Singletons and massless, integral-spin fields on de Sitter space

Christian Fronsdal
Department of Physics, University of California, Los Angeles, California 90024
(Received 20 October 1978)

Gauge-invariant wave equations for massless fields with fixed, but arbitrary, integer spin have been constructed. Extended to include interactions with external sources, the theory remains self-consistent and unitary to lowest order in the coupling. Fields describing states of one or two singletons, with the attendant interesting gauge problems, are studied. An intertwining operator is constructed that explicitly expresses the bilocal two-singleton field in terms of one-particle massless fields.

I. INTRODUCTION

Motivation. The theory of massless fields with higher spins attracts us because it is characterized by interesting and poorly understood gauge problems. One may hope to reach a deeper level of understanding of gauge theories by supplementing the experience gained from the conventional vector-field gauge theories[1] with a study of these unfamiliar phenomena. Lately, it has become fashionable to treat the case of spin-2 massless fields as an instance of Yang-Mills theory,[2] but it is not known whether massless fields of spin 3 or higher can be squeezed into the same framework. Alternatively, one may approach the spin-2 problem in the manner pioneered by Gupta,[3] but the generalized Gupta program advocated by us[4] for dealing in a similar way with higher spins has not yet been carried out. It is possible, and this is a particularly interesting possibility, that massless spin-3 fields can have no interactions without the active participation of massless fields of all spins.

The justification for attacking the problem in the apparently more complicated situation of a curved space-time is that a new interpretation exists that has no analog in flat space. Massless particles in de Sitter space are composite: Each state of a massless particle, with arbitrary spin, may be regarded as a state of two Dirac singletons.[5] Singletons, though represented by scalar fields, are also beset by gauge problems, but these appear to be relatively easy to handle.

The past. Wave equations of the type proposed by Fierz and Pauli,[6] for fields over Minkowski space that describe particles with fixed mass and spin, were obtained by Hagen and Singh.[7] The limiting case of vanishing mass was studied by us.[8] A Fierz-Pauli program for fields over de Sitter space was formulated and carried out by us in the case of spin 1 (Ref. 9) and spin 2 (Ref. 10). In a certain limit, which shall be referred to as the massless case, phenomena appear that are strongly reminiscent of those that characterize massless fields in flat space. In this limit we recovered the correct wave equations for electrodynamics and weak gravitational fields in de Sitter space.

The connection between singletons and massless particles was studied by Flato and Fronsdal.[5] If Rac denotes the integer-spin singleton representation of the de Sitter group discovered by Dirac,[11] then it was proved that the direct product Rac $\otimes$ Rac decomposes into a direct sum over the irreducible representations associated with massless fields. No attempt was made to express massless fields explicitly in terms of bilocal 2-Rac fields, however.

The present. This paper presents, in Sec. II–IV, a derivation of the wave euqations for massless fields with fixed, but arbitrary, integer spin. In Sec. V it is shown that the introduction of interactions with external sources yields a theory that is self-consistent and unitary. In Sec. VI we synthesize all integer spins in preparation for the translation into singleton terms to be carried out later.

Section VII deals with Rac fields and with the novel and interesting gauge problems associated with them. Bilocal fields describe the states of a pair of Racs or anti-Racs. The unwanted Rac-anti-Rac states are eliminated.

In Sec. VIII we construct the intertwining operator that explicitly relates massless fields to bilocal 2-Rac fields. This operator is uniquely defined, up to gauge transformations, on the space of fields that satisfy both wave equations and subsidiary conditions. It is not uniquely defined on the (larger) space of fields subject only to the gauge-invariant wave equations. The ambiguity must be resolved by considerations of utility and aesthetics, but this is still an unsolved problem.

The future. The case of half-integral spins should also be studied. This will bring in the half-integral singletons Di and $\overline{\text{Di}}$ in the combinations Di $\otimes$ Rac, and raises the question whether the combinations Di ⊗ Di must be included in the integer-spin case. The Di and the Rac may be thoroughly integrated by regarding the pair as a supermultiplet. It may be possible to construct a supersymmetric Lagrangian, including interactions, and attempt an interpretation in terms of massless particles. In this case it is rather unlikely that the singleton-antisingleton states can be eliminated. If not, then singletons would assume an existence independent of massless particles. In any case, a theory of interacting singletons will provide an example of interactions between massless fields with higher spins. We emphasize again that this construction has no analog in flat space, though the final result may have a nontrivial flat-space limit.

II. COVARIANT FORMULATION

In this section, h denotes a symmetric tensor field of rank s over Minkowski space or over de Sitter space, with covariant components $h_{\mu}\dots$ with s indices. Indices are raised and lowered by means of the metric tensor g; however, indices will be suppressed or simplified whenever convenient. We denote by h' the trace of h, namely, the symmetric tensor field of rank $s-2$ with components

$$h'_{\lambda}\dots \equiv g^{\mu\nu}h_{\mu\nu\lambda}\dots\,. \tag{2.1}$$

The second trace $h''=(h')'$ is taken to vanish identically.

In flat space, the wave equations for massless fields of integer spin s are of the form[8]

$$Lh=0\,, \tag{2.2}$$

where L is a second-order differential operator,

$$L\equiv BL_0\,, \tag{2.3}$$

with B and L_0 defined by

$$\begin{aligned}(Bh)_{\mu}\dots &= h_{\mu}\dots - \tfrac{1}{2}\textstyle\sum_2 g_{\mu\mu}h'_{\mu}\dots\,,\\ (-L_0h)_{\mu}\dots &= \Box h_{\mu}\dots - \textstyle\sum_1\nabla_{\mu}(\nabla\cdot h)_{\mu}\dots\\ &\quad + \textstyle\sum_2\nabla_{\mu}\nabla_{\mu}h'_{\mu}\dots\end{aligned} \tag{2.4}$$

[Notation: ∇ and $\Box$ stand for the covariant derivative and the covariant d'Alembertian defined by the metric connection and by the metric. Indices are shown simplified and contractions are indicated by a center dot. The sums are over all unequal orderings of the indices; thus $\sum_1$ contains s terms and $\sum_2$ contains $\frac{1}{2}s(s-1)$ terms, since h and g are symmetric.] This expression for L_0 is the only equation in this paper that is restricted to flat space.

Definition 1. A symmetric tensor field of rank s, over Minkowski space or over de Sitter space, is called a *gauge field* if it can be expressed as

$$\tilde{h}_{\mu}\dots = \textstyle\sum_1\nabla_{\mu}\xi_{\mu}\dots\,, \tag{2.5}$$

where ξ is a symmetric tensor field with $\xi'=0$.

Equation (2.2) is "gauge invariant" in the sense that $L_0\tilde{h}$ and *a fortiori* $L\tilde{h}$ vanish identically for every gauge field $\tilde{h}$. This statement holds in flat space only; direct substitution shows that it depends on the commutativity of the covariant derivatives.

de Sitter space is characterized by

$$([\nabla_{\mu},\nabla_{\nu}]h)_{\lambda}\dots = \rho\textstyle\sum_1(g_{\mu\lambda}h_{\nu\lambda}\dots - g_{\nu\lambda}h_{\mu\lambda}\dots)\,, \tag{2.6}$$

where ρ is the curvature parameter. In this case, direct substitution shows that $L_0\tilde{h}$, with L_0 as given above, does not vanish. The simplest way to obtain the correct wave equations in de Sitter space is to modify L_0, by adding compensating terms of order ρ, to recover gauge invariance. This procedure will be justified in Sec. IV. The compensating terms are determined easily and unambiguously, with the following result:

$$\begin{aligned}-L_0h &= \Box h - \textstyle\sum_1\nabla(\nabla\cdot h) + \tfrac{1}{2}\textstyle\sum_1\nabla\textstyle\sum_1\nabla h'\\ &\quad + (s^2-2s-2)\rho h + 2\rho\textstyle\sum_2 gh'\,.\end{aligned} \tag{2.7}$$

Now (2.3) gives

$$\begin{aligned}-Lh &= \Box h - \textstyle\sum_1\nabla(\nabla\cdot h) + \tfrac{1}{2}\textstyle\sum_1\nabla\textstyle\sum_1\nabla h' + (s^2-2s-2)\rho h\\ &\quad + \textstyle\sum_2 g[\nabla\nabla : h - \Box h' - \tfrac{1}{2}\textstyle\sum_1\nabla(\nabla\cdot h') - (s^2-3)\rho h']\,.\end{aligned} \tag{2.8}$$

The wave equations $Lh=0$ are the Euler-Lagrange equations of the Lagrangian

$$\begin{aligned}\mathfrak{L} = \int [&\tfrac{1}{2}g^{\mu\nu}(\nabla_{\mu}h)\cdot(\nabla_{\nu}h) - \tfrac{1}{2}s(\nabla\cdot h)\cdot(\nabla\cdot h) + \tfrac{1}{2}s(s-1)(\nabla h')\cdot(\nabla\cdot h) - \tfrac{1}{4}s(s-1)g^{\mu\nu}(\nabla_{\mu}h')\cdot(\nabla_{\nu}h')\\ &- \tfrac{1}{8}s(s-1)(s-2)(\nabla\cdot h')\cdot(\nabla\cdot h') - \tfrac{1}{2}\rho(s^2-2s-2)h\cdot h + \tfrac{1}{4}s(s-1)(s^2-3)h'\cdot h'](-g)^{1/2}d^4x\,.\end{aligned} \tag{2.9}$$

In the next section we shall introduce more convenient coordinates, defined by the isometric embedding of de Sitter space in $\mathbf{R}^5$. Then in Sec. IV, we shall justify our use of the requirement of gauge invariance to determine the operator L. In Sec. V we begin to study the properties of the wave equation.

III. GLOBAL, ISOMETRIC EMBEDDING

Let V denote the space $\mathbf{R}^5$, endowed with coordinates (y_{α}), $\alpha=0,1,2,3,5$, and the pseudo-

Euclidean metric δ defined by

$$\delta^{\alpha\beta}y_\alpha y_\beta = y_0{}^2 - y_1{}^2 - y_2{}^2 - y_3{}^2 + y_5{}^2 \equiv y^2 .$$

The local, isometric embedding of de Sitter space in V is given by $y^2 = 1/\rho$. To obtain a global embedding we introduce the subspace

$$W \equiv \{ y \in V ; y^2 > 0\} \tag{3.1}$$

and the universal covering space $\overline{W}$ of W. We identify de Sitter space with the image, by the cover map, of the surface $y^2 = 1/\rho$ in W.

There exists a natural bijection between tensor fields over de Sitter space and a certain space of tensor fields over $\overline{W}$. Let U be a domain in W, invariant under the dilations $y_\alpha \to \lambda y_\alpha$, λ in $\mathbf{R}$, and let U_0 denote the intersection of U with the hyperboloid $y^2 = 1/\rho$. Let N be fixed in $\mathbf{C}$, and let k be any tensor field over U that satisfies

$$y^\alpha k_{\ldots\alpha\ldots}(y) = 0, \quad k(\lambda y) = \lambda^N k(y) , \tag{3.2}$$

for λ in $\mathbf{R}$. Let h be the projection of k on U_0, then k is determined by h. One easily extends this bijection to the space of tensor fields over de Sitter space. Locally, the components are related by

$$\begin{aligned} h_{\mu\ldots}(x) &= y_\mu{}^\alpha \cdots k_{\alpha\ldots}(y(x)) , \\ k_{\alpha\ldots}(y) &= (\rho y^2)^{N/2}\, x_\alpha{}^\mu \cdots h_{\mu\ldots}(x(y)) , \end{aligned} \tag{3.3}$$

where $x \to y(x)$ is the embedding map, $y \to x(y)$ the dilatation-invariant projection map, and

$$y_\mu{}^\alpha = \partial y^\alpha / \partial x^\mu , \quad x_\alpha{}^\mu = \partial x^\mu / \partial y^\alpha .$$

Now we can reformulate the results of the preceding section. Let h be a symmetric tensor field over de Sitter space, with $h'' = 0$, and let k be the symmetric tensor field over $\overline{W}$ that is related to h by (3.3). Let the trace k' be defined in analogy with (2.1):

$$k'_{\gamma\ldots} \equiv \delta^{\alpha\beta} k_{\alpha\beta\gamma\ldots} ;$$

then the double trace k'' vanishes. The wave equation (2.2) takes the form

$$Lk = 0, \quad \text{with } L \equiv BL_0 . \tag{3.4}$$

Here we abuse the notation by using the same letters B, L, and L_0 to denote the transforms of these operators by the correspondence (3.3). We find

$$Bk = k - \tfrac{1}{2}\textstyle\sum_2 \theta k' , \tag{3.5}$$

$$\theta_{\alpha\beta} \equiv \delta_{\alpha\beta} - y_\alpha y_\beta / y^2 , \tag{3.6}$$

$$\begin{aligned} -L_0 k = \operatorname{tr}\operatorname{pr}\{ &\partial^2 k - \textstyle\sum_1 \partial(\partial\cdot k) + \textstyle\sum_2 \partial\partial k' \\ &+ (1/y^2)(s - \hat{N} - 2)[(s + \hat{N} + 1)k + \textstyle\sum_2 \delta k']\} . \end{aligned} \tag{3.7}$$

Here tr pr means "the transverse[12] projection of," and

$$\hat{N} \equiv y\cdot\partial \equiv y_\alpha(\partial/\partial y_\alpha) . \tag{3.8}$$

The degree of homogeneity of k is fixed by (3.2), so that

$$(\hat{N} - N)k = 0 . \tag{3.9}$$

This condition is consistent with (3.4).

For the operator $L = BL_0$ we find

$$\begin{aligned} -Lk = \operatorname{tr}\operatorname{pr}\{&\partial^2 k - \textstyle\sum_1 \partial(\partial\cdot k) + \textstyle\sum_2 \partial\partial k' \\ &+ \textstyle\sum_2 \delta[\partial\partial : k - \partial^2 k' - \tfrac{1}{2}\textstyle\sum_1 \partial(\partial\cdot k')] \\ &+ (1/y^2)(s - \hat{N} - 2)(s + \hat{N} + 1)(h - \textstyle\sum_2 \delta h')\} . \end{aligned} \tag{3.10}$$

This operator is formally self-adjoint, and the equations $Lk = 0$ are the Euler-Lagrange equations of a Lagrangian that contains only first-order derivatives.

If $\tilde{h}$ is a gauge field, of the form (2.5), then the corresponding field $\tilde{k}$ on $\overline{W}$ has the form

$$\tilde{k}_{\alpha\ldots} = \textstyle\sum_1 [\partial_\alpha y^2 + (s - \hat{N} - 2)y_\alpha]\zeta_{\alpha\ldots} \tag{3.11}$$

with $y\cdot\zeta = 0$, $\zeta' = 0$, and $\hat{N}\zeta = (N-1)\zeta$.

Definition 2. A symmetric tensor field of rank s over $\overline{W}$ is called a *gauge field* if it can be expressed in the form (3.11), where ζ is a symmetric, traceless, transverse tensor field.

Equation (3.4) is gauge invariant in the sense that $L_0\tilde{k}$ and *a forteriori* $L\tilde{k}$ vanish identically for every gauge field $\tilde{k}$. (3.12)

IV. GROUP-THEORETICAL MEANING OF THE EQUATIONS

The group of motions of de Sitter space is the universal covering of the connected part of SO(3, 2). The group SO(3, 2) is the group of linear transformations of $\mathbf{R}^5$ that leaves the hyperboloid $y^2 = 1/\rho$ invariant; the Lie algebra is generated by a basis $(L_{\alpha\beta})$, $\alpha < \beta = 0, 1, 2, 3, 5$ satisfying standard commutation relations. Any irreducible representation that can be associated with the states of an elementary particle is characterized by an extremal weight $(L_{05}, L_{12}) \to (E_0, s)$. If $E_0 > 0$ $(E_0 < 0)$, then the spectrum of L_{05} is positive definite (negative definite) and we speak of irreducible representations with positive (negative) energies.

The representation $D(E_0, s)$ may be constructed by reduction of the tensor product $D(E_0, 0)\otimes D(s)$, where $D(s)$ is a suitably chosen finite dimensional representation with highest weight $(0, s)$. We limit ourselves to integer s and take $D(s)$ to be the irreducible component with highest weight contained in $\otimes^s D(1)$, where $D(1)$ is the irreducible 5-dimensional representation of SO(3, 2).

Consider a symmetric tensor field of rank s on $\overline{W}$. The action of SO(3, 2) on k is determined by the expressions for the operators $L_{\alpha\beta}$ of the Lie algebra:

$$L_{\alpha\beta}k_{\gamma\ldots} = i(y_\alpha\partial_\beta - y_\beta\partial_\alpha)k_{\gamma\ldots} + i\sum{}_1(\delta_{\alpha\gamma}k_{\beta\gamma\ldots} - \delta_{\beta\gamma}k_{\alpha\gamma\ldots})\,. \quad (4.1)$$

The second-order Casimir operator is

$$Q = \tfrac{1}{2}L^{\alpha\beta}L_{\alpha\beta}\,, \quad (4.2)$$

$$Qk = [\hat{N}(\hat{N}+3) - y^2\partial^2 + s(s+1)]k - 2\sum{}_1 y(\partial\cdot k) + 2\sum{}_1\partial(y\cdot k) - 2\sum{}_2\delta k'\,. \quad (4.3)$$

The case $s=0$ has been investigated already.[13] In this case the other Casimir operator vanishes identically and one expects that the essential step towards the construction of an irreducible representation consists of requiring that Q reduce to a scalar; that is, one imposes the wave equation

$$(Q - \langle Q\rangle)k = 0\,, \quad (4.4)$$

where $\langle Q\rangle$ is the value of Q in $D(E_0, 0)$. The value of Q in $D(E_0, s)$ is

$$\langle Q\rangle = E_0(E_0 - 3) + s(s+1)\,, \quad (4.5)$$

and (4.4) reduces to

$$[(\hat{N}+E_0)(\hat{N}-E_0+3) - y^2\partial^2]k = 0\,. \quad (4.6)$$

The operator $\hat{N}$ is also an invariant, and has to be fixed by imposing (3.9): $(\hat{N}-N)k=0$, the value of N being completely arbitrary. It was found[13] that (4.1) induces, on a space of solutions of (4.6), the unitary, positive-energy representation $D(E_0, 0)$, provided that $E_0 > 1$. [The corresponding negative-energy representation $D(-E_0, 0)$ also appears.]

Let us return now to the general case. When $s \neq 0$ it is not enough to fix the values of Q and $\hat{N}$. If E_0 is large enough, then the representation $D(E_0, s)$ is carried by the subspace of fields that satisfy

$$(Q - \langle Q\rangle)k = 0, \quad (\hat{N}-N)k = 0\,, \qquad \partial\cdot k = 0, \quad y\cdot k = 0\,. \quad (4.7)$$

The last two equations imply that $k' = 0$. The other Casimir operator is fixed by these equations and need not be considered separately.

Another invariant subspace consists of all tensor fields of the form

$$k = \sum{}_1(\partial y^2\zeta + y\eta)\,, \quad (4.8)$$

where ζ and η are symmetric tensor fields of rank $s-1$. The crucial question is now whether these two invariant subspaces have a nontrivial intersection, that is, whether (4.7) admits solutions of the type (4.8). Direct substitution easily yields the following results:

(1) If $E_0 \neq s+1$, then the two subspaces are disjoint. The analysis of (4.7) is a straightforward extension of the special case $s=0$ (Ref. 13) and the result is similar. One finds a space of solutions that carries the irreducible representation $D(E_0, s)$ along with the corresponding negative-energy representation. These are unitary if $E_0 > s+1$; we refer to this case as the "massive" case.

(2) If $E_0 = s+1$, then there exists a space of fields of the form (4.8) that solves (4.7). This is precisely the space of special gauge fields.

Definition 3. A gauge field (definition 2), represented by (3.11), is called *special* if it is conserved in the sense that $\partial\cdot\tilde{k} = 0$.

The special gauge fields form an invariant subspace of the space of solutions of (4.7) and the unitary representation $D(s+1, s)\oplus D(-s-1, s)$ is induced in the quotient space. We refer to this case as the *massless case.* (4.9)

We are now ready to justify our use of general gauge invariance as a means to discover wave equations for massless fields in de Sitter space. We assume, for the sake of this argument, that the equations already obtained for the special case of flat space are the correct ones, and that these equations must be recovered in the limit $\rho\to 0$. In this case the additional requirement of general gauge invariance was seen to imply a unique set of wave equations for each integer spin. We have identified the case $E_0 = s+1$ as the "massless" case because of the appearance, in this case only, of a phenomenon that is a familiar feature of the theory of massless fields in flat space. The results of this section show that any correct set of wave equations for this case must be solved by the special gauge fields; it remains only to justify the strengthening of this requirement by removing the qualifier "special."

The wave equations of electrodynamics in flat space are not unique. Maxwell's equations admit general gauge invariance, but alternative formulations admitting only special gauge invariance are also possible. We believe, nevertheless, that the existence of wave equations with general gauge invariance is crucial to a consistent physical interpretation, though one may prefer not to make direct use of them. In flat space we found that generally gauge-invariant equations exist for arbitrary spin[8,14]; these equations also allowed interactions, self-consistent to lowest order in the coupling to external sources. In this paper we have demonstrated that generally gauge-invariant equations exist in de Sitter space as well, and these are very likely to play the role of Maxwell's equations, generalized to arbitrary spin and to nonvanishing constant curvature. We do not pretend to have proved that our equations are the only correct ones; no *a priori* argument can be complete. The real test is the development of a consistent theory of interacting fields; a beginning will be made in the next section.

Remark 4. We have to verify that every solution of (4.7) actually solves our wave equations (3.4), $Lk=0$, or equivalently $L_0k=0$ with L_0 given by (3.7). That this is indeed the case can be seen by means of the following formulas, obtained from (3.7), (4.3), and (4.5) with $E_0=s+1$ and $y\cdot k=0$:

$$(\langle Q\rangle - Q)k = y^2\partial^2 k + 2\sum{}_1 y(\partial\cdot k) + 2\sum{}_2\delta k' + (s-\hat{N}-2)(s+\hat{N}+1)k\,, \tag{4.10}$$

$$-y^2L_0k = (\langle Q\rangle - Q)k + \sum{}_1[\partial y^2 + (s-\hat{N}-2)y]\zeta\,, \tag{4.11}$$

with

$$\zeta = \tfrac{1}{2}\sum{}_1[\partial y^2 + (s-\hat{N}-5)y](1/y^2)k' - \partial\cdot k\,. \tag{4.12}$$

One checks that $y\cdot\zeta=0$ and $\zeta'=0$.

V. EXTERNAL SOURCES

Equation (3.4), $Lk=0$, with L given by (3.10), is derivable from a real Lagrangian. We now add a term $-t\cdot k$ to the Lagrangian density,[15] determined by a fixed linear functional t that is called the external source. Since $k''=0=y\cdot k$, there is no loss of generality in taking t'' and $y\cdot t$ to vanish; then the Euler-Lagrange equations take the form

$$Lk=t\,. \tag{5.1}$$

The effective interaction, to lowest order in the coupling, is $t\cdot k$, where k is a solution of (5.1). We must study this interaction in order to determine the nature of the long-range correlations of the theory.

Besides the conditions $t''=0=y\cdot t$, self-consistency of the theory requires that the linear functional t vanish on gauge fields

$$t(\tilde{k}) \equiv \int t\cdot\tilde{k}(dy) = 0 \tag{5.2}$$

for every gauge field $\tilde{k}$. Indeed, the fact that $L\tilde{k}$ vanishes identically [see (3.12)] and the fact that L is self-adjoint imply that the transverse, traceless part of $\partial\cdot(Lk)$ vanishes identically. This in turn implies that the transverse, traceless part of $\partial\cdot t$ must vanish, that is,

$$\partial\cdot t = \frac{1}{2(s-1)}\sum{}_2\theta\partial\cdot t' - (1/y^2)\sum{}_1 yt'\,. \tag{5.3}$$

This is equivalent to (5.2) and ensures that the complete Lagrangian is gauge invariant.

Definition 5. A linear functional t will be called an *allowed source* if $t''=0=y\cdot t$ and Eq. (5.3) holds.

Next, it will be shown that the operator L in (5.1) can be replaced by the Casimir operator Q. Let A denote the inverse of the matrix operator B that was defined by (3.5),

$$Ak = k - \frac{1}{2(s-1)}\sum{}_2\theta k'\,, \tag{5.4}$$

and rewrite (5.1) as

$$L_0k = At\,. \tag{5.5}$$

Now Eq. (4.11) shows that

$$y^2L_0k = (Q-\langle Q\rangle)k - \tilde{k}\,,$$

where $\tilde{k}$ is a gauge field, and (5.5) thus becomes

$$(Q-\langle Q\rangle)k = y^2At + \tilde{k}. \tag{5.6}$$

Let $(Q-\langle Q\rangle)^{-1}$ denote an inverse of $Q-\langle Q\rangle$, that is, a propagator incorporating appropriate boundary conditions,[16] and replace (5.6) by the integral equation

$$k = (Q-\langle Q\rangle)^{-1}(y^2At+\tilde{k})\,. \tag{5.7}$$

Now $\tilde{k}$ is a gauge field, and so is $(Q-\langle Q\rangle)^{-1}\tilde{k}$; therefore this term makes no contribution to the effective interaction between allowed sources. In other words, the gauge field $\tilde{k}$ in (5.7) can be chosen to suit our convenience. We shall now attempt to find a choice of $\tilde{k}$ such that $\partial\cdot k$ vanishes; this will make the physical interpretation transparent and allow us to show that the nonlocal part of the interaction is mediated by the states that belong to the unitary representation $D(s+1,s)\oplus D(-s-1,s)$.

Using the representation (3.11) for an arbitrary gauge field, with $y\cdot\zeta=0=\zeta'$, as well as the condition (5.3) satisfied by any allowed source, we find

$$\partial\cdot(At+\tilde{k}) = (\langle Q\rangle - Q)\zeta + y^2\sum{}_1\partial\left(\partial\cdot\zeta - \frac{1}{2(s-1)}t'\right) + \cdots\,,$$

where the unwritten terms vanish on t by virtue of $y\cdot t=0$. We choose our gauge field $\tilde{k}$ by taking for ζ any solution of

$$\partial\cdot\zeta = \frac{1}{2(s-1)}t'\,,\quad y\cdot\zeta=0\,; \tag{5.8}$$

then $\zeta'=0$, and the fields k in (5.7) satisfy

$$\partial \cdot k = -\zeta \,. \tag{5.9}$$

To interpret this result let $t = t_1 + t_2$, where t_1 and t_2 have supports in disjoint regions U_1 and U_2 of de Sitter space.[15] To fix the ideas let us regard t_1 as the emitter and t_2 as the absorber[16]; then t_1 is the source for the field $k = (Q - \langle Q\rangle)^{-1}(y^2 A t_1 + \tilde{k}_1)$ and this field is detected by the value $t_2(k)$ of the linear functional t_2 at k. Now $t_2(k) = \int t_2 \cdot k(dy)$ depends on the values of k in the region $U_2 = \mathrm{supp}(t_2)$ only. We choose the solution ζ of (5.8) so that ζ vanishes in U_2; then (5.9) shows that $\partial \cdot k$ vanishes in U_2. In other words, the field $k_2 \equiv k|U_2$ that is measured by $t_2(k)$ satisfies $\partial \cdot k_2 = y \cdot k_2 = 0$. This means that the observation of freely propagating fields is given by a gauge-invariant linear functional on the space of solutions of the set (4.7), that is, by a linear functional on the quotient space described in (4.9). This quotient space carries the unitary representation $D(s+1, s) \oplus D(-s-1, s)$. Thus we conclude that this representation describes the propagating quanta and that the theory is unitary to the lowest nontrivial order in the coupling.

This concludes our investigation of the physical interpretation of the wave equations (3.4)–(3.10). The next section is a reformulation that prepares the way for the interpretation of massless particles as composite states of Dirac singletons.

VI. SYNTHESIS OF ALL INTEGER SPINS

Fields that describe composite states are expected to be fields over $\overline{W} \times \overline{W}$, or perhaps $\overline{W} \times I$, where I is some "internal space." The internal degree of freedom corresponds to the spin. We must replace the discrete spin label by some new set of variables that will facilitate a bilocal field interpretation.

To this end let (z_α), $\alpha = 0, 1, 3, 5$, be a set of indeterminate quantities and let $K(y, z)$ denote the formal series

$$K(y, z) = \sum_s z^{\alpha_1} \cdots z^{\alpha_s} k_{\alpha_1 \cdots \alpha_s}(y) \,. \tag{6.1}$$

The complete set of wave equations and subsidiary conditions for all integer spins can now be reexpressed in terms of the "field" K. Thus

$$\begin{aligned} &y \cdot k = 0 \iff y \cdot \partial_z K = 0 \text{ (transversality)}, \\ &\partial \cdot k = 0 \iff \partial_y \cdot \partial_z K = 0 \text{ (conservation)}, \\ &k' = 0 \iff \partial_z^2 K = 0 \text{ (tracelessness)}. \end{aligned} \tag{6.2}$$

If each of the fields k (for $s = 1, 2, \ldots$) is a gauge field, then we call (6.1) a gauge field. In other words (see definition 2) we have the following definition:

Definition 6. A formal series (6.1) will be called a *gauge field* if it can be expressed in the form

$$\tilde{K} = [(z \cdot \partial_y) y^2 + (\hat{n} - \hat{N} - 2)(z \cdot y)]\Lambda \,, \tag{6.3}$$

where Λ is a formal series satisfying $y \cdot \partial_z \Lambda = 0 = \partial_z^2 \Lambda$ and where

$$\hat{n} \equiv z \cdot \partial_z, \quad \hat{N} \equiv y \cdot \partial_y \,. \tag{6.4}$$

The gauge field (6.3) is *special* if in addition it is conserved in the sense that $\partial_y \cdot \partial_z \tilde{K} = 0$.

The Casimir operator $\langle Q\rangle - Q$, given by (4.10), takes the form

$$\begin{aligned} (\langle Q\rangle - Q)K = [&y^2 \partial_y^2 + 2y \cdot z \partial_y \cdot \partial_z + z^2 \partial_z^2 \\ &+ (\hat{n} - \hat{N} - 2)(\hat{n} + \hat{N} + 1)]K \,. \end{aligned} \tag{6.5}$$

There is simplification to be gained by taking the degree N of homogeneity of the spin-s field equal to $-s-1$, for in that case

$$(\hat{n} + \hat{N} + 1)K = 0 \,. \tag{6.6}$$

To conclude: The representation $D(s+1, s) \oplus D(-s-1, s)$ was induced by (4.1) on the space of solutions of (4.7), modulo the space of gauge solutions. Therefore, the representation $\oplus_s [D(s+1, s) \oplus D(-s-1, s)]$ is realized on the space of solutions of

$$\begin{aligned} &(\langle Q\rangle - Q)K = 0, \quad (\hat{n} + \hat{N} + 1)K = 0 \,, \\ &\partial_y \cdot \partial_z K = 0, \quad y \cdot \partial_z K = 0 \,, \end{aligned} \tag{6.7}$$

modulo the space of gauge solutions.

VII. RAC FIELDS AND COMPOSITES

The Rac is the unitary, irreducible representation $D(1/2, 0)$; it is one of the four remarkable representations discovered by Dirac.[11] The anti-Rac or $\overline{\text{Rac}}$ is the corresponding negative-energy representation $D(-1/2, 0)$. These representations are related to the "massless" representations $D(s+1, s)$ by the formulas[5]

$$\begin{aligned} &\text{Rac} \otimes \text{Rac} = \bigoplus_s D(s+1, s) \,, \\ &\overline{\text{Rac}} \otimes \overline{\text{Rac}} = \bigoplus_s D(-s-1, s) \,, \end{aligned} \tag{7.1}$$

which tell us that the space of states of 2 Racs can be identified with the space of states of one massless boson.

The Rac field ϕ on $\overline{W}$ is a scalar field with a fixed but arbitrary degree of homogeneity—that satisfies the wave equation

$$(Q + \tfrac{5}{4})\phi = 0 \tag{7.2}$$

and the boundary condition $(r^2 \equiv y_1^2 + y_2^2 + y_3^2)$

$$\lim_{r \to \infty, y^2 \text{ fixed}} r^{1/2} \phi(y) < \infty \,. \tag{7.3}$$

The solutions for which this limit vanishes form an invariant subspace and the representation $\text{Rac} \oplus \overline{\text{Rac}}$ is realized on the quotient space. The Rac states are thus represented, not by fields, but by equivalence classes of fields. This is a novel and striking example of a field theory with gauge problems. We shall find a convenient method to deal with those problems.

Let us take the degree of homogeneity to be $-\frac{1}{2}$; then the field satisfies

$$(\hat{N}+\tfrac{1}{2})\phi=0, \quad \partial_y{}^2\phi=0. \tag{7.4}$$

In this case the boundary condition (7.3) ensures the existence of the limit

$$\psi(a)\equiv\lim_{\lambda\to 1}\phi(\lambda a_0,\vec{a},\lambda a_5) \tag{7.5}$$

for $a^2=0$, $a\neq 0$, which allows the definition of ϕ to be extended by continuity to the universal covering $\overline{C}$ of the cone—compare (3.1)—

$$C\equiv\{y\in V; y^2=0, y\neq 0\}. \tag{7.6}$$

The limit (7.5) defines a mapping from the space of fields on $\overline{W}$ satisfying the boundary conditions (7.3) to a space of fields on $\overline{C}$. The kernel of this map consists of all the fields for which the limit (7.5) vanishes. The Rac states are represented by the restriction to $\overline{C}$ of fields on $\overline{W}\cup\overline{C}$ that satisfy (7.4).[17]

The unitary representations Rac and $\overline{\text{Rac}}$ are defined by (i) the formula $L_{\alpha\beta}\phi=i(y_\alpha\partial_\beta-y_\beta\partial_\alpha)\phi$ for the generators and (ii) specification of a dense, invariant set of analytic vectors for each representation. For the Rac we shall use the space S_+ of analytic vectors that consists of the finite linear combinations of the energy-angular-momentum eigenstates[13]

$$\phi^+_{LM}(y)=(y_+)^{-L-1/2}\mathcal{Y}_{LM}(\vec{y}); \tag{7.7}$$

for the $\overline{\text{Rac}}$ we use the space S_- similarly constructed from

$$\phi^-_{LM}(y)=(y_-)^{-L-1/2}\mathcal{Y}_{LM}(\vec{y}). \tag{7.8}$$

Here $y_\pm=y_0\pm iy_5$ and the $\mathcal{Y}_{LM}$ are the solid spherical harmonics in the variables $\vec{y}=(y_1,y_2,y_3)$.

Remark 7. The eigenfunction $\phi^\pm_{LM}$ is unique up to additional terms that vanish on $\overline{C}$, so $\phi^\pm_{LM}|\overline{C}$ is unique. The choice (7.7) therefore involves a special choice of "gauge."

The 2-Rac field Φ is a scalar field on $(\overline{W}\cup\overline{C})\times(\overline{W}\cup\overline{C})$ or, what we take to mean the same thing, a bilocal, scalar field on $\overline{W}\cup\overline{C}$. The representation $(\text{Rac}\oplus\overline{\text{Rac}})\otimes(\text{Rac}\oplus\overline{\text{Rac}})$ is induced on the restriction to $\overline{C}\times\overline{C}$ of solutions of the system—compare (7.4)—

$$\begin{aligned}&(u\cdot\partial_u+\tfrac{1}{2})\Phi=0=(v\cdot\partial_v+\tfrac{1}{2})\Phi,\\&\partial_u{}^2\Phi=0=\partial_v{}^2\Phi.\end{aligned} \tag{7.9}$$

Here (u_α, v_α) are coordinates for $(\overline{W}\cup\overline{C})\times(\overline{W}\cup\overline{C})$. This representation is equivalent to the direct sum of

$$(\text{Rac}\otimes\text{Rac})\oplus(\overline{\text{Rac}}\otimes\overline{\text{Rac}}) \tag{7.10}$$

and two copies of $\text{Rac}\otimes\overline{\text{Rac}}$. According to (7.1), it is the part (7.10) that describes massless particles, while $\text{Rac}\otimes\overline{\text{Rac}}$ has no such interpretation. We choose a space of analytic vectors that is adapted to the separation of (7.10) from the rest.

A dense set S_{++} of analytic vectors for $\text{Rac}\otimes\text{Rac}$ is given by the set of finite linear combinations of vectors obtained from the set

$$\phi^+_{00}(u)\phi^+_{LM}(v) \tag{7.11}$$

by repeated application of the generators $L_{\alpha\beta}$. Similar spaces S_{--} for $\overline{\text{Rac}}\otimes\overline{\text{Rac}}$, S_{+-} for $\text{Rac}\otimes\overline{\text{Rac}}$, and S_{-+} for $\overline{\text{Rac}}\otimes\text{Rac}$ are defined in a self-evident manner. These choices of analytic vectors are not unique—see remark 7—but they permit a neat separation of (7.10) from the rest. Indeed, it is evident that all bilocal fields of the form (7.11), and therefore all bilocal fields in the space $S_{++}\oplus S_{--}$ that carries (7.10), satisfy the equation

$$\partial_u\cdot\partial_v\Phi=0. \tag{7.12}$$

This equation is not satisfied by the fields $\phi^+_{00}(u)\phi^-_{LM}(v)$ that generate S_{+-}. By a close examination of the eigenstates of energy and angular momentum in $\text{Rac}\otimes\overline{\text{Rac}}$ it may be proved that, in fact, the subspace of the space of solutions of (7.9), characterized by the additional condition (7.12), carries no irreducible component of $\text{Rac}\otimes\overline{\text{Rac}}$. Thus we have the following:

Proposition 8. The space of scalar fields on $(\overline{W}\cup\overline{C})\times(\overline{W}\cup\overline{C})$ that satisfy

$$\begin{aligned}&(u\cdot\partial_u+\tfrac{1}{2})\Phi=0, \quad (v\cdot\partial_v+\tfrac{1}{2})\Phi=0,\\&\partial_u{}^2\Phi=\partial_v{}^2\Phi=\partial_u\cdot\partial_v\Phi=0,\end{aligned} \tag{7.13}$$

restricted to $\overline{C}\times\overline{C}$, carries the unitary representation $(\text{Rac}\otimes\text{Rac})\oplus(\overline{\text{Rac}}\otimes\overline{\text{Rac}})$.

Remark 9. Of course, (7.13) is not gauge invariant, in the sense that if Φ_1 satisfies (7.13) and $(\Phi_1-\Phi_2)|\overline{C}\times\overline{C}=0$ then Φ_2 does not always satisfy (7.13). This reflects the special choice of gauge implied by our definition of S_{++} in terms of (7.11). Compare remark 7.

In anticipation of an intimate relationship between the gauge freedom encountered here and the gauge fields associated with massless particles, we make the following definition.

Definition 10. A bilocal field that satisfies (7.13) is called a *special, bilocal gauge field* if it has the form

$$\tilde{\Phi}=u^2\Psi_1+v^2\Psi_2, \tag{7.14}$$

where Ψ_1 and Ψ_2 are bilocal fields on $\overline{W}\cup\overline{C}$.

VIII. THE INTERTWINING OPERATOR

We shall construct an operator F: $K \to \Phi$ that intertwines between the representation (7.10), realized on bilocal fields satisfying (7.13), and the representation $\oplus_s [D(s+1, s) \oplus D(-s-1, s)]$, realized on a formal series that satisfies (6.7). The final goal is to express our gauge-invariant wave equations in terms of the 2-Rac, bilocal field Φ.

The internal variables (z_α) were introduced in Sec. VI as indeterminate quantities; it is now natural to interpret them as coordinates for an "internal space" I, so that the formal series (6.1) can be regarded as a field over $\overline{W} \times I$, and to express the intertwining operator with the help of a mapping between $\overline{W} \times I$ and $\overline{W} \times \overline{W}$. We identify I with a pseudo-Euclidean space V (Sec. III). The physical picture suggests that the position coordinate y of the massless particle be related symmetrically to the constituent coordinates u and v, and the internal coordinate z to the separation between the constituents; thus

$$y = u + v, \quad z = u - v. \tag{8.1}$$

This does not map $\overline{W} \times \overline{W}$ on $\overline{W} \times I$ globally, but we do not need a global mapping and simply restrict ourselves to a domain in which $(u+v)^2 > 0$, $u^2 > 0$, $v^2 > 0$.

Consider the lowest-energy eigenspace associated with the component $D(s+1, s)$. We write down the corresponding bilocal fields Φ_s, in a gauge compatible with (7.13), and the corresponding series K_s, in a gauge compatible with (6.7):

$$\Phi_s(u, v) = (u_+ v_+)^{-s-1/2} M_s(u_+ \vec{v} - v_+ \vec{u}), \tag{8.2}$$

$$K_s(y, z) = (y_+)^{-2s-1} M_s(z_+ \vec{y} - y_+ \vec{z}), \tag{8.3}$$

where M_s is an s-linear, traceless form:

$$M_s(\vec{\alpha}) = M^{i_1 \cdots i_s} \alpha_{i_1} \cdots \alpha_{i_s} \quad (i_1, \ldots, i_s = 1, 2, 3). \tag{8.4}$$

To relate (8.2) and (8.3) we use

$$(4u_+ v_+)^{-s-1/2} = \sum_k \binom{-s-\frac{1}{2}}{k} (-1)^k (z_+/y_+)^{2k} y_+^{-2s-1}. \tag{8.5}$$

Using the fact that $u \circ \partial_v M_s(u_+ \vec{v} - v_+ \vec{u}) = 0$ one finds that $\Phi_s(u, v)$ can be written, in the domain of convergence of (8.5), as $(F_s K)(y, z)$ with

$$F_s = \sum_k (4^k k!)^{-1} (z \cdot \partial_y)^{2k} s!/(s+k)! \,. \tag{8.6}$$

Let us now consider the formula

$$\Phi(u, v) = (FK)(y, z), \tag{8.7}$$

$$F = \sum_k (4^k k!)^{-1} (z \cdot \partial_y)^{2k} / (\hat{n}+1)(\hat{n}+2) \cdots (\hat{n}+k). \tag{8.8}$$

When $K = K_s$ the operator $\hat{n}$ reduces to multiplication by s and (8.7) expresses Φ_s in terms of K_s, in some domain consistent with (8.1). Thus (8.7) gives us Φ_s in some domain of $\overline{W} \times \overline{W}$ with local coordinates (u_α, v_α) in which $u^2 > 0$, $v^2 > 0$, and $(u+v)^2 > 0$, and in all of $\overline{W} \times \overline{W}$ by analytic continuation. Furthermore, the mapping (8.7) and (8.8) evidently extends to a dense set of analytic vectors for the representations $\oplus_s [D(s+1, s) \oplus D(-s-1, s)]$ on the one hand and for the representation (Rac $\otimes$ Rac) $\oplus$ ($\overline{\text{Rac}} \otimes \overline{\text{Rac}}$) on the other hand.

The domain of applicability of (8.7) and (8.8) can be extended by expressing F as an integral operator, for example, by making use of Fourier and Mellin transforms, but our present purpose does not require it.

IX. PROPERTIES OF THE INTERTWINING OPERATOR

Let K be any formal series of the form (6.1), with coefficients that are symmetric tensor fields without any *a priori* subsidiary conditions. For the moment we impose neither transversality, homogeneity, or tracelessness, but suppose only that the series (6.1) converges for all z and that FK exists. This is certainly the case if (6.1) has a finite number of nonvanishing terms and each tensor field satisfies some mild regularity conditions. In this general context we seek the equations for Φ that characterize the image by F of the subspaces defined by the subsidiary conditions and by the wave equations.

To begin with we note the identity

$$F z_\alpha y \cdot \partial_z = (u - v)_\alpha (u \cdot \partial_u - v \cdot \partial_v) F. \tag{9.1}$$

This shows that our *a priori* condition of transversality, $y \cdot k = 0$ or $y \cdot \partial_z K = 0$, together with the conventional choice (6.6) of the degree of homogeneity, is equivalent to the conventional choice of the degree of homogeneity of Rac fields:

$$\left.\begin{aligned} y \cdot \partial_z K &= 0 \\ (\hat{n} + \hat{N} + 1) K &= 0 \end{aligned}\right\} \Longleftrightarrow \begin{cases} (u \cdot \partial_u + \frac{1}{2}) \Phi = 0, \\ (v \cdot \partial_v + \frac{1}{2}) \Phi = 0. \end{cases} \tag{9.2}$$

From now on we take these conditions for granted.

Next, we record the two expressions for the wave operator $\langle Q \rangle - Q$. From (6.5) and (6.6)

$$(\langle Q \rangle - Q) K = (y^2 \partial_y^{\,2} + 2 y \cdot z \partial_y \cdot \partial_z + z^2 \partial_z^{\,2}) K \tag{9.3}$$

from which we find, using (8.8),

$$F(\langle Q \rangle - Q) K = (u^2 \partial_u^{\,2} + 2u \cdot v \partial_u \cdot \partial_v + v^2 \partial_v^{\,2}) \Phi. \tag{9.4}$$

The subsidiary conditions $\partial \cdot k = 0$ do not seem to be expressible in a simple way in terms of Φ, but on the solution space of the equation $(\langle Q\rangle - Q)K=0$ we find a simple result:

$$\left.\begin{matrix}(\langle Q\rangle - Q)K=0\\ \partial_y\cdot\partial_z K=0\end{matrix}\right\} \Longleftrightarrow \left\{\begin{matrix}\partial_u{}^2\Phi=0=\partial_v{}^2\Phi\,,\\ \partial_u\cdot\partial_v\Phi=0\,.\end{matrix}\right. \tag{9.5}$$

Proposition 11. The operator F defined by (8.7) and (8.8) maps the space of solutions of (6.7) into the space of solutions of (7.13). The mapping is bijective modulo special gauge fields.

Recall—definition 6—that if $\tilde{K}$ is a special gauge field, then it has the form

$$\tilde{K}=[(z\cdot\partial_y)y^2+(2\hat{n}-1)z\cdot y]\Lambda\,, \tag{9.6}$$

with $y\cdot\partial_z\Lambda=\partial_y\cdot\partial_z\Lambda=\partial_y{}^2\Lambda=0$. We expect that $\tilde{\Phi}=F\tilde{K}$ has the form of a special, bilocal gauge field—definition 10—and find that indeed

$$\tilde{\Phi}=u^2\Psi_1+v^2\Psi_2 \tag{9.7}$$

with rather special factors

$$\Psi_1=v\cdot\partial_u\Psi,\quad \Psi_2=-u\cdot\partial_v\Psi\,, \tag{9.8}$$

$$\Psi\equiv\left[\sum_k(4^k k!)^{-1}(z\cdot\partial_y)^{2k}\,/\,(\hat{n}+2)\cdots(\hat{n}+k+1)\right]\Lambda\,. \tag{9.9}$$

The conditions on Λ enumerated after (9.6) imply that $\partial_u{}^2\Psi=\partial_v{}^2\Psi=\partial_u\cdot\partial_v\Psi=0$. These conditions on Ψ_1 and Ψ_2 are sufficient conditions for (9.7) to be a special, bilocal gauge fields, but they are probably not necessary. Thus F maps the special gauge fields into a (probably proper) subset of the special, bilocal gauge fields.

ACKNOWLEDGMENTS

I am grateful to Professor M. Flato for many stimulating discussions, and to J. Fang for checking some of the calculations. I thank Moshe Flato for his hospitality in Paris where this work was carried out. This work was supported by the National Science Foundation.

[1]C. N. Yang and R. Mills, Phys. Rev. 96, 191 (1954); E. S. Abers and B. W. Lee, Phys. Rep. 9C, 1 (1973).
[2]R. Utiyama, Phys. Rev. 101, 1597 (1956).
[3]S. N. Gupta, Phys. Rev. 96, 1683 (1954).
[4]J. Fang and C. Fronsdal, J. Math. Phys. (to be published).
[5]M. Flato and C. Fronsdal, Lett. Math. Phys. 4, 1978.
[6]M. Fierz and W. Pauli, Proc. R. Soc. London A173, 211 (1939).
[7]C. R. Hagen and L. P. S. Singh, Phys. Rev. D 9, 898 (1974).
[8]C. Fronsdal, Phys. Rev. D 18, 3624 (1978).
[9]C. Fronsdal, Phys. Rev. D 12, 3819 (1975).
[10]J. Fang and C. Fronsdal, Lett. Math. Phys. 4, 1978.
[11]P. A. M. Dirac, J. Math. Phys. 4, 901 (1963).
[12]A tensor field k is "transverse" if $y\cdot k=0$.
[13]C. Fronsdal, Phys. Rev. D 10, 589 (1974).
[14]J. Fang and C. Fronsdal, Phys. Rev. D 18, 3630 (1978).
[15]Physical interpretation of the theory is based on the assumption that sources t can be prepared which are localized in space-time.
[16]The causal structure on de Sitter space was discovered by L. Castell, Nucl. Phys. B4, 343 (1967), and described in Ref. 13.
[17]The cone C may be identified with the "cone at infinity" of R. Penrose, Proc. R. Soc. London A284, 159 (1965).