

The Conceptual Basis and Use of the Geometric Invariance Principles

R. M. F. HOUTAPPEL

Instituut Lorentz, Leiden, The Netherlands

H. VAN DAM*

Physics Department, University of North Carolina, Chapel Hill, North Carolina

E. P. WIGNER

Palmer Physical Laboratory, Princeton University, Princeton, New Jersey

Invariance principles are used, in physics, in two distinct manners. First, they are used as superlaws of nature in that, once their validity has been suggested by their consistency with the known laws of nature, they serve as guides in our search for as yet unknown laws of nature. Second, they can serve as tools for obtaining properties of the solutions of the equations provided by the laws of nature. It is desirable for the first use to give a formulation of invariances directly in terms of the primitive concepts of physical theory, i.e., in terms of observations, or measurements, and their results. Invariances which can be so formulated are called geometric invariances. The present paper contains an attempt at such a formulation of geometric invariances. This formulation is then applied, in detail, to the classical mechanics of point particles, to a relativistic mechanics of interacting point particles, and to quantum theory. With the exception of the relativistic mechanics of point particles, these applications form a review, from a single point of view, of earlier work on this subject. The last part of the paper contains a review of the second use of invariances.

CONTENTS

Introduction	595	4.9. Equations of Motion, Conservation Laws.....	622
1. Three Remarks on Invariance Principles.....	596	4.10. The More General Invariances of Section 2.4b....	623
1.1. First Remark: Curie's Observation.....	596	4.11. The Problem of Measurement in Relativistic Quantum Theory.....	625
1.2. Second Remark: The Mathematical Form of the Laws of Nature.....	596	5. Invariance as a Tool for Obtaining Properties of the Solutions of Equations.....	627
1.3. Third Remark: The Empirical Origin of the Invariance Principles.....	597	5.1. Reasons for the Increased Importance of Invariance Principles in Quantum Theory.....	628
2. Laws of Nature and Invariance Transformations.....	597	5.2. Matrix Form of Invariance Operators.....	628
2.1. Extended Role of the Laws of Nature.....	597	5.3. States with Zero Spatial Momentum and Definite Energy.....	629
2.2. Observations.....	598	5.4. The Discrete Spectrum.....	630
2.3. The Laws of Nature.....	598	5.5. The Continuous Spectrum.....	631
2.4. Invariance Transformations.....	600		
2.4a. Invariance Transformations which Correspond to Transformations between Equivalent Reference Frames.....	600		
2.4b. More General Invariance Transformations.....	601		
2.5. Dynamic Invariances.....	603		
3. The Role of the Concepts of Section 2 in Causal Theories.....	603		
3.1. Invariance Transformations in the Classical Mechanics of Point Particles.....	603		
3.1a. The Π Function and the Classical Orbit.....	603		
3.1b. States Suspae.....	604		
3.1c. Invariance Transformations.....	604		
3.1d. The Passive View of Invariance Transformations in Classical Mechanics.....	605		
3.1e. The More Flexible Point of View of Invariance Transformations in Classical Mechanics.....	607		
3.2. Relativistic Mechanics of Point Particles.....	607		
3.2a. Observables.....	607		
3.2b. Equations of Motion.....	608		
3.2c. Invariance Transformations, Conservation Laws.....	609		
4. The Role of the Concepts of Section 2 in Quantum Theory.....	610		
4.1. Introduction.....	610		
4.2. Observations and Decision Observations.....	612		
4.3. The Heisenberg Picture.....	612		
4.4. The Π Function of Quantum Theory.....	613		
4.5. Pure States and Trace 1 Projection Operators.....	614		
4.6. Superselection Rules.....	616		
4.7. The Mappings of the Set of Observations on Itself which Leave the Π Function Invariant.....	616		
4.8. Invariance for Proper Inhomogeneous Lorentz Transformations.....	621		

INTRODUCTION

Invariance principles played an important role in Galileo's and Newton's thinking and writings, but their importance for physics was not fully appreciated until the events which led to the formulation of Einstein's theories of relativity.¹ Even then, most physicists thought of invariance principles as something learned that does not enter the everyday thinking or the day-to-day work. The ease with which invariance principles led to concrete results in quantum mechanics changed this situation. The change did not come suddenly—in the early days of quantum mechanics there was much resistance to the adoption of invariance principles as everyday working tools. This resistance disappeared in the course of the years and is not even understood by the new generation of physicists. This is so much more remarkable since, evidently, the use of invariance

¹The investigations of the properties of crystals provide a notable exception to this statement. See, for instance, W. Voigt's *Lehrbuch der Kristallphysik* (B. G. Teubner, Leipzig, 1910). Chapter 6 of M. von Laue's *History of Physics* (Academic Press Inc., New York, 1950) contains a short, but vivid, history of the thinking on the equivalence of coordinate systems. Cf. also Charles Scribner, Jr., *Am. J. Phys.* **32**, 672 (1964).

* Alfred P. Sloan Foundation Fellow.

principles as guides to find the laws of nature becomes obviated once the laws are found—a situation which is considered to be imminent by many.

1. THREE REMARKS ON INVARIANCE PRINCIPLES

The present section deals with invariance principles in a preliminary and somewhat unsystematic fashion to set the stage for the more detailed, and systematic, and we hope deeper study starting in Sec. 2.

1.1. First Remark: Curie's Observation

The first of our remarks which precedes our discussion goes back in its essentials to P. Curie.² If the physical world were really invariant with respect to a coordinate transformation, such as a displacement through a distance L , this fact could never be discovered because there would be no way to distinguish between two points which are truly equivalent in the sense that the situation at one is exactly the same as at the other. If the world were in all its details invariant with respect to a displacement by L , we ourselves would be (to use the language of quantum mechanics) with equal probabilities at points which differ by this displacement. It is indeed possible to assume that there is a fifth coordinate along which all wave functions are constant, but this assumption would have no observable consequences. It follows that the invariance principles, since we want them to have observable consequences (and they have), postulate the equivalence of coordinate systems only in a certain sense. They are possible only because our knowledge of the physical world has been divided into two categories: initial conditions and laws of nature.³ The state of the world is described by the initial conditions. These are complicated and no accurate regularity has been discovered in them.⁴ In a sense, the physicist is not interested in the initial conditions, but leaves their study to the astronomer, geologist, geographer, etc. Circumstances, which are equivalent in the sense of the invariance principles, may be, and usually are, very different from the point of view of initial conditions, that is, actual physical situations. This is what makes "equivalent" points distinguishable. The invariance principles apply only to the second category of our knowledge of nature; to the so-called laws of nature. Only these are the same in equivalent coordinate systems. The laws of nature describe the further fate of a system, once the initial conditions are given. Hence, an invariance principle holds if two systems with the

same initial conditions in two equivalent coordinate systems develop, from the point of view of the respective coordinate systems, in the same way. We see that invariance principles can be formulated only if one admits the existence of two types of information which correspond in present-day physics to initial conditions and laws of nature. It would be very difficult to find a meaning for invariance principles if the two categories of our knowledge of the physical world could no longer be sharply separated.

1.2. Second Remark: The Mathematical Form of the Laws of Nature

The simplest way to verify an invariance principle would be to create the same initial conditions in two equivalent coordinate systems and to observe whether the further fate of the two systems, from the point of view of the coordinate systems in question, is the same. This cannot be done precisely but it can be done with a sufficient approximation because there are systems of limited extension which are not influenced to any noticeable extent by the state of the world outside them.⁵ The state of the world outside these limited systems suffices to distinguish them but leaves their behavior unchanged.

The observation of the behavior of limited systems which are identical from the point of view of two different coordinate systems is the most obvious way to verify the equivalence of two coordinate systems.⁶ It will be called the first method since it is not the only one. The reason is that the "laws of nature" appear in mathematical language of great generality. As a result, the laws of nature obtained in one coordinate system purport to give not only the behavior of systems which are actually in existence from the point of view of that coordinate system but of an infinity of other systems. Hence the invariance principles can be verified by ascertaining whether actual systems in the second coordinate system behave in accordance with the laws established on the basis of experiments in the first coordinate system. This will be called the second method of the verification of invariance principles. To mention a rather insignificant example: we measure, in our laboratories, the electric field only around a very few charged bodies which are at rest. We infer, nevertheless, that Coulomb's law is valid for all charges at rest, even those which do not exist in our own coordinate system. If we observed that, in a moving coordinate system, the field around a charge at rest in that coordinate system is not given by Coulomb's law, we

² P. Curie, *Oeuvres* (Gauthier-Villars, Paris, 1908), p. 127

³ These two concepts will be replaced, in Sec. 2, by more primitive ones.

⁴ It has indeed been argued (Les Prix Nobel, Stockholm, 1964) that they have—or have had—some inherently random character. Similarly, it has been argued at this place that a sharp separation between initial conditions and laws of nature is possible.

⁵ This point is further elaborated in the article "Invariance in Physical Theory," Proc. Am. Phil. Soc. **93**, 52 (1949). In its fundamentals, the observation goes back to C. S. Pierce. See, for instance *Essays in the Philosophy of Science* (Liberal Arts Press, New York, 1957), p. 237.

⁶ The reader will note that the equivalence of two coordinate systems is considered to be synonymous with the invariance with respect to the transformation which connects them. See Sec. 2.4a for a more detailed discussion.

would declare the two coordinate systems not to be equivalent. We would do this even if there were no charge at rest of the magnitude in question in our own coordinate system: our belief in the possibility of formulating the laws of nature in simple mathematical language is so strong that we unhesitatingly rely on it when judging the equivalence of coordinate systems.⁷

The second method of the verification of principles of invariance is much more common than the first one. It is, obviously, very inconvenient, and often impossible, to create identical conditions in different coordinate systems and the accurate repetition of an experiment is the closest the experimenter ever comes in practice to a verification of an invariance principle by the first method. The repetition of an experiment, if the outcome is the same, verifies the equivalence of two coordinate systems obtained from each other by time displacement.

1.3. Third Remark: The Empirical Origin of the Invariance Principle

The third remark is of a much less subtle nature than the other ones, and it is barely necessary to make it now. The discovery of Lee, Yang, and Wu, showing, among other facts, that the laws of nature are not invariant with respect to charge conjugation,⁸ reminded us of the empirical origin of the laws of invariance in a forcible manner. Before the discoveries of Lee, Yang, and Wu, one could quote Fourier's principle⁹ as an earlier example of an invariance principle which had to be abandoned because of empirical evidence. According to Fourier's principle, the properties of matter should not change no matter how far it is subdivided. With the discovery of the atomic structure this principle had to be discarded.

2. LAWS OF NATURE AND INVARIANCE TRANSFORMATIONS

2.1. Extended Role of the Laws of Nature

In the preceding section our knowledge about the physical world is divided into two categories: the initial conditions and the laws of nature. The initial conditions interest the physicist principally because the laws of

nature do not lead to observable consequences unless the initial conditions are given. In order to test a law of nature, the physicist usually first prepares a system with known initial conditions to which the proposed law of nature is applicable. He then proceeds to ascertain whether the properties of his system change in the way postulated by the law of nature to be tested. Thus the picture of sharply separated acquisition of initial conditions, and subsequent conclusions about the resulting behavior of a system is particularly appropriate for describing physical experimentation which is indeed so designed that this picture be appropriate.

One may want to be able to use one's knowledge about the laws of nature in a more ambitious way, as a help to understand currently what is going on around him. If such a use is attempted, one is forced to abandon the stereotyped picture of first ascertaining the "initial conditions," then drawing consequences from them on the basis of the laws of nature. This picture can be used only for carefully selected, relatively simple systems; in general it is not possible to ascertain all the relevant initial conditions. As a matter of fact, it is part of the skill and ingenuity of the experimenter to prepare systems for which the relevant initial conditions are known. If we do not focus attention on such special carefully selected systems, but consider all our surroundings, the situation is that information which we could not have anticipated reaches us constantly, and such information is essentially added knowledge to what physics calls initial conditions. Similarly, the conclusions which we derive from the laws of nature concerning the behavior of the world around us are tested only partially, but on a continuing basis. Acquisition of unforeseeable information and the drawing of inferences on the basis of the laws of nature are only rarely clearly separated; usually the two processes go on hand in hand.

We wish to avoid using the concept of initial conditions as a primitive concept in our discussion. We are motivated partly by the fact that the acquisition of knowledge about the state of the world on a space-like surface appears artificial from the point of view of the special theory of relativity but, even more strongly, by the conviction that the concepts of physics should increasingly approach the realities of our acquisition of information and knowledge. A concept, "state of a system," will be introduced later which corresponds largely to the "initial conditions" of present day physics. However, we prefer to treat this as a derived concept and use as a primitive concept that of observation or apperception. We are well aware of the fact that the process of observation (in conceptual discussions of quantum mechanics often called measurement) is probably most complex if considered from the point of view of the psychologist. However, we hope that the complexities of the process of observation and apperception (we use these two words as synonyms) can be disregarded when the properties of the inanimate

⁷ See also, "The Unreasonable Effectiveness of Mathematics in the Natural Sciences," *Commun. Pure Appl. Math.* **13**, 1 (1960).

⁸ T. D. Lee and C. N. Yang, *Phys. Rev.* **104**, 254 (1956); C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *ibid.* **105**, 1413 (1957). A fuller realization of the empirical nature of the invariance principles would have saved some embarrassment to the community of physicists in connection with the articles of R. T. Cox, C. G. McIlwraith, and B. Kurrelmeyer. See *Proc. Natl. Acad. Sci. (U.S.)* **14**, 544 (1928).

⁹ J. B. Fourier, *Theorie Analytique de la Chaleur* (Fermin Didot, Paris, 1822; English translation: Cambridge University Press, Cambridge, 1878), articles 85-86. It must be admitted, however, that Fourier is less explicit in stating his invariance principle than is often assumed. For an earlier discussion and rejection of "Fourier's principle" see Galileo's *Dialogues Concerning Two New Sciences* (The Macmillan Company, New York, 1914), p. 130.

world around us are considered. At any rate, the process of observation is clearly a more primitive one than that of initial conditions. In terms of the concept of observations, the laws of nature become correlations between these—and this is indeed what they are.

The fact that the acquisition of knowledge concerning initial conditions, and the testing of the laws of nature, go on hand in hand is not in conflict with the assertion, made already in 1.1, that the two are different in character. The laws of nature are regularities, that is correlations between events, whereas the initial conditions are the unforeseeable events. At any rate, they are events (or the results of observations) and not correlations between results of observations.

If we adopt the attitude just described towards the laws of nature, they provide, in general, not all the information about the behavior of some carefully prepared system but some little information about the whole body of our observations. Connections between our observations obtain, for instance, if we observe the subsequent positions of a stone that is being thrown, or the subsequent positions of a celestial body. The number and effectiveness of the connections between observations on a body increases, in general, as the number of observations increases. In the usual language of physics one says that one ascertains the initial conditions more and more completely.

The preceding definition of the laws of nature applies to more situations than usually implied. However, unless we can attribute some role to the laws of nature similar to the one described, they will serve, effectively, only as subjects for their own verification and as means to enable us to design machines. What we wish is that, in addition, they should constantly support our understanding of what goes on around us and what we experience.

Not all connections between observations are consequences of the laws of nature; there are connections which follow from crude regularities in the initial conditions. Thus, if we see the right side of a lamb, we can pretty well guess what the other side looks like. However, we do believe that the two types of regularities can be sharply separated.

We next turn to a somewhat more detailed specification of the concept of observations.

2.2. Observations

The concept of observation plays a basic role in physics. It is a process which consists of bringing the system to be observed in contact or interaction with something such as an instrument, or light, about the state of which we have some *direct knowledge*. This is, probably, a conceptually very difficult notion but we shall nevertheless take it for granted. The process of observation would be described in modern quantum mechanics as a collision between a measuring instrument and the object, a characteristic of which is to be

measured. The result of the observation is some information about this characteristic, such as position at a given time, or quantum state at another time, or the weight of the object. As was mentioned before, the result of some observations may be foreseeable on the basis of earlier observations and some knowledge of the laws of nature; the results of other observations may be unpredictable but they may help then to foresee the results of further observations. Every observation involves some specific action on the part of the observer—one clearly acts differently depending on whether one wishes to observe the velocity or the color of an object. This action together with the time at which it is performed will be characterized by a label α ; the outcome or result of the observation will be represented by $r(\alpha)$ or r_α . This will stand for a set of numbers. In ordinary language an “observation” yields a qualitative, a “measurement,” a numerical result; following quantum-mechanical practice we use these words as synonyms. When we spoke, in the last section, about correlations between observations, we meant the act of observation, together with its results. Naturally, the correlations are correlations of results, but they depend on the action of the observation which determines “the quantity” which we observe. Let us mention a few examples of “observables.” In the classical theory of point particles the observables are the positions and velocities of the particles at given times. Classical electrodynamics adds to these the electric and magnetic field strengths at all space-time points. The observables in elementary quantum mechanics are represented by self-adjoint operators. It should be noted, though, that self-adjoint operators can not be considered *ipso facto* to correspond to observables. It seems likely that only a small fraction of all self-adjoint operators can be “measured.” This fact will have important bearings on our later considerations. The point is that in order to call something an observable, we must know the acts necessary for the observation. These are surely not known for an arbitrary self-adjoint operator. In fact, it is one of the conceptual weaknesses of quantum mechanics that the prescriptions for the acts which lead to the measurement of any operator are not integrated into the theory. This applies, in our opinion, also to the concepts of quantum field theory. However, this weakness of the conceptual structure of quantum mechanics will have, interestingly enough, little effect on our later considerations.

2.3. The Laws of Nature

The general question which one may pose, and the answer to which should be given by the laws of nature, is as follows: consider a set of measurements $\alpha, \beta, \dots, \epsilon$ together with their results $r_\alpha, r_\beta, \dots, r_\epsilon$. We recall that the symbols α, β, \dots specify not only the quantity which was measured, i.e., the action which constitutes the measurement, but also the time thereof. One then asks for the probability that another set of measure-

ments ζ, η, \dots, ν give the results $r_\zeta, r_\eta, \dots, r_\nu$. This probability will be denoted by

$$\Pi(\alpha, r_\alpha; \beta, r_\beta; \dots; \epsilon, r_\epsilon \mid \zeta, r_\zeta; \eta, r_\eta; \dots; \nu, r_\nu). \quad (2.1)$$

Let us observe that the expression (2.1) is clearly symmetric with respect to interchange of the variable pairs before the bar, and with respect to a similar interchange of the variable pairs after the bar. As a rule, one arranges the observations $\alpha, \beta, \dots, \epsilon$ and ζ, η, \dots, ν in a time-ordered series so that the action necessary for carrying out the measurement α precedes all other measurements before the bar, and ν succeeds all other measurements after the bar. Also, the most interesting and most important assertions given by the function Π are those in which all the measurements before the bar precede the measurements after the bar. For values of the arguments of this character the function Π gives the probabilities of the outcomes of measurements which succeed the measurements the results of which it presupposes. However, as has been emphasized by Watanabe,¹⁰ the theories are used not only for such "predictions" but also for "retrodictions" in which one asks what the properties of a system were before any observation was undertaken thereon. It is most convenient for us not to make any assumption on the time ordering of the measurements $\alpha, \dots, \epsilon, \zeta, \dots, \nu$; those after the bar may even be interspersed between those before the "premiss" measurements, i.e., those before the bar.

The Π are expressions for the correlations between results of observations. They constitute the content of the laws of nature, or of a theory. However, as a rule, the laws of nature are not formulated directly by means of the functions Π . In fact it usually requires much mathematical skill to obtain the value of Π even for a relatively simple set of its variables. Π is surely a very complicated function of its variables and we have the belief and conviction that the laws of nature can be given in a simple and aesthetically appealing form. This form is then, also, cryptic and gives Π very indirectly. In fact, one regards as simple and aesthetically appealing those forms of the laws of nature in which the validity of the invariances is established most easily. Surely, the Π function does not satisfy this requirement. The essential assertions of the laws of nature are, however, expressions for the probabilities (which may amount to certainties) denoted by the Π of (2.1). By using the Π function we can give a formulation of invariances, directly in terms of observations. Such a formulation is attractive since we do not yet fully understand all laws of nature and also because those laws of nature that we do understand are formulated, most simply, in terms of notions which are often radically different for different laws.

The probabilities of observations for which a theory

can give a definite value depend, naturally, on the theory. Thus gravitational theory will not permit the calculation of the value of Π if the observations ζ, η, \dots, ν relate to electrically charged bodies. In other words, the definition domain of Π is, as far as the "variables" $\alpha, \beta, \dots, \epsilon, \zeta, \eta, \dots, \nu$ are concerned, different for different theories. The discussion of the conceptual foundations of quantum mechanics taught us also that observations, which are possible by themselves, may be incompatible—this is not surprising if we recall that they represent actions undertaken at definite times. This gives a fundamental restriction on the definition domain of Π as far as the "variables" α, \dots, ν are concerned. If these variables are indeed defined as actions, these restrictions should follow from their defining description. There may be, however, other restrictions on the definition domain of Π in any theory. A very common such restriction is expressed by the statement that without additional statistical information the results of the observations $\alpha, \beta, \dots, \epsilon$ may not suffice to characterize the physical situation sufficiently to give probabilities for the outcomes of ζ, η, \dots, ν . A familiar example of this obtains in classical mechanics if $\alpha, \beta, \dots, \epsilon$ are position measurements, the ζ, η, \dots, ν simultaneous velocity measurements. This example shows particularly clearly how dependent even the definition domain of Π is on the theory the results of which it expresses. The action which classical theory describes as a position or a velocity measurement is surely very different from the action which quantum mechanics would regard as such a measurement.

It may be useful to give an example of a Π function. In the classical theory of point particles the observables α, β, \dots are the coordinates and velocity components of the particles at definite times. The Π function is defined in the case of an isolated system of n point particles if there are, before the bar, $6n$ independent observations. The value of the Π function will be 1 if r_ζ is indeed the value of the coordinate or velocity component ζ which it assumes for the orbit defined by the data before the bar, and if the same is true for r_η with respect to η , etc. Otherwise, Π will be zero.

There is a set of identities which the probability function Π satisfies which must be valid in any theory because they follow from its definition. Thus, if η is performed later than ζ , the probability that the outcome of η be r_η and the outcome of ζ be r_ζ is equal to the product of the probability that the outcome of ζ be r_ζ and of the conditional probability that, given the outcome r_ζ of ζ , the outcome of η be r_η :

$$\begin{aligned} \Pi(\alpha, r_\alpha; \dots; \epsilon, r_\epsilon \mid \zeta, r_\zeta; \eta, r_\eta) &= \Pi(\alpha, r_\alpha; \dots; \epsilon, r_\epsilon \mid \zeta, r_\zeta) \\ &\times \Pi(\alpha, r_\alpha; \dots; \epsilon, r_\epsilon; \zeta, r_\zeta \mid \eta, r_\eta). \end{aligned} \quad (2.2)$$

In a similar way, all Π can be expressed in terms of Π which ask only for the probability of the outcome of a

¹⁰ S. Watanabe, *Rev. Mod. Phys.* **27**, 179 (1955).

single observation. We also have, naturally

$$\sum_r \Pi(\alpha, r_\alpha; \dots; \epsilon, r_\epsilon | \zeta, r) = 1, \quad (2.3)$$

where the summation over r has to be extended over all possible results of the measurement of ζ . This expresses the fact that the measurement of ζ will surely give some result. A similar equation can be postulated for the more general Π of (2.1)—it also follows from (2.2) and (2.3). Naturally, all Π are positive or zero.

Classical theory presupposes, in addition to (2.2) and (2.3), that

$$\begin{aligned} \Pi(\alpha, r_\alpha; \dots; \epsilon, r_\epsilon | \eta, r_\eta) \\ = \sum_{r_\zeta} \Pi(\alpha, r_\alpha; \dots; \epsilon, r_\epsilon | \zeta, r_\zeta; \eta, r_\eta). \end{aligned} \quad (2.4)$$

This expresses the assumption that the measurement ζ does not affect the system and that, hence, η will give any result r_η with the same probability no matter whether the measurement ζ was undertaken or not. Equation (2.4) is not generally valid in quantum theory and presupposes, at any rate, an idealized type of noninterfering measurement.

The term “measurement” needs some further elucidation. Clearly, *it is an integral part of the theory*, and different theories conceive different observations as possible. The theory should provide a prescription in common language how its measurements are to be carried out. This however, is not done, as a rule, and the actions needed to carry out the measurements are given only tacitly and, at least in quantum mechanics, are not very definite. Usually, the same measurement can be carried out in several different ways (e.g., with measuring rods or interferometrically) and the implication is that it is somehow known which of these ways are valid and equivalent. The symbols α, β, \dots stand, therefore, for equivalence classes of measurements.

2.4. Invariance Transformations

The invariance properties of the laws of nature have played a role in the thinking of the earliest physicists even though their significance became fully apparent only early in the present century. It is surprising, therefore, that there is no generally accepted definition of an invariance principle and one can, in fact, distinguish two schools of thought. According to the first opinion, only those principles are generally valid which postulate the equivalence of reference frames which can physically be changed into each other. (A coordinate system can be rotated or set into motion.) The second view does not define the concept of physical invariance so rigidly but accepts all transformations which leave the known laws of nature invariant and the simplicity of which suggests their universal validity. The difference between the two points of view is partly semantic but not entirely so because the adherents of the first point of view believe that only equivalences of reference frames are generally valid principles.

We begin with a few definitions and observations which will introduce the discussion of both points of view. Since there is a definite, though probably infinite, set of observations, it is natural to consider the one-to-one mappings ($\alpha \leftrightarrow \bar{\alpha}, \beta \leftrightarrow \bar{\beta}, \dots$) of this set upon itself. As all one-to-one mappings, these mappings form a group, the unit element of the group being the mapping of every observation on itself. This group, which may be called the group of all transformations, has little interest in itself. However, it does contain a subgroup, the subgroup of all invariance transformations, which is the main subject of the present article. Before proceeding to the definition of this subgroup, we will have to define another subgroup which will serve us as a stepping stone for defining the subgroup of invariance transformations. The subgroup G in question contains all those mappings $\alpha \leftrightarrow \bar{\alpha}, \beta \leftrightarrow \bar{\beta}, \dots$ which leave the probability function invariant:

$$\begin{aligned} \Pi(\alpha, r_\alpha; \beta, r_\beta; \dots; \epsilon, r_\epsilon | \zeta, r_\zeta; \eta, r_\eta; \dots; \nu, r_\nu) \\ = \Pi(\bar{\alpha}, r_\alpha; \bar{\beta}, r_\beta; \dots; \bar{\epsilon}, r_\epsilon | \bar{\zeta}, r_\zeta; \bar{\eta}, r_\eta; \dots; \bar{\nu}, r_\nu) \end{aligned} \quad (2.5)$$

for all sets of observations $\alpha, \beta, \dots, \epsilon, \zeta, \eta, \dots, \nu$.

The knowledge of the full group G would be almost tantamount to the knowledge of the law of nature embodied in Π . Thus, in a causal theory,¹¹ every transformation which changes a consistent set of observations into a consistent set of observations would be a member of the group. Hence, the elements of G would permit the determination of all sets of consistent data from one such set. The situation is similar in noncausal theories. Hence, if we want the invariance postulates to have greater generality than the law of nature embodied in Π and thus to provide us with a structure of physical theories, they can not contain all the transformations of G . The choice then can be made on the basis of one of the principles given at the beginning of this section.

If our knowledge of the laws of nature were complete, the group of invariance transformations might well coincide with the group G . At the same time, however, the group G would have lost the importance which we now attach to invariance transformations. At present, we regard invariance transformations as superlaws which we expect to hold not only for those laws of nature which we have come to understand but also for all others. Our knowledge of the laws of nature and hence of the Π function is limited.

2.4a. Invariance Transformations which Correspond to Transformations between Equivalent Reference Frames

The transformations of (2.5) are “active” in that the system on which the measurements $\alpha, \beta, \dots, \nu$ are made is in general different from the system on which

¹¹ We use causal in the sense of completely deterministic.

the measurements $\bar{\alpha}, \bar{\beta}, \dots, \bar{\nu}$ are undertaken. On the other hand, all the observations which appear in (2.5) are observations of the same observer, using one definite language to characterize his observations; the symbols $\alpha, \beta, \dots, \bar{\alpha}, \bar{\beta}, \dots$ are all symbols in this language. Passive transformations, on the other hand, describe the same object as it appears to different observers; they characterize the same observation by different symbols. Hence, a passive transformation is, essentially, the translation of the language of one observer into that of another observer.

We shall be interested only in the translation of the languages of "equivalent observers" into each other; the relation of such observers can be derived by means of active transformations which satisfy (2.5). For some of the transformations $\alpha \leftrightarrow \bar{\alpha}, \beta \leftrightarrow \bar{\beta}, \dots$ it is possible to imagine a second observer who is in the same relation to the measurements $\bar{\alpha}, \bar{\beta}, \dots$ as is the original observer to the measurements α, β, \dots . This is possible, for instance, if $\bar{\alpha}, \bar{\beta}, \dots$ represent the same actions as α, β, \dots , except that they are carried out a period later. In this case the second observer will differ from the first one only by his clock being late (showing an earlier time) with respect to that of the first observer. However, the relation of the two observers does not have to be quite as simple. It is necessary only that the transformation $\alpha \leftrightarrow \bar{\alpha}, \beta \leftrightarrow \bar{\beta}, \dots$ satisfy (2.5) and that one should feel satisfied that there could be a second observer for whom the measurements $\bar{\alpha}, \bar{\beta}, \dots$ (or measurements which are equivalent to these in the sense described at the end of the Sec. 2.5) play the same role as do α, β, \dots for the first observer. It is this last condition which singles out invariance transformations from all transformations which satisfy (2.5).

If the second observer uses the same symbols α, β, \dots to specify measurements which are in the same relation to him as the first observer uses for similar measurements, the Π function which he will establish will be identical with the Π function of the first observer. This is simply a restatement of (2.5). However, the object for which his measurements α, β, \dots yield the results r_α, r_β, \dots will not be the same as the object for which the first observer's measurements α, β, \dots gave these results; it will only be in the same relation to him. One says that the two objects are subjectively the same for the two observers. On the contrary, the object which yields the results r_α, r_β, \dots for the measurements which the second observer calls α, β, \dots would yield these results for the measurements $\bar{\alpha}, \bar{\beta}, \dots$ of the first observer: for objectively the same system, subjectively different observations of the two observers yield the same result. Note that unless $\alpha \leftrightarrow \bar{\alpha}, \beta \leftrightarrow \bar{\beta}, \dots$ satisfies (2.5) a relation between the two observers as described is not possible. The correspondence between α and $\bar{\alpha}, \beta$ and $\bar{\beta}$ is then the translation we spoke of: the quantity to which the first observer attributes the symbol $\bar{\alpha}$ is called α by the second observer. The translation may be easy to establish, as is

for instance the translation of the observables of the classical mechanics of point particles, i.e., of the coordinates and velocity components of the particles at definite times. It may be quite difficult to find, even if one has an abstract proof of its existence, as it often is in quantum mechanics.

The usual description of a passive transformation starts with different "reference frames." A reference frame, in this context, is an observer equipped with some apparatus or measuring devices. It is assumed that the apparatus enables the observer to carry out all possible measurements. He can therefore perform also the measurements which another observer can carry out, if not precisely in the same, then at least in an equivalent way. Such equivalent measurements were discussed already at the end of Sec. 2.3. These measurements yield, on the same object, the same results with the same probabilities as those of the first observer, but the second observer attributes another name to them. This again leads us to the problem of translating the measurements of the two observers into each other.

Actually, most reference frames which one calls equivalent do not exist in the sense of having an observer and some apparatus associated with it. The equivalence of these "frames of reference" expresses only the conviction that if such apparatus were made, it could be associated with an observer and brought into a condition in which they would form a reference frame of the nature postulated.

We believe, at present, that all equivalent reference frames can be obtained by proper inhomogeneous Lorentz transformations from one of them. *CP* conjugation can hardly be applied to macroscopic objects, not to mention the observer. The same applies to time inversion (reversal of the direction of motion). The time inverse of the state of a simple system, such as a stable elementary particle, can be obtained relatively easily. In fact in most cases it is equally easy to prepare the state and its time inverse. For macroscopic objects, or more generally objects that we cannot prepare in microscopically definite states, this is practically impossible. The reasons for this are discussed in the statistical theory of thermodynamics and we do not wish to consider them here. Time inversion cannot be interpreted as generated by the equivalence of two reference frames since it is even "more impossible" to change physically a reference frame into its time inverse.

2.4b. More General Invariance Transformations

As was mentioned before, our knowledge of the laws of nature and of the Π function is limited. It is for this reason that we require of an invariance transformation not only that it leave the known laws of nature invariant, but also that it be simple so as to give us the conviction of its general validity. We do not believe in the universality of complicated invariance transformations just as we do not believe in the validity of

complicated laws of nature. Let us take as an example an isolated system which consists of n point particles that attract each other according to Newton's law of gravitation. We can easily find mappings of one solution to Newton's equations into other solutions such that (2.5) is satisfied. These mappings are however, in general, complicated and not universally valid invariance transformations. Somehow we must define the subgroup of invariance transformations, which is contained in the group G by applying simplicity requirements. A very logical such requirement is the one which was introduced in Sec. 2.4a. The limitation that it must be possible to interpret an invariance transformation as generated by the equivalence of two reference frames is, however, very severe and seems to restrict the group of invariance transformations to the proper inhomogeneous Lorentz transformations.

The supporters of the second point of view mentioned in the introduction to this section argue that the restrictions imposed in Sec. 2.4a are too severe and must be replaced by more general requirements of simplicity. They point out that one has transformations, such as CPT , CP conjugation, and time inversion which satisfy (2.5) and which furthermore have a convincing simplicity.¹² Such transformations ought to be considered to be invariance transformations although they do not satisfy the limitations discussed in Sec. 2.4a. The group of more general invariance transformations, in the validity of which the supporters of the second school of thought believe, is not sharply defined. This group is contained in the group of all transformations that satisfy (2.5), and it contains as a subgroup the group of invariances defined in Sec. 2.4a. How large this more general group of invariance transformations is depends, however, on our judgment as to which transformations that satisfy (2.5) can be expected to have general validity.

We wish to add several remarks which apply to all invariance transformations discussed in Sec. 2.4. Our first remark is that invariance principles, as a rule, cannot be tested by the experimental verification of (2.5). It is quite unlikely that the apparatus necessary for carrying out the measurements $\bar{\alpha}$, $\bar{\beta}$, \dots actually exists either as apparatus of the second observer (who would call these measurements α , β , \dots) or as apparatus of the first observer. Even if this were the case, there may be no system in existence for which the measurements $\bar{\alpha}$, $\bar{\beta}$, \dots would give the results r_α , r_β , \dots . Hence, as a rule, our belief in the validity of an invariance transformation is derived by its consistency with the correlations between observations provided by theories or laws of nature. Laws of nature give us the conviction that certain correlations between observations would exist if these observations were

made. Hence, the validity of a relation such as (2.5) can be derived on the basis of a law of nature even if it cannot be, or has not been, derived by direct observation. Such a derivation is what is meant by the "consistency" between invariance principle and law of nature. Just as the laws of nature, their invariance principles do not express only actual facts, but also convictions concerning events which actually could be brought about. Because of their greater generality, and for some other not entirely clear reasons, we have stronger confidence in the validity of invariance principles than in that of the laws of nature the consistency with which is their basis. They do seem to have greater permanence than the latter. This together with the fact that invariance transformations must be valid for all laws of nature is the reason for our formulating these principles directly in terms of observations rather than in terms of the derived concepts of any theory.

Our second remark, which has actually been made before, is that we think of measurements which are concrete and operationally given. Hence, for instance, the replacement of all quantum-mechanical observables Q by $U^{-1}QU$ (where U is unitary) is acceptable only as a member of the group G if the measurement of $U^{-1}QU$ is operationally given, at least in terms of the measurement of Q .

We observe next that the possibility of testing the validity of a theory, i.e., the Π function furnished by it, already implies some minimum of invariance. Otherwise, if every set of observations could be carried out only once, it would not be possible to infer that a particular result of an observation is a necessary consequence of having obtained certain definite results for other observations. The necessity of the existence of invariance transformations, at least for certain isolated systems, is even more evident if a probabilistic theory has to be tested because the mere establishment of a probability law requires repeated observations.

The preceding remark amounts to a restatement of the intimate connections between invariances and laws of nature. If there were no invariances, it would not be possible to formulate and verify laws of nature. One finds this remark often in prefaces to elementary textbooks; our only addition to these remarks is that they must be taken seriously.

Our last observation is somewhat less common. We find it remarkable that even though we have no catalog of the possible measurements and of the laws of nature and even though, at the present state of the theory, we cannot decide whether a quantum-mechanical quantity represented by a self-adjoint operator is truly an observable, we have reason to believe that we know the abstract group of invariances. This statement amounts to the claim that we know something about the structure of the laws of nature, and of the set of possible observations, even though we do not know the laws of nature themselves, nor the set of possible observations.

¹² It should be mentioned, though, that recent experiments of Christenson *et al.* (Ref. 34) have at least cast serious doubt on the validity of CP invariance.

2.5. Dynamic Invariances

There exist invariances which seem to play an important role in physics¹³ but which are not invariances in the sense of Sec. 2.4. An example of such an invariance is the gauge invariance of electrodynamics. Gauge transformations do not affect the observations and correspond to the unit element of the group G . Nevertheless one can derive the interaction between a charge-carrying field and the electromagnetic field from gauge invariance and simplicity requirements. A similar remark applies to the invariance for general coordinate transformations and Einstein's equations for the gravitational field interacting, for instance, with a scalar field.¹⁴ We call invariances of the character of the gauge invariance dynamic, those conforming to the principles of Sec. 2.4, geometric. The last name is derived from their analogy to the invariances of the classical geometries as discussed by F. Klein in the *Erlanger Programm*.¹⁵ Incidentally, as was pointed out by E. Cartan,¹⁶ the manifolds of Riemannian geometry do not, in general, allow any invariances in the sense of Klein. The situation in the theory of general relativity is very similar: the invariance of the theory with respect to general coordinate transformations can not be interpreted as an invariance in the sense of Sec. 2.4. This has been pointed out before by E. Cartan,¹⁶ and even more emphatically by V. A. Fock.¹⁷ It seems to us that the invariance with respect to general coordinate transformations must be interpreted as a dynamic invariance. The present article is restricted to a discussion of geometric invariances.

3. THE ROLE OF THE CONCEPTS OF SECTION 2 IN CAUSAL THEORIES¹⁸

The preceding section discusses the laws of nature and their invariance transformations in general and rather abstract terms. The present section inquires how, and to what extent, classical (Newtonian) and relativistic mechanics of point particles together with their invariances conform with the principles which we postulate. These principles characterized laws of

nature as correlations between observations and invariance principles as correlations between these correlations. The invariance principles are formulated, however, directly in terms of observations independently of the terminology of any specific theory. A formulation of invariance principles directly in terms of the primitive concepts is desirable because their main function, which starts only after their validity has been suggested by their consistence with the known laws of nature, is to serve as guides in our search for as yet unknown laws of nature.

Our procedure is first to find an expression for the probability function Π of (2.1) and then to show how this conforms with the form (2.5) of the various invariance principles. Section 4 contains a discussion of invariances in quantum theory.

3.1. Invariance Transformations in the Classical Mechanics of Point Particles

3.1a. The Π Function and the Classical Orbit

The observables of the classical mechanics of point particles are the positions and velocity components of the particles. If we consider the usual limiting case of infinitely accurate observations, the Π function of (2.1) is easily determined. It will be defined if the input observations, i.e., those before the bar, give $6n$ independent data where n is the number of particles. The values of all the positional and velocity components at a given time are the most usual input variables; they determine in classical mechanics an "orbit." By this we mean the values of the positional and velocity coordinates at all times. This can be represented by a line in the $6n + 1$ dimensional space which is spanned by the aforementioned $6n$ quantities and the time coordinate. Through every point of this space there is one and only one "orbit," i.e., line which obeys the equations of motion.

As a function of the output variables, Π is not a function in the mathematical sense but a distribution because it has δ -function character. The probability Π is 1 if integrated over regions of the output variables which all contain the value given by the orbit; the integral is zero otherwise. The Π function would not have δ -function character if one used observations of finite accuracy; its consideration for such observations would raise interesting problems of principle. However, we shall not be concerned with them.

It is easy to verify the relevant Eqs. (2.2), (2.3), and (2.4) for the Π function. In fact these equations, together with the postulate of the possibility to observe the values of continuous variables, suffice to establish the most important properties of Π as they are discussed above. The situation which we encountered, such as that Π is a distribution with δ -function character, is, therefore, common to all causal theories in which the observables can assume a continuum of values.

¹³ The writers are unaware of any comprehensive and up-to-date review of the dynamic symmetry principles. Their relation to the "geometric" principles, with which the present review deals, is considered by E. P. Wigner, *Proc. Natl. Acad. Sci. (U.S.)* **51**, 956 (1964); also *Rend. Scuola Intern. Fis. Enrico Fermi, Course 29* (Academic Press Inc., New York, 1964); and *Les Prix Nobel*, Stockholm, 1964.

¹⁴ For another example of an interaction which is characterized by a dynamic invariance see C. N. Yang and R. L. Mills, *Phys. Rev.* **96**, 191 (1954); see also, R. Utiyama, *ibid.* **101**, 1597 (1956).

¹⁵ F. Klein, *Math. Ann.* **43**, 63 (1893).

¹⁶ E. Cartan, *l'Enseignement Math.* **26**, 200 (1927).

¹⁷ V. Fock, *The Theory of Space Time and Gravitation* (Perгамon Press, Inc., New York, 1959). The Introduction, pp. 206-211, pp. 350-352, and pp. 367-376 are of special interest in connection with the present article. See also, E. Kretschmann, *Ann. Physik* **53**, 575 (1917); and W. Pauli, *Enzyklopädie der Mathematischen Wissenschaften* (B. G. Teubner, Berlin, 1921), Vol. 2, No. 52.

¹⁸ We use causal in the sense of completely deterministic.

3.1b. States Suspae

The Π function defined in 3.1a is valid for an isolated system, and is valid if the observations referred to therein occur during its period of isolation. It is often useful to imagine that a system remains isolated forever, and has always been isolated. This assumes that the system, with properties as given by the input observations, could have been produced prior to any time and can be kept from interacting with other systems for any desired length of time. For such a system, the Π function obtained in Sec. 3.1a is valid for any output variable; its value for an output variable $q_i(t)$ or $v_i(t)$ is a δ function of this variable which vanishes for every q, v except the one obtained by prolonging the orbit in $6n + 1$ dimensional space, given by the input observations, to the time t . If t is unlimited the Π with the appropriate input variables is said to describe a state *sub specie aeternitatis* or, more briefly, a state suspae. Such a state can be specified by the properties of the system considered (the masses of its constituents and the forces acting between these) and the results of $6n$ independent input observations.

3.1c. Invariance Transformations

An invariance transformation will be, according to the postulates of Sec. 2, a one-to-one mapping $\alpha \leftrightarrow \bar{\alpha}$, $\beta \leftrightarrow \bar{\beta}$, \dots of the set of observations onto itself which satisfies

$$(a) \text{ Eq. (2.5)}$$

and

$$(b) \text{ either the conditions of Sec. 2.4a or those of 2.4b.}$$

Equation (2.5) means in our case that if there is an orbit which is consistent with the results r_α for observation α , the result r_β for observation β , and so on (in which case the Π is different from zero), then there is also an orbit which is consistent with the result r_α for $\bar{\alpha}$, the result r_β for $\bar{\beta}$, and so on. Hence, the mappings which satisfy (2.5) induce a mapping of the set of orbits onto itself and it is not difficult to see that every one-to-one mapping of this set onto itself can be induced by a transformation $\alpha \leftrightarrow \bar{\alpha}$, $\beta \leftrightarrow \bar{\beta}$, \dots satisfying (2.5).

Even though the last statement may be pretty evident, it may be useful to spell out its proof in detail because the proof illustrates the necessity of the restriction (b) on invariance transformations. Let us denote a point in coordinate-velocity space (or in the equivalent phase space) by \mathbf{x} . This is then a vector with $6n$ components; for an orbit $\mathbf{x}=\mathbf{x}(t)$ will be a function of time. Every orbit can be characterized by the vector $\mathbf{x}=\mathbf{x}(0)$, i.e., the positions and velocities of the particles at time 0. Let T be an arbitrary one-to-one mapping of the $6n$ dimensional space onto itself. As a rule, this will not be linear. T will also induce a one-to-

one mapping of orbits into orbits: it maps the orbit, the coordinates and velocities of which at time 0 are characterized by \mathbf{x} (which "goes through the point \mathbf{x} at time 0) into the orbit which goes through the point $\mathbf{x}'=T\mathbf{x}$ at time 0. Let us denote the coordinate-velocity vector of the latter orbit at time t by $\mathbf{x}'(t)$. There is a one-to-one mapping now from $\mathbf{x}(t)$ into $\mathbf{x}(0)=\mathbf{x}$, from this to $\mathbf{x}'(0)=\mathbf{x}'=T\mathbf{x}$, and from this to $\mathbf{x}'(t)$. Hence, there is also a one-to-one mapping T_t which maps $\mathbf{x}(t)$ into $\mathbf{x}'(t)$. It follows that if $\mathbf{x}(t)$ is an orbit, $\mathbf{x}'(t)=T_t\mathbf{x}(t)$ is also an orbit; it will be called the transform of the orbit $\mathbf{x}(t)$. The mappings just defined will permit the explicit definition of a mapping $\alpha \leftrightarrow \bar{\alpha}$, $\beta \leftrightarrow \bar{\beta}$, \dots of the set of classical observables on itself which satisfies (2.5).

Let us consider first an observable α which is a function of the coordinates and velocities measured at time t_α . Such an α can be the measurement of one of the coordinates at time t_α , or of the total kinetic energy at t_α , etc. It can be characterized by a function $A(\mathbf{x})$ of the coordinate-velocity vector \mathbf{x} and the time t at which it is measured. This α will be mapped into another measurement $\bar{\alpha}$ which is also a measurement of a function of the coordinates and velocities at time t ; it is the measurement of that function A' which assumes for the transformed orbits the same values which A assumes for the original orbits

$$A'(T_t\mathbf{x}) = A(\mathbf{x}). \quad (3.1)$$

Similarly, $B'(T_{t'}\mathbf{x})=B(\mathbf{x})$ if β is a measurement at time t' , etc. Since T_t is a one-to-one mapping, (3.1) defines the function A' uniquely, and the other functions B' , \dots are defined in the same way. It is clear, furthermore, that (2.5) will be satisfied for this mapping $\alpha \leftrightarrow \bar{\alpha}$, $\beta \leftrightarrow \bar{\beta}$, \dots : The orbit $\mathbf{x}'(t)$ defined by the equations $A'(\mathbf{x}'(t_\alpha))=r_\alpha$, $B'(\mathbf{x}'(t_\beta))=r_\beta$, \dots , is the transform of the orbit $\mathbf{x}(t)$ defined by

$$A(\mathbf{x}(t_\alpha))=r_\alpha, B(\mathbf{x}(t_\beta))=r_\beta, \dots$$

If $\mathbf{x}(t)$ is in conflict with one of the output equations of (2.5), such as $Z(\mathbf{x}(t_\zeta))=r_\zeta$, the orbit $\mathbf{x}'(t)=T_t(\mathbf{x}(t))$ will be in conflict with the corresponding equation $Z'(\mathbf{x}'(t_\zeta))=r_\zeta$. In this case both sides of (2.5) vanish. If the two sides of Eq. (2.5) are integrated over r_ζ , r_η , \dots , r_ν , the result will be 1 in both cases if the region of integration includes the point $r_\zeta=Z(\mathbf{x}(t_\zeta))$, \dots , $r_\nu=N(\mathbf{x}(t_\nu))$ because this is the same point for which $r_\zeta=Z'(\mathbf{x}'(t_\zeta))$, \dots , $r_\nu=N'(\mathbf{x}'(t_\nu))$ is also valid.

We yet have to define the maps $\bar{\alpha}$ of observables which involve measurements at more than one time t_1, t_2, \dots . An observation of this nature can be characterized by a function $A(\mathbf{x}_1, \mathbf{x}_2, \dots)$ of several coordinate-velocity vectors and the times t_1, t_2, \dots to which these refer. An observable of this nature is, for instance, the product of the kinetic energies at times t_1 and t_2 . Actually, it is neither customary nor necessary to consider observables of this nature. However,

the map $\bar{\alpha}$ of such an observable ought to be given for the sake of completeness: it is characterized by a function $A'(\mathbf{x}_1, \mathbf{x}_2, \dots)$ and the same times t_1, t_2, \dots which characterize α . The function A' is defined by

$$A'(T_{t_1}\mathbf{x}_1, T_{t_2}\mathbf{x}_2, \dots) = A(\mathbf{x}_1, \mathbf{x}_2, \dots). \quad (3.1a)$$

It is evident that (2.5) is valid also if its variables α, β, \dots are of the more general nature now considered. It follows that the mapping $\alpha \leftrightarrow \bar{\alpha}, \beta \leftrightarrow \bar{\beta}, \dots$ described leaves the probability function Π invariant and is an element of the group G , and this is true for every one-to-one mapping T of the $6n$ dimensional space of the vectors \mathbf{x} upon itself.

The fact that the knowledge of all the transformations which satisfy (2.5), together with the knowledge of a single orbit, permits the full determination of all the orbits, clearly indicates that the condition (a) by itself would define too broad a class of transformations $\alpha \leftrightarrow \bar{\alpha}, \beta \leftrightarrow \bar{\beta}, \dots$ as invariance transformations. As the preceding discussion shows, these transformations also depend on the interaction between the particles and since we do not believe we know the exact nature of these interactions, we can not truly specify them. The real principles of invariance will concern transformations of such a nature that they preserve (2.5) for all forms of the interaction which we consider conceivable.

3.1d. *The Passive View of Invariance Transformations in Classical Mechanics*

We adopt in this section the point of view of 2.4a, according to which a transformation $\alpha \leftrightarrow \bar{\alpha}, \beta \leftrightarrow \bar{\beta}, \dots$ is an invariance transformation only if the observations α, β, \dots can be in the same relation to an observer as are $\bar{\alpha}, \bar{\beta}, \dots$ to a second one.

Let us consider first position measurements. If α is such a measurement, $\bar{\alpha}$ is also a position measurement from the point of view of the second observer. Since "position at a time" for one observer has a similar meaning for all observers, it follows that $\bar{\alpha}$ is a position measurement if α is such a measurement.

Let us now consider two states, ϕ and $\bar{\phi}$, the first of which is in the same relation to the first observer as $\bar{\phi}$ is to the second. In classical theory we assume that the relative configuration of the particles, i.e., *their distances from each other at a given time, are measurable and do not depend on the observer*. If a configuration is assumed only once in the course of the history of ϕ , this must be true also for $\bar{\phi}$ with respect to the same relative configuration. The times at which this happens shall be denoted by the first observer by t_0 and \bar{t}_0 . If the configuration of ϕ were different at times $t+t_0$ from that of $\bar{\phi}$ at $t+\bar{t}_0$, this again would constitute a difference independent of the observer because the *time intervals between observations as between any two events, are independent of the observer* in classical theory. It follows then that the configurations given by ϕ and by $\bar{\phi}$, at times $t+t_0$ and $t+\bar{t}_0$, respectively, are identical.

If two configurations are identical, they can be transformed into each other by an Euclidean transformation, that is a rotation followed by a displacement, and possibly by an inversion. This is a geometrical theorem. However, if we postulate the existence of at least four particles, the inversion must be excluded because the *right- or left-handed character of a configuration is also the same for all observers*. Hence, if α is the measurement of a position coordinate at time t , $\bar{\alpha}$ is also a measurement of a position coordinate, but at time $t-t_0+\bar{t}_0$, and in a coordinate system which is obtained from the first one by a Euclidean transformation. This transformation could depend on time but if the first coordinate system is an inertial one, i.e., *if Newton's equations are valid therein with the forces vanishing at very large distances*, the transformation $\alpha \leftrightarrow \bar{\alpha}, \beta \leftrightarrow \bar{\beta}, \dots$ will satisfy (2.5) only if the displacement of the two coordinate systems is a linear function of time. Such a transformation can indeed be considered an invariance even in the more restricted sense of 2.4a because one can well imagine an observer to have been put into uniform motion, even into a uniform motion at very high velocity.

The preceding argument gives only the transformation $\alpha \leftrightarrow \bar{\alpha}$ if α is a position measurement and if one of the observers uses an inertial frame for measuring positions. For these measurements $\alpha \leftrightarrow \bar{\alpha}$ is a Galilei transformation. Since, however, the *velocities can be obtained from the positions by differentiation*, the restriction to position measurements can be dropped. Thus the postulates given in 2.4a reduce the multitude of transformations which satisfy (2.5) to the relatively small set of the proper Galilei transformations. The concrete assumptions which were made in the spirit of 2.4a were set in italics.

The preceding discussion has yet to be completed in an important way. The conditions which we obtained for a transformation to be an invariance transformation are only necessary ones inasmuch as they were shown only to follow from the additional conditions of 2.4a. In order to achieve invariance for Galilei transformations we must add the condition that the mapping of observations into observations, as defined by a Galilei transformation, belongs to the group G , i.e., satisfies (2.5). It is necessary, therefore, to admit within the framework $m\ddot{\mathbf{q}}_k = \mathbf{F}_k$ of Newtonian theory only such forces \mathbf{F}_k which lead to Π satisfying (2.5) for Galilei transformations.

It follows from the discussion in 3.1a that Π will be invariant under Galilei transformations if the set of orbits in $6n+1$ dimensional space is invariant under these transformations and this will be true if it is true for the equations of motion. Since the whole orbit is determined by the positions and velocities of all the particles at one instant of time, the expression for the force $\mathbf{F}_k = m_k\ddot{\mathbf{q}}_k$ can be considered to be a function of these quantities. The postulate of invariance therefore becomes the postulate that the \mathbf{F}_k be Galilei covariant

vector functions of the instantaneous positions and velocities.

This still leaves a great freedom in the fundamental equations of motion so that the Galilei invariance, as here formulated, is not very restrictive. The general expression for the force becomes

$$F_i = \sum_k (\mathbf{q}_{ki} f_{ki} + \dot{\mathbf{q}}_{ki} g_{ki}) + \sum_{k,l} [(\mathbf{q}_{ki} \times \mathbf{q}_{li}) f_{kli} + (\dot{\mathbf{q}}_{ki} \times \dot{\mathbf{q}}_{li}) g_{kli} + (\mathbf{q}_{ki} \times \dot{\mathbf{q}}_{li}) h_{kli}], \quad (3.2)$$

where

$$\mathbf{q}_{ki} = \mathbf{q}_k - \mathbf{q}_i, \quad \dot{\mathbf{q}}_{ki} = \dot{\mathbf{q}}_k - \dot{\mathbf{q}}_i, \quad (3.2a)$$

and the f, g , etc., are Galilei-invariant functions of the coordinates and velocities. These are discussed somewhat more in detail below.

Two remarks may well be made at this point. First, as a rule the forces are assumed to be quite simple functions of the instantaneous positions and velocities. In fact, in classical mechanics, the forces are usually assumed to be derivable from a potential V

$$F_k = - \text{grad}_k V, \quad (3.3)$$

where V is a Galilei-invariant function of the position coordinates only. V then can be expressed as a function of some primitive invariants: the distances $|\mathbf{q}_l - \mathbf{q}_m| = r_{lm}$ between the particles and the oriented volumes of the tetrahedra spanned by them

$$6v_{ijkl} = (\mathbf{q}_j - \mathbf{q}_i) \cdot ((\mathbf{q}_k - \mathbf{q}_i) \times (\mathbf{q}_l - \mathbf{q}_i)). \quad (3.4)$$

Since the distances and oriented volumes are *ipso facto* invariant under a transition to a moving coordinate system, the Galilei-invariant V becomes a time-independent invariant under the Euclidean group, and the Galilei invariance becomes, under the assumption (3.3), a consequence of time displacement and Euclidean invariance.

Naturally, the invariants r_{lm} and v_{ijkl} are not independent of each other. As to the r_{lm} , there are, naturally, only $3n - 6$ independent distances between n particles. The essential identities between the r_{lm} can be expressed¹⁹ as the vanishing of the determinants of four-by-four matrices $a^{(i)}_{kl}$

$$\det a^{(i)}_{kl} = 0,$$

$$a^{(i)}_{kl} = ((\mathbf{q}_k - \mathbf{q}_i) \cdot (\mathbf{q}_l - \mathbf{q}_i)) = \frac{1}{2}(r_{il}^2 + r_{ik}^2 - r_{kl}^2), \quad (3.5)$$

i being an arbitrary one of the points and the indices k and l assuming any four values for the four rows and columns of the determinant. Similarly, the product of any two oriented volumes can be expressed in terms of scalar products and hence of distances between particles. The expression for the product is simplest if one

of the corners coincides

$$36v_{ijkl}v_{i'j'k'l'} = \begin{vmatrix} a^{(i)}_{jj'} & a^{(i)}_{jk'} & a^{(i)}_{jl'} \\ a^{(i)}_{kj'} & a^{(i)}_{kk'} & a^{(i)}_{kl'} \\ a^{(i)}_{lj'} & a^{(i)}_{lk'} & a^{(i)}_{ll'} \end{vmatrix}. \quad (3.6)$$

In particular, the squares of the volumes are all given by (3.6) as functions of the distances, but (3.6) yields indirectly also the product of *any* two oriented volumes. Hence, V can be considered to be a linear function of the oriented volumes (in fact, V is a linear function of one oriented volume for those configurations for which this volume does not vanish). In order to avoid having to deal with this exceptional case, and for reasons of symmetry, we write

$$V = V_1(r_{12}, \dots, r_{n-1,n}) + \left(\sum_{klmn} V_2^{klmn}(r_{12}, \dots, r_{n-1,n}) v_{klmn} \right). \quad (3.7)$$

In this V_1 and the V_2^{klmn} depend on the distances between the particles only. This is, of course, only an analytic expression for the fact that the distances between the particles determine the configuration except that they do not distinguish it from its mirror image.

Let us observe that the conservation laws of linear momentum, energy, and angular momentum follow from (3.7) [and (3.3)] by direct calculation. So does also the uniform motion of the center of mass. They do not follow from Eq. (3.2), and therefore they do not follow from the invariance with respect to Galilei transformations. The connection between invariances and conservation laws is a property of Lagrange's equations rather than of classical mechanics.²⁰ A further illustration of this fact is provided, for instance, by

²⁰ In classical mechanics, the conservation laws do not follow from the invariance principles and Newton's equations of motion, but from the invariance principles and Lagrange's equations. This has been emphasized repeatedly, for instance by the last one of the present authors [Progr. Theoret. Phys. (Kyoto) **11**, 437 (1954)]. Dr. E. Guth kindly acquainted us with his study of the history of the connection between conservation laws and the invariance of the Lagrangian. Apparently, the first one to notice the connection (in 1842) was C. G. J. Jacobi (*Vorlesungen über Dynamik*, Werke, Supplementband, Berlin, 1884) who derived the conservation laws for linear and angular momentum from the Euclidean invariance of the Lagrangian. J. R. Schütz [Gött. Nachr. (1897), p. 110] who, incidentally, did not seem to know Jacobi's considerations, derived the energy principle in a similar fashion. The next important paper is that of G. Hamel, Z. Math. Physik **50**, 1 (1904), who, again, was unaware of his predecessors. The first complete discussion of the derivation of the ten integrals of motion (corresponding to the ten infinitesimal elements of the inhomogeneous Lorentz group) was given by G. Herglotz [Ann. Physik **36**, 493 (1911)]. F. Klein called attention to Herglotz' work and encouraged F. Engel [Gött. Nachr. (1916), p. 270 and (1917), p. 189]; E. Noether [*ibid.* (1918), p. 235]; and E. Bessel-Hagen [Math. Ann. **84**, 258 (1921)] to further explicit these ideas. A more modern treatment of the subject was given by E. L. Hill, Rev. Mod. Phys. **23**, 253 (1951).

¹⁹ The identities in question have been discussed, for instance, by G. Young and A. S. Householder, Psychometrika **3**, 19 (1938).

the equation of motion of a particle in a viscous medium. The energy of such a particle is not conserved although the equation of motion is invariant for time displacements. In quantum theory the connection between invariances and conservation laws is much closer as we shall see in Secs. 4 and 5.

3.1e. The More Flexible Point of View of Invariance Transformations in Classical Mechanics

The more flexible viewpoint of Sec. 2.4b can not be reduced to basic principles as was that of 2.4a in the preceding Sec. 3.1d. It is usually assumed that, in addition to the proper Galilei transformations of 3.1d, there are improper transformations which are also invariances. Whether or not these are precise rules, i.e., whether they satisfy (2.5) is a question of fact which can not be decided by reasoning. The two transformations in question are

$$\mathbf{q}_i(t) \leftrightarrow -\mathbf{q}_i(t), \quad \dot{\mathbf{q}}_i(t) \leftrightarrow -\dot{\mathbf{q}}_i(t), \quad (3.8)$$

and

$$\mathbf{q}_i(t) \leftrightarrow \mathbf{q}_i(-t), \quad \dot{\mathbf{q}}_i(t) \leftrightarrow -\dot{\mathbf{q}}_i(-t). \quad (3.8a)$$

It is well known, from experiments in other fields, that at least the first transformation is not a true invariance transformation.²¹ The two transformations (3.8) and (3.8a), if they were valid, would significantly simplify (3.2).

It may be observed that (3.8a) is automatically valid, if the forces can be derived from the velocity-independent potential V of (3.3) and (3.7). If, in addition, (3.8) is a valid invariance transformation, then only the first term on the right-hand side of (3.7) can be nonzero and V becomes a function of the distances alone. The same is true if one postulates that V has to be invariant under permutations of the particles. The more general equation (3.2) is also drastically simplified if (3.8) and (3.8a) are valid invariance transformations.

One cannot but be impressed how assumptions which have, apparently, little to do with invariance principles, nevertheless have the validity of such principles as a consequence. Thus the validity of the Galilei invariance follows from the assumption that the forces can be derived from a velocity-independent potential, which is invariant under time displacements and the Euclidean group. The same assumption guaranteed the time-inversion invariance. Space-inversion invariance (3.7) follows from the assumption that the force exerted by particles on a particle k is the sum of forces, each of which depends in a Galilei invariant way on the position of only one particle, in addition to the position of the particle k . The *TCP*

²¹ We refer here, of course, to the considerations and experiments of Ref. 8 which prove parity violation. A description of the experiments which shows directly the violation of the reflection invariance implied by (3.8) is given in Fig. 3 of the last author's article in *Rev. Mod. Phys.* **29**, 255 (1957).

theorem of quantum field theories²² is a well-known example for the phenomenon discussed; the situation is even more striking in classical theory.

The Lagrangian equations of motion are often considered to form the basis of classical mechanics. The validity of the conservation laws is then a most remarkable consequence of the Euclidean and time-displacement invariance of the Lagrangian. We have not based our considerations on the Lagrangian theory since we wanted to avoid specific assumptions as much as possible. The fundamental assumption of the classical mechanics of point particles is that the observables are the positions and velocities of the particles, and the causal nature of the theory. These are the properties of the theory which underlie our analysis. It follows from these assumptions that the second time derivatives of the coordinates at any time t can be expressed as functions of $6n$ independent variables. One can choose for these the values and first time derivatives of the coordinates, assumed at the same time t . Equation (3.2) then follows from Galilei invariance. Some postulate of simplicity is necessary, however, to arrive at (3.3) whence the remaining discussion and the conservation laws follow.

3.2. Relativistic Mechanics of Point Particles

3.2a. Observables

The input data of the relativistic mechanics of point particles is the same as that of the nonrelativistic theory: if n particles are present, $6n$ independent position coordinates and velocity components are necessary to obtain inferences for the values of these quantities at other times. Furthermore, the theory is again causal: its results can be given more concisely by characterizing the orbits than by giving the Π function which has again δ -function character. However, whereas in nonrelativistic theory a universal time could be used to specify the position and velocity variables, this would not be an appropriate characterization in relativistic theory. It is customary, rather, to introduce a proper time τ_i for each of the points and give the orbits parametrically, in terms of these proper times:

$$x_{i\alpha} = x_{i\alpha}(\tau_i), \quad (3.9)$$

the component $\alpha=0$, that is x_{i0} , being the time. The parametric description of the orbits (3.9) takes account of the fact that time and space play a much more similar role in relativity theory than in classical mechanics. The proper time parameter τ is so chosen that it measures the Minkowski distance along the path of

²² The *PCT* theorem was first recognized by G. Lüders, *Kgl. Danske Videnskab. Selskab, Mat. Fys. Medd.* **28**, 5 (1954). Its significance was further emphasized in W. Pauli's article in *Niels Bohr and the Development of Physics* (Pergamon Press, Inc., New York, 1955), p. 30. See also, J. Schwinger, *Phys. Rev.* **82**, 914 (1951), and, for more modern treatments, R. Jost, *Helv. Phys. Acta* **30**, 409 (1957) and R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (W. A. Benjamin, Inc., New York, 1964).

the particle, i.e., we have

$$l_\alpha(dx_{i\alpha}/d\tau_i)^2 = 1, \quad (3.10)$$

where $l_0 = -l_1 = -l_2 = -l_3 = 1$, and we assume that the summation over the greek indices need not be indicated.

Naturally, the observables in the Π function are the same as in classical theory: the positions and velocities at given times. In order to obtain these from the orbits (3.9) it is necessary first to solve the equation $x_{i0}(\tau_i) = t$ and substitute the τ_i obtained in this way into the other $x_{i\alpha}(\tau_i)$ and the $dx_{i\alpha}/d\tau_i$. This is, of course, for a particular observer more cumbersome than if the coordinates were given directly in terms of the time coordinate of the observer in question. However, the translation of the orbits as observed in one coordinate system into the language of another observer is much facilitated by using the invariant description (3.9) of the orbits. We shall need such translations since we wish to trace the consequences of Lorentz invariance on the possible equations of motion.

3.2b. Equations of Motion

The equations of motion which determine the orbits can be given, just as in classical mechanics, by giving the accelerations. It is in keeping with the parametric description (3.9) of the orbits to give equations for the second derivatives with respect to proper time, i.e., for $d^2x_{i\alpha}(\tau_i)/d\tau_i^2$. These must then satisfy the condition resulting from (3.10).

$$l_\alpha(dx_{i\alpha}/d\tau_i)(d^2x_{i\alpha}/d\tau_i^2) = 0. \quad (3.11)$$

It must be possible to express the accelerations $d^2x_{i\alpha}(\tau_i)/d\tau_i^2$ as functions of any quantities which determine the orbit, i.e., as functions of any $6n$ independent position and velocity components. In classical theory, it was natural to choose the simultaneous position and velocity components as these quantities. This choice made it also easy to translate the equation of motion from one coordinate system to another one, obtained from it by a Galilei transformation. Incidentally, the same "natural" choice had a great influence on the type of equation, (3.3) or (3.7), which we considered "simple."

If the equations of motion of relativistic mechanics were again given in the form in which the accelerations are expressed in terms of positions and velocities of the same time in some coordinate system, the verification of the Lorentz invariance of these equations would now be very difficult. Since "simultaneous" positions and velocities are not simultaneous any more from the point of view of a moving coordinate system, the verification that the equations are valid in a moving coordinate system would have to include the determination of the positions and velocities which are simultaneous from the point of view of another coordinate system, i.e., essentially the determination of the orbits. There are only two ways out of this difficulty.

The first method is to use as characteristics of the orbit when setting up the equations for $d^2x_{i\alpha}(\tau_i)/d\tau_i^2$,

in addition to $x_{i\alpha}$ and $dx_{i\alpha}(\tau_i)/d\tau_i$, the positions and velocities which the other particles had when they passed through the negative light cone of the point $x_{i\alpha}$. Since the light cone is a relativistically invariant concept, these data can be transformed from one coordinate system to another without solving the equations of motion. Equations for $d^2x_{i\alpha}(\tau_i)/d\tau_i^2$ in terms of the positions and velocities which the other particles had when they passed through the negative light cone of $x_{i\alpha}(\tau_i)$ are said to use retarded potentials.²³ Instead, one can also use advanced potentials, i.e., the positions and velocities on the positive light cone of $x_{i\alpha}$, or both.^{23a} The disadvantage of equations of this nature is that they do not seem to permit the establishment of conservation laws. The only exception under this rule is a variant of the theory in which the particles have to be in contact to interact, but given the fact of a three-dimensional space, such equations lead to no interaction between point particles (as contrasted with fields), because a crossing of the one-dimensional world lines is infinitely improbable in the four-dimensional space-time.

The second type of easily translatable equations expresses the second derivative $d^2x_{i\alpha}(\tau_i)/d\tau_i^2$ in terms of integrals over the paths of the other particles. It uses more quantities than necessary to characterize the orbit quantities which are not independent of each other. In this regard, they are similar to equations which use both retarded and advanced potentials and, similarly to them, they do not give a Cauchy problem for the determination of the orbits but give only implicit equations for them. If we assume that the force due to several particles is the sum of the forces due to these particles individually

$$m_i \frac{d^2x_{i\alpha}(\tau_i)}{d\tau_i^2} = \sum_{k \neq i} F_{ik\alpha}(\tau_i). \quad (3.12)$$

$F_{ik\alpha}$ will be expressed in terms of $x_{i\beta}(\tau_i)$, $dx_{i\beta}(\tau_i)/d\tau_i$, and $x_{k\beta}(\tau_k)$, $dx_{k\beta}(\tau_k)/d\tau_k$ for essentially all τ_k , or at least those τ_k for which $x_{k\beta}(\tau_k)$ has a space-like relation to $x_{i\beta}(\tau_i)$. Equations of this nature are also easily translatable, and we shall postulate that the Lorentz invariance be immediately evident from the translation. Before proceeding to examples for such equations, it is

²³ By eliminating the electromagnetic field from the equations, Sommerfeld obtained equations of motion of this type for a pair of charged particles. These equations are manifestly invariant for inhomogeneous Lorentz transformation but the conservation laws are not satisfied. The last statement is not surprising; one expects the particles to radiate. See J. L. Synge, Proc. Roy. Soc. (London) **177**, 11 (1941); also, "Classical Dynamics," in *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1960), Vol. III/1, p. 11; and W. Pauli, "Relativistatstheorie," in *Enzyklopädie der Mathematischen Wissenschaften* (B. G. Teubner, Berlin, 1921), Vol. 2, p. 645 (p. 92 of the English translation, Pergamon Press, Inc., New York, 1959); A. Sommerfeld, Ann. Physik Leipzig **33**, 665 (1910); A. D. Fokker, Z. Physik **58**, 386 (1929).

^{23a} The interaction of charged particles was formulated in terms of retarded and advanced potentials by J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. **21**, 425 (1949). The equations of these authors correspond to (3.20) with $\phi_{ij}(\rho) = e_i e_j (d/d\rho^2) \delta(\rho^2)$. See also J. W. Dettman and A. Schild, Phys. Rev. **95**, 1059 (1954).

well to restate that they are not unique in the sense that equivalent equations, i.e., equations characterizing the same orbits, can have different forms. This is the consequence of the use of redundant quantities. However, the flexibility inherent in the many different forms of equivalent equations permits the use of one, the invariance of which can be easily demonstrated.

3.2c. *Invariance Transformations, Conservation Laws*

The expression suggested in the preceding section for the force exerted by particle k on particle i is of the form

$$F_{ik\alpha}(\tau_i) = \int d\tau_k \hat{F}_{ik\alpha}(x_{i\beta}(\tau_i), \dot{x}_{i\beta}(\tau_i), x_{k\beta}(\tau_k), \dot{x}_{k\beta}(\tau_k)). \tag{3.13}$$

It is a functional of the path $x_{k\beta}(\tau_k)$; the dot denotes differentiation with respect to the proper time. The requirement that the Lorentz invariance of the equation of motion be immediately evident will be interpreted as the condition that not only the value of the integral (3.13) give the same acceleration in different Lorentz frames of reference but that this be true already for the integrand. This means that the integrand of (3.13) transforms as a vector under Lorentz transformations.

If we adopt the narrower interpretation of Sec. 2.4a for invariance transformations, $\hat{F}_{ik\alpha}$ has to have the properties of a vector only under proper Lorentz transformations; its behavior under reflections and time inversion becomes immaterial. The invariants which can be formed from $x_{i\beta}(\tau_i)$, $\dot{x}_{i\beta}(\tau_i)$, $x_{k\beta}(\tau_k)$, $\dot{x}_{k\beta}(\tau_k)$ become

$$\rho_{ik} = \{-l_\alpha(x_{i\alpha}(\tau_i) - x_{k\alpha}(\tau_k))^2\}^{\frac{1}{2}}, \tag{3.14a}$$

$$\sigma_{ik} = l_\alpha \dot{x}_{i\alpha}(\tau_i) (x_{i\alpha}(\tau_i) - x_{k\alpha}(\tau_k)), \tag{3.14b}$$

$$\zeta_{ik} = l_\alpha \dot{x}_{i\alpha}(\tau_i) \dot{x}_{k\alpha}(\tau_k). \tag{3.14c}$$

Note that $\rho_{ik} = \rho_{ki}$, $\zeta_{ik} = \zeta_{ki}$, but σ_{ik} and σ_{ki} are different. We shall henceforth omit the variables τ_i and τ_k in expressions $x_{i\alpha}(\tau_i)$ and $x_{k\beta}(\tau_k)$. The ‘‘length’’ of the vectors \dot{x}_i and \dot{x}_k is 1 according to (3.10) so that these are not relevant invariants and no ‘‘oriented volume’’ can be formed of the three displacement invariant vectors $x_i - x_k$, \dot{x}_i , and \dot{x}_k . Hence, the general expression for $\hat{F}_{ik\alpha}$ becomes

$$\hat{F}_{ik\alpha} = (x_{i\alpha} - x_{k\alpha})f + \dot{x}_{i\alpha}g + \dot{x}_{k\alpha}h + \epsilon_{\alpha\beta\gamma\delta}(x_{i\beta} - x_{k\beta})\dot{x}_{i\gamma}\dot{x}_{k\delta}j, \tag{3.15}$$

where ϵ is the antisymmetric tensor and f , g , h , and j are functions of the invariants ρ_{ik} , ζ_{ik} , σ_{ik} , and σ_{ki} . Presumably, they vanish unless ρ_{ik} is real, i.e., the points x_i and x_k have a space-like relation.²⁴

²⁴ There exists a widely spread opinion that the interaction between particles must be via ‘‘signals,’’ the velocity of which does not exceed that of light if the theory is to be Lorentz invariant. That this opinion must be revised has also been argued by P. Havas and J. Plebansky, *Bull. Am. Phys. Soc.* **5**, 433 (1960).

If we adopt the more flexible point of view of Sec. 2.4b, we have to investigate the behavior of the quantities $x(\tau)$, $\dot{x}(\tau)$, $\ddot{x}(\tau)$ under space and time reflections. As to space reflections, we can confine our attention to the change of the direction of all three space-like coordinate axes because the other reflection operations can be obtained from this by a succeeding proper Lorentz transformation. We have, hence,

$$x_{i\alpha}(\tau_i) \rightarrow l_\alpha x_{i\alpha}(\tau_i), \tag{3.16}$$

and can obtain by differentiation with respect to τ_i

$$\dot{x}_{i\alpha}(\tau_i) \rightarrow l_\alpha \dot{x}_{i\alpha}(\tau_i), \quad \ddot{x}_{i\alpha}(\tau_i) \rightarrow l_\alpha \ddot{x}_{i\alpha}(\tau_i). \tag{3.16a}$$

Since $l_\alpha^2 = 1$, all invariants of (3.14) are unchanged under space reflection.

In the case of time reflection, we have to change the signs of all τ because we wish to measure τ along the positive time axes. Hence, we have in this case

$$x_{i\alpha}(\tau_i) \rightarrow -l_\alpha x_{i\alpha}(-\tau_i), \tag{3.17}$$

and by differentiation

$$\dot{x}_{i\alpha}(\tau_i) \rightarrow l_\alpha \dot{x}_{i\alpha}(-\tau_i), \quad \ddot{x}_{i\alpha}(\tau_i) \rightarrow -l_\alpha \ddot{x}_{i\alpha}(-\tau_i). \tag{3.17a}$$

Naturally, the \dot{x}_{i0} must remain positive. In this case, $\rho_{ik}(\tau)$ and $\zeta_{ik}(\tau)$ go over into $\rho_{ik}(-\tau)$ and $\zeta_{ik}(-\tau)$, whereas the $\sigma_{ik}(\tau)$ go over into $-\sigma_{ik}(-\tau)$.

The expression for the force (3.13) between two particles is invariant under space reflections (Lorentz transformations with determinant -1 which do not reverse the direction of motion) provided $j=0$. This follows from the fact that ρ_{ik} , ζ_{ik} , σ_{ik} , and σ_{ki} do not change under these transformations.

Under time inversion, the coefficient of f in (3.15) has the right sign. Since the σ_{ik} change sign under time inversion, f must be an even function of the σ_{ik} , whereas g , h , and j must be odd functions of these invariants.

Since our purpose is merely to establish the possibility of a relativistically invariant mechanics of point particles,²⁵ we shall not continue the discussion of (3.15) in its full generality. Rather, we assume that only the first term of (3.15) is present, i.e., that $g=h=j=0$. In

²⁵ It has been pointed out recently that such a theory is not possible within the framework of the canonical representation of the Lorentz group. See D. G. Currie, T. F. Jordan, and E. C. G. Sudershan, *Rev. Mod. Phys.* **35**, 350 (1963); also D. G. Currie, *J. Math. Phys.* **4**, 1470 (1963); and J. T. Cannon and T. F. Jordan, *ibid.* **5**, 299 (1964); also H. Ekstein, *Consistency of Relativistic Particle Theories* (Universit  d’Aix-Marseille, 1964). These ‘‘no interaction’’ theorems apply to a Hamiltonian framework for relativistic mechanics which is due to P. A. M. Dirac; see *Rev. Mod. Phys.* **21**, 392 (1949). In this framework Lorentz transformations are represented by canonical transformations. The ‘‘no interaction’’ theorems can be circumvented if one is willing to give up the condition of the existence of world lines; see L. H. Thomas, *Rev. Mod. Phys.* **17**, 182 (1945); B. Bakamjian and L. H. Thomas, *Phys. Rev.* **92**, 1300 (1953); L. L. Foldy, *ibid.* **122**, 275 (1961). They can also be circumvented by dropping the condition that position be a canonical variable; see E. H. Kerner, *J. Math. Phys.* **6**, 1218 (1965). See also, P. Havas, *Rev. Mod. Phys.* **36**, 938 (1964).

For a more detailed discussion of the equation proposed in Sec. 3.2C, see H. Van Dam and E. P. Wigner, *Phys. Rev.* **138**, B1576 (1965).

order to assure the validity of the law of over-all conservation of linear momentum and energy for collision processes, we further assume

$$\hat{F}_{ik\alpha}(\tau_i, \tau_k) = -\hat{F}_{ki\alpha}(\tau_k, \tau_i), \quad (3.18)$$

i.e., that the contributions to the force exerted by i on k and by k on i , contributed during unit proper times, are oppositely equal.²⁶ The law of over-all conservation of linear momentum and energy is easily established by noting that

$$\begin{aligned} \sum_i \int_{-\infty}^{+\infty} d\tau_i \frac{d}{d\tau_i} \left(m_i \frac{dx_{i\alpha}}{d\tau_i} \right) \\ = \sum_i \int_{-\infty}^{+\infty} d\tau_i \sum_k \int_{-\infty}^{+\infty} d\tau_k \hat{F}_{ik\alpha}(\tau_i, \tau_k) = 0 \end{aligned}$$

provided (3.18) is valid and provided the integrals converge. The angular momentum law follows by a similar direct calculation provided $g=h=j=0$. If one further wishes to satisfy (3.11) so that the physical significance of the proper time, as given by (3.10), remains unchanged, one obtains under rather general assumptions

$$f = 2\xi_{ik}\varphi(\rho_{ik}) + (\sigma_{ik}\sigma_{ki}/\rho_{ik})\varphi'(\rho_{ik}), \quad (3.19)$$

in which $\varphi(\rho)$ is an arbitrary function of the space-time distance ρ of (3.14a) and the prime denotes differentiation with respect to ρ . Just as in classical theory, under similar assumptions, the interaction contains an arbitrary function of a single variable. Evidently (3.19) is an even function of the σ_{ik} so that the equation of motion is time-inversion invariant.

The equation of motion becomes, with (3.19) and dropping the last three terms of (3.15),

$$\begin{aligned} m_i \ddot{x}_{i\alpha}(\tau_i) = \sum_k \int (x_{i\alpha}(\tau_i) - x_{k\alpha}(\tau_k)) (2\xi_{ik}\varphi(\rho_{ik}) \\ + \sigma_{ik}\sigma_{ki}\varphi'(\rho_{ik})/\rho_{ik}) d\tau_k. \end{aligned} \quad (3.20)$$

The invariants ξ_{ik} , σ_{ik} , ρ_{ik} are given by (3.14) as functions of τ_i and τ_k . Equation (3.11) can be verified by transforming the expression $\dot{x}_{i\alpha}\ddot{x}_{i\alpha}$ by partial integration.

4. THE ROLE OF THE CONCEPTS OF SECTION 2 IN QUANTUM THEORY

4.1. Introduction

The observables of quantum theory correspond to self-adjoint operators in Hilbert space. This gives the set of observables a much more complicated structure than they have in macroscopic theories, two examples of which were considered in the preceding sections. The theories considered in Sec. 3 are causal and if the results of a certain number ($6n$ in the examples considered) of

observations is known, the results of all other observations are determined in principle. This is not so in quantum theory; no matter how many observations have been carried out on a system, the results of most subsequent observations remain uncertain and subject to probability laws. It follows that the manifold of states *suspae* is much larger in quantum than in non-quantum theories: quantum theory which started by reducing the manifold of "allowed" states now has multiplied them infinite fold. However, it has also introduced a structure into this manifold (the superposition principle and the existence of a transition probability between different states), which will play a decisive role in our considerations.

It may be worthwhile to illustrate these points on a somewhat schematic example. Let us consider an NH_3 molecule, the H atoms of which are somehow fixed in space. These create a certain potential for the N atom which has two minima, one on each side of the plane through the three H atoms. In classical theory, there are two states with the lowest possible energy; the N atom may be at rest at either position of minimum.

Quantum mechanics postulates that the N atom have a certain amount of "zero-point energy" and perform small vibrations around the equilibrium positions in either state. However, this is far from the full story. If the N atom is, originally at one side, it will slowly "leak over" to the opposite side. At time t , it will be with the probability $\cos^2 t/2T$ on the original side, with the probability $\sin^2 t/2T$ on the opposite side; $2\pi T$ is the period of motion. Hence, starting with the N atom on one side, one can obtain a continuum of states in which it is with varying probabilities on both sides. Furthermore, given an ensemble of identically prepared NH_3 molecules, each of the intermediate states can be distinguished from the others, by an instantaneous measurement in the most orthodox theory, but even actually, by waiting $2\pi T - t$ and then ascertaining that the N atom is again with probability 1 on the original side.

The states of the NH_3 molecule obtained by the passage of time from the state in which the N atom was originally on one side with its zero-point energy are not the only states which correspond to the two classical equilibrium states. There are also two stationary states, in both of which the N atom is, and—since the states are stationary—permanently remains, with equal probabilities on both sides. The states considered before are superpositions of the two stationary states, the ratio of the coefficients of superposition being $\exp it/T$. Actually, the ratio of the coefficients can be any complex number, giving a manifold that can be characterized by two real parameters. Furthermore, as our example shows, the complex phase of the superposition parameter has in this case an important physical significance. The states considered are not statistical ensembles of different states; if so they would not all show a definite property (being at one side of the plane

²⁶ Let us remark that many of the interesting considerations of J. A. Wheeler and R. P. Feynman apply also to the present equations of motion; see *Rev. Mod. Phys.* **17**, 157 (1945).

of the H atoms) at certain times. In quantum mechanics, they are called pure states.

The preceding picture somewhat exaggerates the increase in the number of the states from classical to quantum theory by considering only the states of lowest energy in both. Actually $6n$ real parameters are needed to characterize the state of a classical system of n particles whereas an infinity of such parameters is needed to characterize the state of a quantum-mechanical system, no matter what the number of particles is. Furthermore, whereas there is no structure in the space of the states of a classical system, one can coordinate an infinity of states to any two states of a quantum system: the superpositions of the two states with varying coefficients. Unfortunately from the point of view of the phenomenological nature of the theory, there is no general operational method to obtain the superpositions in question, even if the methods to prepare the individual states are known.

It was mentioned before that the complex nature of the quantum mechanical states is a consequence of the complicated nature of the observations postulated by quantum theory. In Sec. 2, we have described observations as actions. Since the observations are characterized by self-adjoint operators in Hilbert space, it follows that quantum mechanics attributes certain actions, usually called measurements, at least to certain self-adjoint operators. It must again be admitted that, unfortunately from the point of view of the phenomenological nature of the theory, the correspondence between the actions necessary to carry out the measurement, and the operator to which the measurement corresponds, is not spelled out. This is, in our opinion, a serious conceptual weakness of the theory. It is not even known to which operators actions can be attributed, i.e., which operators are actually "measurable." There is a body of opinion according to which only the components of the linear momentum and the energy of individual particles is measurable and this only for particles which are spatially separated from all others. According to this view, the concept of a precise position must also be derived from such momentum and energy measurements. Interestingly enough, it is not necessary to specify, for our purposes, the set of measurable operators. We shall postulate, however, in most of the discussion, that the set of measurable operators have the following properties.

(a) It permits the selection, usually called preparation of a reasonably large set of pure states. The desired set will be specified later; we shall define here only the concept of a pure state.

Let us consider a set of measurements α, β, \dots . If these give the results r_α, r_β, \dots ,

$$\Pi(\alpha, r_\alpha; \beta, r_\beta; \dots | \zeta, r_\zeta; \eta, r_\eta; \dots) \quad (4.1)$$

gives the probability for the results r_ζ, r_η, \dots of the measurements ζ, η, \dots . For given $\alpha, r_\alpha, \beta, r_\beta, \dots$ (4.1) can be considered as a function of the output measure-

ments ζ, η, \dots and their results r_ζ, r_η, \dots . Let us assume that the output measurements are all later than the input measurements. Evidently then the selection of the samples which gave the results r_α, r_β, \dots for the measurements α, β, \dots creates an ensemble the properties of which for later times are given by the Π , considered as a function of the output variables $\zeta, r_\zeta, \eta, r_\eta, \dots$. This ensemble will be called a pure state if (4.1), as a function of the output variables is identical with

$$\Pi(\mu, r_\mu; \alpha, r_\alpha; \beta, r_\beta; \dots | \zeta, r_\zeta; \eta, r_\eta, \dots) \quad (4.1a)$$

for all μ and r_μ provided μ precedes in time all α, β, \dots .

In less technical language this means that the selection $\alpha, r_\alpha; \beta, r_\beta; \dots$ is so far reaching that it does not matter with which state one started. Notice that a similar relation exists also in macroscopic theories: $\alpha, r_\alpha; \beta, r_\beta; \dots$ would be, in this case, $6n$ independent observations. If these are given, any prior observation on the system is redundant and does not yield additional information concerning the future behavior.

Postulate (a) is not at all obvious. It is quite conceivable that, no matter how many observations one used for the selection of an ensemble, further observations, preceding the ones used for the selection, would have produced an ensemble with even more closely defined properties. Hence, (a) demands the possibility of defining a state with a final accuracy on which no further improvement is possible.

Postulate (a) has been questioned recently.²⁷ It has been pointed out, for instance, that, as the wavelength of a light quantum increases, its detection becomes increasingly difficult. It may not be possible, therefore, to ascertain that there is no light quantum present in the system even though it may be possible to ascertain that there is no quantum present with a wavelength above λ , where λ is an arbitrary length. If this were the case, no pure state could be prepared because the properties of a system could be specified increasingly precisely by making statements about light quanta of increasing wavelengths.

It is not yet clear to what extent the reservations concerning the possibility of preparing pure states are valid, and at any rate, the theory of invariance in quantum mechanics has not yet been worked out without the use of the concept of pure states. We shall, therefore, continue to use postulate (a). It is clear that similar objections can be raised also against the concept of isolated systems. There is, in fact, some connection between the two types of objections because in order to guarantee the absence of light of very long wavelength, one must extend the control over very large spatial regions.

In addition to demanding the existence of pure states, we postulate that a sufficient variety of them can be considered states *suspae*. This means that the ensemble

²⁷ R. Haag and D. Kastler, *J. Math. Phys.* **5**, 848 (1964).

which has been selected by the use of the experiments α, β, \dots could have been selected by other experiments, carried out prior to any given time.

(b) If a measurement can be carried out by an observer, the same or an equivalent measurement can be carried out by any other observer. This is the translation of the measurements and of their results between different observers which was discussed already in Sec. 2.

(c) If a measurement can be carried out by one observer, subjectively identical measurements can be carried out by all equivalent observers. This postulate was also discussed in Sec. 2.

4.2. Observations and Decision Observations

It was mentioned before that quantum theory associates measurements with certain self-adjoint operators. We shall say, somewhat loosely, that one can measure A and mean that there is a measurement associated with the operator A . One can think of the measurement as bringing the system on which A is to be measured into interaction with a suitable apparatus which will somehow produce a measurement result; the interaction will also affect the system on which the measurement is carried out. The operator A associated with a measurement is so defined that its characteristic values are the measurement results which the apparatus may produce.

The apparatus used for measuring A and for measuring a function of A , say A^2 , is the same but if A^2 is measured, the measurement result is called λ^2 in those events in which the result is called λ if A is measured. This makes it possible to replace any measurement by a set of simple measurements. If the characteristic values of A are $\lambda_1, \lambda_2, \dots$, one can form the functions $p_1(A), p_2(A), \dots$, where $p_a(\lambda) = 1$ for $\lambda = \lambda_a$, it vanishes if λ is equal to one of the other characteristic values. Hence the two statements: the measurement of $p_a(A)$ gave the result 1, and the other, the measurement of A gave the value λ_a , are essentially synonymous. The measurement of $p_a(A)$ decides whether the measurement of A would have given the value λ_a [and that of $f(A)$ the value $f(\lambda_a)$]. One can, in the way indicated, replace every measurement by a set of decision measurements. Naturally, the results of these decision measurements may not be independent of each other: in the preceding case the measurement of one $p_i(A)$ will give the value 1, the rest of them will give the result 0.

The advantage of restricting ourselves to decision observations is partly that we avoid unrealistic measurements such as measurements with infinite accuracy in a continuous spectrum. These would lead us to states outside Hilbert space. More importantly, the use of decision observations frees us from the arbitrariness inherent in the labeling of outcomes of measurements. As was mentioned before, the same action could have been

called a measurement of A or of $f(A)$; if the same outcome is labeled, respectively, λ_a or $f(\lambda_a)$, where f may be any real function. The operators $p_1(A), p_2(A), \dots$ are, on the other hand, the same no matter whether we started with A or $f(A)$; they depend only on the essential properties of the apparatus. Similarly, the outcomes—0 for all but one of the $p(A)$, and 1 for the this one—do not reflect the arbitrariness of the original labeling of the results.

The operators $p_a(A)$ are called projection operators and are usually denoted by P with a suitable index. They are real functions of self-adjoint operators—hence themselves self-adjoint—their characteristic values are 0 and 1. Hence, they satisfy the equation $P^2 = P$ and more generally $P^n = P$. Furthermore, $1 - P$ is also a projection operator, $(1 - P)^2 = 1 - 2P + P^2 = 1 - 2P + P = 1 - P$. If $P = p_a(A)$ decides whether the value λ_a would have been the result of the measurement of A , the operator $1 - P$ decides whether the result would have been *different* from λ_a . The projection operators $p_1(A), p_2(A), \dots$ commute because they are all functions of the same operator A ; in general, projection operators need not commute.

4.3. The Heisenberg Picture

The time at which a measurement is carried out will be added to the symbol of the measurement. Hence (P, t) is a decision observation carried out at time t . Thus (P, t) is an *action*, whereas P is the operator attributed to it. Let us consider measurements on isolated systems. If (P, t_0) is a decision observation, then it may be possible to devise another action, carried out at time t , which gives the result 1 with the same probability as (P, t_0) if carried out on the same system. This should be valid for all isolated systems. The second observation is then denoted by (P, t) : it is characterized by the same operator P which was used to characterize the original measurement (P, t_0) . A drawback of this notation is that the same action as associated with (P, t_0) , but performed at time t is *not* the action (P, t) . If this notation, the Heisenberg picture, is used, practically the whole content of the theory is contained in the correlation between the observations and the self-adjoint operators. The general Π functions will be given in the next section in terms of the operators P and the problem is, therefore, to correlate these operators with observations. One question of such a correlation concerns the operator which represents the same measurement carried out at time t which is represented by a given operator at time t_0 .

The projection operators of trace 1 will play an important role in our discussion. Such operators project onto a single direction in Hilbert space and may be characterized uniquely by any vector pointing in that direction. They may be characterized also by the set of all vectors in that direction. Such a set of parallel vectors $\alpha\phi$, where ϕ is a unit vector and α an arbitrary

complex number ($\neq 0$) is called a ray and is denoted by Φ . There is, therefore, a one-to-one correspondence between rays and trace 1 projection operators, and we shall label the trace 1 projection operators P_Φ by the symbol Φ of the ray to which they correspond. We have

$$P_\Phi \chi = (\phi, \chi)\phi, \tag{4.2}$$

where χ is an arbitrary vector and ϕ can be any unit vector from the ray Φ because the right-hand side of (4.2) remains unaltered if one substitutes $\omega\phi$ for ϕ with $|\omega| = 1$. Equation (4.2), which may also be expressed as $P_\Phi = \phi \times \phi^*$, shows that P_Φ projects every vector in the direction of the vectors in Φ . We shall often use an abbreviated notation and write P_m for the trace 1 operator associated with the ray which contains ϕ_m and P_{m+n} for the operator associated with the ray which contains $\phi_m + \phi_n$.

Observations represented by (P_Φ, t) , in which P_Φ is a trace 1 projection operator, will be called *complete-decision observations*. The trace 1 operator P_Φ which forms part of the representation of such a complete-decision observation will be called a *c-d operator*. The terms observable trace 1 projection operator and c-d operator will be used as synonyms.

4.4. The Π Function of Quantum Theory

The Π function of quantum theory can be derived from the usual postulate that if the state vector is ϕ , the probability of a positive result (i.e., the result +1) for the measurement of the projection operator P_1 is $(\phi, P_1\phi) = (\phi, P_1^2\phi) = (P_1\phi, P_1\phi)$ and that, if the result is positive, the state of the system, after the measurement, becomes the normalized $P_1\phi$, i.e., it becomes $P_1\phi(P_1\phi, P_1\phi)^{-\frac{1}{2}}$. Similarly, if the result of the measurement is negative (i.e., the result of the measurement of P_1 is 0), the probability for which is $(\phi, (1-P)\phi)$, the state vector becomes parallel to $(1-P)\phi$. If, in the case of a positive result, a second decision measurement is carried out subsequently, the probability that it gives a positive result is

$$\left(P_2 \frac{P_1\phi}{(P_1\phi, P_1\phi)^{\frac{1}{2}}}, P_2 \frac{P_1\phi}{(P_1\phi, P_1\phi)^{\frac{1}{2}}} \right) = \frac{(P_2P_1\phi, P_2P_1\phi)}{(P_1\phi, P_1\phi)}. \tag{4.3}$$

Hence, the probability that *both* of them give a positive result is the product of (4.3) and of the probability that the first measurement gave a positive result, i.e., equal to $(P_2P_1\phi, P_2P_1\phi)$. For several successive decision measurements P_1, P_2, \dots, P_n , carried out in this order, the probability of a positive result throughout is

$$(P_n \dots P_2P_1\phi, P_n \dots P_2P_1\phi). \tag{4.4}$$

Let us assume, first, that the input measurements all precede the output measurements, the former being denoted by $(P_a, t_a), \dots, (P_e, t_e)$, the latter by $(P_f, t_f), \dots, (P_n, t_n)$. Furthermore, these should be already

time-ordered, i.e., the labeling a, \dots, e, f, \dots, n should correspond also to the time order. If some of the measurements are carried out simultaneously, the corresponding projection operators commute and their order in (4.4) is immaterial. Then, the conditional probability that if $(P_a, t_a), \dots, (P_e, t_e)$ gave a positive result, $(P_f, t_f), \dots, (P_n, t_n)$ will do likewise, becomes

$$\frac{(P_n \dots P_f P_e \dots P_a \phi, P_n \dots P_f P_e \dots P_a \phi)}{(P_e \dots P_a \phi, P_e \dots P_a \phi)} = \frac{(\phi, P_a \dots P_e P_f \dots P_n P_n \dots P_f P_e \dots P_a \phi)}{(\phi, P_a \dots P_e P_e \dots P_a \phi)}. \tag{4.5}$$

The right-hand side follows since the P are all self-adjoint.

Let us assume that the positive outcome of the set of measurements $(P_a, t_a), \dots, (P_e, t_e)$ leads to a pure state. Then, the input measurements completely determine the state and (4.5) will be independent of ϕ . Hence, both numerator and denominator can be replaced by a sum of similar expressions in which ϕ is replaced by the members of a complete orthonormal set. Then, both numerator and denominator become the trace of the operator which appears therein and we have $(t_a \leq t_\beta \dots \leq t_e \leq t_f \leq \dots \leq t_n)$,

$$\begin{aligned} \Pi((P_a, t_a), \dots, (P_e, t_e) \mid (P_f, t_f), \dots, (P_n, t_n)) &= \frac{\text{Tr}(P_a \dots P_e P_f \dots P_n P_n \dots P_f P_e \dots P_a)}{\text{Tr}(P_a \dots P_e P_e \dots P_a)} \\ &= \frac{\text{Tr}(Q_{in}^\dagger Q_{out}^\dagger Q_{out} Q_{in})}{\text{Tr}(Q_{in}^\dagger Q_{in})} = \frac{\text{Tr} Q_{out} Q_{in} Q_{in}^\dagger Q_{out}^\dagger}{\text{Tr} Q_{in} Q_{in}^\dagger}, \end{aligned} \tag{4.6}$$

in which

$$Q_{in} = P_e \dots P_a \quad \text{and} \quad Q_{out} = P_n \dots P_f. \tag{4.6a}$$

The last line of (4.6) is obtained by a cyclical interchange of the factors in the trace. The notation in (4.6) differs from that of (2.1) inasmuch as the result of the measurement is not specified. For a decision measurement, this shall be always 1 in the Π function; if the probability for the result 0 is desired, P will be replaced by $1-P$.

Formula (4.6) for one input measurement (P_a, t_a) and one output measurement (P_b, t_β) is

$$\Pi((P_a, t_a) \mid (P_b, t_\beta)) = \frac{\text{Tr} P_a P_b P_b P_a}{\text{Tr} P_a P_a} = \frac{\text{Tr} P_a P_b}{\text{Tr} P_a}. \tag{4.6b}$$

If (P_a, t_a) is not a complete decision this formula includes the assumption of equal *a priori* probabilities for the eigenstates with eigenvalues 1 of P_a ; the appropriate statistical matrix is $P_a (\text{Tr} P_a)^{-1}$. Formula (4.6) can also be derived under the condition of equal *a priori* probability for the eigenstates with eigenvalue 1 of P_a , or, equivalently under the condition of equal

a priori probability for all states. With the latter condition (4.6) is also valid if the time order of the measurements is opposite to that mentioned in (4.6). This condition of equal *a priori* probability is also assumed in the following, more general, expression for the Π function.

If the input measurements do not all precede the output measurements, the expression for the Π function is somewhat more complicated. The numerator as given by the second member of (4.6) remains the same, except that the order of the first set of factors is such that earlier measurements, whether input or output, precede later ones. The opposite is true for the second set of factors. This gives the un-normalized probability that all decision measurements—both input and output—give a positive result. Denoting the denominator by N , we have, therefore, for an arbitrary order of the measurement times

$$\Pi((P_a, t_a), \dots, (P_e, t_e) | (P_f, t_f), \dots, (P_n, t_n)) = N^{-1} \text{Tr } Q^\dagger Q, \quad (4.7)$$

$$Q = T(P_n \dots P_f P_e \dots P_a), \quad (4.7a)$$

T being the time-ordering operator which arranges the succeeding factors according to the times the corresponding measurement has been carried out.

The denominator N is so determined that the sum of Π for all possible outcomes of the output measurements be 1 [cf. (2.3)]. The formal expression is

$$N = \sum_{\sigma_f, \dots, \sigma_n} \text{Tr } Q(\sigma_f, \dots, \sigma_n) Q(\sigma_f, \dots, \sigma_n)^\dagger. \quad (4.7b)$$

All σ assume the values ± 1 and if a σ is -1 the corresponding P in (4.7a) has to be replaced by $1 - P$. Thus

$$Q(\sigma_f, \dots, \sigma_n) = T\left[\left(\frac{1}{2}(1 - \sigma_n) + \sigma_n P_n\right) \dots \left(\frac{1}{2}(1 - \sigma_f) + \sigma_f P_f\right) P_e \dots P_a\right]. \quad (4.7c)$$

The quantum-mechanical equations for Π are quite on a par with the expressions for Π discussed for the classical theory of point particles. In that case, the expression for Π could be given in terms of the orbits, in the present case it is given in terms of the operators (of the Heisenberg picture) which correspond to the various possible observations. Concrete results can be obtained in the classical theory of point particles only if one determines the orbits by solving the equations of motion: in the present case Heisenberg's equations for the time dependence of the operators play a similar role. However, detailed solutions of these equations will not be needed for our discussion of the invariance principles.

We wish to stress the importance of the time ordering in (4.7a) and (4.7c). In quantum theory, every measurement on a system the result of which can not be foretold with certainty, perturbs the system. The order in which measurements are performed is essential. Hence in relativistic quantum theory certain conditions

must be satisfied for any two possible observations (P_a, t_a) and (P_b, t_b) . We must require that P_a and P_b can each be decomposed in projection operators $P_a = \sum_\lambda P_{a\lambda}$, $P_b = \sum_\mu P_{b\mu}$ in such a way that any $(P_{a\lambda}, t_a)$ is in the absolute past or absolute future of any $(P_{b\mu}, t_b)$ with which it does not commute. If $(P_{a\lambda}, t_a)$ is in the absolute past of $(P_{b\lambda}, t_b)$ then the actions $(P_{a\lambda}, t_a)$ and $(P_{b\mu}, t_b)$ must be restricted to a region of finite extension in space.

Newton and Wigner²⁸ have constructed, in the theory of a free relativistic particle with positive energy, self-adjoint operators for the measurement whether the position of the particle is, at a given time, within a certain region. These are projection operators which can be obtained from the operators of their localized states by integrating the expression (4.2) over a finite spatial domain with the ϕ being the wave vector of a state-localized at one point. However, the projection operators obtained in this way do not commute if they refer to different times even if the two regions have an entirely space-like relation to each other. This leads us to conclude that these operators can not represent observations which can be carried out instantaneously within the regions considered. One can infer this also from the nonrelativistic nature of these localized states. In macroscopic theory all observations are compatible and restrictions such as we discussed for quantum theory do not exist.

Formula (4.7) has the remarkable property that the right-hand side does not change if one changes the times $t_a, \dots, t_e, t_f, \dots, t_n$ in such a way that their order remains the same. The observation (P_a, t_a) is in this sense equivalent to (P_a, t'_a) , in which we must remember, however, that the action (P_a, t'_a) is, in general, not the action (P_a, t_a) performed at t'_a (Heisenberg representation).

4.5. Pure States and Trace 1 Projection Operators

Let us consider again output measurements which are all performed later than any of the input measurements. The expression for the Π function is then given by (4.6). When deriving this formula, it was assumed that the succession of input measurements $(P_a, t_a), \dots, (P_e, t_e)$, if they all give a positive result, renders the results of later experiments independent of the original state of the system. This means, phenomenologically, that when considering the future behavior, outcomes of observations made prior to t_a are immaterial: the Π as a function of the output variables is independent of these results. According to the definition of pure states given in Sec. 4.1, this means that the succession of positive results for $(P_a, t_a), \dots, (P_e, t_e)$ defines a pure state.

We wish to show now that the positive result of the succession of $(P_a, t_a), \dots, (P_e, t_e)$ is equivalent, as

²⁸ T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).

far as later measurements are concerned, with the positive outcome of a single trace 1 decision measurement (complete-decision measurement). This means that the state obtained by choosing the samples for which the aforementioned observations gave a positive result, could be chosen also on the basis of a single complete-decision measurement. In ordinary parlance, the state can be described by a wave function.

Clearly, a complete-decision measurement (P_Φ, t) does select a pure state: the probability for the positive outcomes of the measurements $(P_f, t_f), \dots, (P_n, t_n)$ after a positive outcome for (P_Φ, t) is, according to (4.6),

$$\frac{\text{Tr} (P_\Phi Q_{\text{out}}^\dagger Q_{\text{out}} P_\Phi)}{\text{Tr} P_\Phi^2} = (\phi, Q_{\text{out}}^\dagger Q_{\text{out}} \phi). \quad (4.8)$$

If the measurement (P_Φ, t) is preceded by the measurement (P_0, t') with a positive result, the probability for a positive result of the same succession of output measurements becomes

$$\frac{\text{Tr} (P_0 P_\Phi Q_{\text{out}}^\dagger Q_{\text{out}} P_\Phi P_0)}{\text{Tr} P_0 P_\Phi P_\Phi P_0} = \frac{(\phi, Q_{\text{out}}^\dagger Q_{\text{out}} \phi) (\phi, P_0 \phi)}{(\phi, P_0 \phi)}. \quad (4.8a)$$

This is the same as the probability (4.8) obtained for the case that only (P_Φ, t) was observed to give a positive result. In order to obtain the right-hand side of (4.8a), use was made of the invariance of the trace under a cyclic interchange of the factors, of the equation $P_0^2 = P_0$, and of the general formula

$$\text{Tr} P_\Phi A P_\Psi B P_X \dots P_\Omega E = (\phi, A \psi) (\psi, B \chi) \dots (\omega, E \phi) \quad (4.9)$$

which follows from (4.2) directly.

In order to show the converse, that there is a complete decision observation which selects an arbitrary pure state, it is necessary to make an assumption about the set of complete decisions (decisions represented by trace 1 projection operators). We shall assume that for every finite time interval *there exist two sets of complete decisions* $\{(P_\Psi, t_+)\}$ and $\{(P_\Phi, t_-)\}$ so that the decision (P_Ψ, t_+) can be performed later than the time interval under consideration whereas the (P_Φ, t_-) can be performed earlier. We shall further assume that the set of c-d operators (measurable trace 1 operators) $\{P_\Phi\}$ contains a sufficient set of operators so that if $\text{Tr} P_\Phi M = (\phi, M \phi) = 0$ for all c-d operators P_Φ then $M = 0$; a set of operators which satisfies this condition will be called operator complete. *The same assumption is made for the set of c-d operators* $\{P_\Psi\}$.

Let us assume that the measurements $(P_a, t_a), (P_b, t_b), \dots, (P_e, t_e)$ prepare a pure state; these measurements will be assumed to be performed during a finite time interval. We shall now use a (P_Ψ, t_+) as output observation, i.e., let us substitute P_Ψ for $P_n \dots P_f = Q_{\text{out}}$. The fact that the measurement of $P_\Phi, (P_\Phi, t_-)$,

performed prior to the measurements $(P_a, t_a), \dots, (P_e, t_e)$ does not alter the probability of a positive outcome for (P_Ψ, t_+) becomes, using (4.6),

$$\frac{\text{Tr} Q_{\text{in}}^\dagger P_\Psi P_\Psi Q_{\text{in}}}{\text{Tr} Q_{\text{in}}^\dagger Q_{\text{in}}} = \frac{\text{Tr} P_\Phi Q_{\text{in}}^\dagger P_\Psi P_\Psi Q_{\text{in}} P_\Phi}{\text{Tr} P_\Phi Q_{\text{in}}^\dagger Q_{\text{in}} P_\Phi} \quad (4.10)$$

or, by cyclically interchanging the factors in the traces,

$$\frac{\text{Tr} P_\Psi Q_{\text{in}} Q_{\text{in}}^\dagger}{\text{Tr} Q_{\text{in}} Q_{\text{in}}^\dagger} = \frac{\text{Tr} P_\Psi Q_{\text{in}} P_\Phi Q_{\text{in}}^\dagger}{\text{Tr} P_\Phi Q_{\text{in}}^\dagger Q_{\text{in}} P_\Phi}. \quad (4.10a)$$

Since this is valid for all c-d operators P_Ψ [representing a complete decision (P_Ψ, t_+)], it follows from our assumption that the P_Ψ are operator complete

$$\frac{Q_{\text{in}} Q_{\text{in}}^\dagger}{\text{Tr} Q_{\text{in}} Q_{\text{in}}^\dagger} = \frac{Q_{\text{in}} P_\Phi Q_{\text{in}}^\dagger}{(\phi, Q_{\text{in}}^\dagger Q_{\text{in}} \phi)} \quad (4.11)$$

because the two sides of (4.10a) are the traces of P_Ψ times the operators in (4.11).

The preceding calculation assumed that $Q_{\text{in}} \neq 0$, and that Φ was so chosen that $\text{Tr} P_\Phi Q_{\text{in}}^\dagger Q_{\text{in}} \neq 0$. Since $\text{Tr} P_\Phi Q_{\text{in}}^\dagger Q_{\text{in}} = (\phi, Q_{\text{in}}^\dagger Q_{\text{in}} \phi) = (Q_{\text{in}} \phi, Q_{\text{in}} \phi)$, it follows that $Q_{\text{in}} \phi \neq 0$. Applying the right-hand side of (4.11) to any vector f , one finds, using (4.2) to calculate $P_\Phi Q_{\text{in}}^\dagger f$, that the result will be proportional to $Q_{\text{in}} \phi$. It follows that the Hermitean operator (4.11) can have only one characteristic vector with a nonzero characteristic value and that this vector, to be denoted by χ , is a multiple of $Q_{\text{in}} \phi$. Hence, the operator (4.11) is the multiple of a trace 1 projection operator P_χ and since the left-hand side of (4.11) shows that its trace is 1, it is equal to P_χ

$$Q_{\text{in}} Q_{\text{in}}^\dagger / \text{Tr} Q_{\text{in}} Q_{\text{in}}^\dagger = P_\chi. \quad (4.12)$$

Incidentally, as has been shown by von Neumann, if $Q Q^\dagger \phi = (\chi, \phi) \chi$ for all ϕ , then $Q^\dagger \phi = (\chi, \phi) Q^\dagger \chi$ for all ϕ and $Q \eta = (Q^\dagger \chi, \eta) \chi$ for all η . These equations explain why it was immaterial which ϕ was used in the preceding proof as long as $Q_{\text{in}} \phi \neq 0$.

Let us return now to (4.6). Assuming that the input prepares a pure state, it follows that this can be given the form

$$\begin{aligned} \Pi((P_a, t_a), \dots, (P_e, t_e) | (P_f, t_f), \dots, (P_n, t_n)) \\ = \text{Tr} (Q_{\text{out}}^\dagger Q_{\text{out}} Q_{\text{in}} Q_{\text{in}}^\dagger) / \text{Tr} (Q_{\text{in}} Q_{\text{in}}^\dagger) \\ = \text{Tr} (Q_{\text{out}}^\dagger Q_{\text{out}} P_\chi). \end{aligned} \quad (4.6c)$$

Thus, under our assumptions which are italicized, it is indeed true that the pure states, as defined in the Introduction to this chapter, have the same properties with respect to future measurements as those which are characterized by the positive outcome of a complete decision (P_χ, t) , i.e., which can be characterized by a state vector χ .

The preceding argument does not show, of course, that the projection operator $Q_{\text{in}} Q_{\text{in}}^\dagger [\text{Tr} Q_{\text{in}} Q_{\text{in}}^\dagger]^{-1} = P_\chi$ is a c-d operator, that is, corresponds to a complete-

decision (P_X, t), which can be performed instantaneously. In fact, it is not clear at what time it might be measurable since we recall that the measurement of an operator involves different actions depending on the time at which the measurement is carried out (Heisenberg picture). We will assume, however, that the state χ can be considered a state *suspae*, i.e., that it could have been prepared prior to any given time.

4.6. Superselection Rules²⁹

It was assumed in the preceding section that the manifold of measurable trace 1 (or rank 1) projection operators P_Φ is so great that if $\text{Tr } P_\Phi M = (\Phi, M\Phi) = 0$ for all of them, $M=0$. There are situations known in which this condition is not satisfied; the limitations on the Φ for which P_Φ can be measurable are called superselection rules. The superselection rule for electric charge postulates that every observable self-adjoint operator commute with the charge operator. This conclusion was arrived at by analyzing the measurement process. A similar rule holds for the baryon number and, probably, also for the lepton number. One can define an operator V which leaves every state vector representing a state with integer spin unchanged and multiplies the state vectors for half-integer spin by -1 . In a sense, this is the operator of rotation by 2π . It can be deduced from the postulate of the Lorentz invariance of the theory that all measurable operators must commute with this "singlevaluedness" operator. There are similar rules for the so-called type operators.³⁰ All these limitations restrict the set of operators which are measurable and it is well to discuss the consequences of these limitations on the conclusions of the preceding and following sections.

Let us denote the superselection operators by S_1, S_2, \dots . All measurable operators commute with them. We assume that they all have discrete spectra, that they are all measurable at any time, and that such a measurement does not conflict with the measurement of any other quantity. They must then all commute with each other. If the time-displacement operator is measurable, it commutes with them also so that they are independent of time; at any rate we assume that they are. These assumptions are all valid for the superselection operators which we have enumerated. Because of these conditions one can define subspaces of the Hilbert space which will be called coherent subspaces. All the vectors v of the subspace (s_1, s_2, \dots) are characteristic vectors of all S_i , the characteristic value of S_i is s_i so

that $S_i v = s_i v$. There is a coherent subspace for every combination s_1, s_2, \dots of all superselection operators. A vector ϕ , and the corresponding projection operator P_Φ , are said to belong to a coherent subspace if ϕ is a vector in that subspace.

If S is a superselection operator, such as the aforementioned charge operator, $SP_\Phi = P_\Phi S$ must hold for any measurable P_Φ . Applying this to ϕ , one finds $S\phi = P_\Phi S\phi$, i.e., $S\phi$ is a characteristic vector of P_Φ with characteristic value 1. It must be, therefore, parallel to ϕ so that ϕ is also a characteristic vector of S . Since this holds for each S , we conclude that the rays of observable projection operators lie in a single coherent subspace. The same applies to states which can be selected by a succession of measurements, such as the $(P_\alpha, t_\alpha), \dots, (P_e, t_e)$ considered in the preceding section. For an arbitrary measurable operator m it follows from $mS_i = S_i m$ that they transform every vector of a coherent subspace into a vector of the same subspace.

As far as the preceding section is concerned, the principal point that needs elaboration is the one where it was postulated that if $\text{Tr } P_\Psi M = (\Psi, M\Psi) = 0$ for all measurable projection operators P_Ψ , then $M=0$. Actually, because of the limitation on P_Ψ , M need not be zero but could be any operator which transforms each vector of a coherent subspace into a vector of a different subspace. However, the postulate was applied to an M which was the product of measurable operators, i.e., operators which leave all coherent subspaces invariant. For such M the postulate does not conflict with the limitations set by the superselection rules.

The preceding discussion reflects our knowledge on the limitations on the measurability of operators. This is an incomplete knowledge, and it is generally believed that the measurability of most operators is open to question. It is, in fact, not very likely that the limitations on measurability can be all formulated in terms of superselection operators with which all measurable operators commute. However, it is also believed that no incorrect conclusion will be arrived at by assuming the measurability of all self-adjoint operators which commute with all superselection operators.

4.7. The Mappings of the Set of Observations on Itself Which Leave the II Function Invariant

In Sec. 2.4 we have defined the group G of mappings of observations upon observations which leave the II function invariant. The group G maps all observations upon observations, but since all observations can be represented as sets of decision observations, it will suffice to consider the mappings of these observations. Decision observations can have only two results: 1 or 0 and this must hold also for their image if the mapping is to leave II invariant. Since the condition of having only 1 and 0 as possible results is both a necessary and a sufficient condition for the measurement to be repre-

²⁹ G. C. Wick, A. S. Wightman, and E. P. Wigner, *Phys. Rev.* **88**, 101 (1952); A. S. Wightman, *Nuovo Cimento Suppl.* **14**, 81 (1959); J. M. Jauch, *Helv. Phys. Acta* **33**, 711 (1960); A. Galindo, A. Morales, and R. Nuñez-Lagos, *J. Math. Phys.* **3**, 324 (1961). See also Ref. 27.

³⁰ This is discussed further in Sec. 4.10. See also "Unitary Representations of the Inhomogeneous Lorentz Group Including Reflections," by E. P. Wigner, *Summer Conference on Theoretical Physics, Istanbul, 1962* (Gordon and Breach, Inc., New York, 1965), p. 37.

sented by a projection operator, all elements of G will map projection operators on projection operators.

A decision observation can be characterized by two quantities: the self-adjoint projection operator P_a which corresponds thereto, and the time at which it is carried out. Hence, the mapping will be described by two functions: the first of these is "projection valued" and gives the operator $P_{\bar{a}}$ of the measurement which is the image of (P_a, t_α) , the second the time \bar{t}_α of the image measurement. The two functions will be noted by $\bar{a}(a, t_\alpha)$ and $\bar{t}_\alpha(a, t_\alpha)$. Both a and \bar{a} will refer to projections.

We shall carry out the determination of the functions \bar{a} and \bar{t}_α in four steps. First we show that $\bar{a}(a, t_\alpha)$ does not actually depend on t_α . This means that every transformation which belongs to the group G maps not only decision measurements into decision measurements, but also gives a unique map of the set of observable projection operators upon itself. Second, we show that all complete decisions (P_Ψ, t) , the operators P_Ψ of which belong to one coherent subspace, will be mapped into complete decisions the operators of which belong to a single coherent subspace. Third, we conclude that the one-to-one mapping of the set of observable projection operators upon itself, obtained in the first step, can be expressed by $P_{\bar{a}} = OP_aO^{-1}$ for all P_a . The operator O does not depend on a , is essentially unique and either unitary or antiunitary. Fourth, and last, we show that the time ordering of any three mutually noncommuting measurements is either the same or the opposite of that of their images.

In order to arrive at these conclusions it is necessary to make certain assumptions concerning the set of decision observations. One sufficient assumption is that *the set of c-d operators contains, in each coherent subspace, a vector-complete set of mutually orthogonal trace 1 projection operators* $\{P_m^s\}$, i.e., a set the rays Φ_m^s of the members of which contain a complete orthonormal set of vectors ϕ_m^s in the coherent subspace s . *It should, furthermore, contain four further sets* $\{P_{1+m}^s\}$, $\{P_{1+\omega m}^s\}$, $\{P_{1+m+n}^s\}$, $\{P_{1+\omega m+n}^s\}$. Hence, the trace 1 projection operators, the rays of which contain the vectors

$$\phi_m^s, \phi_{1+m}^s, \phi_{1+\omega m}^s, \phi_{1+m+n}^s, \phi_{1+\omega m+n}^s \quad (4.13)$$

(in which ω is a fixed complex number, $\omega \neq \pm 1$, which can depend on s), are assumed to be measurable for all m, n where the ϕ_m^s form a complete orthonormal set in the coherent subspace s . Let us note that these sets are, together, operator complete in the sense that $M=0$ if $\text{Tr } PM=0$ for all P , at least if M leaves the coherent subspaces invariant. Finally, we must assume that the *measurements of the c-d operators represented by (4.13) can be performed at different times* so that expressions such as $\text{Tr } P_m P_{1+m}$ are observable. This innocent looking assumption is quite severe in the Heisenberg

representation which we use. Suppose that (P_m, t) is a decision whether the particle is at x_m at time t , then the actions which one has to perform in order to carry out the measurement (P_m, t') will be different and even unknown in general. The last assumption is satisfied in an approximate sense if we perform the observations in rapid succession.

(i) *The first step.* The operator of the image $\bar{a}(a, t_\alpha)$ is independent of the time t_α of the original observation.

Let us first establish the very plausible fact that the image of a complete-decision observation is again a complete-decision observation. We have already seen that the image of a decision observation is a decision observation (because the outcome can be only 1 or 0) and that the operator of a c-d observation is a trace 1 projection operator (Sec. 4.5). Let us denote by P_a the operator the image of which is the c-d operator P_Φ and let us denote the image of the c-d operator P_Ψ by P_b . Then, using P_a as input, P_Ψ as output, the invariance of Π gives, because of (4.6b),

$$\frac{\text{Tr } P_a P_\Psi}{\text{Tr } P_a} = \frac{\text{Tr } P_\Phi P_b}{\text{Tr } P_\Phi} = \text{Tr } P_\Phi P_b. \quad (4.14)$$

On the other hand, using P_Ψ as input, P_a as output, one obtains

$$\text{Tr } P_\Psi P_a = \frac{\text{Tr } P_b P_\Phi}{\text{Tr } P_b}. \quad (4.15)$$

Since the traces are independent of the orders of the factors, if none of the traces vanishes one obtains

$$\text{Tr } P_a \text{Tr } P_b = 1. \quad (4.16)$$

However, the traces of projection operators are positive integers so that (4.16) can be valid only if

$$\text{Tr } P_a = \text{Tr } P_b = 1, \quad (4.17)$$

i.e., if P_a and P_b are trace 1 projection operators. Hence, the images of c-d measurements are again c-d measurements. Since only two measurements entered the preceding argument, it was not necessary to consider their time ordering; the Π function is independent therefrom. Since a complete decision (P_Ψ, t) , as input, specifies a ray of state vectors Ψ we have in the usual language of quantum mechanics that the elements of G map state vectors into state vectors rather than into "mixtures." Let the map of (P_Ψ, t) be (\bar{P}_Ψ, \bar{t}) , then $\bar{P}_\Psi = P_{\bar{\Psi}}$ for a suitable $\bar{\Psi}$.

In order to establish the independence of the operator of the image $\bar{a}(a, t_\alpha)$ of the time t_α , it is only necessary to use a complete-decision observation (P_Ψ, t) as input and an arbitrary decision observation (P_a, t_α) as output. The time of this may be later or earlier than t . The invariance of Π gives

$$\text{Tr } P_\Psi P_a = \text{Tr } P_{\bar{\Psi}} P_{\bar{a}}. \quad (4.18)$$

The left-hand side is independent of t_α and this applies therefore also to the right-hand side. The $P_{\bar{\Psi}}$ include

all the projection operators with the rays (4.13), because these are observable and therefore occur as images of observable c-d observations. Hence, the traces of $P_{\bar{\Psi}}M$ completely determine an observable M , and the conclusion follows. Equation (4.18) also shows that $P_{\bar{\Psi}}$ is independent of t . It follows that we can replace $\bar{a}(a, t_a)$ by $\bar{a}(a)$.

(ii) *The second step.* The images of the complete decisions in one coherent subspace are complete decisions in a single coherent subspace.

The first step established that if (P_{Ψ}, t) can be performed for several values of t , the image $\bar{\Psi}$ of the ray Ψ does not depend on the time t of the measurement (P_{Ψ}, t) . Since $P_{\bar{\Psi}}$ is, by the definition of the group G , observable, the observable Ψ lies entirely in one coherent subspace, say s' .

If P_{Φ} is in the coherent subspace of Ψ , then it is always possible to find an observable P_X such that both $\text{Tr } P_{\Psi}P_X = |\langle \psi, \chi \rangle|^2 > 0$, $\text{Tr } P_{\Phi}P_X = |\langle \phi, \chi \rangle|^2 > 0$. If the expansion of ψ and ϕ in terms of the ϕ_n of (4.13) both contain at least one ϕ_n with nonvanishing coefficient, this ϕ_n can play the role of χ . If this is not the case, any $\phi_1 + \phi_n + \phi_m$ will do if ϕ_n occurs in the expansion of ψ and ϕ_m in the expansion of ϕ . If, on the other hand, ψ and ϕ are in different coherent subspaces, there can be no observable P_X such that both $\text{Tr } P_{\Psi}P_X$ and $\text{Tr } P_{\Phi}P_X$ be different from zero since X can not be in two different coherent subspaces. Actually the word "coherent" was intended to express just this fact: that if two vectors ψ and ϕ are in the same coherent subspace then there is a c-d operator P_X , a "realizable state" χ , such that both $\langle \psi, \chi \rangle$ and $\langle \phi, \chi \rangle$ are nonzero. The χ then in a sense ties ψ and ϕ together.

In order to prove that the images $P_{\bar{\Psi}}$ and $P_{\bar{\Phi}}$ of the c-d operators P_{Ψ} and P_{Φ} are in the same coherent subspace, we have to find only a c-d operator P_Z such that the traces of $P_{\bar{\Psi}}P_Z$ and of $P_{\bar{\Phi}}P_Z$ be both nonzero. The image $P_{\bar{X}}$ of P_X can serve as this P_Z because then

$$\text{Tr } P_{\Psi}P_X = \text{Tr } P_{\bar{\Psi}}P_{\bar{X}}; \text{Tr } P_{\Phi}P_X = \text{Tr } P_{\bar{\Phi}}P_{\bar{X}}. \quad (4.19)$$

Equation (4.19) follows if one calculates the probability $\Pi((P_X, t') | (P_{\Psi}, t))$ that the measurement (P_X, t') give 1 if the measurement (P_{Ψ}, t) gave 1. According to (4.6a) this probability is

$$\text{Tr } P_{\Psi}P_X = |\langle \psi, \chi \rangle|^2. \quad (4.20)$$

It is usually called transition probability between Ψ and X . Since Π remains unchanged if one replaces (P_{Ψ}, t) and (P_X, t') by their images $(P_{\bar{\Psi}}, \bar{t})$ and $(P_{\bar{X}}, \bar{t}')$, the first of Eqs. (4.19) follows, the second is derived in a similar way. One also sees that if Ψ and Φ are in different subspaces, the same holds for their images. In this case there is no c-d operator P_X so that both $\text{Tr } P_X P_{\Psi}$ and $\text{Tr } P_X P_{\Phi}$ be nonzero. Hence since all c-d operators P_Z are images of some P_X , there can be no P_Z such that both $\text{Tr } P_Z P_{\bar{\Psi}} > 0$, and $\text{Tr } P_Z P_{\bar{\Phi}} > 0$.

The two important elements of the preceding argument were the phenomenological criterion for the coherence of two rays, in the form of a c-d operator into the ray of which both rays have a finite transition probability, and the invariance of this transition probability under the elements of G .

(iii) *The third step.* The mappings of G define a unitary or antiunitary transformation.

The main part of this theorem is of purely mathematical nature. It presupposes the suspaie nature of the states which are described by the rays (4.13) or, equivalently, that the measurements (P_n, t) , (P_{m+n}, t) , etc. can be carried out at different times. Under this assumption, the transition probabilities between these states, and more generally, all transition probabilities into these states, become observable. As a result, if ϕ and ϕ' are vectors of any two rays of the set (4.13), and $\bar{\phi}$ and $\bar{\phi}'$ vectors with the same normalization contained in the rays of the image operators of the rays Φ, Φ' , then

$$|\langle \phi, \phi' \rangle|^2 = |\langle \bar{\phi}, \bar{\phi}' \rangle|^2. \quad (4.21)$$

The left-hand side is $\Pi((P_{\Phi'}, t') | (P_{\Phi}, t))$, the right-hand side $\Pi((P_{\bar{\Phi}'}, \bar{t}') | (P_{\bar{\Phi}}, \bar{t}))$.

Let us consider, first, the rays in the coherent subspace s ; these will be mapped into a single coherent subspace s' which may or may not be identical with s . If P_{Φ} is a c-d operator and $P_{\bar{\Phi}}$ its image, we shall say also that the ray $\bar{\Phi}$ is mapped into the ray of $\bar{P}_{\Phi} = P_{\bar{\Phi}}$ or that the ray $\bar{\Phi}$ of $P_{\bar{\Phi}}$ is the image of Φ . In this way, we obtain also a mapping of the observable rays of s into the observable rays of s' . However, for the time being, we shall omit the superscripts s and s' , it being understood that the Φ are in s , the $\bar{\Phi}$ in s' . Let us choose an arbitrary vector ϕ_n of norm 1 from the ray Φ_n (more precisely, from the ray Φ_n^s) and a similar vector $\bar{\phi}_n$ from the ray $\bar{\Phi}_n$ of the image \bar{P}_n of P_n . The ϕ_n then will form a complete orthonormal set, and, because of (4.21) and their normalization, the $\bar{\phi}_n$ will be also orthonormal. They also will be a complete set because if, for instance,

$$\sum_n \text{Tr } P_m^{s'} \bar{P}_n = \sum_n |\langle \phi_m^{s'}, \bar{\phi}_n \rangle|^2$$

were smaller than 1, the same would hold for

$$\sum_n \text{Tr } \hat{P}_m^{s'} P_n,$$

where $\hat{P}_m^{s'}$ is the projection operator the image of which is $P_m^{s'}$. Hence the $\phi_m^{s'}$ can be expanded in terms of the $\bar{\phi}_n$, and since these form a complete set, the same will apply for any vector of the coherent subspace s' .

Let us now consider an observable ray in the subspace s , and let one of its vectors be $\phi = \sum a_n \phi_n$. We shall expand a vector $\bar{\phi}$ belonging to the map of the ray of ϕ , in terms of the $\bar{\phi}_n$ and denote this by $\sum \bar{a}_n \bar{\phi}_n$.

Assuming that the norm of this latter vector is the same as that of ϕ , we can set $\phi' = \phi_m$ in (4.21) and ob-

tain $|a_n| = |\bar{a}_n|$. Hence we have, in general,

$$\text{ray of } \sum a_n \phi_n \leftrightarrow \text{ray of } \sum a_n \zeta_n \bar{\phi}_n, \quad \text{where } |\zeta_n| = 1, \tag{4.22}$$

and the ζ can yet depend on the coefficients a_n .

In particular, the ray of $\phi_1 + \phi_n$ is mapped into the ray of $\zeta_1 \bar{\phi}_1 + \zeta_n \bar{\phi}_n$. We now define the operator O by stipulating that it map

$$O\phi_1 = \bar{\phi}_1, \quad O\phi_n = (\zeta_n/\zeta_1)\bar{\phi}_n.$$

If we change the notation and denote $(\zeta_n/\zeta_1)\bar{\phi}_n$ by $\bar{\phi}_n$, all preceding equations will be valid and the last one will assume the form

$$O\phi_n = \bar{\phi}_n \tag{4.23}$$

whereas the map of the ray of $\phi_1 + \phi_n$ will contain $\bar{\phi}_1 + \bar{\phi}_n$. There is some arbitrariness in the definition of $O\phi_1$ but this corresponds to the arbitrariness in O : it can be multiplied by a factor of norm 1 and $OP_aO^{-1} = \bar{P}_a$ will remain valid. It is this factor which is fixed by postulating $O\phi_1 = \bar{\phi}_1$. Apart from this, O is completely determined and if our assertion is correct it must follow from (4.21) that if $\sum a_n \phi_n$ is in an observable ray, either $\sum a_n \bar{\phi}_n$ or $\sum a_n^* \bar{\phi}_n$ is in the ray of the image of $\sum a_n \phi_n$. Even more generally, $OP_aO^{-1} = \bar{P}_a$ must follow.

We shall prove these assertions for the $\sum a_n \phi_n$ which are members of the set (4.13). It follows from (4.22) that the image of the ray of the first column of Table I contains a vector of the second column where the σ, ρ, ν are all of absolute value 1. In addition, as a result of the altered meaning of $\bar{\phi}_n$, the image of the ray of $\phi_1 + \phi_n$ contains $\bar{\phi}_1 + \bar{\phi}_n$. Table II shows the conclusions one arrives at by using for ϕ and ϕ' in (4.21) the expressions given in the first two columns thereof. It follows that indeed all ν and hence all s_n are equal. If the latter are 1 the images of the rays of (4.13) contain a vector which is obtained from (4.23) assuming that this is linear, if the s are equal to -1 , the images of the rays of (4.13) contain the vectors obtained from (4.23) assuming that O is antiunitary. In either case, O as applied to these vectors will be defined accordingly and we can indeed extend the definition domain of O to all vectors and write in the former case ($\nu_n = \omega$)

$$O \sum a_n \phi_n = \sum a_n \bar{\phi}_n, \tag{4.24}$$

whereas in the latter case ($\nu_n = \omega^{-1} = \omega^*$) we define

$$O \sum a_n \phi_n = \sum a_n^* \bar{\phi}_n. \tag{4.24a}$$

TABLE I.

Ray of	Image contains
$\phi_1 + \phi_n$	$\bar{\phi}_1 + \bar{\phi}_n$
$\phi_1 + \omega \phi_n$	$\bar{\phi}_1 + \nu_n \bar{\phi}_n$
$\phi_1 + \omega \phi_m + \phi_n$	$\bar{\phi}_1 + \sigma_{mn} \bar{\phi}_m + \rho_{mn} \bar{\phi}_n$

TABLE II.

ϕ	ϕ'	Conclusion
$\phi_1 + \phi_n$	$\phi_1 + \omega \phi_n$	$\nu_n = \omega^{s_n}; s_n = \pm 1$
$\phi_1 + \phi_n$	$\phi_1 + \omega \phi_m + \phi_n$	$\rho_{mn} = 1$
$\phi_1 + \omega \phi_m$	$\phi_1 + \omega \phi_m + \phi_n$	$\sigma_{mn} = \nu_m$
$\phi_1 + \omega \phi_n + \phi_m$	$\phi_1 + \omega \phi_m + \phi_n$	$\nu_m = \nu_n$

The unitary and antiunitary nature of the operators (4.24) and (4.24a) is easily verified.

For the projection operators P_Φ of (4.13),

$$\bar{P}_\Phi \chi = P_{\bar{\Phi}} \chi = (\bar{\phi}, \chi) \bar{\phi} = (O\phi, \chi) O\phi = O(\phi, O^{-1}\chi) \phi \tag{4.25}$$

follows for an arbitrary χ . The last step needs a word of explanation: if O is unitary $(O\phi, \chi) = (\phi, O^{-1}\chi)$ and the other O can be brought in front of the factor $(\phi, O^{-1}\chi)$. If O is antiunitary, $(O\phi, \chi) = (\phi, O^{-1}\chi)^*$, the interchange of O and this factor changes the factor to its conjugate complex. On the other hand,

$$OP_\Phi O^{-1} \chi = O(\phi, O^{-1}\chi) \phi \tag{4.25a}$$

so that indeed for all Φ of (4.13)

$$\bar{P}_\Phi = OP_\Phi O^{-1}. \tag{4.25b}$$

We yet have to show that $\bar{P}_a = OP_a O^{-1}$ holds for an arbitrary observable operator. Using P_Φ with one of the Φ of (4.13) as input and P_a as output, (4.6b) becomes for the original and image measurements

$$\text{Tr } P_\Phi P_a = \text{Tr } \bar{P}_\Phi \bar{P}_a. \tag{4.26}$$

Introduction of (4.25b) gives

$$\text{Tr } P_\Phi P_a = \text{Tr } OP_\Phi O^{-1} \bar{P}_a = \text{Tr } P_\Phi O^{-1} \bar{P}_a O,$$

and since the P_Φ are operator-complete,

$$P_a = O^{-1} \bar{P}_a O \quad \text{or} \quad \bar{P}_a = OP_a O^{-1} \tag{4.27}$$

The preceding argument applies only to projection operators P_a which transform only vectors of the coherent subspace s into other vectors in s but give zero if applied to vectors of other coherent subspaces. These P_a are mapped by the elements of G into the $OP_a O^{-1}$ which transform the vectors of the coherent subspace s' (containing the $\bar{\phi}$) into similar vectors: O^{-1} brings the vectors of s' into s , P_a leaves them there and O returns them to s' . A general projection operator will not be zero in all but one coherent subspace. However, it can be decomposed into such operators and (4.27) will hold for each part. There is no apparent relation between the parts, except that the mapping of coherent subspaces into coherent subspaces must be one-to-one for every element of G . There seems to be no reason even that all parts of a mapping be unitary or all anti-unitary but, so far, mixed mappings have played no role in the theory. There is also a factor arbitrary in each part, i.e., the operator O which corresponds to

an element of G contains an arbitrary factor for each coherent subspace.

This completes the third step of our discussion which is, of course, only a slight modification of similar discussions given before.³¹ The unsatisfactory nature of the proof is the arbitrariness of the set (4.13). Except for its nature to be operator-complete, there is no real reason to choose this set rather than any other. The arbitrariness could be avoided if one knew a sufficiently large set of operators which was, in fact, observable. The lack of our knowledge in this regard was already emphasized as a weakness of the conceptual structure of quantum mechanics. It is interesting to note, however, the similarity between the set (4.13) and the set of quantum-mechanical states described in Sec. 4.1; both show the greater wealth, and the more intricate structure, of the set of states of quantum theory as compared with the classical theories.

It may be useful to remark, finally, that the preceding theorem applies not only to the transformations of the group G but also to every redescription such as accompanying the use of a different language or the use of a different Hilbert space to describe the observables.

(*iiii*) *The fourth step.* Actions with noncommuting operators and images are either in the same, or in opposite time order.

We have, so far, considered the Π function almost exclusively with a single-input and a single-output observation. In this case the time order of the observations is immaterial: the symmetry of (4.6b) shows that the probability connection is the same even if the time order of input and output is interchanged. This is not so if the Π function refers to several observations.

Let us observe first that the condition (*iiii*), together with the conditions derived earlier, are not only necessary but also sufficient conditions for the mapping to leave Π invariant. Indeed, the expression (4.7) is clearly invariant if every P therein is replaced by OPO^{-1} as long as O is unitary because the intermediate O drop out and the first O can be shifted to the end in a trace if it is linear. If O is antiunitary it can be decomposed into the product of a unitary operator and complex conjugation. One easily convinces oneself that the unitary operators cancel as before, and the operation of complex conjugation converts the operators the traces of which appear in (4.7) into their complex conjugates. However, operators QQ^\dagger are Hermitian and their diagonal elements therefore real. If O is unitary when applied to some of the coherent subspaces, antiunitary when applied to others, its unitary part can be applied first, and this followed by complex con-

jugation in some of the coherent subspaces. The former operation was considered already and leaves Π invariant but the same applies to the latter because the diagonal elements are all real. Similarly interchanging the time ordering means, for all Q in (4.7), transition to the Hermitian adjoint and this remains without effect as far as the trace is concerned. Naturally, this does mean that all mappings $(P_a, t_a) \leftrightarrow (\bar{P}_a, \bar{t}_a)$ of the type considered are elements of G .

Condition (*iiii*) is natural physically. Mathematically, also, it would be very surprising if an order of the P , different from the initial or the opposite, would have the expression (4.7) unchanged, considering the large number of possible P that have been postulated. Hence, the proof of (*iiii*) will not be given in full detail. It should be possible, nevertheless, to reproduce it from the description which follows.

One first considers an input measurement which is one of the complete measurements P_Φ postulated in (4.13), and as output measurements an arbitrary P_a followed by a complete measurement $P_{\Phi'}$ [postulated in (4.13)]. The time order is that given. Choosing suitable \bar{P}_n for the complete measurements, one can show that the time order of the maps $\bar{P}_\Phi, \bar{P}_a, \bar{P}_{\Phi'}$ must be either this, or the opposite, unless P_a commutes with all P_n . If this is the case, one replaces the input measurement with a suitable P_{n+m} and finds that (unless $P_a=0$ or 1) the order must be $\bar{P}_\Phi, \bar{P}_a, \bar{P}_{\Phi'}$ in that case also. It is essential for this part of the proof that the P_Φ of (4.13) can be observed both before and after any P_a .

One next considers three complete measurements of the (4.13) type and proves that if the time ordering among any pair is preserved by the mapping, the same will be true for any other pair. This is, perhaps, the most lengthy part of the proof. It then follows, by means of the result of the preceding paragraph, that the same is true for any pair P, P_a one of which is still of the nature postulated in (4.13). As a result, if all (P_Φ, t_0) of (4.13) precede a set of measurements $(P_a, t_a), (P_b, t_b), \dots$, the images $(\bar{P}_\Phi, \bar{t}_0)$ will either all precede or will all succeed all the images $(\bar{P}_a, \bar{t}_a), (\bar{P}_b, \bar{t}_b), \dots$. If one is willing to postulate this, the discussions of the last two paragraphs are unnecessary.

Finally, one considers (P_Φ, t_0) as input, two arbitrary $(P_a, t_a), (P_b, t_b)$ as output, $t_0 < t_a < t_b$. It follows from (4.7) for the case in which the time ordering of P_Φ, P_a and of P_Φ, P_b is preserved that

$$\text{Tr } P_\Phi P_a P_b P_a P_\Phi = \text{Tr } \bar{P}_\Phi \bar{P}_a \bar{P}_b \bar{P}_a \bar{P}_\Phi \text{ or } \text{Tr } \bar{P}_\Phi \bar{P}_b \bar{P}_a \bar{P}_b \bar{P}_\Phi \quad (4.28)$$

depending whether or not the time ordering between P_a and P_b is also preserved, i.e., whether $\bar{t}_a < \bar{t}_b$ or $\bar{t}_a > \bar{t}_b$. The last P_Φ in all expressions (4.28) can be made the first factor and disappears thereupon. Because the \bar{P} are transforms $\bar{P} = OPO^{-1}$ of the P , they

³¹ E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959), p. 251; R. M. F. Houtappel's notes of the lectures of E. P. Wigner, Leiden, 1959 (unpublished); V. Bargmann, *J. Math. Phys.* **5**, 862 (1964); L. O'Raifeartaigh and G. Rasche, *Ann. Phys. (N. Y.)* **25**, 155 (1963). See Bargmann's article for additional references.

can be replaced by the latter in a trace. This then gives

$$\begin{aligned} \text{Tr } P_\Phi P_a P_b P_a &= \text{Tr } P_\Phi P_a P_b P_a \quad \text{for } t_\alpha < t_\beta \\ &= \text{Tr } P_\Phi P_b P_a P_b \quad \text{for } t_\beta < t_\alpha. \end{aligned} \quad (4.28a)$$

Since there is an operator-complete set of P_Φ available, it is possible to conclude, in the case of the second alternative, that

$$P_a P_b P_a = P_b P_a P_b. \quad (4.29)$$

This can be true, however, only if P_a and P_b commute, in which case \bar{P}_a and \bar{P}_b also commute and their time order is irrelevant in the case considered. Had we assumed that the time ordering between P_Φ and P_a or P_b is reversed by the mapping, (4.29) would have followed if the time ordering between P_b and P_a were not reversed.

That $P_a P_b = P_b P_a$ follows from (4.29) can be seen most easily by noting

$$\begin{aligned} (P_a P_b - P_b P_a)^2 &= P_a P_b P_a P_b - P_a P_b P_a - P_b P_a P_b + P_b P_a P_b P_a \\ &= P_a (P_b P_a P_b - P_a P_b P_a) - P_b (P_b P_a P_b - P_a P_b P_a). \end{aligned}$$

Hence, the square of $P_a P_b - P_b P_a = C$ vanishes if (4.29) is fulfilled. It follows for any vector ψ that $(\psi, C^2 \psi) = 0$ and since C is skew Hermitian that $-(C\psi, C\psi) = 0$ or that $C\psi = 0$.

This does not quite complete the proof of (iiii) not only because the case of commuting (P_a, t_α) , (P_b, t_β) should be further considered but also because the existence of superselection rules has not been adequately discussed. However, we leave the completion of the discussion to the reader.

We conclude the section by restating the result. A mapping of decision measurements (P_a, t_α) on decision measurements $(\bar{P}_a, \bar{t}_\alpha)$ will leave the Π function invariant if and only if the two conditions apply: (a) The operator which represents $(\bar{P}_a, \bar{t}_\alpha)$ is $\bar{P}_a = O P_a O^{-1}$, i.e., is a transform of the operator of the original measurement. O is a unitary or antiunitary mapping of every coherent subspace on a coherent subspace such that every such subspace is a map of some subspace. (b) The time order of the measurements is either preserved by a mapping, or inverted. Let us repeat here that the assumptions necessary to arrive at this result were quite far reaching. It was, in particular, necessary to postulate that the c-d operators of the rays (4.13) can be measured prior to any, and after any, definite time.

4.8. Invariance for Proper Inhomogeneous Lorentz Transformations

It is postulated, just as in classical theory, that the mappings of observations onto observations as defined by the proper inhomogeneous Lorentz transformations are invariance transformations, that is, satisfy (2.5). So far this postulate is in agreement with the experi-

mental facts. These Lorentz transformations certainly satisfy the conditions of Sec. 2.4b and they even satisfy those of Sec. 2.4a if one admits the reasonable idealization of observers with arbitrarily high velocities. Therefore they can be considered to be invariance transformations in the sense of either section.

The proper inhomogeneous Lorentz transformations form a group which is called the Poincaré group. If we postulate that every element of the Poincaré group satisfy (2.5), the results of the previous section apply to each transformation of this group. This means that, if certain conditions (as specified in Sec. 4.7) are satisfied, an operator O will correspond to every inhomogeneous proper Lorentz transformation L such that

$$P_L = O_L P O_L^{-1} \quad (4.30)$$

for all observable P . The P_L in (4.30) is the first observer's operator for that measurement which, if carried out at time t on his time scale, appears to the observer removed from the original one by the Lorentz transformation L , as the same measurement as P appears to the original observer if the measurement is carried out at his time t . That this can be valid for all t is already a consequence of the proper Lorentz invariance of the theory.

Since every element of the Poincaré group can be deformed into the unit element in a continuous fashion, it follows, assuming only that one can denumerate the coherent subspaces, that the mapping O_L maps every coherent subspace onto itself. O_L is within each coherent subspace either unitary or antiunitary and unique apart from multiplication by a phase factor. The succession of two Lorentz transformations L and L' can be replaced by their product $L'L$. $O_{L'L}$ must mediate the same mapping as $O_L O_L$. Hence, within each coherent subspace, $O_{L'L}$ and $O_L O_L$ can differ only in a numerical factor of modulus one which, however, can yet depend on L and L' :

$$O_{L'L} O_L = \omega(L', L) O_L O_L. \quad (4.31)$$

Incidentally, as has been remarked by Bargmann,³² it follows from (4.31) at once that the operators O_L must all be unitary since a mapping which is mediated by the square of either a unitary or an antiunitary operator is necessarily unitary and all O_L can be written as such squares.

The discussion of the Poincaré group is simplified by a theorem which applies specifically to this group³³: The operators O_L can be replaced, within each coherent subspace, by new operators $\omega_L O_L$ (with $|\omega_L| = 1$, i.e., O_L and $\omega_L O_L$ mediate exactly the same redescription)

³² V. Bargmann, *Ann. Math.* **59**, 1 (1954).

³³ For a detailed proof of this theorem see E. P. Wigner, *Ann. Math.* **40**, 149 (1939); Ref. 30; A. S. Wightman, "L'invariance dans la mécanique quantique relativiste," in *Dispersion Relations and Elementary Particles* (Les Houches, 1960 Paris, and John Wiley & Sons, Inc., New York, 1960).

so that, for the new operators (writing O_L for $\omega_L O_L$),

$$O_{L'} O_L = \pm O_{L'L} \tag{4.32}$$

holds. This relation is by no means obvious; it amounts to the statement that the functions $\omega(L', L)$ which depend on two Lorentz transformations can be expressed in terms of functions of one Lorentz transformation

$$\omega(L', L) = \pm \omega_{L'L} / \omega_{L'} \omega_L \tag{4.33}$$

Within each coherent subspace either all states have integer angular momentum or all states have half-integer angular momentum (superselection rule of singlevaluedness). In the first case the minus sign in (4.32) can be eliminated. In the second case the minus sign in (4.32) cannot be eliminated and only the operators O_L together with the operators $-O_L$ form a group. The latter group is isomorphic not to the Poincaré group, but to a union (called semidirect product) of the group of parallel displacements in space and time and of the group of two-dimensional complex matrices of determinant 1. The union is quite similar to the union of the group of parallel displacements and of the group of rotations, which forms the Euclidean group. It is homomorphic to the Poincaré group, O_L and $-O_L$ corresponding to the same element L of the latter. Even more compact is the description of this group as a group the elements of which are characterized by two two-dimensional matrices, h and A , the first Hermitian, the second of determinant 1, with the multiplication law

$$(h_1, A_1)(h_2, A_2) = (h_1 + A_1 h_2 A_1^\dagger, A_1 A_2). \tag{4.34}$$

The “spinor transformation” A corresponds to a homogeneous Lorentz transformation and h corresponds to a parallel displacement by x, y, z, t in space and time where

$$h = \begin{pmatrix} t-z & x+iy \\ x-iy & t+z \end{pmatrix}. \tag{4.34a}$$

The group of the $\lambda = (h, A)$ will be called quantum-mechanical Poincaré group. There is a unique unitary operator U_λ associated with each element λ of this group and the invariance of the quantum-mechanical equations with respect to the Lorentz group is expressed by the analogue of (4.30),

$$P_L = U_\lambda P U_\lambda^{-1}, \tag{4.30a}$$

in which P_L is defined as before, L being the Lorentz (or rather Poincaré) transformation to which the element λ of the quantum-mechanical Poincaré group is associated.

In (4.30a), P_L is completely defined conceptually by the knowledge of P . Hence, the two elements of the quantum-mechanical Poincaré group, which correspond to the same element L of the ordinary Poincaré group, must give the same P_L . If one of these elements is λ , the other is $S\lambda$ with $S = (0, -1)$ and it follows from this

that the U_S which corresponds to $(0, -1)$ commutes with all P . It is, therefore, a superselection operator. Since the square $U_S^2 = U_E$, where E is the unit element, and since U_E must be the unit operator, as follows from $(U_E)^2 = U_E^2 = U_E$ for a unitary operator, the characteristic values of U_S are 1 and -1 . The coherent subspaces for which $U_S = 1$ have integer spin, those with $U_S = -1$ half-integer spin.

The operators O_L have, in addition to (4.30) another interpretation. Let us consider a pure state to which, according to Sec. 4.5, a vector or ray Ψ in Hilbert space corresponds such that the probability for the positive outcome of the measurement (P_a, t) is, for the state Ψ in question, $\text{Tr } P_a P_\Psi = (\psi, P_a \psi)$, where ψ is a normalized vector of the ray Ψ . Let us consider then a normalized vector $O_L \psi$ in the ray

$$\Psi_L = O_L \Psi = U_\lambda \Psi. \tag{4.30b}$$

The probability for the positive outcome of $P_L = O_L P O_L^{-1}$ for this state is $(O_L \psi, O_L P O_L^{-1} O_L \psi) = (\psi, P \psi)$. Hence the state $\Psi_L = O_L \Psi$ gives, for the observer removed by the Poincaré transformation L , the same result as does Ψ for the original observer: $O_L \Psi$ is the state Ψ subjected to the transformation L . This is the active interpretation of O_L or of the equivalent U_λ .

4.9. Equations of Motion, Conservation Laws

The invariance operations of the previous section contain also the “equations of motion.” The time and space displacement operators are members of the Poincaré group and the corresponding operators are, therefore, among the U_λ . It is an important consequence of the analysis which led to the elimination of the phase factors $\omega(L, L')$ from (4.31) that the operators of displacement all commute. Let us denote by U_t the operator which corresponds to a time displacement so that it maps every self-adjoint operator P which represents a certain measurement at time t_0 into the operator \bar{P} which represents the same action of measurement, carried out at time $t_0 + t$:

$$\bar{P} = U_t P U_t^{-1}. \tag{4.35}$$

This differs from $(P, t_0 + t)$ which does not represent the same action as (P, t_0) but gives the same result on an isolated system and has, in the Heisenberg picture which we use, the same operator. As a rule, $(P, t_0 + t)$ remains operationally unknown even if (P, t_0) is known. On the other hand, the calculation of the \bar{P} of (4.35) may be difficult, but it represents, operationally, the same action as P , only carried out later. Since the time displacements form a commutative one parameter group we may write

$$U_t = \exp(iHt/\hbar) \tag{4.36}$$

with H self-adjoint. Similarly, the operator of the dis-

placement in the x direction is

$$U_\lambda = \exp(-iP_x x/\hbar). \quad (4.36a)$$

Infinitesimal generators $M_{k\lambda}$ for the homogeneous Lorentz transformations are defined in a similar way. The operators H and P_x, P_y, P_z are identified as the operators for energy and linear momentum. The M_{ij} ($i=1, 2, 3, j=1, 2, 3$) are identified as the angular-momentum operators because they commute with the time-displacement operators (and hence with H). Hence, the operator which corresponds to the measurement of these quantities is independent of time and the probability to find a particular value as a result of this measurement is also independent of time as long as the system is isolated and no measurement thereon takes place in the time interval in question. The conservation laws for energy, linear momentum, and angular momentum are therefore a direct consequence of the relations of the infinitesimal generators of the Poincaré group to the time-displacement operator. The explicitly time-dependent operators $\exp(iHt/\hbar)M_{0i} \times \exp(-iHt/\hbar)$ correspond to the infinitesimal generators for Lorentz transformations which leave the point $(t, 0, 0, 0)$ invariant. It can easily be shown that the $\exp(iHt/\hbar)M_{0i} \exp(-iHt/\hbar)$ are linear functions of time. As a result of the dual role of the infinitesimal operators of the Poincaré group, the relation between invariance and conservation laws is much more direct in quantum than in classical theory.

4.10. The More General Invariances of Section 2.4b

A transformation of the group G will be called an invariance transformation in the sense of Sec. 2.4b if the mapping of observation onto observation is "simple." The simplicity of the mapping must be judged in terms of the mapping of experiment onto experiment and not on the basis of the mathematical framework of the (P_a, t) and formula (4.7). As we have discussed before, the framework does not even distinguish between observable operations (P_a, t) and nonobservable ones; similarly it gives little information on how to measure an observable (P_a, t) . Hence, we must rely first on the experimental determination whether a transformation belongs to the group G , and then decide on the basis of conceptual simplicity whether or not the mapping involved is likely to have universal validity. In the mathematical framework all mappings such as $(P_a, t) \rightarrow (UP_a U^{-1}, t)$, for any U , are equally simple.

Let us nevertheless investigate two particularly simple invariances of the mathematical framework. The first of these is

$$(P_a, t) \leftrightarrow (P_a, -t) \quad \text{for all } (P_a, t). \quad (4.37)$$

The invariance of (4.7) for the mapping has been established in Sec. 4.7, (iii). At first glance, (4.37) seems to be a candidate for time-inversion invariance in

the sense that (P_a, t) and $(P_a, -t)$ may be the same measurement, carried out at different times, at least if P_a refers to a quantity, such as position in a region, or energy within an interval, which are (in contrast to velocities), invariant under the "inversion of the direction of motion." This, however, cannot be the case because a similar statement would apply, because of time-displacement invariance, also to (P_a, t_0+t) and (P_a, t_0-t) . Since t_0 and t are arbitrary, all (P_a, t) would represent the same measurement for the quantities in question. This is possible as far as the energy is concerned but manifestly impossible in the Heisenberg representation for the measurement of the position. We must conclude that the mapping (4.37) has no simple physical interpretation. A similar consideration applies to $(P_a, t) \leftrightarrow (P_a^*, -t)$ (the star denoting the conjugate complex).

In order to find the mapping which may represent the time-inversion invariance, let us consider the operator Q which corresponds to a measurement carried out at time $t=0$ which is not affected by the reversal of the direction of motion. The operators of the same measurement, if this is carried out at times t and $-t$ respectively, are

$$\exp(iHt/\hbar)Q \exp(-iHt/\hbar)$$

and

$$\exp(-iHt/\hbar)Q \exp(iHt/\hbar).$$

These must be mapped on each other by the mapping of time inversion. If this mapping is obtained by transforming all operators by $O=\theta$, we have

$$= \theta \exp(-iHt/\hbar)Q \exp(iHt/\hbar)\theta^{-1}. \quad (4.38)$$

In particular for $t=0$ it follows

$$Q = \theta Q \theta^{-1}. \quad (4.38a)$$

Applying this to $Q=H$, the energy operator, we find that θ and H commute. If θ were unitary, it would commute also with any function of H and one would have to infer from (4.38) and (4.38a) that Q and $\exp(2iHt/\hbar)$ commute, i.e., that Q is a conserved quantity. Since this cannot be true for all operators Q in question, it follows that θ is antiunitary. In this case, it follows from $\theta H = H\theta$ that

$$\theta \exp(-iHt/\hbar)\theta^{-1} = \exp(iHt/\hbar), \quad (4.39)$$

and (4.38) has no other consequence but (4.38a).

Let us observe, before proceeding, that time inversion, as a physical operation, is involutory, i.e., its square must leave all measurable operators invariant. Hence, θ^2 is a superselection operator. It follows that

$$\theta^2 = \bigoplus_{\omega} \omega \mathbf{1}_\omega \quad (4.40)$$

in which \bigoplus stands for the direct sum over the coherent subspaces, $\mathbf{1}_\omega$ is the unit operator in the coherent sub-

space s , and the ω_s are complex numbers of modulus 1. It follows from (4.40) that if θ changes the coherent subspace s into s' , it must also change s' into s : θ can only interchange such subspaces. Furthermore, multiplying (4.40) from the right and from the left with θ and comparing the results one finds $\omega_s = \omega_{s'}^*$. Provided $s' \neq s$ we can make $\theta^2 = 1$ in the subspaces s and s' by replacing θ by the physically equivalent antilinear transformation which consists of first applying θ and then multiplying all vectors in s by the phase factor $(\sqrt{\omega_s})^{-1}$ and those in s' by $(\sqrt{\omega_{s'}})^{-1} = (\sqrt{\omega_s^*})^{-1}$. When $s' = s$, we have $\omega_s = \pm 1$. Concluding we have

$$\omega_s = \pm 1 = \epsilon_{s\theta} \quad (4.40a)$$

in which $\epsilon_{s\theta} = 1$ whenever $s' \neq s$.

Equation (4.39) has a particularly simple interpretation if the operators are interpreted actively, as acting on state vectors. θ^{-1} , just as θ , reverses the direction of motion of the system, $\exp(-iHt/\hbar)$ lets it then progress in time by t , the last θ again reverses the direction of motion. The resulting state of the system is the same as it was at time $-t$. This is natural since the time displacement by t took place after a reversal of the direction of motion. Relations similar to (4.39) can be obtained also for the other elements of the Poincaré group. Thus θ must commute with spatial displacements

$$\theta \exp(i\mathbf{P}\mathbf{x}/\hbar)\theta^{-1} = \exp(i\mathbf{P}\mathbf{x}/\hbar) \quad (4.39a)$$

from which it follows, since θ is antiunitary, that it anticommutes with \mathbf{P} . This is again natural since the direction of the momentum vector is reversed by the reversal of the direction of motion. It is a mathematical theorem that the compatibility relations, such as (4.39) and (4.39a), between the proper Poincaré transformations and the time-inversion operator θ , can be satisfied, in fact in more than one way. This gives rise to the theory of types. Naturally, the existence of an operator θ which leaves Π invariant and satisfies the proper relations, such as (4.39) and (4.39a), with the operators of the Poincaré group is only a necessary condition for time-inversion invariance. The validity of this invariance implies also that θ transforms other measurable operators in the proper way. In particular, if $P_a = Q$ decides whether a particle is at time 0 in a certain region of space, (4.38) must hold for this Q . This, of course, can not be established mathematically.³⁴

There remains a certain arbitrariness in what measurement we expect to be the time inverse of a P_a , even if this refers to the measurement of a simple quantity. We were forcibly reminded of this when it was found⁸ that the operator Σ of space reflection transforms the measurement of a particle into that of an antiparticle

³⁴ That CP conjugation is not a true invariance is suggested by recent experimental data. See J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Phys. Rev. Letters* **13**, 138 (1964). See also T. T. Wu and C. N. Yang, *ibid.* **13**, 380 (1964); R. G. Sachs, *ibid.* **13**, 286 (1964).

(CP) or that, at any rate, the assumption that it transforms the decision that the particle is at \mathbf{r} at time t into the decision that it is at $-\mathbf{r}$ at time t leads to a contradiction with experiment: the operator with that property does not commute with time displacement.

Let us establish relations for the operator Σ of space reflection similar to those obtained for θ . The operator Σ must again commute with H because we want the energy to remain unchanged by space inversion. As a physical operation, Σ also commutes with time displacement which gives an equation

$$\Sigma \exp(iHt/\hbar) = \omega_s(t) \exp(iHt/\hbar) \Sigma \quad (4.41)$$

similar to (4.31). It follows from (4.41) and $\Sigma H = H \Sigma$ that Σ is unitary, just as it follows from (4.38) that θ is antiunitary, at least as long as H has more than one eigenvalue. The transformation Σ is also involutory so that within each coherent subspace

$$\Sigma^2 = \omega_s, \quad |\omega_s| = 1 \quad (4.42)$$

holds. In contrast to θ , which is antiunitary, Σ can be modified, by replacing it with $\Sigma/\omega_s^{\frac{1}{2}}$, so that

$$\Sigma^2 = 1. \quad (4.42a)$$

It is again a mathematical theorem that there is a unitary operator Σ which satisfies (4.42a), (4.41) and all similar relations with the operators of the proper Poincaré group and also with θ . Whether this is an invariance operator again hinges on the question whether it transforms the other observables, such as position, in the expected way. We now know that unless "expected way" means the transformation of a particle into an antiparticle, and conversely, Σ is not an invariance. Hence, if Σ is an invariance it has the effect of what is usually denoted by CP . Note that, provided the superselection rule of charge holds, CP does not give rise to a concept analogous to parity, except for neutral particles.

The succession of a time inversion and a space inversion is again a symmetry operation if the two factors are such operations. The operator I which corresponds to the product can differ from $\Sigma\theta$ by a factor which is different for different coherent subspaces. However, since a similar factor is indeterminate in I , we can write

$$I = \Sigma\theta. \quad (4.43)$$

The physical operations space inversion, time inversion, and space-time inversion form a commutative group which is isomorphic to the four group. Their images Σ , θ , I form, together with the unit element a "projective" representation of that group

$$\theta\Sigma = \omega I, \quad |\omega| = 1, \quad (4.44)$$

etc. Just as in the case of the Poincaré group, one tries to eliminate the phase factors by making use of the arbitrariness in the operators Σ , θ , I , so that they form

a true representation of a group.³⁵ Equation (4.43) is already a step in this direction. The square of I , since it represents, physically, the unit operation, must commute with every observable and is, therefore, a superselection operator. It has a definite value in each coherent subspace and since I is antiunitary, this value can be only ± 1 . It will be denoted by ϵ_{sI} , the index s characterizing the subspace. Whenever s is different from its image s' under the mapping I , we can see to it that $\epsilon_{s'I} = 1$. In case I represents CPT , s' will certainly be different from s for a coherent subspace s the states of which have a nonzero charge; and one can see to it that $\epsilon_{sI} = 1$. The rest of the relations follows from the associative law which applies to both unitary and antiunitary operators and one obtains the following group table for the operators $1, \Sigma, \theta, I$ which correspond physically to no change, space inversion, time inversion, and space-time inversion

	1	Σ	θ	I	
1	1	Σ	θ	I	
Σ	Σ	1	I	θ	
θ	θ	$\epsilon_{sI}\epsilon_{s\theta}I$	$\epsilon_{s\theta}1$	$\epsilon_{sI}\Sigma$	(4.45)
I	I	$\epsilon_{sI}\epsilon_{s\theta}\theta$	$\epsilon_{s\theta}\Sigma$	$\epsilon_{sI}1$	

Unless $\epsilon_{sI} = \epsilon_{s\theta} = 1$, two operators, differing in sign, correspond to each physical operation, just as in the case of the Poincaré group.

Both θ^2 and I^2 commute with every observable; they generate superselection rules—called rules of “type.” There are four different “types” corresponding to the combinations $(+1, +1)$, $(+1, -1)$, $(-1, +1)$, $(-1, -1)$ for θ^2 and I^2 . State vectors which are superpositions of vectors of different types are not realizable. For $\epsilon_{s\theta} = \epsilon_{sI} = 1$, the group of (4.45) is the four group, for $\epsilon_{s\theta} = -\epsilon_{sI} = 1$ and for $-\epsilon_{s\theta} = \epsilon_{sI} = 1$, it is a dihedral group whereas for $\epsilon_{s\theta} = \epsilon_{sI} = -1$, it is the group $C_2 \times C_4$.

The extended Poincaré group is the semidirect product of the Poincaré group and the four-group of $1, \Sigma = CP, T, I = CPT$. The proper Poincaré group $\{L\}$ is a normal subgroup of the extended Poincaré group (this is true for the connected subgroup of every mixed continuous group; see p. 92 of Ref. 31), and the cosets are $\{L\}, CP\{L\}, T\{L\}, CPT\{L\}$. The operators O which represent the transformations of the extended Poincaré group form a projective representation of the group [similar to (4.31)]. Corresponding to the four different types $(1, 1), (1, -1), (-1, 1), (-1, -1)$ one finds, however, that there are four different extended quantum-mechanical Poincaré groups. Each of these is a semidirect product of the quantum-mechanical Poincaré group and one of the groups of (4.45). For a

charge-carrying subspace s , only two types $(1, 1)$, and $(-1, 1)$ occur.

It is evident that the elements of the broader class of invariance transformations, in the sense of Sec. 2.4b, are by no means as sharply defined as those of Sec. 2.4a. Nevertheless, they may represent valid invariances and are, therefore, of interest. In this regard the $I = CPT$ transformation may play a special role because its character as an invariance transformation follows from the postulates of local field theory.²² The situation is analogous to that encountered in Sec. 3.1 on classical mechanics: Galilei invariance and time-inversion invariance were there found to follow from the assumption that the forces can be derived from a velocity-independent potential.

4.11. The Problem of Measurement in Relativistic Quantum Theory

Let us return to the discussion of the Π function of Secs. 2.3, 4.1, and 4.4 to note that the observations which enter into Π are supposed to be carried out at definite times. In fact, as (4.7) shows, the probabilities for the possible outcomes of measurements depend on their ordering in time. This is true even though we use the Heisenberg picture in which the time of measurement of an *operator* is immaterial because the measuring procedure which corresponds to an operator changes in time in such a way as to compensate for the change of the system. As a result, as long as the system on which the measurement is carried out remains isolated the result of the measurement which corresponds to an operator will remain independent of time. However, an intervening measurement affects the system and renders the probabilities for the possible outcomes of a measurement different from what it would have been had the system remained isolated and developed naturally.

The picture which is consistent with the instantaneous nature of the measuring process would be an infinitely short, but very strong, interaction between apparatus and the system on which the measurement is to be undertaken. Only in this way would a measurement at one definite time be possible. The instantaneous very strong interaction is an unrealistic idealization even in nonrelativistic theory. In relativistic theory, if taken seriously, it would contradict the translatability postulate [postulate (b) of Sec. 4.1] because instantaneous from the point of view of one observer is not instantaneous from the point of view of another. This might appear, first, as a superficial difficulty but all ways to avoid it imply significant changes in the concept of measurement which may affect the whole theory. As far as we can see, there are five alternatives.

The first alternative is, naturally, to give up the postulate (b) that all measurements which can be performed by one observer can be performed by all. This is possible, logically, but would reduce the significance of relativistic invariance, in our opinion, to an

³⁵ See E. P. Wigner, Ref. 30. This paper contains a detailed discussion of the material covered in the present section.

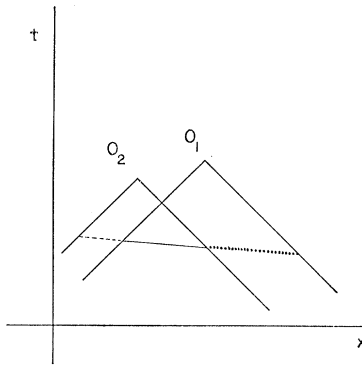


FIG. 1. O_1 and O_2 are points of output measurements; the 45° lines are null-lines, i.e., form the light-cones of O_1 and O_2 . The broken line is inside the past light-cone of O_2 and passes through points which influence the measurement at O_2 . It has a space-like relation to O_1 . Similarly, the dotted line is inside the past light-cone of O_1 and has a space-like relation to O_2 .

undesirable extent. There would be two types of equivalences between observers: the measurements of those who are at rest with respect to each other have a common meaning and are translatable. This would not be the situation for observers moving with respect to each other.

The other modifications of the measurement concept preserve the postulate of translatability. The first possibility is to restrict the concept of measurements in such a way that the results depend only on the conditions at one space-time point. Since space-time points have translatable meanings for all observers, the difficulty which we are considering disappears. This is the philosophy of local quantum field theories if these are followed to their ultimate logical consequences.³⁶ Naturally, the field can be measured at many points—perhaps even at a continuum of points—and these measurements should not interfere with each other as long as the points of measurement are in space-like relation to each other. Inputs are usually considered to be measurements of this nature; the outputs are later similar measurements, again at space-like points with respect to each other. It is usually assumed that the space-time points at which the latter measurements are undertaken are in a time-like relation to the points of the input measurements; otherwise the input measurements interfere with the output ones. If this postulate—the necessity of which is not clear—is valid, there would be no two output measurements involving more than one point in space-time for which the Π function can be calculated without statistical assumptions (see Fig. 1) and, at any rate, no pure states. The reason is that a space-time point, where an output measurement is possible, would have to lie in the positive light cone of all input measurements. It then lies, however, also in the positive light cone of other points where the initial field was not ascertained by the input

³⁶ See, for instance, G. Källén, "Quantenelektrodynamik," in *Handbuch der Physik* (Springer-Verlag, Berlin, 1958), Vol. 1, p. 169.

measurements and influences the field at the output points. It is not clear that this impossibility to have a full set of input measurements is a serious defect of the theory—we constantly rely on not receiving a heavy dose of radiation from the parts of the universe from which are so distant that no radiation could have reached us so far and about which we can know nothing positive—but it does make precise experiments impossible even in principle.

The second possibility which is consistent with the translatability postulate is that alluded to in the preceding discussion; it assumes that measurements are possible also at points which have a space-like relation to points where measurements have taken place. Under this assumption, the input measurements can all precede the output measurements in some frames of reference, in other frames of reference they are later in time. However, as a result of the space-like relation of these input and output measurements, the corresponding operators commute so that (4.7) gives the same result in both coordinate systems. Under the present assumption—which we much prefer to the preceding one—the simple situation to which (4.6) refers remains preserved as long as the input observations are on an entirely space-like surface, and the same holds for the output observations, and the two surfaces do not intersect.

The two alternatives just discussed do not give an equal wealth of observables as one finds in ordinary quantum mechanics because the fields do not enter into the observable operators in arbitrary combinations as do the basic operators p and q of ordinary quantum mechanics. These do permit the measurement of quantities such as

$$\int f(\mathbf{x}, \mathbf{y}) \varphi(\mathbf{x}) \varphi(\mathbf{y}) \varphi(\mathbf{x}) d^3x d^3y$$

because the measurement of this quantity can be carried out by measuring $\varphi(\mathbf{x})$ over all space, and then calculating the integral. This is possible because $\varphi(\mathbf{x})$ and $\varphi(\mathbf{y})$ commute if the relation of \mathbf{x} and \mathbf{y} is space-like and the integral involves only \mathbf{x} and \mathbf{y} of such nature. The two alternatives just considered do not, however, permit the measurement of a quantity such as

$$\int f(\mathbf{x}, \mathbf{y}) \varphi(\mathbf{x}) \pi(\mathbf{y}) \varphi(\mathbf{x}) d^3x d^3y,$$

where φ and π are conjugate field quantities because in this case $\varphi(\mathbf{x})$ and $\pi(\mathbf{y})$ do not commute even if \mathbf{x} and \mathbf{y} have space-like relation. As a result, the measurement of the last integral can not be reduced to the measurement of simultaneously measurable quantities. Most composite expressions in the field quantities are of this nature and are not measurable if only fields are measurable. In ordinary quantum mechanics, composite expressions, such as $p^2 + q^2$ are supposed to be measurable even though their measurement cannot be reduced to the simultaneous measurement of p and q and subsequent calculation of $p^2 + q^2$.

It is very likely, on the other hand, that the instantaneous measurement of such more general operators is impossible, and we believe that even the measurement of simple field quantities takes an irreducible minimum of time. This suggests that the measurability on a space-like surface, which implies an instantaneous measurability, is a dangerous idealization. Hence, the third alternative which is consistent with the translatability postulate assumes that all measurements extend over a finite portion of space-time. Such portions again have a meaning which can be translated from one coordinate system to another. However, it is not clear what the actual limitations on the measurability of fields may be, and the consequences of the limitations of the instantaneous character of measurements have not been explored.

The last possibility concerning measurability gives this process a much more modest scope than any of the preceding alternatives. It is an outgrowth of Heisenberg's observations which led him to propose an S -matrix theory. According to this point of view, only momenta and position coordinates can be measured, the latter only crudely, and even these quantities can be measured only on well-separated particles which do not interact with others.³⁷ If this point of view is adopted, it is immaterial whether or not the measurement is instantaneous: the momenta are, for noninteracting particles, constants of motion and the absolute squares of the S matrix elements are determined as cross sections no matter whether the measurement of the momenta requires a finite time interval or can be done instantaneously (which it probably can not). The measurement of the average position (to obtain the time delay and hence information beyond the absolute value of the S -matrix elements) is possible also because the time variation of the average position coordinates of free particles is a simple function of their momenta. One must admit, in spite of some objections, that this last alternative appears to be the most realistic one and we must admit that our own considerations do not do justice to this point of view. It is, of course, in conflict with the extended role of the laws of nature which we advocated in our Sec. 2. It permits, in essence, only the calculation of the outcomes of individual experiments, namely, of collision processes.³⁸

5. INVARIANCE AS A TOOL FOR OBTAINING PROPERTIES OF THE SOLUTIONS OF EQUATIONS

Invariance principles as defined in Sec. 2.4 and also dynamic invariances are used in physics in two, rather distinct, manners. They can serve as a guide to obtain

³⁷ W. Heisenberg, *Z. Physik* **120**, 513 (1943). For a comprehensive review, see M. L. Goldberger and K. M. Watson, *Collision Theory* (John Wiley & Sons, Inc., New York, 1964).

³⁸ See also G. Källén's remarks to the Discussion générale at the 1961 Solvay Congress [*La Théorie Quantique des Champs* (Interscience Publishers, Inc., New York, 1962)].

the "law of nature" or the "equation of motion." Second, they can be used to solve these equations at least partially, or to obtain properties of the solutions.

Both uses were already possible in classical, that is nonquantum, theory. As is well known, the equations of the general theory of relativity are the simplest, or at least the most natural, equations compatible with Einstein's principle of general covariance.³⁹ As to the help which the invariance principles give for solving the equations of motion, this is provided in classical mechanics, through the medium of the conservation laws. The connection between conservation laws and invariance principles was discussed already in Sec. 3.1. Actually, the original discovery of the conservation laws was independent of the invariance principles and the connection between the two was recognized only later.⁴⁰ Originally, the conservation laws were used to provide "integrals of motion," i.e., as help toward the solution of the equations of mechanics. One has a complete solution of these equations if one has eliminated both the second and the first time derivatives of the coordinates from the equations of motion. Hence, the complete solution of a problem involving n point-masses requires $3n$ second integrals. The conservation laws for momentum, energy, and angular momentum give seven first integrals, i.e., permit the elimination of seven second derivatives. The law of the center of mass, which is related to the invariance with respect to Galilei or Lorentz transformations, gives the motion of the center of mass completely, i.e., gives three second integrals. Thus, the invariance principles of mechanics give, through the conservation laws, a considerable simplification of the problem, particularly if the number of constituents is not too large. In recent times this has received less emphasis than one might expect, perhaps because the difficulty of solving the ordinary (nonpartial) differential equations of the mechanics of point masses is not overwhelmingly great anyway.

The invariance principles proved even more effective in quantum theory. As far as obtaining the basic equations is concerned, two cases are particularly noteworthy: (a) All possible equations for elementary particles could be obtained from the theory of the representations of the inhomogeneous Lorentz group.⁴⁰ This theory does not, however, explain why particles which are described by some of these equations exist whereas most of the equations do not describe existing particles. In particular, the ratio of the masses of the elementary particles remains unaccounted for. (b) The proper form of the weak interaction was obtained by

³⁹ See, for instance, W. Pauli, "Relativitätstheorie," in *Enzyklopädie der Mathematischen Wissenschaften* (B. G. Teubner, Berlin, 1921), Vol. 2, p. 722 (p. 159 of the English translation, Pergamon Press, Inc., New York, 1958); M. von Laue, *Die Relativitätstheorie* (Friedrich Vieweg, Braunschweig, 1953), Vol. 2, p. 100; V. Fock, *The Theory of Space Time and Gravitation* (Pergamon Press, Inc., New York, 1959), p. 171.

⁴⁰ See E. P. Wigner, *Ann. Math.* **40**, 149 (1939); V. Bargmann and E. P. Wigner, *Proc. Natl. Acad. Sci. (U.S.)* **34**, 211 (1948); also R. Shaw, *Nuovo Cimento* **33**, 1074 (1964).

Marshak and Sudarshan, and by Feynman and Gell-Mann, on the basis of invariance principles.⁴¹

The applications of the invariance principles for obtaining properties of the solutions of the quantum-mechanical equations fall, broadly speaking, into two classes. The problems of spectroscopy are concerned with the discrete part of the energy spectrum. It is hardly an exaggeration to say that all qualitative rules of spectroscopy, including the intensity ratios of multiplet spectra, can be derived from invariance principles—some of them exact, others approximate. Similarly, most of the theory of angular distributions and of angular correlations of the products of collisions and of disintegrations, can be based most naturally on the analysis of the invariance properties. A recent example of the use of invariances in spectroscopy is given by the work of Ne'eman and Gell-Mann on the spectroscopy of strongly interacting elementary particles.⁴²

The applications concerning the partial solution of the equations can be best summarized by the statement that an n -parametric group permits in the general case the elimination of n variables; n discrete group elements reduce the variability domain of the variables by a factor n . These statements⁴³ are not complete, however, because, depending on the state, the number of unknown functions is increased. The increase in the number of unknown functions (which is equal to the dimension of the representation encountered) is usually quite small and is particularly small for the low-lying states. Again, the simplification introduced by symmetry principles is much greater than in classical theory in which, for instance, the discrete group elements can be used only in a somewhat esoteric fashion. Naturally, the partial solution of the quantum-mechanical equations by means of invariance principles, and the determination of the properties of these solutions discussed in the preceding paragraph, are not independent of each other.

5.1. Reasons for the Increased Importance of Invariance Principles in Quantum Theory

It is natural to ask for the reasons of the greater effectiveness of the invariance principles in quantum theory than in classical theory. The principal invariances are the same in both theories and are those which were mentioned before: invariance with respect to rotations and Galilei or Lorentz transformations depending on whether one uses nonrelativistic or relativistic theory. The reason for the greater effectiveness

⁴¹ R. E. Marshak and E. C. G. Sudarshan, *Phys. Rev.* **109**, 1860 (1958); R. P. Feynman and M. Gell-Mann, *Phys. Rev.* **109**, 193 (1958); G. S. Gehrstein and A. B. Zeldovitch, *Zh. Eksperim. i Teor. Fiz.* **29**, 698 (1955); J. J. Sakurai, *Nuovo Cimento* **7**, 649 (1958).

⁴² Y. Ne'eman, *Nucl. Phys.* **26**, 222 (1961); M. Gell-Mann, *Phys. Rev.* **125**, 1067 (1962).

⁴³ See E. P. Wigner, *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren* (Friedrich Vieweg, Braunschweig, 1931), Chap. 19. (English transl.: Academic Press Inc., New York, 1959.)

of these principles in quantum theory seems to be that the set of possible states in this theory has a linear structure.⁴⁴ The set of general states is a linear manifold, that is, one can obtain a state by superposing any number of other states, with arbitrary coefficients. This superposition principle can be combined with the invariance principles and renders the latter much more fruitful than they were in nonquantum theory. The superposition principle is lost in that limiting case of quantum theory which corresponds to classical theory because, in the transition from quantum to classical theory, one employs not only a limiting process but also discards all states in which coordinates and velocities do not have, in the limit, sharply defined values. In quantum theory, the invariance principles are most effective for states, such as the low-lying states of atoms, in which rather wide ranges of both velocities and positions have appreciable probabilities.

The preceding discussion indicates that the role of a symmetry principle should become the same in quantum mechanics as in classical theory if a superposition of the initial state and that obtained by the symmetry operation is not meaningful. This is the situation if the two states are separated by a superselection rule. This situation occurs indeed; the only consequence of the symmetry is then, just as in classical theory, that the states connected by the symmetry operation behave in the same way.

It may be of interest to spell out in detail how the linear structure of the underlying Hilbert space renders the use of invariance principles more effective. Consider, for instance, a rotation by π and denote the corresponding operator by O_π . It follows from developments sketched in Sec. 4.8 that, for integer spin, $O_\pi^2 = 1$. Consider now an arbitrary state ψ ; there exists then another state $O_\pi\psi$ the properties of which are related to those of ψ in a rather obvious manner. This is in itself an interesting piece of information; it is valid both in classical and in quantum theory. However, in quantum theory one can infer also the existence of the state $\phi = \psi + O_\pi\psi$ for which $O_\pi\phi = \phi$ is valid. This means that there is a state ϕ which is invariant under the rotation considered. In fact the other state $\phi' = \psi - O_\pi\psi$ is also invariant because $O_\pi\phi' = -\phi'$ represents the same state as ϕ' . Since $\psi = \frac{1}{2}(\phi + \phi')$, this state—and in fact every state—can be written as a superposition of invariant states. These statements, and their generalizations for more, and for more complicated, symmetry operations go much beyond the statement of the existence of a “rotated state” $O_\pi\psi$ and illustrate the power which the linearity of the Hilbert space gives to invariance principles.

5.2. Matrix Form of Invariance Operators

We have introduced, in Sec. 4.8, the operators O_L and U_λ which could be given two interpretations: the

⁴⁴ This point was emphasized already in the article of Ref. 5 and in the lecture of C. N. Yang at the 75th Anniversary of Bryn Mawr College, 1959.

passive one of (4.30) or (4.30a), giving the transformation of observables to a Lorentz-transformed coordinate system, and the active one (4.30b), which defines their effect on state vectors. Let us introduce a complete orthonormal set ψ_1, ψ_2, \dots in one of the coherent subspaces and consider the matrix elements

$$(\psi_i, U_\lambda \psi_k) = D(\lambda)_{ik}. \tag{5.1}$$

One infers from (5.1) and the complete orthonormal nature of the ψ_i that

$$U_\lambda \psi_k = \sum_i D(\lambda)_{ik} \psi_i. \tag{5.2}$$

It is easy to verify, by applying U_μ to both sides that

$$D(\mu)D(\lambda) = D(\mu\lambda). \tag{5.3}$$

If U_μ were anti-unitary and $D(\mu)_{ik}$ is defined as in (5.1), one would have, instead of (5.3)

$$D(\mu)D(\lambda)^* = D(\mu\lambda). \tag{5.4}$$

The unitary nature of the matrices $D(\lambda)$ and $D(\mu)$ follows from (5.2), and the fact that the $\psi_i, U_\lambda \psi_i$ and $U_\mu \psi_i$ are unit vectors, no matter whether the operators U_λ and U_μ are unitary or antiunitary. The set of matrices which satisfy Eqs. (5.3) and (5.4) are said to form a corepresentation⁴⁵ of the group of operators U_λ, U_μ .

The purpose of the preceding calculation in the present context was to give a matrix form to the invariance operators. This renders the following consideration of the use of these operators for the partial solution of definite problems more visualizable.

5.3. States with Zero Spatial Momentum and Definite Energy

The problems toward the solution of which invariance principles contribute most can be formulated in terms of invariance transformations. The most important case is that in which one asks for the states for which:

- (a) the spatial components of the momentum are zero (i.e., the states are invariant under spatial displacements);
- (b) the energy has a definite value E (states, the state vectors of which are multiplied by a factor $\exp(iEt/\hbar)$ under a time displacement by t).

We shall use the example of this set of states to discuss the use of invariance principles for a partial solution of problems. It may be noted, first, that it follows from the linearity of the displacement operators that the set just defined forms a linear manifold. This linear manifold is clearly invariant under the group of all displacements. It is, however, also invariant under the group of purely spatial rotations, proper or improper (if the latter are contained in the symmetry group), and under the combination of these elements with time

inversion. In order to find the linear manifold in question, one can consider those matrices of the representation $D(\lambda)$ which correspond to the subgroup of displacements. These form a representation of the displacement group which is, of course, not irreducible. We may imagine that $D(\lambda)$, as representation of the subgroup, is reduced out and select the states which belong to the representation in which to a displacement by the four-vector t, \mathbf{x} the (one-dimensional) matrix

$$||\exp(iEt/\hbar)|| \tag{5.5}$$

corresponds. [Note that (5.5) is independent of \mathbf{x} ; i.e., the states in question are invariant under displacements by spatial vectors \mathbf{x} .]

It should be noted that, strictly speaking, in the original Hilbert space of the system, there can be no state vectors which satisfy condition (a). The reason from the point of view of physical concepts is that there are no states in which the linear momentum would have a sharply defined value. This can already be inferred from the uncertainty relations. The mathematical counterpart of this fact is that the permissible representations of the Poincaré group, if considered as representations of the subgroup of spatial translations, are integrals rather than sums of the irreducible representations of this subgroup. Similar to the situation encountered in continuous spectra, it is possible to approximate states which belong to an irreducible representation of the subgroup of spatial displacements, but it is not possible to obtain them. Condition (a) requires states which are invariant under the spatial displacement operators. The aforementioned approximation means, in this case, that given a finite set of displacements, it is possible to find a set of state vectors each of which differs arbitrarily little from the state vector obtained by applying a member of any finite set of displacement operators to it. Furthermore, a limiting process can be defined⁴⁶ leading to a new Hilbert space, the vectors of which are strictly invariant under all spatial displacements and are ideal elements (limiting cases) of the vectors of the original Hilbert space. This means that states with zero momentum can be defined, that they form a Hilbert space, and that their properties are limits of the properties of true physical states. We shall be concerned, henceforth, with vectors of this new Hilbert space.

As far as condition (b) is concerned, there are now three cases to be distinguished.

(I) The representation (5.5) is not contained in $D(\lambda)$ if this is considered as representation of the subgroup of displacements. In this case there are no states with zero momentum and energy E .

(II) The representation (5.5) is contained as a discrete representation in $D(\lambda)$. If it is contained l

⁴⁵ See the English translation of the Book of Ref. 43, Chap. 26.

⁴⁶ For the mathematical development of this concept, see J. v. Neumann, *Ann. Math.* **50**, 401 (1949); F. I. Mautner, *ibid.* **51**, 1 (1950); **52**, 528 (1950); G. W. Mackey, *ibid.* **55**, 101 (1952); **58**, 193 (1953); also J. Dixmier, *Les Algèbres d'Opérateurs dans l'Espace Hilbertien* (Gauthier-Villars, Paris, 1957).

times, there are l linearly independent states which satisfy our criteria. In principle, l could be infinite. At any rate, these states belong to the discrete spectrum of the time-displacement operator in the new Hilbert space, or, as is said more commonly but somewhat less completely, they belong to the discrete spectrum of the energy. The study of this case is of principal concern to spectroscopy.

(III) The representation (5.5) is contained as representation of the subgroup of time displacements in that part of $D(\lambda)$ which is an integral, rather than a sum, of irreducible representations of the subgroup of such displacements, even in the new Hilbert space. The corresponding "states" then belong to the continuous spectrum and their consideration involves the problems which are associated with the "states" of the continuous spectrum. It is generally assumed that the absence of true states in the continuous spectrum leads only to formal complications and this can be adequately motivated. The problems of this case belong to collision theory.

It could occur that the representation (5.5) forms part of both the discrete and the continuous spectrum.⁴⁷ In this case, the true states associated with the discrete representations could be considered separately from the "states" associated with the continuous spectrum. The former are the subjects of the study of spectroscopy, the latter the subjects of collision theory.

The important point is that, in any case, the linear manifold obtained is invariant under rotations, reflections and time inversion. This is clear intuitively because a state with linear momentum zero maintains this property under all these operations and the energy is also invariant under them. Arguing more formally, one can note that, as far as rotations are concerned, the corresponding operators are, in the isomorphism to the group given in (4.34), images of $(0, u)$, where u is unitary. The spatial displacements are $(h_s, 1)$, where the trace of h_s is zero. Finally the operators of displacement by t along the time axis are images of $(t1, 1)$. Hence, the linear manifold in question is defined by the equations

$$U_{(h_s, 1)}\psi = \psi \tag{5.6}$$

$$U_{(t1, 1)}\psi = \exp(iEt/\hbar)\psi. \tag{5.7}$$

It follows, furthermore, from (4.34) that

$$(h_s, 1)(0, u) = (h_s, u) = (0, u)(u^{-1}h_s u, 1), \tag{5.8}$$

since $uu^\dagger = 1$. Furthermore, the trace of $u^{-1}h_s u$ is also zero. Hence, since the group relation (4.34) applies also to the operators

$$\begin{aligned} U_{(h_s, 1)}[U_{(0, u)}\psi] &= U_{(h_s, 1)}U_{(0, u)}\psi = U_{(0, u)}U_{(u^{-1}h_s u, 1)}\psi \\ &= U_{(0, u)}[U_{(u^{-1}h_s u, 1)}\psi] = U_{(0, u)}\psi, \end{aligned} \tag{5.9}$$

since the h_s in (5.6) is arbitrary except that its trace is zero. It follows that $U_{(0, u)}\psi$ satisfies the equation

(5.6). Similarly, it follows from (4.34) that

$$(t1, 1)(0, u) = (t1, u) = (0, u)(t1, 1) \tag{5.10}$$

since $ut1u^{-1} = t1$. Since the same equation applies to the operators we have

$$\begin{aligned} U_{(t1, 1)}[U_{(0, u)}\psi] &= U_{(0, u)}U_{(t1, 1)}\psi = U_{(0, u)}\exp(iEt/\hbar)\psi \\ &= \exp(iEt/\hbar)U_{(0, u)}\psi, \end{aligned} \tag{5.11}$$

so that the second equation (5.7) is also satisfied by $U_{(0, u)}\psi$. Hence, the linear manifold defined by (5.6) and (5.7) is invariant under the $U_{(0, u)}$ —which is what was to be proved. The discussion of the effect of the reflection operators will be omitted.

It is to be noted that the invariance of the manifold defined by (5.6) and (5.7) under rotations is not a group-theoretical result. The spatial displacements do not commute with the rotations, only, as (5.10) shows, displacements along the time axis. One could say, rather, that it is a representation theoretical result because it applies to the particular representations of the displacement group in which the spatial displacements are represented by the unit matrix. Hence, if one uses another representation of the displacement group, the corresponding vectors will not be invariant any more under rotations (though they will be invariant under another subgroup of the homogeneous Lorentz group, the one which leaves the corresponding momentum vector unchanged). Similarly, the consideration cannot be carried over to groups, such as the deSitter group, in which the invariance with respect to spatial displacements cannot be equally simply specified.

5.4. The Discrete Spectrum

We shall follow the argument of the preceding section somewhat further because it shows more concretely the reason for the increased usefulness of invariance arguments in quantum theory than the somewhat abstract considerations of Sec. 5.1.

In the preceding section we obtained a linear manifold of states which is invariant under displacements [in a trivial way, see (5.6) and (5.7)], and also invariant under rotations and reflections. The states of this manifold therefore belong to a representation of the group formed by these transformations, just as all the states belong to a representation of the Poincaré group. However, whereas the latter representation is not irreducible, the former one almost invariably is. This means that the reduction of the total Hilbert space to those functions which transform under displacements according to (5.6) and (5.7) reduces the whole linear manifold of the Hilbert space to such a degree that this reduced linear manifold belongs to a single irreducible representation of the very much smaller subgroup which contains only displacements, rotations, and their products, but no true Lorentz transformations. Furthermore, since, according to (5.6) and (5.7), the matrices corresponding to dis-

⁴⁷ L. Fonda, Ann. Phys. (N. Y.) 22, 123 (1963).

placements are all definite multiples of the unit matrix, the representation is essentially determined by the matrices which correspond to rotations

$$U_{(0,u)}\psi_k = \sum_i D(u)_{ik}\psi_i. \quad (5.12)$$

These matrices, as indicated in (5.12), form a representation of the two-dimensional unimodular unitary group (we disregard the reflections for the present). The effect of all operators of the subgroup is contained in Eqs. (5.6), (5.7), and (5.12).

The irreducible representations of the two-dimensional unimodular unitary group can be characterized by an index j which can assume the values $0, \frac{1}{2}, 1, \frac{3}{2}, \dots$; the corresponding representation has $2j+1$ dimensions. Hence, the D in (5.12) should have an upper index $D^{(j)}$; the linear manifold of the states is $2j+1$ dimensional, both k and i can assume $2j+1$ values only. It follows that we have, in (5.12), an infinity of equations (corresponding to the infinite number of matrices u) for a finite number of functions, namely $2j+1$ functions ψ_k . On account of the continuity of the group and its representations $D^{(j)}$ and on account of the finite number (three) of group parameters, these infinitely many equations come to the same thing as a finite number of differential equations. It is natural that these differential equations determine many of the properties of the ψ_k .

It is interesting to compare this situation with the one prevailing in classical theory. In classical theory it is true, also, that a state of zero linear momentum and energy E remains such a state even if subjected to a rotation. However, the number of such states in classical theory is infinite and every rotation of a given state leads to a new state, and there is no further relation between these states. Hence, rotational invariance of the states does not permit one to determine any of the properties of any state, at least not in a simple fashion. In quantum theory, there are also infinitely many states and it is also true, as a rule, that every rotation leads to a different state. However, these states form a linear manifold and, as (5.12) shows, all of them can be expressed linearly in terms of a few ($2j+1$) of them. It is this finite basis for all the states which have zero spatial momentum and a definite energy which makes the rotational invariance in quantum theory so meaningful.

One may wonder whether we have not lost a great deal of information by having restricted ourselves to the operators which correspond to a subgroup, rather than having considered all the operations of the Poincaré group. Since every transformation of the Poincaré group can be obtained as the product of a transformation of the subgroup and of a true Lorentz transformation,⁴⁸ and since we do consider all transformations of the subgroup, only the true Lorentz transformations

need be further considered. However, every true Lorentz transformation transforms a state with zero spatial momentum into a new and linearly independent state so that these transformations yield as little information on the properties of the original functions as *any* of the transformations yield in nonquantum theory. The consideration of the true Lorentz transformations only shows that the original states can be given any velocity. As a matter of fact, the states obtained by a true Lorentz transformation from the states of the linear manifold defined by (5.6), (5.7) give an irreducible representation of the Poincaré group which is characterized by mass E/c^2 and spin j . This representation is one of the irreducible representations contained in the representation of the Poincaré group which is induced by all the possible states of the system. Other irreducible parts are generated by the states which correspond to other discrete energy values and by states which correspond to the continuous spectrum.

5.5. The Continuous Spectrum

The situation in the case of the continuous spectrum differs from that of the discrete spectrum, from the invariance-theoretic point of view, principally by the fact that the representation induced by rotations on the functions satisfying (5.6) and (5.7) (again with a single definite E) is not irreducible. This corresponds to the fact that a continuous spectrum can be present only if one has at least two free particles, that is two particles not bound to each other. The angular momentum of two such particles due to their motion about their center of mass can assume, for any value of the energy, any value $0, \hbar, 2\hbar, \dots$. This manifests itself in the presence of either all the representations of the unimodular two-dimensional unitary group with integer spin j or all those with half-integer spin j , which are present in the whole representation of the Poincaré group, if this is considered as a representation of this subgroup.

The fact that the angular momentum of two particles can assume any value which is a multiple of \hbar is not the only reason for the appearance of infinitely many irreducible parts in the representation (5.12) of the unitary subgroup. In many if not most cases, the existence of several pairs of particles, or even of three or more particles, is possible at energy E . All this multiplies the number of irreducible parts of the representation which appears in (5.12) so that not only an infinity of irreducible representations are present in the $D(U)$ of (5.12), but every one occurs several times—infinity many times if three free particles are possible at the energy E . All this complicates the application of invariance principles in the continuous spectrum, that is, for collision problems. The complications are most severe if three or more free particles can be present at the total energy E of the system, and have been attacked only recently with success.⁴⁹

⁴⁸ We mean by a "true Lorentz transformation" the transition to a moving coordinate system, without rotation. The corresponding matrices in four-dimensional space are symmetric, the A of (4.34) which correspond to them are Hermitian.

⁴⁹ A. J. Macfarlane, *Rev. Mod. Phys.* **34**, 41 (1962); G. C. Wick, *Ann. Phys. (N. Y.)* **18**, 65 (1962).

There are two principal sets of phenomena relating to the continuous spectrum to which invariance principles can be applied most advantageously. The first of these concerns the angular distribution of disintegration products, and even more importantly, the correlations between the directions of motion of successively emitted particles of disintegration. The second area in which invariance principles proved very useful concerns the angular distribution of products resulting from collisions, that is the angular distributions in scattering and reaction processes.

There are good review articles on both these applications⁵⁰ and we shall restrict ourselves to a few remarks concerning a disintegration process of the most simple character. It must be admitted that, even then, the initial state is difficult to define in the framework of orthodox quantum mechanics. One should somehow express the fact that the system—usually a radioactive nucleus—has not yet disintegrated and there is no “physical quantity” with a definite operator attached to it, which would measure the degree of disintegration. However, we shall use only some simple invariant-theoretic properties of the initial state which are well defined: that the spatial components of the momentum are zero, that it has a definite angular momentum j , and that we are dealing with a polarized nucleus, i.e., the component of the angular momentum in a definite direction, usually taken as the direction of the Z axis of the coordinate system, has a definite value m . The mathematical expressions for these conditions are very similar to those considered for the discrete case. The principal difference is that no explicit assumption concerning displacements along the time axis are made, i.e., that the energy is not specified. In fact, the specification of a definite value for the energy would render the properties of the system independent of time whereas those of a disintegrating particle clearly depend on time.

The expressions for the specified properties of the state vector ψ_m^j are

$$U_{(h_z, 1)}\psi_m^j = \psi_m^j, \quad (5.13)$$

and that ψ_m^j is a member of a set of state vectors ψ_n^j with $m = -j, -j+1, \dots, j-1, j$ for which

$$U_{(0, u)}\psi_n^j = \sum_n D^{(j)}(u)_{n'n} \psi_n^j \quad (5.14)$$

holds. These conditions do not completely specify ψ_m^j , something like the energy condition (5.5) would yet be necessary. Instead of this, it is postulated that a very large time displacement bring ψ_m^j asymptotically into a form in which it can be well described in the product space of two irreducible representations, i.e., in which two particles are present. The state vectors in that space can be labeled, most conveniently, by the spatial components $\mathbf{p}_1, \mathbf{p}_2$ of these particles, and by

their polarization variables m_1, m_2 . If we assume that the product particles have zero spin, the spin variables become unnecessary. The state vector will then assume, for a very large time displacement, the form

$$\psi_m(\mathbf{p}_1, \mathbf{p}_2) = \delta(\mathbf{p}_1 + \mathbf{p}_2) \chi_m(|\mathbf{p}_1 - \mathbf{p}_2|, \Omega), \quad (5.15)$$

where Ω is the direction of $\mathbf{p}_1 - \mathbf{p}_2$. The delta-function dependence of ψ_m follows from (5.13) because of (4.36a) in which the components of $\mathbf{p}_1 + \mathbf{p}_2$ have to be substituted for those of \mathbf{P} . The function χ_m has, in typical disintegration processes, a sharp maximum as function of $|\mathbf{p}_1 - \mathbf{p}_2|$, but this is not relevant for our purposes. However, the dependence on Ω , i.e., the angular distribution of the relative momentum $\mathbf{p}_1 - \mathbf{p}_2$ can be deduced from the fact that the operation of (5.14), i.e., the rotations, commute with time displacement. Hence, (5.14) must be valid also for the ψ_m of (5.15) and its partners, obtained from the ψ_n^j by time displacement.

It follows from the physical meaning of the momenta $\mathbf{p}_1, \mathbf{p}_2$ of the two particles that these transform under rotations as vectors. $U_{(0, u)}$ is applied to a ψ , as in (5.14), the variables $\mathbf{p}_1, \mathbf{p}_2$ are replaced therein by the vectors which are obtained from them by the rotation which corresponds to u . This is accomplished most easily by introducing matrices h_1 and h_2 as variables instead of \mathbf{p}_1 and \mathbf{p}_2 . These matrices are formed according to (4.34a) with the components of \mathbf{p}_1 and \mathbf{p}_2 for x, y , and z , and 0 for t . We then have

$$U_{(0, u)}\psi(h_1, h_2) = \psi(u^\dagger h_1 u, u^\dagger h_2 u). \quad (5.16)$$

Setting first Ω parallel to the Z axis, and using a diagonal u , one concludes that only χ_0 is different from zero if Ω is parallel to Z . Applying then (5.14) to a u which diagonalizes the matrix $h = h_1 + h_2$ one sees that χ_m , as function of Ω , is proportional to $D^{(j)}(u)_{0m}$, where u is a unitary matrix which diagonalizes the matrix h . The angular distribution is then proportional to $|D^{(j)}(u)_{0m}|^2$.

The preceding calculation, leading to a concrete result, was carried out to show how such results can be obtained for one problem, involving the continuous spectrum as defined before, from the very general premises from which we started. It is not a particularly useful result, not only because we assumed that the particles produced by the disintegration have zero spin—this assumption can be easily avoided without encountering serious difficulties. Its principal weakness is that it refers to a state which is “polarized,” i.e., in which a component of the angular momentum has a definite value. This is, of course, difficult to achieve experimentally. More useful are angular correlations between successive disintegrations, that is, directions of three or more disintegration products. The reason we chose the present example is that the formulas for such cases, though in practice well established, require for their derivation from our basic assumptions more lengthy considerations than were necessary in the case considered.

⁵⁰ See, for instance, J. M. Blatt, and L. C. Biedenharn, *Rev. Mod. Phys.* **24**, 258 (1952); L. C. Biedenharn and M. L. Rose, *ibid.* **25**, 729 (1953).