

[15] R. C. Tolman's textbook (quoted in reference 6), p. 119. Heitler's lecture note (quoted in reference 5). J. M. Blatt and V. F. Weiskopf, *Theoretical Nuclear Physics* (John Wiley and Sons, Inc., New York, 1952), p. 530.

[16] E. H. Kennard, *Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1938), p. 57.

[17] We should keep in mind that in some cases the chaos hypothesis regarding one variable automatically entails the chaos hypothesis regarding another variable. For instance, the chaos regarding position will result in chaos regarding angular momentum. The same is true for spin and magnetic moment.

[18] Thus, this situation is taken care of by the positional chaos; "force" changes its sign, but "attraction" and "repulsion" are invariant notions.

balance and it is a special case of our principle of averaged balance.

The theorem of semidetailed balance is obviously very convenient when applied to the usual perturbation theory in which plane waves are taken as unperturbed states, i.e., eigenstates of the noninteracting Hamiltonian. Moreover, the spin $i\psi^\dagger\gamma_5\gamma_a\psi$ ($a=1, 2, 3$) in the direction of propagation of the plane wave and the magnetic moment $i\psi^\dagger\gamma_a\gamma_b\psi$($a,b=1, 2, 3$) in a direction perpendicular to this direction of propagation commute with the noninteracting Hamiltonian in the Dirac theory as well as with the momentum operator. The characterization of a state by the momentum and one of these "internal" variables is therefore suitable for discussions in the perturbation theory. But this is only one of the possible modes of description. For instance, the total angular momentum, which is a minus variable, is also a constant of motion of the noninteracting Hamiltonian. It should also be noted that the inverted state of a diverging wave is a converging wave, therefore it is also outside the scope of this theorem.

There is no reason to limit the initial and final states to the eigenstates of the noninteracting Hamiltonian. If, for instance, the probability of existence of particles is more or less localized (wave packet), position variables (minus) will intervene in the description of state, thus we shall again need a hypothesis of chaos regarding these "external" variables. On the other hand, we could avoid internal variables of the minus group by using, for instance, the spin-orbit interaction energy instead of spin (or magnetic moment) itself. If we consider two charged particles in interaction, the magnetic moment is reversed in the inverted state; however, the magnetic field strength, due to the other particle, also changes its sign in the inverted state at the point, so that the magnetic interaction will remain unchanged. Since this description involves more or less localized particles, we shall have to use a chaos hypothesis for "external" variables. We are faced with a kind of complementarity; to avoid one kind of chaos, we have to introduce another kind of chaos. It should also be recalled that there are also internal variables of the plus group (see Table VI).

We thus see that we have to be very cautious in application of Heitler-Coester's principle of semidetailed balance. A general rule, which does not fail, is that we should first determine the classes (plus or minus) of all the variables used in the description of the system, and, if we want to use a theorem of the type of Eq. (5.2), we should assume chaos hypotheses for all the minus class variables (principle of averaged balance). We sometimes encounter, in the existing literature, statements to the effect that every time particles have an internal freedom, we have to perform an averaging with regard to this freedom. But, the coincidence of external and internal variables with plus and minus variables in the case of Heitler-Coester's theorem is only accidental.

We have used the words averaging and chaos in the above exposition without precise definitions. Now we should like to provide such definitions. We take a mixture M_1 of state S and its inverted state S_I with equal weights. Then the average probability (average over S and S_I) of transition of a system in this mixture to S' *or* S_I' is given by

$$W_{12}=\tfrac{1}{2}\{P(S\rightarrow S')+P(S\rightarrow S_I') +P(S_I\rightarrow S')+P(S_I\rightarrow S_I')\}. \quad (5.4)$$

Due to inversibility in Eq. (5.3), this probability is equal to the average transition probability of a system in a mixture M_2 of S' and S_I' with equal weights to S *or* S_I.

$$W_{21}=\tfrac{1}{2}\{P(S'\rightarrow S)+P(S'\rightarrow S_I) +P(S_I'\rightarrow S)+P(S_I'-S_I)\}=W_{12} \quad (5.5)$$

is the exact meaning of Eq. (5.2).

It should not be understood that this general consideration provides any justification for the chaos hypothesis (equal weight of a state and its inverted state). In fact, if a system in mixture M_1 had the same average transition probability to S' *and* to S_I', i.e., if

$$\tfrac{1}{2}\{P(S\rightarrow S')+P(S_I\rightarrow S')\} =\tfrac{1}{2}\{P(S\rightarrow S_I')+P(S_I\rightarrow S_I')\} \quad (5.6)$$

were true, then a "chaos" would remain a "chaos" after transition. By no means is Eq. (5.6) guaranteed by inversibility. One general way of justifying the chaos hypothesis is to take a mixture of *all* the possible states with equal weight; then we can expect that a system in this general mixture has an equal average transition probability to S' and to S_I'. In this case, the equal weight of a state and its inverted state holds as well before as after transition. This property implies that $\sum_i P_{ij}=\sum_i P_{ik}$ where S_j and S_k are inverted states of each other. This condition is guaranteed by the inverse normalization which is always true whenever there is inversibility or reversibility, as we shall see in the next section.

As far as the theorem of averaged balance is concerned, averaging is supposed to be made over each pair of states S and S_I, or $\{Q\}$ and $\{\rho_I Q\}$. But, in case of space coordinates, pairing of $\mathbf{x}$ and $-\mathbf{x}$ with regard to a particular coordinate system does not have an invariant meaning for translation; averaging all the values of coordinates is usually required.

Our derivation of the theorem of averaged balance using inversibility shows that we can also introduce a second theorem of averaged balance by considering reversed states instead of inverted states. Thus, we can obtain a type of relation, as Eq. (5.2), by averaging, or introducing chaos hypotheses, with regard to all the variables involved which have $\rho_R=-1$ (instead of $\rho_I=-1$).

6. BILATERAL NORMALIZATION OF TRANSITION PROBABILITIES

First we shall consider the condition of bilateral normalization of microscopic transition probabilities P, and then consider the same condition in regard to statistical transition probabilities W.

For a given maximally defined state S, we think of a series of maximally defined states S_i ($i=1, 2, 3, \cdots$) such that S is one of them, say, S_i. Let us take the physical system in any condition, and represent the probability of finding it in state S_i by p_i. If

$$\sum_i p_i=1 \tag{6.1}$$

we speak of a "complete" set of states. While in quantum physics there are more than one such set, in classical physics there is only one complete set. In classical physics, the number of possible values of i are usually continuously infinite, and even multi-dimensional. In these variables (i.e., in the phase space) p_i will be a kind of δ function. Thus Eq. (6.1) should be understood as a schematic simplification of the situation.

Take two such complete sets (which may be the same or different) of states S_i and S_j and consider the transition probability:

$$P(S_i \rightarrow S_j; t). \tag{6.2}$$

The definition of P (Sec. 4) results, in virtue of (6.1), in the normalization regarding the final states:

$$\sum_j P(S_i \rightarrow S_j; t)=0. \tag{6.3}$$

However, the normalization with regard to the initial states:

$$\sum_i P(S_i \rightarrow S_j; t)=1 \tag{6.4}$$

is not guaranteed by the definition. We shall see that if the physical system obeys reversibility or inversibility, inverse normalization, Eq. (6.4), follows from the first normalization, Eq. (6.3). In the discussion which immediately follows, we shall speak only of reversibility, but the word "inversibility" can always be substituted for the word "reversibility."

The basis of the demonstration is the fact that if S belongs to a complete set S_i, its reversed state S_R also belongs to set S_i. Because of reversibility, if $S \rightarrow S'$ is a solution of the dynamical law, $S_R' \rightarrow S_R$ is also a solution, implying that if S is a possible state, S_R is also a possible state. In classical physics, there is only one complete set of states, therefore, this means that if S is a member of the set, S_R is also its member.

In quantum physics, a pure state $S=\{Q\}$ can be considered as an eigenstate of a family of Hermitian operators which, though complicated at times, represent some physical quantities belonging to one or other of the four kinds. Therefore, each of them has a definite sign of ρ_R. If the reversed state is a possible state (which is the case here), this state must also be an eigenstate of this family of physical quantities, for it is characterized by $\{\rho_R Q\}$ of the same physical quantities. Hence, S_R belongs to the same complete set as S. We can thus conclude that a complete set of states is composed of self-reversed states and pairs of mutually reversed states.

This being the case, the summation over all the S_i's and the summation over all the S_{iR}'s must mean the same operation. First, by reversibility,

$$\sum_i P(S_i \rightarrow S_j; t)=\sum_i P(S_{jR} \rightarrow S_{iR}; t) \tag{6.5}$$

which, due to the above remarks, equals

$$\sum_i P(S_{jR} \rightarrow S_i; t), \tag{6.6}$$

which is on account of the first normalization condition (6.3) equal to unity. Hence Eq. (6.4), *Q.E.D.*

It is true that the physical meaning underlying the unitarity of transformation matrices in quantum physics is connected with the "completeness" of representation. But here we have derived the bilateral normalization without utilizing specifically quantum-mechanical relations.

We now pass to the bilateral normalization with regard to the statistical transition probabilities W. It is usually the case that states S defined only statistically exhibit also "completeness." In other words, we can consider any nonmaximally defined state S as a member of a series of nonmaximally defined states S_i ($i=1, 2, 3, \cdots$) such that the probability w_i of finding the system in a state S_i of the series obey a normalization condition:

$$\sum_i w_i=1. \tag{6.7}$$

For example, in classical physics, after averaging over all space-coordinates, the state of each molecule of a gas is characterized only by velocities $\mathbf{v}$. Then the velocity space can be divided into small volume elements, which certainly have the property of completeness in the sense of Eq. (6.7). We can also apply this consideration to a pair of molecules, as is usually done in the discussion of collision processes. In case of Heitler-Coester's mode of description, after the averaging over the spin directions, the possible values of the momenta will constitute a complete set.

Now, if such a set of nonmaximally defined states is so chosen that the theorem of averaged balance in Eq. (5.2) is true, then the inverse normalization,

$$\sum_i W(S_i \rightarrow S_j; t)=1, \tag{6.8}$$

follows immediately from the first normalization,

$$\sum_j W(S_i \rightarrow S_j; t)=1, \tag{6.9}$$

which is a consequence of Eq. (6.7). The reversibility relation, Eq. (6.5), or the corresponding inversibility relation, cannot be adapted in the same form to the W's for arbitrarily chosen sets of nonmaximally defined states.

It should be recalled that even in classical physics, the w's and W's are not limited to zero and unity, which is the case for the p's and P's in this form of physics.

7. ERGODIC PROPERTY OF TRANSITION PROBABILITIES

It is well known that the Markoff chain exhibits a particular property which may be called "ergodic." The usual exposition of this subject is too mathematical in nature and often overly simplified by the assumption of detailed balance: $P_{ij}=P_{ji}$.

This section intends to point out that the bilateral normalization of transition probabilities is just necessary and sufficient to derive the "ergodicity" of the Markoff chain; the section also hopes to clarify in what sense we can speak of an ergodic theorem. In the following, we shall discuss the subject in terms of the microscopic, or dynamical probabilities P, but we shall soon find that the main body of argument also applies to the statistical probabilities W.

We take a complete set $\mathfrak{S}$ of maximally defined states S_i, and the indices i, j, etc., of the S's are supposed always to refer to this same set. We limit ourselves to the cases where there are a finite number of states in the set:

$$i=1, 2, 3, \cdots, r. \tag{7.1}$$

In classical physics, there are usually a continuously infinite number of the S's. In this case, the entire argument that follows offers only a mathematical model which may approximate the real physical situation. In quantum physics, Eq. (7.1) does not imply a real limitation, since we need actually consider only a limited region of energy values (microcanonical shell) and we can also assume the space domain to be limited. Then the number of quantum states will become finite.

Among these states S_i ($i=1, 2, \cdots, r$), some will be disconnected from one another due to various conservation laws. For instance, two states belonging to different values of the total angular momentum will allow no transitions from one to the other. Thus, the entire set of S_i will be divided into subsets, in each of which the states are "connected." Hereinafter such a subset will be called a "subshell." A more rigorous definition of subshells will be given soon.

The theory of Markoff chains pertains to the "repeated" transition probability $P_{ij}^{(n)}$ which is defined by

$$P_{ij}^{(n)}=\sum_k P_{ik}^{(n-1)}P_{kj}=\sum_k P_{ik}P_{kj}^{(n-1)} \tag{7.2}$$

where

$$P_{ij}^{(1)}=P_{ij}=P(S_i\rightarrow S_j;\,\tau). \tag{7.3}$$

In classical physics, we have $P_{ij}=1$ or 0, therefore also $P_{ij}^{(n)}=1$ or 0. In quantum physics, $0\leq P_{ij}^{(n)}\leq 1$. In both cases, we have

$$\sum_j P_{ij}^{(n)}=1, \tag{7.4}$$

which follows from Eq. (7.2) in virtue of the first normalization:

$$\sum_j P_{ij}=1. \tag{7.5}$$

In the same manner, the inverse normalization:

$$\sum_i P_{ij}=1 \tag{7.6}$$

will result in

$$\sum_i P_{ij}^{(n)}=1. \tag{7.7}$$

The classical physics is characterized by the fact that

$$P_{ij}^{(n)}=P(S_i\rightarrow S_j;\,n\tau), \tag{7.8}$$

which means that the physical system is not disturbed by observation. In quantum physics this is not the case, in general, unless the operators defining S_i commute with the exact Hamiltonian. In quantum physics, the repeated transition probability in Eq. (7.2) acquires a physical meaning only on assumption that the system is observed every τ seconds with the operators defining S_i. In other words, starting with a pure state S_i, we observe the system after τ seconds, and the result is statistically represented by a mixture (ensemble or density matrix) composed of various S_j with the weight P_{ij}. By repeating this process at each interval of τ seconds, we obtain after $n\tau$ seconds a mixture of S_j's with the respective weights $P_{ij}^{(n)}$. This means that although we start with the microscopic transition probabilities P_{ij}, we have to interpret the repeated transition probabilities $P_{ij}^{(n)}$ in quantum physics in terms of "mixtures." The ergodic theorem discussed in this section thus refers to a chain of repeated observations and should not be confused with the more important ergodic theorem[19] which refers to two observations, one at the initial instant and the other at the final instant.

Closely related to $P_{ij}^{(n)}$, and physically and mathematically more significant than these are the quantities:

$$\Omega_{ij}^{(n)}=\sum_{m=1}^{n} P_{ij}^{(m)}/n. \tag{7.9}$$

They are physically important since $\Omega_{ij}^{(n)}$ represents the "time average" of transition from S_i to S_j during the time $n\tau$ seconds, while $P_{ij}^{(n)}$ represents the transition probability at the instant $n\tau$ seconds after the initial instant. Indeed, the main concept in an ergodic theorem, in physics, is a comparison of "average in time" with "average in microcanonical ensemble." Furthermore they are mathematically useful since $\Omega^{(n)}$

[19] J. von Neumann, Z. Physik **57**, 30 (1929); W. Pauli and M. Fierz, Z. Physik **106**, 572 (1937). For an interpretation of Neumann's ergodic theorem in terms of initial and final observations, see an article by S. Watanabe in the monograph, *Louis de Broglie* (Albin Michel, Paris, 1952), p. 385.

has a better convergence than $P^{(n)}$ for $n\to\infty$. Although $P^{(n)}$ is zero or unity in classical physics, $\Omega^{(n)}$ is not necessarily so: $0\leq\Omega_{ij}^{(n)}\leq 1$.

From the first normalization, Eq. (7.5), follows:

$$\sum_j \Omega_{ij}^{(n)}=1, \tag{7.10}$$

and from the inverse normalization, Eq. (7.6),

$$\sum_i \Omega_{ij}^{(n)}=1. \tag{7.11}$$

Now the "ergodic theorem" which we are going to prove can be enunciated as follows: The time average $\Omega_{ij}^{(n)}(n\to\infty)$ of the probability of finding a system, which started from any S_i, in a state S_j of the same subshell is equal to the *a priori* probability of a state in the subshell, i.e., equal to $1/s$ if s is the number of states in the subshell. This statement is certainly a faithful adaptation of the general ergodic theorem to our simplified case, since the microcanonical ensemble represents a mixture of all the states on an energy shell with equal weights. The main purpose of this section is to show that the inverse normalization is the necessary and sufficient condition for this simplified ergodic theorem, the first normalization being always assumed for the Markoff chain.

First we shall enumerate, without proofs, some of the elementary theorems and definitions regarding the Markoff chain which can be found in any exposition of the subject.[20] We shall denote by $\mathfrak{S}$ the original complete set of states in Eq. (7.1). Only the first normalization is assumed in the following theorems.

Theorem I. The sequence

$$\Omega_{ij}^{(1)},\ \Omega_{ij}^{(2)},\ \cdots,\quad (S_i,\ S_j\epsilon\mathfrak{S}) \tag{7.12}$$

converges to a limit:

$$\lim_{n\to\infty}\Omega_{ij}^{(n)}=\Omega_{ij}^{\infty}. \tag{7.13}$$

Of course, we have

$$0\leq\Omega_{ij}^{\infty}\leq 1,\quad \sum_j \Omega_{ij}^{\infty}=1. \tag{7.14}$$

In the set $\mathfrak{S}$ there can be some states S_j such that the average transition probabilities Ω_{ij}^{∞} to them vanish for any arbitrary initial state S_i.

Definition I. The "vanishing" part $\mathfrak{V}$ of $\mathfrak{S}$ is the set of all states S_j such that

$$\Omega_{ij}^{\infty}=0,\quad (S_i\epsilon\mathfrak{S},\ S_j\epsilon\mathfrak{V}). \tag{7.15}$$

For the rest of the original set: $\mathfrak{S}'=\mathfrak{S}-\mathfrak{V}$ (which can easily be shown not to be empty), we have the following theorem:

[20] Theorems I–IV are given in Husimi's textbook (Husimi, reference 3, p. 280), but their physical applications in physics, including Theorems V and VI, are not given there. Husimi's exposition is based on K. Yosida and S. Kakutani, *Topological Mathematics* (Iwanami, Tokyo, 1939, in Japanese), Vol. 2, p. 20. See also, W. Feller, *Introduction to Probability Theory and its Application* (John Wiley, New York, 1950), p. 307 ff.

Theorem II. If $\Omega_{ij}^{\infty}>0$, then

$$\Omega_{ji}^{\infty}>0,\quad (S_i,\ S_j\epsilon\mathfrak{S}'). \tag{7.16}$$

In other words, if $\Omega_{ij}^{\infty}=0$, then

$$\Omega_{ji}^{\infty}=0,\quad (S_i,\ S_j\epsilon\mathfrak{S}'). \tag{7.17}$$

Using Theorem II, we can divide $\mathfrak{S}'$ into subsets ("subshells") such that Ω_{ij}^{∞} is zero if S_i and S_j belong to different subshells, and $Q_{ij}^{\infty}>0$ for S_i and S_j belonging to the same subshell.

Definition II.

$$\mathfrak{S}'=\mathfrak{E}_1+\mathfrak{E}_2+\cdots+\mathfrak{E}_p, \tag{7.18}$$

$$\Omega_{ij}^{\infty}=0,\quad (S_i\epsilon\mathfrak{E},\ S_j\epsilon\mathfrak{E}'), \tag{7.19}$$

$$\Omega_{ij}^{\infty}>0,\quad (S_i,S_j\epsilon\mathfrak{E}). \tag{7.20}$$

A subshell $\mathfrak{E}$ is disconnected from the vanishing part $\mathfrak{V}$ and from another subshell $\mathfrak{E}'$ not only in terms of Ω_{ij}^{∞} [see Eqs. (7.15) and (7.19)] but also in terms of P_{ij}.

Theorem III.

$$\left.\begin{array}{ll} P_{ij}=0, & (S_i\epsilon\mathfrak{E},\ S_j\epsilon\mathfrak{V}) \\ P_{ij}=0, & (S_i\epsilon\mathfrak{E},\ S_j\epsilon\mathfrak{E}') \\ P_{ij}=0, & (S_i\epsilon\mathfrak{E}',\ S_j\epsilon\mathfrak{E}) \end{array}\right\}. \tag{7.21}$$

Obviously the inverse of this theorem is not true. It can happen that $P_{ij}=0$ even for S_i and S_j belonging to the same subshell, i.e., in spite of $\Omega_{ij}^{\infty}>0$.

We now pass to study the properties of those Ω_{ij}^{∞} whose initial and final states belong to the same subshell $\mathfrak{E}$ consisting of s states S_i:

$$i=1,\ 2,\ 3,\ \cdots,\ s. \tag{7.22}$$

In virtue of Eq. (7.21), we can derive from Eq. (7.4)

$$\sum_{j=1}^{s} P_{ij}=1,\quad (S_i,\ S_j\epsilon\mathfrak{E}). \tag{7.23}$$

Similarly, because of (7.15) and (7.19), we have

$$\sum_{j=1}^{s} \Omega_{ij}^{\infty}=1,\quad (S_i,\ S_j\epsilon\mathfrak{E}). \tag{7.24}$$

The relations (7.23) and (7.24) show that the first normalization Eqs. (7.4), (7.14), remains unchanged when the initial and final states are limited to a subshell.

We are now prepared to introduce an important theorem.

Theorem IV. Ω_{ij}^{∞} $(S_i,\ S_j\epsilon\mathfrak{E})$ is independent of the initial state S_i:

$$\Omega_{ij}^{\infty}=\Omega_j,\quad (S_i,\ S_j\epsilon\mathfrak{E}). \tag{7.25}$$

Of course, we have, due to Eq. (7.24),

$$\sum_{j=1}^{s}\Omega_j=1. \tag{7.26}$$

The discussion up to this point assumes only the first normalization. We now investigate the implication of the inverse normalization. We shall consider the condition,

$$\sum_{i=1}^{s} P_{ij}=1, \quad (S_i, S_j \epsilon \mathfrak{E}), \tag{7.27}$$

which exhibits a symmetry to Eq. (7.23). Because of Eq. (7.21), the summation with regard to S_k in

$$P_{ij}^{(n)}=\sum_k P_{ik}P_{kj}^{(n-1)}, \quad (S_i, S_j \epsilon \mathfrak{E}; S_k \epsilon \mathfrak{S}),$$

actually extends only over $S_k \epsilon \mathfrak{E}$. Hence

$$\sum_i P_{ij}^{(n)}=1, \quad (S_i, S_j \epsilon \mathfrak{E}), \tag{7.28}$$

and, by Eqs. (7.9) and (7.13), also

$$\sum_i \Omega_{ij}^{(n)}=\sum_i \Omega_{ij}^{\infty}=1, \quad (S_i, S_j \epsilon \mathfrak{E}). \tag{7.29}$$

We now propose to show that the inverse normalization as in Eq. (7.27) is equivalent to the condition that Ω_{ij}^{∞} is not only independent of the initial state S_i (Theorem IV) but also independent of the final state S_j. This last condition can be written, in view of (7.26), as

$$\Omega_{ij}^{\infty}=1/s, \quad (S_i, S_j \epsilon \mathfrak{E}). \tag{7.30}$$

Theorem V. The necessary and sufficient condition for Eq. (7.30) is Eq. (7.27).

Proof: From the definition of Ω_{ij}^{∞} in Eqs. (7.9) and (7.13), we can easily obtain

$$\Omega_{ij}^{\infty}=\sum_k \Omega_{ik}^{\infty}P_{kj}, \quad (S_i, S_j \epsilon \mathfrak{E}; S_k \epsilon \mathfrak{S}). \tag{7.31}$$

Because of Eqs. (7.15) and (7.19), the summation over S_k, in reality, extends only over $S_k \epsilon \mathfrak{E}$. If Eq. (7.30) is the case, Eq. (7.31) becomes

$$\frac{1}{s}=\frac{1}{s}\sum_{k=1}^{s} P_{kj}, \quad (S_k, S_j \epsilon \mathfrak{E}) \tag{7.32}$$

showing that Eq. (7.27) is a necessary condition for Eq. (7.30).

Next we shall show that Eq. (7.27) is also a sufficient condition.

If Eq. (7.27) is true, then we have Eq. (7.29), which in view of Eq. (7.25) means

$$\sum_{i=1}^{s} \Omega_{ij}^{\infty}=s\Omega_j=1, \quad (S_i, S_j \epsilon \mathfrak{E}) \tag{7.33}$$

or

$$\Omega_j=\frac{1}{s} \qquad Q.E.D. \tag{7.34}$$

Theorem V is obviously equivalent to the ergodic theorem we stated at the beginning of this section.

It should be noted that our inverse normalization in Eq. (7.27) is not necessarily equivalent to the inverse normalization with regard to the entire set:

$$\sum_i P_{ij}=1, \quad (S_i, S_j \epsilon \mathfrak{S}). \tag{7.35}$$

It is, however, easy to see that if Eq. (7.35) is true then Eq. (7.27) is also true, and that if Eq. (7.27) is true and if the entire set $\mathfrak{S}$ has no vanishing part $\mathfrak{V}$, then Eq. (7.35) is true. Actually, in Eq. (7.27), S_i and S_j can be extended, without any additional assumption, to all the states belonging to $\mathfrak{S}'=\mathfrak{S}-\mathfrak{V}$:

$$\sum_i P_{ij}=1, \quad (S_i, S_j \epsilon \mathfrak{S}'), \tag{7.36}$$

on account of Eq. (7.21). The summation in Eq. (7.36) extends to all the states in $\mathfrak{S}'$. Equation (7.36) is equivalent to Eq. (7.27).

Now, if Eq. (7.35) is true, we shall have Eq. (7.11), with $S_i, S_j \epsilon \mathfrak{S}$. But this contradicts the existence of a vanishing part in Eq. (7.15). Hence, if Eq. (7.35) is true, then $\mathfrak{S}=\mathfrak{S}'$, and Eqs. (7.36) and (7.27) ensue. On the other hand, if we have Eq. (7.36) as a given premise, then the conclusion of Eq. (7.35) can be drawn only with the help of an additional condition $\mathfrak{S}=\mathfrak{S}'$.

From whatever state one may start, ultimately there will be a vanishing probability of having the system in a state belonging to the vanishing part as in Eq. (7.15). Furthermore starting from a state in any one of the $\mathfrak{E}$'s, we have a vanishing single transition probability P_{ij} landing in a state in the vanishing part as in Eq. (7.21). In physical problems, an initial state is, after all, the final state of another chain of observation. We may justifiably exclude states of the vanishing part also as initial states. In any event, symmetry of the physical laws in time (reversibility or inversibility) results in the inverse normalization, Eq. (7.35), which implies nonexistence of the vanishing part.

The ergodic theorem is sometimes expressed as a statement regarding the eventual return to the initial state. For actual physical problems in classical physics, a rigorous return to the initial state is not to be expected, but the return to a state infinitely close to the original state (the so-called quasi-ergodic theorem) is sufficient. However, in our simplified theory, a rigorous return to the initial state can be concluded in classical physics.

Theorem VI. If the values of P_{ij} are limited to zero and unity, then there exists a value of n such that

$$P_{ii}^{(n)}=1, \tag{7.37}$$

except for S_i belonging to the vanishing part.

Proof: Taking $i=j$ in (7.20), we have

$$\Omega_{ii}^{\infty}>0 \tag{7.38}$$

Hence, for large enough values of n, we have

$$\Omega_{ii}^{(n)}>0, \quad (n \geq n_0). \tag{7.39}$$

Comparing Eq. (7.39) with Eq. (7.9), we see that there must be a value of n (indeed there must be an infinite number of such n's) for which

$$P_{ii}^{(n)}>0. \tag{7.40}$$

If P_{ij} is zero or unity as we assume, then $P_{ij}^{(n)}$ is also limited to the values zero and unity. Then Eq. (7.40) means that there is a value of n for which

$$P_{ii}^{(n)}=1. \tag{7.41}$$

Taking the smallest value of such n's, we can further infer, in virtue of Eq. (7.2),

$$1=P_{ii}^{(n)}=P_{ii}^{(2n)}=P_{ii}^{(3n)}=\cdots, \tag{7.42}$$

showing a cyclic return to the initial state.[21] This represents the fundamental fact upon which Boltzmann's theorem, which Tolman calls "cycle of corresponding collisions," is based.[6] Our proof of the cyclic balance as seen in Eq. (7.42) is more general than Tolman's argument, since he (1) assumes without verification the ergodic nature of physical phenomena, (2) utilizes, throughout the chaos, hypothesis regarding the position of molecules, and (3) limits his discussion to collision processes. Admittedly, our proof is conditioned by the assumption that the number of possible states is finite.

It is not surprising that the proof of Theorem VI does not utilize the inverse-normalization, since, in classical physics, exclusion of vanishing part immediately results in inverse normalization. Indeed, states belonging to $\mathfrak{S}'=\mathfrak{S}-\mathfrak{V}$ are connected in this case by a one-to-one correspondence.[22]

The theorem of cyclic balance, in Eq. (7.42), can be considered as a generalization of the theorem of detailed balance, which is a special case of Eq. (7.42) for $n=2$. Indeed, from

$$P_{ij}=P_{ji}, \tag{7.43}$$

follows

$$P_{ii}^{(2)}=\sum_k P_{ik}P_{ki}=\sum_k (P_{ik})^2. \tag{7.44}$$

In classical physics, only one of P_{ik} ($k=1, 2, \cdots$) is different from zero and equal to unity. Thence,

$$P_{ii}^{(2)}=1. \tag{7.45}$$

In quantum physics also, we have Eq. (7.40), but it is not of particular interest. Probably another generalization of Eq. (7.43) may be more useful.

In classical as well as quantum physics, we have

$$\Omega_{ij}^{\infty}=\Omega_{ji}^{\infty}, \tag{7.46}$$

which is an obvious consequence of Theorem V. This means that the time average of transition probability from S_i to S_j is equal to the time average of transition probability from S_j to S_i. This is also equal to the time average of probability of return to the original state: Ω_{ii}^{∞} or Ω_{jj}^{∞}. The theorem presented in Eq. (7.46) may be called the theorem of "long-range balance."

In the entire, foregoing discussions, we used chiefly $\Omega_{ij}^{(n)}$ instead of $P_{ij}^{(n)}$, but it is evident that if $P_{ij}^{(n)}$ ($n\to\infty$) has a limit, this limit is the same as Ω_{ij}^{∞}.

It should also be noticed that we can apply all the foregoing discussions to $W(S_i\to S_j;t)$, if Eq. (6.8) holds. Even in classical physics this quantity is not limited to the values zero and unity. Therefore, what has been stated above, with regard to quantum theoretical P's applies, *mutalis mutandis*, to the W's.

For applications of our results to quantum theoretical problems the following remarks should be kept in mind, as should the comment with regard to chains of repeated observations in connection with Eq. (7.2). If the S's are defined by operators which commute with the exact total Hamiltonian, then $P_{ij}=\delta_{ij}$, and the subshell will reduce to one quantum state. In this case, the entire argument loses its physical interest. Therefore, the essential point in the discussion of ergodicity lies in the tacit assumption that the operators defining the states S_i do not commute with the exact, total Hamiltonian. In fact, this assumption is adopted, explicitly or implicitly, in any version of H-theorem or ergodic theorem in quantum physics.[19] In applications to thermodynamics, it is necessary to introduce the idea of macroscopic cells on the macroscopically defined energy shell.[19] Our derivation, which does not make use of this concept, should therefore be considered as a simplified model which serves only to clarify the mathematical gist underlying more elaborate formulations.

The author would like to thank Dr. Cecile M. DeWitt whose instructive seminar talk partly motivated the author to undertake this work.

[21] Taking the smallest common multiple of the n's for various i, (7.37) and (7.42) will become valid for all the i's.

[22] This means that P_{ij} is actually a permutation, and it is obvious that a finite number of repeated permutations results in the identity transformation. The author's thanks are due Professor S. Kakutani for reading this section before publication and for pointing out various interesting facts pertinent to the subject matter, including the point mentioned in this footnote.