SCALAR QUANTUM FIELD THEORY
IN A CLOSED UNIVERSE OF CONSTANT CURVATURE

Stephen Albert Fulling

A DISSERTATION

PRESENTED TO THE

FACULTY OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY THE

DEPARTMENT OF

PHYSICS

May, 1972

CONTENTS

FIGURES

Theoretical physicists talk much about "fields," but . . . what is a field? Years ago when I was first getting into the subject, and had an intense horror of . . . words which seemed to contribute . . . not at all to the mathematical or observational content of the theory, I confronted Fermi with this blunt question. I was very much relieved that the great man did not respond that this was a stupid or irrelevant question, and that everyone that had any right to think about such matters knew of course what a field was, but appeared to take the question seriously. He stopped to think for a moment, and then responded that, in his judgement, "field" meant most basically the "occupation number formalism." . . . I felt much reassured, if not technically enlightened.

-- I. E. Segal *

---

* Segal (1968), pp. 31-32.

INTRODUCTION

In elementary expositions of quantum field theory (e.g., [Mandl]) physical space is often treated, not as an infinite Euclidean space, but as a large finite "box". Usually periodic boundary conditions are assumed; thus one is effectively identifying opposite points on the boundary of the box and hence imagining the universe to be a flat three-dimensional torus. In a field theory constructed in this way the energy and momentum operators have discrete spectrum and normalizable eigenvectors; for certain purposes this offers conceptual and technical advantages. However, a torus model is not invariant under rotations or Lorentz transformations, and hence some of the most powerful tools of the modern theoretical physicist are not applicable.

Apparently it was Gutzwiller (1956) who first suggested in print that the advantages of a finite universe could be combined with the advantages of a high degree of geometrical symmetry by considering a closed universe of constant curvature.[1] (See also the remarks of Chernikov and Tagirov

---

[1] Following the traditional language of cosmology, we shall call a space-time manifold of dimension s + 1 <u>closed</u> (or <u>finite</u>) if the slice of space corresponding to a given time is a compact manifold, with the topology (usually) of an s-dimensional sphere. The range of the time coordinate is allowed to be infinite. For the definition of "constant curvature" (which we shall never use) see Sec. III.4 below or Sec. 1 of Gutzwiller's paper.

(1968).) There is a space-time (essentially unique -- see Sec. I.1) which satisfies the two criteria of being spatially closed and of having a group of global isometries with the same number of parameters as the Poincaré (inhomogeneous Lorentz) group (i.e., ten parameters in the four-dimensional case). This is the de Sitter universe, which will be described in detail in Chapter I. The associated group, the de Sitter group, is not the same as the Poincaré group, but rather stands to the Poincaré group in a relation like that of the latter to the inhomogeneous Galilei group. Hence one would expect de-Sitter-invariant physics to have as many complications over special relativistic physics as relativistic has over nonrelativistic physics, and one is not disappointed. In de Sitter space spatial translations in different directions do not commute, and time translations are hard even to define. These facts are related to the nonvanishing curvature of de Sitter space, which introduces a modification in the local geometry and physics which has no parallel in the torus model. These and similar complications make Gutzwiller's proposal harder to implement than one might think at first.

In the past decade torus models have been important in the development of constructive quantum field theory. Jaffe (1965) constructed a self-interacting scalar field (with interaction Lagrangian proportional to $\phi(x)^4$) in a two-dimensional box, and this and other kinds of space cutoffs have been used in the process of constructing the ϕ^4 field theory in infinite space (see bibiliography in Glimm and Jaffe (1970)).

Here the purpose of the cutoff is not to make the momentum spectrum discrete but to make the volume of space finite. Then some of the mathematical pathologies characteristic of Euclidean-invariant field theories (see Wightman (1964), Sec. 6) do not arise. (In formal calculations these difficulties show up in the form of divergent integrals over infinite space.)

The method of construction of the ϕ^4 theory in infinite space is not Lorentz-covariant, and one of the hardest tasks of the subject has been to prove that the model finally obtained really possesses the symmetry under Lorentz transformations which it intuitively ought to have. The idea arose, therefore, of constructing an interacting field in the two-dimensional de Sitter space in analogy with the work of Jaffe (1965), but taking care to maintain invariance under the de Sitter group explicitly throughout. This theory would involve the radius of the universe, R, as a parameter. In the limit $R \longrightarrow \infty$ one could hope to recover a clearly Lorentz-invariant theory in flat infinite space by methods like those used in ordinary constructive field theory. This dissertation was intended to accomplish at least the preliminary steps in this program. The Lorentz invariance of the ϕ^4 theory has recently been proved by Streater (1971) by a different method.[2] The original motivation for the present work is thereby weakened. One can

[2] At the time of this writing an error has been discovered in the work which culminated in Streater's proof, so at least for the time being the problem must be regarded as still open.

still hope, however, that de Sitter space will provide an alternative route to the ϕ^4 model with some technical or pedagogical advantages.

However, there is another reason, at present more urgent, for studying quantum field theory in de Sitter space. This is the current interest in possible significant effects due to the quantum nature of matter in relativistic cosmology and astrophysics. A theoretically coherent treatment of particles in interaction with a given gravitational background requires a general theory of quantum fields in Riemannian space-times. De Sitter space is much more like the flat space-time of special relativity than the most general universe is, but it nevertheless possesses some of the features of the general case, such as curvature. Consequently, a well-understood quantum field theory for de Sitter space should be a comparatively easy intermediate goal, and it should also be instructive in relation to the general problem. In this context it is not necessary to consider self-interacting fields to encounter interesting problems; the so-called "free" field already presents considerable problems of mathematical definition and, even more, of physical interpretation.

In the course of three years the research reported here has progressively reoriented itself from the first to the second of these two problem areas. Thus it turns out that this study of the free field in de Sitter space has very little to say about interacting fields in de Sitter and Minkowski space, the concern of constructive quantum field theory, but quite a bit to say about free fields in general space-times, a concern of general relativity theory. For the reader's guidance the work and the document will now be described in more detail.

1. Program.

In this dissertation quantization of a massive neutral scalar field without self-interaction defined on a space-time manifold with given metric is studied, with emphasis on the two-dimensional spatially finite de Sitter universe. As indicated above, the work has been conducted with two purposes in mind:

(1) To lay groundwork for the rigorous study of model interacting fields in de Sitter space, envisioned as a covariant method of introducing a spatial cutoff in constructive quantum field theory.

(2) To make a critical examination of the applicability of the concepts of quantum field and particle in the context of curved space-time -- a problem of current interest in relativistic astrophysics and cosmology.

Let us elaborate on the first point. Note first that the possibility of doing calculations in quantum physics at all is based on the possibility of abstracting a nearly closed system from its surroundings and idealizing its environment. For instance, a scattering process is treated as an encounter between two particles in an otherwise empty universe, even though in reality it takes place in a bubble chamber close to many rapidly moving atoms, which give rise to electromagnetic fields, and so on. The actual problem is hopeless to solve, or even to pose. One must assume (and it is usually taken to be too physically obvious to deserve mention) that the differences between the real problem and the tractable problem are minuscule for any quantities of interest calculated. This is a basic presupposition of the subject, of the same sort as the dogma of stability in classical mechanics as described by [Abraham], pp. 3-4.

The same reasoning is involved in the notion of a cutoff in quantum field theory. A change in the global structure of the spatial universe from a Euclidean space to a torus or a sphere (as in de Sitter space) produces a drastic change in the mathematical structure of the quantum field theory. (Indeed, that is the reason for introducing the cutoff.) Nevertheless, one proceeds on the expectation that if the dimensions of the finite space are taken large enough, the values of observed physical quantities will be indistinguishable from those calculated in an infinite-space theory. For instance, if one works in a box of cosmological size, say length $L = 10^{27}$ cm, the energy and momentum spectra will be discrete, but with spacings of the order of $2\pi\hbar/L = 10^{-16}$ sec^{-1} $= 10^{-43}$ ergs. This quantity is more than 25 orders of magnitude smaller than the smallest energy differences commonly measured in physics, such as the Lamb shift and the K_1-K_2 mass difference. So this momentum quantization, a striking qualitative difference between the finite- and the infinite-space field theories, does not have observable consequences.[3]

In the case of de Sitter space one has introduced not only a global periodicity but also a genuine local effect, curvature. But this, like the weak electromagnetic field near an atom of hydrogen in the bubble chamber, can safely be assumed to be negligible -- if the radius of the universe is sufficiently large. Indeed, the universe we live in is not really a Minkowski space, or a de Sitter space either, but something more complicated; yet we expect laboratory particle events to be adequately described by theories based on the infinite flat space-time of special relativity. With equal justification we might use a de Sitter space-time. (In fact, the real universe may well be spatially closed, in which case a de Sitter space may be the better approximation.) In other words, we have good physical reasons for believing that

(1) any theory of physical processes on the microscopic level ought to come in both de-Sitter-space and Minkowski-space versions;

---

[3] From a certain point of view it can be argued that within a sufficiently small region of space-time the global structure of the space (as opposed to local curvature) should have no effect at all on the dynamics of the quantized field. This will be discussed in Secs. IX.4 and IX.7.

(2) these should make the same experimental predictions in the limit of large radius of the de Sitter space, and, in fact, should already be practically indistinguishable by the time the radius reaches a magnitude characteristic of the actual universe.

This, then, is the physical reasoning underlying the idea of a covariant cutoff. (From the point of view of mathematical technology, of course, the observation that the real universe may be finite is superfluous.) What would one expect to happen on the level of mathematical apparatus as the radius tends to infinity? From general information on the nature of field theories in infinite space (Haag's theorem, etc. -- see Haag (1955) and Wightman (1964)) and the experience with noncovariant cutoffs (Glimm and Jaffe (1970)), we know that it would be unrealistic to expect the state vectors and field operators of a de-Sitter-space theory (with nontrivial interaction) to converge to those of a Minkowski-space theory. The most one can hope for is that the expectation values of the field operators (or of an associated algebra of bounded observables) in some state might converge to distributions which can be interpreted as the vacuum expectation values of a Lorentz-invariant field theory; then the field theory can be reconstructed ([Streater-Wightman], Sec. 3.4). On the other hand, suppose that even this program fails. Then, in view of what was said above, one could still hope to take limits in observable quantities. If one could define an

S-matrix in de Sitter space (a nontrivial task -- see Sec. IV.3 below), an acceptable Lorentz-covariant S-matrix should be extractable from it. Then one would have successful field theories in de Sitter space, though not in Minkowski space, and the demands of practical physics would be satisfied.

The program of constructive field theory in de Sitter space thus falls naturally into two parts: the construction of field theories in de Sitter space and the taking of limits in hopes of recovering field theories in Minkowski space.

We shall approach the first step in two ways, the axiomatic and the constructive. In the axiomatic approach (Chapter IV) an attempt is made to preserve the crucial role played by the symmetry group in standard axiomatic quantum field theory. Then a "free" field in de Sitter space will be constructed (Chapter V, parts of Chapters VIII and X) for comparison. (The comparison will cast doubt on the central role assumed at first for the group -- see next section.) No attempt will be made here to construct a self-interacting field (one satisfying a nonlinear field equation).

A necessary preliminary to the second step seems to be a clarification of the relationship between unitary representations of the de Sitter group and those of the Poincaré group, since the limiting procedure is expected somehow to produce a representation of the one out of a representation of the other. This is the problem of contraction of group representations, discussed in Chapters II and VI.

Any physicist who considers this subject must have some qualms over whether the construction of "the scalar field" in a hypothetical space-time is a well-defined problem, operationally. After all, we live in only one universe. A theory appropriate to a different one cannot be experimentally tested. What, then, does it mean to say that such a theory is correct? Of course, if one is interested in a quantum field theory in de Sitter space only as a mathematical tool, a temporary construction leading eventually to a field theory in flat space, then this question is irrelevant. One feels, however, that the fundamental ideas of quantum field theory should have a unique extension to curved space, much as the Schrödinger equation provides us with what we believe to be the "true" behavior of a particle in an arbitrary potential, even a potential which it is impossible to produce experimentally. From this point of view it was natural to demand a physical interpretation for a quantized field in de Sitter space, and the author resolved to settle this question for the "free" field before proceeding to the construction of interacting fields.

In the absence of experimental tests, one is driven to "internal" criteria for the goodness of a theory. It should be the most natural generalization to the new context of the successful theories with which we are already acquainted. It is more important to generalize the physical ideas than the superficial properties of the mathematical apparatus. To be convincing the result must be as nearly as possible unique; the theory should be somehow "compelling" in its claim to be the right theory. Of course, this goal cannot be completely attained; what is a compelling argument to one person may be unconvincing to others, and at some points arbitrary decisions based on taste will have to be made.

The present author found a variety of approaches to quantum mechanics and field theory in de Sitter space already in the literature (see Sec. I.5). It seemed that these authors either did not carry the project through to the construction of a definite quantum theory (a Hilbert space of state vectors, etc.) or did not realize that their prescriptions were not the only ones possible. In particular, the present author was not convinced that an approach which assigned to the de Sitter group a role as close as possible to that of the Poincaré group in standard field theory was physically justified. He resolved to apply the criterion that a model field theory in de Sitter space must fit coherently into a physically convincing theory of field quantization in general Riemannian[4] space-times. (In the general case, of course, there is no symmetry group; in special cases there may be groups with fewer than the maximal number of parameters.)

This was the origin of the second, and eventually dominant, theme of the research. In the absence of evidence to the contrary, it was assumed as a working hypothesis that the

[4] See footnote 1 of Chapter III.

canonical field quantization precedure of the textbooks should apply to the "free" field on an arbitrary Riemannian manifold. This field satisfies a linear equation (with variable coefficients), so there is no reason to expect the canonical commutation relations to break down, as presumably happens in some nonlinear theories. Also, canonical quantization brings us much closer to the goal of uniqueness than a more general axiomatic scheme, and it is noteworthy that in flat-space theories the canonical commutation relations are closely related to the particle interpretation. It turns out, however, that the canonical structure is not sufficient to determine a particle interpretation uniquely. These matters are studied in Chapters VII through X, and the results are summarized in the next section of the Introduction.

The extension of the scope of the investigation from de Sitter space to general metrics improves the situation with regard to experimental relevance. Field quantization is now being applied to various astrophysical and cosmological problems (e.g., Ruffini and Bonazzola (1969); Parker (1968); Zel'dovich (1970)). It is hoped that the present work will help to put the methods used on a more solid base of general theory. Thus there is a connection between a general theory of field quantization on Riemannian manifolds and actual observations.

2. Results.

The contributions of the work presented here to this program can be summarized in several categories.

(1) The author hopes that this dissertation will be a useful reference for a subject on which previous work has been widely scattered and marked by a certain lack of communication. In this spirit many facts and formulas about de Sitter spaces and groups, various coordinate systems, solutions of the field equation in the two-dimensional case, etc., have been collected; much bibliographical information has been passed along, even on topics which are treated only tangentially here; and (as far as practical) efforts have been made to keep the exposition comprehensible to general relativity theorists without much background in field theory and mathematical physics and vice versa.

(2) Three approaches to the quantization of the "free" field in two-dimensional de Sitter space have been considered: (A) second quantization of a single-particle theory in which the states of the particle support an irreducible representation of the de Sitter group (following Tagirov *et al.* (1967)[5] and Nachtmann (1968b)); (B) canonical quantization of the field in a region where the metric can be regarded as static; (C) canonical quantization

[5] See also Chernikov and Tagirov (1968).

in the whole space, regarded as a homogeneous universe of time-varying radius. (See Chapters V, VIII, and X.) In all cases a Fock-like construction of a Hilbert space for the quantum theory has been attained, but the resulting "particle" interpretations are incompatible (see (4) below). In particular, pair creation from vacuum occurs in theory (C) (at least according to one definition of particle number observables -- see Secs. X.5-6) and not in the others. The representations for the theories (A) and (C) can be shown to be unitarily equivalent (Sec. X.9). Some observations on the field in the spatially open de Sitter universe are included (Secs. III.6 and V.8), with emphasis on the relation between the structure of the space and the self-adjointness of the Hamiltonian of the field theory.

(3) The canonical formalism of (scalar) field quantization has been developed for an arbitrary metric (Chapter VII). For a static metric one can proceed to the standard construction of Fock space, with its particle interpretation (Chapter VIII); the particle number is conserved in such a theory. (See, however, points (4) and (5) below.) In the more interesting case of an "expanding universe" (Chapter X), where pair creation is expected, there is no obvious unique analogue of the Fock representation.[6] In Sec. X.5 a Fock representation at each time

[6] Cf. Parker (1966, 1969).

is defined in a way which most plausibly generalizes the usual condition of positive energy, and the representations for different times are proved unitarily equivalent in the case of a closed two-dimensional universe. This result fails, however, for infinite space and probably for higher dimensions.

(4) Significant negative results appeared when the general theory was applied to simple models. The procedures mentioned in (3) for constructing a representation of the fields as operators in a Hilbert space are based on analogies to the case of the free field in flat space which hold when the metric of the Riemannian space-time has a special form (static or rigidly expanding). If the metric has such properties when expressed in several different coordinate systems, the quantization procedure is not unique. In particular, one can arrive in this way at anomalous representations of the free field in flat space (Chapter IX and Sec. X.2). For the same reason the three methods of quantization for de Sitter space cited in (2) yield different results. These ambiguities affect the physical concepts of the vacuum, particles, and energy density. In Secs. VII.7 and X.7 it is argued that similar problems already arise in principle in flat-space quantum field theories with external potentials, but that in practice these questions are dismissed as operationally meaningless, because an

unambiguious scattering interpretation of the theory is available in the cases of interest.

(5) The author concludes from (4) that the requirement of a unique physical representation of the fields is unrealistic in the general case. This point of view is consistent with the unitary inequivalence of the positive-energy representations at different times in the most general expanding universe. It is suggested that one must work with an abstract algebra of observables associated with the field (Secs. IX.4-5). The quanta associated with a given representation need not have a direct physical interpretation as particles. On the other hand, some representations are probably more directly related to practical observables than others. Some speculation is offered (Sec. IX.7) as to how particle phenomena may arise in a field framework in a way which depends on global boundary conditions. A deeper analysis of the physical process of observation is called for to clarify the physical interpretation of field theory. In Sec. X.8 a proposal is made for a general definition of particle observables, based on a condition of positive energy relative to a geodesic hypersurface, which in flat space reduces to the standard theory. (In a general static space this prescription does not coincide with that suggested by the Fock representation

(see (3) above).[7]) When applied to two-dimensional de Sitter space this ansatz predicts a nonzero but very small rate of creation of pairs out of the vacuum (Sec. X.10).

(6) The axiomatic approach has been applied to field theory in de Sitter space (Chapter IV), with inconclusive results. Most of the standard axioms generalize easily and are satisfied by the free field theory (A) (see Sec. V.6). The most acceptable replacement found for the spectral condition is the general condition of positive energy already referred to in (3) and (5). This can be satisfied by a theory of type (C), which seems to the author to be the most convincing theory physically. In the two-dimensional case theories (A) and (C) are unitarily equivalent, but the invariant vacuum state loses its physical significance if the proposed definition of particle observables is adopted.

(7) The formal correspondences between the irreducible Hermitian representations of the Lie algebra of the de Sitter group and those of the Lie algebra of the Poincaré group are discussed in some detail (Chapter II). It is pointed out that in some contexts, when a discrete index becomes continuous in the contraction, an *ad hoc* distinction between even and odd indices seems to be required in order to

---

[7] Note that from the standpoint of the algebraic approach to quantum theory this does not necessarily mean that the Fock representation is useless.

reproduce the representations of the contracted group completely. This and other unsatisfying or mystifying features of the formal approach to contraction of representations can be clarified by studying, as a concrete example, the contraction of the natural action of the group on the scalar functions on its coset space relative to the contraction subgroup. (In the case of the de Sitter group, this is the universe we are studying.) This idea is developed in detail for the rotation group in Appendix C and discussed briefly for the two-dimensional de Sitter group in Chapter VI.

3. To Prevent a Case of Mistaken Identity.

A few words are in order about what this dissertation does not do. First, there is no attempt to quantize the gravitational field. The metric of space-time is always assumed to be given. In fact, it does not even have to satisfy the Einstein equations to qualify as an interesting model. By the same token, this work is not related to attempts (Deser (1957); Isham et al. (1971) and related papers of A. Salam and coworkers) to cure the diseases of Lagrangian quantum field theory by including an interaction with the gravitational field. The de Sitter model is relevant to only one class of the divergences of field theory[8], those connected with the infinite volume of space, and, as explained above, it is intended in this connection only as a way station on the road to a theory in flat space. Finally, this work has nothing at all to do with attempts to relate the de Sitter group to internal symmetries, the mass spectrum of elementary particles, and so on (e.g., Roman and Aghassi (1966), Böhm (1966), Burcev (1968), Bakri (1969), Vigier (1969), Tait and Cornwell (1971)).

4. Mechanics.

In recognition of the fact that very few people will read this entire volume, every effort has been made to facilitate browsing and random access reference. The research reported in this dissertation has a major theme, the general theory of quantum fields in Riemannian space-time, and a minor theme, the contraction of Lie groups (with its physical application to physics in de Sitter space). The reader interested only in the second will want to stop reading after Chapter VI. The reader interested only in the first should omit Chapters II and VI and Appendix C; indeed, if he is not particularly interested in de Sitter space he might want to start with Chapter VII. The reader in search of ideas (not final answers!) concerning the fundamental question of the meaning of quantum field theory in

---

[8] It has been suggested that the finite volume of the de Sitter universe may be helpful in treating the infrared divergences (E. P. Wigner, private communication). However, not much attention has been paid to massless fields in the present work.

curved space-time will find the importance of the sections to be, on the average, inversely proportional to the density of mathematical symbols. In addition to the qualitative sections, he should read Secs. V.3-6 and X.9-10, where some relevant explicit results for the case of de Sitter space are obtained.

Certain background material has been relegated to appendices. Two appendices contain original material: Appendix C is a digression, a preliminary exercise for the work of Chapters II and VI, and Appendix G is an outgrowth of Secs. IX.5 and IX.7. For the reader who wants to read everything in its logical order the following sequence is recommended: Secs. A.1 and A.2 before Chap. I; Sec. B.2 before Sec. I.4; Sec. A.3 after Chap. I; Apps. B and C before Chap. II; App. D before Chap. III; App. E before Chap. IV; App. F before Chap. VIII; App. G after Chap. IX.

In bibliographical references books are indicated by author's name in brackets, and journal articles, etc., are indicated by author's name and date. "Eq. (1.3)" means the third equation of Section 1 of the current chapter or appendix, "Eqs. (I.1.3-5)" means equations 3 through 5 inclusive of the first section of Chapter I, and so forth.

We shall always use units such that $\hbar = c = 1$. The metric of space-time has one plus sign and three minuses on its diagonal. A^* and $A^\dagger$ are respectively the complex conjugate and the adjoint of A. The letter x denotes sometimes a space-time variable (dimension $n = s + 1$) and sometimes a space variable (dimension s) ; this ambiguity seems less confusing than use of boldface for a quantity which is not, in general, a vector.

5. Acknowledgements.

This research was done under the gentle and patient guidance of Prof. A. S. Wightman and the inspiration of his high standards. We have had many long conversations about the questions it raised. Many of its good points, including the fundamental idea that the behavior of solutions of the field equation in de Sitter space should be interpreted in terms of pair creation, stem from his suggestions. On the other hand, he should not be held responsible for my speculative remarks on the physical interpretation of field theory and the future development of the subject.

The enthusiasm and support of Remo Ruffini and Leonard Parker during a critical period were very important. The last stages of the work benefited greatly from discussions with Prof. Parker.

During the past three years I have discussed my work with many other people. I am especially grateful to B. B. Godfrey, M. C. Reed, and G. J. Zuckerman, who allowed me to pump them frequently for information in their fields of expertise, and to B. Simon, who long ago undertook my general education in mathematical physics. Some of the other conversations were with V. Bargmann, G. Börner, L. Ford, J. Friedman, F. Goldrich, B.-L. Hu, A. Jaffe, K. Kuchař, J. Lewis, C. Misner, F. Narcowich,

C. Newman, C. Rhoades, J. Roberts, L. Rosen, S. Ström, E. Toton, W. Unruh, F. Weinstein, E. Wigner, and J. York.

Among the authors listed in the bibliography I am especially indebted to K. Hannabuss (through J. Lewis), A. Jaffe, M. Lévy-Nahas, L. Parker, and M. Reed for copies of their unpublished dissertations. I. T. Todorov helped me to obtain a copy of a crucial unpublished reference (Tagirov et al. (1967)).

Finally, I am indebted to P. Crean, C. E. Rhoades, Jr., V. Bearg, and especially F. Weinstein in various ways in connection with the computer program used to produce this manuscript.

This work was supported by a National Science Foundation Graduate Fellowship and by AFOSR Contracts AF49-638-1545 and F44620-71-C-0108. Reproduction will be handled under NSF Grant GP30799X, through the generosity of Prof. J. A. Wheeler.

Chapter I

DE SITTER SPACES AND DE SITTER GROUPS

We shall study a spatially finite space-time of constant curvature of dimension n and the symmetry group associated with it. Of course, the case $n = 4$ is most relevant to physics. It is useful to consider also $n = 2$ and $n = 3$, however. The importance of lower-dimensional models in the constructive theory of interacting fields is well known (see Glimm and Jaffe (1970) and references cited there); there one postpones tackling the ultraviolet problem in its full fury while dealing with other aspects of the subject. A rigorous study of interacting fields in curved space would presumably also start in two dimensions. On the other hand, in some contexts the fundamental problems are the same in all dimensions, but the incidental mathematical complexity increases with n. Then the modest approach of starting with dimension 2 and generalizing later to higher dimensions can be a help to both researcher and reader; it allows the basic problem to stand out clearly amid a minimum of inessential algebraic complications. The main contribution of this dissertation is to bring up and, at least partially, to answer several questions of principle which arise and can be studied for the scalar field in two-dimensional space-time just as well as in more complicated cases. Throughout this chapter it will be easy enough to do everything for general n; later we shall usually specialize to $n = 2$ or $n = 4$.

In this chapter we use the terminology and facts concerning pseudo-orthogonal groups set forth in Appendix A.

1. The Closed De Sitter Space of Dimension n.

We define de Sitter space as a certain homogeneous space of the group $SO_o(1,n)$. Recall that if G is a group of transformations (not necessarily linear) on a set M, then M is called a homogeneous space of G if for every two points x and y in M there is an element A of G such that $y = Ax$. M can be identified with G/H, the space of left cosets of G relative to the subgroup H (stability group) of transformations in G which leave a given point of M invariant. (See [Hermann], pp. 3-4.) The n-dimensional Minkowski space is a homogeneous space under the n-dimensional translation group and thus a fortiori under the Poincaré group $ISO_o(1,n-1)$.

We consider the n-dimensional de Sitter group $SO_o(1,n)$ (see Appendix A). Of course, $\mathbb{R}^{n+1}$ is not a homogeneous space of this group, because only vectors with the same length in the metric F(x) (Eq. (A.1.2)) can be connected by transformations in the group. The n-dimensional submanifold M of $\mathbb{R}^{n+1}$ defined by the condition[1]

[1] The coordinates are with respect to an orthonormal basis.

$$F(x) = (x^0)^2 - \sum_{j=1}^{n} (x^j)^2 = -R^2, \tag{1.1}$$

where R is a positive real number, can easily be seen to be homogeneous. We shall call it the (closed) de Sitter space of dimension n. Its induced metric has the signature (+ - ... -) (n - 1 minus signs) appropriate to a space-time model of dimension n. The stability subgroup H of any point in M is isomorphic to the n-dimensional Lorentz group $SO_0(1,n-1)$.

The two-dimensional de Sitter space is sketched in Fig. 1; it is a single-sheeted hyperboloid. The drawing is slightly misleading, because it is hard to visualize the indefinite metric of the enveloping three-dimensional space. For instance, relative distances in the space in various directions may be quite different from what they seem to the eye. Also, contrary to appearance, all points and regions are equivalent in their curvature and other intrinsic geometrical properties, since the space is homogeneous. Geometrical matters will be discussed in more detail in Chapter III.

Consider a point O in M; for instance, the one with coordinates

$$O:\quad (x^0,\dots,x^{n-1},x^n) = (0,\dots,0,R). \tag{1.2}$$

A patch of the space around O whose linear dimensions are very small compared to R will be almost indistinguishable from a piece of n-dimensional Minkowski space (Fig. 2). It is clear that the

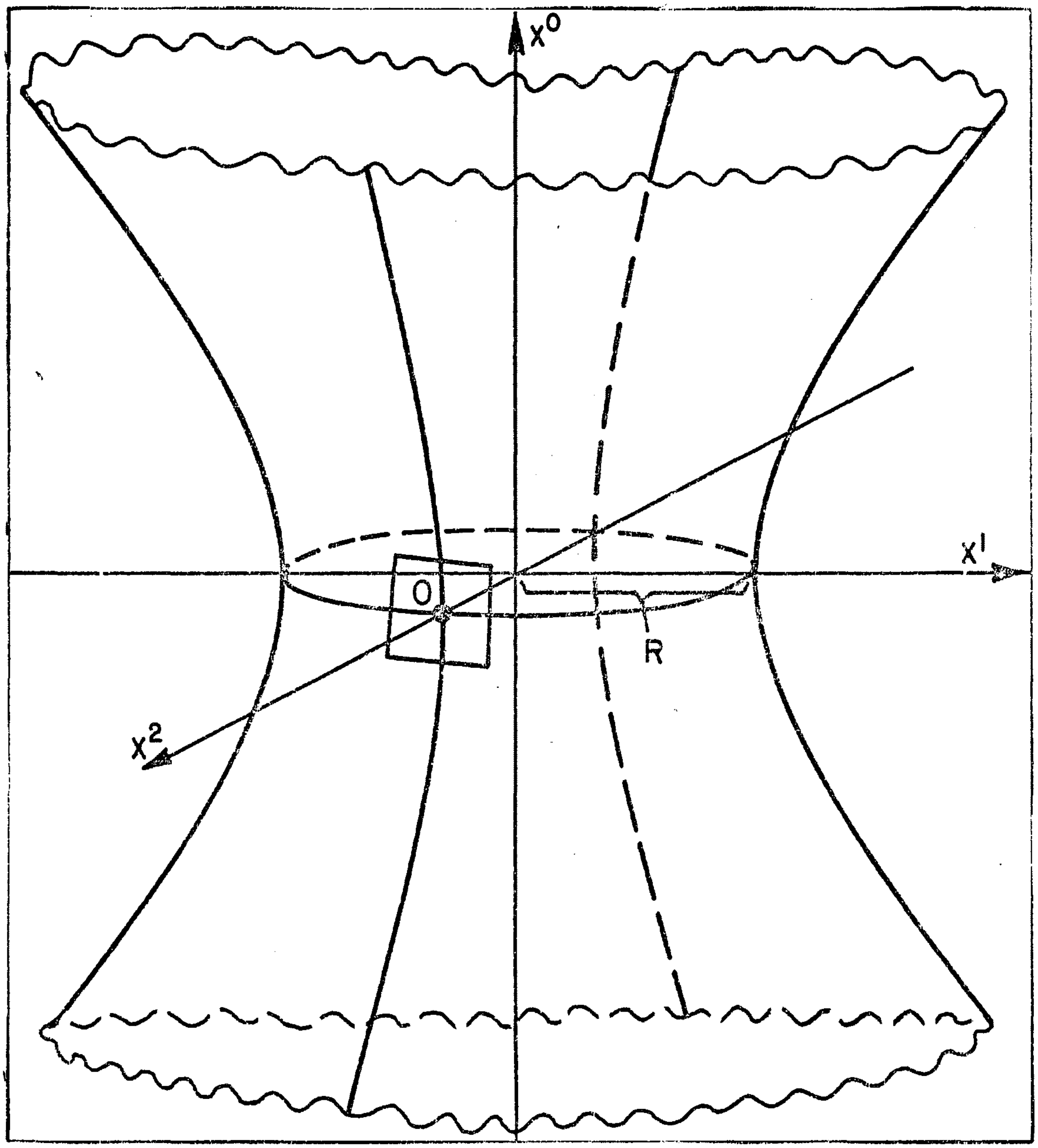


Fig. 1

Two-dimensional closed de Sitter space. The intersections of the surface with the planes $x^0 = 0$ and $x^1 = 0$ and a neighborhood of the point O ($x^2 = +R$) are shown. A left-handed coordinate system is used to make the orientation of Fig. 2 agree with the standard convention. (Figure adapted from Philips (1963).)

transformations which leave 0 invariant (the group H) act on this neighborhood similarly to the action of the n-dimensional homogeneous Lorentz group on flat space.

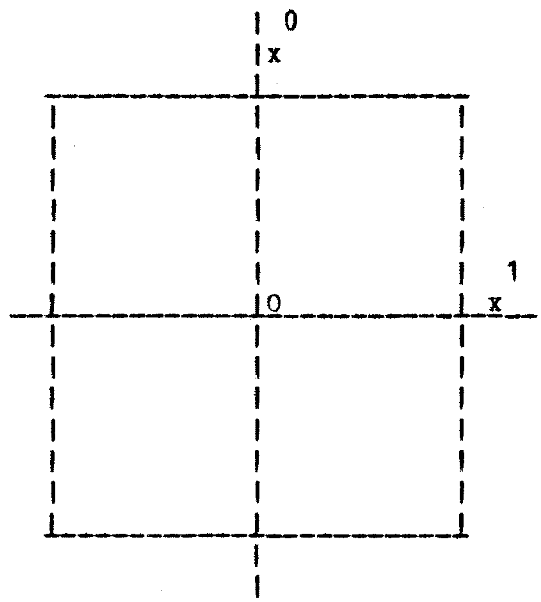


Fig. 2

A neighborhood of the point 0 in the limit of large R.

Physically, the n-dimensional de Sitter space can be interpreted as a universe of n - 1 spacelike and one timelike dimension, finite and closed in the spacelike directions. The submanifold $\{x \mid x^0 = 0\}$, which represents the spatial universe at one instant of time, is a hypersphere of radius R. (There is another way of looking at de Sitter space, according to which the spatial universe is open (infinite) -- see Sec. III.7.)

De Sitter space is a space of constant curvature, and in fact possesses a Lie group of global isometries (viz., $SO_0(1,n)$) with the maximal number of parameters, $n(n+1)/2$. (The latter implies the former, but constant curvature implies maximal symmetry only in a local sense.[2]) The only[3] space-times with this maximal symmetry are Minkowski space, closed de Sitter space, and open de Sitter space (a homogeneous space of $SO_o(2,n-1)$). The closed de Sitter space which we are considering is the only one in which space at each time has finite volume. A group-theoretical determination of all possible space-times of a certain high degree of symmetry was made in the thesis of Hannabuss (1969a) (see his Introduction and Appendix A). See also Calabi and Markus (1962).

2. The Contraction Process.

The geometrical properties of a small neighborhood of any point in de Sitter space are almost the same as those of a neighborhood in Minkowski space. The global properties of the spaces are very different, of course. In the Introduction we argued that one should not expect events which occur on a microscopic scale within a small region of space to be significantly affected by the structure of space at cosmological

---

[2] [Eisenhart], Sec. 27; J. W. York, Jr., private communication. An ordinary circular cylinder is an example of a two-dimensional space of constant curvature without a global three-parameter isometry group.

[3] This statement is not entirely accurate, because of the existence of "covering spaces" with homomorphic isometry groups. First, there is the possibility of considering a smaller space, defined by identifying antipodal points in de Sitter space -- see [Schrödinger], pp. 7-14, or Calabi and Markus (1962). In Chapter VII this space will be rejected for the purposes of field theory because it does not have a consistent time orientation. Conversely, the closed de Sitter space with n = 2 and the open de Sitter spaces of any dimension (see Secs. III.6 and V.8 below) have covering spaces from which they are obtained by identifying points.

distances. On the basis of this plausible physical idea we expect that to every physical theory set in Minkowski space there corresponds a theory in de Sitter space which for sufficiently large R (small curvature) gives virtually identical numerical predictions for observable quantities in any local process (such as elementary particle scattering).

In some sense the theory in de Sitter space should go over smoothly into the Minkowski theory as R approaches infinity. On the other hand, the mathematical structure of the theories is likely to be quite different. For instance, for all finite R one will have the de Sitter group $SO_0(1,n)$ as a symmetry group, but for $R = \infty$ the symmetry group will be the quite different Poincaré group, $ISO_0(1,n-1)$. In field theory, functions on a hypersphere, associated with discrete modes (spherical harmonic expansions), will be replaced in the limit by functions on Euclidean space, associated with continuum Fourier transforms. It is likely, therefore, that for the mathematical apparatus of a de Sitter theory the limit $R \longrightarrow \infty$ will not exist, except perhaps in some very "weak" sense. It may be hard to formulate clearly defined mathematical concepts and rigorous, nontrivial statements concerning the relationship between de Sitter theories (with finite but large R) and Minkowski theories. The problem just posed is a generalization of the problems surrounding the notion of <u>group contraction</u>.[4] By analogy we call the passage

[4] Segal (1951), pp. 254-257; Inönü and Wigner (1953); Saletan (1961); Inönü (1962); [Hermann], pp. 86-101; Bacry and

to the limit $R \longrightarrow \infty$ in a de Sitter theory the <u>contraction</u> of that theory to a theory in Minkowski space. In particular, it will involve a contraction in the usual sense of a representation of the de Sitter group to a representation of the Poincaré group. In this section we discuss contraction as it applies to the groups themselves. Contraction of unitary representations will be discussed in Chapters II and VI.

In investigating the limit $R \longrightarrow \infty$ our general approach will be to make a scaling transformation

$$\bar{x}^{\mu} = x^{\mu} \quad (0 \leq \mu < n), \qquad \bar{x}^{n} = \frac{1}{R} x^{n}, \tag{2.1}$$

starting from the orthonormal system of Eq. (1.1) as the unbarred system, and then to let $R \longrightarrow \infty$ in various equations and expressions connected with de Sitter space, hoping to obtain limits that make sense in Minkowski space. The mathematical rigor of the limiting processes varies with the context; sometimes the limits are intended as aids to the intuition rather than as proofs. It is helpful in keeping track of what is going on to assign to x^j the dimension of length and to note the dimensions of other quantities whenever they first arise. Since we take $\hbar = c = 1$ (dimensionless), we have

---

Lévy-Leblond (1968); Philips and Wigner (1968), pp. 664-666; Lévy-Nahas (1967, 1969).

$$[\text{time}] = [\text{length}] = [\text{mass}]^{-1}.$$

Thus R is a length, $A^j_{\ k}$ -- introduced in Eq. (A.1.3) -- is dimensionless, and $\bar{x}^n$ of Eq. (2.1) is dimensionless.

We begin with a formal algebraic counterpart[5] of the intuitive limit process indicated in Fig. 2. Substituting Eq. (2.1) into Eq. (1.1) and dividing by R^2 yields

$$-(\bar{x}^n)^2 + O(R^{-2}) = -1. \tag{2.2}$$

So the transformation (2.1) seems to be an appropriate one to make when studying the neighborhood of the point O characterized by $\bar{x}^n = +1$ (cf. Eq. (1.2)) in the limit $R \longrightarrow \infty$. Now consider the action of a generator $L \in \mathcal{L}(SO_0(1,n))$:

$$y^j = (L)^j_{\ k} x^k \qquad ((L)_{jk} = -(L)_{kj}). \tag{2.3}$$

Substitute from Eq. (2.1):

$$\bar{y}^\mu = (L)^\mu_{\ \nu} \bar{x}^\nu + (L)^\mu_{\ n} R\bar{x}^n, \tag{2.4a}$$

$$R\bar{y}^n = (L)^n_{\ \mu} \bar{x}^\mu. \tag{2.4b}$$

Therefore, the matrix $\bar{L}$ which represents L with respect to the new basis is

---

[5] Cf. Rosen (1965).

$$(\bar{L})^{\mu}{}_{\nu} = (L)^{\mu}{}_{\nu} , \tag{2.5a}$$

$$(\bar{L})^{\mu}{}_{n} = R(L)^{\mu}{}_{n} , \tag{2.5b}$$

$$(\bar{L})^{n}{}_{\mu} = \frac{1}{R}(L)^{n}{}_{\mu} . \tag{2.5c}$$

$(\bar{L})^{\mu}{}_{n}$ is a length and $(\bar{L})^{\mu}{}_{\nu}$ is dimensionless. So far we have merely written the Lie algebra of the de Sitter group in a different way.

Next, however, we let $R \longrightarrow \infty$ and require that the barred quantities remain finite. Since

$$(\bar{L})^{n}{}_{\mu} = \pm \frac{1}{R^2}(\bar{L})^{\mu}{}_{n} ,$$

in the limit we must have $(\bar{L})^{n}{}_{\mu} = 0$. Now $\bar{L}$ has the form (A.2.6) of a generator of the Poincaré group, and in view of Eq. (2.2) we can think of the de Sitter space M as having been replaced by the hyperplane $\bar{x}^{n} = 1$ (i.e., Minkowski space -- see Sec. A.2). We can extend the range of the variables $\bar{x}^{\mu}$ from $-\infty$ to $+\infty$, even though the limit process just described makes literal sense only for some finite neighborhood of O.

We have contracted a representation of $\mathcal{L}(SO_0(1,n))$ to a representation of $\mathcal{L}(ISO_0(1,n-1))$. The group representations generated by these representations of the Lie algebras are the representations which were used to define the respective groups in Appendix A. So we may speak of a contraction of the n-dimensional de Sitter group to the n-dimensional Poincaré group. The subgroup of $SO_o(1,n)$ with respect to which the contraction takes place is the n-dimensional Lorentz group, $SO_o(1,n-1)$, which leaves the point O invariant, and the resulting Abelian invariant subgroup of $ISO_o(1,n-1)$ is the n-dimensional translation group.

3. Lie Algebras of the De Sitter and Poincaré Groups.

From Appendix A we have the following commutation relations for the basis elements of the Lie algebras (Eqs. (A.1.10), (A.2.8), and (A.2.11)):

De Sitter group $SO_o(1,n)$

$$[L_{ab}, L_{cd}] = 0 \quad \text{if } a, b, c, d \text{ are all distinct.} \tag{3.1a}$$

$$[L_{ab}, L_{ca}] = \eta_a L_{bc} \quad \text{if } b \neq a \neq c. \tag{3.1b}$$

Poincare group $ISO_o(1,n-1)$

$$[\bar{L}_{\alpha\beta}, \bar{L}_{\gamma\delta}] = 0 \quad \text{if } \alpha, \beta, \gamma, \delta \text{ are all distinct.} \tag{3.2a}$$

$$[\bar{L}_{\alpha\beta}, \bar{L}_{\gamma\alpha}] = \eta_\alpha \bar{L}_{\beta\gamma} \quad \text{if } \beta \neq \alpha \neq \gamma. \tag{3.2b}$$

$$[T_\alpha, T_\beta] = 0. \tag{3.2c}$$

$$[\bar{L}_{\alpha\beta}, T_{\gamma}] = 0 \quad \text{if } \alpha \neq \gamma \neq \beta. \tag{3.2d}$$

$$[\bar{L}_{\alpha\beta}, T_{\beta}] = \eta_{\beta} T_{\alpha} \quad \text{if } \alpha \neq \beta. \tag{3.2e}$$

The numbers $(L)^{j}_{k}$ and $(\bar{L})^{j}_{k}$ of Sec. I.2 are the coefficients of the expansions of the elements L of the Lie algebras with respect to these bases. The basis elements transform contragrediently to the coefficients. Thus Eqs. (2.5) suggest relating the basic generators of the Poincaré group to those of the de Sitter group by

$$\bar{L}_{\alpha\beta} = L_{\alpha\beta} , \tag{3.3a}$$

$$T_{\alpha} = \bar{L}_{n\alpha} = \lim_{R\to\infty} \frac{1}{R} L_{n\alpha} , \tag{3.3b}$$

where T_{α} has dimension $[\text{length}]^{-1}$. If one substitutes Eqs. (3.3) into Eqs. (3.1) and divides by appropriate powers of R before taking the limit, one obtains Eqs. (3.2). Group contraction is usually discussed in terms of such a singular transformation on the Lie algebra (see references listed above).

From now on let us omit the bars on the Poincaré generators when there is no chance of confusion. It is helpful to distinguish the generators which play different geometrical-physical roles by different letters instead of indices. Also, in discussing unitary representations in a Hilbert space it is convenient to have Hermitian (rather than skew-Hermitian) generators. To avoid proliferation of notation let us carry out all these reforms at once.

Let capital Latin indices range from 1 to n - 1. In dealing with them we abandon the summation convention and relax the distinction between contravariant and covariant indices. When convenient we shall also use vector notation: $\vec{X} = (X^1,\ldots,X^{n-1})$. Let ϵ_{ABC} be the familiar completely antisymmetric tensor in three-space, with the properties

$$\sum_C \epsilon_{ABC}\,\epsilon_{CDE} = \delta_{AD}\,\delta_{BE} - \delta_{AE}\,\delta_{BD}, \qquad (3.4a)$$

$$\frac{1}{2}\sum_{B,C} \epsilon_{ABC}\,\epsilon_{BCD} = \delta_{AD}. \qquad (3.4b)$$

The indices of ϵ_{ABC} and δ_{AB} are not to be raised with g^{jk}.

With these conventions we define for the de Sitter algebra

$$H = P^0 = P_0 = -\,iL_{n0}, \qquad -P^A = +P_A = -\,iL_{nA},$$

$$K^A = -\,iL_{A0}, \qquad J_{AB} = -\,iL_{AB}, \qquad (3.5)$$

and for the Poincaré algebra

$$H = -iT_0, \qquad -P^A = +P_A = -iT_A,$$

$$K^A = -i\bar{L}_{A0}, \qquad J_{AB} = -i\bar{L}_{AB}. \tag{3.6}$$

In the physical case, n = 4, we let

$$J^A = -\frac{1}{2}\sum_{B,C} \epsilon_{ABC} J_{BC}, \tag{3.7a}$$

or, equivalently,

$$J_{AB} = -\sum_C \epsilon_{ABC} J^C. \tag{3.7b}$$

The J_{AB} generate the group of rotations in space around the point O (Eq. (1.2)). The transformations generated by the P^A can be identified with spatial translations. The J's and $\vec{P}$ together generate the subgroup of isometries which map the spatial universe $\{x \mid x^0 = 0\}$ into itself. In the de Sitter group this subgroup is isomorphic to SO(n). After contraction it becomes ISO(n-1). The J's and $\vec{K}$ generate the "local Lorentz group" of the point O (the group called H in Sec. I.1). The generator H behaves in the neighborhood of O like a generator of time translations, but these are not global time translations, as we shall see in Chapter III.

The sign conventions in these definitions have been chosen to agree with standard usage for the Poincaré group. A translation in the direction of the four-vector b has the form

$$U(b) = \exp\{i b^{\mu} P_{\mu}\} = \exp\{i b^{0} H - i \sum_{A} b^{A} P^{A}\}. \qquad (3.8)$$

The sign in Eq. (3.7a) is necessary for agreement with the standard notation in quantum mechanics (e.g., [Messiah], Chap. XIII) -- see Eq. (3.10a) below. It is not useful to distinguish contravariant and covariant components of $\vec{K}$ and $\vec{J}$, since they are not part of four-vectors.

Bacry and Lévy-Leblond (1968) have introduced an index-free notation for Lie brackets: If Y is a scalar and $\vec{X}$ and $\vec{Z}$ are vectors,

$$[\vec{X}, Y] = \vec{Z} \quad \text{means} \quad [X^{A}, Y] = Z^{A} \qquad (3.9a)$$

(for A = 1, ..., n-1). If $\vec{X}$ and $\vec{Y}$ are vectors and Z is a scalar,

$$[\vec{X}, \vec{Y}] = Z \quad \text{means} \quad [X^{A}, Y^{B}] = \delta_{AB} Z. \qquad (3.9b)$$

If n = 4 and $\vec{X}$, $\vec{Y}$, and $\vec{Z}$ are all vectors,

$$[\vec{X}, \vec{Y}] = \vec{Z} \quad \text{means} \quad [X^{A}, Y^{B}] = \sum_{C} \epsilon_{ABC} Z^{C}. \qquad (3.9c)$$

(Note that $[\vec{Y}, \vec{X}] = [\vec{X}, \vec{Y}] \neq -[\vec{Y}, \vec{X}]$.)

The commutation relations (3.1) and (3.2) for n = 4 can be written

$SO_o(1,4)$	$ISO_o(1,3)$	
$[\vec{J},\vec{J}] = i\vec{J}$	(same)	(3.10a)
$[\vec{J},\vec{K}] = i\vec{K}$	(same)	(3.10b)
$[\vec{K},\vec{K}] = - i\vec{J}$	(same)	(3.10c)
$[\vec{J},H] = 0$	(same)	(3.10d)
$[\vec{J},\vec{P}] = i\vec{P}$	(same)	(3.10e)
$[\vec{K},H] = i\vec{P}$	(same)	(3.10f)
$[\vec{K},\vec{P}] = iH$	(same)	(3.10g)
$[\vec{P},H] = i\vec{K}$	$[\vec{P},H] = 0$	(3.10h)
$[\vec{P},\vec{P}] = i\vec{J}$	$[\vec{P},\vec{P}] = 0$	(3.10i)

If n = 2 there is no $\vec{J}$, and $\vec{P}$ and $\vec{K}$ have only one component each; the Lie algebra reduces to Eqs. (3.10f,g,h). If n = 3, we let $J = - J_{12}$ and have in addition to Eqs. (3.10f,g,h) and (3.10d)

$$[J,K^1] = + iK^2, \qquad [J,K^2] = - iK^1, \tag{3.11a}$$

$$[K^1,K^2] = - iJ, \tag{3.11b}$$

$$[J, P^1] = + iP^2, \qquad [J, P^2] = - iP^1, \tag{3.11c}$$

$$[P^1, P^2] = iJ \quad \text{[or 0, respectively]}. \tag{3.11d}$$

We observe from Eqs. (3.10) that in a representation of one of the Lie algebras some of the operators are completely determined by the others. It follows that some of the commutation relations are redundant. For instance, we have the following propositions:

(1) If operators H, $\vec{K}$, $\vec{P}$, J_{AB} are given, to check that they form a representation of $SO_o(1,n)$ or $ISO_o(1,n-1)$ it suffices to verify Eqs. (3.10b,c,f,g,h) (or their analogues for general n).

(2) If operators H and $\vec{K}$ are given, a representation is obtained by defining J_{AB} and $\vec{P}$ through Eqs. (3.10c) and (3.10f), provided that H and $\vec{K}$ satisfy Eqs. (3.10b,g,h), or, equivalently,

$$[[K^\mu, K^\nu], K^\rho] = 0 \quad \text{if } \mu, \nu, \rho \text{ are distinct}, \tag{3.12a}$$

$$[[K^\mu, K^\nu], K^\mu] = K^\nu \quad \text{if } \mu \neq \nu, \tag{3.12b}$$

where $K^o = H$.

(3) In the case of $SO_o(1,n)$ the analogous statements are valid with the roles of $\vec{P}$ and $\vec{K}$ interchanged.

Proof: Given Eqs. (c,f), Eqs. (b,d,g,h) are equivalent to Eqs. (3.12), and the Jacobi identity shows that Eq. (d) follows from the other cases of Eq. (3.12a). Then, writing J_{Ao} for P_A, we calculate

$$[J_{\mu\nu},J_{\rho\sigma}] = -i\,[J_{\mu\nu},[K^{\rho},K^{\sigma}]]$$

$$= i\,\{[K^{\rho},[K^{\sigma},J_{\mu\nu}]] + [K^{\sigma},[J_{\mu\nu},K^{\rho}]]\}$$

$$= -\,\{[K^{\rho},(\delta_{\mu\sigma}K^{\nu} - \delta_{\nu\sigma}K^{\mu})] + [K^{\sigma},(\delta_{\nu\rho}K^{\mu} - \delta_{\mu\rho}K^{\nu})]\}$$

$$= i\,(\delta_{\mu\sigma}J_{\nu\rho} - \delta_{\nu\sigma}J_{\mu\rho} + \delta_{\nu\rho}J_{\mu\sigma} - \delta_{\mu\rho}J_{\nu\sigma}),$$

which comprises the remaining equations (a,e,i). At the third step we used

$$[J_{\mu\nu},K^{\rho}] = i\,(\delta_{\nu\rho}K^{\mu} - \delta_{\mu\rho}K^{\nu}),$$

which is still another way of writing Eqs. (3.12).

Statements essentially equivalent to these have been used in representation theory for a long time (see, e.g., Hirai (1962a), p. 84). The version given here, however, provides these algebraic facts with physical significance. In particular, consider the failure of statement (3) for $ISO_o(1,n-1)$. The infinitesimal generators of a representation of the Poincaré group by tensor or spinor functions,

$$[U(b,A)\Psi]_\alpha(x) = S(A)_\alpha{}^\beta \Psi_\beta(A^{-1}(x-b)),$$

are

$$H = i\frac{\partial}{\partial x^0}, \qquad P^A = -i\frac{\partial}{\partial x^A}, \tag{3.13}$$

$$K^A = -i(x^A\frac{\partial}{\partial x^0} + x^0\frac{\partial}{\partial x^A}) + w_{0A},$$

$$J_{AB} = i(x^A\frac{\partial}{\partial x^B} - x^B\frac{\partial}{\partial x^A}) + w_{AB}, \tag{3.14}$$

where the w's are infinitesimal operators of the representation S(A). The commutation relations are still satisfied if the w terms in Eqs. (3.14) are dropped. Then we have a direct sum of several copies of the usual representation by scalar functions (see Sec. A.3), where H and $\vec{P}$ are still given by Eqs. (3.13). The statement (3) tells us that this cannot happen for the representations of the de Sitter group. It is impossible to separate "space" and "spin"; in a representation with spin, H and $\vec{P}$ will contain spin terms. This is related to the fact (discussed in Chapter III) that the identification of exp(itH) and exp(itP_A) with translations is valid only in the neighborhood of the point O.

4. Casimir Operators.

The central elements of the universal enveloping algebra of $\mathcal{L}(SO_0(p,q))$ can be found from those of $\mathcal{L}(SO(n))$, $n = p + q$, which were determined by Gel'fand (1950). The result is that the invariants are the "scalars" (expressions with no free indices) which can be formed from L_{ab} by contracting with g^{ab} or the completely antisymmetric tensor $\epsilon^{a_1 a_2 \dots a_n}$. In particular, there are the quantities[6]

$$I^k = L^{a_1}{}_{a_2} L^{a_2}{}_{a_3} \dots L^{a_k}{}_{a_1}, \tag{4.1a}$$

$$D^{2k} = \epsilon^{u_1 u_2 \dots u_{n-2k} a_1 b_1 \dots a_k b_k} L_{a_1 b_1} \dots L_{a_k b_k} \times \epsilon_{u_1 u_2 \dots u_{n-2k} c_1 d_1 \dots c_k d_k} L^{c_1 d_1} \dots L^{c_k d_k}, \tag{4.1b}$$

and, for even n,

$$I^{1/2} = \epsilon^{a_1 \dots a_n} L_{a_1 a_2} \dots L_{a_{n-1} a_n}. \tag{4.1c}$$

An independent set of generators for the center consists of

---

[6] The notation is taken in part from Kihlberg (1965), p. 126. Our D^k is proportional to Kihlberg's Δ^k. The sentence below Eq. (4) of this reference is incorrect as it stands. It should read: "The invariants [of $SO_0(p,q)$] are obtained from those [of $SO(p+q)$] by replacing F_{ij} by iL_{ij} if $i \le p$, $j > p$ or $i > p$, $j \le p$, by $-L_{ij}$ if $i > p$, $j > p$, and by L_{ij} if $i \le p$, $j \le p$."

either I^k or D^k for all even k up to k = n - 1 or n - 2, plus $I^{1/2}$ for even n. In representations these objects are represented by "Casimir operators", which commute with all the group or Lie algebra operators. They are multiples of the identity in any irreducible representation.

Let us write down these invariants for $SO_o(1,n)$, n = 2, 3, 4, in the notation of Eqs. (3.5) and study their behavior under contraction. In all cases we have

$$Q \equiv -\frac{1}{2} I^2 = H^2 + \vec{K}^2 - \vec{P}^2 - \sum_{A<B} J_{AB}^2 \tag{4.2}$$

(which is the Casimir operator in the strict sense). The transformation (3.3) is

$$\bar{\vec{K}} = \vec{K}, \qquad \bar{J}_{AB} = J_{AB}, \tag{4.3a}$$

$$\bar{\vec{P}} = \lim_{R\to\infty} \frac{1}{R} \vec{P}, \qquad \bar{H} = \lim_{R\to\infty} \frac{1}{R} H. \tag{4.3b}$$

Substitution into Eq. (4.2) yields

$$m^2 \equiv \lim_{R\to\infty} \frac{1}{R^2} Q = \bar{H}^2 - \bar{\vec{P}}^2. \tag{4.4}$$

This, of course, is the principal invariant of the Poincaré Lie algebra (3.6).

$SO_o(1,3)$ has another independent invariant, which we take to be

$$Q_2 \equiv -\frac{1}{8} I^{1/2} = JH + K^2 P^1 - K^1 P^2 . \qquad (4.5)$$

Under contraction we have simply

$$S \equiv \lim_{R\to\infty} \frac{1}{R} Q_2 = JH + K^2 P^1 - K^1 P^2 \qquad \text{(bars omitted)}$$

$$= \frac{1}{2} \epsilon^{abc} L_{ab} T_c \qquad (4.6)$$

as the corresponding invariant of the three-dimensional Poincaré group. In an irreducible representation of $ISO_o(1,2)$ (see Sec. B.2) with m = const. > 0, we have at the point of the spectrum where $P^1 = P^2 = 0$ and H = m

$$S = mJ. \qquad (4.7)$$

Thus S characterizes the spin (the representation of the little group). (The irreducible representations of SO(2) are labeled by the possible eigenvalues of J, which are 0, ± 1/2, ± 1, ...)

When n = 4 the simplest fourth-degree invariant is D^4. (I^4 is a linear combination of D^4, Q^2, and Q.) In our notation

$$Q_2 \equiv -\frac{1}{64} D^4 = [\vec{J}H + \vec{P} \times \vec{K}]^2 - (\vec{J}\cdot\vec{P})^2 + (\vec{J}\cdot\vec{K})^2 . \qquad (4.8)$$

In the limit

$$W \equiv \lim_{R\to\infty} \frac{1}{R^2} Q_2 = [\vec{J}H + \vec{P} \times \vec{K}]^2 - (\vec{J}\cdot\vec{P})^2 \qquad \text{(bars omitted)}$$

$$= -\frac{1}{4} \epsilon^{ab_1c_1d_1} L_{b_1c_1} T_{d_1} \epsilon_{ab_2c_2d_2} L^{b_2c_2} T^{d_2}. \qquad (4.9)$$

The analogue of Eq. (4.7) is the well-known relation

$$W = m^2 s(s + 1), \qquad (4.10)$$

where s is the spin of the representation of the little group SO(3). The individual terms which appear squared in Eq. (4.8) are the Casimir operators $I^{1/2}/8$ of various subgroups: $\vec{K}\cdot\vec{J}$ is (up to sign) the operator Q_2 of $SO_0(1,3)$, the group being interpreted now as the homogeneous Lorentz group; $\vec{P}\cdot\vec{J}$ is the analogous thing for the SO(4) subgroup (the spatial isometries).

Since the groups we are concerned with are not compact, their representations are not completely characterized by the values of the Casimir operators, as the results reviewed in Appendix B show.

5. The History of Quantum Theory in De Sitter Space.

Because of its unusual symmetry properties, de Sitter space[7] has been more often studied in relation to quantum

[7] Here we are primarily concerned with the closed de Sitter space of dimension 4, although Dirac and some of the other authors also considered the open space. The work of Philips and Wigner deals with the two-dimensional spaces, as do some of the papers of the Vienna and Dubna groups (see below).

theory than any other curved space-time.

Dirac (1935) proposed theories of the electromagnetic field and the Dirac spinor field in de Sitter space. He employed a very special construction, in which each physical quantity on the de Sitter hyperboloid was regarded as the restriction of a homogeneous function defined in the whole five-dimensional space in which the hyperboloid was imbedded. The degree of the homogeneity was decided for each quantity separately, on various grounds. This approach has been related to the representation theory of the de Sitter group and to more general definitions of wave equations in Riemannian space-times by Gürsey and Lee (Gürsey (1962), Gürsey and Lee (1963)), Hannabuss (1969a,b), and Castagnino (1970).

Philips and Wigner (Philips (1963), Philips and Wigner (1968)) and Hannabuss (1969a,b, 1970) approached particle quantum mechanics in de Sitter space in terms of the irreducible representations of the de Sitter group. Their emphasis was on analyzing the notion of <u>localization</u>, in analogy to the work of Newton and Wigner (1949) and Wightman (1962) in Minkowski space. In these papers, as in the work of Dirac and of Gürsey and Lee cited above, it was tacitly assumed that a covariant wave equation in de Sitter space is an equation governing the wave function of a single particle, whose possible quantum states transform under a representation of the group. (This viewpoint will be disputed in Chapter V.)

The earliest paper known to the present author on quantum field theory in de Sitter space is that of Gutzwiller (1956) (which apparently has gone unnoticed by all later workers in the field except Scarf (1959)). He treated scalar, electromagnetic, and spinor fields. His tools were primarily those of the classical theory of partial differential equations, rather than group theory. However, he demanded (in effect) a stable-particle interpretation of the field theory, since he adopted (at least for the spinor field) a time-independent decomposition of the field into positive- and negative-frequency parts.[8]

The Gruppenpest of the mid-sixties stimulated interest in the de Sitter group among elementary particle physicists (see references in Sec. 3 of the Introduction). In this period scalar field theory in de Sitter space was studied by research groups located at Vienna (Thirring (1967), Nachtmann (1967, 1968a,b)), Dubna (Tagirov et al. (1967), Chernikov and Tagirov (1968)), and Munich (Börner and Dürr (1969), Börner (1970)), and by Castagnino (1969). (Fronsdal (1965) and Castell (1969) considered the open

---

[8] Both Gutzwiller and Philips (1963) consider a definition of positive frequency based on the asymptotic behavior of the solutions of the wave equation -- viz., as $f(t)\exp(\pm imt)$, where m is the mass, t is an appropriate time coordinate, and f is real -- and reject this definition on the grounds that it gives different results when applied at $t \longrightarrow -\infty$ and at $t \longrightarrow +\infty$. Since from the standpoint of the present work such behavior has a natural interpretation in terms of particle creation, the author re-examined the matter. He concluded that this asymptotic behavior comes into play only for wavelengths which, by virtue of the expansion of the universe, have become very large on the laboratory scale; it consequently has very little physical significance, and should not be used to define "incoming" and "outgoing" particles.

de Sitter space.) The definition of particle annihilation and creation operators was treated most cogently by the Vienna and Dubna groups. Combining a maximum of group covariance with a close correspondence to the flat-space theory, they both arrived at the same theory of the scalar field in the two-dimensional closed de Sitter space, one in which particles are not created and destroyed. (This theory is discussed and criticized in a broader context in Secs. V.6, X.4, and X.9 below.)

At this point there appeared the landmark work of Parker (1966, 1968, 1969, 1971) on particle creation in expanding universes, which applies, in particular, to the de Sitter universe. So we must go back to pick up the story of a separate tradition, the theory of field quantization and particle creation in arbitrary space-times.

Schrödinger (1939) observed that in general the solutions of a field equation in a space with time-dependent metric cannot be separated into positive- and negative-frequency solutions; if only one type of "vibration" is present at one time, the other will appear at other times. He also remarked that this behavior could be interpreted in terms of creation and annihilation of particles (cf. Sec. X.3 below); but he considered this an "alarming phenomenon" and concentrated on proving that it did not occur under certain especially simple circumstances.

In the following years not much attention was paid to the physical interpretation of quantum field theory in curved space, although the formalism of fields of arbitrary spin as tensor- and spinor-valued distributions obeying covariant commutation relations was developed to considerable sophistication (Lichnerowicz (1961, 1962, and other papers)). Scarf (1959) reported particle creation in a generalization of the Thirring model[9] to curved space-time, but this work was based on a definition of the S-matrix for the Thirring model in flat space which later was found to be incorrect (A. S. Wightman, private communication). Imamura (1960) defined particle annihilation and creation operators and demonstrated particle creation in the unambiguous and soluble but highly unrealistic case of a universe whose radius is a step function in the time. Sexl and Urbantke (1967, 1969) discussed the cosmological implications of particle creation quite concretely; but their quantitative calculations were theoretically untenable in the most interesting cases (e.g., a closed universe collapsing to a singularity) because they treated only weak gravitational fields against a Minkowski background (by Feynman graph methods).

In the papers cited L. E. Parker developed a theory of field quantization in Robertson-Walker (homogeneous expanding) universes, which is described in part in Chapter X of this dissertation.[10] He shows that, in general, particle creation must occur, although some uncertainty in the identification of

---

[9] The Thirring model in flat space is a theory of a massless spinor field in dimension 2 satisfying an equation with a nonlinear interaction term, which nevertheless can be solved in terms of free fields. See Wightman (1964), pp. 218-231.

[10] A similar but much less thorough treatment was published by Grib and Mamaev (1969). See Sec. X.5 for critical remarks.

particle observables remains (see Secs. X.3 and X.6). This approach has recently been further developed, and applied to cosmologically interesting situations, by Zel'dovich and coworkers (Zel'dovich (1970), Zel'dovich and Pitaevsky (1971), Zel'dovich and Starobinsky (1971)) and by Parker (1972).

Chapter II

CONTRACTION OF THE REPRESENTATIONS OF THE DE SITTER GROUP TO REPRESENTATIONS OF THE POINCARE GROUP: FORMAL APPROACH

The goal of this chapter is to clarify the physical significance of the irreducible unitary (ray) representations of the de Sitter groups by correlating them with the irreducible unitary representations of the Poincaré groups. The representations themselves (which are well known) are described in Appendix B. We shall study the problem of contraction of representations in detail for n = 2 and indicate briefly how our observations extend to higher dimensions. We shall return to the subject by a different approach in Chapter VI.

1. Contraction to the Real-Mass Representations of the Two-Dimensional Poincaré Group.

The infinitesimal form of an irreducible unitary representation of the two-dimensional de Sitter group $SO_o(1,2)$ is given by Eqs. (B.3.8). We want to study the behavior of these formulas under the contraction transformation (I.4.3).[1] If one attempts to take this limit directly in Eqs. (B.3.8), the result

[1] The reader who wishes to follow this discussion down to the last detail should read Appendix C first, since in this section calculations and motivational arguments are sometimes summarized as "analogous to those in Appendix C."

is a (reducible) representation of the two-dimensional Poincaré Lie algebra in which $H = P = 0$, so that in the corresponding group representation the space and time translations are all represented by the identity. This class of representations is not of much physical interest.

The experience of Inönü and Wigner with the rotation group (Sec. C.3) suggests that we should consider a sequence (or a one-parameter family) of inequivalent representations, scaling the value of the Casimir operator along with the parameter R. On the basis of Eq. (I.4.4) we expect to obtain the representation with mass m and timelike momenta if we set

$$m = \frac{\sqrt{Q}}{R} = \text{const.} \tag{1.1}$$

and let the representation range up the principal series (Eq. (B.3.5a)) to $q = \infty$.

Our situation differs from the one studied by Inönü and Wigner (1953) in that the subgroup which is diagonalized in Eqs. (B.3.8) is not the subgroup (viz., exp(iKt)) with respect to which the contraction takes place. (This situation is studied for the rotation group in Sec. C.4.) The subgroup exp(iPt) is more convenient for the study of the representations of the de Sitter group both mathematically (because it is compact) and physically (because in our model it is the symmetry group of the spatial universe at one instant of time).

We express the representation formulas (B.3.8) in terms of $\bar{H}$, $\bar{P}$, m, and

$$\bar{p} = \frac{p}{R}. \tag{1.2}$$

We have

$$\bar{P}\Psi(\bar{p}) = \bar{p}\,\Psi(\bar{p}),$$

$$\bar{H}\Psi(\bar{p}) = \frac{1}{2}\{\sqrt{m^2 + \bar{p}(\bar{p} - \tfrac{1}{R})}\;\Psi(\bar{p} - \frac{1}{R}) + \sqrt{m^2 + \bar{p}(\bar{p} + \tfrac{1}{R})}\;\Psi(\bar{p} + \frac{1}{R})\},$$

$$\bar{K}\Psi(\bar{p}) = \frac{R}{2i}\{\sqrt{m^2 + \bar{p}(\bar{p} - \tfrac{1}{R})}\;\Psi(\bar{p} - \frac{1}{R}) - \sqrt{m^2 + \bar{p}(\bar{p} + \tfrac{1}{R})}\;\Psi(\bar{p} + \frac{1}{R})\}.$$

Then we pass to the limit, under the assumption that $\Psi(p)$ (which is defined only at discrete points, of course) is to be replaced by a reasonably smooth function of a continuous variable, $\Psi(\bar{p})$. The details of the calculation are precisely analogous to those given for SO(3) in Sec. C.4. The result is (bars omitted)

$$P\Psi(p) = p\,\Psi(p), \tag{1.3a}$$

$$H\Psi(p) = \sqrt{m^2 + p^2}\;\Psi(p), \tag{1.3b}$$

$$K\Psi(p) = i\sqrt{m^2 + p^2}\,\frac{d}{dp} + \frac{i}{2}\,\frac{p}{\sqrt{m^2 + p^2}}. \tag{1.3c}$$

Similarly, the scalar product (B.3.7) goes over into

$$(\Psi,\phi) = \int_{-\infty}^{\infty} dp\ \Psi^*(p)\phi(p). \qquad (1.4)$$

The representation is put into a more familiar form by writing

$$\bar{\Psi}(p) = \sqrt{2}\ \sqrt[4]{m^2 + p^2}\ \Psi(p). \qquad (1.5)$$

In terms of the barred functions we have

$$P\bar{\Psi}(p) = p\ \bar{\Psi}(p), \qquad (1.6a)$$

$$H\bar{\Psi}(p) = \sqrt{m^2 + p^2}\ \bar{\Psi}(p), \qquad (1.6b)$$

$$K\bar{\Psi}(p) = i\sqrt{m^2 + p^2}\ \frac{d\bar{\Psi}}{dp}\ ; \qquad (1.6c)$$

$$(H^2 - P^2)\bar{\Psi}(p) = m^2\ \bar{\Psi}(p)\ ; \qquad (1.7)$$

$$(\Psi,\phi) = \int_{-\infty}^{\infty} \frac{dp}{2\sqrt{m^2 + p^2}}\ \bar{\Psi}^*(p)\bar{\phi}(p). \qquad (1.8)$$

This is the irreducible unitary representation of $\mathcal{L}(ISO_0(1,1))$ with mass m and positive energy -- see Eqs. (B.2.5-6) and (B.2.2b).

However, as we know from the case studied in Sec. C.4, in these formal manipulations with Lie algebra matrix elements the choice of the phases of the basis vectors can influence the outcome. In general we could have written

$$\Psi = \sum_p \Psi(p)\,|q;p\rangle e^{i\theta(p)} \tag{1.9}$$

instead of Eq. (B.3.6). Then Eqs. (B.3.8b,c) would be replaced by

$$A^{\pm}\Psi(p) = \sqrt{q + p(p\mp 1)}\;\Psi(p\pm 1)\,e^{i(\theta(p\pm 1) - \theta(p))},$$

$$H = \frac{1}{2}(A^{+} + A^{-}), \qquad K = \frac{1}{2i}(A^{+} - A^{-}). \tag{1.10}$$

If we make the substitutions (I.4.3) and (1.1-2) into Eqs. (1.10) and try to take $R \longrightarrow \infty$, we find that the expression for $\bar{K}$,

$$\bar{K}\Psi_{\rho}(\bar{p}) = \frac{1}{2i}e^{-i\theta(R\bar{p})}\{R\sqrt{m^2 + \bar{p}^2}\;\Psi_{-\rho}(\bar{p})\,(e^{i\theta(R\bar{p}+1)} - e^{i\theta(R\bar{p}-1)})$$

$$- [\sqrt{m^2 + \bar{p}^2}\,\frac{d\Psi_{-\rho}}{d\bar{p}} + \frac{\bar{p}}{2\sqrt{m^2 + \bar{p}^2}}\,\Psi_{-\rho}](e^{i\theta(R\bar{p}+1)} + e^{i\theta(R\bar{p}-1)})$$

$$+ O(R^{-1})\}, \tag{1.11}$$

becomes infinite unless

$$\theta(R\bar{p} - 1) = \theta(R\bar{p} + 1) \qquad (\text{mod } 2\pi) \tag{1.12}$$

for all $\bar{p}$. In arriving at Eq. (1.11) it has been assumed that the $\Psi(p)$, $p = 0, \pm 2, \ldots$, are replaced in the limit by a differentiable function $\Psi_{+1}(\bar{p})$, and that the $\Psi(p)$,

$p = \pm 1, \pm 3, \ldots$, become a possibly different function $\Psi_{-1}(\bar{p})$. Because of Eq. (1.12) we necessarily have

$$e^{i\theta(R\bar{p}+1)} \, e^{-i\theta(R\bar{p})} = c(\rho), \qquad c(+1)\,c(-1) = 1, \tag{1.13}$$

where $\rho = +1$ for even $R\bar{p}$, -1 for odd $R\bar{p}$. Eq. (1.11) and its partner now become (bars omitted)

$$K\Psi_{\rho}(p) = ic(\rho)\left[\sqrt{m^2 + p^2}\,\frac{d\Psi_{-\rho}}{dp} + \frac{p}{2\sqrt{m^2 + p^2}}\,\Psi_{-\rho}(p)\right] \tag{1.14a}$$

and

$$H\Psi_{\rho}(p) = c(\rho)\sqrt{m^2 + p^2}\,\Psi_{-\rho}. \tag{1.14b}$$

Letting

$$\Psi(p,\sigma) = \sqrt[4]{m^2 + p^2}\,[\Psi_{\sigma}(p) + \sigma c(\sigma)\Psi_{-\sigma}(p)], \tag{1.15}$$

we finally obtain

$$K\Psi(p,\sigma) = i\sigma\sqrt{m^2 + p^2}\,\frac{d\Psi}{dp}(p,\sigma), \tag{1.16a}$$

$$H\Psi(p,\sigma) = \sigma\sqrt{m^2 + p^2}\,\Psi(p,\sigma) \tag{1.16b}$$

(along with the obvious generalizations of Eqs. (1.6a), (1.7), and (1.8)). This is a direct sum of two irreducible representations of the Poincaré group, one with positive energy and one with negative energy.

At this stage the concept of contraction of representations is so nebulous that it hardly makes sense to ask which of Eqs. (1.6) and (1.16) represents the "correct" contraction of the de Sitter representation (B.3.8). The problem deserves attention, however. We shall see in later chapters that the major problem of quantum field theory in de Sitter space (and in curved space-time in general) is to find a substitute for the spectral condition (the requirement of positive energy). It is surely relevant to enquire whether a representation of the Poincaré group containing only positive energies can somehow be got out of a representation of the de Sitter group, in which positive and negative energies seem to be inescapably mixed. This last property is related to the "local" nature of time translation in de Sitter space, which we have mentioned already. (This will be explained in Chapter III.) For this reason it is fortunate that we have a geometrical interpretation of the contraction of the representations of the rotation group which is based on shrinking functions defined on the sphere to a point (Secs. C.5-6). We shall apply this idea to the representations of $SO_0(1,2)$ after we study, in Chapter V, the functions on de Sitter space which support an irreducible representation of the group.

The most important conclusion from this discussion, however, is well founded: The real-mass representations of the Poincaré group are related to the principal series of representations of the de Sitter group in the limit of large q.

We should concentrate on the principal series, therefore, when searching for physically relevant theories of particles or fields in de Sitter space using group-theoretical methods.

2. Contraction to the Other Representations of $ISO_0(1,1)$, and a Remark on the Scale of Physical Quantities.

To obtain the representations with spacelike momentum ($H^2 - P^2 = -\mu^2$) we must allow the value of q to become large and negative, by setting, for instance,

$$\mu = \frac{1}{R} k \tag{2.1}$$

in the discrete series (Eq. (B.3.5b)). All the formulas of the previous section apply, with $m^2 = -\mu^2$, but the range of the variable is restricted to $|p| \geq \mu$.

In this case, however, Eqs. (1.6-8) provide only a fragment of a representation of the Poincaré group. To obtain a complete irreducible representation it is mandatory to take the route leading to Eqs. (1.16). This phenomenon is strictly analogous to what happens in the contraction of the representations of the rotation group as described in Sec. C.4, and the discussion there concerning the non-self-adjointness of one of the generators applies to the operator K of Eq. (1.6c). The function $\bar{\Psi}(p)$ is defined on the top branch of the hyperbola $(p^0)^2 = p^2 - \mu^2$. At the point $p^2 = \mu^2$, K is attempting to push $\bar{\Psi}$ over the edge onto the bottom branch, which is missing. Thus K does not generate a unitary operator, and there is no unitary group representation corresponding to Eqs. (1.6).

The representation of $ISO_0(1,1)$ with $m = 0$ can be obtained from either the continuous or the discrete series of $SO_0(1,2)$ by making $|q|$ approach infinity more slowly than R, or not at all. In particular, q could be held constant at any value. Then one is not varying the representation as $R \longrightarrow \infty$, but merely changing the scale of the momentum variable p. This is one of the rare situations where a unitary representation of the contracted group in which the Abelian invariant subgroup is represented nontrivially can be obtained from a fixed representation of the original group by a singular transformation. (Another example was given by Inönü and Wigner (1953).)

This observation brings up an interesting point. In the Introduction and Sec. I.2 we spoke of a family of "de Sitter theories" which should "converge" for large R to an ordinary relativistic flat-space theory. In the case of a system of massless particles, we see now that all the theories in the family could be the same theory, looked at in different ways; namely, one varies the unit of length and confines his attention to a region of space-time of "moderate" dimensions in each length scale. For particles with mass, on the other hand, the theories are expected to be different for each R, since the ratio of the size of the universe to the Compton wavelengths of the particles will change.

For conceptual clarity one should distinguish three natural units of length:

(1) R, the radius of the universe;

(2) L, some length characteristic of the observer or his instruments (say L = 1 cm);

(3) $\lambda \equiv 1/m$, the Compton wavelength of one of the massive particles in the theory. (For simplicity we do not consider the possibility of varying the ratios of the elementary particle masses.)

The general condition for equivalence in practice of a de Sitter and a Poincaré theory is

$$R \gg L. \tag{2.2}$$

However, in a theory of the electron (for instance) one will have

$$\lambda < L, \tag{2.3}$$

and, moreover, the ratio λ/L will probably be held fixed at the observed value throughout the discussion. Then Eq. (2.2) can be replaced by

$$R \gg \lambda, \tag{2.4}$$

and L need never be mentioned explicitly. The situation is different if only massless particles are considered. Of course, Eq. (2.4) is in perfect accord with the ansatz of Sec. II.1 (Eq.

(1.1) with $q \longrightarrow \infty$).

3. Contraction in a "Continuous Basis".

The general element of $\mathcal{L}(SO_0(1,2))$ is of the form

$$L = hH + kK - pP. \qquad (3.1)$$

The one-parameter subgroups of $SO_0(1,2)$ are classified as elliptic, hyperbolic, and parabolic (or nilpotent) when their generators have $h^2 + k^2 - p^2$ negative, positive, and zero, respectively. Subgroups of the same class are geometrically equivalent (conjugate). Hyperbolic and parabolic subgroups are noncompact ($\cong \mathbb{R}$); elliptic subgroups are compact ($\cong SO(2)$). In a representation of the continuous series a generator L of the hyperbolic class has a continuous spectrum of multiplicity 2 extending from $-\infty$ to $+\infty$; the spectrum of a parabolic L has the same range but no doubling. In a representation of the discrete series the spectrum is nondegenerate in both cases and is $-\infty < \lambda < \infty$ in the hyperbolic case and $0 < \lambda < \infty$ or $-\infty < \lambda < 0$ in the parabolic case. (The spectrum of an elliptic element, which is discrete, is described in Appendix B (Eqs. (B.3.5)).)[2]

The representations have been expressed by Mukunda (1967) in a form in which the generator of a hyperbolic subgroup is diagonalized. If

[2] Bargmann (1947), pp. 588-589, 639-640.

$$[P, J_1] = iJ_2, \quad [P, J_2] = -iJ_1, \quad [J_1, J_2] = -iP, \tag{3.2}$$

then for a representation of the principal series (Eq. (B.3.5a), $q \geq 1/4$) with

$$q = s^2 + \frac{1}{4} \tag{3.3}$$

he gives the formulas ($\sigma = \pm 1$, $-\infty < \lambda < \infty$)

$$J_2 \Psi(\lambda, \sigma) = \lambda \Psi(\lambda, \sigma) \ ; \tag{3.4a}$$

$$J_{\pm} \Psi(\lambda, \sigma) = \sigma(\lambda \mp (s + \frac{i}{2})) \Psi(\lambda \mp i, \sigma), \tag{3.4b}$$

$$P = \frac{1}{2}(J_+ + J_-), \qquad J_1 = \frac{1}{2}(J_+ - J_-). \tag{3.5}$$

(The notation has been changed to resemble that of Sec. C.2.) These equations make sense on a domain of analytic functions of λ. Alternatively, one can work in the Fourier transform space:

$$\phi(z, \sigma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\lambda \, e^{iz\lambda} \Psi(\lambda, \sigma) \ ; \tag{3.6}$$

$$J_2 \phi = -i\frac{d\phi}{dz}, \tag{3.7a}$$

$$J_{\pm} \phi = \sigma e^{\mp z} [-i \frac{d}{dz} \pm i(\frac{1}{2} + is)]\phi. \tag{3.7b}$$

If we identify the diagonalized hyperbolic generator J_2 with K and J_1 with H (cf. Eqs. (B.3.1)) and apply the contraction transformation (I.4.3) with

$$m = \frac{s}{R} \tag{3.8}$$

(equivalent in the limit to Eq. (1.1)), we obtain from Eqs. (3.4) or (3.7)

$$K\Psi = \lambda \Psi, \qquad K\phi = -i\frac{d\phi}{dz}; \tag{3.9a}$$

$$J_{\pm}\Psi(\lambda,\sigma) = \mp \sigma m \Psi(\lambda \mp i, \sigma), \qquad J_{\pm}\phi = \mp \sigma m e^{\mp z} \phi. \tag{3.9b}$$

For each of the two values of σ these operators (which are also discussed in Mukunda's paper) are the generators of a representation of $ISO_0(1,1)$ in a "boost basis" (cf. Eqs. (C.2.8,10,12) for the Euclidean group in a rotation basis). This is a close analogue of the Inönü-Wigner contraction of the rotation group (Sec. C.3). Note that from this point of view an irreducible representation of the de Sitter group contracts unambiguously to the direct sum of two representations of the Poincaré group.

On the other hand, one could take J_2 to be H. One would expect that operators obeying the commutation relations of the Poincaré group could be obtained by a contraction analogous to Eqs. (1.1-3) (in which λ as well as s is renormalized). However, it is easy to check that it is impossible to obtain formally convergent expressions, even if the phase of $\Psi(\lambda,\sigma)$ is changed similarly to Eq. (1.9) (with θ an analytic function of λ). Any formal manipulation which led to the desired result would have to be very artificial, since it would have to create a gap in the spectrum in the interval

$$- m < \frac{\lambda}{R} < m. \tag{3.10}$$

It seems unlikely that anything useful can be said about the matter at this level, so we shall drop it until Sec. VI.1.

4. Contraction of Representations in Dimension 3.

To obtain representations of $ISO_0(1,2)$ from representations of $SO_0(1,3)$ we must scale the Casimir invariants so that the quantities m^2 and S of Eqs. (I.4.4,6) have finite limits. From Eq. (B.4.8a) we see that for $m \neq 0$ either k_0 or d must approach infinity proportionally to R. If both do, S becomes infinite, so that case should be excluded.

If we take

$$d = \pm Rm \quad (m > 0), \qquad k_0 = s \operatorname{sgn} d, \tag{4.1a}$$

we have (see Eqs. (B.4.8))

$$\frac{1}{R^2} Q \longrightarrow m^2, \qquad \frac{1}{R} Q_2 \longrightarrow S = ms. \tag{4.1b}$$

So we expect to obtain via Eq. (4.1a) a representation of $ISO_o(1,2)$ with mass m and spin s (see Eq. (I.4.7) and surrounding discussion).

On the other hand, if we let

$$k_0 = R\mu, \qquad d \text{ fixed (real)}, \tag{4.2a}$$

then

$$m^2 = -\mu^2, \qquad S = \mu d. \tag{4.2b}$$

This corresponds to a representation of $ISO_o(1,2)$ with spacelike momentum spectrum. If we set $P^2 = H = 0$ and $P^1 = \sqrt{-m^2} = \mu$ in Eq. (I.4.6), we have, analogously to Eq. (I.4.7),

$$S = \mu K^2 .$$

Therefore, in the limit d is the eigenvalue of K^2, the generator of the little group $SO_o(1,1)$.

Finally, the representations with m = 0 can be reached by (for instance)

$$k_0 = \sqrt{R}\,|a|, \qquad d = \sqrt{R}\,a, \tag{4.3a}$$

$$m^2 = 0, \qquad S = \pm a^2 . \tag{4.3b}$$

The representations with $a \neq 0$ are analogues of the continuous-spin representations of the four-dimensional Poincaré group.

It remains to be shown that the representations of the Lie algebra of $ISO_0(1,2)$ can actually be obtained by formally taking the limit $R \longrightarrow \infty$ in the matrix elements of the representations of $\mathcal{L}(SO_0(1,3))$ (Eqs. (B.4.3-6)). We shall devote the rest of the section to doing this for the case of real mass.

For simplicity assume that d is positive in Eq. (4.1a); then we write $d = Rm$, $k_0 = s$, $k = Rp$. We already know (Secs. C.1-3) that Eqs. (B.4.3) contract (for each p) to the representation (C.2.12) of the two-dimensional Euclidean group (with $M = p$, $m = n$). To investigate the other formulas we expand all the coefficients to the two lowest orders in $1/R$:

$$C(k) = \frac{i}{2p}\sqrt{p^2 + m^2}\ (1 + O(R^{-2})),$$

$$C(k+1) = \frac{i}{2p}\sqrt{p^2 + m^2}\ \left(1 + \frac{1}{R}\left[\frac{p}{p^2 + m^2} - \frac{1}{p}\right] + O(R^{-2})\right),$$

$$\sqrt{(k\mp n+1)(k\mp n+2)} = Rp\ \left(1 + \frac{1}{Rp}\left(\frac{3}{2} \mp n\right) + O(R^{-2})\right),$$

etc.

As in the other cases studied, we take the limit $R \longrightarrow \infty$ in accordance with Eqs. (C.4.3-4). The result is

$$\bar{H}\Psi(p,n) = \sqrt{p^2+m^2}\,\Psi(p,n), \tag{4.4a}$$

$$K_{\pm}\Psi(p,n) = i\sqrt{p^2+m^2}\,\frac{d\Psi}{dp}(p,n\mp 1) + \frac{i}{2}\frac{p}{\sqrt{p^2+m^2}}\Psi(p,n\mp 1)$$

$$+\, i\,\frac{\sqrt{p^2+m^2}}{p}\left(\frac{1}{2}\mp n\right)\Psi(p,n\mp 1) \pm \frac{ims}{p}\Psi(p,n\mp 1). \tag{4.4b}$$

(For ease of writing we follow the approach that leads to the formulas for the positive-energy representation alone. The modification that yields both signs of the energy in analogy to Eqs. (1.9-16) is obvious.)

A few more steps are needed to bring the representation to the form with which we are familiar (Sec. B.2). First, by forming a Fourier series (i.e., reversing the steps of Eqs. (C.2.7-13)), we attain the form

$$J\Psi(p,\phi) = -i\frac{\partial\Psi}{\partial\phi}, \tag{4.5}$$

$$P_{\pm}\Psi(p,\phi) = pe^{\pm i\phi}\,\Psi(p,\phi), \tag{4.6}$$

$$H\Psi(p,\phi) = \sqrt{p^2+m^2}\,\Psi(p,\phi), \tag{4.7}$$

$$K_{\pm}\Psi(p,\phi) = i\sqrt{p^2+m^2}\,\frac{d}{dp}[e^{\pm i\phi}\Psi(p,\phi)]$$

$$+\frac{i}{2}\frac{p}{\sqrt{p^2+m^2}}\,e^{\pm i\phi}\Psi(p,\phi)$$

$$+i\frac{\sqrt{p^2+m^2}}{p}\left(\frac{1}{2}\pm i\frac{\partial}{\partial\phi}\right)[e^{\pm i\phi}\Psi(p,\phi)]$$

$$\pm\frac{ims}{p}\,e^{\pm i\phi}\Psi(p,\phi). \qquad (4.8)$$

Meanwhile, the integration in the scalar product has undergone the contraction

$$\sum_{k=k_0}^{\infty}\sum_{m=-k}^{k} \longrightarrow \int_0^{\infty}dp\sum_{m=-\infty}^{\infty} = \frac{1}{2\pi}\int_0^{\infty}dp\int_0^{2\pi}d\phi. \qquad (4.9)$$

In the final version we would expect a multiple of

$$\int\frac{d\vec{p}}{\sqrt{p^2+m^2}} = \int_0^{\infty}p\,dp\int_0^{2\pi}d\phi\,[p^2+m^2]^{-1/2} \qquad (4.10)$$

(cf. Eq. (B.2.2b)). This, along with the expectation of a spin term in the angular momentum, suggests the transformation

$$\bar{\Psi}(p,\phi) = \frac{1}{\sqrt{\pi}}[m^2+p^2]^{1/4}\,p^{-1/2}\,e^{-is\phi}\,\Psi(p,\phi) \qquad (4.11)$$

as the analogue of Eq. (1.5). This leaves Eqs. (4.6-7) unchanged and converts Eqs. (4.5) and (4.8) to (bars omitted)

$$J\Psi = - i\frac{\partial\Psi}{\partial\phi} + s, \tag{4.12}$$

$$K_{\pm}\Psi = i\sqrt{p^2 + m^2}\, e^{\pm i\phi}\frac{\partial\Psi}{\partial p} \mp \frac{\sqrt{p^2 + m^2}}{p}\, e^{\pm i\phi}\frac{\partial\Psi}{\partial\phi}$$

$$\pm \frac{is}{p}(m - \sqrt{p^2 + m^2})\, e^{\pm i\phi}\,\Psi. \tag{4.13}$$

Eq. (4.13) is equivalent to

$$K^1\Psi = i\sqrt{\vec{p}^2 + m^2}\,\frac{\partial\Psi}{\partial p^1} + s\,(\sqrt{\vec{p}^2 + m^2} + m)^{-1}\, p^2\,\Psi,$$

$$K^2\Psi = i\sqrt{\vec{p}^2 + m^2}\,\frac{\partial\Psi}{\partial p^2} - s\,(\sqrt{\vec{p}^2 + m^2} + m)^{-1}\, p^1\,\Psi. \tag{4.14}$$

A calculation which we omit shows that Eqs. (4.6,7,12,14) are the infinitesimal generators of the $ISO_o(1,2)$ representation (B.2.3) with

$$Q(p_0, e^{-i\theta J})\Psi = e^{-i\theta s}\Psi \tag{4.15}$$

(see Sec. C.2 for sign conventions), provided that we choose p_0 to have components $(1,0,0) \equiv (1,\vec{0})$ and choose $C(q)$ as the pure boost which maps p into q:

$$C(q) = \frac{1}{m}\begin{pmatrix} q^0 & \vec{q} \\ \vec{q} & m + (q^0+m)^{-1}\,\vec{q}\otimes\vec{q} \end{pmatrix} \qquad ((q^0)^2 - \vec{q}^{\,2} = m^2). \quad (4.16)$$

5. Contraction of Representations in Dimension 4, and Some Remarks on the General Case.

The analogous calculations for the de Sitter group, properly so called, ($SO_0(1,4)$) have been carried out by Ström (1965).[3] His results correspond to the cruder of the two approaches we have taken above to $SO_0(1,2)$. That is, he did not treat even and odd values of the index l (see Sec. B.5) separately when replacing 1/R by a continuous variable p, and consequently arrived at expressions involving only one sign of the energy; then the other sign had to be put in "by brute force". It is clear from the previous sections how to improve this procedure.

Ström's results for the correspondence of representations of $SO_0(1,4)$ and $ISO_0(1,3)$ are similar to those we found for dimension 3. The real-mass representations are obtained from the principal series (Eqs. (B.5.3)) by taking

$$r \text{ fixed}, \quad \sigma \longrightarrow \infty, \quad \frac{1}{R^2}\sigma \longrightarrow m^2. \qquad (5.1a)$$

Then from Eqs. (B.5.1) we have

[3] See also Holman (1969) and Böhm (1970).

$$\frac{1}{R^2} Q \longrightarrow m^2, \quad \frac{1}{R^2} Q_2 \longrightarrow m^2 \, r(r + 1),$$

so that (see Eqs. (I.4.4,9,10)) the contracted representation of $ISO_0(1,3)$ should have

$$\text{mass } m \text{ and spin } r. \tag{5.1b}$$

If we set

$$\sigma \text{ fixed}, \quad r \longrightarrow \infty, \quad \frac{1}{R} r \longrightarrow \mu \tag{5.2a}$$

in the principal series, we obtain

$$m^2 = -\mu^2, \quad W = \mu^2 \sigma \tag{5.2b}$$

in the notation of Sec. I.4. These values of the Casimir operators correspond to representations with spacelike momentum (imaginary mass $i\mu$) and a representation of the little group $SO_0(1,2)$ which belongs to the continuous series, with $q = \sigma$ (see Sec. B.3). On the other hand, consider the discrete series of $SO_0(1,4)$, subclass (a) (Eqs. (B.5.6)), and take

$$q \text{ fixed}, \quad r \longrightarrow \infty, \quad \frac{1}{R} r \longrightarrow \mu. \tag{5.3a}$$

Then

$$m^2 = -\mu^2, \quad W = -\mu^2 q(q-1), \tag{5.3b}$$

which corresponds to an imaginary-mass representation with the little-group representation taken from the discrete series. The trivial representation of $SO_o(1,2)$ corresponds to the discrete series, subclass (b) (Eq. (B.5.8)). (Thus there is a correspondence between the representations of $SO_o(1,2)$ and the classes of representations of $SO_o(1,4)$ which differ only in their r values.) Ström also gets the zero-mass continuous-spin representations of $ISO_o(1,3)$ by taking

$$\frac{1}{R}\sigma \longrightarrow a, \quad \frac{1}{R}r^2 \longrightarrow a \tag{5.4a}$$

in the principal series, so that

$$m^2 = 0, \quad W = a^2. \tag{5.4b}$$

The zero-mass representations with discrete helicities are degenerate cases of Eqs. (5.2) or (5.4). Thus all the representations of the Poincaré group are accounted for.

When any of these limits is carried out in the formulas for the representation of the Lie algebra, Ström obtains a representation of $\mathcal{L}(ISO_o(1,3))$ reduced with respect to the subgroup ISO(3) (cf. Eqs. (4.4)). The connections conjectured on the basis of the behavior of the Casimir operators are validated. The internal label l is replaced by the continuous variable

$$p = \lim_{R\to\infty} \frac{1}{R} - 1, \tag{5.5}$$

and the Casimir operators of the Euclidean group take the form

$$\vec{P}^2\,\Psi(p,n;j;m) = p^2\,\Psi(p,n;j;m),$$

$$\vec{J}\cdot\vec{P}\,\Psi(p,n;j;m) = np\,\Psi(p,n;j;m). \tag{5.6}$$

In summary, this chapter has established that the contraction process induces the following correspondences between the various series of representations of the de Sitter and Poincaré groups:

$SO_0(1,n)$		$ISO_0(1,n-1)$
	$n = 2$:	
Continuous	⟶	Timelike
Discrete	⟶	Spacelike
	$n = 3$:	
Continuous	⟶	Timelike
	↘	Spacelike
	$n = 4$:	
Continuous	⟶	Timelike
	↘	
Discrete	⟶	Spacelike

The irreducible representations of $\mathcal{X}(SO_o(1,n))$ for general n have been classified by Hirai (1962a,b), Ottoson (1968), and Schwarz (1971). It is clear that the contraction relationships with the representations of $ISO_o(1,n-1)$ could in principle be determined by studying the Casimir operators as has been done for $n \leq 4$.

We have seen explicitly for $n \leq 4$ that the representations of $SO_o(1,n)$ which contract to the real-mass representations of $ISO_o(1,n-1)$ are those of the principal series, the class of representations which can be <u>induced</u> from unitary representations of the subgroup $(SO(n-1) \otimes \{\exp(itH)\})\cdot N$, where the generators of N are the components of $\vec{K} + \vec{P}$.[4] (N and $\{\exp(itH)\}$ are the nilpotent and Abelian parts of the Iwasawa decomposition of $SO_o(1,n)$; H commutes with $SO(n-1)$, which is the "rotation" part of the maximal compact subgroup $SO(n)$ of $SO_o(1,n)$.) These representations are parametrized by the representations of $SO(n-1) \otimes \{\exp(itH)\}$. Our observation about the classification of the representations of $SO_o(1,4)$ in terms of representations of $SO_o(1,2)$ plus a discrete parameter suggests that a similar relationship must exist between the representations of $SO_o(1,n-2) \otimes \{\exp(itP_1)\}$, say, and the representations of $SO_o(1,n)$ which contract to imaginary-mass representations of $ISO_o(1,n-1)$. This can be verified by enumeration in the cases $n \leq 4$ which we have studied. There is

[4] Takahashi (1963), especially pp. 382-384; Stein (1965); Ström (1971), Chapter VI.

probably a deeper connection in terms of the structure of the representations; it might be useful not only in the study of contraction but in classifying the representations of $SO_o(1,n)$ — themselves. Unfortunately, no exact parallel of the inducing construction exists in the case of the subgroup $SO_o(1,n-2) \otimes SO(2)$, since it cannot be made into a larger group — by adding a nilpotent subgroup (G. Zuckerman, private communication). Thus whatever connection exists must be more subtle.

Chapter III

THE GEOMETRY OF DE SITTER SPACE

Minkowski space and the de Sitter spaces are distinguished from all other space-times by their high degree of symmetry. One way to describe the difference between de Sitter space and Minkowski space is that in the former case the various symmetries do not fit together as well as in flat space. In Chapter I we have noted the noncommutativity of the geometrical isometries which, as far as their effects on the neighborhood of a given point are concerned, are identified as time and space translations. In this chapter several coordinate systems will be introduced, each of which could be considered a natural generalization of Cartesian coordinates in flat space, each of which is especially appropriate for the exploitation of certain of the special geometrical properties of de Sitter space. Each of these ways of looking at de Sitter space suggests an answer to the question: How is physics in the de Sitter universe to be formulated as a dynamical problem?

In later chapters we will find a variety of generalizations of the canonical quantization procedure for the free scalar field in flat space. These are mathematically different -- they are not just transcriptions of one generally covariant theory into terms of various coordinate systems. In the final chapters it is argued that this is a fundamental problem which must be faced by any theory of field quantization in curved space-time, and that its resolution must be sought in a closer analysis of the physical interpretation of field theory in such a context.

The task of this chapter is to introduce these various "pictures" of de Sitter space, to discuss their physical meaning on a geometrical or "classical" level, and to record a few useful facts and formulas related to the associated coordinate systems. We work mostly with the two-dimensional model, but we record enough formulas for the four-dimensional case to establish that the generalization to $n > 2$ is straightforward.

The terminology used for various kinds of canonical coordinate systems and for types of space-time metrics with special properties is explained in Appendix D.

For further discussion of the de Sitter universe from a cosmological point of view see [Schrödinger], pp. 1-40, Rindler (1960), [Tolman], pp. 333-337, 346-360, and [Robertson-Noonan], pp. 365-371. The first of these references is especially sensitive to the sort of question that concerns us in this chapter. A more abstract study of the geometry of the spaces of constant curvature is Calabi and Markus (1962).

1. De Sitter Space as a Closed Robertson-Walker Universe (Geodesic Gaussian Coordinates).

In the two-dimensional de Sitter space,

$$(x^0)^2 - (x^1)^2 - (x^2)^2 = -1, \tag{1.1}$$

we introduce two independent coordinates defined by

$$\begin{aligned} x^0 &= \sinh \tau, \\ x^1 &= \cosh \tau \sin \sigma, \\ x^2 &= \cosh \tau \cos \sigma. \end{aligned} \tag{1.2}$$

As they vary in the range

$$-\infty < \tau = \sinh^{-1} x^0 < \infty, \qquad -\pi < \sigma = \tan^{-1}(x^1/x^2) \leq \pi$$

$$(0 < \sigma < \pi \text{ if } x^1 > 0; \quad -\pi < \sigma < 0 \text{ if } x^1 < 0) \tag{1.3}$$

the whole space (1.1) is covered (see Fig. 3).

The metric of de Sitter space as a Riemannian manifold[1] is induced by the indefinite metric of the

[1] We shall use the word "Riemannian" in the broader sense: it does not imply that the metric is positive definite. Manifolds with the metric signature of space-time are sometimes called "Lorentzian", but this might lead to confusion in field theory, where Lorentz invariance has traditionally been of such great importance.

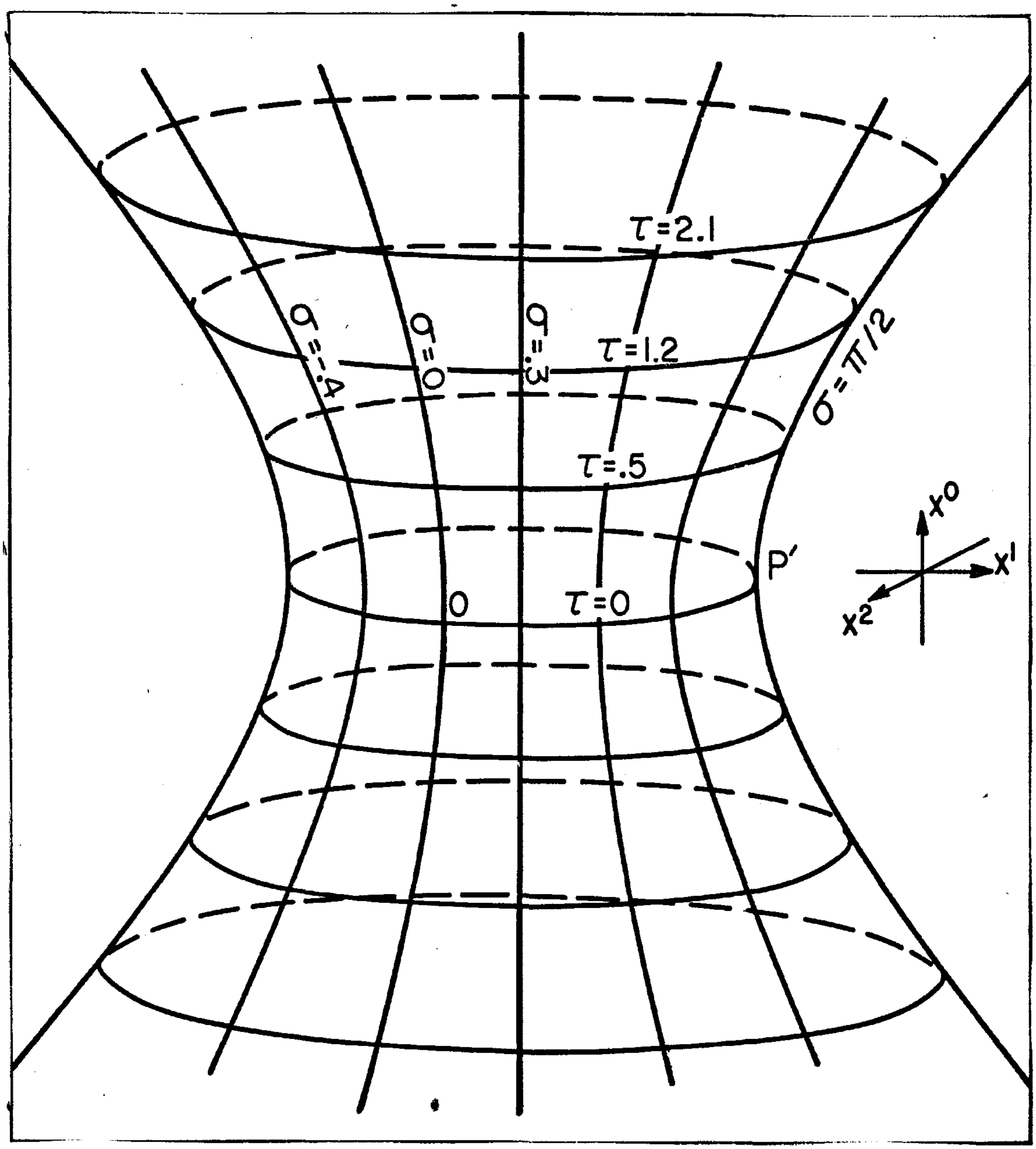


Fig. 3

Geodesic Gaussian coordinates in two-dimensional de Sitter space.

three-dimensional space in which it is imbedded. The line element is easily calculated to be

$$ds^2 \equiv (dx^0)^2 - (dx^1)^2 - (dx^2)^2 = d\tau^2 - \cosh^2 \tau \, d\sigma^2 . \quad (1.4)$$

We observe the following about this metric:

(1) It is orthogonal ($g_{jk} = 0$ for $j \neq k$).

(2) It is a Gaussian metric (Eq. (D.2)). (This is a statement about the coordinate system.) The curve defined by $\tau = 0$ is a geodesic, and σ is an arc-length parameter along it (see Sec. III.4 for details). Thus (τ, σ) is the geodesic Gaussian coordinate system based on the geodesic hypersurface $\{x \mid x^0 = 0\}$.

(3) It is a Robertson-Walker metric (Eq. (D.6)). (This is a statement about the space.) The universe contracts from radius ∞ to radius 1 and then expands again.

(4) The "space translation" $\sigma \longrightarrow \sigma + \sigma_0$ is the element

$$\mathcal{J}(\sigma_0) \equiv e^{-i\sigma_0 P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \sigma_0 & + \sin \sigma_0 \\ 0 & - \sin \sigma_0 & \cos \sigma_0 \end{pmatrix} \quad (1.5)$$

of the de Sitter group (see Eqs. (A.1.11,12a)[2]). This can be seen by direct application to Eqs. (1.2).

The generalization of the geodesic Gaussian coordinate system to higher dimensions is

$$x^0 = \sinh \tau, \tag{1.6a}$$

$$x^j = \cosh \tau \, f^j(\sigma^1, \ldots, \sigma^s) \qquad (j = 1, \ldots, n), \tag{1.6b}$$

where the f^j are such that $(\sigma^1, \ldots, \sigma^s)$ $(s = n - 1)$ form a coordinate system on the s-sphere defined by

$$F(x) = -1, \qquad x^0 = \text{const.} \tag{1.7}$$

(notation of Eq. (I.1.1)). Since necessarily

$$\sum_j (f^j)^2 = 1,$$

$$\sum_{j,k} f^j \frac{\partial f^j}{\partial \sigma^k} d\sigma_k = 0,$$

the metric is

$$ds^2 = d\tau^2 - \cosh^2 \tau \, d\Omega^2, \tag{1.8a}$$

---

[2] When n = 2, P denotes the contravariant momentum iL_{21} (see Eqs. (I.3.5)).

where

$$d\Omega^2 = \sum_{j,k,l=1}^{s} \frac{\partial f^j}{\partial \sigma^k} \frac{\partial f^j}{\partial \sigma^l} d\sigma^k d\sigma^l \qquad (1.8b)$$

is the line element on the s-sphere.

The generalization of property (4) is that each spacelike hypersurface of the form (1.7) is invariant under the SO(n) group generated by $\vec{P}$ and the J_{AB}. The generalization of property (2) is that the geodesics in the sphere $\{x|\tau = 0\}$ are geodesics of the whole space; thus the sphere is a geodesic hypersurface relative to any point in it. Property (3) remains valid, and the coordinates on the sphere can be chosen so as to make the system orthogonal.

For example, when n = 4 we can take Eqs. (1.6b) to be

$$x^1 = \cosh \tau \sin \sigma^1 \qquad (-\frac{\pi}{2} \leq \sigma^1 \leq \frac{\pi}{2}),$$

$$x^2 = \cosh \tau \cos \sigma^1 \sin \sigma^2 \qquad (-\frac{\pi}{2} \leq \sigma^2 \leq \frac{\pi}{2}),$$

$$x^3 = \cosh \tau \cos \sigma^1 \cos \sigma^2 \sin \sigma^3 \qquad (-\pi < \sigma^3 \leq \pi),$$

$$x^4 = \cosh \tau \cos \sigma^1 \cos \sigma^2 \cos \sigma^3 . \qquad (1.9)$$

Then

$$d\Omega^2 = (d\sigma^1)^2 + \cos^2\sigma^1\,(d\sigma^2)^2 + \cos^2\sigma^1\,\cos^2\sigma^2\,(d\sigma^3)^2 . \quad (1.10)$$

In the neighborhood of the distinguished point 0 of Eq. (I.1.2) we have to first order

$$x^0 \sim \tau, \qquad x^1 \sim \sigma^1,$$

$$x^2 \sim \sigma^2, \qquad x^3 \sim \sigma^3. \quad (1.11)$$

So the coordinate system (1.6a,9) "contracts" to Cartesian coordinates in flat space. The transformations $\sigma^3 \longrightarrow \sigma^3 + \sigma_0$ are the space translations generated by P^3.

Alternatively, we could set

$$x^1 = \cosh\tau \sin\sigma_r \sin\sigma_\theta \cos\sigma_\phi , \quad (1.12)$$

$$x^2 = \cosh\tau \sin\sigma_r \sin\sigma_\theta \sin\sigma_\phi \qquad (-\pi < \sigma_\phi \leq \pi),$$

$$x^3 = \cosh\tau \sin\sigma_r \cos\sigma_\theta \qquad (0 \leq \sigma_\theta \leq \pi),$$

$$x^4 = \cosh\tau \cos\sigma_r \qquad (0 \leq \sigma_r \leq \pi),$$

with

$$d\Omega^2 = d\sigma_r^2 + \sin^2\sigma_r\,[d\sigma_\theta^2 + \sin^2\sigma_\theta\, d\sigma_\phi^2]. \tag{1.13}$$

In this case the transformations $\sigma_\phi \longrightarrow \sigma_\phi + \sigma_o$ are the rotations about O generated by J^3. In the neighborhood of O we have

$$x^0 \sim \tau, \qquad x^1 \sim \sigma_r \sin\sigma_\theta \cos\sigma_\phi,$$

$$x^2 \sim \sigma_r \sin\sigma_\theta \sin\sigma_\phi, \qquad x^3 \sim \sigma_r \cos\sigma_\theta, \tag{1.14}$$

which corresponds to a spherical polar coordinate system in Minkowski space.

Of course, all this generalizes the discussion in Secs. C.5-6 of the relation between spherical coordinates on the two-sphere and Cartesian and polar coordiates in the Euclidean plane. Just as there, the coordinate systems can be related to the group parameters through the construction of the homogeneous space concerned as a space of cosets. For instance, the point in Eqs. (1.9) is the coset

$$\exp(i\sigma^3 P_3)\ \exp(i\sigma^2 P_2)\ \exp(i\sigma^1 P_1)\ \exp(i\tau H)\ SO_0(1,3), \tag{1.15a}$$

while that of Eqs. (1.12) is

$$\exp(-i\sigma_\phi J^3)\ \exp(-i\sigma_\theta J^2)\ \exp(-i\sigma_r P^3)\ \exp(i\tau H)\ SO_0(1,3). \tag{1.15b}$$

In the de Sitter space of radius R (Eq. (I.1.1) in place of Eq. (1.1)) we would write instead of Eqs. (1.2)

$$x^0 = R \sinh \frac{\tau}{R},$$

$$x^1 = R \cosh \frac{\tau}{R} \sin \frac{\sigma}{R}, \qquad (1.16)$$

$$x^2 = R \cosh \frac{\tau}{R} \cos \frac{\sigma}{R}.$$

Other definitions in this chapter would be modified similarly. (In Eqs. (1.12), for instance, τ and σ_r should be scaled by 1/R but the angular variables σ_θ and σ_ϕ should not be.) The necessary changes in all the formulas of this chapter to accommodate this generalization are rather obvious. With this definition τ and σ are still the properly normalized arc length parameters on the basic geodesics defining the coordinate system, and under contraction ($R \longrightarrow \infty$) they become Cartesian coordinates in Minkowski space (cf. Eqs. (1.11,14)).

2. De Sitter Space as a Static Universe (Geodesic Fermi Coordinates).

In the neighborhood of the point O in the two-dimensional space (1.1) we can introduce another set of coordinates by

$$x^0 = \sinh \chi \cos \rho,$$

$$x^1 = \sin \rho, \qquad (2.1)$$

$$x^2 = \cosh \chi \cos \rho.$$

Then

$$\chi = \tanh^{-1} (x^0/x^2) \qquad (-\infty < \chi < \infty) \qquad (2.2a)$$

and

$$\rho = \sin^{-1} x^1$$

$$(0 < \rho < \pi \text{ if } x^1 > 0; \quad -\pi < \rho < 0 \text{ if } x^1 < 0) \qquad (2.2b)$$

are defined only if

$$|x^0| < |x^2| \quad \text{and} \quad |x^1| \leq 1. \qquad (2.3)$$

(These are equivalent conditions if the second inequality is made strict.) The region covered when $x^2 > 0$, $-\pi/2 < \rho < \pi/2$, is shown in Fig. 4. Note that on the back side of the de Sitter space, where $x^2 < 0$ and $\pi/2 < |\rho| < \pi$, positive χ corresponds to _negative_ x^0.

In this coordinate system the metric tensor takes the form

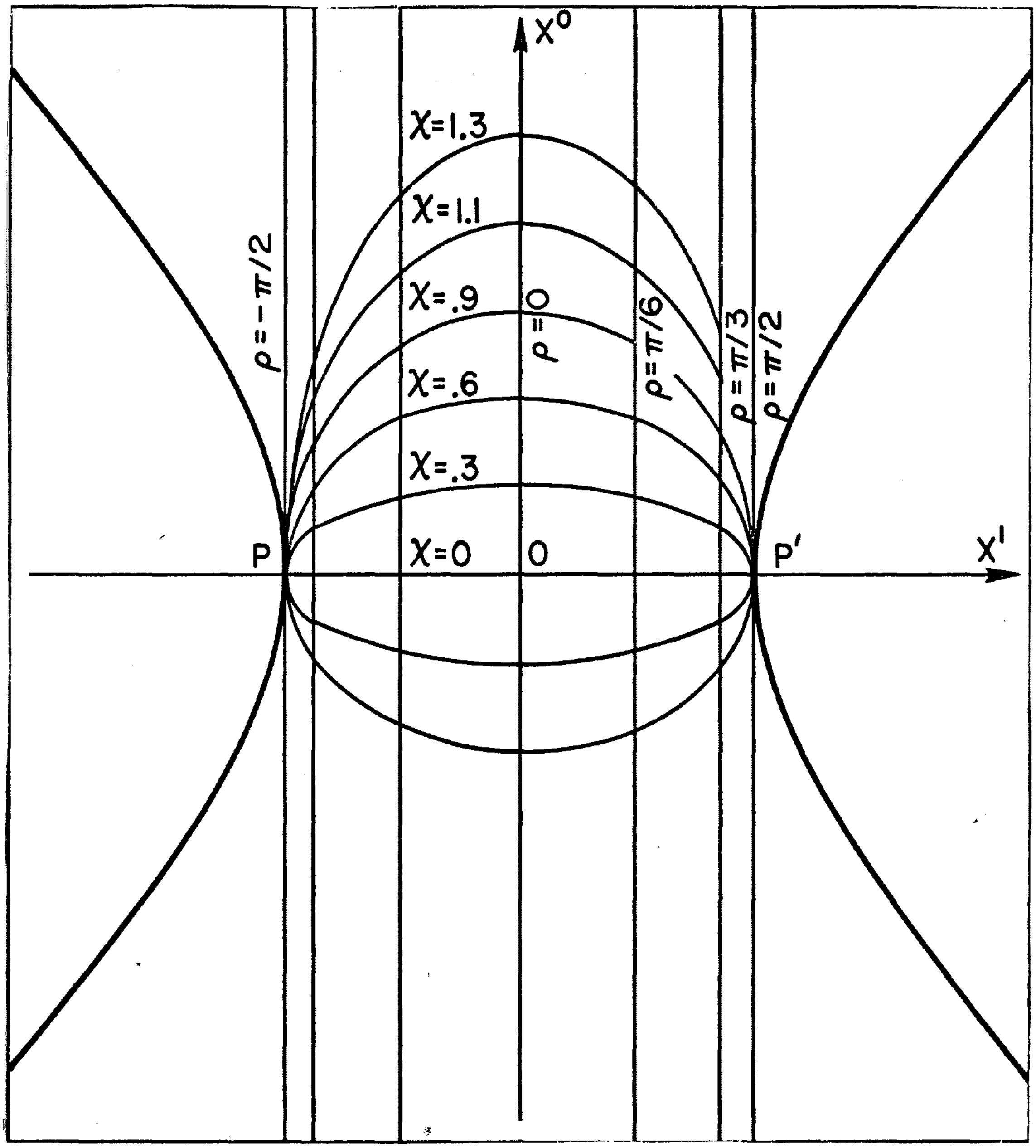


Fig. 4

Geodesic Fermi coordinates in two-dimensional de Sitter space (orthogonal projection onto x^0-x^1 plane).

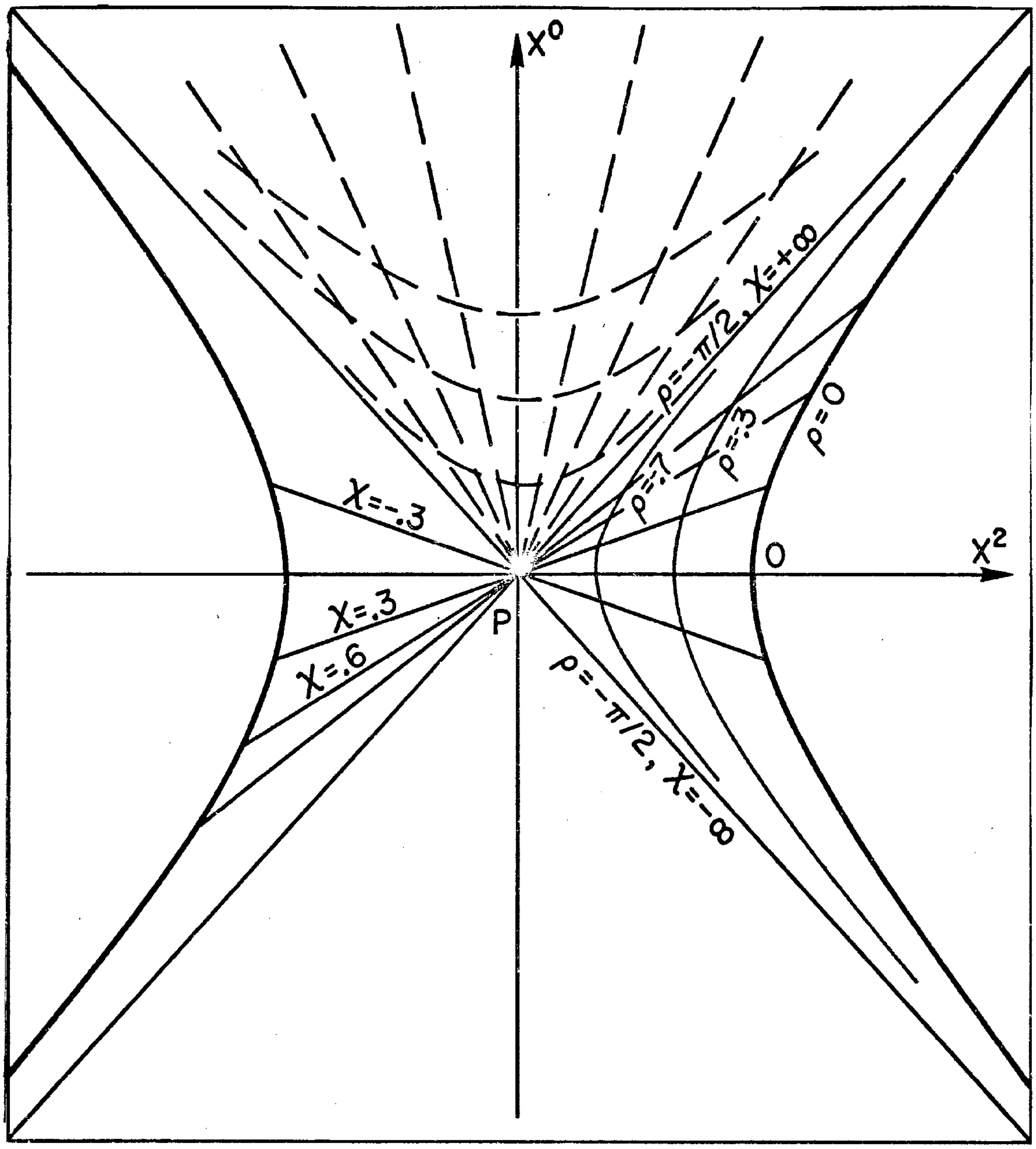


Fig. 5

Geodesic Fermi coordinates (side view).

$$ds^2 = \cos^2 \rho \, d\chi^2 - d\rho^2 . \qquad (2.4)$$

This metric has the following properties:

(1) It is orthogonal (and can be made so in higher dimensions as well).

(2) It is static (Eq. (D.5)). (This is a property of the space.)

(3) (χ, ρ) is the geodesic Fermi coordinate system (Eqs. (D.3)) based on the timelike geodesic defined by $\rho = 0$.

(4) The "time translation" $\chi \longrightarrow \chi + \chi_0$ is the element

$$\mathcal{T}(\chi_0) \equiv e^{+i\chi_0 H} = \begin{pmatrix} \cosh \chi_0 & 0 & \sinh \chi_0 \\ 0 & 1 & 0 \\ \sinh \chi_0 & 0 & \cosh \chi_0 \end{pmatrix} \qquad (2.5)$$

of the de Sitter group.

In the neighborhood of the points P and P' where $|\rho| = \pi/2$ this coordinate system coincides with the polar normal system in the sense of Appendix D (see Fig. 5). The other regions of the space (the triangular regions above and below P and P' in Figs. 4 and 5) appear in the normal coordinate system as a Robertson-Walker universe with a singularity at time zero.

This is the two-dimensional analogue of "Case 5" of Robertson and Noonan (see Appendix D).

In the general case we write

$$x^0 = \sinh \chi \cos \rho, \tag{2.6a}$$

$$x^j = \sin \rho \, g^j(\theta^2, \ldots, \theta^s) \qquad (j = 1, \ldots, s), \tag{2.6b}$$

$$x^n = \cosh \chi \cos \rho, \tag{2.6c}$$

where $(\rho, \theta^2, \ldots, \theta^s)$ form at $\chi = 0$ a polar normal coordinate system (see Appendix D) on the s-sphere in the neighborhood of the point 0. The metric takes the form

$$ds^2 = \cos^2 \rho \, d\chi^2 - d\rho^2 - \sin^2 \rho \, d\Omega^2, \tag{2.7a}$$

where

$$d\Omega^2 = \sum_{j,k,l=2}^{s} \frac{\partial f^j}{\partial \theta^k} \frac{\partial f^j}{\partial \theta^\ell} d\theta^k d\theta^l . \tag{2.7b}$$

This is the geodesic Fermi coordinate system based on the geodesic defined by $\rho = 0$ (cf. Eqs. (D.3)). The hypersurfaces of constant χ are the intersections of the de Sitter hyperboloid with the hyperplanes through the origin defined by

$$x^0/x^n = \text{const.} < 1. \tag{2.8}$$

(Contrast the horizontal slices (1.7) in the other picture.) These are geodesic hypersurfaces; they are s-spheres geometrically similar to $\{x|x^0 = 0\}$. They are mapped into one another by a one-parameter subgroup of $SO_0(1,n)$, which, as in Eq. (2.5), is expressed in the Fermi coordinate system as translation in the variable χ.

There is a coordinate singularity (horizon) at $\rho = \pi/2$. For fixed χ the coordinates $(\rho,\theta^2,\ldots,\theta^s)$ cover half of the s-sphere as ρ ranges between 0 and $\pi/2$. Of course, the region of space where $x^0/x^n \geq 1$ is not covered at all.

In particular, when n = 4 we set

$$x^1 = \sin\rho \sin\theta \cos\phi, \tag{2.9}$$

$$x^2 = \sin\rho \sin\theta \sin\phi \qquad (-\pi < \phi \leq \pi),$$

$$x^3 = \sin\rho \cos\theta \qquad (0 \leq \theta \leq \pi)$$

and have

$$ds^2 = \cos^2\rho \, d\chi^2 - d\rho^2 - \sin^2\rho \, [d\theta^2 + \sin^2\theta \, d\phi^2]. \tag{2.10}$$

If we set $r = \sin\rho$, Eq. (2.10) becomes

$$ds^2 = (1 - r^2)\, d\chi^2 - (1 - r^2)^{-1}\, dr^2$$

$$- r^2 [d\theta^2 + \sin^2\theta\, d\phi^2], \qquad (2.11)$$

the form in which the static de Sitter metric is most often written (e.g., [Tolman], p. 346). Near O we have

$$x^0 \sim \chi, \qquad x^1 \sim \rho \sin\theta \cos\phi,$$

$$x^2 \sim \rho \sin\theta \sin\phi, \qquad x^3 \sim \rho \cos\theta. \qquad (2.12)$$

Consequently, Eqs. (2.9) are just as reasonable as Eqs. (1.12) as a generalization to a finite region of the polar coordinate system in the infinitesimal neighborhood of O. The analogue of Eq. (1.15b) is

$$\exp(i\chi H)\, \exp(-i\phi J^3)\, \exp(-i\theta J^2)\, \exp(-i\rho P^3)\, SO_0(1,3). \qquad (2.13a)$$

A locally Cartesian coordinate system analogous to Eqs. (1.9) corresponds to the coset parametrization

$$\exp(i\chi H)\, \exp(i\theta^3 P_3)\, \exp(-i\theta^2 P_2)\, \exp(i\theta^1 P_1)\, SO_0(1,3). \qquad (2.13b)$$

Identifying $SO_0(1,3)$ with O (Eq. (I.1.2)), we calculate

$$x^0 = \sinh\chi \cos\theta_1 \cos\theta_2 \cos\theta_3 ,$$

$$x^1 = \sin\theta_1 \qquad \left(-\frac{\pi}{2} \le \theta_1 \le \frac{\pi}{2}\right),$$

$$x^2 = \cos\theta_1 \sin\theta_2 \qquad \left(-\frac{\pi}{2} \le \theta_2 \le \frac{\pi}{2}\right),$$

$$x^3 = \cos\theta_1 \cos\theta_2 \sin\theta_3 \qquad (-\pi < \theta_3 \le \pi),$$

$$x^4 = \cosh\chi \cos\theta_1 \cos\theta_2 \cos\theta_3 \qquad (2.14)$$

for the point (2.13b). This coordinate change, of course, maintains the static form of the metric; it just amounts to rotating the spherical coordinate system on the three-sphere so that O becomes an "equatorial" point.

The spatial coordinate system chosen on the s-sphere at each instant of time is not really very important for our considerations; the crucial point is that the <u>separation of time and space</u> is different according to the two ways of looking at de Sitter space -- the "static" point of view of this section or the spatially homogeneous and isotropic "Robertson-Walker" point of view of Sec. III.1.

3. Physical Significance of the Coordinate Systems.

These two types of coordinate system (or, perhaps more accurately, the two definitions of "constant time" reflected in them) provide alternative kinematical frameworks for describing physical processes in de Sitter space. On the classical level these descriptions are equivalent. That is, the behavior of test particles and light rays (namely, motion along geodesics) can be described in terms of intrinsic "geometrical" concepts, and one is free to transcribe their configurations into terms of any space-time coordinate system he likes. It will turn out, however, that fairly convincing generalizations of the canonical quantization procedure for quantum fields lead to different results, depending on which of these kinematical pictures is adopted as basic.

These remarks are relevant to dynamical formulations in which the state of a system is specified by the configuration of the system at an instant of time, and the equations of motion tell how the configuration changes with time. For instance, in the simplest type[3] of scalar field theory one has an operator-valued distribution,

[3] There is evidence (e.g., Powers (1967)) that in more singular models than have been constructed so far -- in particular, in most four-dimensional field theories with interaction -- it will be impossible to define field operators at fixed time (i.e., an integration over a test function depending on time as well as space will be needed in order to get an operator).

$$\phi(\vec{x}) = \int \frac{d^3k}{\sqrt{2\omega_{\vec{k}}}} [a_{\vec{k}} e^{i\vec{k}\cdot\vec{x}} + a^{\dagger}_{\vec{k}} e^{-i\vec{k}\cdot\vec{x}}] = \phi(0,\vec{x}),$$

which determines the expectation values of field measurements in various regions of space at a fixed time, given the state vector of the system. In the Heisenberg picture the field operator at different times is

$$\phi(t,\vec{x}) = e^{+iHt} \phi(\vec{x}) e^{-iHt},$$

where H is a global Hamiltonian operator. In the absence of a better idea, one would like to look at field theory in curved space from this point of view, too. (But see Sec. III.6 below for a complication which must be expected in the general case.)

We are trying to generalize the following picture in Minkowski space: An observer, idealized as a classical point particle moving at some fixed velocity, travels along a timelike geodesic (straight line). In a suitable orthonormal coordinate system (Lorentz frame) this line is the x^0-axis,

$$x^1 = \ldots = x^{n-1} = 0. \tag{3.1}$$

The spacelike hyperplanes orthogonal to this worldline are given by

$$x^0 = t = \text{const.} \tag{3.2}$$

Such a hyperplane represents the universe at time t in this observer's frame.

The natural generalization of these ideas to curved space is the following: An observer travels along some timelike geodesic, or worldline. The arc length t along the geodesic plays the role of time for him. The set of points constituting the "present" of the observer at time t is the hypersurface of points lying on the spacelike geodesics which are orthogonal to the worldline at the point labeled by t. For brevity let us call such a hypersurface an <u>instant</u>, since it represents physical space at an instant of time. Through each point of an instant there is an orthogonal timelike geodesic, which is (potentially) the worldline of an observer "at rest" at that point in space at that instant.

In de Sitter space the curve defined by $\rho = 0$ (or $\sigma_r = 0$) is a timelike geodesic (call it W), and $\tau = \chi$ is the natural time scale along it. Similarly, the hypersurface defined by $\tau = 0$ or $\chi = 0$ is an instant (call it J), and $\rho = \sigma_r$ is the distance of a point on it from the central observer at W. A coordinate system can be constructed which treats in this way the whole family of geodesics parallel to W at J (Gaussian coordinates), or the family of geodesic hypersurfaces perpendicular to W (Fermi coordinates), but not both at once.

In the Gaussian system of Fig. 3 the lines of constant σ can be interpreted as the worldlines of all possible observers who are at rest at the instant $\tau = 0$. The surfaces of constant τ mark off intervals of equal proper time on these worldlines.

That is, for fixed σ, τ is the private time of the observer at σ. Sometimes τ has been treated as a time coordinate in a global sense in de Sitter space (e.g., Philips (1963), Tagirov et al. (1967)). However, this is contrary to the spirit of the group-theoretical approach to dynamics. A translation in τ is not an element of the de Sitter group, and the hypersurfaces τ = const. are geometrically dissimilar -- in particular, they are not geodesics if $\tau \neq 0$. As τ increases these surfaces become increasingly distorted. The whole point of studying de Sitter space instead of some more general manifold is to exploit the existence of a symmetry group with the maximal number of parameters. One would like, therefore, to fit the kinematical description into the group-theoretical framework in analogy to ordinary special relativity, where time translation is an element of the Poincaré group and one instant is just like another, geometrically speaking.[4]

This objection is overcome if we identify time with the coordinate χ of the Fermi system (see Figs. 4-5). On the worldline W χ is equivalent to τ. The equal-χ hypersurfaces are

[4] For the same reason the author disagrees with the statement of Philips (1963), p. 49, that P (in our notation) should be identified with the physical momentum at each point of the two-dimensional de Sitter space. Of course, this is partly an arbitrary matter of definition, but a physically better definition would seem to be the following: Choose a local Lorentz frame at the point Q in question. In dimension 2 this amounts to choosing a timelike geodesic L through Q. Then the momentum at Q relative to this frame is the generator of the subgroup of isometries ($\cong$ SO(2)) which map the spacelike geodesic hypersurface orthogonal to L at Q into itself.

the images of J under the isometries $\mathcal{J}(\chi_0)$ (Eq. (2.5)), and these mappings preserve (ρ,θ) as the position coordinates at each instant (i.e., ρ is always the geodesic distance from W, and the meaning of the angular coordinates is unchanged).

Of course, we have now lost the possibility of describing the spacelike isometries by simple transformations on the spatial coordinates. In fact, the situation is worse: $\mathcal{S}(\sigma_0)$ (Eq. (1.5)) maps the region (2.3) covered by the coordinate system out of itself. Note also that the region covered depends upon both the position and the velocity of the standard observer.

Another disadvantage of this way of looking at de Sitter space is that $\mathcal{J}(\chi_0)$ is not really analogous to a Poincaré time translation except near the worldline W; on the other side of the universe it is a translation in the negative direction, and near $\rho = \pm \pi/2$ it resembles a homogeneous Lorentz transformation (see Fig. 5). This problem reflects geometrical peculiarities of the finite de Sitter space which no choice of coordinate system can completely overcome. It is closely related to the complicated relationship (see Chapters II and VI) between the irreducible representations of the de Sitter group and the Poincaré group.

Consider an (at least approximately) localizable system in a state Ψ for which the expectation value of H (the generator of $\mathcal{J}(\chi_0) = \exp(i\chi_0 H)$) is positive and the system is localized near O. For this state H can with some justification be called the energy. A translated state $\mathcal{J}(\sigma_0)\Psi$ gives a positive expectation value to $\mathcal{J}(\sigma_0) H \mathcal{J}(\sigma_0)^{-1}$, the generator of time translations in the neighborhood of the new point of localization. If the system is now translated in σ by Π, so that it is localized near the antipodal point O', the expectation value of H clearly must be negative, since here the transformations $\exp(i\chi_0 H)$ with $\chi_0 > 0$ move the system in the negative time direction. (In other words, $\mathcal{J}(\Pi) H \mathcal{J}(\Pi)^{-1} = -H$.) This explains why the spectrum of H must run through both positive and negative values in a single irreducible representation.[5] The possibility of contracting a representation to either a positive-energy or a negative-energy representation of the Poincaré group (Sec. II.1) corresponds to the possibility of identifying Minkowski space with the neighborhood of either O or O' in the contraction of spaces described in Sec. I.2. The Hilbert space of states of a localizable system in de Sitter space in some sense becomes limited, under contraction, to the states that are localized near O. By correspondence with standard theory, one expects the contracted operator H to be bounded below, as it is in Eq. (II.1.6b). However, the same formal contraction process must also be capable of yielding the Hilbert space of states localized

[5] This relationship was pointed out by E. P. Wigner. It stimulated the efforts of his students to define localized states in Minkowski and de Sitter space: Newton (1949), Newton and Wigner (1949), Philips (1963, 1964), Philips and Wigner (1968). The last of these references discusses the subject of this paragraph thoroughly.

near O', in which case one would expect H to have the opposite sign. This is achieved by a different choice of phase in Eq. (II.1.9). (See also Chapters II and VI.)

If we accept the Fermi picture -- that is, the definition of the "present" of an observer as the geodesic hypersurface orthogonal to his worldline -- the proper description of the configuration of the system at that instant is likely nevertheless to involve concepts related to the Gaussian picture. As remarked above, for a localized system "energy" is most convincingly associated with translation in the local time, which corresponds locally to translation in τ. At the instant J, differentiation with respect to τ has more physical significance than differentiation with respect to χ. For τ has the same geometrical meaning at all points of J, while χ depends strongly on the position of the observer, which should be irrelevant to describing the state of the whole system at a given instant. This principle will be applied in Secs. IV.2, V.3, and X.8-10.

All the remarks of this section about de Sitter space apply to more general Riemannian space-times, except for the references to isometries. It is always possible, given a point and a velocity through it, to construct Gaussian and Fermi coordinate systems (which, in general, will not cover the whole space) -- see Appendix D. The geometrical reasons for attributing physical significance to them are the same as here. In the general case these canonical coordinate systems will not be orthogonal away from the basic submanifolds W and J.

One might object to this whole discussion along the lines of Schrödinger:[6]

> Some authors hold, or at least favour, the view that the static frame is that of an 'observer' permanently at rest at the spatial origin . . . . But there is no earthly reason for compelling anybody to change the frame of reference he uses in his computations whenever he takes a walk. . . . Let me on this occasion denounce the abuse which has crept in from popular <u>exposes</u>, viz. to connect any particular frame of reference, e.g. in special relativity, with the behaviour (motion) of him who uses it. The physicist's whereabouts are his private affair. It is the very gist of relativity that anybody may use any frame. Indeed, we study, for example, particle collisions alternately in the laboratory frame and in the centre-of-mass frame without having to board a supersonic aeroplane in the latter case.

Of course, a scientist will be aware of points of space-time outside the region covered by his Fermi coordinate system -- he may even reach them himself[7] by changing his velocity -- and he will, when appropriate, describe both these and nearer regions by various kinds of coordinate systems. Conversely, the use of a Fermi coordinate system does not necessarily imply that one's laboratory is located on the central geodesic. Nevertheless, it seems to the author that there are circumstances in which a separation of space-time into space and time is useful, and that the geodesic hypersurface construction is the most reasonable way to define it. In applications of special relativity one does sometimes refer to <u>space</u> and to <u>time</u>, and the meaning of these words <u>does</u> depend on the Lorentz frame -- essentially, on the velocity of the observer. The canonical formalism of field

---

[6] [Schrödinger], p. 20, excerpts from text and footnote.
[7] This assumes a life-span of cosmological magnitude!

theory constrains us to make a separation into space and time. The ways of doing this described in Appendix D and Secs. III.1-2 have intrinsic geometrical significance, once a distinguished position and velocity (most naturally interpreted as those of an observer) are given.

A final remark: We have discussed coordinate systems of two different types based on the same given fundamental geodesics W and J. But even within one type, one must still consider different <u>frames</u>, corresponding to different choices of W and J. (These are analogues to Lorentz frames in special relativity.) It is not obvious that a physical theory defined by, say, a Hamiltonian formulation in a given frame will automatically be equivalent to a theory defined by the same prescription in a different frame. In fact, it will be seen in Chapter IX that the time-translation-invariant quantizations naturally associated with the various static frames in de Sitter space are different in this sense.

4. <u>Geodesics</u>.

The geodesics of the n-dimensional de Sitter space (I.1.1) are its intersections with the planes through the origin in the (n+1)-dimensional imbedding space ([Schrödinger], p. 3; Calabi and Markus (1962)). However, they can also be determined directly from the general definition:

$$\frac{d^2 z^\lambda}{ds^2} + \Gamma^\lambda_{\mu\nu} \frac{dz^\mu}{ds} \frac{dz^\nu}{ds} = 0 \qquad (4.1)$$

([Eisenhart], p. 50).

In an orthogonal coordinate system the Christoffel symbols and the Riemann curvature tensor are easily calculated ([Eisenhart], p. 44). The former are

$$\Gamma^\lambda_{\mu\nu} = 0, \qquad \Gamma^\lambda_{\mu\mu} = -(2g_{\lambda\lambda})^{-1} \frac{\partial g_{\mu\mu}}{\partial x^\lambda},$$

$$\Gamma^\lambda_{\lambda\mu} = (2g_{\lambda\lambda})^{-1} \frac{\partial g_{\lambda\lambda}}{\partial x^\mu}, \qquad \Gamma^\lambda_{\lambda\lambda} = (2g_{\lambda\lambda})^{-1} \frac{\partial g_{\lambda\lambda}}{\partial x^\lambda}. \qquad (4.2)$$

In the Gaussian coordinates (1.2-4) for the two-dimensional de Sitter space of radius 1, therefore, the Christoffel symbols and the geodesic equations (4.1) take the form

$$\Gamma^\tau_{\sigma\sigma} = \sinh\tau \cosh\tau, \quad \Gamma^\sigma_{\sigma\tau} = \Gamma^\sigma_{\tau\sigma} = \tanh\tau \quad \text{(others zero)}, \qquad (4.3)$$

$$\frac{d^2\tau}{ds^2} + \sinh\tau \cosh\tau \left(\frac{d\sigma}{ds}\right)^2 = 0,$$

$$\frac{d^2\sigma}{ds^2} + 2\tanh\tau \frac{d\tau}{ds}\frac{d\sigma}{ds} = 0. \qquad (4.4)$$

In the Fermi system (2.1-4) we have

$$\Gamma^{\rho}_{\chi\chi} = -\sin\rho\cos\rho,\ \Gamma^{\chi}_{\chi\rho} = \Gamma^{\chi}_{\rho\chi} = -\tan\rho \quad \text{(others zero)}, \tag{4.5}$$

$$\frac{d^2\chi}{ds^2} - 2\tan\rho\,\frac{d\chi}{ds}\frac{d\rho}{ds} = 0,$$

$$\frac{d^2\rho}{ds^2} - \sin\rho\cos\rho\left(\frac{d\chi}{ds}\right)^2 = 0. \tag{4.6}$$

In a two-dimensional manifold the curvature tensor has only one independent component. In the case of the two-dimensional de Sitter space of radius R (see Eq. (I.1.1)) it is

$$R_{\sigma\tau\tau\sigma} = R_{\tau\sigma\sigma\tau} = R^2\cosh^2\tau = R^{-2}(g_{\sigma\tau}g_{\tau\sigma} - g_{\sigma\sigma}g_{\tau\tau}). \tag{4.7}$$

This shows explicitly that de Sitter space is a <u>space of constant curvature</u> (see [Eisenhart], pp. 83-84).

From Eqs. (4.4) it follows that the curves

$$\tau = 0, \qquad \sigma = s \text{ [or } as + b\text{]} \tag{4.8}$$

and

$$\sigma = \text{const.}, \qquad \tau = s \tag{4.9}$$

are geodesics, as asserted in Sec. III.1. Eqs. (4.6) show that

$$\rho = 0, \qquad \chi = s \tag{4.10}$$

and

$$\chi = \text{const.}, \qquad \rho = s \tag{4.11}$$

are geodesics, as claimed in Sec. III.2.

The other geodesics through O in the two-dimensional space (with R = 1) can be found by transforming the curves (4.8) and (4.10) by the homogeneous Lorentz transformations at O,

$$\mathcal{B}(\alpha_0) \equiv e^{i\alpha_0 K} = \begin{pmatrix} \cosh \alpha_0 & \sinh \alpha_0 & 0 \\ \sinh \alpha_0 & \cosh \alpha_0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.12}$$

(see Eqs. (A.1.12b) and (I.3.5)). Thus the most general spacelike geodesic through O is

$$x^0 = \sinh \alpha_0 \sin s, \quad x^1 = \cosh \alpha_0 \sin s, \quad x^2 = \cos s, \tag{4.13}$$

and the general timelike geodesic is

$$x^0 = \cosh \alpha_0 \sinh s, \quad x^1 = \sinh \alpha_0 \sinh s, \quad x^2 = \cosh s. \tag{4.14}$$

One could now use Eqs. (1.2-3) or Eqs. (2.1-2) to get expressions for these curves in the coordinate systems (τ,σ) or (χ,ρ). However, one learns more (Wigner (1961)) by graphing Eqs. (4.13-14) in the x^1-x^2 plane. The spacelike geodesics are the ellipses

$$(x^1)^2/\cosh^2 \alpha_0 + (x^2)^2 = 1, \qquad (4.15)$$

and the timelike geodesics are the hyperbolas

$$(x^1)^2/\sinh^2 \alpha_0 - (x^2)^2 = -1. \qquad (4.16)$$

Moreover, the surfaces of constant εs (s = arc-length distance from O, $\varepsilon = +1$ [-1] for timelike [respectively, spacelike] geodesics) are the lines of constant coordinate x^2.

From this it is clear that the lines

$$x^0 = \lambda, \quad x^1 = \pm \lambda, \quad x^2 = 1 \qquad (s = 0) \qquad (4.17)$$

are the lightlike (null) geodesics through O. They are

$$\tau = \sinh^{-1} \lambda, \quad \sigma = \pm \tanh^{-1} \lambda \qquad (4.18a)$$

or

$$\chi = \tanh^{-1} \lambda, \quad \rho = \pm \sin^{-1} \lambda, \qquad (4.18b)$$

which are easily seen to satisfy Eqs. (4.4) and (4.6).

5. Causal Connection, Horizons, Domains of Dependence, Geodesic Completeness.

More important than the explicit form of the geodesics is the information about the geometrical structure of de Sitter space which can be deduced from their qualitative behavior. For the sake of visualizability let us discuss the two-dimensional universe. It will be clear that the structure of the higher-dimensional de Sitter spaces is quite similar. For more information on these subjects see Penrose (1968), pp. 186-196, and Geroch (1970), from which much of the information in this and the next section is taken.

Some care is needed in extending to curved space-time the familiar notions of timelike, spacelike, and lightlike separation of points. Two points will be said to be <u>causally connected</u> if they are connected by a timelike or lightlike <u>curve</u>[8] (not necessarily a geodesic). In the closed de Sitter space, however, all the points causally connected to a point Q lie on causal geodesics through Q. We shall always assume (see remarks at the beginning of Chapter VII below) that the manifold under consideration has a distinguished time orientation. Then the points which are causally connected to Q can be further classified as the <u>past</u> or the <u>future</u> of Q. (These are disjoint unless the space admits closed timelike curves -- see Sec. III.6.)

Just as in flat space, through each point Q there is a cone of lightlike geodesics, which in a two-dimensional space degenerates to two curves, as in Eqs. (4.17-18). The light cone of Q separates the points in the neighborhood of Q which are connected to Q by a timelike geodesic from those which are

---

[8] I.e., a smooth curve with timelike or lightlike tangent vector at each point.

connected by a spacelike geodesic. In de Sitter space the branches of the light cone of O asymptotically approach those worldlines orthogonal to the instant J ($\tau = \chi = 0$) which pass through the points P and P' ($x^1 = \pm R$, $x^0 = x^2 = 0$) located one quarter of the way around the world.

In de Sitter space the points inside and on the light cone of the antipodal point O' ($x^2 = -R$, $x^0 = x^1 = 0$) are not connected to O by any geodesic. Nevertheless, they are connected by nongeodesic spacelike curves; in fact, it is easy to see that any two points in de Sitter space can be connected by a spacelike curve (running all the way around the closed universe if necessary). A timelike or lightlike geodesic through O' can never intersect one through O. Thus, an observer at O and one at O' are completely isolated from each other; not only can they never meet, but they are not acted on by any common influence in the past and cannot both influence any event in the future. This situation has no analogue in Minkowski space. For points inside the future [past] light cone of O' a similar statement can be made about their contact with O in the future [past].

Another interesting separation of the space which has no analogue in special relativity is defined relative to a timelike geodesic. For our standard geodesic W ($\rho = \sigma = 0$) it is marked out by the light cones of P and P', which meet W asymptotically. (These are shown in Figs. 4 and 5.) The parts of the cones which approach W in the future form the <u>event horizon</u>, which separates events which are observable by an observer on W from those which are not. That is, no future-directed lightlike (or timelike) curve through a point beyond the horizon intersects W. (An observer at O can intercept signals from this point, however, if he changes his velocity appropriately, provided the point is not in the future part of the geodesically isolated region discussed above.) The branches of the cones which approach W in the past are called the particle horizon. The points on the near side of the particle horizon are those which an observer on W can influence, in principle; or, those from which he can be observed.

The intersection of the near regions defined by the two horizons is the region (Eq. (2.3) with $x^2 > 0$) covered by a connected patch of the Fermi coordinate system (see Fig. 4). These are the points which at some time or another are contemporaneous with the observer on W in what was argued in Sec. III.3 to be the physically natural sense. These points could be called historical (relative to W), and the others extrahistorical.

Horizons can be defined for any timelike curve in any space-time, but they are located at infinity (in other words, are vacuous) in some cases. For instance, in Minkowski space all points are historical with respect to a timelike geodesic (but not with respect to a timelike hyperbolic curve -- a situation which will be studied in Chapter IX).

We will study the generalized Klein-Gordon equation in de Sitter space in Chapter V and in more general spaces in

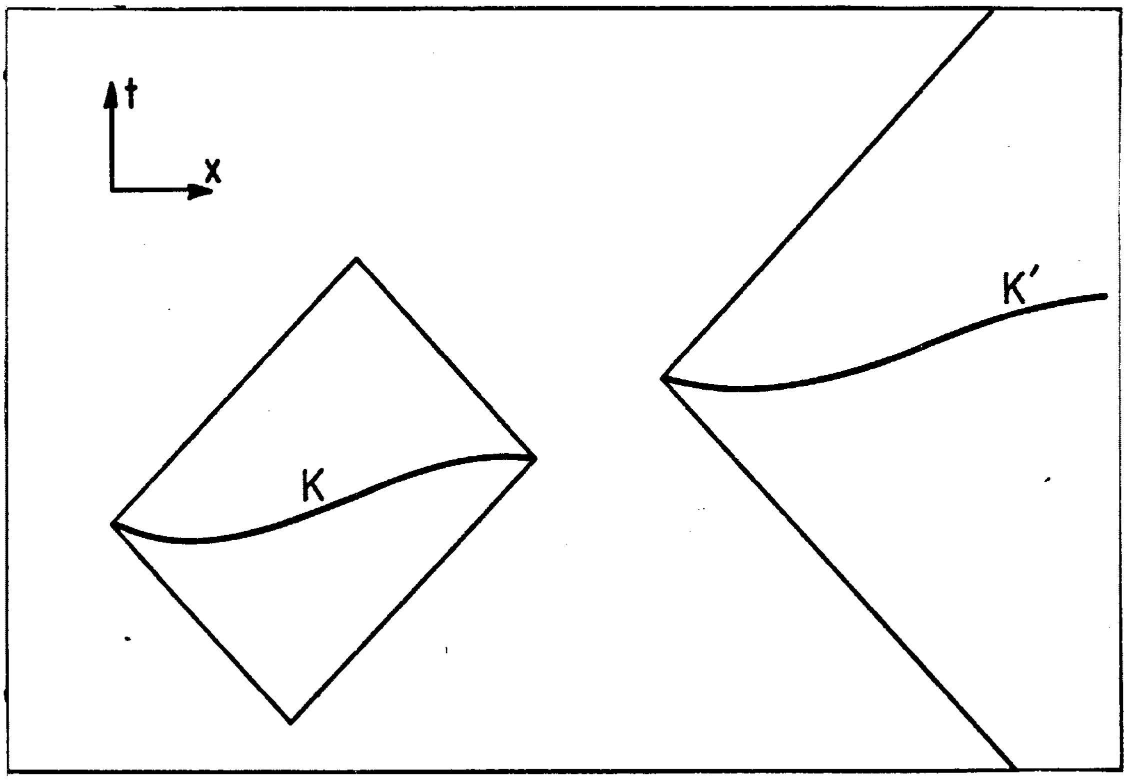


Fig. 6

Domains of dependence and Cauchy horizons in flat space for bounded and unbounded spacelike sets K and K'.

Chapters VII-X. The lightlike geodesics are the characteristics of this equation. Let S be a segment of a spacelike hypersurface. A solution of the equation is completely determined by prescribed values of the function and its time derivative on S within a region called the domain of dependence of S.[9] It is the union of the set of points x such that every timelike curve through x can be extended into the past to intersect S and the set of x satisfying the analogous condition

[9] In general one expects it to be the maximal region with this property, but see Sec. V.8 for a counterexample.

with "past" replaced by "future". If space-time has no closed timelike curves and S is sufficiently nice, the domain of dependence is, as suggested in Fig. 6, a diamond-shaped region outlined by lightlike surfaces called *Cauchy horizons*. In general space-time the domain may have such a boundary even if S extends to infinity -- see the next section.

A hypersurface S whose domain of dependence is the whole space-time is called a *Cauchy surface* for that space-time. Then the Cauchy problem with initial data on S is well-posed. Not every space-time contains a Cauchy surface. Geroch (1970) has proved:

(1) Existence of a Cauchy surface is equivalent to *global hyperbolicity*, a technical condition needed to prove existence and uniqueness theorems for hyperbolic partial differential equations on a manifold (see, e.g., Choquet-Bruhat (1968)).

(2) If a Cauchy surface S exists, the space-time is topologically of the form $S \times \mathbb{R}$, and the "slices" $S \times \{a\}$ can be chosen so that they are all Cauchy surfaces.

In the course of this dissertation we shall consider several examples of the situation described in point (2) (cf. Secs. III.5,6,7, IX.1, X.2). In each case a portion of a space of constant curvature will be covered by a coordinate system of either the Gaussian or the Fermi type, and the surfaces of constant time will be Cauchy surfaces for that region.

From the geometrical information at our disposal we can draw the following conclusions about the Cauchy problem in closed de Sitter space:

(1) A spacelike slice of the form (1.7) is a Cauchy surface for the entire de Sitter space.

(2) The domain of dependence of half of a geodesic spacelike hypersurface (where $|\rho| < \pi/2$ and $\chi = 0$ in some geodesic Fermi coordinate system) is the set of points which are historical (see above) with respect to the timelike geodesic ($\rho = 0$) which passes perpendicularly through the center of that segment of hypersurface.

The latter is an important observation: it means that within the connected region covered by a Fermi coordinate system the Cauchy problem is well-posed for each hypersurface $\{x|\chi = \text{const.}\}$. Thus this patch of space may, as far as the wave equation is concerned, be consistently considered a universe in itself (static!), for which the surfaces of constant χ are Cauchy surfaces.

Although the static de Sitter universe is Cauchy-complete, it is not geodesically complete. This is most vividly explained in terms of the physical pictures introduced in Sec. III.3. Consider one of the worldlines $W' = \{x|\sigma = \text{const.} \neq 0\}$ in relation to the Fermi coordinate system based on the worldline W. At some finite value of τ, the proper time of an observer traveling on W', W' hits the event horizon of W and passes out of the (χ,ρ) universe. Of course, this happens at time $\chi = \infty$ from the point of view of an observer on W.

Geodesic completeness can be characterized intrinsically, without a priori knowledge of whether the space is part of a larger space. One requires that on every geodesic which has been extended as far as possible the values of an affine parameter become arbitrarily large. (See Geroch (1968).) De Sitter space is geodesically complete, because the spacelike geodesics (4.13,15) are periodic and the timelike (4.14,16) and lightlike (4.17,18) geodesics continue to infinity in their affine parameters (e.g., τ and λ respectively).

6. The Open De Sitter Space.

The n-dimensional open de Sitter space is defined by

$$(x^0)^2 - (x^1)^2 - \ldots - (x^{n-1})^2 + (x^n)^2 = + R^2 . \qquad (6.1)$$

In general it is topologically different from the closed space with which we are primarily concerned. When n = 2, however, the open space $(SO_0(2,1)/SO_0(1,1))$ is geometrically identical to the closed space $(SO_0(1,2)/SO_0(1,1))$. Only the physical interpretation is different: time and space are interchanged. The calculations of Secs. III.1,2,4 still apply, but now τ and χ are space coordinates and σ and ρ are time coordinates. One must look at Figs. 3-5 from the side! The physically relevant structural properties of the space, consequently, are quite different from those of the closed space described in the previous section, and provide an interesting comparison.

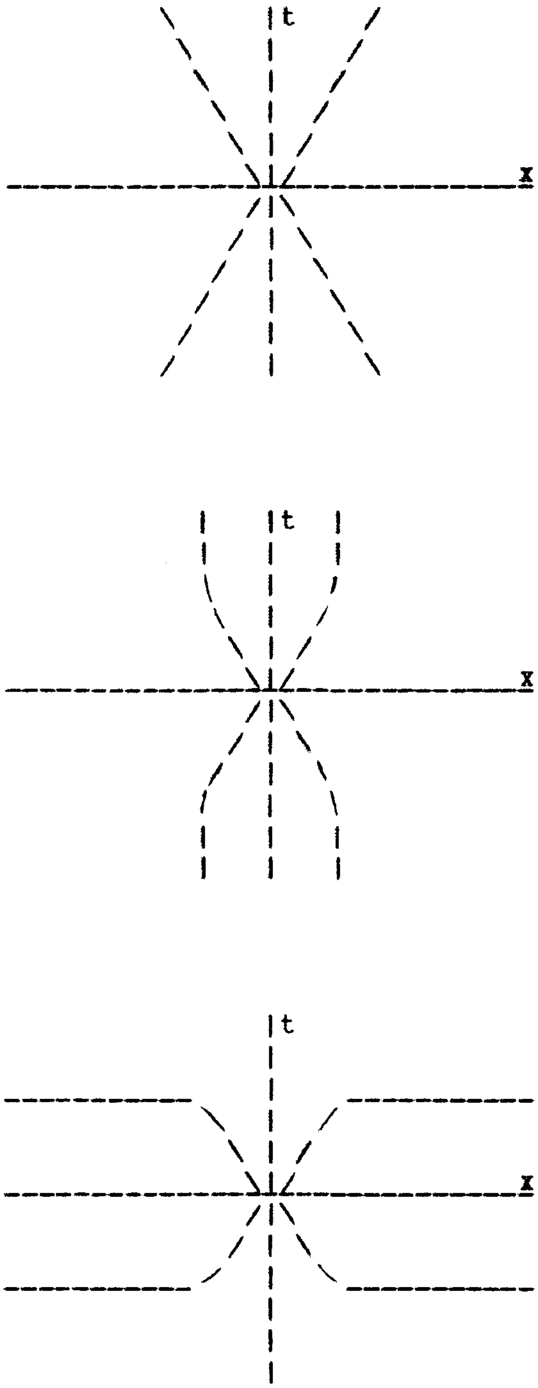
The differences are primarily due to the asymptotic behavior of the light cones, schematically indicated in Fig. 7. In the open space the light cone of a point Q becomes at large spatial distances asymptotically parallel to a geodesic spacelike hypersurface W. The timelike geodesics normal to W all come together at Q (cf. Figs. 4-5). As a consequence a geodesic Gaussian coordinate system (ρ,χ) does not cover the whole space. A geodesic Fermi system (σ,τ) does, however, and it gives rise to a manifestly static metric.

It also follows that there are causal curves which connect points which are not joined by geodesics. In fact, all points in open de Sitter space are causally connected.

It is easy to see that there are no event or particle horizons relative to geodesics in this space.

The formulation of the Cauchy problem in open de Sitter space is very complicated. First, the presence of closed timelike curves wreaks havoc with an initial-value problem. In the present case every point is in a position to influence every other point, even points on the same spacelike hypersurface. There are three possible ways out of this difficulty. (They will all be investigated and related to each other in Sec. V.8.)

First, one could investigate whether imposing the



(a) Minkowski space

(b) Closed de Sitter space

(c) Open de Sitter space

Fig. 7

Behavior of light cones in universes of constant curvature.

condition that the solutions of the wave equation transform under a unitary representation of the de Sitter group allows the boundary condition of periodicity in the time to be reconciled with prescribed initial conditions.

Second, one could restrict attention to the region covered by a Gaussian coordinate system. Here there are no closed timelike curves, and each surface of constant ρ is a Cauchy surface for the restricted space. This model is the two-dimensional analogue of "Case 6" of Robertson and Noonan -- it is a Robertson-Walker universe which expands from a singularity and then contracts again (see Appendix D). This space is not geodesically complete, and, of course, it is not invariant under the action of the de Sitter group.

Finally, one could consider the universal covering space of the hyperboloid (6.1). That is, we allow σ to range from $-\infty$ to $+\infty$ and values of σ which differ by a multiple of 2π are <u>not</u> identified. This space is sketched in Fig. 8. (The spatial infinities have been mapped in to finite locations, and light cones appear as diagonal straight lines.) Wigner (1950) has argued that this is a

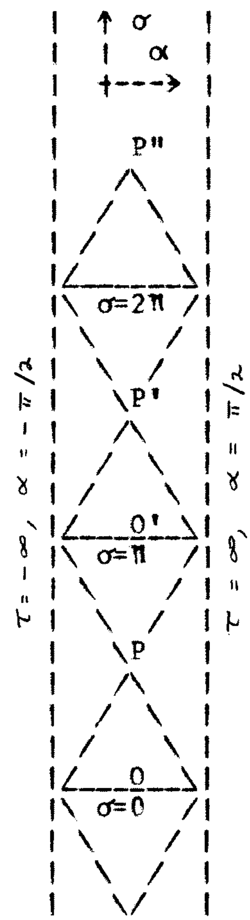


Fig. 8

Covering space of open de Sitter space. The coordinate α is defined in Eq. (V.2.5).

physically reasonable space to study in quantum theory. The symmetry group is the infinite-sheeted universal covering group of $SO_0(2,1)$ (see Sec. B.1). Now the points which lie on spacelike geodesics through O are not causally connected to O. (The σ-translates of these points by a multiple of π, on the other hand, are causally connected but not geodesically connected to O. The timelike geodesics through P are confined to the sequence of diamonds in Fig. 8.)

The domain of dependence of the instant $W = \{x \mid \sigma = 0\}$ is the diamond region covered by the Gaussian coordinates (just discussed). For the whole covering space one would not expect the Cauchy problem to be well-posed. The value of a solution of the wave equation at a point beyond the Cauchy horizon of the initial instant can be influenced by "information which comes in from infinity" along causal curves which do not intersect W. This example points up a complication which will be encountered in applying canonical field quantization to some Riemannian space-times, even with static metrics. In Sec. V.8 the circumstances under which this problem arises in the open de Sitter space will be determined.

7. De Sitter Space as a Euclidean Robertson-Walker Universe (Horospherical Coordinates).

Another coordinate system which has historically been used in the study of the de Sitter universe is the system of Lemaître and Robertson, defined in the two-dimensional case by

$$x^0 = \sinh t + \frac{1}{2} r^2 e^t,$$

$$x^1 = r e^t, \qquad (7.1a)$$

$$x^2 = \cosh t - \frac{1}{2} r^2 e^t.$$

The extension to higher dimensions is simple:

$$x^j = r^j e^t \qquad (1 \leq j \leq n - 1) \qquad (7.1b)$$

defines a Cartesian spatial coordinate system. The range of the coordinates is

$$-\infty < t < \infty, \qquad -\infty < r^j < \infty. \qquad (7.2)$$

The metric is

$$ds^2 = dt^2 - e^{2t} dr^2, \qquad (7.3a)$$

where in higher dimensions

$$dr^2 = (dr^1)^2 + \dots + (dr^{n-1})^2. \qquad (7.3b)$$

The coordinates cover half of the de Sitter hyperboloid (Fig. 9), the region bounded (in dimension 2) by the lightlike geodesics through P and P' which asymptotically approach the worldline W in the far past (the particle horizon of W). Thus

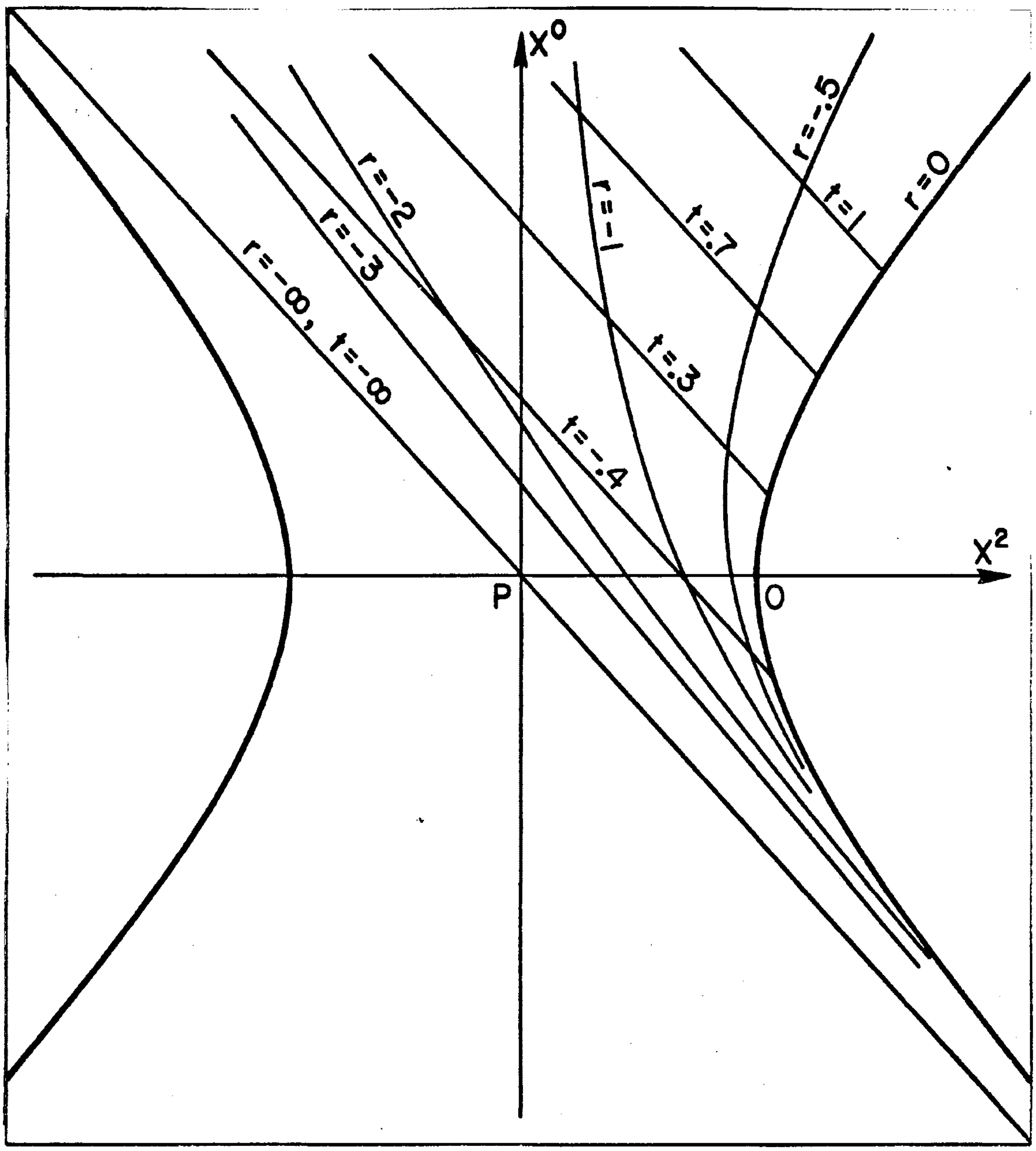


Fig. 9

The Lemaître-Robertson or horospherical coordinate system.

the universe consists of all the points our standard observer (or an observer on any of the geodesics given by r = const.) can in principle observe. The metric (7.3) displays this space as a Robertson-Walker universe which is always expanding. (This property explains the popularity of the Lemaître-Robertson coordinates when de Sitter space is taken seriously as a model of the actual universe -- see [Robertson-Noonan], pp. 365-367.[10]) It is clear that a change of the signs of t and x^0 in Eqs. (7.1) would yield a contracting universe bounded by the event horizon of the central worldline.

The metric of Eq. (7.3) is Gaussian, but none of the spaces of constant time, which are the intersections of the hyperboloid with the planes $x^0 + x^n = \text{const.}$, is a geodesic hypersurface. However, (t,r) can be regarded as the geodesic Gaussian system built on the particle horizon; this cone is a sort of limit as $t \longrightarrow -\infty$ of the spacelike geodesic hypersurfaces orthogonal to W, as we shall see at the end of this section.

The hypersurfaces of constant time, given by

$$F(x) = -1, \qquad x^0 + x^n = \text{const.}, \tag{7.4}$$

are isomorphic to Euclidean (n-1)-space as Riemannian manifolds. That is, they are not only open and infinite (cf. Eqs. (7.2)),

[10] It is specifically this model which these authors call "the de Sitter universe".

but actually flat. This is a trivial statement for $n = 2$, but it is quite significant for $n > 2$, where the spacelike hypersurfaces (1.7) and (2.8) are definitely curved.

The hypersurfaces (7.4) are Cauchy surfaces for the half of de Sitter space which is covered by these coordinates.

The metric (7.3) is obviously invariant under the transformations

$$r \longrightarrow r + r_0 \tag{7.5}$$

and

$$t \longrightarrow t + t_0, \qquad r \longrightarrow e^{-t_0} r. \tag{7.6}$$

The generator of the "spatial translations" (7.5) is

$$L_{10} + L_{21} = i(K - P) \tag{7.7}$$

(see Eqs. (7.1), (A.1.11-12), (I.3.5)). So this is a parabolic subgroup of $SO_0(1,2)$ (see Sec. II.3). The "dilation" (7.6) is just $\mathcal{T}(t_0)$ (Eq. (2.5)). The statement in [Robertson-Noonan] (pp. 347, 348, 365) that "the de Sitter universe is the only nonstatic stationary model" (doubly nonstandard terminology!) refers to the existence of the symmetry (7.6) (see pp. 323 and 346-348 of the book).

The coordinate system (7.1) corresponds to the coset decomposition

$$e^{ir(K - P)} e^{itH} SO_0(1,1). \qquad (7.8a)$$

In higher dimensions this becomes

$$e^{i\vec{r}\cdot(\vec{K} - \vec{P})} e^{itH} SO_0(1,n-1). \qquad (7.8b)$$

This decomposition (which clearly does not cover the whole group) is discussed by Hannabuss (1969a). It is closely related to the Iwasawa decomposition[11] of $SO_0(1,n)$, in which $SO_0(1,n-1)$ is replaced by SO(n). The hypersurfaces (7.4) are the <u>horospheres</u> which are widely used in modern harmonic analysis (e.g., [Gel'fand 5]).

Börner and Dürr (1969) have studied quantum field theory in the four-dimensional de Sitter space using the horospherical coordinate system (7.1-3). Their solution of the eigenvalue problem for the Casimir operator (pp. 681-690) is essentially the decomposition of the quasiregular representation by the horospherical method, described in elementary terms.

To see the relationship of horospherical to geodesic Gaussian coordinates, operate on Eqs. (1.2) with $\mathcal{T}(\chi)$ (Eq. (2.5)):

[11] See, e.g., [Hermann], pp. 40-44.

$$x^0 = \sinh(\tau - \chi) - \sinh\chi \cosh\tau (\cos\sigma - 1),$$

$$x^1 = \cosh\tau \sin\sigma, \qquad (7.9)$$

$$x^2 = \cosh(\tau - \chi) - \sinh\chi \sinh\tau (\cos\sigma - 1).$$

This maps the geodesic hypersurface $\{x|\tau = 0\}$ on which the (τ,σ) system is based back toward $\chi = -\infty$; that is, to the light cone at $t = -\infty$ in the Lemaître-Robertson picture (Fig. 9). Now let

$$t = \tau - \chi, \qquad r = \frac{1}{2}\sigma e^{\chi}, \qquad (7.10)$$

let $\chi \longrightarrow \infty$ (replacing $\cosh\chi$ and $\sinh\chi$ by $e^{\chi}/2$), and expand the trigonometric functions up through order r^2. The result is Eqs. (7.1a). Thus we have exhibited the group of r-translations (7.5,7) as a contraction of the group of σ-translations (1.5), or, better, as a <u>limit</u> of the SO(2) subgroups of $SO_0(1,2)$ in the sense of Hermann ([Hermann], pp. 86-101; see also [Hermann 2]). In the general case we have $ISO_0(1,n-1)$, the symmetry group of the spaces (7.4), as a limit of the SO(n) symmetry groups of the spaces (1.7) and their images under $\mathscr{T}(\chi)$.

As explained in the Introduction, we are interested in de Sitter space as a <u>finite</u> universe of constant curvature. We shall work, therefore, mostly with the coordinate systems of Secs. III.1-2. The horospherical picture has been discussed for the sake of completeness and to emphasize that a given space-time manifold may be split into space and time in several different ways. As we shall see in Chapters IX and X, these alternatives are associated with different procedures of canonical quantization of a field.

Chapter IV

AXIOMS FOR QUANTUM FIELD THEORY IN DE SITTER SPACE

With the geometrical preliminaries out of the way, we are ready to consider the possibility of quantum field theory in de Sitter space. Let us first see how far we can go in rewriting the Wightman axioms for general field theory (see Appendix E) so that they apply to the closed de Sitter space. We shall find the spectral condition to be the major stumbling block. This might have been anticipated from the absence of an obvious analogue of the energy operator in the Lie algebra of the de Sitter group. It is a major contention of this dissertation, however, that this problem is only a special case of a gap in our understanding of the notion of quantum field with respect to curved space-time in general; the claim (see Philips (1963), Fronsdal (1965), Castell (1969)) that there is no such ambiguity in the case of the open de Sitter space (because there is a global time translation group whose generator can be made positive definite) deserves critical examination from a physical point of view.

In later chapters we shall try to construct the quantum theory of a free (i.e., not self-interacting) neutral scalar field in de Sitter space as a special case of such a field in a general Riemannian space-time. The reader should be warned that the conclusion will be that it is not obvious that the most reasonable theory on physical grounds must satisfy the axioms proposed here. In the last section of the chapter the relation between the contents of this chapter and what comes later will be explained further.

As in Chapter I we shall denote the de Sitter space-time manifold by M.

1. Axioms with Straightforward Generalizations.

No change is needed in the first axiom, which deals only with general principles of quantum theory, without reference to space-time:

1. Quantum theory. The states of the theory are described by unit rays in a separable Hilbert space $\mathcal{H}$.

Naturally, one expects to keep the second axiom with a change in the symmetry group:

2. Relativistic invariance. The relativistic transformation law of the states is given by a continuous unitary representation U(A) of the universal covering group of the de Sitter group $SO_o(1,n)$.

See Sec. B.1 for a description of the covering group. One of our major concerns in later chapters will be whether Axiom 2 is consistent with an approach to field theory in curved space which generalizes to arbitrary space-times without any symmetry group.

The next group of axioms deals with the field operators. We are faced first with the problem of choosing a test-function space.

3. <u>Existence and temperedness of the fields</u>. There is a topological vector space $\mathcal{T}$ of functions defined on space-time such that for each f $\in \mathcal{T}$ there exists a set $\phi_1(f), \ldots, \phi_n(f)$ of operators. These operators, together with their adjoints $\phi_1(f)^\dagger, \ldots, \phi_n(f)^\dagger$, are defined on a linear domain D of vectors, dense in $\mathcal{H}$. The $\phi_j(f)$ and $\phi_j(f)^\dagger$ leave D invariant. If $\Phi, \Psi \in D$, then $(\Phi, \phi_j(f)\Psi)$ as a functional of f is a member of the dual space $\mathcal{T}^*$ (a space of distributions).

We shall write symbolically

$$\phi(f) = \int_M d\mu(x)\ \phi(x) f(x) \tag{1.1a}$$

where

$$d\mu(x) = \sqrt{|g|}\ dx^1 \ldots dx^n \tag{1.1b}$$

is the invariant volume element on M. This convention makes both ϕ and f scalar objects if ϕ is a scalar field (cf. Sec. VII.3).

What test-function space $\mathcal{T}$ should we choose? Even in the ordinary relativistic theory this is a somewhat arbitrary choice. In Wightman and Gårding (1965) it is taken to be $\mathcal{D}$, the space of C^∞ functions of compact support; in [Streater-Wightman] it is $\mathcal{S}$, the C^∞ functions of fast decrease. There is no trouble in defining what it means to be a C^∞ function of compact support on M. The likeliest analogue of the condition of rapid decrease is

$$\left|\tau^p d^q f/d\tau^q\right| \longrightarrow 0 \quad \text{as } |\tau| \longrightarrow \infty \text{ for all integers } p,\ q, \qquad (1.2)$$

where τ is the time coordinate of Sec. III.1, which is the geodesic arc length in a certain timelike direction. (Since the universe is spatially finite, no falloff conditions are needed in spacelike directions.) Then the topology in this space and its dual can be defined by seminorms in the usual way (see [Streater-Wightman], pp. 33-34). The condition (1.2) is not manifestly invariant under the de Sitter group, since it depends on a particular Gaussian frame. (Neither is the standard definition of $\mathscr{S}$ manifestly Lorentz-invariant.) But by comparing Eqs. (III.1.2) and (III.7.9) one sees that the time coordinates in two Gaussian frames are asymptotically related by

$$\tau' \sim \tau + f(\sigma), \qquad \frac{d\tau'}{d\tau} \sim 1 \qquad (\tau \longrightarrow \infty), \quad (1.3)$$

where $f(\sigma)$ is bounded, so that the definition is really independent of frame.

We state the transformation law of the fields only for the scalar case:

4s. <u>Tensorial character of the fields</u>. The U(A) leave D invariant, and the equation

$$U(A)\ \phi(f)\ U(A)^{-1} = \phi(Af) \tag{1.4a}$$

is valid when each side is applied to any vector in D, where

$$Af(x) = f(A^{-1}x). \tag{1.4b}$$

There is a variety of ways to generalize the notions of tensor and spinor field to Riemannian space-time in general and especially to de Sitter space (where the de Sitter group and its representations are available to be thrown into the mathematical mix). Which of these formalisms to adopt seems to be at least partly a matter of taste. (For some rival choices in the case of de Sitter space see Dirac (1935), Nachtmann (1967), Börner and Dürr (1969), Hannabuss (1969a). The equivalence of several has been shown by Castagnino (1970). Well-developed formalisms for arbitrary space-times are those of Lichnerowicz (1961) and Penrose (1965).) Since the present work has been limited to scalar fields, we shall not go into this subject.

Every point x in M has an associated light cone which delimits the regions which are related to x in a causal way (see Secs. III.4-5). One expects that measurements at points outside the light cone will be independent of measurements at x.

5. Local commutativity. If there is no pair of points $x \in \text{supp } f$ and $y \in \text{supp } g$ such that x and y are causally connected, then one or the other of

$$[\phi_j(f), \phi_k^{(\dagger)}(g)]_{\pm} = 0 \qquad (1.5)$$

holds for all j and k when the left-hand side is applied to any vector in D.[1]

The three axioms expressing the existence, uniqueness, and cyclicity of the vacuum make sense as they stand:

6. <u>Existence and uniqueness of the vacuum</u>. There is a state Ψ_0, the vacuum, invariant under U, unique up to a phase factor.

7. <u>Cyclicity of the fields</u>. There is a state which is cyclic for the smeared fields; that is, polynomials in the smeared field components, $P(\phi_1(f), \phi_\lambda^{(\dagger)}(g), \ldots)$, applied to this state yield a set D of vectors dense in $\mathcal{H}$.

8. <u>Cyclicity of the vacuum</u>. Ψ_0 is in D and is cyclic.

These axioms, like Axiom 2, are stated only tentatively. It is not obvious that a theory in which the full role of the Poincaré group in special relativistic field theory is attributed to the de Sitter group is the physically most reasonable theory of fields in de Sitter space. Let us be more specific. The geodesic Gaussian (Sec. III.1) and horospherical (Sec. III.7) coordinate systems represent de Sitter space as an <u>expanding</u>

---

[1] The superscript (†) stands for the presence or the absence of an Hermitian conjugation.

universe. It is quite plausible that in such a situation particles will be produced by interaction with the gravitational field. (This problem has been studied by Parker (1966, 1968, 1969, 1971) and is discussed in Chapter X.) Then the no-particle state will not be invariant under the de Sitter group.[2] It is conceivable that there will not be any de-Sitter-invariant state in the theory at all. On the other hand, the de Sitter universe is static from the point of view of the Fermi system (Sec. III.2), but the theory (which has a no-particle state) which is thereby suggested is not de-Sitter-invariant, for reasons to be explained in Chapter IX.

Incidentally, it is not clear that Axiom 7 is equivalent to the statement that the algebra of field operators is irreducible. The usual proof ([Streater-Wightman], p. 141) that cyclicity of the vacuum implies irreducibility depends crucially on the spectral condition.

2. The Spectral Condition.

The crux of the spectral condition (Axiom 9, Appendix E) is that the energy operator P^0 is positive. The generator we have called H (see Secs. I.3 and III.2) is not positive in any irreducible unitary representation of $SO_0(1,2)$ (see Sec. II.3). As we have seen in Sec. III.3, there is a good geometrical reason

---

[2] As it stands this assertion is a *non sequitur*, since a time translation in these coordinate systems is not an element of the group. The claim will be substantiated in Chapters V and X, however.

for this. Philips and Wigner (1968) (pp. 632, 635-639) have shown that <u>no</u> element of $\mathcal{L}(SO_0(1,4))$ can be represented by a positive operator, because each element can be transformed into its negative by an element of the group.

Another aspect of the same problem is the difficulty of separating the solutions of the wave equation in de Sitter space into positive- and negative-energy functions according to any convincing definition (see Philips (1963) and Secs. V.3-6 below). This, however, is a general problem affecting field quantization in all metrics which are not manifestly static (see Chapter X). The impression that it does not arise for the open de Sitter space (see, e.g., Philips (1963)) is due to the existence in that case of a <u>global</u> coordinate system which gives the metric a static form.

As we have seen in Chapter II, there is a correspondence between the irreducible representations of the de Sitter group and those of the Poincaré group for space of the same dimension. In particular, the principal series of representations of $SO_0(1,n)$ contracts to the representations of $ISO_0(1,n-1)$ with timelike momentum spectrum. A plausible first step toward incorporating a spectral condition into a de-Sitter-invariant theory would be to require that only representations of the principal series appear in the decomposition of U(A) into irreducibles. This axiom could not be expected to do the entire job of the spectral condition, because it does not touch the problem that both past-directed and future-directed timelike momenta can be extracted by contraction from the same irreducible representation of the de Sitter group (Secs. II.1 and II.3). In physical terms it would be expected to exclude tachyons but not slower-than-light particles with energy unbounded below.

Unfortunately, even this minimal restriction on the representation $U(A)$ is probably untenable. Pukánszki (1961) has shown in the case of $SO_0(1,2)$ that the tensor product of two representations of the principal series contains representations of the discrete series as direct summands.[3] So the proposed axiom would exclude, for instance, a theory in which the states of one stable particle form an irreducible representation[4] and n-particle states are tensor products of these (second quantization). This is not a conclusive argument, of course, since we have other reasons, explained in later chapters, for being suspicious of this approach of covariant second quantization. However, the complete exclusion of tensor products as subrepresentations of U makes one suspect that this is the wrong track.

Let us, therefore, abandon elegant group-theoretical conjectures and attack the problem by more direct physical reasoning. Positivity of the energy seems to be related to the behavior of a system (or perhaps part of a system) under

---

[3] Nachtmann (1968b) rediscovered this fact and interpreted it as dynamical instability of a covariant second-quantized theory. See Sec. X.4 below.
[4] Cf. Newton and Wigner (1949).

infinitesimal time translation in the positive direction. The failure of group theory has been traced to the absence of any group elements corresponding to time translations in a spatially global sense. However, we do have a notion of local time translation at each point of a geodesic hypersurface; if we choose the latter as the base of a Gaussian coordinate system of the form (III.1.2-3), local time translation is translation in τ. (The gradient of τ is at each point of the hypersurface a unit normal in a distinguished "positive" direction.)

These considerations suggest as a replacement for the spectral condition the following sequence of conditions:

(1) The dynamics of the system is given (at least in the neighborhood of $\{x \mid \tau = 0\}$) by a system of unitary operators:

$$\phi(\tau_2,\sigma) = U(\tau_2,\tau_1)\ \phi(\tau_1,\sigma)\ U(\tau_2,\tau_1)^{-1}. \qquad (2.1)$$

(2) The propagator U is differentiable, in the sense that

$$H(\tau) = -i \lim_{\tau' \to \tau} \frac{U(\tau',\tau) - 1}{\tau' - \tau} \qquad (2.2)$$

exists, so that

$$\frac{\partial \phi}{\partial \tau}(\tau,\sigma) = i[H(\tau),\ \phi(\tau,\sigma)]. \qquad (2.3)$$

(3) H(0) is a positive operator.

Let us not pass judgment on this suggestion until we have had a chance to compare it with some concrete proposals for the quantization of the "free" scalar field in de Sitter space (see Secs. X.8-10).

3. Asymptotic Completeness.

There remains the axiom of asymptotic completeness. One's first impulse is to formulate some condition on the behavior of states (or, rather, observables, in the Heisenberg picture) at large positive and negative times, where "time" is taken to be a coordinate which runs along the axis of the de Sitter hyperboloid, like τ in the Gaussian system. The idea is that in these limits two particles should separate to large distances in some sense (see the classical trajectories in Fig. 10). However, this approach seems inappropriate for several reasons:

(1) The notion of time is highly ambiguous in de Sitter space. As we have seen in Secs. III.2-3, the physically most reasonable definition of time for a given observer assigns infinite time to a surface in de Sitter space (the historical part of his event horizon) which is located at finite time relative to other observers. The physical significance in the large of the time coordinate τ is probably not as great as has sometimes been assumed.

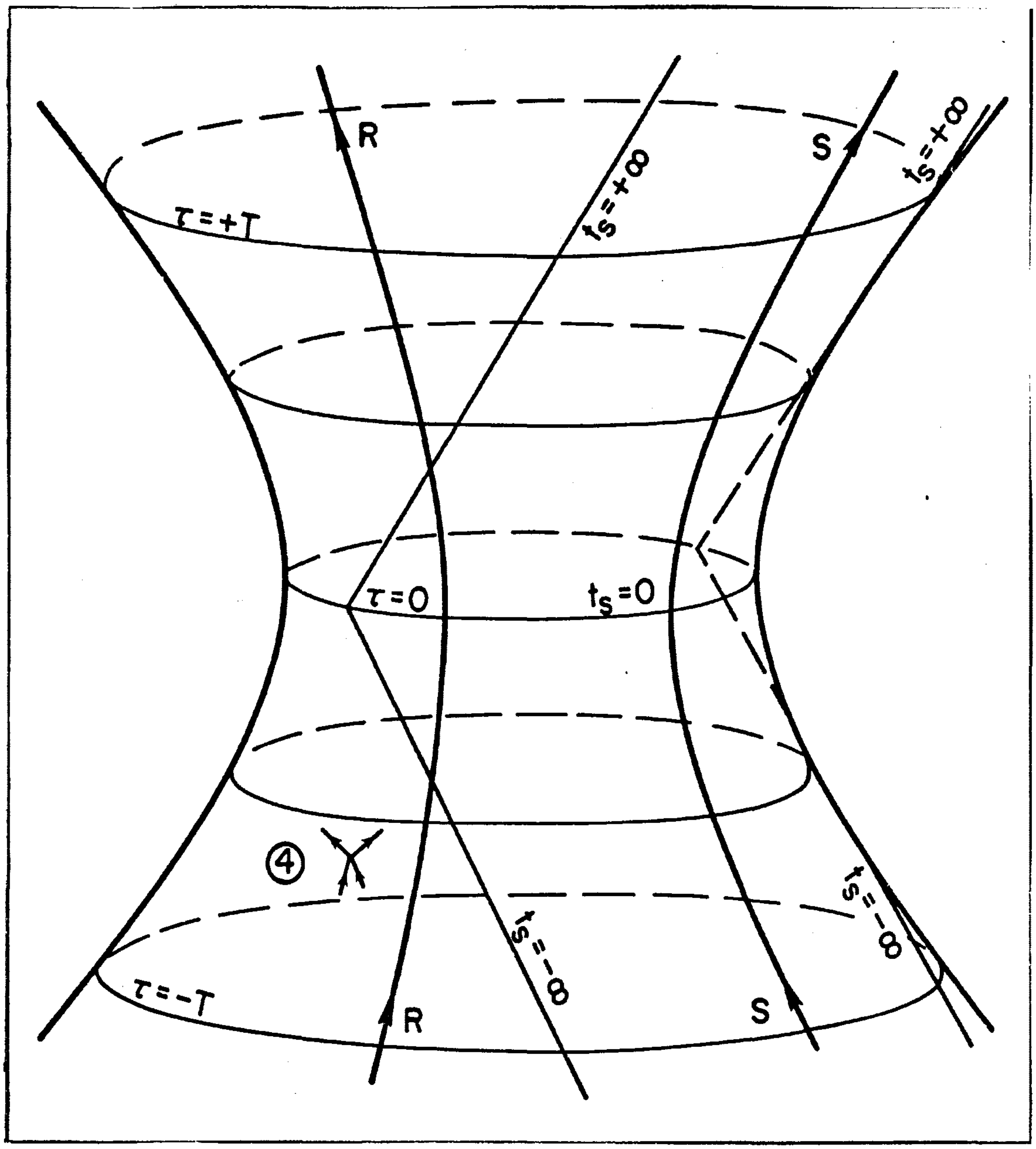


Fig. 10

Scattering in de Sitter space. Shown are (1) a universal time coordinate τ, disparaged in the text; (2) trajectories of two particles, R and S; (3) the curves $t_S = \pm\infty$, where t_S is the proper time of an observer travelling with particle S; (4) an elementary particle scattering event occurring near the trajectory of R.

(2) As remarked at the end of Sec. IV.1, we may have to deal not with a finite number of particles which are stable except insofar as they interact with each other, but with an indefinite, perhaps infinite, perhaps constantly changing, number of particles.

(3) Most importantly, quantum scattering theory deals with phenomena which take place in <u>small</u> regions of space-time. One expects nearly free particle behavior at particle separations which are large compared to the range of the interaction between the particles, but are nevertheless finite -- and, in fact, small -- compared to the scale of the curvature of the universe (see Fig. 10). In flat-space theory an extrapolation to infinite time is made as a mathematical convenience; one could consider only finite times at the cost of dealing with only "approximately free" particle behavior. In de Sitter space the convenience of an extrapolation to infinite time is lost, because the curvature of the universe in the large interferes.

Scattering theory in de Sitter space should be based on an analysis of particle observables at finite separations. The formulas obtained should reduce in the limit of small curvature to the results of asymptotically complete Minkowski-space theories. This task is beyond the scope of the present work, so we shall not consider asymptotic completeness further.

4. Summary and Preview.

In adapting the axioms of field theory to de Sitter space there is no reason to change the principle of general quantum theory. The axioms describing the fields (including locality), the representation of the geometrical symmetry group, and the vacuum have natural analogues; however, those which belong to the "group" side of the picture (see Appendix E) have been asserted with some hesitation. The spectral condition is highly problematical. Asymptotic completeness has been dismissed with some qualitative remarks.

We cannot hope to formulate de-Sitter-space quantum field theory entirely in the abstract. Conjectured general principles must be tested against particular models. One hopes that the study of models will shed light on (1) what, if anything, can be substituted for the spectral condition of Poincaré-invariant theory; (2) how much confidence can be placed in Axioms 2, 6, and 8 (the "group" axioms).

Most of the rest of the dissertation is devoted to the problem of constructing in de Sitter space an analogue of the free scalar field in flat space. In Chapter V we shall study the c-number solutions of the generalized Klein-Gordon equation and consider the possibility of constructing a field theory by second quantization of a single-particle theory. Doubts will arise (related to the old problem of the spectral condition) as to whether this theory is on as sound a physical basis as the ordinary theory for flat space which suggested it. We shall then try a different approach, canonical quantization of the classical field equation. This turns out to be applicable to an arbitrary Riemannian space-time (Chapters VII-X). Evidence will be presented that a "physical" representation of the field operators is not uniquely determined, and that the particle interpretations naively associated with particular representations are of only rather limited physical significance.

In an attempt to suggest a "best" or "most physically significant" representation we arrive at one which for the case of two-dimensional de Sitter space can be shown (Sec. X.9) to satisfy a spectral condition in the sense of Sec. IV.2 (and to possess a reasonable particle interpretation closely related thereto) and also to satisfy the axioms of Sec. IV.1. However, it will do the latter only by accident, as it were: the vacuum state in the group-theoretical sense will have nothing to do with the absence of particles. Moreover, it is doubtful that the theorem that the spectral condition and the group axioms can be satisfied simultaneously can be extended to the four-dimensional de Sitter space. If it should happen that in the four-dimensional case there is one ("covariant") representation which satisfies the axioms of Sec. IV.1 and a different ("positive-frequency") representation which satisfies some kind of spectral condition and fits well into a general theory of field quantization in curved space, then that would be an example of a situation where different representations are useful for different purposes.[5] The first theory would probably be of greater physical interest from the standpoint of general relativity and cosmology, but the second would be the one to use in the program of constructive field theory outlined in the Introduction and Secs. I.2 and VI.3.

---

[5] See Sec. IX.4.

Chapter V

THE FIELD EQUATION IN DE SITTER SPACE AND SECOND QUANTIZATION

We begin the attempt to construct a "free" field in de Sitter space by postulating a generalization of the Klein-Gordon equation. This equation is easily solved in the coordinate systems introduced in Chapter III, and we study the solutions in some detail in the two-dimensional case.

When we try to mimic the familiar construction of the free field in flat space by second quantization, we find ourselves at a loss for an analogue of the notion of "positive-energy solution". The ansatz which seems most reasonable physically turns out to be inconsistent with the idea of a particle as an elementary system whose states transform according to an irreducible representation of the de Sitter group. Further development of this theory is postponed to Chapter X, in the context of a general theory of canonical quantization of fields in Robertson-Walker metrics. Likewise, the quantization suggested by the Fermi coordinate system is absorbed into the theory of quantization in static metrics in Chapter VIII. A theory proposed by Tagirov _et al._ (1967), which maintains the idea of a particle as an elementary system, is discussed from the point of view of Chapter IV.

In the final section the solutions of the field equation in the open de Sitter space are studied briefly. This problem provides an example of subtleties in the physical significance of essential self-adjointness of operators in quantum theory.

1. Differential Form of the Generators and the Casimir Operator.

In Sec. A.3 the basis elements of the Lie algebra of $SO_0(1,n)$ are realized as differential operators in the imbedding space $\mathbb{R}^{n+1}$. These are vector fields which are tangent to the de Sitter hyperboloid (I.1.1). Consequently, they can be expressed as differential operators on de Sitter space. Using Eqs. (III.1.3), they can be expressed in terms of the Gaussian coordinates:

$$P = -i\frac{\partial}{\partial\sigma}, \qquad (1.1a)$$

$$H = i[\cos\sigma\frac{\partial}{\partial\tau} - \tanh\tau\sin\sigma\frac{\partial}{\partial\sigma}], \qquad (1.1b)$$

$$K = i[\sin\sigma\frac{\partial}{\partial\tau} + \tanh\tau\cos\sigma\frac{\partial}{\partial\sigma}]. \qquad (1.1c)$$

(We consider only the two-dimensional space for simplicity.) Similarly, using Eqs. (III.2.2), we find for the Fermi coordinate system

$$P = -i[\cosh\chi \frac{\partial}{\partial\rho} + \tan\rho \sinh\chi \frac{\partial}{\partial\chi}], \tag{1.2a}$$

$$H = i\frac{\partial}{\partial\chi}, \tag{1.2b}$$

$$K = i[\sinh\chi \frac{\partial}{\partial\rho} + \tan\rho \cosh\chi \frac{\partial}{\partial\chi}]. \tag{1.2c}$$

Next we calculate the Casimir operator

$$Q = K^2 + H^2 - P^2. \tag{1.3}$$

In the Gaussian system we obtain

$$Q = -\frac{\partial^2}{\partial\tau^2} - \tanh\tau \frac{\partial}{\partial\tau} + \operatorname{sech}^2\tau \frac{\partial^2}{\partial\sigma^2}, \tag{1.4}$$

and in the Fermi system

$$Q = -\sec^2\rho \frac{\partial^2}{\partial\chi^2} - \tan\rho \frac{\partial}{\partial\rho} + \frac{\partial^2}{\partial\rho^2}. \tag{1.5}$$

From either of these equations, generalized to radius R (see end of Sec. III.1), one can verify that

$$Q = -R^2 \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^\mu}[g^{\mu\nu} \sqrt{|g|} \frac{\partial}{\partial x^\nu}] \equiv -R^2 \Box_c \tag{1.6}$$

($g = \det\{g_{\mu\nu}\}$). By $\Box_c$ we denote the Laplace-Beltrami operator or covariant d'Alembertian ([Adler-Bazin-Schiffer], pp. 71-76).

Being generally covariant, this expression for Q is valid in any coordinate system.

In dimension n the relation (1.6) remains valid, and the higher-order Casimir operators are not independent. For instance, Q_2 of Eq. (I.4.8) is identically zero. So to obtain group representations with "spin", one must consider tensor or spinor functions on space-time, as in the case of the Poincaré group.

2. The Wave Equation.

The Klein-Gordon equation in flat space has the following group-theoretical interpretation:[1] A particle (stable and without internal structure) is an "elementary system" whose possible quantum states support an irreducible unitary representation of the Poincaré group. When the group and its Lie algebra are realized as operators on the space of functions on space-time, the condition that the Casimir operator have a definite value is precisely the Klein-Gordon equation (cf. Eqs. (I.4.4) and (A.3.6a,b)). If a further condition of positive energy is imposed, the solutions form an irreducible representation.

It is natural to attempt to construct the same type of theory for de Sitter space. Accordingly, on the basis of Eq.

[1] See Wigner (1939, 1948, 1956); Bargmann and Wigner (1948); Newton and Wigner (1949); [Streater-Wightman], Chap. 1. Replacing scalar functions by tensor- or spinor-valued functions, one obtains equations for particles of nonzero spin.

(1.6), we postulate the wave equation

$$\frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x^{\mu}}\left[\sqrt{|g|}\,g^{\mu\nu}\frac{\partial}{\partial x^{\nu}}\right]\Psi(x) = -R^{-2}\,q\,\Psi(x), \qquad (2.1)$$

where q is a constant which plays a role like that of the square of the mass in the flat-space theory. (From now on we shall set R = 1 except when contraction (the large-R limit) is being discussed.) This equation is identical to the one which is obtained (without use of group theory) by variation of the simplest Lagrangian density for a scalar field in an arbitrary Riemannian space-time (Sec. VII.1 below).

The only other scalar wave equation which seems to have been seriously considered[2] is one with an extra term

$$\frac{n-2}{4(n-1)}\,R^{\mu}_{\ \mu}\,\Psi = \frac{n(n-2)}{4R^{2}}\,\Psi, \qquad (2.2)$$

where n is the dimension of space-time, R^{μ}_{μ} is the scalar curvature (contracted Ricci tensor), and R is the radius of the de Sitter space. Since in a space of constant curvature this term is a constant, it may be regarded as just a redefinition of the mass. Moreover, it vanishes in the limit of large R. In the context of de Sitter space, therefore, the controversy over whether to include the term (2.2) seems pointless.

In the geodesic Gaussian coordinates (n = 2) the wave

---

[2] See references in Sec. VII.1.

equation has the form (see Eq. (1.4))

$$\frac{\partial^2 \Psi}{\partial \sigma^2} = \left(\cosh \tau \frac{\partial}{\partial \tau}\right)^2 \Psi + q \cosh^2 \tau \, \Psi. \qquad (2.3)$$

An interesting transformation of the equation involves the Gudermannian function ([Gradshteyn-Ryzhik], pp. 43-44), which has the properties

$$\mathrm{gd}\, x \equiv \int_0^x \frac{dt}{\cosh t} = 2 \tan^{-1} e^x - \frac{\pi}{2}, \qquad (2.4a)$$

$$\mathrm{gd}^{-1} x = \int_0^x \frac{dt}{\cos t} = \log \tan \left[\frac{x}{2} + \frac{\pi}{4}\right]. \qquad (2.4b)$$

If we let

$$\alpha = \mathrm{gd}\, \tau \qquad \left(-\frac{\pi}{2} < \alpha < \frac{\pi}{2}\right), \qquad (2.5)$$

then

$$\frac{\partial}{\partial \tau} = \frac{1}{\cosh \tau} \frac{\partial}{\partial \alpha}, \qquad \cosh \tau = \sec \alpha. \qquad (2.6)$$

So the equation becomes

$$\frac{\partial^2 \Psi}{\partial \alpha^2} - \frac{\partial^2 \Psi}{\partial \sigma^2} + q\,(1 + \tan^2 \alpha)\, \Psi = 0. \qquad (2.7)$$

Near 0 the deviation from the ordinary Klein-Gordon equation is quadratic in 1/R if we take $q = R^2 m^2$.

Similarly, from Eq. (1.5) we obtain

$$\frac{\partial^2 \Psi}{\partial \chi^2} = \left(\cos\rho \frac{\partial}{\partial\rho}\right)^2 \Psi - q \cos^2 \rho \, \Psi \tag{2.8}$$

in the static Fermi system. Setting

$$\beta = \mathrm{gd}^{-1} \rho \qquad \left(-\infty < \beta < \infty \text{ for } -\frac{\pi}{2} < \rho < \frac{\pi}{2}\right), \tag{2.9}$$

we have

$$\frac{\partial^2 \Psi}{\partial \chi^2} - \frac{\partial^2 \Psi}{\partial \beta^2} + q\,(1 - \tanh^2 \beta)\, \Psi = 0. \tag{2.10}$$

The possibility of reducing the wave equation to the form of a flat-space Klein-Gordon equation with an external scalar potential term (in other words, a space-time-dependent mass) is peculiar to dimension 2 (see Sec. VII.1).

To solve Eq. (2.3) (Philips (1963)) set

$$z = i \sinh \tau = \sin i\tau, \tag{2.11}$$

$$\Psi(\tau,\sigma) = \sum_{p=-\infty}^{\infty} c(p)\, f_p(z)\, e^{ip\sigma}. \tag{2.12}$$

(Since Ψ is periodic in σ, p takes only integral values.) Then

$$(1 - z^2)\frac{d^2 f_p}{dz^2} - 2z\frac{df_p}{dz} - \frac{p^2}{1 - z^2} f_p = q f_p . \tag{2.13}$$

This is the associated Legendre equation ([N.B.S.], p. 332) with

$$\mu = p \quad \text{and} \quad \nu(\nu + 1) = - q. \tag{2.14}$$

Thus the solutions are the associated Legendre functions $P_\nu^p(z)$ and $Q_\nu^p(z)$ with

$$\nu = -\frac{1}{2} + i\sqrt{q - \frac{1}{4}}. \tag{2.15}$$

The existence of an imaginary part of ν distinguishes the representations of the principal series ($q > 1/4$) from the complementary and discrete series (see Sec. B.3).

Eq. (2.8) can be handled similarly. Let $u = \sin\rho$ and try a solution of the form

$$\Psi_\lambda(\chi,\rho) = f_\lambda(u)\, e^{-i\lambda\chi}. \tag{2.16}$$

Then

$$(1 - u^2)\frac{d^2 f_\lambda}{du^2} - 2u\frac{df_\lambda}{du} - q f_\lambda + \frac{\lambda^2}{1 - u^2} f_\lambda = 0. \tag{2.17}$$

This is Legendre's equation again, with

$$\mu = i\lambda, \qquad \nu = -\frac{1}{2} + i\sqrt{q - \frac{1}{4}}. \tag{2.18}$$

The solutions are $P_{\nu}^{i\lambda}(u)$ and $Q_{\nu}^{i\lambda}(u)$, with $-1 \leq u \leq 1$. It remains to be decided which values of λ occur; this will be discussed in Sec. V.7.

3. Positive-Frequency Solutions in the Geodesic Gaussian System.

In this section we consider the case $q > 1/4$ and attempt to interpret the solutions of the wave equation in terms of particles in the closed de Sitter universe.

It will be convenient for us to use as the basic pair of linearly independent solutions of Eq. (2.13) not the associated Legendre functions but

$$E_p(z) = (1 - z^2)^{-p/2}\, F[-\frac{1}{2}(\nu+p), \frac{1}{2}(1+\nu-p), \frac{1}{2}, z^2], \tag{3.1a}$$

$$O_p(z) = (1 - z^2)^{-p/2}\, z\, F[\frac{1}{2}(1-\nu-p), 1 + \frac{1}{2}(\nu-p), \frac{3}{2}, z^2], \tag{3.1b}$$

where F is the hypergeometric function ([N.B.S.], Chap. 15). The branch of $(1 - z^2)^{-p/2}$ should be chosen continuously along the imaginary axis with the value + 1 at $z = 0$. E_p and O_p are the elementary even and odd solutions of Eq. (2.13):

$$E_p(0) = 1, \qquad E_p'(0) = 0, \tag{3.2a}$$

$$O_p(0) = 0, \qquad O_p'(0) = 1. \tag{3.2b}$$

They are independent of the sign of p. Formulas relating these functions to Legendre functions with the standard branch cut and phase conventions appear in [N.B.S.], pp. 332-333 (Eqs. (8.1.4,7)).

Let

$$P_p(\tau) = E_p(i \sinh \tau) - \sqrt{q + p^2}\, O_p(i \sinh \tau), \tag{3.3a}$$

$$N_p(\tau) = E_p(i \sinh \tau) + \sqrt{q + p^2}\, O_p(i \sinh \tau) = P_p^*(\tau). \tag{3.3b}$$

The series expansion of P_ρ near $\tau = 0$ is

$$P_p(\tau) = 1 - i\sqrt{q + p^2}\,\tau - \frac{1}{2}(q + p^2)\,\tau^2$$

$$+ \frac{i}{6}\sqrt{q + p^2}\,(q + p^2 + 1)\,\tau^3 + O(\tau^4). \tag{3.4}$$

Thus

$$e^{ip\sigma} P_p(\tau) \tag{3.5a}$$

and

$$e^{ip\sigma} N_p(\tau) \qquad (3.5b)$$

agree through order τ^2 with the plane-wave solutions of the Klein-Gordon equation:

$$\exp[ip\sigma \mp i\sqrt{m^2 + p^2}\, \tau] = e^{ip\sigma}[1 \mp i\sqrt{m^2 + p^2}\, \tau$$

$$- \frac{1}{2}(m^2 + p^2)\tau^2 \pm \frac{i}{6}(m^2 + p^2)^{3/2} \tau^3 + O(\tau^4)]. \qquad (3.6)$$

Accordingly, we shall call the function (3.5a) a <u>positive-frequency solution</u>, and (3.5b) a <u>negative-frequency solution</u>. In Sec. X.5 it will be shown that field quantization based on this identification leads to a (time-dependent) Hamiltonian which is a positive operator at $\tau = 0$ (cf. Sec. IV.2). For now we simply gamble, on the basis of the comparison with Eq. (3.6), that Eq. (3.4) demonstrates "positive-energy behavior" in some useful sense.

The most obvious next step would be to interpret the solution (3.5a) as the wave function of a particle with momentum p (cf. Sec. VIII.4 below). We would need a scalar product on the space of all these solutions; let us assume that this has been found (see Sec. V.5). Then we would have a "relativistic wave mechanics" analogous to that based on the ordinary Klein-Gordon equation (as described, e.g., by [Schweber], pp. 54-64, and [Corinaldesi], pp. 25-110). A many-particle theory could be constructed by second quantization ([Schweber], pp. 156-195);

this would be equivalent to the field theory defined by writing[3]

$$\phi(\tau,\sigma) = \sum_{p=-\infty}^{\infty} [2\sqrt{q + p^2}]^{-1/2} \{e^{ip\sigma} P_p(\tau) a_p + e^{-ip\sigma} N_p(\tau) a_p^\dagger\} \quad (3.7)$$

and interpreting a_p and $a_p^\dagger$ as annihilation and creation operators for particles in the state with wave function (3.5a). Unfortunately, this single-particle interpretation seems to be untenable, for reasons which we shall discuss in Sec. V.5.

4. The Representation of the Group in the Space of Solutions.

There is a natural action of the de Sitter group on the solutions of the wave equation: if $\Psi(x)$ is a solution, then so is $U(\Lambda)\Psi(x) = \Psi(\Lambda^{-1}x)$. The expression

$$W(\Psi_1,\Psi_2) = i\int_{-\pi}^{\pi} d\sigma \, [\Psi_1^*(0,\sigma) \frac{\partial\Psi_2}{\partial\tau}(0,\sigma) - \frac{\partial\Psi_1^*}{\partial\tau}(0,\sigma) \Psi_2(0,\sigma)] \quad (4.1)$$

defines an indefinite Hermitian form on solutions whose initial values are sufficiently integrable. This form is invariant under the action of the group:

$$W(U(\Lambda)\Psi_1, U(\Lambda)\Psi_2) = W(\Psi_1,\Psi_2). \quad (4.2)$$

<u>Proof</u>: This will follow from the general theory of

[3] The normalization factor in Eq. (3.7) is determined by the condition that the a and $a^\dagger$ satisfy the correct commutation relations for annihilation and creation operators -- see Sec. X.1.

Secs. VII.3,5. W is the current form,

$$W(\psi_1,\psi_2) = i\int_S d\sigma_\nu \sqrt{|g|}\, g^{\nu\mu} \left[\psi_1^*(x)\, \partial_\mu \psi_2(x) - \partial_\mu \psi_1^*(x)\, \psi_2(x)\right]$$

$$= i\int_{\tau=\text{const.}} d\sigma \quad \cosh \tau \left[\psi_1^*(\tau,\sigma)\frac{\partial\psi_2}{\partial\tau}(\tau,\sigma) - \frac{\partial\psi_1^*}{\partial\tau}(\tau,\sigma)\,\psi_2(\tau,\sigma)\right], \qquad (4.3)$$

evaluated on the hypersurface $\tau = 0$. Since the current form is covariant, the integral has the same value in every Gaussian frame; this is the "passive" interpretation of Eq. (4.2).

Alternate proof: Invariance under σ-translations is obvious. If the functions considered are differentiable, it suffices to consider infinitesimal transformations generated by H and K. One must verify that H and K (the operators of Eqs. (1.1b,c)) are Hermitian with respect to W, or that the ladder operators $A^{\pm} = H \pm iK$ are Hermitian conjugates of each other $(W(A^+\psi_1,\psi_2) = W(\psi_1,A^-\psi_2))$. This is a straightforward calculation, which we omit.

We shall use the form W to decompose the representation of $SO_0(1,2)$ in the space of solutions of Eq. (2.1) into two irreducible unitary representations. A different approach (decomposition of the quasiregular representation) will be considered in Sec. VI.1 and related to this one in Sec. X.4.

We begin by noting that for solutions of the elementary form

$$\Psi_p(\tau,\sigma) = \varphi\, e^{ip\sigma}[A(p)E_p(z) - B(p)O_p(z)] \qquad (\varphi = e/\sqrt{2\pi}) \qquad (4.4)$$

we have (from Eqs. (3.2))

$$W(\Psi_{p_1},\Psi_{p_2}) = \delta_{p_1 p_2}[A^*(p_1)B_2(p_1) + B^*(p_1)A_2(p_1)], \qquad (4.5a)$$

$$W(\Psi_p,\Psi_p) = 2\,\mathrm{Re}(A^*(p)B(p)). \qquad (4.5b)$$

Let us choose a particular solution Ψ_0 of the form (4.4) with $p = 0$ such that $W(\Psi_0,\Psi_0) = 1$. Then by operating repeatedly with $A^{\pm}$ according to Eq. (B.3.2b),

$$\Psi_{p\pm 1} = [q + p(p\pm 1)]^{-1/2} A^{\pm}\Psi_p, \qquad (4.6)$$

we obtain a set of vectors $\{\Psi_p\}$ ($p = 0, \pm 1, \pm 2, \ldots$), of the form (4.4), such that

$$W(\Psi_p,\Psi_{p'}) = \delta_{pp'}.$$

Thus these vectors form an orthonormal basis for a Hilbert space $\mathcal{H}_+$ with scalar product $W(\Psi_1,\Psi_2)$, and the representation of the de Sitter group in this space is the irreducible unitary representation with $Q = q$ (Eq. (B.3.5a)).

<u>Proof</u>: In terms of z the ladder operators are

$$A^{\pm} = H \pm iK = -e^{\pm i\sigma}\{(1-z^2)^{1/2}\frac{\partial}{\partial z} \mp iz(1-z^2)^{-1/2}\frac{\partial}{\partial \sigma}\}. \quad (4.7)$$

The derivatives of the hypergeometric functions may be calculated from formulas (15.2.1,4,6,9) in [N.B.S.], p. 557:

$$(1-z^2)^{1/2}\frac{dE_p}{dz}(z) = [q + p(p+1)]\,O_{p+1}(z) - pz(1-z^2)^{-1/2}E_p(z)$$

$$= [q + p(p-1)]\,O_{p-1}(z) + pz(1-z^2)^{-1/2}E_p(z), \quad (4.8a)$$

$$(1-z^2)^{1/2}\frac{dO_p}{dz}(z) = E_{p+1}(z) - pz(1-z^2)^{-1/2}O_p(z)$$

$$= E_{p-1}(z) + pz(1-z^2)^{-1/2}O_p(z), \quad (4.8b)$$

Hence one obtains

$$A^{\pm}[e^{ip\sigma}E_p(z)] = -[q + p(p\pm 1)]\,e^{i(p\pm 1)\sigma}O_{p\pm 1}(z), \quad (4.9a)$$

$$A^{\pm}[e^{ip\sigma}O_p(z)] = -e^{i(p\pm 1)\sigma}E_{p\pm 1}(z). \quad (4.9b)$$

It follows that

$$W(A^{\pm}\Psi_p, A^{\pm}\Psi_{p'}) = \delta_{pp'}[q + p(p\pm 1)]\,W(\Psi_p, \Psi_p),$$

which demonstrates the orthonormality of the vectors (4.6). The representation of the Lie algebra in this space is obviously that given by Eqs. (B.3.2).

It can easily be seen[4] from Eqs. (4.5) that the complex conjugates of the functions Ψ_p span a Hilbert space $\mathcal{H}_-$ with scalar product $-W(\Psi_1,\Psi_2)$, and that

$$W(\Psi_1,\Psi_1) > 0, \qquad W(\Psi_2,\Psi_2) < 0,$$

$$W(\Psi_1,\Psi_2) = 0 \quad \text{if} \quad \Psi_1 \in \mathcal{H}_+, \ \Psi_2 \in \mathcal{H}_- . \tag{4.10}$$

Finally, $\mathcal{H}_+$ and $\mathcal{H}_-$ exhaust the solutions of the wave equation, in the sense that

$$\Psi(\tau,\sigma) = \sum_p s(p)\Psi_p(\tau,\sigma) + \sum_p t(p)\Psi^*_p(\tau,\sigma)$$

can be solved (by Fourier transformation) for s(p) and t(p) in terms of the Cauchy data $\Psi(0,\sigma)$ and $\partial\Psi/\partial\tau(0,\sigma)$, if the latter are "sufficiently integrable" but otherwise arbitrary. (The vagueness in the integrability condition is discussed in Sec. VII.5.)

The group representations in $\mathcal{H}_+$ and $\mathcal{H}_-$ are <u>equivalent</u>, since we know from Sec. B.3 that there is only one irreducible unitary representation for each value of $q > 1/4$. This is what allowed the initial choice of Ψ_0 to be quite arbitrary. This situation is quite different from the case of

---

[4] Note that (1) $E_p(z)$ is real, $O_p(z)$ is imaginary; (2) $A(-p) = A(p)$, $B(-p) = B(p)$.

the Poincaré group. The solutions of the Klein-Gordon equation in Minkowski space can be separated uniquely into positive-energy and negative-energy functions, which support inequivalent representations of $ISO_0(1,1)$ (cf. Sec. B.2). It follows that from any function which does not lie wholly in one of these subspaces (but is a linear combination) the operators of the Poincaré group generate the entire space of solutions -- the representation space of a reducible representation.

5. The Incompatibility of Positive Frequency and Group Invariance.

In accordance with the remarks at the beginning of Sec. V.2, we would expect, if the notion of a stable spinless quantum-mechanical particle in de Sitter space makes sense at all, that the Hilbert space of states of such a particle would support an irreducible unitary representation of the de Sitter group. This is not the case for the space of positive-frequency solutions defined in Sec. V.3. The positive-frequency solutions of lowest momentum, normalized in the sense of Eq. (4.5b), are

$$(2\sqrt{q})^{-1/2} P_0(\tau) = \frac{1}{\sqrt{2}}[q^{-1/4} E_0(z) - q^{+1/4} O_0(z)] \qquad (5.1)$$

and

$$e^{i\sigma}(2\sqrt{q+1})^{-1/2} P_1(\tau)$$

$$= e^{i\sigma}\frac{1}{\sqrt{2}}[(q+1)^{-1/4} E_1(z) - (q+1)^{+1/4} O_1(z)]. \quad (5.2)$$

But the basis vector with p = 1 in the irreducible representation generated by the vector (5.1) is, by Eqs. (4.6,9),

$$e^{i\sigma}\frac{1}{\sqrt{2}}[(q+2)^{-1/2} q^{+1/4} E_1(z) - (q+2)^{+1/2} q^{-1/4} O_1(z)], \quad (5.3)$$

which is not the same as the function (5.2). _The space of positive-frequency solutions is not invariant under the action of the group._

One possible response to this realization is that our identification of the positive-frequency solutions is wrong: one ought to choose a space of single-particle wave functions which is invariant under the group. This is the approach which has been followed in most previous work on quantum field theory in de Sitter space; it will be reviewed in Sec. V.6.

Another possibility, however, is that the assumption of stable particles is wrong. The curvature of space is equivalent to a time-dependent gravitational field. Now it is well known in the context of external electromagnetic fields that the interaction of a quantized field with a time-dependent external field can produce pairs of particles. (For examples of several different approaches to this problem see Capri (1969), Brezin and

Itzykson (1970), and Narozhny and Nikishov (1970). Moore (1970) studies photon creation due to interaction with a moving reflecting wall.) It is quite reasonable that the same effect should occur in the gravitational case. If so, then a state which describes a universe which contains exactly one particle at time $\tau = 0$ will not necessarily have such a characterization at some other time (more generally, on a different spacelike hypersurface). The same is true for the state in which no particles are present in space at a given time. In particular, if we consider the spaces of constant time in the Fermi picture (see Secs. III.2-3), it follows that the space of one-particle states (and the no-particle state) at $\tau = \chi = 0$ should not be expected to be invariant under the de Sitter group.

However, once we abandon group theory as our guide in the construction of a field theory, the de Sitter universe loses much of its privileged position. We might as well consider any Riemannian space-time, or at least any for which the wave equation can be solved, as in Sec. V.2, by separation of variables. This is what we shall do in Chapter X.

In such a theory the expansion (3.7) of the field will still make sense, but the coefficients $a_{\rho}^{(+)}$ will not be interpreted in terms of stable particles. It is conceivable that the field theory as a whole will be invariant under the de Sitter group in the sense of Sec. IV.1; but the no-particle state at any particular time χ will not be an invariant vacuum satisfying Axiom 6. This conjecture will be verified in Sec. X.9.

Finally, observe that, although the space of positive-frequency solutions (3.5a) does not support a representation of the group, it is a Hilbert space with respect to the form W (whose existence is independent of the symmetry group). If we write an arbitrary (sufficiently integrable -- see Sec. VII.5) solution as

$$\Psi(\tau,\sigma) = \sum_{p=-\infty}^{\infty} e^{ip\sigma} (2\sqrt{q + p^2})^{-1/2} \{\Psi_+(p) P_p(\tau) + \Psi_-(p) N_p(\tau)\}, \quad (5.4)$$

we have

$$W(\Psi,\Psi) = \sum_{p=-\infty}^{\infty} |\Psi_+(p)|^2 - \sum_{p=-\infty}^{\infty} |\Psi_-(p)|^2 . \quad (5.5)$$

Thus the solution space is the W-orthogonal direct sum of a positive-frequency and a negative-frequency Hilbert space with oppositely-signed scalar products (cf. Eqs. (4.10)); the negative-frequency functions are the complex conjugates of the positive-frequency ones. Such a decomposition gives rise to a formally consistent canonical field quantization when the $\Psi_\pm(p)$ are replaced by annihilation and creation operators (see Secs. X.1 and X.5).

It may appear that the inference from loss of invariance under the geometrical symmetry group to particle creation has been drawn rather quickly. Might it not still be possible to interpret the normalizable positive-frequency solutions as the Hilbert space of states of a single particle? If we are willing to contemplate a field theory which *may* turn out not to be invariant (see above), why not a particle theory which is not invariant? The argument in favor of the interpretation offered here can best be presented in terms of an analogy. Consider the Klein-Gordon equation with an external scalar potential (which depends only on time, for simplicity):

$$\Box\Psi(t,x) + V(t)\Psi(t,x) + m^2\Psi(t,x) = 0. \tag{5.6}$$

It has solutions analogous to the functions (3.5a):

$$\Psi(t,x) = e^{ikx} P_k(t), \tag{5.7}$$

where

$$P_k(t) \sim \begin{cases} e^{-i\omega_k t} & \text{as } t \longrightarrow -\infty \qquad (\omega_k = \sqrt{k^2 + m^2}), \\ \alpha^*(k)e^{-i\omega_k t} + \beta^*(k)e^{+i\omega_k t} & \text{as } t \longrightarrow +\infty, \end{cases} \tag{5.8}$$

and $|\alpha(k)|^2 - |\beta(k)|^2 = 1$. It would be folly to insist that Eq. (5.7) gives the wave function of a stable particle, because of its positive-frequency behavior in the past, ignoring its mixed behavior in the future. The physically sensible interpretation of Eq. (5.6) is as the equation of a quantum field. Then the behavior (5.8) indicates that particles are created by the action of the external potential. (This will be explained in detail in a gravitational context in Sec. X.3.) Now it seems most probable that the phenomenon demonstrated in Eqs. (5.1-3) should be interpreted in the same way. The role of a large negative and a large positive time in the example will be taken in de Sitter space by two geodesic hypersurfaces (related by an isometry in the group). Of course, our criterion of positive frequency remains to be justified; all these remarks are relative to that assumption.

6. Quantization Leading to an Invariant One-Particle Space.

Tagirov, Fedyun'kin, and Chernikov (1967) have proposed a quantum theory of a scalar field in two-dimensional de Sitter space in which there are a vacuum state and a space of one-particle states which are both invariant under the de Sitter group.[5] Nachtmann (1968b) independently arrived at the same theory by a different route, emphasizing the representation theory of $SO_0(1,2)$ (see Sec. X.4). Here we shall follow the Dubna group, since their approach is more physical and more easily related to that of Secs. V.3-5 and Chapters VII and X of this dissertation. It is necessary for the calculations in Secs. X.9-10 to express their solutions in terms of our basis solutions (3.1); this requires several special-function identities. (The result is Eq. (6.6) below.)

[5] The four-dimensional case was treated by Chernikov and Tagirov (1968). Bronnikov and Tagirov (1968) treated more general spaces by similar methods.

It will then be verified that this theory fulfills the axioms proposed in Sec. IV.1. To establish uniqueness of the vacuum requires proving a simple but not trivial theorem about the structure of tensor products of representations of $SO_o(1,2)$, which may be of interest in itself.

Tagirov *et al*. expand the scalar field obeying Eq. (2.7) as follows:[6]

$$\phi(\alpha,\sigma) = \sum_{p=-\infty}^{\infty} [e^{ip\sigma} T_p(\alpha)\, b_p + e^{-ip\sigma} T_p^*(\alpha)\, b_p^\dagger], \qquad (6.1)$$

$$T_p(\alpha) = \frac{\sqrt{\Gamma(\nu+1+|p|)\,\Gamma(-\nu+|p|)}}{\sqrt{2}\,|p|!}\, e^{-i|p|\alpha}$$

$$\times\, F[\nu+1, -\nu, |p|+1, \frac{1 - i\tan\alpha}{2}]$$

$$= \frac{1}{\sqrt{2}}\sqrt{\Gamma(\nu+1+|p|)\,\Gamma(-\nu+|p|)}\; P_\nu^{-|p|}(z). \qquad (6.2)$$

(The last equality follows from Eqs. (2.5,11) and formula (8.1.2) of [N.B.S.] (p. 332) -- P is a Legendre function.) They interpret the time-independent operators b_p and $b_p^\dagger$ as annihilation and creation operators for particles; thus their particles are stable, with respect both to a description in terms of the time coordinate α (or τ) and to a description in terms of

[6] The notation has been modified to conform to ours, and a factor of $1/\sqrt{2}$ has been inserted in Eq. (6.2) to correct the normalization (cf. Nachtmann (1968b), Eq. (2.21)).

a time variable defined by de Sitter isometries (such as χ). Using Eqs. (3.1) and formula (8.1.4) of [N.B.S.] (p. 332), we can write Eq. (6.2) as

$$T_p(\alpha) = 2^{-|p|-1/2}\,\pi^{1/2}\sqrt{\Gamma(\nu+1+|p|)\,\Gamma(-\nu+|p|)}\ \times$$

$$\{[\Gamma(\tfrac{1}{2}(1-\nu+|p|))\,\Gamma(1+\tfrac{1}{2}(\nu+|p|))]^{-1}\,E_p(z)$$

$$-\ 2[\Gamma(\tfrac{1}{2}(1+\nu+|p|))\,\Gamma(\tfrac{1}{2}(-\nu+|p|))]^{-1}\,O_p(z)\}. \qquad (6.3)$$

Since

$$\nu + 1 = -\nu^*, \qquad (6.4)$$

$$\Gamma(2z) = \frac{1}{\sqrt{2\pi}}\,2^{2z-1/2}\,\Gamma(z)\,\Gamma(z+\tfrac{1}{2}) \qquad (6.5)$$

([N.B.S.], Eq. (6.1.18) (p. 256)), and $\Gamma(z^*) = \Gamma(z)^*$, this can be simplified to

$$T_p(\alpha) = \frac{1}{2}\left\{\left|\frac{\Gamma(\frac{1}{2}(-\nu+|p|))}{\Gamma(1+\frac{1}{2}(\nu+|p|))}\right| E_p(z) - 2\left|\frac{\Gamma(1+\frac{1}{2}(\nu+|p|))}{\Gamma(\frac{1}{2}(-\nu+|p|))}\right| O_p(z)\right\}. \qquad (6.6)$$

Then the solutions

$$\Psi_p(\alpha,\sigma) = e^{ip\sigma}\,T_p(\alpha) \qquad (6.7)$$

satisfy Eq. (4.6) and the normalization condition $W(\Psi_p, \Psi_p) = 1$ (see Eqs. (4.4-5)). (To verify Eq. (4.6) use Eqs. (4.9) and

$$q + p(p+1) = |\nu+1+p|^2 , \qquad q + p(p-1) = |\nu+p|^2 , \tag{6.8}$$

along with Eq. (6.4) and the well-known relation $z\Gamma(z) = \Gamma(z+1)$.)

We saw in Sec. V.4 that there are many such invariant spaces of solutions on which W defines a positive definite scalar product; they correspond to different choices of A(0) and B(0) in Eq. (4.4). Since the basis vectors can be normalized and their overall phase is arbitrary, the different possibilities can be labeled by two real parameters; Tagirov et al. determine the basis functions in the general case to be

$$\Psi_p = e^{ip\sigma} \left\{ \frac{1}{\sqrt{1 - |\lambda|^2}} T_p(\alpha) + (-1)^p \frac{\lambda^*}{\sqrt{1 - |\lambda|^2}} T^*_p(\alpha) \right\} , \tag{6.9}$$

where λ is complex and $|\lambda| < 1$. To establish that $\lambda = 0$ is the physically relevant choice, they show that only in this case does θ, the phase of Ψ_p, obey in the limit of large $|p|$ the classical Hamilton-Jacobi equation

$$m^2 = g^{\mu\nu} \partial_\mu \theta \, \partial_\nu \theta . \tag{6.10}$$

A more elementary statement of what distinguishes the solutions (6.7) from the other possibilities is that for large $|p|$ they asymptotically approach positive-frequency solutions in the sense of Sec. V.3:

$$T_p(\alpha) \sim [2\sqrt{q + p^2}]^{-1/2} P_p(gd^{-1}\alpha) \quad (|p| \longrightarrow \infty). \qquad (6.11)$$

This can be verified by using

$$\frac{\Gamma(z + a)}{\Gamma(z)} \sim z^a \quad \text{as } |z| \longrightarrow \infty \qquad (6.12)$$

(from [Gradshteyn-Ryzhik], Eq. (8.328.2) (p. 937)). This is rather reasonable physically; it says that the solutions behave like positive-energy plane waves in flat space provided that the wave length is small compared to a length characteristic of the curvature of space-time.

In accordance with the particle interpretation, the algebra of the creation and annihilation operators is realized in the Fock representation. That is, a vacuum vector $|0\rangle$ is postulated such that

$$b_p|0\rangle = 0 \quad \text{for all } p. \qquad (6.13)$$

Then $b_p^\dagger|0\rangle$ is interpreted as the vector of the state of the field in which one particle is present, in the single-particle state with wave function[7]

[7] Tagirov *et al*. call the complex conjugate of Eq. (6.14) the wave function, but this conflicts with the usual terminology for the Klein-Gordon theory in flat space.

$$\langle 0|\phi(\alpha,\sigma)b_p^\dagger|0\rangle = e^{ip\sigma} T_p(\alpha). \qquad (6.14)$$

Similarly, there are two-particle basis states of the form $b_{\rho_1}^\dagger b_{\rho_2}^\dagger|0\rangle$, and so on. The Hilbert space of the theory (Fock space) is the closure of the span of all the n-particle states, $0 \le n < \infty$.

Let us verify that this theory fulfills the axioms proposed in Sec. IV.1, just as the ordinary free Klein-Gordon field obeys the usual Wightman axioms. We have just constructed the Hilbert space (Axiom 1) with cyclic vector $|0\rangle$ (Axiom 7). The representation of the group in the space of one-particle wave functions determines in an obvious way a unitary representation in the whole Fock space (Axiom 2): the representation in the n-particle subspace is the symmetrized n-fold tensor product of the irreducible one-particle representation, and the vacuum is invariant (Axiom 8). The generators of the total representation are clearly

$$P = \sum_{p=-\infty}^{\infty} p\, b_p^\dagger b_p, \qquad (6.15a)$$

$$A^{\pm} = \sum_p \sqrt{q + p(p\pm 1)}\; b_{p\pm 1}^\dagger b_p. \qquad (6.15b)$$

(The Dubna authors obtain these expressions (modulo sign conventions) by integrating the energy-momentum tensor of the field, contracted with the Killing vectors of the isometries, over a spacelike hypersurface.)

We shall not dwell on the technicalities of Axioms 3 and 4. The field (6.1) makes sense as a distribution on suitably smooth test functions. The finite sums of n-particle states with smooth wave functions form an invariant domain. It is easy to see from the definitions that the group operators transform the field operators in the expected way.

The local commutativity (Axiom 5) of any canonically quantized scalar field will be shown in Sec. VII.4. In the present case Tagirov *et al*. have given the commutator in closed form:

$$[\phi(\alpha_2,\sigma_2),\ \phi(\alpha_1,\sigma_1)] = \frac{i}{2}\ \epsilon(\alpha_1-\alpha_2,\ \sigma_1-\sigma_2)\ P_\nu(\gamma), \qquad (6.16)$$

where P_ν is a Legendre function,

$$\epsilon(\alpha,\sigma) = \begin{cases} \operatorname{sgn}\alpha & \text{if } |\alpha| > |\sigma|, \\ 0 & \text{if } |\alpha| < |\sigma|, \end{cases} \qquad (6.17)$$

and

$$\gamma = \frac{\cos(\sigma_1-\sigma_2) - \sin\alpha_1 \sin\alpha_2}{\cos\alpha_1 \cos\alpha_2} \qquad (6.18)$$

(which is the hyperbolic cosine of the geodesic distance between the two points).

All that remains to be checked is the uniqueness part of Axiom 6. That is, does the representation of $SO_0(1,2)$ in the orthogonal complement of $|0\rangle$, when decomposed into irreducibles, contain the trivial representation as a discrete direct summand? It suffices to ask this question of each of the n-particle representations separately, or, in fact, to answer it negatively for the tensor product

$$D_q \otimes D_q \otimes \ldots \otimes D_q \quad \text{(n factors)} \qquad (6.19)$$

(of which the n-particle space is an invariant subspace). Here D_q denotes the irreducible unitary representation of the principal series with $Q = q$.

We start by proving the following theorem:

Let D_1 and D_2 be any two irreducible unitary representations of $SO_o(1,2)$, not both trivial. Then $D_1 \otimes D_2$ does not contain the trivial representation as a discrete direct summand.

<u>Proof</u> (method suggested by Pukánszky (1961), pp. 132-134): Let the Casimir invariants of D_1 and D_2 be q_1 and q_2, respectively. We must show that no vector

$$\Psi = \sum_{p_1, p_2} a_{p_1 p_2} |q_1; p_1\rangle \otimes |q_2; p_2\rangle \quad \text{(infinite sum)}$$

in the tensor product space is invariant under the entire group -- equivalently, annihilated by all the basis elements of the Lie algebra. The condition $P\Psi = 0$ implies that Ψ can be written

$$\Psi = \sum_p a_p \, |q_1;p\rangle \otimes |q_2;-p\rangle.$$

Requiring that $A^{\pm}\Psi = 0$ leads to the equations

$$a_p \sqrt{q_1 + p(p+1)} + a_{p+1} \sqrt{q_2 + p(p+1)} = 0,$$

$$a_p \sqrt{q_2 + p(p+1)} + a_{p+1} \sqrt{q_1 + p(p+1)} = 0$$

for all p. These are consistent only if $q_1 = q_2$. Then

$$a_{p+1} = - a_p, \tag{6.20}$$

except possibly at points where the coefficient $\sqrt{q_1 + p(p+1)}$ vanishes. However, the range of p where a_p is nonzero must (by definition of the tensor product) be a subset of the range of p in some nontrivial irreducible representation (Eqs. (B.3.5)). It follows that nonvanishing a's satisfying Eq. (6.20) must extend to infinity in at least one direction (the interval between two vanishing coefficients being forbidden). Consequently, the sequence $\{a_p\}$ is not square-summable; no normalizable invariant vector can exist.

We can now prove by induction that the representation (6.19) does not contain the trivial representation discretely. For assume that this is true of the direct integral decomposition[8] of the (n-1)-fold tensor product,

---

[8] See, e.g., [Maurin], Chap. V, or Coleman (1968), Sec. IV.

$$\sum_j \int^{\oplus} d\mu_j(r)\, D_r^j,$$

where D_r^j stands for j copies (in direct sum) of the irreducible representation D_r. Then the product (6.19) has the form

$$[\sum_j \int^{\oplus} d\mu_j(r)\, D_r^j] \otimes D_q = \sum_j \int^{\oplus} d\mu_j(r)\, [D_r^j \otimes D_q]$$

$$= \sum_j \int^{\oplus} d\mu_j(r)\, [\sum_k \int^{\oplus} d\mu_k^{(r)}(s)\, D_s^{kj}] = \sum_l \int^{\oplus} d\mu_l(s)\, D_s^l,$$

where the trivial representation never occurs discretely. (We have applied the associative law to the tensor product of a direct sum and integral with another representation and then applied the theorem to each term of the result.) This completes the proof of the uniqueness of the vacuum.

In summary, the second-quantized theory proposed by the Dubna group satisfies the axioms for quantum field theory in de Sitter space proposed in Sec. IV.1. In this respect it is attractive.

From a physical point of view, however, this theory is vulnerable to criticism along the lines of the argument at the end of Sec. V.5. For small p the choice of T_p as the "positive-frequency" solution has not been physically motivated. The argument that it fits into an irreducible representation with functions which have the correct behavior at large p is conclusive only if the existence of an invariant space of particle wave functions is assumed beforehand. For this reason the conclusion of Tagirov *et al*. that particles are not created or destroyed in de Sitter space seems to the present author to be circular. On its face, at least, their definition of annihilation and creation operators does not appear to have much more justification than to define such operators for the field obeying Eq. (5.6) to be the annihilation and creation operators for incoming particles (coefficients of the expansion for the field in terms of the functions (5.7-8)), and to conclude therefrom that particles are not created by the potential V(t). A deeper analysis of the notion of *particle* seems to be necessary to settle this question. We shall return to the subject in Chapter X, but no conclusive resolution will be claimed.

7. *The Solutions in the Static Picture*.

We have seen that the wave equation in Fermi coordinates has solutions of the form (2.16), where $u = \sin\rho = \tanh\beta$ and $f_\lambda(u)$ is an associated Legendre function. To study the completeness and normalization of these eigenfunctions it is convenient to use the form (2.10) of the wave equation. Substituting Eq. (2.16) into Eq. (2.10), we obtain

$$-\frac{d^2 f_\lambda}{d\beta^2} + q \operatorname{sech}^2\beta\, f_\lambda = \lambda^2 f_\lambda \qquad (7.1)$$

(which is equivalent, of course, to Eq. (2.17)). Eq. (7.1) is identical in form to a nonrelativistic one-dimensional Schrödinger equation with a smooth, everywhere positive potential which falls off rapidly to zero at infinity. (Here λ^2 takes the role of the energy eigenvalue.) Consequently, most of the things we need to know about the solutions are well-known results from the one-dimensional barrier penetration problem ([Messiah], Chap. III):

The spectrum of λ^2 is continuous and extends from 0 to $+\infty$. For each value of $k = +\sqrt{\lambda^2}$ there are two linearly independent eigenfunctions, $f_\lambda = \Phi_k(\beta)$ and $f_\lambda = \Phi_{-k}(\beta)$, which respectively correspond in the nonrelativistic scattering problem to beams of particles with energy k^2 incident from the left (momentum k) and from the right (momentum $-k$). There are no "bound states". The eigenfunctions form a complete set (that is, the differential operator in Eq. (7.1) is essentially self-adjoint). When $\lambda^2 \geq q$ the solutions are oscillatory over the whole range of β; when $\lambda^2 < q$ there is a "classically forbidden" region near $\beta = 0$ where the solutions have an approximately exponential behavior. Implications of this observation will be discussed in Secs. VI.1 and VIII.6.

One can check by counting that this spectrum coincides with what one would expect on group-theoretical grounds. On the one hand, we know from the results of Sec. V.4 that the solutions of the wave equation constitute two irreducible representations of the principal series. The spectrum of λ, therefore, should consist of all the real numbers with multiplicity 4 (see Sec. II.3). On the other hand, we know from Sec. III.5 that the solutions are in one-one correspondence with the Cauchy data on an initial geodesic hypersurface. A complete orthonormal set of (generalized) functions on the initial surface must be used twice, to expand the initial value of the function itself and of its time derivative. However, this doubling corresponds precisely to the freedom in the choice of the sign of λ (cf. Secs. VIII.1-2); for a given sign there is a one-one correspondence between solutions and initial values of the function. Now the Ψ_k and Ψ_{-k} cover only the region of the hypersurface where $|\rho| < \pi/2$; another similar set of eigenfunctions is needed to expand functions on the "back side" ($\pi/2 < |\rho| < \pi$). Thus for each λ there are four independent eigenfunctions, as expected.

Completeness and orthogonality mean that every function on the interval $-\infty < \beta < \infty$ can be expanded as

$$f(\beta) = \int_{-\infty}^{\infty} d\mu(k)\ \Psi_k(\beta)\ \tilde{f}(k), \tag{7.2}$$

where

$$\tilde{f}(k) = \int_{-\infty}^{\infty} d\beta\ \Psi_k^*(\beta)\ f(\beta). \tag{7.3}$$

The measure $\mu(k)$ remains to be determined; or, equivalently, the Ψ's must be normalized so that

$$d\rho(k) = dk. \tag{7.4}$$

This can be done easily by referring to the scattering-theory interpretation of the wave functions. The result is[9]

$$\Psi_k(\beta) = -\sqrt{\frac{\pi}{2}}\,\frac{\exp(-\pi k/2)}{\sin \pi(\nu+ik)}\,\frac{\Gamma(1+\nu-ik)}{\Gamma(-ik)\;\Gamma(1+\nu+ik)}\,P_\nu^{ik}(\tanh\beta), \tag{7.5a}$$

$$\Psi_{-k}(\beta) = \Psi_k(-\beta) \qquad (k \geq 0), \tag{7.5b}$$

where ν is given in Eq. (2.18). The rest of this section is devoted to the derivation of Eqs. (7.5).

As the basis $\{\Psi_k;\Psi_{-k}\}$ we have chosen the <u>in</u>-states, each of which is the sum of an "incoming" plane wave and a scattered wave which asymptotically on each side of the barrier shows "outgoing" behavior.[10] The asymptotic form of the sum is then (see [Messiah], Sec. III.11)

$$\Psi_k(\beta) \sim \begin{cases} e^{ik\beta} + R_k\, e^{-ik\beta} & \text{as } \beta \longrightarrow -\infty, \\ S_k\, e^{ik\beta} & \text{as } \beta \longrightarrow +\infty; \end{cases} \tag{7.6a}$$

[9] The branch of $P_\nu^{ik}(u)$ understood here is described in the course of the proof below.

[10] This is an arbitrary choice. One could equally well use wave functions with outgoing plane waves, or some linear combination (cf. Sec. X.4). However, once Ψ_k is chosen, Ψ_{-k} is determined up to phase by the condition of orthogonality.

$$\Psi_{-k}(\beta) \sim \begin{cases} S_{-k}\, \text{ȩ}^{-ik\beta} & \text{as } \beta \longrightarrow -\infty, \\ \text{ȩ}^{-ik\beta} + R_{-k}\, \text{ȩ}^{ik\beta} & \text{as } \beta \longrightarrow +\infty. \end{cases} \tag{7.6b}$$

(see Fig. 11). The cedilla ($\text{ȩ} = e/\sqrt{2\pi}$) indicates that the incoming part has the usual delta-function normalization of a plane wave. It follows by a wave-packet argument that these functions are properly normalized, in the sense of Eqs. (7.2-4). Eq. (7.5b) now follows from the reflection symmetry of the potential, $q \operatorname{sech}^2 \beta$.

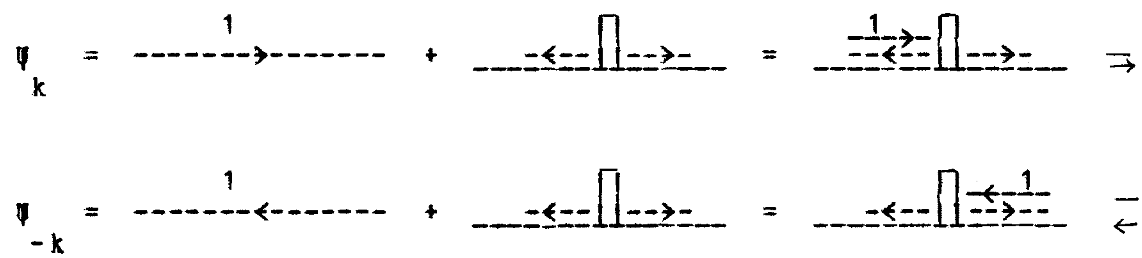


Fig. 11

In-states for the one-dimensional scattering problem. Each directed line stands for one of the terms in Eqs. (7.6). The numeral 1 indicates the normalized component.

To find $\Psi_k(\beta)$ in terms of Legendre functions, we use the relations ([N.B.S.], Eqs. (8.1.2,6) (p. 332))

$$P_\nu^{ik}(u) = \frac{1}{\Gamma(1-ik)} F_1^k(u), \qquad (7.7a)$$

$$e^{\pi k} Q_\nu^{ik}(u) = \frac{1}{2}\Gamma(ik)\, F_1^k(u) + \frac{\Gamma(1+\nu+ik)\ \Gamma(-ik)}{2\Gamma(1+\nu-ik)} F_2^k(u), \qquad (7.7b)$$

$$F_1^k(u) = \left[\frac{u+1}{u-1}\right]^{ik/2} F[-\nu, \nu+1, 1-ik, \frac{1-u}{2}], \qquad (7.8a)$$

$$F_2^k(u) = \left[\frac{u-1}{u+1}\right]^{ik/2} F[-\nu, \nu+1, 1+ik, \frac{1-u}{2}]. \qquad (7.8b)$$

The hypergeometric series converge for $|1 - u| < 2$. The ambiguity in the branch of the other factor (which is not just a phase, since ik is imaginary!) is settled by putting a cut from $+1$ to $-\infty$ and stipulating that $|\arg (u \pm 1)| < \pi$ elsewhere in the plane. Finally, let us define the functions on the interval $-1 < u < 1$ as the limit from above the cut; thus

$$\arg (u + 1) = 0, \quad \arg (u - 1) = +\pi. \qquad (7.9)$$

(Another common definition of $P_\nu^{ik}(u)$ differs by a factor of $\exp(\pi k/2)$ -- see [N.B.S.], Eqs. (8.3.1-2) (p. 333).)

Next we investigate the asymptotic behavior as $\beta \longrightarrow +\infty$, or $u \longrightarrow 1$. Since $F(a,b,c,y) \longrightarrow 1$ as $y \longrightarrow 0$, we need only study

$$\frac{u+1}{u-1} = \frac{\tanh \beta + 1}{\tanh \beta - 1} = - e^{2\beta},$$

where the minus sign must be written as $\exp(-i\pi)$ in accordance with Eqs. (7.9). So we have

$$F_1^k(u) \sim (e^{-i\pi} e^{2\beta})^{ik/2} = e^{+\pi k/2} e^{ik\beta}, \qquad (7.10)$$

a pure transmitted wave. Thus Ψ_k is proportional to F_1^k (cf. Eq. (7.6a)).

To find the normalization constant we must isolate the coefficient of $\exp(ik\beta)$ in F_1^k at $\beta = -\infty$. The formula

$$P_\nu^{ik}(-u+i\varepsilon) = e^{+i\pi\nu} P_\nu^{ik}(u-i\varepsilon) - \frac{2}{\pi} e^{+\pi k} \sin \pi(\nu+ik) Q_\nu^{ik}(u-i\varepsilon) \qquad (7.11)$$

([N.B.S.], Eq. (8.2.3) (p. 333)) yields

$$F_1^k(-|u|+i\varepsilon) = [e^{i\pi\nu} - \frac{1}{\pi}\Gamma(1-ik)\Gamma(ik) \sin \pi(\nu+ik)]\, F_1^k(|u|-i\varepsilon)$$

$$- \frac{1}{\pi} \sin \pi(\nu+ik) \frac{\Gamma(1-ik)\Gamma(-ik)\Gamma(1+\nu+ik)}{\Gamma(1+\nu-ik)} F_2^k(|u|-i\varepsilon). \qquad (7.12)$$

Now because of our branch convention,

$$F_1^k(u-i\varepsilon) = e^{-\pi k} F_1^k(u+i\varepsilon),$$

$$F_2^k(u-i\varepsilon) = e^{+\pi k} F_2^k(u+i\varepsilon). \qquad (7.13)$$

Therefore, as $|u| \longrightarrow 1$, $\beta \longrightarrow -\infty$, we have (using Eq. (7.10) and the analogue for F^k_2)

$$F^k_1(-|u|) \sim \left[e^{i\pi\nu} - \frac{1}{\pi}\Gamma(1-ik)\Gamma(ik)\sin\pi(\nu+ik)\right] e^{-\pi k/2} e^{ik|\beta|}$$

$$- \frac{1}{\pi}\sin\pi(\nu+ik)\, A(\nu,k)^{-1}\, e^{+\pi k/2} e^{-ik|\beta|}, \qquad (7.14)$$

where $A(\nu,k)^{-1}$ is the quotient of gamma functions in Eq. (7.12). The second term is the one which is to be normalized. Consequently,

$$\Psi_k(\beta) = -\sqrt{\frac{\pi}{2}}\,\frac{\exp(-\pi k/2)}{\sin\pi(\nu+ik)}\, A(\nu,k)\, F^k_1(u), \qquad (7.15)$$

which is the assertion (7.5a).

With each Ψ_k there are unambiguously associated a positive-frequency and a negative-frequency solution, given by Eq. (2.16) with positive and negative λ ($f_\lambda(u) = \Psi_k(\beta)$). Hence it seems obvious how to construct a field theory in analogy to the quantization of the free Klein-Gordon field. This can be done equally well for any static space-time; it will be carried out in Chapter VIII. The nonuniqueness of the quantization thus obtained and the associated particle interpretation will be discussed in Chapter IX.

8. Solutions of the Wave Equation in the Open De Sitter Space.

For the two-dimensional open de Sitter space (see Sec. III.6) the wave equation (2.1) again takes the forms (2.3,7) and (2.8,10), except that, because of the interchange of space and time, the physically relevant values of q are negative. (That is, - q plays the role of the square of the mass.) We expect, therefore, to encounter the discrete series of representations of $SO_o(1,2)$.

The functions (2.16) are solutions in the geodesic Gaussian coordinate system (ρ,χ). The $\exp(-i\lambda\chi)$ are a complete set of functions on a spacelike hypersurface. The time dependence is given for each λ by two linearly independent functions obeying Eq. (2.17). Certain linear combinations of these have the correct properties to be interpreted as positive- and negative-frequency functions in analogy to Sec. V.3. The connection with the irreducible representations of $SO_o(1,2)$ or its covering group is not at all clear in this picture.

Now let us turn to the static Fermi coordinate system (σ,τ), and let us, to begin with, consider the wave equation on the covering space (Fig. 8). There are solutions of the form

$$f_p e^{-ip\sigma}, \tag{8.1}$$

where f_ρ is a function of the space variable which satisfies a differential equation which can be written in the various forms

$$-\frac{d^2f}{d\alpha^2} - q\sec^2\alpha\, f = p^2 f, \qquad (8.2a)$$

$$-\left(\cosh\tau \frac{d}{d\tau}\right)^2 f - q\cosh^2\tau\, f = p^2 f, \qquad (8.2b)$$

or Eq. (2.13). The last of these shows that $f(z)$ is an associated Legendre function of order $\mu = p$ (not necessarily an integer) and degree

$$\nu = -\frac{1}{2} + \sqrt{\frac{1}{4} - q}. \qquad (8.3)$$

(We take the positive square root for definiteness. The negative root gives $-\nu - 1$.) Eq. (8.2a) is the easiest to discuss, since it has the familiar form of a Schrödinger equation.

Since the Cauchy problem is not in general well-posed in this space, as explained in Sec. III.6, one might expect that the differential operator in Eq. (8.2a) would not be self-adjoint. That is, in order to distinguish uniquely a complete set of eigenfunctions $\{f_\rho\}$, it should be necessary to impose boundary conditions controlling the disposition of "probability which reaches infinity within a finite time". (See Wightman (1964), Sec. 8.) Different boundary conditions would lead to a different spectrum of p^2 and to different behavior at other times of the solution with given initial data.

However, we shall see that this conjecture is <u>not true</u> except for very small mass ($q > -3/4$). Thus the Cauchy problem for the generalized Klein-Gordon equation with sufficiently large mass <u>is</u> well-posed, at least if one demands that the data on each hypersurface of constant time lie in a certain Hilbert space of pairs of sufficiently integrable functions. A rule of thumb is that the Hamiltonian operator of a quantum-mechanical system is self-adjoint if the corresponding classical particles cannot reach spatial infinity within a finite time.[11] Thus a very crude physical "explanation" of this result is that, although photons, following the light cones, can reach infinity within a finite time (in marked distinction to the case of flat space), the classical trajectories of massive particles (the geodesics) never reach infinity at all (see Secs. III.4-6).

We now turn to the proof of the assertion. First the general theory of self-adjoint extensions of differential operators[12] will be briefly reviewed, since the problem at hand is a special case of one which will arise in the general theory of field quantization in static metrics (see Sec. VIII.1).

A real second-order differential operator of the form

$$-\frac{d^2}{dx^2} + V(x) \qquad (8.4)$$

acting on functions defined on the interval $a < x < b$ may have

[11] Wightman (1964), pp. 266-268. Two qualifications: (1) There are counterexamples to the rule (E. Nelson, unpublished lectures). (2) The rule has always been applied to nonrelativistic mechanics, not, as here, to a relativistic problem (with a different relation between momentum and energy).
[12] See, e.g., [Akhiezer-Glazman], Appendix II.

0, 1, or 2 square-integrable eigenfunctions for a given eigenvalue λ in the complex plane. This number is the same for all nonreal λ; we shall call it the <u>deficiency</u> of the operator. It is the number of boundary conditions which must be added to define the expression (8.4) as a self-adjoint operator in the Hilbert space of L^2 functions on the interval (a,b).

The deficiency is determined by the behavior of V at the endpoints. If an endpoint (a or b) is finite and $|V(x)|$ is integrable in its neighborhood, the endpoint is <u>regular</u>; otherwise, <u>singular</u>. A regular endpoint contributes one unit to the deficiency; however, this can be immediately remedied by imposing a boundary condition at that point on the "acceptable" eigenfunctions. (Examples: (1) $f(a) = 0$ (reflecting wall boundary condition); (2) $f(a) + \gamma f'(a) = 0$; (3) $f(a) = f(b)$, $f'(a) = f'(b)$ (periodic boundary conditions -- applicable if both endpoints are regular).) A singular endpoint may or may not contribute to the deficiency; these are called the <u>limit circle</u> and <u>limit point</u> cases, respectively.

If the potential is symmetric ($V(-x) = V(x)$), the classification of the two endpoints must be the same, so the deficiency can only be 0 (limit point) or 2 (regular or limit circle). In the first case the spectrum is uniquely determined; the only square-integrable eigenfuctions of the operator (8.4) are those belonging to the point spectrum. In the second case <u>all</u> the eigenfuctions (two of them for each complex number as eigenvalue) are square-integrable, and they will not all be mutually orthogonal. Then the boundary conditions pick out a discrete set of real eigenvalues corresponding to a complete orthogonal system of eigenfunctions.

Let us see how the operator (8.2a) fits into this framework. (Recall that the range of α is $-\pi/2 < \alpha < \pi/2$.)

If $q = 0$ the endpoints are regular. The solutions are too well known to require comment. The boundary conditions

$$f(\pi/2) = f(-\pi/2) = 0, \tag{8.5}$$

corresponding to reflection at the boundaries with reversal of phase, yield the spectrum

$$|p| = 1, 2, \ldots$$

appropriate to a representation of the de Sitter group with $q = 0$ (Eqs. (B.3.5b)). Other boundary conditions give rise to Hilbert spaces of solutions of the wave equation which are not invariant under $SO_o(1,2)$*.

If $q \neq 0$, the endpoints are singular. A series solution ([Carrier-Krook-Pearson], pp. 198-202) of Eq. (8.2a) about the point $\alpha = \pi/2$ shows that for any p there are two independent solutions which behave near that point as

$$x^{-\nu} \quad \text{and} \quad x^{\nu+1} \tag{8.6}$$

($x = \alpha - \pi/2$, ν given by Eq. (8.3)). This asymptotic behavior can also be observed from explicit expressions for the Legendre functions at infinity, [N.B.S.], Eqs. (8.1.3,5), p. 332. If $\nu \geq 1/2$ ($q \leq -3/4$), only the second of the solutions (8.6) is square-integrable in the neighborhood of $\pi/2$. This is clearly a limit point case. A normalizable eigenfunction exists only for those discrete values of p such that a solution can be square-integrable at both ends of the interval simultaneously. The spectrum of $|p|$ must be that corresponding to the unitary representations of the covering group $SO_0(1,2)^*$ with $q = -\nu(\nu + 1)$. In particular, if ν is an integer, the solutions (8.1) are periodic in σ and we have the representations (B.3.5b) of $SO_0(1,2)$.

If $0 > q > -3/4$ ($\nu < 1/2$), both solutions (8.6) are square-integrable for all p, and we have the limit circle case. No single-valued representations of $SO_0(1,2)$ (integral ν) fall in this range.

The eigenvalue equation in the form (8.2b) can be transformed by the substitution

$$f = \cosh^{1/2} \tau \, \phi$$

to the form

$$-\frac{d}{d\tau}\left(\cosh^2 \tau \frac{d\phi}{d\tau}\right) + \frac{1}{4}\phi - \left(q + \frac{3}{4}\right) \cosh^2 \tau \, \phi = p^2 \phi,$$

where the operator is Hermitian with respect to the L^2 scalar product. Here the qualitative change at $q = -3/4$ is clearly shown in the behavior of the zeroth-order term.

In the solution spaces which support group representations each admissible value of p appears once, and so does - p. Hence we have a direct sum of the two inequivalent irreducible representations with the given value of q. This count is consistent with the results on the decomposition of the quasiregular representation cited in Sec. VI.1, and with the fact that in the space of solutions in the Gaussian system described at the beginning of this section the spectrum of λ has multiplicity 2 (cf. Sec. II.3).

Chapter VI

CONTRACTION OF THE REPRESENTATIONS OF THE DE SITTER GROUP TO REPRESENTATIONS OF THE POINCARE GROUP: GEOMETRICAL APPROACH

Clearly one cannot rest satisfied with the treatment of contraction of group representations in Chapter II. What it means for unitary representations of two groups to be related by contraction has never been precisely defined. Some intuitive relationships among various series of representations have been pointed out, but the algebraic manipulations employed involved so many *ad hoc* procedures and had to be modified so often to yield the desired results that they can hardly be said to constitute derivations of anything. Finally, the mysterious role of the phases of the basis vectors, which should be arbitrary, cries out for explanation.

In Secs. C.5-6 some of these problems are resolved in the case of the contraction of SO(3) to ISO(2) by paying close attention to the geometrical meaning of contraction. (The basic idea is that the contracted group is a local approximation to the action of the original group on a homogeneous space near a point whose stability group is the subgroup which defines the contraction.) Also, in Sec. III.3 an observation of Wigner (which has been elaborated upon by Philips and Wigner (1968), Sec. VIII) was cited, to the effect that the separation of positive and negative energies in the contraction of an irreducible representation of the de Sitter group must be related to the choice of the point of de Sitter space at which the contraction is regarded as taking place. The author's original intention was to present in this chapter a thorough treatment of the contraction of the representations of $SO_0(1,2)$ from the point of view of functions on the homogeneous space $SO_0(1,2)/SO_0(1,1)$ (the closed de Sitter space). Time and space did not allow this program to be carried out. Nevertheless, in order not to leave the subject hanging, the basic ideas will be presented here in a qualitative way.

1. The Quasiregular Representation of $SO_0(1,2)$.[1]

The natural (quasiregular) action of the de Sitter group and its Lie algebra on the smooth scalar functions of compact support on de Sitter space has been discussed in Secs. A.3 and V.1. Integration with respect to the volume element of the manifold defines an invariant scalar product on this space of functions, with respect to which it can be completed to form a Hilbert space. The unitary representation of the group in this Hilbert space is also called the quasiregular representation.

The quasiregular representation can be decomposed into

---

[1] In this section we shall keep the two-dimensional de Sitter space in mind, but most of the remarks apply to higher dimensions as well.

a direct integral of irreducible representations.[2] Each representation of the principal series occurs twice, as might be expected from our count of the linearly independent solutions of the wave equation in Chapter V (see Secs. V.4 and V.7). Discrete representations also appear (once each in dimension 2 -- cf. Sec. V.8).

Of course, the quasiregular representation of the Poincaré group on Minkowski space can also be decomposed into irreducibles. The familiar Fourier transform does just that. (The plane waves with four-momenta satisfying $p_\mu p^\mu$ = const. support one or more irreducible representations corresponding to that value of the Casimir operator (I.4.4) -- cf. Sec. B.2.) It would also be possible to use basis functions which diagonalize the Lorentz transformations, rather than the translations, within each irreducible representation. In this case the dependence on the coordinate corresponding to geodesic distance from the origin would be given by certain Bessel functions (cf. Chapter IX and Sec. X.2).

The contraction of the quasiregular representation of the de Sitter group to that of the Poincaré group, therefore, provides a natural geometrical setting for the contraction of the respective irreducible representations which are imbedded in them. In fact, this may be used as the definition of the

---

[2] [Gelfand 5] (case n = 3); Molchanov (1966); Limić et al. (1967). Cf. also Börner and Dürr (1969) (see Sec. III.7 above) and Nachtmann (1968b).

contraction relation for the irreducible representations. One proceeds in analogy to the treatment of the rotation group in Appendix C. Consider a function with compact support near the point of contraction; let the radius of the space go to infinity; but "shrink" the function in some systematic (but not unique) way so as to keep the size of the support roughly fixed. (The easiest way to do this is to use one of the standard coordinate systems which, as was explained at length in Chapter III, are closely related to various coset decompositions of the group. Rescale some of the coordinates proportionally to R, and keep the explicit form of the function in terms of the coordinates fixed. See the examples in Secs. C.5-6.) Then, if the relative phases of the basis vectors are correctly chosen, the coefficients of the expansion of the function with respect to the basis elements of the quasiregular representation on the hyperboloid will converge to the expansion coefficients of a function on the plane with respect to a basis for the quasiregular representation of the Poincaré group. In particular, the correspondence of irreducible components of the representations appears in this way; it is manifested in asymptotic expressions for some special functions in terms of others, analogous to Eqs. (C.5.3) and (C.6.6).

For instance, let the point of contraction be the point O in Fig. 3 and use bases of the form (V.6.7) for the principal-series contributions to the quasiregular representation of $SO_0(1,2)$. Then when τ and σ are rescaled as in Eq. (III.1.16) one will recover the Fourier transform in Minkowski space, the Legendre functions being approximated by complex exponentials when τ is small and p is large. (The spacelike momenta are contributed by the representations of the discrete series, in accordance with Sec. II.2.)

The same is true if one uses the static coordinates of Fig. 4 around O, or the coordinates of Fig. 9, although the family of functions through which a given initial function is mapped as R changes will be different in each of the three cases. In connection with the static case it should be noted that the wave functions discussed in Sec. V.7 have "dips" near O when $\lambda^2 < q$, since this is the classically forbidden region inside the potential barrier. Consequently, as a function shrinks down toward O its components along these basis vectors will rapidly vanish. Thus the energy gap between $-m$ and $+m$ in the representations of the Poincaré group arises naturally, even though λ ranges continuously from $-\infty$ to $+\infty$ in the original de Sitter representation (cf. Sec. II.3).

Finally, if one contracts around P in Fig. 5 and uses basis functions adapted to the coordinates indicated in that figure, one obtains an expansion in the plane in terms of the Bessel functions mentioned above.[3]

[3] To prevent misunderstanding, it should be stated again that the author has not actually performed for these cases the detailed manipulations with special functions analogous to those in Secs. C.5-6. Nevertheless, it is obvious that relationships of the type indicated must hold.

2. Geometrical Contraction of an Irreducible Representation.

The indefinite metric of de Sitter space, as contrasted with the positive definite metric of the sphere considered in Appendix C, has an important consequence for the structure of the set of functions on the space which supports an irreducible representation of the principal series. The partial differential equation (V.2.1) which helps determine the set is hyperbolic instead of elliptic, and hence a Cauchy initial value problem can be posed for it. When initial data are specified on a spacelike hypersurface through the point of contraction, all sufficiently well-behaved initial data occur, including those with compact support near the point of contraction. Hence it is possible in this case to apply directly to an irreducible representation a geometrical approach to contraction similar to that of the last section.

Construct an irreducible representation of $SO_o(1,n)$ as in Sec. V.4. That is, choose a particular solution Ψ_o of the wave equation such that $W(\Psi_o,\Psi_o) \neq 0$ (W defined by Eq. (V.4.1)). The space generated from Ψ_o by the natural action of the operators of the group (or of the Lie algebra) consists of vectors with the same sign of W, which thus defines a scalar product. The space becomes a Hilbert space supporting an irreducible unitary representation.

The functions in this space are in one-one correspondence with their initial values on a spacelike hypersurface, which may be taken to be an (n-1)-dimensional hypersphere. Let the radius of the de Sitter space, and hence of the sphere, tend to infinity, confining one's attention to wave functions whose initial values have support within a fixed geodesic distance from the point of contraction. (The entire function, of course, does not have support near the point, and neither does the initial value of its normal derivative on the hypersphere.) As in Secs. C.5-6, the coefficients of the spherical harmonic expansion of the initial value of such a function will converge to the coefficients of an expansion of a function on the (n-1)-dimensional Euclidean plane with respect to a basis of eigenfunctions of the Laplacian.[4] In this process certain discrete parameters will become continuous.

Do these wave functions themselves now converge to solutions of the wave equation on the n-dimensional Minkowskian plane which form the core of an irreducible unitary representation of the Poincaré group? This will be true only if the basis functions corresponding to adjacent values of the confluent discrete paremeters just mentioned agree in the limit. For example, consider the two-dimensional case, where the only parameter is the Fourier series index p. When $R \gg 1$ and q, p, and so on have been rescaled as in Sec. II.1, Eqs. (V.4.6) and

[4] In the case of two-dimensional de Sitter space, "converge" is an understatement. As an initial value g(x) is replaced by g(Rx) and x is rescaled compensatingly, its Fourier series coefficients for various finite values of R and the values of the Fourier transform ($R \longrightarrow \infty$) are all values of the <u>same</u> analytic function, $\hat{g}(k)$. But this is an accident of the low dimension -- a one-dimensional manifold has no intrinsic curvature.

(V.4.9) become approximately

$$\Psi_{p+1} = [m^2 + \bar{p}^2]^{-1/2} A^+ \Psi_p , \tag{2.1}$$

$$A^+ [e^{i\bar{p}x} \bar{E}_p(t)] = - [m^2 + \bar{p}^2] e^{i(\bar{p}+\frac{1}{R})x} \bar{O}_{p+1}(t),$$

$$A^+ [e^{i\bar{p}x} \bar{O}_p(t)] = - e^{i(\bar{p}+\frac{1}{R})x} \bar{E}_{p+1}(t). \tag{2.2}$$

(The bars on the functions E and O indicate that the independent variable (the time) has been appropriately transformed.) The basis vectors Ψ_p are of the form (V.4.4). It is easy to see that the Ψ_p for odd p will be consistent with those for even p only if A(p) and B(p) in Eq. (V.4.4) have very special values. Namely, up to phase, we must have

$$A(p) = [\sqrt{2}\, \sqrt{m^2 + \bar{p}^2}\,]^{-1}, \qquad B(p) = \pm \frac{1}{\sqrt{2}} \sqrt{m^2 + \bar{p}^2}, \tag{2.3}$$

the sign in B corresponding to that of W. The contracted basis functions are then the plane waves which support the representation of the Poincaré group with positive or negative energy, respectively.

Note that according to this version of the geometrical picture an irreducible principal-series representation of the de Sitter group contracts to one irreducible representation of the Poincaré group, not to two.

3. The GNS Construction; Contraction and Reconstruction of Field Theories.

Neither the approach to contraction in terms of Lie algebra matrix elements (Chapter II) nor that in terms of functions on a homogeneous space (Secs. VI.1-2) lends itself easily to the formulation of a precise definition. In this section still another approach is suggested, which has the advantage that a rigorous definition can be stated. A possible application to field theory is described.

A unitary representation of a Lie group G determines an Hermitian representation of its Lie algebra, $\mathcal{L}(G)$, and hence of the associative algebra generated by the Lie algebra, which is the (complex) universal enveloping algebra $\mathcal{U}(G)$. For $\mathcal{U}(G)$ there applies the correspondence between (cyclic) representations and "states" which is given by the Gel'fand-Naimark-Segal construction. The version of the GNS theorem which applies to an algebra of unbounded operators is stated by Powers (1971), Sec. 6. Davies (1971) has recently studied Lie algebras in this way, but he did not consider the subject of contraction.

If $\{A(j)\}$ $(1 \leq j \leq n)$ is a basis for $\mathcal{L}(G)$, then

$$\{\prod_{k=1}^{M} A(j_k)\} \qquad (j_1 \leq j_2 \leq \ldots \leq j_M,\quad M = 0, 1, \ldots)$$

is a basis for $\mathcal{U}(G)$. This statement is true for all Lie algebras of dimension n. So if $\mathcal{L}(G)$ and $\mathcal{L}(G')$ are nonisomorphic Lie algebras of the same dimension, then $\mathcal{U}(G)$ and

$\mathcal{U}(G')$ as vector spaces are the same; only the multiplicative structure (in effect, the rules for expressing products of the A(j) in the "wrong order" in terms of the above basis elements) is different. Of course, the same thing is true if $\mathcal{L}(G)$ and $\mathcal{L}(G')$ are isomorphic -- in other words, if one considers two different bases for $\mathcal{L}(G)$ and then identifies, as vectors, corresponding basis elements. Thus the enveloping algebra regarded just as a vector space (which we can denote by $\mathcal{U}(n)$) provides a fixed arena within which contraction can take place.

The following definition is suggested:

A family {D(R)} ($R \longrightarrow \infty$) of unitary representations of a Lie group G <u>contracts</u> to a representation D' of a Lie group G' if for each R there is a linear functional $\Psi(R)$ on $\mathcal{U}(n)$ such that

(1) $\Psi(R)$ is a vector state from D(R) on $\mathcal{U}(G)$ (identified with $\mathcal{U}(n)$ by a particular choice of basis):

$$\Psi(R)[A] = (\Psi, D(R)[A]\Psi) \qquad (A \in \mathcal{U}(G)),$$

where Ψ is in the Hilbert space where the representation D(R) acts;

(2) There is a vector state Ψ' from a representation D' of $\mathcal{U}(G')$ such that $\Psi(R) \longrightarrow \Psi'$ in the ______ topology as $R \longrightarrow \infty$.

(This is a whole family of definitions, since the topology has not been specified. In the absence of any theorems, it is hard to commit oneself to a particular topology.)

The approach of Chapter II and the geometrical approach could be reformulated in this way, at the cost of studying one particular formal basis vector or function (respectively) instead of all of them at once. (Part of the contracted representation might be lost in this way because of loss of cyclicity.)

The approach to contraction in terms of a distinguished state (or, more precisely, a family of states parametrized by R) may be quite useful in the context of contraction of field theories. By this is meant, as in Sec. I.2, the construction of a Lorentz-covariant theory as a limit of a covariant theory in de Sitter space. The most obvious method of doing this is by means of the Wightman reconstruction theorem ([Streater-Wightman], Sec. 3.4).

First, assume that we have a field theory satisfying the axioms stated in Sec. IV.1. Form the vacuum expectation values (VEVs)

$$W_n(x_1,\ldots,x_n) = (\Psi_0, \phi(x_1) \,.\,.\,.\, \phi(x_n)\Psi_0).$$

It is easy to show, in precise analogy to [Streater-Wightman], Sec. 3.3, that the field's properties of covariance under the de Sitter group, Hermiticity, local commutativity, and positive definiteness of the scalar product are reflected in certain relations satisfied by the VEVs. (One needs to assume that the nuclear theorem holds for $\mathcal{T}$, the space of test functions.)

From the VEVs, which are multilinear distributions on de Sitter space, one may obtain distributions on Minkowski space by passing to the infinitesimal neighborhood of a point by a rescaling of the coordinates (of the type described in Sec. VI.1). The properties of the VEVs expressing Hermiticity, commutativity, and positivity will clearly be preserved in this limit. It is to be expected that the limit VEVs will satisfy the relations corresponding to Poincaré invariance of the contracted theory. A theorem to this effect is needed, however, before the program proposed in the Introduction can be implemented. Contraction of group representations as defined above may turn out to be a useful concept in this connection.

Once Lorentz covariance of the contracted VEVs has been established, it remains to verify the spectral condition and the cluster decomposition property. Then the reconstruction theorem will provide a field theory in flat space satisfying the Wightman axioms with the possible exception of asymptotic completeness.

Chapter VII

THE NEUTRAL SCALAR FIELD IN RIEMANNIAN SPACE-TIME -- GENERAL REMARKS

We shall apply the traditional canonical quantization procedure to a real scalar field satisfying a covariant wave equation on a Riemannian manifold of dimension $s + 1$ and signature $(+ - \ldots -)$ (s minus signs). We require (cf. Choquet-Bruhat (1967), p. 89) that it be possible to define the direction of time globally and continuously on the manifold.[1] The geometry of the manifold is assumed given -- it is an "external gravitational field". That is, there is no attempt to couple the metric of the space to the matter represented by the quantized field. Thus we are studying the quantum-field-theoretical analogue of the problem of determining the motion of test particles in classical general relativity.

We shall find that the classical Lagrangian-Hamiltonian treatment of the scalar field extends immediately to this situation, and that the "smeared fields" needed in quantum theory can be defined in a manifestly covariant way (still on the

---

[1] As we shall see at the end of Sec. VII.4, it is hard to imagine what the commutator of a canonically quantized field would be on a manifold for which there is no distinction between past and future. (One should also note the physical problems associated with the second law of thermodynamics, etc., in such a situation.) This requirement excludes from consideration the de Sitter space with antipodal points identified (see Sec. I.1).

classical level). Then in a formal sense one can proceed immediately to an algebra of quantized fields obeying the canonical commutation relations. However, it is not at all obvious how to realize these objects as operators on a Hilbert space. In the next three chapters we study this problem in two special cases where the wave equation can be solved by separation of variables.

In the fourth section of this chapter we study the distribution solutions of the wave equation which generalize the familiar $\Delta(x_1 - x_2)$, etc., and their relation to the general solutions of the wave equation. It is emphasized that some of these objects are uniquely determined at the formal level and some are not. In Sec. VII.5 the generalization of the indefinite scalar product or current form for the Klein-Gordon equation is discussed. In the last section the problem of quantization is discussed from a more abstract viewpoint.

1. The Classical Canonical Formalism.[2]

We follow the canonical Lagrangian formalism as expounded by, e.g., Hill (1951). In keeping with the spirit of this approach, integrals and derivatives will always be written explicitly in the coordinates. That is, $\int d^{s+1}x$ means $\int dx^0\, dx^1 \ldots dx^s$, and the factor $\sqrt{|g|}$ ($g = \det\{g_{\mu\nu}\} < 0$) needed

[2] This material, and much of the rest of this chapter, is not particularly new, but is treated in detail here in order to establish a consistent framework for scalar quantum field theory in curved space.

to make the volume element covariant must be written explicitly. Similarly, ∂_μ or $\partial/\partial x^\mu$ indicates the simple coordinate derivative, not the covariant (semicolon) derivative ∇_μ.

We start with the generally covariant Lagrangian density

$$\mathcal{L} = \frac{1}{2} \sqrt{|g|}\, [g^{\mu\nu} \partial_\mu \phi\, \partial_\nu \phi - m^2 \phi^2], \qquad (1.1)$$

where ϕ is a real scalar field. (The space-time variable $x = (x^0, x^1, \ldots, x^s)$ is suppressed.) Under coordinate transformations $\mathcal{L}$ transforms as a scalar density (not as a scalar), so that the action integral $I = \int d^{s+1}x\, \mathcal{L}$ is a scalar. (Some authors write $\mathcal{L} = \sqrt{|g|}\, L$ and call the scalar function L the Lagrangian.)

The Euler-Lagrange equation resulting from the variation of I with respect to ϕ is

$$0 = \frac{\partial}{\partial x^\mu} \frac{\partial}{\partial(\partial_\mu \phi)} - \frac{\partial \mathcal{L}}{\partial \phi} = \partial_\mu [\sqrt{|g|} g^{\mu\nu} \partial_\nu \phi] + \sqrt{|g|}\, m^2 \phi.$$

This can be written

$$\Box_c \phi + m^2 \phi = 0 \qquad (1.2)$$

where $\Box_c$ is the Laplace-Beltrami operator:

$$\square_c \phi = \frac{1}{\sqrt{|g|}} \partial_\mu [\sqrt{|g|}\, g^{\mu\nu} \partial_\nu \phi] = g^{\mu\nu} \nabla_\mu \nabla_\nu \phi. \qquad (1.3)$$

(See [Adler-Bazin-Schiffer], Sec. 3.2, for the derivation of the form in terms of covariant derivatives.)

The wave equation (1.2) reduces in flat space to the Klein-Gordon equation. It is not the only covariant generalization of the Klein-Gordon equation to curved space, since there might be other terms which vanish when the curvature is zero, such as $R^{\mu\nu}\nabla_\mu\nabla_\nu\phi$ or $R^\mu_\mu\phi/6$. But it seems to be the simplest, and we take it to be the gravitational analogue of "minimal coupling" in the theory of external electromagnetic fields.

On the other hand, a strong argument has been made[3] that the basic equation should contain a scalar curvature term $R^\mu_\mu\phi/6$ in the case s = 3. (For the general case see Eq. (V.2.2).) Then the equation is conformally invariant when m = 0. This could be accommodated by generalizing the present scheme to allow a scalar potential:

$$\square_c \phi + m(m - V)\phi = 0, \qquad (1.4)$$

V a function of space-time. With the possible exception of one point, mentioned in Sec. VIII.2 below, all the results of this and the next chapter extend directly to the case (1.4). There is

[3] Penrose (1963), pp. 565-566; Penrose (1965), Sec. 2; Chernikov and Tagirov (1968); Tugov (1969).

evidence that the addition of the curvature term makes a significant difference in predictions of particle creation in the early stages of expansion of the universe (Parker (1972)).

When s = 1 it is always possible[4] to choose coordinates in which the metric has the <u>conformally flat</u> form

$$ds^2 = g_{\mu\nu}\, dx^\mu dx^\nu = \alpha(x)\,[(dx^0)^2 - (dx^1)^2].$$

Then Eq. (1.2) becomes

$$\Box\phi + m^2\alpha(x)\phi = 0$$

where $\Box$ is the ordinary d'Alembertian. Thus in two-dimensional space-time the gravitational problem reduces to a scalar potential problem. In higher dimensions this fails for two reasons: not every manifold is conformally flat, and in the conformally flat case $g^{\mu\nu}$ and $\sqrt{|g|}$ do not cancel in Eq. (1.3) as they do when s = 1. However, in some cases the equation can still be reduced to this form at the price of changing the dependent variable (see Grib and Mamaev (1969)).

So far our formulation has been generally covariant. However, the notions of conjugate momentum and Hamiltonian will not make sense unless one coordinate is singled out as the time coordinate. Consequently, we consider from now on only coordinate systems in which the first coordinate, x^0, is timelike

[4] [Eisenhart], Sec. 28.

(see Appendix D). Moreover, we expect that a successful Hamiltonian formulation of the theory will be possible only if for each t the surface $x^0 = t$ is a Cauchy surface for the entire region of space covered by the coordinate system (see Secs. III.5-6). In what follows we frequently denote x^0 by t and the s-dimensional spatial coordinate by x.

The momentum conjugate to ϕ is defined in the usual way:

$$\pi = \frac{\partial \mathcal{L}}{\partial(\partial_0 \phi)} = \sqrt{|g|}\, g^{0\mu} \partial_\mu \phi. \tag{1.5a}$$

In particular, if $g^{0j} = 0$ for $j \neq 0$,

$$\pi = \sqrt{|g|}\, g^{00} \partial_0 \phi = \sqrt{{}^{(s)}g}\, \sqrt{g^{00}}\, \partial_0 \phi \tag{1.5b}$$

where ${}^{(s)}g$ is the determinant of $\{{}^{(s)}g_{jk}\} = \{-g_{jk}\}$ $(j,k \neq 0)$, the metric of space at a fixed time.

Consider a family of coordinate systems in all of which a given spacelike hypersurface has the form t = <u>const</u>. On this hypersurface π acts as a density (proportional to $\sqrt{{}^{(s)}g}$) relative to changes of the space coordinates, and it is unchanged[5] under transformations which leave the space coordinates in the

---

[5] This statement refers to π regarded as a numerical quantity defined on the manifold. The functional form of $\pi(t,x)$ will change, of course, when t and x are expressed in terms of new coordinates. Covariance relative to a distinguished hypersurface is discussed further in Sec. VII.3.

distinguished hypersurface and the direction of time unchanged (because the change in local time scale affects $\sqrt{g^{00}}$ and $\partial_0 g$ in compensating ways). Under time reversal the sign of π changes.

The canonical procedure yields the Hamiltonian density

$$\mathcal{H}(x) = \pi\,\partial_0\phi - \mathcal{L}$$

$$= \frac{1}{2}\sqrt{|g|}\Big[g^{00}(\partial_0\phi)^2 - \sum_{j,k=1}^{3} g^{jk}\,\partial_j\phi\,\partial_k\phi + m^2\phi^2\Big]. \qquad (1.6)$$

(This form is valid even if $g^{0j} \neq 0$.) In terms of π we have

$$\partial_0\phi = \big[\sqrt{|g|}\,g^{00}\big]^{-1}\big[\pi - \sqrt{|g|}\,g^{0j}\,\partial_j\phi\big], \qquad (1.7)$$

$$H = \int d^s x\,\mathcal{H} = \frac{1}{2}\int d^s x\,\sqrt{|g|}\,\Big\{\frac{1}{|g|g^{00}}\big(\pi - \sqrt{|g|}\,g^{0j}\,\partial_j\phi\big)^2 -$$

$$\sum_{j,k} g^{jk}\,\partial_j\phi\,\partial_k\phi + m^2\phi^2\Big\}. \qquad (1.8a)$$

If $g^{0j} = 0$,

$$H = \frac{1}{2}\int d^s x\,\Big[\frac{g_{00}}{\sqrt{|g|}}\pi^2 - \sqrt{|g|}\,g^{jk}\,\partial_j\phi\,\partial_k\phi + \sqrt{|g|}\,m^2\phi^2\Big]. \qquad (1.8b)$$

2. Formal Consistency of Canonical Quantization.[6]

Ultimately we hope to impose the canonical commutation relations (CCRs)

$$[\phi(t,x),\phi(t,y)] = [\Pi(t,x),\Pi(t,y)] = 0, \qquad (2.1a)$$

$$[\phi(t,x),\Pi(t,y)] = i\delta(x - y). \qquad (2.1b)$$

Note that $\delta(x - y)$ is the ordinary delta function in the s coordinates on the equal-time hypersurface (a density, not a scalar). It is easily verified that Eqs. (2.1) and Heisenberg's equation $dA/dt = i[H,A]$, with H given by Eq. (1.8), formally lead to the correct equations of motion (1.7) and (1.2).

It can also be verified by explicit calculation that the CCRs are formally consistent with the equations of motion, in the sense that if Eqs. (2.1) hold at one time, then they continue to hold for all time. We need the following delta-function identities:

$$h(x)\,\frac{\partial}{\partial x}\delta(x - y) = -\frac{\partial}{\partial y}[h(y)\delta(x - y)]; \qquad (2.2a)$$

$$\frac{\partial}{\partial x^j}\left[h^{jk}(x)\,\frac{\partial}{\partial x^k}\delta(x - y)\right] = \frac{\partial}{\partial y^j}\left[h^{jk}(y)\frac{\partial}{\partial y^k}\delta(x - y)\right]$$

$$\text{if } h^{jk} = h^{kj}. \qquad (2.2b)$$

[6] Conclusions similar to those of this section have been published by Urbantke (1969).

To verify Eq. (2.2a), smear each side with a test function[7] of the form f(x)g(y) and integrate by parts:

$$\int dx\, dy\; f(x)g(y)\; \frac{\partial}{\partial y}[h(y)\delta(x-y)] = -\int dx\; f(x)g'(x)h(x)$$

$$= \int dx\; g(x)\; \frac{d}{dx}[f(x)h(x)] = -\int dx\, dy\; f(x)g(y)h(x)\; \frac{\partial}{\partial x}\delta(x-y).$$

(A slight generalization of integration by parts shows that the same identity holds for an arbitrary test function f(x,y).) The proof of Eq.(2.2b) is similar.

We can now see that the derivatives of the canonical commutators vanish by virtue of the equations of motion (whence the assertion follows). Eliminate time derivatives from the commutators by means of Eq. (1.7) and

$$\dot{\Pi} = -\partial_j\left[\frac{1}{g^{00}}g^{j0}\Pi\right] +$$

$$\partial_j\left[\sqrt{|g|}\left(\frac{1}{g^{00}}g^{j0}g^{0k} - g^{jk}\right)\partial_k\phi\right] - \sqrt{|g|}\,m^2\phi, \qquad (2.3)$$

which is equivalent to the wave equation (1.2). (Eqs. (1.7) and (2.3) are the Hamiltonian equations derived from the Hamiltonian (1.8a).) Then

---

[7] At the present level of rigor we do not need to be precise about the test function space, the growth properties of the function h, etc.

$$\frac{d}{dt}[\phi(x),\Pi(y)] = [\dot{\phi}(x),\Pi(y)] + [\phi(x),\dot{\Pi}(y)] =$$

$$- ig^{0j}(x)/g^{00}(x)\frac{\partial}{\partial x^j}\delta(x-y) - i\frac{\partial}{\partial y^j}[g^{j0}(y)/g^{00}(y)\delta(x-y)] = 0.$$

Eq. (2.2b) similarly implies that $\frac{d}{dt}[\Pi(x),\Pi(y)] = 0$, and the $[\phi,\phi]$ case is immediate.

A generalization of the proposition just proved is the following: The canonical commutation relations (2.1) are generally covariant, in the sense that if they are imposed on one spacelike hypersurface, then they hold on any spacelike hypersurface (by virtue of the equations of motion). (This is in marked distinction to the noncovariance of various natural-seeming constructions of representations of the CCRs, which will be a central theme of the rest of this dissertation.) If two hypersurfaces do not intersect, they can both be regarded as equal-time hypersurfaces in one coordinate system, and then the assertion follows immediately from what has just been proved. If they do intersect, one can argue that the CCRs on each are equivalent to the CCRs on an intermediate hypersurface which intersects neither.

3. Covariant Smearing.

In the classical Lagrangian theory of Sec. VII.1 the field $\phi(x)$ has been taken to be a true scalar quantity (rather than a scalar density). On the other hand, we know that in quantum theory $\phi(x)$ will be a distribution which must be integrated over test functions. If the test functions are also taken to be scalars and we wish the result of the smearing to be invariant, we must define the smeared field as follows:

$$\phi(f) = \int d^{s+1}x \sqrt{|g|}\ \phi(x) f(x). \tag{3.1}$$

In the canonical formalism it is natural to smear $\phi(t,x)$ and $\pi(t,x)$ over a spacelike hypersurface of constant time. Let us, therefore, elaborate on the remark in Sec. VII.1 that π is a covariant density on a hypersurface. Call a coordinate system <u>compatible</u> with a hypersurface S if S is defined by an equation of the form t = <u>const</u>. in that system. An expression is S-<u>covariant</u> if it has the same form in all S-compatible coordinate systems. An integral over S is S-covariant if the integrand is the zeroth component of a contravariant vector density. (In the language of differential forms, one constructs an s-form from a contravariant vector by duality.) If $f(x) = f(t,x)$ is a scalar function on space-time, then

$$Df = \sqrt{|g|}\ g^{0\mu} \partial_\mu f \tag{3.2}$$

is such an object. Thus, in particular,

$$\phi(Df) \equiv \int_{t=\text{const.}} d^s x\ \phi(x) \sqrt{|g|} g^{0\mu} \partial_\mu f(x) \tag{3.3a}$$

and

$$\Pi(f) \equiv \int_{t=\text{const.}} d^s x\, \Pi(x) f(x) = \int d^s x\, D\phi(x)\, f(x) \qquad (3.3b)$$

are S-covariant scalars:

$$\int_{t'=\text{const.}} d^s x'\, \Pi'(x') f'(x') = \int_{t=\text{const.}} d^s x\, \Pi(x) f(x),$$

and so on, when the hypersurfaces of integration coincide.

When the coordinate system is not necessarily S-compatible, this type of integral is written

$$\int_S d\sigma_\mu \sqrt{|g|}\, j^\mu(x),$$

where $\sqrt{|g|}\, j^\mu$ is the vector density and

$$d\sigma_\mu = dx^1 \ldots \widehat{dx^\mu} \ldots dx^s .$$

The <u>covariant</u> <u>divergence</u> of a vector function $j^\mu(x)$ is the scalar

$$\text{div } j = \nabla_\mu j^\mu = \frac{1}{\sqrt{|g|}} \partial_\mu (\sqrt{|g|} j^\mu). \qquad (3.4)$$

If div $j = 0$, then it follows from Gauss's theorem that

$$\int_S d^s x \sqrt{|g|}\, j^0(x) \qquad (x^0 = \text{const. on } S) \tag{3.5}$$

is independent of the hypersurface, and thus is completely covariant ([Adler-Bazin-Schiffer], pp. 71-75).

4. <u>The Commutator and Related Distributions</u>.

Let us assume now that the field theory of Sec. VII.1 has been quantized, so that we have operator-valued distributions satisfying the equal-time commutation relations (2.1) and the equations of motion (1.5) and (1.2) (or (2.3)). Let L denote the operator $\Box_c + m^2$. We consider the commutator of the fields at arbitrary times,

$$[\phi(x_2),\ \phi(x_1)] = iG(x_2,x_1) \equiv iG(t_2,x_2;t_1,x_1). \tag{4.1}$$

G is antisymmetric and satisfies the field equation in each variable:

$$L_2\, G(x_2,x_1) = 0 = L_1\, G(x_2,x_1). \tag{4.2}$$

By virtue of Eqs. (2.1) it fulfills the initial conditions

$$G(t,x_2;t,x_1) = 0, \tag{4.3a}$$

$$D_2\, G(t,x_2;t,x_1) = -\, D_1\, G(t,x_2;t,x_1) = -\,\delta(x_2 - x_1). \tag{4.3b}$$

(D_2 indicates the contravariant time differentiation (3.2) acting on the variable x_2. We use the similar conventions L_2, $\sqrt{|g_2|}$,

$\partial_{0\lambda}$, and so on.)

G is the <u>propagator</u> for the Cauchy problem, when it is well-posed. That is, the unique solution[8] of the wave equation (1.2-3) with the initial values (which may be distributions -- see below)

$$F(t,x) = f(x), \qquad DF(t,x) = f'(x) \tag{4.4}$$

on the hypersurface $S = \{x|x^0 = t\}$ is

$$F(x_2) = -\int_S d^s x_1 \; \{G(x_2,x_1)\; D_1 f(x_1) - D_1 G(x_2,x_1)\; f(x_1)\}, \tag{4.5}$$

where $Df = f'$. This formula can be written $(n = s + 1)$

$$F(x_2) = -\int_S d^n x_1 \sqrt{|g_1|}\; G(x_2,x_1)\; f_*(x_1), \tag{4.6a}$$

$$\sqrt{|g_1|}f_*(x_1) = \delta(t_1 - t)f'(x_1)$$

$$+ \partial_{\mu 1}[\delta(t_1 - t)\sqrt{|g_1|}g_1^{0\mu} f(x_1)]; \tag{4.6b}$$

f_* is a distribution in the sense of Eq. (3.1) and is S-covariant.

<u>Proof</u>: It is obvious that $L_\lambda F = 0$ and that $F(t,x_\lambda) = f(x_\lambda)$. Let

---

[8] Whenever numerical (rather than operator-valued) solutions of the wave equation are considered, we allow complex values, even in connection with the theory of an Hermitian field.

$$D_2 F(x_2) = -\int_S d^s x_1 \; D_2 G(x_2, x_1) \; D_1 f(x_1) + A.$$

The first term here becomes the desired $f'(x_2)$. The second term is, with

$$W = \int d^s x_1 \; D_1 G(x_2, x_1) \; f(x_1),$$

$$A = \sqrt{|g_2|}\;\{g_2^{00}(\partial_{02} + \partial_{01} - \partial_{01})W + g_2^{0j}\partial_{j2} W\}$$

$$= \sqrt{|g_2|}\;\{g_2^{00}\frac{2\partial}{\partial(t_1 + t_2)}W\Big|_{(t_1 - t_2 \text{ fixed})} - g_2^{00}\partial_{01} W + g_2^{0j}\partial_{j2} W\}$$

$$\equiv A_1 + A_2 + A_3.$$

When $t_1 - t_2 = 0$, W is independent of $t_1 + t_2$ (by Eq. (4.3b)), so $A_1 = 0$. Using the wave equation for G in the form (2.3) (with $x = x_1$, $\phi = G$, $\Pi = D_1 G$), we have

$$A_2 = \sqrt{|g_2|}\,g_2^{00}\int d^s x_1 \; f(x_1) \; \{\partial_{j1}((g_1^{00})^{-1} g_1^{j0} D_1 G)$$

$$+ \text{ terms in } G \text{ and } \partial_{j1} G\}.$$

But at $t_2 = t_1$ this is

$$A_2 = i\sqrt{|g_2|}\,g_2^{00} \int d^s x_1\, (g_1^{00})^{-1}\, g_1^{j0}\, [i\delta(x_2 - x_1)]\, \partial_j f(x_1)$$

$$= -\sqrt{|g_2|}\,g_2^{0j}\, \partial_j f(x_2) = -A_3 .$$

So $A = 0$.

In the rigorous theory of hyperbolic partial differential equations on a manifold (surveyed by Choquet-Bruhat (1967)) it is proved that if the manifold is globally hyperbolic (cf. Sec. III.5), then the equation

$$L_2 G_{\text{inhom}}(x_2, x_1) = \delta(x_2, x_1) \tag{4.7}$$

(where

$$\int d^n x_1 \sqrt{|g_1|}\,\delta(x_2, x_1)\, f(x_1) = f(x_2) \tag{4.8}$$

-- i.e., δ is a scalar distribution) has a unique solution $G^{ret}(x_2, x_1)$ such that its support in x_2 is "compact toward the past" -- and lies, in fact, inside the future light cone of x_1. This is the <u>retarded Green function</u>. The <u>advanced</u> Green function satisfies Eq. (4.7) and has support in the past light cone of x_1. The (unique) solution of

$$LF(x) = v(x) \tag{4.9}$$

which has support in the future of the support of v (that is, represents outgoing radiation from the source v) is

$$F^{ret}(x_2) = \int d^n x_1 \sqrt{|g_1|}\, G^{ret}(x_2, x_1)\, v(x_1). \qquad (4.10)$$

(Similarly, F^{adv} has support in the past.) G^{ret} is <u>regular</u>: if v is a smooth function, then so is F^{ret}. It can be shown from this that F^{ret} is defined (as a distribution) even if v is a distribution. (Proving this involves interchanging the roles of x_1 and x_2.)

Since the Green functions both satisfy Eq. (4.7),

$$G'' = G^{adv} - G^{ret}$$

is a solution of the homogeneous wave equation (4.2). We shall show that G" = G. The general solution of LF = 0 (which we already have in the form (4.6)) can be written

$$F(x_2) = -\int d^n x_1 \sqrt{|g_1|}\, G''(x_2, x_1)\, v(x_1), \qquad (4.11)$$

where v can be chosen to have support on S (Fourès-Bruhat (1960)). Namely, let

$$v(x_1) = L\,\{\theta(t_1 - t) F(x_1)\} = L\,\{-\theta(t - t_1) F(x_1)\} \qquad (4.12)$$

(which is well-defined since F does not have a singularity in t_1 on S). Then by Eq. (4.10) the right-hand side of Eq. (4.11) is $F^{ret}(x_2) - F^{adv}(x_2)$, where

$$LF^{ret} = LF^{adv} = v$$

and

$$F^{ret[adv]}(x_2) = 0 \text{ if } t_2 < t \; [> t].$$

By the uniqueness theorem, F^{ret} and F^{adv} are respectively the functions in braces in Eq. (4.12); Eq. (4.11) follows immediately. Now from Eq. (4.12) we calculate

$$\sqrt{|g|}\,v(x_1) = \sqrt{|g|}\,m^2\,\theta F + \partial_\mu[g^{\mu 0}\sqrt{|g|}\,\partial_0(\theta F)]$$

$$+ \partial_0[g^{0k}\sqrt{|g|}\,\partial_k(\theta F)] + \partial_j[g^{jk}\sqrt{|g|}\,\partial_k(\theta F)]$$

$$= \theta LF + \partial_\mu[\delta(t_1 - t)\,g^{\mu 0}\sqrt{|g|}\,F] + \delta(t_1 - t)[g^{0\nu}\sqrt{|g|}\,\partial_\nu F]$$

$$= 0 + \sqrt{|g|}\,f_*(x_1)$$

(see Eq. (4.6b)). In particular, this calculation shows that

$$-L_2\{\theta(t_2 - t_1)G(x_2, x_1)\} =$$

$$\frac{1}{\sqrt{|g|}}\,\delta(t_2 - t_1)\,\delta(x_2 - x_1) = \delta(x_2, x_1);$$

since the distribution in the braces has no support in the past, it must be $G^{ret}(x_2, x_1)$.

Thus we have recovered from the formula (4.11) the formulas (4.5-6), with the identifications

$$G(x_2, x_1) = G^{adv}(x_2, x_1) - G^{ret}(x_2, x_1), \tag{4.13}$$

$$G^{ret}(x_2, x_1) = -\theta(t_2 - t_1)\, G(x_2, x_1),$$

$$G^{adv}(x_2, x_1) = \theta(t_1 - t_2)\, G(x_2, x_1) = G^{ret}(x_1, x_2). \tag{4.14}$$

It follows that (in the distribution sense, of course)

$$G(x_2, x_1) = 0 \text{ unless } x_2 \text{ and } x_1 \text{ are causally connected} \tag{4.15}$$

(see Sec. III.5). (In the case of the free field this follows immediately from Eqs. (4.3) by Lorentz invariance.)

Given any state vectors Ψ_1 and Ψ_2, let us define

$$G^{(+)}(\Psi_2, \Psi_1; x_2, x_1) = \langle \Psi_2 | \phi(x_2) \phi(x_1) | \Psi_1 \rangle, \tag{4.16a}$$

$$G^{(-)}(\Psi_2, \Psi_1; x_2, x_1) = \langle \Psi_2 | \phi(x_1) \phi(x_2) | \Psi_1 \rangle, \tag{4.16b}$$

$$G^{(1)}(\Psi_2, \Psi_1; x_2, x_1) = G^{(+)} + G^{(-)}, \tag{4.17}$$

$$G_F(\Psi_2,\Psi_1;x_2,x_1) = \langle\Psi_2|T\{\phi(x_2)\phi(x_1)\}|\Psi_1\rangle, \qquad (4.18)$$

where[9]

$$T\{\phi(x_2)\phi(x_1)\} = \theta(t_2-t_1)\phi(x_2)\phi(x_1) + \theta(t_1-t_2)\phi(x_1)\phi(x_2). \quad (4.19)$$

$G^{(\pm)}$ and $G^{(1)}$ are solutions of the homogeneous equation (4.2); moreover,

$$iG = G^{(+)} - G^{(-)}. \qquad (4.20)$$

$G^{(1)}$ is symmetric in x_1 and x_2. G_F is analogous to the Feynman propagator; iG_F satisfies the inhomogeneous equation (4.7), and

$$G_F = - iG^{ret} + G^{(-)} = - iG^{adv} + G^{(+)}. \qquad (4.21)$$

In the case of the free scalar field in flat space one traditionally chooses $\Psi_2 = \Psi_1 = \Psi_0$, the vacuum. Then the distributions are all functions only of $x_2 - x_1$, because of translation invariance. In this case the relation between our notation and the most widely used one ([Bjorken-Drell 2], pp. 387-390) is

[9] There should be no trouble in defining the distribution products in Eq. (4.19), since, in analogy with the free field, one expects fields smeared in space at sharp time to make sense as operators. See the discussion of the Fock representation in Sec. VIII.3.

$$G(x_2, x_1) = \Delta(x_2 - x_1),$$

$$G^{(\pm)}(\Psi_0, \Psi_0; x_2, x_1) = \Delta_{\pm}(x_2 - x_1) \qquad (G^{(1)} \text{ similar}),$$

$$G^{\text{ret[adv]}}(x_2, x_1) = \Delta_{\text{ret[adv]}}(x_2 - x_1), \tag{4.22}$$

$$G_F(\Psi_0, \Psi_0; x_2, x_1) = i\Delta_F(x_2 - x_1).$$

The important point is that G, G^{ret}, and G^{adv} are unique and intrinsic (determined by the manifold and the wave equation), while $G^{(\pm)}$, $G^{(1)}$, and G_F apparently must be defined in terms of particular states in a quantum theory. Alternatively, one would need some way of splitting G into "positive-frequency" and "negative-frequency" parts (see Eq. (4.20)).

There is one qualification to the statement that G is unique. Time reversal changes the sign of the canonical momentum, and hence (through Eq. (4.3b)) the sign of G. (This is true even for flat space, of course.) The same conclusion can be drawn from the alternative definition (4.13), since the meaning of "ret" and "adv" depends on the direction of time. Note that the uniqueness theorem and the reality of Eq. (4.7) imply that G^{ret} and G^{adv} are real. G is therefore real.

5. The Current Form.

The generalization of the Lagrangian (1.1) to complex ϕ is

$$\mathcal{L} = \sqrt{|g|}\,[g^{\mu\nu}\,\partial_\mu \phi^* \,\partial_\nu \phi - m^2\,\phi^*\,\phi]. \qquad (5.1)$$

It is invariant under the gauge transformation $\phi \longrightarrow \exp(i\gamma)\phi$. The corresponding conserved quantity (obtained through Noether's theorem -- see Hill (1951)) is

$$W(\phi,\phi) = i\int_S d^S x\,[\phi^*\,D\phi - D\phi^*\,\phi]. \qquad (5.2)$$

$W(\phi,\phi)$ can be generalized to

$$W(\phi_1,\phi_2) = i\int_S d^S x\,[\phi^*_1(x)\;D\phi_2(x) - D\phi^*_1(x)\;\phi_2(x)]. \qquad (5.3)$$

For solutions ϕ_1 and ϕ_2 of Eq. (1.2) whose initial values are sufficiently integrable for $W(\phi_1,\phi_2)$ to be defined, it is an indefinite Hermitian (i.e., sesquilinear and conjugate-symmetric) form. It is independent of the hypersurface S (see Sec. VII.3), since the divergence of

$$j^\mu(\phi_1,\phi_2) = ig^{\mu\nu}\,[\phi^*_1\,\partial_\nu\phi_2 - \partial_\nu\phi^*_1\,\phi_2] \qquad (5.4)$$

is easily seen to be zero by virtue of the wave equation. Since the initial values ϕ and $D\phi$ can be chosen independently, W is not degenerate: there is no ϕ' such that $W(\phi,\phi') = 0$ for all ϕ.

If ϕ is a complex quantized field, $W(\phi,\phi)$ can be interpreted as the total charge of the system. For the Hermitian field we are studying, $W(\phi,\phi) = 0$. Hence we shall be interested in W only as a bilinear form defined on complex-valued numerical solutions of the wave equation.

Unlike the case of a positive definite form, there is no unique maximal space of functions on which W is defined. Since there is no Schwarz inequality, it does not follow from $W(\phi_1,\phi_1) < \infty$ and $W(\phi_2,\phi_2) < \infty$ that $W(\phi_1,\phi_2)$ makes sense. Thus the vague phrase "sufficiently integrable" covers a real ambiguity. One obvious possibility is to consider only functions for which both the initial values, $\phi(x)$ and $D\phi(x)$, are L^2 functions. Then the integrals in Eq. (5.3) always converge. However, the case of the free field in flat space shows that this is not necessarily the natural space to consider. In this case (see Segal and Goodman (1965), p. 636) $W(\phi_1,\phi_2)$ is defined for positive-frequency solutions with initial values $\phi(x)$ in the domain of the operator

$$C = [m^2 - \nabla^2]^{1/4} \tag{5.5}$$

(defined via the Fourier transform). Then

$$D\phi(x) = -iC^2\phi(x) \tag{5.6}$$

is a function in the completion of L^2 with respect to the norm

$$\|f\|_C = \|C^{-1} f\|_2 .$$

$W(\phi_1,\phi_2)$ is defined on the direct sum of the positive- and the negative-frequency solutions,

$$\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_- , \tag{5.7}$$

which satisfies Eqs. (V.4.10). The restriction on $D\phi(x)$ is weaker than square-integrability, but the restriction on $\phi(x)$ is (necessarily) stronger. On the other hand, one could define W on a space of the form (5.7) with $\mathcal{H}_+$ defined by the condition

$$D\varphi(x) = -i\phi(x)$$

instead of Eq. (5.6); then all the initial values would be square-integrable.

Generalizations of the construction (5.5-7) are carried out (and applied to field quantization) for de Sitter space in Chapter V, for static universes in Chapter VIII, and for generalized Robertson-Walker universes in Chapter X.

6. Conclusions.

In this chapter we have established the following for a neutral scalar field in an external gravitational potential (in other words, on a given space-time manifold):

(1) The classical Lagrangian-Hamiltonian formalism can be applied in a manifestly covariant way (except that one coordinate is required to be timelike and the direction of time is significant).

(2) Canonical quantization is formally consistent with the equations of motion.

(3) "Smeared" fields can be defined covariantly, either in space-time or on a spacelike hypersurface.

(4) The commutator of a quantized field can be related to the Green functions of the wave equation and to the general solution of the Cauchy problem. The generalization to curved space of the other distribution solutions associated with the Klein-Gordon equation is not unique in the absence of a definition of the vacuum state, or of "positive frequency".

(5) The expression for the conserved current in a theory of a complex field yields a covariant Hermitian form on solutions of the wave equation with sufficiently integrable initial values. This form is positive on some solutions and negative on others, but this fact does not by itself lead to a unique characterization of positive-frequency solutions. In fact, different notions of positive frequency can sometimes lead to different maximal vector spaces on which the form can be defined.

In the standard treatment of the free scalar field in Minkowski space (e.g., [Bjorken-Drell 2], Chaps. 11 and 12) the next step in quantization is the solution of the wave equation by separation of variables and the association of creation and annihilation operators with the Fourier components of ϕ and Π. This leads to the rigorous construction of the fields as operator-valued distributions in Fock space. In Chapter VIII we shall show that this entire procedure goes through for the case of a static gravitational field; we have an ambiguity, however, if the metric has a static form in more than one coordinate system (Chapter IX). In Chapter X we attack the nonstatic case and encounter a more complicated situation.

7. The Axiomatic Approach.

In Chapter IV an attempt was made to adapt the general principles of quantum field theory (Appendix E) to de Sitter space. Here we shall briefly discuss to what extent this can be done for an arbitrary Riemannian space-time. This approach is logically independent of the rest of the chapter. The axioms stated should apply to self- or mutually interacting fields in curved space as well as to the "free" field described in the rest of the chapter.

The word "free" is placed in quotation marks here because the deviation from the ordinary special relativistic free field equations due to the nonconstant metric coefficients is every bit as drastic as that due to, say, an interaction with an applied electromagnetic potential. One is tempted to describe the field on a Riemannian manifold as a field coupled to an external gravitational field, and thus to absorb it into the general class of external potential problems. There is a distinction, of course: In special relativistic external potential problems one thinks of the external force as something superimposed on a fundamental flat space; there are distinguished systems of Cartesian coordinates. In the gravitational case, because of the universality of the gravitational interaction (the principle of equivalence), it is operationally impossible to define distinguished global coordinate systems, or to split the tensor which appears in the equations of motion uniquely into a space-time metric and an "external" field. Also, the topology of the space may be different from that of Minkowski space. These added complications do not affect the point that the metric of a curved space, or even of flat space treated in curvilinear coordinates, enters the dynamics of a field as an external potential. We shall therefore broaden our discussion to include external potentials in general.

We may define an external potential interaction, as opposed to a self-interaction or mutual interaction of quantized fields, as any interaction described by a quadratic term in the Lagrangian or Hamiltonian, and hence by a linear term in the equations of motion of the quantum fields. (We could also consider "external source" problems, where the interaction term is linear in the Hamiltonian and constant in the equations.)

These include external electromagnetic (four-vector) potentials and many other types of less relevance to nature, such as the scalar potential introduced in Eq. (V.5.6).

Field theories involving only external potentials (no true field interactions) are much less pathological mathematically than interacting fields. Nevertheless, they present a problem from the point of view of framing general principles like those in Appendix E. The reason is that nontrivial external potential interactions are in general not Lorentz-invariant. For instance, an applied electromagnetic field must point in some direction, and the field strength may vary from point to point in space and time. Under these circumstances it is not possible to make the assumption of Poincaré invariance (Axiom 2) which is normally imposed on relativistic field theories. It follows that all the axioms which fall below Axiom 2 in the graph of Fig. 20 must at least suffer re-examination; in fact, most of them become meaningless.

There is less difficulty with the "field" axioms. In the case of an external potential in Minkowski space the meaning of Axioms 3, 5, and 7 is clear, and there is no visible reason not to keep them. In Riemannian space some modifications are necessary along the lines indicated in Sec. IV.1. Local commutativity, in particular, must be defined with respect to the causal structure of the space (see Sec. III.5); with some risk of oversimplification, we can say that the light cone at a point divides those points which are causally related to it from those that are not. Finally, note that the various kinds of tensor and spinor fields can be defined even though there is no longer a relation with a group representation of the type described by Axiom 4; in accordance with the ground rule laid down in Sec. IV.1, we shall not discuss them.

The main problem, then, is to find reformulations of or substitutes for the spectral condition, the vacuum axioms, and the axiom of asymptotic completeness, or to get along without them. The extent to which this can be done depends on the properties of the potential. Therefore, we shall discuss some special cases.

If the potential, along with any other interactions in the model, is independent of time, almost everything can be recovered. Axiom 2 can be restated, with ISL(2,$\mathbb{C}$) replaced by just the time translation group, or possibly some larger symmetry group containing it. Then there is a self-adjoint generator of the time translations, H. One can require that the spectrum of H be bounded below (substitute for Axiom 9), that its ground state be discrete and nondegenerate (for Axiom 6), and that this state be cyclic for the fields (Axiom 8). If the potential does not fall off to zero at spatial infinity, however, asymptotic completeness has no clear meaning, and there may be a difficulty in giving the theory a particle interpretation. The static situation will be discussed in the gravitational context in Chapters VIII and IX.

The next best case seems to be that of a potential which vanishes, or at least becomes static, asymptotically in time. Then during the periods when the interaction is "turned off" the states of the system can be classified in terms of the appropriate free, or static, Hamiltonian. Thus the conditions of positivity of the energy, etc., can be imposed at each end. The only trouble is that the equations of motion may imply that the field operator defined in this way at early times (*in*-representation), evaluated at late times, is not unitarily equivalent to the field operator defined according to the axioms at late times. Physically, one can say in such a case that infinitely many particles are produced by the interaction. (Note that even it this does not happen, the *in*- and *out*-vacuums will not usually be the same state -- that is, there is some probability for the creation of finitely many particles.) This sort of situation will be discussed further in Secs. X.3 and X.7.

In the general case it is not at all obvious that any remnant of the spectral condition, etc., can be salvaged. Perhaps one can define a spectral condition at each time by the method sketched in Sec. IV.2; but, for reasons similar to those just discussed (for the asymptotically static case), one would not expect this procedure applied at different times to yield the same representation of the fields.[10]

[10] If the potential is quite smooth and falls off rapidly in both space and time, a satisfactory theory (at least for spins 0 and 1/2) has been developed, in which the *in*- and *out*-representations are equivalent. See Capri (1967, 1969),

Why have these difficulties not been considered more thoroughly in the literature of quantum field theory in connection with external potential problems? The most important reason seems to be that in "laboratory" or "terrestrial" applications of field theory one can always assume that the external field is finite in spatial extent and either static or asymptotically vanishing in time. (Even when the expression which is explicitly written down does not have these properties, it is argued that it is an idealized approximation to something which does.) Hence the states of the system have an asymptotic particle structure which enables one to recover most of the consequences of positive energy. In particular, the theory has a physical interpretation in terms of particles. Another reason for the lack of attention to the problems which arise in singular external potential problems is the feeling that whatever pathologies appear are a punishment for treating the external field classically instead of as a quantized field in its own right; they are expected to disappear in the complete, Lorentz-invariant theory of the future.

In gravitational problems on the astrophysical or cosmological scale, however, one cannot take these ways out. (With respect to the first point, see the remarks of Secs. IV.2 and VIII.4. The suggestion that a coherent treatment of elementary particle processes in a gravitational background must

---

Capri <u>et al</u>. (1971), Schroer <u>et al</u>. (1970), Wightman (1968, 1971), Seiler (1972).

await success in the conceptually murky project of "quantizing the gravitational field" is unwelcome, to say the least.) It seems quite probable, therefore, that the growing interest in applications of quantum field theory to astrophysics and cosmology may force quantum field theory to confront certain fundamental issues which up to now have largely been evaded.

Chapter VIII

QUANTIZATION IN A STATIC GRAVITATIONAL FIELD

In this chapter we assume that there exists a coordinate system in which the metric is static (see Appendix D).[1] That is, all the components $g_{\mu\nu}$ are independent of t (= x^o), and $g_{oj} = 0$ for $j \neq 0$. Topologically, space-time is $I \times M$, where $I \subset \mathbb{R}$ is the time axis and M is a manifold covered by the s spacelike coordinates. (In this chapter the letter x will stand for just the spacelike coordinates of a point.)

Although what is done in this chapter is a very straightforward generalization of the familiar quantization of the free scalar field in Minkowski space, it does not appear to have been written out in much detail before. Special cases have been treated briefly, of course. For instance, Bonazzola and Pacini (1966) quantized a scalar field in the general spherically symmetric static background metric in preparation for a self-consistent treatment of a system of many particles in their own strong gravitational field; this theory and the analogous one for fermions were applied in neutron star calculations by Ruffini and Bonazzola (1969). Also, since we shall find in Secs. VIII.3-4 that the field theory constructed here is equivalent

---

[1] The static coordinates may cover only part of space-time, in which case we temporarily forget about the rest. See Sec. VIII.6 and Chapter IX.

(under most circumstances) to a single-particle wave mechanics, all attempts to do relativistic quantum mechanics in a static gravitational field can be considered instances of this theory or its higher-spin generalizations. (For example, Unruh (1971) has studied the Dirac equation in the Schwarzschild metric.)

In short, the theory presented here seems to be equivalent to what anyone would naturally do, and some theorists have done, when confronted with the problem of describing matter in a static gravitational field quantum-theoretically. This circumstance lends interest to the fact which will be established in the next two chapters -- namely, that this quantization is not unique. In fact, in Sec. X.8 the present author will suggest a method of quantization which does not agree, in general, with this one. To what extent the ambiguities affect observable quantities has not been fully determined. (See Secs. IX.4-5, IX.7, and X.7-8.)

1. Solution of the Wave Equation by Separation of Variables.

In the static case the wave equation (VII.1.2,3) is

$$-\partial_0^2 \phi = \frac{g_{00}}{\sqrt{|g|}}\, \partial_j [\sqrt{|g|}\, g^{jk}\, \partial_k \phi] + g_{00} m^2 \phi \equiv K\phi. \qquad (1.1)$$

The ansatz

$$\phi(t,x) = \Psi_j(x)\, e^{-iE_j t} \qquad (1.2)$$

leads to

$$K\Psi_j(x) = E_j^2 \Psi_j(x). \tag{1.3}$$

The operator K is Hermitian in the scalar product[2]

$$(F_1, F_2) = \int d^s x \sqrt{|g|}\, g^{00} F_1^*(x) F_2(x). \tag{1.4}$$

It is also positive:

$$(F_1, KF_1) = \int dx\, F_1^*(x)\, \partial_j [\sqrt{|g|}\, g^{jk} \partial_k F_1(x)] + \int dx \sqrt{|g|}\, m^2 |F_1(x)|^2$$

$$\geq + \int dx \sqrt{|g|}\; {}^{(s)}g^{jk}\, \partial_j F_1^*(x)\, \partial_k F_1(x) \geq 0$$

since $\{{}^{(s)}g^{jk}\} = \{-g^{jk}\}$ is a positive definite matrix. The operator therefore has self-adjoint extensions (see [Reed-Simon], Sec. 8.6, and [Kato], Secs. VI.1-2).

If the Cauchy problem is well-posed in this coordinate system in the sense of Sec. III.5 (i.e., each hypersurface t = const. is a Cauchy surface for the region of space-time covered by the coordinates), one expects on physical grounds that K (with some obvious boundary conditions, if necessary) will be essentially self-adjoint, since particles cannot leak in or out

---

[2] Note that in Secs. V.7-8 coordinates were tacitly chosen so that $\sqrt{|g|}\, g^{00} = 1$.

of the space. This can be verified in particular cases. ("Obvious boundary conditions" refers to those which are necessitated by the topology of the manifold M, such as periodicity in an angular coordinate. In contrast, in the open de Sitter space (Secs III.6 and V.8) the Cauchy problem is not well-posed, and for $q > -3/2$ a boundary condition which is not physically obvious is needed.)

From now on we assume that K (on a suitable domain) is a self-adjoint operator in the Hilbert space $\mathcal{H}$ of functions of s variables with the norm

$$\|F\|^2 = \int d^s x \sqrt{|g|}\, g^{00}\, |F(x)|^2 .$$

The spectral representation of K gives a unitary correspondence, analogous to the Fourier transform, between $\mathcal{H}$ and another Hilbert space $L^2(\mu)$ of functions $\tilde{f}(j)$, in terms of which K is "diagonal": $K\tilde{f}(j) = E_j^2 \tilde{f}(j)$ (see, e.g., [Reed-Simon], Chaps. 7 and 8.)

For convenience we shall assume (as usual in elementary quantum mechanics) that the numbers in the spectrum can be classified as <u>point spectrum</u> σ_p or <u>continuous spectrum</u> σ_c (or both), and that a corresponding complete set of generalized eigenfunctions exists.[3] That is, an arbitrary function in $\mathcal{H}$

---

[3] These assertions have been established only for certain classes of differential operators. (For instance, Ikebe (1960) has treated the ordinary Schrödinger equation with a potential vanishing at infinity.) In the general case the eigenfunction notation should be regarded as formal shorthand for a rigorous

can be expanded as

$$F(x) = \int d\mu(j)\ \tilde{f}(j)\Psi_j(x), \tag{1.5}$$

where $\Psi_j(x)$ are the solutions (not square-integrable if E_j^2 is in the continuous spectrum) of the eigenvalue equation (1.3). Here μ is the measure which defines the scalar product in $L^2(\mu)$:

$$\|\tilde{f}\|^2 = \int d\mu(j)\,|\tilde{f}(j)|^2 . \tag{1.6}$$

For instance, if the $\Psi_j(x)$ are suitably normalized, $\int d\mu(j)$ means

$$\sum_{j\in\sigma_p} + \int_{\sigma_c} dj \ .$$

In any case, for consistency of Eqs. (1.4) and (1.6) we must choose the normalization so that

$$\int d^s x\ \sqrt{|g|}\,g^{00}\ \Psi^*_j(x)\Psi_k(x) = \delta(j,k), \tag{1.7}$$

where $\int d\mu(k)\,\delta(j,k)\,\tilde{f}(k) = \tilde{f}(j)$. Then we have the inversion formula

$$\tilde{f}(j) = \int d^s x\ \sqrt{|g|}\,g^{00}\ \Psi^*_j(x)F(x) \tag{1.8}$$

and the completeness relation

---

treatment in terms of spectral projections, or perhaps of rigged Hilbert spaces.

$$\int d\mu(j)\, \Psi^*_j(x) \Psi_j(y) = [\sqrt{|g|}\, g^{00}]^{-1}\, \delta(x - y). \qquad (1.9)$$

(Eqs. (1.7) and (1.9) just say that the mappings (1.5) and (1.8) are inverse to each other.)

The general solution of the wave equation (1.1), expanded in terms of the eigenfunctions (1.2), is

$$\phi(t,x) = \int d\mu(j) [Q^-_j \Psi_j(x) e^{-iE_j t} + Q^+_j \Psi_j(x) e^{iE_j t}]$$

$$+ Q^{(1)}_0 \Psi_0(x) + Q^{(2)}_0 \Psi_0(x) t. \qquad (1.10)$$

The last two terms occur only if $E_j^2 = 0$ is in the point spectrum of K. (To save writing it is assumed that this eigenvalue is not degenerate.) This happens, for example, in the case of the free massless scalar field quantized in a finite "box" with periodic boundary conditions. Since K is a positive operator, E_j^2 is always nonnegative. We take $E_j \geq 0$ by definition.

2. Creation and Annihilation Operators.

Without loss of generality we may choose the $\Psi_j(x)$ real. In some contexts it may be convenient to choose complex basis functions (momentum eigenfunctions, for instance); they will be considered at the end of this section.

The general solution (1.10) is determined by the Cauchy initial data $\phi(0,x)$ and $\Pi(0,x)$, which we can expand in eigenfunctions:

$$\phi(0,x) = \int d\mu(j)\ q_j\ \psi_j(x), \tag{2.1a}$$

$$\pi(0,x) \equiv \sqrt{|g|}g^{00}\ \dot{\phi}(0,x) = \sqrt{|g|}g^{00} \int d\mu(j)\ p_j\ \psi_j(x). \tag{2.1b}$$

Then

$$q_j = \int dx\ \sqrt{|g|}g^{00}\ \psi^*_j(x)\phi(0,x), \tag{2.2a}$$

$$p_j = \int dx\ \psi^*_j(x)\pi(0,x). \tag{2.2b}$$

(As written these formulas apply also to complex ψ_j.) If ϕ and π are Hermitian quantum fields satisfying the CCRs (VII.2.1), then q_j and p_j are Hermitian (for ψ_j real!), and

$$[q_j, p_k] = \int dx \int dy\ \sqrt{|g|}g^{00}\ \psi_j(x)\psi_k(y)\ i\delta(x - y) = i\delta(j,k).$$

(Here and in what follows we use without comment the elementary formulas recorded in the preceding section.)

Comparing Eqs. (2.1) with Eq. (1.10) and its time derivative at t = 0, we find

$$Q_j^- = \frac{1}{2} [q_j + i\, p_j / E_j],$$

$$Q_j^+ = \frac{1}{2} [q_j - i\, p_j / E_j] = (Q_j^-)^\dagger,$$

and

$$Q_0^{(1)} = q_0, \quad Q_0^{(2)} = p_0, \tag{2.3}$$

and hence, for $E_j > 0$,

$$[Q_j^-, Q_k^+] = (2E_j)^{-1} \delta(j,k), \quad [Q_j^-, Q_k^-] = 0.$$

This suggests that for $E_j > 0$ we set

$$a_j = \sqrt{2E_j}\, Q_j^-, \quad a_j^\dagger = \sqrt{2E_j}\, Q_j^+, \tag{2.4}$$

in analogy to the familiar quantization of the free field of mass m. Then

$$[a_j, a_k] = 0, \quad [a_j, a_k^\dagger] = \delta(j,k). \tag{2.5}$$

Substituting Eqs. (2.1) into Eq. (VII.1.8b), we have, after integration by parts and use of Eq. (1.3),

$$H = \frac{1}{2}\int d\mu(j)\,(p_j^2 + E_j^2 q_j^2) = \int du(j)\, H_j\,. \tag{2.6}$$

The contribution of each mode with $E_j > 0$ to the Hamiltonian is

$$H_j = \frac{1}{2}(p_j^2 + E_j^2 q_j^2) = E_j(a_j^\dagger a_j + \frac{1}{2})\,. \tag{2.7a}$$

In order for the total energy H to converge in the Fock representation to be constructed in the next section, we must discard the constant term $E_j/2$ by normal ordering with respect to a_j and $a_j^\dagger$. Then

$$H_j = E_j a_j^\dagger a_j \equiv E_j N_j\,. \tag{2.7b}$$

In any representation the operator of Eq. (2.7b) has the discrete spectrum 0, E_j, $2E_j$, ... , nE_j, ... Therefore N_j is regarded as the number operator for quanta of the type j, each of which carries energy E_j.

When $E_j = E_0 = 0$, $H_0 = p_0^2/2$ has continuous spectrum from 0 to $+\infty$ with multiplicity two. To obtain a complete set of commuting operators, H_0 can be supplemented by sgn p_0; this is equivalent to using the spectral representation of p_0. Alternatively one might use H_0 and σ, the parity under change of sign of p_0 and q_0 (equivalently, of $\phi(x)$). The physical interpretation of this spectrum will be discussed in Sec. VIII.5.

Since K is positive, we do not have to worry about "jelly modes" with $E_j^2 < 0$ (Schiff <u>et al</u>. (1940); Schroer and

Swieca (1970); Schroer (1971)). In external potential problems these give contributions $H_j = \frac{1}{2}(p_j^2 - |E_j|^2 q_j^2)$, which are not bounded below, to the energy, and exponential terms $\Psi_j(x)\exp(\pm|E_j|t)$ to the field. Their physical interpretation is obscure. It is conceivable that this complication might arise in the conformally invariant theory (see Sec. VII.1) for sufficiently large negative curvature.

Let us now consider the possibility of complex eigenfunctions. The complex conjugates $\Psi_j^*(x)$, like the $\Psi_j(x)$, make up a complete set of generalized eigenvectors. In general Ψ_j and Ψ_k^* are not orthogonal, but they are if they correspond to different eigenvalues ($E_j \neq E_k$). (For the free field, where j is the momentum vector $\vec{k}$, $\Psi_{\vec{k}}^*(x)$ is $(2\pi)^{-3/2}\exp(-i\vec{k}\cdot x) = \Psi_{-\vec{k}}(x)$.) In Sec. VIII.4 scattering states will be defined for field theories in static space-times which are asymptotically flat; in this case $\Psi_{\vec{k}}^{in}*(x) = \Psi_{-\vec{k}}^{out}(x)$, and

$$\int d^s x\sqrt{|g|}g^{00}\,\Psi_{\vec{k}}^{in}(x)\Psi_{\vec{l}}^{in}(x) = (\Psi_{\vec{k}}^{in}*,\Psi_{\vec{l}}^{in}) = S(-\vec{k},\vec{l}), \qquad (2.8)$$

an element of the S-matrix.)

If complex basis functions are allowed and the terms for $E_j = 0$ are dropped, Eq. (1.10) (with the definitions (2.4) and (2.3)) generalizes to

$$\phi(t,x) = \int \frac{d\mu(j)}{\sqrt{2E_j}} [\psi_j(x) e^{-iE_j t} a_j + \psi_j^*(x) e^{iE_j t} a_j^\dagger]. \qquad (2.9)$$

The Hermiticity of ϕ forces the occurrence of ψ_j^* in the second term. It is clear that the modes will not decouple as neatly as in the treatment above for real eigenfunctions, but it is equally clear that this complication is merely an inessential notational nuisance. The canonical momentum is now

$$\pi(t,x) =$$

$$- i\sqrt{|g|}g^{00} \int d\mu(j) \sqrt{\frac{E_j}{2}} [\psi_j(x) e^{-iE_j t} a_j - \psi_j^*(x) e^{iE_j t} a_j^\dagger]. \qquad (2.10)$$

The last two equations are easily inverted to yield

$$a_j = \frac{1}{\sqrt{2}}(\sqrt{E_j}\, q_j + i\, p_j/\sqrt{E_j})$$

$$= \frac{1}{\sqrt{2}}[\sqrt{E_j} \int dx \sqrt{|g|}g^{00}\, \psi_j^*(x)\phi(0,x) + \frac{i}{\sqrt{E_j}} \int dx\, \psi_j^*(x)\pi(0,x)]; \qquad (2.11)$$

$a_j^\dagger$ is given by the adjoint of this expression. A byproduct of this calculation is the observation that the creation operator corresponding to the basis $\{\psi_j^*\}$ is

$$\bar{a}_j^\dagger = \frac{1}{\sqrt{2}}(\sqrt{E_j}\, q_j - i\, p_j/\sqrt{E_j}) = \int d\mu(k)\, (\psi_k, \psi_j^*)\, a_k^\dagger. \qquad (2.12)$$

The generalization of Eqs. (2.7) (with normal ordering) is

$$H_j = \frac{1}{2}:(p_j^\dagger p_j + E_j^2 q_j^\dagger q_j): = E_j a_j^\dagger a_j . \tag{2.13}$$

So

$$H = \int d\mu(j)\; E_j a_j^\dagger a_j . \tag{2.14}$$

A generalization of Eq. (1.9) is the fact that

$$\int_{E_j=E} d\mu(j)\, \Psi_j^*(x)\Psi_j(y) = \int_{E_j=E} d\mu(j)\, \Psi_j(x)\Psi_j^*(y)$$

is the kernel of the projection onto the space of vectors of eigenvalue E. It follows that

$$\int d\mu(j)\, \Psi_j(x)\Psi_j^*(y)\, A(E_j) = \int d\mu(j)\, \Psi_j^*(x)\Psi_j(y)\, A(E_j) \tag{2.15}$$

when A depends on j only through E_j. Eq. (2.15) is often useful in manipulating complex basis functions.

3. The Fock Representation.

In this section and the next we assume that the point spectrum does not contain E = 0.

The formulas of the preceding sections do not yet constitute a quantum theory in the common sense of the term, since the fields have not been realized as operators on a Hilbert space. It is well known that there are many inequivalent ways of doing this. (See, e.g., Wightman and Schweber (1955).) In the static case, however, as in the case of the free field, one representation stands out as a leading candidate for "correct" or "physical" representation. It is determined by the requirement that there be a cyclic vector (the vacuum or no-particle state) which is annihilated by all the operators a_j of Sec. VIII.3. ("Cyclic" means that all the vectors in the Hilbert space are limits of sums of vectors obtained by acting on the vacuum by products of the field operators. In other words, we choose the smallest space containing the vacuum consistent with the action of the fields.) This is the Fock representation. It has a particle interpretation (Sec. VIII.4).

Let $\mathcal{H}$ stand for the space $L^2(\mu)$ introduced in Sec. VIII.1. Let $\mathcal{H}^{\otimes n}$ be the Hilbert-space closure of the symmetrized n-fold tensor product of $\mathcal{H}$; its elements are the μ-square-integrable functions of n variables $j_1, \ldots, j_n$. An element of the Fock space $\mathcal{F}$ is a sequence

$$f = (\tilde{f}_0, \tilde{f}_1(j), \tilde{f}_2(j_1,j_2), \ldots, \tilde{f}_n(j_1,\ldots,j_n), \ldots) \equiv \{\tilde{f}_n\}$$

with $\tilde{f}_n \in \mathcal{H}^{\otimes n}$ and $\|f\|^2 = \sum_{n=0}^{\infty} \|\tilde{f}_n\|^2 < \infty$. Let $\mathcal{N}$ be the sequences in $\mathcal{F}$ with $\tilde{f}_n = 0$ for all n greater than some N. Let $\mathcal{T} \subset \mathcal{H}$ be the space of functions $\tilde{f}(j)$ which vanish whenever E_j is greater than some maximum value (functions of compact support in the energy), and let $\mathcal{T}^{\otimes n}$ be the symmetrized n-fold tensor product of this space (no closure implied). Let $\mathcal{D}$ be the sequences in $\mathcal{H}$ with each $\tilde{f}_n \in \mathcal{J}^{\otimes n}$.[4]

We are ready to construct the operators of the representation.

Creation and annihilation operators. For $f = \{\tilde{f}_n(j_1,\dots,j_n)\} \in \mathcal{H}$ and $\mathfrak{g} \in \mathcal{U}$, let

$$a(\mathfrak{g})f = \{\sqrt{n+1}\int d\mu(j)\mathfrak{g}(j)\tilde{f}_{n+1}(j,j_1,\dots,j_n)\}, \tag{3.1a}$$

$$a^\dagger(\mathfrak{g})f = \{\sqrt{n}\ \mathrm{sym}[\mathfrak{g}(j_1)\tilde{f}_{n-1}(j_2,\dots,j_n)]\}. \tag{3.1b}$$

In the standard way one verifies that the adjoint of $a(\mathfrak{g})$ is (an extension of) $a^\dagger(\mathfrak{g}^*)$, and that

$$[a(\tilde{f}),\ a^\dagger(\mathfrak{g})] = (\tilde{f}^*,\mathfrak{g}), \qquad [a(\tilde{f}),\ a(\mathfrak{g})] = 0. \tag{3.2a}$$

Setting $a^{(\dagger)}(\mathfrak{g}) = \int d\mu(j)a_j^{(\dagger)}\mathfrak{g}(j)$, we have

$$[a_j,\ a_k^\dagger] = \delta(j,k), \quad \text{etc.}, \tag{3.2b}$$

$$a_k\{\tilde{f}_n(j_1,\dots,j_n)\} = \{\sqrt{n+1}\ \tilde{f}_{n+1}(k,j_1,\dots,j_n)\}, \tag{3.3a}$$

[4] The choice of $\mathcal{J}$ and $\mathcal{H}$ is arbitrary. Most of our results could be proved for larger domains; e.g., those built out of functions of fast decrease in E_j or sequences of fast decrease in n. On the other hand, one might in some contexts want to restrict the functions in $\mathcal{J}$ by some condition of smoothness in j.

$$a_k^\dagger\{\tilde{f}_n(j_1,\ldots,j_n)\} = \{\sqrt{n}\ \mathrm{sym}[\delta(k,j_1)\tilde{f}_{n-1}(j_2,\ldots,j_n)]\}, \quad (3.3b)$$

$$a_k|0\rangle = 0, \quad (3.4)$$

$$\{0,\ 0,\ \ldots,\ \tilde{f}_n(j_1,\ldots,j_n),\ 0,\ \ldots\} =$$

$$\frac{1}{\sqrt{n!}}\int d\mu(j_1)\ldots d\mu(j_n)\ \tilde{f}_n(j_1,\ldots,j_n)a_{j_1}^\dagger\ \ldots\ a_{j_n}^\dagger\ |0\rangle \quad (3.5)$$

($|0\rangle = \{1,\ 0,\ 0,\ \ldots\}$).

<u>Hamiltonian and number operator</u>. With these definitions the normal-ordered Hamiltonian (2.14) makes sense:

$$Hf = \{(E_{j_1} + \ldots + E_{j_n})\ \tilde{f}_n(j_1,\ldots,j_n)\}. \quad (3.6)$$

H is manifestly self-adjoint on $\mathcal{D}$. Similarly, a total number operator $N = \int d\mu(j)\,a_j^\dagger a_j$ is defined and is essentially self-adjoint on $\mathcal{H}$; $N\Psi = n\Psi$ if Ψ is the vector in Eq. (3.5).

<u>Fields</u>. Consider a function $F(x)$ and define its transform $\tilde{f}(j)$ by Eq. (1.8); let $\bar{f}(j)$ stand for the complex conjugate of the transform of $F^*(x)$. (Note that the map $\tilde{f} \longrightarrow \bar{f}$ is linear (not antilinear). If the basis functions $\Psi_j(x)$ are real, $\bar{f} = \tilde{f}$. If F is real, $\bar{f} = \tilde{f}^*$.) Now define

$$\phi(t,F) = \int d^s x \sqrt{|g|}\, g^{00}\, \phi(t,x) F(x)$$

$$= \int \frac{d\mu(j)}{\sqrt{2E_j}} [a_j \bar{\tilde{f}}(j) e^{-iE_j t} + a_j^\dagger \tilde{f}(j) e^{iE_j t}]$$

$$= a(\bar{\tilde{f}} e^{-iEt}/\sqrt{2E}) + a^\dagger(\tilde{f} e^{iEt}/\sqrt{2E}) \qquad (3.7)$$

and similarly

$$\Pi(t,F) = - i[a(\sqrt{\frac{E}{2}}\, \bar{\tilde{f}} e^{-iEt}) - a^\dagger(\sqrt{\frac{E}{2}}\, \tilde{f} e^{iEt})]. \qquad (3.8)$$

The field operators will not be defined (as unbounded operators in $\mathcal{F}$ with domain $\mathcal{N}$) unless

$$\frac{1}{\sqrt{E}} \tilde{f} \in \mathcal{H} \quad \text{for} \quad \phi(F), \qquad (3.9a)$$

$$\sqrt{E}\, \tilde{f} \in \mathcal{H} \quad \text{for} \quad \Pi(F). \qquad (3.9b)$$

(If the lower bound of the spectrum is not $E = 0$, the first of these conditions holds for all $\tilde{f} \in \mathcal{H}$.) If these restrictions are met, $\mathcal{N}$ is an invariant domain for the operators. $\mathcal{D}$ is an invariant domain if $\tilde{f} \in \mathcal{T}$. Note that $\tilde{f} \in \mathcal{T}$ implies Eq. (3.9b) (but not (3.9a)). If $\tilde{f}$ is in $\mathcal{T}$ and F is real we can show that $\phi(F)$ ($\equiv \phi(0,F)$) and $\Pi(F)$ are essentially self-adjoint on $\mathcal{D}$ and that the Weyl relations (exponentiated CCRs)

$$e^{i\phi(f)}\, e^{i\Pi(G)} = e^{-i(F,G)}\, e^{i\Pi(G)}\, e^{i\phi(F)} \tag{3.10}$$

are satisfied. The easiest method (just as for the free field) is to show that the n-particle vectors (3.5) are analytic vectors for $\phi(F)^2 + \Pi(G)^2$ and then to apply the theorems of Nelson (1959).

Time translation group. Let

$$U_t f = e^{+iHt} f = \{\exp(+i(E_{j_1}+\dots+E_{j_n})t)\, \tilde{f}_n(j_1,\dots,j_n)\}. \tag{3.11}$$

Then it is easy to verify that

$$U_t\, \phi(0,F)\, U_t^{-1} = \phi(t,F) \tag{3.12}$$

and similarly for $\Pi(t,F)$. Equivalently, commutation with the generator H yields rigorously the equations of motion discussed in Chapter VII. In the usual way one can pass from the Heisenberg to the Schrödinger picture with the propagator exp(-iHt).

The Fock representation is irreducible. This follows from the positivity of H just as for Lorentz-invariant field theories ([Streater-Wightman], p. 141).

Vacuum expectation values (n-point functions). Unlike the general case discussed in Sec. VII.4, the field theory in a static universe has an obvious distinguished state, the vacuum $|0\rangle$. We define (cf. Eq. (VII.4.16a))

$$G^{(+)}(t_2,x_2;t_1,x_1) = G^{(+)}(t_2-t_1,x_2,x_1) = \langle 0|\phi(t_2,x_2)\phi(t_1,x_1)|0\rangle$$

$$= \int \frac{d\mu(j)}{2E_j} \Psi_j(x_2)\Psi_j^*(x_1) \exp[-iE_j(t_2-t_1)]. \qquad (3.13)$$

It is easy to calculate, as for a free or generalized free field[5], the expectation value of an arbitrary number of field operators:

$$W^n(x_n,\ldots,x_1) = \langle 0|\phi(x_n) \ldots \phi(x_1)|0\rangle \qquad (x_n \equiv (t_n,x_n)); \qquad (3.14)$$

$$W^{2n+1}(x_{2n+1},\ldots,x_1) = 0, \qquad (3.15a)$$

$$W^{2n}(x_{2n},\ldots,x_1) = \sum G^{(+)}(x_{i(2n)},x_{i(2n-1)}) \ldots$$

$$\times G^{(+)}(x_{i(2)},x_{i(1)}), \qquad (3.15b)$$

where the sum is over all partitions of the indices $(1,\ldots,2n)$ into pairs with $i(2k) > i(2k-1)$. The field theory can be reconstructed from the vacuum expectation values in the usual way ([Streater-Wightman], Sec. 3.4).

<u>Configuration space and propagators</u>. One can identify the elements of the space $\mathcal{H}$ with the (normalizable)

[5] Greenberg (1961), pp. 161-163; [Streater-Wightman], p. 116.

positive-frequency solutions of the wave equation (1.1) through the formula

$$f(t,x) = \int \frac{d\mu(j)}{\sqrt{2E_j}} \Psi_j(x) e^{-iE_j t} \tilde{f}(j). \qquad (3.16)$$

We write

$$f(x) = f(0,x), \qquad \dot{f}(x) = \frac{\partial f}{\partial t}(0,x). \qquad (3.17)$$

The scalar product takes the form

$$(f_1, f_2) = \int d\mu(j) \tilde{f}^*_1(j) \tilde{f}_2(j)$$

$$= i \int d^s x \sqrt{|g|} g^{00} [f^*_1(x) \dot{f}_2(x) - (\dot{f}_1)^*(x) f_2(x)]$$

$$\equiv i \int d^s x \sqrt{|g|} g^{00} f^*_1(x) \overleftrightarrow{\partial}_0 f_2(x) = W(f_1, f_2), \qquad (3.18)$$

which is the covariant current form of Sec. VII.5. (In contrast, the transform (1.5) leads to the noncovariant scalar product (1.4).)

The solution (3.16) is uniquely determined by __either__ of its initial values (3.17). Indeed, inversion of Eq. (3.16) yields

$$f(t_2,x_2) = 2i\int d^s x_1 \sqrt{|g|}g^{00} G^{(+)}(t_2,x_2;t_1,x_1)\dot{f}(t_1,x_1)$$

$$= -2i\int d^s x_1 \sqrt{|g|}g^{00} \frac{\partial}{\partial t_1}G^{(+)}(t_2,x_2;t_1,x_1) f(t_1,x_1)$$

$$= i\int d^s x_1 \sqrt{|g|}g^{00} G^{(+)}(t_2,x_2;t_1,x_1) \overleftrightarrow{\partial_0} f(t_1,x_1). \quad (3.19)$$

For this reason $G^{(+)}$ could be called the <u>forward propagator</u>. Of course, in the last form of Eq. (3.19) f and $\dot{f}$ cannot be chosen independently; any positive-frequency solution is characterized by

$$\dot{f}(x) = -i\int d\mu(j)E_j \Psi_j(x) \int dy \sqrt{|g|}g^{00} \Psi_j^*(y) f(y). \quad (3.20)$$

In contrast, the full propagator of Sec. VII.4,

$$iG(t_2,x_2;t_1,x_1) = [\phi(t_2,x_2), \phi(t_1,x_1)]$$

$$= \int \frac{d\mu(j)}{2E_j} \Psi_j(x_2)\Psi_j^*(x_1) [e^{-iE_j(t_2-t_1)} - e^{iE_j(t_2-t_1)}], \quad (3.21)$$

gives a <u>general</u> solution in terms of a full (independent) set of Cauchy initial data:

$$f(t_2,x_2) = -\int d^s x_1 \sqrt{|g|}g^{00} G(t_2,x_2;t_1,x_1) \overleftrightarrow{\partial_0} f(t_1,x_1). \quad (3.22)$$

The commutator iG is determined by the CCRs and the wave equation, but the two-point vacuum expectation value $G^{(+)}$ depends on the representation. The latter fact is one aspect of a problem which will occupy our attention throughout most of the rest of this dissertation.

Using Eqs. (3.16) and (2.11), we calculate

$$a^{\dagger}(\tilde{f}) \equiv \int d\mu(j)\, a_j^{\dagger}\, \tilde{f}(j) = i[\phi(\dot{f}) - \pi(f)]$$

$$= i\int d^s x\, \sqrt{|g|}\, g^{00}\, \phi(x) \overleftrightarrow{\partial}_0 f(x), \qquad (3.23a)$$

$$a(\bar{f}) = i[\phi(\dot{f}) + \pi(f)] \;(= a(\tilde{f}) \text{ if } \Psi_j(x) \text{ are real}). \qquad (3.23b)$$

(The bar in Eq. (3.23b) is the price of requiring $a(\tilde{f})$ to be linear in $\tilde{f}$ -- see Sec. F.1. Note that complex conjugation in the f representation is equivalent to complex conjugation in the F representation used earlier, so $\bar{f}$ is unambiguous.) These formulas appear to define $a(\tilde{f})$ and $a^{\dagger}(\tilde{f})$ in a manifestly covariant way (see Sec. VII.3). However, one must remember that $\dot{f}(x)$ is defined in terms of $f(x)$ (which we take as given) through Eq. (3.19), and our definition of the forward propagator $G^{(+)}$ depends crucially on the splitting of the general solution of the wave equation in a static coordinate system into positive- and negative-frequency parts. Consequently, on the one hand, Eqs. (3.23) are of no use in the general case (nonstatic metric) unless we can give a more general definition of the forward propagator. On the other hand, our procedure is not well-defined if the metric takes a static form in two different coordinate systems. Such a situation will be studied in detail in the next chapter.

At first sight Eqs. (3.23a) and (3.23b) may seem contradictory, since the first implies that (since ϕ and π are Hermitian fields)

$$a^\dagger(\tilde{f})^\dagger = i[-\phi((\dot{f})^*) + \pi(f^*)],$$

while the second equation says that

$$a^\dagger(\tilde{f})^\dagger = a(\widetilde{f^*}) = i[+\phi((f^*)^{\cdot}) + \pi(f^*)].$$

This brings out a subtle point: In our present notation $f^*(t,x)$ is defined in terms of the initial value $f^*(x)$ and the positive-frequency time propagation; thus it is not equal to $[f(t,x)]^*$. In fact, in this context differentiation with respect to time anticommutes with complex conjugation:

$$(f^*)^{\cdot} = \frac{d}{dt}\left[\int \frac{d\mu(j)}{\sqrt{2E_j}}\, \Psi_j(x)\, e^{-iE_j t}\, \widetilde{f^*}(j)\right]_{t=0}$$

$$= -i\int d\mu(j)\, \sqrt{\frac{E_j}{2}}\, \Psi_j(x)\, \widetilde{f^*}(j),$$

but

$$(\dot{f})^* = i\int d\mu(j) \sqrt{\frac{E_j}{2}}\, \Psi_j^*(x)\tilde{f}^*(j) = + i\int d\mu(j) \sqrt{\frac{E_j}{2}}\Psi_j(x)\widetilde{f^*}(j).$$

(The last step uses

$$f(x) \equiv \int\frac{d\mu(j)}{\sqrt{2E_j}}\, \Psi_j(x)\tilde{f}(j) = \int\frac{d\mu(j)}{\sqrt{2E_j}}\, \Psi_j^*(x)\widetilde{f^*}^*(j), \qquad (3.24)$$

which is also used in deriving Eq. (3.23b). Substitute f^* for f in the first equality and complex conjugate to obtain the second equality.)

4. Particle Interpretation.

Each of the n-particle spaces is carried into itself by the action of the time translation group. In other words, the particle number N is a constant of the motion. Also, the particles present in an n-particle state do not interact with each other; they behave entirely independently, except for the restrictions of Bose statistics. Consequently, this field theory, like the theory based on the ordinary Klein-Gordon equation, is essentially the second quantization (see, e.g., [Schweber], pp. 156-195) of a single-particle theory.

In fact, one could have started with a one-particle theory, or "relativistic wave mechanics", based on Eq. (1.1). In this approach the equation is to be solved for a complex-valued numerical function instead of an Hermitian operator field. The positive-frequency solutions are the possible wave functions of a particle. The solutions have the general form (3.16), with the scalar product (3.18). The wave functions for a system of n identical particles are obtained from these by the symmetrized tensor product. The annihilation and creation operators can be defined as mappings between the n-particle spaces with adjacent values of n. Finally, the Hilbert space of all possible states of the world (when only particles of this one type are considered) is defined as the direct sum of all the n-particle spaces (including a no-particle state $|0\rangle$). The operator of the scalar field in configuration space can be defined by Eq. (2.9) (or (3.7)).

Conversely, one can recover the single-particle theory from the field theory by studying the one-particle states:

$$|f\rangle = \{0,\ \tilde{f}(j),\ 0,\ \ldots\} = a^\dagger(f)|0\rangle. \tag{4.1}$$

The matrix element of the field between $|f\rangle$ and the vacuum gives the x-space wave function:

$$\langle 0|\phi(t,x)|f\rangle = \langle 0|\int \frac{d\mu(j)}{\sqrt{2E_j}}\,\psi_j(x)\,e^{-iE_j t}\,a_j \int d\mu(k)\ \tilde{f}(k)\,a_k^\dagger|0\rangle$$

$$= \int \frac{d\mu(j)}{\sqrt{2E_j}}\,\psi_j(x)\,e^{-iE_j t}\,\tilde{f}(j) = f(t,x). \tag{4.2}$$

Eq. (4.2) can be loosely regarded as the scalar product of $|f\rangle$ with a continuum basis state $\phi(t,x)|0\rangle$ associated with the point x. However, these generalized states are not orthonormal, since

$$\langle 0|\phi(0,x)\phi(0,y)|0\rangle = \int \frac{d\mu(j)}{2E_j}\, \Psi_j(x)\Psi_j^*(y) \neq \delta(x-y).$$

On the other hand, if we define an operator distribution

$$\Phi(x) = \int d\mu(j)\, \Psi_j(x) a_j, \tag{4.3}$$

we have

$$\langle 0|\Phi(x)\Phi^\dagger(y)|0\rangle = \delta(x-y), \tag{4.4}$$

$$\langle 0|\Phi(x)|f\rangle = \int d\mu(j)\, \Psi_j(x)\tilde{f}(j) = F(x), \tag{4.5}$$

$$a^\dagger(f) = \int d^s x \sqrt{|g|}\, g^{00}\, F(x)\Phi(x) \tag{4.6}$$

(cf. Eqs. (1.5,8)). An object analogous to Φ was introduced by Friedrichs under the name "modified annihilation operator" in the study of a field interacting with an external potential ([Friedrichs], pp. 189-191). When the one-particle states are represented by the functions $F(x)$, the scalar product takes the simple form (1.4) and multiplication by x becomes a self-adjoint operator. It is analogous to the Newton-Wigner position operator for the free field in Minkowski space (Newton and Wigner (1949); Wightman and Schweber (1955)). Correspondingly, $\Phi^\dagger(x)\Phi(x)$ can be interpreted as a particle number density operator in the field theory. Although this operator x has the correct mathematical properties to be interpreted as a position observable in the universe under consideration, it is not thereby self-evident that it has anything to do with the particle behavior observed in actual experiments.[6] Note that F(x) is still a solution of the wave equation when the momentum components are given their natural time dependence; so is the Hermitian field

$$\frac{1}{\sqrt{2}}[\Phi(x) + \Phi^{\dagger}(x)]. \qquad (4.7)$$

This object, however, does not commute at spacelike separations (and in the case of flat space it is not Lorentz-invariant).

The results presented in the next chapter show that there are limitations on how seriously the particle interpretation of field theory in static space-time developed in this chapter can be taken. However, there is one situation in which the particle picture seems to be beyond dispute -- the case of an asymptotically flat space metric. This case fits into the familiar framework of quantum-mechanical scattering theory.[7]

That is, the spectrum of the "squared single-particle Hamiltonian" K (see Eq. (1.1)), except for possible discrete bound states, coincides with the spectrum of K for flat space. In particular, the parameter j labeling the modes can be chosen to be a momentum vector $\vec{k}$. In configuration space the behavior

---

[6] Cf. remarks of Wightman (1962), p. 846.

[7] The brief discussion here cannot do justice to this powerful but subtle framework of thought. The reader who is unfamiliar with it is referred to the first three sections of Brenig and Haag (1959) and to [Messiah], pp. 369-380.

of wave functions is as follows: There are (possibly) bound states, which always remain localized near the region where the metric is not flat. For any given normalizable state orthogonal to the bound states there is a time before which, and a time after which, the particle it describes is essentially out of the range of the curved part of the metric and closely follows "free" or "flat" dynamics. Therefore, these scattering states can be labeled by the configuration of the particles in the remote past (the <u>in</u> representation) or the remote future (the <u>out</u> representation). There are corresponding operators $a^{in}_{\vec{k}}$, $a^{out}_{\vec{k}}$, etc. The momenta labeling the annihilation and creation operators in the <u>in</u> and <u>out</u> representations are associated with the directions and energies with which the observable particles approach and leave the potential, respectively. The transformation between these two bases is given by the S-matrix:

$$|\vec{k}\ \mathrm{in}\rangle = \int d^s\vec{j}\ |\vec{j}\ \mathrm{out}\rangle\langle\vec{j}\ \mathrm{out}|\vec{k}\ \mathrm{in}\rangle = \int d^s\vec{j}\ |\vec{j}\ \mathrm{out}\rangle\ S(\vec{j},\vec{k}). \quad (4.8)$$

(From the properties of the Schrödinger equation under time reversal it follows that

$$\Psi^{\mathrm{out}}_{\vec{k}}(x) = \Psi^{\mathrm{in}}_{-\vec{k}}{}^*(x). \quad (4.9)$$

Thus the <u>in</u>- and <u>out</u>-states form a pair of mutually conjugate basis sets, as discussed at the end of Sec. VIII.2. In particular, Eq. (2.8) holds, and Eq. (2.12) implies that

$$a_{\vec{j}}^{\text{in}\dagger} = \int d^s\vec{k}\, (\Psi_{\vec{k}}^{\text{out}}, \Psi_{-\vec{j}}^{\text{out}}{}^*)\, a_{\vec{k}}^{\text{out}\dagger} = \int d^s\vec{k}\, S(\vec{k},\vec{j})\, a_{\vec{k}}^{\text{out}\dagger} \quad (4.10)$$

(as was to be expected from Eq. (4.8)).)

The essential point here is that as long as the particles are out of range of the gravitational field the physical system essentially reduces to the ordinary free field, whose physical interpretation, especially in momentum space, is well established. Thus one's accceptance of the asymptotic particle interpretation of the field theory considered here should be as strong as one's faith in the free field.

In most problems to which quantum field theory is applied the range of the interaction is microscopic, and the only feasible experiments are scattering experiments (and perhaps measurements of bound state energies). In gravitational problems, however, one normally has a gravitational field in a region of macroscopic or even cosmological dimensions, and experiments may take place entirely inside it. The emphasis, therefore, in the physical interpretation of field theory can be expected (even in the case of an asymptotically flat universe) to shift to observables which have something to do with local phenomena in x-space. As the remarks of the next two chapters show, this is a subject which still remains to be satisfactorily developed.

5. The Case of Zero Frequency and the Infrared Case.

If there is an eigenvector of K with E = 0, it has been remarked in Sec. VIII.2 that the contribution of that mode to the total energy becomes continuous. (As the prototype of this situation we may take a free scalar field with m = 0 in a finite "box" with periodic boundary conditions.) The field has the expansion (1.10). The most obvious representation is the tensor product of the unique irreducible representation of the canonical operators $\{q_0, p_0\}$ with the Fock representation (Eqs. (2.4), (3.1)) of all the higher modes. (Let us denote the Hilbert space of this Fock representation by $\mathcal{F}$, as before.) In the tensor product representation there is no (normalizable) vacuum state of zero energy. For each state in $\mathcal{F}$ (with energy E, say) there exists in the tensor product space a continuum of states with energy running from E to $+\infty$. This spectrum has multiplicity 2, corresponding to the two possible parities σ of a state under reversal of sign of the field ϕ.

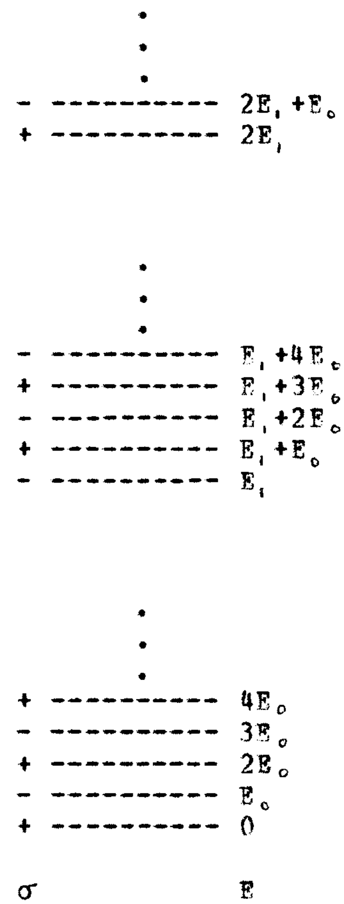


Fig. 12

Second quantization of the spectrum (VIII.5.1).

This structure is not at all surprising if one considers the theory as a limiting case of a theory for which the lowest point of the spectrum of K is a discrete eigenvalue E_0 slightly greater than 0. Let

$$0 < E_0 \ll E_1 , \tag{5.1}$$

where E_1 is the next lowest point of the spectrum of K. The energy spectrum of the field theory is shown in Fig. 12. Each band $E + nE_0$ extends to infinity. The quantum number σ is the parity of the number of particles (of all modes) in the state. As $E_0 \longrightarrow 0$ all the discrete states with energies $E + nE_0$ collapse into the level E. But what is left is a continuum of states, with σ-parity + and - for each energy above E. The state of energy E which marked the bottom of the band (corresponding to a vector in $\mathcal{F}$) is washed out into the continuum. To the extent that particle language makes sense at all when $E_0 = 0$, every state of the theory contains an infinite number of zero-energy particles. The "amount of E_0 present" can be measured (by its energy p_0^2) but not counted.

A related situation, which arises more often in practice, is that 0, although not an eigenvalue, is a limit point of the spectrum of K (usually, the lower endpoint of a continuous spectrum). The theory of massless scalar particles in Euclidean space is the simplest example. In this "infrared" case the Fock representation as defined in Sec. VIII.3 makes perfect sense. However, two points should be kept in mind in dealing with free infrared fields (in addition to the difficulties to be encountered for interacting fields, well known from quantum electrodynamics).

First, there exist other representations besides the Fock representation in which there is a continuous unitary time translation group whose generator, the energy operator H, is nonnegative (Borchers et al. (1963)). Intuitively, this reflects the possibility of states of the field containing infinitely many particles but, nevertheless, finite energy.

Second, as already remarked, Eq. (3.9a) becomes a nontrivial restriction on the test functions in x-space for which the smeared field operators are defined. For example, in the case of the massless scalar field in space-time of dimension 2 (but not higher dimensions) $\phi(f)$ cannot be defined for all $f(t,x)$ in the space $\mathscr{S}$ (see Sec. IV.1); it is necessary to require also that the Fourier transform of f vanish at zero momentum (see Wightman (1964), pp. 204-212). In this case the integral which defines the two-point function (cf. Eq. (3.13)),

$$G^{(+)}(t_2,x_2;t_1,x_1)$$

$$= \frac{1}{2\pi}\int_0^{\infty} \frac{dp}{|p|} \cos[ip(x_2-x_1)] \exp[-i|p|(t_2-t_1)], \quad (5.2)$$

does not converge in any usual sense. But $G^{(+)}$ does make sense as a functional on the restricted test function space just described, and Eq. (5.2) can then be given a meaning by formal interchange of the order of integration. One should not be surprised if similar phenomena occur in other theories with the infrared property, such as those discussed in Sec. VIII.6 and Chapter IX.

6. Summary; Application to De Sitter Space.

In this chapter we have considered static metrics for which the squared single-particle Hamiltonian operator K (Eq. (1.1)) is essentially self-adjoint. In the case when K does not have the eigenvalue 0 we have arrived at the following conclusions:

(1) Because there is a basis of solutions of the elementary form (1.2), there is a clear notion of "positive frequency". Hence a "forward propagator" can be defined.

(2) A Fock representation can be constructed in close analogy to the theory of the free field. The two-point vacuum expectation value in this representation is the forward propagator.

(3) This theory has an obvious particle interpretation. In fact, since the particles are not created or destroyed, the theory can be reinterpreted as a "wave mechanics" for a single particle. A position observable of the Newton-Wigner-Wightman-Schweber type can be defined.

An example of a space-time to which this theory applies is the portion of two-dimensional de Sitter space covered by a geodesic Fermi coordinate system (see Secs. III.2, III.3, III.5, V.2, and V.7). The eigenfunctions $\Psi_j(x)$ for this case are given in Eqs. (V.7.5). From this point of view the effect of the curvature of space shows up as a smooth potential hill, and the eigenfunctions display a nonvanishing probability for both reflection and transmission of particles.[8]

A classical free particle in de Sitter space follows a timelike geodesic. Depending on whether the initial velocity of the particle is high or low relative to the central worldline of the Fermi coordinate system, such a path will seem to "penetrate" or to "reflect" from the center of the universe in terms of the Fermi space coordinate β or ρ. (Imagine the geodesics labeled "r = - 3" and "r = - .5" in Fig. 9 (Sec. III.7) superimposed on Fig. 4 (Sec. III.2). These are the paths of particles with high and low velocity, respectively.) One would expect, therefore, that in the quantum theory the transmission probability will be very large for large momentum k and very small for small k, relative to the mass parameter q. Of course, this is to be expected in general from a barrier penetration problem. An explicit calculation, based on comparison of the appropriate

---

[8] Although the time dependence of the wave function is given here by an ultrarelativistic energy-momentum relation, the qualitative behavior of wave packets is the same as in nonrelativistic quantum mechanics, and hence the interpretation of the coefficients in the eigenfunctions as reflection and transmission amplitudes remains valid.

coefficients in Eqs. (V.7.10) and (V.7.14), yields for the ratio of the reflection and transmission amplitudes

$$\frac{|\text{Reflection}|}{|\text{Transmission}|} = e^{-\pi k}\left| e^{i\pi\nu} - \frac{1}{\pi}\Gamma(1-ik)\Gamma(ik)\ \sin \pi(\nu+ik)\right|$$

$$= \frac{\cosh \pi\gamma}{\sinh \pi k}, \qquad (6.1)$$

where

$$\gamma = \operatorname{Im} \nu = \sqrt{q - \tfrac{1}{4}}. \qquad (6.2)$$

This expression shows the expected behavior, although the transition from almost total transmission to almost total reflection is perhaps not as abrupt as one might expect.

We shall not study the motion of wave packets in de Sitter space in more detail, but in Sec. IX.2 we shall investigate a closely analogous case more quantitatively.

7. Stationary Metrics.

If, in a distinguished coordinate system, the metric coefficients are independent of time but the space-time orthogonality condition (D.4) does not hold, then the metric is called stationary. An example of current interest is the exterior Kerr metric (Kerr (1963)), the gravitational field outside a rotating massive star or black hole. It would be very surprising if the results of this chapter did not extend almost verbatim to such a case; however, there will be some complications of the formalism.

When the substitution (1.2) is made into the wave equation (VII.1.2,3) with a stationary metric, one obtains (dropping the index on the energy and the eigenfunction)

$$- E^2 g^{00} \psi - 2iE g^{0j} \partial_j \psi - iE \frac{1}{\sqrt{|g|}} \partial_j [\sqrt{|g|} g^{0j}] \psi$$

$$+ \frac{1}{\sqrt{|g|}} \partial_j [\sqrt{|g|} g^{jk} \partial_k \psi] + m^2 \psi = 0. \quad (7.1)$$

This is not an eigenvalue equation in the usual sense, since E occurs both linearly and quadratically. The same situation arises in the study of the Klein-Gordon equation with an external electrostatic field (Snyder and Weinberg (1940)). In the electrostatic case two approaches have been followed:

(1) to work directly with the solutions of the equation analogous to Eq. (7.1), which are not orthogonal for different E (*ibid.*);

(2) to convert the wave equation, by a change of variables whose analogue here would be (notation of Eq. (VII.3.2))

$$\psi = \frac{1}{\sqrt{2}} (v + u), \qquad D\psi = \frac{m}{\sqrt{2}} (v - u), \qquad (7.2)$$

into a pair of first-order equations, which leads to a true eigenvalue problem (Feshbach and Villars (1958); [Corinaldesi], Chaps. 3, 5, 6; Veselić (1970)).[9]

These methods presumably would work when applied to the stationary gravitational field.

The subject will not be pursued here, since, unlike the study of time-dependent metrics, it does not seem to involve any new matters of principle.

[9] A variant of method (2) is to use the covariant five-component Petiau-Duffin-Kemmer formalism (see [Umezawa], pp. 85-91, 197).

Chapter IX

THE FREE FIELD IN RINDLER COORDINATES

The theory presented in Chapter VIII is compelling, because it is such a natural generalization of the familiar quantization of the free field. However, the construction of the Fock representation has been based on a particular eigenfunction expansion associated with a coordinate system in which the wave equation separates. If there is more than one coordinate system in which the metric takes a static form, one must ask whether and in what sense the corresponding field theories are equivalent.

These can be compared most easily in the case of two coordinate systems which have one equal-time hypersurface in common. For instance, in two-dimensional de Sitter space each timelike geodesic is associated with a field theory, as described in Sec. VIII.6. The Fermi coordinate systems based on two geodesics O_1 and O_2 (see Fig. 13) cover different portions of the space, but (in general) there is a region of overlap. The spacelike curves of constant time coincide at the instant of closest approach of O_1 and O_2. By symmetry, each of the two Fock field theories is equally valid for the description of phenomena occurring in the overlap region. To the extent that they disagree, either both theories must be rejected, or the concepts and quantities involved in the disagreement must be shown to be

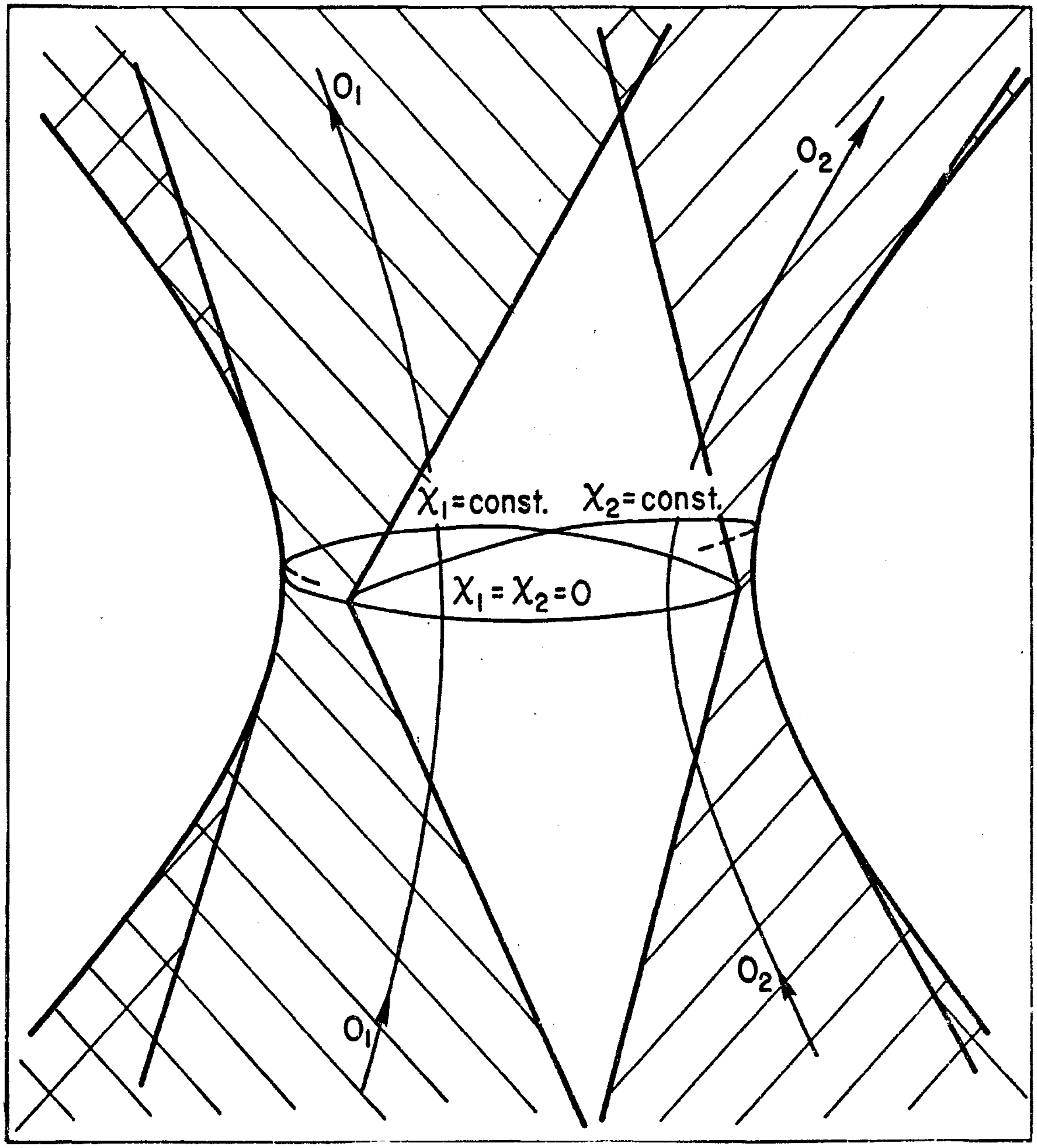


Fig. 13

Historical regions associated with two freely moving observers in de Sitter space. Regions extrahistorical for O_1 are shaded ////; regions extrahistorical for O_2 are shaded \\\\.

without true observational significance, or the theories must be interpreted as applying to different physical situations.

In this chapter we shall study in detail an even simpler example, which exhibits the ambiguity of Fock quantization in a very striking way. The physical situation involved is a very familiar and, seemingly, well-understood one: the free scalar field in ordinary flat space.

1. The Rindler Model.

We consider the region $\{(t,x) \mid |t| < x\}$ of two-dimensional Minkowski space[1], and define coordinates (v,z) by

$$t = z \sinh v, \qquad x = z \cosh v \tag{1.1}$$

(see Fig. 14). Then

$$v = \tanh^{-1} \frac{t}{x}, \qquad -\infty < v < \infty, \tag{1.2a}$$

$$z = \sqrt{x^2 - t^2}, \qquad 0 < z < \infty, \tag{1.2b}$$

and when $t = v = 0$, z coincides with x. (Therefore $\Pi(0,x) = \Pi(0,z)$ is the same quantity in both systems -- see Sec. VII.1.) We calculate

$$ds^2 \equiv g_{\mu\nu}\, dx^\mu\, dx^\nu = z^2\, dv^2 - dz^2, \tag{1.3}$$

---

[1] It should not be hard to extend the results of this chapter to four dimensions.

$$\sqrt{|g|} = z, \qquad g^{00} = z^{-2}. \tag{1.4}$$

This metric is static. The physical reason for this is that the operation of translation in the time coordinate v is simply a homogeneous Lorentz transformation, which is a symmetry of Minkowski space. We have restricted ourselves to the region where the Killing vector of this isometry is timelike and future-directed. Each surface v = const. is a Cauchy surface (see Sec. III.5) for this region, so a self-contained field theory within the region should exist.

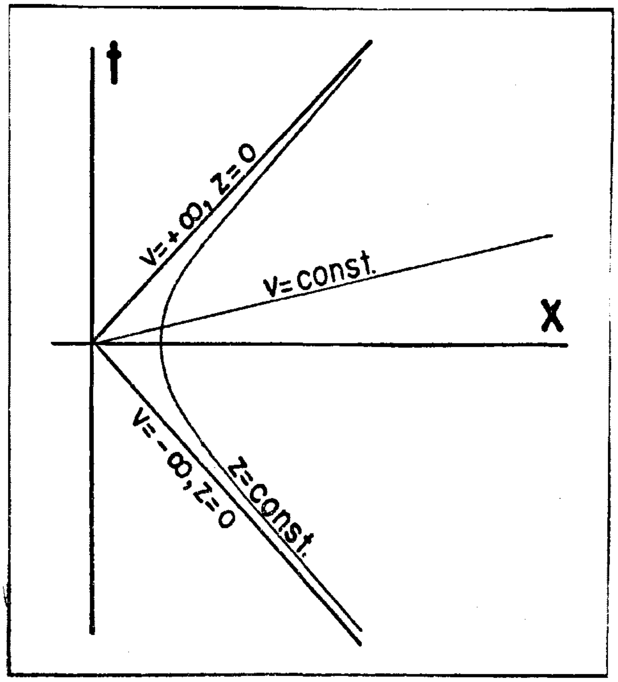


Fig. 14

Rindler coordinates.

The coordinates (1.2) are the Fermi coordinates (see Appendix D) relative to a hyperbolic timelike curve z = const. This, of course, is not a geodesic, but it is the worldline of an observer who undergoes a constant acceleration and, consequently, experiences a constant gravitational field. (For instance, to some degree of approximation an observer on the surface of the earth has such a Fermi coordinate system; the exterior

Schwarzschild metric with the usual static coordinate system would be a better approximation for that situation.) The numerical value of this acceleration is 10^{32} cm/sec times 1/z, where z is measured in Compton wavelengths. Thus, if m is a typical elementary particle mass, z = 1 corresponds to an acceleration of 10^{29} g's, and an observer accelerating at 1 g would be 10^{13} kilometers distant from the horizon (z = 0) of his Fermi coordinate system.

We shall call these <u>Rindler coordinates</u>, because they have been discussed most thoroughly by W. Rindler[2] (in the four-dimensional case). He points out that the relation of this system to Cartesian coordinates is very similar to the relation between Schwarzschild and Kruskal coordinates for the space surrounding an isolated point mass.[3] It is important to realize that Rindler coordinates are just as appropriate for the description of the region of flat space which they cover as Schwarzschild coordinates are for the study of the space around a massive body outside the radius r = 2M. If the theory of Chapter VIII fails in this test case, it must also be rejected for the Schwarzschild metric, and it cannot be applied as a general method (but see Secs. IX.6-7 below).

---

[2] Rindler (1966) (also [Rindler 2], pp. 184-195); earlier papers by others are cited in Refs. 1 and 2 of his paper. The physics of the situation ("uniformly accelerated rigid rod") is discussed in [Rindler], pp. 41-43 (or [Rindler 2], pp. 61-64).
[3] The Schwarzschild solution is discussed in most textbooks on general relativity. The fundamental modern paper on the subject is Kruskal (1960).

It is also clear from a comparison of Fig. 14 with Fig. 5 (Sec. III.2) that the situation under discussion here is very similar to the relation between a geodesic Fermi coordinate system in de Sitter space and a geodesic Gaussian system (or any system which is regular in the neighborhood of one of the two singular points of the Fermi system). The discussion of this chapter applies with very little change to the situation in de Sitter space.

We proceed to quantize the scalar field along the lines of Chapter VIII. The eigenvalue equation (VIII.1.3) (K defined in Eq. (VIII.1.1)) is in this case a Bessel equation

$$\left(z^2 \frac{d^2}{dz^2} + z \frac{d}{dz} - m^2 z^2 + E_j^2\right) \Psi_j(z) = 0, \qquad (1.5)$$

and the volume element in the scalar product (VIII.1.4) is

$$\int dz \sqrt{|g|}\, g^{00} = \int_0^\infty \frac{dz}{z}. \qquad (1.6)$$

We shall consider only the case $m \neq 0$. The solution of this eigenvalue problem is given in [Titchmarsh], Sec. 4.15. The spectrum of E_j^2 extends from 0 to $+\infty$ with unit multiplicity. We can therefore use E_j itself as the parameter j; we have $0 \leq j < \infty$. The eigenfunctions are

$$\Psi_j(z) = \frac{1}{\pi}[2j \sinh(\pi j)]^{1/2} K_{ij}(mz), \qquad (1.7)$$

where K_{ij} is the Macdonald function (modified Bessel function) of imaginary order. The functions (1.7) are normalized so that (see Eqs. (VIII.1.7,9))

$$\int_0^\infty \frac{dz}{z} \Psi_j^*(z)\Psi_k(z) = \delta(j-k), \qquad (1.8a)$$

$$\int_0^\infty dj\, \Psi_j^*(z)\Psi_j(y) = z\,\delta(z-y). \qquad (1.8b)$$

From now on we set $m = 1$. This is no loss of generality; it means we choose the unit of length to be the Compton wavelength of the particle.

The expansion of the field in annihilation and creation operators (Eq. (VIII.2.9)) is

$$\phi(v,z) = \int_0^\infty \frac{dj}{\sqrt{2j}} \Psi_j(z)[e^{-ijv} a_j + e^{ijv} a_j^\dagger]. \qquad (1.9)$$

The canonical momentum is

$$\Pi(v,z) = \frac{1}{z}\dot{\phi}(v,z) \quad \left(= \frac{d\phi}{dt} \right) =$$

$$-\frac{i}{z}\int_0^\infty dj \sqrt{\frac{j}{2}}\, \Psi_j(z)[e^{-ijv} a_j - e^{ijv} a_j^\dagger], \qquad (1.10)$$

and Eq. (VIII.2.11) becomes

$$a_j = \frac{1}{\sqrt{2}}[\sqrt{j}\int_0^\infty \frac{dz}{z}\Psi_j(z)\phi(0,z) + \frac{i}{\sqrt{j}}\int_0^\infty dz\, \Psi_j(z)\Pi(0,z)]. \qquad (1.11)$$

A field theory with a vacuum can be constructed as described in Sec. VIII.3. Since the spectrum of j extends all the way down to 0, this model falls into the infrared category (see Sec. VIII.5), even though $m \neq 0$.

2. Quasiclassical Behavior of Positive-Frequency Solutions.

According to the theory proposed in Sec. VIII.4, the wave function of a particle in Rindler space is of the form

$$\Psi(v,z) = \int_0^\infty dj\, \tilde{f}(j)\, \Psi_j(z) e^{-ijv}. \qquad (2.1)$$

This is not a positive-frequency solution of the Klein-Gordon equation in the usual sense when transcribed back into terms of t and x, but rather a superposition of positive- and negative-frequency solutions. (This can be demonstrated by a calculation essentially the same as that carried out for the quantized field in the next section.) Therefore, we are considering a relativistic theory of a single free spinless particle which differs in its details from the usual one. It is of interest to verify that wave packets of the form (2.1) approximately follow classical trajectories. If this should turn out not to be the case, one would be inclined to throw out the theory as physically unreasonable. However, we shall see that the quasiclassical behavior is correct, which is an argument in favor of taking the theory seriously despite its difference from the usual one.

If we let

$$u = \log z, \tag{2.2}$$

Eq. (1.5) becomes

$$-\frac{d^2\Psi_j}{du^2} + e^{2u}\Psi_j = j^2\Psi_j \qquad (\Psi_j = \Psi_j(e^u)), \tag{2.3}$$

which has the form of a nonrelativistic Schrödinger equation with potential e^{2u}. We know, therefore, that its solutions will decay rapidly as $u \longrightarrow +\infty$ and will approach a sum of incoming and outgoing plane waves as $u \longrightarrow -\infty$, the oscillatory behavior beginning at the classical turning point, $e^u = j$. Explicitly, there is the series expansion[4] (see [Vilenkin], p. 270)

$$K_{ij}(z) = \operatorname{Re}\left[\left(\frac{z}{2}\right)^{ij} \sum_{k=0}^{\infty} (-1)^k \,\Gamma(-k-ij)/k!\, \left(\frac{z}{2}\right)^{2k} \right], \tag{2.4}$$

which gives for small z

$$\Psi_j(e^u) = \varphi\, e^{iju}\, e^{-ij\log 2} \frac{\Gamma(-ij)}{|\Gamma(-ij)|} \left[1 + \frac{1}{4}(1+ij)^{-1} e^{2u} + O(e^{4u}) \right] + \text{complex conjugate}, \tag{2.5}$$

since

---

[4] This expansion could have been used to determine the normalization constant in Eq. (1.7), after the fashion of Sec. V.7.

$$|\Gamma(\pm ij)| = \sqrt{\frac{\pi}{j \sinh \pi j}} \, . \tag{2.6}$$

For large z we have the asymptotic expansion

$$\Psi_j(z) \sim \frac{1}{\sqrt{\pi}} \sqrt{j \sinh \pi j}\, z^{-1/2} e^{-z} \{1 + \frac{\mu - 1}{8z} + \frac{(\mu - 1)(\mu - 9)}{2(8z)^2}$$

$$+ \frac{(\mu - 1)(\mu - 9)(\mu - 25)}{3!(8z)^3} + \ldots\} \quad (\mu = -4j^2) \tag{2.7}$$

([N.B.S.], Eq. (9.7.2) (p. 378)). One of the eigenfunctions is graphed in Fig. 15.

Substituting Eq. (2.5) into Eq. (2.1) and applying the principle of stationary phase to the incoming and outgoing waves yields

$$\begin{aligned} v &= u - \alpha - \beta(j) \quad \text{as } v \longrightarrow -\infty, \\ v &= u - \alpha + \beta(j) \quad \text{as } v \longrightarrow +\infty, \end{aligned} \tag{2.8}$$

where

$$\alpha = -\frac{d}{dj}\{\arg \tilde{f}(j)\}, \tag{2.9a}$$

$$\beta(j) = \log 2 + \frac{d}{dj}\arg \Gamma(ij) + \frac{1}{4}\,\frac{1 - j^2}{(1 + j^2)^2}\, e^{2u} + O(e^{4u}). \tag{2.9b}$$

(It is assumed that $\tilde{f}(j)$ is peaked around a particular value of j, and α can be chosen independently of j.) Now simple algebra

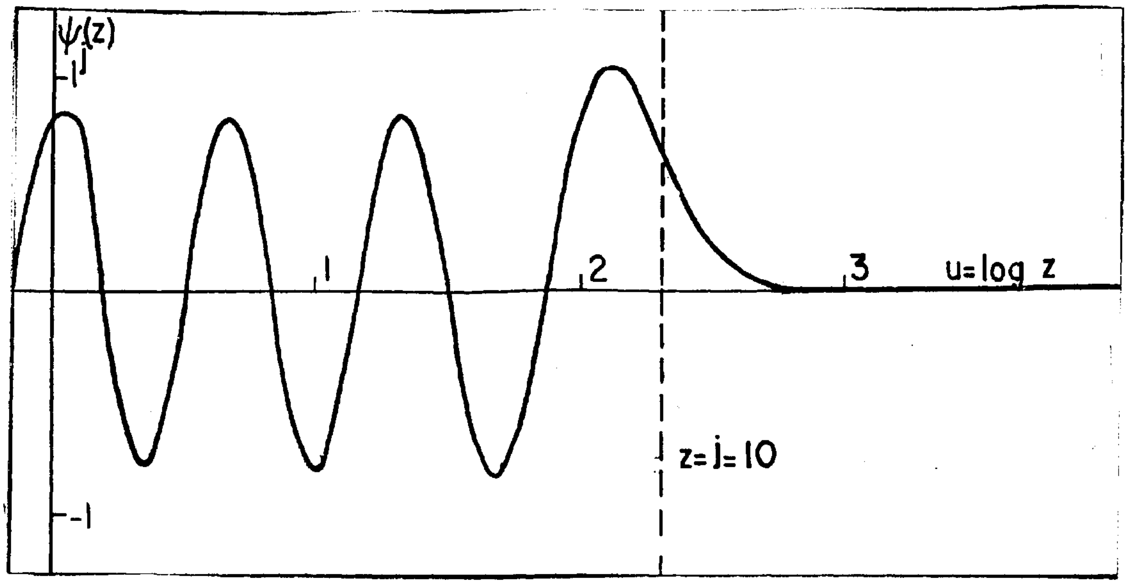
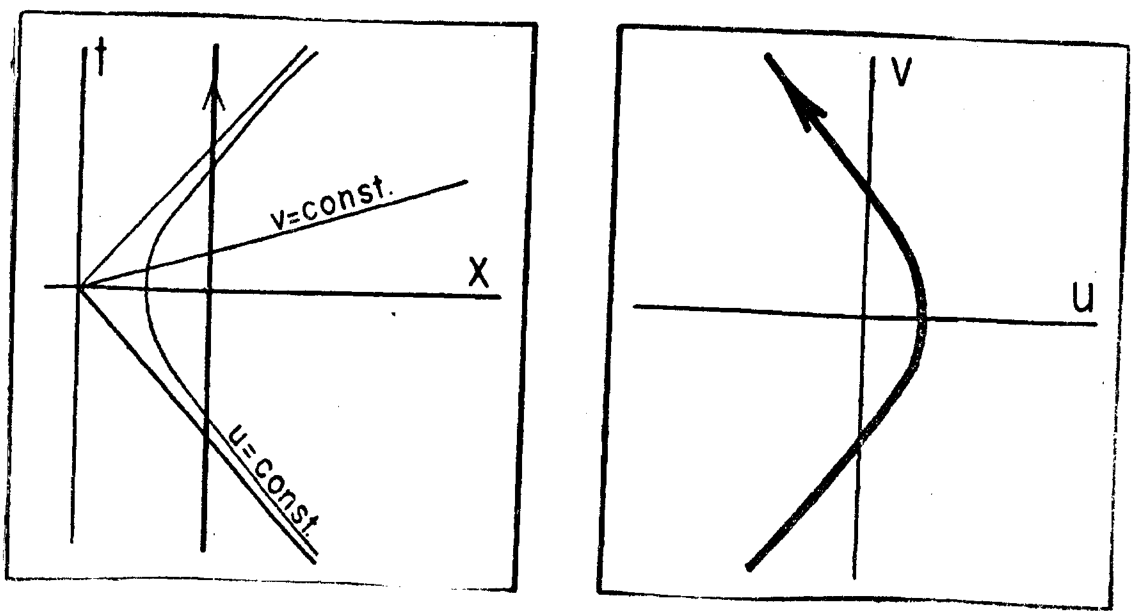


Fig. 15

The eigenfunction $\Psi_{10}(z)$.



(a) Cartesian coordinates (b) Rindler coordinates

Fig. 16

Trajectory of a free particle in flat space.

shows that the curve defined by $u = v + a$ is just the straight line $x = t + e^{a}$. At first glance, then, Eqs. (2.8) appear to be an absurd result; they seem to say that the (free, massive) particle enters the Rindler region at the velocity of light, bounces (at a point depending on j), and departs at the speed of light.

However, let us take a closer look at the general form of the trajectory we were expecting,

$$x = bt + c \qquad (|b| < 1, \quad c = e^{a} > 0). \tag{2.10}$$

Substituting from Eqs. (1.1), we obtain

$$\frac{1}{2}\{(1 - b)e^{v} + (1 + b)e^{-v}\} = ce^{-u}, \tag{2.11}$$

which can be regarded as a quadratic equation in either e^{v} or e^{-v}. The vanishing of the discriminant gives the maximum value of u attained on the trajectory:

$$e^{u} = j_{cl} \equiv c(1 - b^{2})^{-1/2}. \tag{2.12a}$$

(Trajectories with the same value of $j_{c\ell}$ but different values of c are images of each other under Lorentz boosts, just as trajectories with the same slope b are related by time translation. We have already seen that in the quantum theory the turning point is related to j, the variable conjugate to v as energy is conjugate to t.) One root of the solution for e^v yields, for large negative u,

$$v = \log \{\frac{1}{1-b}[ce^{-u} + [c^2 e^{-2u} - (1-b^2)]^{1/2}]\}$$

$$\sim \log \{\frac{1}{1-b} ce^{-u} [2 - (1-b^2)c^{-2} e^{2u}]\}$$

$$\sim -u + a + \log \frac{2}{1-b} - \frac{1}{4} j_{cl}^{-2} e^{2u}$$

as $v \longrightarrow +\infty$. To find an expansion for $v \longrightarrow -\infty$ we use one of the solutions for e^{-v} in the same way. Finally, we can add and subtract 2a from the former expression, using

$$a = \log j_{cl} + \frac{1}{2}[\log (1+b) + \log (1-b)]. \tag{2.12b}$$

The results are

$$v \sim u - a - \log \frac{2}{1+b} + \frac{1}{4} j_{cl}^{-2} e^{2u} \quad \text{as } v \longrightarrow -\infty, \tag{2.13}$$

$$v \sim -u - a - \log \frac{2}{1+b} - \frac{1}{4} j_{cl}^{-2} e^{2u} + 2 \log 2j_{cl} \quad \text{as } v \longrightarrow +\infty.$$

Thus an exponentially small deviation in u-v space from a line with unit slope corresponds in x-t space to a finite change in slope, and hence Eqs. (2.8) are not incorrect.

If we identify j with j_{cl}, comparison of Eqs. (2.13) and (2.8) suggests

$$\alpha = a - \log(1 + b) - \log j, \qquad (2.14a)$$

$$\beta(j) = -\frac{1}{4} j^{-2} e^{2u} + \log 2j. \qquad (2.14b)$$

Now for large j, when the most classical behavior is expected because of the short wavelengths[5], we can write

$$\frac{1 - j^2}{(1 + j^2)^2} \sim - j^{-2},$$

$$\frac{d}{dj} \arg \Gamma(ij) = \frac{d}{dj} \operatorname{Im} \log \Gamma(ij)$$

$$= \frac{d}{dj} \operatorname{Im} [(ij - \frac{1}{2}) \log ij - ij + \frac{1}{2} \log 2\pi + O(j^{-1})] \sim \log j$$

([N.B.S.], Eq. (6.1.41) (p. 257)). So Eq. (2.9b) becomes identical to Eq. (2.14b), and the quasiclassical behavior is verified. Eq. (2.14a) then gives the j-dependent relationship between α and a.

The trajectories are sketched in Fig. 16.

---

[5] Note, however, that large j corresponds to large distance from the coordinate singularity as well as high energy.

3. Comparison of the Two Representations.

We now have two field-theoretical descriptions of the behavior of scalar particles in the wedge of space-time where $|t| < x$, the theory of Sec. IX.1 and the ordinary textbook theory of the free scalar field. Are these in some sense the same theory?

The standard expansion of the free field at $t = 0$ and its conjugate momentum into annihilation and creation operators is ($\omega_k = \sqrt{k^2 + m^2}$)

$$\phi(x) = \int \frac{dk}{\sqrt{2\omega_k}} [e^{ikx} b_k + e^{-ikx} b_k^\dagger], \qquad (3.1a)$$

$$\pi(x) = -i \int dk \sqrt{\frac{\omega_k}{2}} [e^{ikx} b_k - e^{-ikx} b_k^\dagger]. \qquad (3.1b)$$

Substituting into the formula (1.11) for a_j, we find

$$a_j = \frac{1}{2} \{\int_0^\infty dy\, \Psi_j(y) \int_{-\infty}^\infty dk\, e^{iky} [-\frac{1}{y}\sqrt{\frac{j}{\omega_k}} + \sqrt{\frac{\omega_k}{j}}]\, b_k$$

$$+ \int_0^\infty dy\, \Psi_j(y) \int_{-\infty}^\infty dk\, e^{-iky} [-\frac{1}{y}\sqrt{\frac{j}{\omega_k}} - \sqrt{\frac{\omega_k}{j}}]\, b_k^\dagger\}. \qquad (3.2)$$

The kernel in the second term does not vanish. (We shall study it in more detail shortly.)

The presence of creation operators in this formula has drastic consequences. It means that a vector which is annihilated by the b's is not annihilated by the a's, and vice versa. So the vacuum of the Rindler-space theory is not the ordinary vacuum of the free field. One-particle states in one theory are not one-particle states in the other theory, and so on. The notion of a particle is completely different in the two theories. The particles or quanta of the Rindler Fock representation cannot be identified with the physical particles described by the usual quantum theory of the free field.

The minimal conclusion which must be drawn from this observation is the following: In the context of the general static universe treated in Chapter VIII the particle concept does not have the same physical significance as in free field theory. The theory of quantization in a static metric amounts to the following: Given a manifold with a timelike Killing vector, we have constructed a representation of the field algebra in which the symmetry generated by the Killing vector field is implemented by a group of unitary operators. Also, the generator of this unitary group has been required to be a positive operator, and we have used it like the Hamiltonian in special-relativistic theories. We found that the eigenstates of this operator can be labeled in a way which is quite similar to the particle structure of the states of the free field. This doesn't necessarily mean, however, that these eigenstates have anything to do with physical particles in the usual sense, things that trigger detectors and so on. It might be better to use the term "quanta" (or "virtual particles") instead of "particles".[6]

---

[6] Of course, when the metric is asymptotically flat as well as

The difference between the two theories shows up in the associated single-particle wave mechanics, because the definition of "positive-frequency solution" is different. A positive-frequency solution in the ordinary sense ($\exp(-i\omega_k t)$) is a superposition of positive- and negative-frequency solutions in the sense of $\exp(\pm ijv)$. This probably leads to slight differences in the way wave packets diffuse, and so forth, even though, as shown in the previous section, the qualitative behavior of a wave packet is correct in the Rindler theory.

In pondering the significance of the disconcerting appearance of two different quantizations of the free field, it would be helpful to know whether the two representations involved are equivalent in the mathematical sense. In other words, can the Hilbert spaces of the two theories be identified in a natural way, even though the vacuum vectors (and associated particle structure) are different?[7] The question makes sense only for field operators for which both representations are defined. For (s+1)-dimensionally smeared fields of the form (VII.3.1) (the type usually considered in axiomatic field theory) this means that the question should be asked for an algebra of field operators with test functions $f = f(v,z)$ with support inside the

---

static, the identification of the quanta with physical particles is convincing -- see Secs. VII.7 and VIII.4.

[7] More precisely, two representations $A_1(f)$ and $A_2(f)$ of an algebra of elements f, on Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, are <u>unitarily equivalent</u> if there is a unitary operator U from $\mathcal{H}_1$ onto $\mathcal{H}_2$ such that $UA_1(f)U^{-1} = A_2(f)$ for all f.

Rindler region (and with sufficient smoothness properties, etc., for the field operators to be defined in both representations). Alternatively, for a canonical theory one can consider s-dimensional smearing of the equal-time (or Schrödinger-picture) fields:

$$\phi(f) = \int dx\ \phi(0,x)\,f(x) = \int dz\ \phi(0,z)\,f(z) \qquad (3.3)$$

and similarly for $\pi(f)$[8]; then we are interested in f's with support on the positive part of the x-axis. In view of the explicit canonical construction employed here, the latter framework is easier to study in the present case.

Let us call the standard representation of the free field (in either the one- or the two-dimensional sense) <u>the ΦOK representation</u> (to distinguish it from the general notion of <u>a Fock</u> representation used in Secs. VIII.3 and F.1). The representation of Sec. IX.1 will be called the Rindler representation.

On the basis of an abstract argument it can be shown that the Rindler representation is at best a <u>subrepresentation</u> of the ΦOK representation. The representation within ΦOK space of the subalgebra of field operators with support in the Rindler region is reducible, for these operators commute with all the fields with support in the symmetrically opposite region of space-time, $\{(t,x) \mid x < -|t|\}$. The Rindler representation,

---

[8] Eq. (3.3) and its partner can be written in the covariant form (VII.3.3).

however, is irreducible (Sec. VIII.3). (Hence it should have been obvious from the beginning that the two vacuum states are not identical; for if the Rindler representation contained the ΦOK vacuum, its Hilbert space would be the entire ΦOK space, by the Reeh-Schlieder theorem ([Streater-Wightman], pp. 138-139), which would contradict its irreducibility.)

The possibility remains that ΦOK space contains a vector, Ψ_0, which is annihilated by all the Rindler annihilation operators a_j (Eq. (3.2)) and, consequently, can be identified with the Rindler vacuum. Then the Hilbert space of the Rindler theory would be identified with a certain subspace of ΦOK space (perhaps not uniquely). Evidence against this possibility is the fact, which will be demonstrated below, that the kernel in the second term of Eq. (3.2) is not square-integrable, and so does not represent a Hilbert-Schmidt operator. If the Bogolubov transformation (3.2) were invertible, the theorem stated and proved in Sec. F.3 would imply that the vacuum of one representation cannot lie in the Fock space of the other. Since the theorem has not been extended to noninvertible transformations, the argument offered here does not rigorously establish that no Rindler vacuum Ψ_0 exists in ΦOK space, but it makes this conclusion more likely. (If Ψ_0 exists, it must lie outside the quadratic-form domain of the ΦOK number operator, since the expectation value of the number operator is the Hilbert-Schmidt norm of the kernel in question plus a manifestly positive term involving the field on the negative x-axis.)

Let us turn to the evaluation of the integrals in Eq. (3.2). We have, at least formally,

$$a_j = \int_{-\infty}^{\infty} dk\ U(j,k)\ b_k + \int_{-\infty}^{\infty} dk\ V(j,k)\ b_k^{\dagger}, \qquad (3.4)$$

$$U(j,k) = \frac{1}{2}\left[A^*(j,k) + B^*(j,k)\right], \qquad (3.5a)$$

$$V(j,k) = \frac{1}{2}\left[A(j,k) - B(j,k)\right], \qquad (3.5b)$$

$$A(j,k) = \sqrt{\frac{j}{\omega_k}} \int_0^{\infty} dy\ \frac{1}{y}\ \Psi_j(y)\ e^{-iky}, \qquad (3.6a)$$

$$B(j,k) = \sqrt{\frac{\omega_k}{j}} \int_0^{\infty} dy\ \Psi_j(y)\ e^{-iky}, \qquad (3.6b)$$

where $\Psi_j(y)$ is given by Eq. (1.7). (We still take $m = 1$ for convenience.) Eq. (3.4) really stands for a transformation of smeared fields of the type (F.2.3) (with a and b interchanged). At first it is not obvious that the order of integration in Eq. (3.2) can be changed to yield <u>bona fide</u> integral operators, as implied in Eq. (3.4). We shall find, however, that $U(j,k)$ and $V(j,k)$ are simple smooth functions, given in Eqs. (3.12) (to which the reader may skip if not interested in the details of the integration).

Formulas (6.699.3,4) of [Gradshteyn-Ryzhik] (p. 747) yield

$$\int_0^\infty dy\, y^\lambda K_{ij}(y) \sin ky = 2^\lambda k \left|\Gamma\left(\frac{2+\lambda+ij}{2}\right)\right|^2 \times$$

$$F\left[\frac{2+\lambda+ij}{2}, \frac{2+\lambda-ij}{2}, \frac{3}{2}, -k^2\right] \qquad (\mathrm{Re}\,\lambda > -2), \qquad (3.7a)$$

$$\int_0^\infty dy\, y^\lambda K_{ij}(y) \cos ky = 2^{\lambda-1} \left|\Gamma\left(\frac{1+\lambda+ij}{2}\right)\right|^2 \times$$

$$F\left[\frac{1+\lambda+ij}{2}, \frac{1+\lambda-ij}{2}, \frac{1}{2}, -k^2\right] \qquad (\mathrm{Re}\,\lambda > -1). \qquad (3.7b)$$

In the integral (3.6a) we must use Eq. (3.7b) with $\lambda = -1$, which is outside the stated range of validity. So the convergence of this integral deserves close attention. Let us write

$$I(j,k;\lambda) = \int_0^\infty dy\, y^\lambda\, \Psi_j(y)\, e^{-iky}$$

$$= \int_{-\infty}^\infty du\, e^{(1+\lambda)u}\, \Psi_j(e^u)\, \exp(-ike^u)/\sqrt{2\pi}\,.$$

Near $u = -\infty$ the integrand behaves like

$$-\frac{1}{\pi} e^{(1+\lambda)u} \sin ju,$$

and as $u \longrightarrow +\infty$ it falls off faster than exponentially (see Eqs. (2.5) and (2.7)). In effect we are taking Fourier transforms of a family of tempered distributions which has a limit as $\lambda \longrightarrow -1$

from above. We can write $I(j,k;\lambda)$ as

$$-\frac{1}{\pi}\int_{-\infty}^{0} du\, e^{(1+\lambda)u} \sin ju \tag{3.8}$$

plus a function of j which is smooth in the whole range $-\infty \leq j \leq \infty$ (being the sum of two Fourier transforms of a distribution of rapid decrease). As $\lambda \longrightarrow -1$, the distribution (3.8) converges to the principal value of $1/(\pi j)$. This pole is included in the expression obtained by setting $\lambda = -1$ directly in Eq. (3.7b) (see Eq. (3.9a) below); there is no need to add a singular term. So Eq. (3.7b) remains valid in a distribution sense for $\lambda = -1$, and hence Eq. (3.6a) defines $A(j,k)$ as a distribution in j for fixed k, which happens to be a smooth function in both j (away from 0) and k. Of course, the same is true of B, for which there is no problem of convergence in the integral.

The expressions for A and B can be reduced to elementary functions using formulas (8.332.1,2) of [Gradshteyn-Ryzhik] (p. 937), (15.1.11,12) of [N.B.S.] (p. 556), and (15.2.20,21,25) of [N.B.S.] (p. 558). The results are

$$A(j,k) = \frac{1}{2}[\omega_k \pi \sinh \pi j]^{-1/2} \Big[\cosh\frac{\pi j}{2} \{[\ +\ ]^{ij} + [\ -\ ]^{ij}\}$$

$$- \sinh\frac{\pi j}{2} \{[\ +\ ]^{ij} - [\ -\ ]^{ij}\}\Big], \tag{3.9a}$$

$$B(j,k) = \frac{1}{2}[\omega_k \pi \sinh \pi j]^{-1/2} \left[\sinh \frac{\pi j}{2} \{[+]^{ij} + [-]^{ij}\} - \cosh \frac{\pi j}{2} \{[+]^{ij} - [-]^{ij}\}\right], \qquad (3.9b)$$

where

$$[\pm] = \sqrt{1 + k^2} \pm k = [\mp]^{-1}. \qquad (3.10)$$

The relation

$$\frac{d}{dk}\left[\sqrt{\frac{\omega_k}{j}}\, A(j,k)\right] = -i \sqrt{\frac{j}{\omega_k}}\, B(j,k), \qquad (3.11)$$

suggested by Eqs. (3.6), is satisfied by these functions. Finally, one has (for general m)

$$U(j,k) = [2\pi \omega_k (1 - e^{-2\pi j})]^{-1/2} \left[\frac{\omega_k + k}{m}\right]^{ij}, \qquad (3.12a)$$

$$V(j,k) = [2\pi \omega_k (e^{2\pi j} - 1)]^{-1/2} \left[\frac{\omega_k + k}{m}\right]^{ij}. \qquad (3.12b)$$

Then we have

$$\int_{-\infty}^{\infty} dk \int_0^{\infty} dj\, |V(j,k)|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \int_0^{\infty} dj\, [\sqrt{k^2 + m^2}\, (e^{2\pi j} - 1)]^{-1},$$

which diverges at large k and at small j. Therefore, V is not a Hilbert-Schmidt operator.

4. Can We Live With Two Different Quantizations?

In Chapter VII it was pointed out that the formalism of field theory does not uniquely determine an explicit representation of the fields as operators. (In the general case this ambiguity has little to do with freedom in the choice of coordinate system. It is misleading, therefore, to speak of the problem we are facing as primarily a breakdown of general covariance.) In the static case there appeared to be an obvious choice of representation, based on the notion of positive frequency. Now, however, we have seen that in a few special cases, such as flat space and de Sitter space, the availability of several rival static coordinate systems makes even the natural prescription for quantization in a static universe ambiguous. For the field in a certain region of Minkowski space a heretical quantization has been proposed. What attitude should one adopt toward it?

One possibility is to reject it outright. One could argue as follows: We understand the free field in flat space. The physically relevant representation of the fields is the ΦOK representation; the definitions of the vacuum and the number-of-particles operator are unambiguous. If any other proposed theory disagrees with this one, so much the worse for that theory!

More specifically, one might object that the Rindler coordinate system covers only a part of space-time; that it has a singularity at $z = 0$ which has nothing to do with the intrinsic structure of the space; that it is not an inertial frame, in the sense that the curves of constant z are not geodesics[9]. For these reasons, the critic would argue, it is not surprising that a naive imitation of the quantization of the free field leads to unphysical results in this context.

This, however, would be a very unwelcome conclusion. All three of the above aspersions upon the Rindler coordinate system also apply to the Schwarzschild system. The analogy between these two situations has already been pointed out in Sec. IX.1. It was argued there that Rindler space must be regarded as a test case for any general theory of field quantization in static metrics. In the case of a general static metric we do not have an underlying flat space to tell us what the "right answer" is. The Fock quantization (in the sense of Chapter VIII) is a natural generalization from the theoretical ideas which have evolved in the study of free fields and of external potential problems. If it is wrong, the interpretation, both physical and mathematical, of the formalism of quantum field theory is left obscure.

The suggestion that a trustworthy field quantization can only be performed on "the whole space" is especially frightening.[10] Many of the solutions of Einstein's equations

---

[9] This last objection does not apply to the situation in de Sitter space, illustrated in Fig. 13. There one can set up many different static coordinate systems based on timelike geodesics, which can be regarded as the worldlines of unaccelerated observers. These are related to one another much as Rindler coordinates are related to Cartesian, and none of them has any

studied in modern general relativity are quite complicated (multiply connected, and so forth). Must one really construct a quantized field on the entire manifold in order to treat exactly the particle phenomena in a small region? Furthermore, given a manifold with a metric, it is sometimes hard to say whether it constitutes "the whole space" (see, e.g., Geroch (1968)).

On the other hand, the conclusion is hard to accept on general physical grounds. Quantum mechanics is physically a local theory; it has to do with phenomena that happen on a microscopic scale. As argued in the Introduction and Sec. IV.3, it is hard to believe that the global structure of space has more than a negligible effect on any quantities that are physically observable. But the construction of a Fock space is inherently global, because it is based on momentum space (the Fourier decomposition).[11] If, as seems to be the case, we run into trouble when we try to do quantum theory in terms of local coordinate patches, then perhaps that is evidence for the inadequacy of our present formulation of quantum field theory, rather than for a breakdown of the principle of general covariance.

These considerations suggest that before we discard the theory of the free field in the Rindler wedge as physically wrong, we should try to make sense out of it, adjusting our

---

reason to be preferred to the others.
[10] We shall return to this subject in Secs. IX.6-7.
[11] See further remarks in Sec. X.7.

preconceived ideas if necessary. If we succeed, we may learn something from this model that will help us to understand field theory in nonstatic spaces, the subject of the next chapter. Such a reinterpretation, if it is not to be manifestly inconsistent with the established interpretation of the ΦOK representation, must somehow weaken the direct physical significance of the quanta of the general Fock representation.

Let us start from the realization that these quanta cannot be the basic observables of the theory. What, then, is observable? On physical grounds one can argue that observations take place via interactions of the system studied with other physical systems; therefore, one ought to study the currents by which our field might couple with other fields (the expressions through which the field can occur in possible interaction terms in the Hamiltonian or Lagrangian). In particular, in cosmological and astrophysical problems the energy-momentum tensor, which couples to the gravitational field, is presumably the most important object. Unfortunately, as will be demonstrated in the next section, the ambiguities in particle creation and annihilation operators carry over to the definition of at least some current operators.

We are left with the fields themselves as observables. The time evolution of the field operators (or their expectation values) from given initial values is given by a classical formula (Eq. (VII.4.5)), independent of representation. Similarly, the expectation value of the product of n field operators is a distribution which satisfies the field equation in each of its n arguments, and hence is determined by its initial values within a domain of dependence. This suggests that, if the fields alone are the basic observables, it is not necessary to choose a representation. One can think of the fundamental dynamical problem as the prediction of the outcome of field measurements at later times on the basis of known results of measurements at earlier times. A quantum state is just an intermediary apparatus which summarizes (idealized) earlier measurements. No practical set of measurements can completely determine the state, or distinguish between inequivalent representations (see below). It is proposed, then, that we should reject the demand for a unique "physical" representation. That is, we should be prepared to admit all representations as possibly physically relevant, and to give up the search for some absolute definition of the number of particles in a general space-time. (In special situations some analogue of the familiar particle notion may still have a limited physical significance -- see below. See also Secs. IX.7 and X.7.)

Of course, this point of view leaves the interpretation of the field theory in terms of observable quantities still quite vague. We are used to thinking of quantum processes in terms of particles. In practice we never measure field strengths as such (except for macroscopic electromagnetic and gravitational fields, which are outside the quantum domain). But to label the states in terms of a particle structure requires a definite representation of the canonical commutation relations, either for the fields of the Heisenberg picture themselves or for asymptotic free fields.[12] There is much work to be done in clarifying the physical interpretation of quantum field theory in situations where the asymptotic particle interpretation does not apply. This dissertation claims only to pose and clarify the problem and to suggest a program for future research.

Although the relation of field operators to particle detectors is not obvious, we do know that the fields in a region ($\int dx\ \phi(x)\ f(x)$, support of f in the region) have something to do with experimental operations performed in the region. If the region is rather small, this localization may be the most relevant fact about the experimental operation.[13]

Thus we have been led to the approach to quantum physics in terms of local algebras, proposed in the fundamental papers of Segal (1947) and Haag and Kastler (1964). Unfortunately for our purposes, much of the recent work in this area depends crucially on the assumptions of Poincaré covariance (covariance under space-time translations and Lorentz transformations) and the existence of an invariant vacuum state, ingredients which are missing here.

---

[12] For the purposes of interpretation we are concerned not just with the unitary equivalence class of the representation but also with the identification of the vacuum state, the one-particle states, etc. (cf. Sec. X.2 below).

[13] These ideas are basic to the work of R. Haag *et al.* (see Araki and Haag (1967) and earlier papers and lecture notes of Haag) on the relation between local fields and asymptotic particle observables.

However, the basic ideas behind the algebraic approach are applicable.[14] A state of a physical system is taken as any functional ω on the algebra of observables which can be interpreted[15] as an expectation value; it does not have to be related to a vector in a particular Hilbert space. Each state is related as a vector state,

$$\omega(A) = \langle\Psi|A|\Psi\rangle \qquad (|\Psi\rangle \in \mathcal{H}), \quad (4.1)$$

to some representation of the algebra as operators in a Hilbert space, but there are many inequivalent representations. It is argued, however, that all the faithful representations are physically equivalent, because every representation contains a state which is consistent with any given set of practical observations.[16] (This means that, given a list of results of a finite set of measurements, these results can be reproduced to arbitrary accuracy by a weighted average (density matrix) of the expectation values with respect to certain vectors in any one of these Hilbert spaces.) The mathematical basis for this claim is

---

[14] There is space here only for a sketchy discussion. The reader is urged to read the paper of Haag and Kastler (1964).
[15] In particular, a state is required to be positive: $\omega(A^\dagger A) \geq 0$ for all A in the algebra.
[16] For a similar conclusion in the context of the canonical formalism see Komar (1964). (Note, however, that in this paper the work of Wightman and Schweber (1955) was misinterpreted: in the standard approach to field theory the Hilbert space is separable, consisting of just one of the equivalence classes referred to by Komar. Note also that the representations which Komar considers explicitly are not all of the representations of the canonical commutation relations, and not even all of the tensor product representations.)

a theorem proved by Fell (1960).

To forestall possible misunderstanding, it should be emphasized that the doctrine of physical equivalence of representations does not say that the vacuum of the Rindler Fock representation is an approximation to the vacuum of the ΦOK representation. These are two different states, yielding different expectation values for operators, and corresponding to different notions of particle. The claim is that the Rindler representation contains other vectors which approximate the ΦOK vacuum with respect to any given finite set of observables.

The preceding discussion has emphasized the arbitrariness in the choice of representation. The other side of the coin is that frequently one representation recommends itself as the best one to work with, because it has some especially nice feature.[17] For instance, in Poincaré-covariant theories the representation generated by an invariant vacuum state plays a distinguished role. This is in analogy to the observation that when studying a physical theory in flat space it would be folly to refuse to use Cartesian coordinates for a calculation because of the possibility of writing all the equations in a generally covariant form.

In the external gravitational field problem there are several special cases where a "nice" representation is suggested by special properties of the metric. If the metric is

---

[17] The author is grateful to J. E. Roberts for a conversation in which he emphasized this point.

asymptotically flat, or if it becomes actually flat in the remote past and future, we have an asymptotic particle interpretation (see Secs. VII.7 and VIII.4), and the natural representations to use are the <u>in</u>- and <u>out</u>-representations, in which there are finitely many particles coming in or coming out in each state. On the other hand, if the space has a symmetry group, we should consider a representation which is invariant under the group. (That is, in this representation the symmetry is implemented by a unitary operator.) In particular, if there is a time translation group, we have the static Fock representation of Chapter VIII. Whatever its relation to physical observables may be, in the general static case one might expect the "particle" structure based on the existence of a timelike symmetry to be the most convenient way to label the states of the theory.[18] (Rindler space is a very special case in which there is another way of looking at the space which makes additional symmetries manifest and, consequently, leads to a more useful notion of particle.)

---

[18] However, a contrary view will be tentatively developed in Secs. IX.7 and X.7. The position of the author is that our present theoretical resources do not allow us confidently to generalize quantum field theory to curved space-time. A given Riemannian geometry may suggest several methods of quantization, perhaps none of them entirely satisfactory. A decision among these approaches cannot be made on the basis of pure thought. A great deal of research is needed on particular models, to clarify how the predictions of these approaches differ and how they compare with astrophysical observation. It is to be hoped that a coherent theory will develop, including (1) an understanding of the relation of the mathematical apparatus to observation, and (2) a practical understanding of what representation of the field algebra it is either necessary or prudent to use for a given purpose.

Finally, in the case of a nonstatic universe, one might be able to define particle observables in such a way that in one representation there are finitely many particles present in a finite volume at each time (cf. Secs. X.5-6 below).

5. Bounded Observables, Currents, and the Energy-Momentum Tensor.

In this section we shall push a little farther the study of the association of field operators with physical observables, on which the algebraic interpretation of the field theory formalism espoused in the last section depends. First, the rigorous technical work on this subject deals with algebras of bounded observables (C*-algebras), whereas the field operators (whether smeared in s or s + 1 dimensions) and the annihilation and creation operators are necessarily unbounded. For a neutral scalar field there is a variety of ways of defining from the field a C*-algebra of bounded observables[19]; the matter will not be discussed further here.

When charged (complex) fields or spinor fields are under consideration, however, a difficulty which is more a problem of principle arises. Such fields cannot be observables at all, because they do not commute with the superselection rules (see [Streater-Wightman], Sec. 1.1). In the C*-algebra approach a remedy is to form a C*-algebra from all the fields and then to

[19] See, e.g., [Segal], Kastler (1965), Manuceau (1968), Dell'Antonio (1968), Wilde (1971), Slawny (1972).

distinguish a subalgebra of observables by their formal property of commuting with the relevant gauge transformations or superselection operators (see, e.g., Wilde (1971), Chapter 2).

On a more intuitive level, however, the observables associated with charged and spinor fields are usually assumed to be currents -- quadratic (or higher-order) combinations of the fields and their derivatives. Examples are the charge-current vector (VII.5.4) of a charged scalar field, the famous five tensors formed from the Dirac field ([Messiah], Sec. XX.14), and the energy-momentum tensor[20], $T^{\mu\nu}(x)$. These quantities are also important because they appear in the interaction terms of the Lagrangians and Hamiltonians of the nonlinear theories of interacting fields. Indeed, they are assumed to be observable precisely because it is through them that a field interacts with other physical systems -- such as experimental apparatus.

In astrophysics and cosmology, of course, the energy-momentum tensor is surely the object of greatest interest, since it is through it that the matter represented by the quantized field interacts with the gravitational field, according to the theory of general relativity. Observations of the

---

[20] In general one must distinguish between the canonical energy-momentum tensor suggested by the canonical formalism, of which the Hamiltonian density (VII.1.6) is one component, and the symmetrized, covariant tensor appropriate to general relativity, which is obtained by variation of the action with respect to the metric tensor (Belinfante (1940)). For our neutral scalar field with minimal gravitational coupling these are the same. For the rival neutral scalar field theory (see Secs. V.2 and VII.1) they are already different (Chernikov and Tagirov (1968)).

presence of matter through its gravitational effects are more likely than particle detection events of the kind familiar in terrestrial laboratories. So one might be happy to forego particle observables if the energy-momentum tensor could be unambiguously defined. On the other hand, $T^{\mu\nu}$ compares favorably with the field itself as a plausible physical observable (even on the microscopic level), given the absence of "pion field-strength meters". Can one, then, regard this tensor field as the basic observable in an algebraic (representation-free) theory, rather than the field itself as in Sec. IX.4?

Unfortunately, a current naively defined in terms of products of fields does not, in general, define a finite operator-valued distribution in most representations of the fields. The obvious extension of the procedure normally used to make sense out of currents in the ΦOK representation of a free field is the following: In a representation of the Fock type (i.e., characterized by annihilation operators which annihilate a certain state of no quanta) one is to <u>normal order</u> the formal expression for the current by changing all terms of the form $a_i a_j^\dagger$ to $a_j^\dagger a_i$ and discarding any constant terms. Then the expression will formally annihilate the no-quantum state, and it will now make sense as a bilinear form on a dense domain. The important point is that this definition obviously depends upon the representation (and upon the no-quantum state). It does not provide an intrinsic algebraic object.

For instance, consider the time-time component of the energy-momentum tensor of the free scalar field in two-dimensional Minkowski space,

$$T^{00}(x) = \frac{1}{2}:[\pi^2 + (\frac{\partial\phi}{\partial x})^2 + m^2\phi^2]:, \qquad (5.1)$$

where the colons indicate normal ordering with respect to some representation. We wish to compare the normal orderings corresponding to the ΦOK and the Rindler representation (at $t = v = 0$). Note that in both cases we consider a component of the tensor T with respect to the same field of basis vectors at each point (tetrads); for convenience the familiar orthonormal basis associated with Cartesian coordinates has been chosen. The difference between the two definitions of T^{oo} under discussion has nothing to do with the transformation of tensor components from one coordinate system or frame to another. (The contravariant component T^{oo} with respect to the tetrads canonically associated with the Rindler coordinate system would be a <u>different</u> physical quantity from that of Eq. (5.1), and we are not interested in it.)

The comparison proceeds in analogy to a simpler case which is discussed in detail in Sec. G.2. The fields in Eq. (5.1) are expressed in terms of the operators b_k and $b_k^\dagger$ (cf. Eqs. (3.1)), and the resulting expression is normal ordered. (This is the standard energy density for the free field.) Then the b's are re-expressed in terms of the a's.[21] Of course, no true inverse of Eq. (3.2) exists, since the a's span only a subalgebra of the whole equal-time scalar field algebra (the part associated with the positive axis in x-space). One calculates

$$b_k = \int_0^{\infty} dj\ U^*(j,k)\, a_j - \int_0^{\infty} dj\ V(j,k)\, a_j^{\dagger}$$

$$+ \sqrt{\frac{\omega_k}{2}} \int_{-\infty}^{0} dy\ e^{-iky}\, \phi(y) + \frac{i}{\sqrt{2\omega_k}} \int_{-\infty}^{0} dy\ e^{-iky}\, \Pi(y). \quad (5.2)$$

When Eq. (5.2) is substituted into the expression for $T^{00}(x)$ ($x > 0$), all the terms involving negative y cancel, as they must, since $T^{00}(x)$ depends only on the field and its derivatives at x. What is left is a bilinear expression in the Rindler operators which is not normal ordered. Our interest centers on the constant[22] which must be subtracted to make the expression coincide with the Rindler-normal ordered version of Eq. (5.1). Formally this is

$$\langle T^{00}(x)\rangle = \frac{1}{2} \int_0^{\infty} dj\ \{ \int_{-\infty}^{\infty} dk \int_{-\infty}^{\infty} dl\ (\omega_k \omega_l)^{-1/2}\ e^{i(k-l)x} \times$$

---

[21] One could, of course, work from the other direction, using the transformation (3.2). The method chosen here leads to a result free of Bessel functions.

[22] By this is meant a c-number (multiple of the identity operator); it may depend upon x.

$$[(\omega_k \omega_l - kl - m^2) U(j,l) V(j,k)$$

$$+ (\omega_k \omega_l + kl + m^2) V^*(j,l) V(j,k)]\}, \quad (5.3)$$

where U and V have been calculated in Sec. IX.3.

Eq. (5.3) is not a determinate expression. (It is not even unambiguously infinite -- its divergent parts conceivably could formally cancel, as in the integral $\int_{-\infty}^{\infty} k\, dk$.) It can be regularized, in analogy to Eq. (G.2.6), with the aid of smooth test functions. But it is hard to tell whether even this "smeared" quantity is zero, finite, or infinite.[23] However, the burden of proof is certainly upon him who would assert that it is zero. It seems most implausible that the energy density can be unambiguously defined by normal ordering.

In Sec. IX.7 another contrast of quantizations of the free field will be developed (Euclidean space vs. a finite box). In Sec. G.2 it is shown that the difference between the local energy densities of these two representations is infinite, according to a reasonable interpretation of this statement.

For still another example we anticipate the results of the next chapter. There it will be necessary to consider different representations at different times, with their respective annihilation-creation operators related by equations

---

[23] An apparent divergence at the lower limit of the j integration is seen upon closer inspection to disappear for test functions with support on the strictly positive axis.

of the form (X.5.6) (where $\alpha(k)$ and $\beta(k)$ depend on the initial and final times, t_1 and t_2). One easily calculates that $T^{00}(x)$, normal ordered at t_1, has with respect to the no-particle state at t_2 the expectation value

$$\langle T^{00}(x)\rangle = \sum_k \omega_k |\beta(t_2, t_1; k)|^2 . \qquad (5.4)$$

This quantity is generally nonzero[24] and possibly infinite.

Hawking (1970) has pointed out that if $:T^{00}:$ is defined by normal ordering with respect to a different no-particle state at each time, then it is not obvious that it will satisfy the usual divergence condition,

$$\nabla_\mu T^\mu = 0. \qquad (5.5)$$

His argument that this equation __must__ fail is erroneous, however, since it assumes that an expectation value $\langle\psi|T^{\mu\nu}|\psi\rangle$ must satisfy the classical condition

$$T^{00} \geq |T^{\mu\nu}|, \qquad (5.6)$$

and this is not generally true in quantum field theory.[25]

[24] This is true of any theory in which particle creation occurs, not only that of Sec. X.5. See the discussion in Sec. X.6.

[25] That nonpositivity of the energy density is inevitable in a field theory of the usual type was proved by Epstein __et al.__ (1965).

Indeed, for the free field let $\Psi = |0\rangle + \lambda|2\rangle$, where $|0\rangle$ is the vacuum and $|2\rangle$ is a two-particle state. Then $\langle\Psi|:T^{00}(x):|\Psi\rangle$ is of the form $\lambda A(x) + \lambda^2 B(x)$, where A is contributed by the aa and $a^\dagger a^\dagger$ terms in $:T^{00}:$ and B by the $a^\dagger a$ terms. (A, B, and λ are real.) Choose Ψ so that $A \neq 0$. For λ sufficiently small and of opposite sign from A, $\langle\Psi|:T^{00}(x):|\Psi\rangle$ is negative! That Hawking's argument should be regarded as disproving Eq. (5.6) rather than Eq. (5.5) has also been pointed out by Zel'dovich and Pitaevsky (1971), who have shown in perturbation theory how Eq. (5.6) is violated during particle creation in a universe with weak but nonstatic curvature.

Very recently Zel'dovich and Starobinsky (1971) have used a renormalization technique to define (without reference to normal ordering except at an <u>initial</u> <u>time</u>) a finite energy-momentum tensor which obeys Eq. (5.5). Although the rationale for their procedure is far from clear to the present author, this work may point the way to a solution of the problem of defining $T^{\mu\nu}$ as far as practical calculations are concerned.

It is noteworthy that the free-field momentum density components, $T^{0j}(x)$, seem to be largely immune to the ambiguity noted here for T^{00}. When field expansions are substituted into

$$T^{01} = \frac{1}{2}\left[\frac{\partial\phi}{\partial x}\frac{\partial\phi}{\partial t} + \frac{\partial\phi}{\partial t}\frac{\partial\phi}{\partial x}\right], \tag{5.7}$$

the result is <u>automatically</u> <u>normal</u> <u>ordered</u> in all representations the author has had reason to consider. For the standard Fourier decompositions this well-known result (which is due to cancellations from positive and negative k) is independent of the mass of the field (see Sec. F.4) and of whether the quantization is performed in a box or in infinite space (see Sec. IX.7 and Appendix G). It also holds for the Rindler representation, where one finds for the vacuum expectation value

$$\langle T^{01}\rangle = \frac{1}{2}\left\{\int \frac{dj}{\sqrt{2j}}\,\frac{\partial \Psi_j}{\partial z}\,\frac{i}{z}\sqrt{\frac{j}{2}}\,\Psi_j(z) + -\frac{i}{z}\int dj\,\sqrt{\frac{j}{2}}\,\Psi_j(z)\,\frac{1}{\sqrt{2j}}\,\frac{\partial \Psi_j}{\partial z}\right\}$$

$$= 0. \qquad (5.8)$$

The same thing happens for the charge-current density of a charged field, if one writes the time component in the symmetrized form

$$j^0 = \frac{i}{2}\left[\phi^{\dagger}\frac{\partial \phi}{\partial t} - \frac{\partial \phi^{\dagger}}{\partial t}\phi + \frac{\partial \phi}{\partial t}\phi^{\dagger} - \phi\frac{\partial \phi^{\dagger}}{\partial t}\right]. \qquad (5.9)$$

Thus Eqs. (5.7) and (5.9) may provide intrinsic definitions of these quantities, which yield the same result as normal ordering in all representations which are "natural" in some sense. Since the Bogolubov transformation for a charged field analogous to Eq. (3.2) relates particle annihilation operators to <u>antiparticle</u> creation operators, it is not surprising that the vacuum states of the two representations are charge-free in the same sense. Similarly, the result for the momentum in the various ΦOK-like representations may be attributed to the fact that in such transformations as (F.4.8) and (G.2.3) the new annihilation operators involve creation operators only for quanta with the opposite momentum.[26] The result (5.8) is rather surprising, however, since the transformation (3.2) is not manifestly diagonal in the momentum, and, in fact, the adoption of the Rindler coordinate system disrupts and obscures the translation invariance of the space.

6. Geodesic Completeness and the Feynman Path Integral.

The reaction of many people to the troublesome development described in this chapter has been that the responsibility for it somehow lies with the fact that the Rindler coordinate system does not cover the whole Minkowski space-time. The integral transformations (such as Eq. (1.11) and the Fourier transformation) involved in the decomposition of the field into modes (on which the Fock quantizations are based) are global operations. It is not surprising, therefore, that widely separated regions of space turn out to be mixed up with each other in the construction of quantum fields. It is urged that field quantization should be attempted only on geodesically complete manifolds (see Sec. III.5), or at least that quantization on an incomplete space should be regarded as a distinct physical situation from quantizing on a complete space in which that region is embedded and then restricting attention

[26] Note that the phenomenon seems to be linked to physical quantities which, unlike energy, can be carried with opposite signs by particles.

to the region. This viewpoint, if established, would refute our assumption (cf. Sec. VIII.1) that only completeness of the Cauchy type is relevant to the dynamics of fields. For reasons mentioned in Sec. IX.4, working only with complete spaces would be a considerable nuisance in practice. Of course, that is not a convincing argument against its necessity!

However, the argument of Sec. IX.3 is of a very general type -- we shall meet it again in Sec. X.2 -- and it seems that in equations like (3.2) the vanishing of the kernel of the second term will be very much the exception rather than the rule. It seems to the author, therefore, that the local distortion of the "(3 + 1)" structure of space-time, rather than the global mutilation of the space, is sufficient to lead to the phenomenon. Unfortunately, there is no way to test this claim, since apparently there is no example of a manifold with two linearly independent globally timelike Killing vectors. (A timelike Killing vector is needed to make the metric take a static form, so that a criterion is available to define a Fock representation uniquely.) On the other hand, admittedly, one can give an example (Sec. IX.7) which shows that a change of the global structure without a distortion of the time scale is sufficient to change the definition of the vacuum.

An argument in favor of the "global" viewpoint has been offered by L. H. Ford (private communication). If one formulates quantum particle dynamics, following Feynman, in terms of a sum over virtual paths, one would expect to have a nonzero (although small) contribution from partially spacelike paths. In this sense it is possible for the quantum particle to leave and re-enter the Rindler wedge, despite the latter's Cauchy completeness. One might conjecture (pending an explicit calculation, which the author has not attempted) that the entire difference between the single-particle theories associated with the two quantizations is due to neglect of these paths in one case.

This suggests that it may be worthwhile to generalize to curved space-time Feynman's path-integral construction of the propagator for the Klein-Gordon equation (Feynman (1950), Appendix A), and to study its implications in various models. Such a project is beyond the scope of this thesis, but some of the things which might be investigated in the future can be outlined here.

Since the Lagrangian

$$\mathcal{L} = -\frac{1}{2} g_{\mu\nu} \frac{dz^{\mu}}{du} \frac{dz^{\nu}}{du} \tag{6.1}$$

yields the classical equation of motion of a particle, Eq. (III.4.1), it seems clear[27] that the generalization of Feynman's Eq. (4A) should be

---

[27] Since the quadratic terms in the Lagrangian have nonconstant coefficients, different ways of breaking the action integral into steps may yield different results -- see Feynman (1948), pp. 376-377.

$\phi(x_n, u_0) =$

$$\int \prod_{j=0}^{n-1} \frac{\sqrt{|g|}\, d^4 x_j}{4\pi^2 \varepsilon^2 i} \exp\{-\frac{i}{2\varepsilon} \sum_{j=1}^{n} g_{\mu\nu} (x_i^\mu - x_{i-1}^\mu)(x_i^\nu - x_{i-1}^\nu)\} \, \phi(x_0, 0)$$

$$\equiv \int d^4 x_0 \, K(x_n, u_0; x_0, 0) \, \phi(x_0, 0) \qquad (6.2)$$

However, there seem to be obstacles in showing that this formula is equivalent to a proper-time version of the wave equation (analogue of Feynman's Eq. (2A)). If one has such a propagator in u-space, the physically relevant propagator is to be found by Fourier transforming:

$G_F(x_2, x_1) =$

$$\int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} du_0 \, \exp[-im^2 (u-u_0)/2] \, K(x_2, u; x_1, u_0). \qquad (6.3)$$

This appears to be a generally covariant construction.

Then it would be interesting to study the following questions:

(1) If the metric is static, does $G_F(x_2, x_1)$ coincide with $G_F(\Psi_0, \Psi_0; x_2, x_1)$, defined in Eq. (VII.4.18), where Ψ_0 is the Fock vacuum? (In particular, is Ford's conjecture about Rindler space true?)

(2) In the general case, define $G^{(+)}$ in terms of G_F and the intrinsically defined G^{adv} by Eq. (VII.4.21). Is $G^{(+)}$ of positive type, so that the reconstruction theorem can be used to construct a representation with a cyclic vector Ψ? If so, what is the physical significance of Ψ?

(3) On the other hand, if the metric is asymptotically static (see Sec. X.3 below), is $G_F(x_2,x_1)$ proportional to $G_F(\Psi_0^{out},\Psi_0^{in};x_2,x_1)$ (definition of Eq. (VII.4.18))? Cf., e.g., Wightman (1968), p. 296. If so, does this fact have any physically sensible extension to the general case?

Similar questions could be asked of other proposed definitions of the Feynman propagator -- e.g., that of De Witt (1963), pp. 738-741, or that of Duistermaat and Hörmander (1971).

7. Further Thoughts on the Particle Concept in Static Spaces.

In this chapter and Sec. VIII.6 we have considered certain proper subregions of Minkowski space and de Sitter space and have tried to treat them as "universes" in their own right, because they are causally closed from the point of view of the classical theory of fields (the Cauchy problem). We have had cause to wonder whether this procedure is legitimate, and whether the theory thus obtained is really physically equivalent to a theory which treats the entire space. Could we be inadvertently imposing some "boundary condition", so that the theory describes the behavior of matter in an isolated geodesically incomplete space, but is not appropriate for a region which is actually part of a larger space?

To shed some light on these questions let us consider the representation of the algebra of fields in a bounded region of space-time when it is regarded as a part of (a) an ordinary Minkowski space-time, and (b) a "box" universe of length L with periodic boundary conditions. (As usual we consider a two-dimensional space without loss of generality.) Of course, these are <u>different</u> universes, and one would not expect the quantized fields in the entire region $\{(t,x) \mid 0 \leq x < L\}$ to be the same in the two theories. In the box (or torus) case a boundary condition holds which means, physically, that a particle which reaches one end of the box does not disappear into another region of space, but re-enters at the other end of the box.

Nevertheless, this large-scale behavior should be irrelevant to what happens inside the domain of dependence, D, of an interval

$$I: \quad a < x < b \qquad (0 < a,\ b < L). \quad (7.1)$$

(D is a diamond-shaped region, as in Fig. 6.) I is a Cauchy surface for D. Thus, if the point of view tentatively espoused in Sec. IX.4 is correct, the field in D presents a self-contained dynamical problem. The outcome of measurements in D should be predictable (in the statistical sense in which predictions are possible in quantum theory) on the basis of measurements in (or near) I. The dynamics within this region should depend only on the field equation, the canonical commutation relations, and whatever local interaction between the field and a measuring apparatus gives the theory its physical content.

Of course, the last of these, the measurement interaction, is the weakest link in the chain. The whole argument hinges on the assumption that <u>observations and other experimental operations performed in a region correspond strictly to field operators</u>[28] <u>whose test functions have support in the region</u>. (In contrast, it necessarily follows that localized <u>particle</u> observables, like the Newton-Wigner operator (see Sec. VIII.4), correspond to <u>nonlocal</u> measurements and state preparations.[29]) This is the assumption which is usually made; it is the motivation for the axiom of local commutativity (or anticommutativity) of fields. Conversely, if it does not hold, one would expect violations of the principle that information does not travel faster than light. It is not, of course, an unchallengeable article of faith. If it turns out to be impossible to make physical sense out of field theories interpreted in this way, we will have to change our way of thinking. In the meantime, however, the statement stands as a description of the type of theory we are trying to construct.

---

[28] More precisely, to <u>observable</u> operators in the algebra generated by the fields of the region -- see Sec. IX.5.
[29] It is probably more realistic to say that, to the extent that these "observables" can be measured at all, they can only be measured <u>approximately</u>, by local operations.

So the infinite- and finite-space quantizations of the free field provide two representations of the same algebra, the algebra of fields in D; and we are claiming also that the physical situation, as far as observations entirely within D are concerned, is also in some sense the same in the two contexts. It is interesting, therefore, to compare the two representations. We shall see that they are not the same.

As explained in Sec. VIII.3, the representation is determined by the two-point function (VIII.3.13). For the box the two-point function is

$$\sum_k (2\omega_k)^{-1} \varphi^{ikx_2} \varphi^{-ikx_1} \exp\{-i\omega_k(t_2-t_1)\} \qquad (\varphi = e/\sqrt{L}), \quad (7.2a)$$

the sum being over the lattice $\{2\pi n/L\}$, and for infinite space (the ΦOK representation) it is

$$\int \frac{dk}{2\omega_k} \varphi^{ikx_2} \varphi^{-ikx_1} \exp\{-i\omega_k(t_2-t_1)\} \qquad (\varphi = e/\sqrt{2\pi}). \quad (7.2b)$$

These distributions are not the same, even for test functions with compact support inside D (see Appendix G).

The conclusion is that for the fields in D we have two distinct vacuum states[30], each of which, by virtue of its origin in a respectable free field theory, has a good claim to be a state in which no particles are present. What is the physical

[30] Here "state" is to be understood in the algebraic sense of a linear functional. Actually, of course, we have a continuous

origin of these two different notions of (absence of) particles? How can this ambiguity be reconciled with the fact that particle detection seems to be a well-defined experimental concept?

One way out is to conclude that the experimental act of particle detection is actually different in the infinite and the finite universe -- that it somehow inherently involves the whole space. Indeed, it is known (as a corollary of the Reeh-Schlieder theorem) that an operator which annihilates the vacuum cannot be a member of the local algebra associated with a bounded region. For this reason Araki and Haag (1967) and Steinmann (1968) explicitly associate a particle detector with an operator which is only quasilocal (but annihilates the vacuum exactly).

However, it seems to the present author that this approach to the problem is backwards. Instead of taking the hallowed concept of the vacuum to be the fundamental starting point, one should model the measurement process in terms of literally local operators and study to what extent particle concepts (such as a no-particle state) can then be extracted. Of course, not just any local operator will do. The intuitive notion of a particle must somehow be input, to lead us to the kind of structure we want to get out. We know by observation (describable in crude terms) that there are entities which move roughly in straight lines, except when they interact with each other or with macroscopic bodies (cf. Steinmann (1968), Sec. 1).

---

family of states, one for each value of L greater than the length of I.

In an attempt to refine the notion of observation of particles, we are led to the idea of a detector as a system which interacts with particles in its vicinity, but makes no response when there are no particles present.

On the basis of what was said above, it seems probable that no such ideal detector is possible -- that is, every real detector (which is surely contained in some finite region) has some probability of making a response, whatever the state of the quantum field system.[31] However, we can understand the definition of the no-particle state in terms of the following analogy. The classical concept of length is abstracted from the stability and mutual consistency of the ways in which it is observed to be possible to juxtapose various material objects ("measuring rods"), although each of these objects is only imperfectly rigid when re-examined in terms of this very definition of length. Similarly, the concept of time is abstracted from observational comparison of many natural processes which individually manifest approximate regularity.[32] In the same way, various systems (detector candidates) are observed to behave roughly in the way expected of an ideal detector. In particular, under certain experimental conditions

---

[31] Probably for any actual detector this effect, which is inherent in quantum field theory, is lost in the noise of the concrete experimental arrangement.
[32] In particular, when measuring very short times (or lengths) one encounters limitations due to the quantum uncertainty principle; these are quite analogous to the problems of particle detection.

they are observed to give almost always (or perhaps always, within some finite experimental error) no response. These experimental conditions are taken to define a state of the field in which no particles are present in the vicinity of the test system, and the systems are regarded as particle detectors. A true no-particle state for a region D would cause no response in a detector placed anywhere in D.

Now it is plausible that for D there may be more than one pair of the form (state + class of successful detector candidates) with these properties.[33] (Recall that any normalized positive linear functional on the algebra of observables is a state. The rival states mentioned here need not be vector states in the same cyclic representation.) Which of these pairs will the community of scientists define to be (vacuum + detectors)? Surely the one such that the first member (the state) is most likely to be encountered in actual experimental practice. Our experience is that there is precisely one particle concept of this type which is arrived at by natural and straightforward experimental operations of state preparation and particle detection.

The vacuum defined in this way seems to be a kind of equilibrium state[34], of which the other observed states are

---

[33] It is known from the classic work of Bohr and Rosenfeld (1933) that the vacuum is not just sheer emptiness as far as the field is concerned, but is full of fluctuations. Could there not be more than one such state of fluctuation with the qualitative experimental properties of a vacuum?

[34] This is consistent with the fact that when the laws of

excitations. (The analogy of phonons in a crystal is helpful here.) In fact, in a field theory of the usual type the vacuum is the ground state, and hence precisely the zero-temperature equilibrium state in the sense of statistical mechanics. The experimenter's production of a "good vacuum" in a region by pumping out almost all the particles (localizable field excitations) is basically the same as the process of bringing a system such as a crystal close to absolute zero temperature by extracting energy.

We can now conjecture the following picture: There is an equilibrium pure state of minimal excitation of the field, which corresponds to the Fock vacuum for the universe under consideration. There is a class $\tilde{S}$ of physical systems which are (practically) inert when placed in the vacuum. There are other observed states for which the systems in $\tilde{S}$ are not inert. Given a state (defined by the procedure for its experimental preparation), study of the correlations among the responses of these systems when placed at various positions reveals (at least for some subset $S \subset \tilde{S}$) a pattern of "events" which can be interpreted as triggerings by particles (cf. Steinmann (1968)). The systems in S are then called particle detectors. The particle structure of the states, it is conjectured, turns out to correspond to that defined by the creation operators of the Fock representation. (One hopes that how this happens will eventually

---

motion of the field are invariant under time translation the vacuum state is also expected to be invariant.

be worked out rather explicitly.)

This picture makes the existence of many different mathematically and physically acceptable vacuum states, associated with different global structures of the universe, more understandable. In the infinite universe free particles move indefinitely to new frontiers; in a torus universe they keep recirculating through the finite space. It is not too surprising that the equilibrium state of the field in the region D should be different in the two cases. If so, our earlier conclusions about the self-contained nature of the field dynamics in a Cauchy-complete region must be qualified: It is true that the time development of expectation values is completely determined by initial conditions, but <u>which initial conditions actually occur</u> depends on the global setting of the region D. The approach to equilibrium is something which happens over the entire previous history of the universe, so there is plenty of time for information about the global geometry to reach D without exceeding the velocity of light.

It remains to be explained how these different equilibrium situations yield qualitatively the same physics. Consider one of the pure equilibrium states of the algebra of D in terms of the quanta of another vacuum state, which we call the "original" one. It has a structure something like that indicated in Eqs. (F.3.7,11). It is a "soup" of virtual particles or field fluctuations, a medium so uniform that nothing in particular ever happens in it. Only excitations superimposed on this field substratum can be detected. There is a class of detectors which are inert with respect to this new vacuum. (They might be the same in their physical construction as the original detectors, except that they have come to equilibrium with the soup by attracting a cloud of virtual particles.) The equilibrium can be disturbed by adding a particle (or several). These excitations travel through the medium in qualitatively the same way as free particles travel through the original vacuum (cf. Sec. IX.2). They are detected as particles by the new or "dressed" detectors. If these conjectures are true, then the same basic physical laws (laws of motion of the field, laws of interaction with the detectors) lead to qualitatively the same phenomenological physics in the infinite and the finite universe, but with differences in detail which in principle could be observed. (Again, justification of these conjectures by means of explicit models of the measurement interaction would be highly desirable.)

What are the implications of all these considerations for the quantization of fields in the Rindler wedge and in the analogous patches of de Sitter space which can be treated as static universes? They tell us that we should not have expected the Fock vacuum in such a theory to be the physical, observed vacuum state appropriate to the embedding of the region in a larger, geodesically complete, universe. Such an embedding could be done in different ways, and the physically appropriate vacuum is probably different in each case. Our Fock vacuum might not correspond to <u>any</u> physically reasonable situation.[35] The earlier argument (Sec.IX.4) that all representations of the field algebra are physically equivalent remains valid, if one accepts the Fell-Haag-Kastler argument for general quantum systems. However, the evidence presented in this section strongly suggests that (and offers the beginning of an explanation why), in the case of Rindler space, the particle structure of the ΦOK - representation has an especially direct connection to the observable phenomena in our world, more so than the quanta of the Rindler representation. An approach to field quantization in general space-times which incorporates this idea will be suggested in Sec. X.8, after we have discussed the problems of canonical quantization in time-dependent metrics.

---

[35] It is tempting to say that it corresponds to the region as an isolated universe, as suggested at the beginning of this section. But this notion presents difficulties for a geodesically incomplete space. There are classical particle paths which leave and, even more disturbing, enter the space at finite proper times; to and from where? The system seems underdetermined without some boundary condition. In the quantum theory this means that we do not understand, physically, how the equilibrium state of the field is determined in such a situation.

Chapter X

FIELD QUANTIZATION IN AN EXPANDING UNIVERSE

Next in complexity after the static universes comes the class of Gaussian metrics which are defined in Appendix D as "generalized Robertson-Walker metrics". In these cases (see Eq. (D.6)) the time dependence and the space dependence of the metric are "separable", so that the field equation can still be solved by reduction to uncoupled one-dimensional modes. Now, however, the time dependence of the solution is not a simple complex exponential, but a solution of a more general second-order linear ordinary differential equation (Eq. (1.14)). Consequently, there are two important differences between this and the static case. First, there is a possibility of particle creation. Since the background is not static, there is no reason to expect to have stationary-state solutions which can be interpreted as n-particle states in the manner of Chapter VIII. There are both physical and mathematical reasons to believe that if particle number observables can be defined at each time, then they will not be constants of the motion. A great obstacle to making this idea precise and quantitative is the second new feature, the uncertainty in how to define particle observables at a fixed time. For the solutions of the general equation (1.14) there is no obvious analogue of the division into positive- and negative-frequency functions, which led so naturally to the introduction of the particle concept in the static case.

It seems unlikely that the study of the scalar field in more general space-times (which will involve, in general, coupled equations, or worse, for the time dependence) will lead to any fundamentally new physical phenomena or conceptual difficulties. (Investigations of particle creation in anisotropic universes, based on tentative assumptions about the interpretation of the field operators in terms of particles, are already in progress: Zel'dovich (1970); Zel'dovich and Starobinsky (1971); B.-L. Hu, dissertation, Princeton University, in preparation.)

The time-dependent case will not be treated here as systematically as the static case was above because of lack of space and because the subject has been treated thoroughly by Parker (1966, 1968, 1969, 1971, 1972). The purposes of this chapter are to set up a framework, slightly different from Parker's, for studying the solutions of the field equation, to summarize critically the prevailing points of view on the quantization problem, to present a few new technical results (see particularly Secs. X.5 and X.9-10), and to discuss the implications of the observations of this and the preceding chapters for future work toward the acceptable physical interpretation and mathematical definition of quantum field theories in curved space-time.

1. Solution of the Wave Equation; the Ambiguity of Quantization.

We study a metric of the form (D.6):

$$ds^2 = (dx^0)^2 - R^2(x^0)\, h_{jk}(x)\, dx^j dx^k , \tag{1.1}$$

so that

$$\sqrt{|g|} = R^s \sqrt{h}, \qquad g^{00} = 1 \qquad (h \equiv \det \{h_{jk}\}). \tag{1.2}$$

(As always, $s + 1$ is the dimension of space-time.) For convenience this is called an "expanding universe", without any implication that R must be an increasing function. Specializing from the formalism of Sec. VII.1, we have

$$\mathcal{L} = \frac{1}{2} R^s \sqrt{h}\, [(\partial_0 \phi)^2 - R^{-2} h^{jk} \partial_j \phi\, \partial_k \phi - m^2 \phi^2], \tag{1.3}$$

$$\pi = R^s \sqrt{h}\, \partial_0 \phi, \tag{1.4}$$

$$\partial_0^2 \phi + s\, R'/R\, \partial_0 \phi - R^{-2} \Delta_c \phi + m^2 \phi = 0, \tag{1.5}$$

where

$$\Delta_c \phi = \frac{1}{\sqrt{h}} \partial_j [\sqrt{h}\, h^{jk}\, \partial_j \phi], \tag{1.6}$$

$R' = dR/dx^0$, and $\{h^{jk}\}$ is the inverse of $\{h_{jk}\}$.

The occurrences of the time-dependent quantity R in the formalism can be minimized by introducing a new time variable

$$t = \int^{x^0} R^{-s}\, dx^0 , \tag{1.7}$$

so that

$$ds^2 = R^{2s}\, dt^2 - R^2 h_{jk}\, dx^j dx^k . \tag{1.8}$$

This is no longer a Gaussian metric, but we have now a time-independent three-space volume element

$$\sqrt{|g|}\, g^{tt} = R^{2s} \sqrt{h}\, R^{-2s} = \sqrt{h} \tag{1.9}$$

instead of Eqs. (1.2). Then Eqs. (1.4-5) are replaced by

$$\pi = \sqrt{h}\, \partial_t \phi \tag{1.10}$$

and

$$\partial_t^2 \phi - R^{2s-2} \Delta_c \phi + R^{2s} m^2 \phi = 0. \tag{1.11}$$

Note that the first-order term has been eliminated.

We shall assume, as in Chapter VIII, that Δ_c has a complete set of generalized eigenfunctions $\phi_j(x)$:

$$\Delta_c \phi_j(x) = - e_j^2 \phi_j(x) \qquad (e_j > 0). \qquad (1.12)$$

The substitution

$$\phi(t,x) = \phi_j(x) \psi_j(t) \qquad (1.13)$$

yields

$$\partial_t^2 \psi_j(t) + R^{2s} [e_j^2/R^2 + m^2] \psi_j(t) = 0. \qquad (1.14a)$$

(In terms of x^0, of course, we have the equation

$$\partial_0^2 \psi_j + s R'/R \, \partial_0 \psi_j + [e_j^2/R^2 + m^2] \psi_j = 0.) \qquad (1.14b)$$

From now on let us for simplicity consider primarily the case

$$h_{jk} = \delta_{jk}, \qquad \Delta_c = \nabla^2, \qquad j = e_j = \vec{k}, \qquad \phi_{\vec{k}}(x) = \varphi^{i\vec{k}\cdot x}. \qquad (1.15)$$

The general case can be treated in exactly the same way; the specialization is made primarily to make the formulas readable with less effort. The range of the spatial variables in the special case can be either infinite Euclidean space or a finite box; let us write "$\int dk$" and interpret it as a sum or as an integral appropriately in each case. (Likewise, the implicit normalization factor in φ depends on the volume of the box.) Also, the vector symbol over k will henceforth be omitted. Then the general real or Hermitian solution of the field equation is

$$\phi(t,x) = \int dk\, [e^{ik\cdot x}\, \psi_k(t) a_k + e^{-ik\cdot x}\, \psi^*_k(t) a^\dagger_k]. \tag{1.16}$$

Our attention must now center on the equation (1.14a). Eq. (1.16) does not define a_k and $a^\dagger_k$ until a particular solution $\psi_k(t)$ is chosen from the two-dimensional complex vector space of possibilities. The choice is not entirely arbitrary if one wants the creation and annihilation operators to satisfy the canonical commutation relations. From Eqs. (1.10) and (1.16) we have

$$\pi(t,x) = \int dk\, [e^{ik\cdot x}\, \dot{\psi}_k(t) a_k + e^{-ik\cdot x}\, \dot{\psi}^*_k(t) a^\dagger_k], \tag{1.17}$$

where the dot now indicates differentiation with respect to t. It is reasonable to assume (or demand, according to one's point of view) that $\psi_k(t)$ depends only on k^2 (more generally, that ψ_j depends only on e_j).[1] Then we have

$$\int d^s x\, e^{-ik\cdot x}\, \phi(t,x) = \psi_k(t) a_k + \psi^*_k(t) a^\dagger_{-k},$$

$$\int d^s x\, e^{-ik\cdot x}\, \pi(t,x) = \dot{\psi}_k(t) a_k + \dot{\psi}^*_k(t) a^\dagger_{-k},$$

[1] Alternatively, in order to avoid making this assumption for the purpose of the calculation below, one could work, as in the first part of Sec. VIII.2, with real basis functions ψ_j (here, $\sin k\cdot x$ and $\cos k\cdot x$).

and hence

$$a_k = w_k^{-1} [\dot{\psi}^*_k(t) \int dx\, e^{-ik\cdot x} \phi - \psi^*_k(t) \int dx\, e^{-ik\cdot x} \pi], \quad (1.18)$$

where w_k is the Wronskian

$$w_k = \psi_k(t) \dot{\psi}^*_k(t) - \dot{\psi}_k(t) \psi^*_k(t). \quad (1.19)$$

The form of Eq. (1.14a) implies that w_k is independent[2] of t; it is obviously imaginary. When $a_k^\dagger$ is found from Eq. (1.18) (or solved for directly), it is easily seen that

$$[a_k, a_{k'}^\dagger] = \frac{i}{w_k} \delta(k - k').$$

So ψ_k must be chosen so that

$$w_k = i. \quad (1.20)$$

We can write

$$\psi_k(t) = A_k E_k(t) + B_k O_k(t), \quad (1.21)$$

where E_k and O_k are standard solutions of Eq. (1.14a) satisfying

[2] With the original time coordinate x^0 one would have $w_k \propto R^{-s}$, and, at the last step, $w_k = iR^{-s}$.

$$E_k(0) = 1, \qquad \dot{E}_k(0) = 0, \tag{1.22a}$$

$$O_k(0) = 0, \qquad \dot{O}_k(0) = 1. \tag{1.22b}$$

(This notational convention is slightly different from that of Sec. V.3, where an imaginary time coordinate was used for technical convenience.) Note that

$$A_k = \Psi_k(0), \qquad B_k = \dot{\Psi}_k(0). \tag{1.23}$$

The Wronskian of E_k and O_k is equal to 1. Therefore, the condition (1.20) is equivalent to

$$2 \operatorname{Im} A_k B_k^* = 1. \tag{1.24}$$

If $\phi_k(t,x)$ and $\phi_j(t,x)$ are two solutions of the elementary type (1.13), the current form (VII.5.3) is

$$W(\phi_k, \phi_j) = i\int d^s x \, [\phi_k^* \dot{\phi}_j - \dot{\phi}_k^* \phi_j]$$

$$= i(2\pi)^{-s} \int dx \, [e^{i(j-k)\cdot x} \Psi_k^*(t) \dot{\Psi}_j(t) - e^{-i(j-k)\cdot x} \dot{\Psi}_k^*(t) \Psi_j(t)]$$

$$= - i w_k \, \delta(k,j) = \delta(k,j). \tag{1.25}$$

Thus the condition (1.20) ensures that these functions span a space of solutions of positive norm with respect to W, and form an orthonormal basis in that space (cf. Secs. V.4-5).

2. A Flat Expanding Space-Time.

A tentative approach to the quantization of the field equation solved in the last section is the following: The Hilbert space is the Fock space of the set of operators $\{a_k, a_k^\dagger\}$ given by Eq. (1.18) with some choice of the functions Ψ_k (Eq. (1.21)). (Cf. Secs. V.3-5 and VIII.3-4.) The Ψ_k should be chosen (how?) so that the a_k and $a_k^\dagger$ are the annihilation and creation operators for physical particles present (in mode k) in the state of the universe at some time t_0.

The purpose of this section is to show that the adoption of this point of view has serious implications. We consider a simple model complementary to the one studied in Chapter IX.[3] In two-dimensional Minkowski space with coordinates (x^0, x) let

$$x^0 = e^t \cosh y, \qquad x = e^t \sinh y. \tag{2.1}$$

The coordinates (t,y) $(-\infty < t,\ y < \infty)$ cover the region where $x^0 > |x|$. The metric is of the type (1.8):

[3] This space has two four-dimensional analogues. One, which is spherically symmetric, is one of the Robertson-Walker universes of constant curvature (see Appendix D). The other is defined by a transformation of the form (2.1) on one space coordinate, the other two being unchanged; this is the degenerate Kasner universe mentioned in passing by Zel'dovich (1970) and Zel'dovich and Starobinsky (1971). Both of these models are merely patches of Minkowski space in disguise.

$$ds^2 = e^{2t}(dt^2 - dy^2) \qquad (R(t) = e^t). \quad (2.2)$$

The equation (1.14a) for the time dependence of a mode of the scalar field becomes

$$\partial_t^2 \psi_k(t) + e^{2t}[k^2 e^{-2t} + m^2]\psi_k(t) = 0, \qquad (2.3a)$$

or, in terms of the Gaussian time coordinate denoted in the previous section by x^0, which in this case is $e^t = R$,

$$\partial_R^2 \psi_k(R) + \frac{1}{R}\partial_R \psi_k(R) + [k^2/R^2 + m^2]\psi_k(R) = 0 \qquad (2.3b)$$

(cf. Eqs. (1.7) and (1.14b)). The solutions of Eq. (2.3b) are linear combinations of the Bessel functions $J_{ik}(R)$ and $N_{ik}(R)$, but we shall not use this information in what follows. Instead, let us continue to denote by $\psi_k(t)$ a generic solution of Eq. (2.3a) satisfying Eqs. (1.21-24).

We set t = 0 in the formula (1.18) for a_k:

$$a_k = -i[\dot{\psi}_k^*(0)\int dy\, e^{-iky}\phi(0,y) - \psi_k^*(0)\int dy\, e^{-iky}\pi(0,y)],$$

and substitute for ϕ and π the expressions in terms of Fourier components in Cartesian coordinates:

$$\phi(0,y) = \phi(x^0=\cosh y,\ x=\sinh y) =$$

$$\int \frac{dp}{\sqrt{2\omega_p}} [\varphi^{ip \sinh y} e^{-i\omega_p \cosh y} b_p + \varphi^{-ip \sinh y} e^{i\omega_p \cosh y} b_p^\dagger],$$

$$\pi(0,y) = -i \int dp \sqrt{\frac{\omega_p}{2}} [\varphi^{ip \sinh y} e^{-i\omega_p \cosh y} b_p$$

$$- \varphi^{-ip \sinh y} e^{i\omega_p \cosh y} b_p^\dagger],$$

where $\omega_p = \sqrt{p^2 + m^2}$. The result is

$$a_k = \int dp\ J(k,p)\ [\psi^*_k(0)\sqrt{\omega_p} - i\dot{\psi}^*_k(0)\frac{1}{\sqrt{\omega_p}}]\ b_p$$

$$- \int dp\ J^*(-k,p)\ [\psi^*_k(0)\sqrt{\omega_p} + i\dot{\psi}^*_k(0)\frac{1}{\sqrt{\omega_p}}]\ b_p^\dagger, \qquad (2.4)$$

where

$$J(k,p) = \frac{1}{\sqrt{2}} \int dy\ \varphi^{-iky}\ \varphi^{ip \sinh y}\ e^{-i\omega_p \cosh y}$$

$$= \frac{1}{\sqrt{2}}\ \frac{1}{2\pi} e^{-ik\theta_p} \int_{-\infty}^{\infty} dy\ e^{-iky}\ e^{-im \cosh y} \qquad (p = m \sinh \theta_p),$$

or

$$J(k,p) = -\,i\,2^{-3/2}\,e^{-ik\theta_{p}}\,e^{\pi k/2}\,H^{(2)}_{ik}(m) \qquad (2.5)$$

([Gradshteyn-Ryzhik], Eq. (8.421.2) (p. 955)). No choice of $\Psi_k(0)$ and $\dot{\Psi}_k(0)$ (i.e., of A_k and B_k in Eqs. (1.21,24)) can make the kernel in the last term of Eq. (2.4) vanish identically in p (as a distribution), so that the vacuum of the free field be annihilated by the a_k.

The conclusion is that for the free field in "expanding Minkowski space" <u>none</u> of the tensor product representations proposed at the beginning of this section coincides with the standard representation of the free field.[4] In previous work on field quantization in expanding universes, to the best of the author's knowledge, it has usually been assumed that some splitting of the Fourier components of the field into annihilation and creation operators for physical particles is possible, the problem being to determine which splitting is correct.[5] The argument above drives one to one of two conclusions:

---

[4] What is at issue here is not unitary equivalence but strict identity of representations (better, identity of vacuum states), leading to identical particle interpretations. Of course, only a subalgebra of the full algebra of field operators in Minkowski space would be involved in this identity.

[5] In his 1969 paper (p. 1064) Parker has emphasized, however, that if particles are being produced as a result of metric expansion, then the concept of particle number at a given time is operationally fuzzy because of the uncertainty principle. In Parker (1966) particle number is defined only for a slowly expanding universe.

(1) Quantization based on such a splitting is physically wrong -- at least in some cases.

or

(2) There is no unique physically correct representation of the fields (cf. Sec. IX.4).

In either case the interpretation of the operators a_k of Eq. (1.18) in terms of physical particles in weakened.

3. Asymptotically Static Metrics.

For contrast to the negative result of the last section we turn to a situation where the field has a clear interpretation in terms of particles -- the case

$$R(t) = R_+ \quad \text{for } t > t_0, \qquad R(t) = R_- \quad \text{for } t < -t_0. \qquad (3.1)$$

(With proper attention to technicalities one could make the same statements about a metric for which R(t) merely approaches constant values "sufficiently fast" in the past and future.) We shall call this behavior asymptotically static. (Parker's term is "statically bounded".) Of course, in the special case (1.15) considered here the space-time is actually flat in the asymptotic regions.

In the region of space-time where $t > t_0$ the equation of motion is that of a free field. The general solution can be written

$$\phi(t,x) = \int dk\, [2\omega_k R_+^s]^{-1/2} [\varphi^{ik\cdot x} \exp\{-i\omega_k R_+^s t\}\, a_k^{out}$$

$$+ \varphi^{-ik\cdot x} \exp\{i\omega_k R_+^s t\}\, a_k^{out\dagger}] \qquad (t > t_0), \quad (3.2)$$

with

$$\omega_k^2 = k^2/R_+^2 + m^2. \quad (3.3)$$

It is hard to believe that in the quantum theory the coefficients a_k^{out} and $a_k^{out\dagger}$ should not be interpreted as annihilation and creation operators for physical, observable particles. In this region of space-time we have simply a free field, whose physical interpretation is well understood.

Similar statements hold, of course, for the region where $t < -t_0$. We have (extending the analogue of Eq. (3.2) to general t)

$$\phi(t,x) = \int dk\, [\varphi^{ik\cdot x}\, \psi_k^{in}(t)\, a_k^{in} + \varphi^{-ik\cdot x}\, \psi_k^{in}{}^*(t)\, a_k^{in\dagger}], \quad (3.4)$$

where $\psi_k^{in}(t)$ $(= \psi_{-k}^{in}(t))$ satisfies Eq. (1.14a) and ($\omega_k^- = \sqrt{k^2/R_-^2 + m^2}$)

$$\psi_k^{in}(t) = [2\omega_k^- R_-^s]^{-1/2} \exp\{-i\omega_k^- R_-^s t\} \quad \text{for } t < -t_0, \quad (3.5)$$

$$\Psi_k^{in}(t) = [2\omega R_{k+}^s]^{-1/2} [\alpha^*(k) \exp\{-i\omega R_{k+}^s t\}$$

$$+ \beta^*(k) \exp\{i\omega R_{k+}^s t\}] \quad \text{for } t > t_0 ; \qquad (3.6)$$

$\alpha(k)$ and $\beta(k)$ are certain coefficients which depend on the whole function $R(t)$. The solution (3.5) is normalized according to Eqs. (1.19-20). Since the Wronskian is independent of t, we find from Eq. (3.6)

$$|\alpha(k)|^2 - |\beta(k)|^2 = 1. \qquad (3.7)$$

Comparing Eqs. (3.2) and (3.4,6), we find

$$a_k^{out} = \alpha^*(k) a_k^{in} + \beta(k) a_{-k}^{in\dagger}. \qquad (3.8)$$

This is a Bogolubov transformation (Secs. F.2,4). Eq. (3.7) is precisely the condition which assures that canonical commutation relations for both the in- and the out-operators are consistent. When $\beta(k) \neq 0$, the import of Eq. (3.8) is that pairs of particles in the modes k and - k are created and destroyed during the expansion of the universe. To be explicit, the operator of the number of particles in the mode k after t_0 is

$$N_k^{out} \equiv a_k^{out\dagger} a_k^{out} = |\alpha(k)|^2 N_k^{in} + |\beta(k)|^2 (N_{-k}^{in} + \delta(-k,-k))$$

$$+ \alpha(k)\beta(k) a_k^{in\dagger} a_{-k}^{in\dagger} + \alpha^*(k)\beta^*(k) a_{-k}^{in} a_k^{in} . \quad (3.9)$$

(See remarks below on the interpretation of $\delta(-k,-k)$.) Thus, for example, if there are no particles before the expansion, the expectation value of the number of particles in this mode afterwards is

$$\langle 0\ in|N_k^{out}|0\ in\rangle = |\beta(k)|^2 \quad (3.10a)$$

if space at fixed time is a finite box with periodic boundary conditions (torus). If one particle is present initially, we have

$$\langle k\ in|N_k^{out}|k\ in\rangle = \langle 0\ in|a_k N_k^{out} a_k^{\dagger}|0\ in\rangle$$

$$= |\alpha(k)|^2 + |\beta(k)|^2 , \quad (3.10b)$$

and so on. For more details see Parker (1969), Sec. C.

In the case of a metric which is time-independent in the limit of large positive and negative times but is not of the generalized Robertson-Walker form (1.1), the field equations in conjunction with the free field interpretation in the periods before and after the expansion will again predict particle creation, but the various modes will be coupled; that is, Eq. (3.8) will be replaced by a Bogolubov transformation of the more general type discussed in Secs. F.2-3.

According to Eq. (F.3.2),

$$\int dk\, \delta(k,k)\, |\beta(k)|^2 < \infty \tag{3.11}$$

is a necessary and sufficient condition for the existence of a unitary operator S such that for all k

$$a_k^{in} = S a_k^{out} S^{-1} . \tag{3.12}$$

If space is infinite (k is a continuous variable), then δ in Eq. (3.11) is a Dirac delta function, and the condition (3.11) <u>always</u> fails. (This is an infinite-volume divergence.) If space is finite (k is a discrete variable $2\pi n/L$), then δ is a Kronecker delta function, and Eq. (3.11) is a statement about the ultraviolet behavior of $\beta(k)$ (cf. Sec. F.4).

If S exists, the Fock space of the <u>out</u>-operators can be identified with the Fock space of the <u>in</u>-operators. The <u>in</u>-vacuum is

$$|0\ in\rangle = S|0\ out\rangle, \tag{3.13}$$

which is a linear combination of $|0\ out\rangle$, two-particle <u>out</u>-states of the form

$$a_k^{out\dagger} a_{-k}^{out\dagger} |0\ out\rangle,$$

four-particle out-states, etc., such that Eqs. (3.10) hold (see Secs. F.3-4 or Parker (1969), pp. 1061-1062)). This is a Heisenberg-picture approach; one looks at the late-time observables, such as N_k^{out}, with respect to the fixed state vector $|0\ in\rangle$ (or another in-state). One can also look at the system in the interaction picture, where S is interpreted as the limiting propagator $U(\infty,-\infty)$, which maps a state (of the free field) initially containing no particles into a final state of the form just described, and so forth.

If no S exists, it is no longer possible to regard the Fock space of the in-operators as the "arena" in which the time development of the system takes place (interaction picture). The Heisenberg picture, in a fixed representation, does still make sense; however, the possible states, as functionals on the algebra of field operators (over all space-time), will be different in the two asymptotic representations (cf. Sec. IX.4). But this "pathology" is entirely reasonable physically; it means that infinitely many particles are produced in the expansion, as is to be expected from an interaction extending uniformly throughout an infinite space.[6] In the author's opinion we have here strong additional evidence that, as suggested in Sec.

---

[6] In this case convergence of the integral (3.11) for the analogous box case is of interest, since it has the significance that the density of particles created is finite (cf. Parker (1969), pp. 1062-1063).

IX.4, the mathematically convenient framework of a single irreducible representation of the field algebra is too narrow for some applications of quantum field theory.

The important conclusion of this section is that, as Parker (1966, 1969) has strongly emphasized, a time-dependent space-time metric leads unambiguously to particle creation in quantum field theory. This creation is independent of any particle interpretation of the field theory during the period of expansion.[7]

This kind of particle creation must be clearly distinguished from two similar effects which have been discussed in the literature. One is the production of new kinds of particles in high-energy elementary particle processes, such as

$$\gamma + \gamma \longrightarrow e^{+} + e^{-} \quad (\text{or } p + \bar{p}, \text{ etc.}),$$

whose importance in early cosmology and in some astrophysical situations has long been recognized. This is simply the conversion of one type of particle into another via strong, electromagnetic, or weak interactions, and has nothing to do with the appearance of particles in what was initially the vacuum state of a free field theory. Another kind of particle creation has been predicted in de Sitter space by Nachtmann (1968a,b). He interprets the free field as a theory of stable particles, as in

[7] For another statement of the same argument in a different context see Moore (1970), Sec. VII.

Sec. V.6. Then he finds that the addition of a $\lambda\phi^4$ interaction leads in perturbation theory to creation of particles relative to this definition of particle, at least if some particles are already present. (That is, a one-particle state can evolve into a three-particle state, and so on.) Unlike Parker's effect, Nachtmann's depends on a nonlinear interaction term in the field equation.

4. Nachtmann's Ansatz.

Nachtmann (1968b) proposes to define positive-frequency solutions of the wave equation in two-dimensional de Sitter space in the following way. The Hilbert space of the quasiregular representation of the de Sitter group (see Secs. V.4 and VI.1) can be decomposed as a direct integral, the spectral representation of the Casimir operator Q:

$$\mathcal{H} = \int^{\oplus} d\mu(q) \ \mathcal{H}_q . \tag{4.1}$$

$\mathcal{H}_q$ can be identified with the space of generalized eigenvectors of Q with eigenvalue q -- i.e., the sufficiently integrable solutions of the wave equation (V.2.1). The scalar product in $\mathcal{H}_q$ is defined up to a factor. The current form (V.4.1) is a bounded Hermitian form on $\mathcal{H}_q$. Therefore, it determines an Hermitian operator N by the formula

$$W(\Psi_1, \Psi_2) = (\Psi_1, N\Psi_2). \tag{4.2}$$

We pass to the spectral representation of N: $\mathcal{H}_q$ is a direct sum

$$\mathcal{H}_q = \mathcal{H}_+ \oplus \mathcal{H}_-, \tag{4.3}$$

where

$$N\Psi = \pm \Psi \quad \text{for } \Psi \in \mathcal{H}_\pm$$

for some choice of the arbitrary constant in the scalar product. Eq. (4.3) provides a <u>distinguished</u> decomposition of the type (V.4.10). $\mathcal{H}_+$ is taken to be the space of positive-frequency solutions. It turns out to coincide with the space chosen by Tagirov <u>et al</u>. (Sec. V.6).

Nachtmann remarks that this definition can be generalized to an arbitrary Riemannian manifold. What is intended is presumably the following. The spectral decomposition of the Laplace-Beltrami operator (VII.1.3) equips the solutions of the wave equation (VII.1.2) with a positive definite scalar product. This added structure suffices, as above, to remove the ambiguity in the classification of the solutions as positive or negative under the current form (VII.5.3): the positive functions are those in the positive piece of the spectral resolution (4.2-3) of the form. (The positive solutions can be interpreted as particle wave functions and used to build a Fock space as in Sec. V.6 and Secs. VIII.1-4. Clearly, there is no particle creation in such a theory.)

The physical relevance of this construction may be challenged for the reason stated at the end of Sec. V.6. To sharpen this argument we shall now investigate what Nachtmann's prescription yields in the case of an asymptotically static universe, where we already have a convincing physical interpretation of the solutions. Unfortunately, for lack of space the exposition will be rather sketchy.

Consider a two-dimensional asymptotically static metric (Eqs. (1.1), (3.1)) with $R_+ = R_- = 1$. The Laplace-Beltrami operator $\Box_c$ is Hermitian in the Hilbert space of functions on the whole manifold square-integrable with respect to the volume element $\sqrt{|g|}\,dt\,dx = R^2 dt\,dx$. The corresponding differential equation, which is the wave equation (1.11), is

$$R^{-2}[\partial_t^2 - \partial_x^2]\phi + m^2\phi = 0, \tag{4.4}$$

where m^2 is the eigenvalue. There are eigensolutions of the form

$$\phi_{m,k,\sigma}(t,x) = e^{ikx}\,\psi_{m,k,\sigma}(t) \qquad (\sigma = \pm), \tag{4.5}$$

where for given m and $|k|$ there are two independent solutions $\psi_{m,k,\pm}$ of

$$R^{-2}[\ddot{\Psi} + k^2 \Psi] + m^2 \Psi = 0, \tag{4.6}$$

one of which can be chosen to have the asymptotic form (3.4-5). The problem is to normalize $\Psi_{m,k,\sigma}$ so that

$$\int R^2 \, dt \, dx \, \phi^*_{m,k,\sigma} \, \phi_{m',k',\sigma'} = \delta_{\sigma\sigma'} \, \delta(k - k') \, \delta(m - m'); \tag{4.7}$$

then upon discarding the last delta function on the right-hand side of this formula we will obtain a scalar product on the space of solutions of the wave equation for a fixed m (by taking the ϕ's as an orthonormal[8] basis).

Eq. (4.6) is identical to the Schrödinger equation for a one-dimensional scattering problem ($t \equiv x$; $V(t) = - R^2 m^2$). In the present problem, however, we must regard k^2 as a fixed parameter and m^2 as a varying eigenvalue (the reverse of the situation for scattering), and the relevant scalar product involves $\int R^2 dt$ instead of Lebesgue integration, $\int dt$. Nevertheless, it can be shown that the orthonormalization of the eigenfunctions is the same as in the familiar problem. Namely, we can choose

$$\Psi_{m,k,+} = \Psi^{in}_{\omega}, \qquad \Psi_{m,k,-} = \Psi^{in}_{-\omega} \qquad (\omega = \sqrt{k^2 + m^2}), \tag{4.8}$$

the functions with the structure indicated in Fig. 11 (Sec. V.7)

---

[8] "Orthonormal" is meant in the strict or the generalized sense, depending on the nature of the spectrum of k.

or Eqs. (V.7.6) with obvious notational substitutions. An alternate basis is provided by their complex conjugates $\Psi^{out}_{-\omega}$ and Ψ^{out}_{ω}, whose structure is schematically indicated by the first and second line of Fig. 11, respectively, with the arrows reversed. The notation here is that appropriate to the scattering analogy, not to the actual physical situation; the function called Ψ^{in}_{k} in Eqs. (3.5-6) is proportional to $\Psi^{in}_{-\omega}$ under the present convention. It is crucial to the following argument (in particular, to the equality in Eq.(4.10b)) that the absolute value of the constant of proportionality (equivalently, of the transmission coefficients S in Eqs. (V.7.6)) is independent of the sign of $\pm\omega$ (see [Messiah], p. 107).

For fixed values of both k and m we have a two-dimensional complex Hilbert space of solutions (4.5) with the scalar product induced by Eq. (4.7). That is,

$$(\phi_{m,k,\sigma}, \phi_{m,k,\sigma'}) = \delta_{\sigma\sigma'} \tag{4.9}$$

for either of the two bases just defined, and the scalar product of linear combinations of these is then determined. The current form or Wronskian[9], however, is not diagonal in either of these bases; rather, we have

$$W(\Psi^{in}_{-\omega}, \Psi^{out}_{\omega}) = 0, \tag{4.10a}$$

[9] In the scattering analogy, $-W(\Psi,\Psi)/2$ is the flux in the beam.

$$W(\Psi_{-\omega}^{in}, \Psi_{-\omega}^{in}) = - W(\Psi_{\omega}^{out}, \Psi_{\omega}^{out}) > 0, \qquad (4.10b)$$

and similarly for $\Psi_{-\omega}^{out}$ and Ψ_{ω}^{in}. Our task is to find a new basis $\{\Psi_{-\omega}, \Psi_{\omega}\}$ which is orthogonal with respect to both the scalar product and W.

It can be shown that if ϕ, θ, and χ are defined by[10]

$$S_k = e^{i\theta} \cos \phi, \quad R_k = e^{i\chi} \sin \phi \qquad (0 < \phi < \frac{\pi}{2}), \quad (4.11)$$

so that

$$\Psi_{\omega}^{out} = e^{-i\theta} \cos \phi \, \Psi_{\omega}^{in} - e^{-2i\theta} e^{i\chi} \sin \phi \, \Psi_{-\omega}^{in}, \qquad (4.12)$$

then a solution of the problem is

$$\Psi_{\omega} = \cos \frac{\phi}{2} \Psi_{\omega}^{in} - e^{i(\chi - \theta)} \sin \frac{\phi}{2} \Psi_{-\omega}^{in}, \qquad (4.13a)$$

$$\Psi_{-\omega} = \Psi_{\omega}^{*}; \qquad (4.13b)$$

$$(\Psi_{\omega}, \Psi_{\omega}) = (\Psi_{-\omega}, \Psi_{-\omega}) = 1, \quad (\Psi_{\omega}, \Psi_{-\omega}) = 0, \qquad (4.14a)$$

$$W(\Psi_{-\omega}, \Psi_{-\omega}) = - W(\Psi_{\omega}, \Psi_{\omega}) > 0, \quad W(\Psi_{\omega}, \Psi_{-\omega}) = 0. \qquad (4.14b)$$

Thus the new basis is obtained by "rotating half way" from the

---

[10] See Eq. (V.7.6a) for the definition of S_k and R_k.

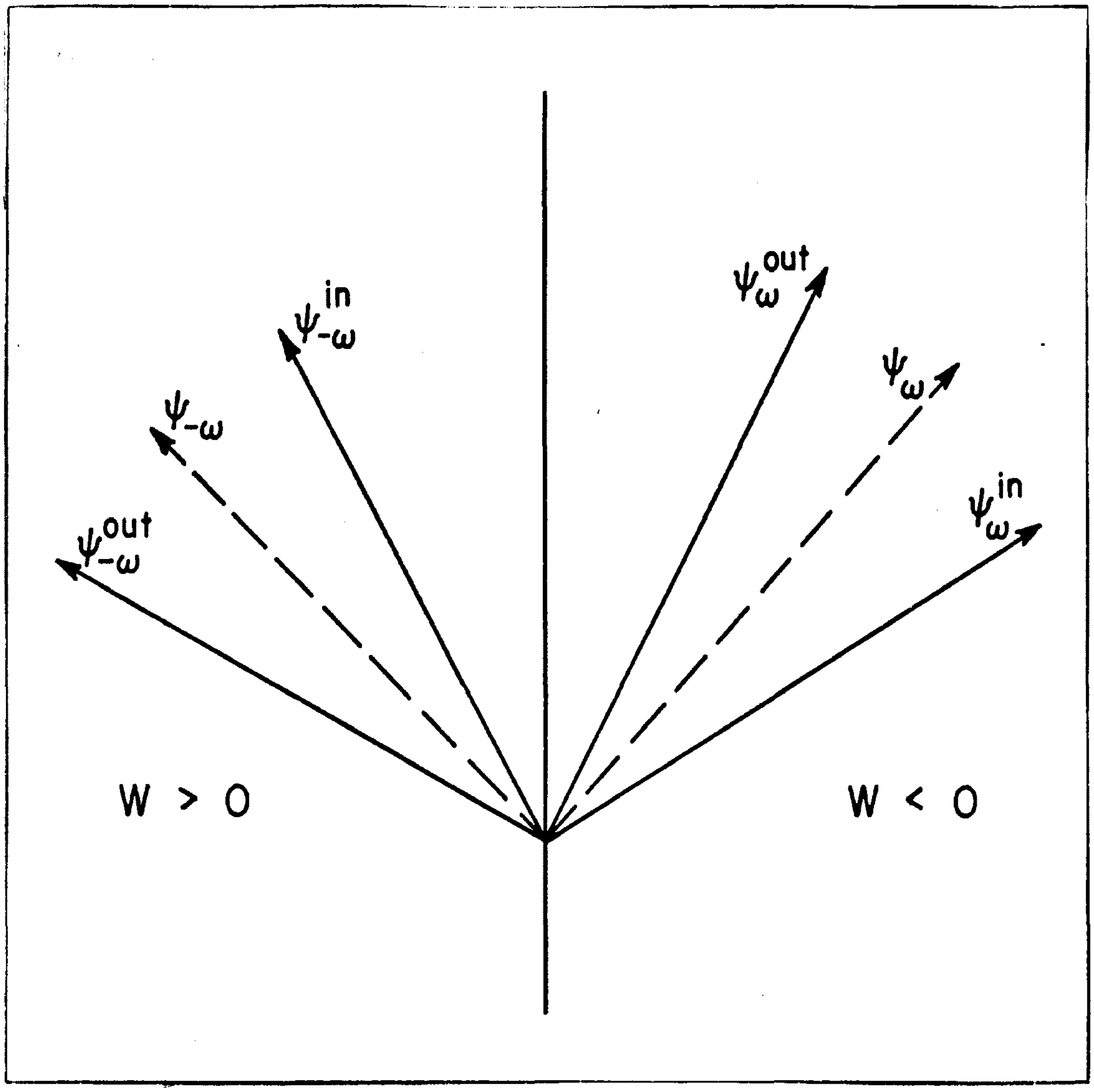


Fig. 17

Relation of Nachtmann's basis to the asymptotic bases. Perpendicular vectors in the figure are orthogonal in the positive definite scalar product (,). Vectors situated symmetrically with respect to the central vertical axis are mutually complex conjugate and are orthogonal with respect to the current form W(,).

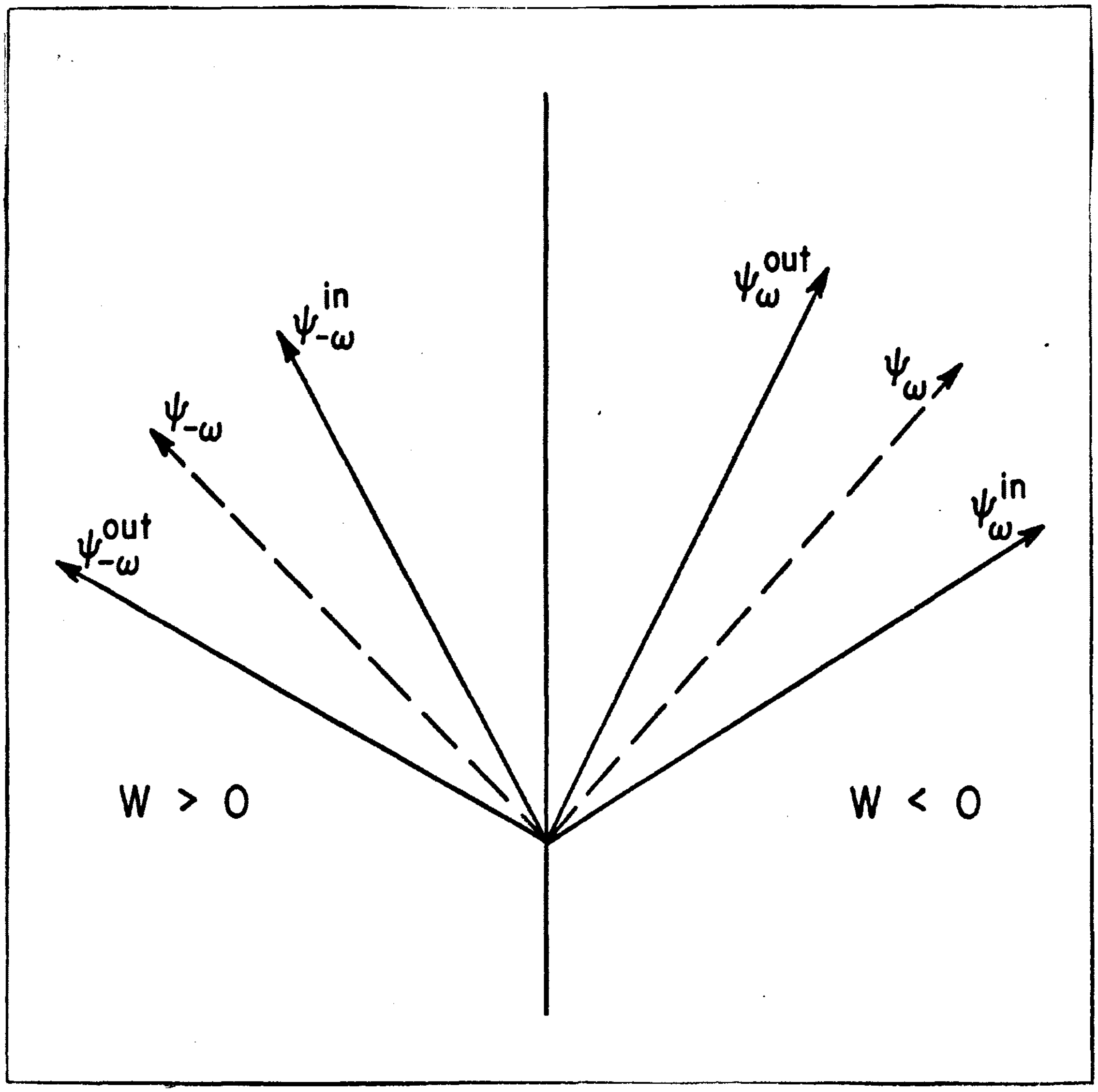


Fig. 17

Relation of Nachtmann's basis to the asymptotic bases. Perpendicular vectors in the figure are orthogonal in the positive definite scalar product (,). Vectors situated symmetrically with respect to the central vertical axis are mutually complex conjugate and are orthogonal with respect to the current form W(,).

in-basis to the out-basis, as indicated in Fig. 17. The asymptotic behavior of $\Psi_\omega(t)$ is

$$\cos\frac{\phi}{2}\, e^{i\omega t} + e^{i\chi}\sin\frac{\phi}{2}\, e^{-i\omega t} \tag{4.15a}$$

in the past $(t < -t_0)$, and

$$e^{i\theta}\cos\frac{\phi}{2}\, e^{i\omega t} - e^{i\chi}e^{-i\theta}\sin\frac{\phi}{2}\, e^{-i\omega t} \tag{4.15b}$$

in the future.

What is the significance of these functions? To interpret the functions

$$e^{ikx}\,\Psi_{-\omega}(t)$$

as wave functions of a stable particle seems much less convincing than the interpretation of the field expounded in Sec. X.3. As remarked earlier, in the static portion of space-time one has a free-field situation, and the usual interpretation of the free field surely applies. Consequently, the validity of this approach to other (not asymptotically static) manifolds, such as de Sitter space, is questionable. The same objection obviously applies to any attempt to define a unique notion of positive frequency for solutions of the wave equation in a nonstatic space-time (for instance, the definition via a distinguished even solution $G^{(1)}$ (cf. Eq. (VII.4.17)) by Lichnerowicz (1962)).

On the other hand, Nachtmann's prescription gives an interesting classification of the solutions which is intrinsic (independent of time, independent of any coordinate system). If, as the results of this and the last chapter seem to indicate, the notion of particle is a rather weak and ambiguous one in curved space-time, then this ansatz may give as good a definition as any of particle states for the free scalar field in an arbitrary Riemannian universe. That is, it may be a mathematically convenient way of classifying the states in a certain representation of the field algebra. (For instance, in a theory with interactions it may be useful to do perturbation theory starting from the Nachtmann states of the noninteracting theory as unperturbed states.) The fact that for de Sitter space the method leads to what was independently shown to be the only group-covariant definition of particle which seems to be at all physically sensible (Sec. V.6) is a point in its favor.

5. A Definition of Positive Frequency at a Given Time, and Some Theorems on Equivalence of Representations.

Up to now we have succeeded in splitting the field operator into annihilation and creation operators for particles only in cases where the metric is constant, at least in some finite interval of time. The criterion has been that the time dependence of the annihilation operators for a mode characterized by energy E should have the time dependence exp(-iEt). Is it possible to place some related condition on the behavior of the functions $\Psi_k(t)$ (of Eqs. (1.16,21)) near a fixed value of t, so that the a_k in Eq. (1.16) correspond to annihilation operators for particles at that time? In a sense the result of Sec. X.2 answers this question in the negative, unless the example studied there can be argued away as illegitimate (because the space is extendable, for instance -- cf. Sec. IX.6). However, in accordance with the author's often-repeated opinion that a particle notion may be of some value even if it is not unique, definitions of this type will now be considered.

Let us take t = 0 to be the time at which particles are to be defined. Then the problem is to make a sensible choice of the coefficients A_k and B_k in Eqs. (1.21-24), one which reduces when R = <u>const</u>. in the neighborhood of 0 to

$$A_k = [2E_k R^s]^{-1/2}, \qquad B_k = -\frac{i}{2} A_k^{-1} \tag{5.1}$$

(cf., e.g., Eq. (3.5)). The annihilation operator a_k will then be determined by Eq. (1.18).

The most obvious approach is to diagonalize the instantaneous Hamiltonian H(0).[11] That is, we are to choose

---

[11] See, for instance, Grib and Mamaev (1969). (This paper seems to the present author to contain several confusions. In particular, Eq. (12) as written is not a solution of the field equation; it becomes one (and the theory becomes equivalent to that of this section) if the argument η of $\lambda(\eta,k)$ is replaced by a separate parameter η_0. Also, the assertion (p. 724) that the particle creation violates charge conservation is incomprehensible to the present author; it seems to be based on a confusion of particle and antiparticle creation operators.)

the coefficients so that the contribution of the mode k to the Hamiltonian is proportional to $a_k^\dagger a_k$, which is regarded as a number operator. Substituting the expressions of Sec. X.1 into the Hamiltonian (VII.1.8b), we obtain

$$H = \frac{1}{2}\int dk\ \{[B_k^2 a_k a_{-k} + B_k^{*2} a_k^\dagger a_{-k}^\dagger + |B_k|^2 (a_k a_k^\dagger + a_k^\dagger a_k)]$$

$$+ R^{2s} [k^2/R^2 + m^2][A_k^2 a_k a_{-k} + A_k^{*2} a_k^\dagger a_{-k}^\dagger$$

$$+ |A_k|^2 (a_k a_k^\dagger + a_k^\dagger a_k)]\}. \qquad (5.2)$$

(This applies to the case (1.15); in the general case, but with real eigenfunctions, one would have e_k^2 for k^2 and a_k^2 instead of $a_k a_{-k}$, etc.) So we must have

$$- B_k^2 = R^{2s} [k^2/R^2 + m^2] A_k^2. \qquad (5.3)$$

Eqs. (5.3) and (1.24) together provide three real equations for two complex quantities; the solution is unique up to an overall phase. If we arbitrarily require A_k to be real and positive, the solution is

$$A_k = \frac{1}{\sqrt{2}} R^{-s/2} [k^2/R^2 + m^2]^{-1/4},$$

$$B_k = -i \frac{1}{\sqrt{2}} R^{+s/2} [k^2/R^2 + m^2]^{+1/4} = -\frac{i}{2} A_k^{-1}, \tag{5.4}$$

where R stands for R(0). This makes the Hamiltonian take the form

$$H(0) = R^s \int dk\, [k^2/R^2 + m^2]^{1/2} (a_k^\dagger a_k + \frac{1}{2}\delta(k,k)), \tag{5.5}$$

where the delta function term is an infinite constant which may be discarded; H(0), thus normal ordered, is manifestly a positive self-adjoint operator in the Fock space of the operators a_k. (If we had used the original Gaussian time scale x^0 (see Eq. (1.7)), there would be no factor of R^s in H(0).)

Positive-frequency solutions in de Sitter space were defined in this way in Sec. V.3. There it was remarked that these solutions resemble the static-space positive-frequency solutions up through order t^2. In the general case this correspondence holds only through first order. In de Sitter space $\dot{R}(0)$ is zero, and hence the behavior of the system is static to one higher order in t. This slightly increases one's confidence in the physical relevance of the particle number defined in this way on a geodesic hypersurface in de Sitter space.

Local particle observables of the Newton-Wigner-

Wightman-Schweber type can be defined by Fourier-transforming a_k back to x-space without the factor A_k, in analogy to Eq. (VIII.4.3).

Obviously, the procedure applied here at $t = 0$ can be applied at any time. For different times it will, in general, yield different definitions of a_k; this is the phenomenon of particle creation. The operators a^1_k and a^2_k appropriate to different times t_1 and t_2 will be related by Bogolubov transformations of the form

$$a^2_k = \alpha^*(k)a^1_k + \beta(k)a^{1\dagger}_{-k}, \qquad (5.6)$$

where the coefficients obey Eq. (3.7). Let us calculate these coefficients. For j = 1 and 2 define

$$\Psi^j_k(t) = A_k(t_j)E^{t_j}_k(t) + B_k(t_j)O^{t_j}_k(t), \qquad (5.7)$$

where $E^{t_j}_k$ and $O^{t_j}_k$ obey the obvious generalization of Eqs. (1.22) and $A_k(t_j)$ and $B_k(t_j)$ are given by Eqs. (5.4) with $R = R(t_j)$. We have

$$\phi(t_2,x) = \int dk\,[e^{ik\cdot x}\Psi^1_k(t_2)a^1_k + e^{-ik\cdot x}\Psi^1_k{}^*(t_2)a^{1\dagger}_k], \qquad (5.8)$$

$$\pi(t_2,x) = \int dk\,[e^{ik\cdot x}\dot\Psi^1_k(t_2)a^1_k + e^{-ik\cdot x}\dot\Psi^1_k{}^*(t_2)a^{1\dagger}_k].$$

But $\phi(t_2,x)$ and $\Pi(t_2,x)$ can also be expressed in terms of the quantities with the index 2, and those equations can be inverted as in Eq. (1.18)[12]:

$$a_k^2 = -i[-B_k(t_2)\int dx\, e^{-ik\cdot x}\,\phi - A_k(t_2)\int dx\, e^{-ik\cdot x}\,\Pi].$$

Substituting, we obtain

$$\alpha^*(k) = i[B_k(t_2)\Psi_k^1(t_2) + A_k(t_2)\dot{\Psi}_k^1(t_2)], \qquad (5.9a)$$

$$\beta(k) = i[B_k(t_2)\Psi_k^{1*}(t_2) + A_k(t_2)\dot{\Psi}_k^{1*}(t_2)]. \qquad (5.9b)$$

If $\beta(k) \equiv 0$, Eq. (5.6) is just a phase change and there is no particle creation. Otherwise, the Fock vacuum states of the operators corresponding to different times must be different. The question arises whether the Hilbert spaces of these Fock representations are the same; in other words, whether there is a unitary operator $U(t_2,t_1)$ such that

$$a_k^2 = U(t_2,t_1)\, a_k^1\, U(t_2,t_1)^{-1} \quad \text{for all } k. \qquad (5.10)$$

The statements about the S-operator in Sec. X.3 apply to U as

---

[12] Recall that $A_k^* = A_k$, $B_k^* = -B_k$, that $\Psi_k^j(t_j) = A_k(t_j)$, $\dot{\Psi}_k^j(t_j) = B_k(t_j)$, and that $\Psi_{-k} = \Psi_k$.

well. If the volume of space is finite and

$$\sum_k |\beta(k)|^2 < \infty, \qquad (5.11)$$

then the Hilbert spaces are the same and U exists. In this case the initial vacuum $|0\ t_1\rangle$ can be expanded in terms of the Fock basis at t_2, and the coefficients are to be interpreted as probability amplitudes for finding (finitely many) particles distributed in the various modes at t_2 if the universe was empty at t_1. Alternatively, $U(t_2,t_1)$ can be regarded as the time evolution operator of the states in a Schrödinger picture where the particle operators a_k, N_k (= $a_k^\dagger a_k$, the number operator), etc., are held fixed.[13] Note, however, that for reasons about to be explained the notion of a Schrödinger picture for the a_k is not quite the same as that of a Schrödinger picture for the field operators.

Eq. (5.10) does not mean, unless $R(t_2) = R(t_1)$, that

$$\phi(t_2,x) = U(t_2,t_1)\ \phi(t_1,x)\ U(t_2,t_1)^{-1}; \qquad (5.12)$$

because $\phi(t_j,x)$ and a_k^j are related by a <u>time-dependent</u> Fourier transform:

---

[13] In Sec. X.3 the phases of the a_k in the asymptotic region were tacitly "run back" to t = 0, and so an interaction picture instead of a Schrödinger picture was mentioned.

$$\phi(t_j,x) = \int dk\, A_k(t_j)\, [\varphi^{ik\cdot x} a_k^j + \varphi^{-ik\cdot x} a_k^{j\dagger}], \qquad (5.13a)$$

$$\Pi(t_j,x) = \int dk\, B_k(t_j)\, [\varphi^{ik\cdot x} a_k^j + \varphi^{-ik\cdot x} a_k^{j\dagger}], \qquad (5.13b)$$

where A_k and B_k depend on t_j through $R(t_j)$ (see Eqs. (5.4)).[14] Eq. (5.12) would give us not Eq. (5.10) but unitary equivalence of a_k^{1} with canonical operators $\bar{a}_k$ defined from $\phi(t_2,x)$ by means of an equation like Eq. (5.13a) (j = 2), but with $A_k(t_1)$ in place of $A_k(t_2)$ (and a similar equation for $\Pi(t_2,x)$). Any operator U satisfying Eq. (5.10) would have to be a composition of the U in Eq. (5.12) with a unitary operator implementing the automorphism $\bar{a}_k \longrightarrow a_k^2$.

We shall now determine when such an operator exists. Analogously to Eqs. (5.6,9) one finds that

$$a_k^2 = \frac{1}{2}\left[\frac{D(k;2)}{D(k;1)} + \frac{D(k;1)}{D(k;2)}\right]\bar{a}_k + \frac{1}{2}\left[\frac{D(k;2)}{D(k;1)} - \frac{D(k;1)}{D(k;2)}\right]\bar{a}_{-k}^{\dagger},$$

where

$$D(k;j) = R^{s/2}(t_j)\,[k^2/R^2(t_j) + m^2]^{1/4}.$$

Let

[14] Cf. Parker (1969), p. 1061.

$$\beta(k) = \frac{1}{2}\left[\frac{D(k;2)}{D(k;1)} - \frac{D(k;1)}{D(k;2)}\right].$$

If s = 1 and we are working in finite space, the sum (5.11) is precisely the one treated in Sec. F.4, which converges; therefore, unitary equivalence of the fields is equivalent to unitary equivalence of the particle operators. If $R(t_2) \neq R(t_1)$ and $s \neq 1$, $\beta(k)$ approaches a nonzero constant as $k \longrightarrow \infty$; thus the sum (5.11) diverges, and hence equations of the forms (5.10) and (5.12) cannot both hold. The factors $R^{s/2}$ make the crucial difference for s = 2 or 3 between this situation and the one discussed in Sec. F.4.

The distinction between the two types of equivalence is sufficiently confusing to justify a restatement. Our basic ansatz, expressed in Eqs. (5.2-5), defines at each time t_1 a representation of the field operator $\phi(t_1, x)$ (a distribution in the s-dimensional space variable x), which we may call the Fock representation for time t_1. However, the dynamics given by the field equation defines in the Hilbert space of the Fock representation for time t_1 a representation of the field at time t_2, $\phi(t_2, x)$, given by Eq. (5.8). One may ask whether this representation is equivalent to the Fock representation for time t_2; an affirmative answer is Eq. (5.10). An entirely different question is whether the time evolution indicated in Eq. (5.8) is unitarily implementable within the Fock representation for time t_1; this is the content of Eq. (5.12).

In an asymptotically static universe with $R_+ \neq R_-$

unitary equivalence in the sense (5.10) means that the <u>in</u> Fock representation is equivalent to the <u>out</u> Fock representation, whereas in the case (5.12) the <u>in</u>-representation is a "strange" representation of the <u>out</u>-operators. This seems to indicate that the more physically relevant type of equivalence is that of Eq. (5.10).

Another argument to the same point is the following. When $R(t_1) \neq R(t_2)$, the spatial universes at the two times are geometrically different, and it is not obvious that the field observables defined in these two spaces should be considered to be the same algebra. In a generalized Robertson-Walker universe the spaces at different times have the same "shape"; it is only this fact which allows one to make a natural identification of the points, so that Eq. (5.12) makes sense. In a more general universe there is no preferred coordinate system, and there is no particular reason to expect Eq. (5.12) to hold for an arbitrary system. Indeed, the range of the spatial variable x could be different at different times, as in the Kruskal metric (see footnote 3 of Chapter IX). On the other hand, in this general context it is still possible to define a Fock representation for each spacelike hypersurface by diagonalizing the Hamiltonian. One can then ask whether the Fock representations corresponding to different times are equivalent; instead of Eqs. (5.6-9) one will have to consider a more complicated Bogolubov transformation, like that discussed in Sec. IX.3. Equivalence in this context simply means that the Fock representation for time

t_2 is the same as the representation of the field algebra at time t_2 defined by the dynamics (in analogy to Eq. (5.8)) in the space of the Fock representation for time t_1. This condition may no longer be expressible as a momentum-space unitary equivalence of the form (5.12), since the structure of the spectrum of the Hamiltonian will, in general, be different at different times.

Finally, it will now be shown that Eq. (5.10) does hold for any two-dimensional closed generalized Robertson-Walker universe. (In the two-dimensional case one can always choose the spatial coordinate so that Eqs. (1.15) hold. The assumption that the space is closed (has finite circumference) is necessary to make k a discrete variable.) It is assumed that R(t) is bounded in the interval $t_1 \leq t \leq t_2$ and sufficiently well-behaved that the solutions of the wave equation (see Eq. (5.15) below) exist and are bounded in this interval.

We must investigate the sum (5.11) for $\beta(k)$ given by Eq. (5.9b); that is (taking $t_1 = 0$, $R(0) = 1$, $t_2 = t$, $R(t) = R$),

$$\beta(k) = \frac{1}{2}[k^2 + m^2R^2]^{1/4}[k^2 + m^2]^{-1/4} P^*_k(t)$$

$$+ \frac{i}{2}[k^2 + m^2R^2]^{-1/4}[k^2 + m^2]^{1/4} \frac{1}{\sqrt{k^2 + m^2}} \frac{dP^*_k(t)}{dt}, \qquad (5.14)$$

where $P_k(t)$ is the solution of Eq. (1.14a),

$$\frac{d^2 P_k}{dt^2} + [k^2 + m^2 R^2] P_k = 0, \qquad (5.15)$$

satisfying

$$P_k(0) = 1, \qquad \frac{dP_k(0)}{dt} = -i\sqrt{k^2 + m^2} \qquad (5.16)$$

(cf. Eqs. (5.7) and (5.4)). Thus

$$\beta(k) = \frac{1}{2}[1+O(k^{-2})]\, P^*_k(t) + \frac{1}{2}[1+O(k^{-2})]\, \frac{i}{\sqrt{k^2 + m^2}} \frac{dP^*_k(t)}{dt}. \qquad (5.17)$$

(The expansion of the radicals is as in Sec. F.4.) Hence if, for a fixed t, $P_k(t)$ and $\dot{P}_k(t)$ are bounded and

$$X_k(t) = P_k(t) - \frac{i}{\sqrt{k^2 + m^2}} \frac{dP_k(t)}{dt} \qquad (5.18)$$

vanishes as $k \longrightarrow \infty$ at least as fast as $1/k$, we will have

$$|\beta(k)|^2 = O(k^{-2}),$$

and the sum will converge, which was to be proved.

Now P_k obeys the integral equation

$$P_k(t) = e^{-i\omega t} - \frac{1}{\omega}\int_0^t ds \sin[\omega(t-s)]\, m^2 (R^2(s)-1)\, P_k(s), \qquad (5.19)$$

where $\omega = \sqrt{k^2 + m^2}$. (Eq. (5.19) is constructed from the Green function for Eq. (5.15) with R = 1 appropriate to the initial conditions (5.16). Eqs. (5.15-16) can be verified by differentiation of this equation.) Therefore, the maximum value of $P_k(s)$ for $0 \leq s \leq t$ is

$$|P_k|_{max} \leq 1 + \frac{t}{\omega} m^2 |R^2 - 1|_{max} |P_k|_{max} = 1 + A|P_k|_{max} .$$

For sufficiently large k (t fixed), $A < 1/2$ (say), and so

$$|P_k|_{max} \leq \frac{1}{1 - A} < 2.$$

The same argument applied to the derivative of Eq. (5.19) shows that $\dot{P}_k/\omega$ is bounded for large k. Also, we have

$$X_k(t) = - e^{i\omega t} \frac{1}{\omega} \int_0^t ds \, P_k(s) \, m^2 (R^2(s)-1) \, e^{-i\omega s} .$$

Since $P_k(s)$ and $R^2(s) - 1$ are bounded, it follows that

$$|X(t)| < \frac{\text{const.}}{\omega} \sim \frac{1}{k},$$

as desired.

It is clear that this proof cannot be extended to higher dimensions without some better estimate on the decrease of X(k) at infinity. Also, the theorem is certainly not true for two-dimensional Robertson-Walker universes of infinite spatial extent, for the reason indicated in connection with Eqs. (3.11-12).

6. Critique of the Definition; Parker's Alternatives.

The viewpoint of the last section, if pressed to its extreme, is that the physical particles present in the universe at time t are given by the expansion of the state in terms of the stationary states of the instantaneous Hamiltonian H(t), these eigenstates being given the particle interpretation they would have if H(t) were a static Hamiltonian (see Chapter VIII). This is equivalent to the assumption that in studying particle creation between times t_1 and t_2 one may replace the actual metric by the asymptotically static one for which

$$R(t) = R(t_1) \text{ if } t < t_1, \quad R(t) = R(t_2) \text{ for } t > t_2,$$

and R(t) coincides with the original R(t) in between, and one may give the field in the static regions the usual interpretation.

It has been questioned (e.g., Parker (1969), Sec. F) whether this ansatz is justified physically. For one thing, it is not obvious that the fact that the metric is changing in time is irrelevant to the way in which the excitation of the field manifests itself in particlelike behavior. After all, the equivalent metric mentioned above involves a violent sudden change in the behavior of R(t), which might be expected to affect the particle number discontinuously. Also, since any actual measurement takes a finite time, it is not clear that the instantaneous particle "observables" have any operational meaning when the metric is changing rapidly. "There is no reason why a precisely defined operator should correspond to the physical particle number when [the creation rate] does not vanish."[15] Finally, Parker (1966, 1969) finds that the particle number density defined by diagonalization of H(t) diverges when summed over all modes (for s = 3) and has rapid oscillations which are in principle unobservable.

Parker has suggested replacements for the instantaneous particle operators of Sec. X.5 which avoid these divergences and oscillations. They are based on the observation that

$$\bar{\psi}_k(t) = A_k(t) \exp\{- i\int^t dt' \, A_k(t')^{-2}/2\} \qquad (6.1)$$

is a first-order adiabatic approximation to $\psi_k(t)$. He rewrites the field expansion (1.16) in terms of $\bar{\psi}_k$ (in effect), thus factoring out and isolating the deviation from adiabatic behavior of the time dependence of the field. In his thesis (Parker (1966)) a second-order adiabatic approximation was used to define an approximate number operator which is constant during the time of a measurement. In the published version (Parker (1969)) he introduced instead the postulate that $A_k(t)$ in Eq. (6.1) should be replaced by a new function, chosen to minimize the expectation

[15] Ibid. Cf. Moore (1970): "In [a period when the external conditions are time-dependent] the very concept of photons becomes muddy, just because the absence of photons (namely, a time-translationally invariant vacuum state) cannot be defined."

values of the derivatives of the number operator for mode k.[16]

The opinions of the present author concerning particle operators at fixed time are the following:

(1) It is quite likely, for the reasons mentioned, that the creation and annihilation operators defined in Sec. X.5 have very little to do with physical particles which would actually be detected by some experimental apparatus. In fact, the very idea of a particle may not be applicable to the behavior of the field when R(t) is rapidly changing.

(2) Nevertheless, the particles defined in terms of the instantaneous creation and annihilation operators at a fixed time may provide the most convenient way of labeling the states when the evolution of the system between finite times is studied. These quanta should be called <u>virtual particles</u>. If the virtual particle concept turns out to be useful in this context, how much reality one attributes to these particles when they are, strictly speaking, unobservable is largely a matter of taste.

(3) What operators, if any, correspond to real, observable particles cannot be decided on the basis of mathematical properties alone. Any identification of

---

[16] This is to be done (L. Parker, private communication) consistently with the discussion surrounding Eq. (32) of the paper, where W(k,t) is determined by R(t) and its derivatives. Dr. Parker believes that for a slowly expanding universe this procedure will agree with the one in his thesis.

observables must ultimately be tested against some (perhaps very crude) physical analysis of the measurement process itself. Such a program is somewhat circular, because one needs a complete theory, including interactions, before a model of interaction of the field with some other system can be studied with confidence. Nevertheless, this does not mean that progress cannot be made.

(4) Since the practical interest in our subject relates mostly to relativistic astrophysics and cosmology, it may be wise to concentrate on the energy-momentum tensor as the observable to be analyzed (cf. Sec. IX.5), rather than the response of a hypothetical apparatus to detect individual particles. (The investigation might or might not still be conducted in terms of a particle formalism.)

(5) A possible starting hypothesis for such an investigation would be that physical particles (which are detected directly or which indirectly (for instance, through a normal-ordering prescription) enter the definition of observable quantities such as energy) correspond to some kind of smoothed-out particle operators such as those defined by Parker.

Such a project far exceeds the scope of the present work. In the next section, however, are collected a few reflections concerning how the particle concept arises in traditional field theories and why it threatens to lose its validity in the situations of interest to us.

7. Remarks toward an Analysis of the Particle Concept.

The problem we face has to do with both the mathematical definition and the physical interpretation of a quantum field theory in a Riemannian space-time. As to the first, what constitutes a field theory, we early rejected the naive notion that a field equation and a commutator function are enough, and we set about trying to define a Hilbert space of state vectors in which the field would be represented as an operator-valued distribution. When it turned out that the choice of representation was problematical, a sophisticated cousin of the naive idea presented itself: the doctrine of physical equivalence of all faithful representations of an abstract algebra of observables. This point of view, however, leaves us still embarrassed with respect to interpretation.

The interplay of the mathematical and the observational aspect of the problem is hinted at in the anecdote of I. E. Segal which stands at the beginning of this dissertation. Segal has interpreted "the occupation number formalism" to mean the algebraic structure of the canonical commutation relations, to which he gives an abstract formulation; most physicists would think rather of the interpretation of the states of the theory in terms of configurations of particles; but they are concerned with the same problem. As long as our experiments involve observations of particle events rather than measurements of field strengths, it seems to be necessary to add some structure to the abstract framework of quantum field theory in order to complete the link between theory and the world. Indeed, in one of the papers in which the Wightman axioms (see Appendix E) were proposed, it was stated, "Practically, most measurements are made on particles not fields, and a relativistic quantum theory is not really complete unless it includes some kind of particle observables."[17]

Why and how are field theories interpreted in terms of particles? There seem to be two elements involved, a physical one and a formal one. Let us review how the particle concept is introduced into the theory of a free field, where the problem seems to be completely under control. It arises first from the physical fact that corpuscular behavior is observed in the situations where free field theory is applicable.[18] In particular, one observes motion in straight lines at constant speeds, which, together with the conservation observed in interactions (interactions which our theory does not attempt to describe, of course), leads to the concepts of momentum and energy. Secondly, the solution of the field equation by

[17] Wightman and Gårding (1965), p. 156. See also Wightman and Schweber (1955).

[18] This regime can be characterized only by a circular statement such as "Particles behave essentially freely when they are far from one another."

separation of variables leads to an expansion in terms of creation and annihilation operators:

$$\phi(t,x) = \int \frac{d^3k}{\sqrt{2\omega_k}} \left[e^{i\vec{k}\cdot x} e^{-i\omega_k t} a_k + e^{-i\vec{k}\cdot x} e^{i\omega_k t} a_k^\dagger \right]. \quad (7.1)$$

The operator $N_k = a_k^\dagger a_k$ has the spectrum appropriate to a number operator. It is natural to associate the index k with the physical momentum of the particles "counted" by N_k. This interpretation has been found entirely consistent and satisfactory. (See any textbook on quantum field theory for details.)

In external potential problems, including the gravitational, this treatment can be imitated only partially, and only in special cases.[19] First of all, if the external field is static[20], one can define eigenstates of the Hamiltonian analogous to the momentum eigenstates in the case of the free field. Thus the theory has a particle, or quantal, structure (cf. Chapter VIII). However, if the potential does not vanish asymptotically one might question the identification of these quanta with physical particles, especially in view of the phenomena pointed out in Chapter IX.

---

[19] The reader will note a tendency for this section to repeat Sec. VII.7. This is simply a manifestation of the point emphasized above; a discussion of the mathematical structure of field theory tends to parallel a discussion of its meaning.
[20] It must also be such that the single-particle squared Hamiltonian is self-adjoint and has no "jelly modes" -- see Secs. VIII.1-2.

On the other hand, the asymptotic approach applies if the potential falls off in space and at least becomes asymptotically static in time[21], or if it falls off in time. Then there is a representation (out-representation) such that every vector state can be interpreted as describing a configuration of particles which, after a sufficient time, are either out of the range of the potential or in stable bound states. Similarly, there is an in-representation; if these two representations are unitarily equivalent (cf. Sec. X.3), one has a distinguished representation for the fields and a particle interpretation which is adequate for the description of scattering processes.

In both these approaches the particle interpretation is closely tied to the canonical structure. This is obvious for the first approach, which works from the formalism to the interpretation. The asymptotic method works in the other direction. Although in this case the canonical structure of the Heisenberg-picture field does not play a role (and is not even expected to be valid for general interacting fields), the particle structure of the asymptotic states leads to asymptotic fields, which are free fields obeying the canonical commutation relations (see [Streater-Wightman], pp. 26-27).

It is the asymptotic approach which is usually taken in field theory, both for external potential problems and for

[21] The latter condition is imposed to avoid the problem of a time-dependent potential which continually emits particles.

interacting fields. Wightman and Gårding (1965) say (p. 157):

> In view of the current state of ignorance [concerning the relation between particle observables and (nontrivially interacting) fields], there is no alternative to setting aside the notion of particle observable for further study and accepting something much weaker, that of asymptotic particle observable. This procedure is also advantageous from another point of view. It is quite possible that relativistic field theories exist in which asymptotic particle observables can be defined but not particle observables at each fixed time. This would not be unreasonable physically for it might be that the notion of particle can only be defined in some limiting sense in which the particles are far from one another.

Hence the axiom of asymptotic completeness (see Appendix E).

In Sec. VII.7 it was argued that in gravitational external potential problems one cannot be satisfied with an asymptotic particle interpretation. We must either return to the unfinished business of defining true particle observables, or make physical sense out of field observables without the particle concept.

The author has come to believe that the second of these alternatives deserves serious consideration. We know from the analysis of Wigner (see Sec. V.2) that the particle concept is closely related to the representation theory of the Poincaré group, the symmetry group of flat space. It seems quite reasonable that the notion of particles should weaken in an external potential situation which departs greatly from the free Lorentz-invariant one. Particle behavior is recovered in "asymptotic" theories because there is a region of space-time within which the dynamics is approximately free. If there is no such region, if it is impossible for the excitations in the quantized field to separate themselves from the irregularities in the environment represented by the external field, then, as Wightman and his collaborators have said, the particle concept may simply not apply. Elementary particles may be very much like the quasiparticles of solid state physics. The quantum theory of the physical system composed of the atoms or the electrons in a crystal predicts particle-like excitations (phonons or plasmons). But these quasiparticles are stable only for a perfect crystal. The more severe the impurities or dislocations or external perturbations, the faster the quasiparticles decay away, until finally the concept becomes useless. Then one must go back to the substratum: in solid state physics, the atoms and electrons; in more fundamental physics, the field.

In the case of quantum fields in curved space-time one can expect on general physical grounds four regimes with respect to the usefulness of the particle and field concepts, depending on the properties of the metric:

(1) The single-particle domain: a convincing definition of particle observables exists according to which no particles are created, or a negligible number. Then a satisfactory single-particle quantum theory exists, and the apparatus of field theory is not really needed. This would be the case for the static theories of Chapter VIII, to the extent that the definition of particle therein is regarded as trustworthy. It is also true of time-dependent models where

the creation is negligible.[22]

(2) <u>The many-particle domain</u>: a particle interpretation exists, and the particle number is nonconstant to a nonnegligible extent. This is the regime which we have implicitly assumed to be of interest earlier in this chapter. It remains to be shown, however, how wide this band is, or even that it exists at all.

(3) <u>The field domain</u>: the concept of an external gravitational field[23] applies, but there is no particle

---

[22] The literature on relativistic wave mechanics (e.g., Feshbach and Villars (1958)) often implies that the Klein-Gordon or Dirac equation has an obvious one-particle interpretation for "sufficiently weak" external potentials, without any indication of a crucial difference between time-independent and time-dependent potentials. However, the observations of Secs. X.1 and X.3-5 apply to time-dependent electromagnetic (etc.) as well as gravitational potentials: There is no obvious way to separate the solutions of the wave equation into positive- and negative-frequency parts, and, moreover, any physically plausible definition of the splitting which is adopted will probably give different results at different times -- that is, it will predict particle creation. The origin of the conventional wisdom seems to lie in the fact that if the high-frequency Fourier components of an asymptotically static potential are very small, the vacuum-to-vacuum S-matrix element, $\langle 0 \text{ out} | 0 \text{ in} \rangle$, is very close to 1. That is, the <u>out</u>- and <u>in</u>-vacuums can be identified, and no "real" particles are created. (This effect is necessary for conservation of energy; the energy required to create a pair ($\geq 2mc^2$) must be extracted from the modes of the external field which carry sufficiently large energy, $\hbar\omega$.) In such a case the virtual particles occurring at finite times in a field-theoretical treatment can be disregarded, and then the field equation can be taken as the equation for the wave function of a single particle. To the extent that the vacuum-to-vacuum amplitude differs from 1, this kind of theory is troubled by a version of the Klein paradox.

[23] Among the external gravitational field problems one should include not only models in which the metric is prescribed once and for all, but also theories in which the (classical)

interpretation of the quantized scalar field theory.

(4) The domain of quantum general relativity: the interaction of the gravitational field, as a dynamical object in its own right, with the quantized matter fields must be taken into account. For consistency it will probably be necessary to treat the gravitational field in a quantum-theoretical manner; yet it cannot be an ordinary quantum field, since it itself determines the manifold on which the fields must be defined. Let us leave this problem to the physicists of the future and return to the external field framework.

We have been speaking of "real" particles -- modes of observable behavior which conform to our intuitive notion of particle, which is derived from observation of particles which are almost free. There are also definitions of virtual particles which arise out of the field-theoretical formalism. Any splitting of the field into annihilation and creation parts, as in Eq. (1.16), gives rise to a notion of virtual particles. Some of these are more likely than others to be useful, just as in ordinary quantum mechanics some choices of basis in the Hilbert space are more convenient than others. (Compare the discussion of "useful representations" near the end of Sec. IX.4.) A

---

gravitational field is influenced by the matter field through some "self-consistency" scheme, such as the work of Ruffini and Bonazzola (1969).

virtual particle language may be a useful way of classifying the states of a field theory for certain purposes even if the virtual particles have nothing to do with real particles. On the other hand, a "good" notion of virtual particles may turn out to correspond to a weakened notion of real particles (as just described), or may yield a physically relevant notion of current or energy density (see Sec. IX.5). The nonuniqueness in plausible definitions of virtual particles is not a conclusive counterargument against this hope, since as the notion of real particles weakens, there is more room for ambiguity in it. In this spirit a general definition of particle observables at fixed time for an arbitrary Riemannian manifold will be offered in the next section.

Let us consider an example of how the particle concept becomes fuzzy, and at the same time ambiguous, in a sequence of external potential problems. Consider a neutral scalar field of mass m interacting with a scalar potential; that is, a wave equation of the form

$$\Box\varphi + m(m - V(t,x))\varphi = 0. \tag{7.2}$$

Assume at first that the potential has compact support in time:

$$V(t,x) = 0 \quad \text{for } |t| > T. \tag{7.3}$$

There are clear definitions of particles in the asymptotic regions (the _in_- and _out_-representations); the particle operators are defined in terms of the field operators ϕ and π through a Fourier transformation (cf. Eq. (3.2)). If the potential also falls off sufficiently rapidly in space, Schroer *et al.* (1970) have proved that the time evolution of the fields is unitarily implementable in the *in*-representation (which is therefore equivalent to the *out*-representation). It follows that if creation and annihilation operators at each time t are defined in terms of $\phi(t)$ and $\pi(t)$ by a Fourier transformation (cf. Eqs. (IX.3.1)), then the Fock representations of all these sets of particle operators are equivalent (to each other and to the asymptotic representations). If this is taken as the definition of particles, the relation between the observables and the field at a given time is the same as for the free field. In particular it does not depend on the time. The definition of the vacuum (as a Heisenberg state) does not change with time. (Of course, the evolution of the system will take the vacuum, as a Schrödinger-picture state, into states with a probability for the presence of particles at another time.)

On the other hand, consider a potential which is independent of t. Then one is led, as in Chapter VIII, to define particle operators via an expansion in the eigenfunctions of

$$-\nabla^2 + m(m - V(x)), \qquad (7.4)$$

rather than of the Laplacian alone, as above. It is crucial to note that this natural procedure for the time-independent case, followed by Schroer and Swieca (1970), is fundamentally different from that just described for the case of compact support in time, adopted in the adjacent paper by the same authors (Schroer *et al.* (1970)).

We can confront these methods with each other by considering a potential satisfying Eq. (7.3) and also

$$V(t,x) = V_0(x) \quad \text{for } |t| < T' < T. \tag{7.5}$$

We may imagine T' to be many times larger than the age of the actual universe, so that it would be absurd to claim that there is a significant difference between this situation and the previous one with respect to the physics around the time $t = 0$. But then for the potential of Eq. (7.5) in the vicinity of $t = 0$ the quantization based on a V-dependent integral transform, as for a static potential, is a serious competitor of the quantization described above for a general potential of compact support, based on the Fourier transform. For some potentials these representations are unitarily inequivalent; in fact, Schroer *et al.* remark that the former may have "jelly" (or indefinite metric) troubles while the latter (being an ordinary Fock representation) is perfectly normal.

This example demonstrates that the ambiguity in the definition of virtual particles can assert itself already in external potential problems; it is not inseparably connected to the general covariance of the gravitational problem. Which of these particle notions, if either, corresponds to real, observable particles? It seems to the author that if T' is very large, a case can be made for the V-dependent definition, which yields a vacuum (and one-particle eigenstates, etc.) which are stationary states during the long static period. On the other hand, if T is rather small, the V-independent particle concept may make more sense, since in that picture (if the potential is of fast decrease in space) there are always only finitely many particles present, before, during, and after the interaction. Obviously, there must be a regime in between where the particle notion becomes fuzzy.

There is nothing to keep one from applying the V-dependent quantization at each time, even if V is not constant in a finite time interval, although its physical relevance is more questionable in this situation. (This procedure can also be specified as follows: At each time, choose the representation which makes the instantaneous Hamiltonian a positive operator by explicitly diagonalizing it into the form of a linear combination (or integral) of number operators. The method of Sec. X.5 was of this type, and it will be recommended again in the next section (Eq. (8.1)). The other kind of quantization considered here for the external scalar potential does not have a plausible analogue in the generally covariant gravitational context, since it essentially depends on a fixed Cartesian coordinate system.) If this is done, the instantaneous vacuum is not constant, even in the Heisenberg representation, because the relation between the particle observables and the field operators depends on time.[24] The vacuum may even wander through mutually inequivalent representations. Consequently, if the vacuum corresponding to each time is to be a vector in the Hilbert space of the system, the Hilbert space cannot coincide with the cyclic space generated by any one of these vacuum states, but is much larger. Thus one would not have, as usually assumed, a single irreducible representation of the fields, within which the time evolution is unitarily implemented.

Now let us consider potentials which do not satisfy Eq. (7.3), and in fact fall off so slowly in space and time that no asymptotic representations can be defined. Then if we are to have particle observables at all, they must be definable at finite times. In analogy to what has gone before, two definitions offer themselves: (1) a simple Fourier transform, the potential being ignored; (2) a V-dependent decomposition at each time.

Let us look at the implications of these two approaches in the very special case

$$V(t,x) = \text{const.} \neq 0. \qquad (7.6)$$

From point of view (2), the field is just a free field[25] of

[24] Compare the distinction which was drawn in Sec. X.5 between two types of equivalence of representations.
[25] Is there no operational distinction between a free field of mass M and a field of mass m interacting with a constant potential V? The answer depends on whether it is possible to have an apparatus which detects the quanta of mass m, as opposed

mass $M = \sqrt{m(m - V)}$. If the approach (1) is followed, however, one has pair creation of the wildest sort: the representation of the fields for which the $a_k(0)$ have a vacuum state is not equivalent to that for which the $a_k(t)$ have a vacuum state ($a_k(s)$ being defined at each time s by a Fourier transform of the form (IX.3.1) with the mass m). This statement follows[26] from the discussion of Wightman (1964), pp. 251-255, if one (1) interchanges the role of "free" (or "interaction-picture") and "interacting" fields and (2) notes that the m-particle operators at different times are equivalently represented if and only if the corresponding M-particle operators are, since the connection between m- and M-particle operators at equal times is independent of time. (The latter point also implies immediately that $a_k(0)$ and $a_k(t)$ <u>are</u> unitarily equivalent in the Fock representation of the M-operators. That is, there is a representation (which is not the Fock representation of the $a_k(0)$!) in which the time evolution of the a_k is unitarily implementable. This, however, is a special property of the interaction (7.6), which will not

---

to the quanta of mass M, given that the field obeys Eq. (7.2). (Recall that absence of quanta of one mass does not mean absence of quanta of the other mass -- see Sec. F.4.) It seems that the answer has traditionally been assumed to be <u>no</u>, since fields with Lorentz-invariant quadratic interactions (of which this is a special case) are regarded as trivial since they are equivalent to systems of free fields (see Jaffe (1965), Chapter XII, and Wightman (1964), pp. 180-182). The author believes, but has not been able to demonstrate in a model, that any reasonable detector will "come to equilibrium" with the physical (stable) vacuum state and hence will be sensitive only to M-quanta (cf. Sec. VII.7).

[26] See also Grib (1969).

persist for general potentials.)

Grib (1969) has described the situation which exists for the m-operators (and exists in general when there is a time evolution which is not unitarily implementable within a given irreducible representation) as follows: The Heisenberg picture exists, but the Schrödinger picture does not. This statement may be interpreted in two ways, depending on what kind of quantum theory one has in mind:

(1) If one insists on an irreducible representation, it must be chosen arbitrarily (say as the Fock representation for the particle operators defined at one particular time). Then within this Hilbert space the time evolution of the fields is defined as an automorphism, but there is no unitary propagator $U(t_2,t_1)$,

$$a_k(t_2) = U(t_2,t_1)a_k(t_1)U(t_2,t_1)^{-1} , \qquad (7.7)$$

by means of which a Schrödinger picture could be defined.

(2) If one allows a direct sum of many representations, a propagator $U(t_2,t_1)$ can be defined, but since it mixes the representations it will not be differentiable[27] in t_2, and hence an infinitesimal form of

[27] If $\Psi(t)$ and $\Psi(t + \Delta t)$ are unit vectors in two different components of the orthogonal direct sum, their difference has norm $\sqrt{2}$. Thus the difference quotient in the definition of the derivative does not converge as $\Delta t \longrightarrow 0$.

the Schrödinger picture,

$$i\frac{d\Psi}{dt} = H(t)\Psi \quad \text{where} \quad \frac{dU(t_2,t_1)}{dt_2}\bigg|_{t_2=t_1} = -\,iH(t_1), \qquad (7.8)$$

will not exist.

Another possible approach to systems of this type, advocated by Kristensen *et al.* (1967), is to identify the state at each time with a nonnormalizable distribution in a rigged Hilbert space associated with the Hilbert space of some irreducible representation (such as that corresponding to an initial time).

Other examples for which a similar analysis could be made are

$$V = 0 \quad \text{for } t < 0, \qquad V = \text{const.} \neq 0 \quad \text{for } t > 0 \qquad (7.9)$$

and

$$V = 0 \quad \text{for } |t| > T, \quad V = \text{const.} \neq 0 \quad \text{for } |t| < T. \qquad (7.10)$$

These will be left for the reader's contemplation. It should now be clear that in the case of a general scalar potential the splitting of the second term in Eq. (7.2) into a mass part and a potential part is essentially arbitrary. (This ambiguity is analogous to that in the gravitational problem which is associated, at least in part, with general covariance. In the electromagnetic case similar games may be played with the freedom in the choice of gauge.) Moreover, (in general) no choice should be expected to yield a Fock representation within which the time evolution is unitarily implemented. Thus it is impossible to maintain a belief in a unique irreducible "physical" representation.

All the virtual particle concepts are nonlocal, in the sense that they are based on eigenfunction expansions of the field which involve integrations over a whole hypersurface. The resulting concepts of vacuum necessarily have a very global character, as shown by the Reeh-Schlieder theorem (see Sec. IX.3) and by the observations of Sec. IX.7. This situation is in conflict with our intuitive notion of particles. As has been stated above many times, an asymptotic interpretation of field theory in Riemannian space-time is unsuitable; that is why we have stuck so closely to the canonical formalism, in the hope of extracting a more local definition of particles. Perhaps, though, the canonical formalism itself is still too global. In the language of J. A. Wheeler, physical intuition tells us that the universe is a vast haystack[28], but the canonical formalism forces us to regard it as a stack of automobile fenders[29]. This is an additional argument for the necessity of an identification of observables directly in terms of fields, avoiding the intermediate concept of particles.

However, in the absence of a brilliant idea which would

---

[28] Marzke and Wheeler (1964), p. 42.
[29] Wheeler (1963), p. 346.

tell us how to develop the field domain (3), we have little choice but to continue trying to work in the many-particle domain (2), at least in order to discover its limits. Out of the many possible ways to define virtual particles, which has the best chance of corresponding to something like real particles? An answer is suggested in the next section.

8. A Proposal.

What follows does not contain any new ideas which will resolve the impasse described in the previous section. Rather, it combines some simple ideas which have guided the whole of this dissertation -- namely, the straightforward extension of the canonical formalism to a manifold and the geometrical or kinematical ideas (stated in Appendix D and Sec. III.3) which motivate Gaussian and Fermi coordinate systems -- so as to define, with as little ambiguity as possible, a quantization of the scalar field. By a quantization is meant a splitting of the field at each time into annihilation and creation parts, in terms of which a representation of the Fock type can be constructed as in Chapter VIII and in the earlier sections of this chapter. Thus it may also be regarded as a definition of particle observables at each time. Admittedly, it may be that this approach is too naive, and that some totally new idea is needed. What is intended here is to specify the physically most reasonable ansatz of this type as a starting point for further research (see the remarks at the end of the section).

The first half of the proposal is:

> To define particle observables on a spacelike hypersurface, express all quantities in terms of the Gaussian coordinate system based on that hypersurface, and construct the Fock representation which makes the corresponding instantaneous Hamiltonian a manifestly positive self-adjoint operator. (8.1)

(In other words, the Hamiltonian to be made positive is defined in terms of a normalized normal derivative to the hypersurface -- see Sec. IV.2.) This procedure has already been carried out in Secs. X.1 and X.5 for the case that the field equation can be solved by separation of variables; a similar construction was performed in Secs. VIII.1-3 in greater detail under the assumption that the metric is static (but not necessarily Gaussian). The important point to notice is that as long as only <u>one</u> hypersurface is considered at a time, these restrictions are unimportant. In the general case one can still expand the initial values of ϕ and π on the hypersurface in terms of eigenfunctions ψ_j of the Laplace-Beltrami operator on the hypersurface in such a way that the coefficients a_j, $a_j^\dagger$ satisfy the commutation relations for creation and annihilation operators and the Hamiltonian (VII.1.8b) assumes the "diagonal" form

$$H(0) = \int d\mu(j)\ E_j\ a_j^\dagger a_j$$

(cf. Sec. VIII.2).

The second suggestion is:

Define particle observables on the <u>geodesic</u> hypersurfaces orthogonal to a given timelike curve. (8.2)

(These are the surfaces of constant time in some Fermi coordinate system.) The point of this requirement is to take the distortion out of the hypersurfaces as much as possible. In the neighborhood of the point where it cuts the central worldline a geodesic hypersurface is as near to a flat hyperplane as one can come in a curved space. Unfortunately, such a hypersurface may not be geodesic relative to another point[30], so this notion of flatness is not absolute, but relative to a point. (Of course, a timelike direction at the point is also needed to determine the hypersurface uniquely.) The imbedding of a given geodesic hypersurface into a family of hypersurfaces orthogonal to a given curve does not place any additional restriction on it, and hence is not really necessary in the statement of the procedure. But this seems to be the natural way to fit hypersurfaces together into a full kinematical scheme in which to describe the history of the system. The curve (which is not necessarily a geodesic)

---

[30] That is, a geodesic curve tangent to the hypersurface at a point other than the original point may not lie entirely in the hypersurface. See Appendix D.

can be interpreted as the worldline of an observer (see the discussion in Sec. III.3).

For the instruction (8.1) to make sense, the Laplace-Beltrami operator on each geodesic hypersurface must be self-adjoint. Presumably -- at least if the metric coefficients are smooth and bounded at finite points -- this will be true if and only if the hypersurface is geodesically complete.

The energy-momentum tensor can now be unambiguously defined by normal ordering the formal expression for it with respect to the a_j, $a_j^\dagger$ appropriate to a hypersurface on which it is to be evaluated. From the point of view of cosmological and astrophysical applications, $T^{\mu\nu}(x)$ is presumably more important than the particle observables themselves. As remarked in Sec. IX.5, an energy density defined this way will not be positive as an operator.

One must be prepared for the possibility that the representations defined by different hypersurfaces of the family are inequivalent, and even that the difference between the respective normal-ordered energy densities is infinite (does not make sense even as a distribution). This would have to be either taken as sufficient reason for rejecting the theory, or interpreted as creation of an infinite density of particles. In the latter case, obviously, there will be difficulties in using $T^{\mu\nu}$ in the Einstein equations. In the former case one might try to define smoothed-out particle observables like Parker's (Sec. X.6), or to develop a new kind of field observable, as urged in the last section. It is too early to say whether these measures are necessary.

A strong argument in favor of the proposal (8.1-2) is that it disposes of the ambiguities of quantization in Minkowski space in the most conservative and reassuring way: it specifies the standard free field quantization to be the correct quantization. The type of alternative representation considered in Sec. X.2 is rejected because the hypersurfaces of constant time violate the condition (8.2). It is legitimate to take the hypersurfaces of constant time to be the hyperplanes orthogonal to the worldline of a uniformly accelerated observer (the lines of constant v in Chapter IX); however, one is told by the rule (8.1) to use the ordinary ΦOK representation on each of these hyperplanes, rather than the Rindler representation of Chapter IX. (Since the ΦOK representation is Lorentz-covariant, the successive hyperplanes yield the same definition of the vacuum in this case.)

In the case of de Sitter space, the principles (8.1-2) tell us to reject the covariant quantization of Sec. V.6 and the "static" quantization of Sec. VIII.6 (and also, for instance, a quantization (of the type of Sec. X.5) based on the horospherical coordinate system of Sec. III.7). Instead, we are to use the representation of Sec. X.5 (and Secs. V.3,5), but _only in the neighborhood of a geodesic hypersurface_[31]. The history of the

---

[31] Recall that a geodesic hypersurface in two-dimensional de Sitter space is the "neck" of the hyperboloid in Fig. 3, or any

system is most naturally given in terms of the time coordinate of the Fermi system of an inertial observer. In contrast to the static theory, in this theory there may be particle creation when the particle observables on different hypersurfaces of the Fermi family are compared. This effect will be calculated in Sec. X.10. In Sec. X.9 it will be shown that the representation prescribed here for each time is unitarily equivalent to the covariant representation; so the convenient properties of the latter may be exploited as a technical tool, even though the associated particle interpretation is not the physically correct one according to our present point of view.

Other simple models on which this method could be tried out are the Schwarzschild solution and the various Friedmann cosmological solutions. In each of these cases the prescription (8.1-2) will give different results from the previously discussed approaches.

In the Schwarzschild case quantization is to be based on a Gaussian time coordinate relative to a hyperplane extending through the entire Schwarzschild-Kruskal solution, rather than on the usual time coordinate t, defined in the exterior region alone, with respect to which the exterior metric is static. One therefore expects particle creation, <u>even in the exterior region</u> (where one naturally compares observables on different hypersurfaces of constant t, in accordance with Eq. (8.2)). This

---

curve isometrically related to it, such as the family of ellipses of constant χ in Figs. 4-5.

would reflect the dynamic nature of the Schwarzschild-Kruskal solution as a whole. It would be interesting to see whether this effect is significant and whether it is physically reasonable. One should make calculations in this theory and compare with the quantization of the field in the exterior region viewed as a static space, which does not predict particle creation.

The Friedmann universes are special cases of the Robertson-Walker metrics discussed in Sec. X.1, and previous work (see Parker (1968, 1969) and Grib and Mamaev (1969)) has been carried out in the coordinate system in which the Robertson-Walker form is manifest.[32] These hypersurfaces of constant time are not, in general, geodesic hypersurfaces, however. Therefore, the prescription (8.2) indicates a different approach; a comparison would be interesting.

9. Unitary Equivalence of Covariant and Positive-Frequency Quantizations in Two-Dimensional De Sitter Space.

Among the two-dimensional closed Robertson-Walker universes is the de Sitter universe described in geodesic Gaussian coordinates, which has been discussed in Secs. III.1, III.3, IV.2, and V.1-6. We can now apply the theory of this chapter to that case.

For each Gaussian frame (i.e., each choice of a spacelike geodesic to serve as the basic hypersurface $\tau = 0$ of a

[32] Parker also used a different definition of particle observables -- see Sec. X.6 above.

Gaussian coordinate system (III.1.2)) we have, following Sec. X.5, a definition of particle operators a_k^τ at each time τ. (This is the theory that was adumbrated in Secs. V.3 and V.5.) The theorem proved at the end of Sec. X.5 shows that the Fock representations defined in this way at different times are unitarily equivalent. In more physical terms, if there are finitely many field quanta present at one time, there are finitely many at all times. Of course, the no-particle states at different times will not be identical: there will be a particle creation effect, to the extent that these a_k^τ-quanta can be identified with physical particles at time τ.

The obvious next question is whether the representations of the field built in this way on different Gaussian frames are unitarily equivalent. Also, one would like to know whether the representation corresponding to a given Gaussian frame is equivalent to the covariant representation of Tagirov _et al_. and Nachtmann (see Secs. V.6 and X.4). An affirmative answer to the second question is also an answer to the first, since equivalence is transitive. (There is, of course, only one covariant representation, independent of the particular Gaussian frame used in the explicit construction of it in Sec. V.6.) In this section this equivalence will be proved, and its implications will be discussed.

The Gaussian time coordinate, called x^0 in Sec. X.1, is τ, and $R(\tau) = \cosh \tau$ (see Eq. (III.1.4)). The coordinate t, therefore, is what was called α in Chapter V (according to Eqs.

(1.7) and (V.2.4a,5)). Denote the annihilation operators in the positive-frequency representation at $\tau = 0$ by a_p and those of the covariant representation by b_p. The expansion of the field in terms of the a_p is Eq. (V.3.7), where P_p and N_p are defined by Eqs. (V.3.3). That is, in the notation of Eqs. (5.7-8)

$$\Psi^0_p(\alpha) = [2\sqrt{q + p^2}]^{-1/2} P_p(\tau(\alpha)). \tag{9.1}$$

Note that

$$\frac{dO_p(0)}{d\alpha} = i \tag{9.2}$$

(when O_p is defined as in Chapter V). The covariant expansion of the field is Eq. (V.6.1), with $\Psi_p(\alpha) = T_p(\alpha)$ defined by Eq. (V.6.6). Eq. (V.6.11) shows that Ψ_p and Ψ^0_p approach each other in the limit of large $|p|$, which is certainly a necessary condition for unitary equivalence. Sufficiency depends on the speed of this convergence.

The inverse of a field expansion is given by Eq. (1.18) (with Eq. (1.20)). Combining this with the formulas cited above, one obtains

$$a_p = i[\overset{0}{\Psi}{}^*_p(0)\,\dot{T}_p(0) - \overset{.0}{\Psi}{}^*_p(0)\,T_p(0)]\,b_p$$

$$+ i[\overset{0}{\Psi}{}^*_p(0)\,\dot{T}^*_p(0) - \overset{.0}{\Psi}{}^*_p(0)\,T^*_p(0)]\,b^{\dagger}_{-p}, \quad (9.3)$$

where

$$T_p(0) = \frac{1}{2}\left|\frac{\Gamma(\frac{1}{2}(-\nu+|p|))}{\Gamma(1+\frac{1}{2}(\nu+|p|))}\right|, \qquad \dot{T}_p(0) = -\frac{i}{2}\,T_p(0)^{-1}, \quad (9.4)$$

$$\overset{0}{\Psi}{}^*_p(0) = \frac{1}{\sqrt{2}}[q + p^2]^{-1/4}, \qquad \overset{.0}{\Psi}{}^*_p(0) = \frac{i}{\sqrt{2}}[q + p^2]^{+1/4}. \quad (9.5)$$

That is, in notation analogous to that of Eq. (5.6),

$$\beta(p) = \frac{1}{2}\{\sqrt{2}[q + p^2]^{1/4}\,T_p(0) - \frac{1}{\sqrt{2}}[q + p^2]^{-1/4}\,T_p(0)^{-1}\}. \quad (9.6)$$

Now it follows from the asymptotic formula (6.1.47) of [N.B.S.] (p. 257) for the quotient of two gamma functions that

$$T_p(0) \sim \left|\left(\frac{|p|}{2}\right)^{-\nu-1}[1+O(|p|^{-2})]\right| = \left(\frac{|p|}{2}\right)^{-1/2}[1+O(|p|)^{-2})] \quad (9.7)$$

for large p. Combining this with the by now quite familiar Taylor expansion of the fourth root (cf. Secs. F.4 and X.5), one sees that the leading term in $\beta(p)$ is at most of order $|p|^{-2}$. Therefore certainly

$$\sum_{p=-\infty}^{\infty} |\beta(p)|^2 < \infty, \tag{9.8}$$

which proves the equivalence, according to the theorem of Sec. F.3.

The meaning of this theorem is that the two ways of quantizing the scalar field considered in Chapter V yield the same field, as an operator-valued distribution in an abstract Hilbert space. Only the physical interpretation of the states differs. General mathematical properties, such as those discussed in Chapter IV, which do not concern fixed-time observables can be established in either framework. For instance, we know now that this field satisfies the axioms proposed in Sec. IV.1, because this was proved in Sec. V.6 for the theory in its covariant guise. We also know, on the other hand, that the theory satisfies the spectral condition proposed at the end of Sec. IV.2, because the Hamiltonian has been put into the manifestly positive form (5.5) (with the divergent term discarded). (That $U(\tau_2, 0)$ of Eqs. (IV.2.1-2) exists in the present case follows from the two unitary equivalence theorems of Sec. X.5. It is clear from the construction in Sec. F.3 that U is differentiable in τ_2 provided that the operators of the Bogolubov automorphism are. The derivative of U must be H(0) (up to a constant), since H(0) exists as a self-adjoint operator and the formal calculation mentioned at the beginning of Sec. VII.2 shows that it has the correct commutation relations to generate the dynamics (cf. Eq. (IV.2.3)). In the present case, where the

Bogolubov transformation is diagonal, Eq. (IV.2.2) can be derived directly by approximating U by a unitary operator affecting only finitely many modes (see Reed (1968), pp. 13-14) and applying an "$\varepsilon/3$" argument.)

In Sec. III.3 it was argued that the most natural way to frame a dynamical problem in de Sitter space is not to compare the state of the world at various times τ in a Gaussian frame, but to look at the family of geodesic hypersurfaces orthogonal to a given timelike geodesic. These are the various times χ in a Fermi frame, or the instants $\tau = 0$ in a certain one-parameter family of <u>different</u> Gaussian frames. It was also suggested, however, that the state of the system at a particular instant should be characterized in terms of quantities defined in terms of the Gaussian frame corresponding to that time, rather than the (fixed) Fermi frame of the observer (see Secs. III.3 and IV.2). In the present case this means that the particle operators on a geodesic hypersurface should be defined in the manner of this chapter rather than in the way considered in Secs. V.7 and VIII.6, where positive frequency was defined with respect to Fermi coordinates. As remarked near the beginning of this section, the existence of a unitary propagator implementing this kind of dynamics is an immediate corollary of the theorem just proved. One can think of the no-particle state (as a function of the time, χ) as "precessing" around the vacuum vector of the covariant representation.

What has just been described is the specialization to two-dimensional de Sitter space of the general proposal of Sec. X.8. In the next section a calculation will be done in this framework.

10. Estimate of the Particle Creation in Two-Dimensional De Sitter Space.

In the spirit of Sec. X.8 let us define particle number observables on a geodesic hypersurface in terms of diagonalization of the Hamiltonian, H(0), defined with respect to the Gaussian coordinate system associated with that surface, and let us frame the dynamical problem in terms of the family of geodesic hypersurfaces orthogonal to a geodesic worldline. The operator of the (Fermi) time translation along the geodesic defined by $\sigma = 0$ is $\exp(i\chi H)$ -- where H, as in Secs. I.3, B.3, III.2, etc., is an element of the de Sitter Lie algebra, not to be confused with the Hamiltonian H(0). For our present purposes we might as well pass to a Schrödinger picture and study the operator $\exp(-i\chi H)$ as a mapping of the configuration of the system at time 0 into the configuration at time χ.

The calculation will be an approximation valid for large R, where R, as in the earlier chapters, is the radius of the de Sitter hyperboloid, not to be confused with the coordinate-dependent quantity R(t) used elsewhere in this chapter. As discussed in Sec. II.2, large R signifies two things:

(1) The natural time scale is $T = R\chi$. A reasonable

interval of T corresponds to a very small interval in the dimensionless "angular" coordinate χ. Thus one should be able to get away with expanding $\exp(-i\chi H)$ to first order.

(2) The mass parameter q (which takes the place of m^2 in the general formulas of this chapter -- cf. Sec. X.9) is large. The physical mass is $M = \sqrt{q}/R$. The assumption that $q \gg 1$ will be used in the calculation to simplify the expressions.

The transformation between the covariant particle operators b_p and the operators a_p of the positive-frequency representation (which we are tentatively accepting as the physical particle operators) is given in Eqs. (9.3-6). Here $\alpha(p)$ and $\beta(p)$ are real and even in p. The inverse transformation is

$$b_p = \alpha(p)a_p - \beta(p)a^\dagger_{-p} \qquad (10.1)$$

and the Hermitian conjugate of this equation. Tagirov <u>et</u> <u>al</u>. have given the expression for H in the covariant Fock space (see Eq. (V.6.15b)):

$$H = \frac{1}{2}(A^+ + A^-) =$$

$$\frac{1}{2}\sum_p \sqrt{q + p(p+1)}\; b^\dagger_{p+1} b_p + \frac{1}{2}\sum_p \sqrt{q + p(p-1)}\; b^\dagger_{p-1} b_p\,. \qquad (10.2)$$

We shall express H in terms of a_ρ and $a_\rho^\dagger$ by means of Eq. (10.1) and normal order the result. The normal ordering affects only the phase of $\exp(-i\chi H)$, which is physically arbitrary. This convention makes $H|0\rangle$ orthogonal to $|0\rangle$, where $|0\rangle$ is the initial no-particle state; the first-order calculation of the action of $\exp(-i\chi H)$ on $|0\rangle$ is thus optimized.

The result of the substitution, after some rearrangement, is

$$H = \frac{1}{2}\sum_{p\geq 0}\sqrt{q+p(p+1)}\,[\alpha(p+1)\alpha(p) + \beta(p+1)\beta(p)]$$

$$\times\{a^\dagger_{p+1}a_p + a^\dagger_{-(p+1)}a_{-p} + a^\dagger_p a_{p+1} + a^\dagger_{-p}a_{-(p+1)}\}$$

$$-\frac{1}{2}\sum_{p\geq 0}\sqrt{q+p(p+1)}\,[\alpha(p+1)\beta(p) + \alpha(p)\beta(p+1)]$$

$$\times\{a^\dagger_{p+1}a^\dagger_{-p} + a^\dagger_{-(p+1)}a^\dagger_p + a_{p+1}a_{-p} + a_{-(p+1)}a_p\}. \qquad (10.3)$$

Now α and β are to be read off from Eqs. (9.3-5). A considerable simplification occurs in the particular combinations which appear in Eq. (10.3) because (for $p \geq 0$)

$$T_p(0)\,T_{p+1}(0) = \frac{1}{4}\left|\frac{\Gamma(\frac{1}{2}(\nu+p+1))}{\Gamma(1+\frac{1}{2}(\nu+p+1))}\right|$$

$$= \frac{1}{2}\,|\nu + p + 1|^{-1} = \frac{1}{2}\,\frac{1}{\sqrt{q + p(p+1)}} \qquad (10.4)$$

(see Eq. (V.6.8)). One obtains

$$\alpha(p+1)\alpha(p) + \beta(p+1)\beta(p) =$$

$$\frac{1}{2}\left\{\left[\frac{\sqrt{q+(p+1)^2}\,\sqrt{q+p^2}}{q+p(p+1)}\right]^{1/2} + \left[\frac{q+p(p+1)}{\sqrt{q+(p+1)^2}\,\sqrt{q+p^2}}\right]^{1/2}\right\}, \qquad (10.5a)$$

$$\alpha(p+1)\beta(p) + \alpha(p)\beta(p+1) =$$

$$\frac{1}{2}\left\{\left[\frac{\sqrt{q+(p+1)^2}\,\sqrt{q+p^2}}{q+p(p+1)}\right]^{1/2} - \left[\frac{q+p(p+1)}{\sqrt{q+(p+1)^2}\,\sqrt{q+p^2}}\right]^{1/2}\right\}. \qquad (10.5b)$$

Let

$$Q = q + p^2. \qquad (10.6)$$

Since $q \gg 1$, we may assume that p and 1 are small compared to Q. We expand everything in sight in Taylor series in $1/Q$. The result is, through second order,

$$\left[\frac{\sqrt{q+(p+1)^2}\,\sqrt{q+p^2}}{q+p(p+1)}\right]^{1/2} = 1 + \frac{1}{4Q} - \frac{1}{4Q^2}\left(p^2 + 2p + \frac{3}{8}\right), \qquad (10.7a)$$

square

$$\left[\frac{q+p(p+1)}{\sqrt{q+(p+1)^2}\,\sqrt{q+p^2}}\right]^{1/2} = 1 - \frac{1}{4Q} + \frac{1}{4Q^2}\left(p^2 + 2p + \frac{5}{8}\right). \qquad (10.7b)$$

We are now ready to compute the action of $\exp(-i\chi H) = 1 - i\chi H + O(\chi^2)$ on the initial no-particle state. The first-order term is

$$\Psi_1 = -i\chi H|0\rangle = +\frac{i\chi}{2}\sum_{p\geq 0}\sqrt{q+p(p+1)}\,[\alpha(p+1)\beta(p) + \alpha(p)\beta(p+1)]$$

$$\times\,[a^\dagger_{p+1}a^\dagger_{-p}|0\rangle + a^\dagger_{-(p+1)}a^\dagger_{p}|0\rangle]$$

$$\sim \frac{i\chi}{8}\sum_{p\geq 0} Q^{-1/2}\,[a^\dagger_{p+1}a^\dagger_{-p}|0\rangle + a^\dagger_{-(p+1)}a^\dagger_{p}|0\rangle].$$

The basis vectors which appear in the sum (inside the second pair of brackets) are all orthogonal and normalized. Hence we have

$$\|\Psi_1\|^2 = \frac{\chi^2}{32}\sum_{p\geq 0}[q+p^2]^{-1} \approx \frac{\pi}{64}\frac{\chi^2}{\sqrt{q}}. \qquad (10.8)$$

In the last step the sum has been approximated by

$$\int_0^\infty dp\,[q+p^2]^{-1} = \frac{\pi}{2\sqrt{q}}.$$

The number given in Eq. (10.8) is the probability (to lowest order in both χ and $1/q$) that the state of the field is other than vacuum at time χ if it was vacuum at time 0. Introducing the physical units of time and mass, we have

$$\|\Psi_1\|^2 \approx \frac{1}{20}\,\frac{T^2}{R^3 M}. \qquad (10.9)$$

(As always, $\hbar = c = 1$.) Let us choose R to be a typical cosmological distance, 10^{27} cm, and M to be 10^{13} cm^{-1}, a typical elementary particle inverse Compton wavelength. Then

$$\|\Psi_1\|^2 = T^2 \times 10^{-95}\ \mathrm{cm}^{-2} = T^2 \times (10^{-59}\ \mathrm{yr}^{-2}). \qquad (10.10)$$

So it appears that 10^{28} years is needed before there is a significant probability (10^{-3}) of "decay" of the vacuum. (We shall find that this is just an upper bound.)

One should question for what range of T this calculation is valid. An attempt to estimate the second-order term leads to an infinite result, because $|0\rangle$ is not in the domain of H^2. However, inspection shows that the divergence comes entirely from the number-preserving term of H (the first term in Eq. (10.3)). The number-changing term is bounded as an operator on each n-particle subspace of the Fock space. This suggests that a better calculation of the creation probability would proceed by standard time-dependent perturbation theory (e.g., [Messiah], pp. 722-739), with H_0, the first term of Eq. (10.3), as unperturbed Hamiltonian and V, the second term of Eq.

(10.3), as the perturbation. (The previous calculation, in effect, treated all of H as a perturbation.)

The initial state $|0\rangle$ is an eigenstate of H_0 with eigenvalue 0. The transition probability from $|0\rangle$ to the two-particle space is ($\hbar = 1$)

$$W = \int |V_{b0}|^2 \, f(\chi, E_b) \, d\mu(b), \tag{10.11}$$

where

$$f(\chi, E_b) = 2(1 - \cos E_b \chi)/E_b^2, \tag{10.12}$$

$$V_{b0} = \langle b|V|0\rangle, \tag{10.13}$$

and

$$H_0 |b\rangle = E_b |b\rangle \tag{10.14}$$

for each two-particle state $|b\rangle$. The usual "golden rule" approximation is[33]

[33] Unlike Eqs. (10.8-10), Eq. (10.15) exhibits the familiar linear time dependence of a first-order transition probability. The quadratic time dependence of our earlier result is due to the use of an "unperturbed Hamiltonian" which is identically zero. Since the spectrum was entirely degenerate, the narrowing of the peak of the function (10.12) did not have its usual effect.

$$W = 2\pi \int |V_{b0}|^2 \, d\mu_0(b) \; \chi, \tag{10.15}$$

where the integration is now over the states with $E_b = 0$. (The "density of states" is absorbed into the measures μ and μ_0.)

Since the maximum value of the function (10.12) is χ^2, an upper bound on W of Eq. (10.11) is

$$\chi^2 \int |V_{b0}|^2 \, d\mu(b) = \chi^2 \, \| V|0\rangle \|^2 = \| \Psi_1 \|^2 ,$$

the result previously obtained (Eq. (10.8)). Our only concern, therefore, is whether the exact creation rate is much smaller than indicated by Eq. (10.10).

Inspection of Eqs. (10.3,5,7) shows that H_0 consists of (1) a term H* which, in terms of the a_p, is identical in form to H, in terms of the b_p (Eq. (10.2)), and (2) a residual term, H_0 - H*, which decreases even faster in q than V does. The second term can be neglected. The spectrum of H* in the two-particle space is the spectrum of a noncompact generator of $SO_0(1,2)$ in the direct product of two irreducible representations of the group of the principal series. Therefore, if the $|b\rangle$ are properly normalized,

$$d\mu(b) = \sum \int_{-\infty}^{\infty} dh_1 \int_{-\infty}^{\infty} dh_2 , \tag{10.16}$$

where h_1 and h_2 represent the spectrum of the generator in the factor representations, and $\sum$ indicates a sum over a discrete variable which assumes four values, accounting for the multiplicity of these spectra (see Sec. II.3). In Eq. (10.15) we are to hold $h_1 + h_2 = E_b$ fixed; hence

$$d\mu_0(b) = \frac{1}{2} \sum \int_{-\infty}^{\infty} d(h_1 - h_2). \qquad (10.17)$$

To evaluate W we need the matrix elements, with respect to the generalized eigenfunctions of H*, of V, which is expressed in terms of the eigenbasis of the compact generator of $SO_0(1,2)$ (with spectrum p). The expansion with respect to the p-basis of the eigenfunctions of a noncompact generator within an irreducible representation is available (Lindblad and Nagel (1970), Eq. (4.13)). The direct-product vectors can be constructed from these. In principle, therefore, the desired matrix elements V_{b0} can be calculated from Eq. (10.3). However, quite a bit of work, analytic or numerical, would be needed to extract a number. Our two-dimensional model, which has no direct connection with experiment, does not warrant such treatment; its purpose is just to show that particle creation can be calculated in principle and is not unreasonably large in the theory proposed in the last two sections.

Instead of an exact evaluation of the expression (10.15), then, let us stop with a rough estimate. We note from the previous calculation that $\|\Psi_1\|^2$ decreases as $1/\sqrt{q}$ for large q, while the individual matrix elements in the p-representation go down as 1/q. Let us assume that

$$\int |V_{b0}|^2 \, d\mu_0(b) \sim q^{-\sigma/2}, \qquad 1 \leq \sigma \leq 2. \qquad (10.18)$$

Then by dimensional analysis we may write down for comparison with Eq. (10.9)

$$W \sim \frac{T}{R}(RM)^{-\sigma} = T \times 10^{-40\sigma} \times (10^{-9}\ \text{yr}^{-1}). \qquad (10.19)$$

Recall that the quadratic estimate (10.9) is better than the above for small T, when it is the smaller. If $\sigma = 1$, we find that Eq. (10.9) is good up to $T = 10^{10}$ yr, and W reaches 10^{-3} at $T = 10^{46}$ yr. If $\sigma = 2$, the estimate (10.9) is already too large for $T > 10^{-23}$ sec, and $W = 10^{-3}$ only when $T = 10^{86}$ yr.

What about particle creation (or annihilation) when matter is present initially? At one extreme is the vacuum of the covariant theory, which is invariant under $\exp(-i\chi H)$ -- hence all expectation values are constant. On the other hand, consider a state containing exactly N particles. From Eq. (10.3) and the formulas (F.1.4) for the action of creation and annihilation operators in Fock space one can put a crude upper bound on its decay probability of N^2 times the vacuum value. If the estimate (10.10) is used for the latter, which, we have seen, is probably much too liberal, a cosmologically reasonable density of one particle per centimeter would raise the probability of a transition to a state with N + 1 or N - 1 particles to $T^2 \times (10^{-5}\ \text{yr}^{-2})$ at most. For any realistic matter distribution the exact probability is probably much smaller. (Also, of course, states differing by only one particle in the entire universe will not be very distinguishable.)

Although a number derived from a two-dimensional model should not be taken too seriously, these estimates increase one's confidence that our reasoning, based on the formalism of quantum field theory, will not predict absurdly large creation rates. In fact, the creation effect is so negligible that we seem to be still in the domain where single-particle quantum mechanics is quite adequate for any practical purpose (cf. Sec. X.7). It must be remembered, however, that the theorem of Sec. X.9, on which this calculation hinges, may not hold for the four-dimensional case.

11. Summary of Chapter X.

The major points established in this chapter are:

(1) In a rigidly expanding (i.e., generalized Robertson-Walker) universe there is a natural decomposition of the field into modes, but the lack of an obvious splitting into positive- and negative-frequency parts means that the associated "Fock" representation is not unique.

(2) A region of Minkowski space can be cast into the Robertson-Walker form in such a way that none of these representations (tensor products of the modes) coincides with the standard quantization of the free field.

(3) Examination of the asymptotically static case suggests that a different representation is appropriate at each time, and that it should be chosen on physical grounds (e.g., particle interpretation).

(4) Attempts to give an abstract time-independent definition of the representation will clash with the obvious physical interpretation of the asymptotically static situation.

(5) The most natural generalization from the static case is to define a Fock representation at each time by diagonalizing the Hamiltonian into a linear combination of number operators. The normal-ordered Hamiltonian is then a positive self-adjoint operator. This procedure can be extended to the general case, where the field equation is not solvable by separation of variables. It is this prescription which is assumed in the following points.

(6) In general the representations at different times will not be unitarily equivalent. In particular, this is true of generalized Robertson-Walker universes of infinite spatial extent.

(7) For an expanding universe there are two distinct notions of unitary implementability of the dynamics. These are coextensive for dimension 2 but completely incompatible

for higher dimensions, except when the radius of the universe is constant.

(8) In a two-dimensional expanding universe of finite circumference, the dynamics is unitarily implementable. (I.e., the representations at different times are equivalent.)

(9) A completely satisfactory approach to field theory in curved space-time may require abandoning the particle concept. Also, the canonical formalism on hypersurfaces may be inappropriate.

(10) The ambiguities encountered here are instances of a difficulty which afflicts external potential problems in general, when the potentials do not fall off sufficiently fast to allow an unambiguous asymptotic particle interpretation.

(11) It is proposed that representations defined as in (5) are more likely to be physically appropriate if the hypersurfaces involved are geodesic.

(12) In two-dimensional closed de Sitter space the representation specified by (5) and (11) is unitarily equivalent to the one which is covariant under the de Sitter group. The particle interpretations, however, are different. This model satisfies both the "group" axioms

and the spectral condition formulated in Chapter IV.

(13) The particle creation in the two-dimensional de Sitter model is extremely small.

Appendix A

THE PSEUDO-ORTHOGONAL GROUPS

Much of this material is taken from Bargmann (1947), especially pp. 585-586, and Bargmann (1954), especially pp. 34-36.

1. Homogeneous Groups.

Consider n-dimensional real space R^n with a nondegenerate but possibly indefinite scalar product given by a (constant) metric tensor g_{jk} $(0 \leq j,k < n)$. We use the convention that an index is to be summed over when it appears once in contravariant and once in covariant position, and we use g_{jk} and its inverse g^{jk} to lower and raise indices in the standard way. We may choose a basis for the vector space (orthonormal basis) with respect to which for some index p the metric has the form

$$g_{jj} \equiv \eta_j = +1 \quad \text{if } 0 \leq j < p,$$

$$g_{jj} \equiv \eta_j = -1 \quad \text{if } p \leq j < n, \tag{1.1}$$

$$g_{jk} = 0 \quad \text{if } j \neq k.$$

All orthonormal bases yield the same p. We set $q = n - p$.

The pseudo-orthogonal group $O(p,q)$ is defined as the group of linear transformations which leave invariant the quadratic form

$$F(x) = g_{jk} x^j x^k . \tag{1.2}$$

The connected component of this group containing the identity is denoted by $SO_0(p,q)$. Clearly $SO_0(p,q)$ is isomorphic to $SO_0(q,p)$. $SO_0(p,0)$ is $SO(p)$, the p-dimensional rotation group. We will be particularly concerned with the group $SO_0(1,q)$, which we call the q-dimensional (closed) de Sitter group or the (q+1)-dimensional (homogeneous) Lorentz group, depending on the physical context in which it is considered.

In a particular basis a transformation $A \in SO_0(p,q)$ is represented by a matrix: if $y = Ax$, then

$$y^j \equiv (Ax)^j = A^j_{\ k} x^k . \tag{1.3}$$

The condition $F(y) = F(x)$ yields

$$A_k^{\ j} A^k_{\ l} = \delta^j_l \qquad (\text{where } A_k^{\ j} = g_{mk} A^m_{\ n} g^{nj}). \tag{1.4}$$

Let $\mathcal{L}(SO_0(p,q))$ be the Lie algebra of $SO_0(p,q)$. An element of $SO_0(p,q)$ close to the identity can be written to first order as

$$A = 1 + tL \tag{1.5}$$

where $L \in \mathcal{L}(SO_o(p,q))$. Then the condition (1.4) is equivalent to

$$(L)_{jl} + (L)_{lj} = 0. \tag{1.6}$$

The matrices satisfying Eq. (1.6) are precisely the linear combinations of the matrices L_{ab} defined by

$$(L_{ab})^j{}_k = \delta^j{}_a g_{bk} - \delta^j{}_b g_{ak} . \tag{1.7}$$

Since

$$L_{ba} = - L_{ab} \quad \text{(and hence } L_{aa} = 0), \tag{1.8}$$

only $n(n-1)/2$ of the L_{ab} are linearly independent. The L_{ab} with $a < b$ form a basis for $\mathcal{L}(SO_o(p,q))$. Therefore, $SO_o(p,q)$ is a connected Lie group of dimension $n(n-1)/2$.

The commutators of the L_{ab} are

$$[L_{ab}, L_{cd}] = g_{bc} L_{ad} - g_{ac} L_{bd} + g_{ad} L_{bc} - g_{bd} L_{ac} . \tag{1.9}$$

This is easily proved by working in the defining representation (1.7). Eq. (1.9) becomes more transparent when an orthonormal coordinate system is used and special cases are considered. First note that we may assume $a \neq b$ and $c \neq d$, since otherwise one of the L's is zero. Also, if the pair (a,b) is equal to the pair (c,d) in either order, the commutator is trivially zero. If a, b, c, d are all distinct, Eq. (1.9) tells us that

$$[L_{ab}, L_{cd}] = 0, \tag{1.10a}$$

since the metric tensor is diagonal. The remaining cases have exactly one index of the first pair equal to one index of the second pair. Using Eq. (1.8), we may assume without loss of generality that a = d (and the indices are otherwise distinct). Then

$$[L_{ab}, L_{ca}] = \eta_a L_{bc}. \tag{1.10b}$$

If $\eta_a = \eta_b$, the one-parameter subgroup generated by L_{ab},

$$A(t) = \exp(tL_{ab}), \tag{1.11}$$

is of the "rotation" type, with matrices of the form

$$\begin{matrix} a \to \\ b \to \\ j \neq a, b \to \end{matrix} \begin{pmatrix} \cos t & \eta_b \sin t & 0 \\ -\eta_b \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.12a}$$

If $\eta_a = -\eta_b$, it is of the "boost" type,

$$\begin{pmatrix} \cosh t & \eta_b \sinh t & 0 \\ \eta_b \sinh t & \cosh t & 0 \\ 0 & 0 & 1 \end{pmatrix}. \qquad (1.12b)$$

Some additional general information about the groups $SO_o(p,q)$ and their unitary representations can be found in Kihlberg (1965).

2. Inhomogeneous Groups.

Let G be a group defined by an n-dimensional real representation (such as $SO_o(p,n-p)$, Eqs. (1.3-4)). The corresponding inhomogeneous group, which we call IG, is the semidirect product of G with the additive group of the n-dimensional space (the translation group). That is, an element of IG is a pair (b,A) $(b \in \mathbb{R}^n,\ A \in G)$, and the defining realization is

$$y^\mu \equiv ((b,A)x)^\mu = (Ax + b)^\mu \equiv b^\mu + A^\mu{}_\nu x^\nu . \qquad (2.1)$$

It follows that the group product is

$$(b,A)(b',A') = (b + Ab',\ AA') \qquad ((Ab')^\mu = A^\mu{}_\nu b'^\nu) \qquad (2.2)$$

and the inverse is

$$(b,A)^{-1} = (-A^{-1}b,\ A^{-1}). \tag{2.3}$$

If $G = SO_0(1,n-1)$ (the n-dimensional Lorentz group), then IG is the n-<u>dimensional Poincaré group</u>.

The realization (2.1) is not a linear representation since the translations are not (homogeneous) linear transformations. However, it is entirely equivalent to the following (n+1)-dimensional representation. Identify $\mathbb{R}^n$ with the hyperplane $x^n = 1$ in $\mathbb{R}^{n+1}$. (Recall our convention that indices in $\mathbb{R}^n$ run from 0 to n - 1. In what follows Greek indices will be understood as running from 0 to n - 1, Latin indices from 0 to n.) Then the matrices $*A$ defined by

$$\begin{aligned} *A^{\mu}{}_{\nu} &= A^{\mu}{}_{\nu}, \qquad & *A^{n}{}_{n} &= 1, \\ *A^{\mu}{}_{n} &= b^{\mu}, \qquad & *A^{n}{}_{\nu} &= 0, \end{aligned} \tag{2.4}$$

are a representation of IG, and they map the hyperplane $x^n = 1$ into itself in accordance with Eq. (2.1):

$$\begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} Ax + b \\ 1 \end{pmatrix}. \tag{2.5}$$

It is easy to see that the elements of $\mathcal{L}(IG)$ are represented by matrices of the form

$$L = \begin{pmatrix} L' & b \\ 0 & 0 \end{pmatrix} \qquad (L' \in \mathcal{L}(G),\ b \in \mathbb{R}^n). \qquad (2.6)$$

As a basis for $\mathcal{L}(IG)$ we can take a basis for $\mathcal{L}(G)$ (represented by matrices of the form (2.6) with b = 0) and add n basis elements T_α $(0 \le \alpha < n)$ defined relative to some basis in $\mathbb{R}^n$ by

$$(T_\alpha)^j{}_\mu = 0 = (T_\alpha)^n{}_n,$$

$$(T_\alpha)^\mu{}_n = \delta^\mu_\alpha. \qquad (2.7)$$

The T_α commute among themselves:

$$[T_\alpha, T_\beta] = 0. \qquad (2.8)$$

In $SO_0(p,n-p)$ the other commutators can be expressed in any of the equivalent forms

$$[L, T_\alpha] = (L)^\beta{}_\alpha T_\beta \qquad (L \in \mathcal{L}(G)); \qquad (2.9)$$

$$[L_{\alpha\beta}, T_\gamma] = g_{\beta\gamma} T_\alpha - g_{\alpha\gamma} T_\beta; \qquad (2.10)$$

or

$$[L_{\alpha\beta}, T_\gamma] = 0 \qquad (\gamma \neq \alpha, \beta), \tag{2.11a}$$

$$[L_{\alpha\beta}, T_\beta] = \eta_\beta T_\alpha . \tag{2.11b}$$

(Eqs. (2.11), like Eqs. (1.10), assume an orthonormal basis.)

3. <u>Representation of the Lie Algebras by Differential Operators</u>.

Let M be a homogeneous space under a group G. Consider the vector space $\mathcal{D}(M)$ of complex-valued C^∞ functions of compact support on M. The <u>quasiregular representation</u> of G is defined in $\mathcal{D}(M)$ as follows: If $A \in G$, its representation $U(A)$ is

$$[U(A)\Psi](x) = \Psi(A^{-1}x) \qquad (\Psi \in \mathcal{D}(M),\ x \in M). \tag{3.1}$$

Pictorially speaking, $U(A)$ moves the function around bodily in M; the value which once was assigned to $x \in M$ is now attached to Ax.

Let $L \in \mathcal{L}(G)$ have the matrix $(L)^j{}_k$ with respect to some basis. Then the quasiregular representation of L is

$$U(L) = - (L)^j{}_k \, x^k \frac{\partial}{\partial x^j}, \tag{3.2}$$

as can be seen by expanding Eq. (3.1) to first order in the parameter of the subgroup generated by L.

In the case of $SO_o(p,q)$ (with an orthonormal basis) we obtain from Eq. (1.7)

$$U(L_{ab}) = x_a \frac{\partial}{\partial x^b} - x_b \frac{\partial}{\partial x^a}. \qquad (3.3)$$

(Note that the indices have been lowered on the factors x_j.) For $ISO_0(p,q-1)$ we have (in addition to generators of the form (3.3)) the translation generators (see Eq. (2.7))

$$U(T_\alpha) = - x^n \frac{\partial}{\partial x^\alpha} \quad (= - \frac{\partial}{\partial x^\alpha} \text{ on the hyperplane } x^n = 1). \qquad (3.4)$$

According to Eq. (I.3.3b), this operator must be the contraction of

$$U(\frac{L_{n\alpha}}{R}) = - \bar{x}^n \frac{\partial}{\partial \bar{x}^\alpha} - \frac{\bar{x}_\alpha}{R} \frac{\partial}{\partial \bar{x}^n}. \qquad (3.5)$$

The connection is evident.

As coordinates in de Sitter space (Eq. (I.1.1)) the x^a are not independent. Nevertheless, the transformations generated by the operators (3.3) map the space into itself, since $U(L_{ab})\ F(x) = 0$.

In the notation of Sec. I.3, Eqs. (3.3) for the de Sitter group $SO_0(1,n)$ become (U omitted)

$$H = i\ (x^n \frac{\partial}{\partial x^0} + x^0 \frac{\partial}{\partial x^n}) \qquad [i \frac{\partial}{\partial x^0}], \qquad (3.6a)$$

$$P^A = - i\ (x^n \frac{\partial}{\partial x^A} - x^A \frac{\partial}{\partial x^n}) \qquad [- i \frac{\partial}{\partial x^A}], \qquad (3.6b)$$

$$K^A = i\left(x^A \frac{\partial}{\partial x^0} + x^0 \frac{\partial}{\partial x^A}\right) \quad (= x^A H - x^0 P^A), \tag{3.6c}$$

$$J^A = i \sum_{B,C} \epsilon_{ABC}\, x^C \frac{\partial}{\partial x^B} \quad (= \sum_{B,C} \epsilon_{ABC}\, x^B P^C). \tag{3.6d}$$

The operators in brackets in Eqs. (3.6a,b) are the corresponding generators of the associated Poincaré group $ISO_0(1,n-1)$. It is easy to verify that all these operators have the correct commutation relations.

In Chapter III we express these differential operators in terms of coordinate charts on the manifold M. Then the Casimir operator Q (Eq. (I.4.2)) becomes a second-order differential operator which defines a scalar wave equation on M.

Appendix B

IRREDUCIBLE UNITARY RAY REPRESENTATIONS OF THE DE SITTER AND POINCARE GROUPS

1. Ray Representations.

In quantum theory a group of symmetries of a physical system corresponds, in general, to a unitary *ray* representation in the Hilbert space of state vectors. That is, every element g of the group G is implemented by a unitary operator $U(g)$, and

$$U(g_1)\, U(g_2) = \omega(g_1, g_2)\, U(g_1 g_2), \tag{1.1}$$

where $\omega(g_1, g_2)$ is a complex number of modulus 1. Bargmann (1954) showed that for connected Lie groups the classification of the ray representations of G can be reduced to the study of the true representations ($\omega \equiv 1$) of G and some groups related to it (namely, its universal covering group G^* and the nontrivial one-parameter central extensions of G^*, if any). In the same paper Bargmann applied his method to the homogeneous and inhomogeneous pseudo-orthogonal groups. His findings are summarized here.

(1) Every factor ω for $SO_0(p,q)$ with $p + q \geq 2$ and every factor ω for $ISO_0(p,q)$ with $p + q \geq 3$ is equivalent to unity. It follows that every ray representation of one of

these groups is equivalent to a true unitary representation of its universal covering group (which we call an "ordinary representation").

(2) In the case $ISO_o(p,q)$, $p + q = 2$, the set of equivalence classes of factors has dimension 1. Hence these groups possess ray representations which are not equivalent to ordinary representations.

(3) The kernel of the homomorphism of the universal covering group of $SO_o(p,q)$ or of $ISO_o(p,q)$ onto the group itself is a direct product $C(p) \otimes C(q)$, where $C(p)$ and $C(q)$ are cyclic groups. $C(0)$ and $C(1)$ are groups of order one, $C(2)$ is infinite, and $C(p)$ is of order two if $p \geq 3$. Therefore, in particular, the ray representations of $SO_o(1,n-1)$ (and $ISO_o(1,n-1)$ if $n \geq 3$) are single-valued or double-valued representations if $n \geq 4$, are single-valued if $n = 2$, and may be many-valued if $n = 3$.

In the physical context we are considering, "dimension n" refers to the de Sitter group $SO_o(1,n)$ and the Poincaré group $ISO_o(1,n-1)$. So the conclusions of the analysis are: If $n \geq 4$, the only ray representations of either group other than the true representations are the familiar double-valued spinor representations. If $n = 3$, the de Sitter group again has only true representations and spinor representations, but the Poincaré group has many-valued representations. If $n = 2$, the de Sitter group has many-valued representations; the Poincaré group has no many-valued ordinary representations (not even double-valued ones) since it is its own covering group, but in view of (2) it has ray representations which are not equivalent to ordinary representations. For the most part we shall consider only the true (single-valued) representations in the case n = 2 and only the single- and double-valued representations for n = 3.

2. Representations of the Poincaré Groups.

The single- and double-valued irreducible unitary representations of the four-dimensional Poincaré group are well known, as is the Frobenius-Wigner "little group" method of deriving them (see Wigner (1939), Wightman (1959, 1960)). The same method works for any $ISO_o(p,q)$. Let us summarize the results for $ISO_o(1,n-1)$.

The space $\mathbb{R}^n$ (regarded as its own dual, momentum space) is divided into "orbits" or homogeneous spaces (see Sec. I.1) under the action of $SO_o(1,n-1)$. Typical of the conditions defining an orbit is

$$p^{\mu}p_{\mu} = m^2, \qquad p^0 > 0 \qquad\qquad (p \in \mathbb{R}^n). \quad (2.1)$$

Let p_o be an arbitrarily chosen point on an orbit M. The "little group" is the subgroup of $SO_o(1,n-1)$ which leaves p_o invariant. An irreducible representation is defined in a Hilbert space of functions on M with values in a "little" Hilbert space (spin space) which supports an irreducible unitary representation (or spinor representation) $Q(p_o,A)$ of the little group. The scalar product is

$$(\Psi,\phi) = \int_M d\Omega\ (\Psi(p),\phi(p)), \qquad (2.2a)$$

where $(\Psi(p),\phi(p))$ is the scalar product in spin space and Ω is an invariant measure on M. For the orbit (2.1) we have

$$(\Psi,\phi) = \int d^n p\ \delta(p^\mu p_\mu - m^2)\ \theta(p^0)\ (\Psi(p),\phi(p))$$

$$= \int \frac{d^{n-1}\vec{p}}{2\sqrt{\vec{p}^2 + m^2}}\ (\Psi(\vec{p}),\phi(\vec{p})); \qquad (2.2b)$$

in the last form Ψ and ϕ are regarded as functions of the independent variables $\vec{p} = (p^1,\dots,p^{n-1})$. The representation is given by the formula

$$[U(b,A)\Psi](p) = e^{ip_\mu b^\mu}\ Q(p_0,C^{-1}(p)AC(A^{-1}p))\ [\Psi(A^{-1}p)] \qquad (2.3)$$

for each element (b,A) of $ISO_o(1,n-1)$* (see Sec. A.2). Here C(q) is a canonically chosen transformation in $SO_o(1,n-1)$ which maps p_o into q.

In the case n = 2 there are nine classes of irreducible representations corresponding to the nine orbits of two-dimensional Minkowski space (Fig. 18) (cf. Dubin (1970)). The representations associated with orbits of the type labeled 1

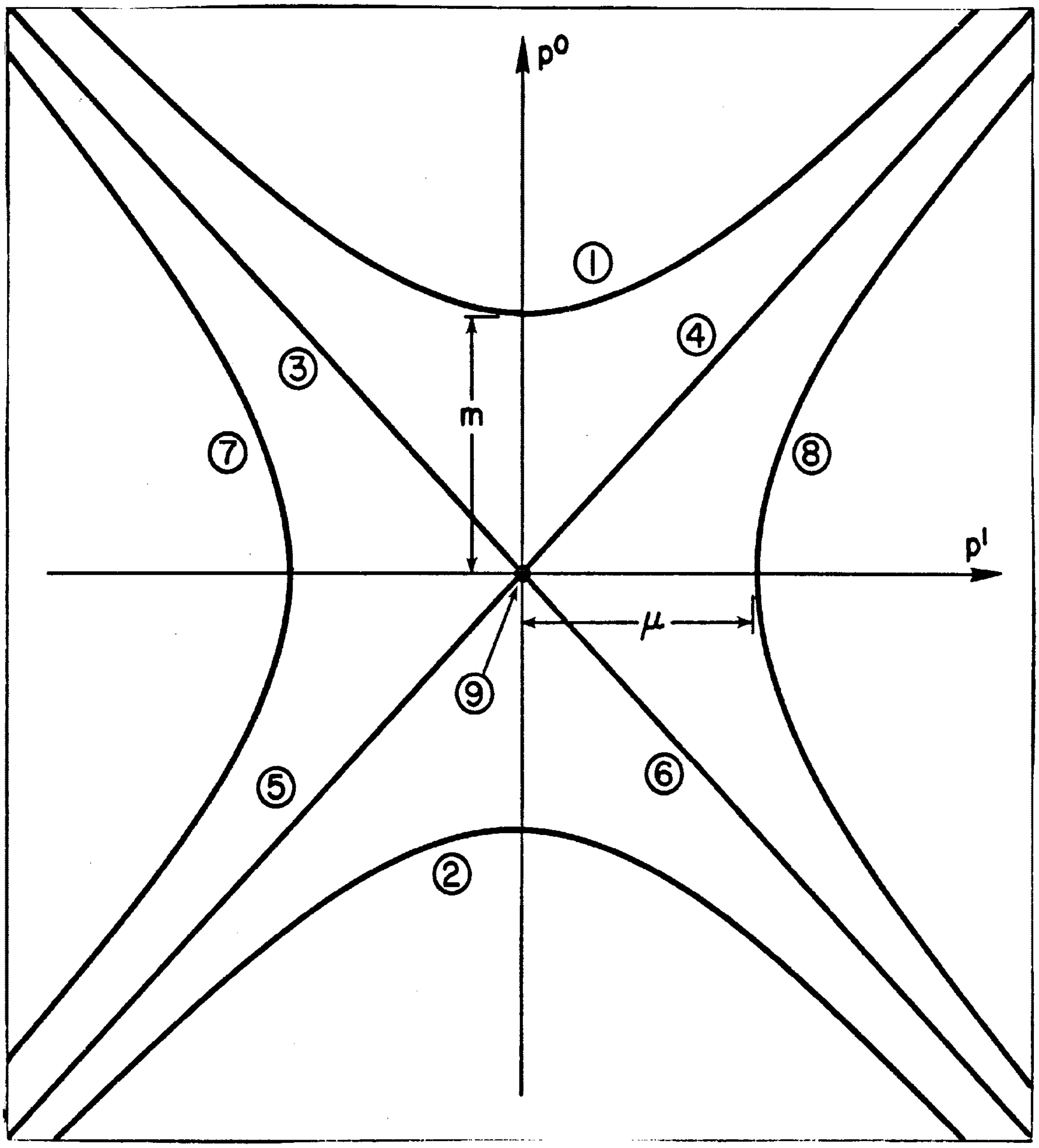


Fig. 18

Orbits of the Lorentz group in Minkowski space. In dimension greater than 2 the pairs of orbits 3-4, 5-6, and 7-8 are connected.

in Fig. 18 (to which Eqs. (2.1) and (2.2b) apply) are analogues of the positive-mass positive-energy representations of the four-dimensional Poincaré group; the parameter m defining them is the mass. Note that the representation on the positive-energy part of the light cone splits into two irreducible parts (3 and 4) because of topological peculiarities of the two-dimensional space; the spacelike momenta (7 and 8) are likewise split.

For each $m > 0$ there is one representation of class 1. It has the form

$$U(b,\Lambda)\Psi(p) = e^{ib^{\mu}p_{\mu}}\,\Psi(\Lambda^{-1}p), \qquad (2.4)$$

where Ψ is an L^2 function on the hyperbola of Eq. (2.1). There is no little-group representation involved since the little group of a vector on the hyperbola is the group with one element, which has only the trivial representation. The generators of the representation (2.4) are (cf. Eqs. (I.3.6) and (A.1.12b))

$$H\Psi(p) = -\,i\frac{\partial}{\partial b^{0}}\,[U(b,1)\Psi]_{b=0}(p) = +\,p^{0}\,\Psi(p), \qquad (2.5a)$$

$$P\Psi(p) = +\,i\frac{\partial}{\partial b^{1}}\,[U(b,1)\Psi]_{b=0}(p) = +\,p^{1}\,\Psi(p), \qquad (2.5b)$$

$$K\psi(p) =$$

$$- i\frac{\partial}{\partial\theta}\psi(\cosh\theta\, p^0 - \sinh\theta\, p^1, - \sinh\theta\, p^0 + \cosh\theta\, p^1)\Big|_{\theta=0}$$

$$= i\{p^1 \frac{\partial\psi}{\partial p^0} + p^0 \frac{\partial\psi}{\partial p^1}\}. \qquad (2.5c)$$

If $p^0 = \sqrt{(p^1)^2 + m^2}$ is regarded as a function of p^1, Eq. (2.5c) becomes

$$K\psi(p) = ip^0 \frac{d\psi}{dp^1}. \qquad (2.6)$$

In higher dimensions there are only six types of orbits. Fig. 18 may be viewed as a plane cross-section of the momentum space. Then sets 3 and 4 belong to the same orbit; likewise 5 and 6, and 7 and 8.

When n = 3 the little group is SO(2) for timelike momenta and is isomorphic to the real line in the spacelike and lightlike cases. The irreducible representations of these groups are one-dimensional (but not trivial). In general the little group is SO(n-1) in the timelike case, $SO_o(n-2,1)$ in the spacelike case, and the common contraction of these groups, ISO(n-2), on the light cone.

3. Representations of the Two-Dimensional De Sitter Group.

In this and the next two sections we record the irreducible self-adjoint[1] representations of the Lie algebras of the groups $SO_0(1,n)$ which correspond to unitary representations of the groups or their covering groups. By "unitary representation" we always mean a (weakly) continuous representation; then an associated self-adjoint representation of the Lie algebra exists and can be found by differentiating the one-parameter subgroups (see, e.g., Bargmann (1947), pp. 598-600). The converse requires an additional condition: Nelson (1959) proved that a Hermitian representation of the elements of a basis for the Lie algebra[2] corresponds to a unitary representation of the covering group if and only if the sum of the squares of the basis elements is a densely defined operator with a unique self-adjoint extension. Now the representations of $\mathcal{L}(SO_0(1,n))$ are usually constructed in a form in which the representation of the maximal compact subgroup $SO(n)$ is explicitly decomposed into irreducibles (see Eqs. (3.2) and (4.3) below). Nelson's operator (for the usual choice of basis in the Lie algebra) is the sum of the Casimir operator (I.4.2), which is a multiple of the identity in an irreducible representation, and twice the Casimir operator of the compact subgroup, which is diagonalized and therefore manifestly self-adjoint on an

---

[1] We use the convention of Sec. I.3: a one-parameter subgroup is $\exp(itL)$, $L \in \mathcal{L}(G)$.
[2] That is, a set of Hermitian operators with the proper commutation relations (e.g., Eqs. (3.1)).

appropriate domain. Thus there is no problem in showing that the Hermitian representations of $\mathcal{L}(SO_0(1,n))$ actually generate group representations.

The representations of $SO_0(1,2)$ were found by Bargmann (1947) (see especially pp. 598-609). The commutation relations are (Eqs. (I.3.10f,g,h)))

$$[K,H] = iP, \qquad [K,P] = iH, \qquad [P,H] = iK. \tag{3.1}$$

Eqs. (3.1) correspond to the usual parametrization of $SO_0(1,2)$, according to which $\exp(2\pi iP) = 1$. For a single-valued representation, therefore, the spectrum of P must consist of integers. The irreducible representations are found by a method parallel to the familiar derivation of the representations of $\mathcal{L}(SO(3))$. One arrives at the formulas $\langle q;p|q;p\rangle = 1$,

$$P|q;p\rangle = p|q;p\rangle, \tag{3.2a}$$

$$A^{\pm}|q;p\rangle = \sqrt{q + p(p \pm 1)}\ |q;\ p\pm 1\rangle, \tag{3.2b}$$

$$Q|q;p\rangle = q|q;p\rangle \qquad (Q = K^2 + H^2 - P^2), \tag{3.3}$$

where

$$H = \frac{1}{2}(A^+ + A^-), \qquad K = \frac{1}{2i}(A^+ - A^-). \tag{3.4}$$

The representation is at least partially labeled by q, the constant value of the Casimir operator Q. The requirement that H and K be Hermitian restricts the possible values of q and the values of p which occur in an irreducible representation for a given q.[3] The result is that there are three classes of single-valued representations:

(1) Trivial representation: $q = 0$; $p = 0$.

(2) Continuous series: $q > 0$; p = all integers. (3.5a)

(3) Discrete series: $q = -k(k-1)$, $k = 1, 2, \ldots$;

$$p = k, k+1, \ldots \quad \text{or} \quad p = -k, -k-1, \ldots \tag{3.5b}$$

(two irreducible discrete representations for each k).

The continuous representations with $q \geq 1/4$ are called the principal series; those with $0 < q < 1/4$ are called the complementary (or supplementary, or exceptional) series. The matrix elements of the group operators in a representation of the complementary series have a different qualitative behavior. As a result, these representations do not occur in the decomposition of the quasiregular representation on the hyperboloid (see Sec. VI.1). (See also Bargmann (1947), pp. 609-639.)

There are other representations of $\mathcal{L}(SO_0(1,2))$ which correspond to representations of a covering group of $SO_0(1,2)$. In such a case the p's need not be integers, but they still vary in integral steps within an irreducible representation. The

[3] This analysis is performed by a clear and elegant graphical method in Philips (1963).

representations can be classified in the same way as above, with similar results except that the possibilities become more complicated in the interval $0 \leq q \leq 1/4$ (Philips (1963), Chap. 5).

The representation (3.2) can be rewritten in a form more convenient for our purposes. A general vector in the Hilbert space can be written

$$\Psi = \sum_p \Psi(p)\,|q;p\rangle \qquad \left(\sum_p |\Psi(p)|^2 < \infty\right). \quad (3.6)$$

The scalar product of two vectors is

$$(\Psi,\phi) = \sum_p \Psi^*(p)\ \phi(p). \quad (3.7)$$

Eqs. (3.2) are equivalent to

$$P\Psi(p) = p\Psi(p), \quad (3.8a)$$

$$H\Psi(p) = \frac{1}{2}\sqrt{q + p(p-1)}\ \Psi(p-1) + \frac{1}{2}\sqrt{q + p(p+1)}\ \Psi(p+1), \quad (3.8b)$$

$$K\Psi(p) = \frac{1}{2i}\sqrt{q + p(p-1)}\ \Psi(p-1) - \frac{1}{2i}\sqrt{q + p(p+1)}\ \Psi(p+1). \quad (3.8c)$$

4. Representations of the Three-Dimensional De Sitter Group.

The symmetry group of the three-dimensional closed universe of constant curvature has been much studied under its alias, the homogeneous Lorentz group. The irreducible unitary representations were found by Gel'fand and Naimark (1946, 1947) and Bargmann (1947), pp. 570-571 and 640. The following information is taken from [Naimark], Sec. 8, with some changes in notation. (In particular, the relative phase of the vectors corresponding to different values of the parameter k has been changed, to smooth the process of contraction in Sec. II.4.)

Let

$$P_{\pm} = P^{1} \pm iP^{2}, \qquad K_{\pm} = K^{1} \pm iK^{2}. \tag{4.1}$$

An irreducible representation of the group is characterized by two numbers, d and k_0. It acts in a direct sum of vector spaces $\mathcal{H}_k$ ($k = k_0, k_0 + 1, \ldots$) in each of which the SO(3) subgroup generated by J and $\vec{P}$ acts according to the irreducible representation of spin k. We shall write the representation of the Lie algebra in the form analogous to Eqs. (3.6-8). A general vector has the form

$$\Psi = \sum_{k=k_0}^{\infty} \sum_{n=-k}^{k} \Psi(k,n) \; |d,k_0;k;n\rangle. \tag{4.2}$$

Then the representation of the generators is

$$J\Psi(k,n) = n\,\Psi(k,n), \tag{4.3a}$$

$$P_{\pm}\Psi(k,n) = \sqrt{k(k+1) - n(n\mp 1)}\;\Psi(k,n\mp 1), \tag{4.3b}$$

$$H\Psi(k,n) = -\, i\sqrt{(k+n+1)(k-n+1)}\; C(k+1)\; \Psi(k+1,n)$$

$$-\, n\, A(k)\; \Psi(k,n) - i\sqrt{(k+n)(k-n)}\; C(k)\; \Psi(k-1,n), \quad (4.4)$$

$$K_{\pm}\Psi(k,n) = \sqrt{(k\mp n+1)(k\mp n+2)}\; C(k+1)\; \Psi(k+1,n\mp 1)$$

$$\mp\, i\sqrt{(k\pm n)(k\mp n+1)}\; A(k)\; \Psi(k,n\mp 1)$$

$$-\sqrt{(k\pm n)(k\pm n-1)}\; C(k)\; \Psi(k-1,n\mp 1), \quad (4.5)$$

where

$$A(k) = \frac{-\,k_0 d}{k(k+1)}, \quad (4.6a)$$

$$C(k) = \frac{i}{k}\left[(k^2 - k_0^2)(k^2 + d^2)/(4k^2 - 1)\right]^{1/2}. \quad (4.6b)$$

There are two kinds of unitary representations:

(1) Principal series: d real, $k_0 = 0,\ 1/2,\ 1,\ \ldots.$ (4.7a)

If $k_0 = 0$, d can be taken positive.

(2) Complementary series: $k_0 = 0$, d imaginary,

$0 \leq \mathrm{Im}\, d \leq 1.$ (4.7b)

Of course, there is also a trivial one-dimensional representation. The Casimir operators (I.4.2) and (I.4.5) take the values

$$Q = -k_0^2 + d^2 + 1, \tag{4.8a}$$

$$Q_2 = k_0 d \tag{4.8b}$$

([Naimark], p. 167). Unlike $SO_o(1,n)$ with n even, $SO_o(1,3)$ does not have a discrete series of representations.

5. Representations of the Four-Dimensional De Sitter Group.

Ström (1965) has put the irreducible representations[4] of $\mathcal{L}(SO_o(1,4))$ into the most convenient form for comparison with the representations of $\mathcal{L}(ISO_o(1,3))$. He denotes our $\vec{K}$ by $-\vec{N}$. The basis vectors have the labels $|r,\sigma;l,n;j;m\rangle$, where r and σ are related to the constant values of the Casimir operators (I.4.2,8) by

$$Q = -r(r+1) + \sigma + 2,$$
$$Q_2 = r(r+1)\sigma. \tag{5.1}$$

The representation is reduced with respect to the SO(4) subgroup generated by $\vec{J}$ and $\vec{P}$, which in turn is reduced with respect to the SO(3) subgroup generated by $\vec{J}$. The indices l and n label the SO(4) representations which occur[5]; the relationship to the

[4] The original references are Thomas (1941), Newton (1949, 1950), Dixmier (1961), Takahashi (1963).
[5] SO(4) is the direct product of the SO(3) groups generated by $(\vec{J} + \vec{P})/2$ and $(\vec{J} - \vec{P})/2$. In an irreducible representation where these factor groups have respective spins k' and k, one sets $l = k' + k + 1$, $n = k' - k$.

Casimir operators of SO(4) is

$$\vec{J}^2 + \vec{P}^2 = l^2 + n^2 - 1,$$

$$\vec{J}\cdot\vec{P} = ln. \quad (5.2)$$

Then j and m have the usual meaning with respect to the $\vec{J}$ subgroup. The smallest value of j which occurs in a representation of SO(4) is $|n|$.

The formulas for the representation of the Lie algebra are complicated, so we shall not repeat them.[6] For instance, $K_3|l,n;j;m\rangle$ has 12 terms, the j index taking the values $j - 1$, j, $j + 1$ and the (n,l) pair taking the values $(n, l \pm 1)$ and $(n \pm 1, l)$.

The classification of the unitary irreducible representations is the following (in addition to the trivial representation):

(1) Continuous series:

$$r = 0, \frac{1}{2}, 1, \ldots; \quad (5.3a)$$

[6] Dr. Ström has supplied the following correction to his paper: In the expression for $\alpha(j,l,n)$ (p. 461) the factor $l^2 - (j - 1)^2$ should be $l^2 - j^2$.

$$\sigma > \begin{cases} -2 \text{ if } r = 0 \\ 0 \text{ if } r = 1, 2, \ldots \\ \frac{1}{4} \text{ if } r = \frac{1}{2}, \ldots . \end{cases} \qquad (5.3b)$$

The range of the internal parameters is

$$l = r + 1, r + 2, \ldots, \qquad (5.4)$$

$$- r \leq n \leq r. \qquad (5.5)$$

The representations with $\sigma > 1/4$ are called the (<u>first</u>) <u>principal series</u>.

(2) <u>Discrete series</u>:

(a) $$\sigma = - q(q - 1), \qquad q = \frac{1}{2}, 1, \ldots; \qquad (5.6a)$$

$$r = q, q + 1, \ldots . \qquad (5.6b)$$

The range of the internal parameters is given by Eq. (5.4) and

$$q \leq n \leq r \qquad \text{or} \qquad - q \geq n \geq - r \qquad (5.7)$$

(two irreducible representations for each value of q and r).

(b) $\sigma = 0, \ r = 1, 2, \ldots; \qquad n \equiv 0.$ (5.8)

The range of l is given by Eq. (5.4).

Half-integral values of r imply half-integral values of j, and hence double-valued representations of $SO_o(1,4)$.

Note the similarity of this classification to the one in Sec. B.3, σ playing the role of the q of $SO_o(1,2)$. (Case (2b) corresponds to the trivial representation of $SO_o(1,2)$.) In the contraction to representations of $ISO_o(1,3)$ this parallelism shows up in an interesting way (Sec. (II.5)).

Appendix C

A STUDY OF THE CONTRACTION OF THE REPRESENTATIONS OF SO(3)

For comparison with the discussion in Chapters II and VI of the contraction of representations of $SO_0(1,2)$ to representations of $ISO_0(1,1)$, we examine the analogous problem for a more familiar pair of groups, SO(3) (the three-dimensional rotation group) and ISO(2) (the two-dimensional Euclidean group, often denoted E(2)).

1. Irreducible Unitary Representations of $\mathcal{L}$(SO(3)).

In analogy with Eqs. (B.3.6-8) we re-express the standard formulas for the irreducible representations of the Lie algebra of SO(3) in terms of a Hilbert space of functions. For the representation of dimension 2j + 1 we write

$$\Psi = \sum_{m=-j}^{j} \Psi(m)\,|j;m\rangle \; ; \qquad (1.1)$$

then the representation is

$$J_3 \Psi(m) = m\,\Psi(m), \qquad (1.2a)$$

$$J_1 \Psi(m) = \frac{1}{2} \{\sqrt{j(j+1) - m(m-1)}\ \Psi(m-1)$$

$$+ \sqrt{j(j+1) - m(m+1)}\ \Psi(m+1)\}, \qquad (1.2b)$$

$$J_2 \Psi(m) = \frac{1}{2i} \{\sqrt{j(j+1) - m(m-1)}\ \Psi(m-1)$$

$$- \sqrt{j(j+1) - m(m+1)}\ \Psi(m+1)\}. \qquad (1.2c)$$

The representation is characterized by the value of the Casimir operator

$$\vec{J}^{\,2} = \sum_{A=1}^{3} J_A^{\,2} : \qquad (1.3)$$

$$\vec{J}^{\,2} \Psi(m) = j(j + 1)\ \Psi(m). \qquad (1.4)$$

The scalar product is

$$(\Psi,\phi) = \sum_{m=-j}^{j} \Psi^*(m)\phi(m). \qquad (1.5)$$

2. <u>Irreducible Unitary Representations of</u> $\mathcal{L}$(ISO(2)).

The representations of ISO(2) are similar to those of $ISO_0(1,1)$ (see Sec. B.2). There are two types of orbits, the origin ($p_1 = p_2 = 0$) and the circles

$$(p_1)^2 + (p_2)^2 = M^2 . \qquad (2.1)$$

In the first case the translations are represented trivially, and the formula (B.2.3) reduces to a representation of the little group, which is the rotation group SO(2). All these representations are one-dimensional:

$$e^{-i\theta J}|m\rangle = e^{-im\theta}|m\rangle \qquad (2.2)$$

(m = integer or half-integer).

In the case (2.1) the little group has order 1. We can choose p_1 ($-M \leq p_1 \leq M$) and $\sigma = \text{sgn}\, p_2$ as independent variables. The scalar product (B.2.2a) is

$$(\Psi,\phi) = \sum_{\sigma} \int_{-M}^{M} \frac{dp_1}{2\sqrt{M^2 - p_1^2}} \Psi^*(p_1,\sigma)\phi(p_1,\sigma) . \qquad (2.3)$$

The representation is given by a formula identical to Eq. (B.2.4), where in the present case $b^\mu p_\mu = b^1 p_1 + b^2 p_2$ and Λ is a rotation:

$$\Lambda \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \equiv e^{-i\theta J} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} \cos\theta\, p_1 - \sin\theta\, p_2 \\ +\sin\theta\, p_1 + \cos\theta\, p_2 \end{pmatrix} . \qquad (2.4)$$

Then the generators are

$$P_1 \Psi(p) = p_1 \Psi(p) , \qquad P_2 \Psi(p) = \sigma\sqrt{M^2 - p_1^2}\, \Psi(p) , \qquad (2.5a)$$

$$J\Psi(p) = + i\frac{d}{d\theta}\Psi(\Lambda^{-1}p)\Big|_{\theta=0} = + i\sigma\sqrt{M^2 - p_1^2}\,\frac{d\Psi}{dp_1}. \qquad (2.5b)$$

The representation is characterized by

$$\vec{P}^2\,\Psi = M^2\,\Psi. \qquad (2.6)$$

This representation can be expressed in a form in which the operator J is diagonalized; in other words, in which the restriction of the representation to the SO(2) subgroup exp(-iθJ) is explicitly reduced into a direct sum of irreducible SO(2) representations. If we make the change of variables

$$p_1 = M\cos\phi, \qquad p_2 = M\sin\phi, \qquad (2.7)$$

it is easy to see that

$$J\Psi(\phi) = - i\frac{\partial\Psi}{\partial\phi}, \qquad (2.8a)$$

$$e^{-i\theta J}\,\Psi(\phi) = \Psi(\phi - \theta). \qquad (2.8b)$$

The eigenfunctions of J are

$$\Psi_m(\phi) = e^{im\phi} \qquad (J\Psi_m(\phi) = m\,\Psi_m(\phi)). \qquad (2.9)$$

Then

$$P_1 \Psi_m(\phi) = M \cos\phi\, \Psi_m(\phi) = \frac{M}{2}[\Psi_{m+1}(\phi) + \Psi_{m-1}(\phi)],$$

$$P_2 \Psi_m(\phi) = M \sin\phi\, \Psi_m(\phi) = \frac{M}{2i}[\Psi_{m+1}(\phi) - \Psi_{m-1}(\phi)]. \tag{2.10}$$

Finally, writing

$$\Psi = \sum_m \Psi(m)\, \Psi_m(\phi), \tag{2.11}$$

we put the representation into the form

$$J\Psi(m) = m\, \Psi(m), \tag{2.12a}$$

$$P_1 \Psi(m) = \frac{M}{2}[\Psi(m-1) + \Psi(m+1)],$$

$$P_2 \Psi(m) = \frac{M}{2i}[\Psi(m-1) - \Psi(m+1)]. \tag{2.12b}$$

The scalar product (2.3) is

$$(\Psi,\phi) = \int_0^{2\pi} d\phi\, \Psi^*(\phi)\phi(\phi) = 2\pi\sum_m \Psi^*(m)\phi(m). \tag{2.13}$$

From Eqs. (2.10) it is clear that the spectrum of J in an irreducible representation consists of all the numbers which are separated by integers from some m . The group representation is single-valued if m is an integer.

3. Formal Contraction of the Irreducible Representations with Respect to the Diagonalized Subgroup.

Substituting

$$P_1 = \frac{1}{R} J_1, \quad P_2 = \frac{1}{R} J_2, \quad J = J_3 \tag{3.1}$$

into the commutation relations of $\mathcal{L}(SO(3))$ (Eq. (I.3.10a)) and taking $R \longrightarrow \infty$ leads to

$$[P_1, P_2] = 0, \quad [J, P_1] = iP_2, \quad [J, P_2] = -iP_1, \tag{3.2}$$

the Lie algebra of ISO(2). Similarly, Eq. (1.3) yields

$$\vec{P}^2 = \lim_{R \to \infty} \frac{1}{R^2} \vec{J}^2. \tag{3.3}$$

Thus ISO(2) is a contraction of SO(3) with respect to a one-parameter subgroup.

When the SO(3) representation formulas (1.2) are expressed in terms of $\vec{P}$ and J, an immediate passage to the limit $R \longrightarrow \infty$ yields a representation of $\mathcal{L}(ISO(2))$ in which the translations are represented trivially ($\vec{P} \equiv 0$). It is a direct sum of 2j + 1 irreducible representations of the form (2.2).

Inönü and Wigner (1953) pointed out that a more interesting relationship between the representations of SO(3) and those of ISO(2) can be observed by letting the SO(3) representation vary with R so that the value of the Casimir operator approaches a definite nonvanishing limit. Express Eqs. (1.2), (1.4), and (1.5) in terms of $\vec{P}$, J, and $M \equiv j/R$. Take the limit

$$j \longrightarrow \infty, \quad R \longrightarrow \infty, \qquad M \longrightarrow \text{const.} \neq 0, \tag{3.4}$$

multiplying each equation by the power of R necessary to keep both sides finite and not identically zero. The results are Eqs. (2.12), (2.6), and (2.13). (The limits in Eqs. (1.2b,c) are not uniform in m; the prescription is to act as if $m/R \ll M$.) $\mathcal{L}(ISO(2))$ representations with integral and with half-integral m values are obtained in this way. The fastest convergence ($O(R^{-2})$ instead of $O(R^{-1})$) is obtained by taking

$$M = \frac{1}{R}\left(j + \frac{1}{2}\right). \tag{3.5}$$

As this example illustrates, "contraction of representations" typically refers to a relation between whole families or sequences of representations of the two groups.

4. <u>Formal Contraction with Respect to a Subgroup which Is Not Diagonalized</u>.

Suppose that in place of Eqs. (3.1) we set

$$P_1 = \frac{1}{R} J_3, \quad P_2 = \frac{1}{R} J_1, \quad J = J_2 \tag{4.1}$$

and attempt to carry out the limit (3.4) in Eqs. (1.2). We find that to get a finite expression for P_1 out of Eqs. (1.2a) we must absorb a factor of R by setting

$$p_1 = \frac{m}{R}. \qquad (4.2)$$

But then in the limit the distance between the values of p_1 vanishes. We tentatively postulate, therefore, that for $R = \infty$ p_1 is a continuous variable ranging from $-M$ to $+M$, and that the vectors of the representation space are functions of p_1. The scalar product (1.5), if divided by R, goes into

$$(\Psi,\phi) = \int_{-M}^{M} dp_1 \, \Psi^*(p_1)\phi(p_1).$$

Next we attack Eqs. (1.2b,c). We make the ansatz

$$\frac{1}{2}[\Psi(p_1 - \frac{1}{R}) + \Psi(p_1 + \frac{1}{R})] \longrightarrow \Psi(p_1), \qquad (4.3)$$

$$\frac{R}{2}[\Psi(p_1 - \frac{1}{R}) - \Psi(p_1 + \frac{1}{R})] \longrightarrow -\frac{d}{dp_1}\Psi(p_1). \qquad (4.4)$$

Then

$$P_2\Psi(p_1) = \frac{1}{2}\sqrt{M^2 - p_1^2}\,[\Psi(p_1 - \frac{1}{R}) + \Psi(p_1 + \frac{1}{R})] + O(R^{-1})$$

$$\longrightarrow \sqrt{M^2 - p_1^2}\,\Psi(p_1),$$

$$J\Psi(p_1) = \frac{R}{2i}\{[\sqrt{M^2 - p_1^2} + \frac{M + p_1}{2R\sqrt{M^2 - p_1^2}} + O(R^{-2})]\,\Psi(p_1 - \frac{1}{R})$$

$$- [\sqrt{M^2 - p_1^2} + \frac{M - p_1}{2R\sqrt{M^2 - p_1^2}} + O(R^{-2})]\,\Psi(p_1 + \frac{1}{R})\}$$

$$= \frac{R}{2i}\sqrt{M^2 - p_1^2}\,[\Psi(p_1 - \frac{1}{R}) - \Psi(p_1 + \frac{1}{R})] +$$

$$\frac{1}{4i}\,\frac{p_1}{\sqrt{M^2 - p_1^2}}\,[\Psi(p_1 - \frac{1}{R}) + \Psi(p_1 + \frac{1}{R})] + O(R^{-1})$$

$$\longrightarrow\ i\sqrt{M^2 - p_1^2}\,\frac{d}{dp_1}\Psi(p_1) - \frac{i}{2}\,\frac{p_1}{\sqrt{M^2 - p_1^2}}.$$

If we now set

$$\bar{\Psi}(p_1) = \sqrt{2}\,\sqrt[4]{M^2 - p_1^2}\;\Psi(p_1) \qquad (4.5)$$

and rewrite all the formulas in terms of $\bar{\Psi}$, we have

$$P_1\bar{\Psi}(p_1) = p_1\,\bar{\Psi}(p_1),$$

$$P_2\bar{\Psi}(p_1) = \sqrt{M^2 - p_1^2}\;\bar{\Psi}(p_1),$$

$$J\bar{\Psi}(p_1) \equiv \sqrt{2}\,\sqrt[4]{M^2 - p_1^2}\,[J\Psi](p_1)$$

$$= i\sqrt{M^2 - p_1^2}\,\frac{d\bar{\Psi}}{dp_1} + i(M^2 - p_1^2)^{3/4}\,\frac{d}{dp_1}[(M^2 - p_1^2)^{-1/4}]\,\bar{\Psi}$$

$$- \frac{i}{2}\,\frac{p_1}{\sqrt{M^2 - p_1^2}}\,\bar{\Psi} = i\sqrt{M^2 - p_1^2}\,\frac{d}{dp_1}\bar{\Psi}(p_1),$$

$$(\Psi,\emptyset) = \int_{-M}^{M} \frac{dp_1}{\sqrt{M^2 - p_1^2}}\,\bar{\Psi}^*(p_1)\bar{\emptyset}(p_1).$$

Comparing with Sec. C.2, we observe that we have come up with only half of a representation of ISO(2), the part with $\sigma = +1$. Our operators have the expected commutation relations, but they are not the infinitesimal generators of a representation of the group.

According to Nelson (1959), in order to have a unitary group representation P_1, P_2, J, and

$$\Delta = P_1^2 + P_2^2 + J^2 \qquad (4.6)$$

must be essentially self-adjoint (see Sec. B.3). In the present case the functions in the Hilbert space are defined on a semicircle in the upper half of the p_1-p_2 plane (Fig. 19). The infinitesimal transformation $1 + \theta J + \ldots$ is attempting to rotate the functions onto the other half of the circle. J cannot be integrated to a finite unitary operator (i.e., J cannot be made self-adjoint) unless we either enlarge the Hilbert space to include the bottom half of the circle or impose a boundary condition which, for instance, requires the operator to feed the functions back in at the right end of the semicircle as soon as it pushes them out at the left.[1] The nonexistence of representations of ISO(2) with only one sign of p_2 shows that the latter cannot be done in such a way that Δ is self-adjoint.

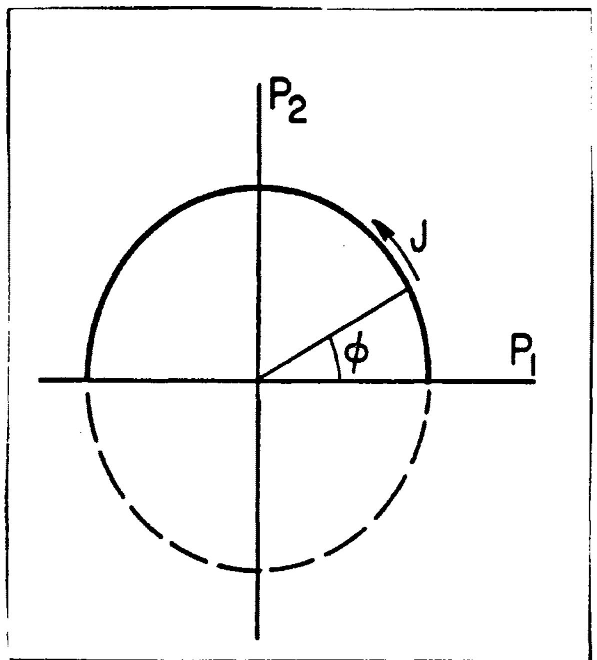


Fig. 19

The $\vec{p}$ plane for a nonintegrable representation of $\mathcal{L}$(ISO(2)).

But why did our heuristic manipulation give us the top half of the circle rather than the bottom? Did we slip in an unconscious assumption that P_2 is positive? The resolution of this perplexity is amusing. The choice of the relative phases of the basis vectors in an irreducible representation of $\mathcal{L}$(SO(3)), although almost universal, is really arbitrary. Suppose we were to change the sign of all the $|j;m\rangle$ with odd m. Then we would have

[1] Cf. Wightman (1964), pp. 264-266.

$$J_{\pm}\,|j;m\rangle = -\sqrt{j(j+1) - m(m\pm 1)}\;|j;m\pm 1\rangle \qquad (4.7)$$

instead of the standard formula without the minus sign. There will be a corresponding sign change in Eqs. (1.2b,c). Then the same procedure as above would lead to the opposite sign for the operators P_a and J -- so we would obtain the part of the representation with $\sigma = -1$. More generally, note that most choices of the phases of $|j;m\rangle$ yield expressions for P_a and J which do not converge at all as $R \longrightarrow \infty$. (Also, the contraction

$$J = J_1, \quad P_1 = \frac{1}{R} J_2, \quad P_2 = \frac{1}{R} J_3$$

with the canonical choice of phases leads to this sort of trouble.)

Let us scrutinize the argument to see if it can be fixed up to give a whole representation instead of a just a fragment of one. The expressions on the left of Eqs. (4.3-4) involve evaluations of Ψ at points separated by $2/R$; that is, points corresponding to values of m that are both odd or both even. Our reasoning was that in the limit, when these points coalesce, the values of Ψ at adjacent points must also approach each other, so that the sequence $\Psi(m)$ is replaced by a differentiable function $\Psi(p_1)$. However, the utility of Eqs. (4.3-4) will not be affected if we postulate that as $R \longrightarrow \infty$ the sequence $\Psi(m)$ with m even flows together into a smooth function $\Psi_{+1}(p_1)$ and the sequence $\Psi(m)$ with m odd becomes a <u>different</u>

function $\bar{\Psi}_{-1}(p_1)$. Then the same calculation as before results in

$$P_1 \bar{\Psi}_\rho(p_1) = p_1 \bar{\Psi}_\rho(p_1), \tag{4.8a}$$

$$P_2 \bar{\Psi}_\rho(p_1) = \sqrt{M^2 - p_1^2}\, \bar{\Psi}_{-\rho}(p_1), \tag{4.8b}$$

$$J\bar{\Psi}_\rho(p_1) = i\sqrt{M^2 - p_1^2}\, \frac{d}{dp_1} \bar{\Psi}_{-\rho}(p_1), \tag{4.8c}$$

$$(\Psi,\phi) = \sum_\rho \int_{-M}^{M} \frac{dp_1}{2\sqrt{M^2 - p_1^2}}\, \bar{\Psi}^*_\rho(p_1)\bar{\phi}_\rho(p_1). \tag{4.8d}$$

Now let

$$\Psi(p_1,1) = \frac{1}{\sqrt{2}}[\bar{\Psi}_{+1}(p_1) + \bar{\Psi}_{-1}(p_1)],$$

$$\Psi(p_1,-1) = \frac{1}{\sqrt{2}}[\bar{\Psi}_{-1}(p_1) - \bar{\Psi}_{+1}(p_1)]. \tag{4.9}$$

Then for $\sigma = \pm 1$, $p = (p_1,\sigma)$, we obtain equations identical with Eqs. (2.5) and (2.3).

So, starting with the advantage of knowing the answer beforehand, we have pulled the representations of the Euclidean group out of the representations of the rotation group by hook and crook. In what follows, some of the *ad hoc* features of this discussion will be given a geometrical interpretation.

5. Contraction of the Quasiregular Representation at the Pole.

Contraction is fundamentally a geometrical notion. Intuitively, the action of SO(3) on a sphere is approximated in the neighborhood of a point O by the contraction of SO(3) with respect to the subgroup of rotations about the axis through O, which we shall take to be the z-axis in the following discussion. This can be made precise in terms of a natural action of the contracted group (ISO(2)) on the tangent plane to the sphere at O (see [Talman], pp. 206-209). It is reasonable to expect that the contraction of the representations can be interpreted in terms of this geometrical picture.

Such a connection was already established in the original paper of Inönü and Wigner (1953). (See also [Vilenkin], pp. 228-230.) They pointed out that the realization of the irreducible representations of SO(3) by basis vectors which are the functions on the sphere

$$|j;m\rangle = Y_j^m(\theta,\phi) = (-1)^m \left[\frac{2j+1}{4\pi}\,\frac{(j-m)!}{(j+m)!}\right]^{1/2} P_j^m(\cos\theta)\, e^{im\phi} \qquad (5.1)$$

is related (near the polar point, where $\theta = 0$) to the realization of the irreducible representations of ISO(2) by the basis functions in the plane

$$|M;m\rangle = J_m(M\theta)\, e^{im\phi} \qquad (5.2)$$

via the (previously known) formula

$$\lim_{j\to\infty} j^{-m} P_j^m(\cos \frac{x}{j}) = J_{-m}(x) = (-1)^m J_m(x). \qquad (5.3)$$

(In Eq. (5.2) J_m is a Bessel function, θ and ϕ are polar coordinates in a plane:

$$x = \theta \cos \phi, \qquad y = \theta \sin \phi, \qquad (5.4)$$

and M^2 is the value of the Casimir operator for the representation (2.12). The |M;m> obey Eqs. (2.10) with

$$P_1 = + i\frac{\partial}{\partial y}, \qquad P_2 = - i\frac{\partial}{\partial x}; \qquad (5.5)$$

these identifications are easily seen to be in keeping with the geometrical picture.) A generalization of Eq. (5.3) ([Vilenkin], p. 229) relates the matrix elements of the SO(3) representations (Jacobi polynomials) to the matrix elements of the ISO(2) representations (Bessel functions).

Eq. (5.3) relates the basis functions of the representations of the two groups near the pole. The irreducible representation of spin j, however, comprises functions which are nonzero in regions all over the sphere.[2] There is no reason, therefore, to expect a representation of ISO(2) to be an overall approximation to a representation of SO(3). It seems more to the point to study the group action in the set of all functions which

[2] The Y_j^m with m << j tend to be concentrated near the poles, those with m ≈ j near the equator. This is obviously related to the nonuniform convergence observed in Sec. C.3.

are concentrated near the pole, and this requires considering an expansion with respect to j.

The functions (5.1) are the eigenfunctions (of eigenvalue $j(j+1)$) of the Casimir operator of SO(3) as a differential operator on the sphere (cf. Sec. A.3 and Sec. V.1). Any L^2 function on the sphere can be expanded in terms of this complete orthonormal set. This provides a decomposition of the quasiregular representation (see Sec. A.3) of SO(3) into irreducible representations. Similarly, the functions of Eq. (5.2) are eigenfunctions of the negative of the Laplacian in the plane with eigenvalue M^2, and the same statements apply with the obvious changes.

Let us write down the spherical harmonic expansion for functions on a sphere of radius R, using a polar angle coordinate θ which is scaled so as to measure the geodesic distance from the pole in constant units (independent of R).[3] The range of the variables is

$$0 \le \theta \le \pi R, \qquad -\pi < \phi \le \pi.$$

Define (cf. Eq. (3.5))

[3] That is, instead of shrinking a neighborhood of the pole down to a point, we keep the dimensions of the neighborhood constant and expand the radius of the sphere. These two viewpoints are obviously equivalent, but the one chosen is simpler to handle algebraically and also is more in keeping with the cosmological motivation of our problem.

$$Z^m_M(\theta,\phi) = \sqrt{R}\, Y^m_{RM-1/2}(\theta/R,\ \phi). \tag{5.6}$$

Let $g(\theta,\phi)$ be a function on the sphere with support in the region where $\theta < \pi R_0$ (R_0 fixed, $R_0 < R$ for all values of R considered). Define the transform of g by

$$\tilde{g}(M,m) = \int_{-\pi}^{\pi} d\phi \int_0^{\pi R} \sin\frac{\theta}{R}\, d\theta\ Z^{*m}_M(\theta,\phi)\, g(\theta,\phi). \tag{5.7}$$

The inverse transformation is[4]

$$g(\theta,\phi) = \frac{1}{R^2} \sum_M \sum_m Z^m_M(\theta,\phi)\,\tilde{g}(M,m), \tag{5.8a}$$

$$-\left(RM - \frac{1}{2}\right) \le m \le RM - \frac{1}{2}, \qquad M - \frac{1}{2R} = 0,\ \frac{1}{R},\ \frac{2}{R},\ \ldots. \tag{5.8b}$$

Note that $\tilde{g}(M,m)$ is defined (in fact, analytic) for all positive M, even though only a discrete set of values enters the inversion formula. The scalar product is

$$R\int d\phi \int \sin\frac{\theta}{R}\, d\theta\ |g(\theta,\phi)|^2 = \frac{1}{R} \sum_M \sum_m |\tilde{g}(M,m)|^2. \tag{5.9}$$

It should be noted that the rescaling of θ has a direct connection with the contraction transformation (3.1). The sphere, being a homogeneous space, can be identified with the

[4] The orthonormality and completeness relations for the spherical harmonics are given in [Messiah], p. 495.

cosets $SO(3)/S$, where $S = \{\exp(-itJ_3)\}$ is the stability subgroup of the pole O. The elements of the group can be given the Euler angle parametrization:

$$\exp(-i\phi J_3)\,\exp(-i\theta J_2)\,\exp(-itJ_3). \qquad (5.10)$$

The cosets are then labeled by θ and ϕ; this is precisely the familiar spherical coordinate system. (The elements of the coset (θ,ϕ) map O into the point (θ,ϕ).) Clearly, scaling J_2 as in Eq. (3.1) redefines θ by a factor of R.

Combining Eqs. (5.6) and (5.1) and a refined version[5] of Eq. (5.3), we find that for $M \gg 1/R$, $M \gg |m|/R$, and $\theta \le R_0 \ll R$,

$$\frac{1}{R} Z_M^m(\theta,\phi) = \frac{1}{\sqrt{2\pi}} \sqrt{M}\, J_m(M\theta)\, e^{im\phi} + O(R^{-2}). \qquad (5.11)$$

(The 1/2 in Eq. (3.5) is essential to eliminate a term of order R^{-1}.) Consequently, as $R \longrightarrow \infty$ the integral (5.7) approaches a limit:

$$\overset{(R)}{\tilde{g}}(M,m) \longrightarrow \overset{(\infty)}{\tilde{g}}(M,m) =$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} d\phi \int_0^{\infty\,[\text{or } R_0]} \theta\, d\theta\, \sqrt{M}\, J_m(M\theta)\, e^{-im\phi}\, g(\theta,\phi). \qquad (5.12a)$$

---

[5] [Gradshteyn-Ryzhik], 8.722.1 (p. 1003). The formula contains a misprint: the exponent ν on the left-hand side should be μ.

Likewise,

$$g(\theta,\phi) = \frac{1}{\sqrt{2\pi}} \int_0^\infty dM \sum_{M=-\infty}^{\infty} \sqrt{M}\, J_m(M\theta)\, e^{im\phi}\, \tilde{g}(M,m). \qquad (5.12b)$$

So we have recovered the eigenfunction expansion on the plane, which is summarized by the standard formula

$$\int_0^\infty dM\, M\, J_m(M\theta) J_m(M\theta') = \frac{1}{\theta} \delta(\theta - \theta') \qquad (5.13)$$

(e.g., [Jackson], p. 77). The scalar product converges to

$$\int d\phi \int \theta\, d\theta\, |g(\theta,\phi)|^2 = \int dM \sum_m |\tilde{g}(M,m)|^2. \qquad (5.14)$$

6. Contraction to an Equatorial Point.

Another parametrization which yields the same coordinate system as Eq. (5.10) is

$$\exp(-i\phi J_3)\, \exp(+i\psi J_2)\, \exp(-itJ_1) \qquad (\psi = \frac{\pi}{2} - \theta). \qquad (6.1)$$

Here the sphere is exhibited as the space of cosets relative to the stability group of the origin of the (ϕ,ψ) coordinates. In the realization of the irreducible representations in terms of the spherical harmonics, the J_3 subgroup, which is the group of "translations" in ϕ, is diagonalized. Hence Eq. (6.1) provides a natural setting for a contraction of the type of Sec. C.4.

However, as remarked in Sec. C.4, the standard phases for the spherical harmonics lead to divergent expressions when contracted around the x_1-axis. To achieve success one must either change the phases or move to the x_2-axis. We shall do the latter. Then in place of Eq. (6.1) we have

$$\exp(-i\omega J_3)\ \exp(-i\Psi J_1)\ \exp(-itJ_2), \tag{6.2a}$$

where

$$\omega = \phi - \frac{\pi}{2}, \qquad \Psi = \frac{\pi}{2} - \theta. \tag{6.2b}$$

Contraction according to Eq. (4.1) induces a rescaling of both ω and Ψ by a factor of R; from now on we employ the rescaled variables. Let

$$X^{p}_{M}(\omega,\Psi) = (-\ i)^{RM-1/2}\ Y^{Rp}_{RM-1/2}\left(\frac{\pi}{2} - \frac{\Psi}{R}, \frac{\pi}{2} + \frac{\omega}{R}\right). \tag{6.3}$$

In analogy to Eqs. (5.7,8a) we have a transform

$$\tilde{g}(M,p) = \int_{-\pi R}^{\pi R} d\omega \int_{-\pi R/2}^{\pi R/2} \cos\frac{\Psi}{R}\, d\Psi\ X^{*p}_{M}(\omega,\Psi)\, g(\omega,\Psi), \tag{6.4}$$

$$g(\omega,\Psi) = \frac{1}{R^2} \sum_M \sum_p X^{p}_{M}(\omega,\Psi)\, \tilde{g}(M,p). \tag{6.5}$$

The sum is over the range (5.8b), with $p = m/R$ (m = integer or half-integer). In what follows it is assumed that g has support where ω, $\Psi < \pi R_0$, $R_0 \ll R$.

Now for θ not too far from $\pi/2$ and j large one has

$$P_j^m(\cos\theta) \approx j^{m-1/2}\sqrt{\frac{2}{\pi \sin\theta}}\cos\left[\left(j+\frac{1}{2}\right)\theta - \frac{\pi}{4} + \frac{m\pi}{2}\right] \qquad (6.6)$$

([Gradshteyn-Ryzhik], Eq. (8.721.4) (p. 1003)). This formula is valid only for $|m| \ll j$. A more uniform approximation can be found, but the algebraic complications of carrying through the following discussion in terms of it are enormous and would obscure the main point. So the corrections needed for $m \approx j$ will only be stated at the end. One finds from Eqs. (6.3), (5.1), and (6.6) that an approximation to X_M^p analogous to Eq. (5.11) is

$$X_M^p(\omega,\Psi) \approx (-i)^{j+m}\,\frac{1}{\pi}\cos\left[-M\Psi + (j+m)\frac{\pi}{2}\right]e^{ip\omega}$$

$$\left(j = RM - \frac{1}{2},\ m = Rp\right).$$

That is,

$$X_M^p(\omega,\Psi) \approx \frac{1}{\pi}\cos M\Psi\, e^{ip\omega} \qquad \text{if } j+m \text{ is even,} \qquad (6.7a)$$

$$X_M^p(\omega,\Psi) \approx -i\,\frac{1}{\pi}\sin M\Psi\, e^{ip\omega} \qquad \text{if } j+m \text{ is odd.} \qquad (6.7b)$$

So, considering even and odd $j + m$ separately, one is led from Eq. (6.4) to two quantities,

$$\mathfrak{g}_1(M,p) = \frac{1}{\pi}\int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\Psi \cos M\Psi \, e^{-ip\omega} g(\omega,\Psi),$$

$$\mathfrak{g}_2(M,p) = \frac{i}{\pi}\int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\Psi \sin M\Psi \, e^{-ip\omega} g(\omega,\Psi).$$

The inversion formula (6.5) becomes

$$g(\omega,\Psi) = \frac{1}{2\pi}\int_0^{\infty} dM \int_{-M}^{M} dp \sum_{\rho=\pm 1} \mathfrak{g}_\rho(M,p) \left\{ \begin{array}{c} \cos M\Psi \\ -\sin M\Psi \end{array} \right\} e^{ip\omega}$$

(cos for $\rho = +1$, - sin for $\rho = -1$). Let

$$\mathfrak{g}(M,p,\sigma) = \frac{1}{2}[\mathfrak{g}_\sigma(M,p) + \sigma\,\mathfrak{g}_{-\sigma}(M,p)]$$

as in Eqs. (4.9). Then

$$\mathfrak{g}(M,p,\sigma) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\Psi \, e^{i\sigma M\Psi} e^{-ip\omega} g(\omega,\Psi), \tag{6.8}$$

$$g(\omega,\Psi) = \frac{1}{2\pi}\int_0^{\infty} dM \int_{-M}^{M} dp \, e^{ip\omega} [e^{-iM\Psi}\mathfrak{g}(M,p,1) + e^{iM\Psi}\mathfrak{g}(M,p,-1)].$$

By now it should be obvious that our destination is the ordinary Fourier transform in the Euclidean plane. If a more uniform asymptotic expression had been used in place of Eq. (6.6), there would have been two essential changes in Eqs. (6.8): the M in $\exp(-iM\Psi)$ would be replaced by $|q| = \sqrt{M^2 - p^2}$, and a factor of $1/\sqrt{|q|}$ would appear in each equation. Then the standard Fourier transform is recovered by changing variables from (M,σ) to q (with sgn $q = -\sigma$) and adjusting the normalization of $\tilde{g}$ by a factor of $\sqrt{|q|}$ (cf. Eq. (4.5)).

The most important point of this exercise is that in the contraction of the quasiregular representation the distinction between indices of different parity which was introduced by fiat in Eqs. (4.8) _arises naturally from the behavior of the basis functions in the neighborhood of the point of contraction_. Also, the choice of relative phases of the basis vectors (6.3) is crucial. The phases of adjacent (in j and m) functions of the same parity of m must be coherent near the point of contraction in order for the integral transform to make sense in the limit of large R, when the variables become continuous. These same functions will not have coherent phases with respect to any other point except the antipodal point; this is consistent with the results of attempts at formal contraction of Lie algebra representations. Although the discussion in these last two sections has not been very precise as to the nature of the limits taken, it does indicate that the seemingly arbitrary elements introduced in formal discussions of contraction of irreducible group representations have perfectly clear geometrical counterparts in the structure of representations by functions on homogeneous spaces.

Appendix D

TYPES OF METRICS AND COORDINATE SYSTEMS

In Chapter III and later reference is often made to coordinate systems in which the explicit form of the metric tensor of space-time has certain convenient properties, and to special classes of metrics which take on especially simple forms in certain coordinate systems. In this appendix some terminology is introduced, which is partly standard and partly idiosyncratic. The approach to normal and Fermi coordinates via a polar form is unconventional, but it brings out the geometrical motivation behind the constructions.

We are considering Riemannian manifolds (see footnote 1 of Chapter III) of dimension s + 1 and signature (+ - ... -) (s minus signs). We always consider coordinate systems in which one coordinate, x^0, is timelike[1] and the others are spacelike. Thus, in a system in which the mixed time-space components are zero, the metric will have the form

$$ds^2 = g_{00} (dx^0)^2 + g_{jk} dx^j dx^k , \qquad (1)$$

where $g_{00} > 0$ and $\{g_{jk}\}$ is a negative definite matrix at each

[1] That is, all the tangent vectors to the hypersurfaces $\{x \mid x^0 = \text{const.}\}$ are spacelike.

point. Our coordinate systems are not required to cover the entire space; the desirable properties imposed locally may force coordinate singularities to develop which mark a natural boundary to the region covered.

It is well known that a point and a direction through it (the latter specified, for example, by a vector of length 1 in the timelike case) uniquely determine a geodesic (curve) through the point in the given direction. A geodesic hypersurface is defined similarly. For example, consider the family of geodesics generated by all the spacelike vectors normal to a given timelike vector at a given point P. The set of points obtained in this way is an s-dimensional spacelike hypersurface, which we shall call a geodesic hypersurface relative to P. (The hypersurface is not necessarily geodesic relative to any other point in it, if s > 1. For instance, a surface defined by t = const. in the three-dimensional metric

$$ds^2 = dt^2 - dr^2 - r^2 f(t)d\theta^2$$

is generated by the geodesics θ = const. through the point r = 0, but these are the only geodesics which lie entirely in the surface. In de Sitter space, however, because of the symmetry, a geodesic hypersurface is geodesic relative to all its points -- see Sec. III.1.)

The construction which is about to be described is most easily visualized in a space with definite metric, such as the geodesic hypersurface just discussed. Given a distinguished point P, we define the radial coordinate $x^1 \equiv r$ of each other point Q as the geodesic distance from P to Q; that is, the arc length of the segment of geodesic joining Q to P (which is unique, at least locally). To complete the coordinate system we assign the same "angular" coordinates $x^2, \ldots, x^s$ to all the points on a given geodesic through P. Now the geodesics are in one-to-one correspondence to their unit tangent vectors at P. Furthermore, the geometry of the space of tangent vectors is Euclidean, and so, given an orthonormal basis in the tangent space, one can assign angular variables to the tangent vectors in a standard way (e.g., spherical coordinates). We call these <u>normal coordinates in polar form</u>.

Using the standard formulas relating spherical and Cartesian coordinates, one can pass to a <u>quasi-Cartesian</u> system with origin at P corresponding to the orthonormal basis chosen. At a finite distance from P these coordinates $(y^1, \ldots, y^s)$ will not be orthogonal, in general. (That is, the metric will contain terms in $dy^1 dy^2$ and so on.) It is coordinates of this type which are usually called <u>normal</u> (cf. [Synge], pp. 76-77).

In a space with indefinite metric an analogous construction can be carried out. Instead of spherical one will use "hyperboloidal" coordinates, related to Cartesian coordinates by formulas involving hyperbolic functions (cf. Secs. IX.1 and X.2). The polar form of this normal coordinate system is not very useful, because it is singular not only at the origin but over the entire light cone of the origin.

Gaussian coordinates ([Adler-Bazin-Schiffer], pp. 59-62) are associated with a given spacelike hypersurface S (an s-dimensional submanifold with a timelike normal vector at each point) and a given coordinate system $(x^1,\ldots,x^s)$ on S. The curve of points in space-time whose spatial coordinates are $(x^1,\ldots,x^s)$ is defined to be the geodesic through the point $(x^1,\ldots,x^s)$ in S in the direction normal to S. The time coordinate of a point on one of these geodesics is (up to sign) its geodesic distance from S. Then it can easily be shown that the metric has the form

$$ds^2 = (dx^0)^2 + g_{jk}\, dx^j dx^k , \tag{2}$$

where g_{jk} may be a function of x^0 as well as the x^i. Note that the hypersurfaces $x^0 = $ *const.* are not, in general, geodesic hypersurfaces, even if S is geodesic.

Generalized Fermi coordinates are associated with a given timelike curve C. The hypersurfaces of constant time are the geodesic hypersurfaces normal to C at each point. (Any monotonic parametrization of C can provide the numerical value of the time coordinate.) In each of these hypersurfaces we choose a system of normal coordinates (polar or quasi-Cartesian). In the polar case the metric takes the form

$$ds^2 = g_{00}(dx^0)^2 + 2\sum_{j=2}^{s} g_{0j}\,dx^0 dx^j$$

$$- dr^2 + \sum_{j,k=2}^{s} g_{jk}\,dx^j dx^k . \quad (3)$$

(The coefficients may be functions of any of the coordinate variables.) The metric has, relative to any hypersurface r = const., the Gaussian form (2) (generalized in an obvious way to hypersurfaces with spacelike normals).

In Fermi coordinates[2], properly so called, the angular (or quasi-Cartesian) coordinates at each time are determined by those at an initial time. Fermi defined an angle-preserving mapping of the unit normals attached to one point of C to the unit normals at each other point of C (see [Synge], pp. 12-15). Fermi's transport law gives a definite meaning to the intuitive requirement that the coordinate axes should not rotate.

A further specialization is to geodesic Fermi coordinates[3], where the curve C is a geodesic and x^0 is the arc length along it. Fermi's transport law reduces in this case to ordinary parallel transport. In analogy to the situation with Gaussian coordinates, it should be noted that the curves r = const., x^2 = const., ..., x^s = const. (other than C itself)

[2] [Synge], pp. 83-85; Schild (1965), pp. 54-55. These are Schild's "Fermi coordinates of the second kind".
[3] Schild (1965), p. 55; Manasse and Misner (1963). Manasse and Misner call these Fermi normal coordinates.

are not generally geodesics.

Analogously, we define a <u>geodesic Gaussian coordinate system</u> as a Gaussian system such that

(1) the initial hypersurface S is a geodesic hypersurface;

(2) the coordinate system given <u>a priori</u> in S is a normal system.

Geodesic Gaussian and geodesic Fermi coordinate systems are very natural for physical applications. We may think of the (instantaneous) vantage-point of an observer as being represented by his position (a point P in space-time) and his velocity (a timelike unit vector v). It is natural for this observer to think of the geodesic C generated by v as the time axis ("here") and the geodesic hypersurface S normal to it at P as "now". These identifications are consistent with a normal coordinate system based on P, a geodesic Gaussian system based on S, or a geodesic Fermi system based on C (although in general these systems will not coincide elsewhere). Of course, to define each of these systems uniquely requires specifying a complete orthonormal set of spacelike vectors normal to v. The normal system is the natural extension to a finite region of the "local Lorentz frame" determined infinitesimally at P by v. However, in the context of a theory which depends heavily on a distinction between space and time, such as quantum mechanics in a Hamiltonian formulation, the Gaussian and Fermi systems may be expected to play roles at least as important. This subject is discussed further in Secs. III.3 and X.8.

So far we have discussed forms into which any metric can be cast by a proper choice of coordinate system. We turn now to two forms which put nontrivial restrictions on the metric. Thus they define intrinsic properties of the metric itself. The importance of these classes of metrics for us is that for them the scalar wave equation can be solved by separation of variables -- see Chapters V, VIII, and X.

A <u>static metric</u>, as its name implies, is independent of time; it is also required that

$$g_{0k} = 0, \qquad k = 1, \ldots, s. \tag{4}$$

Thus

$$ds^2 = g_{00} (dx^0)^2 + g_{ij}\, dx^j dx^k \tag{5}$$

with the coefficients functions of $x^1, \ldots, x^s$ only.

We call a <u>generalized Robertson-Walker metric</u> any metric of the form

$$ds^2 = (dx^0)^2 - R^2(x^0)\, h_{jk}\, dx^j dx^k, \tag{6}$$

where the h_{jk} are functions of $x^1, \ldots, x^s$ alone. This is a special kind of Gaussian metric, describing a universe which may expand or contract but does not change its "shape".

A <u>Robertson-Walker metric</u>[4] in the strict sense is required to be homogeneous and isotropic at each time. This is a restriction on the s-tensor h_{jk}, or the manifold it describes. In the case s = 3 there are three classic possibilities: Euclidean space, the three-sphere, and the three-dimensional analogue of Lobachevsky space.

In their textbook Robertson and Noonan have listed all the four-dimensional <u>Robertson-Walker universes of constant curvature</u>; there are six.[5] Models with different Robertson-Walker coordinate systems are regarded as distinct, even if (at least locally) they have the same four-dimensional geometry. In this dissertation two-dimensional analogues of all six of these models will be encountered:[6]

(1) Ordinary Minkowski universe.

(2) Expanding Minkowski universe: Sec. X.2.

(3) De Sitter universe, proper: Sec. III.7 (de Sitter space in horospherical coordinates).

(4) Lanczos universe: Sec. III.1 (de Sitter space in geodesic Gaussian coordinates).

---

[4] [Adler-Bazin-Schiffer], pp. 338-349.
[5] [Robertson-Noonan], pp. 362-371. See also pp. 335-348.
[6] The number and the name given first are those of [Robertson-Noonan].

(5) : Sec. III.2 (a portion of de Sitter space in polar normal coordinates).

(6) : Sec. III.6 (a portion of open de Sitter space).

Appendix E

THE AXIOMS OF RELATIVISTIC QUANTUM FIELD THEORY

The general principles of quantum field theory stated, for instance, in [Streater-Wightman], pp. 96-102 and 29-30, or in Wightman and Gårding (1965), involve two basic mathematical elements: the operator fields defined on space-time and the representation of the Poincaré group. The second of these is not available in the theory of quantized fields coupled to fixed external (c-number) fields, including the theory of fields in curved space-time, where the curvature can be regarded as an external gravitational field (see Sec. VII.7). (In the latter case the structure of space-time is changed, and some corresponding minor changes in the notion of local field operators are needed in addition to some compensation for the loss of Poincaré invariance.) In the de Sitter spaces one might expect the role of the Poincaré group to be taken over by the appropriate de Sitter group, but the results of the present work tend to a contrary conclusion.

In discussing these problems in Chapters IV and VII, therefore, it will be helpful to refer to a version of the axioms in which the roles of group and field are clearly separated. In this appendix the axioms are divided into ten statements which, as indicated in Fig. 20, fall into four classes depending on whether they deal with neither, one or the other, or both of these elements. An axiom which comes below another in the graph of Fig. 20 either implies it and renders it redundant (as in the case of Axiom 8) or tacitly assumes it (e.g., Axiom 9).

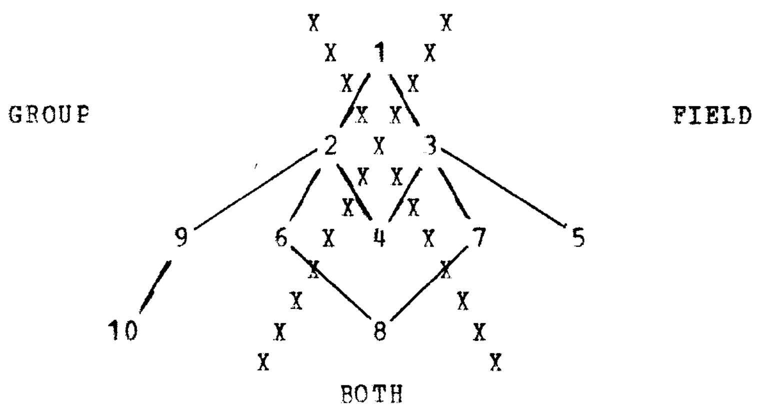


Fig. 20

Logical interdependence of the axioms

These are the axioms:

1. Quantum theory. The states of the theory are described by unit rays in a separable Hilbert space $\mathcal{H}$.

2. Relativistic invariance. The relativistic transformation law of the states is given by a continuous unitary representation of ISL(2,$\mathbb{C}$), the universal covering group of the Poincaré group: $\{a,A\} \longrightarrow U(a,A)$.

3. Existence and temperedness of the fields. For each test function $f \in \mathcal{S}$ (the space of smooth functions of rapid decrease

-- see [Streater-Wightman], Sec. 2.1) defined on space-time there exists a set $\phi_1(f), \ldots, \phi_n(f)$ of operators. These operators, together with their adjoints $\phi_1(f)^\dagger, \ldots, \phi_n(f)^\dagger$, are defined on a linear domain D of vectors, dense in $\mathcal{H}$. The $\phi_j(f)$ and $\phi_j(f)^\dagger$ leave D invariant. If $\Phi, \Psi \in D$, then $(\Phi, \phi_j(f)\Psi)$ is a tempered distribution, regarded as a functional of f.

4. Tensorial character of the fields. The $U(a,A)$ leave D invariant, and the equation

$$U(a,A)\, \phi_j(f)\, U(a,A)^{-1} = \sum S_{jk}(A^{-1})\, \phi_k(\{a,A\}f)$$

is valid when each side is applied to any vector in D. Here S is a representation of SL(2,$\mathbb{C}$), and

$$\{a,A\}f(x) = f(A^{-1}(x - a)).$$

5. Local commutativity. If the support of f and the support of g are spacelike separated, then one or the other of

$$[\phi_j(f), \phi_k^{(\dagger)}(g)]_{\pm} = 0$$

(anticommutator or commutator) holds for all j and k when the left-hand side is applied to any vector in D.

6. Existence and uniqueness of the vacuum. There is a state Ψ_0, the vacuum, invariant under U, unique up to a phase

factor.

7. Cyclicity of the fields. There is a state which is cyclic for the smeared fields; that is, polynomials in the smeared field components, $P(\phi_1(f), \phi_2^{(+)}(g), \ldots)$, applied to this state yield a set D_0 of vectors dense in $\mathcal{H}$.

8. Cyclicity of the vacuum. Ψ_0 is in D and is cyclic.

9. Spectral condition. The eigenvalues of P^μ lie in or on the plus cone (i.e., $P^\mu P_\mu \geq 0$), where $U(a, 1) = \exp(iP_\mu a^\mu)$.

10. Asymptotic completeness. The decomposition of U into irreducible representations is one appropriate to a theory of noninteracting particles of various masses and spins. In fact, the states (rays in $\mathcal{H}$) are in correspondence with all the possible incoming (alternatively, outgoing) configurations of the stable particles described by the theory. (A more explicit formulation of this axiom would take too much space here. See Haag (1955) (Sec. I) or the Wightman references above.)

Appendix F

REPRESENTATIONS OF THE CANONICAL COMMUTATION RELATIONS

1. Definitions.

The formal structure of the canonical commutation relations (CCRs), Eqs. (VII.2.1) or (VIII.2.5), can be treated rigorously either in terms of an algebra of bounded operators satisfying the so-called Weyl relations (VIII.3.10), or in terms of an algebra of unbounded operators satisfying the "naive" CCRs on a common invariant domain (cf. Powers (1971)). Reed (1969) has proved that every representation of the first type is associated with one of the second type, but a famous example of E. Nelson (see [Reed-Simon], Sec. 8.5) shows that the converse if false, even for one degree of freedom and even if the field operators are required to be self-adjoint on a common dense invariant domain. The first approach facilitates the proof of abstract theorems, but the second is more intuitive and more convenient for concrete calculations. Here we shall be primarily concerned with the Fock representation and representations defined from it by a certain kind of transformation, so a formulation of the second type is sufficient. Also, for present purposes it will not be necessary to impose any condition of continuity in the test function.

Let $\mathcal{H}$ (the one-particle space) be a complex Hilbert space with an involution operation, $f \longrightarrow \bar{f}$. Without loss of generality we shall take $\mathcal{H}$ to be $L^2(M,\mu)$, where (M,μ) is a measure space and the involution is complex conjugation. Let $\mathcal{T}$ (the <u>test function space</u>) be a dense subspace of $\mathcal{H}$. Let $\mathcal{F}$ be a Hilbert space and $\mathcal{D}$ a dense subspace of $\mathcal{F}$. Let $a(\ )$ be a linear map of $\mathcal{T}$ into the (unbounded) closable linear operators on $\mathcal{F}$ with domain $\mathcal{D}$, such that

(1) the $a(f)$ $(f \in \mathcal{T})$ and their adjoints $a^\dagger(f) \equiv a(\bar{f})^\dagger$ leave $\mathcal{D}$ invariant;

(2) $[a(f), a(g)] = 0$ and

$$[a(f), a^\dagger(g)] = \int d\mu(x)\ f(x)g(x) = (\bar{g},f) = (\bar{f},g) \qquad (1.1)$$

on $\mathcal{D}$.

Such a system will be called a <u>representation of the canonical commutation relations</u>. Formally we may write[1]

$$a(f) = \int a(x) f(x)\ d\mu(x), \qquad (1.2)$$

$$[a(x), a^\dagger(y)] = \delta(x - y). \qquad (1.3)$$

In particular, the <u>Fock representation</u> over $\mathcal{H}$ is defined by taking (cf. Sec. VIII.3)

---

[1] One may call $a(x)$ an "operator-valued distribution", although no continuity condition in terms of a topology on $\mathcal{T}$ has been stated. The delta function in Eq. (1.1) has meaning as a bilinear form on $\mathcal{H}$.

$\mathcal{T} = \mathcal{H}$;

$\mathcal{N}$ = space of finite sequences of the form

$$\Psi = \{\Psi_0, \Psi_1(x_1), \Psi_2(x_1,x_2), \ldots, \Psi_N(x_1,\ldots,x_N), 0, \ldots\}$$

with scalar product given by $\|\Psi\|^2 = \sum_{n=0}^{N} \|\Psi_n\|^2$, where Ψ_n is in $\mathcal{H}^{\odot n}$, the Hilbert-space completion of the symmetrized n-fold tensor product of $\mathcal{H}$;

$\mathcal{F}$ = completion of $\mathcal{N}$, with typical member

$$\Psi = \{\Psi_0, \Psi_1(x_1), \ldots, \Psi_n(x_1,\ldots,x_n), \ldots\} \equiv \{\Psi_n(x_1,\ldots,x_n)\};$$

$$a(f)\ \Psi = \{\sqrt{n+1} \int d\mu(x)\ f(x)\ \Psi_{n+1}(x,x_1,x_2,\ldots,x_n)\}, \tag{1.4a}$$

$$a^\dagger(f)\ \Psi = \{\sqrt{n}\ \mathrm{sym}\ f(x_1)\Psi_{n-1}(x_2,\ldots,x_n)\}. \tag{1.4b}$$

In the last equation sym denotes the symmetrizer

$$\mathrm{sym}\ \phi(x_1,\ldots,x_n) = \frac{1}{n!} \sum \phi(x_{i_1},\ldots,x_{i_n}) \tag{1.5}$$

(sum over all permutations); in the present case it reduces to $\frac{1}{n}\sum_{j=1}^{n} f(x_j)\Psi_{n-1}(x_1,\ldots,\hat{x}_j,\ldots,x_n)\}$. The vacuum sequence $\{0, 0, \ldots\}$ is denoted by $|0\rangle$.

These definitions can be formulated abstractly, without reference to a concrete L^2 realization or even a distinguished involution. If the involution is abandoned, however, it is best to take the map $a^\dagger(\ )$ as the basic object. Then the commutation relations are

$$[a^\dagger(f),\ a^\dagger(g)] = 0, \qquad [(a^\dagger(f))^\dagger,\ a^\dagger(g)] = (f,g), \tag{1.6}$$

and the Fock representation has the abstract characterization

$$a^\dagger(f)\,\Psi = a^\dagger(f)\{\Psi_n\} = \{\sqrt{n}\ \mathrm{sym}\ (f \otimes \Psi_{n-1})\}, \tag{1.7a}$$

$$(a^\dagger(f))^\dagger\,\Psi = \{\sqrt{n+1}\ (f,\Psi_{n+1})\}, \tag{1.7b}$$

where $(f,\phi_1\otimes\phi_2\otimes\ldots\otimes\phi_n) = (f,\phi_1)\,\phi_2\otimes\ldots\otimes\phi_n$, <u>sym</u> is defined on vectors of the form $\phi_1\otimes\ldots\otimes\phi_n$ in the obvious way, and these operations are extended to all of $\mathcal{H}^{\otimes n}$ (and hence $\mathcal{H}^{\odot n}$) by linearity and continuity. Now for <u>each</u> realization of $\mathcal{H}$ as an L^2 space one can define $a(f) = a^\dagger(\bar{f})^\dagger$ and recover the commutation relations and Fock representation formulas (1.1) and (1.4). Note that the meaning of $a(f)$, unlike that of $a^\dagger(f)$, depends on the realization (more precisely, on the involution).[2]

Hermitian operators may be formed from these annihilation and creation operators in two ways. The first way,

---

[2] In physical terms, the decision to take $a^\dagger(\ )$ rather than $a(\ )$ as the realization-independent object is forced by the demand that the realization-independent description of a one-particle state be linear in the wave function ($a^\dagger(f)|0\rangle$), not antilinear ($a(f)^\dagger|0\rangle$).

which is how field operators are usually related to annihilation and creation operators in quantum field theory, is realization-dependent, but the fields depend linearly on the complex function f:

$$\phi(f) = \frac{1}{\sqrt{2}}\,(a(f) + a^{\dagger}(f)), \quad \pi(f) = \frac{-i}{\sqrt{2}}\,(a(f) - a^{\dagger}(f));$$

$$[\phi(f),\, \pi(g)] = i(\bar{f},g); \qquad (1.8)$$

$$\phi(f)^{\dagger} = \phi(\bar{f}), \quad \phi(if) = i\phi(f), \quad \text{etc.}$$

The other way is independent of realization, but the operators are only <u>real</u> linear in f (and always Hermitian):

$$Q(f) = \frac{1}{\sqrt{2}}\,(a^{\dagger}(f)^{\dagger} + a^{\dagger}(f)), \quad P(f) = \frac{-i}{\sqrt{2}}\,(a^{\dagger}(f)^{\dagger} - a^{\dagger}(f));$$

$$[Q(f),\, P(g)] = i(f,g); \qquad (1.9)$$

$$Q(f)^{\dagger} = Q(f), \quad Q(if) = P(f), \quad \text{etc.}$$

If M is chosen to be a set of discrete points -- this amounts to choosing an orthonormal basis in $\mathcal{H}$ -- the CCR algebra (1.3) is that appropriate to a collection of one-dimensional harmonic oscillators (cf. [Messiah], Chap. XII). A <u>tensor-</u>

product representation[3] is constructed by considering a formal product vector

$$\Psi = \bigotimes_{x=1}^{\infty} \Psi_x ,$$

where each Ψ_x is some vector in the state space of the oscillator with index x, and generating a Hilbert space by acting on Ψ with all the elements of the algebra in the obvious way. (Different product vectors can yield unitarily equivalent representations. This happens if and only if the vectors are "weakly equivalent" -- see the references.) The Fock representation is the tensor-product representation in which each Ψ_x is the ground state (annihilated by a_x).

2. Bogolubov Transformations.

Consider a representation of the CCRs as defined above, and let $\mathcal{H}'$ be another L^2 space. (If $\mathcal{H}$ is regarded abstractly, $\mathcal{H}'$ may be $\mathcal{H}$ itself in a different realization. In this case it is important to remember that the involution $\bar{f}$ will depend, in general, on whether f is regarded as a member of $\mathcal{H}$ or of $\mathcal{H}'$.) Let U and V be operators from $\mathcal{H}$ to $\mathcal{H}'$. For the moment we assume U and V to be bounded (and defined everywhere). Then $U^\dagger$ and $V^\dagger$ are bounded operators from $\mathcal{H}'$ to $\mathcal{H}$. Also, we define the complex conjugate $\bar{U}$: $\mathcal{H} \longrightarrow \mathcal{H}'$ and the transpose U^T: $\mathcal{H}' \longrightarrow \mathcal{H}$

[3] Klauder et al. (1966); Streit (1967); Reed (1968, 1970). The fundamental paper on infinite tensor products is von Neumann (1938).

of an operator by

$$\overline{U}g = \overline{U\overline{g}}, \tag{2.1}$$

$$U^T = \overline{U}^{\dagger} = \overline{U^{\dagger}}. \tag{2.2}$$

Under these conditions

$$b(g) = a(U^T g) + a^{\dagger}(V^T g) \tag{2.3}$$

is an operator on $\mathcal{F}$ if $U^T g$ and $V^T g$ are both in $\mathcal{F}$, and the corresponding adjoint is

$$b^{\dagger}(g) \equiv b(\overline{g})^{\dagger} = a^{\dagger}(U^{\dagger} g) + a(V^{\dagger} g). \tag{2.4}$$

By direct calculation we find that b( ) is a representation of the CCRs on $\mathcal{F}$ with test function space $U^{T-1}(\mathcal{F}) \cap V^{T-1}(\mathcal{F})$ (which we assume dense in $\mathcal{H}'$) if and only if

$$UU^{\dagger} = 1 + VV^{\dagger} \tag{2.5}$$

and

$$UV^T = VU^T. \tag{2.6}$$

The transformation (2.3) is called a <u>Bogolubov transformation</u>, in reference to an application in the theory of superconductivity.

Suppose that, in addition, the a's can be re-expressed in terms of the b's, at least for a dense set of f's in $\mathcal{H}$:

$$a(f) = b(X^T f) + b^\dagger(Y^T f). \tag{2.7}$$

Then

$$a(f) = a((U^T X^T + V^\dagger Y^T)f) + a^\dagger((V^T X^T + U^\dagger Y^T)f).$$

It is easy to see that necessarily

$$V^T X^T + U^\dagger Y^T = 0, \quad U^T X^T + V^\dagger Y^T = 1.$$

So (using Eq. (2.6)) we have $-UU^\dagger Y^T = UV^T X^T = VU^T X^T = V - VV^\dagger Y^T$, and hence (using Eq. (2.5)) $Y^T = -V$. Similarly, using the complex conjugates of Eqs. (2.5-6), we have $-\bar{V}V^T X^T = \bar{U} - \bar{U}U^T X^T$ and hence $X^T = \bar{U}$. So Eq. (2.7) becomes

$$a(f) = b(\bar{U}f) - b^\dagger(Vf), \tag{2.8a}$$

$$a^\dagger(f) = b^\dagger(Uf) - b(\bar{V}f). \tag{2.8b}$$

On the other hand, Eq. (2.7) must be a Bogolubov transformation itself, so Eqs. (2.5-6) applied to Eqs. (2.8) yield

$$U^\dagger U = 1 + V^T \bar{V}, \tag{2.9}$$

$$U^\dagger V = V^T \bar{U}. \tag{2.10}$$

Eqs. (2.9-10) are necessary and sufficient for the invertibility of the transformation (2.4).

3. <u>The Theorem on Equivalence of Representations Related by Bogolubov Transformations</u>.[4]

Let $\{a_k\}$ be a system of annihilation operators in the Fock representation on a Hilbert space $\mathcal{F}$, and $\{b_j\}$ another set of operators which also satisfy the CCRs and are related to the a's and $a^\dagger$'s by a linear transformation, which may be schematically indicated by

$$b_j = \sum_k [U_{jk} a_k + V_{jk} a_k^\dagger]. \tag{3.1}$$

Then, roughly speaking, the representation of the b's in $\mathcal{F}$ is the Fock representation if and only if the "matrix" V_{jk} is Hilbert-Schmidt:

$$\sum_j \sum_k |V_{jk}|^2 < \infty. \tag{3.2}$$

This condition is quite reasonable, since the expression on the left-hand side of the inequality is the expectation value in the a-vacuum of the total b-number operator $\sum b_j^\dagger b_j$. In general, j and k are continuous variables, and the sums in the condition (3.2)

[4] [Friedrichs], Part V; Shale (1962); [Berezin], Chapter II; Kristensen *et al.* (1967). The present exposition follows the last two references, but generalizes them by allowing different realizations $L^2(M,\mu)$ of the one-particle Hilbert space at the two ends of the transformation.

are to be interpreted as integrals.

In practice the kernels U_{jk} and V_{jk} usually are obtained by some formal calculation (cf. Secs. IX.3 and X.2), and their mathematical status may be dubious. Unfortunately, the criterion stated above has been proved rigorously only when these kernels define bounded operators, U and V, in a Hilbert space. Then we have a Bogolubov transformation in the sense of Sec. F.2. (Eq. (2.3) is Eq. (3.1) smeared with a test function.) Moreover, the proof assumes that the transformation is invertible (i.e., Eq. (3.1) can be solved for a_k, or Eqs. (2.7-10) hold).

<u>Theorem</u>: Assume the following:

(1) a( ) is a Fock representation with one-particle space $\mathcal{H}$ and Fock space $\mathcal{F}$.

(2) U and V are bounded operators from $\mathcal{H}$ to another Hilbert space (with involution) $\mathcal{H}'$.

(3) $$UU^{\dagger} = 1 + VV^{\dagger} \quad \text{and} \quad UV^{T} = VU^{T} \tag{3.3}$$

(so that $b(g) = a(U^{T}g) + a^{\dagger}(V^{T}g)$ defines a representation in $\mathcal{F}$ of the CCRs with one-particle space $\mathcal{H}'$).

(4) $$U^{\dagger}U = 1 + V^{T}\overline{V} \quad \text{and} \quad U^{\dagger}V = V^{T}\overline{U} \tag{3.4}$$

(so that a( ) can be expressed in terms of b( ): $a(f) = b(\overline{U}f) - b^{\dagger}(Vf)$).

Then the condition

(5) V is a Hilbert-Schmidt operator

is necessary and sufficient for the conclusion:

There is a vector $\Psi_{(0)} \in \mathcal{F}$ such that $b(g)\Psi_{(0)} = 0$ for all $g \in \mathcal{H}'$.

In this case the representation $b(\ )$ is unitarily equivalent to a Fock representation.

<u>Proof</u>:

(1) U <u>has a bounded inverse</u>: U and $U^\dagger$ have the polar decompositions

$$U = (UU^\dagger)^{1/2} U_0 , \qquad U^\dagger = U_0^\dagger (UU^\dagger)^{1/2} \tag{3.5}$$

where U_0 is a partial isometry ([Kato], pp. 334-335). The kernel of U_0 is the kernel of $U^\dagger U$, and the range of U_0 is the closure of the range of U; similarly, the kernel of $U_0^\dagger$ is that of $UU^\dagger$, and its range is the closure of the range of $U^\dagger$.

On the other hand, the equation $UU^\dagger = 1 + VV^\dagger$ shows that $UU^\dagger$ is a (strictly) positive definite self-adjoint operator. There follows: (a) $(UU^\dagger)^{-1/2}$ exists and is bounded; (b) $U_0^\dagger$ is injective, since $UU^\dagger$ is. The same reasoning starting from $U^\dagger U = 1 + \bar{V}^\dagger \bar{V}$ shows that U_0 is injective. So U_0 is actually a unitary operator, and $U_0^\dagger = U_0^{-1}$. Consequently, U has the bounded inverse $U^{-1} = U_0^{-1}(UU^\dagger)^{-1/2}$. In passing we have proved that the ranges of U and $U^\dagger$ are dense.

(2) The operator $\mathcal{Y}$: Define

$$\mathcal{Y} = -U^{-1}V. \tag{3.6}$$

Since U^{-1} is bounded, $\mathcal{Y}$ is Hilbert-Schmidt if and only if V is. Since $UV^T = VU^T$, one has $\mathcal{Y}^T = \mathcal{Y}$.

(3) $\|\mathcal{Y}\| < 1$:
$$U^\dagger U(1 - \mathcal{Y}\mathcal{Y}^\dagger) = U^\dagger U(1 - \mathcal{Y}\bar{\mathcal{Y}})$$

$$= U^\dagger U - U^\dagger UU^{-1}V\bar{U}^{-1}\bar{V} = U^\dagger U - U^\dagger V\bar{U}^{-1}\bar{V} = U^\dagger U - V^T\bar{U}\bar{U}^{-1}\bar{V}$$

$$= U^\dagger U - V^T\bar{V} = 1.$$

(Eqs. (2.9-10) have been used.) Since we know $U^\dagger U$ has an inverse, this shows that $1 - \mathcal{Y}\mathcal{Y}^\dagger = (U^\dagger U)^{-1}$, or $\mathcal{Y}\mathcal{Y}^\dagger = 1 - U^{-1}U^{\dagger -1}$. Thus for all $f \in \mathcal{H}$

$$\|\mathcal{Y}^\dagger f\|^2 = \|f\|^2 - \|U^{\dagger -1} f\|^2 < \|f\|^2.$$

(The inequality is strict since $U^{\dagger -1}$ is injective.) Therefore, the operator norm of $\mathcal{Y}$ is

$$\|\mathcal{Y}\|_{op} = \|\mathcal{Y}^\dagger\|_{op} < 1.$$

(4) Solution for $\Psi_{(0)}$: Consider the equation

$b(g)\Psi_{(0)} = 0$, where

$$\Psi_{(0)} = \{\Psi_n(x_1,\ldots,x_n)\}. \tag{3.7}$$

(We are using an explicit realization of $\mathcal{H}$ as an L^2 space.) Written out in terms of the definitions (2.3) and (1.4), the equation is

$$\{\sqrt{n+1}\int d\mu(x)\ (U^T g)(x)\ \Psi_{n+1}(x,x_1,\ldots,x_n)\}$$

$$= -\{\sqrt{n}\ \mathrm{sym}\ (V^T g)(x_1)\ \Psi_{n-1}(x_2,\ldots,x_n)\}. \tag{3.8}$$

Let us consider the various component equations.

$$n = 0: \qquad \int d\mu(x)\ (U^T g)(x)\ \Psi_1(x) = 0.$$

If this is to hold for all g, we must have $\Psi_1(x) = 0$, since U^T $(= \overline{U^+})$ has dense range.

$$n = 1: \quad \sqrt{2}\int d\mu(x)\ (U^T g)(x)\ \Psi_2(x,x_1) = -(V^T g)(x_1)\ \Psi_0.$$

Substitute $U^{-1T}g$ for g:

$$\sqrt{2}\int d\mu(x)\ g(x)\ \Psi_2(x,x_1) = (\gamma^T g)(x_1)\ \Psi_0.$$

If $\Psi_0 \neq 0$, this equation states that γ is an integral operator with the symmetric kernel

$$\gamma(x_1,x_2) = \sqrt{2}\,\Psi_0^{-1}\,\Psi_2(x_1,x_2) \in L^2(dx_1 dx_2). \qquad (3.9)$$

This is equivalent to the Hilbert-Schmidt and symmetry properties of the operator γ.

General n: Ψ_{n+1} is related homogeneously to Ψ_{n-1}. By induction, $\Psi_n = 0$ for all odd n. It follows that $\Psi_0 \neq 0$ for nonzero $\Psi_{(0)}$; let us set $\Psi_0 = 1$. For even n we obtain

$$\sqrt{n+2}\int d\mu(x)\; g(x)\; \Psi_{n+2}(x,x_1,\ldots,x_{n+1}) =$$

$$\sqrt{n+1}\;\mathrm{sym}\;(\gamma^T g)(x_1)\;\Psi_n(x_2,\ldots,x_{n+1}),$$

or

$$\Psi_{n+2}(x_1,\ldots,x_{n+2}) =$$

$$\frac{1}{\sqrt{(n+1)(n+2)}}\sum_{p=2}^{n+2}\gamma(x_1,x_p)\;\Psi_n(x_2,\ldots,\hat{x}_p,\ldots,x_{n+2}). \qquad (3.10)$$

The solution of this recursion is (Kristensen et al. (1967))

$$\Psi_n(x_1,\ldots,x_n) = \sqrt{n!}\;2^{-n/2}\left[\left(\frac{n}{2}\right)!\right]^{-1}$$

$$\times\;\mathrm{sym}\,[\gamma(x_1,x_2)\gamma(x_3,x_4)\ldots\gamma(x_{n-1},x_n)] \qquad (3.11)$$

for n even, which is a symmetric L^2 function.

(5) $\Psi_{(0)}$ <u>is</u> <u>normalizable</u>: Kristensen <u>et</u> <u>al</u>. (1967) calculate that

$$\|\Psi_{(0)}\|^2 = \prod_j \left[\sum_{\nu=0}^{\infty} \binom{-1/2}{\nu} (-\gamma_j^2)^\nu \right], \tag{3.12}$$

where the γ_j are the eigenvalues of $(\gamma^\dagger\gamma)^{1/2}$. The maximum γ_j equals $\|\gamma\|_{op}$. If all γ_j are less than 1, the series converge, and

$$\|\Psi_{(0)}\|^2 = \prod_j (1 - \gamma_j^2)^{-1/2} < \infty .$$

Otherwise the expression is infinite. Consequently, $\|\gamma\|_{op} < 1$ is a necessary and sufficient condition for $\Psi_{(0)}$ to be a member of $\mathcal{F}$.

To summarize, it has been shown so far that the following are equivalent:

a. V is Hilbert-Schmidt.

b. γ is Hilbert-Schmidt and $\|\gamma\|_{op} < 1$.

c. There is a $\Psi_{(0)} \in \mathcal{F}$ such that $b(g)\Psi_{(0)} = 0$ for all $g \in \mathcal{H}'$.

When $\Psi_{(0)}$ exists, it is unique except for a constant factor.

(6) <u>The representation</u> b( ) <u>is Fock in the cyclic subspace generated by</u> $\Psi_{(0)}$: Let $\mathcal{F}_1$ be the closed linear span of the vectors of the form

$$\frac{1}{\sqrt{n!}}\, b^{\dagger}(g_1) \ldots b^{\dagger}(g_n)\, \Psi_{(0)} \qquad (n = 0, 1, \ldots). \qquad (3.13)$$

If such a vector is identified with the sequence $\{0, 0, \ldots, g_1(x_1)\ldots g_n(x_n), 0, \ldots\}$, the action of $b(g)$ and $b^{\dagger}(g)$ in $\mathcal{F}_1$ is seen to be that of the operators of the Fock representation.

(7) <u>The cyclic subspace is all of $\mathcal{F}$</u>: Repeating the entire argument with a and b interchanged, we find (since $-V^T$ is Hilbert-Schmidt) that there is a unique vector $|0\rangle \in \mathcal{F}_1$ which is annihilated by all the $a(f)$, and that the cyclic subspace generated by it is a subset of $\mathcal{F}_1$. Since this vector is necessarily the original Fock vacuum, $\mathcal{F} \subset \mathcal{F}_1$. Thus $\mathcal{F}_1 = \mathcal{F}$.

If $\mathcal{H}$ and $\mathcal{H}'$ are the same, the unitary equivalence of the two Fock representations is implemented by the unitary operator which maps each basis vector

$$\frac{1}{\sqrt{n!}}\, a^{\dagger}(g_1) \ldots a^{\dagger}(g_n)\, |0\rangle$$

into the corresponding vector (3.13).

4. <u>Diagonal Bogolubov Transformations</u>.

A special case of Eq. (3.1) (or (2.3)) is[5]

[5] In the applications in Chapter X (and below in this section), where k is a momentum variable, $a^{\dagger}_{-k}$ appears instead of $a^{\dagger}_{k}$. The extension of the following remarks to this case is easy.

$$b_k = \alpha^*(k) a_k + \beta(k) a_k^\dagger. \qquad (4.1)$$

The condition (2.5) becomes

$$|\alpha(k)|^2 - |\beta(k)|^2 = 1. \qquad (4.2)$$

(Eq. (2.6) is trivial here.) If k is a discrete variable, the representation of the b's in the Fock space of the a's is a tensor-product representation. The theorem of the last section certainly applies to the case (4.1). (The transformation has the inverse

$$a_k = \alpha(k) b_k - \beta(k) b_k^\dagger. \qquad (4.3)$$

If either operator (of multiplication by α or β) is unbounded, then β is not Hilbert-Schmidt; but in this case γ of Eq. (4.4) below is not normalizable, so the conclusion of the theorem holds.) The solution for the b-vacuum is given by Eq. (3.11) with

$$\gamma(k_1, k_2) = -\frac{\beta(k_1)}{\alpha^*(k_1)} \delta(k_1 - k_2). \qquad (4.4)$$

Note that if k is a continuous variable, γ is not normalizable, and hence the representations are inequivalent. The test for equivalence in the discrete case is

$$\sum |\beta(k)|^2 < \infty \tag{4.5}$$

(from Eq. (3.2)).

As an example of the application of the theorem let us compare the Fock representations corresponding to different choices of the mass for the scalar field in a box.

Write[6]

$$\phi(0,x) = \frac{1}{\sqrt{2}} \sum_k [k^2 + m^2]^{-1/4} [e^{ik\cdot x} a_k + e^{-ik\cdot x} a_k^\dagger],$$

$$\pi(0,x) = -\frac{i}{\sqrt{2}} \sum_k [k^2 + m^2]^{+1/4} [e^{ik\cdot x} a_k - e^{-ik\cdot x} a_k^\dagger], \tag{4.6}$$

and also

$$\phi(0,x) = \frac{1}{\sqrt{2}} \sum_p [p^2 + M^2]^{-1/4} [e^{ip\cdot x} b_p + e^{-ip\cdot x} b_p^\dagger],$$

$$\pi(0,x) = -\frac{i}{\sqrt{2}} \sum_p [p^2 + M^2]^{+1/4} [e^{ip\cdot x} b_p - e^{ip\cdot x} b_p^\dagger]. \tag{4.7}$$

(These equations are Fourier expansions of the canonical field operators at a fixed time; they _define_ the a's and b's. Such an expansion certainly makes sense (except possibly in a representation with a very unusual test function space) regardless of whether the mass parameter involved is related to the dynamics of the field, which has not been specified.) Invert

---

[6] Here x and k are s-dimensional vectors.

Eqs. (4.7) and substitute from Eqs. (4.6):

$$b_p = \frac{1}{2}\sum_k \{[\sqrt[4]{\frac{p^2+M^2}{k^2+m^2}} + \sqrt[4]{\frac{k^2+m^2}{p^2+M^2}}]\delta(p-k)a_k$$

$$+ [\sqrt[4]{\frac{p^2+M^2}{k^2+m^2}} - \sqrt[4]{\frac{k^2+m^2}{p^2+M^2}}]\delta(p+k)a_k^\dagger\}. \quad (4.8)$$

Thus

$$V(p,k) = \frac{1}{2}[\sqrt[4]{\frac{p^2+M^2}{k^2+m^2}} - \sqrt[4]{\frac{k^2+m^2}{p^2+M^2}}]\delta(p+k) \equiv \beta(p)\delta(p+k), \quad (4.9)$$

where at the last step one sets $k = -p$ in the first factor. The delta function is a Kronecker delta, since p and k are discrete variables ($=2\pi n/L$, where L is the length of the box).

We wish to know whether the Fock representation of the a operators, which is always used when the field is to satisfy the Klein-Gordon equation with mass m, is equivalent to the Fock representation of the b operators, appropriate to mass M. So let us apply the criterion (4.5). At large p we have

$$[\frac{p^2+M^2}{p^2+m^2}]^{1/4} = 1 + (M^2 - m^2)/4p^2$$

$$- (3M^4 + 2M^2m^2 - 5m^4)/32p^4 + \ldots,$$

$$\beta(p) = (M^2 - m^2)/4p^2 + O(p^{-4}),$$

$$\sum_p |\beta(p)|^2 = \text{finite term} + \frac{1}{16}(M^2 - m^2)^2 \sum_p p^{-4} . \qquad (4.10)$$

This sum converges if (and only if) $s \leq 3$; that is, in space-time of dimension 2, 3, or 4.

So in a finite flat space (torus) of physical dimension or smaller, the Fock representations for different masses are equivalent. In other words, the representation of the b's of Eqs. (4.7) in the usual Fock space for a free field of mass m ($\neq$ M) is just the Fock representation. Of course, the vacuum (better, no-particle) states and the rest of the particle structure are different. Thus each mass determines a virtual particle concept (see Sec. X.7); in the usual theory of the free field the one related to the mass which appears in the field equation or the Hamiltonian corresponds to real particles.

The sum (4.10) grows proportionally to the volume as the size of the box approaches infinity (since the points of the momentum lattice become denser). Hence the inequivalence of the representations in infinite space is not surprising, and is an infinite-volume effect. For $s > 3$, however, one has inequivalence even in a finite region -- an ultraviolet divergence.

The equivalence of the a and b representations can also be decided by determining whether the no-particle states are weakly equivalent in the sense of infinite tensor products (see Sec. F.1 and references cited there). The calculations necessary to test for weak equivalence are essentially identical to those of Jaffe (1965), pp. 197-200, for the slightly more complicated case of two quadratically coupled fields, although he does not use the language of tensor products. (See also Haag (1955), Sec. II.1.) One is led again to a sum in which the terms fall off as p^{-4}, so the conclusions are consistent.

Appendix G

PROOF OF ASSERTIONS CONCERNING THE VACUUM OF THE FREE FIELD IN A BOX

1. The Two-Point Function Depends on Global Boundary Conditions.

Let $G_L^{(+)}(x_2, x_1)$ be the two-point function of the free scalar field in a closed two-dimensional flat universe of length L (with spatial coordinate $-L/2 \leq x \leq L/2$). In Sec. IX.7 it is asserted that

$$E_L(f,g) \equiv \int d^2x_2 \int d^2x_1 \; f^*(x_2)\, g(x_1)\, G_L^{(+)}(x_2, x_1) \qquad (1.1)$$

is not equal to the analogous expression formed from the two-point function of the scalar field in infinite two-dimensional space, even when f(x) and g(x) have support in a causal diamond with base inside the interval $(-L/2, L/2)$. To establish this claim (for at least one L) it clearly suffices to prove that $E_L \neq E_{L'}$ when $L \neq L'$ (for at least one pair of L's). Since three-dimensional smearing is allowed in the standard Fock representations, we may set $t_1 = t_2 = 0$ and let x_1 and x_2 denote the one-dimensional space variables. Then

$$G_L^{(+)}(x_2, x_1) = \frac{1}{L}\sum_k \frac{1}{2\omega_k} \exp[ik(x_2 - x_1)],$$

$$k = 0, \pm 2\pi/L, \ldots; \quad (1.2)$$

$$E_L(f,g) = \frac{2\pi}{L}\sum_k \frac{1}{2\omega_k} \hat{f}^*(k)\hat{g}(k), \quad (1.3)$$

$$\hat{g}(k) = \int_{-L/2}^{L/2} dx\, e^{-ikx}\, g(x) \qquad (g(t,x) = g(x)\delta(t)). \quad (1.4)$$

Take f(x) and g(x) to be the characteristic function of the interval (- π/2, π/2). Then

$$\hat{g}(k) = \int_{-\pi/2}^{\pi/2} dx\, e^{-ikx} = \sqrt{\frac{2}{\pi}}\, \frac{\sin k\pi/2}{k}. \quad (1.5)$$

Note that

$$|\hat{g}(0)|^2 = \frac{\pi}{2},$$

$$|\hat{g}(k)|^2 = 0 \text{ if } k \text{ is a nonzero even integer},$$

$$|\hat{g}(k)|^2 = \frac{2}{\pi} k^{-2} \text{ if } k \text{ is an odd integer}, \quad (1.6)$$

$$|\hat{g}(k)|^2 = \frac{1}{\pi} k^{-2} \text{ if } k \text{ is half an odd integer}.$$

It follows that

$$E_{2\pi}(g,g) = \frac{\pi}{4m} + \sum_k [\pi\,\omega_k\,k^2]^{-1}, \tag{1.7}$$

and

$$E_{4\pi}(g,g) = \frac{\pi}{8m} + \sum_k [2\pi\,\omega_k\,k^2]^{-1} + \sum_k [4\pi\,\omega_{\frac{k}{2}}(\tfrac{k}{2})^2]^{-1}, \tag{1.8}$$

where, in all the sums, $k = \pm 1, \pm 3, \ldots$.

For sufficiently small m the expansion

$$\omega_k^{-1} \sim \frac{1}{k}\left(1 - \frac{1}{2}\left(\frac{m}{k}\right)^2 + \ldots\right)$$

is valid. Therefore, as $m \longrightarrow 0$ the sums in Eqs. (1.7) and (1.8) remain bounded and the $k = 0$ terms, which approach infinity, dominate. But these differ from each other by a factor of 2. Thus $E_{2\pi}(g,g) \neq E_{4\pi}(g,g)$ for sufficiently small m. Since it is easy to see that each $E_L(g,g)$ is analytic in m for m positive, they cannot coincide even for large m, except possibly at some discrete points.

2. The Difference between the Energy Densities Is Infinite.

As in the previous section, we regard the box of length $L = 2\pi$ as embedded in the infinite universe as the interval $(-\pi, \pi)$. Then the field algebra of the box is a subalgebra of the complete field algebra, and in analogy to Sec. IX.3 we can solve for the box annihilation operators, which appear in the expansions

$$\phi(0,x) = \sum_k (2\omega_k)^{-1/2} [e^{ikx} a_k + e^{-ikx} a_k^{\dagger}], \qquad (2.1)$$

$$\pi(0,x) = -i \sum_k (\omega_k/2)^{1/2} [e^{ikx} a_k - e^{-ikx} a_k^{\dagger}]$$

(k integral), in terms of the annihilation and creation operators $b_\rho^{(\dagger)}$ of the ΦCK representation (terminology of Sec. IX.3). The expression of the a_k in terms of the field is (cf. Eq. (VIII.2.11))

$$a_k = \frac{1}{\sqrt{2}} [\sqrt{\omega_k} \int_{-\pi}^{\pi} dx\, e^{-ikx} \phi(x) + \frac{i}{\sqrt{\omega_k}} \int_{-\pi}^{\pi} dx\, e^{-ikx} \pi(x)]$$

$$= \frac{1}{\sqrt{2}} [\sqrt{\omega_k} \int_{-\infty}^{\infty} dx\, e^{-ikx} u(x)\phi(x) + \frac{i}{\sqrt{\omega_k}} \int_{-\infty}^{\infty} dx\, e^{-ikx} u(x)\pi(x)],$$

where u is the characteristic function of the box. The Fourier transform of u is

$$\tilde{u}(p) \equiv \sqrt{2\pi}\, \hat{u}(p) = \int_{-\infty}^{\infty} dx\, e^{-ipx} u(x) = \sqrt{2\pi}\, \frac{\sin \pi p}{\pi p}. \qquad (2.2)$$

Hence, using the convolution theorem, we find

$$a_k = \frac{1}{\sqrt{2}} [\sqrt{\omega_k} \int_{-\infty}^{\infty} dp\, \hat{u}(k-p)\hat{\phi}(p) + \frac{i}{\sqrt{\omega_k}} \int_{-\infty}^{\infty} dp\, \hat{u}(k-p)\hat{\pi}(p)], \qquad (2\wedge)$$

or

$$a_k = \int_{-\infty}^{\infty} dp\, \tilde{U}(k - p) \times$$

$$\{\frac{1}{2}[\sqrt{\omega_k/\omega_p} + \sqrt{\omega_p/\omega_k}]\, b_p + \frac{1}{2}[\sqrt{\omega_k/\omega_p} - \sqrt{\omega_p/\omega_k}]\, b^{\dagger}_{-p}\}. \quad (2.3)$$

Since the second term does not vanish, the physical quantities $N_k = a^{\dagger}_k a_k$ (k integral) which correspond to the quanta of the Fock representation for the box are represented in the ΦOK representation by operators which do not annihilate the vacuum. In fact, it is easy to see that the kernel of the creation term in Eq. (2.3) is not Hilbert-Schmidt. That is, the box number operator does not have finite vacuum expectation value in the ΦOK representation, and vice versa. (Cf. Sec. IX.3 and Appendix F.)

The energy density of the scalar field, $T^{00}(x)$, which has the classical expression (IX.5.1), is ordinarily made into a quantum-theoretical operator in each of these Fock representations by normal ordering with respect to the appropriate set of annihilation and creation operators. This procedure can be described as the discarding of an infinite numerical term (c-number) in each case. We ask whether the difference between these two infinite quantities is in some sense finite, or even zero. We can take the expression for T^{00}, normal ordered in terms of the a operators, and substitute from Eq. (2.3). The resulting expression will contain terms in $b_\rho b^{\dagger}_\rho$, which contribute a vacuum expectation value in the ΦOK representation. It is this quantity which we wish to investigate. (Let us denote it $\langle T^{00}\rangle$.)

The direct formal calculation just outlined leads to a hopelessly indeterminate expression of the form

$$\sum_{k_1,k_2} A(k_1,k_2),$$

where each $A(k_1,k_2)$ is a divergent integral whose phase depends on the k's. So $\langle T^{00}\rangle$ must be defined in a subtler way, taking account of the distribution nature of the operator T^{00}. Let us consider

$$T^{00}(g,g) = \frac{1}{2}\int dy_1 \int dy_2\, g(y_1)g(y_2)\,[:\pi(y_1)\pi(y_2)$$

$$+ \frac{\partial\phi}{\partial x}(y_1)\frac{\partial\phi}{\partial x}(y_2) + m^2\,\phi(y_1)\phi(y_2):], \quad (2.4)$$

where the normal ordering is with respect to the a's, and g is a smooth function with support in the box. One would naturally define $T^{00}(x)$ as the limit, if any, of this object as $g(y) \longrightarrow \delta(y - x)$. If $\langle T^{00}(g,g)\rangle$, the vacuum expectation value of $T^{00}(g,g)$ in the b_p-representation, is not finite, there is little hope of interpreting the more singular $\langle T^{00}(x)\rangle$ as a finite quantity (even as a distribution).

Defining the Fourier transform of a function by Eq. (2.2), we find from Eqs. (2.4) and (2.1)

$$T^{00}(g,g) = \sum_{k_1}\sum_{k_2} (4\sqrt{\omega_1\omega_2})^{-1} \{(\omega_1\omega_2 + k_1k_2 + m^2)$$

$$\times [\bar{g}(-k_1)\bar{g}(k_2)a^\dagger_{k_2}a_{k_1} + \bar{g}(k_1)\bar{g}(-k_2)a^\dagger_{k_1}a_{k_2}]$$

$$- (\omega_1\omega_2 + k_1k_2 - m^2)[\bar{g}(-k_1)\bar{g}(-k_2)a_{k_1}a_{k_2}$$

$$+ \bar{g}(k_1)\bar{g}(k_2)a^\dagger_{k_1}a^\dagger_{k_2}]\} \quad (2.5)$$

(where $\omega_1^2 = k_1^2 + m^2$, etc). Let us substitute from Eq. (2.3) and isolate the b-vacuum term, writing the p integration on the outside:

$$\langle T^{00}(g,g)\rangle =$$

$$\frac{1}{8}\int_{-\infty}^{\infty} dp \sum_{k_1}\sum_{k_2} \bar{u}(k_1 - p)\bar{u}(k_2 - p)(\omega_1\omega_2)^{-1/2}\, \bar{g}(-k_1)\bar{g}(k_2) \times$$

$$\{(\omega_1\omega_2 + k_1k_2 + m^2)[\frac{\omega_p}{\sqrt{\omega_1\omega_2}} + \frac{\sqrt{\omega_1\omega_2}}{\omega_p} - \sqrt{\frac{\omega_1}{\omega_2}} - \sqrt{\frac{\omega_2}{\omega_1}}]$$

$$+ (\omega_1\omega_2 - k_1k_2 - m^2)[\frac{\omega_p}{\sqrt{\omega_1\omega_2}} - \frac{\sqrt{\omega_1\omega_2}}{\omega_p} + \sqrt{\frac{\omega_1}{\omega_2}} - \sqrt{\frac{\omega_2}{\omega_1}}]\}. \quad (2.6)$$

The sums converge for fixed p, since $\bar{g}(k)$ decreases rapidly. The integral over p converges for all terms except those proportional to ω_p, which can be written

$$\frac{1}{4\pi^2}\int_0^{\infty} dp \; \omega_p \sin^2 \pi p \left| \sum_{k=-\infty}^{\infty} (-1)^k \frac{\tilde{g}(k)}{k-p} \right|^2 \qquad (2.7)$$

if $g(y)$ is real. This integral diverges at the upper limit, since the integrand falls off only as $1/p$. (It can easily be checked that the sum does not vanish identically as $\tilde{g}(k)$ approaches the Fourier coefficients of a delta function.)

Hence the quantum-field-theoretical $:T^{00}(x):$ (unlike $:T^{01}(x):$) seems to depend strongly on the representation with respect to which the normal ordering is defined. This example shows that normal ordering is highly suspect as a method of defining a local energy density suitable for the purposes of general relativity.

BIBLIOGRAPHY

I. Books

Abraham
Abraham, R., Foundations of Mechanics, New York, W. A. Benjamin, Inc., 1967.

Adler-Bazin-Schiffer
Adler, R., Bazin, M., and Schiffer, M., Introduction to General Relativity, New York, McGraw-Hill Book Co., 1965.

Akhiezer-Glazman
Akhiezer, N. I., and Glazman, I. M., Theory of Linear Operators in Hilbert Space, New York, F. Ungar Publ. Co., 1961.

Berezin
Berezin, F. A., The Method of Second Quantization, New York, Academic Press, 1966.

Bjorken-Drell 2
Bjorken, J. D., and Drell, S. D., Relativistic Quantum Fields, New York, McGraw-Hill Book Co., 1965.

Carrier-Krook-Pearson
Carrier, G. F., Krook, M., and Pearson, C. E., Functions of a Complex Variable: Theory and Technique, New York, McGraw-Hill Book Co., 1966.

Corinaldesi
Corinaldesi, E., Relativistic Wave Mechanics, ed. by F. Strocchi, Amsterdam, North-Holland Publ. Co., 1963.

Eisenhart
Eisenhart, L. P., Riemannian Geometry, 2nd ed., Princeton, N.J., Princeton Univ. Press, 1950.

Friedrichs
Friedrichs, K. O., Mathematical Aspects of the Quantum Theory of Fields, New York, Interscience Publs., Inc., 1953.

Gel'fand 5
Gel'fand, I. M., Graev, M. I., and Vilenkin, N. Ya., Generalized Functions, vol. 5, Integral Geometry and Representation Theory, New York, Academic Press, 1966.

Gradshteyn-Ryzhik
Gradshteyn, I. S., and Ryzhik, I. M., Table of Integrals, Series, and Products, New York, Academic Press, 1965.

Hermann
Hermann, R., Lie Groups for Physicists, New York, W. A. Benjamin, Inc., 1966.

Hermann 2
Hermann, R., Fourier Analysis on Groups and Partial Wave Analysis, New York, W. A. Benjamin, Inc., 1969.

Jackson
Jackson, J. D., Classical Electrodynamics, New York, John Wiley and Sons, Inc., 1962.

Kato
Kato, T., Perturbation Theory for Linear Operators, New York, Springer Verlag, 1966.

Mandl
Mandl, F., Introduction to Quantum Field Theory, New York, Interscience Publs., Inc., 1959.

Maurin
Maurin, K., General Eigenfunction Expansions and Unitary Representations of Topological Groups, Warsaw, Polish Scientific Publishers, 1968.

Messiah
Messiah, A., Quantum Mechanics, vols. I-II, Amsterdam, North-Holland Publ. Co., 1961.

Naimark
Naimark, M. A., Linear Representations of the Lorentz Group, Oxford, Pergamon Press, 1964.

N.B.S.
Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables (National Bureau of Standards Applied Mathematics Series, No. 55), ed. by M. Abramowitz and I. A. Stegun, 7th printing, Washington, D.C., U. S. Government Printing Office, 1968.

Reed-Simon
Reed, M., and Simon, B., _Methods of Modern Mathematical Physics_, vol. 1, _Functional Analysis_, New York, Academic Press, 1972.

Rindler
Rindler, W., _Special Relativity_, 2nd ed., Edinburgh, Oliver and Boyd, 1966.

Rindler 2
Rindler, W., _Essential Relativity_, New York, Van Nostrand Reinhold Co., 1969.

Robertson-Noonan
Robertson, H. P., and Noonan, T. W., _Relativity and Cosmology_, Philadelphia, W. B. Saunders Co., 1968.

Schrödinger
Schrödinger, E., _Expanding Universes_, Cambridge, The University Press, 1956.

Schweber
Schweber, S. S., _An Introduction to Relativistic Quantum Field Theory_, New York, Harper and Row, Inc., 1961.

Segal
Segal, I. E., _Mathematical Problems of Relativistic Physics_, Providence, R. I., American Mathematical Society, 1963.

Streater-Wightman
Streater, R. F., and Wightman, A. S., _PCT, Spin and Statistics, and All That_, New York, W. A. Benjamin, Inc., 1964.

Synge
Synge, J. L., _Relativity: The General Theory_, Amsterdam, North-Holland Publ. Co., 1960.

Talman
Talman, J. D., _Special Functions: A Group Theoretic Approach_ (based on lectures by E. P. Wigner), New York, W. A. Benjamin, Inc., 1968.

Titchmarsh
Titchmarsh, E. C., _Eigenfunction Expansions Associated with Second-order Differential Equations_, 2nd ed., Part I, Oxford, Clarendon Press, 1962.

Tolman

Tolman, R. C., Relativity, Thermodynamics, and Cosmology, Oxford, Clarendon Press, 1934.

Umezawa

Umezawa, H., Quantum Field Theory, Amsterdam, North-Holland Publ. Co., 1956.

Vilenkin

Vilenkin, N. J., Special Functions and the Theory of Group Representations (Transl. Math. Monographs, vol. 22), Providence, R.I., American Mathematical Society, 1968.

II. Papers, Lecture Notes, and Dissertations

Araki, H., and Haag, R. (1967)
Collision Cross Sections in Terms of Local Observables, Commun. Math. Phys. 4, 77.

Bacry, H., and Lévy-Leblond, J.-M. (1968)
Possible Kinematics, J. Math. Phys. 9, 1605.

Bakri, M. M. (1969)
De Sitter Symmetric Field Theory. I. One-Particle Theory, J. Math. Phys. 10, 298.

Bargmann, V. (1947)
Irreducible Unitary Representations of the Lorentz Group, Ann. Math. 48, 568.

Bargmann, V. (1954)
On Unitary Ray Representations of Continuous Groups, Ann. Math. 59, 1.

Bargmann, V., and Wigner, E. P. (1948)
Group Theoretical Discussion of Relativistic Wave Equations, Proc. Natl. Acad. Sci. U.S. 34, 211.

Belinfante, F. J. (1940)
On the Current and the Density of the Electric Charge, the Energy, the Linear Momentum and the Angular Momentum of Arbitrary Fields, Physica 7, 449.

Böhm, A. (1966)
The Dynamical Group of a Simple Particle Model, Lectures in Theoretical Physics, vol. IXB, High Energy and Particle Physics (Theoretical Physics Institute, Univ. of Colorado, 1966), ed. by W. E. Brittin and A. O. Barut (New York, Gordon and Breach, 1967), pp. 327-336.

Böhm, A. (1970)
Generalized Eigenvectors and Group Representations -- the Connection Between Representations of SO(4,1) and the Poincaré Group (lectures at Istanbul Summer Institute for Mathematical Physics, 1970), preprint CPT-89 (Feb., 1971), Center for Particle Theory, University of Texas at Austin.

Bohr, N., and Rosenfeld, L. (1933)
Zur Frage der Messbarkeit der electromagnetischen Feldgrössen, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. 12, No. 8.

Bonazzola, S., and Pacini, F. (1966)
Equilibrium of a Large Assembly of Particles in General Relativity, Phys. Rev. 148, 1269.

Borchers, H. J., Haag, R., and Schroer, B. (1963)
The Vacuum State in Quantum Field Theory, Nuovo Cimento 29, 148.

Börner, G. (1970)
Mass Zero Particles in De Sitter Space, preprint, Max-Planck-Institut für Physik und Astrophysik, Munich [to appear in Lectures in Theoretical Physics, vol. XIII, De Sitter and Conformal Groups (Symposium, University of Colorado, 1970), ed. by A. O. Barut, to be published by Gordon and Breach].

Börner, G., and Dürr, H. P. (1969)
Classical and Quantum Fields in de Sitter Space, Nuovo Cimento A64, 669.

Brenig, W., and Haag, R. (1959)
General Quantum Theory of Collision Processes, Quantum Scattering Theory, ed. by M. Ross (Bloomington, Ind., Indiana Univ. Press, 1963), pp. 13-108 [tr. of Fortschr. Phys. 7, 183].

Brezin, E., and Itzykson, C. (1970)
Pair Production in Vacuum by an Alternating Field, Phys. Rev. D2, 1191.

Bronnikov, K. A., and Tagirov, E. A. (1968)
Quantum Theory of Scalar Field in an Isotropic World, preprint P2-4151, Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna [in Russian].

Burcev, P. (1968)
De Sitter Model for Stable Particles, Nuovo Cimento 56A, 795.

Calabi, E., and Markus, L. (1962)
Relativistic Space Forms, Ann. Math. 75, 63.

Capri, A. Z. (1967)
External Field Problem for Higher Spin Particles, dissertation, Princeton University.

Capri, A. Z. (1969)
Electron in a Given Time-Dependent Electromagnetic Field, J. Math. Phys. 10, 575.

Capri, A. Z., Labonté, G., Menon, D., and Shamaly, A. (1971)
The Reconstruction Theorem with External Sources, Nuovo Cimento 3B, 233.

Castagnino, M. (1969)
Deuxième quantification des champs scalaires réels dans l'espace-temps de De-Sitter, Compt. Rend. 268, A1157.

Castagnino, M. (1970)
Champs de spin entier dans l'espace-temps de De Sitter, Ann. Inst. Henri Poincaré 13, 263.

Castell, L. (1969)
Goldstone Particles in de-Sitter Space, Nuovo Cimento A61, 585.

Chernikov, N. A., and Tagirov, E. A. (1968)
Quantum Theory of Scalar Field in De Sitter Space-Time, Ann. Inst. Henri Poincaré 9, 109.

Choquet-Bruhat, Y. (1967)
Hyperbolic Partial Differential Equations on a Manifold, Battelle Rencontres: 1967 Lectures in Mathematics and Physics, ed. by C. M. De Witt and J. A. Wheeler (New York, W. A. Benjamin, Inc., 1968), pp. 84-106.

Coleman, A. J. (1968)
Induced and Subduced Representations, Group Theory and Its Applications, ed. by E. M. Loebl (New York, Academic Press, 1968), pp. 57-118.

Davies, E. B. (1971)
Hilbert Space Representations of Lie Algebras, Commun. Math. Phys. 23, 159.

Dell'Antonio, G. F. (1968)
Structure of the Algebras of Some Free Systems, Commun. Math. Phys. 9, 81.

Deser, S. (1957)
General Relativity and the Divergence Problem in Quantum Field Theory, Rev. Mod. Phys. 29, 417.

De Witt, B. S. (1963)
Dynamical Theory of Groups and Fields, Relativity, Groups and Topology (Summer School of Theoretical Physics, Les Houches, 1963), ed. by C. DeWitt and B. DeWitt (New York, Gordon and Breach, 1964), pp. 585-820.

Dirac, P. A. M. (1935)
The Electron Wave Equation in De-Sitter Space, Ann. Math. 36, 657.

Dixmier, J. (1961)
Représentations intégrables du groupe de De Sitter, Bull. Soc. Math. France 89, 9.

Dubin, D. A. (1970)
The Group-Theoretical Structure of Free Quantum Fields in Two Dimensions, Nuovo Cimento 67B, 39.

Duistermaat, J. J., and Hörmander, L. (1971)
Fourier Integral Operators. II, preprint, to be published in Acta Math. [This material also appeared in lectures of Hörmander at the Summer Symposium on Partial Differential Equations, Berkeley, Calif., 1971, to be published by American Mathematical Society.]

Epstein, H., Glazer, V., and Jaffe, A. (1965)
Nonpositivity of the Energy Density in Quantized Field Theories, Nuovo Cimento 36, 1016.

Fell, J. M. G. (1960)
The Dual Spaces of C*-Algebras, Transac. Am. Math. Soc. 94, 365.

Feshbach, H., and Villars, F. (1958)
Elementary Relativistic Wave Mechanics of Spin 0 and Spin 1/2 Particles, Rev. Mod. Phys. 30, 24.

Feynman, R. P. (1948)
Space-Time Approach to Non-Relativistic Quantum Mechanics, Rev. Mod. Phys. 20, 267.

Feynman, R. P. (1950)
Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction, Phys. Rev. 80, 440.

Fourès-Bruhat, Y. (1960)
Propagateurs et solutions d'équations homogènes hyperboliques, Compt. Rend. 251, 29.

Fronsdal, C. (1965)
Elementary Particles in a Curved Space, Rev. Mod. Phys. **37**, 221.

Gel'fand, I. M. (1950)
The Center of an Infinitesimal Group Ring, Mat. Sb., N.S., **26**, 103 [in Russian].

Gel'fand, I., and Neumark [Naimark], M. (1946)
Unitary Representations of the Lorentz Group, J. Phys. (U.S.S.R.) **10**, 93.

Gel'fand, I. M., and Naimark, M. A. (1947)
Unitary Representations of the Lorentz Group, Izv. Akad. Nauk, Ser. Mat., **11**, 411 [in Russian].

Geroch, R. (1968)
What is a Singularity in General Relativity?, Ann. Phys. (N.Y.) **48**, 526.

Geroch, R. (1970)
Domain of Dependence, J. Math. Phys. **11**, 437.

Glimm, J., and Jaffe, A. (1970)
Quantum Field Theory Models, _Statistical Mechanics and Quantum Field Theory_ (Summer School of Theoretical Physics, Les Houches, 1970), ed. by C. DeWitt and R. Stora (New York, Gordon and Breach, Inc., 1971), pp. 1-108.

Greenberg, O. W. (1961)
Generalized Free Fields and Models of Local Field Theory, Ann. Phys. (N.Y.) **16**, 158.

Grib, A. A. (1969)
On a Possible Difference between the Heisenberg and Schrödinger Pictures in Quantum Field Theory, Vestnik Leningrad. Univ. **1969**, No. 10, 61 [in Russian].

Grib, A. A., and Mamaev, S. G. (1969)
Contribution to the Field Theory in Friedmann Space, Sov. J. Nucl. Phys. **10**, 722 [tr. of Yad. Fiz. **10**, 1276].

Gürsey, F. (1962)
Introduction to the De Sitter Group, _Group Theoretical Concepts and Methods in Elementary Particle Physics_ (Istanbul Summer School of Theoretical Physics, 1962), ed. by F. Gürsey (New York, Gordon and Breach, Inc., 1964), pp. 365-389.

Gürsey, F., and Lee, T. D. (1963)
Spin 1/2 Wave Equation in De-Sitter Space, Proc. Natl. Acad. Sci. U.S. **49**, 179.

Gutzwiller, M. (1956)
Quantum Theory of Wave Fields in a Curved Space, Helv. Phys. Acta **29**, 313.

Haag, R. (1955)
On Quantum Field Theories, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **29**, No. 12.

Haag, R., and Kastler, D. (1964)
An Algebraic Approach to Quantum Field Theory, J. Math. Phys. **5**, 848.

Hannabuss, K. C. (1969a)
Quantum Dynamics in De Sitter Space, thesis, Oxford University.

Hannabuss, K. C. (1969b)
The Dirac Equation in de Sitter Space, J. Phys. A **2**, 274.

Hannabuss, K. C. (1970)
The Localisability of Particles in de Sitter Space, preprint, Mathematical Institute, Oxford.

Hawking, S. (1970)
The Conservation of Matter in General Relativity, Commun. Math. Phys. **18**, 301.

Hill, E. L. (1951)
Hamilton's Principle and the Conservation Theorems of Mathematical Physics, Rev. Mod. Phys. **23**, 253.

Hirai, T. (1962a)
On Infinitesimal Operators of Irreducible Representations of the Lorentz Group of n-th Order, Proc. Japan Acad. **38**, 83.

Hirai, T. (1962b)
On Irreducible Representations of the Lorentz Group of n-th Order, Proc. Japan Acad. **38**, 258.

Holman, W. J. (1969)
Representation Theory of SO(4,1) and E(3,1): An Explicit Spinor Calculus, J. Math. Phys. **10**, 1888.

Ikebe, T. (1960)
Eigenfunction Expansions Associated with the Schroedinger Operators and their Applications to Scattering Theory, Arch. Rat. Mech. Anal. 5, 1.

Imamura, T. (1960)
Quantized Meson Field in a Classical Gravitational Field, Phys. Rev. 118, 1430.

Inönü, E. (1962)
Contraction of Lie Groups and their Representations, Group Theoretical Concepts and Methods in Elementary Particle Physics (Istanbul Summer School of Theoretical Physics, 1962), ed. by F. Gürsey (New York, Gordon and Breach, Inc., 1964), pp. 391-402.

Inönü, E., and Wigner, E. P. (1953)
On the Contraction of Groups and Their Representations, Proc. Natl. Acad. Sci. U.S. 39, 510; correction, On a Particular Type of Convergence to a Singular Matrix, 40, 119 (1954).

Isham, C. J., Salam, A., and Strathdee, J. (1971)
Infinity Suppression in Gravity-Modified Quantum Electrodynamics, Phys. Rev. D3, 1805.

Jaffe, A. M. (1965)
Dynamics of a Cut-Off $\lambda\phi^4$ Field Theory, dissertation, Princeton University.

Kastler, D. (1965)
The C*-Algebras of a Free Boson Field. I. Discussion of the Basic Facts, Commun. Math. Phys. 1, 14.

Kerr, R. P. (1963)
Gravitational Field of a Spinning Mass as an Example of Algebraically Special Metrics, Phys. Rev. Letters 11, 237.

Kihlberg, A. (1965)
On the Unitary Representations of a Class of Pseudo-Orthogonal Groups, Arkiv Fysik 30, 121.

Klauder, J. R., McKenna, J., and Woods, E. J. (1966)
Direct-Product Representations of the Canonical Commutation Relations, J. Math. Phys. 7, 822.

Komar, A. (1964)
Undecidability of Macroscopically Distinguishable States in Quantum Field Theory, Phys. Rev. 133B, 542.

Kristensen, P., Mejlbo, L., and Poulsen, E. T. (1967)
Tempered Distributions in Infinitely Many Dimensions. III. Linear Transformations of Field Operators, Commun. Math. Phys. 6, 29.

Kruskal, M. D. (1960)
Maximal Extension of Schwarzschild Metric, Phys. Rev. 119, 1743.

Lévy-Nahas, M. (1967)
Deformation and Contraction of Lie Algebras, J. Math. Phys. 8, 1211.

Lévy-Nahas, M. (1969)
Sur les déformations and contractions d'algèbres de Lie et de leurs représentations, thesis, University of Paris.

Lichnerowicz, A. (1961)
Propagateurs et commutateurs en relativité générale, Institut de Hautes Etudes Scientifiques, Publications Mathématiques, No. 10.

Lichnerowicz, A. (1962)
Propagateurs et quantification en relativité générale, Conférence internationale sur les théories relativistes de la gravitation (Warsaw and Jabłonna, 1962), ed. by L. Infeld (Paris, Gauthier-Villars, 1964), pp. 177-188.

Limić, N., Niederle, J., and Rączka, R. (1967)
Eigenfunction Expansions Associated with the Second-Order Invariant Operator on Hyperboloids and Cones. III, J. Math. Phys. 8, 1079.

Lindblad, G., and Nagel, B. (1970)
Continuous Bases for Unitary Irreducible Representations of SU(1,1), Ann. Inst. Henri Poincaré A13, 27.

Manasse, F. K., and Misner, C. W. (1963)
Fermi Normal Coordinates and Some Basic Concepts in Differential Geometry, J. Math. Phys. 4, 735.

Manuceau, J. (1968)
C*-algèbre de relations de commutation, Ann. Inst. Henri Poincaré A8, 139.

Marzke, R. F., and Wheeler, J. A. (1964)
Gravitation as Geometry -- I: The Geometry of Space-Time and the Geometrodynamical Standard Meter, Gravitation and Relativity, ed. by H.-Y. Chiu and W. F. Hoffmann (New York, W. A. Benjamin, Inc., 1964), pp. 40-64.

Molčanov [Molchanov], V. F. (1966)
Harmonic Analysis on a Hyperboloid of One Sheet, Sov. Math. -- Dokl. 7, 1553 [tr. of Dokl. Akad. Nauk 171, 794].

Moore, G. T. (1970)
Quantum Theory of the Electromagnetic Field in a Variable-Length One-Dimensional Cavity, J. Math. Phys. 11, 2679.

Mukunda, N. (1967)
Unitary Representations of the Group O(2,1) in an O(1,1) Basis, J. Math. Phys. 8, 2210.

Nachtmann, O. (1967)
Quantum Theory in de-Sitter Space, Commun. Math. Phys. 6, 1.

Nachtmann, O. (1968a)
Continuous Creation in a Closed World Model, Z. Physik 208, 113.

Nachtmann, O. (1968b)
Dynamische Stabilität im de-Sitter-Raum, Sitzber. Österr. Akad. Wiss., Math.-naturw. Klasse, 167, 363.

Narozhnyĭ [Narozhny], N. B., and Nikishov, A. I. (1970)
The Simplest Processes in a Pair-Producing Electric Field, Sov. J. Nucl. Phys. 11, 596 [tr. of Yad. Fiz. 11, 1072].

Nelson, E. (1959)
Analytic Vectors, Ann. Math. 70, 572.

Newton, T. D. (1949)
Localized States for Elementary Systems, dissertation, Princeton University.

Newton, T. D. (1950)
A Note on the Representations of the de Sitter Group, Ann. Math. 51, 730.

Newton, T. D., and Wigner, E. P. (1949)
Localized States for Elementary Systems, Rev. Mod. Phys. 21, 400.

Ottoson, U. (1968)
A Classification of the Unitary Irreducible Representations of $SO_o(N,1)$, Commun. Math. Phys. 8, 228.

Parker, L. E. (1966)
The Creation of Particles in an Expanding Universe, thesis, Harvard University.

Parker, L. (1968)
Particle Creation in Expanding Universes, Phys. Rev. Letters 21, 562.

Parker, L. (1969)
Quantized Fields and Particle Creation in Expanding Universes. I, Phys. Rev. 183, 1057.

Parker, L. (1971)
Quantized Fields and Particle Creation in Expanding Universes. II, Phys. Rev. D3, 346.

Parker, L. (1972)
Particle Creation in Isotropic Cosmologies, Phys. Rev. Letters 28, 705.

Penrose, R. (1963)
Conformal Treatment of Infinity, Relativity, Groups and Topology (Summer School of Theoretical Physics, Les Houches, 1963), ed. by C. DeWitt and B. DeWitt (New York, Gordon and Breach, 1964), pp. 563-584.

Penrose, R. (1965)
Zero Rest-Mass Fields Including Gravitation: Asymptotic Behavior, Proc. Roy. Soc. (London) A284, 159.

Penrose, R. (1967)
Structure of Space-Time, Battelle Rencontres: 1967 Lectures in Mathematics and Physics, ed. by C. M. De Witt and J. A. Wheeler (New York, W. A. Benjamin, Inc., 1968), pp. 121-235.

Philips, T. O. (1963)
Localized States in de Sitter Space, dissertation, Princeton University.

Philips, T. C. (1964)
Lorentz Invariant Localized States, Phys. Rev. 136B, 893.

Philips, T. C., and Wigner, E. P. (1968)
De Sitter Space and Positive Energy, Group Theory and Its Applications, ed. by E. M. Loebl (New York, Academic Press, 1968), pp. 631-676.

Powers, R. T. (1967)
Absence of Interaction as a Consequence of Good Ultraviolet Behavior in the Case of a Local Fermi Field, Commun. Math. Phys. 4, 145.

Powers, R. T. (1971)
Self-Adjoint Algebras of Unbounded Operators, Commun. Math. Phys. 21, 85.

Pukánszky, L. (1961)
On the Kronecker Products of Irreducible Representations of the 2 X 2 Real Unimodular Group. I, Transac. Am. Math. Soc. 100, 116.

Reed, M. C. (1968)
On the Self-Adjointness of Quantum Fields and Hamiltonians, dissertation, Stanford University.

Reed, M. C. (1969)
A Gårding Domain for Quantum Fields, Commun. Math. Phys. 14, 336.

Reed, M. C. (1970)
On Self-Adjointness in Infinite Tensor Product Spaces, J. Func. Anal. 5, 94.

Rindler, W. (1960)
Remarks on Schrödinger's Model of de Sitter Space, Phys. Rev. 120, 1041.

Rindler, W. (1966)
Kruskal Space and the Uniformly Accelerated Frame, Am. J. Phys. 34, 1174.

Roman, P., and Aghassi, J. J. (1966)
On the Energy Spectrum of de Sitter World Theories, Nuovo Cimento 42A, 193.

Rosen, J. (1965)
The Inhomogeneous Pseudo-Rotation Group as a Limiting Case of a (Homogeneous) Pseudo-Rotation Group of One More Dimension, Nuovo Cimento 35, 1234.

Ruffini, R., and Bonazzola, S. (1969)
Systems of Self-Gravitating Particles in General Relativity and the Concept of an Equation of State, Phys. Rev. 187, 1767.

Saletan, E. J. (1961)
Contraction of Lie Groups, J. Math. Phys. 2, 1.

Scarf, F. L. (1959)
A Soluble Quantum Field Theory in Curved Space, Les théories relativistes de la gravitation (conference, Royaumont, 1959) (Paris, Centre National de la Recherche Scientifique, 1962), pp. 421-428.

Schiff, L. I., Snyder, H., and Weinberg, J. (1940)
On the Existence of Stationary States of the Mesotron Field, Phys. Rev. 57, 315.

Schild, A. (1965)
Lectures on General Relativity Theory, Lectures in Applied Mathematics, vol. 8, Relativity Theory and Astrophysics 1: Relativity and Cosmology (Fourth Summer Seminar on Applied Mathematics, Cornell Univ., 1965), ed. by J. Ehlers (Providence, R. I., American Mathematical Society, 1967), pp. 1-104.

Schrödinger, E. (1939)
The Proper Vibrations of the Expanding Universe, Physica 6, 899.

Schroer, B. (1971)
Quantization of $m^2 < 0$ Field Equations, Phys. Rev. D3, 1764.

Schroer, B., Seiler, R., and Swieca, J. A. (1970)
Problems of Stability for Quantum Fields in External Time-Dependent Potentials, Phys. Rev. D2, 2927.

Schroer, B., and Swieca, J. A. (1970)
Indefinite Metric and Stationary External Interactions of Quantized Fields, Phys. Rev. D2, 2938.

Schwarz, F. (1971)
Unitary Irreducible Representations of the Groups $SO_0(n,1)$, J. Math. Phys. 12, 131.

Segal, I. E. (1947)
Postulates for General Quantum Mechanics, Ann. Math. 48, 930.

Segal, I. E. (1951)
A Class of Operator Algebras which Are Determined by Groups, Duke Math. J. 18, 221.

Segal, I. E. (1968)
The Mathematical Theory of Quantum Fields, _Lectures in Modern Analysis and Applications II_ (Lecture Notes in Mathematics, vol. 140), ed. by C. T. Taam (Berlin, Springer-Verlag, 1970), pp. 30-57.

Segal, I. E., and Goodman, R. W. (1965)
Anti-Locality of Certain Lorentz-Invariant Operators, J. Math. & Mech. 14, 629.

Seiler, R. (1972)
Quantum Theory of Particles with Spin Zero and One Half in External Fields, Commun. Math. Phys. 25, 127.

Sexl, R. U., and Urbantke, H. K. (1967)
Cosmic Particle Creation Processes, Acta Phys. Austriaca 26, 339.

Sexl, R. U., and Urbantke, H. K. (1969)
Production of Particles by Gravitational Fields, Phys. Rev. 179, 1247.

Shale, D. (1962)
Linear Symmetries of Free Boson Fields, Transac. Am. Math. Soc. 103, 149.

Slawny, J. (1972)
On Factor Representations and the C*-Algebra of Canonical Commutation Relations, Commun. Math. Phys. 24, 151.

Snyder, H., and Weinberg, J. (1940)
Stationary States of Scalar and Vector Fields, Phys. Rev. 57, 307.

Stein, E. M. (1965)
A Survey of Representations of Non-Compact Groups, _High-Energy Physics and Elementary Particles_ (Seminar, International Center for Theoretical Physics, Trieste, 1965) (Vienna, International Atomic Energy Agency, 1965), pp. 563-584.

Steinmann, O. (1968)
Particle Localization in Field Theory, Commun. Math. Phys. 7, 112.

Streater, R. F. (1971)
On the Lorentz Invariance of $P(\phi)_2$, preprint, Bedford College, London.

Streit, L. (1967)
Test Function Spaces for Direct Product Representations of the Canonical Commutation Relations, Commun. Math. Phys. 4, 22.

Ström, S. (1965)
Construction of Representations of the Inhomogeneous Lorentz Group by Means of Contraction of Representations of the (1 + 4) De Sitter Group, Arkiv Fysik 30, 455.

Ström, S. (1971)
Introduction to the Theory of Groups and Group Representations, preprint CPT-120, Center for Particle Theory, University of Texas at Austin.

Tagirov, E. A., Fedyun'kin, E. D., and Chernikov, N. A. (1967)
Quantum Theory of a Scalar Field in Two-Dimensional De Sitter Space, preprint P2-3392, Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna [in Russian].

Tait, W., and Cornwell, J. F. (1971)
Coupling of de Sitter Space-Time and Internal Symmetry, J. Math. Phys. 12, 1651.

Takahashi, R. (1963)
Sur les représentations unitaires des groupes de Lorentz généralisés, Bull. Soc. Math. France 91, 289.

Thirring, W. (1967)
Quantum Field Theory in De Sitter Space, Special Problems in High Energy Physics (6th Internationale Universitätswochen für Kernphysik, Schladming, Austria, 1967), ed. by P. Urban (Vienna, Springer-Verlag, 1967), pp. 269-287.

Thomas, L. S. (1941)
On Unitary Representations of the Group of De Sitter Space, Ann. Math. 42, 113.

Tugov, I. I. (1969)
Conformal Covariance and Invariant Formulation of Scalar Wave Equations, Ann. Inst. Henri Poincaré 11, 207.

Unruh, W. G. (1971)
Dirac Particles and Geometrodynamical Charge in Curved Geometries, dissertation, Princeton University.

Urbantke, H. K. (1969)
Remark on Noninvariance Groups and Field Quantization in Curved Space, Nuovo Cimento 63B, 203.

Veselić, K. (1970)
A Spectral Theory for the Klein-Gordon Equation with an External Electrostatic Potential, Nucl. Phys. A147, 215.

Vigier, J.-P. (1969)
Unification of External and Internal Motions within $SO_{6,1}$ and Possible Mass Splitting of SU_3 Baryon Multiplets without Symmetry Breaking, Lett. Nuovo Cimento 1, 445.

Von Neumann, J. (1938)
On Infinite Direct Products, Compos. Math. 6, 1.

Wheeler, J. A. (1963)
Geometrodynamics and the Issue of the Final State, Relativity, Groups and Topology (Summer School of Theoretical Physics, Les Houches, 1963), ed. by C. DeWitt and B. DeWitt (New York, Gordon and Breach, 1964), pp. 315-520.

Wightman, A. S. (1959)
Relativistic Invariance and Quantum Mechanics, Nuovo Cimento Suppl. 14, 81.

Wightman, A. S. (1960)
L'invariance dans la mécanique quantique relativiste, Relations de dispersion et particules élémentaires (Université de Grenoble, Ecole d'été de physique théorique, Les Houches, 1960), ed. by C. DeWitt and R. Omnes (Paris, Hermann, 1960), pp. 159-226.

Wightman, A. S. (1962)
On the Localizability of Quantum Mechanical Systems, Rev. Mod. Phys. 34, 845.

Wightman, A. S. (1964)
Introduction to Some Aspects of Relativistic Dynamics of Quantum Fields, Cargèse Lectures in Theoretical Physics: High Energy Electromagnetic Interactions and Field Theory, ed. by M. Lévy (New York, Gordon and Breach, 1967), pp. 171-291.

Wightman, A. S. (1968)
The Stability of Representations of the Poincaré Group, Symmetry Principles at High Energy (5th Coral Gables Conference, 1968), ed. by A. Perlmutter, C. A. Hurst, and B. Kursunoglu (New York, W. A. Benjamin, Inc., 1968), pp. 291-314.

Wightman, A. S. (1971)
Relativistic Wave Equations as Singular Hyperbolic Systems, lectures at the Summer Symposium on Partial Differential Equations, Berkeley, Calif., 1971, to be published by American Mathematical Society.

Wightman, A. S., and Gårding, L. (1965)
Fields as Operator-Valued Distributions in Relativistic Quantum Field Theory, Arkiv Fysik 28, 129.

Wightman, A. S., and Schweber, S. S. (1955)
Configuration Space Methods in Relativistic Quantum Field Theory. I, Phys. Rev. 98, 812.

Wigner, E. P. (1939)
On Unitary Representations of the Inhomogeneous Lorentz Group, Ann. Math. 40, 149.

Wigner, E. P. (1948)
Relativistische Wellengleichungen, Z. Physik 124, 665.

Wigner, E. P. (1950)
Some Remarks on the Infinite De Sitter Space, Proc. Natl. Acad. Sci. U.S. 36, 184.

Wigner, E. P. (1956)
Relativistic Invariance in Quantum Mechanics, Nuovo Cimento 3, 517.

Wigner, E. P. (1961)
Geometry of Light Paths between Two Material Bodies, J. Math. Phys. 2, 207.

Wilde, I. F. (1971)
Algebraic Quantum Field Theory, thesis, University of London.

Zel'dovich, Ya. B. (1970)
Particle Production in Cosmology, JETP Lett. 12, 307 [tr. of ZhETF Pis. Red. 12, 443].

Zeldovič [Zel'dovich], Ya. B., and Pitaevskiĭ [Pitaevsky], L. P. (1971)
On the Possibility of the Creation of Particles by a Classical Gravitational Field, Commun. Math. Phys. 23, 185.

Zel'dovich, Ya. B., and Starobinsky, A. A. (1971)
Particle Creation and Vacuum Polarization in the Anisotropic Gravitational Field, Zh. Eksp. Teor. Fiz. 61, 2161 [in Russian; preliminary English tr., preprint, U.S.S.R. Academy of Sciences, Institute of Applied Mathematics].

ABSTRACT

Quantization of a massive neutral scalar field without self-interaction defined on a space-time manifold with given metric is studied, with emphasis on the two-dimensional de Sitter space. Applications in both general relativity and constructive quantum field theory are envisaged.

The canonical formalism is developed for an arbitrary metric, and for special classes of metrics a Fock space can be constructed in analogy to the case of flat space. However, in this way one is led to different theories for the same manifold, with different definitions of particle observables and energy density. In particular, two nonstandard quantizations of the free field in flat space are exhibited, and three approaches to the two-dimensional de Sitter space are compared: a covariant theory in which the states of a particle transform according to a representation of the symmetry group of the space, a quantization exploiting the static nature of a portion of the universe bounded by horizons, and an "expanding universe" theory in which the particle observables diagonalize the field Hamiltonian at each time and the particle number is not constant. The representations of the canonical commutation relations in the first and third cases are unitarily equivalent.

It is concluded that in this context choice of a unique physical representation of the fields is impossible. One must

deal with an abstract algebra of observables associated with the field. Nevertheless, some representations are more likely to be useful than others. In this spirit a proposal is made for a definition of particle observables based on diagonalization of the Hamiltonian on geodesic hypersurfaces. In Minkowski space this condition distinguishes the standard theory from the others. In two-dimensional de Sitter space such a theory predicts finite and reasonably small creation of particles.

The relation of "contraction" between the irreducible unitary representations of the de Sitter group and those of the Poincaré group of the same dimension is discussed in some detail. It is indicated that some arbitrariness can be removed from the treatment by considering concrete realizations of the representations by functions on the respective homogeneous spaces. The analogous case of the three-dimensional rotation group and the Euclidean group of the plane is treated in an appendix.