1. INTRODUCTION

1.1. Introductory remarks on supersymmetry. The subject of *supersymmetry* (SUSY) is a part of the theory of elementary particles and their interactions and the still unfinished quest of obtaining a unified view of all the elementary forces in a manner compatible with quantum theory and general relativity. Supersymmetry was discovered in the early 1970's, and in the intervening years has become a major component of theoretical physics. Its novel mathematical features have led to a deeper understanding of the geometrical structure of spacetime, a theme to which great thinkers like Riemann, Poincaré, Einstein, Weyl, and many others have contributed.

Symmetry has always played a fundamental role in quantum theory: rotational symmetry in the theory of spin, Poincaré symmetry in the classification of elementary particles, and permutation symmetry in the treatment of systems of identical particles. Supersymmetry is a new kind of symmetry which was discovered by the physicists in the early 1970's. However, it is different from all other discoveries in physics in the sense that there has been no experimental evidence supporting it so far. Nevertheless an enormous effort has been expended by many physicists in developing it because of its many unique features and also because of its beauty and coherence[1]. Here are some of its salient features[2]:

- It gives rise to symmetries between bosons and fermions at a fundamental level.
- Supersymmetric quantum field theories have "softer" divergences.
- Supersymmetric string theory (superstrings) offers the best context known so far for constructing unified field theories.

The development of supersymmetry has led to a number of remarkable predictions. One of the most striking of these is that every elementary particle has a SUSY partner of opposite spin parity, i.e., if the particle is a boson (resp. fermion), its partner is a fermion (resp. boson). The partners of electrons, neutrinos, and quarks are called selectrons, sneutrinos, and squarks, the partner of the photon is a fermion named photino, and so on. However the masses of these partner particles are in the TeV range and so are beyond the reach of currently functioning accelerators (the Fermilab has energies in the 1 TeV range). The new LHC being built at CERN and expected to be operational by 2005 or so, will have energies > 10 TeV and it is expected that perhaps some of these SUSY partners may be found among the collisions that will be created there. Also SUSY predicts a mass for the Higgs particle in the range of about several hundred times the mass of the proton whereas there are no such bounds for it in the usual standard model.

For the mathematician the attraction of supersymmetry lies above all in the fact that it has provided a new look at geometry, both differential and algebraic, beyond its conventional limits. In fact, supersymmetry has provided a surprising continuation of the long evolution of ideas regarding the concept of space and more generally of what a geometric object should be like, an evolution that started with Riemann and was believed to have ended with the creation of the theory of schemes by Grothendieck. If we mean by a geometrical object something that is built out of local pieces, and which in some sense reflects our most fundamental ideas about the structure of space or spacetime, then the most general such object is a *superscheme*, and the symmetries of such an object are *supesymmetries* which are described by *super group schemes*.

1.2. Classical mechanics, the electromagnetic, and gravitational fields. The temporal evolution of a deterministic system is generally described by starting with a set S whose elements are the "states" of the system, and giving a one-parameter group

$$D : t \longmapsto D_t \qquad (t \in \mathbf{R})$$

of bijections of S. D is called the *dynamical group* and its physical meaning is that if s is the state at time 0, then $D_t[s]$ is the state at time t. Usually S has some additional structure and the D_t would preserve this structure, and so the D_t

would be "automorphisms" of S. If thermodynamic considerations are important, then D will be only a semigroup, defined for $t > 0$; the D_t would then typically be only *endomorphisms* of S, i.e., not invertible, so that the dynamics will not be reversible in time. Irreversibility of the dynamics is a consequence of the second law of thermodynamics which says that the entropy of a system increases with time and so furnishes a direction to the arrow of time. But at the microscopic level all dynamics are time reversible, and so we will always have a dynamical group. If the system is relativistic, then the reference to time in the above remarks is to the time in the frame of an (inertial) observer. In this case one requires additional data that describe the fact that the description of the system is the same for all observers. This is usually achieved by requiring that the set of states should be the same for all observers, and that there is a "dictionary" that specifies how to go from the description of one observer to the description of another. The dictionary is given by an action of the Poincaré group P on S. If

$$P \times S \longrightarrow S, \qquad g, s \longmapsto g[s]$$

is the group action, and O, O' are two observers whose coordinate systems are related by $g \in P$, and if $s \in S$ is the state of the system as described by O, then $s' = g[s]$ is the state of the system as described by O'. We shall see examples of this later.

Finally, physical observables are represented by real-valued functions on the set of states and form a real commutative algebra.

Classical mechanics. In this case S is a *smooth* manifold and the dynamical group comes from a *smooth* action of $\mathbf{R}$ on S. If

$$X_s := X_{D,S} = \left(\frac{d}{dt}\right)_{t=0} (D_t[s]) \qquad (s \in S)$$

then $X(s \longmapsto X_s)$ is a vector field on S, the *dynamical vector field*. In practice only X is given in physical theories and the construction of the D_t is only implicit. Strictly speaking, for a given X, the D_t are not defined for all t without some further restriction on X (compact support will do, in particular if S is compact). The D_t are however defined uniquely for small time starting from points of S, i.e., we have a *local flow* generated by X. A key property of this local flow is that for any compact set $K \subset S$ there is $\varepsilon > 0$ such that for all points $s \in K$ the flow starting from s at time 0 is defined for all $t \in (-\varepsilon, +\varepsilon)$.

In most cases we have a manifold M, the so called "configuration space" of the system. For instance, for a system consisting of N point masses moving on some

manifold U, U^N is the configuration space. There are then two ways of formulating classical mechanics.

Hamiltonian mechanics. Here $S = T^*M$, the cotangent bundle of M. S has a canonical 1-form ω which in local coordinates $q_i, p_i (1 \leq i \leq n)$ is $p_1 dq_1 + \ldots + p_n dq_n$. In coordinate-free terms the description of ω is well-known. If $s \in T^*M$ is a cotangent vector at $m \in M$ and π is the projection $T^*M \longrightarrow M$, and if ξ is a tangent vector to T^*M at s, then $\omega(\xi) = \langle d\pi_m(\xi), s \rangle$. Since $d\omega = \sum dp_i \wedge dq_i$ locally, $d\omega$ is nondegenerate, i.e., S is *symplectic*. At each point of S we thus have a nondegenerate bilinear form on the tangent space to S at that point, giving rise to an isomorphism of the tangent and cotangent spaces at that point. Hence there is a natural map from the space of 1-forms on S to the space of vector fields on S, $\lambda \longmapsto \lambda^\sim$. In local coordinates we have $dp_i^\sim = \partial/\partial q_i, dq_i^\sim = -\partial/\partial p_i$. If H is a real function on S then we have the vector field $X_H := (dH)^\sim$ which generates a dynamical group (at least for small time locally). Vector fields of this type are called *Hamiltonian,* and H is called the *Hamiltonian* of the dynamical system. In local coordinates (q, p) the equations of motion for a path $x(t \longmapsto x(t))$ are given by

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i} \qquad (1 \leq i \leq n).$$

Notice that the map

$$H \longmapsto X_H$$

has only the space of constants as its kernel. Thus the dynamics determines the Hamiltonian function up to an additive constant. The function H is constant on the dynamical trajectories and so is a *preserved quantity*; it is the *energy* of the system. More generally, physical observables are real functions, generally smooth, on T^*M, and form a real commutative algebra. If U is a vector field on M, then one can view U as a function on T^*M which is linear on each cotangent space. These are the so-called *momentum observables*. If (u_t) is the (local) one-parameter group of diffeomorphisms of M generated by U, then U, viewed as a function on T^*M, is the momentum corresponding to this group of symmetries of M. For $M = \mathbf{R}^N$ we thus have linear and angular momenta, corresponding to the translation and rotation subgroups of diffeomorphisms of M.

More generally S can be any symplectic manifold and the D_t symplectic diffeomorphisms. Locally the symplectic form can be written as $\sum dp_i \wedge dq_i$ in suitable local coordinates (Darboux's theorem). For a good introduction see the book of Arnold[4].

Lagrangian Mechanics. Here $S = TM$, the tangent bundle of M. Physical observables are the smooth real-valued functions on S and form a real commutative

algebra. The dynamical equations are generated once again by functions L on S, called *Lagrangians.* Let L be a Lagrangian, assumed to be smooth. For any path x defined on $[t_0, t_1]$ with values in S, its *action* is defined as

$$\mathcal{A}[x] = \int_{t_0}^{t_1} L(x(t), \dot{x}(t))dt.$$

The dynamical equations are obtained by equating to 0 the variational derivative of this functional for variations of x for which the values at the end points t_0, t_1 are fixed. The equations thus obtained are the well-known *Euler–Lagrange equations.* In local coordinates they are

$$\frac{\partial L}{\partial q_i} = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) \qquad (1 \leq i \leq n).$$

Heuristically one thinks of the actual path as the one for which the action is a minimum, but the equations express only the fact that that the path is an *extremum*, i.e., a stationary point in the space of paths for the action functional. The variational interpretation of these equations implies at once that the dynamical equations are coordinate independent. Under suitable conditions on L one can get a diffeomorphism of TM with T^*M preserving fibers (but in general not linear on them), and a function H_L on T^*M, such that the dynamics on TM generated by L goes over to the dynamics on T^*M generated by H_L under this diffeomorphism (*Legendre transformation*).

Most dynamical systems with finitely many degrees of freedom are subsumed under one of these two models or some variations thereof (holonomic systems); this includes celestial mechanics. The fundamental discoveries go back to Galilei and Newton, but the general coordinate independent treatment was the achievement of Lagrange. The actual solutions of specific problems is another matter; there are still major unsolved problems in this framework.

Electromagnetic field. Maxwell's equations. This is a dynamical system with an *infinite* number of degrees of freedom. In general such systems are difficult to treat because the differential geometry of infinite dimensional manifolds is not yet in definitive form, except in special cases. The theory of electromagnetic fields is one such special case because the theory is *linear*. Its description was the great achievement of Maxwell who built on the work of Farady. The fundamental objects are the electric field $\mathbf{E} = (E_1, E_2, E_3)$ and the magnetic field $\mathbf{B} = (B_1, B_2, B_3)$ which are functions on space depending on time and so may be viewed as functions on spacetime $\mathbf{R}^4$. In vacuum, i.e., in regions where there are no sources present,

these are governed by Maxwell's equations (in units where c, the velocity of light in vacuum, is 1):

$$\frac{d\mathbf{B}}{dt} = -\nabla \times \mathbf{E}, \qquad \nabla \cdot \mathbf{B} = 0 \tag{1}$$

and

$$\frac{d\mathbf{E}}{dt} = \nabla \times \mathbf{B}, \qquad \nabla \cdot \mathbf{E} = 0. \tag{2}$$

Here the operators ∇ refer only to the space variables. Notice that the equations (1) become the equations (2) under the duality transformation

$$(\mathbf{E}, \mathbf{B}) \longmapsto (-\mathbf{B}, \mathbf{E}).$$

To describe these equations concisely t is customary to introduce the *electromagnetic tensor* on spacetime given by the 4×4 skewsymmetrix matrix

$$\begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -B_3 & B_2 \\ -E_2 & B_3 & 0 & -B_1 \\ -E_3 & -B_2 & B_1 & 0 \end{pmatrix}.$$

It is actually better to work with the exterior 2-form

$$F = E_1 dt \wedge dx + \ldots - B_1 dy \wedge dz - \ldots$$

where $\ldots$ means cyclic summation in x, y, z. Then it is easily verified that the system of equations (1) is equivalent to $dF = 0$.

To describe the duality that takes (1) to (2) we need some preparation. For any vector space V of dimension n over the reals equipped with a nondegenerate scalar product $(\cdot, \cdot)$ of arbitrary signature, we have nondegenerate scalar products defined on all the exterior powers $\Lambda^r(V) = \Lambda^r$ by

$$(v_1 \wedge \ldots \wedge v_r, w_1 \wedge \ldots \wedge w_r) = \det((v_i, w_j))_{1 \leq i,j \leq r}.$$

We choose an orientation for V and define $\tau \in \Lambda^n$ by

$$\tau = v_1 \wedge \ldots \wedge v_n$$

where (v_i) is an oriented orthogonal basis for V with $(v_i, v_i) = \pm 1$ for all i; τ is independent of the choice of such a basis. Then the Hodge duality $*$ is a linear isomorphism of Λ^r with Λ^{n-r} defined by

$$a \wedge *b = (a, b)\tau \qquad (a, b \in \Lambda^r).$$

If M is a pseudo Riemannian manifold which is oriented, the above definition gives rise to a $*$-operator smooth with respect to the points of M that maps r-forms to $n-r$-forms and is linear over $C^\infty(M)$. In our case we take V to be the dual to $\mathbf{R}^4$ with the quadratic form $(x^0)^2-(x^1)^2-(x^2)^2-(x^3)^2$ where we write the dual basis as dx^μ. Then for $\tau = dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3$ we have

$$*dx^\mu \wedge dx^\nu = \varepsilon_\mu \varepsilon_\nu dx^\rho \wedge dx^\sigma$$

with $(\mu\nu\rho\sigma)$ an even permutation of (0123), the ε_μ being the metric coefficients, being 1 for $\mu = 0$ and -1 for $\mu = 1,2,3$. Now we regard $\mathbf{R}^4$ as a pseudo Riemannian manifold with metric $dt^2-dx^2-dy^2-dz^2$, and extend the $*$-operator defined above to a $*$–operator, linear over $C^\infty(\mathbf{R}^4)$ and taking 2-forms to 2-forms. In particular

$$*dt \wedge dx = -dy \wedge dz, \qquad *dy \wedge dz = dt \wedge dx$$

with similar formulae obtained by cyclically permuting x, y, z. Then $*F$ is obtained from F by the duality map $(\mathbf{E}, \mathbf{B}) \longmapsto (-\mathbf{B}, \mathbf{E})$. So the two sets of Maxwell equations are equivalent to

$$dF = 0, \qquad d*F = 0.$$

In this coordinate independent form they make sense on any pseudo Riemannian manifold of dimension 4. F is the *electromagnetic field.*

The Maxwell equations on $\mathbf{R}^4$, or more generally, on any convex open set $\Omega \subset \mathbf{R}^4$, can be written in a simpler form. First, all closed forms on Ω are exact and so we can write $F = dA$ where A is a 1-form. It is called the *four vector potential.* It is not unique and can be replaced by $A + d\alpha$ where α is a scalar function. The classical viewpoint is that only F is physically significant and the introduction of A is to be thought of merely as a mathematical device. A functional dependent on A will define a physical quantity only if it is unchanged under the map $A \longmapsto A + d\alpha$. This is the *principle of gauge invariance.* The field equations are the Euler–Lagrange equations for the action

$$\mathcal{A}[A] = -\frac{1}{2}\int (dA \wedge *dA)\; d^4x = \frac{1}{2}\int \left(E^2 - B^2\right)\; dtdxdydz.$$

The Maxwell equations on Ω can now be written in terms of A. Let us take the coordinates as $(x^\mu)(\mu = 0,1,2,3)$ where x^0 denotes the time and the $x^i (i = 1,2,3)$ the space coordinates. Then

$$A = \sum_\mu A_\mu dx^\mu, \qquad F = \sum_{\mu<\nu} F_{\mu\nu} dx^\mu \wedge dx^\nu \qquad F_{\mu\nu} = A_{\nu,\,\mu} - A_{\mu,\,\nu}$$

with the usual convention that $f_{,\,\mu} = \partial f/\partial x^\mu$. Then, writing $F^{\mu\nu} = \varepsilon_\mu \varepsilon_\nu F_{\mu\nu}$ with the ε_μ as above, the equation $d * F = 0$ can be checked to be the same as

$$\sum_\nu F^{\mu\nu}_{\ ,\,\nu} = \sum_\nu \frac{\partial F^{\mu\nu}}{\partial x^\nu} = 0 \qquad (\mu = 0, 1, 2, 3).$$

Let us now introduce the *Lorentz divergence* of $f = (f_\mu)$ given by

$$\mathrm{div}_L f = \sum_\mu \varepsilon_\mu \frac{\partial f_\mu}{\partial x_\mu}.$$

Then, writing

$$\mathcal{D} = \partial_0^2 - \partial_1^2 - \partial_2^2 - \partial_3^2, \qquad \partial_\mu = \frac{\partial}{\partial x^\mu},$$

the Maxwell equations become

$$\mathcal{D} A_\mu = (\mathrm{div}_L A)_{,\,\mu} \qquad (\mu = 0, 1, 2, 3).$$

Now from general theorems of PDE one knows that on any convex open set Ω, any constant coefficient differential operator $P(D)$ has the property that the map $u \longmapsto P(D)u$ is surjective on $C^\infty(\Omega)$. Hence we can find α such that $\mathcal{D}\alpha = -\mathrm{div}_L A$. Changing A to $A + d\alpha$ and writing A in place of $A + d\alpha$, the Maxwell equations are equivalent to

$$\mathcal{D} A_\mu = 0, \qquad \mathrm{div}_L A = 0, \qquad (\mu = 0, 1, 2, 3).$$

The condition

$$\mathrm{div}_L A = 0$$

is called the *Lorentz gauge.* Notice however that A is still not unique; one can change A to $A + d\alpha$ where $\mathcal{D}\alpha = 0$ without changing F while still remaining in the Lorentz gauge.

In classical electrodynamics it is usually not emphasized that the vector potential A may not always exist on an open set Ω unless the second De Rham cohomology of Ω vanishes, i.e., $H^{2,\mathrm{DR}}(\Omega) = 0$. If this condition is not satisfied, the study of the Maxwell equations have to take into account the *global topology* of Ω. Dirac was the first to treat such situations when he constructed the electrodynamics of a stationary magnetic monopole in a famous paper[1]. Then in 1959 Aharanov and Bohm suggested that there may be quantum electrodynamic effects in a nonsimply connected region even though the electromagnetic field is 0. They suggested that this is due to the fact that although the vector potential is locally zero, because of

its multiple-valued nature, the topology of the region is responsible for the physical effects and hence that the vector potential must be regarded as having physical significance. Their suggestion was verified in a beautiful experiment done by Chambers in 1960[3]. This link between electrodynamics and global topology has proved to be a very fertile one in recent years.

Returning to the convex open Ω above, the invariance of the Maxwell equations under the Poincaré group is manifest. However we can see this also in the original form involving F:

$$dF = 0, \qquad d * F = 0.$$

The first equation is invariant under *all* diffeomorphisms. The second is invariant under all diffeomorphisms that leave $*$ invariant, in particular under diffeomorphisms preserving the metric. So there is invariance under the Poincaré group. But even more is true. It can be shown that diffeomorphisms that change the metric by a positive scalar function also leave the Maxwell equations invariant. These are the *conformal transformations.* Thus the Maxwell equations are invariant under the *conformal group.* This was first noticed by Weyl and was the starting point of his investigations that led to his discovery of gauge theories.

Conformal invariance of Maxwell's equations. It may not be out of place to give the simple calculation showing the conformal invariance of the Maxwell equations. It is a question of showing that on a vector space V with a metric g of even dimension $2n$ and of arbitrary signature, the $*$-operators for g and $g' = cg$ ($c > 0$), denoted by $*$ and $*'$, are related on k-forms by

$$*' = c^{k-n} * \qquad (*)$$

so that, when $k = n$ we have

$$*' = *.$$

Thus if M, M' are oriented pseudo Riemannian manifolds of even dimension $2n$ and $f(M \simeq M')$ is a conformal isomorphmism, then for forms F, F' of degree n on M and M' respectively with $F = f^*(F')$, we have

$$f^*(*F') = *F.$$

So

$$d * F' = 0 \Leftrightarrow d * F = 0$$

which is what we want to show. To prove $(*)$ let (v_i) be an oriented orthogonal basis of V for g with $g(v_i, v_i) = \pm 1$ and let $\tau = v_1 \wedge \ldots \wedge v_{2n}$. Let $g' = cg$ where

$c > 0$. Then $(v'_i = c^{-1/2} v_i)$ is an orthogonal basis for g' with $g'(v'_i, v'_i) = \pm 1$ and $\tau' = v'_1 \wedge \ldots \wedge v'_{2n} = c^{-n}\tau$. Hence if a, b are elements of $\Lambda^k V$, then

$$a \wedge *'b = g'(a,b)\tau' = c^{k-n} g(a,b)\tau = c^{k-n} a \wedge *b$$

so that

$$a \wedge *'b = c^{k-n} a \wedge b.$$

This gives $(*)$ at once.

The fact that the Maxwell equations are not invariant under the Newtonian (Galilean) transformations connecting inertial frames was one of the major aspects of the crisis that erupted in fundamental classical physics towards the end of the 19^{th} century. Despite many contributions from Lorentz, Poincaré, and others, the situation remained murky till Einstein clarified the situation completely. His theory of special relativity, special because only inertial frames were taken into account, developed the kinematics of spacetime events on the sole hypothesis that the speed of light does not depend on the motion of the light source. Then spacetime becomes an *affine* space with a distinguished nondegenerate quadratic form of signature $(+,-,-,-)$. The automorphisms of spacetime are then the elements of the Poincaré group and the Maxwell equations are invariant under these. We shall take a more detailed look into these matters later on in this chapter.

Gravitational field. Einstein equations. Special relativity was discovered by Einstein in 1905. Immediately afterward Einstein began his quest of freeing relativity from the restriction to inertial frames so that gravitation could be included. The culmination of his efforts was the creation in 1917 of theory of *general relativity*. Spacetime became a smooth manifold with a pseudo Riemannian metric $ds^2 = \sum_{\mu\nu} g_{\mu\nu} dx^\mu dx^\nu$ of signature $(+,-,-,-)$. The most fantastic aspect of the general theory is the fact that *gravitation is now a purely geometric phenomenon, a manifestation of the curvature of spacetime*. Einstein interpreted the $g_{\mu\nu}$ as the *gravitational potentials* and showed that in matter-free regions of spacetime they satisfy

$$R_{ij} = 0$$

where R_{ij} are the components of the Ricci tensor. These are the *Einstein equations*. Unlike the Maxwell equations they are *non linear* in the $g_{\mu\nu}$. Physicists regard the Einstein theory of gravitation as the most perfect physical theory ever invented.

1.3. Principles of quantum mechanics. The beginning of the 20^{th} century also witnessed the emergence of a second crisis in classical physics. This was in the realm of atomic phenomena when refined spectroscopic measurements led to

results that showed that the stability of atoms, and hence of all matter, could not be explained on the basis of classical electrodynamics; indeed, according to classical electrodynamics, a charged particle revolving around a nucleus will radiate and hence continually lose energy, forcing it to revolve in a steadily diminishing radius, so that it will ultimately fall into the nucleus. This crisis was resolved only in 1925 when Heisenberg created quantum mechanics. Shortly thereafter a number of people including Heisenberg, Dirac, and Schrödinger established the fundamental features of this entirely new mechanics, which was more general and more beautiful than classical mechanics and gave a complete and convincing explanation of atomic phenomena.

The most basic feature of atomic physics is that when one makes a measurement of some physical observable in an atomic system, the act of measurement disturbs the system in a manner that is not predictable. This is because the measuring instruments and the quantities to be measured are both of the same small size. Consequently measurements under the same conditions will not yield the same value. The most fundamental assumption in quantum theory is that we can at least obtain a *probability distribution* for the values of the observable being measured. Although in a completely arbitrary state this probability distribution will not have zero (or at least a small dispersion), in principle one can make sure that the dispersion is zero (or at least arbitrarily small); this is called *preparation of state*. However once this is done with respect to a particular observable, some other observables will have probability distributions whose dispersions are not small. This is a great departure from classical mechanics where, once the state is determined exactly (or nearly exactly), all observables take exact (or nearly exact) values. *In quantum theory there is no state in which all observables will have zero* (*or arbitrarily small*) *dispersion*. Nevertheless the mathematical model is such that the states still evolve causally and deterministically as long as measurements are not made. This mode of interpretation, called the *Copenhagen interpretation* because it was first propounded by the Danish physicist Niels Bohr and the members of his school such as Heisenberg, Pauli, and others, is now universally accepted. One of the triumphs of quantum theory and the Copenhagen interpretation was a convincing explanation of the wave-particle duality of light. We recall that in Newton's original treatise *Optiks* light was assumed to consist of particles; but later on, in the 18^{th} and 19^{th} centuries, diffraction experiments pointed unmistakably to the wave nature of light. Quantum theory resolves this difficulty beautifully. It says that light has both particle and wave properties; it is the structure of the act of measurement that determines which aspect will be revealed. In fact quantum theory goes much further and says that *all matter* has both particle and wave properties. This is a broadening of the famous Bohr principle of *complementarity*. In the remarks below

we shall sketch rapidly the mathematical model in which these statements make perfectly good sense. For discussions of much greater depth and scope one should consult the beautiful books of Dirac, Von Neumann and Weyl[4].

States, observables, and probabilities. In quantum theory states and observables are related in a manner entirely different from that of classical mechanics. The mathematical description of any quantum system is in terms of a *complex separable* Hilbert space $\mathcal{H}$; the states of the system are then the points of the *projective space* $\mathbf{P}(\mathcal{H})$ of $\mathcal{H}$. Recall that if V is any vector space, the projective space $\mathbf{P}(V)$ of V is the set of one-dimensional subspaces (*rays*) of V. Any one dimensional subspace of $\mathcal{H}$ has a basis vector ψ of norm 1, i.e., a *unit vector*, determined up to a scalar factor of absolute value 1 (called a *phase factor*). So the states are described by unit vectors with the proviso that unit vectors ψ, ψ' describe the same state if and only if $\psi' = c\psi$ where c is a phase factor. The observables are described by *self adjoint operators* of $\mathcal{H}$; we use the same letter to denote both the observable and the operator that represents it. If the observable (operator) A has a pure discrete simple spectrum with eigenvalues $a_1, a_2, \ldots$ and corresponding (unit) eigenvectors $\psi_1, \psi_2, \ldots$, then a measurement of A in the state ψ will yield the value a_i with probability $|(\psi, \psi_i)|^2$. Thus

$$\mathrm{Prob}_\psi(A = a_i) = |(\psi, \psi_i)|^2 \qquad (i = 1, 2, \ldots).$$

The complex number (ψ, ψ_i) is called the *probability amplitude,* so that quantum probabilities are computed as squares of absolute values of complex probability amplitudes. Notice that as (ψ_i) is a ON basis of $\mathcal{H}$ we must have

$$\sum_i |(\psi, \psi_i)|^2 = 1$$

so that the act of measurement is certain to produce some a_i as the value of A. It follows from many experiments (see Von Neumann's discussion of the Compton-Simons scattering experiment[3], pp. 211–215) that a measurement *made immediately after* always leads to this value a_i, so that we know that the state after the first measurement is ψ_i. In other words, while the state was arbitrary and undetermined before measurement, once we make the measurement and know that the value is a_i, we know that the state of the system has become ψ_i. This aspect of measurement, called the *collapse of the wave packet*, is also the method of *preparation of states*. We shall elucidate this remarkable aspect of measurement theory a little later, using Schwinger's analysis of Stern-Gerlach experiments. If the Hilbert space is infinite dimensional, self adjoint operators can have continuous spectra and the probability statements given above have to make use of the more sophisticated spectral theory of

such operators. In the case of an arbitrary self adjoint operator A, one can associate to it its *spectral measure* which is a projection valued measure that replaces the notion of eigenspaces. The relationship between A and P^A is given by

$$A = \int_{-\infty}^{+\infty} \lambda dP^A(\lambda).$$

In this case

$$\text{Prob}_\psi(A \in E) = ||P_E^A \psi||^2 = (P_E^A \psi, \psi) \qquad (E \subset \mathbf{R}).$$

The operators representing position and momentum are of this type, i.e., have continuous spectra. For the expectation value and dispersion (variance) of A in the state ψ we have the following formulae:

$$E_\psi(A) = (A\psi, \psi), \qquad \text{Var}_\psi(A) = ||(A - mI)\psi||^2 \qquad (m = E_\psi(A)).$$

As an extreme example of this principle, the quantity

$$|(\psi, \psi')|^2 \qquad (\text{resp.}(\psi, \psi'))$$

is the probability (resp. probability amplitude) that when the system is in the state ψ and a measurement is made to determine if the state is ψ', the state will be found to be ψ'.

The most impressive aspect of the discussion above is that the states are the *points of a projective geometry.* Physicists call this the *principle of superposition of states.* If ψ_i $(i = 1, 2, 3)$ are 3 states, ψ_3 is a *superposition* of ψ_1 and ψ_2 if and only if $[\psi_3]$ is on the line in the projective space $P(\mathcal{H})$ joining $[\psi_1]$ and $[\psi_2]$ (here $[\psi_i]$ is the point of $\mathbf{P}(\mathcal{H})$ represented by the vector ψ_i). In terms of vectors this is the same as saying that ψ_3 is a linear combination of ψ_1 and ψ_2. One should contrast this with the description of classical systems, where states are points of a set where no superposition is possible; there one can say that the states are the points of a *Boolean algebra.* The transition

$$\text{Boolean algebra} \longrightarrow \text{projective geometry}$$

is the mathematical essence of the change of description from classical to quantum that allows a mathematically and physically consistent scheme rich enough to model the unique features of quantum theory like the wave-particle duality of all matter, and more generally, the principle of complementarity.

In classical *statistical mechanics* the states are often *probability measures* on the phase space. However this is due to the fact that the huge number of degrees of freedom of the system makes it impossible to know the state exactly, and so the probability measures are a reflection of the incomplete knowledge of the actual state. The statistical nature of the description thus derives from parameters which are "hidden". By contrast, in quantum mechanics the states are already assumed to be determined with maximal precision and the statistical character is entirely intrinsic. The maximally precise states are often called *pure states* and these are the ones we have called states. In quantum statistical mechanics we encounter states with less than maximal precision, the so-called *mixed states.* These are described by what are called *density operators*, namely, operators D which are bounded, self adjoint, positive, and of trace 1. If A is an observable, its expectation value in the state D is given by

$$E_D(A) = \text{Tr } (DA) = \text{Tr } (D^{1/2}AD^{1/2}).$$

These mixed states form a *convex* set, whose extreme points are the pure states; in this picture the pure states correspond to the density operators which are the projection operators $P_{[\psi]}$ on the one-dimensional subspaces of the Hilbert space. However it should be remembered that the representation of a mixed state as a convex combination of pure states is not always unique, making the physical interpretation of mixtures a very delicate matter.

For a long time after the discovery of quantum mechanics and the Copenhagen interpretation, some people refused to accept them on the grounds that the statistical description in quantum theory is ultimately due to the *incompleteness* of the quantum state, and that a fuller knowledge of the state will remove the probabilities. This is called the *hidden variables interpretation.* Among the subscribers to this view was Einstein who never reconciled himself to the new quantum theory ("God does not play dice"), although he was one of the central figures in the quantum revolution because of his epoch-making work on the photon as a light quantum. Among his most spectacular attempts to reveal the incomplete nature of the quantum mechanical description of nature is the EPR paradox, first suggested in a famous paper by Einstein, Padolsky, and Rosen. However his views were refuted by Niels Bohr convincingly. Nowadays there is no paradox in the EPR experiment; experiments conducted everyday in high energy physics laboratories confirm convincingly that things happen as quantum theory predicts.

At the mathematical level one can ask the question whether the results of the quantum theory can be explained by a hidden parameter model. The answer is a resounding "no". The first such theorem was proved by Von Neumann; since then a

galaxy of people have examined this question under varying levels of assumptions: Mackey, Gleason, Bell,.... However the question is not entirely mathematical. For a discussion of these aspects see my book as well as the other references contained in the monumental book of Wheeler and Zurek[5] (which has reprints of most of the fundamental articles on the theory of measurement, including a complete extract of Von Neumann's treatment of the thermodynamic aspects of measurement from his book[4].

Stern–Gerlach experiments. Finite models. The discussion above is very brief and does not do full justice to the absolutely remarkable nature of the difference between classical and quantum physics. It is therefore reasonable to ask if there is a way to comprehend better these remarkable features, for instance, by a discussion that is closer to the experimental situations but somewhat simpler from a mathematical standpoint. The Hilbert space $\mathcal{H}$ of quantum theory is usually infinite dimensional because many observables of importance (position coordinates, momenta etc) have values which form a continuous range, and any discussion of the features of quantum theory rapidly gets lost among technicalities of the mathematical theory. To illustrate the striking features of quantum theory most simply and elegantly one should look at *finite models* where $\mathcal{H}$ is finite dimensional. Such models go back to Weyl[6] in the 1930's; they were revived in the 1950's by Schwinger[7], and resurrected again[8] in the 1990's. For a beautiful treatment of the foundations of quantum mechanics from this point of view see Schwinger's book[9], in particular the prologue.

The simplest such situation is the measurement of spin or the magnetic moment of an atom. The original experiments were done by Stern and Gerlach and so such measurements are known as *Stern-Gerlach measurements*. In this experiment silver pellets are heated in an oven to a very high temperature till they are vapourized, and then they are drawn out through an aperture in the oven and refined by passing through several slits. The beam is then passed through a magnetic field and then stopped on a screen. Since the silver atoms have been heated to a high temperature it is a natural assumption to make that their magnetic moments are distributed randomly. So one should expect a continuous distribution of the magnetic moments on the screen; instead one finds that the atoms are concentrated in two sharp piles of moments $+\mu$ and $-\mu$.

This kind of experiment is a typical spin measurement with two values; the measuring apparatus, in this case the magnetic field oriented in a specific direction, measures the magnetic moment along that. Of course the direction of the magnetic field is at one's disposal so that we have an example of a system where all observables have either one or two values. If we decide to stop only the $-$ beam, the $+$ beam

will pass through undeflected through a second magnetic field parallel to the first. Then one knows that the atoms in the + beam all have their spins aligned in the given direction. This is an example of what we defined earlier as preparation of state. Measurements in different directions will then lead to a more or less complete enumeration of the observables of this system. Moreover, when repeated measurements are made, we can see quite explicitly how the measurement changes the state and destroys any previous information that one has accumulated about the state. The fact that one cannot make the dispersions of all the observables simultaneously small is very clearly seen here. This is the heart of the result that the results of quantum theory do not have an interpretation by hidden variables. Indeed, the experiments suggested by Bohm for elucidating the EPR paradox are essentially spin or polarization measurements and use finite models. In fact one can even show that all states that possess the features of the EPR phenomenon are of the Bohm type or generalizations thereof[10].

From the mathematical point of view, these spin systems are examples of systems where all observables have at most N values (N is a fixed integer) and generic observables have exactly N values. The Hilbert space can then be taken to be $\mathbf{C}^N$ with the standard scalar product. The observables are then $N \times N$ Hermitian matrices whose spectra are the sets of values of these observables. The determination of states is made by measurements of observables with exactly N distinct values. If A is a Hermitian matrix with distinct eigenvalues $a_1, \ldots, a_N$ and eigenvectors $\psi_1, \ldots, \psi_N$, and a measurement of A yields a value a_i, then we can say with certainty that the state is ψ_i immediately after measurement, and it will evolve deterministically under the dynamics till another measurement is made. This is the way states are determined in quantum theory, by specifying the values (= quantum numbers) of one or more observables even in more complicated systems. Suppose B is another Hermitian matrix with eigenvalues b_i and eigenvectors ψ'_i. If A is measured and found to have the value a_i, an immediately following measurement of B will yield the values b_j with probabilities $|(\psi_i, \psi'_j)|^2$. Suppose now (this is always possible) we select B so that

$$|(\psi_i, \psi'_j)|^2 = \frac{1}{N} \qquad (1 \leq i, j \leq N),$$

then we see that in the state where A has a specific value, all values of B are equally likely and so there is minimal information about B. Pairs of observables like A and B with the above property may be called *complementary.* In the continuum limit of this model A and B will (under appropriate conditions) go over to the position and momentum of a particle moving on the real line, and one will obtain the Heisenberg uncertainty principle, namely that there is no state in which the dispersions of the position and momentum measurements of the particle are *both* arbitrarily small.

In a classical setting, the model for a system all of whose observables have at most N values (with generic ones having N values) is a set X_N with N elements, observables being real functions on X_N. The observables thus form a *real algebra* whose dimension is N. Not so in quantum theory for a similarly defined system: the states are the points of the projective space $\mathbf{P}(\mathbf{C}^N)$ and the observables are $N \times N$ Hermitian matrices which *do not form an algebra.* Rather, they are the *real* elements of a *complex* algebra with an involution $*$ (adjoint), real being defined as being fixed under $*$. The dimension of the space of observables has now become N^2; the extra dimensions are needed to accommodate complementary observables. The complex algebra itself can be interpreted, as Schwinger discovered[9], in terms of the measurement process, so that it can be legitimately called, following Schwinger, the *measurement algebra*. Finally, if A and B are two Hermitian matrices, then AB is Hermitian if and only if $AB = BA$ which is equivalent to the existence of an ON basis for $\mathbf{C}^N$ whose elements are simultaneous eigenvectors for both A and B; in the corresponding states both A and B can be measured with zero dispersion. Thus commutativity of observables is equivalent to *simultaneous observability.* In classical mechanics *all observables are simultaneously observable.* This is spectacularly false in quantum theory.

Although the quantum observables do not form an algebra they are the real elements of a complex algebra. Thus one can say that *the transition from classical to quantum theory is achieved by replacing the commutative algebra of classical observables by a complex algebra with involution whose real elements form the space of observables of the quantum system*[11]. By abuse of language we shall refer to this complex algebra itself as the *observable algebra.*

The preceding discussion has captured only the barest essentials of the foundations of quantum theory. However, in order to understand the relation between this new mechanics and classical mechanics it is essential to encode into the new theory the fact which is characteristic of quantum systems, namely, that they are really microscopic; what this means is that the quantum of action, namely *Planck's constant* $\hbar$ really defines the boundary between classical and quantum. In situations where we can neglect $\hbar$, quantum theory may be replaced by classical theory. For instance, the commutation rule between position and momentum, namely

$$[p, q] = -i\hbar$$

goes over to

$$[p, q] = 0$$

when $\hbar$ is 0. Therefore a really deeper study of quantum foundations must bring in $\hbar$ in such a way that the noncommutative quantum observable algebra depending

on $\hbar$, now treated as a *parameter,* goes over in the limit $\hbar \to 0$ to the commutative algebra of classical observables (complexified). Thus *quantization,* by which we mean the transition from a classically described system to a "corresponding quantum system", is viewed as a *deformation* of the classical commutative algebra into a noncommutative quantum algebra. However one has to go to *infinite dimensional* algebras to truly exhibit this aspect of quantum theory[12].

Remark. Occasionally there arise situations where the projective geometric model given above has to be modified. Typically these are contexts where there are *superselection observables.* These are observables which are simultaneously measurable with *all observables.* (In the usual model above only the constants are simultaneously measurable with every observable.) If all superselection observables have specific values, the states are again points of a projective geometry; the choice of the values for the superselection observables is referred to as a *sector.* The simplest example of such a situation arises when the Hilbert space $\mathcal{H}$ has a decomposition

$$\mathcal{H} = \bigoplus_j \mathcal{H}_j$$

and only those operators of $\mathcal{H}$ are considered as observables that commute with all the orthogonal projections

$$P_j : \mathcal{H} \longrightarrow \mathcal{H}_j.$$

The *center* of the observable algebra is then generated by the P_j. Any real linear combination of the P_j is then a superselection observable. The states are then rays which *lie in some* $\mathcal{H}_j$. So we can say that the states are points of the *union*

$$\bigcup_j \mathbf{P}(\mathcal{H}_j).$$

This situation can be generalized. Let us keep to the notation above but require that for each j there is a $*$-algebra $\mathcal{A}_j$ of operators on $\mathcal{H}_j$ which is isomorphic to a full finite dimensional matrix $*$-algebra such that the observables are those operators that leave the $\mathcal{H}_j$ invariant and whose restrictions to $\mathcal{H}_j$ commute with $\mathcal{A}_j$. It is not difficult to see that we can write

$$\mathcal{H}_j \simeq V_j \otimes \mathcal{K}_j \qquad (\dim(V_j) < \infty)$$

where $\mathcal{A}_j$ acts on the first factor and observables act on the second factor, with $\mathcal{A}_j$ isomorphic to the full $*$–algebra of operators on V_j, so that the observable algebra

on $\mathcal{H}_j$ is isomorphic to the full operator algebra on $\mathcal{K}_j$. In this case the states may be identified with the elements of

$$\bigcup_j \mathbf{P}(\mathcal{K}_j).$$

Notice that once again we have a *union* of projective geometries. Thus, between states belonging to different $\mathbf{P}(\mathcal{K}_j)$ *there is no superposition.* The points of $\mathbf{P}(\mathcal{K}_j)$ are the *states in the $\mathcal{A}_j$-sector.*

The above remarks have dealt with only the simplest of situations and do not even go into quantum mechanics. More complicated systems like quantum field theory require vastly more mathematical preparation.

One final remark may be in order. The profound difference between classical and quantum descriptions of states and observables makes it important to examine whether there is a deeper way of looking at the foundations that will provide a more natural link between these two pictures. This was done for the first time by Von Neumann and then, after him, by a whole host of successors. Let $\mathcal{O}$ be a complex algebra with involution $*$ whose real elements represent the bounded physical observables. Then for any state of the system we may write $\lambda(a)$ for the *expectation value* of the observable a in that state. Then $\lambda(a^n)$ is the expectation value of the observable a^n in the state. Since the moments of a probability distribution with compact support determine it uniquely it is clear that we may *identify* the state with the corresponding functional

$$\lambda : a \longmapsto \lambda(a).$$

The natural assumptions about λ are that it be linear and positive in the sense that $\lambda(a^2) \geq 0$ for any observable a. Both of these are satisfied by complex linear functions λ on $\mathcal{O}$ with the property that $\lambda(a^*a) \geq 0$. Such functionals on $\mathcal{O}$ are then called *states*. To obtain states one starts with a *$*$-representation* ρ of $\mathcal{O}$ by operators in a Hilbert space and then define, for some unit vector ψ in the Hilbert space, the state by

$$\lambda(a) = (\rho(a)\psi, \psi).$$

It is a remarkable fact of $*$-representations of algebras with involution that under general circumstances *any* state comes from a pair (ρ, ψ) as above, and that if we require ψ to be *cyclic*, then the pair (ρ, ψ) is unique up to unitary equivalence. Thus quantum the descriptions of states and observables are essentially inevitable; the only extra assumption that is made, which is a natural simplifying one, is that there is a *single* representation, or a *single* Hilbert space, whose vectors represent the states. For more details see my book[5].

1.4. Symmetries and projective unitary representations. The notion of a symmetry of a quantum system can be defined in complete generality.

Definition. A *symmetry* of a quantum system with $\mathcal{H}$ as its Hilbert space of states is any bijection of $\mathbf{P}(\mathcal{H})$ that preserves $|(\psi, \psi')|^2$.

For any $\psi \in \mathcal{H}$ which is nonzero let $[\psi]$ be the point of $\mathbf{P}(\mathcal{H})$ it defines and let

$$p([\psi], [\psi']) = |(\psi, \psi')|^2.$$

Then a symmetry s is a bijection

$$s : \mathbf{P}(\mathcal{H}) \longrightarrow \mathbf{P}(\mathcal{H})$$

such that

$$p(s[\psi], s[\psi']) = p([\psi], [\psi']) \qquad (\psi, \psi' \in \mathcal{H}).$$

Suppose U is a unitary (resp. antiunitary) operator of $\mathcal{H}$; this means that U is a linear (resp. antilinear) bijection of $\mathcal{H}$ such that

$$(U\psi, U\psi') = (\psi, \psi') \qquad ((U\psi, U\psi') = (\psi', \psi).$$

Then

$$[\psi] \longmapsto [U\psi]$$

is a symmetry. We say that the symmetry is *induced* by U; the symmetry is called unitary or antiunitary according as U is unitary or antiunitary. The fundamental theorem on which the entire theory of symmetries is based is the following[13]:

Theorem (Wigner) 1.4.1. *Every symmetry is induced by a unitary or antiunitary operator of $\mathcal{H}$ which moreover is determined uniquely upto multiplication by a phase factor. The symmetries form a group and the unitary ones a normal subgroup of index* 2.

This theorem goes to the heart of the reason why quantum theory is *linear*. The ultimate reason for this is the superposition principle or the fact that the states form the points of a projective geometry, so that the automorphisms of the set of states arise from linear or conjugate linear transformations. Recently people have been exploring the possibility of *nonlinear extensions of quantum mechanics*. Of course such extensions cannot be made arbitrarily and must pay attention to the remarkable structure of quantum mechanics. Some of these attempts are very interesting[14].

Let us return to Wigner's theorem and some of its consequences. Clearly the square of a symmetry is always unitary. The simplest and most typical example of an antiunitary symmetry is the map

$$f \longmapsto f^{\mathrm{conj}} \qquad (f \in L^2(\mathbf{R})).$$

Suppose that G is a group which acts as a group of symmetries and that G is generated by squares. Then every element of G acts as a unitary symmetry. Now, if G is a Lie group, it is known that the connected component of G is generated by elements of the form $\exp X$ where X lies in the Lie algebra of G. As $\exp X = (\exp X/2)^2$ it follows that every element of the connected component of G acts as a unitary symmetry. We thus have the corollary:

Corollary 1.4.2. *If G is a connected Lie group and $\lambda : g \longmapsto \lambda(g)(g \in G)$ is a homomorphism of G into the group of symmetries of $\mathcal{H}$, then for each g there is a unitary operator $L(g)$ of $\mathcal{H}$ such that $\lambda(g)$ is induced by $L(g)$.*

If one makes the choice of $L(g)$ for each $g \in G$ in some manner, one obtains a map

$$L : g \longmapsto L(g) \qquad (g \in G)$$

which cannot in general be expected to be a unitary representation of G in $\mathcal{H}$. Recall here that to say that a map of a topological group G into $\mathcal{U}(\mathcal{H})$ of a Hilbert space $\mathcal{H}$ is a representation is to require that L is a continuous homomorphism of G into the unitary group $\mathcal{U}(\mathcal{H})$ of $\mathcal{H}$ equipped with its strong operator topology. The continuity is already implied (when G and $\mathcal{H}$ are separable) by the much weaker and almost always fulfilled condition that the maps $\varphi, \psi \mapsto (L(g)\varphi, \psi)$ are Borel. In the case above, we may obviously assume that $L(1) = 1$; as $\lambda(gh) = \lambda(g)\lambda(h)$ we have

$$L(g)L(h) = m(g,h)L(gh) \qquad (|m(g,h)| = 1).$$

Now, although $L(g)$ is not uniquely determined by $\lambda(g)$, its image $L^{\sim}(g)$ in the projective unitary group $\mathcal{U}(\mathcal{H}/\mathbf{C}^{\times}1$ is well-defined. We shall always assume that the action of G is such that the map $L^{\sim}$ is continuous. The continuity of $L^{\sim}$, and hence the continuity of the action of G, is guaranteed as soon as the maps $g \longmapsto |(L(g)\varphi, \psi)|$ are Borel. Given such a continuous action one can always choose the $L(g)$ such that $g \longmapsto L(g)$ from G to $\mathcal{U}(\mathcal{H})$ is *Borel*. L is then called a *projective unitary representation* of G in $\mathcal{H}$. In this case the function m above is Borel. Thus symmetry actions correspond to projective unitary representations of G. The function m is called the *multiplier* of L; since we can change $L(g)$ to $c(g)L(g)$ for each g, c being a Borel map of G into the unit circle, m is only significant upto

multiplication by a function $c(g)c(h)/c(gh)$, and L will be called *unitarizable* if we can choose c so that cL is a unitary representation in the usual sense.

If $G^\sim$ is a locally compact second countable topological group and $C \subset G^\sim$ is a closed normal subgroup and $G = G^\sim/C$, then any unitary representation of $G^\sim$ which takes elements of C into scalars (scalar on C) gives rise to a projective unitary representation of G because for any $g \in G$ all the unitaries of elements above g differ only by scalars. If C is *central*, i.e., if the elements of C commute with all elements of $G^\sim$, and if the original representation of $G^\sim$ is irreducible, then by Schur's lemma the representation is scalar on C and so we have a projective unitary representation of G. $G^\sim$ is called a *central extension* of G if $G = G^\sim/C$ where C is central. It is a very general theorem that for any locally compact second countable group G every projective unitary representation arises only in this manner, C being taken as the circle group, although $G^\sim$ will in general depend on the given projective representation of G.

Suppose G is a connected Lie group and $G^\sim$ is its simply connected covering group with a given covering map $G^\sim \longrightarrow G$. The kernel F of this map is a *discrete* central subgroup of $G^\sim$; it is the fundamental group of G. Although every irreducible unitary representation of $G^\sim$ defines a projective unitary representation of G, not every projective unitary representation of G can be obtained in this manner; in general there will be irreducible projective unitary representations of G which are not unitarizable even after being lifted to $G^\sim$. However in many cases we can construct a *universal central extension* $G^\sim$ such that *all* projective irreducible representations of G are induced as above by unitary representations of $G^\sim$.

This situation is in stark contrast with what happens for finite dimensional representations, unitary or not. A projective finite dimensional representation of a Lie group G is a smooth morphism of G into the projective group of some vector space, i.e., into some $\mathrm{PGL}(N, \mathbf{C})$. It can then be shown that the lift of this map to $G^\sim$ is renormalizable to an ordinary representation, which will be unique upto multiplication by a character of $G^\sim$, i.e., a morphism of $G^\sim$ into $\mathbf{C}^\times$.

Projective representations of finite groups go back to Schur. The theory for Lie groups was begun by Weyl but was worked out in a definitive manner by Bargmann for Lie groups and Mackey for general locally compact second countable groups[5,15].

We shall now give some examples that have importance in physics to illustrate some of these remarks.

$G = \mathbf{R}$ or the circle group S^1: A projective unitary representation of S^1 is also one for $\mathbf{R}$ and so we can restrict ourselves to $G = \mathbf{R}$. In this case *any projective unitary representation can be renormalized to be a unitary representation*. In

particular, the *dynamical evolution*, which is governed by a projective unitary representation D of $\mathbf{R}$, is given by an ordinary unitary representation of $\mathbf{R}$; by Stone's theorem we then have

$$D : t \longmapsto e^{itH} \qquad (t \in \mathbf{R})$$

where H is a self adjoint operator. Since

$$e^{it(H+k)} = e^{itk} e^{itH}$$

where k is a real constant, the change $H \longmapsto H + kI$ does not change the corresponding projective representation and so does not change the dynamics. However this is the extent of the ambiguity. H is the *energy* of the system (recall that self-adjoint operators correspond to observables). Exactly as in Hamiltonian mechanics, dynamical systems are generated by the energy observables, and the observable is determined by the dynamics up to an additive constant.

$G = \mathbf{R}^2$: It is no longer true that all projective unitary representations of G are unitarizable. Indeed, the commutation rules of Heisenberg, as generalized by Weyl, give rise to an *infinite dimensional irreducible* projective unitary representation of G. Since irreducible unitary representations of an abelian group are of dimension 1, such a projective unitary representation cannot be unitarized. Let $\mathcal{H} = L^2(\mathbf{R})$. Let Q, P be the position and momentum operators, i.e.,

$$(Qf)(x) = xf(x), \qquad (Pf)(x) = -i\frac{df}{dx}.$$

Both of these are *unbounded* and so one has to exercise care in thinking of them as self adjoint operators. The way to do this is to pass to the unitary groups generated by them. Let

$$U(a) : f(x) \longmapsto e^{iax} f(x), \qquad V(b) : f(x) \longmapsto f(x+b) \qquad (a, b \in \mathbf{R}).$$

These are both one-parameter unitary groups and so by Stone's theorem they can be written as

$$U(a) = e^{iaQ'}, \qquad V(b) = e^{ibP'} \qquad (a, b \in \mathbf{R})$$

where Q', P' are self-adjoint; we define $Q = Q', P = P'$. A simple calculation shows that

$$U(a)V(b) = e^{-iab} V(b)U(a).$$

So, if

$$W(a, b) = e^{iab/2} U(a)V(b),$$

(the exponential factor is harmless and is useful below) then we have

$$W(a,b)W(a',b') = e^{i(a'b-ab')/2}W(a+a',b+b'),$$

showing that W is a projective unitary representation of $\mathbf{R}^2$. If a bounded operator A commutes with W, its commutativity with U implies that A is multiplication by a bounded function f, and then its commutativity with V implies that f is invariant under translations, so that f is constant, i.e., A is a scalar. So W is irreducible.

The multiplier of W arises directly out of the symplectic structure of $\mathbf{R}^2$ regarded as the *classical phase space* of a particle moving on $\mathbf{R}$. Thus *quantization* may be viewed as passing from the phase space to a projective unitary representation canonically associated to the symplectic structure of the phase space. This was Weyl's point of view.

$G = \mathrm{SO}(3)$, $G^\sim = \mathrm{SU}(2)$: Rotational symmetry is of great importance in the study of atomic spectra. $G^\sim = \mathrm{SU}(2)$ operates on the space of 3×3 Hermitian matrices of trace 0 by $g, h \longmapsto ghg^{-1}$. The Hermitian matrices of trace 0 can be written as

$$h = \begin{pmatrix} x_3 & x_1 - ix_2 \\ x_1 + ix_2 & -x_3 \end{pmatrix}.$$

Since

$$\det(h) = -(x_1^2 + x_2^2 + x_3^2)$$

is preserved, the action of any element of SU(2) lies in O(3) and so we have a map $G^\sim \longrightarrow \mathrm{O}(3)$. Its kernel is easily checked to be $\{\pm 1\}$. Since $G^\sim$ is connected, its image is actually in SO(3) and as the kernel of the map has dimension 0, the image of SU(2) is also of dimension 3. As SO(3) also has dimension 3 the map is surjective. We thus have an exact sequence

$$1 \longrightarrow \{\pm 1\} \longrightarrow \mathrm{SU}(2) \longrightarrow \mathrm{SO}(3) \longrightarrow 1.$$

Now SU(2) consists of all matrices of the form

$$\begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix} \qquad (a\overline{a} + b\overline{b} = 1)$$

and so topologically $\mathrm{SU}(2) \simeq S^3$. Thus SU(2) is simply connected and the above exact sequence describes the universal covering of SO(3). If we omit, in the description of elements of SU(2), the determinant condition, we get the quaternion algebra by the identification

$$\begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix} \longmapsto a + bj \qquad (i^2 = -1, j^2 = -1, ij = -ji)$$

so that SU(2) may be viewed as the group of elements of unit norm of the quaternion algebra. For dimens ions $N > 3$ a similar description of the universal covering group of $\mathrm{SO}(N)$ is possible; the universal covering groups are the *spin groups* $\mathrm{Spin}(N)$, and they appear as the unit groups of the *Clifford algebras* which generalize quaternion algebras.

$G = \mathrm{SO}(1,3)^0$, $G^\sim = \mathrm{SL}(2,\mathbf{C})$*:* G is the *connected Lorentz group,* namely, the component of the identity element of the group $\mathrm{O}(1,3)$ of all nonsingular matrices g of order 4 preserving

$$x_0^2 - x_1^2 - x_2^2 - x_3^2.$$

Also $\mathrm{SL}(2,\mathbf{C})$ must be viewed as the *real* Lie group underlying the complex Lie group $\mathrm{SL}(2,\mathbf{C})$ so that its real dimension is 6 which is double its complex dimension which is 3; we shall omit the subscript $\mathbf{R}$ if it is clear that we are dealing with the real Lie group. We have the action $g, h \longmapsto ghg^*$ of $G^\sim$ on the space of 2×2 Hermitian matrices identified with $\mathbf{R}^4$ by writing them in the form

$$h = \begin{pmatrix} x_0 + x_3 & x_1 - ix_2 \\ x_1 + ix_2 & x_0 - x_3 \end{pmatrix}.$$

The action preserves

$$\det(h) = x_0^2 - x_1^2 - x_2^2 - x_3^2$$

and so maps $G^\sim$ into $\mathrm{O}(1,3)$. It is not difficult to check using polar decomposition that $G^\sim$ is connected and simply connected and the kernel of the map $G^\sim \longrightarrow G$ is (± 1). As in the unitary case, as $\dim\ G = \dim\ \mathrm{SO}(1,3)^0 = 6$, we have the exact sequence

$$1 \longrightarrow \{\pm 1\} \longrightarrow \mathrm{SL}(2,\mathbf{C}) \longrightarrow \mathrm{SO}(1,3)^0 \longrightarrow 1.$$

Representations of $\mathrm{SU}(2)$ **and** $\mathrm{SL}(2,\mathbf{C})$**:** Any irreducible projective unitary representation of $\mathrm{SO}(3)$ is finite dimensional and arises from an ordinary irreducible representation of $\mathrm{SU}(2)$ via the covering map $\mathrm{SU}(2) \longrightarrow \mathrm{SO}(3)$. The general representation of $\mathrm{SU}(2)$ is parameterized by a half-integer $j \in \frac{1}{2}\mathbf{Z}$ and is of dimension $2j+1$. It is the representation obtained on the space of homogeneous polynomials p in z_1, z_2 of degree $2j$ from the natural action of $\mathrm{SU}(2)$ on $\mathbf{C}^2$. It is usually denoted by D^j. The representation $D^{1/2}$ is the basic one. The parameter j is called the *spin* of the representation. The element -1 of $\mathrm{SU}(2)$ goes over to $(-1)^{2j}$ and so the representations of $\mathrm{SO}(3)$ are those for which j is itself an integer. These are the odd dimensional ones. For applications one needs the formula

$$D^j \otimes D^k = D^{|j-k|} \oplus D^{|j-k|+1} \oplus \ldots \oplus D^{j+k}.$$

This is the so-called *Clebsch–Gordan formula.*

Let us go back to the context of the Stern-Gerlach experiment in which atoms are subjected to a magnetic field. The experiment is clearly covariant under SO(3) and the mathematical description of the covariance must be through a projective unitary representation of SO(3). But the measurements of the magnetic moment are all two-valued and so the Hilbert space must be of dimension 2. So the representation must be $D^{1/2}$. Notice that *the use of projective representations is essential since* SO(3) *has no ordinary representation in dimension* 2 *other than the direct sum of two trivial representations which obviously cannot be the one we are looking for*. The space of $D^{1/2}$ is to be viewed as an *internal space* of the particle. It is to be thought of as being attached to the particle and so should move with the particle. In the above discussion the symmetry action of SU(2) is *global* in the sense that it does not depend on where the particle is. In the 1950's the physicists Yang and Mills introduced a deep generailzation of this global symmetry which they called *local symmetry.* Here the element of SU(2) which describes the internal symmetry is *allowed to depend on the spacetime point where the particle is located.* These local symmetries are then described by *functions on spacetime with values in* SU(2); they are called *gauge symmetries* and the group of all such (smooth) functions is called the *gauge group.* The fact that the internal vector space varies with the point of spacetime means that we have a *vector bundle* on spacetime. Thus the natural context for treating gauge theories is a vector bundle on spacetime.

Internal characteristics of particles are pervasive in high energy physics. They go under names such as spin, isospin, charm color, flavor, etc. In gauge theories the goal is to work with equations which are gauge-invariant, i.e., invariant under the group of gauge symmetries. Since the gauge group is infinite dimensional, this is a vast generalization of classical theory. Actually the idea of a vector space attached to points of the spacetime manifold originated with Weyl in the context of his unification of electromagnetism and gravitation. Weyl wrote down the gauge invariant coupled equations of electromagnetism and gravitation. The vector bundle in Weyl's case was a *line bundle* and so the gauge group is the group of smooth functions on spacetime with values in the unit circle, hence an abelian group. The Yang-Mills equations however involve a nonabelian gauge group[16].

Suppose now $G = \mathrm{SL}(2, \mathbf{C})$. We must remember that we have to regard this as a topological rather than a complex analytic group, or, what comes to the same thing, view it as a real Lie group. So to make matters precise we usually write this group as $\mathrm{SL}(2, \mathbf{C})_{\mathbf{R}}$, omitting the subscript when there is no ambiguity. Notice first of all that the representations D^j defined earlier by the action of SU(2) on the space of homogeneous polynomials in z_1, z_2 of degree $2j$ actually make sense for the

complex group $\mathrm{SL}(2,\mathbf{C})$; we denote these by $D^{j,0}$ and note that the representing matrices (for instance with respect to the basis $(z_1^r z_2^{2j-r})$) have entries which are *polynomials* in the entries a, b, c, d of the element of $SL(2,\mathbf{C})$. They are thus *algebraic* or *holomorphic* representations. If C is the complex conjugation on the space of polynomials, then $D^{0,j} := CD^{j,0}C^{-1}$ is again a representation of $\mathrm{SL}(2,\mathbf{C})$ but with *antiholomorphic* matrix entries. It turns out that the representations

$$D^{j,k} := D^{j,0} \otimes D^{0,k}$$

are still *irreducible* and that they are *precisely all the finite dimensional irreducible representations of* $\mathrm{SL}(2,\mathbf{C})_{\mathbf{R}}$. None of them except the trivial representation $D^{0,0}$ is unitary. This construction is typical; if G is a complex connected Lie group and $G_{\mathbf{R}}$ is G treated as a real Lie group, then the irreducible finite dimensional representations of $G_{\mathbf{R}}$ are precisely the ones

$$D \otimes \overline{E}$$

where D, E are holomorphic irreducible representations of the complex group G. In our case the restriction of $D^{0.k}$ to $\mathrm{SU}(2)$ is still D^k and so the restriction of $D^{j,k}$ to $\mathrm{SU}(2)$ is $D^j \otimes D^k$ whose decomposition is given by the Clebsch-Gordan formula.

1.5. Poincaré symmetry and particle classification. Special relativity was discovered by Einstein in 1905. Working in virtual isolation as a clerk in the Swiss patent office in Berne, Switzerland, he wrote one of the most famous and influential papers in the entire history of science with the deceptive title *On the electrodynamics of moving bodies*, and thereby changed forever our conceptions of space and time. Using beautiful but mathematically very elementary arguments he demolished the assumptions of Newton and his successors that space and time were absolute. He showed rather that time flows differently for different observers, that moving clocks are slower, and that events that are simultaneous for one observer are not in general simultaneous for another. By making the fundamental assumption that the speed of light in vacuum is constant in all (inertial) frames of reference (i.e., independent of the speed of the source of light), he showed that the change of coordinates between two inertial observers has the form

$$x' = Lx + u \qquad (x, u \in \mathbf{R}^4),$$

where L is a 4×4 real invertible matrix which preserves the quadratic form

$$(x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2;$$

here $x^0 = ct$ where t is the time coordinate, $x^i (i = 1, 2, 3)$ are the space coordinates; if the units are chosen so that the speed of light in vacuum is 1, then x^0 is the time coordinate itself. Such L are called *Lorentz transformations* and form a group denoted by $\mathrm{O}(1,3)$. The fact that the distinction between space and time, which had been a part of all of our thinking for centuries, is *dependent on the observer*, follows from these formulae. It also follows that no particle can travel with a speed greater than the speed of light in vacuum. The transformations for changing from one inertial frame to another given above form the so-called *inhomogeneous Lorentz group*; this is the set of pairs (u, L) with multiplication defined by

$$(u, L)(u', L') = (u + Lu', LL').$$

It is the *semidirect product* $\mathbf{R}^4 \times' \mathrm{O}(1,3)$. The term *Poincaré group* is usually reserved for $\mathbf{R}^4 \times' \mathrm{SL}(2,\mathbf{C})_{\mathbf{R}}$ where $\mathrm{SL}(2,\mathbf{C})_{\mathbf{R}}$ is viewed as a covering group of $\mathrm{SO}(1,3)^0$, acting on $\mathbf{R}^4$ through the covering map. $\mathrm{SO}(1,3)^0$ itself is the group of Lorentz matrices $L = (h_{\mu\nu})$ such that $\det(L) = 1$ and $h_{00} > 0$ (since $h_{00}^2 - h_{01}^2 - h_{02}^2 - h_{03}^2 = 1$, $|h_{00}| \geq 1$ always and so on $\mathrm{SO}(1,3)^0$ it is ≥ 1).

It may be of interest to add a few remarks to this brief discussion of special relativity. The idea that an observer can describe the surrounding world by 4 coordinates is the starting point of the mathematical treatment of phenomena. This description applies especially convincingly in the part of the world that is close to the observer. Already Kepler and Copernicus had realized that the laws governing the planetary movements take a simple form only when viewed against the background of the distant fixed stars. This meant that a special class of coordinate frames were singled out, namely those in which the distant stars appear to be fixed or moving with uniform speed (certainly not rotating as they would be if seen from the frame of the rotating earth). These are the so-called *inertial frames,* the ones in which Galilei's law of inertia holds, namely, objects (such as the distant stars) on which no forces are acting, are at rest or are in uniform motion in a straight line. Nowadays such frames are commonplace, for instance the frame of a rocket ship which is moving outside the earth's gravitational field so that all objects inside it are weightless, and Galelei's law of inertia is satisfied for all objects in it. Observers defining such frames are called inertial also. If now two inertial observers observe the world, the change of coordinates between their respective frames must be such that the linear character of the trajectories of objects moving uniformly without acceleration must not change. It is a consequence of results in projective geometry that such a transformation has to be *affine,* i.e., of the form

$$x' = Lx + u$$

where u refers to spacetime translation and is a vector in $\mathbf{R}^4$ and L is a real 4×4 invertible matrix. Thus *spacetime is an affine manifold.* It is important to remember that this much is already true without any assumptions on speeds of signals.

For Newton, space and time were absolute, and the space part, consisting of events that are simultaneous, formed a Euclidean space of dimension 3. Thus space time was *layered* by equal-time slices. The group of transformations between Newtonian (or Galilean) inertial frames is then the 10-parameter *Galilean group* in which L above is restricted to the group generated by spatial orthogonal transformations and *boosts.* Boosts refer to the transformations linking an observer to another who is moving with uniform velocity with respect to the first. They are of the form

$$(x^0)' = x^0, \qquad \xi' = \xi + x^0 v$$

where ξ refers to the space coordinates, and v is the velocity vector. However in the last years of the 19^{th} century there already appeared cracks in the structure of the Newtonian view of the world. The *Michelson-Morley experiment*, designed to discover the relative velocity of the earth in the ether, came up with the result that the relative velocity was 0. Many different mechanistic hypotheses were put forward to reconcile this with known theories, such as the *Lorentz-Fitzgerald contraction* which asserted that all objects contracted in the ratio

$$1 : \sqrt{1 - \frac{v^2}{c^2}}$$

along the direction of motion, v being the speed of motion and c the constant velocity of light in vacuum. On the other hand, Poincaré observed that the Maxwell equations are *not* invariant under the Galilean group but rather light behaves as if its invariance group is really the inhomogeneous Lorentz group. So a number of people sensed that some drastic changes were necessary in order to get a consistent picture of the physical world that would include electromagnetic phenomena. It was Einstein who took the decisive step; with a few simple strokes he painted a coherent picture of space and time that has been vindicated by countless experiments over the past century. Indeed, the experiments in high energy laboratories confirm everyday the assumptions of Einstein. He banished the ether, abandoned mechanistic assumptions to "justify" physical laws, and ushered in the era in which the role of the physicist was limited to building mathematical models that explain and predict phenomena. The revolution in thinking that he started was an absolutely essential prerequisite for the second great revolution in 20^{th} century science, namely quantum theory.

Spacetime with the affine structure given by $\mathbf{R}^4$ and equipped with the basic quadratic form

$$(x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2,$$

is called *Minkowski spacetime* because its clear mathematical formulation as well as a coordinate independent treatment of electrodynamics in it was first given by Minkowski in a famous talk[17]. At each point of Minkowski spacetime, the future is represented by all the points in a cone with vertex at that point, the so-called *forward light cone* which is the set of all points that can be reached by a signal emitted at that point. (In Galilean spacetime the future is the half-space of points whose time coordinate is greater than the time coordinate of the given point.) One can show (this is a beautiful result of A. D. Aleksandrov[18]) that any bijection of Minkowski spacetime which preserves this cone structure is necessarily affine

$$x' = a(Lx + u)$$

where a is a nonzero constant, L a Lorentz transformation, and $u \in \mathbf{R}^4$. The constant a cannot be asserted to be 1 if one uses only light signals in analyzing the structure of spacetime; indeed, one cannot pin down the basic quadratic form except up to a multiplicative scalar, because the points reached from the origin by light signals satisfy the equation $(x^0)^2-(x^1)^2-(x^2)^2-(x^3)^2 = 0$, which is unaltered by scalar multiplication. But if we consider material particles as well, one can show that the quadratic form is determined absolutely. Thus the transformation between two inertial observers O, O' where O' is moving uniformly in the positive direction of the x-axis of O with velocity $v > 0$ is given in the units where $c = 1$ by

$$x^{0\prime} = \frac{1}{\sqrt{1-v^2}}(x^0 - vx^1), \qquad x^{1\prime} = \frac{1}{\sqrt{1-v^2}}(-vx^1 + x^1).$$

To get the formula in the usual units one must replace v by v/c and $x^{0\prime}, x^0$ by ct', ct. It follows from this that the constant a in the earlier formula must be 1. If the direction of motion of O' is arbitrary, the transformation formula is more complicated; it was first obtained by Herglotz. All the remarkable properties of moving observers such as time dilation, space contraction, relativistic composition formula for velocities, and so on, can be derived from the above formula[19].

The fact that in the treatment of light the quadratic form is determined only up to a scalar means that description of radiation phenomena must be invariant under the much larger *conformal group.* Globally it is nothing more than adding the dilations to the Poincaré group; but conformal transformations can be fully treated only after compactifying spacetime, and then the conformal group becomes

$\mathrm{SO}(2,4)$. We shall discuss this later on. The reader can in the meantime think of the corresponding situation in complex geometry where the group of transformations $z \longmapsto az + b$ of $\mathbf{C}$ enlarges to the group of the *fractional linear transformations* $z \longmapsto (az + b)/(cz + d)$ of the extended complex plane $\mathbf{C} \cup \infty$.

Let us now return to quantum theory. To describe a quantum system in a manner compatible with special relativity means that we must have a projective unitary representation of the Poincaré group on the Hilbert space $\mathcal{H}$ of the system. It was proved by Wigner[20] in a famous paper in 1939 that *any projective unitary representation of the Poincaré group is unitarizable.* It was also in the same paper that he classified the *physically relevant* irreducible unitary representations of the Poincaré group. If G is a semidirect product $\mathbf{R}^N \times' H$ where H is a simply connected semisimple group acting on $\mathbf{R}^N$ in such a manner that there are no nonzero skewsymmetric invariant bilinear forms (if the action of H on $\mathbf{C}^N$ is irreducible and admits a nonzero symmetric invariant bilinear form then this condition is satisfied), then all projective representations of G are unitarizable; Wigner's theorem is a special case of this[5]. However there are groups for which the unitarizability theorem is not true, such as the additive groups of vector spaces of dimension ≥ 2, and more significantly, the simply connected covering group of the Galilean group. Indeed, for a given simply connected Lie group to have the property that all projective unitaries are unitarizable, the second cohomology of the group with coefficients in the circle group must vanish.

It follows from these remarks that relativistic invariance of a quantum system is encoded by a unitary representation of the Poincaré group. It is natural to postulate that if the system is that of an elementary particle then the corresponding representation should be irreducible. Thus, a classification of irreducible unitary representations of the Poincaré group will yield a classification of elementary particles that is compatible with special relativity. We shall now describe how the irreducible unitary representations of the Poincaré group are constructed.

Before taking this up I should point out that physicists do not describe symmetries as we have done using unitary representations explicitly. Most of the time the Hilbert spaces they work with contain only the most important states of the system, for instance those that are obtained by repeated application of certain key operators (creation, annihilation) on certain key states (vacuum); this core is usually invariant under the operators of the Lie algebra of the symmetry group and so only these Lie algebra operators are specified. In certain cases the symmetry group or rather its Lie algebra is infinite dimensional, such as the Virasoro or affine Lie algebras; in this case there is no natural Lie group and the symmetry is only infinitesimal.

Let P be the Poincaré group,

$$P = \mathbf{R}^4 \times' \mathrm{SL}(2, \mathbf{C}).$$

Wigner's determination of the irreducible unitary representations of P in 1939 was extended in great depth to a vast class of locally compact semidirect product groups by Mackey (the "Mackey machine"). But the basic ideas already go back to Frobenius who was a great pioneer in the theory of representations of finite groups. Let

$$G = A \times' H$$

where A is abelian, H acts on A through automorphisms. The irreducible representations are then classified by a very general scheme, depending on two "parameters" O, σ where O is an orbit of the (dual) action of H on the character group $\widehat{A}$ of A, and σ is an irreducible unitary representation of the stability subgroup in H of a point $\chi \in O$. In order to motivate the Mackey theory better I shall first discuss the case when G is finite where there are no technical complications.

Let then G be finite and L an irreducible unitary representation of G. We identify A and H with subgroups of G by the maps $a \longmapsto (a, 1)$ and $h \longmapsto (1, h)$. A will then be normal in G and H will act on it by $h, a \longmapsto hah^{-1}$. The restrictions of L to A and H are unitary representations of these groups which we write as U and V. Thus

$$L(ah) = U(a)V(h), \qquad V(h)U(a)V(h)^{-1} = U(hah^{-1}) \qquad (a \in A, h \in H).$$

Conversely, if we start with unitary representations U, V of A and H in the same Hilbert space such that

$$V(h)U(a)V(h)^{-1} = U(hah^{-1}) \qquad (a \in A, h \in H) \qquad (*)$$

then

$$L : ah \longmapsto U(a)V(h)$$

is a unitary representation of G. Thus we must try to build pairs (U, V) satisfying $(*)$ which are irreducible.

Let $\mathcal{H}$ be the (finite dimensional) Hilbert space of U, V. Since the $U(a)(a \in A)$ is a set of commuting unitaries, there is an ON basis in which all the $U(a)$ are diagonal. If v is a basis vector, $U(a)v = \chi(a)v$ and it is clear that $\chi \in \widehat{A}$, where, as we mentioned earlier, $\widehat{A}$ is the group of characters of A. So we can write

$$\mathcal{H} = \oplus_{\chi \in F} \mathcal{H}_\chi \qquad U(a)v = \chi(a)v \qquad (v \in \mathcal{H}_\chi)$$

where F is a subset of $\widehat{A}$ and the $\mathcal{H}_\chi \neq 0$. The action of H on A gives rise to the dual action of H on $\widehat{A}$ given by $h, \chi \longmapsto h \cdot \chi$ where $(h \cdot \chi)(a) = \chi(h^{-1}ah)$. Since $U(a)V(h)v = V(h)U(h^{-1}ah)v$ it follows that each $V(h)(h \in H)$ moves $\mathcal{H}_\chi$ into $\mathcal{H}_{h\cdot\chi}$. This shows that F is stable under H and that if O is an orbit contained in F, the space $\oplus_{\chi\in O}\mathcal{H}_\chi$ is invariant under both A and H and so invariant under (U, V). Since (U, V) is irreducible this means that F is a single orbit, say O. Let us fix a $\chi \in O$ and let H_χ be the subgroup of all $h \in H$ such that $h \cdot \chi = \chi$. Since $V(h)$ takes $\mathcal{H}_\chi$ to $\mathcal{H}_{h\cdot\chi}$ we see that $\mathcal{H}_\chi$ is stable under H_χ and so defines a unitary representation σ of H_χ. If W is a subspace of $\mathcal{H}_\chi$ invariant under σ, it is clear that $S[W] := \oplus_{h\in H}L(h)[W]$ is stable under V. If $W' \perp W$ is another σ-invariant subspace of $\mathcal{H}_\chi$ then $S[W] \perp S[W']$; indeed, if $hH_\chi \neq h'H_\chi$ then $V(h)[W]$ and $V(h')[W']$ are orthogonal because they belong to different $\mathcal{H}_\xi$, while for $hH_\chi = h'H_\chi$ they are orthogonal because $V(h')[W'] = V(h)[W'] \perp V(h)[W]$ from the unitarity of $V(h)$. These remarks prove that σ is irreducible. We have thus defined O, σ corresponding to L.

Notice that we can think of $\mathcal{H}$ as the collection of vector spaces $(\mathcal{H}_\xi)$ parameterized by $\xi \in O$, i.e., as a *vector bundle* over O. A *section* of the bundle is a family $(v(\xi))$ where $v(\xi) \in \mathcal{H}_\xi$ for all $\xi \in O$. Under componentwise addition these sections form a vector space which is isomorphic to $\mathcal{H}$ by the map $(v(\xi)) \mapsto \sum_\xi v(\xi)$. The action of V on $\mathcal{H}$ takes $\mathcal{H}_\xi$ to $\mathcal{H}_{h\cdot\xi}$ and so can be viewed as an action of H on the bundle compatible with its action on O. The stabilizer H_χ then acts on $\mathcal{H}_\chi$. Thus irreducible pairs (U, V) are completely determined by such vector bundles on the orbits of H. We call them H-bundles.

Suppose that we have two H-bundles $(\mathcal{H}_\xi)$ and $(\mathcal{K}_\xi)$ on O such that the representations of H_χ on $\mathcal{H}_\chi$ and $\mathcal{K}_\chi$ are equivalent by an isomorphism $v \longmapsto v'$. We claim that we can extend this to an isomorphism of the bundles that commutes with the action of H. In fact there is just one possible way to define the extension: it should take $h[v]$ to $h[v']$ for $h \in H$ and $v \in \mathcal{H}_\chi$. So the claim will be proved if we show that this is well-defined. But suppose that $h[v] = h_1[v_1]$ were $h, h_1 \in H$ and $v, v_1 \in \mathcal{H}_\chi$. Then $h \cdot \chi = h_1 \cdot \chi$ and so $h_1 = hk$ for some $k \in H_\chi$. But then $h[v] = h_1[v_1] = hk[v_1]$ so that $v = k[v_1]$; and so

$$h_1[v_1'] = hk[v_1'] = h[(k[v_1])'] = h[v'].$$

It only remains to show that any pair (O, σ) gives rise to a H-bundle for which these are the corresponding objects. We define the vector bundle $\mathcal{V}$ over O as the quotient of the trivial bundle $H \times \mathcal{H}_\chi$ on H by a natural equivalence relation which will make the quotient a vector bundle over O. More precisely,

$$\mathcal{V} = H \times \mathcal{H}_\chi / \sim$$
$$(h, v) \sim (h', v') \iff h' = hk, v' = \sigma(k)^{-1}v \text{ for some } k \in H_\chi \ .$$

Note that $(h, v) \longmapsto hH_\chi$ gives a well defined map of $\mathcal{V}$ to H/H_χ and allows us to view $\mathcal{V}$ as a vector bundle over H/H_χ. The map

$$h, (h', v) \longmapsto (hh', v)$$

the defines an action of H on $\mathcal{V}$ and converts $\mathcal{V}$ into a H-bundle.

The subgroup H_χ is called the *little group.* Thus the irreducible representations of G correspond bijectively (up to equivalence of course) to H-bundles $\mathcal{V}$ on orbits O of H in $\widehat{A}$ such that the action of the little group H_χ at a point $\chi \in O$ on the fiber at χ is irreducible. The scalar product on the fiber vector space at χ which makes the representation σ of H_χ unitary can then be transported by H to get a covariant family of scalar products $((\cdot)_\xi)$ on the fibers of $\mathcal{V}$. $\mathcal{V}$ is thus a *unitary H–bundle.* The sections of $\mathcal{V}$ then form a Hilbert space for the scalar product

$$(s, t) = \sum_{\xi \in O} (s(\xi), t(\xi))_\xi.$$

The representation L of G is then given by

$$\begin{aligned} L(ah) &= U(a)V(h) \\ (U(a)s)(\xi) &= \xi(a)s(\xi) \ (\xi \in O) \\ (V(h)s)(\xi) &= h[s(h^{-1}(\xi))] \ (\xi \in O). \end{aligned}$$

The vector bundle on $O \simeq H/H_\chi$ is determined as soon as σ is given. Indeed, we can replace H_χ by any subgroup H_0 of H. Thus, given any subgroup H_0 of H and a vector space F which is a H_0-module, there is a vector bundle $\mathcal{V}$ on H/H_0 which is a H-bundle whose fiber at the coset H_0 is F. The H-action on $\mathcal{V}$ gives rise to a representation of H on the space of sections of $\mathcal{V}$. If F has a unitary structure then $\mathcal{V}$ becomes a unitary bundle and the representation of H is unitary. This is called the *representation of H induced by σ.* We shall now give a definition of it without the intervention of the vector bundle; this will be useful later on in situations where the vector bundles are not so easily constructed. Recall that we have defined $\mathcal{V}$ as the set of equivalence classes (h, v) with $h \in H, v \in F$. A section is then a map of H into F,

$$s : h \longmapsto s(h) \qquad (h \in H, s(h) \in F)$$

such that

$$s(hk) = \sigma(k)^{-1} s(h) \qquad (k \in H_0, h \in H),$$

the corresponding section being

$$hH_0 \longmapsto [(h, s(h))]$$

where $[(h, v)]$ represents the equivalence class of (h, v). A simple calculation shows that the action of $h \in H$ on the section becomes the action

$$s \longmapsto s', \qquad s'(h') = s(h^{-1}h') \qquad (h' \in H).$$

Thus the space of sections is identified with the space F^σ of functions s from H to F satisfying

$$s(hk) = \sigma(k)^{-1}s(h) \qquad (h \in H, k \in H_0)$$

and the representation $V = V^\sigma$ is just left translation:

$$(V(h)s)(h') = s(h^{-1}h') \qquad (h, h' \in H).$$

The defining condition for F^σ is on the right and so the action from the left does not disturb it. The representation is unitary (if F is unitary) for the scalar product

$$(s, t) = \sum_{h \in H/H_0} (s(h), t(h)).$$

The sum is over the coset space H/H_0 as $(s(h), t(h))$ is really a function on H/H_0.

Apart from technical measure theoretic points the theory is the same when G is locally compact and second countable. The second countability is strictly speaking not necessary but is satisfied in all applications and so there is no sense in not imposing it. In this case the dual group $\widehat{A}$ is also locally compact abelian and second countable, and the action of H on $\widehat{A}$ is continuous. What has to be faced however is that there are in general continuum many orbits of H in $\widehat{A}$, and the *space of orbits* may not have good properties. As a result we can only say that while a given orbit and an irreducible representation of the little group of a point on that orbit still define an irreducible unitary representation of G, there will be still others if the orbit space is not nice in a measure theoretic sense. So there has to be an additional condition of regularity.

What do we mean by the requirement that the orbit space is nice? Let X be a second countable locally compact Hausdorff space on which a second countable locally compact group L acts continuously. Both X and L are separable metric and have their σ-algebras (Borel structures) of Borel sets. These σ-algebras are standard in the sense that X and L, equipped with their Borel structures, are Borel

isomorphic to the Borel space of the real line. We now introduce the space Y of the L-orbits in X and the natural map $\pi : X \longrightarrow Y$ that sends any point to the orbit containing it. We can equip Y with the σ-algebra of sets with the property that their preimages are Borel in X. One way to formulate the niceness of Y is to require that Y, with this Borel structure is standard. A more heuristic idea is to require that we can enumerate the orbits in some way, namely, that there is a Borel set in X that meets each orbit exactly once. The central theorem in the subject is a remarkable criterion due to Effros for the space of orbits to be nice in any one of these senses. We shall formulate it by first giving a definition. The action of L on X is said to be *regular* if the orbits of L in X are all locally closed. Recall here that a subset Z of a topological space Y is *locally closed* if the following equivalent conditions are satisfied:

(i) Z is open in its closure.

(ii) $Z = C \cap U$ where C is closed and U is open in Y.

(iii) For any $z \in Z$ there is an open neighborhood V of z in Y such that $Z \cap V$ is closed in V.

The significance of the condition of regularity is contained in the following theorem which is a special case of Effros's work[21] on group actions in the Polonais category.

Theorem (Effros) 1.5.1. *Let X be a locally compact second countable Hausdorff space and L a locally compact second countable group acting continuously on X. Then the following are equivalent:*

(i) *All L-orbits in X are locally closed.*

(ii) *There exists a Borel set $E \subset X$ which meets each L-orbit exactly once.*

(iii) *If $Y = L\backslash X$ is equipped with the quotient topology, the Borel structure of Y consists of all sets $F \subset Y$ whose preimages in X are Borel, and the space Y with this Borel structure is standard.*

(iv) *Every invariant measure on X which is ergodic for L is supported on an orbit and is the unique (up to a normalizing scalar) invariant measure on this orbit.*

Remark. The conditions (ii) through (iv) are the ones that say that the space of orbits is nice in a measure theoretic sense. The real depth of the Effros theorem is that this property of niceness of the orbit space, which is *global,* is equivalent to the condition (i) which is essentially *local,*; it can be verified by looking at each orbit without worrying about the others. If the group L is *compact,* then *all* orbits are closed and so the semidirect product is always regular. The action of $\mathbf{Q}$ on $\mathbf{R}$ by

$q, r \longmapsto q + r$ is not regular as Lebesgue first pointed out; indeed any set meeting each orbit exactly once is not even Lebesgue measurable. There are many other examples of this kind.

To relate this definition of regularity in our set up we shall say that the *semidirect product* $G = A \times' H$ *is regular* if the action of H on $\widehat{A}$ is regular. In order to state the main result elegantly we need the concept of *induced representations* in this general context. Let H be a locally compact second countable group and H_0 a closed subgroup; let $X = H/H_0$. For simplicity we shall assume that H/H_0 has a H-invariant measure, although everything goes through with suitable modifications in the general case. Given a unitary representation σ of H_0 in a Hilbert space F (with norm $|\cdot|$ and scalar product $(\cdot, \cdot)$) we define F^σ to be the space of all (Borel) functions (up to equality almost everywhere) s from H to F such that

$$s(hk) = \sigma(k)^{-1} s(h)$$

for each $k \in H_0$ for almost all $h \in H$, and

$$||s||^2 = \int_{H/H_0} |s(h)|^2 \, d\overline{h} < \infty$$

where $d\overline{h}$ is the invariant measure on H/H_0. Under the scalar product

$$(s|t) = \int_{H/H_0} (s(h), t(h)) d\overline{h}$$

F^σ is a Hilbert space. If

$$(V^\sigma(h)s)(h') = s(h^{-1}h') \qquad (h, h' \in H)$$

then V^σ is a unitary representation of H; it is the representation *induced* by σ.

Under additional assumptions it is possible to exhibit a more geometric definition of the induced representation. Let H be a Lie group and let σ be a unitary *finite dimensional* representation of H_0 in F. Then one can construct a smooth vector bundle on X with fibers isomorphic to F in the same way as we did in the finite case. The fact that the action of H on X has local sections implies that we have a smooth vector bundle $\mathcal{V}^\sigma$ on H/H_0 admitting an action $h, u \longmapsto h[u]$ of H such that the action of H_0 on the fiber at H_0 is just σ. Using the scalar products on F we can define the structure of a unitary bundle on $\mathcal{V}$. If we assume that X has a H-invariant measure then we can define the notion of square integrability of

sections and form F^σ, the Hilbert space of square integrable sections of $\mathcal{V}^\sigma$. Let us define

$$(V^\sigma(h)s)(x) = h[s(h^{-1}(x))] \quad (h \in H, s \in F^\sigma, x \in H/H_0).$$

Then V^σ is the induced representation we defined earlier.

Theorem (Mackey) 1.5.2. *Let $G = A \times' H$ and let O be an orbit of H in $\widehat{A}$. Fix a point χ in O and let H_χ be the stabilizer of χ in H. Let σ be a unitary irreducible representation of H_χ and let $V = V^\sigma$ be the induced representation of H. For any $a \in A$ let $U(a)$ be the unitary operator on F^σ defined by*

$$(U(a)s)(h) = (h \cdot \chi)(a)s(h) \qquad (s \in F^\sigma, h \in H, a \in A).$$

If we define

$$L(ah) = U(a)V^\sigma(h) \qquad (a \in A, h \in H)$$

then $L = L_{O,\sigma}$ is a unitary representation of G which is irreducible. If G is a regular semidirect product then every irreducible unitary representation of G is of this form. The choice of a different χ in O leads to equivalent representations. Finally,

$$L_{O,\sigma} \simeq L_{O',\sigma'} \iff O = O', \sigma \simeq \sigma'$$

(for the same choice of χ).

The subgroup H_χ is called the *little group* at χ.

Remark. Suppose that G is not a regular semidirect product. Then by the theorem of Effros there is a H-invariant measure on $\widehat{A}$ which is ergodic but gives measure 0 to all orbits. Let μ be such a measure and let $\mathcal{H} = L^2(\mu)$. Define U, V and L by $L(ah) = U(a)V(h)$ and

$$(U(a)f)(\xi) = \xi(a)f(\xi), \quad (V(h)f)(\xi) = f(h^{-1} \cdot \xi) \quad (f \in \mathcal{H}, a \in A, h \in H).$$

Then L is an irreducible unitary representation of G which does not arise from any orbit.

The Poincaré group. Here $A = \mathbf{R}^4$ with coordinates (x_μ), H is the group $\mathrm{SO}(1,3)^0$, and

$$P = \mathbf{R}^4 \times' \mathrm{SO}(1,3)^0,$$

the Poincaré group. We identify $\widehat{A}$ with a copy of $\mathbf{R}^4$ which we write as $\mathbf{P}^4$ with coordinates (p_μ), by the map $p = (p_\mu) \longmapsto \chi_p$ where $\chi_p(x) = e^{i\langle x,p\rangle}$, with

$$\langle x, p\rangle = x_0p_0 - x_1p_1 - x_2p_2 - x_3p_3.$$

$\mathbf{P}^4$ is the *momentum space*. The dual action of $\mathrm{O}(1,3)$ on $\mathbf{P}^4$ is then the same as its action on $\mathbf{R}^4$. There is one invariant, namely the quadratic form

$$p_0^2 - p_1^2 - p_2^2 - p_3^2$$

and so the level sets of this form are certainly invariant and fill up the $\mathbf{P}^4$. The orbits are obtained by splitting these level sets.

The orbits $X_m^{\pm}$: The sets $X_m^{\pm}$ are defined by

$$X_m^{\pm} = \{p_0^2 - p_1^2 - p_2^2 - p_3^2 = m^2,\ p_0 >< 0\} \qquad (m > 0).$$

These are hyperboloids inside the forward or the backward cone at the origin. Note that $p_0^2 = p_1^2 + p_2^2 + p_3^2 + m^2 > m^2$ on the orbits and so p_0 is either $> m$ or $< -m$ on any of these orbits. The point $(m, 0, 0, 0)$ is the rest frame of a particle of mass m since all the momenta are 0. The little group at $(m, 0, 0, 0)$ is the preimage of $\mathrm{SO}(3)$ in $\mathrm{SL}(2, \mathbf{C})$ and so is $\mathrm{SU}(2)$. The representations of $\mathrm{SU}(2)$ are the $D^j (j \in \frac{1}{2}\mathbf{Z})$ and the corresponding representations of P are denoted by $L_{m,j}^{\pm}$. There is an antiunitary isomorphism of $L_{m,j}^{+}$ with $L_{m,j}^{-}$ allowing the interpretation of the representations defined by the latter as the antiparticle with opposite charge. We write $L_{m,j}$ for the representation $L_{m,j}^{+}$. It describes a *massive particle,* of mass m and spin j (and, by convention, of negative charge). The representation $L_{m,1/2}$ describes any massive particle of spin $1/2$ such as the *electron*. We note also that there is an invariant measure on the orbit. There are several ways of seeing this. The simplest is to note that in the region $F = \{p_0^2 - p_1^2 - p_2^2 - p_3^2 > 0\}$ the change of coordinates

$$q_0 = p_0^2 - p_1^2 - p_2^2 - p_3^2 > 0,\ q_i = p_i\ (i = 1, 2, 3)$$

is a diffeomorphism and we have

$$d^4p = \frac{1}{2(q_0 + q_1^2 + q_2^2 + q_3^2)^{1/2}} d^4q.$$

Since q_0 is invariant under $SO(1,3)^0$ we see that for any $m > 0$ the measure

$$d\mu_m^{+} = \frac{d^3p}{2(m^2 + p_1^2 + p_2^2 + p_3^2)^{1/2}}$$

is an invariant measure on X_m^{+} where we use the $p_i (i = 1, 2, 3)$ as the coordinates for X_m^{+} through the map

$$(p_0, p_1, p_2, p_3) \longmapsto (p_1, p_2, p_3)$$

which is a diffeomorphism of X_m^+ with $\mathbf{R}^3$.

The orbits $X_0^\pm$: The sets $X_0^\pm$ are defined by

$$X_0^\pm = \{p_0^2 - p_1^2 - p_2^2 - p_3^2 = 0,\ p_0 >< 0\}.$$

We determine the little group at $(1,0,0,1)$ (as before we ignore the orbit where $p_0 < 0$). The points of X_0^+ represent particle travelling with the speed of light. Classically the only such particles are the *photons.* There is no frame where such particles are at rest, contrary to the case of the massive particles. We choose for convenience the point $(1,0,0,1)$. In our identification of $\mathbf{P}^4$ with 2×2 Hermitian matrices it corresponds to the Hermitian matrix

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

which goes into

$$\begin{pmatrix} 2a\overline{a} & 2a\overline{c} \\ 2\overline{a}c & 2c\overline{c} \end{pmatrix}$$

under the action of

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

So the little group is the group of all matrices

$$e_{a,b} = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \qquad (a, b \in \mathbf{C},\ |a| = 1).$$

This is also a semidirect product of the group of all elements $e_{1,b}$ which is isomorphic to $\mathbf{C}$, and the group of all elements $e_{a,0}$ which is isomorphic to the circle group S; the action defining the semidirect product is

$$a, b \longmapsto a^2 b.$$

So the little group at $(1,0,0,1)$ is the 2-fold cover of the Euclidean motion group of the plane, the plane being identified with $\mathbf{C}$. The only *finite dimensional* unitary irreducible representations of the little group are

$$\sigma_n : \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \longmapsto a^n \qquad (n \in \mathbf{Z}).$$

The corresponding representations are denoted by $L_{0,n}$. The representations

$$L'_{0,n} = L_{0,n} \oplus L_{0,-n} \qquad (n > 0)$$

are called *representations with helicity* $|n|$; they are not irreducible. The representation $L'_{0,2}$ describes the *photon.* The orbit X_0^+ also has an invariant measure (seen by letting $m \to 0+$ in the formula for the invariant measure on X_m^+), namely

$$d\mu_0^+ = \frac{d^3p}{2(p_1^2 + p_2^2 + p_3^2)^{1/2}}$$

is an invariant measure on X_0^+ where we use the $p_i (i = 1, 2, 3)$ as the coordinates for X_0^+ through the map

$$(p_0, p_1, p_2, p_3) \longmapsto (p_1, p_2, p_3)$$

which is a diffeomorphism of X_0^+ with $\mathbf{R}^3 \setminus \{0\}$.

The orbits Y_m: These are defined by

$$Y_m = \{p_0^2 - p_1^2 - p_2^2 - p_3^2 = -m^2\} \qquad (m > 0).$$

The little groups are not compact and these are unphysical as we shall explain a little later.

The orbit (0): The orbit is the single point 0, the origin of $\mathbf{P}^4$. The little group is all of $\mathrm{SL}(2, \mathbf{C})$, and the corresponding representations are just the irreducible unitary representations of $\mathrm{SL}(2, \mathbf{C})$ viewed as representations of P via the map $P \longrightarrow P/\mathbf{R}^4 \simeq \mathrm{SL}(2, \mathbf{C})$. These are also unphysical except for the trivial one dimensional representation which models the vacuum.

Let O denote any one of these orbits and H_0 the little group at the point described above of the orbit in question (base point). We shall presently construct smooth vector bundles V over O which are $\mathrm{SL}(2, \mathbf{C})$-bundles, namely which admit an action by $\mathrm{SL}(2, \mathbf{C})$, written $h, v \mapsto h[v]$, compatible with its action on the orbit, such that the action of the little group H_0 on the fiber at the corresponding base point is a specified irreducible unitary representation of the little group. Let μ be an invariant measure on O. Since the representation of the little group is unitary, the scalar product on the fiber at the base point can be transported by the action of H to scalar products $((\cdot, \cdot)_p, |\cdot|_p)$ on the fibers at all points of O which vary

covariantly under the action of H. V thus becomes a unitary bundle. Then the Hilbert space of the representation is the space of sections s such that

$$||s||^2 = \int_O |s(p)|_p^2 d\mu(p) < \infty$$

and the representation L of the Poincaré group is given as follows:

$$(L(a)s)(p) = e^{i\langle a,p\rangle}s(p) \qquad (a \in \mathbf{R}^4,\ p \in O)$$
$$(L(h)s)(p) = h[s(h^{-1}p)] \qquad (h \in SL(2,\mathbf{C}), p \in O).$$

In this model spacetime translations act as multiplication operators. If e_μ is the vector in spacetime with components $\delta_{\mu\nu}$, then

$$(L(te_\mu)s)(p) = e^{itp_\mu}s(p)$$

so that the momentum operators are multiplications:

$$P_\mu : s \longmapsto p_\mu s.$$

We have

$$P_0^2 - P_1^2 - P_2^2 - P_3^2 = m^2$$

which is the relativistic energy momentum relation. Thus the parameter m may be identified with the *mass* of the particle. This identification makes clear the reason why we excluded the orbits Y_m; they lead to particles with *imaginary* mass. The representations corresponding to the orbit (0) are such that the spacetime translations act trivially in them. So the energy is 0 and the only representation of this type that is physical is the trivial 1-dimensional representation, which represents the *vacuum.*

There is however a more serious omission in our discussion for the case $m = 0$. We have considered only the characters σ_n of the little group H_0. This group is a semidirect product

$$P_0 = \mathbf{C} \times' S$$

where S is the circle group acting on $\mathbf{C}$ by

$$a[b] = a^2 b,\ |a| = 1, b \in \mathbf{C}.$$

The Mackey theory can be used to determine all of its unitary representations. The orbits are the circles $|b| = \beta^2$ for $\beta \geq 0$. The orbit $\beta = 0$ corresponds to the

representations σ_n, these being the characters of S viewed as representations of H_0. At the points β the little group is (± 1) which has two characters, the trivial one and the one which takes -1 to -1. So there are two irreducible unitaries $\lambda_{\beta,\pm}$ for each $\beta > 0$. Associated to these we have representations $L_{0,\beta,\pm}$ which define particles of mass 0 and infinite helicity, i.e., possessing an internal space of *infinite* dimension. These have also to be excluded because of their unphysical nature.

Representations of the Poincaré group of Minkowski space of arbitrary dimension. The theory described above goes over with virtually no changes to the case of the Minkowski space $V = \mathbf{R}^{1,D-1}$ of dimension D. Thus

$$G = \mathbf{R}^{1,D-1} \times' H, \qquad H = \mathrm{Spin}(1, D-1)$$

where the spin group $\mathrm{Spin}(1, D-1)$ is the universal (= 2-fold) cover of $\mathrm{SO}(1, D-1)^0$ (see Chapter 5 for notation and results on real spin groups). The orbits are classified as before. For the orbits $X_m^{\pm}$ the little group at $(m, 0, \ldots, 0)$ is $\mathrm{Spin}(D-1)$. The orbits have the invariant measure

$$d\mu_m^+ = \frac{d^{D-1}p}{(m^2 + p_1^2 + \ldots + p_{D-1}^2)^{1/2}}.$$

The orbits $X_0^{\pm}$ require a little more care because our earlier description of the little groups for the case $D = 4$ used the special model of Hermitian matrices for spacetime.

We write $(e_\mu)_{0 \le \mu \le D-1}$ for the standard basis of $V = \mathbf{R}^{1,D-1}$, with $(e_0, e_0) = -(e_j, e_j) = 1 (1 \le j \le D-1)$. We wish to determine the little group at the point $q = e_0 + e_{D-1}$. Let ℓ be the line $\mathbf{R} \cdot q$ and let H_q be the little group at q, the subgroup of H fixing q. We write H_q' for the stabilizer of q in the group $V \times' \mathrm{SO}(1, D-1)^0$ so that H_q is the lift of H_q' inside G. Clearly H_q fixes $\ell^\perp$ and so we have the H_q-invariant flag

$$\ell \subset \ell^\perp \subset V.$$

Now ℓ is the radical of the restriction of the metric to $\ell^\perp$ and so the induced metric on $E := \ell^\perp / \ell$ is *strictly negative definite*. We shall now show that there is a natural map

$$H_q' \simeq E \times' \mathrm{SO}(E).$$

Let $h \in H_q'$. Then h induces an element $h^\sim$ of $\mathrm{O}(E)$. We claim first that $h^\sim \in \mathrm{SO}(E)$ and that h induces the identity on $V/\ell^\perp$. Since $\det(h) = 1$ and $\det(h^\sim) = \pm 1$, we see that h induces ± 1 on $V/\ell^\perp$ and so it is enough to prove that h induces $+1$ on $V/\ell^\perp$. Now $e_0 \notin \ell^\perp$ and $h \cdot e_0 = a e_0 + u$ where $a = \pm 1$ and $u \in \ell^\perp$. Then

$(e_0, q) = (h{\cdot}e_0, q) = a(q, e_0)$ so that $a = 1$. Since $h{\cdot}e_0 - e_0 \in \ell^\perp$ its image in E is well-defined; we write $t(h)$ for it. We thus have a map

$$H'_q \longrightarrow E \times' \mathrm{SO}(E), \qquad h \longmapsto (t(h), h^\sim).$$

It is easy to check that this is a morphism of Lie groups. We assert that this map is injective. Suppose that h is in the kernel of this map so that $h{\cdot}u = u + a(h)q$ for all $u \in \ell^\perp$ and $h{\cdot}e_0 = e_0 + b(h)q$. Then $(e_0, e_0) = (h{\cdot}e_0, h{\cdot}e_0) = (e_0, e_0) + 2b(h)(q, e_0)$, giving $b(h) = 0$. Also $(u, e_0) = (h{\cdot}u, h{\cdot}e_0) = (u, e_0) + a(h)(q, e_0)$, giving $a(h) = 0$. Thus $h = 1$. A simple calculation with the Lie algebra shows that $\mathrm{Lie}(H'_q)$ has the same dimension as $E \times' \mathrm{SO}(E)$. Therefore the map above is an isomorphism of H'_q with $E \times' \mathrm{SO}(E)$.

Let H_q be the stabilizer of q in $V \times' \mathrm{Spin}(1, D-1)$. We shall show that H_q is connected if $D \geq 4$. Let $x = e_1 e_2$ and $a_t = \exp tx$. Since $(e_1, e_1) = (e_2, e_2) = -1$, we have $x^2 = -1$ and so $a_t = \cos t{\cdot}1 + \sin t{\cdot}x$. It is obvious that a_t fixes q and so lies in H_q for all t. But for $t = \pi$ we have $a_\pi = -1$. Thus H_q^0 contains the kernel of the map from $\mathrm{Spin}(1, D-1)$ to $\mathrm{SO}(1, D-1)^0$, proving that $H_q = H_q^0$. Thus finally

$$H_q = H_q^0 \simeq E \times' \mathrm{Spin}(E).$$

We have thus shown that for $D \geq 4$, the little group of any point q of X_0^+ is the 2-fold cover of the Euclidean motion group of $\ell^\perp/\ell$ where $\ell = \mathbf{R}q$, exactly as in the case $D = 4$.

1.6. Vector bundles and wave equations. The Maxwell, Dirac, and Weyl equations. Two things remain to be done. The first is to construct the representations explicitly by describing the corresponding vector bundles. This will give a description of the states in what is called the *momentum picture*, in which the momentum operators are diagonalized. The physicists also use frequently a description where the states are represented by functions on spacetime and the spacetime group acts naturally on them. Indeed such descriptions are very useful when treating interactions of the particles with other systems such as an external electromagnetic field. In the spacetime picture the states will be formally singled out by a *wave equation.* This description can be obtained from the momentum space representation by taking *Fourier transforms.* Throughout this section Fourier transforms are taken with respect to the Lorentz-invariant scalar product

$$\langle x, p \rangle = \sum \varepsilon_\mu x_\mu p_\mu$$

so that

$$\widehat{u}(x) = \int e^{-i\langle x, p\rangle} u(p) d^4 p.$$

In particular, multiplication by p_μ goes over to $i\varepsilon_\mu\partial_\mu$:

$$p_\mu \longrightarrow i\varepsilon_\mu\partial_\mu \qquad (\partial_\mu = \partial/\partial x_\mu).$$

Klein-Gordon equation. As our first example let us take the simplest particle, one of mass $m \geq 0$ and spin 0. It could be charged or neutral. Here there is no internal space and so the bundle is trivial; the Hilbert space is

$$\mathcal{H}_m^\pm = L^2(X_m^\pm, \mu_m^\pm)$$

where $\mu_m^\pm$ is the invariant measure on $X_m^\pm$. The action of the Poincaré group is as follows:

$$(L(a)f)(p) = e^{i\langle a,p\rangle} f(p) \qquad (a \in \mathbf{R}^4)$$
$$(L(h)f)(p) = f(h^{-1}p) \qquad (h \in \mathrm{SL}(2,\mathbf{C})).$$

To take Fourier transforms of the f we view them as distributions on $\mathbf{R}^4$,

$$fd\mu_m^\pm : \varphi \longmapsto \int_{\mathbf{P}^4} f\varphi d\mu_m^\pm \qquad (\varphi \in \mathcal{D}(\mathbf{P}^4))$$

where $\mathcal{D}(\mathbf{P}^4)$ is the space of smooth compactly supported functions on $\mathbf{P}^4$. It is not difficult to show that these distributions, which are actually complex measures, are *tempered*. Indeed, this follows from the easily established fact that the $\mu_m^\pm$-measure of a ball of radius R grows at most like a power of R, actually like R^3 in this case. Since the $fd\mu_m^\pm$ live on $X_m^\pm$ it is immediate that they satisfy the equation

$$(p_0^2 - p_1^2 - p_2^2 - p_3^2 - m^2) \cdot (fd\mu_m^\pm) = 0.$$

Taking Fourier transforms and writing $\psi = \widehat{fd\mu_m}$, we have

$$(\partial_0^2 - \partial_1^2 - \partial_2^2 - \partial_3^2 + m^2)\psi = 0$$

which is the so-called *Klein-Gordon equation.* One can say that the states of the scalar massive particle of mass $m > 0$ are the *tempered* solutions of the K-G equation. On the other hand, if we are given a tempered solution ψ of the K-G equation, it is not difficult to see that $\psi = \widehat{u}$ where u is a distribution which *lives on* X_m. Whether the support of u is confined to one of the two orbits $X_m^\pm$ is not easily decidable in terms of ψ alone. At the same time, from the formula for the action of the spacetime translations we see that the energy operator P_0 is multiplication by p_0 and so the spectrum of P_0 is $\geq m$ on $\mathcal{H}_m^+$ and $\leq -m$ on $\mathcal{H}_m^-$ (the so-called negative energy states). Nowadays, following Dirac (see below), the space $\mathcal{H}_m^-$ is

viewed as *antiparticle* charged oppositely to the original particle described by $\mathcal{H}_m^+$. We can combine the two Hilbert spaces $\mathcal{H}_m^\pm$ into one,

$$\mathcal{H}_m = \mathcal{H}_m^+ \oplus \mathcal{H}_m^- = L^2(X_m, \mu_m)$$

where μ_m is the measure on X_m coinciding with $\mu_m^\pm$ on $X_m^\pm$, and allow the full symmetry group

$$\mathbf{R}^4 \times' \mathrm{O}(1,3)$$

to act on $\mathcal{H}_m$. Thus the K-G spinless particle-antiparticle of mass m has this complete symmetry group and the distributions $\psi = \widehat{u}$ ($u \in \mathcal{H}_m$) satisfy the K-G equation. For any tempered solution ψ we have $\psi = \widehat{u}$ where u lives on X_m; but to define an actual state u must be a measure on X_m absolutely continuous with respect to μ_m and $du/\mu_m \in f \in L^2(X_m, \mu_m)$, the L^2-norm of this derivative being the norm of the state.

Dirac equation. During the early stages of development of relativistic quantum mechanics the K-G equation was the only equation that described relativistic particles. But Dirac was dissatisfied with this picture. For various reasons connected with difficulties in defining probability densities and currents he felt that the wave equation should be of the *first order* in time, and hence, as time and space coordinates are to be treated on the same footing, *it should be of the first order in all variables.* He therefore looked for an equation of the form

$$i\left(\sum_\mu \gamma_\mu \partial_\mu\right)\psi = m\psi.$$

Of course the K-G equation was not to be abandoned; it was required to follow as a consequence of this equation. Dirac therefore assumed that

$$\left(\sum_\mu \gamma_\mu \partial_\mu\right)^2 = \partial_0^2 - \partial_1^2 - \partial_2^2 - \partial_3^2.$$

In this way the differential operator he was looking for would be a sort of *square root* of the K-G operator. Dirac's assumption leads to the relations

$$\gamma_\mu^2 = \varepsilon_\mu, \qquad \gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 0 \ (\mu \neq \nu)$$

where

$$\varepsilon_\mu = \begin{cases} 1 & \text{if } \mu = 0 \\ -1 & \text{if } \mu = 1,2,3. \end{cases}$$

It is now clear that the γ_μ cannot be scalars. Dirac discovered that there is a solution with 4×4 matrices and that this solution is unique upto a similarity. But then the operator

$$D = i \sum_\mu \gamma_\mu \partial_\mu$$

has to operate on *vector functions with* 4 *components* so that the Dirac particle has automatically an internal space of dimension 4! D is the famous *Dirac operator.*

We shall follow this procedure of Dirac in constructing the vector bundle on the *full* orbit X_m. We look for objects γ_μ such that

$$\left(\sum_\mu \gamma_\mu p_\mu \right)^2 = \sum \varepsilon_\mu p_\mu^2 = \sum_\mu p_\mu p^\mu \qquad (p^\mu = \varepsilon_\mu p_\mu)$$

giving the relations

$$\gamma_\mu^2 = \varepsilon_\mu, \qquad \gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = 0 \ (\mu \neq \nu).$$

We consider the algebra $\mathcal{C}$ generated by the γ_μ with the relations above. It is called the *Clifford algebra* which is a generalization of the quaternion algebra. It is of dimension 16, and is isomorphic to a full matrix algebra (this will follow from our explicit formula for the γ's below). Hence it has a *unique* irreducible representation in dimension 4; any representation of $\mathcal{C}$ is a direct sum of copies of this representation. The uniqueness of the 4-dimensional representation means that if γ_μ, γ'_μ are 4×4 matrices satisfying the above relations, there is a 4×4 invertible matrix S such that

$$\gamma'_\mu = S \gamma_\mu S^{-1}$$

for all μ. S is unique up to a scalar multiplier because if S' is another such, then $S' S^{-1}$ commutes with all the γ's and so must be a scalar by Schur's lemma. As a useful application of this principle we note that given a set (γ_μ), the matrices $(-\gamma_\mu)$ also satisfy the same relations and so there is $S \in \mathrm{GL}(4, \mathbf{C})$ such that

$$-\gamma_\mu = S \gamma_\mu S^{-1}.$$

As $\gamma_0^2 = 1$ and γ_0 and $-\gamma_0$ are similar, we see that γ_0 has the eigenvalues ± 1 with eigenspaces of dimension 2 each. The same is true of $i\gamma_j (j = 1, 2, 3)$. The γ_μ are the famous *Dirac gamma matrices.* They are a part of a whole yoga of *spinor calculus* (see Chapter 5).

At the risk of being pedantic let us write Λ for the covering morphism from $\mathrm{SL}(2,\mathbf{C})$ onto $\mathrm{SO}(1,3)^0$. Consider now a variable point $p=(p_\mu)$. Fix a set of 4×4 gamma matrices γ_μ. Write $p^\mu=\varepsilon_\mu p_\mu$. If $h=(h_{\mu\nu})\in \mathrm{O}(1,3)$ and $q=hp$, we have

$$\left(\sum_\mu p_\mu\gamma_\mu\right)^2=\sum_\mu p_\mu p^\mu=\sum_\mu q_\mu q^\mu=\left(\sum_\mu q_\mu\gamma_\mu\right)^2=\left(\sum_\mu p_\nu\gamma'_\nu\right)^2$$

where

$$\gamma'_\nu=\sum_\mu h_{\mu\nu}\gamma_\mu.$$

Thus the γ'_μ also satisfy the basic relations and hence there is $S(h)\in \mathrm{GL}(4,\mathbf{C})$ such that

$$S(h)\gamma_\mu S(h)^{-1}=\sum_\nu \gamma_\nu h_{\nu\mu}$$

or, equivalently,

$$S(h)(p\cdot\gamma)S(h)^{-1}=(hp)\cdot\gamma,\qquad p\cdot\gamma=\sum_\mu p_\mu\gamma_\mu.$$

From the uniqueness up to a scalar of $S(h)$ and the calculation

$$S(k)S(h)\gamma_\mu S(h)^{-1}S(k)^{-1}=\sum_\rho \gamma_\rho k_{\rho\nu}h_{\nu\mu}=S(kh)\gamma_\mu S(kh)^{-1}$$

we see that $S(k)S(h)$ and $S(kh)$ differ by a scalar. So S defines a homomorphism of $\mathrm{O}(1,3)$ into the projective group $\mathrm{PGL}(4,\mathbf{C})$. We shall show presently that its restriction to $\mathrm{SL}(2,\mathbf{C})$ comes from a representation of $\mathrm{SL}(2,\mathbf{C})$ and that this representation is unique.

For this we shall exhibit a set of γ's and compute the representation S explicitly. Since we want to involve the full symmetry group $\mathrm{O}(1,3)$ rather than its connected component we shall first enlarge $\mathrm{SL}(2,\mathbf{C})$ to a group $\mathrm{O}(1,3)^\sim$ so $\mathrm{SL}(2,\mathbf{C})$ is the connected component of $\mathrm{O}(1,3)^\sim$ and we have a natural 2-fold covering map Λ from $\mathrm{O}(1,3)^\sim$ to $\mathrm{O}(1,3)$. To do this notice that $\mathrm{O}(1,3)$ is the semidirect product

$$\mathrm{O}(1,3)=\mathrm{O}(1,3)^0\times' I$$

where

$$I\simeq \mathbf{Z}_2\oplus\mathbf{Z}_2=\{1,I_s,I_t,I_{st}\}$$

the I's being the inversions in space, time, and spacetime. Since $\mathrm{SL}(2,\mathbf{C})$ is simply connected we can view I (uniquely) as acting on it compatibly with the covering map onto $\mathrm{O}(1,3)^0$. This means that for any inversion $I_r(r = s,t,st)$, $g \longmapsto I_r[g]$ is an automorphism of $\mathrm{SL}(2,\mathbf{C})$ such that $\Lambda(I_r[g]) = I_r\Lambda(g)I_r$. We can then define

$$\mathrm{O}(1,3)^{\sim} = \mathrm{SL}(2,\mathbf{C}) \times' I$$

and get a 2-fold cover

$$\Lambda : \mathrm{O}(1,3)^{\sim} \longrightarrow \mathrm{O}(1,3).$$

Let us introduce the Pauli spin matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then

$$\sigma_j^2 = 1, \qquad \sigma_j\sigma_k + \sigma_k\sigma_j = 0 \ (j \neq k).$$

If we then take

$$\gamma_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \gamma_j = \begin{pmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{pmatrix} \qquad (j = 1,2,3)$$

where 1 refers to the 2×2 identity matrix, then we have a set of γ's satisfying the relations we need. It is easy to check that the γ_μ act irreducibly.

Let us write $\mathbf{p} = (p_1,p_2,p_3)$ and $p = (p_0,\mathbf{p})$ and let $\mathbf{s} = (\sigma_1,\sigma_2,\sigma_3)$. Then, writing $\mathbf{p}\cdot\mathbf{s} = p_1\sigma_1 + p_2\sigma_2 + p_3\sigma_3$ we have

$$p\cdot\gamma = \sum_\mu p_\mu\gamma_\mu = \begin{pmatrix} 0 & p_0 1 + \mathbf{p}\cdot\mathbf{s} \\ p_0 1 - \mathbf{p}\cdot\mathbf{s} & 0 \end{pmatrix}.$$

On the other hand

$$p_0 1 + \mathbf{p}\cdot\mathbf{s} = \begin{pmatrix} p_0 + p_3 & p_1 - ip_2 \\ p_1 + ip_2 & p_0 - p_3 \end{pmatrix}$$

so that, with $*$ denoting adjoints,

$$g(p_0 1 + \mathbf{p}\cdot\mathbf{s})g^* = q_0 1 + \mathbf{q}\cdot\mathbf{s}, \qquad q = \Lambda(g)p \qquad (g \in \mathrm{SL}(2,\mathbf{C})).$$

Now $\det(p_0 1 + \mathbf{p}\cdot\mathbf{s}) = p^2$ where $p^2 = p_0^2 - p_1^2 - p_2^2 - p_3^2$ and so

$$(p_0 1 + \mathbf{p}\cdot\mathbf{s})^{-1} = (p^2)^{-1}(p_0 1 - \mathbf{p}\cdot\mathbf{s})$$

from which we get

$$g^{*-1}(p_0 1 - \mathbf{p}\cdot\mathbf{s})g^{-1} = q_0 1 - \mathbf{q}\cdot\mathbf{s}.$$

From this we get at once that

$$\begin{pmatrix} g & 0 \\ 0 & g^{*-1} \end{pmatrix} \left(\sum_\mu p_\mu \gamma_\mu \right) \begin{pmatrix} g^{-1} & 0 \\ 0 & g^* \end{pmatrix} = \sum_\mu q_\mu \gamma_\mu \qquad q = \Lambda(g)p.$$

Since this is precisely the defining relation for $S(g)$ we get

$$S(g) = \begin{pmatrix} g & 0 \\ 0 & g^{*-1} \end{pmatrix}.$$

We would like to extend S to include the inversions also. A simple calculation shows that we can take

$$S(I_s) = \pm \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \qquad S(I_t) = \pm \begin{pmatrix} 0 & iI \\ -iI & 0 \end{pmatrix}, \qquad S(I_{st}) = \pm \begin{pmatrix} iI & 0 \\ 0 & -iI \end{pmatrix}.$$

The uniqueness of S follows from the fact that $\mathrm{SL}(2, \mathbf{C})$ has only the trivial representation in dimension 1. Notice that with any choices $S(I_s)S(I_t) = -S(I_t)S(I_s)$ so that these choices always define the unique irreducible projective representation of $I \simeq \mathbf{Z}_2 \oplus \mathbf{Z}_2$ in dimension 2 tensored by $\mathbf{C}^2$. A simple calculation shows that

$$S(I_r[g]) = S(I_r)S(g)S(I_r)$$

since both sides are examples of a representation S' satisfying the relations

$$S'(g)(p\cdot\gamma)S'(g)^{-1} = (\Lambda(I_r[g])p)\cdot\gamma.$$

If we define

$$S(I_r g) = S(I_r)S(g) \qquad (r = s, t, st, g \in \mathrm{SL}(2, \mathbf{C}))$$

we see that S is a double-valued representation of $\mathrm{O}(1,3)^\sim$ that restricts on $\mathrm{SL}(2, \mathbf{C})$ to a representation.

Lemma 1.6.1 *Let γ_μ be defined as above. Then there is a double-valued (d.v.) representation S of $\mathrm{O}(1,3)^\sim$ in dimension 4 restricting to a representation on $\mathrm{SL}(2, \mathbf{C})$ such that*

$$S(h)(p\cdot\gamma)S(h)^{-1} = \sum_\mu q_\mu \gamma_\mu \qquad (q = \Lambda(h)p).$$

The restriction of S to $\mathrm{SL}(2,\mathbf{C})$ is unique and is given by

$$S(g) = \begin{pmatrix} g & 0 \\ 0 & g^{*-1} \end{pmatrix}.$$

The d.v. representation S defines a d.v. action of $\mathrm{O}(1,3)^\sim$ on the trivial bundle

$$T = X \times \mathbf{C}^4 \qquad (X = X_m^+)$$

by

$$g, (p, v) \longmapsto (\Lambda(g)p, S(g)v).$$

Define now

$$\mathbf{D}_m(p) = \Big\{ v \in \mathbf{C}^4 \;\Big|\; (p{\cdot}\gamma)\, v = mv \Big\}.$$

If $p = \Lambda(g)p^0$ where p^0 is the base point with coordinates $(m,0,0,0)$, we have $S(g)(m\gamma_0)S(g)^{-1} = \sum_\mu p_\mu \gamma_\mu$. Hence $\sum p_\mu \gamma_\mu$ is semisimple for all (p_μ) with $\sum_\mu p_\mu p^\mu = m^2 > 0$ and its eigenspaces for the eigenvalues $\pm m$ are of dimension 2. In particular all the spaces $\mathbf{D}_m(p)$ have dimension 2 and

$$S(g)[\mathbf{D}_m(p)] = \mathbf{D}_m(\Lambda(g)p).$$

This shows that the spaces $\mathbf{D}_m(p)$ define a *subbundle* $\mathbf{D}_m$ of T of rank 2, stable under the d.v. action of $\mathrm{O}(1,3)^\sim$ given by

$$(p, v) \longmapsto (\Lambda(g)p, S(h)v) \qquad (h \in \mathrm{O}(1,3)^\sim).$$

One may call $\mathbf{D}_m$ the *Dirac bundle* on X_m.

The stabilizer of $\pm p^0 = (\pm m, 0, 0, 0)$ within $\mathrm{SL}(2,\mathbf{C})$ is $\mathrm{SU}(2)$ and it acts by

$$\begin{pmatrix} g & 0 \\ 0 & g \end{pmatrix} \qquad (g \in \mathrm{SU}(2)).$$

It commutes with γ_0 and so leaves invariant the spaces $\mathbf{D}_m(\pm p^0)$ where it acts like the representation **2**. The standard scalar product on $\mathbf{C}^4$ is invariant under $\mathrm{SU}(2)$ and so induces an invariant scalar product on $\mathbf{D}_m(\pm p^0)$. The inversions I_r either preserve the spaces and are unitary on them ($r = st$) or exchange them in a unitary manner ($r = s, t$). We may then transport this scalar product to all the fibers $D_m(\pm p)$ on X_m covariantly. We thus obtain a Hermitian bundle on X_m on which

the action of $\mathrm{SL}(2,\mathbf{C})$ is unitary. The inversions preserve this Hermitian structure and so the action of the entire group $\mathrm{O}(1,3)^\sim$ is unitary.

The Hilbert space of square integrable sections of the bundle $\mathbf{D}_m$ then carries a projective unitary representation of $\mathrm{O}(1,3)^\sim$ whose restriction to $\mathrm{SL}(2,\mathbf{C})$ is

$$L_{m,1/2} := L^+_{m,1/2} \oplus L^-_{m,1/2}.$$

Identifying sections s with measures $sd\mu_m$ and taking Fourier transforms we get the *Dirac equation*

$$i\left(\sum_\mu \varepsilon_\mu \gamma_\mu \partial_\mu\right)\psi = m\psi$$

or

$$i\left(\sum_\mu \gamma^\mu \partial_\mu\right)\psi = m\psi.$$

As before we shall regard $\mathcal{H}_m$ as describing the particle-antiparticle of mass m.

Write any section ψ of T in the form

$$\psi = \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}, \qquad \psi_j : X_m \longrightarrow \mathbf{C}^2.$$

Since

$$(p_0 1 + \mathbf{p}\cdot\mathbf{s})(p_0 1 - \mathbf{p}\cdot\mathbf{s}) = p^2 1$$

it follows that

$$\begin{pmatrix}\psi_1(p)\\ \psi_2(p)\end{pmatrix} \in \mathbf{D}_m(p) \Leftrightarrow \psi_2(p) = m^{-1}(p_0 1 - \mathbf{p}\cdot\mathbf{s})\psi_1(p).$$

Hence

$$\begin{pmatrix}v_1\\ v\end{pmatrix} \longmapsto v_1$$

gives a bundle isomorphism of $\mathbf{D}_m$ with the *trivial bundle* $V_m = X_m \times \mathbf{C}^2$ in such a manner that the action of the Poincaré group on $\mathbf{D}_m$ goes over to the action L'_m on V_m defined by

$$(L'_m(u,g)\psi_1)(p) = e^{i\langle u,p\rangle} g\psi_1(\Lambda(g)^{-1}p).$$

The spinor field ψ_1 which is a section of the $\mathrm{SL}(2,\mathbf{C})$-bundle V_m is usually called a 2-*component spinor*. It was first treated systematically by van der Waerden.

Holes and antimatter. Let us go back to the description of the states of the electron by the Dirac wave equation

$$i\sum_{\mu}\gamma^{\mu}\partial_{\mu}\psi = m\psi.$$

The Hilbert space $\mathcal{H}_m$ carries a (projective) action of the full group of automorphisms of Minkowski spacetime. Now $\mathcal{H}_m = \mathcal{H}_m^+ \oplus \mathcal{H}_m^-$ and it is clear as in the case of the K-G equation that the spectrum of the energy operator P_0, which is multiplication by p_0, is > 0 on $\mathcal{H}_m^+$ and < 0 on $\mathcal{H}_m^-$. The states in $\mathcal{H}_m^{\pm}$ are usually called the *positive and negative energy states*. As long as the electron is free its state will be in $\mathcal{H}^+$, but as soon as it is placed in a magnetic field, transitions to negative energy states cannot be excluded. That this does not happen was a big problem to be solved at the time Dirac proposed his equation. It was in order to explain the meaning of the negative energy states that Dirac invented his *hole theory* which asserts that all the negative energy states are occupied, and transition to them is possible only when one of these states becomes available for occupation as a *hole.* The holes were then interpreted by him as positive energy particles of charge opposite to that of the electron. This led him to predict the existence of a new particle, the *positron.* Shortly after Dirac made his prediction, the positron was discovered by Anderson. Eventually, with the discovery of the antiproton and other antiparticles it became clear that all particles have their *antiparticles* which are constituents of *antimatter.* (However the overwhelming preponderance of matter over antimatter in the universe probably depends on conditions that were prevalent in the early evolution of the universe.) The discovery of antimatter is regarded by physicists as one of the greatest achievements of physics of all time and consequently the stature of Dirac in the physics pantheon rivals that of Newton and Einstein.

As an interesting historical footnote, when Dirac proposed that particles of positive charge should correspond to the holes he thought that these should be protons which were the only particles of positive charge known at that time (1929c); it was Weyl who pointed out that *symmetry requirements force the hole to have the same mass as the electron and so the new particle cannot be the proton but a positively charged particle with the same mass as the electron*, nowadays called the *positron*. Eventually this prediction of Dirac was confirmed when Anderson exhibited the track of a positron. In retrospect one knows that at the time of Anderson's discovery Blackett apparently had three tracks of the positron in his experiments but was hesitant to announce them because he felt more evidence was necessary. Anderson at Caltech had only one track and had no hesitation in announcing it!

Zero mass bundles and equations. We shall now construct the bundles for the representations $L^{\pm}_{0,N}$.

Maxwell equation for the photon. We consider first the case $N = 2$. We start with the *tangent bundle* F of the cone X_0^+. The action of the Lorentz group on the cone lifts to an action on F. The tangent space at $(1, 0, 0, 1)$ consists of all $(\xi_0, \xi_1, \xi_2, \xi_0)$. The ambient metric on this space is $-(\xi_1^2 + \xi_2^2)$ which is ≤ 0 but degenerate, and the null vectors are multiples of $(1, 0, 0, 1)$. In the basis

$$v_0 = (1, 0, 0, 1), \qquad v_1 = (0, 1, 0, 0), \qquad v_2 = (0, 0, 1, 0)$$

the action of the little group at $p^0 = (1, 0, 0, 1)$ is

$$\begin{pmatrix} e^{i\theta} & b \\ 0 & e^{-i\theta} \end{pmatrix} : v_0 \mapsto v_0, \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \mapsto \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

Let R be the subbundle of F whose fiber at p is the line $\mathbf{R}p$; this is the line bundle whose fiber at p is the space of null vectors at p for the induced metric on the tangent space F_p at p. Let F^+ be the quotient bundle F/R. The metric on the fibers of F then descends to a *positive definite* metric on the fibers of F^+ and the representation carried by the square integrable (with respect to μ_0^+) sections of F^+ is $L_{0,2} \oplus L_{0,-2}$. We may regard the sections of F^+ as vectors $a = (a_\mu)$ with 4 components satisfying

$$\sum \varepsilon_\mu p_\mu a_\mu = 0$$

but identifying sections $a = (a_\mu)$ and $a' = (a'_\mu)$ by

$$a \sim a' \iff p \wedge (a - a') = 0.$$

Taking Fourier transforms and writing $A_\mu = \varepsilon_\mu \widehat{a_\mu}$, we get

$$\mathcal{D} A_\mu = 0, \qquad \text{div}_{\,L} A = 0$$

with

$$A \sim A' \iff d(A - A') = 0.$$

These are just the Maxwell equations in the Lorentz gauge. It is thus natural to call $L_{0,2} \oplus L_{0,-2}$ the *photon representation.* Thus *the one particle photon equations are already the Maxwell equations.* However one must remember that the Maxwell equations deal with real vector potentials and the photon equations deal with complex potentials. But because the tangent bundle is real, the real sections define a real form of $L_{0,2} \oplus L_{0,-2}$, and so our identification of the two equations is quite reasonable. The helicity of the photon is ± 1 and the two values correspond to left and right circularly polarizations.

The fact that the equations describing the photon are the same as Maxwell's equation is very important. In Dirac's theory of radiation he quantized the classical wave equation of Maxwell and found that the states of the (free) quantized electromagnetic field thus obtained were the same as one would obtain by treating a system of photons with Bose-Einstein statistics, i.e., by replacing the one-photon Hilbert space by the *symmetric algebra* over it (see Section 7 below). This was then interpreted by him as an expression of the wave-particle duality of light. Since the Maxwell equations are already the equations describing a free photon, the process of going from a single photon to a system of several photons was called *the second quantization*.

Weyl equation for the neutrino. One can make $m = 0$ in the Dirac bundle and get the bundle N on the light cone. However more care is necessary because for $p \in X_0 = X + 0^+ \cup X_0^-$ the operator $p{\cdot}\gamma$ is nilpotent: $(p{\cdot}\gamma)^2 = 0$. Let

$$N(p) = \{v \in \mathbf{C}^4 \mid (p{\cdot}\gamma)v = 0\}.$$

For $p \in X_0$, $p{\cdot}\gamma$ is conjugate by an element of $\mathrm{SL}(2,\mathbf{C})$ to $\pm(\gamma_0 + \gamma_3)$. But $(\gamma_0 + \gamma_3)^2 = 0$ and its null space is spanned by e_0, e_3 in $\mathbf{C}^4$. Hence $\dim(N(p)) = 2$. Thus the $N(p)$ define a smooth subbundle N of $X_0 \times \mathbf{C}^4$ stable under $\mathrm{O}(1,3)^\sim$.

For $p \in X_0$ we have

$$\begin{pmatrix} v_- \\ v_+ \end{pmatrix} \in N(p) \Leftrightarrow (p_0 1 \pm \mathbf{p}{\cdot}\mathbf{s})v_\pm = 0 \qquad (v_\pm \in \mathbf{C}^2).$$

We write

$$\ell_\pm(p) = \{v \in \mathbf{C}^2 \mid (p_0 1 \pm \mathbf{p}{\cdot}\mathbf{s})v = 0\}.$$

Since we have, for $g \in \mathrm{SL}(2,\mathbf{C})$,

$$g(p_0 1 + \mathbf{p}{\cdot}\mathbf{s})g^* = q_0 1 + \mathbf{q}{\cdot}\mathbf{s}, \quad {g*}^{-1}(p_0 1 - \mathbf{p}{\cdot}\mathbf{s})g^{-1} = q_0 1 - \mathbf{q}{\cdot}\mathbf{s} \qquad (q = \Lambda(g)p),$$

it follows that

$$v \in \ell_-(p) \Longleftrightarrow gv \in \ell_-(\Lambda(g)p), \qquad v \in \ell_+(p) \Longleftrightarrow {g^*}^{-1}v \in \ell_+(\Lambda(g)p).$$

This shows that the $\ell_\pm(p)$ define *line bundles* $\mathbf{W}_{0,\pm}$ which are homogeneous for the action of $\mathrm{SL}(2,\mathbf{C})$ defined by

$$(\mathbf{W}_{0,-}) : (p,v) \longmapsto (\Lambda(g)p, gv), \qquad (\mathbf{W}_{0,+}) : (p,v) \longmapsto (\Lambda(g)p, {g^*}^{-1}v).$$

We then have an isomorphism

$$\ell_-(p) \oplus \ell_+(p) \simeq N(p), \qquad (u, v) \longmapsto \begin{pmatrix} u \\ v \end{pmatrix}$$

which gives a $\mathrm{SL}(2, \mathbf{C})$-equivariant bundle isomorphism

$$\mathbf{W}_{0,-} \oplus \mathbf{W}_{0,+} \simeq N.$$

We shall identify $\mathbf{W}_{0,\pm}$ as subbundles of N and denote their restrictions to $X_0^\pm$ by $\mathbf{W}_{0,\pm}^\pm$. The bundles $\mathbf{W}_{0,\pm}^\pm$ may be appropriately called the *Weyl bundles* since the equations satisfied by the Fourier transforms of their sections were first discovered by Weyl and proposed by him as the equations that the neutrinos should satisfy.

Let us compute the little group actions at $\pm p^0 = \pm(1, 0, 0, 1)$. The little group at p^0 is

$$\begin{pmatrix} a & b \\ 0 & \overline{a} \end{pmatrix} \qquad (|a| = 1).$$

Further $\ell_\pm(p^0)$ are spanned by e_3 and e_0 respectively, and the actions are easily computed to be

$$\begin{pmatrix} a & b \\ 0 & \overline{a} \end{pmatrix} : v \longmapsto a^\mp v.$$

So the representations defined by the $W_{0,\pm}^\pm$, the restrictions of $W_{0,\pm}$ to $X_0^\pm$, are $L_{0,\mp 1}^\pm$. The calculations are the same at $-p^0$. The restriction to the one-dimensional spaces $\ell_\pm(\pm p^0)$ of the standard norm in $\mathbf{C}^4$ transported by the group action now gives the invariant Hermitian structures on the Weyl bundles which is invariant under the action of the Poincaré group.

It must be noticed that the *Weyl bundles are invariant under spacetime inversion but not invariant under the action of either space or time inversions*. In fact we have

$$I_s, I_t : \mathbf{W}_{0,\pm}^\pm \longrightarrow \mathbf{W}_{0,\mp}^\pm, \qquad I_{st} : \mathbf{W}_{0,\pm}^\pm \longrightarrow \mathbf{W}_{0,\pm}^\pm.$$

Let us now take a closer look at the elements of $\ell_\pm(p)$. We have

$$u \in \ell_+(p) \Longleftrightarrow (\mathbf{p}\cdot\mathbf{s})u = -p_0 u.$$

For $p_0 > 0$ or $p_0 < 0$ respectively we have $p_0 = \pm|\mathbf{p}|$ and so we have

$$u \in \ell_+(p) \Longleftrightarrow (\mathbf{p}\cdot\mathbf{s})u = \begin{cases} -|\mathbf{p}|u & \text{if } p_0 > 0 \\ +|\mathbf{p}|u & \text{if } p_0 < 0 \end{cases}$$

showing that the direction of the spin is *antiparallel to the momentum* for $p_0 > 0$ and *parallel to the momentum* for $p_0 < 0$. Similarly for $u \in \ell_+(p)$ we have the spin and momentum are parallel for $p_0 > 0$ and antiparallel for $p_0 < 0$. Let us refer to the case where the spin and momentum are antiparallel (resp. parallel) as *lefthanded* (resp. *righthanded*). It follows that the bundles $\mathbf{W}^+_{0,+}$, $\mathbf{W}^-_{0,+}$ represent respectively the lefthanded neutrinos and righthanded antineutrinos, while $\mathbf{W}^+_{0,-}$, $\mathbf{W}^-_{0,-}$ represent respectively the righthanded neutrinos and lefthanded antineutrinos.

By taking Fourier transforms of sections of these bundles we get the 2-*component Weyl equations for the neutrino-antineutro pairs*, namely

$$(\partial_0 - \nabla\cdot\mathbf{s})\psi_+ = 0$$

for the wave functions of the leftneutrino-rightantineutrino pairs and

$$(\partial_0 + \nabla\cdot\mathbf{s})\psi_- = 0$$

for the wave functions of the rightneutrino-leftantineutrino pairs. Under space inversion the two equations are interchanged.

Weyl proposed these 2-component equations for the zero mass spin 1/2 particles in 1929. At that time they were rejected by Pauli because of their lack of invariance with respect to space inversion. Indeed it was always a basic principle that the wave equations should be invariant under *all* Lorentz transformations, not just those in the connected component. In particular, invariance under space inversion, also called *parity conservation*, was demanded. In the mid 1950's, in experiments performed by Wu following a famous suggestion by Yang and Lee that the neutrinos did not have the parity conservation property, it was found that the neutrinos emitted during beta decay had a preferred orientation. Experimental evidence further indicated that the spin is always antiparallel to the momentum for the neutrinos so that the neutrinos are *always lefthanded*. After Wu's experiment, Landau and Salam proposed that the Weyl equation, namely

$$(\partial_0 - \nabla\cdot\mathbf{s})\psi_\pm = 0,$$

for the lefthanded neutrino-righthanded antineutrino pairs be restored as the equation satisfied by the neutrino. It is this equation that now governs massless particles, not only in Minkowski spacetime but also in curved spacetime.

It is clear from the entire discussion that in the course of quantization, classical particles acquire internal spaces and symmetries (little groups). Thus classically only the photons travel with the speed of light but quantum theory allows many

more, such as the neutrinos (although there are some recent indications that the neutrinos have a very small but positive mass).

The direct approach to wave equations of L. Schwartz and Hilbert spaces of distributions on spacetime. The method of first getting the bundles in momentum space and then obtaining the wave equations by Fourier transforms that we have followed above is indirect. It is natural to ask if one can construct the wave equations and the Hilbert spaces directly on spacetime. This was carried out by L. Schwartz in a beautiful memoir[22]. Schwartz determined all Hilbert subspaces $\mathcal{H}$ of the space $\mathcal{D}'(M)$ of distributions on Minkowski spacetime M, with scalar or vector values such that

(i) The natural inclusion $\mathcal{H} \hookrightarrow \mathcal{D}'(M)$ is continuous.

(ii) The natural action of the Poincaré group on $\mathcal{D}'(M)$ leaves $\mathcal{H}$ invariant and induces a unitary representation on it.

Not surprisingly his classification is the same as the Wigner one. However by focussing attention on distributions on spacetime his analysis reveals how restrictive the requirements of Hilbert structure, unitarity and Poincaré invariance are. For instance translation invariance already implies that all elements of $\mathcal{H}$ are tempered.

The analysis of Schwartz does not exhaust the subject of wave equations. Indeed, the same representation is obtained by wave equations which look very different formally, and the different versions are important in interpretation. One can formulate the general notion of a relativistic wave equation and try to classify them. Many people have worked on this problem and the results in some sense are still not definitive. For a discussion of these aspects see[23].

1.7. Bosons and fermions. The concept of bosons and fermions arises when one wishes to treat a system of *identical particles* quantum mechanically. If $\mathcal{S}_i (1 \leq i \leq N)$ are quantum systems, then the natural way to represent the states of $\mathcal{S}$, the system composed of the $\mathcal{S}_i$, is to take its Hilbert space as $\mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_N$ where $\mathcal{H}_i$ is the Hilbert space of $\mathcal{S}_i$. Thus if $\mathcal{S}_i$ is in the state ψ_i, then $\mathcal{S}$ is in the state $\psi_1 \otimes \ldots \otimes \psi_N$. However if $\mathcal{S}_i = \mathcal{S}_0$ is the systems of a particle such as the electron or the photon, the quantum theoretic description of $\mathcal{S}$ must take into account the purely quantum theoretic phenomenon that the *particles are indistinguishable.* For instance, the theory must allow for the *Pauli exclusion principle* according to which two electrons cannot occupy the same state. It was found that the correct way to describe an N-electron system is to use the space $\Lambda^N(\mathcal{K})$ of *antisymmetric tensors* in $\mathcal{K}^{\otimes N}$, $\mathcal{K}$ being the space of states of a single electron. Similarly, in dealing with a system of N photons the correct space was found to be $S^N(\mathcal{K})$, the space of *symmetric tensors* in $\mathcal{K}^{\otimes N}$ where $\mathcal{K}$ is now the space of states of a single photon.

Let P^a, P^s be the orthogonal projection from the full tensor product onto the space of antisymmetric and symmetric tensors. If ψ is in $\mathcal{K}$, $P^a(\psi \otimes \ldots \otimes \psi)$ is the state in which all the electrons are in the state ψ, and as this is 0 for $N \geq 2$ we see that this model is compatible with the exclusion principle. But for photons $\psi \otimes \ldots \otimes \psi$ is already symmetric and nonzero and represents the state where all the photons are in the state ψ. There is nothing which forbids this, and in fact this possibility is crucial in the construction of the laser.

Experimentally it has been found that all particles belong to one of these two categories, namely, those whose N-particle systems are modeled by $\Lambda^N(\mathcal{K})$, and those whose N-particle systems are modeled by $S^N(\mathcal{K})$. The former type of particles are called *fermions* after the great Italian physicist E. Fermi, and the latter kind *bosons*, after the great Indian physicist S. N. Bose.

Let us now look more closely into the mathematical description of systems of identical particles without assuming anything except the indistinguishability of the particles. Let $\mathcal{K}$ be the Hilbert space of states of a single particle. If there are N particles, then to start with the Hilbert space of states of the N–particle system may be taken as $\mathcal{H}_N = \mathcal{K}^{\otimes N}$. This space carries an obvious action of the group S_N, the group of permutations of $\{1, 2, \ldots, N\}$. The indistinguishability of the particles may then be expressed by saying that the observable algebra is the centralizer of S_N, the algebra $\mathcal{O}$ of all bounded operators commuting with S_N. We shall now decompose $\mathcal{H}_N$ with respect to the action of S_N. For any irreducible representation π of S_N of dimension $d(\pi)$ let P_π be the operator

$$P_\pi = \frac{d(\pi)}{N!} \sum_{s \in S_N} \chi_\pi(s)^{\mathrm{conj}} s$$

where we write s for the operator corresponding to s and χ_π is the character of π. It is easy to see that P_π is a projection, and in fact, it is the projection on the span of all subspaces that transform according to π under the action of S_N. Let

$$\mathcal{H}_N[\pi] = P_\pi \mathcal{H}_N.$$

If M is any subspace of $\mathcal{H}_N$ transforming according to π and $L \in \mathcal{O}$, then $L[M]$ is either 0 or transforms according to π and so $\mathcal{H}_N[\pi]$ is stable under L. Thus any element of the observable algebra $\mathcal{O}$ commutes with each P_π. We now have a decomposition

$$\mathcal{H}_N[\pi] \simeq V[\pi] \otimes \mathcal{K}_\pi$$

where:

(i) $V[\pi]$ is a model for π.

(ii) An operator of $\mathcal{H}_N[\pi]$ lies in $\mathcal{O}$ if and only if it is of the form $1 \otimes A$ where A is an operator of $\mathcal{K}_\pi$.

Hence the observable algebra $\mathcal{O}$ has the decomposition

$$\mathcal{O} = \bigoplus_\pi (1 \otimes \mathcal{O}_\pi)$$

where $\mathcal{O}_\pi$ is the full algebra of all bounded operators on $\mathcal{K}_\pi$. This is a situation which we have discussed earlier. After that discussion it is clear that the states may now be identified with

$$\bigcup_\pi \mathbf{P}(\mathcal{K}_\pi).$$

We thus have *superselection sectors* corresponding to the various π. There will be no superposition between states belonging to different sectors. For fixed π if we take the Hilbert space $\mathcal{K}_\pi$ as the Hilbert space of states we get a model for treating N identical particles *obeying π-statistics.*

The group S_N has two basic representations: the trivial one and the alternating one, the latter being the representation in dimension 1 that sends each permutation s to its signature sgn(s). We then get the two projections

$$\frac{1}{N!}\sum_s s, \qquad \frac{1}{N!}\sum_s \mathrm{sgn}(s).$$

The corresponding spaces $\mathcal{H}_N[\pi]$ are respectively

$$S^N(\mathcal{K}), \qquad \Lambda^N(\mathcal{K})$$

where $S^N(\mathcal{K})$ is the space of *symmetric tensors* and $\Lambda^N(\mathcal{K})$ is the space of *antisymmetric tensors.* In physics only these two types of statistics have been encountered. Particles for which the states are represented by $S^N(\mathcal{K})$, the bosons, are said to obey the *Bose-Einstein statistics*, while particles for which the states are represented by $\Lambda^N(\mathcal{K})$, the fermions, are said to obey the *Fermi-Dirac statistics*.

The essential question at this stage is the following. Can one tell, from the properties of a *single* particle, the type of statistics obeyed by a system consisting of several particles of the same type? It turns out, and this is a consequence of special relativity, that the statistics are completely determined by the spin of the particle. This is the so-called *spin-statistics theorem* in relativistic quantum field theory; it says that particles with half-integral spin are fermions and obey

the statistics corresponding to the signature representation (Fermi-Dirac statistics), while particles with integral spin are bosons and obey the statistics corresponding to the trivial representation (Bose-Einstein statistics). Thus for a system of N particles with half-integral spin we use $\Lambda^N(\mathcal{K})$ as the Hilbert space of states and for a system of N particles with integral spin we use $S^N(\mathcal{K})$ as the Hilbert space of states. This distinction is of crucial importance in the theory of superconductivity; properties of bulk matter differ spectacularly depending on whether we are dealing with matter formed of particles of integral or half-integral spin.

1.8. Supersymmetry as the symmetry of a $\mathbf{Z}_2$–graded geometry. In a quantum field theory which contains interacting particles of both spin parities, the Hilbert space $\mathcal{K}$ of 1-particle states has a decomposition

$$\mathcal{K} = \mathcal{K}_0 \oplus \mathcal{K}_1$$

where $\mathcal{K}_0$ (resp. $\mathcal{K}_1$) is the space of states where there is one boson (resp. one fermion). The N-particle space is then

$$\mathcal{H}_N = \bigoplus_{1 \le d \le N} S^d(\mathcal{K}_0) \otimes \Lambda^{N-d}(\mathcal{K}_1).$$

The full Hilbert space in which the particle number is not fixed is then

$$\mathcal{H} = S(\mathcal{K}_0) \otimes \Lambda(\mathcal{K}_1).$$

People slowly realized that it would be advantageous to have a single unified framework in which there would be no necessity to treat separately the bosonic and fermionic cases* and that the unified treatment would result in increased clarity and understanding. Eventually the algebraic aspects of such a unified theory came to be seen as a linear theory where all (linear) objects are systematically graded by $\mathbf{Z}_2$, just as the Hilbert space of 1-particles above was graded into bosonic and fermionic parts. In the meantime, in the early 1970's, several groups of physicists (Gol'fand, Likhtman, and Volvov, and Akulov, Zumino, Wess) almost simultaneously came up with a notion of infinitesimal symmetry of such graded spaces, and viewed it as a type of symmetry not encountered hitherto, namely a symmetry that sent bosonic states into fermionic states and vice versa. These symmetries were called *supersymmetries*, and remarkably, they depended on parameters consisting of both usual variables and variables from a *Grassmann algebra.* The appearance

* Separate but equal facilities are inherently discriminatory!

of the Grassmann or exterior algebra is related to the circumstance that in quantum field theory the Fermi fields obey not commutation rules but anticommutation rules. It was soon realized (Salam and Strathdee) that a systematic theory of spaces with usual and Grassmann coordinates could be developed in great depth, and that classical field theory on these *superspaces* would lead, upon quantization, to supersymmetric quantum field theories and gauge theories (Wess, Zumino, Ferrara, Salam, Strathdee). Then in 1976 a supersymmetric extension of Einstein's theory of gravitation (supergravity) was discovered by Ferrara, Freedman and van Nieuwenhuizen, and a little later, by Deser and Zumino. With this discovery supersymmetry became the natural context for seeking a unified field theory[24].

The infinitesimal supersymmetries discovered by the physicists would become the *super Lie algebras* and their corresponding groups the *super Lie groups.* A systematic theory of super Lie algebras culminating in the classification of simple super Lie algebras over an algebraically closed field was carried out by V. Kac shortly after the first papers on supergroups and algebras appeared in the physics literature[25]. Of course as long as one can work with the infinitesimal picture the theory of super Lie algebras is perfectly adequate and it is immediately accessible because it is a linear theory and is modeled after the well-known theory of simple Lie algebras; but for a fuller understanding the deeper (nonlinear) theory of supermanifolds and super Lie groups cannot be evaded. First introduced by Salam and Strathdee, the concept of supermanifolds and super Lie groups was developed by the physicists. Among mathematicians one of the earliest pioneering efforts was that of F. A. Berezin[26] who tried to emphasize the idea that this was a new branch of algebra and analysis. Among the important more recent works exposing the theory for mathematicians were the articles and books of B. De Witt, D. Leites, and Yu. Manin as well as the expositions of P. Deligne and J. Morgan[27], and the lectures of D. Freed[28].

Informally speaking, a supermanifold is a manifold in which the coordinate functions are smooth functions of the usual coordinates as well as the so-called odd variables. The simplest example of this is $\mathbf{R}^p$ on which the coordinate functions form the algebra $C^\infty(\mathbf{R}^p) \otimes \mathbf{R}[\theta_1, \ldots, \theta_q]$ where $\theta_j (1 \leq j \leq q)$ are odd variables which are anticommuting, i.e., satisfy

$$\theta_j \theta_k + \theta_k \theta_j = 0 \qquad (1 \leq j, k \leq q).$$

Such a space is denoted by $\mathbf{R}^{p|q}$, and the general supermanifold is obtained by gluing spaces which locally look like $\mathbf{R}^{p|q}$. While this definition imitates that of smooth manifolds with obvious variants in the analytic and holomorphic categories, there is a striking difference: the odd variables are *not numerical* in the sense that they all have the value 0. So they are more subtle, and a supermanifold is more like a scheme

of Grothendieck on which the rings of the structure sheaf have nilpotent elements; indeed, any odd element in the structure sheaf of a supermanifold is nilpotent. So a supermanifold is a generalization of a manifold at a fundamental level. However the techniques for studying supermanifolds did not have to be freshly created; one could simply follow the ideas of Grothendieck's theory of schemes. Supermanifolds are more general than schemes because the coordinate rings are not commutative but *supercommutative*, a mildly noncommutative variant of commutative rings. If we drop the smoothness requirement in a supermanifold we obtain a superscheme which is the most general geometric object yet constructed. Super Lie groups, and more generally super group schemes, are the symmetries of these objects.

REFERENCES

[1] It is not often that one speaks of beauty and coherence in the physical description of the world around us. Supersymmetry, as a theory with almost no experimental confirmation, is very much like mathematics in that it relies on internal esthetics to a much larger extent than traditional physical theories, although this situation may change with the advent of supercolliders in the TeV range. The mathematical path to physical insight was the modus operandi for one of the greatest physicists of all time, P. A. M. Dirac. In his famous paper on monopoles,

P. A. M. Dirac, *Quantized singularities in the electromagnetic field,* Proc. Roy. Soc. Lond., **A133** (1931), 60–72,

Dirac has this to say on this issue:

"The steady progress of physics requires for its theoretical formulation a mathematics that gets continually more advanced. This is only natural and to be expected... Non-euclidean geometry and noncommutative algebra, which were at one time considered to be purely fictions of the mind and pastimes for logical thinkers, have now been found to be very necessary for the description of general facts of the physical world. It seems likely that this process of increasing abstraction will continue in the future and that advance in physics is to be associated with a continual modification and generalisation of the axioms at the base of the mathematics rather than with a logical development of any one mathematical scheme on a fixed foundation.

The theoretical worker in the future will therefore have to proceed in a more indirect way. The most powerful method of advance that can be suggested at present is to employ all

the resources of pure mathematics in attempts to perfect and generalise the mathematical formalism that forms the existing basis of theoretical physics, and after each success in this direction, to try to interpret the new mathematical features in terms of physical entities...

Here is another quotation from Y. Nambu:

Y. Nambu, *Broken Symmetry: Selected Papers of Y. Nambu*, 1995, World Scientific, pp. 383–389, eds. T. Eguchi and K. Nishijima.

The Dirac mode is to invent, so to speak, a new mathematical concept or framework first, and then try to find its relevance in the real world, with the expectation that (in a distorted paraphrasing of Dirac) a mathematically beautiful idea must have been adopted by God. Of course the question of what constitutes a beautiful and relevant idea is where physics begins to become an art.

I think this second mode is unique to physics among the natural sciences, being most akin to the mode practiced by the mathematicians. Particle physics, in particular, has thrived on the interplay of these two modes. Among examples of this second approach, one may cite such concepts as

Magnetic monopole

Non-Abelian gauge theory

Supersymmetry

On rare occasions, these two modes can become one and the same, as in the cases of Einstein gravity and the Dirac equation...

[2] See

S. Ferrara, *Supersimmetria*, in *L'Italia al CERN*, 369. F. Menzinger ed. Ufficio Pubblicazioni INFN, Frascati, 1995.

[3] The suggestion of Y. Aharanov and Bohm is contained in

Y. Aharonov and D. Bohm, *Significance of electromagnetic potentials in quantum theory,* Phys. Rev., **115** (1959), 485–491.

For the experiment of Chambers see

R. G. Chambers, *Shift of an electron interference pattern by enclosed magnetic flux,* Phys. Rev. Lett., **5** (1960), 3–5.

[4] The books treating classical physics that are close to the spirit of our discussions are the following.

Classical mechanics:

V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer Verlag, 1978.

Electromagnetism and Gravitation:

H. Weyl, *Space, Time, and Matter*, Dover, 1950.

Quantum Mechanics.

P. A. M. Dirac, *Principles of Quantum Mechanics*, Oxford, 1958.

J. Von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton, 1955.

H. Weyl, *The Theory of Groups and Quantum Mechanics*, Dover, 1930.

[5] The references for foundational matters in quantum mechanics are the following:

E. G. Beltrametti and G. Cassinelli, *The Logic of Quantum Mechanics*, Vol 15, Encyclopedia of Mathematics and its Applications, ed. Gian-Carlo Rota, 1981.

V. S. Varadarajan, *Geometry of Quantum theory*, Springer Verlag, 1985 (Second Edition).

J. A. Wheeler and W. H. Zurek, *Quantum theory and Measurement*, Princeton Series in Physics, 1983.

[6] See[4]

[7] In the 1950's J. Schwinger wrote a series of papers on the foundations of quantum mechanics which were summarized in a book:

J. Schwinger, *Quantum Kinematics and Dynamics*, Benjamin, New York, 1970.

He discusses his measurement algebra in his paper given on the occasion of the centenary marking Weyl's birth:

J. Schwinger, *Hermann Weyl and Quantum Kinematics,* in *Exact Sciences and Their Philosophical Foundations,* Verlag Peter Lang, Frankfurt, 1988, 107–129.

[8] T. Digernes, V. S. Varadarajan and S. R. S. Varadhan, *Rev. Math. Phys.,* **6** 1994, 621.

[9] J. Schwinger, *Quantum Mechanics: Symbolism of Atomic Measurements*, Springer, 2001. See also his article[7] on the occasion of the Weyl centenary.

[10] R. Arens and V. S. Varadarajan, *On the concept of Einstein-Podolsky-Rosen states and their structure*, Jour. Math. Phys., **41** (2000), 638.

[11] See the preface to Weyl's book[4] and Schwinger's Weyl centenary address[7].

[12] The pioneering papers are:

J. E. Moyal, Proc. Camb. Phil. Soc., **45**, 1949, 99.

F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz and D. Sternheimer, Ann. Phys. (N. Y.), **111**, 111.

The subject of quantization as deformation took off in the 1980's and led directly to the theory of quantum groups.

[13] For Wigner's theorem see my book[5], 104. For more recent discussions see

S. Weinberg, *The Quantum Theory of Fields*, Vol I, 91.

G. Cassinelli, P. Lahti, and A. Levrero, *Symmetry groups in quantum mechanics and the theorem of Wigner on the symmetry transformations*, Rev. Math. Phys., **9** (1997), 921.

[14] S. Weinberg, *Testing quantum mechanics*, Ann. Phys., **194** (1989), 336.

[15] See Weyl's book[4] for the beginnings of the theory of projective representations of Lie groups and the projective representations of finite groups going back to I. Schur. For the complete theory see my book[5] and

K. R. Parthasarathy, *Multipliers on locally compact groups*, Lecture Notes in Mathematics, Springer, No. 93, Berlin, 1969.

The seminal papers are:

V. Bargmann, *Unitary ray representations*, Ann. Math., (1954).

G. W. Mackey, *Unitary representations of group extensions*, I. Acta Math., **99** (1958), 265.

G. W. Mackey, *Les ensembles Boreliens et les extensions des groupes*, Jour. de Math. Pures et. Appl., **36** (1957), 171.

For more recent contributions and references, see

G. Cassinelli, E. De Vito, P. Lahti and A. Levrero, *Symmetries of the quantum state space and group representations*, Rev. Math. Phys, **10** (1998), 893.

P. T. Diwakaran, *Symmetries and quantisation*, Preprint; also Rev. Math. Phys., **6** (1994), 167.

[16] For a general historical and mathematical discussion of Weyl's role in the discovery of gauge theory see

V. S. Varadarajan, *Vector Bundles and Connections in Physics and Mathematics: Some Historical Remarks*, to appear in *A Tribute to C.S.Seshadri-A Collection of articles on geometry and representation theory*, Hindustan Book Agency, New Delhi, India, 2003.

The book

L. O'Raifeartaigh, *The Dawning of Gauge Theory*, Princeton Series in Physics, 1997

contains the reprints of some of the most fundamental papers in the subject.

[17] H. Minkowski, *Space and Time*, in The Principles of Relativity, Dover 1952.

Minkowski's address started with these famous words:

The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. They are radical. Henceforth, space by itself, and time by itself, are doomed to fade away in mere shadows, and only a kind of union of the two will preserve an independent reality.

[18] A. D. Aleksandrov, Canad. J. Math., **19** (1967), 1119.

[19] The basic reference (which also contains the herglotz formula) is

W. Pauli, *The Theory of Relativity*, Dover, 1981.

For a beautiful exposition of the basic ideas see

J. Schwinger, *Einstein's Legacy*, Scientific American Library, 1986.

[20] Wigner's great 1939 paper is in

E. P. Wigner, *Unitary representations of the inhomogeneous Lorentz group*, Ann. Math., **40** (1939), 149.

[21] See

E. G. Effros, *Transformation Groups and C^*-algebras,* Ann. Math.,**81** (1965), 38–55.

For further and more recent work, see

H. Becker and A. H. Kechris, *The Descriptive Set Theory of Polish Group Actions*, London Math. Soc. Lecture Note Series **232**, Camb. Univ. Press, 1996.

[22] See

L. Schwartz, *Application of Distributions to the Theory of Elementary Particles in Quantum Mechanics*, Gordon and Breach, 1968.

[23] G. Velo and A. S. Wightman (eds.) *Invariant Wave Equations*, Springer-Verlag, 1978.

[24] There are two comprehensive collections of basic articles on supersymmetry which are very valuable:

S. Ferrara (ed.) *Supersymmetry*, Vols. 1 and 2, World Scientific, 1987.

A. Salam and E. Sezgin, *Supergravities in Diverse Dimensions*, World Scientific, 1987.

[25] V. Kac, *Lie superalgebras*, Adv. in Math., **26** (1977), 8.

[26] F. A. Berezin, *Introduction to Superanalysis*, D. Reidel, 1987.

[27] The recent expositions close to the spirit of our discussions are as follows.

B. Kostant, *Graded Manifolds, Graded Lie Theory, and Prequantization*, Lecture Notes in Mathematics, **570** (1977), Springer.

D. A. Leites, *Introduction to the theory of supermanifolds*, Uspeki Mat. Nauk **35** (1980), no. 1, 3; Russian Math. Surveys, **35** no. 1, (1980), 1.

Yu. I. Manin, *Gauge Field Theory and Complex Geometry*, Springer Verlag, 1988.

P. Deligne and J. W. Morgan, *Notes on Supersymmetry* (*following Joseph Bernstein*). *Quantum Fields and Strings*: *A Course for Mathematicians*, 41. Amer. math. Soc., 1999.

[28] D. S. Freed, *Five Lectures on Supersymmetry*, Amer. Math. Soc., 1999.