PHYSICAL REVIEW D VOLUME 33, NUMBER 6 15 MARCH 1986

Second-order wave equation for spin-$\frac{1}{2}$ fields. II. The Hilbert space of the states

Nicola Cufaro Petroni
Istituto Nazionale di Fisica Nucleare e Dipartimento di Fisica dell'Università, via Amendola 173, 70126, Bari, Italy

Philippe Gueret
Institut de Mathématiques Pures et Appliées, Université P. et M. Curie, 4, Pl. Jussieu, 75230, Paris, France

Jean Pierre Vigier and Anastasios Kyprianidis
Institut Henri Poincaré, Laboratoire de Physique Théorique, 11, rue P. et M. Curie, 75231, Paris, France
(Received 23 September 1985)

It is shown that, for spin-$\frac{1}{2}$ fields ruled by a second-order wave equation, it is possible to define a conserved current density whose zero component is positive definite. Hence one can (1) give a coherent statistical interpretation of the wave function, and (2) define a Hilbert space of the states with all the usual quantum-mechanical formalism. A new linearized form of the wave equation, completely equivalent to the second-order one, is finally presented.

I. INTRODUCTION

It is a general opinion among theoreticians that it is impossible to give a coherent statistical interpretation for fields obeying second-order, relativistic, quantum wave equations. This idea follows from the fact that, for differential equations containing second-order derivatives in time, the usual form of the conserved current density has as zero component (namely, as conserved density) a quantity that is not positive definite.[1] This has two correlated main consequences: (a) the said zero component of the conserved current cannot be directly interpreted as a probability density; (b) the scalar product, defined by means of this current, cannot be utilized to define coherently a norm on the vector space of the states of the quantum system.

It is thus apparently very difficult, if not impossible, to build on this ground a relativistic generalization of the usual quantum formalism, with its Hermitian operators acting in a Hilbert space. As is also well known,[1] the first-order Dirac equation is born exactly in order to give an answer to this problem, and consists in a linearization (obtained by means of a spinorial formalism) of the second-order, relativistic wave equation, which yields a positive probability density.

On the other hand, we must remark that the relativistic, second-order wave equation still constitutes the most natural generalization of Schrödinger's wave mechanics. From this standpoint, despite the very impressive quantity of results directly obtained from the Dirac equation, it is a bit unsatisfactory that this linearization cuts off a substantial part of the solutions of the second-order equation. Nothing similar, indeed, is usually done on the Klein-Gordon or on the Proca equation. Moreover, it is well known, for example, that, in their theory of the Fermi interaction, Feynman and Gell-Mann[2] described the fermion states by means of spinors that are not solutions of the Dirac equation, and yield results (such as parity violation) which have to be superimposed on Dirac's solutions.

In the spirit of these remarks, this paper will be devoted to show that, for spin-$\frac{1}{2}$ fields ruled by a second-order, relativistic wave equation, it is possible to define a conserved current density, whose zero component is always positive definite. Of course, that will allow us to define coherently a statistical interpretation, a Hilbert space of states, and all the ordinary machinery of usual quantum mechanics, without restricting ourselves to the solutions of the Dirac equation. Moreover, the formalism is defined in such a way that we will recover all the usual results, when we consider the particular case of the Dirac solutions.

We briefly recall here the notation and the results of the preceding papers[3] devoted to this argument. If we write

$$D_\mu=\frac{1}{mc}\left[i\hbar\partial_\mu-\frac{e}{c}A_\mu\right], \tag{1.1}$$

the second-order, relativistic wave equation we are talking about is the four-spinor equation (also called second-order Dirac equation)

$$(I-\not{D}^2)\psi(x)=0\,, \tag{1.2}$$

with $\not{D}=\gamma_\mu D^\mu$, where γ_μ represent the usual 4×4 Dirac matrices. In this notation the Dirac equation is written in the form

$$(I-\not{D})\psi(x)=0\,. \tag{1.3}$$

Along with Eq. (1.3) we will also consider the associated equation

$$(I+\not{D})\psi(x)=0\,. \tag{1.4}$$

We will denote by $\mathscr{F}$, $\mathscr{D}_-$, and $\mathscr{D}_+$ the sets of spinors $\psi(x)$ which are solutions, respectively, of (1.2), (1.3), and (1.4). Let us now summarize some preliminary results

that were proved in the preceding papers.[3]

(a) The only spinor common to $\mathscr{D}_+$ and $\mathscr{D}_-$ is the identically zero spinor $\psi(x)=0$.

(b) $\mathscr{D}_+$ and $\mathscr{D}_-$ are vector subspaces of the vector space $\mathscr{F}$.

(c) There is a one-to-one relation between $\mathscr{D}_+$ and $\mathscr{D}_-$, in the sense that, for each ψ_+ of $\mathscr{D}_+$, there exists one and only one ψ_- in $\mathscr{D}_-$, such that $\psi_+=\gamma_5\psi_-$ (with $\gamma_5=i\gamma_0\gamma_1\gamma_2\gamma_3$), and vice versa.

(d) $\mathscr{F}$ is the direct sum of $\mathscr{D}_+$ and $\mathscr{D}_-$, in the sense that, for an arbitrary ψ of $\mathscr{F}$, we can always construct two spinors

$$\psi_+=\tfrac{1}{2}(I-\not{D})\psi\ ,\quad \psi_-=\tfrac{1}{2}(I+\not{D})\psi\ , \tag{1.5}$$

respectively, in $\mathscr{D}_+$ and $\mathscr{D}_-$, such that $\psi=\psi_++\psi_-$. Vice versa, each linear combination of elements of $\mathscr{D}_+$ and $\mathscr{D}_-$ is an element of $\mathscr{F}$.

We remark that, at this stage, concepts such as scalar product, norm, orthogonality, and so on, are not defined. Moreover, the above-mentioned properties show that, if we want to deal with the complete space $\mathscr{F}$ of the solutions of (1.2), we must take into account both Eqs. (1.3) and (1.4), and not only the Dirac equation (1.3). In the subsequent sections we will show that this can be done without destroying the positivity of the conserved density. This result is obtained by mixing the Dirac spinorial formalism with a procedure similar to that introduced by Feshbach and Villars[4] for the scalar field equation. A more precise comparison with this approach will be discussed in the following sections of the paper.

II. THE POSITIVE CONSERVED DENSITY

As is well known in the general field theory,[5] it is possible to define several conserved current densities, by means of suitable local transformations which leave invariant the Lagrangian density of the given physical system. In order to point out the existence of a particular current with a positive, conserved zero component, let us start with a generalized (i.e., nonreal) Lagrangian density leading, as Euler-Lagrange equations, to (1.2), namely,

$$\mathscr{L}=\bar{\phi}\psi-\overline{\not{D}\phi}\not{D}\psi\ . \tag{2.1}$$

This unusual utilization of a complex Lagrangian is introduced here only because it allows us to define a generalized form of the conserved current density containing, as a particular case, not only the well-known form of the current, but also the form we are looking for. If one dislikes this approach, however, we must immediately remark that the conservation of the proposed currents is a straightforward consequence of Eq. (1.2), and it can be proved also by direct calculation, without using generalized complex quantities at all.

From (2.1) we find that the usual form of the generalized conserved current density is

$$J_\mu(x)=\frac{imc}{2\hbar}\left[\frac{\partial\mathscr{L}}{\partial(\partial^\mu\psi)}\psi-\bar{\phi}\frac{\partial\mathscr{L}}{\partial(\partial^\mu\bar{\phi})}\right]$$

$$=\tfrac{1}{2}(\overline{\not{D}\phi}\gamma_\mu\psi+\bar{\phi}\gamma_\mu\not{D}\psi)\ . \tag{2.2}$$

Its conservation can be directly verified, under the condition that ϕ,ψ satisfy (1.2). Of course (2.2) becomes real when $\phi=\psi$, and then reduces to the usual form of the Dirac current

$$J_\mu(x)=\bar{\psi}\gamma_\mu\psi \tag{2.3}$$

when $\phi=\psi$ satisfies the Dirac equation (1.3). In the real case $\phi=\psi$, it is well known that the zero component of (2.2), i.e.,

$$J_0(x)=\mathrm{Re}(\psi^\dagger\not{D}\psi)\ , \tag{2.4}$$

is not positive definite, unless ψ is a solution of the Dirac equation. It is on this ground that an expression like (2.2) is not considered as a suitable starting point to define a scalar product, and cannot be utilized to work out a statistical interpretation of the second-order equation (1.2).

To overcome this difficulty we remark now that it is possible to obtain a real form for (2.2) also when $\phi\neq\psi$. Indeed, it is very easy to show that, if ψ is a solution of (1.2), so is $\chi=\not{D}\psi$ since we have

$$(I-\not{D}^2)\chi=(I-\not{D}^2)\not{D}\psi=\not{D}\psi-\not{D}\psi=0\ . \tag{2.5}$$

In fact, the operator $\not{D}$ defines a one-to-one map between the elements of $\mathscr{F}$, since we have also $\not{D}\chi=\not{D}^2\psi=\psi$. We can thus substitute $\not{D}\psi$ for ψ in (2.2) and get another generalized current density

$$j_\mu(x)=\tfrac{1}{2}(\overline{\not{D}\phi}\gamma_\mu\not{D}\psi+\bar{\phi}\gamma_\mu\psi)\ , \tag{2.6}$$

whose conservation equation can be verified by direct calculation, if only ϕ and ψ satisfy (1.2). If we take $\phi=\psi$ in (2.6), we now obtain a real conserved current density, with

$$j_0(x)=\tfrac{1}{2}[(\not{D}\psi)^\dagger\not{D}\psi+\psi^\dagger\psi]\geq 0\ . \tag{2.7}$$

Of course, this reduces to the usual Dirac expression when $\phi=\psi$ is a solution of (1.3). The positive conserved density (2.7) can thus now be considered as a probability density, and hence can constitute a coherent starting point for a statistical interpretation of the relativistic quantum fields ruled by Eq. (1.2). Moreover, it will be used as basis for the definition of a scalar product for the state vectors.

In fact, we know[6] that the Lorentz scalar

$$\int_\sigma F_\mu(x)d\sigma^\mu(x)\ , \tag{2.8}$$

obtained by integrating on a spacelike hypersurface σ a given vector field $F_\mu(x)$, is a quantity independent of the particular choice of σ, if $F_\mu(x)$ is conserved in the sense that $\partial_\mu F^\mu(x)=0$. As usual, $d\sigma^\mu(x)$ represents the pseudovector surface element on σ, whose components are

$$d\sigma^\mu=(dx^1dx^2dx^3,dx^0dx^2dx^3,dx^0dx^1dx^3,dx^0dx^1ds^2)\ . \tag{2.9}$$

It then follows that the scalar, Hermitian, bilinear form defined on the vector space of the four-spinors as

$$(\phi,\psi)=\int_\sigma j_\mu(x)d\sigma^\mu(x)$$

$$=\tfrac{1}{2}\int_\sigma(\overline{\not{D}\phi}\gamma_\mu\not{D}\psi+\bar{\phi}\gamma_\mu\psi)d\sigma^\mu(x) \tag{2.10}$$

is independent of the particular σ chosen for its evaluation. If, for example, σ is the hyperplane orthogonal to

the x^0 axis we will have

$$(\phi,\psi)=\tfrac{1}{2}\int[(\not{D}\phi)^\dagger\not{D}\psi+\phi^\dagger\psi]d^3\mathbf{r}\,, \tag{2.11}$$

so that

$$(\psi,\psi)=\tfrac{1}{2}\int[(\not{D}\psi)^\dagger\not{D}\psi+\psi^\dagger\psi]d^3\mathbf{r}\geq 0\,, \tag{2.12}$$

where the case $(\psi,\psi)=0$ is verified only if the spinor $\psi(x)$ is identically zero. The bilinear form $(\cdot,\cdot)$ can, consequently, be considered as the starting point to define a norm on the vector space of the states, as we will see in the following section.

III. THE HILBERT SPACE OF THE STATES

If we want to reproduce the construction of a Hilbert space of the states by following the standard procedures, we must now remark that the state of a system described by the second-order equation (1.2), at a fixed time x^0, cannot be considered as completely specified by the spinor $\psi(x)|_{x^0}$ as a function of its spatial coordinates x^k. Indeed, unlike what happens in the Schrödinger and the Dirac equations, the knowledge of $\psi(x)$ at a fixed time is not sufficient to determine the subsequent time evolution of the spinor. In other words, the knowledge of $\psi(x)$ in all space-time contains more information than Eq. (1.2) associated with the incomplete initial condition on $\psi(x)|_{x^0}$. Hence, if we want to study the quantum mechanics of our relativistic system at a fixed time and define the Hilbert space of all its states, we cannot deal only with a spinor $\psi_0(\mathbf{r})$ at a given time, since it, by itself, constitutes an incomplete determination of the state.

Equation (1.2) is a differential equation which contains second-order derivatives in time. As a consequence, the determination of the time evolution of a spinor needs two initial conditions, namely,

$$\begin{aligned}&\psi(x)|_{x^0=0}=\chi_1(\mathbf{r})\,,\\ &\partial_0\psi(x)|_{x^0=0}=\chi_2(\mathbf{r})\,,\end{aligned} \tag{3.1}$$

where $\chi_1(\mathbf{r})$ and $\chi_2(\mathbf{r})$ are two arbitrary and independent four-component functions of the space coordinates only. The more general and covariant way to express the conditions (3.1) is to fix the values of $\psi(x)$ and $\partial_0\psi(x)$ on an arbitrary, spacelike hypersurface σ_0. In this case[6] we will replace the succession of time instants x^0, by a continuous slicing of spacelike hypersurfaces σ, and Eq. (1.2) [associated with the conditions for $\psi(x)$ and $\partial_0\psi(x)$ on the given σ_0] will fix the evolution of the spinor $\psi(x)$ in all space-time. Moreover we remark that the initial conditions (3.1) can be also given in the equivalent form

$$\begin{aligned}&\psi(x)|_{x^0=0}=\phi_1(\mathbf{r})\,,\\ &\not{D}\psi(x)|_{x^0=0}=\phi_2(\mathbf{r})\,,\end{aligned} \tag{3.2}$$

in the sense that we can always deduce (3.1) and (3.2) from each other. The formulation (3.2) (or its covariant form on arbitrary spacelike hypersurface σ_0) has the advantage, with respect to (3.1), that ϕ_1 and ϕ_2 are both four-spinors, so that it can be considered as the more Lorentz-symmetric form of the initial-value problem for Eq. (1.2).

In light of these remarks, we will say that, at a given time x^0 (or on a given hypersurface σ_0), the state of our system is specified by an ordered couple of spinors $[\phi_1(\mathbf{r}),\phi_2(\mathbf{r})]$, in the sense that this couple, along with Eq. (1.2), completely determines the spinor $\psi(x)$ in all space-time. Hence, we take as vector space of the states of our system (on a hypersurface σ or, in particular, at a given time x^0) the space $\mathscr{H}$ of the double-spinors

$$\Psi(\sigma)=\frac{1}{\sqrt{2}}\begin{bmatrix}\psi_1(\sigma)\\ \psi_2(\sigma)\end{bmatrix}\,, \tag{3.3}$$

where $\Psi(\sigma)$ means $\Psi(x)|_{x\in\sigma}$ and so on. $\mathscr{H}$ becomes a Hilbert space if we adopt as scalar product

$$\begin{aligned}\langle\Psi\,|\,\Phi\rangle&=\int_\sigma d\sigma^\mu\overline{\Psi}\,\mathbb{C}_\mu\Phi\\ &=\tfrac{1}{2}\int_\sigma d\sigma^\mu(\bar{\psi}_1\gamma_\mu\phi_1+\bar{\psi}_2\gamma_\mu\phi_2)\,,\end{aligned} \tag{3.4}$$

where $\mathbb{C}_\mu$ are the 8×8 matrices

$$\mathbb{C}_\mu=\begin{bmatrix}\gamma_\mu & 0\\ 0 & \gamma_\mu\end{bmatrix} \tag{3.5}$$

and

$$\overline{\Psi}=\Psi^\dagger\mathbb{C}_0=[\bar{\psi}_1,\bar{\psi}_2]\,. \tag{3.6}$$

If we consider now the time (or σ) evolution of a state, we recall that it must be completely contained in only one spinor function $\psi(x)$ that is a solution of (1.2) and given on all the space-time. As a consequence, the state $\Psi(x)$ on all space-time cannot be compounded by two independent spinors, but must be in a one-to-one correspondence with the elements $\psi(x)$ of $\mathscr{F}$. The preceding analysis of the initial conditions suggests that we associate to each $\psi(x)$ of $\mathscr{F}$ the double-spinor

$$\Psi(x)=\frac{1}{\sqrt{2}}\begin{bmatrix}\psi(x)\\ \not{D}\psi(x)\end{bmatrix}\,, \tag{3.7}$$

so that the generalized initial-value problem becomes

$$\Psi(\sigma_0)=\Phi_0\,, \tag{3.8}$$

where Φ_0 is an arbitrary double-spinor given on σ_0. The expression (3.7) evidently clarifies the connection with the discussion of Sec. II. In fact, if $\Psi(x)$ and $\Phi(x)$ are double-spinors of the form (3.7), the scalar product (3.4), on an arbitrary σ, coincides with the Hermitian bilinear form (2.10) defined on the vector space of the four-spinors:

$$\begin{aligned}\langle\Psi\,|\,\Phi\rangle&=\tfrac{1}{2}\int_\sigma d\sigma^\mu(\bar{\psi}\gamma_\mu\phi+\overline{\not{D}\psi}\gamma_\mu\not{D}\phi)\\ &=(\psi,\phi)\,.\end{aligned} \tag{3.9}$$

By recalling now that (ψ,ϕ) is independent of the σ chosen to calculate it, we deduce also that the scalar product $\langle\Psi\,|\,\Phi\rangle$ does not depend on this choice. Of course, this corresponds to the complete arbitrariness of the σ chosen to give the initial conditions (3.8).

Moreover, the initial-value problem (3.8), which does not contain explicit conditions on the derivatives, suggests

that $\Psi(x)$ must satisfy some first-order differential equation. In fact, it is well known that a second-order equation can, in general, be made equivalent to a system of two first-order equations.[4] Now, the general double spinor

$$\Psi(x)=\frac{1}{\sqrt{2}}\begin{bmatrix}\psi(x)\\ \phi(x)\end{bmatrix} \tag{3.10}$$

is a quantity defined on all the space-time by means of two four-spinors, so that the equation

$$(\mathbb{C}_\mu D^\mu-\mathbb{C})\Psi(x)=\frac{1}{\sqrt{2}}\begin{bmatrix}\not{D} & -I\\ -I & \not{D}\end{bmatrix}\begin{bmatrix}\psi(x)\\ \phi(x)\end{bmatrix}=0\,, \tag{3.11}$$

with

$$\mathbb{C}=\begin{bmatrix}0 & I\\ I & 0\end{bmatrix}, \tag{3.12}$$

is a synthetic form for a system of two equations which is equivalent to (1.2). In fact, it is easy to derive from (3.11) that

$$\phi=\not{D}\psi\,,\quad (I-\not{D}^2)\psi=0\,, \tag{3.13}$$

namely, that $\Psi(x)$ must have the form (3.7), where $\psi(x)$ is a solution of (1.2).

We see here a striking analogy of the preceding procedure with the method adopted by Feshbach and Villars,[4] in reducing the second-order Klein-Gordon equation for a scalar field to a system of the first-order (in time) equations, on a two-component function, constituted by the field and its time derivative. Here, however, the positivity of the conserved density and the perfectly Lorentz-symmetric form of the double spinor (3.7) allow us to give a coherent reformulation of the relativistic quantum mechanics for particles ruled by Eq. (1.2). First of all, the states of our quantum system are represented by double-spinors at a given time (or on a spacelike hypersurface) and their time evolution is ruled by Eq. (3.11). The vector space $\mathscr{H}$ of these double-spinors is a Hilbert space, where the norm is defined by means of the scalar product (3.4), which is independent of the σ chosen for its evaluation. The statistical interpretation, in the usual sense, is based on a conserved probability density of the form $\Psi^\dagger\Psi$, and the observables are represented by Hermitian operators in the given Hilbert space $\mathscr{H}$. If, for example, we maintain the ordinary association between the more common observables and operators, we can also identify a simple Hamiltonian operator. In fact, if $i\hbar\partial/\partial t$ is, as usual, the energy operator, we deduce from (3.11) that

$$\mathbb{C}_0 D_0\Psi(x)=(\mathbb{C}+\mathbb{C}_k D_k)\Psi(x)\,, \tag{3.14}$$

and hence

$$i\hbar\partial_0\Psi(x)=\left[\frac{e}{c}A_0+mc(\mathbb{B}+\mathbb{A}_k D_k)\right]\Psi(x)\,, \tag{3.15}$$

with

$$\mathbb{B}=\mathbb{C}_0\mathbb{C}=\begin{bmatrix}0 & \beta\\ \beta & 0\end{bmatrix},\qquad \mathbb{A}_k=\mathbb{C}_0\mathbb{C}_k=\begin{bmatrix}\alpha_k & 0\\ 0 & \alpha_k\end{bmatrix}, \tag{3.16}$$

so that the Hamiltonian operator is

$$\mathbb{H}=eA_0+mc^2(\mathbb{B}+\mathbb{A}_k D_k)\,. \tag{3.17}$$

IV. THE ENERGY EIGENVALUE PROBLEM

If the Hamiltonian $\mathbb{H}$ is independent of time (i.e., if the external electromagnetic field A_μ is constant in time), the energy eigenvalue problem

$$i\hbar\frac{\partial}{\partial t}\Psi_E(x)=E\Psi_E(x) \tag{4.1}$$

is solved by the stationary states

$$\Psi_E(x)=e^{-iEt/\hbar}\Phi_E(\mathbf{r})\,, \tag{4.2}$$

where, of course, $\Phi_E(\mathbf{r})$ is a solution of

$$\mathbb{H}\,\Phi_E(\mathbf{r})=E\Phi_E(\mathbf{r})\,. \tag{4.3}$$

Hence, from (3.17), we can see that $\Psi_E(x)$ will have the form (3.7), where now the higher four-spinor

$$\psi_E(x)=e^{-iEt/\hbar}\phi_E(\mathbf{r}) \tag{4.4}$$

is a stationary solution of (1.2), and the lower four-spinor is

$$\not{D}\psi_E(x)=e^{-iEt/\hbar}\left[\frac{\beta}{mc^2}(E-eA_0-mc^2\alpha_k D_k)\phi_E(\mathbf{r})\right], \tag{4.5}$$

so that we will have for (4.3)

$$\Phi_E(\mathbf{r})=\frac{1}{\sqrt{2}}\begin{bmatrix}\phi_E(\mathbf{r})\\ \dfrac{\beta}{mc^2}(E-eA_0-mc^2\alpha_k D_k)\phi_E(\mathbf{r})\end{bmatrix}. \tag{4.6}$$

In this sense, it is perfectly equivalent to have a stationary state (4.2) that is a solution of (3.11), or a four-spinor (4.4) that is a solution of (1.2). The stationary character of (4.2) is easily seen from the fact that the conserved probability density

$$\bar{\Psi}_E\,\mathbb{C}_0\Psi_E=\Phi_E^\dagger(\mathbf{r})\Phi_E(\mathbf{r}) \tag{4.7}$$

is a completely time independent function of the spatial coordinates.

In order to better discuss the energy eigenvalue problem, let us remark now that the proposed linearization (3.11) of the second-order equation (1.2) is quite different from the well-known Dirac linearization. In fact, the set of the double-spinors $\Psi(x)$ that are a solution of (3.11) is in a one-to-one correspondence with all the set $\mathscr{F}$ of the solutions of (1.2), and not only with the subset $\mathscr{D}_-$ of the solutions of the Dirac equation (1.3). In other words (3.11) contains both Eqs. (1.3) and (1.4). To directly see that, it is sufficient to remark that the matrix $\mathbb{C}$ commutes with $(\mathbb{C}-\mathbb{C}_\mu D^\mu)$, and has as eigenvalues 1 and -1, and as eigenvectors

$$\Psi_{\pm}(x)=\frac{1}{\sqrt{2}}\begin{bmatrix}\psi(x)\\ \pm\psi(x)\end{bmatrix}, \tag{4.8}$$

that are solutions of the equation

$$\mathbf{C}\,\Psi_{\pm}(x)=\pm\Psi_{\pm}(x)\,. \tag{4.9}$$

Hence, the solutions of (3.11) can be simultaneous eigenvectors of $\mathbf{C}$: if we substitute (4.8) in (3.11) we immediately find that $\psi(x)$ must be a solution of (1.3) or of (1.4), following the choice of the sign in (4.8). In this sense, the Dirac linearization consists in taking on only one of the two eigenvalues of $\mathbf{C}$, and more precisely the value $+1$. It is also very easy to verify that

$$\mathbf{P}_{\pm}=\frac{\mathbf{I}\pm\mathbf{C}}{2} \tag{4.10}$$

are the projection operators that project each double-spinor $\Psi(x)$, that is a solution of (3.11), in the double-spinors $\Psi_{\pm}(x)$ of the type (4.8). In this case we will have also that

$$\Psi(x)=\Psi_{+}(x)+\Psi_{-}(x)\,. \tag{4.11}$$

These remarks show that the relations among the sets of the double-spinors $\Psi(x),\Psi_{+}(x),\Psi_{-}(x)$ are exactly the same as the relations among the corresponding sets of four-spinors $\mathscr{F},\mathscr{D}_{+},\mathscr{D}_{-}$. As a consequence, we can, without confusion, consider $\mathscr{F}$ as the vector space containing the double-spinors solutions of (3.11), and $\mathscr{D}_{\pm}$ as the subspaces relative to the projections $\mathbf{P}_{\pm}$. Indeed, all the relations recalled in Sec. I still hold, and now we can also prove that $\mathscr{D}_{+}$ and $\mathscr{D}_{-}$ are two orthogonal subspaces of $\mathscr{F}$. In fact, if

$$\Psi_{+}(x)=\frac{1}{\sqrt{2}}\begin{bmatrix}\psi(x)\\ \psi(x)\end{bmatrix},\quad \Phi_{-}(x)=\frac{1}{\sqrt{2}}\begin{bmatrix}\phi(x)\\ -\phi(x)\end{bmatrix} \tag{4.12}$$

are two arbitrary double-spinors, belonging, respectively, to $\mathscr{D}_{+}$ and $\mathscr{D}_{-}$, we will have from (3.8) that

$$\begin{aligned}\langle\Psi_{+}\,|\,\Phi_{-}\rangle&=\int_{\sigma}d\sigma^{\mu}\bar{\Psi}_{+}\,\mathbf{C}_{\mu}\Phi_{-}\\ &=\tfrac{1}{2}\int_{\sigma}d\sigma^{\mu}(\bar{\psi}\gamma_{\mu}\phi-\bar{\psi}\gamma_{\mu}\phi)=0\,.\end{aligned} \tag{4.13}$$

In other words, each double-spinor of $\mathscr{F}$ can always be decomposed in two double-spinors, belonging to the orthogonal subspaces $\mathscr{D}_{+}$ and $\mathscr{D}_{-}$. The one-to-one connection between elements of $\mathscr{D}_{+}$ and $\mathscr{D}_{-}$ is here obtained by means of the matrix

$$\mathbf{C}_{5}=\begin{bmatrix}\gamma_5 & 0\\ 0 & -\gamma_5\end{bmatrix}. \tag{4.14}$$

On the ground of these remarks, we can now show that the energy spectrum of a quantum system, obtained in the framework of the second-order equation (1.2) [or, equivalently, of Eq. (3.11)], is always exactly coincident with that obtained from the first-order Dirac equation (1.3). In fact, it is very easy to verify that the relations connecting the elements of $\mathscr{F},\mathscr{D}_{+},\mathscr{D}_{-}$ are obtained by means of operations, like (4.10), (4.11), (4.14), which preserve the form (4.2) of a stationary state, since they consist in linear combinations of components, all multiplied by the same phase factor $e^{-iEt/\hbar}$. The difference is that now, for each given eigenvalue, we must take as the eigenstate in $\mathscr{F}$ both the eigenvectors of $\mathscr{D}_{+}$ and $\mathscr{D}_{-}$. This means that we can always approach an energy eigenvalue problem, by discussing it by means of the Dirac equation (1.3): the spectrum will always be the same for the complete equations (1.2) or (3.11). Then, we can get, from the solutions of $\mathscr{D}_{-}$, the solutions of $\mathscr{D}_{+}$ belonging to the same eigenvalue, by means of the matrix (4.14). The complete set of the eigenstates in $\mathscr{F}$ will be constituted by both the solutions belonging to $\mathscr{D}_{+}$ and $\mathscr{D}_{-}$. It is for that reason, for example, that we get the complete hydrogen atom spectrum also if we work out only the solutions of the Dirac equation.

V. THE FREE PARTICLE

For a free particle Eq. (3.11) can be written as

$$(i\hbar\,\mathbf{C}_{\mu}\partial^{\mu}-mc\,\mathbf{C})\Psi(x)=0\,. \tag{5.1}$$

The plane-wave solutions of this equation have the form

$$\Psi_{\epsilon,\mathbf{p}}(x)=N\,e^{-i\epsilon p\cdot x/\hbar}\Xi\,, \tag{5.2}$$

where N is a normalization factor, Ξ is a constant (in the space-time) double-spinor, and $\epsilon=\pm1$ is the sign of the energy. Indeed, as usual in the relativistic quantum mechanics, the energy can take positive and negative values and, if we use the expression

$$cp_0=E=mc^2\left[1+\left[\frac{\mathbf{p}}{mc}\right]^2\right]^{1/2}>0 \tag{5.3}$$

to represent the absolute value of the energy, we can factorize its sign by means of ϵ.

If we substitute (5.2) in (5.1), we find that the double-spinor Ξ is not completely arbitrary, since it must satisfy the equation

$$(\epsilon\,\mathbf{C}_{\mu}p^{\mu}-mc\,\mathbf{C})\Xi=0\,. \tag{5.4}$$

If we pose

$$\Xi=\frac{1}{\sqrt{2}}\begin{bmatrix}\xi\\ \eta\end{bmatrix}, \tag{5.5}$$

we get from (5.4) that, if $p_{\mu}p^{\mu}=m^2c^2$, ξ is an arbitrary spinor and

$$\eta=\epsilon\frac{\not{p}}{mc}\xi\,, \tag{5.6}$$

so that, the most general form for a plane wave is

$$\Psi_{\epsilon,\mathbf{p}}(x)=\frac{N}{\sqrt{2}}e^{-i\epsilon p\cdot x/\hbar}\begin{bmatrix}\xi\\ \epsilon\dfrac{\not{p}}{mc}\xi\end{bmatrix}. \tag{5.7}$$

In order to obtain the orthonormality relations (in the generalized sense of the δ functions), we consider the following scalar product between double-spinors of the form (5.7):

$$\langle \Psi_{\epsilon,\mathbf{p}} | \Psi_{\epsilon',\mathbf{p}'} \rangle = \int d^3\mathbf{r}\, \bar{\Psi}_{\epsilon,\mathbf{p}}(x) \mathbb{C}_0 \Psi_{\epsilon',\mathbf{p}'}(x)$$

$$= NN'\delta^3(\epsilon\mathbf{p}-\epsilon'\mathbf{p})(2\pi\hbar)^3 e^{-i(\epsilon E-\epsilon' E')t/\hbar} \bar{\xi} \frac{m^2c^2\gamma_0+\epsilon\epsilon' \not{p}\gamma_0 \not{p}'}{2m^2c^2} \xi' , \tag{5.8}$$

that, from the usual properties of the γ matrices, becomes

$$\langle \Psi_{\epsilon,\mathbf{p}} | \Psi_{\epsilon',\mathbf{p}'} \rangle = NN'(2\pi\hbar)^3 \frac{E}{mc^2} \bar{\xi} \frac{\not{p}}{mc} \xi' \delta_{\epsilon\epsilon'} \delta^3(\mathbf{p}-\mathbf{p}') . \tag{5.9}$$

To further calculate, we remark that the spinor ξ can be chosen as the eigenstate of a commutative set of operators. For example, if we fix p_μ, we can consider the matrix

$$G(p) = \frac{\not{p}}{mc} , \tag{5.10}$$

such that $G^2(p)=I$, with eigenvalues $\lambda=\pm 1$. Since the space of the spinors ξ is a four-dimensional, complex, Euclidean space, we need another matrix to remove all the degeneracies. If we take now a vector s_μ, with $s_\mu s^\mu = -1$ and $p_\mu s^\mu = 0$, which plays the role of a spin polarization vector, it is easy to verify that the matrix

$$\Sigma(s) = \gamma_5 \not{s} \tag{5.11}$$

has as eigenvalues $\tau=\pm 1$, since $\Sigma^2(s)=I$, and that

$$[G(p), \Sigma(s)] = 0 . \tag{5.12}$$

Hence we can take, for fixed p_μ and s_μ, as a basis the spinors $\xi_{\lambda\tau}$ ($\lambda,\tau=\pm 1$) that are simultaneous solutions of the following eigenvalue equations:

$$G(p)\xi_{\lambda\tau} = \lambda \xi_{\lambda\tau} , \quad \Sigma(s)\xi_{\lambda\tau} = \tau \xi_{\lambda\tau} . \tag{5.13}$$

It is a nice exercise to show that the explicit form of these spinors is

$$\xi_{\lambda\tau} = \left[\frac{p_0+mc}{2mc} \right]^{1/2} \times \begin{cases} \begin{bmatrix} u_\tau \\ \dfrac{\mathbf{p}\cdot\boldsymbol{\sigma}}{p_0+mc} u_\tau \end{bmatrix} & (\lambda=+1) , \\ \begin{bmatrix} \dfrac{\mathbf{p}\cdot\boldsymbol{\sigma}}{p_0+mc} u_{-\tau} \\ u_{-\tau} \end{bmatrix} & (\lambda=-1) , \end{cases} \tag{5.14}$$

where u_τ is a two-component spinor solution of

$$\mathbf{s}'\cdot\boldsymbol{\sigma} u_\tau = \tau u_\tau \quad (\tau=\pm 1) , \tag{5.15}$$

with

$$\mathbf{s}' = \mathbf{s} - s_0 \frac{\mathbf{p}}{p_0+mc} . \tag{5.16}$$

If we take the orthonormalized solutions of (5.15), namely, if

$$u_\tau^\dagger u_{\tau'} = \delta_{\tau\tau'} , \tag{5.17}$$

we can show that the following relations hold:

$$\bar{\xi}_{\lambda\tau}\xi_{\lambda'\tau'} = \lambda\delta_{\lambda\lambda'}\delta_{\tau\tau'} . \tag{5.18}$$

As a consequence we get also

$$\bar{\xi}_{\lambda\tau} \frac{\not{p}}{mc} \xi_{\lambda'\tau'} = \bar{\xi}_{\lambda\tau} G(p) \xi_{\lambda'\tau'} = \delta_{\lambda\lambda'}\delta_{\tau\tau'} . \tag{5.19}$$

Going back now to the plane wave (5.7), we can easy deduce that the set of double-spinors

$$\Psi_{\lambda,\tau,\epsilon,\mathbf{p}}(x) = \left[\frac{mc^2}{E(2\pi\hbar)^3} \right]^{1/2} e^{-i\epsilon p\cdot x/\hbar} \frac{1}{\sqrt{2}} \begin{bmatrix} \xi_{\lambda\tau} \\ \lambda\epsilon\xi_{\lambda\tau} \end{bmatrix} \tag{5.20}$$

is an orthonormalized set, because from (5.9), (5.19), and (5.20) we get

$$\langle \Psi_{\lambda,\tau,\epsilon,\mathbf{p}} | \Psi_{\lambda',\tau',\epsilon',\mathbf{p}'} \rangle = \delta_{\lambda\lambda'}\delta_{\tau\tau'}\delta_{\epsilon\epsilon'}\delta^3(\mathbf{p}-\mathbf{p}') . \tag{5.21}$$

By comparing (5.20) with (4.8) we obtain also that these plane waves are all elements either of $\mathscr{D}_+$ or of $\mathscr{D}_-$. Of course the most general solution of (5.1) is a wave packet of the form

$$\Psi(x) = \sum_{\lambda,\tau,\epsilon} \int d^3\mathbf{p}\, c_{\lambda,\tau,\epsilon}(\mathbf{p}) \Psi_{\lambda,\tau,\epsilon,\mathbf{p}}(x) , \tag{5.22}$$

and it is not, in general, a solution of the Dirac equation. The wave packet (5.22) will also be normalized if

$$\langle \Psi | \Psi \rangle = \sum_{\lambda,\tau,\epsilon} \int d^3\mathbf{p}\, | c_{\lambda,\tau,\epsilon}(\mathbf{p}) |^2 = 1 . \tag{5.23}$$

VI. CONCLUSIONS

The authors consider, in some sense, astonishing the fact that spin-$\frac{1}{2}$ particles should be described by means of a first-order equation, whereas the spin-0 and the spin-1 fields are usually ruled by second-order equations, which are the most natural quantum analog of the relativistic energy-momentum relations. Of course, they know that the Dirac form of the wave equation is extensively used because of its enormous amount of very good results and predictions. However, they tried to show, in the present paper, that these results and predictions can be preserved also in a theory ruled by the second-order equation (1.2). In fact, in this case, it is perfectly possible to define a coherent statistical interpretation, a Hilbert space of the states and all we need in order to have a quantum mechanics in the ordinary sense. Moreover, it was shown in Sec. IV that all the energy eigenvalue problems have the same spectra in the two approaches and that the energy eigensolutions $\Psi_E(x)$ are doubled by means of the matrix $\mathbb{C}_5$. Despite the fact that these additional eigenstates have the same probability density as the original states, since

$$(\mathbb{C}_5\Psi_E)^\dagger \mathbb{C}_5 \Psi_E = \Psi_E^\dagger \Psi_E , \tag{6.1}$$

this doubling of the eigenstates points out two facts.

(a) It is possible, by means of linear combinations, to find complete orthonormal systems of eigenstates that are

quite different from the Dirac system, and hence it is possible to define complete systems of commuting observables that are not available in the Dirac theory.

(b) The space $\mathscr{F}$ is, is some sense, more symmetric than $\mathscr{D}_{\pm}$: in fact the matrix γ_5 [that does not commute with the Dirac operator (1.3) and that transforms $\mathscr{D}_+$ in $\mathscr{D}_-$ and vice versa] is a symmetry operation in $\mathscr{F}$. It corresponds to the fact that the space of our spinors is no longer a representation space for the Lorentz group only, but it contains also discrete symmetries. This can be considered as an advantage in the sense that, for example, $\mathscr{F}$ contains also subspaces corresponding to the two-component theory of Feynman and Gell-Mann.[2]

However, if a relativistic quantum mechanics is possible for spin-$\frac{1}{2}$ particles, even on the ground of a second-order wave equation, we must also ask how it is possible to extend this program to the bosonic case. In fact, if we try to reproduce our present method for a scalar field ruled by a Klein-Gordon equation, we immediately find some difficulty. If, for example, we start with a generalized Klein-Gordon current density

$$I_\mu(x)=\frac{i\hbar}{2mc}(\phi^*\partial_\mu\psi-\partial_\mu\phi^*\psi)-\frac{e}{mc^2}A_\mu\phi^*\psi\ , \qquad (6.2)$$

we do not find a form for ϕ and ψ such that I_μ is real with $I_0\geq 0$. That is connected with the following remark: the generalized current (2.2) for the spin-$\frac{1}{2}$ case is not immediately analogous to (6.2). Indeed, we have from (2.2) that its Gordon decomposition is

$$J_\mu(x)=\frac{i\hbar}{2mc}(\bar{\phi}\partial_\mu\psi-\partial_\mu\bar{\phi}\psi)-\frac{e}{mc^2}A_\mu\bar{\phi}\psi + \frac{\hbar}{2mc}\partial^\nu(\bar{\phi}\sigma_{\mu\nu}\psi)\ , \qquad (6.3)$$

with a third term which has no correspondent in (6.2). This term, which is separately conserved by symmetry reasons, plays an important role, in the sense that we could not obtain a current with a positive zero component, if we did not take it into account. In fact, if we take $\psi=\not{D}\phi$, the first two terms of (6.3) have no positive zero component. Hence, the problem arises if it is possible to find a more general form of the Klein-Gordon current that allows us to define a positive conserved density. On the other hand, we must remark that the reduction of a second-order equation to a first-order one does not guarantee in itself the possibility of building a conserved positive density, as the Feshbach and Villars approach shows.[4]

However, one thing seems to be clear to us: we would not succeed in defining a positive conserved density for (1.2) without passing through the Dirac formalism with its four-component wave functions, and its decomposition of the second-order operator by means of the γ matrices. Maybe, it means that, for bosonic wave equations also, we should pass in the framework of the linearized wave equations given, for example, in the Kemmer spinorial formalism.[7] The discussion of this problem will be the argument of a forthcoming paper.

---

[1] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).

[2] R. P. Feynman and M. Gell-Mann, Phys. Rev. **109**, 193 (1958).

[3] N. Cufaro Petroni, Ph. Gueret, and J. P. Vigier, Nuovo Cimento **81B**, 243 (1984); Phys. Rev. D **30**, 495 (1984); N. Cufaro Petroni, Ph. Gueret, J. P. Vigier, and A. Kyprianidis, *ibid.* **31**, 3157 (1985).

[4] H. Feshbach and F. Villars, Rev. Mod. Phys. **30**, 24 (1958); V. A. Rizov, H. Sazdjian, and I. T. Todorov, Ann. Phys. (N.Y.) **165**, 59 (1985).

[5] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965).

[6] S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Petersen, Evanston, New York, 1961); O. Costa de Beauregard, *Précis de Mécanique Quantique Relativiste* (Dunod, Paris, 1967).

[7] N. Kemmer, Proc. R. Soc. London **A173**, 91 (1939); N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Wiley, New York, 1959).