UNITARITY IN THE $N\theta\theta$ SECTOR
OF SOLUBLE MODEL WITH INDEFINITE METRIC

T. D. LEE *
CERN, Geneva, Switzerland

and

G. C. WICK
Columbia University, New York, N.Y. **

Received 17 February 1969

Abstract: Assuming that the $N\theta$ sector of a certain soluble model with an indefinite metric contains no stable $V$ state, the unitarity condition of the $S$-matrix is explicitly verified in the $N\theta\theta$ sector.

1. INTRODUCTION

In a preceding paper [1], quoted as I in the following, we have promised a detailed study of the $N\theta\theta$ sector of the soluble model [2], by means of the explicit type of solution discovered by Amado [3] and investigated by many other authors [4] for other cases of the soluble model ***. The presence, in our case, of two complex roots of the eigenvalue condition in the $N\theta$ sector requires certain modifications, but the method goes through all the same. Although the unitarity of the $S$-matrix has already been inferred from general considerations [1], it seems of some interest to verify it directly in a case which has been shown to give trouble [5] for different values of the parameters, i.e. when a stable $V$ and a 'ghost' exist†.

In order to avoid unnecessary repetitions, we shall refer as much as is practical to the earlier work of Pagnamenta [4]; the Hamiltonian and nota-

* On leave from Columbia University.
** This research was supported in part by the U.S. Atomic Energy Commission and the National Science Foundation.
*** The list of papers in ref. [4] could be much longer. See the last paper in the reference for further literature. Pagnamenta's two papers of ref. [4] will be quoted as AP1 and AP2 respectively.
† We recall that a 'ghost' state is defined to be an eigenstate of the total Hamiltonian with a real eigenvalue, but of a negative, or zero, square length. Our basic requirement, as given by eq. (3) below, that there should be no stable $V$ excludes the existence of such a 'ghost' state, in contrast to the case considered by Källen and Pauli [5].
tions are taken over from eqs. (2.2) - (2.8) of our paper [1]. We go immediately to the limit $R = \infty$; see eqs. (2.32) and (2.33) of I. For a quicker comparison with Pagamenta's formulae, we shall designate the function $h_\infty$ in I, eq. (2.14), as $H(z)$ (z a complex value of the energy parameter $E$) and its limiting value on the upper rim of the cut on the real axis as $H^+(\omega) = H(\omega + i0)$. Similarly $H^-(\omega) = H(\omega - i0)$; unlike Pagamenta we attach the $\pm$ designation to the function, not to some specific variable (a confusing habit, in our opinion). Thus we have:

$$H(z) = -m_V^0 + z - \frac{1}{\pi} \int_\mu^\infty \frac{\sigma(\omega)}{\omega - z} \, d\omega,$$

(1)

where

$$\sigma(\omega) = + (4\pi)^{-1} g^2 k u_k^2 ; \quad k = (\omega^2 - \mu^2)^{\frac{1}{2}},$$

(compare eqs. (2.14) and (2.17) of I). Of course,

$$\text{Im} H^+(\omega) = -\sigma(\omega), \quad (\omega > \mu).$$

(2)

We emphasize that, in spite of superficial differences due to renormalizations and subtractions, adopted by Pagamenta and other authors, the function $H(z)$ defined by the unsubtracted relation (1) corresponds exactly to $H(z)$ in AP1, 2.

As pointed out in I, one can easily see that if $m_V^0$ is such that on the uncut part of the real axis,

$$-\infty < \omega < \mu , \quad H(\omega) < 0 ,$$

(3)

then the equation $H(z) = 0$ has a complex root $E_+$ in the upper half-plane, eq. (2.18) of I, and a root $E_- = E_+^* = E_-$ in the lower half-plane. We shall need, in the following, a formula analogous to eq. (14) of AP1 for the reciprocal function $[H(z)]^{-1}$. In our case this function has simple complex poles at $z = E_\pm$, and is $O(z^{-1})$ at $\infty$. It is easy to establish, therefore, the Cauchy formula:

$$[H(z)]^{-1} = (z - E_+)^{-1} \Gamma_+ + (z - E_-)^{-1} \Gamma_- + \frac{1}{\pi} \int_\mu^\infty \frac{\text{Im}[H^+(\omega)]^{-1} \, d\omega}{\omega - z},$$

(4)

where

$$\Gamma_\pm^{-1} = H'(E_\pm).$$

Of course

$$\text{Im}[H^+(\omega)]^{-1} = \sigma(\omega)/|H^+(\omega)|^2.$$

We now go over to the scattering problem.
2. THE Nøø SECTOR

Consider a state consisting of an incident wave $N^\dagger A_p^\dagger A_q^\dagger|\text{vac}\rangle$ plus additional terms which in the asymptotic region reduce to outgoing waves only. In the notation of I, this state is designated by $|N_{\theta_p}^{\text{out}}\rangle$. In view of I, eq. (2.11), the energy of this state is

$$E = \omega_p + \omega_q .$$  \hspace{1cm} (5)

In the following, we use the abbreviations $\omega$, $\omega'$, $\omega_1$, $\omega_2$ for $\omega_k$, $\omega_k'$, $\omega k_1$, $\omega k_2$. The corresponding values of the form-factor $u_k$, etc. will be abbreviated likewise as $u$, $u'$, $u_1$, $u_2$.

The state vector in question may be expanded:

$$|N_{\theta_p}^{\text{out}}\rangle = \int \int \int \int \omega^2_1 \omega^2_2 u\alpha(\omega) A^\dagger_k V^\dagger |\text{vac}\rangle$$

$$+ \int \int \int \omega_2 \phi(\omega_1, \omega_2) A^\dagger_{k_1} A^\dagger_{k_2} N^\dagger |\text{vac}\rangle .$$  \hspace{1cm} (6)

From the Schrödinger equation one obtains two relations between $\phi$ and $\alpha$:

$$(E - \omega_1 - \omega_2)\phi(\omega_1, \omega_2) = -(g/4\pi)u_1 u_2(\omega_1 \omega_2)^{1/2}[\alpha(\omega_1) + \alpha(\omega_2)] , $$  \hspace{1cm} (7)

and

$$(E - m_\nu^0 - \omega)^{1/2} u\alpha(\omega) = (g/2\pi) \int \int \int \omega^{1/2} k' u'[\phi(\omega, \omega') + \phi(\omega', \omega)]d\omega' . .$$  \hspace{1cm} (8)

The first relation may also be written as

$$\phi(\omega_1, \omega_2) = \delta(\omega_1 - \omega_p)\delta(\omega_2 - \omega_q)$$

$$- (g/4\pi)u_1 u_2(\omega_1 \omega_2)^{1/2}(E + i0 - \omega_1 - \omega_2)^{-1}[\alpha(\omega_1) + \alpha(\omega_2)] . $$  \hspace{1cm} (9)

Here the 'inhomogeneous' term on the right represents the incident wave, while the energy denominator in the second term is in the form corresponding to the assumed asymptotic behavior of the state (6). The second relation after eliminating $\phi$ by means of eq. (9) takes the form of an integral equation for $\alpha(\omega)$, which after some manipulations (compare, e.g., AP1, eqs. (58), (59), (60) and (II)) can be reduced to the following form. First introduce a new unknown function $\chi^{-}(\omega)$ through the substitution:

$$\frac{1}{2}g u_p u_q(\omega_p \omega_q)^{1/2} H^+(\omega_p)H^+(\omega_q)\alpha(\omega)$$

$$= \{\sigma(\omega_p)H^+(\omega_q)\delta(\omega - \omega_q) + (p = q)\} - \frac{1}{\pi} \sigma(\omega_p)\sigma(\omega_q)\chi^{-}(\omega) , $$  \hspace{1cm} (10)

then $\chi^{-}$ is a solution of the integral equation.
\[ H^+(E - \omega)\chi^-(\omega) = \frac{1}{\pi} \int_{\mu}^{\infty} \sigma(\omega')\chi^-(\omega')d\omega' + \frac{H^+(\omega_q)}{\omega_p + i0 - \omega} + \frac{H^+(\omega_p)}{\omega_q + i0 - \omega}. \]  

(11)

This is to be regarded as an integral equation for a function \( \chi^- \) of a real variable \( \omega \), over the interval

\[ \mu < \omega < \infty. \]  

(12)

If \( \chi^-(\omega') \) in the integrand on the right is regarded as known, however, one can regard the right-hand side of eq. (11) as the boundary value for \( \chi^- \) of a function of a complex variable \( z \). Since this is also true of the factor \( H^+(E - \omega) \) on the left, we can set

\[ \chi^-(\omega) = \chi(\omega - i0), \]  

(13)

where \( \chi(z) \) is defined by

\[ H(E - z)\chi(z) = \frac{1}{\pi} \int_{\mu}^{\infty} \sigma(\omega')\chi^-(\omega')d\omega' + \frac{H^+(\omega_q)}{\omega_p - z} + \frac{H^+(\omega_p)}{\omega_q - z}. \]  

(14)

One sees that \( \chi(z) \) is meromorphic in the whole \( z \)-plane, except for a cut on the real axis, \( \omega = E \) (real) such that

\[ -\infty < \omega < E - \mu. \]  

(15)

Notice that since \( E = \omega_p + \omega_q > 2\mu \), the cut (15) overlaps the integration interval (12) which is also the cut for the function \( H(z) \). All these circumstances have been discussed in detail by Pagnamenta, Sommerfield, etc.; therefore we summarize. One difference here is that the factor \( H(E - z) \) on the left of eq. (14) has zeroes at

\[ z = z_{\pm} = E \pm E_{\pm}. \]  

(16)

Hence \( \chi(z) \) has two poles in the cut plane. In addition, of course, the last two terms in eq. (14) indicate that \( \chi(z) \) has poles at \( z = \omega_p \) and \( \omega_q \), but owing to eq. (5) these poles lie on the boundary of the cut plane

\[ \mu < \omega_p, \quad \omega_q < E - \mu. \]  

If we approach one of these poles, for example \( \omega_p \), from below we see that \( H(E - z) \) on the left tends to \( H(E - \omega_p + i0) = H^+(\omega_q) \). As a result we calculate a residue \(-1\) for \( \chi(z) \) at these poles

\[ \chi(z) = -(z - \omega_p)^{-1} - (z - \omega_q)^{-1} + \ldots. \]  

(17)

At first sight one obtains a different residue, however, if one approaches the poles from above; a more careful consideration shows, however, that in addition to the 'explicit' pole terms on the right-hand side of eq. (14), we must also examine the singularities which may arise from the integral term, when the integration path (the real interval from \( \mu \) to \( +\infty \)) is 'pinched' between a pole of \( \chi^-(\omega') \), say \( \omega' = \omega_q + i0 \), and the zero \( \omega' = E - z \) of the denominator. This pinching occurs when \( z \) approaches \( E - \omega_q = \omega_p \) (respec-
tively \( E - \omega_p = \omega_q \) from above. Thus while the residues calculated from below, eq. (17), require no correction, the different residues obtained from above turn out to be erroneous. The correction is easily calculated by means of eq. (1) (compare AP1) and the result is, quite simply, that eq. (17) holds above the cut as well.

We now notice with Pagamenta that \( \chi(z) \) may be determined from the Cauchy formula, in our case:

\[
\chi(z) = \frac{R_+}{z - z_+} + \frac{R_-}{z - z_-} + \frac{1}{2\pi i} \int_{-\infty}^{E-\mu} \frac{\Delta \chi(\omega')d\omega'}{\omega' - z},
\]  

(18)

if the residues \( R_+ \) and \( R_- \), and the discontinuity across the cut (15),

\[
\Delta \chi(\omega) = \chi(\omega + i0) - \chi(\omega - i0),
\]  

(19)

are known. In order to evaluate the latter we follow more or less the procedure of AP2, eqs. (17) - (22), i.e. we calculate eq. (14) both above and below the cut and take the difference, thus:

\[
H^-(E - \omega)\chi^-(\omega) - H^+(E - \omega)\chi^+(\omega) = -2i\sigma(E - \omega)\chi^-(E - \omega) + 2\pi i \{H^+(\omega q)\delta(\omega - \omega_p) + H^+(\omega p)\delta(\omega - \omega_q)\},
\]  

(20)

and replace \( H^-(E - \omega) \to H^+(E - \omega) + 2i\sigma(E - \omega) \); we then divide by \( H^+(E - \omega) \) obtaining

\[
\Delta \chi(\omega) = -2i\sigma(E - \omega)[H^+(E - \omega)]^{-1} \{\chi^+(\omega) + \chi^-(E - \omega)\}
\]

\[
+ 2\pi i [\delta(\omega - \omega_p) + \delta(\omega - \omega_q)].
\]  

(21)

The \( \delta \)-terms when inserted into eq. (18) give, of course, the pole terms (17); they need no further attention. The key to the solution is the fact (compare AP1, 2 and Sommerfield's paper of ref. [4]) that the expression between braces \( \ldots \) is the boundary value \( M^+(\omega) \) of the function \( M(z) = \chi(z) + \chi(E - z) \) for which one obtains the remarkably simple result

\[
M(z) = \chi(z) + \chi(E - z) = \frac{L}{H(z)H(E - z)},
\]  

(22)

where \( L \) is a constant! To this end one shows that

\[
L(z) = H(z)H(E - z)M(z),
\]  

(23)

is regular analytic (holomorphic) in the whole \( z \)-plane. Since it is, rather obviously, bounded at \( \infty \), it must be a constant, hence eq. (22). To see that \( L(z) \) is holomorphic, one notices first that the poles (17) disappear when one forms the symmetric combination \( \chi(z) + \chi(E - z) \). After that, there are still poles at \( z_{\pm} = E - E_{\pm} \), see eq. (18), and at \( E - z_{\pm} = E_{\pm} \). These poles, however, are zeroes of \( H(E - z) \) and \( H(z) \) respectively; the product (23), therefore, has no poles whatsoever. Finally, we must still show that \( L(z) \) is continuous across the real axis; \( M(z) \), of course, has both cuts (12) and (15), but if we manipulate eq. (20) in a different way, eliminating \( \sigma \) instead of \( H^- \), we obtain: 

---

123
\[ H^-(E - \omega)[\chi^+(\omega) + \chi^-(E - \omega)] = H^+(E - \omega)\{\chi^-(\omega) + \chi^-(E - \omega)\} + 2\pi i \{H^+(\omega)\delta(\omega - \omega_p) + H^+(\omega_p)\delta(\omega - \omega_q)\} \text{.} \] (21')

We multiply on both sides by \(H^+(\omega)\), whereupon the right-hand side is clearly seen to be invariant under the substitution \(\omega \rightarrow E - \omega\), while the left-hand side becomes

\[ H^+(\omega)H^-(E - \omega)M^+(\omega) = L^+(\omega) \text{.} \]

Thus we get \(L^+(\omega) = L^+(E - \omega)\). On the other hand, clearly \(L(\omega) = L(E - \omega)\), hence \(L^+(E - \omega) = L^-(\omega)\). Finally,

\[ L^+(\omega) = L^-(\omega) \text{.} \] (24)

The proof of eq. (22) is thus completed.

One easily sees that this equation also allows one to express the residues \(R_\pm\), eq. (18), in terms of the constant \(L\)

\[ R_\pm = -L[H(z_\pm)H'(E_\pm)]^{-1} \text{.} \] (25)

When the results (21), (22) and (25) are inserted into eq. (18), one obtains, remembering eq. (2),

\[ \chi(z) = (\omega_p - z)^{-1} + (\omega_q - z)^{-1} + L\{r_+(z_+ - z)^{-1} + r_-(z_- - z)^{-1} + I(E - \omega)\} \text{,} \] (26)

where

\[ r_\pm = [H(z_\pm)H'(E_\pm)]^{-1} \text{,} \] (27)

and

\[ I(z) = \frac{1}{\pi} \int_{-\infty}^{\infty} \omega' \frac{1}{E - z - \omega'} \frac{1}{H^+(\omega')} \text{Im} \frac{1}{H^+(E - \omega')} \text{d}z \]

\[ = \frac{1}{\pi} \int_{\mu}^{\infty} \frac{1}{\omega' - z} \frac{1}{H^+(E - \omega')} \text{Im} \frac{1}{H^+(\omega')} \text{d}z \text{.} \] (28)

We still have to determine the constant \(L\). To do this, one can (see AP1, eq. (31), etc.) substitute eq. (26) back into the integral equation, (11) or (14), and verify that the equation is satisfied for a certain value of \(L\). We can avoid this rather tedious calculation, if we notice that, according to eq. (14), \(\chi(z) = O(z^{-2})\) when \(|z| \rightarrow \infty\). The terms of \(O(z^{-1})\) in eq. (26) must therefore vanish. This gives the condition:

\[ L[A(E) - r_+ - r_-] = 2 \text{,} \] (29)

where
\[ A(E) = \frac{1}{\pi} \int_{\mu}^{\infty} \frac{d\omega}{H^+(E-\omega)} \text{Im} \frac{1}{H^+(\omega)} . \] (30)

Our solution is now complete.

3. THE S-MATRIX

The wave-function (9) could be labeled more properly \( \phi(\omega_1\omega_2; \omega_p\omega_q) \); also \( \alpha(\omega) \), of course, depends on \( \omega_p \) and \( \omega_q \), although this is not indicated explicitly. The S-matrix element

\[ \langle N_\theta_1 \delta^{|\text{in}} \mid \eta \mid N_\theta_2 \delta^{|\text{out}} \rangle = 2 \left( \frac{k_1k_2\rho q}{\omega_1\omega_2\omega_p\omega_q} \right)^{1/2} S(\omega_1\omega_2; \omega_p\omega_q) \] (31)

can be calculated from the expansion (6). More rapidly \( S(\omega_1\omega_2; \omega_p\omega_q) \) can be obtained from Möller's wave matrix \( \phi(\omega_1\omega_2; \omega_p\omega_q) \). The usual recipe is: multiply \( \phi \) by the square root of the final velocities \( v_i = k_i\omega_i^{-1} (i = 1, 2) \) and divide by the square root of the initial velocities \( p\omega_1^{-1} \) and \( q\omega_2^{-1} \); then replace the denominator \( (E+i0 - \omega_1 - \omega_2)^{-1} \) with \( -2\pi i \delta(E - \omega_1 - \omega_2) \); in addition the 'unit operator' (the first term on the right of eq. (9)) must be symmetrized, as will automatically happen if we calculate eq. (31). Thus, remembering eq. (5),

\[ S(\omega_1\omega_2; \omega_p\omega_q) = \Delta - 2\pi i \delta(\omega_1 + \omega_2 - \omega_p - \omega_q) T , \] (32)

where

\[ \Delta \equiv \Delta(\omega_1\omega_2; \omega_p\omega_q) = \frac{1}{2} \left[ \delta(\omega_1 - \omega_p)\delta(\omega_2 - \omega_q) + \delta(\omega_1 - \omega_q)\delta(\omega_2 - \omega_p) \right] , \] (33)

\[ T \equiv T(\omega_1\omega_2; \omega_p\omega_q) = \frac{1}{(g/4\pi)\nu_1 u_2} \left[ \omega_p\omega_q k_1k_2/(\rho q) \right]^{1/2} \left[ \alpha(\omega_1) + \alpha(\omega_2) \right] . \]

We decompose \( T = T_a + T_b \) corresponding to the two terms on the right of eq. (10); then

\[ \delta(\omega_1 + \omega_2 - \omega_p - \omega_q) T_a = - \frac{1}{\pi} \left[ \frac{\sigma(\omega_p)}{H^+(\omega_p)} + \frac{\sigma(\omega_q)}{H^+(\omega_q)} \right] \Delta . \] (34)

These terms, as one easily sees, correspond to disconnected diagrams in which one of the \( \theta \)-particles goes through without interacting, while the other one is scattered elastically (see I, eq. (2.16)).

In order to calculate \( T_b \) we see from eqs. (9) and (10) that we need an expression for \( \chi^-(\omega_1) + \chi^-(\omega_2) \). Since \( \omega_2 = E - \omega_1 \) (on the mass shell), this is just the expression between braces on the right-hand side of eq. (21'), while on the left side \( \chi^+(\omega) + \chi^-(E - \omega) = M^+(\omega) \). We calculate therefore easily, with the help of eq. (22):

\[ \chi^-(\omega) + \chi^-(E - \omega) = -2\pi i [\delta(\omega - \omega_p) + \delta(\omega - \omega_q)] + \frac{L}{H^+(\omega)H^+(E - \omega)} . \] (35)
Hence

\[
\delta(E - \omega_1 - \omega_2)T_B = -\frac{2i}{\pi} \frac{\sigma(\omega_p)\sigma(\omega_q)}{H^+(\omega_p)H^+(\omega_q)} \Delta
\]

\[
+ \delta(\omega_1 + \omega_2 - \omega_p - \omega_q) \frac{1}{2\pi^2} \left[ \sigma(\omega_1)\sigma(\omega_2)\sigma(\omega_p)\sigma(\omega_q) \right]^{\frac{1}{2}} \frac{L}{H^+(\omega_1)H^+(\omega_2)H^+(\omega_p)H^+(\omega_q)}.
\]

(36)

Let us now put all the terms together. Notice the simplification

\[
1 + 2i \frac{\sigma(\omega_p)}{H^+(\omega_p)} + 2i \frac{\sigma(\omega_q)}{H^+(\omega_q)} - 4 \frac{\sigma(\omega_p)\sigma(\omega_q)}{H^+(\omega_p)H^+(\omega_q)} = e^{2i(\delta_p + \delta_q)}
\]

(37)

where \(\delta_p\) is the phase shift in the \(N^\theta\) sector

\[
e^{2i\delta_p} = H^-(\omega_p)/H^+(\omega_p).
\]

(38)

Compare I, eq. (2.16). Then, with the abbreviations

\[
\sigma_i = \sigma(\omega_i), \quad H_i = H^+(\omega_i), \quad (i = 1, 2, p, q),
\]

(39)

we have

\[
S(\omega_1\omega_2; \omega_p\omega_q) = e^{i(\delta_p + \delta_q + \delta_1 + \delta_2)} \left\{ \Delta(\omega_1\omega_2; \omega_p\omega_q)
\right.
\]

\[
- \frac{i\pi^{-1} L}{\pi} \delta(\omega_1 + \omega_2 - \omega_p - \omega_q) \left[ \sigma_1 \sigma_2 \sigma_p \sigma_q \right]^{\frac{1}{2}} |H_1H_2H_pH_q|^{-1} \}.
\]

(40)

We now check unitarity

\[
\int_{\mu}^{\infty} d\omega_1 \int_{\mu}^{\infty} d\omega_2 S(\omega_1\omega_2; \omega_p\omega_q) S^*(\omega_1\omega_2; \omega_r\omega_s) = \Delta(\omega_r\omega_s; \omega_p\omega_q)
\]

\[
+ e^{i(\delta_p + \delta_q - \delta_r - \delta_s)} \left[ \sigma_p \sigma_q \sigma_r \sigma_s \right]^{\frac{1}{2}} |H_pH_qH_rH_s|^{-1} \frac{1}{\pi} \delta(\omega_p + \omega_q - \omega_r - \omega_s)
\]

\[
\times \left\{ -i(L - L^*) + \frac{1}{L} \int_{\mu}^{\infty} \frac{\sigma_1 \sigma_2}{|H_1|^2|H_2|^2} \delta(E - \omega_1 - \omega_2) d\omega_1 d\omega_2 \right\}.
\]

(41)

Now from eqs. (29) and (30) we see that the expression between braces, divided by \(|L|^2\), is equal to

\[
-\text{Im} \frac{2}{L} + \frac{1}{\pi} \int_{\mu}^{\infty} \frac{\sigma(\omega)\sigma(E - \omega)}{|H^+(\omega)|^2|H^+(E - \omega)|^2} d\omega
\]

\[
= -\text{Im} A(E) + \frac{1}{\pi} \int_{\mu}^{\infty} \text{Im} \frac{1}{H^+(\omega)} \text{Im} \frac{1}{H^+(E - \omega)} d\omega = 0.
\]

(42)
The residues \( r_+ \) and \( r_- \) do not contribute because of the reality condition \( H(z^*) = [H(z)]^* \). As a result we have proved the unitarity condition

\[
\int d\omega_1 d\omega_2 S(\omega_1 \omega_2; \omega p \omega q) S^*(\omega_1 \omega_2; \omega q \omega s) = \frac{1}{2} \left\{ (\omega p - \omega q) \delta(\omega q - \omega s) + (p = q) \right\} .
\]  

(43)

The proof is, of course, essentially the same as would apply to the situation considered in AP2, sect. V. The interest here was in showing that the existence of complex roots does not create any difficulties in extending unitarity to the situation envisaged here.

4. NON-EXISTENCE OF A BOUND STATE IN THE N\( \theta \theta \) SECTOR

To complete our discussion of the N\( \theta \theta \) sector of the soluble model, we shall prove that in the present case the only eigenstates with real eigenvalues in the N\( \theta \theta \) sector are the scattering states \(|N\theta \theta q_{\text{out}}\rangle\) or \(|N\theta \theta q_{\text{in}}\rangle\).

To establish this, we start from the time-dependent Schrödinger equation, but, instead of eq. (5), require that the energy value

\[
E < 2\mu .
\]  

(44)

Such a state, if it did exist, would correspond to a bound state; however, as the following discussion will show, it in fact cannot exist. Formally, one may express such a state in terms of the right-hand side of eq. (6) where the functions \( \phi(\omega_1, \omega_2) \) and \( \alpha(\omega) \) still satisfy eqs. (7) and (8). For the bound state, instead of eq. (9), one has simply

\[
\phi(\omega_1, \omega_2) = -(g/4\pi) u_1 u_2 (\omega_1 \omega_2)^{\frac{1}{2}} (E - \omega_1 - \omega_2)^{-1} [\alpha(\omega_1) + \alpha(\omega_2)] .
\]  

(45)

After eliminating \( \phi(\omega_1, \omega_2) \), the integral equation for \( \alpha(\omega) \) becomes a homogeneous one which can be written as

\[
\int_{\mu}^{\infty} K(\omega, \omega') \alpha(\omega') d\omega' = 0 ,
\]  

(46)

where

\[
K(\omega, \omega') = -\sigma(\omega) H(E - \omega) \delta(\omega - \omega') + \frac{1}{\pi} \frac{\sigma(\omega) \sigma(\omega')}{\omega + \omega' - E} .
\]  

(47)

We recall that \( \sigma(\omega) \) is, by definition, real and positive. For \( E < 2\mu \), because of eq. (3), \(-H(E - \omega)\) is also real and positive. Thus, the matrix

\[
K = (K(\omega, \omega')) ,
\]  

(48)

is real and symmetric. By using the identity

\[
\int_{\mu}^{\infty} d\omega \int_{\mu}^{\infty} d\omega' \alpha^*(\omega) \alpha(\omega') \sigma(\omega) \sigma(\omega')(\omega + \omega' - E)^{-1} = \int_0^{\infty} d\lambda \left| \phi(\lambda) \right|^2 ,
\]  

(49)
where

\[ \phi(\lambda) = \int_{-\infty}^{\infty} d\omega \sigma(\omega) \sigma(\omega) \exp[-\lambda(\omega - \frac{1}{2}E)] , \]  

one sees that the matrix K is also positive definite. Consequently, eq. (46) has no solution except \( \sigma(\omega) = 0 \), which completes the proof.

REFERENCES

Finite Theory of Quantum Electrodynamics

T. D. Lee and G. C. Wick
Physics Department, Columbia University, New York, New York 10027
(Received 15 May 1970)

A new theory of quantum electrodynamics is presented, which is relativistically invariant, gauge invariant, unitary, and free of divergences. In this theory, mass renormalization, charge renormalization, and wave-function renormalization are all finite. Experimental consequences are discussed, and theoretical implications, especially those related to causality, are analyzed.

I. INTRODUCTION

In spite of the spectacular success of the renormalized theory of quantum electrodynamics, there remain unsatisfactory aspects concerning the inherent mathematical ambiguities in manipulating divergent expressions. Even apart from questions of mathematical rigor, there are serious difficulties when one tries to extend the renormalization process to hadronic systems. Consider, for example, the problem of the mass difference between $\pi^+$ and $\pi^0$. It seems reasonable to assume that this mass difference is due purely to the electromagnetic interaction. Yet, according to the usual theory, this mass difference is calculated to be infinite; furthermore, it can be shown that under rather general assumptions such an infinity, being closely connected with the equal-time commutator between the current operator and its derivatives, cannot be removed through strong interactions. The same divergence difficulties exist for all observed (therefore finite) mass differences between different hadrons in the same isospin multiplet; all such mass differences are found to be infinite according to the usual theory of quantum electrodynamics. Similarly, by making the universality assumption about the weak interaction, one expects that at least the ratio between the radiative correction to the Fermi constant $G_F$ in a $\beta$ decay and that to the $\mu$-decay constant $G_\mu$ should be finite. Again, it is infinite in the conventional theory.

Now, the various fractional mass differences between different hadrons in the same isospin multiplet are certainly finite, and are all of the order of the fine structure constant $\alpha$. By using the Cabibbo theory, the amount of the radiative correction to $G_F/G_\mu$ can be deduced from the observed rates of $\beta$ decay and $\mu$ decay; it is, of course, finite, and also of the order of $\alpha$. The fact that all these quantities are of the right order of magnitude strongly indicates that they are indeed due to second-order electromagnetic processes. It appears then that there must be fundamental changes in our basic formulation of quantum field theory, so that unrenormalized masses and unrenormalized coupling constants can become finite.

Recently, it has been found that there exists a general class of field theories in which the $S$ matrix is fully unitary, but the Lagrangian is not Hermitian. This makes it possible to construct relativistic local field theories which satisfy the unitarity requirement and are free from divergences. In particular, by replacing in the electromagnetic interaction the usual zero-mass photon field $A_\mu$ by a complex field

$$\phi_\mu = A_\mu + i B_\mu,$$

where $B_\mu$ is a massive boson field associated with a negative metric, it is possible to remove all infinities from the electromagnetic mass differences between hadrons, as well as those associated with radiative corrections to weak decays.

The purpose of the present paper is, on the one hand, to supply some further details of this new form of quantum electrodynamics, and on the other hand to extend a similar modification also to the fermion fields, so that the unrenormalized electric charge is also finite. The resulting theory is then completely free from divergent expressions.

In order to gain proper perspective, we shall first review briefly in Sec. II some elementary properties of quantum theories with indefinite metric, and then proceed to analyze the relations between the usual commutation (or anticommutation) relations and the metric of the system. It is shown that, under rather general conditions, in the case of Fermi statistics the positive definiteness, or indefiniteness, of the metric is uniquely determined by the equal-time anticommutation relations. Thus, once the Lagrangian for a fermion system is given, the metric is completely specified (of course, up to a canonical transformation, connected...
with possible changes of basis vectors). For the case of Bose statistics, the usual commutation relations do not uniquely determine the metric; there remains a free choice whether the metric is positive definite, or indefinite.

In Sec. III, we give the details for the modified photon field \( A_\pm + i B_\pm \). The interaction between the negative-metric \( B_\pm \) field and the charged lepton and hadron pair states requires that the modified photon propagator have, besides the usual photon pole at

\[
-(\text{four-momentum})^2 = 0, \tag{1.2}
\]

also complex poles at

\[
-(\text{four-momentum})^2 = (m_\beta \pm \frac{1}{2} i \gamma_\beta)^2, \tag{1.3}
\]

where, to \( O(\alpha) \), the partial width \( (\gamma_\beta)_\text{lepton} \) due to charged lepton pairs can be calculated, and is given by

\[
(\gamma_\beta)_\text{lepton} = \frac{3}{2} m_\beta. \tag{1.4}
\]

As will be discussed, in order to maintain unitarity, it is necessary that the poles due to the negative-metric \( B_\pm \) field are, indeed, off the real axis.

It is well known that in order to regularize the usual charge renormalization, one cannot, because of gauge invariance, simply modify, say, the electron propagator by introducing convergence factors. This difficulty is resolved in Sec. IV. We introduce two new Dirac spinor fields; their charges are imaginary \( \pm i e \), and their masses are complex

\[
m_F \pm \frac{1}{2} i \gamma_F, \tag{1.5}
\]

where \( m_F \) and \( \gamma_F \) are both real and not zero. Both fields obey Fermi statistics. As will be shown, in this theory, the metric is uniquely determined by the canonical anticommutation relations and is indefinite; the gauge invariance and, therefore, current conservation are fully satisfied by introducing the usual minimal electromagnetic interaction; the unitarity is maintained because of the complex masses; and the charge renormalization becomes finite since, because of their \( \pm i e \) charges, the loop diagram due to these new fermion fields is of the opposite sign from the usual one.

The modified Feynman rule is given in Sec. V. Explicit evaluations of charge renormalization and photon propagator are carried out in Sec. VI. Experimental consequences of this new finite theory of quantum electrodynamics are discussed in Sec. VII, and the causality problem is analyzed in Sec. VIII.

II. INDEFINITE METRIC

In this section we collect, for the convenience of the reader, some well-known facts about vector spaces with an indefinite metric and self-adjoint operators, and we discuss the relations between the usual canonical commutation (or anticommutation) rules and the general structure of metric that can be used in a quantum theory.

Let us consider a complex vector space \( \mathcal{H} \); let \( |x\rangle, |y\rangle, \ldots \) be vectors in \( \mathcal{H} \), and let \( |1\rangle, |2\rangle, \ldots \) be a basis, so that

\[
|x\rangle = \sum_i x_i |i\rangle, \tag{2.1}
\]

where \( x_1, x_2, \ldots \) are complex numbers. We define the scalar product \( \langle x | y \rangle \) to be the Hermitian form

\[
\langle x | y \rangle = \sum_{i,j} \eta_{ij} x_i^* y_j, \tag{2.2}
\]

where an asterisk denotes complex conjugation, and the matrix

\[
\eta = (\eta_{ij}) \tag{2.3}
\]

is assumed to be a non-singular Hermitian matrix. One calls the metric indefinite if \( \eta \) is not a positive definite matrix. From (2.2), one has, of course,

\[
\langle i | j \rangle = \eta_{ij}. \tag{2.4}
\]

A linear operator \( A_{ij} \) can be specified by the equations

\[
A_{ij} |i\rangle = \sum_j A_{ij} |j\rangle, \tag{2.5}
\]

where the \( A_{ij} \)'s are complex numbers; the adjoint operator \( \bar{A}_{ij} \) is defined by the equation

\[
\langle x | \bar{A}_{ij} | y \rangle = \langle y | A_{ij} | x \rangle^* \tag{2.6}
\]

for all vectors \( |x\rangle \) and \( |y\rangle \), where \( \langle x | A_{ij} | y \rangle \) denotes the scalar product between \( |x\rangle \) and \( A_{ij} |y\rangle \). According to (2.2) and (2.4), we have

\[
\langle x | A_{ij} | y \rangle = \sum_{i,j,k} x_i^* \eta_{ik} A_{jk} y_k. \tag{2.7}
\]

The expectation value \( \langle x | A_{ij} | x \rangle \) of a self-adjoint operator, defined by

\[
A_{ij} = \bar{A}_{ij}, \tag{2.8}
\]

is, therefore, real for all vectors \( |x\rangle \).

All these equations can, of course, be expressed in terms of the appropriate matrix notations. For example, Eqs. (2.2) and (2.7) can be simply written as

\[
\langle x | y \rangle = \bar{z} y \tag{2.9}
\]

and

\[
\langle x | A_{ij} | y \rangle = \bar{z} A_{ij} \tag{2.10}
\]

where \( y \) is a column matrix whose \( j \)th matrix element is \( y_j \), \( z \) is a row matrix whose \( i \)th matrix element is

\[
\bar{z}_i = \sum_j z_j^* \eta_{ij}, \tag{2.11}
\]

and \( A \) is the square matrix

\[
A = (A_{ij}). \tag{2.12}
\]
Note that the matrix element $A_{ij}$ should not be confused with the scalar product between the basis vector $|i\rangle$ and the vector $A_{op}|j\rangle$, which is given by

$$\langle i|A_{op}|j\rangle = (\eta A)_{ij}.$$  \hspace{1cm} (2.13)

It is worthwhile to point out that the well-known transformation law of the matrix $\eta$ implies that $\eta$ does not represent a \textit{bona fide} operator. In fact, if one chooses new basis vectors $|1\rangle$, $|2\rangle$, $\ldots$, then in terms of the new basis, (2.1) becomes

$$|x\rangle = \sum_i x_i |i\rangle \Rightarrow \sum_i x'_i |i\rangle.$$  \hspace{1cm} (2.14)

Thus $x_i = \sum_j T_{ij} x'_j$, or, symbolically,

$$x' = T^{-1} x,$$  \hspace{1cm} (2.15)

where, again, $x$ and $x'$ are column matrices, and $T^{-1}$ is the inverse of the matrix $T = (T_{ij})$. The transformed matrix $A'$ for the operator $A_{op}$, defined by

$$A_{op}|i\rangle = \sum_j A_{ij}|j\rangle,$$  \hspace{1cm} (2.16)

is given by

$$A' = (A_{ij}) = T^{-1} A T.$$  \hspace{1cm} (2.17)

On the other hand, relative to the new basis, the metric is obviously given by the matrix $\eta' = (\eta_{ij})$, where

$$\eta' = (\langle i|\langle j\rangle),$$  \hspace{1cm} (2.18)

so that

$$\eta' = T^* \eta T.$$  \hspace{1cm} (2.19)

In view of the difference between (2.18) and (2.20) it is worthwhile to make clear distinction between $\eta$ and an \textit{operator}. For example, algebraic equations involving $\eta$ are not in general independent of the choice of basis. \textit{Bona fide} operators, of course, do not have this defect.\footnote{For this reason, we have chosen the notation in which $\eta$ is omitted in the scalar-product symbol, given by the left-hand side of (2.2) (being “absorbed” in the definition of $\langle i|\rangle$). Such a notation has, of course, already been used occasionally, for example, by W. Heisenberg [Nucl. Phys. 4, 532 (1957)]. It differs from the more commonly used notation in the literature, for example, in Pauli [Rev. Mod. Phys. 15, 175 (1943)], and also in our earlier papers on the subject (Refs. 4–6). The present notation of not explicitly displaying $\eta$ has the advantage of avoiding unnecessary noncovariant equations.}

Similarly, the eigenvalues of an operator are basis independent. Those of $\eta$ are not. This has its good side: One can not only transform $\eta$ to a diagonal form, but also assume that the diagonal matrix elements are equal to $\pm 1$ (zero being excluded by the assumption that $\eta$ is nonsingular). At any rate, the eigenvalues of $\eta$ have no invariant meaning. The only invariant property of $\eta$ is the “signature” which (in a finite dimensional space) is defined to be the number of its positive eigenvalues minus that of its negative eigenvalues.

We now turn to the general question of different classes of the metric $\eta$ allowed in a quantum theory. [All $\eta$'s related by the transformation (2.20) are said to belong to the same class.] Through the usual quantization rules, one has, to begin with, the appropriate commutation, or anticommutation, relations for a set of operators $a$ and their adjoints $\bar{a}$; furthermore, one is often only interested in the case in which the matrix representation of the entire set of these operators $a$ and $\bar{a}$ is irreducible. As we shall see, in such a case, these commutation, or anticommutation, relations then completely determine the possible classes of the metric $\eta$.

We note that in any basis, because of (2.6) and $\eta \rightarrow \eta'$, the matrix representations of $a$ and $\bar{a}$ are related by

$$\bar{a} = \eta^{-1} a \eta,$$  \hspace{1cm} (2.20)

and therefore the over-all sign of $\eta$ cannot be determined by any set of algebraic equations between $a$ and $\bar{a}$. Of course, such an over-all sign has no physical meaning, since it merely changes the entire $\eta$ matrix by a minus sign; in any quantum theory, only the relative phase has no physical significance. Thus, it is convenient to take care of this trivial over-all sign choice by generalizing the definition of the same class of $\eta$ to include also

$$\eta \rightarrow -\eta,$$  \hspace{1cm} (2.21)

in addition to those related by (2.20); i.e.,

$$\eta \rightarrow T^* \eta T.$$  \hspace{1cm} (2.22)

The mathematical details for the determination of different classes of $\eta$ are given in Appendices A and B. Here, we shall discuss some of the main conclusions. Consider, for example, the case of a single harmonic oscillator obeying Fermi-Dirac statistics. We have

$$a^2 = 0 \quad \text{and} \quad \bar{a}^2 = 0.$$  \hspace{1cm} (2.23)

There are two different possibilities for the anticommulation relation between $a$ and $\bar{a}$:

(i) $\bar{a} a + a \bar{a} = 1$;  \hspace{1cm} (2.24)

(ii) $\bar{a} a + a \bar{a} = -1$.  \hspace{1cm} (2.25)

As will be shown in Appendix A, in case (i), by using (2.22), the metric $\eta$ can always be set to be positive definite; this corresponds to the usual positive-metric case, in which

$\langle x| x \rangle$ is positive  \hspace{1cm} (2.26)

for all vectors $|x\rangle$, and therefore one can always choose
the basis so that
\[ \eta = 1. \] (2.26')
In case (ii), there is also only a single class for the metric \( \eta \); by using (2.22), one can always require, instead of (2.26), that
\[ \langle x | (-1)^S | x \rangle \text{ is positive} \] (2.27)
for all vectors \( | x \rangle \). The metric \( \eta \) is therefore indefinite. If one wishes, one may choose a specific basis so that
\[ \eta = (-1)^S. \] (2.27')
We note that (2.26) and (2.27) are valid for all bases, while (2.26') and (2.27') are, of course, basis dependent. In either case, the quantum of this oscillator is said to be of positive or negative metric, depending on whether case (i) or case (ii) holds.

In the case of a single harmonic oscillator obeying Bose-Einstein statistics, the commutation relation
\[ a \dot{a} - a \dot{a} = 1 \] (2.28)
can be reduced to
\[ a \dot{a} - a \dot{a} = 1 \] (2.29)
by interchanging the roles of \( a \) and \( \dot{a} \). As will be shown in Appendix B, this commutation relation limits the metric \( \eta \) to one of the following three classes.

(i) **Definite metric case.** In this case (2.26) holds and therefore (2.26') is always a possible choice.
(ii) **Indefinite metric case (normal).** In this case (2.27) holds, and therefore (2.27') is always a possible choice.
(iii) **Abnormal case.** In this case the metric \( \eta \) is indefinite; unlike the situation in cases (i) and (ii), the eigenvalues of \( a \dot{a} \) are not integers, and they have neither upper nor lower bound.

In the following, we shall impose the condition that the operator \( a \dot{a} \) should be bound either from above, or from below; this would ensure, at least for the free oscillator, a lower bound to the energy spectrum. This additional condition then rules out the abnormal case (iii). The class of metric is then uniquely determined by specifying whether \( \eta \) is definite or indefinite. The quantum of this oscillator is said to be of positive or negative metric, depending on whether \( \eta \) is definite or indefinite.

For physical applications of a quantum theory with indefinite metric, we choose the Hamiltonian \( H \) to be a self-adjoint operator. Let \( | r \rangle \) denote any eigenvector of \( H \) with a real eigenvalue. We recall the elementary, but important, property\(^{10}\) that the \( S \) matrix is unitary if
\[ \langle r | r \rangle \text{ is positive} \] (2.30)
for all such eigenvectors \( | r \rangle \) with real eigenvalues. In such a theory, each physical observable is represented by a self-adjoint operator \( \hat{A} \), and each physical state \( \langle \rangle \) must be a linear superposition of only eigenaestates \( | r \rangle \) belonging to the real eigenvalues. The expectation value of \( \hat{A} \) is then given by the usual formula,
\[ \langle A \rangle = \langle | A | \rangle / \langle | \rangle \]. \] (2.31)

### III. PHOTON FIELD

In this section we shall review briefly the basic formalism of the modified photon field
\[ \phi_\mu = A_\mu + i B_\mu. \] (3.1)
For clarity, we consider first a simple system consisting of three fields: \( A_\mu, B_\mu \), and a charged lepton field \( \psi \), which can be either the usual electron field or the usual muon field. (The inclusion of other fermion fields of an unusual type will be discussed in Sec. IV.)

The Lagrangian density of this simple system is given by
\[ \mathcal{L}_{\text{free}}(\phi) + \mathcal{L}(l, \phi), \] (3.2)
where
\[ \mathcal{L}_{\text{free}}(\phi) = -\frac{1}{4} (G_{\mu\nu}^2 + F_{\mu\nu}^2) - \frac{1}{2} (m_\phi^2 B_\mu)^2, \] (3.3)
\[ \mathcal{L}(l, \phi) = -\bar{\psi} \gamma_\mu \left( \gamma^\mu \frac{\partial}{\partial x_\mu} - i e \phi \right) + m_l^2 \psi, \] (3.4)
\[ F_{\mu\nu} = \frac{\partial}{\partial x_\mu} A_\nu - \frac{\partial}{\partial x_\nu} A_\mu, \] (3.5)
\[ F_{\mu\nu} = \frac{\partial}{\partial x_\mu} B_\nu - \frac{\partial}{\partial x_\nu} B_\mu, \] (3.6)
\( e_0 \) is the unrenormalized charge, and \( m_\phi^2 \) and \( m_l^2 \) are the unrenormalized masses of the \( B_\mu \) and \( \psi \) fields. In the above expressions, \( \bar{\psi} \) denotes the adjoint operator of \( \psi \), and \( \phi_\mu \) is a self-adjoint vector field\(^{10}\) i.e.,
\[ \phi_\mu = + \phi_\mu \text{ for } \mu \neq 4 \]
and
\[ \phi_4 = - \phi_4. \]
Consequently, \( A_4 \) is a self-adjoint field,
\[ A_4 = + A_4 \text{ for } \mu \neq 4 \] (3.7)
and
\[ A_4 = - A_4, \] (3.8)
but \( B_\mu \) is not,\(^{11}\)
\[ B_\mu = - B_\mu \text{ for } \mu \neq 4 \] (3.9)
and
\[ B_4 = + B_4. \] (3.10)

The Hamiltonian \( H \) and its quantization can be carried out by following the usual canonical treatment. We may choose as generalized coordinates the lepton field \( \psi \), the transverse electromagnetic vector potential
\(^{10}\) Throughout the paper, the subscripts \( \mu \) and \( x \) denote the space-time indices; \( \mu = 4 \) is the time component with \( x_4 = t \). All boldface letters denote three-dimensional vectors.
\(^{11}\) If one wishes, one may replace \( B_\mu \) by \( B'_\mu \), which is a self-adjoint vector field. Then, of course, the free Lagrangian for \( B'_\mu \) appears in (3.3) with an unusual sign.
\( A_i^\mu \) (in the Coulomb gauge\(^{13}\)), and the spatial component \( B_j \) of the massive Boson field; their conjugate momenta are, respectively, \( i\phi_t^\pm, -E^\mu, \) and

\[
\Pi_j = i\xi_j. \tag{3.11}
\]

At equal time, one has

\[
\{\psi_i(x,t), \bar{\psi}_j(x',t)\} = \delta(x-x'), \tag{3.12}
\]

\[
[-E^\mu(x,t), A_i^\nu(x',t)] = -i(\delta_{ij} - \nabla_i \nabla_j) \delta(x-x'), \tag{3.13}
\]

and

\[
[\Pi_j(x,t), B_k(x',t)] = -i\delta_{jk} \delta(x-x'). \tag{3.14}
\]

It is convenient to introduce the usual Fourier components of these fields. For example, we may write

\[
B = \sum_{k,t} (2\Omega_0\omega)^{-1/2}(a_k e^{ik\cdot x} - \bar{a}_k e^{-ik\cdot x})\eta_k^t,
\]

\[
+ \sum_{k}(2\Omega_0\omega)^{-1/2}(a_k e^{ik\cdot x} + \bar{a}_k e^{-ik\cdot x})(\omega_k/k/m_0^2)
\]

and

\[
\Pi = \sum_{k,t} (2\Omega_0\omega)^{-1/2}(a_k e^{ik\cdot x} + \bar{a}_k e^{-ik\cdot x})(-i\omega_k\eta_k^t),
\]

where \( \delta^t \), \( \delta^s \), and \( k = |k|^2 k \) form a right-handed orthonormal set of unit vectors,

\[
\omega = [k^2 + (m_0^2)^2]^{1/2},
\]

and \( \Omega \) is the volume of the system. The commutation relation (3.14) becomes then

\[
[a_k^\alpha, \bar{a}_{k'}^\beta] = -\delta_{k,k'} \delta_{\alpha\beta}, \tag{3.18}
\]

where \( a, b \) denote either \( t=1, 2, \) or \( l \). As already discussed in the previous section, the anticommutation relation determines the metric of \( \psi_i \) to be positive; the commutation relations (3.13) and (3.14) limit the metric of these Boson fields to two possibilities: either positive or negative. We will now specify the metric of the system by requiring\(^{13}\)

\[
\langle (\{-1\}^N \psi \rangle \text{ to be positive} \tag{3.19}
\]

for all vectors \( \psi \), where

\[
N_B = -\sum_k \delta_k^t a_k^t - \sum_k \delta_k^s a_k^s. \tag{3.20}
\]

That is, \( \phi_h \) is of positive metric and \( B_\mu \) is of negative metric.

It is easy to verify that the total Hamiltonian \( H \) is self-adjoint. The equations of motion can be derived by using the usual Heisenberg equations, and the compatibility of the equal-time commutation relations (3.12)–(3.14) with relativistic invariance can be established following the usual arguments. For the unitarity property, we note that for the free field (\( \epsilon_0 = 0 \)), our basic condition (2.30) is not satisfied since there exist negative-metric eigenstates of the free Hamiltonian with real eigenvalues, e.g., states with \( N_B = \) odd integers. [For the free system, the condition (2.30) is, of course, irrelevant since the \( S \) matrix is the unit matrix.] In order to satisfy our basic condition (2.30), it is necessary to have \( \epsilon_0 > 0 \). Consider, for example, the negative-metric state \( |B \rangle \) of a single free \( B_\mu \) quantum, i.e., \( N_B = 1 \). For \( m_0^2 < 2m_0^2 \), such a state, in the limit \( \epsilon_0 = 0 \), is degenerate\(^{14}\) with the positive-metric lepton-pair states \( |l^+l^-\rangle \). It can be readily verified that this degeneracy is removed in the presence of the interaction. Furthermore, the resulting eigenvalues are\(^{15}\)

\[
m_{\notB} = \pm \frac{1}{2} \gamma_8, \tag{3.21}
\]

where \( \gamma_B \) and \( m_B \) are both real, denoting the width and the renormalized mass of the \( B_\mu \) quantum. The corresponding eigenvectors \( |\pm \rangle \) of these two complex eigenvalues consist of a coherent mixture of \( |B \rangle \) and \( |l^+l^-\rangle \); these eigenvectors both have zero norm,

\[
\langle + | + \rangle = \langle - | - \rangle = 0, \tag{3.22}
\]

and can be normalized so that

\[
\langle + | - \rangle = 1. \tag{3.23}
\]

The remaining eigenstates \( |\tau \rangle \) with real eigenvalues can then be shown to have positive norm, and therefore (2.30) is satisfied. Further details will be given in Secs. V and VI.

Since the propagator of the modified photon field \( \phi_h = A_\mu + iB_\mu \) can be easily seen to be proportional to \( k^{-4} \) in the high-momentum limit as \( k \to \infty \), one finds that, through a straightforward power counting, except for charge renormalization all higher-order processes of the Lagrangian (3.2) are finite. In order to render charge renormalization finite, additional new fermion fields have to be introduced; these will be discussed in Sec. IV.

\(^{13}\) The Coulomb gauge is chosen here purely for convenience. In the Coulomb gauge, \( A_\mu \) is given by \( A_\mu = A_\mu^\text{C} + i\phi_\mu \), where \( A_\mu^\text{C} \) satisfies \( \delta A_\mu^\text{C}/\delta \phi_\nu = 0 \) and \( \phi_\nu \) is the solution of \( \Delta \phi = -eQ/4\pi \). Correspondingly, the electric field is given by \( E_\mu = E_\mu^\text{C} + E_\mu^\text{m} \), where \( E_\mu^\text{m} = -\delta\phi_\mu/\partial \phi_\mu \). In deriving the canonical momentum of \( A_\mu^\text{C} \) as \(-E_\mu^\text{C}\), we have, as usual, first set in the Lagrangian the spatial integral \( \int d^4x E_\mu^\text{m} i\not{x} \psi = 0 \), and then taken its derivative with respect to \( \delta A_\mu^\text{C}/\delta \phi_\mu \).

\(^{14}\) We emphasize that (3.19), as well as other equations in this section, are all basis independent. Of course, it follows from (3.19) that, if one wishes, one can always choose a specific basis so that \( \eta = (-1)^N \).

\(^{15}\) If \( m_0^2 < 2m_0^2 \), then, in the limit \( \epsilon_0 = 0 \), this negative-metric state is degenerate with the positive-metric three-photon state. The introduction of \( \epsilon_0 \neq 0 \) would also reduce this degeneracy and make it possible to satisfy (2.30).

After the completion of this paper, Norman Kroll kindly drew our attention to the fact that in connection with his investigation on \textit{ad hoc} modifications of quantum electrodynamics [Nuovo Cimento 45, 65 (1966)], he has also noted the possibility of such complex poles and some of their consequences. The so-called "causality difficulty" mentioned in Kroll's paper, however, has been resolved in Ref. 4. (See also Sec. VIII of the present paper.)
IV. FERMION FIELDS

Let

\[ \psi_p = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \]  

(4.1)
denote a new fermion field, where \( \psi_1 \) and \( \psi_2 \) are both four-component Dirac fields, and

\[ \bar{\psi}_p = (\bar{\psi}_1 \bar{\psi}_2) \]

denotes the adjoint field, where the bar is used in the sense of Eq. (2.6). The Lagrangian density for a system consisting of \( A_\mu \), \( B_\mu \), the usual charged lepton fields \( \psi_\nu \), \( \bar{\psi}_\nu \) and this new fermion field \( \psi_p \) is given by

\[ \mathcal{L} = \mathcal{L}_{\text{bare}}(\phi) + \mathcal{L}(F, \bar{\phi}) + \sum_{\nu = 1}^{n} \mathcal{L}(l, \phi), \]

(4.2)

where \( \mathcal{L}_{\text{bare}}(\phi) \) and \( \mathcal{L}(l, \phi) \) are given, respectively, by (3.3) and (4.4),

\[ \mathcal{L}(F, \bar{\phi}) = -\bar{\psi}_p \gamma_\mu \gamma_5 \left[ \gamma_\mu \left( \frac{\partial}{\partial x_\mu} + e_\phi \phi_\lambda \right) + M \rho \right] \psi_p, \]

(4.3)

and \( \tau_\sigma \), \( \tau_\rho \) are two of the usual \((2 \times 2)\) Pauli matrices which commute with all Dirac matrices \( \gamma_\mu \) and satisfy

\[ \tau_\sigma^2 = 1, \]

(4.4)

\[ \tau_\rho^2 = 1, \]

(4.5)

\[ \tau_\sigma \tau_\rho = 0. \]

(4.6)

In (4.3), \( M \rho \) is also a \((2 \times 2)\) matrix, given by

\[ M \rho = m_\rho \gamma_5 + i \gamma_\rho \gamma_\lambda \tau_\lambda \]

(4.7)

where \( m_\rho \) and \( \gamma_\rho \) are real numbers. The total Lagrangian density is therefore self-adjoint, because of (4.5) and (4.6). From (4.3), one finds that the conjugate momentum of \( \psi_p \) is \( i \bar{\psi}_p \gamma_\mu \). The canonical quantization rule gives then, in addition to (3.12)–(3.14),

\[ \{ \psi_p(x), \bar{\psi}_p(x') \} = i \delta^4(x-x'). \]

(4.8)

By following the standard procedure of passing from Lagrangian to Hamiltonian, one finds the Heisenberg equation for the new fermion field,

\[ \gamma_\mu \left( \frac{\partial}{\partial x_\mu} + e_\phi \phi_\lambda \right) \psi_p + M \rho \psi_p = 0, \]

(4.9)

which implies

\[ \frac{\partial \bar{\psi}_p}{\partial x_\lambda} - \gamma_\mu \gamma_\lambda \bar{\psi}_p \gamma_\mu \left( e_\phi \phi_\lambda + M \rho \right) = 0. \]

(4.10)

We have, therefore, two conservation laws:

\[ \frac{\partial}{\partial x_\lambda} \left( i \bar{\psi}_p \gamma_\mu \gamma_\lambda \psi_p \right) = 0, \]

(4.11')

which correspond to the two similar conservation laws for the usual charged lepton fields,

\[ \sum_{\nu = 1}^{n} \frac{\partial}{\partial x_\lambda} (i \bar{\psi}_\nu \gamma_\mu \gamma_\lambda \psi_\nu) = 0 \]

(4.11)

and

\[ \frac{\partial}{\partial x_\lambda} (i \bar{\psi}_p \gamma_\mu \gamma_\lambda \psi_p - i e_0 \sum_{\nu = 1}^{n} M \rho \gamma_\lambda \psi_\nu) = 0. \]

(4.11')

The total electromagnetic current density \( j_\mu \) is given by

\[ j_\mu = e_0 \bar{\psi}_p \gamma_\mu \gamma_\lambda \psi_p - i e_0 \sum_{\nu = 1}^{n} \bar{\psi}_\nu \gamma_\lambda \psi_\nu, \]

(4.12)

and it is clearly conserved:

\[ \partial j_\mu / \partial x_\lambda = 0. \]

(4.13)

Under the gauge transformation

\[ A_\mu \rightarrow A_\mu + \partial \Lambda / \partial x_\mu, \]

(4.14)

and

\[ B_\mu \rightarrow B_\mu, \]

(4.15)

the charged lepton field \( \psi_\nu \) \((\nu = e, \mu)\) has the usual transformation

\[ \psi_\nu \rightarrow e^{i e_\phi \phi_\lambda} \psi_\nu, \]

(4.16)

and therefore

\[ \bar{\psi}_p \rightarrow \bar{\psi}_p e^{-i e_\phi \phi_\lambda}, \]

(4.17)

the new fermion field \( \psi_p \) transforms according to

\[ \psi_p \rightarrow e^{-i e_\phi \phi_\lambda} \psi_p, \]

(4.18)

By using (4.2), it can be readily verified that the theory is gauge invariant.

Just as in (3.19), we will now specify the metric of the system by requiring

\[ \langle (1-N^{\rho} + N^\nu) \rangle \]

(4.19)

for all vectors \( | \rangle \), where \( N_B \) is given by (3.20) and \( N_F \) is given by Eq. (C16) of Appendix C. By following the general discussions given in Sec. II, it is easy to see that the factor \((-1)^{N_F}\) is completely determined by the anticommutation relation (4.8), just as is the factor \((-1)^{n_\rho}\) in (2.27). The details are given in Appendix C. In the same appendix, the free Hamiltonian of the system is explicitly diagonalized. As will be shown there, the frequency spectrum of the free \( \psi_p \) field is given by

[\( \mathbf{k}^2 + (m_\rho^2 + \not{\mathbf{B}})^2 \mathbf{k}^2 \)].

We note that in the boson case, the bare mass \( m_B^0 \) of \( B_\rho \) is assumed to be real; the width \( \Gamma_B \) is acquired through the electromagnetic interaction between \( B_\rho \) and the charged lepton pairs, and therefore it is not an
independent parameter, as will be calculated in Sec. VI. On the other hand, in the fermion case, because of the conservation laws (4.10) and (4.10'), in order to ensure unitarity we have to assume a nonzero width $\gamma_\rho^0$ for the free $\psi_\rho$ field.

V. FERMATIAN DIAGRAMS

In the present theory, because of complex singularities such as (3.21) and (4.19), the Green’s function $U(t, -t)$ of the time-dependent Schrödinger equation in, say, the interaction representation diverges exponentially in the limit $t \to \infty$. On the other hand, the $S$ matrix is, of course, well defined in terms of the eigenvectors of the total Hamiltonian with real eigenvalues; its matrix elements are given by

$$S_{\rho, \sigma} = \langle \rho' | \rho \rangle \langle \sigma | \sigma' \rangle, \quad (5.1)$$

where the superscript “out” (or “in”) denotes states consisting of only plane waves and outgoing (or incoming) waves. Thus, one does not have the usual relation between the $S$ matrix and the limit $U(t, -t)$ at $t = \infty$. Nevertheless, as pointed out in Ref. 4, it is possible to separate $U(t, -t)$ into a sum

$$U(t, -t) = U^{\text{res}}(t, -t) + U^{\text{scat}}(t, -t), \quad (5.2)$$

where as $t \to \infty$, $U^{\text{scat}}(t, -t)$ diverges exponentially, but

$$\lim_{t \to \infty} U^{\text{res}}(t, -t) = S. \quad (5.3)$$

From (5.1) and (5.3), one obtains a set of modified Feynman rules which will be briefly reviewed in this section.

In general, in a quantum field theory with an indefinite metric, any $S$-matrix element can always be given by a sum over an appropriate set of Feynman graphs, just as in the usual theory with a definite metric. Each graph still stands for a multidimensional integral, integrated over a domain of some virtual four-momenta, which will be labeled collectively as $k'$; the integration over the space component $k$ remains, as usual, over the entire real region, but that over the time component $k_0$ is now along a complex path. We note that, unlike the case in the usual theory, the integrand now has complex singularities. To obtain the correct integration path in the complex $k_0$ plane, it is convenient to first regard all imaginary parts $\gamma_i$ of these complex singularities as independent parameters. For $\gamma_i = 0$, we have the usual Feynman rule: At fixed $k$, the integration over $k_0$ is along the real axis, and the positions of the singularities of the integrand are determined by the usual $ie$ rule. Alternatively, one may regard (still for $\gamma_i = 0$) all singularities to be on the real axis, but the Feynman path $C$ is slightly detoured so as to go either above or below the appropriate singularities. Now as $\gamma_i$ increases from zero to its final value, these singularities will move continuously. We shall require the integration path $C$ for $dk_0$ to be continuously deformed in such a way that none of the singularities ever crosses $C$. It is clear that $C$ is unambiguously defined only if there is no pinching along the path; i.e., if complex singularities on different sides of the path do not coalesce. As we shall see, such pinching occurs only over an integration domain of "zero" measure, and therefore does not contribute to the Feynman integral.

For clarity, let us consider some specific diagrams for, say, the elastic scattering of $e^+e^-$ given in Fig. 1. The dashed line refers to the propagator of

$$\phi_\mu = A_\mu + iB_\mu,$$

which, as will be explicitly calculated in Sec. VI, contains complex poles at

$-(\text{four-momentum})^2 = M_B^2$ and $((M_B^*)^2$, \quad (5.4)

where

$$M_B = m_B + \frac{1}{2}i\gamma_B. \quad (5.5)$$

Thus, the integrand for either diagram in Fig. 1 has complex poles at, among others,

$$k_0 = (k^2 + M_B^*)^{1/2},$$

and

$$k_0 = p_0 - [(p - k)^2 + (M_B^*)^2]^{1/2}, \quad (5.7)$$

where $p_0$ denotes the total momentum of $e^+$ and $e^-$. According to the above "modified" Feynman rule, the pole (5.6) should be below the integration path $C$ and the pole (5.7) should be above $C$. This is always possible except when these two poles coalesce, which occurs when

$$p_0 = (k^2 + M_B^*)^{1/2} + [(p - k)^2 + (M_B^*)^2]^{1/2}. \quad (5.8)$$

Except for the special center-of-mass reference frame, in any reference frame the total momentum $p$ is not zero, and therefore Eq. (5.8) represents two independent conditions in order to satisfy both its imaginary and real parts:

$$|k| = |p - k| \quad (5.9)$$

and

$$p_0 = \frac{1}{2} \text{Re}(k^2 + M_B^*)^{1/2}, \quad (5.10)$$

since $p_0$ denotes the total energy of $e^+$ and $e^-$, which must be real.

In $k$ space, the points that satisfy these two conditions lie on a circle whose center is at $\frac{1}{2} |p|$. For all other points, there is no pinching and the Feynman integral over $k_0$ is well defined; the result of this $k_0$ integration is
a function of \( \mathbf{k} \) which is singular at the above-mentioned circle, but with singularities no worse than

\[
\left( \rho_0 - (k^* + M_{*})^{1/2} \right)^2 \left[ (p - k^*)^2 + (M_{*})^{1/2} \right]^{(2)} - 1.
\]  
(5.11)

The subsequent integration over \( \mathbf{k} \) is therefore absolutely convergent and well defined\(^{14}\); the ambiguity of the integrand over such a set of points of zero measure (a circle) in \( \mathbf{k} \) space does not lead to any ambiguity in the final result. Therefore, in any system (except the center-of-mass system), the Feynman integral can be obtained by continuous deformation from the usual Feynman path. For the particular center-of-mass frame \( p = 0 \), we require the corresponding Feynman integral to be obtained as the limit \( p \to 0 \).

The actual Feynman integration for the above diagrams can be carried out in exactly the same way as done in Sec. 5 of Ref. 5. The result is that, apart from the usual cut corresponding to the normal two-zero-mass-photon exchange, there is no additional imaginary part in the scattering amplitude due to the virtual two complex-mass heavy boson states. Consequently, unitarity holds. The usual condition of analyticity is, of course, violated.

The above discussions can be readily carried out for more complicated diagrams involving multiphoton exchanges. If the intermediate state consists of three or more particles, the equation for pinching always gives two independent conditions in any reference frame, including the center-of-mass frame. Take, for example, the case of three particles of energies, say,

\[
\omega_3 = (k^* + M_{*,*})^{1/2}, \quad \omega_4 = (q^* + (M_{*,*})^{1/2})^{1/2},
\]

and \( q_{p-k-q} = \mid p - k - q \mid \). Just as in Eq. (5.8), the condition for pinching occurs when

\[
\rho_0 = \omega_3 + \omega_4 + r_{p - k - q},
\]  
(5.12)

where \( p \) and \( \rho_0 \) denote the total momentum and energy of the system. Equation (5.12) implies two separate conditions,

\[
\mid k \mid = \mid q \mid
\]  
(5.13)

and

\[
\frac{1}{x} (p_0 - x_{p - k - q}) = \omega_3 \omega_4,
\]  
(5.14)

independently of whether \( p = 0 \) or not. Just as in the above case of two-photon exchange, the points that satisfy these two conditions (5.13) and (5.14) give zero contribution to the Feynman integral.

In these Feynman integrals, since the domain of all three-momenta is kept real, but that of the fourth components is complex, the question of relativistic invariance naturally arises, especially for more complicated diagrams involving several complex masses. In this connection, there exists an alternative, but manifestly covariant, prescription given by Cutkosky, Landshoff, Olive, and Polkinghorne\(^{15}\); in their prescription, whenever there are a pair of regularized photon lines, one always first sets the complex masses in one line to be \( M_{*} \) and \( M_{*,*} \), and in the other line \( M_{*,*} \) and \( M_{*,*} \). The physical case \( M_{*} = M_{*,*} \) is obtained as the limit \( M_{*,*} \to M_{*} \). Relativistic invariance then becomes obvious; for example, in the two-photon-exchange diagrams discussed, for arbitrary \( M_{*,*} \neq M_{*,*} \), the Feynman integral is well defined in any system \( p = 0 \) or \( p = 0 \). For simple diagrams, these two prescriptions give identical results. For more complicated diagrams there may be some differences,\(^{16}\) in which case one should adopt the prescription of Cutkosky et al. with (whenever necessary) some further specifications as to the correct order of limits. For example, Cutkosky et al. found that, for the so-called double ice cream cone diagram (which corresponds to a diagram of order at least \( e^{20} \)), depending on the order of limits, their prescription leads to two different expressions for the \( S \) matrix, each being fully relativistic and unitary. The mathematical complexities involved in these diagrams of rather high order have prevented us from appreciating fully their arguments. Nevertheless, it seems reasonable to expect that such an ambiguity can be resolved by noting the Bose statistics nature of photons. All limits must be taken symmetrically with respect to different internal photon lines. This would then lead to a unique answer for the double ice cream cone diagram.

VI. PHOTON PROPAGATOR

As an explicit example, we shall evaluate the propagator \( D \) of the modified photon field

\[
\phi_\mathbf{\rho} = A_\mathbf{\rho} + i B_\mathbf{\rho}.
\]  
(6.1)

It is convenient to represent all propagators between the two fields \( A_\mathbf{\rho} \) and \( B_\mathbf{\rho} \) by a \((2 \times 2)\) matrix

\[
D = \begin{pmatrix}
D_{AA} & D_{AB} \\
D_{BA} & D_{BB}
\end{pmatrix},
\]  
(6.2)

where the two subscripts denote, respectively, the initial and final states to be either \( A_\mathbf{\rho} \) or \( B_\mathbf{\rho} \). The propagator for the coherent mixture \( A_\mathbf{\rho} + i B_\mathbf{\rho} \) is then given by

\[
D = \frac{1}{2} \bar{\xi} \gamma \xi ,
\]  
(6.3)

where
\[ \xi = \begin{pmatrix} 1 \\ i \end{pmatrix} \] (6.4)

and \( \xi \) is the transpose. For the free fields, we have
\[ D = D_{\text{tree}} = -i \left( \frac{(m_0^2)^2}{k^2 + (m_0^2)^2} \right) \delta_{\mu\nu} + (\cdots k, k) \cdot \delta_{\mu\nu} \] (6.6)

and therefore becomes
\[ D_{\text{tree}} = -i \left( \frac{(m_0^2)^2}{k^2 + (m_0^2)^2} \right) \delta_{\mu\nu} + (\cdots k, k) \cdot \delta_{\mu\nu} \] (6.6)

where \((\cdots k, k)\) denotes gauge-dependent terms, which are always proportional to \( k \times k \).

In Fig. 2, we introduce
\[ i(k^2 \delta_{\mu\nu} - k_\mu k_\nu) P \] (6.7)

to represent the sum of all irreducible self-energy graphs in the propagator \( D_{AA} \); because of the coherent mixture (6.1) that appears in all electromagnetic interactions, identical sets of irreducible self-energy graphs also appear in the other propagators \( D_{AB}, D_{BA}, \) and \( D_{BB} \). Let \( \Pi \) be a \((2 \times 2)\) matrix, defined by
\[ \Pi = i(k^2 \delta_{\mu\nu} - k_\mu k_\nu) P \] (6.8)

One, then, has
\[ D = D_{\text{tree}} + D_{\text{tree}} \Pi D_{\text{tree}} + D_{\text{tree}} \Pi D_{\text{tree}} \Pi D_{\text{tree}} + \cdots \] (6.9)

which leads to
\[ D = \frac{-i(m_0^2)^2 \delta_{\mu\nu}}{k^2 + (m_0^2)^2 (1 - P)} + (\cdots k, k) \cdot \delta_{\mu\nu} \] (6.10)

In Fig. 2, the dashed line represents \( D \) itself. It is easy to see that the sum of the first four lowest-order graphs is finite, although each of them diverges logarithmically. With respect to the other graphs of higher order in \( \alpha \), since \( D \) is proportional to \( k^{-4} \) as \( k \to \infty \), they are all finite, as may be shown by means of the "power counting theorem." In this respect, the gauge invariance of our perturbation is crucial, since it leads to the cancellation of terms that would otherwise lead to divergences.

To calculate this sum, we shall follow the modified Feynman rule given in Sec. V. The result is
\[ P(z) = -\frac{1}{\alpha} \ln(M \rho M \rho^*/m_0) + f(z, m_0) \] (6.11)

where \( \alpha = (137)^{-1} \), \( M \rho = m_\rho + \frac{1}{2} i \eta \),
\[ z = -k^2 \] (6.12)

\[ f(z, \lambda) = (3z)^{-1} \left[ 4(\lambda^2 + 2z) [1 + (2z - \lambda)^2] J(z, \lambda) \right] \] (6.13)

\[ J(z, \lambda) = \int_0^1 \frac{dx}{\lambda^2 - x(1 - x)} \] (6.14)

and
\[ F = \frac{z^2 - 4\lambda^2}{1^2} \] (6.15)

We note that in (6.14) the integration path for \( x \) is along the real axis from 0 to 1; this follows from the modified Feynman rule, after some transformations and calculations. In (6.11), we have omitted the superposed 0 in all masses, since to \( O(\alpha) \) it is immaterial whether one uses the bare masses or the observed masses.

For \( \lambda = m_l \) (\( l = e \) or \( \mu \)), \( J(z, m_l) \) has a branch point at \( z = 4m_l^2 \), and we may choose the cut to be along the real axis from \( z = 4m_l^2 \) to \( \infty \). Outside the cut, the integral representation (6.14) is valid everywhere. The imaginary part of \( f(z, m_l) \) along the cut is given by
\[ \text{Im} f(z, m_l) = \pm \frac{z + 2m_l^2}{3z} \left( z - 4m_l^2 \right)^{1/2} \] (6.16)

where the + or - sign depends on \( z \) being slightly above or below the cut. For \( \lambda = M \rho, \) or \( M \rho^*, \) \( f(z, \lambda) \) is analytic everywhere except at the branch point \( z = 4M \rho^2, \) or \( 4(M \rho^*)^2 \). Furthermore, it follows from (6.14) that
\[ f(z, \lambda)^* = f(z^*, \lambda^*) \] (6.17)

\[ i(k^2 \delta_{\mu\nu} - k_\mu k_\nu) P = m_\rho + m_\mu + m_\nu + M \rho + M \mu + M \nu + \cdots \] (6.18)

Fig. 2. Sum of irreducible photon self-energy graphs. The charges carried by the four fermion lines, which are labeled by their masses \( m_\alpha, m_\beta, M_\rho, \) and \( M_\rho^* \), are, respectively, \( e_\alpha, e_\beta, i\sigma_\alpha, \) and \( -i\sigma_\beta \).

---

18 For a proof of this theorem, without postulating the validity of contour rotations in the energy variables, see W. Zimmermann, Commun. Math. Phys. 11, 1 (1968).

19 We note that each of the first four graphs in Fig. 2 becomes logarithmically divergent (and therefore the sum becomes finite) only if the integration is performed after the gauge-invariant term \((A_\mu^\alpha - k_\mu k_\alpha)\) is factored out. Quite often, it is much more convenient to do the integration before any factoring. In such a case, in order to avoid any mathematical ambiguity, one may define \( \Pi(m_\rho^2, m_\mu^2, M_\rho^*, M_\rho^*) \) to be the formal sum of all irreducible self-energy graphs, as before, but replace (6.8) by
\[ i \xi \xi \Pi(A_\mu^\alpha - k_\mu k_\alpha) \] and
\[ -\Pi(A_\mu^\alpha + m_\rho^2, m_\mu^2, M_\rho^*, M_\rho^*) \].

By taking the limit \( \Lambda \to \infty \) after the integration, one can then unambiguously arrive at (6.11).
and therefore \( f(\varepsilon, M^2) = f(\varepsilon, M^2) \) is real along the entire real axis, in accordance with the unitarity condition. By using (6.10), (6.11), and (6.16), one finds that besides the usual photon pole \( k^2 = 0 \), \( D \) has poles at

\[
-k^2 = (m_B \pm \frac{1}{2} \gamma_B)^2, \tag{6.18}
\]

where to \( O(\alpha) \) and neglecting \((m/m_B)^{32}\) as compared to unity,

\[
\gamma_B = \frac{3}{\alpha} m_B \tag{6.19}
\]

and

\[
m_B = (m_B^2)[1 - \frac{1}{3} \text{Re} F(m_B^2)]. \tag{6.20}
\]

The ratio of the renormalized charge \( e \) and the unrenormalized charge \( e_0 \) is, by definition, given by the behavior of \( D \) as \( \varepsilon \to 0 \), i.e.,

\[
D \to -i(k^2/e_0)^2 \text{ as } k^2 \to 0. \tag{6.21}
\]

Since \( f(\varepsilon, \lambda^2) \to \frac{1}{2} \) as \( \varepsilon \to 0 \), one finds for charge renormalization,

\[
\left( \frac{e_0}{e} \right)^2 = 1 + \frac{\alpha}{3\pi} \ln \left( \frac{M_{eF} M_{eP}}{m_{\mu} m_{\nu}} \right)^2 + O(\alpha^2), \tag{6.22}
\]

which, of course, is finite. As \( k^2 \to \infty \) and slightly above the real axis, one finds

\[
\text{Re} P \to (2\alpha/\pi)(M_{eF}^2 + M_{eP}^2 - m_{\mu}^2 - m_{\nu}^2),
\]

\[
\text{Im} P \to -\frac{3}{\alpha}. \tag{6.23}
\]

Therefore, one finds [after neglecting higher-order terms in both \( O(\alpha/k^4) \) and \( O(\alpha^2) \) that in the timelike region, as \(-k^2 \to \infty + i\varepsilon\), where \( \varepsilon \to 0+\),

\[
D \to -i(m_B^2)^2 \left[ k^2 + (m_B^0)^2 + \frac{3}{4} \alpha (m_B^0)^2 \right]^{-1}, \tag{6.24}
\]

while in the spacelike region, as \( k^2 \to \infty \),

\[
D \to -i(m_B^0)^2 \left[ k^2 + (m_B^0)^2 \right]^{-1}. \tag{6.25}
\]

Except for the constant imaginary parts in (6.23), both limiting expressions are exactly the same as \( D_{\text{tree}} \).

VII. EXPERIMENTAL CONSEQUENCES

The following remarks concern some experimental consequences of the theory.

(1) The modification of the photon propagator from the usual \(-ik^2\) to \(D(k)\), given by (6.10), in the present theory can be observed among any electromagnetic processes sensitive to the high \( k^2 \) behavior of the photon propagator, such as \( e^+ e^- \) and \( \mu^+ \mu^- \) scatterings for spacelike \( k^2 \), or pair productions of \( e^+ e^- \) and \( \mu^+ \mu^- \) from any high-energy collision processes involving leptons and/or hadrons for timelike \( k^2 \). In the timelike region, for \(-k^2 \to m_H^2 \), the transition probability, being proportional to \( |D(k)|^2 \), would exhibit a \( k^2 \) dependence identical to the standard Breit-Wigner resonance formula; on the other hand, the transition amplitude should have a phase that is of an opposite sign \((-90^\circ \text{ instead of } +90^\circ)\). At \(-k^2 \text{ away from } m_B^2\), one may use the zeroth-order expression for \(D(k)\):

\[
D(k) = \frac{-im_B^2}{k^2 + m_B^2 - \frac{3}{4} \alpha (m_B^0)^2 + \frac{1}{2} i(k^2 + m_B^2)}. \tag{7.1}
\]

Consider, e.g., the differential cross section of

\[
e^+ e^- \to \mu^+ \mu^- \tag{7.2}
\]

in the present theory; it differs from that in the usual quantum electrodynamics by a simple multiplicative factor

\[
\left[ m_B^2/(k^2 + m_B^2) \right]^2. \tag{7.3}
\]

It is of interest to note that this factor can be rather substantial even if \(-k^2\) is quite far from the resonance region. For example, for \( m_B \sim 20 \text{ GeV} \) the width \( \gamma_B \) is only \(-200 \text{ MeV} \); yet, at, say, \((-k^2)^{1/2} = 10 \text{ GeV}, \) the above factor is \(-1.8\).

(2) There should be a deviation in the gyromagnetic ratio \( g \) of the muon from the usual expression due to such a modification in the photon propagator:

\[
\delta g = -(3\alpha)^{-1} \alpha(m_\mu/m_B^2). \tag{7.4}
\]

From the present experimental result,\(^{20}\) one concludes that

\[
m_B > 5 \text{ GeV}. \tag{7.5}
\]

This limit is also consistent with the present high-energy experimental results\(^{19}\) on

\[
\beta + \text{uranium} \to \beta + \mu^+ \mu^- \cdots. \tag{7.6}
\]

(3) In principle, \( m_B \) can also be determined from the finite value of mass differences between hadrons in the same isospin multiplet and from radiative corrections to weak decays. In practice this is difficult, since all these terms depend on \( m_B \) only logarithmically, and none of these terms can be calculated accurately. For example, the mass difference \( \Delta m_B \), between \( s^\pm \) and \( s^+ \) in the usual chiral SU\(_2\)×SU\(_2\) phenomenological Lagrangian method is infinite, while in the present theory, it is of course finite. By using the same approximation for the strong-interaction vertex, one finds

\[
\Delta m_B = \frac{3\alpha m_B}{8\pi m_{\mu}} \left[ 2 \ln(1 + \delta^2) - \frac{m_{\mu}^2}{m_B^2} - \frac{m_B^2}{m_{\mu}^2} + O\left( \frac{m_B^2}{m_{\mu}^2} \right) \right]
\]

\[
\approx \left[ 1 + 0.003(1 + \delta^2) \ln\left( \frac{m_B^2}{m_{\mu}^2} \right) + O\left( \frac{m_B^2}{m_{\mu}^2} \right) \right] \text{MeV}, \tag{7.7}
\]

where the first term was first derived by Das et al.,\(^{21}\) and \( \delta \) is the anomalous gyromagnetic ratio of the \( A_1 \) meson, and


\(O(m_\gamma^2/m_e^2)\) denotes terms proportional to \((m_\gamma^2/m_e^2)\) but remaining finite even in the limit \(m_\gamma \to \infty\). The entire expression (7.7) is, of course, identical to those obtained by Gerstein et al., except for replacing their ad hoc cutoff parameter \(\Lambda\) by \(m_\gamma\). It is clear that once \(\Delta m_\gamma\) is made finite, it becomes rather insensitive to the precise value of \(m_\gamma\).

For the radiative correction to weak decays, the dependence on \(m_\gamma\) is again only logarithmical. The result is similar to those calculated by using the charged intermediate vector boson \(W^\pm\), except for replacing \(m_W\) by \(m_\gamma\). If such calculations involving strong-interaction vertices could be made accurate, then it would be possible to determine \(m_\gamma\) from the observed value of the Cabibbo angle \(\theta\). This is, of course, far from the actual case. Assuming that \(\theta = 2.02\), the best estimate for \(m_\gamma\) in the intermediate boson theory is

\[\ln(m_W/m_\gamma) = 2.8\pm 0.8 \text{ if quark algebra is applicable} \quad \text{and} \quad \ln(m_W/m_\gamma) = 3.5\pm 1.0 \text{ if field algebra is applicable.}\]

This leaves a large admissible range for \(m_\gamma\) (assuming \(m_W\) from about 7 to 90 GeV.

Thus, the best way to determine \(m_\gamma\) is through the direct observation of possible deviations from the conventional quantum electrodynamics predictions at high energy.

(4) By using (6.10) and (6.11), one sees that both \(m_\rho\) and \(\gamma_\rho\), and therefore also the finite value of charge renormalization \((e_\rho/e)^2\), can be determined by accurately measuring the photon propagator \(D(k)\) at high \(k^4\). However, these measurements are more difficult since it is then necessary to measure \(D(k)\) at least to an accuracy comparable to, or better than, \(O(\alpha)\).

(5) We remark that an attractive, but highly speculative, idea is to regard \(B_\alpha\) as the neutral component of the hypothetical charged intermediate boson field \(W^\pm\) for the weak interaction, in which case one would expect \(m_\gamma \sim m_W\). The further speculation that the semi-weak interaction coupling constant \(g^2/4\pi\) is, in fact, the same as the fine structure constant \(\alpha\) leads to

\[m_\gamma \sim m_W \sim (4\pi\alpha/G_V)^{1/2} \sim 100 \text{ GeV.}\]  

(7.8)

### VIII. CAUSALITY

The presence of complex singularities near the physical region, but on the physical sheet, has an unusual effect on the propagation of wave packets in a collision process. As has been discussed in Ref. 4, because of the uncertainty principle, such unusual effects disappear if one studies only the average position of the wave packet; this then automatically removes all of the so-called "causality difficulties" or, more precisely, those difficulties that could be directly related to a classical description. In quantum physics, as we shall see, in the first place, there is no general agreement as to the precise meaning of causality. While in the present theory, unusual effects do occur if one analyzes the shape of the wave packet, such effects, though unusual, are not in contradiction with anything known at present about the physical world. Furthermore, it should be clear that such effects cannot ever lead to logical difficulties (i.e., self-contradictory predictions), since they are the mathematical consequences of a set of well-defined self-consistent equations. In the following, we shall briefly review these unusual effects and their connection with the causality question.

In most papers on causality, attempts have been made to transform the somewhat ill-defined problem of causality to that of relativistic invariance, which can be stated with precision. In the classical derivation of the dispersion formula for light waves, one begins by assuming a sharp wave front for a physical signal in the theory; e.g., the signal is zero for the space-time region specified by, say, \(t<0\) and \(x<0\). The subsequent requirement that this wave front should not travel faster than the velocity of light leads to the well known classical dispersion formula which, in fact, rules out complex singularities such as those that appear in the present theory. We recall that in quantum field theory, it is not possible to construct such a sharp wave front for the incoming wave for all \(t<0\), since that would require the superposition of plane waves of all frequencies \(\nu\), negative as well as positive. Even for the zero-mass photon field, a coherent mixture of many photon states necessarily covers only the positive energy range, and therefore \(\nu \geq 0\), by using the time-dependent Schrödinger equation; for a massive field, the range of physically allowable frequencies is even smaller, since \(\nu \geq h^{-1} \times (\text{mass}) > 0\). The impossibility of constructing a wave packet with a sharply defined front makes it impossible to apply such a classical argument.

Another approach is to use local field theory. One requires the commutators between any two field operators at points separated by a spacelike distance to be zero. This requirement is satisfied in the present theory. The derivation of the usual analyticity condition is not applicable to our theory because it assumes the energy spectrum to be real (or, the underlying Lagrangian to be Hermitian), which is obviously not true in

---

9 For a more recent discussion, see A. Sirlin, in Proceedings of the Fourteenth International Conference on High-Energy Physics, Vienna, 1969, edited by J. C. Prüntki and J. Steinberger (CERN, Geneva, 1968). The fact that, in the intermediate boson theory of the weak interaction, the \(O(\alpha)\) radiative correction to the ratio \((G_V/G_a)\) is finite in the conventional form of quantum electrodynamics has, of course, been known for quite some time [T. D. Lee, Phys. Rev. 128, 899 (1962)].
10 The values quoted are based on the recent calculations by A. Sirlin (private communication). We thank Dr. Sirlin for making these values available to us.
12 N. G. Van Kampen, Phys. Rev. 89, 1072 (1953); 91, 1267 (1953).
the present theory. The violation of the usual analyticity condition is, however, totally consistent with the requirement of relativistic invariance.

Still another description of causality that has been used in the literature8 is one deriving from the study of the average motion of a wave packet. For example, in the case of a simple $S$-wave elastic scattering, if the incoming wave packet has an average relative position $\langle r_{\text{in}} \rangle = -\tau \delta$ for time $t < 0$, then the outgoing wave packet, for $t > 0$, has an average relative position

$$\langle r_{\text{out}} \rangle = \tau \delta + 1,$$

where

$$l = -2 \left( \frac{\partial \phi}{\partial \delta} \right)_{\delta = 0},$$

(8.2)

$\delta$ denotes the phase shift, and $k$ is the relative momentum. A $-90^\circ$ resonance, such as the one required by the $B_s$ field in our theory, would contribute a positive value for $l$, and therefore give rise to an advancement of the outgoing wave packet. However, it can be shown,30 under very general conditions, that one has the inequality

$$l < l_{\text{max}} = O(\Delta^{-1}),$$

(8.3)

where $\Delta$ denotes the momentum width of the incoming wave packet. Therefore, it is not possible to draw any strong conclusion concerning causality by studying the average position of a wave packet.

Rather unusual behavior of wave packets can, nevertheless, be demonstrated to exist in the present theory. In general, these unusual properties concern the detailed shape of the wave packet. Consider, for example, the elastic collision of $e^+$ and $e^-$ in the center-of-mass system at the resonance energy $m_{B_s}$. For clarity, let us assume the radial dependence of the incoming wave to be given by

$$\phi_{\text{in}}(r, \beta) \propto r^{-1} e^{-\Delta (t-r)}$$

(8.4)

at large relative distances $r$. As shown in Ref. 4, the presence of a complex pole, such as $(m_{B_s} - \frac{1}{2} \gamma_{B_s})$ in the photon propagator, implies that the outgoing wave has a radial dependence given by

$$\phi_{\text{out}}(r, \beta) \propto r^{-1} e^{-\Delta (t-r)}$$

for $t > r$, but

$$\phi_{\text{out}}(r, \beta) \propto r^{-1} \left[ e^{-\Delta (t-r)} - \frac{4\gamma \Delta}{(2\Delta + \gamma)^2} e^{-\frac{t}{2} \gamma_{B_s}} \right]$$

(8.5)

for $t < r$.

(8.6)

where $\gamma = \gamma_{B_s}$, the first term $e^{-\Delta (t-r)}$ merely reproduces the shape of the initial wave, but the second term $e^{-\frac{t}{2} \gamma_{B_s}}$ is quite unusual. Nevertheless, the presence of such a term is, of course, perfectly compatible with the requirement of relativistic invariance as well as with all existing experiments.

We note that at present in any high-energy experiment almost nothing is known concerning the shape of wave packets. Without some detailed knowledge of the shape of wave packets, one can study only the average positions ($\langle r \rangle$), which, as mentioned above, are insensitive to the complex singularities. In order to see the unusual tail $e^{-\frac{t}{2} \gamma_{B_s}}$, we may consider a measurement which can differentiate the time-advanced region, say $t > r$, from the time-retarded region $t < r$, where $r$ represents the experimental space-time resolution. At present, the best value of the time resolution in any high-energy experiment is $\sim 10^{-10}$ sec. Assuming, for example, $m_{B_s} \sim 20$ GeV and therefore $\gamma_{B_s} \sim 200$ MeV, the full intensity of the unusual tail, integrated over the entire time-advanced region from $t = r$ to $t = \infty$, comes out to be

$$\sim \exp(-\gamma_{B_s}) \sim \exp[-3 \times 10^{12}].$$

(8.7)

The smallness of this probability32 makes it unlikely that we can realistically detect such an unusual effect in any near future (assuming that we can reach a center-of-mass energy $\sim$ the resonance energy $m_{B_s}$). Of course, in principle, this effect should be measurable.

In any quantum theory, what one really studies are only correlations between various events occurring at different space-time regions. The impossibility of constructing during $t < 0$ a sharp wave front for the incoming wave makes it also not possible to give a strict causal interpretation to such correlations. Thus, there does not exist a sharply defined causality principle. (We regard requirements such as the usual zero commutator of two local field operators separated by a spacelike distance as simply an expression of relativistic invariance and local canonical quantum field theory, but not of causality.23) The attribution as to which effect should be regarded as “noncausal,” therefore, has a certain degree of arbitrariness, except in the classical limit.25 It seems nevertheless appropriate to call the above described unusual tail $e^{-\frac{t}{2} \gamma_{B_s}}$ in the outgoing wave packet “noncausal,” although one must emphasize that there is no logical difficulty in having this particular


9 See Ref. 4. The arguments used in Ref. 4 can be extended to the nonresonance region as well. The inequality (8.3) holds, provided the integral $\int \langle \partial \phi / \partial \delta \rangle \delta \phi$ is integrated over the entire momentum width of the incoming wave, $\sigma_{\text{in}}$.

10 The two-photon exchange processes can lead to a larger value, though still much too small to be detected at present. See T. D. Lee (Ref. 5) and Cutkosky et al. (Ref. 17).

11 A similar opinion is also expressed by G. Wanders [Nuovo Cimento 14, 168 (1959)], whose conclusions on causality seem to agree with ours also in other respects. General questions of causality have also been discussed by B. Ferretti, Nuovo Cimento 43, 506 (1966); 43, 516 (1966).

12 Since a macroscopic body does not necessarily imply the validity of a classical limit (e.g., the superfluidity phenomenon), there also does not exist in quantum physics any general macrocausality principle that can be sharply defined without further qualifications.
kind of “noncausal” effect; the question of whether it indeed exists in nature can only be resolved by future experimentation.

ACKNOWLEDGMENTS

We wish to thank N. Christ, S. Coleman, R. Jackiw, and B. Zumino for helpful discussions.

APPENDIX A

Let us study anticommutation relations for one oscillator, assuming

\[ a^2 = a^* = 0, \quad a a^* + a^* a = \pm 1. \]  

(A1)

The case with the + sign is, of course, well known, but we can study the two cases simultaneously if we define \( N = \pm a^* a \); then

\[ N^2 = N. \]  

(A2)

As is well known, \( N \) satisfying (A2) can always be diagonalized. In fact, any vector \( \psi \) can always be decomposed into a sum of two vectors (one of which may be zero) \( N \psi \) and \( (1-N) \psi \) which are eigenvectors of \( N \), corresponding to the eigenvalues 1 and 0, respectively. Both eigenvalues certainly exist; for if we assume a vector \( \psi \) with the eigenvalue 0, \( N \psi = 0 \), then from \( a a^* \psi = (\psi - N \psi) = -\psi \) we see that \( \psi = a^* \psi \); furthermore, \( N \phi = \pm a^* \phi = \phi \); and therefore \( \phi \) belongs to the eigenvalue 1. Conversely, if \( \phi \) is an eigenvector belonging to the eigenvalue 1, \( N \phi = \phi \), then \( a \phi \) belongs to the eigenvalue 0, for \( N a \phi = \pm a \phi \phi = 0 \); notice again that \( a \phi \) cannot be zero, since \( N \phi = \pm a \phi \phi \neq 0 \). These calculations are well known from the ordinary case, where \( a = a^* \) and the + sign holds in Eq. (A1). We have repeated them here, merely to remind the reader of the fact that the analysis makes no use whatsoever of the metric; it is based entirely on the algebraic structure of (A1). Pursuing the analysis further, one sees that the whole vector space \( \mathcal{S} \) is the direct sum of two subspaces \( \mathcal{S}_0 = \mathcal{S}_0 \phi \otimes \mathcal{S}_1 \), in one of which the operator \( N = 0 \), while in the other it acts as if the identity \( N = 1 \). We have seen that if \( \psi \) belongs to the first subspace \( \mathcal{S}_0 \), then \( \phi = a \psi \) belongs to \( \mathcal{S}_1 \), and we have the relations

\[ \tilde{a} \phi = \phi, \quad a \phi = \pm \psi, \]  

\[ a \phi = 0, \quad a^* \phi = 0. \]  

(A3)

The same relations are obtained if we start from any vector \( \phi \) of \( \mathcal{S}_1 \), and define \( \psi = a \phi \). From this we see that \( \tilde{a} \) maps \( \mathcal{S}_0 \) into \( \mathcal{S}_1 \), and \( \mathcal{S}_0 \) into \( \mathcal{S}_1 \). Regarded as a mapping of \( \mathcal{S}_0 \) into \( \mathcal{S}_1 \), \( \tilde{a} \) has \( \pm a \) as a left inverse. Likewise \( a \tilde{a} \) maps \( \mathcal{S}_1 \) into \( \mathcal{S}_0 \), and \( \mathcal{S}_1 \) into \( \mathcal{S}_0 \); regarded as a mapping of \( \mathcal{S}_1 \) into \( \mathcal{S}_0 \), \( a \) has \( \pm \tilde{a} \) as a left inverse. It follows that \( \mathcal{S}_0 \) and \( \mathcal{S}_1 \) have the same dimensionality.

Choosing bases \( \psi_1, \ldots, \psi_n \) in \( \mathcal{S}_0 \) and \( \phi_1, \ldots, \phi_n \) in \( \mathcal{S}_1 \) such that \( \psi_i \phi_j \) and \( \phi_i \phi_j \) are related to each other in the same way as \( \psi \) and \( \phi \) in (A3), \( a \) and \( \tilde{a} \) are reduced to the canonical form

\[ a = \begin{pmatrix} 0 & \pm I \\ 0 & 0 \end{pmatrix}, \quad \tilde{a} = \begin{pmatrix} 0 & 0 \\ I & 0 \end{pmatrix}, \]  

(A4)

where \( I \), of course, is the \( n \)-dimensional unit matrix. This may be obtained from the usual case, if in (A1) one identifies \( a \) with \( \pm a^* \). Notice again, however, that no use has been made of Hermiticity.

In order to determine uniquely the metric \( \eta \), we shall impose the condition that the set of matrices \( a \) and \( \tilde{a} \) should be irreducible. It follows, then, from (A4) that both \( a \) and \( \tilde{a} \) are \( (2 \times 2) \) matrices, i.e., there is only one \( \psi \) and one \( \phi \). To conform to the notations used in Sec. II, we shall denote \( \psi \) and \( \phi \) by \( |0 \rangle \) and \( |1 \rangle \), respectively. Equation (A3) becomes

\[ a |0 \rangle = |1 \rangle, \quad a |1 \rangle = \pm |0 \rangle, \]  

(A5)

and

\[ a |0 \rangle = 0. \]  

(A6)

Thus, one finds

\[ \langle 1 | 0 \rangle = \pm (\langle 1 | a | 1 \rangle = \pm (\langle 1 | a^* | 1 \rangle^* = 0, \]  

\[ \langle 0 | 1 \rangle = 0. \]  

(A6)

and

\[ \langle 1 | 1 \rangle = 0 \]  

(A7)

By using transformation (2.22), we can always choose \( \langle 0 | 0 \rangle \) to be +1, and therefore there are only two classes specified by (2.26) and (2.27).

The preceding result can be applied immediately to the case of \( n \) “oscillators” with anticommuting variables \( a_1, a_2, \ldots, a_n \). We assume

\[ \{ a_i, a_j \} = 0, \quad \{ a_i, a_j^* \} = \epsilon_{ij} \]  

(A8)

while

\[ \{ a_i, a_i^* \} = \epsilon_{ir} \]  

(A9)

Then the above analysis can be immediately applied, say to the first pair of operators, \( a_1 \) and \( a_1^* \). We may in fact cut short all remaining calculations, if we notice that if we set \( a_i^* = \epsilon_{ir} a_r \), the above equations have the usual form. The notation does not require that \( a_i^* \) be Hermitian conjugate to \( a_r \); it is simply another operator.

The result is the customary one; if the set of matrices is assumed irreducible, the span has \( 2^n \) dimensions. A basis may be chosen, in which the basis vectors are

\[ |n_1, n_2, \ldots, n_n \rangle, \]  

where each \( n_k = 0 \) or 1. These vectors satisfy

\[ a_k |0, 0, \ldots, 0 \rangle = 0, \]  

(A9)

\[ \epsilon_{kr} a_k |n_1, n_2, \ldots, n_r \rangle = n_k |n_1, n_2, \ldots, n_r \rangle, \]  

(A10)

and

\[ |n_1, n_2, \ldots, n_n \rangle = (a_1)^{n_1} (a_2)^{n_2} \cdots (a_n)^{n_n} |0, 0, \ldots, 0 \rangle. \]  

(A11)

We note in passing, that in a somewhat more general formulation (A8) could be replaced by

\[ \{ a_r, a_i \} = K_{ri}, \]  

(A12)
where the matrix $K = (K_{\gamma \nu})$ is a nonsingular Hermitian matrix. This greater generality is only fictitious; a linear transformation $b = \sum \gamma \alpha \mu \nu$, $\tilde{b} = \sum \gamma \alpha \mu \nu$, $\gamma' = \sum \gamma \alpha \mu \nu$, will change the matrix $K$ to $TKTK$. As is well known, one can always choose $T$ so that (A12) reduces to (A8), with $\epsilon_{\gamma} = \pm 1$. Although trivial, this generalization is nevertheless useful in practice, as we shall see in Appendix C.

Just as in the case of a single oscillator, the metric $\eta$ is completely determined, up to a real proportionality constant. By using (A9)–(A11), one easily sees that

$$\langle n_1', n_2', \ldots, n_r' | n_1, n_2, \ldots, n_r \rangle = \prod_{r=1}^{r} \delta_{n_r', n_r} \times \langle 0, 0, \ldots, 0 | 0, 0, \ldots, 0 \rangle.$$  \hspace{1cm} (A13)

**APPENDIX B**

We examine here the consequences of the commutation relation

$$[a, a\dagger] = -1,$$  \hspace{1cm} (B1)

which, by interchanging the roles of $a$ and $a\dagger$, is of course completely equivalent to the alternative form

$$[a, a\dagger] = +1.$$  

Just as in the case for the anticommutation relation, we need the following assumption: (i) The algebra of matrices generated by $a, a\dagger$ is irreducible, i.e., there is no invariant subspace. In addition, we assume (ii) there is at least one eigenvector $|\psi\rangle$ of the operator $a\dagger a$:

$$a\dagger a |\psi\rangle = \lambda |\psi\rangle.$$  \hspace{1cm} (B2)

(For brevity, in this section, a vector is denoted by the symbols $|\psi\rangle$ or $\psi$ as seems most convenient.)

The metric $\eta$, which is to be determined, is, as always, assumed to be Hermitian and nonsingular; therefore, there is no vector $\phi \neq 0$ such that

$$\langle \phi | = 0$$  \hspace{1cm} (B3)

for all vectors $|\psi\rangle$.

The argument that follows is fairly trivial, as it relies on the customary construction of a string of eigenvectors of $a\dagger a$,

$$\cdots, a\dagger \psi, a\dagger \phi_1, a\dagger \phi_2, a\dagger \phi_3, \ldots,$$  \hspace{1cm} (B4)

and to the eigenvalues

$$\ldots, \lambda - 2, \lambda - 1, \lambda, \lambda + 1, \lambda + 2, \ldots.$$  \hspace{1cm} (B5)

The eigenvalue equations

$$a\dagger \phi_p = (\lambda + p) \phi_p, \hspace{1cm} \phi_p = a\phi_p$$  \hspace{1cm} (B6)

and

$$a\dagger X_p = (\lambda - p) X_p, \hspace{1cm} X_p = a\dagger X_p$$

follow, of course, from (B1) by recursion, with the qualification that the string may stop on the right at $\phi_p$, if $a\dagger \phi_p = 0$ (or on the left at $X_p$, if $a\dagger X_p = 0$). In any event, by a well-known argument, the eigenvectors in (B4) form an independent system, their eigenvalues in (B5) being all different. In addition one easily sees that

$$a\phi_p = a\phi_{p+1}, \hspace{1cm} a\phi_p = a\phi_{p-1} = (\lambda + p - 1) \phi_p,$$

$$a\dagger \phi = a\dagger \phi_{p+1}, \hspace{1cm} a\dagger \phi_p = a\dagger \phi_{p-1} = (\lambda - p) \phi_p.$$  \hspace{1cm} (B7)

All these equations hold for $p = 1, 2, \ldots$ provided one defines $\phi_0 = X_0 = \phi$. Thus the string of vectors (B4) spans our invariant subspace, which according to our assumption (i) must be the whole space. Clearly the string may not stop on both sides, since (B1) cannot hold in a finite dimensional space (by the usual trace argument). We now have three possibilities:

1. The string stops at the left: $X_{p+1} = 0$ but $X_p \neq 0$.

2. The string stops at the right: $\phi_{p+1} = 0$ but $\phi_p \neq 0$ (where $p$ may also be $= 0$). But then from (B7), $0 = a\dagger \phi_{p+1} = (\lambda + p) \phi_p$; hence $\lambda = - p$.

3. The string extends to infinity in both directions: In this case, as we shall see, $\lambda$ is an arbitrary real number but not an integer.

It is now easy to see that in all three cases the metric of the space is completely determined, apart from a real proportionality constant. Two distinct vectors in (B4) are mutually orthogonal, since they belong to different eigenvalues of a self-adjoint operator. None of them can have zero norm, or we would have degeneracy, Eq. (B3). From (B6) and (B7) we have finally

$$\langle \phi_{p+1} | \phi_{p+1} \rangle = \langle \phi_p | a\dagger \phi | \phi_p \rangle = (\lambda + p) \langle \phi_p | \phi_p \rangle$$

and

$$\langle X_{p+1} | X_{p+1} \rangle = \langle X_p | a\dagger X_p | X_p \rangle = (\lambda - p - 1) \langle X_p | X_p \rangle.$$  \hspace{1cm} (B8)

These equations determine the norm of all vectors up to a common proportionality constant. They show that $\lambda$ must be real. Furthermore, if a vector in the string (B4) has a negative eigenvalue, then the vector to the right of it (if it exists) has a square norm of the opposite sign. This never occurs in case (1), which is therefore recognized as the ordinary oscillator with positive definite metric by setting $a\dagger = b$ and $a = b^\dagger$. Case (2) is likewise recognized as Dirac's or Pauli's oscillator with indefinite metric. Choosing suitable normalization factors, one can reduce Eqs. (B7) to one of the standard forms commonly employed.

We shall not mention case (1) any further. In case (2) or (3) we may (by a suitable selection of the vector called $\psi$ in the string) assume that

$$0 \leq \lambda < 1.$$  

Case (2) corresponds to $\lambda = 0$, in this case the vectors $\phi_0, \phi_1, \ldots$ do not exist, and the eigenvalues are $-2, -1, 0$. From (B8), and by choosing

$$\langle X_0 | X_0 \rangle$$

one finds (2.27), i.e.,

$$\langle | (-1)^{\delta a} \rangle$$

for all vectors $| \rangle$. 

142
In case (3), $0 < \lambda < 1$, we may define new basis vectors, setting
\[ |\lambda\rangle = \psi, \]
\[ |\lambda + \rho\rangle = |\lambda\rangle (\lambda + 1) \cdots (\lambda + \rho - 1) |\lambda\rangle^{-1/2} a^\rho |\lambda\rangle \quad (B9) \]
and
\[ |\lambda - \rho\rangle = \iota^\rho \{ (1 - \lambda) (2 - \lambda) \cdots (\rho - \lambda) \} |\lambda\rangle^{-1/2} d^\rho |\lambda\rangle. \]
Then we get, denoting by $|\mu\rangle$ one of the new basis vectors ($\mu = \lambda$ or $\lambda \pm \rho$),
\[ a |\mu\rangle = \mu^{1/2} |\mu + 1\rangle \]
and
\[ d |\mu\rangle = (\mu - 1)^{1/2} |\mu - 1\rangle. \quad (B10) \]
Notice again that, when $\mu < 0$, $\mu^{1/2} = +i |\mu|^{1/2}$ by definition. Thus, one finds
\[ (\mu + 1) |\mu + 1\rangle = (\mu/|\mu|) |\mu\rangle |\mu\rangle, \quad (B11) \]
i.e., all $|\mu\rangle$'s have the same sign for $\mu > 0$, but alternate signs for $\mu < 0$.

**APPENDIX C**

In this appendix, we discuss the explicit diagonal form of the free Hamiltonian for the $\psi_F$ field discussed in Sec. IV. According to (4.3), the free Lagrangian density is
\[ L_{trac} = -\bar{\psi} \gamma^\mu \sigma \frac{\partial}{\partial x_\mu} \psi_F, \quad (C1) \]
where
\[ M \rho^\mu = m \rho^\mu + \frac{1}{2} i \gamma^\mu \rho^\rho \tau^\rho, \quad (C2) \]
and for convenience, we have set
\[ \tau^a = \tau^a \quad \text{and} \quad \tau^b = \tau^b. \quad (C3) \]
The free Hamiltonian is then given by
\[ H_{trac} = \int d^x \bar{\psi} \gamma^\mu \sigma \partial^\mu \psi_F, \quad (C4) \]
in which, as usual, $\gamma^\mu = \beta$ and $\gamma^\rho = -i \delta^\rho_{\tau^\rho}$. It is useful to introduce the $c$ number 4-component spinor functions $u_\lambda(p, s)$ and $v_\lambda(p, s)$, defined by
\[ (a \cdot p + \beta M^s) u_\lambda(p, s) = E_\lambda u_\lambda(p, s), \]
\[ (a \cdot p + \beta M^s) v_{-\lambda}(p, s) = E_{\lambda}^* v_{-\lambda}(p, s), \quad (C5) \]
and
\[ (a \cdot p + \beta M^s) v_{-\lambda}(p, s) = -E_{\lambda}^* v_{-\lambda}(p, s), \]
where
\[ E_\lambda = (p^2 + M^2)^{1/2}, \quad (C6) \]
\[ M = m p^0 + \frac{1}{2} i \gamma^\rho p^\rho, \quad (C7) \]
and $s = \pm 1$ denotes the usual helicity. From (C5), it can be readily verified that the following orthogonality relations hold:
\[ u_\lambda^* (p, s) u_{-\lambda}(p, s) = \delta_{\lambda, -\lambda}(p, s) = 0. \quad (C8) \]
By a suitable choice of the normalization factors, one can set
\[ \sum_{\lambda} [u_\lambda(p, s) u_{-\lambda}(p, s) + v_{-\lambda}(p, s) v_{-\lambda}(p, s)] = I, \quad (C9) \]
where $I$ is the $(4 \times 4)$ unit matrix.

At any given time,
\[ \psi_F = (\psi_F^1, \psi_F^2) \]
can be expanded in terms of these spinors
\[ \psi_1 = \sum_{p, s} (2\Omega)^{-1/2} [a_s(p, s) u_s(p, s) + a_{-s}(p, s) u_{-s}(p, s)] e^{i p \cdot s}, \quad (C10) \]
and
\[ \psi_2 = \sum_{p, s} (2\Omega)^{-1/2} [a_s(p, s) u_s(p, s) - a_{-s}(p, s) u_{-s}(p, s)] e^{i p \cdot s}, \quad (C11) \]
where $\Omega$ is the volume of the system. Correspondingly, the adjoint operators are given by
\[ \bar{\psi}_1 = \sum_{p, s} (2\Omega)^{-1/2} [\bar{a}_s(p, s) u_s(p, s) + \bar{a}_{-s}(p, s) u_{-s}(p, s)] e^{-i p \cdot s}, \quad (C12) \]
and
\[ \bar{\psi}_2 = \sum_{p, s} (2\Omega)^{-1/2} [\bar{a}_s(p, s) u_s(p, s) - \bar{a}_{-s}(p, s) u_{-s}(p, s)] e^{-i p \cdot s}, \quad (C13) \]
By using (C9) and the quantization rule (4.8) (setting $\tau^a = \tau^a$), one finds
\[ \{ a_s(p, s), \bar{a}_{-s}(p', s') \} = \{ b_s(p, s), \bar{b}_{-s}(p', s') \} = \{ b_s(p, s), \bar{b}_{-s}(p', s') \} = \delta_{ss'} \delta_{pp'}, \quad (C14) \]
and all other equal-time anticommutators are zero. These anticommutators are seen to correspond to the more general type of Eq. (A12), rather than the "diagonal" case of Eqs. (A7) and (A8) of Appendix A. The relations may be diagonalized by introducing new operators
\[ a_1(p, s) = [a_s(p, s) + a_{-s}(p, s)]/\sqrt{2}, \quad (C15) \]
\[ a_2(p, s) = [a_s(p, s) - a_{-s}(p, s)]/\sqrt{2}, \quad (C16) \]
and similar expressions for the $b$ operators. Then

$$\{a_\lambda(p,s),a_\lambda'(p',s')\} = \{b_\lambda(p,s),b_\lambda'(p',s')\} = \epsilon_s \delta_{ss'} \delta_{pp'}, \quad (C14')$$

where $\rho$, $s = 1$ or 2, and $\epsilon_1 = 1$, $\epsilon_2 = -1$. All other equal-time anticommutators are zero. It then follows from the arguments given in Appendix A that the metric $\eta$ is uniquely determined, up to the transformations (2.20') and (2.22). We find, for a system consisting only of the fermion field $\psi_p$,

$$\langle \langle | (-1)^{\nu}\psi | \rangle \rangle$$

for all vectors $\langle \rangle$, where

$$N_p = -\sum_{p,s} \left[ \delta_{ss'} \delta_{pp'} + b_\lambda(p,s)b_\lambda'(p',s') \right]. \quad (C16)$$

In terms of the Fourier components $a_\lambda$ and $b_\lambda$, the free Hamiltonian becomes

$$H_{\text{free}} = \sum_{p,s} \left[ E_p \left[ a_{\lambda'}(p,s) a_{\lambda}(p,s) - b_{\lambda'}(p,s) b_{\lambda}(p,s) \right] + E_p^* \left[ a_{\lambda'}(p,s) a_{\lambda}(p,s) - b_{\lambda'}(p,s) b_{\lambda}(p,s) \right] \right]. \quad (C17)$$

We remark that, if we follow the prescriptions of Appendix A, we will naturally arrive at a basis in which the operators $a_\lambda a_\lambda$, $b_\lambda b_\lambda$ are diagonal; in this basis $a_\lambda a_\lambda$, etc., are not diagonal. We notice, however, that the metric is positive definite with respect to the degrees of freedom with $\rho = 1$ and indefinite for $\rho = 2$ [see Eq. (C14')]. Furthermore, in the basis we have described, because of (A13), the matrix representation $\eta$ of the metric, defined by (2.3) and (2.4), is diagonal, and the matrices representing $a_\lambda$, $a_\lambda$, $b_\lambda$, $b_\lambda$ (we omit the indices $p, s$ for simplicity) obey the rule

$$\bar{a}_\lambda = \epsilon_s a_\lambda^\dagger, \quad b_\lambda = \epsilon_s b_\lambda^\dagger. \quad \text{ (C18)}$$

In this basis, therefore,

$$a_\lambda = a_\lambda^\dagger, \quad b_\lambda = b_\lambda^\dagger, \quad \bar{a}_\lambda = a_\lambda^\dagger, \quad \bar{b}_\lambda = b_\lambda^\dagger, \quad \text{(C19)}$$

and Eqs. (C14) take the standard form for fermion operators,

$$\{a_\lambda(p,s),a_\lambda'(p',s')\} = \{b_\lambda(p,s),b_\lambda'(p',s')\} = \delta_{pp'} \delta_{ss'}, \quad \text{(C20)}$$

where $\lambda = + \text{ or } -$ and all other equal-time anticommutators remain zero. Similarly, the free Hamiltonian becomes

$$H_{\text{free}} = \sum_{\gamma} \left[ E_p \left[ a_{\gamma}(p,s) a_{\gamma}(p,s) - b_{\gamma}(p,s) b_{\gamma}(p,s) \right] + E_p^* \left[ a_{\gamma}(p,s) a_{\gamma}(p,s) - b_{\gamma}(p,s) b_{\gamma}(p,s) \right] \right]. \quad (C21)$$

Now, technically we are still in the basis in which $\bar{a}_\lambda a_\lambda = \epsilon_s a_\lambda^\dagger a_\lambda$ and $\bar{b}_\lambda b_\lambda = \epsilon_s b_\lambda^\dagger b_\lambda$, and therefore $H_{\text{free}}$ is not yet diagonal. However, through a familiar unitary transformation, one can easily transform the basis vectors into the eigenvectors of $a_\mu^\dagger a_\mu$, $a_\mu^\dagger a_\mu$, $b_\mu^\dagger b_\mu$, and $b_\mu^\dagger b_\mu$. Because this is a unitary transformation, the transformation law (2.20) governing the metric is the same as that for a bona fide operator. Therefore, in this new basis, Eqs. (C18)–(C21) remain valid and the free Hamiltonian (C21) is diagonal, though the matrix $\eta$ is no longer diagonal.

It is useful, especially in the presence of interactions, to define

$$\psi_{\pm} = 2^{-1/2} (\psi_1 \pm \psi_2). \quad \text{ (C22)}$$

The free Hamiltonian (C4) then becomes

$$H_{\text{free}} = \int \left[ \psi_{\pm}(-i\alpha \cdot \nabla + \beta M) \psi_{\pm} + \psi_{\pm}(-i\alpha \cdot \nabla + \beta M^*) \psi_{\mp} \right] d^4 \tau. \quad \text{ (C23)}$$

By using (4.3), one finds that the electromagnetic interaction is given by

$$H_{\text{int}} = \int e_0 (\bar{\psi}_{\gamma} \gamma \gamma_{\gamma} \psi_{\gamma} - \bar{\psi}_{\gamma} \gamma \gamma_{\gamma} \psi_{\gamma}) d^4 \tau. \quad \text{ (C24)}$$

Furthermore, according to (4.8),

$$\langle \psi_{\gamma}(p,s) \bar{\psi}_{\gamma}(p',s') \rangle = \langle \psi_{\gamma}(p,s) \bar{\psi}_{\gamma}(p',s') \rangle = \delta^4(p - p') \quad \text{(C25)}$$

and

$$\langle \psi_{\gamma}(p,s) \bar{\psi}_{\gamma}(p',s') \rangle = \langle \psi_{\gamma}(p,s) \bar{\psi}_{\gamma}(p',s') \rangle = 0. \quad \text{(C26)}$$

From these relations and Eqs. (C10)–(C13), it is easy to see that $\psi_{\gamma}$ annihilates “particles” of mass $M$ and charge $ie_0$, and creates “antiparticles” of mass $M$ and charge $-ie_0$, while $\bar{\psi}_{\gamma}$ creates “particles” of mass $M$ and charge $ie_0$, etc. By using (C8) and (C9), one can then proceed to derive the appropriate propagators and vertex functions for these fermion fields. The resulting expressions are just like the usual ones, except that the masses $M$ and $M^*$ are complex and the charges $ie_0$ and $-ie_0$ are imaginary.
Questions of Lorentz Invariance in Field Theories with Indefinite Metric

T. D. Lee and G. C. Wick
Columbia University, New York, New York 10027
(Received 1 December 1970)

It is pointed out that, contrary to the view expressed by Nakanishi, there is no real reason why one should rigidly follow the so-called "real space-momentum" prescription when "pinching" occurs in a field theory with indefinite metric. Since by adopting the modification introduced by Cutkosky et al., a relativistic and unitary S matrix can be obtained, that should be the obvious choice.

The paper by Nakanishi makes a valid technical point about a suggestion made in one of our papers, but derives from it conclusions far beyond what the author has proved. The following comments seem all the more necessary, since one of us, in a certain sense, has been directly cited as a witness (albeit perhaps unintentionally) by the closing sentences of the above-mentioned paper. Here, then, is our opinion.

The question revolves around the proper definition of integration paths in Feynman integrals in a field theory with indefinite metric; since the usual "iπ" prescription fails, one has to find a suitable generalization compatible with unitarity, relativistic invariance, etc. The problem was posed, but not solved, in our first two papers on the subject, since no attempt was made there to discuss relativistic invariance. Later on, a simple solution, or so it seemed at the time, was suggested by one of us. Let us refer to this, in view of later experience, as the "naive" prescription. Soon afterwards, it was pointed out in a paper by Cutkosky et al. that, due to pinching, this naive prescription would lead to violations of relativistic invariance in some higher-order diagrams. A manifestly covariant, but more sophisticated, prescription for defining the Feynman integrals when pinching occurs in our theory was then described by CLOP in the same paper.

Section II of Nakanishi's paper discusses again the question of relativistic invariance of the naive prescription; it shows that noninvariance difficulties can occur in some simpler graphs than the ones pointed out earlier in the CLOP paper. In our finite theory of quantum electrodynamics, the simplest case that corresponds to Nakanishi's example would be some of the fourth-order e+e− scattering diagrams; for these diagrams, violation of relativistic invariance occurs only if one includes effects due to the square of the width of the B0 quantum, which is in turn proportional to a2. Previously, a comparison between the two above-mentioned prescriptions has been made for the same type of diagrams, but only to first order in the width. To that order, these two prescriptions do give identical results. Owing to our incorrect impression that there was no discrepancy between the two prescriptions in these fourth-order diagrams, the naive form was allowed to persist in the discussion of these diagrams, given in Sec. V of our recent paper, simply because, for pedagogical reasons, it seemed easier to present. This particular discussion should now, of course, be stricken out completely; whenever pinching occurs, one should follow the CLOP prescription. Fortunately, this has no effect on any other parts of that paper, nor does it affect any of the few actual calculations that have been performed so far, such as the vacuum-polarization calculation given in Sec. VI.

In the Introduction and in Sec. III of Nakanishi's paper, however, some very general conclusions are stated, for which no real foundation has been laid. The notion of a "nonanalytic barrier" with a "definite shape" following from the Hamiltonian or Lagrangian structure of the theory seems rather nebulous. True, for a relativistic field theory with complex masses, the contour deformation required by the CLOP prescription has not been derived from the Lagrangian formulation (though heuristic arguments can be given concerning their compatibility). For that matter, neither has the naive "real space-momentum" prescription. It may be worthwhile to recall that this naive prescription was originally suggested by one of us, based only on an ad hoc application of the usual quantization rule to the Fourier components of a field with complex masses. Even for the

---

* Research supported in part by the U. S. Atomic Energy Commission.


¶ R. E. Cutkosky, P. V. Landshef, D. Olive, and J. C. Polkinghorne, Nucl. Phys. B12, 281 (1969). This paper will be referred to as CLOP hereafter. In fairness to these authors, it should be pointed out that their involvement in the present controversy is entirely unintentional, the purposes of their investigation, as clearly stated in the introduction to their paper, being rather different from ours.

— The existence of difficulties was kindly mentioned to us by R. Cutkosky prior to the appearance of the CLOP paper (private communication). Nakanishi, however, in Ref. 6 of his paper, denies the validity of the CLOP counter-examples; although we do not quite follow his reasoning, we shall not argue this (by now somewhat academic) point any further.


* See, e.g., lectures by T. D. Lee, in Proceedings of the International School of Physics "Ennio Majorana," Erice, Italy, 1970, edited by A. Zichichi (Academic, New York, to be published). The formal argument given there, which is based on Lagrangian field theory, leads automatically to the completely symmetric limit of the CLOP prescription, and thereby resolves the ambiguity encountered in the CLOP paper for the so-called "double ice cream cone" diagram.
free field, these Fourier components (i.e., plane-wave solutions) would lead to an explicitly noncovariant representation in which the three-momentum is always real, but the energy complex. In the coordinate space, the equation of a free field with complex masses is, of course, manifestly covariant; however, its plane-wave solutions diverge exponentially in the asymptotic region. (One notes that, even for those solutions that diverge only along the time direction in a specific Lorentz frame, the same solutions viewed in other systems of reference would diverge in the asymptotic region along the spatial directions as well, thus violating the condition for the validity of the Fourier theorem.) The mathematical procedure of applying the usual quantization rules to these Fourier components must, therefore, be regarded as a purely formal one; its general validity is clearly questionable. Thus, both the naive and the CLOP prescriptions are, at present, merely recipes for evaluating the $S$ matrix, connected only heuristically to the Lagrangian field theory.

Between these two prescriptions, the one that leads to a relativistically invariant unitary $S$ matrix is clearly to be preferred.\(^4\) The alternative raised by Nakanishi, either to sacrifice the Lorentz invariance or the Lagrangian field-theoretical formulation, appears to be an artificially created issue. His novel suggestion that there may be some merit in a "very slightly" noninvariant $S$ matrix, obtained by rigidly adhering to the original naive prescription, is not, at any rate, a line of thought we would like to encourage.

\(^4\) See also the remarks given near the end of Sec. V of Ref. 6.
Vacuum stability and vacuum excitation in a spin-0 field theory

T. D. Lee and G. C. Wick
Columbia University, New York, New York 10027
(Received 17 January 1974)

The theoretical possibility that in a limited domain in space the expectation value \( \langle \phi(x) \rangle \) of a neutral spin-0 field may be abnormal (that is to say quite different from its normal vacuum expectation value) is investigated. It is shown that if the \( \phi \) coupling is sufficiently large, then such a configuration can be metastable, and its physical size may become substantially greater than the usual microscopic dimension in particle physics. Furthermore, independent of the strength of the \( \phi \) coupling, if \( \langle \phi(x) \rangle \) inside a very heavy nucleus can become the minimum-energy state, at least within the tree approximation; in such a state, the “effective” nucleon mass inside the nucleus may be much lower than the normal value. Both possibilities may lead to physical systems that have not yet been observed.

I. INTRODUCTION

In a relativistic field theory, the vacuum state is defined to be the lowest energy level of the system. In analogy with other quantum-mechanical systems, however, a relativistic field may possess a degenerate lowest state. Perhaps the best known and simplest analogy is to Heisenberg’s infinite ferromagnet, in which case the degeneracy of the ground state is due to rotational invariance. The assumption of a degeneracy of the vacuum state, connected with a symmetry group of the Lagrangian, obviously has some far-reaching consequences, the most alluring of which is the possibility to “understand” that puzzling aspect of particle physics, the existence of broken symmetries. As is well known, this has given rise to a host of interesting theoretical speculations.

Besides spontaneous symmetry breaking,\(^1\) and other well-known consequences\(^2\) related to it (Goldstone bosons, the Higgs phenomenon, etc.), the assumption of vacuum degeneracy, or near degeneracy, probably has other striking consequences, which have received little attention so far. We describe in the following an investigation of various questions which arise naturally out of the virtual existence, within a given dynamical scheme, of states which could play the same role as the observed vacuum state, but are nevertheless different from it. We shall see that, depending on the details of the theory and on the values of certain physical parameters, which are not too well known experimentally, there may or may not be consequences that are just as drastic as the already-known features of this kind of theory.

All the schemes so far considered in the literature have two assumptions in common:

(a) The Lagrangian of the system is invariant (or sometimes nearly invariant) under a certain group of transformations of the field variables.

(b) In the (observed) lowest state of the system, some of the field variables have expectation values which are not invariant under the transformations of the symmetry group. Because of (a) we must envisage the existence of other possible lowest states, or nearly lowest states, in which the expectation values of some of the fields are different; such states represent the abnormal vacuum states.

This is, of course, what is referred to in the literature as degeneracy of the vacuum; at the same time we are often reminded of the essential difference between this phenomenon and the common variety of degenerate ground state encountered in finite systems: In the latter case all the states of a degenerate multiplet have the same degree of physical reality; the system can easily be induced to make transitions from one substate to the others. On the other hand, only one vacuum state is realized in our world; all the others are unphysical.

On second thought, the difference is not as profound as it seems. For the sake of clarity, and at the cost of repeating familiar things, let us recall in somewhat loose terms what is really implied. In a field theory of this type, the system possesses several “equivalent” configurations of minimum potential energy; in the observed lowest state the system performs small zero-point oscillations about one of these configurations. When the system is excited the configuration will deviate more strongly, but in any event only locally, from the basic equilibrium configuration. Fundamentally the stability of the situation is attributed to the infinite nature of the system; owing to this, the system will never
flip over as a whole from the normally observed minimum configuration to one of the others whose existence is required by the symmetry group. (As an example, the reader may recall what is usually said about the Heisenberg ferromagnet, spin waves, etc., and in particular the physical impossibility of rotating all the spins of an infinite ferromagnet simultaneously.)

Now in certain attempts at a sharp mathematical formulation of this state of affairs, it has even been asserted (perhaps on quite sound mathematical grounds) that in the limit of an infinite system one can construct a Hilbert space which contains only one "vacuum state," e.g., the observed one, and the excited states built upon it by local excitations. In this Hilbert space the physical quantities corresponding to local measurements are represented by well-defined operators; some global quantities such as the total energy or momentum are also represented, we hope, but the global generators of the group are not.

It may seem, at first sight, that in this way one has neatly thrown the abnormal degenerate vacuum states out the window, but physically it does not make so much difference, since in a certain sense they can reappear in the form of local excitations. In ferromagnetism the phenomenon is well known under the name of domains of magnetization. More generally we argue as follows: Suppose the configuration of the system flips over from the ordinary one to an abnormal equilibrium configuration, but only in a finite though large domain. As a volume effect, this will cost nothing; the difference in energy will be a relatively unimportant surface effect. In the case of a ferromagnet, for example, a very weak external field applied to a sufficiently large volume can easily cause the transition. Physical common sense suggests that any system with analogous features in the structure of the Lagrangian can exhibit similar phenomena under suitable circumstances. The absolute stability of the asymmetric vacuum state is therefore a relative thing.

In this paper we intend to investigate the general question of vacuum stability, and in particular to inquire whether it is experimentally possible in a limited domain in space to "excite" (flip) the ordinary vacuum to an abnormal one. As we shall see, our discussion can be readily extended to include also theories that have no vacuum degeneracy, but only other local minima in the field energy. For definiteness, we shall first consider the simple theory of a renormalizable spin-0 Hermitian field $\phi$. The Lagrangian density is

$$\mathcal{L} = -\frac{1}{2} \left( \frac{\partial \phi}{\partial x} \right)^2 - U(\phi) + \text{counterterms},$$

where

$$U(\phi) = \frac{1}{2} a \phi^2 + \frac{1}{3} b \phi^3 + \frac{1}{4} c \phi^4,$$

(1.2)

$\phi$ denotes the renormalized field operator, and $a, b, c$ are the appropriately defined renormalized constants. As usual, the counterterms are for renormalization purposes; their precise definitions are given in Sec. III and in Appendix A. In $U(\phi)$, the constant $c$ is assumed to be $> 0$ so that the energy spectrum has a lower bound. Through the transformation $\phi(x) \rightarrow \phi(x) + \text{constant}$, one may always assume that for the vacuum state

$$\langle \text{vac} \mid \phi(x) \mid \text{vac} \rangle = 0,$$

(1.3)

Thus, $U(\phi)$ does not contain a term linear in $\phi$. [Note that in order to maintain (1.3) there is a linear term in the counterterms.] Furthermore, since the vacuum state is assumed to be the lowest-energy state, the constant $a$ is also $> 0$. For convenience, by using the transformation $\phi(x) \rightarrow \phi(x)$, we may also choose the constant $b$ to be $> 0$. As a result, but without any loss of generality, the three constants $a, b, c$ are all assumed to be positive.

To study the question whether there are other abnormal vacuum states, i.e., either degenerate or "excited" vacuumlike states, we find it convenient to first quantize the system in a box of a finite volume $\Omega$ with the periodic boundary condition, and then let $\Omega \rightarrow \infty$ in the end. A useful concept is to define an energy density function $\mathcal{S}(\overline{\phi})$:

$$\mathcal{S}(\overline{\phi}) = \lim_{\Omega \rightarrow \infty} \Omega^{-1} (\text{minimum } \langle \mid H \mid \rangle),$$

(1.4)

where $H$ is the total Hamiltonian and the minimum is taken among all states $\mid \rangle$ under the constraint

$$\Omega^{-1} \int \langle \mid \phi(x) \mid \rangle d^3r = \overline{\phi},$$

(1.5)

The value $\overline{\phi} = 0$ is, by definition, the minimum of $\mathcal{S}(\overline{\phi})$. Furthermore, it is convenient to adjust the constant part of the counterterms in (1.1) such that at the minimum $\overline{\phi} = 0$

$$\mathcal{S}(0) = 0.$$

(1.6)

The question whether there are other, either degenerate or "excited," vacuumlike states then reduces simply to the investigation of the function $\mathcal{S}(\phi)$ for $\phi \neq 0$, which turns out to have some rather interesting properties.

As will be shown in the next section, the dependence of $\mathcal{S}(\phi)$ on $\phi$ bears a certain resemblance to the dependence of the Helmholtz free energy on the specific volume in thermodynamics. Just as in thermodynamics, when there is a phase transition, the Helmholtz free energy exhibits a
straight-line dependence on the specific volume, its slope being the negative of the pressure; here, depending on the values of the renormalized constants \(a\), \(b\), and \(c\), the function \(\delta(\bar{\phi})\) may also contain a straight section, say between \(\phi_a < \bar{\phi} < \phi_b\). The existence of such a straight section appears to be a general feature of the theory, provided that the \(\phi^3\) coupling constant \(b\) is sufficiently large. It exists even in the approximation of neglecting all loop diagrams; in such an approximation, one has \(\delta(\bar{\phi}) = U(\bar{\phi})\) outside the straight section, where \(U\) is given by (1.2). [Note that \(U(\bar{\phi})\) does not contain any straight section.] Along the straight section \(\phi_a < \bar{\phi} < \phi_b\), the system actually comprises two phases, in analogy to the phase-transition phenomenon in thermodynamics. Outside the straight section, \(\bar{\phi} < \phi_a\) or \(\bar{\phi} > \phi_b\), the system exists only in a single phase. The "true" vacuum state \(\bar{\phi} = 0\) is included in the region \(\bar{\phi} > \phi_b\), as illustrated in Fig. 1.

The inclusion of loop diagrams does not alter the general character of the energy density curve \(\delta(\bar{\phi})\). The explicit contributions of all one-loop and two-loop diagrams are given in Sec. III. From these results, one expects that the function \(\delta(\bar{\phi})\) defined in either one of the two single-phase regions, say \(\bar{\phi} < \phi_a\), can be analytically continued beyond the point \(\bar{\phi} = \phi_a\) to the region \(\bar{\phi} > \phi_a\); its analytic continuation, called \(\delta_a(\bar{\phi})\), is, of course, different from \(\delta(\bar{\phi})\) in the two-phase region. [This phenomenon is again in close analogy to the familiar gas-liquid transition in statistical mechanics; the analytic continuation of the gas (or liquid) phase is the supercooled gas (or superheated liquid) region, not the two-phase region.] Similarly, one may analytically continue the function \(\delta(\bar{\phi})\), defined in the other single-phase region \(\bar{\phi} > \phi_b\), to the region \(\bar{\phi} < \phi_b\) and call its analytic continuation \(\delta_b(\bar{\phi})\). In general, one expects the function \(\delta_a(\bar{\phi})\) to have a minimum at

\[
\bar{\phi} = \phi_{\text{vex}}, \quad (1.7)
\]

where the subscript "vex" denotes the vacuum excitation state.

In the case of the degenerate vacuum, both the true vacuum state \(\bar{\phi} = 0\) and the vacuum excitation state \(\bar{\phi} = \phi_{\text{vex}}\) appear as the end points of the straight section \(\phi_a < \bar{\phi} < \phi_b\), i.e.,

\[
\phi_a = \phi_{\text{vex}}, \quad \phi_b = 0,
\]

and because of (1.6)

\[
\delta(\phi_a) = \delta(\phi_b) = 0. \quad (1.8)
\]

From (1.2) one sees that if all loop diagrams are neglected, then the degeneracy occurs at

\[
3ac = b^2. \quad (1.9)
\]

As we shall discuss in Sec. III, there is a simple and convenient way to define the renormalization constants so that (1.9) is the exact condition for degeneracy when all the loop diagrams are also included. Consequently, in order that the absolute minimum energy level is at \(\bar{\phi} = 0\), we must have

\[
3ac \gg b^2; \quad (1.10)
\]

otherwise, the role of the states \(\bar{\phi} = 0\) and \(\bar{\phi} = \phi_{\text{vex}}\) will be interchanged.

In Sec. IV, we study the question of the lifetime of the system in the excited state \(\bar{\phi} = \phi_{\text{vex}}\). We

![Graphs of \(\delta(\bar{\phi})\) and \(J(\bar{\phi})\)](image-url)

**FIG. 1.** Examples of graphs of \(\delta(\bar{\phi})\) and \(J(\bar{\phi})\)

\(- (d\delta/d\bar{\phi})\) in the tree approximation. The two-phase region is between the points \(\alpha\) and \(\beta\); \(\phi_a\) and \(\phi_b\) are their abscissas. In (a), outside the interval \(\phi_a < \bar{\phi} < \phi_b\), \(\delta(\bar{\phi}) = U(\bar{\phi})\); inside the interval, the solid line refers to \(\delta_b\), and the dashed curve to \(U\). In (b), the two areas \(\delta(\alpha, \beta)\) and \(\delta(\beta, \phi_{\text{vex}})\) (between the dashed curve and the solid line) are equal.
shall show that in the nondegenerate case \(3ac > b^2\), as the volume \(V \to \infty\), the lifetime becomes zero. On the other hand, there may exist metastable states which satisfy approximately

\[
\langle | \phi(x) | \rangle = \phi_{\text{vac}}
\]  

(1.11)

in a finite volume \(L^3\), where \(L \gg m^{-1}\) and \(m^{-1}\) denotes the relevant microscopic length in the problem; \(m\) can be either \(O(\hbar)\) or \(O(a^{-1/2})\). Outside the volume, except over a surface region of a volume \(O(L^2 m^2)\), one has \(\langle | \phi(x) | \rangle = 0\). The excitation energy of such a state in its rest frame is given by

\[
M_{\text{exc}} = L^3 \hbar_e(\phi_{\text{vac}}) + O(L^2 m^3),
\]

(1.12)

where \(O(L^2 m^3)\) denotes the surface energy and \(\hbar_e(\phi_{\text{vac}})\) is the aforementioned analytic continuation of \(S(\phi)\). The lifetime \(\tau\) of such a state is given by

\[
\tau = L^3,
\]

(1.13)

provided \(\ln(Lm)\) is not too large, though \((Lm)\) must be \(\gg 1\). Only in the special case of a vacuum degeneracy, i.e., \(\hbar_e(\phi_{\text{vac}}) = 0\), can the size \(L\) be arbitrarily large; its rest mass is determined completely by the surface energy. In general, the ratio of the width to the rest mass of such vacuum excitations in either the degenerate or the nondegenerate case is exceedingly small, given by

\[
(\frac{\tau M_{\text{exc}}}{L^3})^{-1} \lesssim [L^3 \hbar_e(\phi_{\text{vac}}) + O(L^2 m^3)]^{-1} \ll 1.
\]

(1.14)

In Sec. V, we discuss the classical solutions corresponding to the vacuum excitations. The most interesting aspect of these solutions occurs when there is an extended external source. For definitiveness, we may treat approximately the effect of a heavy nucleus as that of an "external source," assuming that there is a strong interaction \(g\phi \gamma_\mu \gamma_5 \phi\) between the scalar field \(\phi\) and the nucleon field \(\phi\). As we shall see, within the tree approximation, if the surface energy can be neglected, then when \(g\) is sufficiently strong, or when the nuclear density is sufficiently high, the lowest-energy state becomes one in which the expectation value \(\langle \phi(x) \rangle\) inside the nucleus can be quite different from its normal vacuum expectation value (which is zero, by our convention). Furthermore, inside the nucleus the "effective" mass of the nucleon becomes \(m_N + g\langle \phi \rangle\), which can also be quite different from its normal value \(m_N\).

A concrete example of such a strong interaction is given by the well-known \(\sigma\) model. This is examined in Sec. VI. It appears that, within the tree approximation, if the mass of the \(\sigma\) particle is \(\leq 1\) GeV, there may well exist a new family of metastable, or even stable, superheavy nuclei.

By taking the zero pion mass limit, we can readily extend our discussion of the \(\sigma\) model to theories with Goldstone bosons; with some further minor modifications, it can also be applied to fields with Higgs mechanisms.

II. ENERGY DENSITY FUNCTION

To evaluate the energy density function \(S(\phi)\), defined by (1.4), we apply the standard Lagrangian multiplier method to take into account the constraint (1.5). Let \(H\) be a new Hamiltonian, defined by

\[
H = H + J \int \phi(x) d^3x,
\]

(2.1)

where \(J\) is the Lagrangian multiplier and \(H\) is the original Hamiltonian, which according to (1.1) is given by

\[
H = \int \left[ \frac{1}{2} \Pi^2 + \frac{1}{2} \langle \phi \rangle^2 + U(\phi) + \text{counterterms} \right] d^3x,
\]

(2.2)

where \(\Pi\) is the conjugate momentum of \(\phi\). Let the lowest eigenvalue of \(H\) be \(\Omega \lambda_J\), i.e.,

\[
H \left| \lambda_J \right\rangle = \Omega \lambda_J \left| \lambda_J \right\rangle,
\]

(2.3)

By using (2.1), (1.4), and (1.5), we find the energy density function \(S(\phi)\) to be given by the Legendre transformation

\[
S(\phi) = \lambda_J - J \phi,
\]

(2.4)

where

\[
\frac{\partial \lambda_J}{\partial \phi} = \frac{\partial S(\phi)}{\partial \phi},
\]

(2.5)

and

\[
J = - \frac{\partial S(\phi)}{\partial \phi}.
\]

(2.6)

To calculate \(\lambda_J\), let us decompose

\[
H = H_0 + H_1,
\]

(2.7)

where

\[
H_0 = \frac{1}{2} \int \left[ \Pi^2 + U(\phi) + \frac{1}{2} \lambda_J \phi^2 \right] d^3x,
\]

(2.8)

and regard \(H_1\) as a perturbation. The power-series expansion of \(\lambda_J\) in terms of the constants \(J, b, c\) and \(c\) can be readily derived. Following the treatment given by Coleman and Weinberg\(^4\) (which is also formally analogous to some of the analysis developed in statistical mechanics and many-body problems\(^3\)), we may reorganize the perturbation-series expansion of \(\lambda_J\) into sums of tree diagrams,
one-loop diagrams, two-loop diagrams, etc. The systematics of these loop diagrams will be given in the next section. Here we only discuss the tree approximation. It is not difficult to see that in the tree approximation $\lambda_\tau$ is given by the absolute minimum of

$$U_\tau(\phi) = J\phi + U(\phi)$$

$$= J\phi + \frac{1}{2}a\phi^2 + (3!)^{-1}b\phi^3 + (4!)^{-1}c\phi^4,$$  \hspace{1cm} (2.9)

and $\phi = \bar{\phi}$ is the minimum point. (For completeness, a proof is given in Appendix A.) At $J=0$, one has $U_\tau = U$. Since we are interested in the case where the function $U(\phi)$ in the original Lagrangian (1.1) has more than one local minimum, the $\phi^4$ coupling constant $b$ cannot be too small:

$$b^2 \geq \frac{9}{4}ac.$$  \hspace{1cm} (2.10)

On the other hand, because of our convention that the absolute minimum of $U(\phi)$ should be at $\phi = 0$, we have

$$b^2 < 3ac.$$  \hspace{1cm} (2.11)

[The apparently narrow region defined by these two inequalities may be deceptive. Actually, only (2.10) is the relevant one. If $b^2 > 3ac$, then the absolute minimum of $U$ is not at $\phi = 0$. By using the transformation $\phi = \phi$ constant, this absolute minimum can be shifted back to $\phi = 0$. Under such a transformation, only the coupling constant $c$ is invariant; the new constants $a$ and $b$ now satisfy $b^2 < 3ac$.]

Next, we consider the equation $\partial U_\tau / \partial \phi = 0$; i.e., on account of (2.9),

$$J = -\frac{\partial U}{\partial \phi} = -a\phi - \frac{1}{2}b\phi^2 - (3!)^{-1}c\phi^3,$$  \hspace{1cm} (2.12)

which at $J=0$ has three roots:

$$\phi = 0$$

and

$$\phi = \phi_\tau = \pm \frac{3}{2c} \left[ -b \pm (b^2 - \frac{9}{4}ac)^{1/2} \right].$$  \hspace{1cm} (2.13)

Among these, $\phi = 0$ is the absolute minimum of $U(\phi)$, $\phi = \phi_\tau$ is a local maximum, and $\phi = \phi_\tau$ is the other local minimum. As $J$ increases, these two minima will move, and the corresponding values of $U(\phi)$ will also change. There is a critical value $J_\tau$ at which these two minima become degenerate. As illustrated in Fig. 1, we may determine graphically the value $J = J_\tau$ by using Maxwell’s rule of equal area. The absolute minimum $\phi = \bar{\phi}$ makes a sudden jump from $\phi = \phi_\tau$ at $J = J_\tau$ to $\phi = \phi_\tau$ at $J = J_\tau$. By using (2.4) we find in the tree approximation

$$\delta(\bar{\phi}) = U(\bar{\phi})$$

in the region

$$\phi \gg \phi_\tau$$

(2.14)

But in $\phi_\tau < \phi < \phi_\tau$, $\delta(\bar{\phi})$ is a linear function of $\bar{\phi}$, which is simply the common tangent line of $U(\bar{\phi})$ at $\phi = \phi_\tau$ and $\phi_\tau$.

Such behavior is analogous to the problem of phase transition in statistical mechanics. In the statistical analog, the roles of $J$, $\bar{\phi}$, $\delta(\bar{\phi})$, and $\lambda_\tau$ are replaced by those of pressure, specific volume, Helmholtz free energy density, and Gibbs free energy density, respectively. The straight section $\phi_\tau < \phi < \phi_\tau$ denotes the two-phase region. As already noted in the Introduction, the function $\delta(\bar{\phi})$ in either one of the single-phase regions, $\phi < \phi_\tau$ or $\phi > \phi_\tau$, can be analytically continued into the two-phase region. In the tree approximation, these two analytic continuations are identical and both lead to $U(\phi)$. This is again analogous to the Van der Waals approximation used in statistical mechanics. In statistical mechanics, the analytic continuations of the thermodynamical functions of the liquid and the gas phases are respectively those of the superheated liquid and the supercooled gas, which should be different functions, but they reduce to the same expression in the Van der Waals approximation.

In the present problem, except for the degenerate vacuum case, the energy density function $\delta(\bar{\phi})$ has only one minimum at $\bar{\phi} = 0$, and that is the true vacuum state. On the other hand, if the $\phi^3$ coupling constant $b$ is not too small, the analytic continuation of $\delta(\bar{\phi})$ is expected to have another minimum at $\phi = \phi_m$, which denotes the vacuum excitation. In the above, this property has been established in the tree approximation; as we shall see in the next section, if the coupling $c$ is not too large, this property remains correct at least to every order in the loop expansion.

III. LOOP DIAGRAMS

The reduction of the perturbation-series expansion of $\delta(\bar{\phi})$ into a sum of three diagrams, one-loop diagrams, etc. has been given in Ref. 4. In this section we shall first briefly review the procedure, and then discuss some new properties.

A. Prototype diagrams

By using the free-field Hamiltonian $H_m$ defined by (2.8), as the unperturbed Hamiltonian, one can readily expand the energy density function $\delta(\bar{\phi})$ as a power series in $b$, $c$, and $\bar{\phi}$. As will be shown
in Appendix A, we may separate $\mathcal{S}(\vec{p})$ into a sum of tree diagrams and loop diagrams:

$$\mathcal{S}(\vec{p}) = [\mathcal{S}(\vec{p})]_{\text{tree}} + \sum_{\text{all}} [\mathcal{S}(\vec{p})]_{\text{loop}},$$  

(3.1)

where $[\mathcal{S}(\vec{p})]_{\text{loop}}$ represents the summation over all one-particle irreducible scattering diagrams that have $l$ loops and in which every external line carries a zero 4-momentum and contributes a factor $\mathcal{S}$ to the Feynman integral. For the tree diagrams (away from the two-phase region), one has

$$[\mathcal{S}(\vec{p})]_{\text{tree}} \mathcal{S}(\vec{p}),$$  

(3.2)

where $\mathcal{S}$ is given by (1.2), provided that the renormalized constants $a$, $b$, and $c$ in $U(\vec{p})$ are related to the appropriate scattering amplitudes at zero momentum. (See Sec. III B and Appendix A for further discussions of renormalization.)

For $l \neq 0$, it is useful to introduce $D(k)$, defined to be the propagator of the spin-0 particle moving in a given constant external field $\Phi_{\text{ext}}$, whose value happens to be given by $\Phi_{\text{ext}} = \vec{p}$. Thus, $D(k)$ is identical to a propagator of a free particle, but with its (mass) $\sqrt{\Phi}$ given by $(\Phi^2/\sqrt{\Phi})$; i.e.,

$$D(k) = -i[(k^2 + a(1 + \Delta))]^{-1},$$  

(3.3)

where

$$\Delta = \frac{\Phi}{b + \frac{1}{2} c \Phi} / a.$$  

(3.4)

Let us first consider the sum of all one-loop diagrams, and differentiate $[\mathcal{S}(\vec{p})]_{\text{one-loop}}$ with respect to $a$, keeping $b$, $c$, and $\Phi$ fixed. We obtain

$$\frac{\partial}{\partial a}[\mathcal{S}(\vec{p})]_{\text{one-loop}} = \frac{1}{2} \int (2\pi)^{-4} d^4k [D(k) + \text{subtraction term}],$$  

(3.5)

which can be readily established by first expanding both sides as a power series of $\Phi$, then noting that graphically the differentiation $\partial/\partial a$ on the one-loop diagram is just like cutting open one of its internal lines; this turns each loop diagram into a propagator diagram. Thus, diagram by diagram, both sides of (3.5) are equal. The subtraction term in (3.5) is needed to eliminate divergences. (The details of the subtraction term will be given in Sec. III B.) From Eq. (3.5), it follows that

$$[\mathcal{S}(\vec{p})]_{\text{one-loop}} = \frac{1}{2} \int (2\pi)^{-4} d^4k$$

$$\times \{ \ln [iD(k)] + \text{subtraction term} \}.$$  

(3.6)

Throughout the paper, $k^2 = k^2 \Phi$ and $d^4k$ is real. It is straightforward to express the higher-order loop diagrams in terms of $D(k)$. In this way all external lines attached to a three-point vertex and all pairs of external lines attached to a four-point vertex are implicitly accounted for. We need only consider those $l$-loop diagrams, called "prototype diagrams", in which all external lines, if they exist, must be attached separately to different four-point vertices; i.e., every three-point vertex $b \Phi^2$ connects only internal lines and every four-point vertex $c \Phi^3$ connects at most one external line to the diagram. For any given $l$, there are only a finite number of such prototype diagrams. We shall evaluate these prototype diagrams according to the standard Feynman rule, except that each internal line gives a factor $D(k)$, not $-i[(k^2 + a)]^{-1}$, to the Feynman integral. Otherwise, all the remaining factors in the Feynman integral are as usual, i.e., we assign factors $b$, $c$, and $\Phi$ respectively for a three-point vertex, a four-point vertex, and an external line. Except for the subtraction terms that are needed for renormalization purposes (and which will be discussed in Sec. III B), the function $[\mathcal{S}(\vec{p})]_{\text{one-loop}}$ for $l=1$ is simply given by the sum over the finite set of all different prototype $l$-loop diagrams. As an example, for $l=2$, there are only four different prototype diagrams; these are given by diagrams (i)-(iv) in Fig. 2. [Because of renormalization, one must combine these four diagrams together with diagrams (ii)' (iii)' and (iv)' in Fig. 2. The explicit value of these two-loop diagrams is given in Sec. III C.]

B. Renormalization

In (3.6) the integral $\int d^4k \ln iD(k)$ is quadratically divergent; therefore three subtractions are needed to eliminate the infinities. The corresponding subtraction term should be at least a quadratic function in $\Phi$. However, it is entirely a matter of choice whether or not one should also subtract the finite $\Phi^2$ and $\Phi^4$ terms from the integral. Similar ambiguities also exist for higher-order loop diagrams. This problem is closely tied to the original freedom in defining the renormalized constants $a$, $b$, and $c$. Any finite loop-diagram contribution to the $\Phi^0$ and $\Phi^4$ terms can either be included in the renormalized constants i.e., already included in the $\Phi^2$ and $\Phi^4$ terms in the original $U(\vec{p})$ function given by (1.2) or not included. If they are included, then a corresponding subtraction is necessary in the relevant loop calculation to avoid double counting, but otherwise this is not necessary. As it turns out, there is a particularly convenient way to decide on which choice to make.

Let us first consider the special case of degenerate vacuum. If $3ac = b^2$, the function $U(\Phi)$ in the original Lagrangian (1.1) is symmetric with respect to the transformation
(3.7), and that implies a degenerate vacuum. In the following, the requirement (3.8) will be imposed also for the general case, even when there is no degeneracy.

With this requirement, and the convention that \( \bar{\phi} = 0 \) denotes the true vacuum, we derive the inequality

\[
b^2 < 3ac,
\]

which is the same as (2.11), but is now valid with the inclusion of all loop-diagram corrections, not just in the tree approximation.

C. Loop expansion

In order to understand the nature of the loop expansion, we establish first the following theorem:

**Theorem 1.** At any \( l \geq 1 \), \( \left[ \langle \bar{\phi} \phi \rangle \right]_{\text{loop}} \) can be written in terms of \( l \) dimensionless functions \( F_{1,1}, F_{1,2}, \ldots F_{1,l} \), which depend only on \( \Delta \):

\[
\left[ \langle \bar{\phi} \phi \rangle \right]_{l, \text{loop}} = \frac{a^l}{\Delta} \sum_{k=1}^{l} c^{-\Delta k} \left[ a^{-1} (b^2 + 2ac) \right]^{l-k} F_{1,k}(\Delta),
\]

(3.10)

where \( \Delta \) is given by (3.4).

**Proof.** Let us consider an \( l \)-loop prototype diagram with \( N \) three-point vertices, \( M \) four-point vertices, \( E \) external lines, and \( I \) internal lines. From the explicit Feynman rules given above, it follows that the corresponding Feynman integral for \( \langle \bar{\phi} \phi \rangle \) is of the form

\[
b^k c^\Delta \langle \bar{\phi} \phi \rangle f(a, \Delta). \]

(3.11)

Since the total number of loops is given by \( l = I - N - M + 1 \) and since \( (2I + E) \) is equal to \( (3N + 4M) \), we have

\[
l = \frac{1}{2} N + M - \frac{1}{2} E + 1.
\]

(3.12)

The \( a \) dependence in (3.11) can be easily obtained from a simple dimensional analysis. Because \( \Delta \) and \( c \) are both dimensionless, \( a, b^2, \bar{\phi}^2 \), and \( [\langle \bar{\phi} \phi \rangle]^{1/2} \) are of the same dimension (mass)\(^2\), we obtain

\[
f(a, \Delta) = a^{(N + E)/2} F(\Delta),
\]

(3.13)

where \( F \) is dimensionless. For the special case of \( E = 0 \) (i.e., those prototype diagrams with no external line \( \bar{\phi} \)), by using (3.11)–(3.13) we find that the Feynman integral of such a diagram is of the form

\[
a^{\Delta^2} \langle \bar{\phi} \phi \rangle f(a, \Delta)\]

(3.14)

Now, from the definition of prototype diagrams we see that any \( E \neq 0 \) prototype diagram can be transformed into an \( E = 0 \) prototype diagram by simply replacing all four-point vertices that are
attached to external lines by three-point vertices, but keeping all internal lines and other vertices unchanged. Formally, we may represent such a replacement by

$$c\phi a = b\phi a,$$  \hspace{1cm} (3.15)

where \(\phi a\) denotes the appropriate internal line and \(\phi\) the external line. Thus, the sum over all different prototype diagrams that can be transformed into the same \(E=0\) prototype diagram through (3.15) is equal to the Feynman integral of the \(E=0\) diagram, provided we change \(b=\phi a\); therefore, (3.14) becomes

$$[S(\mathcal{F})^2]_{\text{two-loop}} = \frac{a^2}{2(32\pi)^2} \left\{ \frac{1}{2}(1+\Delta)\ln(1+\Delta) - \Delta \right\} + \frac{2a(\beta^2 + 3ac\Delta)}{(32\pi)^2} \left\{ \left(\frac{1}{2}(1+\Delta)\ln(1+\Delta) - \Delta \right)^2 - 2(1+\Delta)\ln(1+\Delta) + 2\Delta + \frac{1}{2} \Delta^2 \right\}. \hspace{1cm} (3.16)$$

Proof. The evaluation of \([S(\mathcal{F})]_{\text{two-loop}}\) follows readily from (3.6) and (3.8); the result is (3.16). (If \(b=0\), that is in the pure \(\phi^4\) theory, the above expression for \([S(\mathcal{F})]_{\text{one-loop}}\) reduces to the form derived by Coleman and Weinberg.) The two-loop prototype diagrams are listed in Fig. 2. These diagrams can be calculated according to the general rules given in the previous sections. The calculation is somewhat involved because of renormalization. The details are given in Appendix B, and the result is (3.17).

The evaluation of higher-order loop diagrams is complicated partly because of the large number of diagrams and partly because of the renormalization procedures required to eliminate infinities. For simplicity, we shall consider the special case \(c=0\). In such a case, there are only the \(b\phi\) vertices, and the theory is super-renormalizable. The Feynman integrals of the majority of the \(l\)-loop diagrams are convergent. In the following theorem, we shall restrict our discussion to these convergent diagrams, or "primitively divergent" diagrams as in the case of \(l=3\). (A primitively divergent diagram, as defined by Dyson, is one whose Feynman integral, though divergent, becomes convergent when any one of its internal momenta is held fixed; here, the only example is in \(l=3\).)

Theorem 3. If \(c=0\) and if we include only convergent, or primitively divergent, diagrams, then

$$[S(\mathcal{F})]_{\text{three-loop}} = \text{(constant)} \delta \left[ \ln(1+\Delta) - \Delta + \frac{1}{2} \Delta^2 \right] \hspace{1cm} (3.18)$$

and for \(l>3\)

$$a^2c(a^{-1}b^2 + 2ac\Delta)l^{l-1}F(\Delta).$$

Since \(M\) can vary from 0 to \(l-1\), Theorem 1 is proved.

Remarks. According to (3.9), \(b^2 \ll 3ac\); we may regard the loop expansion as a power-series expansion in \(c\), but treating \(\Delta\) and \((b^2/ac)\) [and therefore also \((b\phi/a)\) and \((c\phi^2/a)\)] as \(\ll 1\).

Theorem 2.

$$[S(\mathcal{F})]_{\text{one-loop}} = \frac{a^2}{32\pi} \left\{ \frac{1}{2}(1+\Delta)\ln(1+\Delta) - \frac{1}{2} \Delta - \frac{1}{2} \Delta^2 \right\}. \hspace{1cm} (3.16)$$

(3.17)

$$[S(\mathcal{F})]_{\text{two-loop}} = \text{(constant)} a^2 b^2 (32\pi)^2 \left\{ \frac{1}{2}(1+\Delta)\ln(1+\Delta) - \frac{1}{2} \Delta - \frac{1}{2} \Delta^2 \right\}.$$  \hspace{1cm} (3.18)

This theorem is proved in Appendix C.)

Remarks. From Theorem 2 and Theorem 3, it follows that every term in the loop expansion is singular at \(\Delta = -1\), i.e.,

$$\phi = c^{-1/2}(b^2 - 2ac)^{1/2}, \hspace{1cm} (3.20)$$

which are the points of inflection \(A\) and \(B\) of the function \(U(\phi)\), as illustrated in Fig. 1. This implies that the energy density function \(S(\phi)\) can be analytically continued from either one of the two single-phase regions, \(\phi < \phi_A\) or \(\phi > \phi_B\), to the two-phase region. Let \(S_\alpha(\phi)\) denote the analytic continuation from \(\phi < \phi_A\) and \(S_\beta(\phi)\) that from \(\phi > \phi_B\).

If the loop expansion is used, then \(S_\alpha(\phi)\) has a singularity at \(A\), and \(S_\beta(\phi)\) a singularity at \(B\). [At \(\Delta = -1\), the propagator \(D(k)\) is that of a zero-mass particle. Thus, physically, it seems reasonable that there should be such singularities for these analytic continuations, independent of the loop expansion.] The true vacuum is at \(\phi = 0\), and therefore it lies in the single phase region \(\phi > \phi_B\). The vacuum excitation \(\phi = \phi_m\) denotes the minimum of the analytic continuation \(S_\alpha(\phi)\). From Fig. 1, one sees that the point \(\phi = \phi_m\) lies in between \(\phi = \phi_A\) and \(\phi = \phi_B\) corresponding to \(A\).

In Fig. 3 we plot the modification of the \(J \leftrightarrow \phi\) curve due to the one-loop diagram for the special case \(b^2 = 3ac\). Because of the symmetry under the transformation (3.7), the two-phase region, with the inclusion of the loop-diagram correction, remains given by
but in the vacuum excitation state $\bar{\phi} = \phi_{\text{max}}$. As remarked before, only in the case of a degenerate vacuum do both $\bar{\phi} = 0$ and $\bar{\phi} = \phi_{\text{max}}$ lie on the energy density curve $\mathcal{S}(\bar{\phi})$. In the nondegenerate case, while the true vacuum state $\bar{\phi} = 0$ is on the energy density curve $\mathcal{S}(\bar{\phi})$, the vacuum excitation state $\bar{\phi} = \phi_{\text{max}}$ lies on the analytical continuation of $\mathcal{S}(\bar{\phi})$, denoted by $\mathcal{S}_a(\bar{\phi})$, as illustrated in Fig. 1.

### A. Nondegenerate case ($b^2 < 3ac$)

We assume that at time $t = 0$ the system is in the vacuum excitation state $|\bar{\phi}\rangle$ which satisfies

$$\langle |\phi(x)| \rangle = \phi_{\text{max}}$$

at every point $x$ in the volume $\Omega$. For convenience, let us take $\Omega$ to be a cube, which will be divided into $N$ smaller cubes, each of a linear size $L$; all adjacent cubes are separated by a distance $\delta$.

Hence,

$$\Omega = N(L + \delta)^3,$$

where $\delta$ is of the order of the microscopic length of the problem, but $L$ is much larger and may even be of a macroscopic dimension; e.g.,

$$\delta \sim O(a^{-1/2}) \text{ or } O(b^{-1})$$

and

$$L \gg \delta.$$  

Let $p(t)$ be the probability that at a later time $t$ the system is either in a state in which

$$\langle |\phi(x)| \rangle = \begin{cases} 0 \text{ in one of the cubes } L^3 \\ \text{arbitrary in the surface region } -O(L^2 \delta) \\ \phi_{\text{max}} \text{ outside} \end{cases}$$

or in states that differ from (4.4) by some additional high-energy quantum excitations inside the cube $L^3$ that has been singled out. In the nondegenerate vacuum case, one has $\mathcal{S}_a(\phi_{\text{max}}) > \mathcal{S}(0)$, where $\mathcal{S}_a$ denotes the analytical continuation of $\mathcal{S}$. These states can have the same energy as the initial state, provided

$$L^3 \mathcal{S}_a(\phi_{\text{max}}) = L^3 \mathcal{S}(0) + \text{excitation energy}.$$  

Since $L$ is $\gg O(a^{-1/2})$ or $O(b^{-1})$, there is a large number of such states that satisfy (4.5); their entropy is proportional to $L^3$. Thus, by using the standard calculation of transition rates, one finds

$$p(t) = 1 - \exp(-\lambda_c t),$$

where $\lambda_c \neq 0$, and at fixed $L$ and $\delta$ the probability $p(t)$ is independent of $N$. As shown in Appendix D,
a lower bound in \( \lambda_c \) can be easily estimated; we find for \( L \) sufficiently large

\[
\ln \lambda_c > -\kappa L^2, \quad (4.7)
\]

where \( \kappa \) is positive-definite and depends only on the renormalized constants \( a, b, \) and \( c. \)

Since the \( N \) cubes are arranged to be physically separated from each other, they can be regarded as independent systems. For an initial state (4.1), the probability that at a later time \( t \) the system remains in the same state is

\[
[1 - p(t)]^N = \exp(-N\lambda_c t), \quad (4.8)
\]

which, at a fixed \( L \), approaches zero as \( N \) (and therefore \( \Omega \)) becomes \( \infty \). Thus, if the vacuum excitation state extends over an infinite volume, its lifetime is zero.

However, the lifetime of a vacuum excitation in a limited volume \( v \) is quite a different matter. Let us consider a finite volume \( v \) and a surface region \( s \) that surrounds \( v \). The domain \( v + s \) is, of course, inside the bigger volume \( \Omega \) of the entire quantum system; for simplicity, one may assume \( \Omega \) to be infinite. Let the vacuum excitation be described by the state \( |\text{vex}\rangle \) that satisfies

\[
\langle \text{vex} | \phi(x) | \text{vex} \rangle = \begin{cases} 
\phi_{\text{vex}} \ln v & \text{if } v > 0 \\
0 & \text{outside } v + s \\
\text{arbitrary, though smooth,} & \text{inside } s.
\end{cases} \quad (4.9)
\]

Furthermore, we assume that in its rest system (i.e., \( |\text{vex}\rangle \) of zero 3-momentum) the shape of \( v \) is one in which the linear dimension is \( \sim O(v^{1/2}) \) in all directions. Thus, because of (1.6), the rest mass of \( |\text{vex}\rangle \) is

\[
M_{\text{vex}} = v S_0(\phi_{\text{vex}}) + \text{surface energy}. \quad (4.10)
\]

Such a state can decay through meson emissions. There are two dominant modes of decay: One is via the surface contraction, and the other is via the decay law (4.8), provided that \( v \) is sufficiently large. The latter resembles a “boiling” mechanism; we may first imagine that \( v \) is divided into \( n \) smaller volumes, \( v = n(L + \delta)^3 \), and then each smaller volume \( L^2 \) decays exponentially as \( \exp(-\lambda_c t) \). Let \( \tau_c \) and \( \tau_b \) be, respectively, the time scales for surface contraction and for boiling. It is clear that

\[
\tau_c \sim v^{-1/3} \quad \text{and} \quad \tau_b \sim (n\lambda_c)^{-1}. \quad (4.11)
\]

For \( v \) small, the decay time is determined by \( \tau_c \), and for \( v \) sufficiently large by \( \tau_b \). To have a rough idea of the critical volume size when \( \tau_\sigma = \tau_\tau \), we may use the lower bound (4.7) as an estimate of \( \lambda_c \). As shown in Appendix D, this lower bound is derived by using the WKB approximation; we may write

\[
-\ln(\lambda_c \delta) = P \sim 2L^3 \left( U(\phi) - U(\phi_{\text{vex}}) \right)^{1/2} d\phi, \quad (4.12)
\]

in which \( \delta \) is, as before, \( \sim O(\sigma^{-1/2}) \) or \( O(\delta^{-1}) \), \( U(\phi) \) is given by (1.2), and the integral extends from \( \phi_{\text{vex}} \) to \( \phi_\delta \), where \( U(\phi_\delta) = U(\phi_{\text{vex}}) \). Because \( L \gg \delta \), we expect \( P \) to be quite large, and therefore at \( \tau_c \sim \tau_b \) the critical volume \( v_c \) should also be rather large. For example, if we arbitrarily assume \( L \sim 10^5 \), \( \delta \sim 10^{-15} \) cm, and \( P \sim 10^3 \), then \( v_c \sim 1 \) mm\(^3\); the corresponding lifetime of the vacuum excitation state \( |\text{vex}\rangle \) is \( \sim 3 \times 10^{-12} \) sec. Since the theory is Lorentz-invariant, such a state can acquire a non-zero momentum; of course, its shape would then undergo a Lorentz contraction, and its lifetime a time dilatation.

B. Degenerate vacuum (\( b^2 = 3ac \))

In this case, the system is invariant under the transformation

\[
\phi \to \frac{b + c}{c} \phi - \frac{b}{c} \quad (4.13)
\]

The states \( \phi = 0 \) and \( \phi = -(2b/c) \) are therefore completely symmetrical with respect to each other. We observe that any classical path in the functional space \( (\phi(x)) \) that connects these two states must pass through a potential barrier whose height is at least proportional to \( O^2/\Omega \), where \( \Omega \) is the volume of the entire system. The transition matrix element between these two states becomes zero as \( \Omega \) approaches \( \infty \). Consequently, in an infinite volume, the states \( \phi = 0 \) and \( \phi = -(2b/c) \) are degenerate, and are both stable.

Next, we examine the lifetime of a vacuum excitation that extends over only a limited volume \( v \) (but \( \Omega \) is still \( \infty \)). Let \( |\text{vex}\rangle \) be such a vacuum excitation state defined by (4.9), where \( \phi_{\text{vex}} = -(2b/c) \). In this case, the rest mass consists of only the surface energy, and the lifetime is determined completely by surface contraction. It is not possible to have “boiling” inside \( v \), because of energy conservation. Near the surface, “boiling” is possible, but then there is no clear distinction between that and surface contraction.

In both the degenerate and the nondegenerate cases, we see that the vacuum excitation can, in principle, extend over a domain of macroscopic sizes. In the degenerate case, there is no limit to its size; the larger its dimension is, the bigger its mass but the smaller its width, and therefore the sharper the definition of the state. In the nondegenerate case, the same holds only if the “boil-
V. CLASSICAL SOLUTION

Some knowledge of the actual shape of the vacuum excitation state in space may be obtained by studying its classical solution; this is especially useful if its size may extend over a macroscopic region. For simplicity, we concentrate mainly on the degenerate case \( b^2 = 3ac \) in this section. With slight modifications, the method used below can be readily applied to the nondegenerate case \( b^2 \neq 3ac \) as well.

A. One spatial dimension

It is convenient to introduce the dimensionless variables
\[
x = a^{-1/2} \xi, \\
t = a^{-1/2} \tau,
\]
and
\[
\phi = \frac{b}{c} (\chi - 1).
\]
The wave equation for the degenerate case \( b^2 = 3ac \) in a one-dimensional space becomes
\[
-\frac{\partial^2 \chi}{\partial \xi^2} + \frac{\partial^2 \chi}{\partial \tau^2} + \frac{1}{2} \chi (1 - \chi^2) = 0.
\]
(5.2)

We first examine the time-independent solution. From (5.2) it follows that if \( \partial \chi / \partial \tau = 0 \), then
\[
\frac{dK}{d\xi} = 0,
\]
(5.3)
where
\[
K = \frac{1}{2} \left( \frac{dx}{d\xi} \right)^2 - \frac{1}{2} (1 - \chi^2) p.
\]
(5.4)

Thus, if we regard \( \xi \) as a fictitious "time," the problem becomes identical to one in elementary mechanics, in which there is a point particle at \( \chi \) moving in a potential
\[
W = -\frac{1}{2} (1 - \chi^2) p
\]
and \( K \) is the total energy of the particle. The explicit solution \( \chi = \chi(\xi) \) can then be readily obtained.

To illustrate the different types of solutions in this problem, we may consider, for example, the special case \( K = 0 \). The solutions are
\[
\chi = \pm 1
\]
(5.6)
and
\[
\chi = \pm \tanh \frac{1}{2} (\xi - \zeta_0),
\]
(5.7)
where \( \zeta_0 \) is a constant. In terms of the mechanical analog, (5.6) is the solution that the particle is at one of the two peaks of \( W \), and (5.7) is the solution such that the particle goes from one peak to the other. In the field-theory problem, the two solutions in (5.6) represent simply the two degenerate vacuum states \( \phi = 0 \) and \( \phi = -(2b/c) \). The solution in (5.7) gives the details of the transition from \( \phi = 0 \) at, say, \( x = +\infty \) to \( \phi = -(2b/c) \) at \( x = -\infty \).

Through a Lorentz transformation, the solution (5.7) can be easily transformed to one in which the transition region moves with a velocity \( u \). The explicit form is
\[
\chi = \pm \tanh \theta,
\]
(5.8)

where
\[
\theta = \frac{1}{2} (1 - u^2)^{-1/2} (\xi - ut + \text{constant}).
\]

B. Three-dimensional case

For simplicity, we consider only the spherically symmetrical solution. Again, we introduce the dimensionless variables
\[
y = a^{-1/2} \rho, \\
t = a^{-1/2} \tau,
\]
and
\[
\phi = \frac{b}{c} (\chi - 1).
\]
(5.9)

For the degenerate case \( b^2 = 3ac \), the wave equation becomes
\[
-\frac{\partial^2 \chi}{\partial \tau^2} + \frac{1}{\rho^2} \frac{\partial}{\partial \rho} \left( \rho^2 \frac{\partial \chi}{\partial \rho} \right) + \frac{1}{2} \chi (1 - \chi^2) = 0.
\]
(5.10)

For the time-independent solution \( \partial \chi / \partial \tau = 0 \), one has now, instead of (5.3),
\[
\frac{dK}{d\rho} = \frac{2}{\rho} \left( \frac{dx}{d\rho} \right)^2,
\]
(5.11)

where, as before,
\[
K = \frac{1}{2} \left( \frac{dx}{d\rho} \right)^2 - \frac{1}{2} (1 - \chi^2) p.
\]
(5.12)

Again, we may consider the mechanical analog by regarding \( \rho \) as the "time" and \( \chi \) as the "position" of a particle. The "potential" \( W \) is again given by (5.5). But now, because
\[
\frac{dK}{d\rho} \leq 0,
\]
(5.13)
the particle is in a dissipative system, with a "time"-dependent frictional force. The motion of the particle can be discussed in the standard way by plotting the \( K = (\text{constant}) \) contours in the phase space (with \( \chi \) and \( dx/d\rho \) as the coordinates). Since a regular solution at \( \rho = 0 \) implies that \( \chi(0) \) is finite and \( (dx/d\rho)_{+\infty} \) is zero, at \( \rho = 0 \) the trajectory must
begin at a point on the real axis (i.e., \( d\chi / d\rho = 0 \)) in the phase space. As \( \rho \) increases, because of (5.13), the value of \( K \) along the trajectory must keep on decreasing. From Fig. 4, one sees that the \( K = 0 \) contour divides the entire phase space into one closed region \( \delta \) and four open regions. Thus, depending on the initial value \( \chi(0) \), there are three types of solutions:

1. **Stationary solution.** If \( \chi(0) = 1 \) or \(-1 \), then at all \( \rho > 0 \)
\[
\chi(\rho) = 1 \text{ or } -1.
\]

2. **Runaway solution.** For \( \chi(0) > 1 \) or \(< -1 \), the trajectory in the phase space moves toward points at infinity as \( \rho \) increases.

3. **Spiral solution.** If \(-1 < \chi(0) < 1 \), the trajectory lies within the closed region \( \delta \) bounded by the \( K = 0 \) contour. Inside \( \delta \), the minimum \( K \) is at the origin. As illustrated by the dashed curve in Fig. 4, a typical trajectory would begin at a point on the real axis at \( \rho = 0 \), then spiral in, and eventually approach the origin as \( \rho \rightarrow \infty \).

Returning to the field-theory problem, one sees that the two stable solutions given by (5.14) correspond to the two degenerate vacuum states \( \bar{\phi} = 0 \) and \( \bar{\phi} = -(2\hbar/c) \). Both the runaway solution and the spiral solution have a field-energy content \( \int d^4x [\frac{1}{2}(\nabla \phi)^2 + U(\phi)] \) that is infinite. Thus, they are unphysical.

This situation is quite different from the one-dimensional case; as shown in the previous section, there is a time-independent solution (5.7) in which \( \chi \) is not a constant, and the solution has a finite field-energy. In three dimensions, a similar transition from \( \chi \approx -1 \) at, say, \( \rho \ll R \) to \( \chi = +1 \) at \( \rho \approx R \) gives rise to a surface energy which can always be reduced by decreasing \( R \). Thus, such a solution cannot be stable (i.e., time-independent) as in the one-dimensional case.

### C. Constant external source

It is therefore of interest to examine the three-dimensional time-independent classical solutions which may exist in the presence of an external source \( J(x) \). For example, we may assume that the spin-0 field \( \phi(x) \) is of parity +1 and interacts with a spin-\( \frac{1}{2} \) nucleon field \( \psi \) through a scalar coupling \(-g\psi^\dagger \gamma_4 \phi \). The Lagrangian density is given by
\[
\mathcal{L} = -\frac{1}{2} \left( \frac{\delta \phi}{\delta x_\mu} \right)^2 - U(\phi) - \phi^\text{\dagger} \gamma_4 \frac{\delta}{\delta x^\mu} \phi + m_x \phi^\text{\dagger} \gamma_4 \phi + \text{counterterms},
\]
where \( U(\phi) \) is given by (1.2), \( m_x \) is the physical mass of the nucleon, \( \psi^\text{\dagger} \) is the Hermitian conjugate of \( \psi \), and \( g \) is the renormalized coupling constant.

The wave equation is now of the form
\[
\frac{\partial^2 \phi}{\partial x^\mu \partial x^\mu} - \frac{dU}{d\phi} - J = 0,
\]
where (neglecting the counterterm)
\[
J = g \psi^\text{\dagger} \gamma_4 \psi.
\]

In this section, we shall assume that in regions occupied by nuclear matter, the source \( J \) is a constant. Physically, we may assume either \( g \) weak or \( m_x \) large, so that
\[
m_x^2 \gg (g \phi \text{\dagger} \phi)^2.
\]

[The case \( m_x^2 \ll (g \phi \text{\dagger} \phi)^2 \) will be considered in the next section.] Thus, when \( \phi \) changes from 0 to \( O(\phi_m) \), the coupling term \( g \phi^\text{\dagger} \gamma_4 \phi \) remains much smaller than the nucleon-mass term \( m_x \phi^\text{\dagger} \gamma_4 \phi \). The perturbation on \( \psi \) due to the variation of \( \phi \) may therefore be neglected. So far as the classical solution is concerned, we may then regard \( J(x) \) as a given function. For definiteness, we consider \( J(x) \) to resemble the nucleon distribution in, say, a spherical heavy nucleus; it will be assumed to be time-independent and of the form
\[
J(x) = \begin{cases} 0 & \text{if } \rho > R, \\ (ab/c) & \text{if } \rho < R, \end{cases}
\]
where \( \rho \) is defined by (5.9); \( R \) and \( j \) are both dimensionless constants.

By using the dimensionless variables introduced in (5.9), we find that for the degenerate vacuum case \( (b^2 = 3ac) \), the time-independent spherically symmetric equation is, as before,
\[
\frac{1}{\rho^2} \frac{d}{d\rho} \left( \rho^2 \frac{d\phi}{d\rho} \right) + \frac{\chi(1 - \chi^2)}{\rho^4} = 0
\]
in the outside region \( \rho > R \); it is

---

**FIG. 4.** Phase-space diagram for the mechanical analogy discussed in Sec. V B. Inside the region \( \delta \), the minimal \( K \) is \(-\frac{1}{2} \) at the origin. The dashed curve illustrates a spiral solution.
\[
\frac{1}{\rho} \frac{d}{d\rho} \left( \rho^2 \frac{dx}{d\rho} \right) + \frac{1}{2} \chi(1 - \chi^2) = j
\]

(5.19)
in the inside region \( \rho < R \). At \( \rho = R \), the outside and inside solutions are joined together so that \( \chi \) and \( dx/d\rho \) are both continuous. The solution is then determined by requiring \( \chi \) to be regular at the origin and at infinity.

The solutions that we are interested in are those in which \( R \) is large and \( \chi \) is nearly a constant either inside or outside \( \rho = R \); only near the boundary \( \rho \approx R \) does \( \chi \) have any significant variation. In order to have the "true" vacuum \( \phi = 0 \) at infinity, we require that in the outside region \( \chi = 1 \) as \( \rho \to \infty \); the next term in the asymptotic expansion of \( \chi \) is then exhibited in

\[
\chi = 1 - \lambda \rho^{-1} e^{-\rho},
\]

(5.20)
where \( \lambda \) is a constant. In the inside region, we require that as \( \rho \to 0 \),

\[
\chi \to \chi_0 + \epsilon \rho^{-1} \sinh(\epsilon \rho),
\]

(5.21)
where \( \epsilon \ll 1 \), \( \chi_0 \) satisfies

\[
x_0(1 - \chi_0^2) = 2j
\]

(5.22)
and

\[
\kappa^2 = \frac{1}{3}(3x_0^2 - 1).
\]

(5.23)

It can be readily verified that in the outside region the asymptotic solution (5.20) satisfies the differential equation (5.18) to first order in \( (\chi - 1) \); similarly, in the inside region the corresponding limiting solution (5.21) satisfies (5.19) to first order in \( (\chi - \chi_0) \). The exact determination of these parameters \( \lambda \) and \( \epsilon \) in terms of \( j \) and \( R \) is rather involved, but some of the general characteristics can be derived without detailed calculations.

For \( j \ll \frac{1}{R^2} \), Eq. (5.22) has three real roots, \( \chi_0 = \chi_{s,0}, \chi_{s,1}, \chi_{t,1} \), given by

\[
x_0 = \frac{2}{3} \cos \left( \frac{\sqrt{3}}{2} \right),
\]

\[
x_s = \frac{2}{3} \cos \left( \frac{\sqrt{3}}{2} \right),
\]

\[
x_t = \frac{2}{3} \cos \left( \frac{4\sqrt{3}}{3} \right),
\]

(5.24)
and

\[
\cos \delta = -3\sqrt{3} j.
\]

We choose \( \delta \approx \delta_0 \approx 0 \), and therefore \( \chi_0 = \chi_s < \chi_t < \chi_t \). By following the same argument given in Sec. VB, one can show that for \( j < 0 \) there is no solution which satisfies the desired boundary conditions (5.20) and (5.21). At \( j = 0 \), the three roots are \( \chi_0 = -1 \), \( \chi_s = +1 \), and \( \chi_t = 0 \), but there is only one solution that satisfies the boundary conditions (5.20) and (5.21): \( \chi(\rho) = 1 \) at all \( \rho \).

At a fixed \( R \), as \( j \) increases gradually from zero, the inside solution assumes (except near the surface \( \rho = R \)) the form (5.21) with \( \chi_0 = \chi_s \). Because of the continuity condition at \( \rho = R \), the value of \( \epsilon \) is \( \sim O(e^{-\epsilon \rho}) \). For \( R \gg 1 \), which is the case of physical interest for the classical solution, \( \epsilon \) is exceedingly small. Thus, \( \chi \approx \chi_s \ll 1 \) near the origin. At larger \( \rho \), the inside solution increases very slowly. It makes a rapid rise only when near the surface \( \rho = R \). At the surface, it connects with the outside solution, and then approaches unity asymptotically as \( \rho \to \infty \). According to (5.22), as \( j \) increases beyond \( j = 1/3\sqrt{3} \), the root \( \chi_0 = \chi_s \) ceases to exist, and therefore the solution disappears. Physically, this means that inside \( \rho < R \), as \( j \) increases adiabatically from zero, the state shifts from \( \phi(\rho) = 0 \) to \( \phi(\rho) < 0 \), until \( j \) reaches the value at point \( B \) in Fig. 3. Beyond that, \( \phi(\rho) \) has to make a jump to a completely different solution which represents the vacuum excitation state.

To obtain this other solution, let us first consider the case \( R \gg 1 \) and \( j \ll 1 \). We assume that the solution is approximately given by (5.21) in the region \( \rho < (R - d) \) where \( \chi_0 = \chi_s \approx -1 \) and \( d = O(1) \). In the region \( \rho > (R + d) \), we assume that the solution is approximately given by (5.20). In the transition region \( (R - d) \ll \rho < (R + d) \), we may neglect both \( R^{-1} \) and \( j \) as a zeroth approximation; thus, according to (5.7), we have

\[
\chi = \tanh \frac{1}{2}(\rho - \rho_0)
\]

(5.25)
where \( \rho_0 \) lies within the transition region. From the continuity condition, it follows that \( \epsilon = O(e^{-\epsilon \rho}) \), and therefore \( \chi \approx -1 \) in the region \( \rho < (R - d) \).

Similarly, we find \( \chi \approx +1 \) in the region \( \rho > (R + d) \). By multiplying (5.18) and (5.19) by \( dx/d\rho \) and then integrating over all space, we find

\[
\int_0^\infty \rho^{-1} \left( \frac{dx}{d\rho} \right)^2 d\rho,
\]

(5.26)
where

\[
\delta \chi = \chi(R) - \chi(0).
\]

(5.27)

In terms of the mechanical analog discussed in the previous section, (5.26) implies simply that the energy dissipated by the "frictional force" equals the work done by the "external force" \( j \). To evaluate approximately the "energy dissipation," we need only to consider the transition region. By using (5.25), we find the right-hand side of (5.26) to be approximately given by \( 4\delta R \). Since for \( j \ll 1 \) \( \delta \lambda \) is \( \ll 2 \), we derive the approximate condition

\[
j \gamma \frac{2}{3R}
\]

(5.28)
in order to have the vacuum excitation solution in-
side $\rho = R$.

Next, we examine its field-energy content

$$4\pi \int H \rho^2 d\rho,$$

where $H$ is

$$H = \frac{1}{2} \left( \frac{d}{dp} \right)^2 + \frac{1}{8} (1 - \chi^2)^2 + \begin{cases} j_\omega & \text{for } \rho < R, \\ 0 & \text{for } \rho > R. \end{cases}$$

By using the above solution, which is valid for $R \gg 1$ and $j \ll 1$, we find that to first order in $j$ the integral of the Hamiltonian density $H$ in the inside region, $\rho < R - d$, is given by

$$(4\pi) \int_0^{R-d} j_\omega \rho^2 d\rho \approx -\frac{4\pi}{3} R^2 j.$$

The energy content in the transition region is approximately given by

$$4\pi R^2 \int_{R-d}^{R-d} \left[ \frac{1}{2} \left( \frac{d}{dp} \right)^2 + \frac{1}{8} (1 - \chi^2)^2 \right] d\rho \approx \frac{8\pi}{3} R^2.$$

To the same order, we may neglect the energy content in the outside region $\rho > (R + d)$. The total field energy content is therefore

$$\frac{4\pi}{3} R^2 (2 - Rj). \quad (5.29)$$

This is to be compared with the approximate energy content

$$\frac{4\pi}{3} R^2 j \quad (5.30)$$

of the other solution ($\chi \equiv \chi_0 - 1$ inside $\rho < R$). Thus, for $R \gg 1$, by comparing (5.29) with (5.30), we find that the vacuum excitation solution has a lower energy if $j > 1/R$.

To summarize: For $R \gg 1$, as $j$ gradually increases from 0, the solution changes continuously from $\chi = 1$ everywhere to one in which $\chi \equiv \chi_0 - 1$ in the inside region $\rho < R$, but $\chi$ remains $\equiv 1$ in the outside region $\rho \gg R$. As $j$ becomes larger than $2/3R$, there appears another solution, called the vacuum excitation solution, in which $\chi \equiv \chi_0 < -1$ for $\rho \ll R$, though $\chi$ is still $\equiv +1$ for $\rho \gg R$. If $j$ becomes $>1/R$, then the vacuum excitation solution has a lower energy. When $j$ exceeds $1/3\sqrt{3}$, the vacuum excitation becomes the only form of time-independent solution.

D. External source (free nucleon gas)

We now turn to the case in which the coupling constant $g$ in (5.15) is assumed to be sufficiently strong that (5.17) may not hold. We recall that in the "true" vacuum, because of our convention (1.3), $\bar{\phi} = 0$; by definition, the nucleon mass is $m_N$. However, in states with $\bar{\phi} \neq 0$, the nucleon mass is $m_N + g\phi$. In discussing the classical equation, if the solution $\phi(x)$ is slowly varying, we may expect $\phi(x)$ to replace locally the role of $\bar{\phi}$ in the quantum-mechanical treatment. Thus, the "effective" mass of the nucleon becomes $m_N + g\phi$, which in the present case may be quite different from $m_N$. For definiteness, let us again consider the example of a heavy nucleus. Inside the nucleus, we have

$$\langle \psi^\dagger \psi \rangle = n, \quad (5.31)$$

where $n$ is the nucleon density, and $\langle \rangle$ denotes the expectation value. However, as we shall see, when $g$ is strong (or relatively speaking, $m_N$ not too large), in contrast with (5.17), $\langle \psi^\dagger \gamma_\alpha \psi \rangle$ is constant and must depend on $\phi$.

To discuss the classical solution of the spin-0 field, we shall assume the nucleons to be approximately described by a degenerate Fermi distribution, characterized by a maximum Fermi momentum $k_F$. In the simple example of an equal number of protons and neutrons, $k_F$ is given by

$$k_F = \left( \frac{3\pi^2 n}{2} \right)^{1/3}. \quad (5.32)$$

Since the classical solution $\phi(x)$ is expected to be slowly varying inside the nucleus, one may treat $m_N + g\phi(x)$ as the "effective" mass of the nucleon at $x$; the density of the kinetic energy of nucleons is therefore given by

$$U_n = \frac{2}{\sqrt{2}} \int_0^{k_F} k^2 (k^2 + M^2)^{1/2} dk$$

$$= (2\pi^2)^{-1/2} \left[ k_F (k_F^2 + M^2)^{1/2} (k_F^2 + 1/2M^2) - 1/2M^4 \frac{k_F^2 + (k_F^2 + M^2)^{1/2}}{M} \right]. \quad (5.33)$$

where $M^2 = (m_N + g\phi)^2$. The nuclear density $n$ is determined both by the usual short-range nuclear forces (generated through the exchange of high-frequency virtual mesons) and by the long-range "classical" potential $\phi(x)$ (which, in the time-independent solution, is of zero frequency). In the following, we shall consider two models: (i) the free-gas model, to be discussed in this section, and (ii) the incompressible-fluid model, which will be discussed in Sec. VE. The actual physical situation should lie somewhere in between these two extreme possibilities.
Free-nucleon-gas model

In this model, we neglect all short-range nuclear forces, as well as the electromagnetic interaction between nucleons. The nucleons are treated as a free degenerate Fermi gas moving in a classical field $\phi(x)$. To derive the time-independent field equation, we consider the minimum of the field energy $E$, defined by

$$E = \int \left[ \frac{1}{2} (\nabla \phi)^2 + U_\phi + U_\mu \right] d^3r$$

(5.34)

but subject to the constraint that the total number of nucleons $N$ is a constant, where for a system of equal number of neutrons and protons,

$$N = \frac{2}{3\pi^2} \int k_F^3 dk^3,$$

(5.35)

$U_\mu$ is given by (5.33) and $U_\phi$ is given by (1.2); i.e.,

$$U_\phi = \frac{1}{2} \alpha \phi^2 + (31)^{-1} b \phi^3 + (41)^{-1} c \phi^4.$$

(5.36)

By setting, at constant $k_F$, the variational derivative of $E$ with respect to $\phi$ equal to zero, we derive

$$-\nabla^2 \phi + \frac{d}{d\phi} U_\phi = \left( \frac{\partial}{\partial \phi} U_\mu \right)_{k_F} = 0.$$  

(5.37)

Next, let us consider the variation of $E$ with respect to $k_F$ at constant $\phi$ and under the constraint (5.35). By using the standard Lagrangian-multiplier method, we find that in order to have $E$ minimum,

$$k_F \left[ (k_F^2 + M^2)^{1/2} - \text{constant} \right] = 0,$$

(5.38)

where the constant is the Lagrangian multiplier. Thus, at any point in space, either there is no nuclear matter, hence $k_F = 0$, or since $M = m_N + g\phi$, $k_F$ is related to $\phi$ by

$$k_F^2 + (m_N + g\phi)^2 = \omega^2 m_N^2 = \text{constant},$$

(5.39)

which implies that the top energy of the degenerate Fermi sea is a constant. Together, (5.37) and (5.38) determine the classical time-independent equation for $\phi$.

The most remarkable consequence of the above field equation is the possibility that it may have solutions in which the $N$ nucleons can be bound together in a region of finite and nonzero volume, even though the nucleons are treated as free gas particles without any short-range forces. Furthermore, these solutions exhibit typical "saturation" properties; i.e., for $N$ sufficiently large, the volume is proportional to $N$ and the binding energy per nucleon is independent of $N$. In such solutions, the classical field $\phi \neq 0$ at infinity, so that, in accordance with our convention (1.3), we have the usual vacuum at infinity. However, the constant $\omega$ in (5.39) is chosen to be $< 1$, so that there can be a finite volume in space in which $g\phi$ is negative and $-m_N(1 - \omega)$. The nuclear matter will be confined in this volume, whose boundary is defined by

$$g\phi(x) = -m_N(1 - \omega) < 0.$$  

(5.40)

As we shall see, if $g$ is sufficiently large, one has

$$m_N + g\phi \equiv 0,$$

(5.41)

inside the bound volume, except in a small region near the boundary; therefore, because of (5.39), inside the volume

$$k_F = \omega m_N;$$  

(5.42)

i.e., the nuclear density $n \approx 2(3\pi^2)^{-1}(\omega m_N)^3$ is also nearly a constant inside. Furthermore, because of (5.41), the "effective" mass of the nucleon is $m_N$. Thus, the field energy $E$ for such a bound solution is given by

$$E = \left[ \frac{(\omega m_N)^2}{2\pi^2} + U_\phi \left( -\frac{m_N}{g} \right) \right] \Omega_N + \text{surface energy},$$

where $\Omega_N$ denotes the volume of the bound solution and $U_\phi(-m_N/g)$ is the value of $U_\phi$ at $\phi = -m_N/g$. Because of (5.35), $k_F \propto \Omega_N^{1/2}$. Therefore, if one neglects the surface energy, the minimum of $E$ occurs at $\Omega N = \theta E/\theta \Omega_N = 0$; i.e.,

$$\omega m_N)^3 = 6\pi^2 U_\phi(-m_N/g).$$

(5.43)

By using (5.35), one finds

$$N = 2(3\pi^2)^{-1}(\omega m_N)^2 \Omega_N + O(\Omega_N^{3/2}).$$

The minimum energy $E$ of the bound solution is given by

$$N^{-\frac{1}{2}} E = \omega m_N + O(N^{1/3}).$$

(5.44)

This is to be compared with the lowest energy $N m_N$ of the unbound solution [in which $\phi = 0$ and $k_F = 0$ everywhere, but one retains (5.35) by having the particles at infinity]. Now since, according to (5.36) $U_\phi = 0$ at $\phi = 0$, the bound solution has a lower energy than the unbound solution, provided $g$ is sufficiently large that

$$6\pi^2 U_\phi(-m_N/g) < m_N^2,$$

(5.45)

and therefore $\omega < 1$; in addition, $N$ must be sufficiently large that the surface energy can be neglected. The binding energy per nucleon is

$$(1 - \omega) m_N.$$  

[Note. One may wonder, as $\phi$ varies from 0 to a nonzero value, what happens to the energy change of the Dirac sea of negative energy states. Naively, one might expect the increment $\Delta E$ in energy density to be given by the difference

161
\[ d = \Omega^{-1} \left( \sum (p^2 + m^2)^{3/2} - \sum (p^2 + (m_n + g\phi)^2)^{3/2} \right), \]

where \( \Omega \) is the volume of the system. Since the summation extends over all negative energy states, the expression for \( d \) is \( \sim \) if \( \phi = 0 \). Actually, \( \Delta S \neq d \). By following the argument given in Sec. III, it can be shown that \( \Delta S \) should be given by the appropriate sum of all nucleon loop diagrams. If we include only the one-nucleon-loop approximation, then \( \Delta S \) is equal to the above expression \( d \), but only after a subtraction of a fourth-order polynomial in \( \phi \); its explicit expression, after summing over both proton and neutron, is

\[ \Delta S = \frac{1}{8\pi^2} (m_n + g\phi)^4 \ln \left( \frac{(m_n + g\phi)^2}{m_n^2} \right) \]

+a finite fourth-order polynomial.

The exact form of the polynomial depends on the definition of the renormalized constants \( a \), \( b \), and \( c \), just as in the case of the boson loop discussed in Sec. III B. If one wishes, one may determine the polynomial by requiring that as \( \phi \to 0 \), \( \Delta S \) is \( O(\phi^2) \); in that case

\[ \Delta S = \frac{1}{8\pi^2} \left( (m_n + g\phi)^4 \ln (m_n + g\phi)^2/m_n^2 \right) \]

\[ -2g\phi(m_n^2 + g\phi)^2 + \frac{4}{3} m_n (g\phi)^2 \]

\[ + \frac{4}{3} \left( g\phi \right)^4, \]

which gives \( \Delta S = (16\pi^2)^{-1} m_n^4 \) at \( m_n + g\phi = 0 \). In a complete treatment, one should, of course, include \( \Delta S \) in \( U_\phi (m_n^2/g) \), together with other corrections due to higher-order fermion loops as well as all boson loops. (Note that the above expression gives \( \delta \Delta S = \frac{1}{4} m_n^4 < \Phi^4 \), therefore the inequality (5.45) is not violated.) In this section, however, our discussion is restricted to the semiclassical equation, without taking into account any loop correction.

We emphasize that, unlike the other topics discussed in this paper, the existence of this rather unusual type of heavy "nucleus" is independent of the existence of another local minimum in \( U_\phi \) (besides \( \phi = 0 \)); it may occur even if the \( \phi \) coupling \( b = 0 \).

To illustrate more explicitly the details of such bound solutions, let us consider the degenerate vacuum case \( b^2 = 3ac \). In addition, for simplicity we shall also assume

\[ m_n = gb/c \]

so that both \( U_\sigma \) and \( U_\phi \) are symmetric under the transformation (3.7): \( \phi \to b/c - (\phi + b/c) \). By using the dimensionless variables introduced in (5.9), one has

\[ r = a^{1/2} \rho \quad \text{and} \quad \chi = 1 + (g\phi/m_n) \]

(5.47)

Let \( \rho = R \) be the boundary of a spherically symmetric solution, representing a heavy nucleus. Because of (5.37)-(5.39), the corresponding time-independent equation is given by

\[ \frac{1}{\rho^2} \frac{d}{dp} \left( \rho^2 \frac{d\chi}{dp} \right) + \frac{1}{2} \chi (1 - \chi^2) = 0, \quad \rho > R \]

(5.48)

\[ = j_n, \quad \rho < R \]

(5.49)

where

\[ j_n = 2\beta \chi \left( \frac{\omega^2 - \omega}{\omega^2} \right)^{1/2} \omega - \frac{1}{2} \chi^2 \ln \left( \frac{\omega + (\omega^2 - \omega^2)^{1/2}}{\chi} \right) \]

(5.50)

\[ \beta = 3g^{1/2} \]

(5.51)

and

\[ k_\rho/m_n = (\omega^2 - \omega^2)^{1/2} \]

(5.52)

The boundary \( \rho = R \) is determined by

\[ \chi (R) = \omega \]

As \( \rho \to \infty \), the asymptotic behavior of \( \chi \) is given by (5.20);

\[ \chi - 1 - \lambda \rho^{-1} e^{-\rho} \]

(5.53)

where \( \lambda \) is a constant. As \( \rho \to 0 \), we have

\[ \chi - \epsilon \rho^{-1} \sinh \kappa \rho \]

(5.54)

where

\[ \kappa^2 = 2\beta \omega^2 - \frac{1}{2} \]

(5.55)

which is assumed to be \( > 0 \). Both \( \chi \) and \( d\chi/d\rho \) are continuous at \( \rho = R \). Therefore, the constant \( \epsilon \) is \( O(e^{-\kappa R}) \); consequently, except when near the boundary, inside the nucleus \( \chi - O(e^{-\kappa R}) \approx 0 \), provided that \( R \) is large.

Let \( \chi_{\text{out}} \) and \( \chi_{\text{in}} \) be, respectively, the solutions of (5.48) and (5.49) that satisfy the conditions (5.53) and (5.54). To study how these two solutions can be connected at \( \rho = R \), we note that at \( \rho \gg 1 \),

\[ \rho^{-2} \frac{\delta}{\delta \rho} \left( \rho^2 \frac{\delta}{\delta \rho} \chi \right) \approx \frac{\delta^2 \chi}{\delta \rho^2} \]

Thus, when \( R \) is sufficiently large, (5.48) implies that at \( \rho = R \) the outside solution \( \chi_{\text{out}} \) satisfies

\[ \frac{d\chi_{\text{out}}}{d\rho} = \frac{1}{2} (1 - \chi_{\text{out}}^2) \]

(5.56)

which can be easily derived by following the same steps leading to (5.4). Similarly, from (5.49) one concludes that in the transition region near the boundary,

\[ R > \rho > (R - d) \gg 1 \]

(5.57)
the inside solution $\chi = \chi_{in}$ satisfies
\[ \frac{1}{2} \left( \frac{dx}{dp} \right)^2 - \frac{1}{2} (1 - \chi^2)^2 - \int j_\mu d\chi = \text{constant}, \] (5.58)
where
\[ \int j_\mu d\chi = \frac{1}{2} \beta \left[ \omega \left( 5 \chi^2 - 2 \omega^2 \right) (\omega^2 - \chi^2)^{1/2} - 3 \chi^4 \ln \frac{\omega + (\omega^2 - \chi^2)^{1/2}}{\chi} \right]. \] (5.59)

The width $d$ of the transition region (5.57) is $\sim O(1)$; it is chosen such that at $\rho = R - d$, $\chi_{in}$ and $d\chi_{in}/dp$ are $\approx 0$. Since at $\rho = R$, $\chi_{in} = \omega$, (5.59) is zero; we obtain for $R$ sufficiently large, at $\rho = R$,
\[ \frac{d\chi_{in}}{dp} = \left[ \frac{1}{2} \left( \frac{\beta}{2} + 3 \right) \chi_{in}^2 - \frac{1}{2} \right]^{1/2} \chi_{in}. \] (5.60)
The intersection of (5.58) and (5.60) determines $\chi$ and $d\chi/dp$ at $\rho = R$. We find
\[ \chi(R) = \omega = \left( \frac{3}{8 \beta} \right)^{1/4}, \] (5.61)
provided that $R$ is sufficiently large. This result, of course, agrees with (5.43). If we neglect the surface energy, then the binding energy per nucleon is
\[ m_N (1 - \omega). \] (5.62)
As we shall see in Sec. VI, in the $\sigma$ model the constant $\beta$ is given by
\[ \beta \approx \frac{g^2}{2 \pi^2} \left( \frac{m_N}{m_o} \right)^2. \]

[See Eq. (6.13).] This leads to a value $\beta \approx 10$ if $(4\pi)^{-1} g^2 \approx 15.7$ and $m_o \approx m_N$. The corresponding value of $\omega$ is $\approx 0.44$. In Fig. 5, the two solid curves labeled "outside (R = \infty)" and "inside (R = \infty)" refer respectively to (5.56) and (5.60) with $\beta = 10$. These are to be compared with the dashed curves for $R = 10$, determined by the numerical solutions of (5.48) and (5.49). As a further illustration, the numerical solution of $\chi(\rho)$ is plotted in Fig. 6 for $R = 20$ and $\beta = 10$; the corresponding value of $N$ is $\approx 210$ and that of $\omega$ is $\approx 0.46$, which is to be compared with the asymptotic value $\omega = 0.44$ if $R = \infty$.

E. External source (incompressible nucleon fluid)

In this model, we assume the short-range nuclear force to be so strong that the nuclear density $n$ is a constant. The nuclear matter resembles an incompressible fluid. Thus, if we retain the approximation that the nucleon density is still related to the Fermi momentum $k_F$ by (5.32) and that the kinetic energy of the nucleons remains given by

![FIG. 5. $\chi$ vs $d\chi/dp$ at $\rho = R$. The "outside" curves refer to solutions of (5.48), integrating from $\rho = \infty$ to $R$. The "inside" curves refer to solutions of (5.49), integrating from $\rho = 0$ to $R$, with $\beta = 10$.](image)

\[ (5.33), \text{then the time-independent equation for } \phi \text{ is} \]
\[ -\nabla^2 \phi + \frac{d}{d\phi} U_\phi = 0 \] (5.63)
outside the nucleus, and
\[ -\nabla^2 \phi + \frac{d}{d\phi} U_\phi + \left( \frac{\partial}{\partial \phi} U_N \right)_{k_F} = 0 \] (5.64)
inside the nucleus, which is the same as (5.37), except that instead of (5.38) we have now
\[ k_F = \text{constant}. \] (5.65)

To illustrate the main feature of the model, let us consider again the degenerate vacuum case

![FIG. 6. Numerical solution of $\chi(\rho)$ in the free nucleon gas model. The total number of nucleons is $\approx 210$ and the top Fermi energy is $\omega m_N \approx 0.46 m_N$. The nuclear radius is $R = 20$, and the nucleon density $n$ is zero outside the nucleus, but $a (\omega^2 - \chi^2)^{1/2}$ inside.](image)

for $\rho < R = 20$,
\[ n \propto \left[ \omega^2 - \chi^2 \right]^{1/2}. \]
\[ b^2 = 3ac, \text{ and let us assume that (5.46) holds. For} \]
the spherically symmetric case, in terms of the dimensionless variables \( \rho \) and \( \chi \) introduced in
(5.47), Eqs. (5.63) and (5.64) become
\[
\frac{1}{\rho^2} \frac{d}{d\rho} \left( \rho^2 \frac{dV}{d\rho} \right) + \frac{1}{2} \chi (1 - \chi^2) = 0, \quad \rho > R \quad (5.66)
\]
where \( \frac{d}{d\chi} V_N, \quad \rho < R \quad (5.67) \)
\[
V_N(\chi) = \beta \left\{ \frac{\alpha (\alpha^2 + \chi^2)^{1/2}}{\chi} \right\} - \frac{1}{4} \chi^4 \ln \left[ \frac{\alpha + (\alpha^2 + \chi^2)^{1/2}}{\chi} \right], \quad (5.68)
\]
\[ \rho = R \text{ is the radius of the nucleus, and } \alpha, \beta \text{ are both constants, given by} \]
\[ \alpha = \hbar/R \quad \text{and } \beta = 3 \hbar^2 / 2 \pi^2 c. \quad (5.69) \]
The field energy of the system is given by
\[ E = E_{\text{out}} + E_{\text{in}}, \quad (5.70) \]
where apart from a common multiplicative factor
\[
E_{\text{out}} = \int_0^R \rho^2 d\rho \left[ \frac{1}{2} \left( \frac{dV_N}{d\rho} \right)^2 + \frac{1}{4} (1 - \chi^2)^2 \right], \quad (5.71)
\]
\[
E_{\text{in}} = \int_0^R \rho^2 d\rho \left[ \frac{1}{2} \left( \frac{dV}{d\rho} \right)^2 + V(\chi) \right], \quad (5.72)
\]
and
\[ V(\chi) = \frac{1}{2} (1 - \chi^2)^2 + V_N(\chi) - V_N(1), \quad (5.73) \]
in which the constant term \(-V_N(1)\) is arbitrarily added, such that for the true vacuum \( \phi = 0, \chi = 1 \),
one has \( V(1) = 0 \). In contrast with the previous free-gas model, the nuclear radius \( R \) is predetermined
by the given constant \( n \) and the given number of nucleons. By varying \( E_{\text{out}} \) and \( E_{\text{in}} \) independently,
we derive the field equations (5.66) and (5.67).

Outside the nucleus, the solution has the same form as that in the previous section; e.g.,
the asymptotic solution remains given by (5.53) as \( \rho \to \infty \). However, as will be analyzed,
the solution \( \chi \) inside the nucleus changes its character depending on the physical parameters. In the weak-coupling limit,
as expected, the equation becomes identical to that in the constant-current model, discussed in Sec.
V C. Similar behavior also occurs in the low-nucleon-density limit, even though the coupling constant \( g \) may be strong. But when the nucleon density is sufficiently high and \( g \) is strong, the solution resembles that in the free-gas model.

We first observe that as \( g \to 0 \), the minimum of \( V \) is at \( \chi = \pm 1 + O(g) \); therefore, to zeroth order in \( g \), the current
\[
\frac{dV_N}{d\chi} = 2 \beta \left\{ \alpha (\alpha^2 + 1)^{1/2} - \ln \left[ \alpha + (\alpha^2 + 1)^{1/2} \right] \right\} \quad (5.74)
\]
is a constant inside the nucleus. Equation (5.67) reduces to the previous equation, (5.19). Next,
we consider the case where \( g \) is strong, but the nucleon density \( n \to 0 \), and therefore also \( \chi \to 0 \). At
\( \chi = 0 \), one has \( dV/d\chi = 0 \) and \( d^2 V/d\chi^2 = 2 \beta \alpha^2 - \frac{1}{2} \); consequently as \( \chi \to 0 \), the point \( \chi = 0 \) is a local
maximum of \( V \). The minimum of \( V \) remains at \( \chi = \pm 1 \); the solution then retains the character
of the constant-current model. However, when \( \alpha \) increases to
\[ 2 \beta \alpha^2 > \frac{1}{2}, \]
the point \( \chi = 0 \) becomes a local minimum of \( V \).
When the nucleon density becomes sufficiently high, \( \chi = 0 \) becomes the absolute minimum of \( V \).
Thus, it resembles the free-gas model when \( g \) is strong and nuclear density is sufficiently high.
(This is in contrast with the situation in the constant-current model, in which \( \chi = 0 \) is always the
local maximum of the field energy.) The corresponding solution inside the nucleus can be readily
obtained by using (5.67).

As \( \rho \to 0 \), the solution satisfies
\[ \chi = -\epsilon \rho^{-1} \sinh(\epsilon \rho), \quad (5.75) \]
where
\[ \kappa^2 = 2 \beta \alpha^2 - \frac{1}{2}. \quad (5.76) \]
Because \( \chi \) is continuous at \( \rho = R \), and because outside
the nucleus, according to (5.53), \( \chi \approx \epsilon \), one finds that \( \epsilon \) is \(-O(\epsilon^n \beta)\). Thus, if \( R \) is sufficiently
large, for the most part inside the nucleus, the value of \( \chi \) is near zero. As \( \rho \) approaches \( R \), \( \chi \)
begins to increase. If one neglects \( O(R^{n+1}) \), then one has for \( \rho \) near \( R \) but inside the nucleus
\[ \frac{1}{2} \left( \frac{dV_N}{d\rho} \right)^2 - V(\chi) = -V(0); \quad (5.77) \]
for \( \rho \) near \( R \) and outside the nucleus, one has
\[ \frac{1}{2} \left( \frac{dV_N}{d\rho} \right)^2 - \frac{1}{4} (1 - \chi^2)^2 = 0. \quad (5.78) \]
Consequently, at \( \rho = R \), \( \chi \) satisfies
\[ V_N(\chi) = \beta \alpha^2 + \frac{1}{2}. \quad (5.79) \]
In order for \( \chi = 0 \) to be the absolute minimum of \( V \), we must have \( V(0) < V(1) \), i.e.,
\[ V_N(1) - \beta \alpha^2 > \frac{1}{2}. \quad (5.80) \]
If \( \beta > 1 \), this inequality can be satisfied for a relatively small \( \alpha \), and therefore also a relatively
low nuclear density. Since for \( \alpha \) small \( V_\eta(1) \)
\approx \frac{1}{2} \beta a^3/\Lambda^4 \), (5.80) can be satisfied if \( \alpha \) is above a critical value \( \alpha_c \),
\[
\alpha_c = \left( \frac{3}{32\beta} \right)^{1/3},
\]
(5.81)
provided that \( \beta \) is sufficiently large; the corresponding critical density is \( \approx \alpha_c^3 \approx \beta^{-1} \).

The above discussions, after some minor changes, can be extended to cases where \( b^2 \approx 3ac \) and \( m_\mu \neq g/b/c \).

### VI. \( \sigma \) MODEL

It is not our purpose here to start a complete reinvestigation of the \( \sigma \) model\(^1\) of strong interactions; such a project clearly lies outside the scope of the present paper. However, as we shall see, there are some rather new and interesting properties in the \( \sigma \) model when a sizable chunk of nuclear matter is present; these properties are closely related to those discussed above. In this section, we shall give only a brief survey of these new features. Our discussion will be restricted to the tree approximation.

The \( \sigma \) model consists of a spin-\( \frac{1}{2} \) nucleon field \( \psi \), a spin-0 (even-parity) field \( \sigma \), and the usual pseudoscalar pion field \( \pi \). The Lagrangian density is given, apart from the counterterms for renormalization, by
\[
\mathcal{L} = -\bar{\psi}(\gamma_\mu \gamma_5 \psi) \frac{\partial}{\partial x^\mu} - \frac{1}{2} \left[ \left( \frac{\partial \sigma}{\partial x^\mu} \right)^2 + \left( \frac{\partial \pi}{\partial x^\mu} \right)^2 \right] - U_\sigma,
\]
(6.1)
where
\[
U_\sigma = \frac{1}{4} \lambda^2 \left[ (\sigma^2 + \pi^2) - (\mu/\lambda)^2 \right]^2 - C_\pi \sigma.
\]
(6.2)
For convenience, we assume the parameters \( C_\pi \), \( \mu \), and \( \lambda \) to be all positive. The minimum of the \( c \)-number function \( U_\sigma \) occurs at \( \sigma = \sigma_0 \) and \( \pi = 0 \), where \( \sigma_0 > (\mu/\lambda) \) and satisfies
\[
C_\pi = \sigma_0 (\lambda^2 \sigma_0^2 - \mu^2).
\]
(6.3)
In the tree approximation, the renormalized constants \( \lambda, \mu, \) and \( C_\pi \) are related to the physical masses \( m_\pi, m_\sigma, \) and \( m_\rho \) of the particles by
\[
m_\pi = g\sigma_0 ,
\]
\[
m_\rho = m_\sigma \sigma_0 ,
\]
\[
m_\rho^2 = \lambda^2 \sigma_0^2 - \mu^2 ,
\]
(6.4)
and
\[
m_\sigma^2 = 3\lambda^2 \sigma_0^2 - \mu^2 .
\]

The vacuum state satisfies
\[
\langle \text{vac}|\sigma(x)|\text{vac} \rangle = \sigma_0 ,
\]
(6.5)
and \( \langle \text{vac}|\pi(x)|\text{vac} \rangle = 0 \). In the \( \sigma \) model, the constant \( g \) is given by the well-known \( \pi \)-nucleon coupling, \( (4\pi)^{-1}g^2 \approx 15.7 \). The only unknown parameter is \( m_\sigma \). However, from the absence of any \( 0^+ \) resonance that has been positively identified experimentally, we may conclude \( m_\sigma \gg m_\pi \), and may perhaps be \( 11 \sim O(m_\pi) \).

We note that if \( \pi = 0 \), then \( U_\sigma \) reduces to the form (1.2) with \( \phi \propto (\sigma - \sigma_0) \). Owing to the smallness of \( m_\pi \), and therefore also of \( C_\pi \), the function \( U_\sigma \) has a local maximum at \( \sigma \) near zero and a local minimum, besides \( \sigma = \sigma_0 \) at \( \sigma \) near \( -\sigma_0 \). However, when \( \pi \) is now allowed to vary, this local minimum at \( \sigma \) near \( -\sigma_0 \) turns into a saddle point; it is connected to the absolute minimum point \( \sigma = \sigma_0 \) by a smooth path, \( \sigma^2 + \pi^2 = \sigma_0^2 \), without passing through any potential barrier. Thus, in the absence of nuclear matter, the \( \sigma \) model is quite different from the system discussed in the previous sections. On the other hand, when there is nuclear matter present in a certain region, then for a sufficiently large nuclear density and the region not too small, the \( \sigma \) model exhibits almost exactly the same property as that discussed in the previous sections.

It is convenient to introduce, similar to (5.9), the dimensionless variables
\[
\rho = \sqrt{2} \nu \rho \quad \text{and} \quad \chi = \lambda \sigma /\mu ;
\]
(6.6)
i.e., on account of (6.4),
\[
\rho = (m_\rho^2 - 3m_\pi^2)^{1/2} \eta \rho
\]
and
\[
\chi = (g\sigma/m_\pi)(m_\rho^2 - m_\pi^2)^{1/2}(m_\pi^2 - 3m_\pi^2)^{-1/2}.
\]
For simplicity, let us consider a spherical nucleus of radius \( \rho = R \). Furthermore, just as in Secs. VD and VE, we assume for the nucleons a degenerate Fermi distribution with a maximum Fermi momentum \( k_F \), given by (5.32). By following exactly the same discussion given in the previous two sections, we find that outside the nucleus the classical time-independent spherically symmetric equation for \( \sigma \) (with \( \pi = 0 \)) is
\[
\frac{1}{\rho^2} \frac{d}{d\rho} \left( \rho^2 \frac{dV_\sigma}{d\rho} \right) - \frac{d}{d\rho} V_\sigma(\chi) = 0 ,
\]
(6.7)
where
\[
V_\sigma(\chi) = \frac{1}{\chi} (1 - \chi^2)^2 - \eta \chi ,
\]
(6.8)
and where because of (6.4) \( \eta \) is given by
\[ \eta = m_s^2 (m_s^2 - m_\pi^2)^{1/2} (m_s^2 - 3 m_\pi^2)^{3/2} \ll 1. \]  
(6.9)

Inside the nucleus, the corresponding equation is

\[ \frac{1}{\rho^2} \frac{d}{d\rho} \left( \rho^2 \frac{d}{d\rho} \right) V_\rho (\chi) = j_B (\chi). \]  
(6.10)

The function \( j_B (\chi) \) depends on the nuclear model.

Under the assumption of the free-gas model, we have \( j_B = j_B \text{gas} \), where, just as in (5.50),

\[ (j_B \text{gas}) = 2 \beta x \left( \frac{\omega^2 - \chi^2}{\omega^2} \right)^{1/2} \omega \left( \frac{\omega^2 - \chi^2}{\chi} \right) \left( \frac{\omega^2 - \chi^2}{\chi} \right)^{1/2} \right)^2, \]  
(6.11)

in which \( \omega \) is a constant, related to the value of \( \chi \) at \( \rho = R \) by

\[ \chi(R) = \omega, \]  
(6.12)

and, because of (6.4),

\[ \beta = \frac{g^2 m_n^2}{2 \pi^2 (m_\sigma^2 - m_\pi^2)^2}. \]  
(6.13)

On the other hand, if we assume the incompressible-fluid model then \( j_B = j_B \text{fluid} \), where, just as in (5.67) and (5.68),

\[ (j_B \text{fluid}) = 2 \beta x \left( \frac{\alpha}{\alpha^2 + \chi^2} \right)^{1/2} \omega \left( \frac{\alpha}{\alpha^2 + \chi^2} \right)^{1/2} \right)^2 \]  
(6.14)

in which \( \beta \) is given by (6.13) and \( \alpha \) is a constant related to the Fermi momentum \( k_F \) by

\[ \alpha = \frac{k_F}{m_n} \left( \frac{m_n^2 - m_\pi^2}{m_n^2 - 3 m_\pi^2} \right)^{1/2}. \]  
(6.15)

In the limit \( m_\pi = 0 \), one has \( \eta = 0 \) and \( \alpha = m_n^{-1} k_F \); Eq. (6.10) reduces identically to either (5.49) or (5.67).

In the \( \sigma \) model, (4.21) \[ \omega = 10 (m_\pi/m_\sigma)^2. \]  
(6.16)

In the free nucleon gas model, by using (5.61) we find

\[ \omega = 0.44 (m_\pi/m_\sigma)^{1/2}. \]  
(6.17)

If we neglect the surface energy, then according to (5.62) the binding energy per nucleon is

\[ 1 - \omega m_n. \]  

Thus, in this model, if \( m_n \) is less than \( -5 \) \( m_n \) there will be a new type of stable heavy nucleus, provided that the short-range nuclear force can be neglected and that the nucleon number is sufficiently large.

If we assume the incompressible-fluid model, then the field energy is given by (5.70)-(5.72), except that \( V(\chi) \) is now

\[ V(\chi) = V_\rho (\chi) + V_\nu (\chi) - V_\gamma (1), \]  
(6.18)

where \( V_\rho \) is given by (6.8), but \( V_\nu \) remains given by (5.68). The above expression reduces to (5.73) in the limit \( m_\pi = 0 \). As noted in Sec. V E, when the nucleon density is sufficiently high, the minimum energy state of a very heavy nucleus flips from the "normal" solution (in which \( \chi \) is near unity and the nucleon mass \( m_n \)) to an "abnormal" one, in which both \( \chi \) and the effective nucleon mass are near 0. In order to produce the flip to the abnormal solution, (5.80) must be satisfied. By using (5.81) and (6.16), one finds that the critical density is approximately determined by

\[ \alpha_c = \left( \frac{k_F}{m_n} \right) = 0.21 \left( \frac{m_\sigma}{m_n} \right)^{3/2}. \]  
(6.19)

If \( m_\sigma = m_n \), then the critical density is about the usual nuclear density

\[ n_c = \left( \frac{4 \pi}{3} \left( 1.3 \times 10^{-13} \text{ cm}^3 \right) \right)^{-1}. \]  
(6.20)

If \( m_\sigma = m_n \), then the critical density \( n_c \) varies approximately as

\[ n_c \approx n_c \left( \frac{m_\sigma}{m_n} \right)^2, \]  
(6.21)

provided that \( m_\sigma \) is not too large.

In Fig. 7, the function \( V(\chi) \) is plotted for \( m_\sigma = m_n \) and the usual nuclear density \( n = n_c \), with \( m_n \neq 0 \). From the plot one sees more explicitly that under these conditions, if the nucleus is sufficiently heavy (so that surface energy can be neglected), then, as expected, the abnormal solution has an energy comparable to but slightly higher than that of the normal solution. For \( m_\sigma \\sim m_n \), one may produce the abnormal nuclear state by increasing the nucleon density through, say, high-energy collisions between very heavy nuclei. From Fig. 7 one observes that there is practically no potential barrier between the normal and the abnormal configurations once the critical density is reached; the corresponding production probability should, therefore, be not too small.

VII. REMARKS

In this paper we have investigated, among other things, the possibility that over a limited region in space the expectation value \( \langle \phi \rangle \) of a spin-0 even-parity field \( \phi(x) \) may be different from its "normal" vacuum expectation value (which can be chosen, by convention, to be zero). This investigation leads us to a study of several different physical problems, each containing some rather
FIG. 7. $V(x)$ in the $\sigma$ model for a nucleon density $n = \{4\times(1.3 \times 10^{-13} \text{ cm}^3)^{1/3}\}$ and for $n = 0$. See Eq. (6.10) for the definition of $V(x)$.

interesting properties. However, not all of them have been fully examined in this paper.

If the spin-0 field has a strong interaction with some matter field, say the nucleon field with a large coupling $g$, then whenever there is a sizable bulk of nuclear matter present there is a tendency to have $\langle \phi(x) \rangle = -(m_\sigma/g)$ in the region occupied by nuclear matter. This would reduce the "effective" nucleon mass to $\approx 0$, and thereby lower the kinetic energy of the nucleons. Within a certain range of the relevant physical parameters, this unusual solution may even become the lowest energy state. Thus, if such a strongly interacting scalar field does exist, there would be the possibility of a large class of "stable" or "metastable" super-heavy nuclei, hitherto undiscovered.

As a mathematical model, such a possibility suggests also a possible extension to the bound-state description of a single nucleon, by replacing the role of the nucleon by a nucleon, and that of nucleons by a mixture of quarks plus a suitable quark-antiquark continuum. Since the effective quark mass might be near zero inside the bound state (though heavy outside), one could hope to resolve some of the present theoretical difficulties in such a description.

If the spin-0 field has a large $\phi^2$ coupling constant $b$, then the function $U(\phi)$, defined by (1.2), can have another local minimum at $\phi = \phi_{\text{opt}} = 0$. In this case, even without the presence of nuclear matter, there could be the possibility of a pure vacuum excitation state, in which the expectation value $\langle \phi(x) \rangle = \phi_{\text{opt}}$ over an extended region in space. This leads naturally to the physical picture that the so-called vacuum actually more resembles a medium whose properties can be changed. If this is true, which of course we do not know at present, it must ultimately lead to rather striking physical consequences.

ACKNOWLEDGMENTS

We wish to thank N. H. Christ, A. H. Mueller, L. M. Lederman, and M. A. Ruderman for discussions.

APPENDIX A

In this appendix we give the details of the graphical representation of the energy density function $\delta(\phi)$, which is defined by (2.4). It is convenient to introduce the unrenormalized field operator $\phi_0$,

$$\phi_0 = Z^{1/2} \phi,$$

where $\phi$ is the renormalized field operator, as before. The Lagrangian density (1.1) may be written as

$$\mathcal{L} = -\frac{1}{2} \left( \frac{\partial \phi_0}{\partial x_\mu} \right) \cdot \left( \frac{\partial \phi_0}{\partial x_\mu} \right) - \phi_0 \delta J - \frac{1}{2} \phi_0^2 (a + \delta a) - (31)^{-1} \phi_0^2 (b + \delta b) - (41)^{-1} \phi_0 (c + \delta c),$$

where $\delta J$, $\delta a$, $\delta b$, and $\delta c$ are counterterms; together with $(Z^{1/2} - 1)$, these terms are needed to cancel the infinities.

The counterterm $\delta J$ is determined by requiring

$$\langle \text{vac} | \phi_0(x) | \text{vac} \rangle = 0.$$  

(A3)

The precise definitions $\delta a$, $\delta b$, and $\delta c$ will be given below [after Eq. (A23)].

From (2.1) and (2.2), the Hamiltonian $H_0$ may be written as the sum of a zeroth-order term $H_0$ and a perturbation term $H_1$:

$$H = H_0 + H_1,$$

$$H_0 = \frac{1}{2} \int \left[ (\nabla \phi_0)^2 + a_0 \phi_0^2 \right] d^3 \gamma,$$

$$H_1 = \int \left[ (J_0 + \delta J) \phi_0 + (31)^{-1} b_0 \phi_0^2 + (41)^{-1} c_0 \phi_0 \right] d^3 \gamma,$$

(A6)

where $\Pi_0$ is the conjugate momentum of $\phi_0$. 

167
\[ a_0 = a + \alpha a, \]
\[ b_0 = b + \delta b, \quad (A7) \]
\[ c_0 = c + \delta c, \]
and the constant \( J_0 \) is related to \( J \), introduced in (2.1), by
\[ J_0 = JZ^{-1/2}. \quad (A8) \]
Since the counterterm \( \delta J \) is determined by (A3), in which \( |\text{vac}\rangle \) is defined to be the lowest-energy eigenstate of \( H_I \) with \( J_0 = 0 \), there should be a non-zero expectation value of \( \phi(x) \) in the lowest-energy eigenstate \( |\rangle \) of \( H_I \) when \( J_0 \neq 0 \). We define
\[ \phi_0 = \Omega^{-1} \int \langle |\phi_0(x)| \rangle d^n x. \quad (A9) \]
Both \( \phi_0 \) and the corresponding lowest eigenvalue \( \lambda_J \) of \( \Omega^{-1} H_J \) may be evaluated by regarding \( H_0 \) as the unperturbed Hamiltonian and \( H_I \) as the perturbation. The perturbation series of \( \lambda_J \) is the sum of all connected Feynman graphs that have no external line. We may write
\[ \lambda_J = (\lambda_J)_{\text{tree}} + (\lambda_J)_{\text{one-loop}} + (\lambda_J)_{\text{two-loop}} + \cdots, \quad (A10) \]
in which \( (\lambda_J)_{\text{tree}} \) denotes the partial summation of all such diagrams that are trees and \( (\lambda_J)_{\text{one-loop}} \) denotes the partial summation of all such diagrams that have \( l \) loops.

From (A6), (A9), and (2.3) one sees that, keeping \( a_0 \), \( b_0 \), and \( c_0 \) fixed,
\[ \frac{\delta \lambda_J}{\delta J_0} = \phi_0. \quad (A11) \]
We recall that according to (2.4)
\[ i (\lambda_J)_{\text{tree}} = \quad + \quad \quad + \quad H + X + \cdots \]
\[ \frac{\delta}{\delta J_0} (\lambda_J)_{\text{tree}} = \quad + \quad \quad + \quad H + X + \cdots \]
\[ i \frac{\delta^2}{\delta J_0^2} (\lambda_J)_{\text{tree}} = \quad + \quad \quad + \quad H + X + \cdots \]
\[ + \quad \quad + \quad \cdots \]

FIG. 8. Tree diagrams for \( \lambda_J \) and its derivatives. All lines carry zero 4-momentum. For the Feynman amplitude, there is a factor \(-i/(2\pi)^4\) to each line, \(-i\cdot k_{\mu}\) to each one-point vertex, \(-i\cdot q_{\nu}\) to each three-point vertex, and \(-i\cdot c_0\) to each four-point vertex. Each open circle denotes a differentiation with respect to \((-i\cdot k_{\mu})\).

\[ \delta(\phi_0) = \lambda_J - J_0 \phi_0. \quad (A12) \]
Thus, keeping \( a_0 \), \( b_0 \), and \( c_0 \) fixed, we have
\[ \frac{\delta \phi}{\delta \phi_0} = -J_0 \quad (A13) \]
and
\[ \frac{\delta^2 \phi}{\delta \phi_0^2} \frac{\delta^2 \lambda_J}{\delta J_0^2} = -1. \quad (A14) \]

1. Tree diagrams

In Fig. 8, we list the sum \((\lambda_J)_{\text{tree}}\) of all the tree diagrams. In these diagrams, there is no external line. Every internal line carries a zero 4-momentum, so it gives to the Feynman amplitude a factor \(-i/(2\pi)^4\) with \( k = 0 \). Every one-point vertex gives a factor \(-i\cdot J_0\) every three-point vertex a factor \(-i\cdot b_0\) and every four-point vertex a factor \(-i\cdot c_0\).

From Fig. 8, it follows that, keeping \( a_0 \), \( b_0 \), and \( c_0 \) fixed, (A11) holds within the tree approximation, i.e., \((\delta \lambda_J)/\delta J_0)_{\text{tree}} = \phi_0\). Furthermore, in the same tree approximation, the full propagator of \( \phi_0 \) at zero 4-momentum is simply \( i(\delta^2 \lambda_J/\delta J_0^2)_{\text{tree}}\). Thus, one derives
\[ \left( \frac{\delta^2 \lambda_J}{\delta J_0^2} \right)_{\text{tree}} = -\left( k^2 + a_0 + b_0 \phi_0 + \frac{1}{2} c_0 \phi_0^2 \right). \quad (A15) \]
Because of (A14), this leads to
\[ \left( \frac{\delta^2 \phi}{\delta \phi_0^2} \right)_{\text{tree}} = a_0 + b_0 \phi_0 + \frac{1}{2} c_0 \phi_0^2. \quad (A16) \]
Again from Fig. 8, one sees that as \( J_0 \to 0 \), \((\lambda_J)_{\text{tree}} \to 0\) and \((\delta \lambda_J/\delta J_0)_{\text{tree}} \to 0\). Therefore, as \( \phi_0 \to 0 \), one must have \((\delta \phi)/\phi_0 \to 0\) and \((\delta \phi/\phi_0)_{\text{tree}} \to 0\). Consequently,
\[ \delta (\bar{\phi}_0) = \frac{1}{2} a_0 \bar{\phi}_0^2 + (31)^{-1} b_0 \bar{\phi}_0^3 + (41)^{-1} c_0 \bar{\phi}_0^4. \] (A17)

2. General expression

To find the general expression of \( \delta (\bar{\phi}_0) \), let us consider the scattering of \( n \) zero-momentum mesons whose interaction is given by the Lagrangian density (A2); \( n \) may vary from 2 to \( \infty \). We define \( [S(\bar{\phi}_0)]_{\text{loop}} \) to be the sum of all such one-particle irreducible scattering diagrams that are not trees; in these diagrams, each external line carries zero 4-momentum and gives a factor \( \bar{\phi}_0 \) to the Feynman integral. The corresponding factors for the internal line, the three-point vertex, and the four-point vertex are, respectively, \(-i(k^2 + a_0)^{-1} \), \(-ib_0\) and \(-ic_0\). (Note that there is no one-point vertex in these scattering graphs.) We shall now establish

\[ \delta (\bar{\phi}_0) = [\delta (\bar{\phi}_0)]_{\text{tree}} + [iS(\bar{\phi}_0)]_{\text{loop}}. \] (A18)

To prove this, we consider the sum (A10) and note that, similarly to (A15),

\[ \delta \frac{\partial^2 \Sigma}{\partial J_{\phi}^2} = [\Sigma(\phi)]_{\text{h.o.}} = 0, \] (A19)

where \( \Sigma(\phi) \) is the full propagator of the meson field \( \phi \) in a theory in which the Hamiltonian is given by (A4). We may write

\[ [i\Sigma(\phi)]^{-1} = [k^2 + a_0 + i\Sigma(\phi)], \] (A20)

where \( \Sigma(\phi) \) is, by definition, the sum of all proper self-energy diagrams. Let us separate in \( \Sigma(\phi) \) the \( J_0 \)-dependent part \( \Sigma_0(\phi) \) from the \( J_0 \)-independent part \( \Sigma(\phi) \):

\[ \Sigma(\phi) = \Sigma_0(\phi) + \Sigma_0(\phi), \] (A21)

where as \( J_0 = 0 \), \( \Sigma_0(\phi) = 0 \) and therefore \( \Sigma(\phi) = \Sigma_0(\phi) \). According to (A6), the dependence on \( J_0 \) is completely due to the one-point vertex. Thus, every diagram in \( \Sigma_0(\phi) \) is one-particle reducible — i.e., it is possible to separate every diagram in \( \Sigma_0(\phi) \) into two disconnected parts by cutting an internal line; one of these two disconnected parts contains the external momentum \( k_0 \), and the other does not. By repeating this cutting procedure and keeping only the part that contains \( k_0 \), we can reduce each of these diagrams to a one-particle irreducible diagram in which there is no \( J_0 \) vertex, but besides the two external lines that carry \( k_0 \), we have also other zero-momentum external lines (as the remainder of the cutting). If we assign to each of these additional zero-momentum external lines a factor \( \bar{\phi}_0 \) to the Feynman amplitude, we find that \( \Sigma_0(\phi) \), introduced in (A21), is equal to the sum over the set of all such different one-particle irreducible (proper self-energy) diagrams. In this set, for \( k \neq 0 \) every diagram has at least one zero-momentum external line. Among these diagrams, there are only two diagrams without any loop; these are simply \(-ib_0\bar{\phi}_0\) and \(-ic_0\bar{\phi}_0\). The rest all have some loops.

Next, we note that for \( k \neq 0 \) the \( J_0 \)-independent part \( \Sigma_0(\phi) \), defined in (A20), consists of all one-particle irreducible proper self-energy diagrams that do not have any zero-momentum external line. Together, \( \Sigma(k) = \Sigma_0(k) + \Sigma_0(k) \) is then the sum of all one-particle irreducible proper self-energy diagrams which may or may not have additional zero-momentum external lines. It is now straightforward to show that \( [iS(\bar{\phi}_0)]_{\text{loop}} \), defined above, is related to \( \Sigma(k) \) at \( k = 0 \) by

\[ i\Sigma(0) = \frac{b_0\bar{\phi}_0}{2} + \frac{c_0\bar{\phi}_0^2}{2} \] (A22)

By using (A14) and the boundary condition that at \( \bar{\phi}_0 = 0 \), both \( \delta \) and \( \partial \delta / \partial \bar{\phi}_0 \) vanish. We establish (A18).

Equations (A17) and (A18) still differ from (3.1) and (3.2) by being expressed in terms of \( a_0 \), \( b_0 \), \( c_0 \), and \( \bar{\phi}_0 \) rather than the corresponding renormalized quantities. We note that whatever may be the precise definitions of these renormalized quantities, the counterterms \( \phi_0 \), \( \phi_0 \), \( \phi_0 \), and \( \Sigma(k) \) can always be expressed formally as sums over the appropriate set of diagrams in which only the renormalized quantities \( a_0 \), \( b_0 \), \( c_0 \), and \( \bar{\phi}_0 \) appear.

Every one of these diagrams must have loops. By redefining "loop" to include also these loops in the counterterm, we derive (3.1) and (3.2).

3. Renormalized constants

To define the wave-function renormalization constant \( Z \), we may follow the standard procedure: Set \( J_0 = 0 \). The full propagator of the \( \phi_0 \) field becomes then

\[ \Sigma_0(\phi) = -i[k^2 + a_0 + \Sigma_0(\phi)]^{-1}, \] (A23)

where \( \Sigma_0(\phi) \) is defined in (A21). Let \( k^2 = -m_\phi^2 \) be the zero of \( [\Sigma_0(\phi)]^{-1} \). We require as \( k^2 = -m_\phi^2 \)

\[ [\Sigma_0(\phi)]^{-1} = iZ^{-1}(k^2 + m_\phi^2). \] (A24)

Thus, \( Z \) is defined and \( m_\phi \) is the physical mass of the meson. The renormalized constant \( a \) is defined by

\[ \Sigma_0(\phi) = -iZ/a \text{ as } k^2 = 0. \] (A25)

Consequently,

\[ a = [a_0 + \Sigma_0(0)] Z. \] (A26)

From (A18), (A22), and the fact that \( \Sigma(k) = \Sigma_0(k) \) as
\( \bar{\phi} - 0 \), we obtain in the same limit, \( \bar{\phi} - 0 \),
\[
\delta(\bar{\phi}) - \frac{i}{2} a \bar{\phi}^2 + O(\bar{\phi}^3).
\]
(A27)

We may expand the scattering amplitude \( S(\bar{\phi}, \phi) \) as a power series in \( \bar{\phi} \):
\[
[S(\bar{\phi}, \phi)]_{\text{loop}} = \sum_{n=2}^{\infty} (n!)^{-1} S_n \bar{\phi}^n,
\]
(A28)

In which \( n \) denotes the number of external mesons in the scattering amplitude. From (A21), (A22), and (A26), it follows that
\[
a = [a_o + i S_o] Z.
\]
(A29)

Since \( S_o \) contains a quadratically divergent Feynman integral, two counterterms \( \delta a \) and \( (Z - 1) \) are needed to render \( (A29) \) finite. The renormalized coupling constants \( b \) and \( c \) are related to the scattering amplitudes \( S_b \) and \( S_c \) by
\[
b + \text{finite term} = [b_o + i S_b] Z^{3/2}
\]
(A30)

and
\[
c + \text{finite term} = [c_o + i S_c] Z^2
\]
(A31)

Since \( S_b \) and \( Z \) both contain only logarithmically divergent integrals, one counterterm \( \delta b \) is sufficient to render \( (A30) \) finite; similarly, one counterterm \( \delta c \) is sufficient to render \( (A31) \) finite. The precise values of the finite terms in \( (A30) \) and \( (A31) \) are determined by imposing \( (3.8) \), as discussed in Sec. III B.

If one wishes, one may alter the above definition of \( Z \) by an arbitrary finite multiplicative factor,
\[
\begin{align*}
A_{(i)} &= \frac{-i b^2}{2(31)(2\pi)^3} \int \frac{d^4k d^4q}{[k^2 + a(1 + \Delta)][q^2 + a(1 + \Delta)][(k + q)^2 + a(1 + \Delta)]} + \text{subtraction term},
\end{align*}
\]
(B4)

\[
A_{(i)'} &= \frac{-i c}{2(2\pi)^3} \int \frac{d^4k}{k^2 + a(1 + \Delta)} + \text{subtraction term},
\]
(B5)

and, to the lowest order,
\[
\delta a = \frac{-i b^2}{2(2\pi)^3} \int \frac{d^4q}{(q^2 + a)^2},
\]
(B6)

where, according to \( (3.8) \), the subtraction terms must be quadratic functions of \( \Delta \). Since both integrals in \( (B4) \) and \( (B5) \) are not primitively divergent, even with the subtraction terms included, \( (i) \) and \( (i)' \) are still logarithmically divergent. It is convenient to introduce another diagram, diagram \( (v) \) in Fig. 2, in which the dashed line denotes the propagator \( -i(k^2 + a)^{-1} \). The solid line remains \( -(k^2 + a + a\Delta)^{-1} \).

Integrals in \( (B7) \) and \( (B8) \) are not primitively divergent, even with the subtraction terms included, \( (ii) \) and \( (ii)' \) are still logarithmically divergent. It is convenient to introduce another diagram, diagram \( (v) \) in Fig. 2, in which the dashed line denotes the propagator \( -i(k^2 + a)^{-1} \). The solid line remains \( -(k^2 + a + a\Delta)^{-1} \).
where \( x_i \geq 0 \) and the function \( F \) is

\[
F(y) = (1 + y) \log(1 + y) - y - \frac{1}{2} y^2.
\]

Similarly, we find

\[
A_{(i)} - A_{(y)} = \frac{iab^2}{12(16\pi^2)^3} \int \prod_{i=1}^3 dx_i \delta \left( \sum_{i=1}^3 x_i - 1 \right) (x_1 x_2 + x_2 x_3 + x_3 x_1) \left[ F(\Delta) - \sum_{j=1}^3 F(x_j, \Delta) \right].
\]

Both expressions are now finite. It is straightforward to verify that

\[
A_{(i)} + A_{(y)} = \frac{ia b^2}{4(16\pi^2)^3} \left[ F(\Delta) - G(\Delta) \right],
\]

where

\[
G(\Delta) = \frac{1}{2} (1 + \Delta)[\ln(1 + \Delta)]^2 - (1 + \Delta) \ln(1 + \Delta) + \Delta.
\]

By using Theorem 1, one sees that diagrams (iii) + (iii)' and (iv) + (iv)' are related to (ii) + (ii)' in a simple way. Their entire sum is given by (B11), provided one substitutes \( b^2 \) by \( b^2 + 2a c_{\Delta} \), but keeping \( a \) and \( \Delta \) fixed. Equation (3.17) is then established, and this completes the proof of Theorem 2.

**APPENDIX C**

To establish Theorem 3 (stated in Sec. III C) we shall consider first the case \( I = 3 \) and \( c = 0 \). One can readily verify that in this case there is no primitively divergent prototype diagram; consequently, we need only consider the convergent ones. A typical example is given by diagram (i) in Fig. 9. Let \( I \) and \( V \) be, respectively, the number of internal lines and vertices in the diagram. We have

\[
2I = 3V, \quad I = I - V + 1
\]

and therefore

\[
V = 2(I - 1).
\]

Since each vertex carries a factor \( b \), the corresponding Feynman amplitude is proportional to \( b^2 \). The dimension of the energy density function is

\[
(\text{mass})^4.
\]

Thus, from a simple dimensional consideration and by using (C1), one sees that the amplitude should be proportional to \( a^2 (b^2/a)^{I-1} \).

Now, according to the rules for the prototype diagram given in Sec. III A, the parameter \( a \) appears only in the product \( a(1 + \Delta) \); this implies that the amplitude is proportional to

\[
a^2 (b^2/a)^{I-1} (1 + \Delta)^{I-1}.
\]

Since the diagram is a convergent one, one finds the proportionality constant to be finite and independent of \( \Delta \). Equation (3.19) now follows because of (3.8).

Next, we consider the case \( I = 3 \) and \( c = 0 \). In this case, there is only one primitively divergent prototype diagram, given by (ii) in Fig. 9. By writing down explicitly the corresponding Feynman amplitude, one can readily derive (3.18). Theorem 3 is then proved.

**APPENDIX D**

In this appendix, we give an estimate of a lower bound for the decay rate \( \lambda_{\phi} \) defined in (4.6). Let us expand the field operator \( \phi(\vec{r}, t) \) in terms of the Fourier series in the volume \( L^3 \):

\[
\phi(\vec{r}, t) = q_0 + \sum_{k=0} q_k e^{ik \cdot \vec{r}}.
\]

The Lagrangian for the system inside \( L^3 \) is

\[
\int L d^3 r = \frac{1}{2} L^3 \left[ \left( \frac{dq_0}{dt} \right)^2 - U(q_0) + \cdots \right],
\]

where \( U \) is given by (1.2), and \( \cdots \) is \( q_0 \)-dependent \((k \neq 0)\). The conjugate momentum of \( q_0 \) is

\[
p_0 = L^3 \left( \frac{dq_0}{dt} \right).
\]

Therefore, the Hamiltonian is

\[
H = \frac{1}{2} [L^2 p_0^2 + L^3 U(q_0)] + \cdots.
\]

According to (4.1), at time \( t = 0 \) the system is at \( q_0 = \phi_{min} \), which is only a local minimum of \( U(q_0) \). There is a potential barrier that separates this local minimum from the absolute minimum of \( U(q_0) \), which is at \( q_0 = 0 \). To estimate the barrier-penetration probability, we shall use the WKB
method for the $q_0$ degree of freedom, but suppress all other $k \neq 0$ degrees of freedom (i.e., set $q_k = 0$ for $k \neq 0$). The result is

$$\lambda_L \sim \omega \exp \left\{ -2L^2 \int \left[ U(q_0) - U(\phi_{\text{ext}}) \right]^{1/2} dq_0 \right\},$$

(D5)

where

$$\omega^2 = a + b \phi_{\text{ext}} + \frac{1}{2} c \phi_{\text{ext}}^2.$$  

(D6)

In (D5), the integration is from $q_0 = \phi_{\text{ext}}$ to $q_0^*$, where $U(q_0^*) = U(\phi_{\text{ext}})$ and $\phi_{\text{ext}} < q_0^* < 0$. Such an estimation of $\lambda_L$ is obviously an underestimation, since by using the other $k \neq 0$ degrees of freedom one can easily show that there are other paths leading from the local minimum $q_0 = \phi_{\text{ext}}$ to regions near the absolute minimum $q_0 = 0$, but passing through a much lower potential barrier.

---

*Research supported in part by the U. S. Atomic Energy Commission.

1For a history of this subject see Y. Nambu, Fields and Quanta 1, 33 (1970).


3T. D. Lee and C. N. Yang, Phys. Rev. 87, 404 (1952); 87, 417 (1952).


Notions essentially identical to the "prototype diagram" have been used in many-body problems and statistical mechanics. See Ref. 5.

7F. J. Dyson, Phys. Rev. 75, 1736 (1949).


10For the present experimental status, see (e.g.) the review article by S. Protopopescu et al., in Experimental Meson Spectroscopy—1972, edited by K.-W. Lai and A. H. Rosenfeld (A.I.P., New York, 1972).
Feynman rules of quantum chromodynamics inside a hadron

T. D. Lee
Columbia University, New York, New York 10027
and Brookhaven National Laboratory, Upton, New York 11973
(Received 6 September 1978)

We start from quantum chromodynamics in a finite volume of linear size $L$ and examine its color-dielectric constant $\kappa_L$, especially the limit $\kappa_L$ as $L \to \infty$. By choosing as our standard $\kappa_L = 1$ when $L = \text{some hadron size } R$, we conclude that $\kappa_L$ must be <1; furthermore, from the fact that a free quark has not been observed we can estimate an upper bound: $\kappa_L < 1.3 \times 10^{-3} \alpha$ where $\alpha$ is the fine-structure constant of QCD inside the hadron. A permanent quark confinement corresponds to the limit $\kappa_L = 0$. The hadrons are viewed as small domain structures (with color-dielectric constant = 1) immersed in a perfect, or nearly perfect, color-dia-electric medium, which is the vacuum. The Feynman rules of QCD inside the hadron are derived; they are found to depend on the color-dielectric constant $\kappa_L$ of the vacuum that lies outside. We show that, when $\kappa_L \to 0$, the mass of any color-nonsinglet state becomes $\infty$, but for color-singlet states their masses and scattering amplitudes remain finite. These new Feynman rules also depend on the hadron size $R$. Only at high energy and large four-momentum transfer can such $R$ dependence be neglected and, for color-singlet states, these new rules be reduced to the usual ones.

I. INTRODUCTION

In a recent paper, it was emphasized that in order to give quantum chromodynamics a well-defined meaning, a convenient method is to first contain the whole system within a volume of size $L$. At a finite $L$, there is the usual perturbation series which is finite to every order of the renormalized coupling constant $\alpha$. Assuming that the limit $L \to \infty$ exists, one expects the existence of a long-range order in the vacuum for an infinite volume. Because of relativistic invariance, such a long-range order must be a Lorentz scalar. It is then suggested that such a long-range order can be expressed in terms of the (color) dielectric constant $\kappa$ of the vacuum. In this approach, a permanent quark confinement is simply viewed as the vacuum of an infinite volume being a perfect “dielectric” substance with its dielectric constant $\kappa \to 0$, while the “vacuum” inside a hadron is normal ($\kappa = 1$). As we shall see, such a description leads to a set of Feynman rules that is quite different from the usual ones given in the literature. The propagator of the vector gauge field inside the hadron has an explicit dependence on the dielectric constant $\kappa$ of the vacuum that lies outside the hadron. In the limit $\kappa \to 0$, the mass of any color-nonsinglet state, such as a single quark or a single vector gauge particle, becomes infinite. On the other hand, as will be shown in this paper, for color-singlet states, their masses and scattering amplitudes remain finite when $\kappa \to 0$.

In the following, we begin in Sec. II with a brief summary of what is currently known about the (color) dielectric constant of the vacuum. In order to take into account the long-range order of the vacuum and its long-wavelength fluctuations, we adopt the standard soliton description by introducing a scalar field to represent the dynamics of such collective motion. The related phenomenological Lagrangian is given in Sec. III. This enables us in Sec. IV to write down the appropriate Feynman rules and to analyze their properties. These new Feynman rules are different from the usual ones because they are applicable to configurations near a soliton (or bag) solution, while the usual ones are for perturbations near a spatially homogeneous solution, such as a pure infinite vacuum without the soliton (or bag). Since in our picture all hadrons are solitons (or bags), it is difficult to have small perturbations around such a pure vacuum. As we shall see, the Feynman rules derived here have an explicit dependence on the hadron size $R$. Only at very high energy and four-momentum transfer can one neglect the $R$ dependence and thereby reduce these new rules to the usual ones as an approximation.

In a relativistic theory, the dielectric constant $\kappa$ of the vacuum is always the inverse of its magnetic susceptibility $\mu$ because the velocity of light $c/(\kappa \mu)^{1/2}$ must equal $c$ itself. Thus, $\kappa \propto 1/\mu$, and a zero dielectric constant is the same as an infinite magnetic susceptibility. It is interesting to ask whether there exists a critical temperature at which the infinite-volume (QCD) system can undergo a phase transition in its dielectric constant (or magnetic susceptibility). This and other questions related to the long-range order, the
scale determination of the hadron mass (or radius), and the experimental limit of the (color) dielectric constant are briefly discussed in Sec. V.

II. DIELECTRIC CONSTANT

For simplicity, let us consider in this section a pure QCD system consisting of only color gauge fields \( V^a_\mu \) (\( a=1,2,\ldots,8 \)). We first contain the system within a finite volume \( L^3 \). So long as \( L \) is fixed, the (color) dielectric constant \( \kappa_L \) of the vacuum has no absolute meaning, since the transformation

\[
V^a_\mu \rightarrow \kappa^{1/2} V^a_\mu, \\
g \rightarrow \kappa^{1/2} g
\]

brings the covariant field derivative \( V^a_\mu \rightarrow \kappa^{1/2} V^a_\mu \), and therefore the Lagrangian density

\[
-\frac{1}{4} V^a_\mu V^a_\nu \rightarrow -\frac{1}{4} \kappa V^a_\mu V^a_\nu,
\]

where \( g \) is the renormalized coupling constant, the subscripts \( \mu, \nu \) denote the space-time indices, and the superscript \( a \) is the color index. But when we consider two different volumes, say of sizes \( L \) and \( L' \), the ratio \( \kappa_L/\kappa_L \) is, of course, physically meaningful. Because of (1), this ratio is equal to the inverse of the corresponding ratio of the square of the renormalized coupling constants. By using the well-known properties of the \( \beta \) function, one can readily derive

\[
\kappa_L < \kappa_L \text{ if } L' > L.
\]

Furthermore, if

\[
\beta(g) = 0 \text{ only at } g = 0,
\]

then

\[
\lim_{L \rightarrow \infty} \kappa_L = 0.
\]

Throughout our discussions, we assume \( \kappa_L \) is a smooth function of \( L \), and that the limit \( L \rightarrow \infty \) exists. For completeness, a proof is given in Appendix A. [Although the mathematics given there is essentially identical to that used in deriving asymptotic freedom, its application to the color dielectric constant focuses on a new aspect of the physics involved. As we shall see, this leads to the conclusion that domain structures (soliton solutions) should develop whenever there are quarks and antiquarks present; thereby one bypasses the usual difficulty of strong coupling in the so-called "infrared slavery."]

For small \( L \), because \( g^2 \ll 1 \), \( \kappa_L \) is clearly \( \neq 0 \). As a convention, we may define when \( L = \) some hadronic radius \( R \),

\[
\kappa_R = 1.
\]

Let

\[
\kappa = \lim_{L \rightarrow \infty} \kappa_L.
\]

From (3) and (5), it follows that the dielectric constant of the vacuum for an infinite volume must be less than unity, i.e.,

\[
\kappa < 1.
\]

Furthermore, \( \kappa_L \) is zero if (4) holds. In the following, we assume

\[
\kappa = 0
\]

or at least

\[
\kappa \ll 1.
\]

In the former, the vacuum for an infinite system is a perfect dia-electric, in the latter, a nearly perfect dia-electric. By following the arguments given in Sec. II of Ref. 1 (and as will also be shown in the following), one sees that the mass of any color-nonsinglet state diverges when \( \kappa \rightarrow 0 \). Thus, (8a) implies a permanent quark confinement, while (8b) implies a nearly permanent confinement.

At present, it is not known whether \( \beta(g) = 0 \) has only a real root at \( g = 0 \). From (7), we know that \( \kappa_L \) must be \( < 1 \). But as yet, we are not able to decide by pure theoretical deduction how small \( \kappa_L \) actually is.\(^6\) Now the fact that a single free quark has never been observed puts a lower limit on its mass \( m_q \). As will be shown in Sec. V, if we set \( m_q > 5 \text{ GeV} \), which is a rather lenient lower bound of the free quark mass, then we can estimate an upper limit of \( \kappa_L \):

\[
\kappa = 1.3 \times 10^{-6} \alpha,
\]

where \( \alpha = (4\pi)^4 \beta \), with \( R \) chosen such that \( \kappa_R = 1 \), in accordance with our convention (5). Thus, we can regard that as an experimental determination \( \kappa_L \) must be \( \ll 1 \). On esthetic grounds, one may conjecture \( \kappa_L = 0 \). The otherwise puzzling phenomenon of quark confinement, or near confinement, now receives a natural "explanation".

III. SOLITON (OR BAG) MODEL

A. Phenomenological Lagrangian

In order to incorporate the long-range order of the vacuum and its long-wavelength fluctuations into the dynamics of the hadron, we adopt the soliton description\(^1\) through the use of collective coordinates. The dielectric constant \( \kappa \) will now be represented by a scalar field \( \sigma \). The simplest form is

\[
\sigma = 1 - \kappa.
\]

If we assume (8a), then
\[ \frac{\sigma}{\sigma_{\text{vac}}} = 1 - \kappa, \]
so that for an infinite volume, \( \kappa = 0 \) and therefore \( \sigma = \sigma_{\text{vac}} \). In the soliton picture, one has outside the hadron \( \sigma = \sigma_{\text{vac}} \), but inside the hadron \( \sigma = 0 \) which gives \( \kappa \approx 1 \), consistent with our convention (5).

The phenomenological Lagrangian density is assumed to be (see, however, the modification given in Sec. III D). \[ \mathcal{L} = -\frac{1}{4} F_{\mu \nu}^a F^{a}_{\mu \nu} - \psi^\dagger \gamma_4 (\gamma_5 D_\mu + m) \psi \]
\[ - \frac{1}{2} \left( \frac{\partial \sigma}{\partial x_m} \right)^2 - U(\sigma) + \text{counterterms}, \tag{10} \]

where the superscript dagger denotes the Hermitian conjugation, \( x_m = (\tau, \vec{x}) \),
\[ V_{\mu}^a = \frac{\partial}{\partial x_\mu} v_{\mu}^a - \frac{\partial}{\partial x_\mu} v_{\mu}^b + g C^{abc} v_{\mu}^b v_{\mu}^c, \]
\[ D_\mu = \frac{\partial}{\partial x_\mu} - \frac{i}{2} g \lambda^a v_{\mu}^a, \tag{11} \]

the \( \lambda^a \)'s are the standard Gell-Mann matrices related to \( C^{abc} \) by the commutation relation
\[ [\lambda^a, \lambda^b] = 2i \epsilon^{abc} \lambda^c, \]

the \( \gamma_\mu \)'s are the usual Hermitian Dirac matrices, and the function \( U(\sigma) \) has an absolute minimum at \( \sigma = \sigma_{\text{vac}} \) and a local minimum at \( \sigma = 0 \) with \( U(\sigma_{\text{vac}}) = 0 \)

and
\[ U_0 = -\rho > 0. \tag{12} \]

As we shall see, the detailed form of \( U(\sigma) \) is not important. If one wishes, one may assume \( U(\sigma) \) to be simply a quartic function of \( \sigma \). In (10) and (11), \( \psi \) denotes the quark field which, besides being a color triplet, also has \( F \) flavors, \( m \) is the mass matrix for quarks inside the hadron, and \( g \) and \( f \) are both renormalizable coupling constants. Since \( \sigma \) is only a phenomenological field, describing the long-range collective effects of QCD, its short-wavelength components do not exist in reality. The counterterms in \( \mathcal{L} \) are for renormalization; they consist only of those due to loop diagrams of the vector gauge field \( V_{\mu}^a \) and the quark field \( \psi \). In the following, we shall ignore all \( \sigma \) loops, i.e., \( \sigma \) will be approximated by a classical field.

In our phenomenological Lagrangian density, the \( \sigma \) field is coupled to the quarks in two ways: One is through the quark-\( V_{\mu}^a \) coupling \( g \) and \( -\frac{i}{2} k V_{\mu}^a v_{\mu}^a \), which, because of (9), couples \( V_{\mu}^a \) with \( \sigma \), the other is a direct quark-\( \sigma \) coupling \( f \). So far as the quark confinement problem is concerned, there is no

need to have the direct quark-\( \sigma \) coupling; it suffices to have the vacuum be a perfect dia-electric (\( \kappa = 0 \)). The origin of the \( f \) coupling lies in the dichotomy that only inside the hadron is the coupling between the quark and the vector gauge field really \( g \), which is relatively small; outside the hadron, because of (1), it is actually \( g/\kappa_{\text{vac}}^{1/2} \). For \( \kappa_{\text{vac}} \ll 1 \), \( g/\kappa_{\text{vac}}^{1/2} \) becomes very large; on the one hand, this has the desirable effect of preventing the quarks from moving outside the hadron, but on the other hand it also presents a technical difficulty for a diagram analysis in powers of \( g^2 \). In (10) the direct quark-\( \sigma \) coupling \( f \) with

\[ f \sigma_{\text{vac}} \approx \text{hadron mass} \tag{13} \]

is introduced phenomenologically to give a convenient alternative formulation of the same physical effect, but bypassing the above difficulty. As we shall see, because of (13) the \( f \) coupling restricts the quarks to staying always inside the hadron, and that enables us to take full advantage of the relative smallness of the quark-\( V_{\mu}^a \) coupling \( g \).

With the \( f \) coupling we may now expand any physical observable, say the hadron mass \( M \), in a power series of \( \alpha = (4\pi)^{-1} g^2 \):
\[ M = M_0 + \alpha M_1 + \alpha^2 M_2 + \cdots. \tag{14} \]

If \( f \) were zero, it would be difficult to derive the zeroth-order term \( M_0 \), since when \( \alpha = 0 \) both \( g \) and \( g/\kappa_{\text{vac}}^{1/2} \) should be zero; hence, without the \( f \) coupling quarks would be unconfined. (See Sec. III D for further discussions.)

B. Zeroth-order approximation

With the direct quark-\( \sigma \) coupling \( f \), we can now in the zeroth-order calculation neglect the exchange of \( V_{\mu}^a \). The description of the hadron reduces to that of a simple soliton model, consisting of the scalar \( \sigma \) field and the quark \( \psi \) field. The relevant part of the Lagrangian density (10) now consists simply of

\[ -\phi^4 \left( \frac{\partial}{\partial x_\mu} + m + \sigma \right) \phi \left( \frac{\partial}{\partial x_\mu} \right)^2 - U(\sigma). \tag{15} \]

As mentioned before, since \( \sigma \) is only a phenomenological field that has no short-wavelength components, we shall neglect all \( \sigma \)-loop diagrams. The remaining \( \sigma \) diagrams are all tree diagrams, which correspond to the quasiclassical approximation that has been extensively studied in the literature. Here we give only a summary of the results for light quark hadrons, with \( m = 0 \). In this case, \( \psi \) and \( \sigma \) can be reduced to c-number functions which satisfy

\[ (-i \vec{\sigma} \cdot \vec{\nabla} + f \sigma) \psi = c \psi \]
and
\[ -\nabla^2 \sigma + \mathcal{U}^*(\sigma) = -N \phi \psi' \psi, \]
where \( \mathcal{U}^*(\sigma) = dU/d\sigma \), \( \gamma \) and \( \beta \) are the usual Dirac matrices, \( \epsilon > 0 \), \( \psi \) satisfies \( \psi' \psi d^3r = 1 \), and \( N \) is the total number of quarks and antiquarks, \( N = 2 \) for mesons and \( N = 3 \) for baryons.

As shown in Ref. 7, for \( f \sigma_{\text{vac}} \gg \) hadron mass and under very general assumptions, the hadron acquires a well-defined surface \( \mathcal{S} \). One has

\[ \psi = 0 \]

and

\[ \sigma = \sigma_{\text{vac}} \text{ outside } \mathcal{S}; \]

on the inner side of the surface \( \mathcal{S} \),

\[ -i\beta \cdot \mathbf{\sigma} n = \psi, \]

where \( \mathbf{n} \) is the unit vector normal to \( \mathcal{S} \), and inside \( \mathcal{S} \) the field \( \sigma \) is \( \equiv 0 \), with its deviation from zero proportional to \( \psi' \psi \). For the \( s_{1/2} \) orbit, \( \mathcal{S} \) is a spherical surface of radius \( R \). Inside \( \mathcal{S} \), (16) can be further reduced, through scaling, to a simple system of two coupled first-order differential equations:

\[ \frac{du}{d\rho} = (-1 + u^2 - v^2)u \]

and

\[ \frac{dv}{d\rho} + \frac{2v}{\rho} = (1 + u^2 - v^2)u, \]

where \( \rho \) is related to the radius \( r \) by \( \rho = \epsilon r \),

\[ \psi \propto \begin{pmatrix} u \\ (i\mathbf{\sigma} \cdot \mathbf{\mathcal{F}}/r) v \end{pmatrix} \zeta, \]

\[ \zeta = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \]

\( \mathcal{F} \) is the usual Pauli spin matrix, and the variables \( \rho, u, \) and \( v \) are all dimensionless. Although (19) does not contain any explicit parameters, its solutions form a one-parameter family. As \( \rho \to 0 \), one has \( v \to 0 \) and \( u \to \infty \). For every \( \epsilon \) between 0 and a critical value \( \epsilon_c = 1.7419 \), there is a solution of (19). The solution can be obtained by direct integration from \( \rho = 0 \) to \( \rho = \rho_0 \). At \( \rho = \rho_0 \), one has \( u(\rho_0) = v(\rho_0) \) and therefore the boundary condition (18) is satisfied. The radius of the hadron is given by

\[ R = \rho_0/\epsilon. \]

A convenient parameter to label these solutions can be either \( u(0) \) or the integral

\[ n = \int (u^2 + v^2) d^3\rho, \]

over the region \( \rho < \rho_0 \). As the initial value \( u(0) \to 0 \), one has \( n \to 0 \), but as \( u(0) - u_c = 1.7419 \), \( n \to \infty \).

The physical description of the soliton solution then resembles that of a gas bubble (i.e., the hadron) inside a medium (i.e., the vacuum). The hadron mass is determined by three parameters:

\[ \rho, \ s, \ \text{and} \ n, \]

where \( \rho \) is defined by (12) which represents the pressure of the medium on the bubble, \( s \) is the surface tension which arises because \( \sigma \) changes from 0 to \( \sigma_{\text{vac}} \) across the soliton surface, and \( n \) determines the gas pressure inside the bubble which is due to both the kinetic energy of quarks and the excitation energy of \( \sigma \). In either of the limits \( n \to 0 \) or \( n \to \infty \) the hadron mass has the form, in the notation of (14),

\[ M_\sigma = \frac{N\rho_0}{R} + \frac{1}{4} \pi R^2 \rho + 4\pi R^2 s, \]

where

\[ \rho_0 = 2.0428 \text{ when } n \to 0 \]

and \( \rho_0 = 1 \) when \( n \to \infty \).

The double limit \( n \to 0 \) and \( s \to 0 \) gives the MIT bag,\(^{1,3}\) and the double limit \( n \to \infty \) and \( \rho \to 0 \) gives the SLAC bag.\(^{10}\)

When \( n \to 0 \), the field \( \sigma \) assumes a simple form:

\[ \sigma = \sigma_{\text{vac}} \text{ outside } \mathcal{S} \]

and

\[ \sigma = 0 \text{ inside } \mathcal{S}. \]

For \( n \neq 0 \), though \( \sigma = \sigma_{\text{vac}} \text{ outside } \mathcal{S}, \sigma \) is only \( \equiv 0 \) inside; as mentioned before, its deviation from 0 is proportional to \( \psi' \psi \).

C. Coulomb gauge

To simplify our discussions, in the following we shall regard \( \sigma(\mathbf{r}) \), and therefore also \( \kappa(\mathbf{r}) \), as a given function of \( \mathcal{F} \). Let us consider a single hadron system so that its surface \( \mathcal{S} \) is simply connected. We assume further that \( \sigma(\mathbf{F}) \) has the simple form given by (23); hence, because of (9)

\[ \kappa(\mathbf{F}) = \kappa_n \to 0 \text{ outside } \mathcal{S} \]

and

\[ \kappa(\mathbf{F}) = 1 \text{ inside } \mathcal{S}, \]

where the precise form of the surface \( \mathcal{S} \) is to be determined by minimizing the energy of the hadron state under consideration, as is done in the above.
zeroth-order calculation. [It is quite straightforward to extend our analysis to the more complicated case in which $\kappa(\tilde{F})$ is only $\equiv 1$ inside $\mathcal{S}$, as in the general $n \neq 0$ solutions examined in the preceding section.]

From (10), it follows that $V_\mu^a$ satisfies (neglecting the counterterms)
\[
\frac{\partial}{\partial x^\mu}(\kappa V_\mu^a) = -gJ_\mu^a, \tag{25}
\]
where
\[
J_\mu^a = j_\mu^a + \kappa C^{abc}V_\mu^b V_\mu^c \tag{26}
\]
and
\[
j_\mu^a = \frac{1}{2} i \psi \gamma_\mu \lambda^a \psi. \tag{27}
\]

It is convenient to introduce the three-vectors $\tilde{F}_\mu^a$, $\tilde{B}_\mu^a$, and $\tilde{H}_\mu^a$, as in the case of the usual electromagnetic field in a medium:
\[
V_{4a} = i\tilde{E}_\mu^a, \quad V_{ij}^a = e_{ijk}\tilde{B}_k^a, \tag{28}
\]
\[
\tilde{F}_\mu^a = \kappa \tilde{B}_\mu^a, \quad \text{and} \quad \tilde{H}_\mu^a = \kappa \tilde{B}_\mu^a,
\]
where the roman subscripts denote the space components, $e_{ijk} = +1$ if $ijk$ is an even permutation of $123$, $-1$ if it is an odd permutation, and $0$ otherwise. Equation (25) becomes
\[
\tilde{\nabla} \times \tilde{F}_\mu^a = gJ_\mu^a \tag{29}
\]
and
\[
\tilde{\nabla} \times \tilde{H}_\mu^a = \tilde{E}_\mu^a = gJ_\mu^a,
\]
where the dot denotes a time derivative, $J_\mu^a = (\tilde{F}_\mu^a, \tilde{J}_\mu^a)$,
\[
J_\mu^a = \frac{1}{2} \phi \lambda^a \psi - \kappa C^{abc}(V_\mu^a \tilde{E}_\nu^b + \tilde{V}_\mu^a \times \tilde{B}_\nu^b). \tag{30}
\]

In the Coulomb gauge, $\tilde{V}_\mu^a$ is chosen to be divergence-free,
\[
\tilde{\nabla} \cdot \tilde{V}_\mu^a = 0, \tag{31}
\]
and it satisfies
\[
\tilde{V}_\mu^a \cdot \hat{n} = 0 \text{ on } \mathcal{S}, \tag{32}
\]
where, as before, $\hat{n}$ is the normal of $\mathcal{S}$. Because the Lagrangian density (10) is locally gauge-invariant, one sees readily that (as will also be proved in Appendix B) both (31) and (32) can be satisfied. According to (24), $\kappa(\tilde{F})$ has a finite discontinuity across $\mathcal{S}$, which together with (32) gives
\[
\tilde{V}_\mu^a \cdot \tilde{\nabla} \kappa = 0. \tag{33}
\]

The function $V_\mu^a$ will now be regarded as a functional of $\tilde{V}_\mu^a$, $\tilde{V}_\mu^a$, and
\[
j_\mu^a = \frac{1}{2} \phi \lambda^a \psi \tag{34}
\]
through
\[
-\tilde{\nabla} \cdot (\kappa \tilde{V}_\mu^a) = gJ_\mu^a, \tag{35}
\]
where on account of (28)–(33)
\[
\tilde{F}_\mu^a = J_\mu^a + \kappa C^{abc}\tilde{V}_\mu^a \times \tilde{V}_\mu^b \tag{36}
\]
Because $\kappa C^{abc}\tilde{V}_\mu^a \times \tilde{V}_\mu^b = \tilde{\nabla} \cdot (\kappa C^{abc}\tilde{V}_\mu^a)$, the total "color" charge is
\[
Q^a = \int J_\mu^a d\mathcal{S} = \int \tilde{F}_\mu^a d\mathcal{S}; \tag{37}
\]

furthermore, it is a constant of motion. Let $G(\tilde{F}, \tilde{F}')$ be the Green’s function defined by
\[
-\tilde{\nabla} \cdot [\kappa \tilde{V}_\mu G(\tilde{F}, \tilde{F}')] = \delta^{2}(\tilde{F} - \tilde{F}'), \tag{38}
\]
where $\tilde{\nabla}$ operates on $\tilde{F}$. The solution $V_\mu^a$ of (35) satisfies
\[
V_\mu^a(\tilde{F}) = g \int G(\tilde{F}, \tilde{F}') \tilde{F}''(\tilde{F}') d\mathcal{S}'. \tag{39}
\]
Since $\tilde{F}$ depends on $V_\mu^a$, the functional dependence of
\[
V_\mu^a = V_\mu^a(\tilde{V}_\mu^a, \tilde{V}_\mu^a, j_\mu^a) \tag{40}
\]
can be obtained from (36) and (39) by iteration. From (24), one sees that as $\kappa \to 0$, the Green’s function $G(\tilde{F}, \tilde{F}') \to \infty$ everywhere. We may expand $G(\tilde{F}, \tilde{F}')$ in powers of $\kappa$:
\[
G(\tilde{F}, \tilde{F}') = \kappa^{-1} G_0(\tilde{F}, \tilde{F}') + G_1(\tilde{F}, \tilde{F}') + G_2(\tilde{F}, \tilde{F}') + \cdots. \tag{41}
\]

As shown in Appendix B, for arbitrary surface $\mathcal{S}$,
\[
G_0(\tilde{F}, \tilde{F}') = \text{constant} \tag{42}
\]
when both $\tilde{F}$ and $\tilde{F}'$ are inside $\mathcal{S}$; furthermore, $G_0(\tilde{F}, \tilde{F}')$ is independent of $\tilde{F}$ when $\tilde{F}$ is inside $\mathcal{S}$ (but $\tilde{F}'$ may be outside). Likewise, owing to symmetry, $G_0(\tilde{F}, \tilde{F}')$ is independent of $\tilde{F}'$ when $\tilde{F}'$ is inside $\mathcal{S}$. Hence for a color-singlet hadron, because $Q^a = 0$ and $j_\mu^a = \tilde{V}_\mu^a = 0$ outside $\mathcal{S}$,
\[
\lim_{\kappa \to 0} V_\mu^a(\tilde{F}) = \text{finite}. \tag{43}
\]

The Green’s function $G(\tilde{F}, \tilde{F}')$ for a spherical surface of radius $R$ can be readily derived (and is also given in Appendix B), from which one sees that, e.g., when $\tilde{F}$ and $\tilde{F}'$ are both inside the surface $G_0(\tilde{F}, \tilde{F}') = (4\pi R)^{-1}$ and

178
where the magnitudes of \( \mathbf{F} \) and \( \mathbf{F}' \) are \( r \) and \( r' \), respectively, the angle between them is \( \theta \), and the \( P_i \)'s are the standard Legendre polynomials. [The sum in (44) can be easily carried out. See (B22) and (B23) of Appendix B.]

In the Coulomb gauge, \( \hat{\mathbf{S}} \) and \( \hat{\mathbf{V}} \) are regarded as independent generalized coordinates. Their conjugate momenta are, respectively, \( \hat{\mathcal{P}} \) and

\[
\hat{\Pi} = [\delta_{ij} - (\nabla \hat{\mathbf{V}}/\nabla^2)] \hat{D}_{ij} + \mathcal{Q} \frac{1}{R} \left[ -1 + \frac{l+1}{l} \left( \frac{\mathbf{r} r'}{R^2} \right)^l P_l(\cos \theta) \right] .
\]

(45)

It is useful to resolve \( \hat{\mathbf{E}} \) and \( \hat{\mathbf{D}} \) into their transverse and longitudinal components:

\[
\hat{\mathbf{E}} = \hat{\mathbf{E}}_T + \hat{\mathbf{E}}_{\text{load}}
\]

and

\[
\hat{\mathbf{D}} = \hat{\mathbf{D}}_T + \hat{\mathbf{D}}_{\text{load}} ,
\]

where \( \hat{\mathbf{D}}_T = \hat{\mathbf{E}}_T = 0 \) and \( \hat{\mathbf{D}} \times \hat{\mathbf{D}}_{\text{load}} = \hat{\mathbf{E}} \times \hat{\mathbf{E}}_{\text{load}} = 0 \). Equation (45) becomes

\[
\hat{\Pi} = -\hat{\mathbf{D}}_T .
\]

(47)

By using (35) and (36), one can readily verify that

\[
\int \hat{\mathbf{D}}_{\text{load}} \cdot \hat{\mathbf{E}}_{\text{load}} d^3r = g \int V^v_0 j_0^v + C^v \nabla \hat{\mathbf{V}}^v \cdot \hat{\Pi}^v d^3r ,
\]

(46)

and therefore, through partial integrations, the Hamiltonian density can be set to be

\[
\mathcal{H} = \frac{1}{2} \hat{\mathbf{E}}_T \cdot \hat{\mathbf{E}}_T + \frac{1}{2} \hat{\mathbf{D}}_{\text{load}} \cdot \hat{\mathbf{E}}_{\text{load}} + \frac{1}{2} \hat{\mathbf{D}}_T \cdot \hat{\mathbf{D}}_T + \frac{1}{2} \hat{\mathbf{E}}_T \cdot \hat{\mathbf{D}}_T + \frac{1}{2} \hat{\mathbf{E}}_{\text{load}} \cdot \hat{\mathbf{D}}_{\text{load}}
\]

\[
+ \frac{1}{g} \left( \hat{\mathbf{E}}_T \cdot \hat{\mathbf{D}}_T + j_0^v + m \right) \hat{\mathcal{P}} - \frac{1}{2} \hat{\mathbf{V}}^v \cdot \hat{\mathbf{V}}^v + \frac{1}{2} \hat{\Pi}^v \hat{\Pi}^v + U(\sigma) ,
\]

(49)

which is valid in the approximation that \( \sigma(x) \) is a given time-independent classical field.

D. A modification

As mentioned in Sec. III A, the quark-\( V_0^v \) coupling has the relatively small magnitude \( g \) only inside the hadron; outside, it is \( g/\kappa \omega^{1/2} \) which \( \to 0 \) when \( \omega \to 0 \). In order to develop a series expansion in \( g \), but not in \( g/\kappa \omega^{1/2} \), we introduce a direct quark-\( \sigma \) coupling \( f \) in the Lagrangian density (10). When \( j_0^v \gg \text{hadron mass} \), the quark field \( \phi \) becomes 0 outside the hadron surface \( \delta \), in accordance with (17); thereby we avoid the very large coupling \( g/\kappa^{1/2} \). The same problem also arises between the vector gauge fields themselves because of self-coupling. Likewise, we would also like to restrict

\[
V_\phi = 0 \quad \text{outside } \delta .
\]

(50)

For simplicity, let us assume that, as in the preceding section, \( \sigma(x) \) is given by (23). Within the context of a relativistic local theory, the above restriction can be achieved most simply by modifying the Lagrangian density from \( \mathcal{L} \) to

\[
\mathcal{L}' = \mathcal{L} - \frac{1}{2} \kappa (f' \sigma)^2 V_0^v V_\phi \nonumber \]

(51)

where, as before, \( \kappa \) and \( \sigma \) are related by (9) and \( \delta \) is given by (10). Like (13), this additional coupling \( f' \) satisfies

\[
f' \sigma \gg \mu \gg \text{hadron mass} .
\]

(52)

The modified Lagrangian density \( \mathcal{L}' \) is no longer locally gauge invariant, though it remains globally invariant. Hence the total "color" charge \( Q^v \) is still conserved. Furthermore, as we shall see, since

\[
f' = 0 \quad \text{inside } \delta ,
\]

(53)

we retain almost all the physical consequences of the locally gauge-invariant Lagrangian density \( \mathcal{L} \).

[One may wonder: Why not directly set in the Lagrangian density (10) \( f' = 0 \), or in (51) \( f' = 0 \), and simply impose \( \sigma = 0 \) and \( V_\phi = 0 \) outside \( \delta \) as constraints? Although such an approach will be discussed in Sec. IV, in principle there are several advantages in starting from a Lorentz-invariant Lagrangian density. (i) By using its soliton solution, one can then derive these constraints. This way it ensures the self-consistency of the constraints, especially since \( V_\phi \) is not an independent variable. For example, without the \( f' \) coupling, even if \( \phi = \text{const} \) outside \( \delta \), according to (39) there is still a tail of nonzero \( V_\phi \) outside \( \delta \), and that might lead to a Van der Waals-type force between hadrons which would be in violation of experimental observations. However, in the classical limit, one can show that when \( \omega \to 0 \) the long-range force between hadrons must vanish. (ii) A covariant Lagrangian theory allows one to study the motion of the hadron surface \( \delta \) and to ensure relativistic covariance.]

Instead of (25), \( V_\phi \) now satisfies

\[
\frac{\partial}{\partial \xi^\mu} (\kappa V_\phi) - \kappa (f' \sigma)^2 V_\phi = - g d^\nu ,
\]

(54)
and instead of (35), \( V_\alpha^\phi \) satisfies
\[
- \nabla_i (\kappa \nabla^i \phi^\alpha) + \kappa (\partial_i \phi^\alpha) \nabla_i \phi^\alpha = g \phi^\alpha, \tag{55}
\]
where \( \psi_\alpha^r \) and \( \phi^\alpha \) remain given by (26) and (36), respectively. Because the local gauge invariance is broken in \( \mathcal{L}' \), it is not possible to impose the divergence condition (31); all three components of \( \nabla^2 \) are independent variables. However, since \( \nabla^2 \phi^\alpha \) is absent in \( \mathcal{L}' \), only \( \psi \) and \( \nabla^2 \) can be regarded as independent generalized coordinates; \( V_\alpha^\phi \) remains a dependent variable. Just as in (39), we may write
\[
V_\alpha^\phi(\mathcal{F}) = g \int G(\mathcal{F}, \mathcal{F}') n^r(\mathcal{F}') d^4r', \tag{56}
\]
where, instead of (38), \( G(\mathcal{F}, \mathcal{F}') \) is now defined by
\[
- [\nabla_i (\kappa \nabla^i \phi^\alpha) + \kappa (\partial_i \phi^\alpha)] G(\mathcal{F}, \mathcal{F}') = \delta^i_4(\mathcal{F} - \mathcal{F}'). \tag{57}
\]
At any fixed \( \mu = \gamma \sigma_{\text{max}} \), when \( \kappa_{\text{max}} = 0 \), this new Green's function \( G(\mathcal{F}, \mathcal{F}') \rightarrow -\infty \) everywhere, as before. We may again expand \( G(\mathcal{F}, \mathcal{F}') \) in powers of \( \kappa_{\text{max}} \):
\[
G(\mathcal{F}, \mathcal{F}') = \kappa_{\text{max}} G_\lambda(\mathcal{F}, \mathcal{F}') + G_\mu(\mathcal{F}, \mathcal{F}') + \cdots, \tag{58}
\]
It will be proved in Appendix C that (42) and (43) remain valid, i.e.,
\[
G_\mu(\mathcal{F}, \mathcal{F}') = \text{const}, \tag{59}
\]
when \( \mathcal{F} \) and \( \mathcal{F}' \) are both inside \$, \$ and for a color-singlet hadron, \( V_\alpha^\phi(\mathcal{F}) \) finite when \( \kappa_{\text{max}} = 0 \). Furthermore, for a spherical surface of radius \( R \) and for \( \mathcal{F}, \mathcal{F}' \) inside the sphere, \( G_\lambda(\mathcal{F}, \mathcal{F}') \) remains given by (44), apart from an additive constant. Since for a color singlet the additive constant in \( G(\mathcal{F}, \mathcal{F}') \), or \( G_\lambda(\mathcal{F}, \mathcal{F}') \), is of no importance, we may rewrite (44) as
\[
G_\lambda(\mathcal{F}, \mathcal{F}') = (4\pi)^{-1} \left[ \frac{1}{|\mathcal{F} - \mathcal{F}'|} + \sum_{\ell = 1}^{\infty} \frac{1}{L^2} \frac{R^{\ell+1}}{R^{\ell+1}} P_{\ell}(\cos \theta) \right]. \tag{60}
\]
Outside \$, \$, this new \( G_\lambda \) decreases exponentially; hence, it is quite different from the old \( G_\lambda \), which has only a power dependence on \( \mathcal{F} \) and/or \( \mathcal{F}' \). [For nonspherical surfaces, the new \( G_\lambda \) can be quite different from the old \( G_\lambda \), even inside \$. \$ See expression (C17) of Appendix C.] Inside \$, \$, since \( \sigma = 0 \) we have \( \kappa = 1 \) and \( \mathcal{L}' = \mathcal{L} \); hence, \( \nabla^2 = \nabla^2 \mathcal{F} \) and our Lagrangian density is locally gauge invariant. We may again adopt the Coulomb gauge and require (31) and (32), i.e., \( \nabla \cdot \nabla^2 = 0 \) inside the surface and \( \nabla \cdot \nabla^2 = 0 \) on the surface. As in (46), \( \nabla^2 \) may be decomposed into \( \nabla^2 = \nabla^2_{\text{tr}} + \nabla^2_{\text{max}} \)
\[
\begin{align*}
\nabla_{\text{tr}} &= - \nabla V_{\phi}^\phi + \nabla \phi^\alpha, \\
\nabla_{\text{max}} &= \nabla V_{\phi}^\phi - \nabla \phi^\alpha, \tag{61}
\end{align*}
\]
and
\[
\nabla^2 \phi^\alpha = - (g \nabla^2 + (\gamma^2 \nabla^2 \nabla^2)). \tag{62}
\]
As will be shown in Appendix C, one can always choose \( \phi^\alpha \) such that \( \nabla_{\text{tr}} \phi = 0 \) on \$. \$
\[
\nabla^2 \phi^\alpha = - \nabla^2 = 0. \tag{63}
\]
Equation (47) now becomes
\[
\vec{\Pi}^\phi = - \vec{\Pi}_{\text{tr}}^\phi \text{ inside } \$. \$
\]
Outside the surface, when \( f_\sigma \rightarrow -\infty \) and \( f' \sigma \rightarrow -\infty \), we have
\[
\psi = \nabla^2 = 0. \tag{64}
\]
Therefore, the modified Lagrangian (51) provides a self-consistent device to exclude the quark and the gauge fields from the outside region.

IV. FEYNMAN DIAGRAMS

A. Reduced Hamiltonian

To simplify our discussions, we again ignore the surface motion,\(^{12}\) and assume \( \sigma \) and \( \kappa \) as given by (23) and (24), respectively. Inside the surface \$, \$, the equations of motion reduce to
\[
\frac{\partial}{\partial x_\mu} V_{\mu \nu}^\phi = - g J_{\nu}^\phi \tag{65}
\]
and
\[
(g D_\mu + m) \phi = 0, \tag{66}
\]
where \( D_\mu \) and \( J_\mu^\phi \) are given by (11) and (26). Outside \$, \$, we impose the constraints \( \phi = \nabla^2 = 0 \), in accordance with (64). We shall adopt the Coulomb gauge: \( \nabla \cdot \nabla = 0 \) inside \$. \$ In addition, \( \phi \) and \( \nabla^2 \) satisfy the boundary conditions (18) and (32), respectively.

By following the discussions given in the preceding section (and as will also be shown in Appendix C), the Hamiltonian of the system can be reduced to the form
\[
H_0 = \int_\$ x d^3r + s A, \tag{65}
\]
where \( H_0 \) refers to the volume within \$, \$
\[
x = \frac{1}{2} \vec{\Pi}^2 + \frac{1}{2} g V_{\phi}^\phi + C_{\text{vec}} \vec{\Pi}^2 + \frac{1}{2} \vec{B}^2 + \vec{E}^2 + \vec{\phi}^2 \left( \frac{1}{4} \nabla^2 + m \right) \vec{\Psi} + \vec{J} \cdot \vec{\Psi} + P, \tag{66}
\]
p is given by (12), \( s \) is the surface tension defined in (20); and \( A \) is the area of \$. \$ As before, \( \vec{\Pi}^\phi \)}
\begin{align*}
\mathbf{I}^a &= \int A^a d\mathbf{r}', \\
\mathbf{V}^a(\mathbf{r}) &= \sum_{\phi}^\mathbf{a} \mathbf{V}^a(\mathbf{r}) d\mathbf{r}',
\end{align*}

where \(G(\mathbf{F}, \mathbf{F}')\) is given by (57).

As shown in Appendix C, for color nonsinglets, \(Q^a = 0\),

\begin{align*}
H_0 &= (2\pi a)^{1/2} G_0 \mathbf{Q}^a Q^a, \\
H_0 &= (2\pi a)^{-1/2} G_0 \mathbf{Q}^a Q^a - \infty,
\end{align*}

where according to (69) \(G_0\) is a constant. But for a color singlet, \(H_0\) is finite and (65) becomes

\begin{align*}
H_0 &= \int \mathbf{\Phi}^a d\mathbf{r} + sA
\end{align*}

where \(\mathbf{\Phi}\) is the same as \(\mathbf{\Phi}\), provided \(V_n'\) is replaced by

\begin{align*}
V_n^a(\mathbf{r}) &= \int G_1(\mathbf{F}, \mathbf{F}') \mathbf{F}'(\mathbf{r})' d\mathbf{r}',
\end{align*}

where \(G_1\) is defined by (58). The explicit form of \(\mathbf{\Phi}\) is

\begin{align*}
\mathbf{\Phi} &= \mathbf{A}^a = \frac{1}{2} \mathbf{A}^a + \frac{1}{2} \mathbf{V}^a(\mathbf{F}^a + C^{ab} \mathbf{V}^b) + \frac{1}{2} \mathbf{B}^a + \mathbf{\Phi}^a, \\
&= \mathbf{A}^a + \mathbf{\Phi}^a, \\
&= \mathbf{A}^a + \mathbf{\Phi}^a + \mathbf{\Phi}^a + \mathbf{\Phi}^a + \mathbf{\Phi}^a
\end{align*}

From (69) and (70), we conclude that when \(\mathbf{\Phi}^a = 0\), the Hamiltonian operator is always divergent for color nonsinglets, but for color singlets, it is always finite.

B. Feynman rules

In the Coulomb gauge, the derivation from the Lagrangian to the Hamiltonian follows the standard canonical procedure. However, because the Lagrangian contains nonlinear terms which are \(\mathbf{\Phi}^a\) dependent, there must be some additional action which gives rise to new loop diagrams. The simplest derivation is to follow the path-integration method of Faddev and Popov.\(^{13,14}\) The result can be expressed in terms of an additional Hamiltonian

\begin{align*}
H_1 &= -\frac{1}{2} i \mathbf{r} \ln(1 - \Delta^a) + i \mathbf{r} \ln(1 - \Delta^a)N,
\end{align*}

where \(\mathbf{r}\) is the unit matrix

\begin{align*}
\langle a, \mathbf{F} | b, \mathbf{F} \rangle &= \delta^{ab} S(\mathbf{F} - \mathbf{P}),
\end{align*}

and the matrices \(\Delta, M, \) and \(N\) are given by

\begin{align*}
\langle a, \mathbf{F} | 1, b, \mathbf{F} \rangle &= \delta^{ab} \mathbf{P}(\mathbf{F} - \mathbf{P}), \\
\langle a, \mathbf{F} | M, b, \mathbf{F} \rangle &= g C^{abc} \mathbf{V}^c(\mathbf{F}) + \mathbf{V}^c(\mathbf{P}) + g C^{abc} \mathbf{V}^c(\mathbf{F}) - \mathbf{V}^c(\mathbf{P}),
\end{align*}

where \(g\) is the usual Kronecker symbol, and \(\mathbf{V}\) acts on \(\mathbf{F}\). The Feynman diagrams can now be constructed by using the "effective Hamiltonian"
\( \hat{n} \cdot \hat{\nabla} \phi = 0 \)

and

\[ \hat{n} \times (\hat{\nabla} \times \phi) = 0 \quad \text{on} \quad \mathcal{S}, \]

where, as before, \( \hat{n} \) is the normal vector. (See Appendix C for further discussions.)

In the special case that \( \mathcal{S} \) is a spherical surface of radius \( R \), the usual TE and TB modes give the complete solution. These radiation modes can be expressed in terms of the scalar solution of

\[ \nabla^2 \phi + k^2 \phi = 0, \]

which may in turn be written as

\[ \phi_{k,1,\alpha}(\vec{r}) = \text{constant} \times j_1(kr)Y_{1,\alpha}^{\alpha}(\phi, \beta), \]

where \( r, \alpha, \beta \) are the spherical coordinates of \( \vec{r} \), \( Y_{1,\alpha}^{\alpha} \) is the spherical harmonics, \( j_1 \) is the spherical Bessel function, \( k > 0 \), and \( \lambda = E \) or \( B \) depending on the TE or TB mode. For the TE mode

\[ \nabla_{k,1,\alpha}(\vec{r}) = \hat{\nabla} \times (\hat{\nabla} \phi_{k,1,\alpha}^{\alpha}(\vec{r})) \]

and \( k \) is determined by

\[ \frac{d}{dr} (r j_1(kr)) = 0 \quad \text{at} \quad r = R. \]

For the TB mode

\[ \nabla_{k,1,\alpha}(\vec{r}) = \hat{\nabla} \times (\hat{\nabla} \phi_{k,1,\alpha}^{\alpha}(\vec{r})) \]

and \( k \) is determined by

\[ j_1(kR) = 0. \]

The constant in (78) is chosen so that the integral of \( |\nabla_{k,1,\alpha}(\vec{r})|^2 \) over the volume \( r < R \) is unity. At \( r = R \), in either mode \( \nabla_{k,1,\alpha}^{\alpha} \) has only tangential components while \( \hat{\nabla} \nabla_{k,1,\alpha}^{\alpha} \) has only a normal component, in accordance with (77).

The Feynman propagator \( D_{k,1,\alpha}^{\alpha}(x, x') \) of the radiation field is given by

\[ [D_{k,1,\alpha}^{\alpha}(x, x')]_{ij} = \delta^{\alpha \beta} \sum_{k,1,\alpha} (2k)^{-2} \int (\nabla_{k,1,\alpha}(\vec{r}')) \cdot e^{i(kr' - kr)}, \]

where the minus sign in the exponent is for \( t > t' \) and the plus sign for \( t < t' \).

To derive the propagator of the quark field, we need the complete set of \( c \)-number solutions \( \chi_{n,\alpha} \) of the Dirac equation in a cavity:

\[ \left( \frac{\alpha}{i} - \frac{1}{2} \hat{\nabla} + \beta m \right) \chi_{n,\alpha} = \pm \epsilon_n \chi_{n,\alpha}, \]

where \( \epsilon_n > 0 \) and \( \chi_{n,\alpha} \) satisfies the boundary condition (18). The Feynman propagator \( S(x, x') \) of the quark field is given by, for \( t > t' \),

\[ S(x, x') = \sum_n (\chi_n(\vec{r}), \chi_n^{\dagger}(\vec{r}') \beta) e^{i\epsilon_n(t' - t)}. \]

and for \( t < t' \),

\[ S(x, x') = -\sum_n (\chi_n(\vec{r}), \chi_n^{\dagger}(\vec{r}') \beta) e^{i\epsilon_n(t - t')}. \]

The various \( g \)-dependent terms in \( H_5 + H_1 \), defined by (75), give directly the great variety of vertices in this problem. These vertices, together with the propagators \( D_{k,1,\alpha}^{\alpha} \), \( D_{1,1,\alpha}^{\alpha} \), and \( S \) given above, complete our discussion of Feynman rules.

We emphasize that the quark-vector and vector-vector interactions are local in character, while the color singletness of the hadron state is a global property. This is why in (76a) there is an explicit dependence on \( \kappa_m \) in the propagator \( D_{k,1,\alpha}^{\alpha}(x, x') \).

When \( \kappa_m = 0 \), our Feynman rules explicitly forbid the appearance of external color nonsinglets, such as true quarks or free vector gauge particles. The propagators \( D_{k,1,\alpha}^{\alpha} \), \( D_{1,1,\alpha}^{\alpha} \), and \( S \) also depend on the linear size \( R \) of the hadron surface. Only at high energy and large four-momentum transfer can one neglect the effect of hadron surfaces; in that case, one recovers the "usual" Feynman rules used in the current literature on QCD.

V. REMARKS

A. Phase transition

In our picture, at low temperature, an infinite-volume QCD exists only in a single phase. A good analogy is to think of an infinite ferromagnet below the Curie temperature, which has a long-range order and also exists in a single phase. The hadrons are then analogous to some small domain structures within the infinite ferromagnet. Another analogy that has been used frequently in the soliton (or bag) model is to regard the vacuum as a liquid and hadrons as bubbles. It seems reasonable to assume that there should exist a critical temperature at which the infinite-volume QCD system can undergo a phase transition in its dielectric constant \( \kappa \) (or magnetic susceptibility \( \kappa = 1/k \)).

The value of the critical temperature and the nature of the phase transition depend sensitively on the excitation curve of hadron spectroscopy. If one follows Chodos et al.\(^3\) and approximates the excited hadron system as noninteracting bubbles of an ideal relativistic gas, then not only does there exist a critical temperature \( T_\kappa \), but \( T_\kappa \) is also the maximum temperature that the system can attain.\(^{1,17}\) However, as the size \( l \) of such bubbles grows, its dielectric constant \( \kappa \) should decrease, and therefore the effective coupling \( c/k \) would increase, which makes the quarks behave less and less like an ideal gas. Furthermore, as the bubbles increase in size, their mutual interactions should become more important. Hence, it remains an
interesting open question whether the idea of a maximum temperature is really correct.

B. Long-range order

Let us leave our problem for the moment and consider a ferromagnet of volume \( L^3 \) at a nonzero temperature below the Curie temperature. The magnetization per unit volume \( M(H, L) \) under an external magnetic field has the following familiar properties: At any finite volume, \( M(H, L) \) is an analytic function of \( H \) and \( \lim_{H \to 0} M(H, L) = 0 \). Thus,

\[
\lim_{H \to 0} M(H, L) = 0. \tag{83}
\]

On the other hand, if we take the infinite-volume limit first and then the \( H \to 0 \) limit, a finite non-zero magnetization results:

\[
\lim_{H \to 0} M(H, L) \neq 0; \tag{84}
\]

this is why we usually say the infinite ferromagnet carries a long-range order, which is characterized by its magnetization and given by the double limit (84).

In our problem, we may take the ratio of the (color) dielectric constant between two volumes of sizes \( l^3 \) and \( L^3 \) with \( l < L \). By using (A2), (A5), (A7), and (A16) in Appendix A, we find, after neglecting \( O(\alpha^2) \),

\[
\frac{\kappa_L}{\kappa_t} = \left[ 1 + \frac{11}{2\pi^2} \ln(L/l) \right]^{-1}, \tag{85}
\]

where \( \alpha = 4\pi^2 |g|^2. \) The approximation of neglecting \( O(\alpha^2) \) is, of course, not a good one, but some insight into the long-range order in QCD may be obtained even in such a crude approximation. Let us keep the small size \( l \) always fixed. At any finite \( L \), we have, according to (85), \( \lim_{\alpha \to 0} (\kappa_L/\kappa_t) = 1 \) and therefore

\[
\lim_{\alpha \to 0} \lim_{L \to \infty} [1 - (\kappa_L/\kappa_t)] = 0. \tag{86}
\]

On the other hand, \( \lim_{L \to \infty} (\kappa_L/\kappa_t) = 0 \) at any \( \alpha \neq 0 \).

Hence,

\[
\lim_{L \to \infty} [1 - (\kappa_L/\kappa_t)] = 1 \tag{87}
\]

in analogy with (83) and (84). Hence, we may regard the long-range order in QCD to be characterized by the perfect dia-electric property of its vacuum when the volume becomes infinite. (See Ref. 1 for further discussions.)

C. Scale of hadron radius

Let us consider a QCD system in the limit of quark mass \( = 0 \) inside the hadron. At first sight, since there seems to be no mass scale in the problem, it may be difficult to see how the physical hadron radius can possibly emerge. To explain this apparent contradiction, let us introduce, as in Appendix A, an ultraviolet momentum cutoff \( \Lambda \) and denote \( g_0 \) as the unrenormalized coupling constant. Inside a hadron of radius \( R \), let \( \alpha \) be the relevant fine-structure constant of the color gauge field. Clearly, \( \alpha \) must be a function of \( R \Lambda \) and \( g_0 \):

\[
\alpha = f(R\Lambda, g_0).
\]

When \( \Lambda \to \infty \), one must vary \( g_0 \) accordingly so that the limiting function

\[
\alpha(R) = \lim_{\Lambda \to \infty} f(R\Lambda, g_0) \tag{88}
\]

exists. This limiting function must contain a length scale since \( R \) has a dimension but \( \alpha \) does not. (In a realistic theory, because the quark masses inside the hadron are not really zero, there are additional mass scales.)

In a quasiclassical soliton (or bag) model, the actual value of \( R \) is determined by minimizing the hadron energy spectrum, of which the lowest level is the pion. Its mass in a quasiclassical calculation is

\[
M_\pi \approx |\frac{4\pi^2}{R} + 4\pi s R^2 + 2M_p - \frac{\alpha(R)}{R} \frac{\eta_s}, \tag{89}
\]

where \( p, s, \rho_0 \) are given by (21) and (23), and \( \eta_s \) is a positive number. As discussed in Ref. 1, since the minimum of \( M_\pi = 0 \), \( \alpha(R) \) should be near a critical value \( \alpha_c \); when \( \alpha(R) = \alpha_c \), \( M_\pi = 0 \), and that in turn determines the pion radius \( R \).

D. Experimental upper limit of \( \kappa_{\infty} \)

As mentioned in Sec. II, according to (7) \( \kappa_{\infty} \) must be \( < 1 \); but as yet, no one has been able to determine through pure theoretical deduction just how small \( \kappa_{\infty} \) actually is. Within our picture, the fact that a single free quark has never been observed sets a lower limit on the free quark mass \( m_\pi \). As shown in Appendix D,

\[
\kappa_\infty \leq 0.11 \alpha(m_p/m_\pi)^{1/3}, \tag{90}
\]

where \( m_p \) is the proton mass and \( \alpha = g^2/4\pi \) is the fine structure for hadrons. Even without any detailed search of the literature, we may set \( m_\pi > 5 \) GeV, from which we determine an experimental upper limit of \( \kappa_{\infty} \):

\[
\kappa_\infty < 1.3 \times 10^{-2} \alpha. \tag{91}
\]

ACKNOWLEDGMENTS

I wish to thank N. H. Christ and A. H. Mueller for many very helpful discussions. The kind hospitality of the Physics Department of BNL during the summer of 1978 is gratefully acknowledged. This research was supported in part by the U. S. Department of Energy.
APPENDIX A

The proof of properties (3) and (4) concerning the (color) dielectric constant of the vacuum, which are stated in Sec. II, will be given in this appendix. As in Sec. II, let us consider a pure QCD system consisting of only color gauge fields $V_a^\mu$. For clarity, we introduce an ultraviolet momentum cutoff $\Lambda$, and denote $g_0$ as the unrenormalized coupling constant. The whole system is then enclosed in a finite volume $L^3$. Let $\kappa_L$ and $g_L$ be, respectively, the (color) dielectric constant of the vacuum and the renormalized coupling constant. Both $\kappa_L$ and $g_L$ are functions of $g_0$, $L$, and $\Lambda$. From dimensional considerations, they must depend only on $g_0$ and the product $L\Lambda$, i.e., $\kappa_L = F_L \Lambda, g_0)$ and

$$g_L = G(L\Lambda, g_0). \quad (A1)$$

Let us consider two different volumes $L^3$ and $L'^3$, but with the same $\Lambda$ and $g_0$. From (1), it follows that

$$\frac{\kappa_L}{\kappa_L} = \left(\frac{g_L}{g_L}\right)^3, \quad (A2)$$

where, as in (A1),

$$g_L = G(L\Lambda, g_0) \quad (A3)$$

or its inverse $g_0 = g_0(G_L, g_L)$. Eliminating $g_0$ between (A1) and (A3), we may express $g_L$ in terms of $L\Lambda$, $\Lambda$, and $g_L$:

$$g_L = G(L\Lambda, g_0(l\Lambda, g_L)). \quad (A4)$$

Since the theory is a renormalizable one, the limit $\Lambda \to \infty$ of (A4) should exist; in this limit,

$$g_L = \lim_{\Lambda \to \infty} G(L\Lambda, g_0(l\Lambda, g_L)), \quad (A5)$$

which may be written as

$$g_L = g(\lambda, g_0), \quad (A6)$$

where

$$\lambda = L^3/L. \quad (A7)$$

At $\lambda = 1$, (A5) becomes

$$g_1 = g(1, g_0). \quad (A8)$$

Because $g_2 = G(L\Lambda, g_0)$ is independent of $I$, one has $(\partial g_2/\partial I)|_{\lambda, g_0} = 0$; hence, $g(\lambda, g_0)$ satisfies the familiar renormalization-group equation (but without the usual $\gamma$ and $\delta$ functions)

$$[-(\delta^2/\delta \ln \lambda) + \beta(\delta/\delta g)] g(\lambda, g) = 0, \quad (A9)$$

and

$$\beta = -\delta g/\delta \ln \lambda. \quad (A10)$$

So far, the finite volume is just a device for the infrared cutoff. Presumably, our physical result should not be sensitive to the precise form in which this infrared cutoff is introduced. In order to derive the dependence of $\beta$ on $g_1$, we assume $r_0 = T = 1$ the usual momentum value $M$ chosen for renormalization. The $\beta$ function as a power series of $g_1$ is then given by

$$\beta(g_1) = - \frac{11(16\pi^2)}{4} g_1^3 - \frac{102(16\pi^2)^2}{4} g_1^5 + O(g_1^7) \quad \text{(A11)}$$

for a pure SU(3) gauge field system. The solution of (A8) has the standard form

$$g(\lambda, g_0) = f(\lambda), \quad (A12)$$

where

$$z = \ln \lambda + \int_{\lambda_1}^{\lambda} \frac{dg'}{\beta(g')} \quad (A13)$$

and $g_1$ is an arbitrary constant, which will be chosen so that

$$0 < g_1 < 1 \quad \text{and} \quad \beta(g') < 0 \text{ for } 0 < g' < g_1. \quad (A14)$$

Let $f^{-1}(z)$ be the inverse of $f(\lambda)$, i.e., $f^{-1}(f(\lambda)) = \lambda$. From (A11), one has $z = f^{-1}(g(\lambda, g_0))$, which at $\lambda = 1$, because of (A7), becomes $z = f^{-1}(g_1)$. Now $z$ is also given by (A12). By setting $\lambda = 1$, one derives

$$f^{-1}(g_1) = \int_{\lambda_1}^{\lambda} \frac{dg'}{\beta(g')} \quad (A15)$$

For $g_1$ sufficiently small, the $O(g_1)$ term in (A10) may be neglected. Hence,

$$f^{-1}(g_1) \approx \frac{8}{3} \pi^2 (g_1^2 - g_1^{-2}), \quad (A16)$$

its inverse function is

$$f(z) \approx \left(\frac{8\pi^2}{3} 11z + g_1^2 \right)^{1/2},$$

and therefore

$$g(\lambda, g_0) \approx \left[\frac{8\pi^2/11}{(z^2 + g_1^2)} \right]^{1/2}, \quad (A17)$$

Since $g(\lambda, g_0)$ is an even function of $g_0$, we need only consider positive values of $g_0$. From (A14), we see that (i) the physical value of $g_0$ can vary from 0 to $\infty$ if $\beta(g') = 0$ only at $g' = 0$, otherwise (ii) $g_1$ can only vary between 0 and $\infty$ where $\beta$ is the smallest positive-definitive root of $\beta(g) = 0$.

When $g_1 = g$, the integral in (A14) diverges. In either case, setting $g_0 = x$, one has in the physical region $\frac{d^2f}{dx} = y/\beta(y) < 0$; hence, the inverse function $f(z)$ satisfies

$$\beta = -\delta g_1/\delta \ln \lambda. \quad (A18)$$
\[
\frac{df(z)}{dz} < 0. \tag{A17}
\]

By using (A11) and (A12), one finds
\[
\frac{\partial}{\partial \lambda} g(\lambda, g_L) < 0. \tag{A18}
\]

Since \(\lambda = l/L\), this means for \(L > l\), \(g_L > g_l\), and therefore, because of (A2), which is \(\kappa_L < \kappa_l\), the inequality (3) of Sec. II.

According to (A14) and (A15), as \(g_f\) varies from 0 to \(g_l\), \(f^1(g_f)\) decreases from \(\infty\) to 0. When \(g_f\) increases to \(g_L\), \(f^1(g_f)\) becomes negative. In the above case (i), as \(g_f \to \infty\) we have either \(f^1(\infty) = -A\) = finite or \(f^1(\infty) = -\infty\); in the former the inverse function \(f(z) = \infty\) as \(z = -A\), which means that because of (A11) and (A12), at any fixed \(g_f, g_L = g(\lambda, g_f)\) becomes singular at a finite \(\lambda = l/L\), in violation of our assumption that \(\kappa_L\) (therefore, also \(g_L\)) is a smooth function of \(L\) when \(L \to \infty\). Consequently, we are left only with the latter: \(f^1(\infty) = -\infty\) or \(f(\infty) = \infty\). Since according to (A12), \(\lambda \to -\infty\) means \(\lambda L \to -\infty\), we conclude that in case (i) as \(L \to \infty\), \(g_L \to -\infty\), and therefore \(\kappa_L \to 0\). Statement (4) of Sec. II is then established.

In case (ii), \(f^1(g_f) = -\infty\) as \(g_f \to g_L\). Thus, \(f(z) \to -\infty\) as \(z \to -\infty\), and that gives \(\lim_{z \to -\infty} g_L = g\).

Adopting convention (5) of Sec. II, we find
\[
\kappa = (g/g_L)^{\frac{1}{2}}, \tag{A19}
\]
where, as before, \(\alpha\) is some hadron radius chosen to set the scale of \(g = 1\), and \(g\) is the smallest positive-definite root of \(\beta(g) = 0\).

**APPENDIX B**

In this appendix, we shall prove a number of technical points mentioned in Sec. III concerning the Coulomb gauge. As we shall see, all of these proofs are quite elementary.

1. Conditions (31) and (32)

Suppose \(\vec{V}^2\) satisfies neither (31) nor (32). Under an infinitesimal gauge transformation, \(\vec{V}^2\) becomes \(\vec{V}^2(\theta)\) where
\[
\vec{V}^2(\theta) = \vec{V}^2 + C^{ab} g^a \partial^b \vec{V}^2 - g^{-1} \partial^a g^a \tag{B1}
\]
and \(\theta^a\) is an infinitesimal. We first show that it is always possible to choose \(\theta^a\) such that
\[
\nabla^2 g^2 - g^2 \vec{V} \cdot (C^{ab} g^a \partial^b \vec{V}^2) = g^2 \vec{V} \cdot \vec{V}^2 \tag{B2}
\]
and therefore
\[
\vec{V} \cdot \vec{V}^2(\theta) = (1 - \epsilon) \vec{V} \cdot \vec{V}^2, \tag{B3}
\]
where \(\epsilon = 0\). This can be readily established by expanding
\[
\theta^a = g \theta^a + g^2 \theta^a + g^3 \theta^a + \cdots \tag{B4}
\]
and noting that the inhomogeneous Laplace equations
\[
\nabla^2 \theta^a - \epsilon \vec{V} \cdot \vec{V}^a \tag{B5}
\]
and
\[
\nabla^2 \theta^a = C^{ab} \vec{V} \cdot (\partial^b \vec{V}^c) \text{ for } i > 1 \tag{B6}
\]
all have solutions. By using (B3) and varying \(\epsilon\) continuously, we can transform \(\vec{V} \cdot \vec{V}^a\) from any value to 0. Hence, (31) is satisfied.

Next, we assume \(\vec{V}^2\) to satisfy (31) but not (32), i.e., \(\vec{V} \cdot \vec{V}^2 = 0\), but \(\vec{V} \cdot \vec{V}^2 \neq 0\) on \(\delta\). Again, we consider an infinitesimal gauge transformation \(\vec{V} = \vec{V}^2(\theta)\), but now we want to choose \(\theta^a\) such that (31) remains satisfied,
\[
\vec{V} \cdot \vec{V}^2(\theta) = 0, \tag{B7}
\]
and, in addition
\[
\hat{n} \cdot \vec{V}^2(\theta) = (1 - \epsilon) \hat{n} \cdot \vec{V}^a \text{ on } \delta, \tag{B8}
\]
where \(\epsilon = 0\). By using (B4), we note that in order to achieve (B7) we must have
\[
\nabla^2 \theta^a = 0 \tag{B9}
\]
and
\[
\hat{n} \cdot \vec{V} \cdot \vec{V}^2(\theta) = \vec{V} \cdot \vec{V}^2 \text{ on } \delta. \tag{B9}
\]
Furthermore, besides (B6), we also need (for \(i > 1\))
\[
\hat{n} \cdot \vec{V} \cdot \vec{V}^a = C^{ab} \vec{V} \cdot (\partial^b \vec{V}^c) \text{ on } \delta. \tag{B10}
\]
For simplicity, let us first assume \(\delta\) to be a spherical surface of radius \(\alpha\). The general solution of (B8) can be written in terms of the spherical coordinates \((r, \alpha, \beta)\)
\[
\theta^a = \sum A_{lm}^a \rho^l Y_{lm}(\alpha, \beta), \tag{B10}
\]
where \(Y_{lm}\) is the usual spherical harmonics. Let
\[
\hat{n} \cdot \vec{V} = \sum B_{lm} Y_{lm}(\alpha, \beta) \text{ at } r = \alpha,
\]
where
\[
B_{lm} \alpha \equiv \int_{\alpha} \hat{n} \cdot \vec{V} R^{l-1} \sin \alpha \, d\alpha \, d\beta
\]
and
\[
= \int_{\alpha} \hat{n} \cdot \vec{V} R^{l-1} = 0.
\]
By choosing
\[
A_{lm} = \epsilon B_{lm} / (R^{l-1}) \text{ for } l > 1
\]
and leaving \(A_{l=0}\) arbitrary, the above \(\theta^a\) satisfies both (B8) and (B9). Let \(B_{lm}\) by any particular solution of
\[
\nabla^2 \theta^a = C^{ab} \vec{V} \cdot (\partial^b \vec{V}^c)
\]
and

185
\( \alpha_i^* = \theta_i^* = \frac{\theta_i}{r} \).

In order to have (B6) and (B10) for \( i = 1 \), \( \alpha_i^* \) must satisfy

\[ \nabla^2 \alpha_i^* = 0 \]

and

\[ \mathbf{n} \cdot \nabla \alpha_i^* = n \cdot (C_{\theta_0} \theta_{\phi} \nabla r - \nabla \theta_{\phi}^*), \]

which have exactly the same form as (B8) and (B9). Consequently, \( \alpha_i^* \) exists and so does \( \theta_i^* \). Likewise, we can construct \( \theta_{i_1}^*, \theta_{i_2}^*, \ldots, \) and therefore (B7) is satisfied. By varying \( \epsilon \) continuously, we can transform \( \mathbf{n} \cdot \nabla \mathbf{a} \) from any value to 0 on \( \mathcal{S} \). (When \( \mathcal{S} \) is not a spherical surface, by choosing the appropriate curvilinear coordinates, we can apply identical arguments as those given above.) Conditions (31) and (32) are then established.

2. Equations (42)-(44)

Consider a simple problem in classical electrostatics: A unit point charge is placed at \( \mathbf{r} = \mathbf{r}_0 \) inside a simply connected surface \( \mathcal{S} \). The dielectric constant is \( \kappa = 1 \) inside \( \mathcal{S} \), but \( \kappa = \kappa_\mathcal{S} = 0 \) outside. The electrostatic potential \( \psi(\mathbf{r}) \) in this problem is the Green’s function of \( \mathbf{F}, \mathbf{F}_0 \) defined by (33). Let \( V(\mathbf{F}) = V_{\text{in}}(\mathbf{F}) \) or \( V_{\text{out}}(\mathbf{F}) \) depending on whether \( \mathbf{F} \) is inside or outside \( \mathcal{S} \). Hence,

\[ -\nabla^2 V_{\text{in}}(\mathbf{F}) = 0 \]

and

\[ -\nabla^2 V_{\text{out}}(\mathbf{F}) = \delta(\mathbf{r} - \mathbf{F}_0). \]

The boundary conditions at the surface \( \mathcal{S} \) are

\[ V_{\text{in}}(\mathbf{F}) = V_{\text{out}}(\mathbf{F}) \]

and

\[ \mathbf{n} \cdot \nabla V_{\text{in}}(\mathbf{F}) = \kappa_\mathcal{S} \mathbf{n} \cdot \nabla V_{\text{out}}(\mathbf{F}), \]

where, as before, \( \mathbf{n} \) is the normal vector of \( \mathcal{S} \). We now expand \( V_{\alpha} \) (\( \alpha = \text{in or out} \)) in powers of \( \kappa_\mathcal{S} \):

\[ V_{\alpha} = \kappa_\mathcal{S}^{-1} V_{\alpha}^{(0)} + \kappa_\mathcal{S} V_{\alpha}^{(1)} + \kappa_\mathcal{S}^2 V_{\alpha}^{(2)} + \cdots. \]  

Equation (B11) becomes

\[ -\nabla^2 V_{\text{in}}^{(0)} = 0, \quad -\nabla^2 V_{\text{out}}^{(0)} = \delta(\mathbf{F} - \mathbf{F}_0), \]

\[ -\nabla^2 V_{\text{in}}^{(1)} = 0 \quad \text{for} \quad l \geq 2 \]

and

\[ -\nabla^2 V_{\text{out}}^{(m)} = 0 \quad \text{for} \quad m \geq 1. \]

The boundary condition (B12) requires that at the surface \( \mathcal{S} \)

\[ V_{\text{in}}^{(l)} = V_{\text{out}}^{(l)} \quad \text{for} \quad l, \]

\[ \mathbf{n} \cdot \nabla V_{\text{in}}^{(m)} = 0, \]

and

\[ \mathbf{n} \cdot \nabla V_{\text{out}}^{(m)} = \mathbf{n} \cdot \nabla V_{\text{out}}^{(m-1)} \quad \text{for} \quad m \geq 1. \]

Let \( \mathbf{E}_{\alpha}^{(l)} \) be the electrostatic field associated with \( V_{\alpha}^{(l)} \):

\[ \mathbf{E}_{\alpha}^{(l)} = \frac{-\nabla V_{\alpha}^{(l)}}{\kappa_\mathcal{S}}. \]

Now \( \mathbf{E}_{\alpha}^{(0)} \) is, by definition, irrotational; from (B14) it is also divergence-free, and from (B15) its normal component is zero on \( \mathcal{S} \). Hence, \( V_{\text{in}}^{(0)} = 0 \) and therefore

\[ V_{\text{in}}^{(0)}(\mathbf{F}) = \text{constant}. \]

So far, \( \mathbf{F}_0 \) is assumed to be inside \( \mathcal{S} \). When \( \mathbf{F} \) is also inside \( \mathcal{S} \), the Green’s function \( G_{\mathcal{S}}(\mathbf{F}, \mathbf{F}_0) = V_{\text{in}}(\mathbf{F}) \) and \( G_{\mathcal{S}}(\mathbf{F}, \mathbf{F}_0) = V_{\text{out}}(\mathbf{F}) \) where \( G_{\mathcal{S}} \) is defined by (41). Thus, (B17) shows that when \( \mathbf{F} \) and \( \mathbf{F}_0 \) are both inside \( \mathcal{S} \), \( G_{\mathcal{S}}(\mathbf{F}, \mathbf{F}_0) \) is independent of \( \mathbf{F} \); by symmetry \( G_{\mathcal{S}}(\mathbf{F}, \mathbf{F}_0) \) is also independent of \( \mathbf{F}_0 \), and therefore (42) follows. By following exactly the same argument, but for \( \mathbf{F}_0 \) outside \( \mathcal{S} \), we find \( G_{\mathcal{S}}(\mathbf{F}, \mathbf{F}_0) \) to be independent of \( \mathbf{F} \), so long as \( \mathbf{F} \) is inside \( \mathcal{S} \). These properties may be summarized as follows:

\[ G_{\mathcal{S}}(\mathbf{F}, \mathbf{F}') \] is independent of \( \mathbf{F} \) (or \( \mathbf{F}' \)) whenever \( \mathbf{F} \) (or \( \mathbf{F}' \)) is inside \( \mathcal{S} \).

To determine the constant in (B18), we have to examine \( V_{\text{out}}^{(0)} \). Since according to (B14) and (B15),

\[ -\nabla^2 V_{\text{out}}^{(0)} = 0 \quad \text{outside} \quad \mathcal{S} \quad \text{and} \quad V_{\text{out}}^{(0)} = V_{\text{out}}^{(0)} = \text{constant on} \quad \mathcal{S}, \]

one sees that \( V_{\text{out}}^{(0)}(\mathbf{F}) \) is the same as the electrostatic potential in free space outside a perfect conductor of surface \( \mathcal{S} \). Let \( \sigma(\mathbf{F}) \) be the surface charge density of that problem. We have \( -\nabla^2 V_{\text{out}}^{(0)} = \sigma(\mathbf{F}) \)

\[ = \mathbf{n} \sigma \quad \text{on} \quad \mathcal{S}. \]

From (B14) and (B15) we see that

\[ -\nabla^2 V_{\text{in}}^{(l)} = \delta(\mathbf{F} - \mathbf{F}_0) \]

and

\[ -\nabla^2 V_{\text{out}}^{(l)} = 0 \quad \text{on} \quad \mathcal{S}. \]

Hence,

\[ \mathbf{n} \cdot \nabla V_{\text{in}}^{(l)} = 0, \]

\[ \mathbf{n} \cdot \nabla V_{\text{out}}^{(l)} = 0. \]
\[ \int_{\Omega} dS \cdot \nabla^2 \nabla_{\Omega} \frac{dS}{\Omega} = 1 \]

where \( dS \) is the surface element of \( \Omega \) and \( \Omega \) is the volume inside \( \Omega \). From (41), it follows that \( G_0(\vec{r}, \vec{r}_0) = V_{\Omega}^{(0)}(\vec{r}) \) when \( \vec{r} \) is inside \( \Omega \). Consequently, inside \( \Omega \), \( G_0(\vec{r}, \vec{r}_0) \) is the electrostatic potential generated by a positive unit charge at \( \vec{r} = \vec{r}_0 \) with the boundary condition (B19) on \( \Omega \).

If \( \psi = \vec{V} = 0 \) outside \( \Omega \), then \( j^0 = j^\rho = 0 = \text{also outside} \). Hence, according to (37), (39), (41), and (B18), as \( \kappa_\omega \to 0 \),

\[ V_{\Omega}^{(0)}(\vec{F}) = \kappa_\omega^{-1} g_0(\vec{F}, \vec{R}) Q^2 \left[ 1 + \frac{4}{3} \kappa_\omega \right] + O(\kappa_\omega), \tag{B20} \]

where \( \vec{R} \) is any position vector on \( \Omega \). According to (B18), \( G_0(\vec{F}, \vec{F}') = G_0(\vec{F}, \vec{R}) \) whenever \( \vec{F}' \) is inside \( \Omega \). Now, for a color singlet \( Q^2 = 0 \), \( V_{\Omega}^{(0)}(\vec{F}) \) is finite as \( \kappa_\omega \to 0 \). Equation (43) then follows.

To derive (44), we assume \( \omega \) to be a spherical surface of radius \( R \). As before, a unit charge is placed at \( \vec{r} = \vec{r}_0 \) inside \( \Omega \). It is straightforward to show that for \( r < R \)

\[ G(\vec{F}, \vec{r}_0) = V_{\text{in}}(\vec{F}) = V_{\text{in}}(\vec{F}) = (4\pi)^{-1} \sum_{l=0}^{\infty} \frac{2l+1}{l} \frac{\sin \frac{r_0}{R}}{l + 1 + \frac{l}{\kappa_\omega}} P_l(\cos \theta), \tag{B21} \]

and for \( r > R \)

\[ G(\vec{F}, \vec{r}_0) = V_{\text{out}}(\vec{F}) = V_{\text{out}}(\vec{F}) = (4\pi)^{-1} \sum_{l=0}^{\infty} \frac{2l+1}{l} \frac{\sin \frac{r_0}{R}}{l + 1 + \frac{l}{\kappa_\omega}} P_l(\cos \theta), \tag{B22} \]

where \( r \) and \( r_0 \) are, respectively, the magnitudes of \( \vec{F} \) and \( \vec{r}_0 \), \( \theta \) is the angle between them, and the \( P_l \)'s are Legendre polynomials. Taking the limit \( \kappa_\omega \to 0 \) and by using (41), we obtain (44).

As \( \kappa_\omega \to 0 \), an alternative form of (44) can be given in terms of image charges. We observe that for \( r < R \), (B21) becomes

\[ G(\vec{F}, \vec{r}_0) = V_{\text{in}}(\vec{F}) = V_{\text{in}}(\vec{F}) = (4\pi)^{-1} \sum_{l=0}^{\infty} \frac{2l+1}{l} \frac{\sin \frac{r_0}{R}}{l + 1 + \frac{l}{\kappa_\omega}} P_l(\cos \theta), \tag{B23} \]

and for \( r > R \)

\[ G(\vec{F}, \vec{r}_0) = V_{\text{out}}(\vec{F}) = V_{\text{out}}(\vec{F}) = (4\pi)^{-1} \sum_{l=0}^{\infty} \frac{2l+1}{l} \frac{\sin \frac{r_0}{R}}{l + 1 + \frac{l}{\kappa_\omega}} P_l(\cos \theta), \tag{B24} \]

Thus, as \( \kappa_\omega \to 0 \), \( V_{\text{in}}(\vec{F}) \), apart from an additive constant, is the electrostatic potential due to a unit charge at \( \vec{r} = \vec{r}_0 \) inside the sphere, an image point charge of magnitude \( (R/r_0) \) at \( \vec{r} = R^2 \vec{r}_0 / r_0^2 \) outside the sphere, and a continuous line distribution, also outside the sphere, along the \( \vec{F} \) direction from \( \vec{F} = R^2 \vec{r}_0 / r_0^2 \) to \( \infty \);

\[ V_{\text{out}}(\vec{F}) \] is that due to two point charges, one of magnitude \( 1 - 2\kappa_\omega \) at \( \vec{r} = \vec{F} \) and the other of magnitude \( 2\kappa_\omega \) at \( \vec{r} = \vec{r}_0 \), plus a continuous line distribution of image charges along \( \vec{r}_0 \) from \( \vec{F} = \vec{r}_0 \) to \( \vec{F} \), all inside the sphere. The integrals in (B23) can be readily integrated. We find for \( r < R \)

\[ G(\vec{F}, \vec{r}) = V_{\text{in}}(\vec{F}) = V_{\text{in}}(\vec{F}) = (4\pi)^{-1} \left\{ 1 - \frac{2\kappa_\omega}{1 + \frac{1}{\kappa_\omega R}} + \frac{R r_0}{|\vec{F} - \vec{r}_0|} + \frac{R r_0}{|\vec{r} - \vec{F}|} - \frac{R}{|\vec{F} - \vec{r}_0|} \right\} + O(\kappa_\omega) \]

and for \( r > R \)

\[ G(\vec{F}, \vec{r}) = V_{\text{out}}(\vec{F}) = V_{\text{out}}(\vec{F}) = (4\pi)^{-1} \left\{ 1 - \frac{2\kappa_\omega}{1 + \frac{1}{\kappa_\omega R}} + \frac{R r_0}{|\vec{F} - \vec{r}_0|} + \frac{R r_0}{|\vec{r} - \vec{F}|} - \frac{R}{|\vec{F} - \vec{r}_0|} \right\} + O(\kappa_\omega). \]

APPENDIX C

Several properties of the modified Lagrangian density \( \mathcal{L}' \) and the related Green's function \( G(\vec{F}, \vec{F}') \) are stated in Sec. III D and Sec. IV A; these will be established in this appendix.

1. Equations (59) and (60). The proof is essentially identical to that given in Appendix B2. Let
\[ V(\mathbf{F}) \text{ be the solution of} \]
\[ [ - \nabla \cdot (k \nabla) + k(f' \alpha) ] V(\mathbf{F}) = 0 \]
that equals the Green's function \( G(\mathbf{F}, \mathbf{F}_0) \), defined by (57),
\[ G(\mathbf{F}, \mathbf{F}_0) = V(\mathbf{F}). \]
Denote \( V(\mathbf{F}) = V_{\text{in}}(\mathbf{F}) \) or \( V_{\text{out}}(\mathbf{F}) \) depending on whether \( \mathbf{F} \) is inside or outside the surface \( s \). For \( \mathbf{F}_s \) inside \( s \), because of (23), (24), and (52), (C1) becomes
\[ (\nabla^2 - \mu^2) \mathbf{V}_{\text{out}}(\mathbf{F}) = 0 \]
and
\[ -\nabla^2 V_{\text{in}}(\mathbf{F}) = \delta^3(\mathbf{F} - \mathbf{F}_0) \]
with the boundary conditions at the surface \( s \):
\[ V_{\text{in}}(\mathbf{F}) = V_{\text{out}}(\mathbf{F}) \]
and
\[ \cdot \mathbf{n} \cdot \nabla V_{\text{in}}(\mathbf{F}) = \kappa_\alpha \cdot \mathbf{n} \cdot \nabla V_{\text{out}}(\mathbf{F}), \]
where, as in (B12), \( \mathbf{n} \) is the normal vector of \( s \). Just as in (B13), we may expand \( V_\alpha (\alpha = \text{in or out}) \) in powers of \( \kappa_\alpha \):
\[ V_\alpha = \kappa_\alpha^{-1} V_{\text{in}}^{(0)} + V_{\text{in}}^{(1)} + \kappa_\alpha^{-2} V_{\text{in}}^{(2)} + \cdots, \]
where because of (C3)
\[ -\nabla^2 V_{\text{in}}^{(0)} = 0, \quad -\nabla^2 V_{\text{in}}^{(1)} = \delta^3(\mathbf{F} - \mathbf{F}_0), \]
\[ -\nabla^2 V_{\text{in}}^{(l)} = 0, \quad \text{for } l \geq 2 \]
and
\[ (\nabla^2 - \mu^2) V_{\text{out}}^{(m)} = 0, \quad \text{for } m \geq 0. \]
The boundary condition (C4) becomes
\[ V_{\text{in}}^{(l)} = V_{\text{out}}^{(l)}, \quad \text{for all } l \]
\[ \cdot \mathbf{n} \cdot \nabla V_{\text{in}}^{(0)} = 0, \]
\[ \cdot \mathbf{n} \cdot \nabla V_{\text{in}}^{(m)} = \cdot \mathbf{n} \cdot \nabla V_{\text{out}}^{(m-1)}, \quad \text{for } m \geq 1 \]
at the surface \( s \).
Since \( V_{\text{in}}^{(l)} \) satisfies exactly the same equations here as in Appendix B2, by following identical arguments as those given after (B15), one establishes, as in (B17),
\[ V_{\text{in}}^{(0)}(\mathbf{F}) = \text{constant}, \]
and, as in (B18),
\[ G_\alpha(\mathbf{F}, \mathbf{F}') \text{ is independent of } \mathbf{F} \text{ (or } \mathbf{F}') \text{ whenever} \]
\( \mathbf{F} \) (or \( \mathbf{F}' \)) is inside \( s \).
Hence, (59) follows.
If \( \cdot \mathbf{F} = 0 \) outside \( s \), then \( J'_0 = J'_e = 0 \) also outside \( s \). Hence, (56) reduces to (B20), i.e.,
\[ V_{\text{in}}^{(0)}(\mathbf{F}) = \kappa_\alpha^{-1} g G_0(\mathbf{F}, \mathbf{F}_0) Q^a \]
\[ + g \int G_1(\mathbf{F}, \mathbf{F}') d^3r' \partial^3 r' + O(\kappa), \]
where, as before, \( \mathbf{R} \) is any position vector on \( s \).
According to (C9), \( G_0(\mathbf{F}, \mathbf{F}') = G_0(\mathbf{F}, \mathbf{R}) \) whenever \( \mathbf{F} \) is inside \( s \). For a color singlet, \( Q^a = 0 \); hence (C10) becomes
\[ V_{\text{in}}^{(0)}(\mathbf{F}) = g \int G_1(\mathbf{F}, \mathbf{F}') d^3r' \partial^3 r' + O(\kappa), \]
which is finite when \( \kappa_\alpha = 0 \).
To derive (60), let us assume \( s \) to be a spherical surface of radius \( R \). By using (C6)–(C8), one sees that \( V_{\text{in}}^{(0)}(\mathbf{F}) \) is given by
\[ V_{\text{in}}^{(0)} = \frac{q e^{-r}/r}{2}, \]
where \( q \) is a constant. The function \( V_{\text{in}}^{(0)} \) satisfies
\[ -\nabla^2 V_{\text{in}}^{(0)} = \delta^3(\mathbf{F} - \mathbf{F}_0) \]
and
\[ -\mathbf{n} \cdot \nabla V_{\text{in}}^{(0)} = \sigma = \text{constant on } s. \]
Because of the Gauss theorem, \( \sigma = (4\pi R^{-1}) \) when \( s \) is a spherical surface of radius \( R \). Since \( \cdot \mathbf{n} \cdot \nabla V_{\text{in}}^{(0)} = \cdot \mathbf{n} \cdot \nabla V_{\text{out}}^{(0)} \) on \( s \), we determine the constant in (C12): \( q = (4\pi)^{-1}(1 + \mu R)^{-1} \exp(\mu R) \). For \( r < R \),
\[ G_0(\mathbf{F}, \mathbf{F}_0) = V_{\text{in}}^{(0)}(\mathbf{F}) = (4\pi R^{-1})^{-1}(1 + \mu R)^{-1} \]
and for \( r > R \)
\[ G_0(\mathbf{F}, \mathbf{F}_0) = V_{\text{in}}^{(0)}(\mathbf{F}) = (4\pi R^{-1})^{-1}(1 + \mu R)^{-1} e^{-r}(r-R). \]
By setting \( \mu = 0 \), we recover the previous form of \( G_0(\mathbf{F}, \mathbf{F}_0) \) discussed in Appendix B.
In the present case of a spherical surface, (C13) is identical to (B19); the solution \( V_{\text{in}}^{(0)} \) must also be the same, apart from an additive constant which is not determined by (C13). Thus, (60) is established.
For an arbitrary surface and assuming \( \mu \gg \text{hadron mass} \), one has
\[ \cdot \mathbf{n} \cdot \nabla V_{\text{out}}^{(0)} = -\mu V_{\text{out}}^{(0)} \text{ on } s. \]
In this approximation, since \( \cdot \mathbf{n} \cdot \nabla V_{\text{out}}^{(0)} = \cdot \mathbf{n} \cdot \nabla V_{\text{in}}^{(0)} \) and \( V_{\text{in}}^{(0)} = V_{\text{in}}^{(0)} \) is constant on \( s \), the normal electric field defined by (C13) is a constant for any surface. Hence for \( \mathbf{F} \) and \( \mathbf{F}_0 \) both inside \( s \), \( G_0(\mathbf{F}, \mathbf{F}_0) = V_{\text{in}}^{(0)}(\mathbf{F}) \) is the electrostatic potential generated by a positive unit charge at \( \mathbf{F} = \mathbf{F}_0 \), with the normal component of its electric field being a constant on the surface.
We note that for nonspherical surfaces, even inside the Green's function $G_1(\mathbf{r}, \mathbf{r}_p)$ determined by (C17) can be quite different from that derived in Appendix B; however, their difference must be of the form $f(\mathbf{F}) + f(\mathbf{F}_p)$ due to the reciprocity relation. [Compare (C17) with the derivation following (B19) in Appendix B.] As mentioned earlier, outside the $G_1(\mathbf{r}, \mathbf{r}_p)$ function of Appendix B has a long tail, which leads to a Van der Waals-type force between hadrons, in violation of experimental results.

2. Equation (62). Inside $s$, since $\sigma = 0$ and $\kappa = 1$, the Lagrangian density is locally gauge invariant. In the Coulomb gauge, according to (31) and (32), $\nabla \cdot \mathbf{V} = 0$ inside $s$ and $\hat{n} \cdot \mathbf{V} = 0$ on $s$.

To derive (62), we first show that it is always possible to choose a solution $\phi^s$ of the last equation of (61),

$$\nabla \cdot \phi^s = 0 \quad \text{on } s,$$  

(C18)

such that

$$\hat{n} \cdot \nabla \phi^s = 0 \quad \text{on } s.$$  

(C19)

To see this, let us assume that $\phi^s \equiv \phi^s_0$ is a solution of (C18) inside $s$, i.e.,

$$\nabla \cdot \phi^s_0 = -g \nabla \cdot (C^{\alpha \beta} \mathbf{V}^\alpha),$$  

(C20)

but $\hat{n} \cdot \nabla \phi^s_0 \neq 0$ on $s$. For simplicity, let $s$ be a spherical surface of radius $R$ and let $(r, \alpha, \beta)$ be the spherical coordinates of $F$. Expand

$$\hat{n} \cdot \nabla \phi^s_0 = \sum_{i, m} A^s_{i m} Y_{i m}(\alpha, \beta) \quad \text{at } r = R,$$  

(C21)

where the $Y_{i m}$'s are the spherical harmonics. Because of (C20) and (32),

$$A^s_{i m} = \int_{\Omega} \nabla \phi^s_0 d^3r = 0,$$  

(C22)

where $\Omega$ denotes the volume inside $s$. Now choose $\phi^s$ to be a solution of $\nabla \phi^s = 0$ inside $s$. Hence

$$\int_{\Omega} \bar{E}_0 \cdot \bar{E}_0 d^3r = \int_{\Omega} V_0^2 \mathbf{V} \cdot \mathbf{V} d^3S + \int_{\Omega} V_0^2 (j_0 + C^{\alpha \beta} \nabla \cdot \mathbf{V}) d^3r$$

and

$$\int_{\Omega} (\mathbf{E}_0 \cdot \mathbf{V}^s + i \varphi \psi - \mathcal{L}_N) d^3r = \int_{\Omega} \mathcal{K} d^3r - \frac{1}{2} \int (V_0^2 \hat{n} \cdot \nabla V_0^s)_{in} d^3S,$$  

(C31)

where $\mathcal{K}$ is given by (66), $d^3S$ is the surface element of $s$, and $(\mathcal{L}_N)$ refers to the value of the inner side of $s$. Now, outside $s$, because of (C25), Eq. (55) becomes $(\varphi^2 - \mu^2) V_0^s = 0$. Hence,

$$\int_{\Omega} \mathcal{L}_N d^3r = \frac{1}{2} \int_{\Omega} (V_0^2 \hat{n} \cdot \nabla V_0^s)_{out} d^3S.$$  

(C32)
Furthermore, at the surface \((V^\nu)^{\text{out}}_a = (V^\nu)^{\text{in}}_a\) and 
\(\kappa_a(\vec{\nabla} \cdot \nabla) (V^\nu)^{\text{out}}_a = (\vec{\nabla} \cdot \nabla) (V^\nu)^{\text{in}}_a\). Hence, \(H\) is the same as \(H_0\) given by (65).

4. "Radiation field" inside the hadron. The Lagrangian density \(\mathcal{L}'\) given by (51) is not locally gauge invariant. As mentioned before, inside the hadron, \(\sigma = 0\) and therefore \(\mathcal{L}' = \mathcal{L}\) given by (10), which is locally gauge invariant. Outside, \(\sigma = \sigma_{\text{vac}} \neq 0\). Hence, local gauge invariance is broken. So long as we restrict ourselves to the inside region, we can perform local gauge transformations. However, in order to make a connection with the outside solution, it is sometimes necessary to transform the inside solution to a specific gauge; then one can satisfy the continuity condition at the surface.

As an example, let us consider the free radiation fields \(\vec{V}_a^\pm, \vec{V}_b^\pm\) and \(\vec{V}_b^\pm, \vec{V}_b^\pm\) of the TE and TB modes discussed in Sec. IV. B. For these free field solutions, we may set \(g = 0\). Equation (64) becomes
\[
\frac{\partial}{\partial x_\mu} V_{a\mu} = 0 \quad \text{inside} \delta,
\]
and
\[
\frac{\partial}{\partial x_\mu} V_{a\mu} - \mu^2 V_a^\pm \equiv 0 \quad \text{outside} \delta,
\]
where for simplicity the color index is omitted and since \(g = 0\), \(V_{a\mu} = (\partial V_a^\mu/\partial x_\mu) - (\partial V_a^\mu/\partial x_\mu)\). Assume \(\delta\) to be a spherical surface of radius \(R\). For \(r < R\), the solution for the TE or TB, mode may be written as \((\lambda = E \text{ or } B)\)
\[
\vec{V}(r) = \vec{V}_a(r, \theta, \phi) e^{i(kr)}
\]
and
\[
V_a^\pm(r) = 0,
\]
where \(\vec{V}_a(r, \theta, \phi)\) is given by (79) or (80). In spherical coordinates \(\vec{r} = (r, \theta, \phi)\), since \(\vec{r} \cdot \vec{r} = 0\) we have
\[
\frac{\partial V_a^\pm}{\partial r} + \frac{1}{r} \frac{\partial V_a^\pm}{\partial \theta} + \frac{1}{r \sin \theta} \frac{\partial V_a^\pm}{\partial \phi} = 0
\]
and at \(r = R\), because of (77),
\[
V_a^\pm = 0.
\]
In the outside region, the lower equation in (C34) gives
\[
\frac{\partial V_a^\pm}{\partial x_\mu} = 0;
\]
for \(\mu > R^{-1}\), its solution near the surface \((r = R\pm)\) must be of the form
\[
V_a^\pm(r, \theta, \phi, t) = (R/r) V_a(R, \theta, \phi, t) e^{-i\omega r - ia},
\]
where
\[
M = \mu [1 + O(\mu R)^{-2}]]
\]
In order to connect the outside solution with the inside solution, we set the outside solution \(V_a^\pm\),
\[
V_a^\pm = (V_a^\pm)_{\text{in}}
\]
outside \(\delta\),
\[
V_a^\pm = (V_a^\pm)_{\text{in}}
\]
and
\[
V_a^\pm = -\frac{1}{\mu} \left( \frac{\partial V_a^\pm}{\partial r} \right)_{R},
\]
where \(V_a^\pm\) is the inside solution given by (C35), and \((\partial V_a^\pm/\partial r)_{R}\) refers to its value at \(r = R\). The last term in (C40) is needed so that at \(r = R\), neglecting \(O(1/\mu R)\), \(\partial V_a/\partial x_\mu = 0\). To the zeroth order in \((\mu R)^{-1}\), (C40) and (C35) give the same field at \(r = R\). Hence, to order the continuity condition at the surface is satisfied.

It is of interest to examine the first-order correction. To the first order in \((\mu R)^{-1}\), we must modify the inside solution (C35) by a local gauge transformation (which is allowed since the inside Lagrangian density is locally gauge invariant):
\[
\vec{V} = \vec{V}_a^\pm + \mu^{-1} \vec{\chi}
\]
and
\[
V_a^\pm = -\mu^{-1} \chi = \chi + \mu^{-1}(ik) \chi,
\]
where \(\chi\) is analytic inside the sphere \((r < R)\) and it should satisfy the following boundary conditions at the surface:
\[
\chi(r, \theta, \phi, t) = 0
\]
and
\[
\left( \frac{\partial \chi}{\partial r} \right)_{R} = -\left( \frac{\partial V_a^\pm}{\partial r} \right)_{R}.
\]
Consequently, (C40) and (C41) satisfy the continuity condition at \(r = R\). [That such an \(\chi\) exists can be easily seen by assuming, e.g., \(\chi \propto r^l(r^2 - R^2)\times Y_m(\theta, \phi)\) for the appropriate TE or TB solution.]

We note further that as \(\kappa_a \to 0\) the integral \(\kappa_a(f^a)^2 \vec{V}_a^\pm \overline{\vec{V}}^\pm\) over the outside region becomes zero. Hence, the constraint \(\vec{V}_a^\pm = 0\) outside \(\delta\) is derived by using the Lagrangian (51) provided that we take the limit \(\kappa_a \to 0\) first and then \(f^a\sigma_mc = \mu \to -\infty\).

APPENDIX D

In this appendix, we give the derivation of (90), which leads to the experimental upper limit of \(\kappa_a\) discussed in Sec. V.D. From (69), the mass of a free quark is given approximately by
\[
m_q \approx \frac{2k^2 G}{3 \kappa_a} + \frac{4\pi}{3} p R^3,
\]
where \(p\) is defined in (12). Here, for simplicity, we neglect the surface tension \(s\). According to (C14)
where we discuss the two limiting cases: 
\[ \mu \ll (4\pi \rho / m_p)^{1/2} \] and 
\[ \mu \gg (4\pi \rho / m_p)^{1/2}. \]

1. \( \mu \ll (4\pi \rho / m_p)^{1/2} \). In this case, \( G_0 \approx (4\pi R)^{-1} \).

By setting \( dm / dr = 0 \), we find

\[ m_\rho = \frac{4}{3} \left( \frac{2\alpha}{3K_{\pi \rho}} \right)^{3/4} \left( 4\pi \rho / m_p \right)^{1/4}, \]

where \( \alpha = (4\pi)^{-1}g^2 \). Likewise, the proton mass is

\[ m_p = \frac{4}{3} m_\rho \left( \frac{4\pi \rho}{m_p} \right)^{1/4}, \]

where \( \rho_0 \) is given by (22). Hence,

\[ \kappa = \frac{2}{9\rho_0} \left( \frac{m_p}{m_\rho} \right)^{1/4}. \]
Operator ordering and Feynman rules in gauge theories

N. H. Christ and T. D. Lee
Columbia University, New York, New York 10027
(Received 10 April 1980)

The ordering of operators in the Yang-Mills Hamiltonian is determined for the \( V_\alpha = 0 \) gauge and for a general noncovariant gauge \( \chi (V_\alpha) = 0 \), with \( \chi \) a linear function of the spatial components of the gauge field \( V_\alpha \). We show that a Cartesian ordering of the \( V_\alpha = 0 \) gauge Hamiltonian defines a quantum theory equivalent to that of the usual, covariant-gauge Feynman rules. However, a straightforward change of variables reduces this \( V_\alpha = 0 \) gauge Hamiltonian to a \( \chi (V_\alpha) = 0 \) gauge Hamiltonian with an unconventional operator ordering. The resulting Hamiltonian theory, when translated into Feynman graphs, is shown to imply new nonlocal interactions, even in the familiar Coulomb gauge.

I. INTRODUCTION

Gauge theories have a wide range of applications in physics, from quantum electrodynamics (QED) to quantum chromodynamics (QCD) and the unifying theory of weak and electromagnetic interactions. Because the gauge transformations contain arbitrary functions of time, the usual canonical quantization procedure can only be carried out in a specific gauge. It is natural to inquire whether there are rules to ensure that the quantum theories in different gauges are indeed the same. As we shall see, this question is closely connected with the ordering problem of operators, especially in the non-Abelian Yang-Mills theory because of the intrinsic nonlinear nature of the interaction.

In this paper, we will show that the Yang-Mills theory has a Cartesian realization in the \( V_\alpha = 0 \) gauge; in this gauge the naive ordering of operators in the Hamiltonian is correct. We may then go from the \( V_\alpha = 0 \) gauge to other gauges such as the axial gauge, Coulomb gauge, or covariant gauges via either the operator or the path-integration formalism, thereby resolving whatever ambiguities may arise in these gauges.

The fact that the Euler-Lagrange equations of motion may allow arbitrary functions of time is by no means restricted to relativistic field theories. For convenience of nomenclature, all such theories will be referred to as gauge theories. In the next section, we give one of the simplest examples of such a system.

In Sec. III we review the operator formulation of the Yang-Mills theory in the \( V_\alpha = 0 \) gauge and establish our notation. Next, in Sec. IV, we exploit the residual symmetry of the \( V_\alpha = 0 \) gauge to make a change of coordinates from the Cartesian basis provided by the gauge potentials \( V_\alpha (x) \) to new variables \( \phi_\alpha (x) \) and \( A_\mu (x) \): \( \phi_\alpha (x) \) are pure gauge variables while the gauge potentials \( A_\mu (x) \) obey the constraint \( \chi (A_\mu) = 0 \). When expressed in terms of these new variables, the gauge-invariant sector of the theory is recognized as identical to the Yang-Mills theory formulated in a noncovariant, \( \chi (A_\mu) = 0 \) gauge with a specific ordering of operators in the Hamiltonian. This discussion is in precise analogy with the treatment of the simple mechanical system given in Sec. II. As shown in the Appendix, many of our formulas are also identical to those of rigid-body rotation, when one changes from the laboratory frame to the rotating-body frame.

In Sec. V we justify our assertion that the \( V_\alpha = 0 \) gauge provides a Cartesian realization of the quantum Yang-Mills theory: Using functional integration we demonstrate the equivalence of that theory and the usual, covariant-gauge Feynman rules.

The Hamiltonian formulation of the quantum Yang-Mills theory obtained in Sec. IV is not a convenient one for weak-field perturbation theory. In Sec. VI, this Hamiltonian operator theory is translated into the Lagrangian, path-integral language with careful attention paid to the question of operator ordering. This path-integral description implies new nonlocal interactions, called \( U_\alpha + U_\beta \) in our paper, that must be added to the usual Feynman rules. Finally, in the conclusion we show explicitly how to equate our Hamiltonian to that obtained by Schwinger for the Coulomb gauge. Although the \( U_\beta \) term is new, the \( U_\alpha \) term was derived by Schwinger in 1962; both have been left out in the conventional treatment of Coulomb-gauge Feynman rules.

II. A SIMPLE MECHANICAL EXAMPLE

Let us consider a point particle in a three-dimensional space at position \( \mathbf{r} \). Its Lagrangian is

\[
L = \frac{1}{2} \left( \dot{\mathbf{r}} - \vec{\mathbf{q}} \times \dot{\mathbf{r}} \right)^2 - V(r),
\]

(2.1)

where \( \vec{\mathbf{q}} \) is another coordinate vector, but \( \mathbf{q} \) is absent in \( L \). As usual, the dot denotes the time de-
rivative and \( r = |\mathbf{r}| \). From (2.1), one sees immediately that \( L \) is invariant under the transformation
\[
\mathbf{r} \rightarrow \mathbf{r} + \epsilon \times \mathbf{r}
\]
and
\[
\mathbf{q} = \mathbf{q} + \epsilon \times \mathbf{q}
\]  
(2.2)
where \( \mathbf{q} = \mathbf{q}(t) \) can be an arbitrary infinitesimal vector function of time \( t \). Except for the \( \frac{1}{2} \mathbf{q} \times \mathbf{r} \) term, this would be the problem of a nonrelativistic charged particle moving in a central potential and under the influence of an external magnetic field.

We may further simplify the problem by imposing the constraint that \( \mathbf{r} \) lies in the \((x, y)\) plane and \( \mathbf{q} = \hat{z}q \), where \( \hat{z} \) is the unit vector along the \( z \) axis. Equation (2.1) becomes then
\[
L = \frac{1}{2} (x^2 + y^2) - (x\dot{y} - y\dot{x})q + \frac{1}{2} q^2 - V(r),
\]
(2.3)
where \( x \) and \( y \) are the Cartesian coordinates of \( \mathbf{r} \) which is now a two-dimensional vector. In terms of the polar coordinates \( x = r \cos \theta \) and \( y = r \sin \theta \), (2.3) can be written as
\[
L = \frac{1}{2} [\dot{r}^2 + r^2 (\dot{\theta} - q)^2] - V(r)
\]  
(2.4)
and (2.2) is simply the Abelian group of transformations
\[
\theta \rightarrow \theta + \epsilon(t)
\]
and
\[
q \rightarrow q + \epsilon(t),
\]
(2.5)
where \( \epsilon(t) \) can now be any finite function of \( t \).

The invariance group of this simple example shares with the gauge groups of QED or QCD the special feature that its elements contain arbitrary functions of \( t \). Consequently, the canonical procedure from the Lagrangian to the Hamiltonian and to quantization requires a specific choice of the "gauge." The Lagrange equations of motion can, of course, be written down without specifying the gauge. We find in polar coordinates
\[
\dot{q} - q = 0
\]  
(2.6)
and
\[
\dot{r} + \frac{dV}{dr} = 0.
\]
(2.7)

A. \( q = 0 \) gauge

Because of (2.5), any orbit \( \mathbf{r} = \mathbf{r}(t) \) and \( q = q(t) \) can be transformed to one in which \( q = 0 \) at all times. In this gauge, \( L = \frac{1}{2} \dot{r}^2 - V(r) \), the momentum \( p \) is \( \mathbf{r} \) and the Hamiltonian \( H \) is \( \frac{p^2}{2} + V(r) \). Thus, in quantum mechanics, \( \hat{p} = -i\hbar \hat{\nabla} \) and
\[
H = -\frac{i}{2} \mathbf{\nabla}^2 + V(r).
\]  
(2.8)

The angular momentum operator
\[
I = -i\left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right)
\]
(2.9)
commutes with \( H \). To be consistent with the equation of motion (2.6), only eigenstates of \( H \) with \( I = 0 \) should be accepted. These eigenstates are all \( \theta \) independent, and that leads to
\[
H = -\frac{1}{2r} \frac{d}{dr} \left( r \frac{d}{dr} + V(r) \right).
\]  
(2.10)

B. \( y = 0 \) gauge

From (2.5), we see that any orbit \( \mathbf{r} = \mathbf{r}(t) \) and \( q = q(t) \) can also be transformed to one with \( y = 0 \) at all times, albeit there are two branches: \( x \) can be \( > 0 \) or \( < 0 \), with \( x = 0 \) being the point that the Jacobian of the transformation is zero. In the \( y = 0 \) gauge, \( r^2 = x^2 \) and the Lagrangian (2.3) becomes
\[
L = \frac{1}{2} \dot{x}^2 + \frac{1}{2} q^2 - V(x),
\]  
(2.11)
where, for definiteness, we choose the branch of positive \( x \). Since the above \( L \) does not contain \( q \), we may follow the standard procedure to eliminate \( q \) through \( \delta L/\delta q = 0 \), which in the present example is simply \( \dot{x}q = 0 \). Hence, (2.11) becomes
\[
L = \frac{1}{2} \dot{x}^2 - V(x).
\]  
(2.12)

The conjugate momentum \( p \) is \( \dot{x} \) and the classical Hamiltonian is
\[
H = \frac{1}{2} p^2 + V(x).
\]  
(2.13)

In passing over to quantum mechanics, in order that the spectrum of this Hamiltonian be identical to that of (2.10), it is important not to treat \( x \) in the \( y = 0 \) gauge as a Cartesian coordinate; in (2.13) \( p^2 \) is the operator
\[
-\frac{1}{x} \frac{d}{dx} \left( x \frac{d}{dx} \right)
\]
and not \( -d^2/dx^2 \).

The \( q = 0 \) gauge is the analog of the \( V_c = 0 \) gauge in QED or QCD; the \( y = 0 \) gauge corresponds to either the usual axial gauge or the Coulomb gauge. In this simple example, the choice of which coordinate in what gauge is Cartesian can be determined by the one who makes up the problem. In the Yang-Mills theory, however, one is guided by the requirement of relativistic invariance. As we shall see, that leads to the choice of the \( V_c = 0 \) gauge as the starting point.

Before leaving this example, we note that if one wishes, one may choose a more general gauge in which an arbitrary function \( \chi(x, y, q) = 0 \), provided that any point in the \((x, y, q)\) space can be transformed onto the surface \( \chi(x, y, q) = 0 \) through the
gauge transformation (2.5). Since this mechanical example is such a simple one, we shall refrain from exhibiting further details, except to remark that in passing over to the path-integration formalism, it is useful to absorb a factor of $r^{1/2}$ into the state vectors so that $H$, given by (2.10), becomes

$$H = r^{1/2} H' r^{-1/2} = \frac{1}{2} \frac{d^2}{dr^2} + V(r) - \frac{1}{2} \frac{1}{\sigma^2},$$  
(2.14)

which changes the volume element from $r dr$ to $dr$ and adds to the potential $V(r)$ a new term $-(8 \pi^2)$. 

III. $V_0 = 0$ GAUGE

For definiteness, let us consider an SU$_3$ gauge theory consisting of a spin-$\frac{1}{2}$ fermion field $\psi$ which belongs to the $N$-dimensional representation of the SU$_3$ group and a spin-1 gauge field $V'_{\mu}$, with $\mu = 1, 2, \ldots, M$ and $M = N^2 - 1$. The Lagrangian density is

$$\mathcal{L} = \frac{1}{2} V'_{\mu} \nabla_{\nu} V'_{\nu} - \bar{\psi} i \gamma_{\nu} \gamma_{\mu} D_{\nu} \psi,$$  
(3.1)

where all repeated indices are summed over, a dagger denotes Hermitian conjugation,

$$V'_{\mu} = \frac{\partial}{\partial x_{\mu}} V'_r - \frac{1}{2} \frac{g}{\sigma} \text{Tr}(T^i T^m) D_{\mu} \psi,$$  
(3.3)

$$[T^i, T^m] = i \sigma^m T^i,$$

$$\text{with } \sigma^m \text{ the antisymmetric structure constants of the group algebra}. \text{ For simplicity, the fermion mass has been set equal to zero. The electric and magnetic fields } E'_{\mu} \text{ and } B'_{\mu} \text{ are given by the usual expressions}

$$E'_{\mu} = \frac{\partial}{\partial x_{\mu}} \bar{\psi} V_r = - \bar{\psi} i \gamma_{\nu} V'_{\nu} + g f^{kmn} V'^{k}_{\nu} V'^{m}_{\nu},$$

$$\text{and}

\epsilon_{ijk} B'_{ij} = \nabla_{i} V'_{j} - \nabla_{j} V'_{i} + g f^{kmn} V'^{k}_{\nu} V'^{m}_{\nu},$$

where $V'_{\nu} = -i V'_{\nu}$ and the subscripts $i, j, k$ denote the space indices which vary from 1 to 3. The Lagrangian equations of motion are

$$\gamma_{\nu} D'_{\mu} \psi = 0,$$  
(3.4)

$$\frac{\partial}{\partial x_{\mu}} V'_{\nu} + g f^{kmn} V'^{k}_{\nu} V'^{m}_{\nu} + I'_{\mu} = 0,$$  
(3.5)

where

$$I'_{\mu} = i \bar{\psi} \gamma_{\nu} \gamma_{\mu} T^i \psi.$$

It is convenient to introduce the matrix function

$$V_{\mu} = T^i V'_{\mu},$$  
(3.6)

The Lagrangian density (3.1) is invariant under the SU$_3$ transformation

$$V_{\mu} - u V'_{\mu} + \frac{i}{g} u \frac{\partial u^i}{\partial x_{\mu}}$$  
(3.7)

and

$$\psi - u \psi,$$  
(3.8)

where $u = u(x)$ is any $N \times N$ unitary matrix function of $x$ with $\det u = 1$. From any configuration $V_u = F_u (\vec{r}, t)$, we may choose $u$ to be the following time-ordered function:

$$u' (\vec{r}, t) = T \exp \left[ -i \int_0^t g F \delta (\vec{r}, t') \right],$$  
(3.9)

where $F = -i F_0$ and $T$ is the time-ordering operator. Hence, $u'$ satisfies

$$\frac{\partial u^i}{\partial t} = -i g \delta (\vec{r}, t).$$  
(3.10)

The transformation (3.7) then brings $V_u$ from the configuration $F_u (x)$ to the gauge in which

$$V_0 (x) = 0.$$  
(3.11)

In the $V_0 = 0$ gauge, $E'_{\mu} = - \bar{\psi} V'_{\mu}$ and (3.1) becomes

$$\mathcal{L} = \frac{1}{2} (\bar{\psi} V'_{\mu} - B'_{\mu} B^R_{\mu}) - \bar{\psi} i \gamma_{\nu} D_{\nu} \psi.$$  
(3.12)

The conjugate momentum of $V'_{\mu}$ is simply

$$\Pi'_{\mu} = \frac{\partial}{\partial x_{\mu}} \psi,$$  
(3.13)

and the Hamiltonian density is

$$\mathcal{H} = \frac{1}{2} \text{Tr} \left[ \Pi'_{\mu} \Pi^R_{\mu} - \epsilon \delta^m \text{Tr} (\delta^{i<m} T^i - \bar{\psi} \gamma_{\nu} \gamma_{\mu} \psi) \right].$$  
(3.14)

The usual canonical quantization procedure leads to

$$[V'_{\mu} (\vec{r}, t), \Pi^R_{\nu} (\vec{r}, t')] = i \delta_{\mu \nu} \delta^m \text{Tr} (\delta^{i<m} \vec{T} - \bar{T}).$$  
(3.15)

$$[\bar{\psi} (\vec{r}, t), \psi' (\vec{r}, t')] = \delta^m \text{Tr} (\delta^{i<m} \vec{T} - \bar{T}).$$  
(3.16)

The equal-time commutators between the $V'_{\mu}$'s and between the $\Pi'_{\mu}$'s are zero; likewise, the equal-time anticommutators between the $\psi$'s and between the $\psi'$'s are also zero.

Remnants of the original gauge transformations (3.7) and (3.8) remain important. In accordance with (3.6), we denote

$$V_{\mu} = T^i V'_{\mu},$$  
(3.17)

and

$$\Pi_{\mu} = T^i \Pi^R_{\mu}.$$  

It can be readily verified that the Hamiltonian den-
sity $\mathcal{K}$ and the commutation relations are invariant under a time-dependent $SU_N$ transformation
\begin{align}
V_i - uV_i u^t - \frac{i}{g} (\nabla_i u) u^t, \\
\Pi_i - u \Pi_i u^t,
\end{align}
and
$$\psi - u \psi,$$
where $u = u(\tilde{\mathcal{F}})$ can be any $N \times N$ unitary matrix function of $\tilde{\mathcal{F}}$ with $\det u = 1$. Since the $u(\tilde{\mathcal{F}})$'s are time-independent, the invariance group $\{u(\tilde{\mathcal{F}})\}$ is generated by the $\tilde{\mathcal{F}}$-dependent operators $\tilde{g}^t$ which are conserved:
$$\tilde{g}^t = J^t + \tilde{g}^t T^t \psi,$$
where
$$J^t = \frac{1}{g} D_i^{\mu} \Pi_i^\mu,$$
and
$$D_i^{\mu} = \delta^{\mu i} \nabla_i - g \tilde{g}^{i \nu} \nabla_i^\nu.$$
It is straightforward to verify
$$[J^t(\tilde{\mathcal{F}}, t), J^t(\tilde{\mathcal{F}'}, t)] = i f^{i \nu} \delta^t(\tilde{\mathcal{F}} - \tilde{\mathcal{F}'}) J^t(\tilde{\mathcal{F}}),$$
$$[g^t(\tilde{\mathcal{F}}, t), g^t(\tilde{\mathcal{F}'}, t)] = i f^{i \nu} \delta^t(\tilde{\mathcal{F}} - \tilde{\mathcal{F}'}) g^t(\tilde{\mathcal{F}}),$$
$$[g^t(\tilde{\mathcal{F}}, t), \phi^t(\tilde{\mathcal{F}'}, t)] = - \delta^t(\tilde{\mathcal{F}} - \tilde{\mathcal{F}'}) T^t \phi^t(\tilde{\mathcal{F}}),$$
and
$$[g^t(\tilde{\mathcal{F}}, t), \Pi^t(\tilde{\mathcal{F}'}, t)] = i f^{i \nu} \delta^t(\tilde{\mathcal{F}} - \tilde{\mathcal{F}'}) \Pi^t(\tilde{\mathcal{F}}),$$
where $\nabla_i$ is the differential operator with respect to $\tilde{\mathcal{F}}$. Consequently, $\tilde{g}^t$ commutes with the Hamiltonian $H = \int \mathcal{K} d^3 \tilde{r}$ and
$$\tilde{g}^t(\tilde{\mathcal{F}}, t) = i[H, \tilde{g}^t(\tilde{\mathcal{F}}, t)] = 0.$$
By commuting $H$ with $\psi$, we obtain the equation of motion (3.4) for $\psi$. Likewise, by commuting $H$ with $V_i$ and $\Pi_i$, we derive (3.5) for $\nu = i$ which can be 1, 2, or 3 but not 4. We note that $i \tilde{g}^t / g$ is identical to the left-hand side of (3.5) when $\nu = 4$. Thus, in order to be consistent with all the Lagrangian equations of motion, in the $V_0 = 0$ gauge we require all state vectors $| \rangle$ to satisfy
$$\tilde{g}^t | \rangle = 0.$$
In the Schrödinger picture the operators $V_i = V_i^t(\tilde{\mathcal{F}})$ and $\Pi_i = \Pi_i^t(\tilde{\mathcal{F}})$ are all $t$ independent. The state vector in the $V_i^t$ representation is the functional
$$\Psi(V_i) = \langle V_i^t | \rangle.$$
In this representation, the Hamiltonian $H$ is
$$H = \mathcal{K} + \mathcal{V},$$
where
$$\mathcal{K} = -\frac{1}{2} \int \frac{\delta}{\delta V_i^t(\tilde{\mathcal{F}})} \frac{\delta}{\delta V_i^t(\tilde{\mathcal{F}})} d^3 \tilde{r}$$
and
$$\mathcal{V} = \int \left( \frac{1}{2} B_i^t B_i^t - g l_i^t V_i^t + \phi^t \gamma_i \nabla_i \phi^t \right) d^3 \tilde{r}.$$ 
In Sec. IV we shall see how the introduction of curvilinear coordinates can be used to eliminate the constraint (3.27), in complete analogy to the passage from (2.8) to (2.10) in the simple mechanical example discussed in the previous section.

**IV. NONCOVARIANT GAUGES**

Let us start from the $V_0 = 0$ gauge quantum theory of Sec. III and show how to reach other noncovariant gauges such as the axial or Coulomb gauges. For notational clarity the gauge field in these other gauges will be referred to as $A_\mu = T^t A_i^\mu$. The spatial components of $A_\mu$ obey a gauge condition
$$\chi(A_\mu) = 0.$$
Among possible choices for $\chi$ one has
$$\chi(A_\mu) = \begin{cases} A_3 & \text{in axial gauge}, \\ \nabla_i A_i & \text{in Coulomb gauge}. \end{cases}$$
For simplicity we will treat $\chi$ as a linear homogeneous functional of $A_i$. As in (4.2), we assume $\chi = T^t \chi^{\prime t}$ to be an $N \times N$ Hermitian matrix with zero trace. Thus at any given space-time point, (4.1) expresses $M = N^2 - 1$ conditions:
$$0 = \chi'(A_i, \tilde{\mathcal{F}}) = \int d^3 \tilde{r} \langle \tilde{\mathcal{F}}, l | \Gamma_l | \tilde{\mathcal{F}'}, m \rangle A_i^\gamma(\tilde{\mathcal{F}'}, l),$$
where the matrix element of $\Gamma_l$ is real, and as before, the parameters $l$ and $m$ can vary from 1 to $M$. In addition, we will assume that for every field configuration $V_0(\tilde{\mathcal{F}}, t)$ in the $V_0 = 0$ gauge there exists a unique gauge-transformation $u(\tilde{\mathcal{F}}, t)$ such that
$$V_i = u A_i u^{-1} + \frac{i}{g} u \nabla_i u^{-1}.$$ 
If we view the $N \times N$ matrix $u$ as a function $u(\phi_\alpha)$ of $M$ group parameters $\phi_\alpha$, $1 \leq \alpha < M$, then the gauge transformation $u(\tilde{\mathcal{F}}, t)$ in turn specifies $M$ functions $\phi_\alpha(\tilde{\mathcal{F}}, t)$ such that
$$u(\tilde{\mathcal{F}}, t) = u(\phi_\alpha(\tilde{\mathcal{F}}, t)).$$
Equation (4.4) can thus be viewed as expressing the gauge field $V_i(\tilde{\mathcal{F}}, t)$ in terms of the curvilinear

196
coordinates \( A_i(\bar{F}, t), \phi_i(\bar{F}, t) \). For example, if our group is \( \text{SU}_2 \), then \( M = 3 \) and we might choose the \( \phi_i \)'s to be simply the three Euler angles. (See the Appendix for details.)

Because of (3.30), the \( V_i \)'s can be regarded as the Cartesian coordinates. We recall that in any coordinate transformation from a set of Cartesian coordinates \( q_1, q_2, \ldots \) to one of curvilinear coordinates \( Q_1(q), Q_2(q), \ldots \), the standard kinetic energy term in the Lagrangian may be written as

\[
K = \frac{1}{2} \sum_{\alpha, \beta} \dot{Q}_\alpha M_{\alpha \beta} \dot{Q}_\beta .
\]  

(4.6)

In the quantum Hamiltonian, the corresponding operator in the coordinate representation is

\[
\mathbf{H} = \frac{1}{2} |M|^{-1/2} \int d^3 \tau \int d^3 \tau' [A(\bar{F}) |M|^{1/2} M'(\bar{F}, \bar{F}') \phi(\bar{F}) + \phi(\bar{F}) |M|^{1/2} M'(\bar{F}, \bar{F}') A(\bar{F}')] + P(\bar{F}) |M|^{1/2} M'(\bar{F}, \bar{F}') \phi(\bar{F}) + \phi(\bar{F}) |M|^{1/2} M'(\bar{F}, \bar{F}') A(\bar{F}')] .
\]  

(4.10)

where \( \phi_\alpha \) and \( P_\alpha \) are the momenta conjugate to \( \phi_\alpha \) and \( A_\alpha \). In the coordinate representation of the Schrödinger picture, \( \phi_\alpha(\bar{F}) \) and \( A_\alpha(\bar{F}) \) are all \( t \) independent. Clearly,

\[
\dot{\phi}_\alpha(\bar{F}) = -i \frac{\delta}{\delta \phi_\alpha(\bar{F})} .
\]  

(4.11)

However, because of the constraints (4.2) obeyed by \( A_\alpha(\bar{F}) \), its conjugate momentum must be defined more carefully. We can expand \( A_\alpha(\bar{F}) \) in terms of a complete set of real orthonormal functions \( f_\alpha(\bar{F}) \), each obeying the same set of constraints

\[
\int d^3 \tau' \langle \tau | \Gamma_{\alpha} | \tau' \rangle \phi_{\alpha}(\bar{F}) f_{\alpha}(\bar{F}) = 0
\]  

(4.12)

for all \( \Gamma_\alpha \), \( \bar{F} \), and \( l \). The expansion is

\[
A_\alpha(\bar{F}) = \sum_\alpha \phi_\alpha f_\alpha(\bar{F}) .
\]  

(4.13)

The \( \phi_\alpha \)'s may now be viewed as the independent generalized coordinates. The momentum \( P_\alpha(\bar{F}) \) is then given by

\[
P_\alpha(\bar{F}) = \sum_\alpha f_\alpha(\bar{F}) \left( -i \frac{\delta}{\delta \phi_\alpha} \right) .
\]  

(4.14)

Thus, \( P_\alpha(\bar{F}) \) also obeys the constraints

\[
\chi'(P_\alpha, \bar{F}) = \int d^3 \tau' \langle \tau | l | \tau' \rangle \phi_{\alpha}(\bar{F}) \phi_{\alpha}(\bar{F}) = 0.
\]  

(4.15)

The following construction for the functions \( f_\alpha(\bar{F}) \) will be useful. We may view \( \langle \bar{F}, l | \Gamma_{\alpha} | \bar{F}', l' \rangle \) as a transformation matrix which maps a vector |\( \xi \rangle \) in the functional space \( h = \{ |\bar{F}, l' \rangle \} \) to a vector \( |\tilde{\xi} \rangle \) in a larger space \( h = \{ |i, \bar{F}, \bar{F} \rangle \} \), where \( |\tilde{\xi} \rangle \) is given by

\[
\langle \bar{F}, l | \Gamma_{\alpha} | \tau \rangle \langle \tau | \bar{F}', l' \rangle = \langle i, \bar{F}, \bar{F} | \tilde{\xi} \rangle .
\]  

(4.16)

Let \( |N\rangle \) be a vector in \( h \) that is orthogonal to all such \( |\tilde{\xi} \rangle \)'s. If we denote

\[
\langle i, \bar{F}, \bar{F} | \tilde{\xi} \rangle = \langle i, \bar{F}, l | N \rangle ,
\]  

(4.17)

then since

\[
\langle \bar{F}, l | N \rangle = \int \langle \xi | \Gamma_{\alpha} | \bar{F}, l \rangle \delta f_\alpha(\bar{F}) d^3 \tau = 0
\]  

(4.18)

for all the \( |\xi \rangle \) vectors in \( h \), (4.12) follows.

The inverse of the mass matrix \( M \) appearing in (4.10) has been divided into four blocks with

\[
M^{-1}(\bar{F}, \bar{F}) = \int d^3 \tau' \frac{\delta \phi_\alpha(\bar{F})}{\delta \phi_{\alpha}(\bar{F})} \frac{\delta \phi_{\alpha}(\bar{F})}{\delta \phi_{\alpha}(\bar{F})} ,
\]  

(4.19)

\[
M^{-1}(\bar{F}, \bar{F}) = \int d^3 \tau' \frac{\delta A_\alpha(\bar{F})}{\delta A_{\alpha}(\bar{F})} \frac{\delta A_{\alpha}(\bar{F})}{\delta A_{\alpha}(\bar{F})} .
\]  

Next, let us determine these functional derivatives in terms of the gauge-fixing function \( \chi(A_\alpha, \bar{F}) \). Solving Eq. (4.4) for \( A_\alpha \), and considering small variations of \( A_\alpha, \phi_\alpha \), and \( V_i \) we obtain

\[
\delta A_\alpha = 0 \delta V_i + u \delta \phi_\alpha + u^{-1} \delta V_i \delta u,
\]  

(4.20)
By using (3.3), we may convert the above equation into the form

$$\delta A_i^I = U^m\delta V^m_i + \frac{1}{g} \delta V^m_i (\lambda_i^m \phi_i),$$

(4.21)

where $\lambda_i^m$ is defined by

$$iu^{-1} \frac{\partial u}{\partial \phi_a} = \lambda_a^I = T^a \lambda^I_a,$$

(4.22)

$U^m$ satisfies

$$u^{-1} T^m \mu = U^m T^I,$$

(4.23)

and $D^I_m$ represents the covariant-derivative operator containing the field $A_i^I$

$$D^I_m = \delta^I_m \nabla_i - g T^a \eta^m A^I_a.$$

(4.24)

Because of the first two equations in (3.3), the matrix $U = (U^m)$ is real and orthogonal, and therefore also unitary.

The requirement that $\chi(\lambda^I) = \partial \lambda^I$ should also vanish then relates $\delta V_i^m$ and $\delta \phi_i$:

$$0 = \int d^3r' \langle \xi, l | I_i^I | F'\rangle, m \left[ C^{(m)}(F') \delta V^m_i(F') \right.$$  
$$
+ \frac{1}{g} \delta V^m_i (\lambda_i^m \phi_i(F')) \bigg]_I,$$

(4.25)

which can be solved for $\delta \phi_i(F)$, thereby determining

$$\frac{\delta \phi_i(F)}{\delta V^m_i(F')} = - g \chi^{-1}(F) \delta V^m_i(F'),$$

(4.26)

where $(\Gamma, \delta_i)^{-1}$ is the inverse of the matrix $\Gamma_i \delta_i$, whose matrix elements are

$$\langle \xi, l | \Gamma_i \delta_i | F'\rangle, m \rangle = \int d^3r' \langle \xi, l | F'\rangle, m \rangle \times$$
$$\times \langle \xi, l | \Gamma_i \delta_i | F'\rangle, m \rangle,$$

(4.27)

with $\delta_i$ as the antisymmetric matrix, defined by

$$\langle \xi, l | \delta_i | F'\rangle, m \rangle = \delta^{(m)}(F' - F).$$

(4.28)

The matrix $\lambda^{-1} (F)$ is the inverse of the $M \times M$ matrix $\lambda (F) = (\lambda_i^m)$, with $\lambda^{-1})_i^m = \delta^{(m)}$. If (4.26) is used in (4.21), we can also obtain

$$\frac{\delta A^I_i(F)}{\delta V^m_i(F')} = U^m \delta V^m_i(F') -$$
$$- \langle \xi, l | \Gamma_i \delta_i | F'\rangle, m \rangle \times \delta^{(m)}(F' - F).$$

(4.29)

Note that (4.29) automatically obeys

$$\int d^3r' \langle \xi, l | I_i^I | F', m \rangle \frac{\delta A^I_i(F)}{\delta V^m_i(F')} = 0.$$  

(4.30)

The determinant $|M|$ in (4.10) can now be determined explicitly. From (4.9), we see that $|M|$ is invariant under an orthogonal transformation among the Cartesian coordinates $q_a$. By using the notation of (4.13) and (4.16), we may consider the variations

$$\delta A^I_i(F) = \sum_k f^I_k(F) \delta Q^k,$$

(4.31)

and

$$\delta V^m_i(F') = L^m_k [\delta V^m_i(F'), \delta V^m_i(F')],$$

(4.32)

with

$$\delta V^m_i(F') = \sum_k (F', l | I_i^I | \Gamma_i^I \Gamma^I_j)^{-1/2} | \xi \rangle \delta q^k,$$

(4.33)

and

$$\delta V^m_i(F') = \sum_k f^I_k(F') \delta Q^k,$$

(4.34)

where $\xi$ runs over any complete set of orthonormal basis vectors $|\xi\rangle$ in $\hbar$. Hence, the matrix $(\delta Q^m_\alpha \delta q_\beta)$ in (4.9) now takes a $2 \times 2$ block form. We find

$$|M|^{-1/2} = \det \begin{bmatrix} - g \chi^{-1}(\Gamma_i \delta_i)^{-1}(\Gamma_i \Gamma_i)^{-1/2} & 0 \\ 1 & 1 \end{bmatrix}$$

(4.35)

where $X$ is a rectangular matrix whose matrix elements are

$$X_{\eta k} = \frac{\partial \phi_i(F)}{\partial \xi},$$

(4.36)

From (4.35), we see that

$$|M|^{1/2} = \frac{\det(-g^{-1}(\Gamma_i \delta_i)^{-1/2} \Gamma_i \delta_i)}{\det \chi},$$

(4.37)

and $\det(-g^{-1}(\Gamma_i \delta_i)^{-1/2})$ which is independent of both $A_i^I(F)$ and $\phi_i(F)$.

Our partial derivative formulas (4.26) and (4.29) may also be substituted into the expression (4.10) for $X$ yielding

$$X = \frac{1}{2} \chi^{-1} \int d^3r' P^I_i(F) \frac{d V^{m^I}_i(F')}{\delta V^m(F')} + \frac{1}{2} \chi^{-1} | \lambda |^{-1} \int d^3r' d^3r' [- P^I_i(F) \delta_i - g \phi_i(F)(\lambda^{-1})_i^r]$$
$$\times \langle \Gamma_i \delta_i | F', l | (\delta i \Gamma_i^I \Gamma^I_j)^{-1/2} | \xi \rangle \delta q^k,$$

(4.38)
where in the double integral the matrix element \( \langle \tilde{F}, J | 0 | \tilde{F}', J' \rangle \) between any pair of momenta is simply written as \( O \). The above formula appears quite similar to the familiar Coulomb-gauge formula, \( P'(F) \) replacing the transverse momenta, the quantity in square brackets the charge density, and the central matrix element the Coulomb Green's function. However, the analogy is not yet completely precise because of the presence of the angle variables \( \phi_\alpha'(F) \), their conjugate momenta \( p_\alpha'(F) \), and the matrix \( \lambda \).

In order to eliminate the angular dependence, it is useful to first construct the group generators in terms of the \( p_\alpha \)'s. Let us introduce the following parallel expression to (4.22):

\[
i \frac{3}{\hbar} \mathcal{U}^{-1} \mathcal{U} = \Lambda_a = T^+ \Lambda_a'.
\]

Because of (4.23), \( \Lambda_a' \) and \( \lambda_\alpha' \) are related by

\[
\Lambda_a' = U^m \lambda_\alpha'.
\]

which implies

\[
det \Lambda = \det \lambda.
\]

Since \( u \) is unitary and \( T^+ \) Hermitian, both \( \Lambda_a' \) and \( \lambda_\alpha' \) are real. We define two sets of operators \( \{ j^+ (F) \} \) and \( \{ j^+ (F) \} \):

\[
p_\alpha = \lambda_\alpha' J^+ = \lambda_\alpha' J^+.
\]

Hence,

\[
J^+ = (\lambda^{-1})_\alpha' p_\alpha,
\]

\[
J^+ = U^m J^+ = U^m J^+.
\]

On account of (4.5) and (4.11), the \( \Phi \) dependence of \( J^+ \), \( J^+ \), \( \lambda_\alpha' \), and \( \lambda_\alpha' \) is entirely through their dependence on \( \phi_\alpha(F) \) and \( p_\alpha(F) \). By using (3.3) and by differentiating (4.22) and (4.39) with respect to \( \phi_\alpha \), we can verify that

\[
[j^+(F), j^-(F)] = -i \lambda^\alpha \eta_\alpha(F) \delta^3(F - F')
\]

and

\[
[j^+(F), j^-(F)] = i \eta^\alpha \lambda_\alpha(F) \delta^3(F - F').
\]

Similarly, through differentiation of (4.23), it follows that

\[
[j^+(F), U^m(F)] = i \lambda^\alpha \eta_\alpha(F) \delta^3(F - F'),
\]

\[
[j^+(F), U^m(F)] = -i \lambda^\alpha \eta_\alpha(F) \delta^3(F - F'),
\]

which together with (4.43)–(4.45) lead to

\[
[j^+(F), j^-(F)] = 0.
\]

Furthermore, by using

\[
|\lambda|^{-1} \frac{\partial}{\partial \phi_\alpha} |\lambda| = (\lambda^{-1})_\alpha' \frac{\partial}{\partial \phi_\alpha} \lambda_\alpha'
\]

and by differentiating (4.22), we can derive, after some manipulation, a useful commutation relation

\[
[\rho_\alpha, (\lambda^{-1})_\alpha' |\lambda|] = 0.
\]

Thus, for arbitrary functions \( f(\phi_\alpha) \) and \( g(\phi_\alpha) \), \( J^+ \) satisfies the Hermiticity condition

\[
\int f^* \eta_\alpha J^+ d\tau_\alpha = (\int g^* \eta_\alpha J^+ d\tau_\alpha)^*,
\]

where \( d\tau_\alpha = |\lambda| d\phi_\alpha. \) Likewise, \( J^+ \) is also Hermitian. That there should exist two sets of operators \( \{ j^+(F) \} \) and \( \{ -j^+_\alpha(F) \} \), both satisfying the same group algebra and mutually commuting, has a simple geometrical meaning. In the case of the SU_2 group, this situation is identical to the familiar problem of rigid-body rotation, with \( J^+ \) as the angular momentum operator in the laboratory frame and \( J^+ \) that in the body frame. The details will be given in the Appendix.

We shall now show that the operator \( J^+ \) defined by (4.43) is equal to \( g^{-1} D^m_\alpha \Pi^m_\alpha \), given by (3.20). From (4.4), (4.23), and (4.39) we find

\[
V^\alpha = U^m_\alpha \lambda^\alpha - \frac{1}{2} \langle J^+ \rangle V^\alpha p_\alpha,
\]

which together with (3.20) and (4.24) leads to

\[
D^m_\alpha V^\alpha = U^m_\alpha \lambda^\alpha.
\]

The operator \( D^m_\alpha \Pi^m_\alpha \) can be written as a linear combination of the momenta \( p_\alpha \) and \( P_\alpha \):

\[
D^m_\alpha \Pi^m_\alpha = -i D^m_\alpha \Pi^m_\alpha \frac{\delta}{\delta V^\alpha_{\gamma}} = \int d^3 \eta' \left[ D^m_\alpha \Pi^m_\alpha \frac{\delta \phi_\alpha(F')}{\delta V^\alpha_{\gamma}(F')} p_\alpha(F') + D^m_\alpha \Pi^m_\alpha \frac{\delta \lambda^\alpha(F')}{\delta V^\alpha_{\gamma}(F')} P_\alpha(F') \right].
\]

The substitution of Eqs. (4.26) and (4.29) simplifies this considerably. For example, by using the transpose of (4.51), we find

\[
D^m_\alpha \Pi^m_\alpha \frac{\delta \lambda^\alpha(F')}{\delta V^\alpha_{\gamma}(F')} = \langle F', \eta | U^m \eta | F', \eta \rangle = 0
\]

\[
= -\langle F', \eta | [\delta_{\alpha, \gamma} - [\eta, X_\alpha, X_\gamma]^{-1} \gamma_\alpha, Y^\alpha, \eta] \rangle U^m_\gamma | F, \eta \rangle = 0
\]

in which the matrix \( \langle F, \eta | U | F', \eta \rangle = U^m_\eta(F) \delta^3(F - F') \) is real, and \( \langle F, \eta | D^m_\alpha | F', \eta \rangle = -D^m_\alpha(F) \delta(F - F') \) is antisymmetric, like \( \delta_\alpha \). Likewise,

\[
D^m_\alpha \Pi^m_\alpha \frac{\delta \phi_\alpha(F')}{\delta V^\alpha_{\gamma}(F')} = g^{-1}(F') U^m_\alpha \delta^3(F - F').
\]

Consequently, (4.52) becomes

\[
D^m_\alpha \Pi^m_\alpha = g(\lambda^{-1})_\alpha' p_\alpha = g J^+,
\]
which establishes the identity between the two definitions of $J_i$, (3.20) and (4.43).

Let us finally examine the requirement (3.27) that the physical subspace of the $V_P = 0$ gauge Hilbert space be annihilated by the operator $\mathfrak{g}$' = $j'$

$+ \phi^i T^i$. Under the transformation (4.4), the state vector (3.28) becomes a functional of $A_i$ and $\phi_i$. Equation (3.27) can be written as

$$\left[ \frac{\delta}{\delta \phi_i} + \mathcal{A}_i \phi^T \phi \right] \Psi[A_i, \phi_i] = 0,$$

(4.56)
an equation which is easily solved:

$$\Psi[A_i, \phi_i] = \psi(\phi_i) \tilde{\Psi}(A_i).$$

(4.57)

Here $\tilde{\Psi}(A_i)$ is any state vector depending on the vector potential $A_i$ and the fermionic degrees of freedom, but independent of $\phi_i$. $\psi(\phi_i)$ is the unitary transformation acting on the fermionic degrees of freedom which represents the group element $\psi(\phi)$ has been chosen as a generic function for a general class of functions.

Since the generators of $u(\phi_i)$ are simply $\phi^i T^i$, the equation

$$\frac{\delta}{\delta \phi_i} u^{-1} = -i \mathcal{A}_i \phi^T \phi$$

(4.58)

is the Hilbert-space analog of (4.39) and this equation implies directly that the state $\Psi[A_i, \phi_i]$ defined in Eq. (4.57) does indeed satisfy Gauss' law. Eq. (4.58). [In the SU$_3$ case, if we represent $u$ in terms of the Euler angles $a$, $b$, $c$ as in (A1) of the Appendix,]

$$\bar{H}(P, \phi) = \frac{1}{2} \mathcal{J} \int d^3r \mathcal{F}(\mathcal{F}) \delta \mathcal{F}(\mathcal{F})$$

$$+ \frac{1}{2} \mathcal{J} \int d^3r \int d^3r' \left[ -P_i(\mathcal{F}) D_i + \phi^T \phi \right] \delta \left\{ (\Gamma_i \mathcal{D}_i)^{-1} D_i \Gamma_i^T \left[ \left( \Gamma_i \mathcal{D}_i \right)^{-1} + \phi^T \phi \right] \right\}$$

$$+ \int d^3r [\frac{1}{2} A_j B^j + \phi^i A_j (\phi^T \phi)]$$

(4.62)

where, as in (4.37), $\mathcal{J} = \det \mathcal{J} \Gamma_i \mathcal{D}_i$. The angle variables have now been completely eliminated from the problem. The Yang-Mills quantum theory, when restricted to the states $\Psi(A_i)$, becomes very close to the canonical theory that would have been naively proposed for the gauge $\chi[A_i]$). However, the precise ordering of the operators in (4.62), in particular, the appearance of the Jacobian $\mathcal{J}$, is not the conventional one. In fact, as will be shown in Sec. VI, the operator ordering in (4.62) yields additional vertices in the Feynman rules. Although for the case of the Coulomb gauge, the operator ordering implied by (4.62) can be shown to be identical to that proposed by Schwinger et al as, which will be discussed in Sec. VII, the derivation presented above appears to be particularly simple and clear, the kinetic energy in (4.62) being essentially the familiar formula (4.7) for the Laplacian in curvilinear coordinates.

Let us conclude this section by specializing our resulting Hamiltonian operator in (4.62) to the Coulomb gauge for the group SU$_3$. In that case the matrix $\Gamma_i$ becomes

$$\langle \mathcal{F}, l | \mathcal{F}', m \rangle = \nabla_j \left[ \phi^i (\mathcal{F} - \mathcal{F}') \delta \phi^i \right].$$

(4.63)
The operator $\Gamma_i \mathcal{D}_i$ becomes

$$\Gamma_i \mathcal{D}_i = \nabla_j \mathcal{D}_i = \nabla_j \left[ \nabla_j \delta \phi^i - \epsilon g \mathcal{J} \phi^i \right]$$

(4.64)

so that the Jacobian $\mathcal{J}$ of (4.37) can be recognized as the familiar, Coulomb-gauge, Faddeev-Popov determinant. Equation (4.62) becomes
\[ H = \int d^4r \left[ \frac{1}{2} g^2 \Delta_\overline{\Delta} \phi^2 + \frac{1}{2} g^2 \Delta_\overline{\Delta} \phi^2 + \frac{1}{2} g^2 \Delta_\overline{\Delta} \phi^2 \right] \]

\[ + \frac{1}{2} g^2 \int d^4r \, d^4r' \, g^{-1} \left( \left[ - \vec{\Delta}_n \phi^2 (F) \times \vec{\Delta}_n \phi^2 (F) \right] + \frac{1}{2} \phi^2 (F) \phi^2 (F) \right) \]

\[ \cdot \left[ \Delta_n (F') \times \Delta_n (F') + \frac{1}{2} \phi^2 (F') \phi^2 (F') \right] \right] \mathbf{R}^{m'} \left( F', m' \right), \tag{4.65} \]

where \( T^I = \frac{1}{2} A^I \), all SU vectors are indicated by arrows, and \( \mathbf{R}^I \) is replaced by the more familiar notation \( \vec{\Delta}_n \phi^2 \). Both \( \vec{\Delta}_n \) and \( \mathbf{R}_n \) are transverse. Note that the matrix \( \left( \vec{F}, m \right) \left( \vec{F}, m \right)^{-1} \left( \vec{F}, m \right)^{-1} \left( \vec{F}, m \right)^{-1} \) has reduced to precisely the usual non-Abelian, Coulomb, Green's function.

V. COVARIANT GAUGES

We will now show that the Cartesian operator ordering of the \( \mathcal{V}_n = 0 \) gauge Hamiltonian (3.29)–(3.31) defines a quantum theory identical to the theory determined by the usual covariant-gauge Feynman rules. The transformation of the \( \mathcal{V}_n = 0 \) gauge quantum theory of Sec. III into a covariant gauge is most easily done using the Feynman path-integral formalism.\(^6\)

For the case of a single coordinate \( q \) and conjugate momentum \( p \) the path integral for the Schrödinger Green's function applied to a state \( \psi(1) \) is obtained by writing

\[ \langle q' | e^{-iH(t'-t)} | q(n) \rangle = \lim_{\Xi \to \infty} \prod_{n=1}^{N} \int dq(n) \langle q(n+1) | 1 - i \epsilon H(p, q) | q(n) \rangle \langle q(1) |, \tag{5.1} \]

where \( q(n) \) denotes the coordinate \( q \) at time \( t_n = t + (n-1)\epsilon \) with \( \epsilon = (t' - t)/\Xi \) and \( q(\Xi + 1) = q' \). Each matrix element in the product on the right-hand side of (5.1) is represented by

\[ \langle q(n+1) | 1 - i \epsilon H(p, q) | q(n) \rangle = \int \frac{dp(n)}{2\pi} \, e^{i(p(n) \epsilon + (n+1) - \epsilon(n))} \left[ 1 - i \epsilon H(p(n), \frac{q(n) + q(n+1)}{2}) \right], \tag{5.2} \]

so that (5.1) becomes

\[ \langle q' | e^{-iH(t'-t)} | q(n) \rangle = \lim_{\Xi \to \infty} \prod_{n=1}^{N} \int dq(n) dp(n) \frac{dp(n)}{2\pi} \exp \left[ i \epsilon \left[ H(p(n), \frac{q(n+1) + q(n)}{2}) \right] - H \left( p(n), \frac{q(n+1) + q(n)}{2} \right) \right] \langle q(1) |, \tag{5.3} \]

the familiar Hamiltonian path integral for a one-dimensional problem. Our substitution of the variable \( \frac{1}{2} \left[ q(n+1) + q(n) \right] \) for the operator \( q \) in \( H(p, q) \) of (5.1) corresponds to the Weyl-ordered form of the Hamiltonian. As can be readily verified, for a classical Hamiltonian \( p^2 f(q) + p \mu(q) + v(q) \), the Weyl form of the quantum-mechanical Hamiltonian is

\[ H = \frac{1}{2} \left[ p^2 f(q) + 2pf(q)p + f(q)p^2 \right] + \frac{1}{2} \left[ \mu(q) + \mu(q)p + v(q) \right], \tag{5.4} \]

where \( q \) and \( p \) are operators. By substituting this expression into the left-hand side of (5.2), one sees that it leads precisely to the right-hand side.

Equation (5.3) can be immediately generalized to represent the \( \mathcal{V}_n = 0 \) gauge Schrödinger Green's function by a path integral for the gauge theory without fermions:

\[ \langle V' | e^{-iH(t'-t)} | V \rangle = \lim_{\Xi \to \infty} \int d[V(n)] d[\Pi(n)] \]

\[ \times \exp \left[ \frac{1}{\Xi} \int d^4r \, \text{tr} \left( \Pi_{i}(n) \left( V_{i}(n+1) - V_{i}(n) \right) \right) \epsilon^{-1} \right. \]

\[ \left. \left. - \frac{1}{2} \Pi_{i}(n) \Pi_{j}(n) - \frac{1}{2} B_{i}(n) B_{j}(n) \right) \right] \Psi[V_{i}(1)], \tag{5.5} \]

where \( \Psi[V_{i}] \) is introduced by (3.28), \( V_{i}(n) \) and \( \Pi_{i}(n) \) refer to the matrix form (3.17) of the fields \( V_{i}(\overline{F}, t_{n}) \) and \( \Pi_{i}(\overline{F}, t_{n}) \), with \( t_{n} = t + (n-1)\epsilon \). For simplicity, the position dependence is suppressed, \( V_{i}(\overline{F}, t_{n}) = V_{i}' \), and the differentials \( d[V(n)] \) and \( d[\Pi(n)] \) stand, respectively, for the products.
\[ \Pi_{\alpha_1, \ldots, \alpha_r} d\tilde{q}_r(\vec{r}, t_r) \text{ and } \Pi_{\alpha_1, \ldots, \alpha_r} d\tilde{q}_r(\vec{r}, t_r). \]

To include fermions in the path-integral description, it is necessary to introduce a representation of the fermion Hilbert space as a space of polynomials of generators \( \tilde{q}_1, \tilde{q}_2, \ldots \) of a Grassmann algebra:

\[ \{ \tilde{q}_a, \tilde{q}_b \} = 0 \]

for all \( a \) and \( b = 1, 2, \ldots \). Their differentials \( d\tilde{q}_a \) and derivative operators \( \partial / \partial \tilde{q}_a \) satisfy

\[ \left( \frac{\partial}{\partial \tilde{q}_a}, \tilde{q}_b \right) = \left( \frac{\partial}{\partial \tilde{q}_b}, \tilde{q}_a \right) = \delta_{ab} \] (5.6)

and

\[ \{ d\tilde{q}_a, \tilde{q}_b \} = \{ d\tilde{q}_a, d\tilde{q}_b \} = 0 . \] (5.7)

In addition, there are the usual integration rules

\[ \int d\tilde{q}_a = 0 \quad \text{and} \quad \int \tilde{q}_a d\tilde{q}_a = \delta_{aa} . \] (5.8)

Let

\[ \{ \psi_a(\vec{r}) \} \] (5.9)

be a complete orthonormal set of c-number single-particle spinor functions. By introducing for each \( \psi_a(\vec{r}) \) a Grassmann generator \( \tilde{q}_a \), we can represent a multiparticle state vector \( | \psi \rangle \) with the probability amplitude \( C_r(\alpha_1, \ldots, \alpha_r) \) for the states \( \psi_{\alpha_1}, \ldots, \psi_{\alpha_r} \) to be occupied by the polynomial

\[ | \psi \rangle = \sum_{\alpha_1} \sum_{\alpha_2} \cdots \sum_{\alpha_r} C_r(\alpha_1, \ldots, \alpha_r) \prod_{\beta = \alpha_1}^{\alpha_r} \tilde{q}_\beta \] (5.10)

in which, for definiteness, the product of the \( \tilde{q}_\beta \)'s is arranged in the order of increasing \( \beta \).

Equation (5.10) is the \( \tilde{q} \) representation of the bra vector \( \langle \psi \rangle \); its ket vector (\( | \rangle \) in the \( \tilde{q} \) representation then assumes the form

\[ | \psi \rangle = \sum_{\alpha_1} \sum_{\alpha_2} \cdots \sum_{\alpha_r} \pm C_r(\alpha_1, \ldots, \alpha_r) \tilde{q}_{\alpha_1} \tilde{q}_{\alpha_2} \cdots \tilde{q}_{\alpha_r} , \] (5.11)

where the \( \pm \) sign is determined by the normalization condition

\[ \langle \psi | \psi \rangle = \int \prod_{\alpha} d\tilde{q}_\alpha | \psi \rangle \langle \psi | \] (5.12)

With the above definition, a multiplication \( \tilde{q}_a \) onto \( \langle \tilde{q} \rangle \) becomes the annihilation operator for the \( a \)th state, while the differentiation \( \partial / \partial \tilde{q}_a \) is the corresponding creation operator.

The "Fourier transform" of any polynomial \( \psi(\tilde{q}) \) is given by

\[ \phi(\tilde{q}) = \int \exp(\sum_{\alpha} \tilde{q}_\alpha \tilde{q}_\alpha) \psi(\tilde{q}) \prod_{a} d\tilde{q}_a , \] (5.13)

where the \( \tilde{q}_a \)'s anticommute with each other as well as with the \( \tilde{q}_a \)'s. By using (5.6) and (5.8), we see that

\[ \tilde{q}_a \phi(\tilde{q}) = \int \exp(\sum_{\alpha} \tilde{q}_\alpha \tilde{q}_\alpha) \frac{\partial \phi(\tilde{q})}{\partial \tilde{q}_a} \prod_{a} d\tilde{q}_a , \] (5.14)

and therefore the usual partial-integration rule holds. The "\( 0 \) function" is defined by

\[ \langle \tilde{q} | \tilde{q} \rangle = \delta(\tilde{q} - \tilde{q}') \]

\[ = \int \prod_{a} d\tilde{q}_a \exp(\sum_{\alpha} \tilde{q}_a \tilde{q}_a) \prod_{\alpha} d\tilde{q}_a = \delta(\tilde{q}_a - \tilde{q}_a) \] (5.15)

since, for arbitrary \( \tilde{q}(\tilde{q}) \),

\[ \psi(\tilde{q}) = \int \delta(\tilde{q} - \tilde{q}') \prod_{a} d\tilde{q}_a \psi(\tilde{q}') , \] (5.16)

where the order of product \( \prod_{a} \) is the transpose of that in \( \Pi_{\alpha} \); i.e., if in (5.15) \( \prod_{\alpha} d\tilde{q}_a = d\tilde{q}_1 d\tilde{q}_2 \cdots \), then \( \delta(\tilde{q} - \tilde{q}') = (\tilde{q}'_1 - \tilde{q}_1)(\tilde{q}'_2 - \tilde{q}_2) \cdots \) in the same order, but in (5.16) \( \prod_{\alpha} d\tilde{q}_a = \cdots d\tilde{q}_2 d\tilde{q}_1 \) in the transposed order. Hence, the inverse transform of (5.13) is

\[ \psi(\tilde{q}) = \int \exp(\sum_{\alpha} \tilde{q}_\alpha \tilde{q}_\alpha) \phi(\tilde{q}) \prod_{a} d\tilde{q}_a . \] (5.17)

As an illustration, we first examine the case of a single mode with only one generator \( \tilde{q} \). The corresponding basis vectors of the Hilbert space can be chosen to be \( | \tilde{q} \rangle \), where \( \tilde{q} \) is the occupation number \( = 0 \) or \( 1 \). In accordance with (5.10)–(5.12), we may write

\[ \langle \tilde{q} | \tilde{q} \rangle = 0, \quad \langle \tilde{q} | 1 \rangle = 1 , \] (5.18)

\[ \langle 0 | \tilde{q} \rangle = -1, \quad \langle 1 | \tilde{q} \rangle = \tilde{q}_a . \]

Therefore, we have the usual orthonormality relation

\[ \langle 1 | 1 \rangle = \int \langle 1 | \tilde{q} \rangle d\tilde{q} \langle \tilde{q} | 1 \rangle = \delta_{11} \]

and the completeness condition

\[ \sum_{\tilde{q} = 0, 1} \langle \tilde{q} | 1 \rangle \langle 1 | \tilde{q} \rangle = \delta(\tilde{q} - \tilde{q}) . \] (5.19)
If a Hamiltonian $H$ has $|i\rangle$ as its eigenvector with eigenvalue $E_i = i E$, then the matrix representation of $e^{-iHt}$ is

$$\langle \hat{q}' | e^{-iHt} | \hat{q} \rangle = \sum_{\ell = 0}^{\infty} \langle \hat{q}' | i \ell \rangle e^{-i\ell t} \langle i \ell | \hat{q} \rangle$$

$$= -\hat{q}' + e^{-i\ell t} \hat{q}.$$  

(5.20)

Returning to the general case, with the help of these equations we can easily write down the fermionic analog of (5.2):

$$\langle \hat{q}(n+1) | 1 - t \in H | \hat{q}(n) \rangle = \int \prod_{\alpha} dq_{\alpha}(n) \exp \left[ \sum_{\alpha} q_{\alpha}(n)(\hat{q}_{\alpha}(n+1) - \hat{q}_{\alpha}(n)) \right]$$

$$\times \left[ 1 + i \sum_{\gamma, \gamma'} q_{\gamma}(n) H_{\gamma, \gamma'} \hat{q}_{\gamma'}(n) \right],$$

(5.21)

where in our case, because of (3.1),

$$H_{\gamma, \gamma'} = -i \int d\tau \phi_{\gamma'}^\dagger(\bar{\tau}) \alpha D_{\gamma'} \phi_{\gamma}(\bar{\tau})$$

(5.22)

with $\phi_{\gamma}(\bar{\tau})$ as a member of the $c$-number spinor set (5.9). Thus, if we define the anticommuting functions

$$\phi_{\alpha}(n) = \phi_{\alpha}(t, n) = \sum_{\gamma} \phi_{\alpha}(\bar{\tau}) q_{\gamma}(n),$$

(5.23)

$$\bar{\phi}_{\alpha}(n) = \bar{\phi}_{\alpha}(t, n) = -\sum_{\gamma} \bar{\phi}_{\gamma}(\bar{\tau}) q_{\gamma}(n),$$

where $\phi_{\alpha}(n) = \psi_{\alpha}(t, n)$, $t = t + (n - 1)\epsilon$ as before, and write $\int \prod_{\alpha} dq_{\alpha}(n)$ as $\int d[\psi(n)]$ and $\int d[\bar{\psi}(n)]$, respectively, the fermionic Green's function becomes

$$\langle \psi'|e^{-iH(t'-t)}|\psi(n)\rangle = \lim_{\hbar \rightarrow 0} \int \prod_{n=1}^{N} d[\bar{\psi}(n)] d[\psi(n)]$$

$$\times \left\{ \exp \left[ \sum_{n=1}^{N} \int d\tau [-\bar{\psi}(n)\gamma_4 (\psi(n+1) - \psi(n)) - i\epsilon \bar{\psi}(n)\gamma_j D_j \psi(n)] \right] \right\} \Psi[\psi(1)],$$

(5.24)

where $\Psi[\psi(1)] = \langle \psi(1) \rangle$. Equations (5.5) and (5.24) can be combined to give a path-integral expression for the Yang-Mills and fermion Green's function. By using the Cartesian form $\Pi_4(n)$ $\Pi_4(n)$ for the kinetic energy in (5.5), with no additional terms we have explicitly incorporated the operator ordering of the $V_o=0$ gauge Hamiltonian (3.29)-(3.31).

We can now transform the $V_o=0$ gauge quantum theory defined by (5.5) and (5.24) into a noncovariant gauge, $\chi(A_f) = 0$, by a simple change of variables. Of course, we must be careful when transforming $V(n+1) - V(n)$ and $\psi(n+1) - \psi(n)$ to keep all terms in the exponent through order $\epsilon$ if we intend the transformed theory to be equivalent to the original one. Such a procedure is completely equivalent to the operator manipulations of Sec. IV and leads to the same result.8

In this section we wish to use the path-integral formulation to transform to a covariant gauge. This can be done most economically if we introduce into the integrand of (5.5) the factor

$$1 = \int \prod_{n=1}^{N} d[\phi_{\alpha}(n)] \bar{\Psi} \left( \exp \left[ \sum_{n=1}^{N} \int d\tau \frac{-i\epsilon}{2\hbar} \left[ \gamma_4 (A_1(\bar{\tau})) + \frac{i}{\hbar} (A_0(\bar{\tau}) \gamma_4 u(n)) \right]^2 \right) \right],$$

(5.25)

where $A_1(n) = A_1(t, n)$ and $A_0(n) = A_0(\bar{\tau}, t)$ are Hermitian matrices, defined by

$$A_1(n) = u(n)^{-1} V_1(n) u(n) + \frac{i}{\hbar} u(n)^{-1} A_0(n) u(n)$$

(5.26)

and

$$A_0(n) = -\frac{i}{\hbar \epsilon} \ln [u(n) u(n+1)^{-1}] = -\frac{i}{\hbar \epsilon} u(n)^{-1} (u(n+1) - u(n)) + \frac{i}{2\hbar \epsilon} [u(n)^{-1} u(n+1) - u(n)] u(n),$$

(5.27)

with

$$u(n) = u(\phi_{\alpha}(n))$$

(5.28)
the $N \times N$ unitary matrix function of the $M$ group parameters $\phi_\alpha(n) = \phi_\alpha(F, t_\alpha)$ in (4.5), and the differential
d$[\phi_\alpha(n)] = \Pi_{\alpha i} d[\phi_\alpha(F, t_\alpha)]$. Except for a numerical factor, the Jacobian $[B]$ is given by

$$
[B] = \det(F, t_\alpha) \frac{\det(\phi_\alpha(F', t_\alpha))}{\det(F, t_\alpha)} = \det \frac{\delta \phi_\alpha(n')}{\phi_\alpha(n)} \left[ \nabla_i A_i(n') + \frac{(A_i(n' + 1) - A_i(n'))}{\epsilon} \right].
$$

(5.29)

In order to match the final configuration of (5.5), we set $A_i(\mathbb{R} + 1) = V_i(\mathbb{R} + 1) = V'_i$, i.e.,

$$
u(\mathbb{R} + 1) = 1.
$$

(5.30)

Furthermore, since the integrand of (5.25) is invariant under $A_i(n) - A_i(n') + \text{constant}$ for all $n = 1, 2, \ldots, \mathbb{R} + 1$, we may choose $A_i(\mathbb{R} + 1) = 0$.

We now examine the region where the summand in the exponent of (5.25) is of order 1. Because in the summand the coefficient of $[A_i(n + 1) - A_i(n)]^2$ is proportional to $\epsilon^{-1}$, we expect only those configurations with $A_i(n + 1) - A_i(n)$ of order $\epsilon^{1/2}$ to contribute to the integral over $\phi_\alpha(n)$. Thus since $A_i(\mathbb{R} + 1) = 0$, $A_i(n)$ should be of order $(\mathbb{R} - n)^{1/2} \epsilon^{1/2} = O(1)$ and, from (5.27), $u(n + 1) - u(n)$ is only of order $\epsilon$.

If we perform $\int d[\Pi(n)]$ in (5.5) we find, up to a numerical factor,

$$
\int d[\Pi(n)] \exp \left( i \sum_n \int d^2r \left[ \Pi_i(n) \left( V_i(n + 1) - V_i(n) \right) - \frac{i}{\epsilon} \Pi_i(n) \Pi_i(n) \right] \right)
= \exp \left( i \sum_n \int d^2r \left[ V_i(n + 1) - V_i(n) \right]^{2} \epsilon^{-1} \right).
$$

(5.31)

Next we can use (5.26) to express this exponent in terms of $A_i(n)$ and $u$. A particularly symmetrical form is obtained if the quantity in square brackets in (5.31) is conjugated with the matrix $[u(n) - u(n + 1)]^{1/2} u(n + 1)^{-1}:

$$
\epsilon^{-1} \text{tr} \left[ V_i(n + 1) - V_i(n) \right]^{2} = \epsilon^{-1} \text{tr} \left\{ [u(n) - u(n + 1)]^{1/2} A_i(n + 1) [u(n) - u(n + 1)]^{-1/2}
+ \frac{i}{\epsilon} [u(n) - u(n + 1)]^{1/2} \nabla_i [u(n) - u(n + 1)]^{1/2}
- [u(n) - u(n + 1)]^{-1/2} A_i(n) [u(n) - u(n + 1)]^{1/2}
- \frac{i}{\epsilon} [u(n) - u(n + 1)]^{-1/2} \nabla_i [u(n) - u(n + 1)]^{1/2} \right\}.
$$

(5.32)

Using (5.27) to replace $u(n) - u(n + 1)$ by a function of $A_i(n)$,

$$
u(n) - u(n + 1) = 1 + \frac{i}{\epsilon} g \epsilon A_i(n) + O(\epsilon^2),
$$

we can expand the right-hand side of Eq. (5.32) through order $\epsilon$:

$$
\epsilon^{-1} \text{tr} \left[ V_i(n + 1) - V_i(n) \right]^{2} = \epsilon \epsilon^{-1} \left( A_i(n + 1) - A_i(n) \right) + \frac{i}{\epsilon} g A_i(n, A_i(n)] + \nabla_i A_i(n)]^{2} + O(\epsilon^2)
$$

(5.33)

where we treat $A_i(n + 1) - A_i(n)$ as of order $\epsilon^{1/2}$. Thus, if we change integration variables from $V_i(n)$, $\phi(n)$, and $\tilde{\phi}(n)$ to the unitarily equivalent set $A_i(n)$,

$$
\psi(n) = u(n)^{-1} \phi(n),
$$

and

$$
\tilde{\psi}(n) = \tilde{\phi}(n) u(n),
$$

the exponentials in (5.5), (5.24), and (5.25) combine to give the usual covariant-gauge action up to terms vanishing with $\epsilon$:

$$
i \epsilon \sum_n \int d^2r \left\{ \frac{i}{2} \left( A_i(n + 1) - A_i(n) \right) (A_i(n) - A_i(n + 1) + \text{g} A_i(n, A_i(n)] + \nabla_i A_i(n)]^{2}
\right.
- \frac{i}{2} \left[ B_i(n) \right]^{2} - \frac{1}{2 \epsilon} \left[ \nabla_i A_i(n) + (A_i(n + 1) - A_i(n)) \epsilon^{-1} \right]^{2}
\left.
\right) ,
$$

$$
i \epsilon \sum_n \int d^2r \left\{ \tilde{\psi}(n) \left[ \frac{\gamma_{4} \tilde{\psi}(n + 1) - \tilde{\psi}(n)}{i \epsilon} + g A_i(n) \nu_{4} \psi(n + 1) + \nu_{4} \left( \nabla_i - i \epsilon A_i(n) \psi(n) \right) \right] \right\}.
$$

(5.35)
Finally, we must use Eq. (5.27) to change the integration variable \( \phi_{\alpha}(n) \) to \( A_{\alpha}^{a}(n) \) where \( T^{a}A_{\alpha}^{a}(n) = A_{\alpha}(n) \). The resulting Jacobian \( \frac{\delta A_{\alpha}^{a}(n)}{\delta \phi_{\alpha}(n')} \) can then be worked out explicitly. By using (5.27), we see that the matrix \( \frac{\delta A_{\alpha}^{a}(n)}{\delta \phi_{\alpha}(n')} \) can be resolved into two terms:

\[
\frac{\delta A_{\alpha}^{a}(n)}{\delta \phi_{\alpha}(n')} = \frac{1}{g^2} \left( \rho_{\alpha}(G(n)|m) + \delta_{\alpha 1} \rho_{\alpha}(G'(n)|m) \right),
\]

where, on account of (4.22),

\[
\rho_{\alpha}(G(n)|m) = \gamma_{\alpha}^{a}(n) - \frac{1}{2} g \epsilon^{abc} \gamma_{\alpha}^{c}(n) A_{\beta}^{b}(n) + O(\epsilon^3).
\]

(5.37)

Because of the boundary condition (5.30), \( A_{\alpha}^{a}(\mathcal{M}) \) depends only on \( u(\mathcal{M}) \). Hence, we have

\[
\rho_{\alpha}(G'(n)|m) = 0,
\]

and therefore

\[
\rho_{\alpha}(G(n)|m) = \frac{1}{g^2} \frac{\delta A_{\alpha}^{a}(n)}{\delta \phi_{\alpha}(n'} \rho_{\alpha}(G(n)|m).
\]

(5.38)

From (5.37), it follows that

\[
\langle V' \mid e^{-iH(t-t')} \rangle = \int d[A_{\alpha}] d[\bar{\psi}'] d[\bar{\psi}'] \exp \left\{ i \int \frac{d^4 x}{g^2} \operatorname{Tr} \left[ -\frac{1}{2} F_{\mu\nu}^2 - \bar{\psi}' \gamma_{\alpha} A_{\alpha}^\mu \psi' - \frac{1}{2} \left( g A_{\alpha}^\mu \right)^2 \right] \right\},
\]

(5.43)

where

\[
d[A_{\alpha}] d[\bar{\psi}'] d[\bar{\psi}'] = \lim_{n \to \infty} \prod_{\alpha} d[A_{\alpha}(n)] d[\bar{\psi}(n)] d[\bar{\psi}(n)].
\]

VI. NONCOVARIANT-GAUGE FEYNMAN RULES

The quantum Yang-Mills theory in the gauge \( \lambda(A_{\alpha}) = 0 \) is completely specified by the Hamiltonian operator (6.52) derived in Sec. IV. However, if this Hamiltonian is divided into free and interacting pieces,

\[
H_{\text{free}} + H_{\text{int}},
\]

(6.1)

the resulting Dyson-Wick perturbation theory in \( H_{\text{int}} \) will be complicated because of the quadratic dependence of \( H_{\text{int}} \) on the canonical momenta \( P_{\alpha} \). In the interaction representation the \( P_{\alpha} \) become \( A_{\alpha}^\mu \), the Wick contractions of \( A_{\alpha}^\mu(n) \) or \( A_{\alpha}^\mu(y) \) contain noncovariant \( \delta(x_{\alpha} - y_{\alpha}) \) terms, these \( \delta \)-function terms can be systematically resummed and the result represented as new, nonpolynomial additions to \( H_{\text{int}} \). The vertices implied by this modified \( H_{\text{int}} \) when joined with "naive" Feynman propagators, then provide the correct Feynman perturbation series for the original quantum theory.

In this section we will derive these simplified

\[
\text{det}(a | G(n) | m) = \left[ 1 - \frac{i}{g} \epsilon^{abc} \lambda_{\alpha}^{a}(n) A_{\beta}^{b}(n)(\lambda_{\alpha}^{c}(n)) \right] \cdot \text{det}(a | n) + O(\epsilon^2)
\]

\[
= \text{det}(a | n) + O(\epsilon^2),
\]

(5.39)

where \( \lambda(n) \) stands for the \( M \times M \) matrix \( \lambda_{\alpha}^{a}(n) \). Since \( \lambda = O(\epsilon^{-1}) \), when \( \epsilon \to 0 \) we can drop the \( O(\epsilon^2) \) term in (5.39) and derive

\[
\text{det}(a | G(n) | m) = \text{det}(a | n) \times \prod_{\alpha} \text{det}(\lambda_{\alpha})(n).
\]

(5.40)

By following steps very similar to those that led to (4.36), we may express \( \langle \beta \rangle \) of (5.29) as

\[
\langle \beta \rangle = \text{const} \times \frac{\text{det}(a | G(n) | m)}{\text{det}(a | n) \times \prod_{\alpha} \text{det}(\lambda_{\alpha})(n)}
\]

(5.41)

which when multiplied by (5.40) gives, in the limit \( \epsilon \to 0 \), a constant times the usual, covariant-gauge Faddeev-Popov determinant

\[
\delta_{\alpha} = \text{det}(a_{\alpha} | a_{\alpha}).
\]

(5.42)

Consequently, apart from a constant multiplicative factor, the original \( V_{\alpha} = 0 \) gauge Schrödinger Green's function has been transformed precisely into the usual covariant-gauge expression

\[
\text{Feynman rules for the noncovariant-gauge Hamiltonian (6.42). In practice, these Feynman rules are most easily obtained using functional integration.}^6 \text{ We first represent the Schrödinger Green's function by an integral over classical trajectories in phase space as in (5.1). Instead of immediately evaluating the Gaussian integral over } P_{\alpha}, \text{ we next replace the term in } H_{\text{int}} \text{, quadratic in } P_{\alpha}, \text{ by a Gaussian integral with respect to a new variable } A_{\alpha}^\mu \text{ coupling linearly with } P_{\alpha}. \text{ Finally, the integration over } P_{\alpha} \text{ is carried out leaving a Lagrangian functional integral over } A_{\alpha}^\mu \text{ and } A_{\alpha}^\mu \text{ which, if the original Hamiltonian was Weyl ordered, directly specifies the proper Feynman rules.}
\]

The first step in this procedure is the most difficult. We must rearrange the operators in the Hamiltonian (6.42) into Weyl order. The Weyl ordering of the Laplacian (4.7) in curvilinear coordinates,

\[
\mathbf{x} = -\frac{1}{2} \sum_{\alpha, \beta} |M|^{-1/2} \frac{\partial}{\partial q_{\alpha}} |M|^{-1/2} \frac{\partial}{\partial q_{\beta}},
\]

(4.7)

is quite straightforward. The result is particularly simple if we extract a factor of \( |M|^{-1/4} \) from the state vectors:
\[ |M|^{1/4} |\bar{M}|^{-1/4} = -\frac{1}{2} \left\{ \left. \frac{\partial}{\partial \phi_\alpha} \frac{\partial}{\partial \phi_\beta} M^{\sigma \sigma} \right|_{\phi_\alpha = 0, \phi_\beta = 0} + \frac{1}{2} \left. \frac{\partial}{\partial \phi_\alpha} \frac{\partial}{\partial \phi_\beta} M^{\sigma \epsilon} \right|_{\phi_\epsilon = 0} + \frac{1}{4} \left. \frac{\partial}{\partial \phi_\alpha} \frac{\partial}{\partial \phi_\beta} \frac{\partial}{\partial \phi_\gamma} \right|_{\phi_\gamma = 0} \right\} + \frac{1}{6} \left( \frac{\partial}{\partial \phi_\alpha} \frac{\partial}{\partial \phi_\beta} \frac{\partial}{\partial \phi_\gamma} \right) \left( \frac{\partial}{\partial \phi_\delta} \frac{\partial}{\partial \phi_\epsilon} \right) \left( \frac{\partial}{\partial \phi_\sigma} \frac{\partial}{\partial \phi_\tau} \right). \] (6.2)

in which the differentiations appearing inside the curly brackets are arranged in Weyl order, as in (5.4).

However, complications arise because what we are interested in is the Weyl ordering of the Hamiltonian (4.62) which, having been simplified by restriction to gauge-invariant states, (4.57), no longer has the form (4.7). In order to apply this simple formula to the case at hand, we must return to the form (4.38) for the gauge-theory Hamiltonian in which the angle variables \( \phi_\alpha \) and their conjugate momenta \( p_\alpha \) appear. Equation (6.2) then gives a Weyl-ordered form for \( H \), from which \( \phi_\alpha \) and \( p_\alpha \) must again be eliminated as follows: The terms in curly brackets in (6.2) which contain derivatives with respect to \( \phi_\alpha \) are

\[ -\frac{1}{2} \int d^4r \int d^4r' \left\{ \langle \bar{F}, \bar{F} \mid (\Gamma_\alpha \delta_{\alpha \beta})^{-1} \Gamma_\epsilon \bar{F}_\epsilon (\Gamma_\delta \delta_{\delta \epsilon})^{-1} \mid \bar{F}, \bar{F} \rangle \right\} \]

\[ \times \left[ \frac{1}{4} \frac{\partial}{\partial \phi_\alpha} \frac{\partial}{\partial \phi_\beta} \lambda^{-1}(F'_\epsilon) \lambda^{-1}(F'_\delta) + \frac{1}{2} \frac{\partial}{\partial \phi_\alpha} \frac{\partial}{\partial \phi_\epsilon} \lambda^{-1}(F'_\epsilon) \lambda^{-1}(F'_\delta) \right] \] 

\[ + \frac{1}{4} \lambda^{-1}(F'_\epsilon) \lambda^{-1}(F'_\delta) \frac{\partial}{\partial \phi_\alpha} \frac{\partial}{\partial \phi_\beta} \left( \frac{\partial}{\partial \phi_\epsilon} \right) \]

\[ + i(\partial_P P(F,F,\bar{F},\bar{F} \mid (D_{\alpha} (\Gamma_\alpha \delta_{\alpha \beta})^{-1} \Gamma_\epsilon \bar{F}_\epsilon (\Gamma_\delta \delta_{\delta \epsilon})^{-1} \mid \bar{F}, \bar{F} \rangle + \langle \bar{F}, \bar{F} \mid D_{\alpha} (\Gamma_\alpha \delta_{\alpha \beta})^{-1} \Gamma_\epsilon \bar{F}_\epsilon (\Gamma_\delta \delta_{\delta \epsilon})^{-1} \mid \bar{F}, \bar{F} \rangle P_\alpha (\bar{F}) \rangle \]

\[ \times \left[ \frac{1}{2} \frac{\partial}{\partial \phi_\alpha} \frac{\partial}{\partial \phi_\beta} \lambda^{-1}(F'_\epsilon) \lambda^{-1}(F'_\delta) \right] \right\}. \] (6.3)

If this operator is applied to the state

\[ |\lambda|^{1/2} \Psi(\phi_\alpha) \Psi(\alpha), \] (6.4)

the result is

\[ -\frac{1}{2} |\lambda|^{1/2} \Psi(\phi_\alpha) \int d^4r \int d^4r' \left\{ \langle \bar{F}, \bar{F} \mid (\Gamma_\alpha \delta_{\alpha \beta})^{-1} \Gamma_\epsilon \bar{F}_\epsilon (\Gamma_\delta \delta_{\delta \epsilon})^{-1} | \bar{F}, \bar{F} \rangle \right\} \]

\[ \times \left[ -\phi'(F') T' \psi(F') T' \psi(F') + \frac{1}{4} \left( \frac{\partial}{\partial \phi_\alpha} \frac{\partial}{\partial \phi_\epsilon} \lambda^{-1}(F') \right) \right] \]

\[ + i \left( \partial_P (P(F,F,\bar{F},\bar{F} \mid \bar{F}, \bar{F} \rangle \lambda^{-1}(F') \lambda^{-1}(F') \right) \]

\[ + \langle \bar{F}, \bar{F} \mid (D_{\alpha} (\Gamma_\alpha \delta_{\alpha \beta})^{-1} \Gamma_\epsilon \bar{F}_\epsilon (\Gamma_\delta \delta_{\delta \epsilon})^{-1} \mid F', F \rangle P_\alpha (\bar{F}) \rangle \] (6.5)

In obtaining this expression we have made repeated use of the identities (4.48) and (4.60). The state (6.4) above is simply (4.57) multiplied by \( |\lambda|^{1/2} \), the \( \phi_\alpha \)-dependent part of the factor \( |M|^{1/4} \) introduced in (6.2).

Next let us examine the remainder term in (6.2):

\[ \left( \frac{\partial}{\partial \phi_\alpha} \right) \left( \frac{\partial}{\partial \phi_\beta} \right) \left( \frac{\partial}{\partial \phi_\gamma} \right) \left( \frac{\partial}{\partial \phi_\delta} \right) \left( \frac{\partial}{\partial \phi_\tau} \right) \] (6.6)

The last term in (6.6) can be evaluated quite explicitly if we use the commutation relation (4.44) in the

206
form
\[ \lambda^{+1}(F_{\lambda} \delta \lambda^{+1}(F_{\lambda})'') - \lambda^{-1}(F_{\lambda} \delta \lambda^{-1}(F_{\lambda})') \] 
\[ = \frac{1}{\phi_{\mu}(F_{\lambda})} \int \frac{d^{3}r}{(2\pi)^{3}} \left\{ \langle \tilde{F}, \tilde{F} \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \Gamma_{m}^{-1} \Gamma_{m}' \mid \tilde{F}' \rangle \right\} \left( \frac{\delta}{\delta \phi_{\mu}(F_{\lambda})} \lambda^{(F_{\lambda})'} \right) \left( \frac{\delta}{\delta \phi_{\mu}(F_{\lambda})} \lambda^{(F_{\lambda})} \right) \] 
\[ - \langle \tilde{F}, \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \Gamma_{m}^{-1} \Gamma_{m}' \mid \tilde{F} \rangle \delta^{(F_{\lambda})} \delta^{(F_{\lambda})'} \right\}, \] 
(6.7) with the result
\[ \frac{1}{2} \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \int d^{3}r' \left\{ \langle \tilde{F}, \Gamma_{m} \mid \tilde{F}' \rangle \right\} \right) \left( \frac{\delta}{\delta \phi_{\mu}(F_{\lambda})} \lambda^{(F_{\lambda})'} \right) \left( \frac{\delta}{\delta \phi_{\mu}(F_{\lambda})} \lambda^{(F_{\lambda})} \right) \] 
\[ - \langle \tilde{F}, \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \Gamma_{m}^{-1} \Gamma_{m}' \mid \tilde{F} \rangle \delta^{(F_{\lambda})} \delta^{(F_{\lambda})'} \right\}, \] 
(6.8)
where \((\Gamma_{\mu})_{mn} = -i\delta_{nm}\) is the adjoint representation analog of the generator \(\Gamma_{\mu}\). The \(\phi\)-dependent term in this result cancels the \(\phi\) dependence of (6.5) so the combination of (6.5) and (6.8) will not depend on \(\phi\).

The second term in (6.6), linear in \(P_{0}^{i}(\tilde{F})\), is also not hard to simplify. Because the vector \(\tilde{E}\) defined by
\[ \langle i, \Gamma, l \mid \tilde{E} \rangle = \langle i, \Gamma, l \mid \delta_{\mu} - \delta_{\mu}(\Gamma_{m}^{-1} \Gamma_{m}) \rangle \] 
\[ \langle \tilde{E}, \Gamma, l \rangle = \langle \tilde{E}, \Gamma, l \rangle \] 
(6.9)
is automatically orthogonal to \(\tilde{E}\),
\[ \langle i, \Gamma, l \mid \tilde{E} \rangle = \langle i, \Gamma, l \rangle \]
given by (4.16), we can replace
\[ P_{0}^{i}(\tilde{F}) = \sum_{n} f_{n}^{i} \left( \frac{\delta}{\delta \phi_{\mu}} \right) \] 
by \(-i\delta/\delta A_{\mu}^{i}(\tilde{F})\) so that the second term becomes
\[ \frac{1}{2} g^{2} \int d^{3}r \int d^{3}r' \left\{ \langle \tilde{E}, \Gamma, l \mid \delta_{\mu} - \delta_{\mu}(\Gamma_{m}^{-1} \Gamma_{m}) \rangle \right\} \] 
\[ + \langle \tilde{E}, \Gamma, l \rangle \] 
(6.10)
This result can be combined with the second, \(\phi\)-independent term in (6.8) to yield
\[ \mathcal{U}_{1}(A) = \frac{1}{2} g^{2} \int d^{3}r \int d^{3}r' \left\{ \langle \tilde{E}, \Gamma, l \mid \delta_{\mu} - \delta_{\mu}(\Gamma_{m}^{-1} \Gamma_{m}) \rangle \right\} \] 
\[ + \langle \tilde{E}, \Gamma, l \rangle \] 
(6.11)
which, on account of the identity, valid for arbitrary states \(|X\rangle, \left| Y \right\rangle, \) and \(|Z\rangle, \)
\[ \int d^{3}r f^{*}(\tilde{E}, a) \langle \tilde{E}, a \mid \delta_{\mu} - \delta_{\mu}(\Gamma_{m}^{-1} \Gamma_{m}) \rangle \] 
\[ = \langle \tilde{E}, a \rangle \] 
\[ + \langle \tilde{E}, a \rangle \] 
(6.12)
can be further reduced to
\[ \mathcal{U}_{1}(A) = \frac{1}{2} g^{2} \int d^{3}r \left\{ \langle \tilde{E}, \Gamma, l \mid \delta_{\mu} - \delta_{\mu}(\Gamma_{m}^{-1} \Gamma_{m}) \rangle \right\} \] 
\[ \times \langle \tilde{E}, \Gamma, l \rangle \] 
\[ \times \langle \tilde{E}, \Gamma, l \rangle \] 
(6.13)
and therefore

\[ [P_{0}^{i}(F_{\lambda}), \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle] \] 
\[ = \frac{1}{2} \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ + \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ - \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ \times \langle \tilde{F}, \Gamma, l \rangle \] 
\[ \times \langle \tilde{F}, \Gamma, l \rangle \] 
(6.14)

\[ + \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ - \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ \times \langle \tilde{F}, \Gamma, l \rangle \] 
\[ \times \langle \tilde{F}, \Gamma, l \rangle \] 
(6.15)

\[ + \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ - \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ \times \langle \tilde{F}, \Gamma, l \rangle \] 
\[ \times \langle \tilde{F}, \Gamma, l \rangle \] 
(6.16)

\[ + \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ - \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ \times \langle \tilde{F}, \Gamma, l \rangle \] 
\[ \times \langle \tilde{F}, \Gamma, l \rangle \] 
(6.17)

\[ + \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ - \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ \times \langle \tilde{F}, \Gamma, l \rangle \] 
\[ \times \langle \tilde{F}, \Gamma, l \rangle \] 
(6.18)

\[ + \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ - \frac{g^2}{\phi_{\mu}(F_{\lambda})} \int d^{3}r \left\{ \langle \tilde{F}, \Gamma, l \mid \Gamma_{(\lambda_{\mu})}^{-1} \Gamma_{j} \Gamma_{j}' \rangle \right\} \] 
\[ \times \langle \tilde{F}, \Gamma, l \rangle \] 
\[ \times \langle \tilde{F}, \Gamma, l \rangle \] 
(6.19)
The results (6.5), (6.8), (6.12), and (6.14) can now be combined, giving a Weyl-ordered form for the $\chi(A_x) = 0$ gauge Hamiltonian

$$
\bar{H}(P, A) = g^{1/2} \tilde{H}(P, \alpha) g^{-1/2}
$$

$$
= \frac{i}{\hbar} \int d^2 r \left[ \left( - P^a_i(\bar{F}) P^a_i(\bar{F}) + \mathbf{U}_1(A) + \mathbf{U}_2(A) \right)
+ \frac{1}{\hbar} \int d^2 r \left[ d^2 r' \left( [-P^a_i(\bar{F}) D_i + g \psi T^i \phi][\Gamma_{x \delta} \Delta x^i]^{-1} \Gamma_{x \rho} \Gamma_{x \lambda} \Gamma_{x \beta}^{-1} \left[ \mathbf{D}_s P^a_i(\bar{F}) + g \psi T^i \phi \right] \right] \right)
+ \int d^2 r [\frac{1}{2} B^a_i B^a_i + \psi^a \alpha^a \left( i \nabla_j - g A^a_j T^i \phi \right)],
\right. \tag{6.15}
$$

where $\beta$ and $\beta_i(P, A)$ are given by (4.37) and (4.62), respectively, and $\{ \}$ indicates that the enclosed operators are to be Weyl-ordered, i.e., arranged in the same order as those inside the curly brackets in (6.2). Thus, with the substitution

$$
\tilde{\Psi}(A) = g^{1/2} \tilde{\Psi}(A),
$$

(6.16)

the time-dependent Schrödinger equation $i \partial \tilde{\Psi}(A) = \bar{H}(P, A) \tilde{\Psi}(A)$ becomes

$$
i \partial \tilde{\Psi}(A) = \bar{H}(P, A) \tilde{\Psi}(A).
$$

(6.17)

As was discussed in Sec. V, the Schrödinger Green's function, applied to the state $\tilde{\Psi}(A)$, can be written as

$$
\langle A' | e^{-iH(t' - t)} \rangle = \lim_{n \to +\infty} \prod_{n=1}^n \left[ d[A(n)] d[P(n)] d[\tilde{\Psi}(n)] d[\tilde{\Phi}(n)] \right]
\times \left( \exp \left\{ i \epsilon \sum_{n=1}^n \int d^2 r \left[ P^a_i(n) A^a_i(n+1) - A^a_i(n) \right] \epsilon^i - \frac{\tilde{H}(P(n), A(n+1) + A(n))}{2} \right\} \right)
\mathcal{N}(A_1(1)),
$$

(6.18)

where all notations are the same as those in (5.5) and (5.24), except for the replacements of $V(n)$ and $\Pi(n)$ by $A(n)$ and $P(n)$; hence, $A_i(n+1) = A_i(n) = \Pi_{l, l', l''} A^a_i(\bar{r}, \bar{t}, \bar{r}', \bar{t}'), A^a_i(n) = A_i(\bar{r}, \bar{t})$, $\tilde{\Psi}(n) = \tilde{\Psi}(\bar{r}, \bar{t})$, $\dot{\Phi}(n) = \dot{\Phi}(\bar{r}, \bar{t})$, $\ldots$. The quadratic interaction of the momenta $P_i(\bar{r})$ can be removed by introducing a Gaussian integral over a new variable $A_s^a(\bar{r}; \bar{t})$:

$$
\exp \left\{ - i \frac{1}{2} \sum_{n=1}^n \int d^2 r \int d^2 r' \left[ - P^a_i(\bar{r}) D_i + g \psi T^i \phi \right][\Gamma_{x \delta} \Delta x^i]^{-1} \Gamma_{x \rho} \Gamma_{x \lambda} \Gamma_{x \beta}^{-1} \left[ D_s P^a_i(\bar{r}) + g \psi T^i \phi \right] \right\}
= \int \prod_{n=1}^n \left[ d[A(n)] d[\tilde{\Psi}(n)] d[\tilde{\Phi}(n)] \right] \exp \left\{ i \epsilon \sum_{n=1}^n \left[ \int d^2 r \left[ P^a_i(n) \mathbf{D}_s A^a_s(\bar{r}, \bar{t}) - g \psi T^i \phi n_\alpha A^a_s(\bar{r}, \bar{t}) \right] \right] \right.
+ \frac{1}{2} \int d^2 r \int d^2 r' A^a_s(\bar{r}, \bar{r}', \bar{t}, \bar{t}', \bar{r}, \bar{t}) \left[ 1 - A^a_s(\bar{r}, \bar{t}) \right] \mathcal{N}(A_1(1)),
$$

(6.19)

where, apart from a constant factor, $\beta = \det |\Gamma_{x \delta} \Delta x^i|$ as in (4.37), which differs from

$$
[\det (\bar{\Omega} \Gamma_{x \delta} \Gamma_{x \rho} \Gamma_{x \lambda} \Gamma_{x \beta}^{-1} \bar{\Gamma}_{x \rho} \bar{\Gamma}_{x \lambda} \bar{\Gamma}_{x \beta}^{-1} \bar{\bar{D}}_x) ]^{1/2}
$$

only by another $A_s^a$-independent factor. One should note that just as $\bar{H}$ in (6.18) is evaluated at the symmetric point $\frac{1}{2} [A(n+1) + A(n)]$, the same rule must also apply to the operator $\mathbf{D}_s$ wherever it appears in (6.19). Finally, we can perform the $P_i$ integrations. Because $P_i$ obeys the constraint (4.15), $T_i P_j = 0$, we have

$$
\int d[P(n)] \exp \left\{ i \epsilon \sum_{n=1}^n \left[ \int d^2 r \left( P^a_i(n) \left( A^a_i(n+1) - A^a_i(n) \right) \epsilon^i + \Delta_s A^a_s(n) \right) - \frac{1}{2} P^a_i(n) P^a_i(n) \right] \right\}
= \exp \left\{ i \epsilon \left[ \int d^2 r \left( T^i \left( A^a_i(n+1) - A^a_i(n) \right) \epsilon^i + \Delta_s A^a_s(n) \right) \right]
- \frac{1}{2} \int d^2 r d^2 r' A^a_s(\bar{r}, \bar{r}', \bar{t}, \bar{t}', \bar{r}, \bar{t}) \left[ 1 - A^a_s(\bar{r}, \bar{t}) \right] \mathcal{N}(A_1(1)),
$$

(6.20)

Substituting (6.19) into (6.18), applying (6.20) to the $P_i$ integrations and taking the limit $\epsilon \to 0$, we find
Perturbation expansion of (6.21) gives the usual Feynman rules for the gauge \( \chi(A'_t) = 0 \) with the addition of the new term \( \mathcal{U}_1 + \mathcal{U}_2 \).

In order to justify this statement, we must recall the connection between occurrence of the symmetric combination \( \frac{1}{2} [A(n+1) + A(n)] \) in our definition of (6.18) and (6.19) and the usual Feynman rules. This connection can be most easily illustrated by considering the case of a simple one-dimensional harmonic oscillator with frequency \( \omega \).

Let the "contraction" between any two functions \( F(t) \) and \( G(t) \) of the coordinate \( q(t) \) and the velocity \( dq/dt \) be given by the standard expression

\[
F(t)G(t') = \frac{\int [dq(t')dq(t)] F(t')G(t') exp \left( -i \int L(q, dq/dt) dt \right)}{\int [dq(t')] F(t')G(t') exp \left( -i \int L(q, dq/dt) dt \right)}
\]

(6.22)

With our convention of replacing \( q(t) \) by the symmetric form \( \frac{1}{2} [q(n+1) + q(n)] \) and \( dq/dt \) by the anti-symmetric form \( [q(n+1) - q(n)] \), and then taking the limit \( \epsilon \to 0 \), it is straightforward to verify that

\[
\frac{dq(t)}{dt} \cdot \frac{dq(t')}{dt'} = D(t - t') = \frac{i}{\omega} \int e^{-ik_0(t-t')} \frac{\partial}{\partial k_0} dk_0,
\]

(6.23)

and

\[
\frac{dq(t)}{dt} \cdot \frac{dq(t')}{dt'} = -\frac{d^2}{dt^2} D(t - t')
\]

(6.24)

\[\omega D(t - t' + i\delta(t - t')) = \omega D(t - t') + i\delta(t - t').\]  

(6.25)

These equations correspond to the usual "covariant" Feynman propagators.

We note that in (6.22) when \( t = t' \), the product \( F(t)G(t) \) for \( F = dq/dt \) and \( G = q \) is, because of our choice,

\[
\frac{1}{2\epsilon} [q(n+1) - q(n)][q(n+1) + q(n)] = \frac{1}{2\epsilon} [q(n+1)^2 - q(n)^2],
\]

which, after the \( dq(n) \) and \( dq(n+1) \) integrations, clearly gives zero, in accordance with the right-hand side of (6.24).

\[
\frac{d}{dt} D(0) = \frac{1}{2\epsilon} \int \frac{k_0^2}{k_0^2 - \omega^2} dk_0 = 0.
\]

For \( F = G = dq/dt \), the factor \( F(t)G(t) \) is

\[
\frac{1}{\epsilon^2} [q(n+1) - q(n)]^2
\]

whose integrated value is \( \frac{1}{\epsilon} + i\epsilon^2 \), which in the limit \( \epsilon \to 0 \) agrees with the right-hand side of (6.25). This is to be contrasted with the usual Wick definition: the vacuum expectation value of \( \langle \langle dq(t)/dt \rangle dq(t'/dt') \rangle \), or \( T\{\langle dq(t')dq(t)\rangle \} \), which is finite when \( t = t' \) and differs from (6.25) by \( i\delta(t - t') \).

It is this difference that, when summed directly in the Dyson-Wick perturbation formalism, leads to the Jacobian term or Faddeev-Popov ghost term.

Our result (6.21) implies that for the gauge \( \chi(A'_t) = 0 \), in addition to the Jacobian \( e^{i\alpha} \), the usual Feynman rules must be augmented by the potential-like term \( \mathcal{U}_1 + \mathcal{U}_2 \) given in (6.12) and (6.14). Although the derivation is somewhat lengthy, the physical origin of both terms is the same as that of \( \mathcal{U}_1 \) and \( \mathcal{U}_2 \) in (2.14), for the simple mechanical example.

The Jacobian is usually converted into an additional term \( -i\delta(0) [n\eta] \) in the Lagrangian. In contrast, the new \( \mathcal{U}_1 + \mathcal{U}_2 \) terms have very different characteristics; they are both real and without the \( 5(0) \) factor. When expanded in power series of \( A'_t \), each can be written as

\[
\mathcal{U}_a = O(g^4A^4) + O(g^6A^6) + O(g^8A^8) + \cdots,
\]

(6.26)

where \( a = 1 \) or 2. [There is an \( O(g^3) \) constant term in \( \mathcal{U}_2 \), which can be dropped.] For example, in the Coulomb gauge and for SU(2), to order \( g^4 \) we have

\[
\mathcal{U}_1(A) = -\frac{g^4}{4} \int d^3r d^3r' d^3r''
\]

\[
K_{ij} (\vec{r} - \vec{r}') \hat{A}_{ij} (\vec{r}' - \vec{r}'' \hat{A}_{ij} (\vec{r}'') \cdot \hat{A}_{ij} (\vec{r}''))
\]

(6.27)

where

\[
K_{ij}(\vec{p}) = \left[ \frac{1}{2\epsilon} \delta^3(\vec{p}) - \frac{1}{4\pi \rho} (3\rho_i \rho_j - \rho^2 \delta_{ij}) \right] \frac{1}{4\pi \rho},
\]

(6.28)

where \( \rho = |\vec{p}| \) and the components of \( \hat{A}_{ij} \) are \( A'_t \).

The corresponding expression for \( \chi(A) \) is similar, but more complicated. In (6.28), the singular term \( \delta^3(\vec{p})/\rho \) gives rise to infinities which are presumably relevant for the cancellation of divergences from the usual two-loop \( g^4 \)-Feynman graphs; the remainder is finite when \( \rho = 0 \) and leads to new nonlocal interactions between the gauge fields which could be of physical importance.

VII. CONCLUSION

The \( \chi(A_t) = 0 \) gauge generating function \( ZU_t \) implied by the discussion of Sec. VI is
The final two terms in the exponent make up the usual classical action which, if a factor of $g$ is absorbed into $A^a$, is of order $g^{2d}$. The third Faddeev–Popov term is of order $g^2$ and peculiar to the functional–integral description, makes an essential contribution to the Feynman rules. The final two terms, of order $g^2$, have been the object of the discussion in this paper, and are given explicitly by (6.12) and (6.14). If Planck’s constant is distinguished from one, then these three classes of terms are of order $\hbar^4$, $\hbar^0$, and $\hbar$, respectively.

As we have seen, the extra terms $U_1$ and $U_2$ are precisely those which arise when the term in the quantum Hamiltonian $\hat{H}$, quadratic in the conjugate momenta $P(\vec{r})$, is Weyl ordered. In the pure Yang-Mills case, Eq. (6.15) can be written as

$$B(P,A) = \frac{1}{2} \int d^3 r \left( |E(\vec{r})|^2 + |B(\vec{r})|^2 \right) + U_1 + U_2,$$

(7.2)

where

$$E(\vec{r}) = P(\vec{r}) + \int d^3 \nu(\vec{r}, \nu) \Gamma(\vec{r}) \nu \cdot \Gamma(\vec{r}, \nu) P(\vec{r}, \nu),$$

(7.3)

and, as before, the subscript $W$ denotes Weyl ordering. For the case of the Coulomb gauge, our result (7.2) can be directly compared with that of Schwinger. In our notation, Schwinger’s Hamiltonian takes the form

$$\frac{1}{2} \int d^3 r \left( |E(\vec{r})|^2 + |B(\vec{r})|^2 \right) + U_1,$$

(7.4)

where the operator $E(\vec{r})$ defined by (7.3) is first to be Weyl ordered and then squared. The equality of the expression (7.4) and $\hat{H}$ in (7.2) follows from the identity

$$\frac{1}{8} \left[ g \left( \frac{\delta Q^a}{\delta F^a} + \frac{\delta F^a}{\delta Q^a} \right) \right]^2 = \frac{1}{8} \left[ g \left( \frac{\delta Q^a}{\delta Q^b} + \frac{\delta F^a}{\delta Q^b} \right) + 2 \frac{g}{\delta Q^a} \frac{\delta F^a}{\delta Q^b} + \frac{g}{\delta Q^a} \frac{\delta F^a}{\delta Q^b} \right]^2,$$

(7.5)

in which we may let $Q^a$ be that given by (4.13) and $F^a$ be

$$\frac{\delta Q^a}{\delta V^a(\vec{r})} = \int d^3 r' \left( \frac{\delta Q^a}{\delta V^a(\vec{r})} \right) \Gamma(\vec{r}, \nu) \cdot \Gamma(\vec{r}, \nu),$$

(7.6)

The last term on the right-hand side of (7.5) can then be written as

$$\int d^3 r \left( \frac{\delta Q^a}{\delta V^a(\vec{r})} \right) \frac{\delta F^a}{\delta Q^a},$$

(7.7)

which is precisely the first term inside the curly brackets in (6.6), the quantity defined as $U_2$. Thus, $[E^2(\vec{r})] = [E^2(\vec{r})] + U_2$ and the Hamiltonians (7.2) and (7.4) are identical.

Although the detailed arguments used in this paper to deduce the Hamiltonian operator $\hat{H}$ differ significantly from the method used by Schwinger, there is a close relationship between the initial physical assumptions. While Schwinger determines $\hat{H}$ so that the Lorentz group generators obey the proper commutation relations, we show that $\hat{H}$ is equivalent to a $V_g = 0$ gauge Hamiltonian with Cartesian operator ordering. Of course, the Lorentz group generators are gauge invariant and can be easily seen to obey the proper commutation relations when expressed in the $V_g = 0$ gauge. Hence, these two methods lead to the same Hamiltonian operator. However, it is the Weyl–ordered $\hat{H}$ of (7.2) given in this paper that must be used to deduce the Feynman rules.

Note added in proof. We wish to thank J.-L. Gervais, I. Muzinich, and T. N. Tudron for informing us of the following papers which have also discussed the transformation from the $V_g = 0$ gauge to the Coulomb gauge along lines similar to those in Sec. IV of our paper: V. N. Gribov, lecture at the 12th Winter School of the Leningrad Nuclear Physics Institute, 1977 (unpublished); J.-L. Gervais and B. Sakita, Phys. Rev. D 18, 453 (1978); M. Creutz, I. Muzinich, and T. N. Tudron, ibid. 19, 531 (1979); T. N. Tudron, Syracuse University Report No. SU-4217-156 (unpublished). Finally, we are indebted to R. Marmelius for bringing the work of R. Utiyama and J. Sakamoto [Prog. Theor. Phys. 55, 1631 (1976)] to our attention. These authors discuss the Coulomb–gauge operator–ordering problem from a viewpoint quite similar to ours. However, their technique for solving the Gauss constraint is significantly different from ours and the result seems more complex.
ACKNOWLEDGMENTS

We thank Edward Yao and Charles Sommerfield for bringing the work of J. Schwinger and R. P.  Treat to our attention. This research was supported in part by the U. S. Department of Energy.

APPENDIX

In this appendix, we consider the special case of SU(3) and give a simple physical interpretation of the operators $j^i, J^i$ and their commutation relations (4.44)–(4.47). There are three group parameters $\phi$, which may be chosen to be the Eulerian angles $\alpha, \beta, \gamma$. Thus, the $u$ matrix of (4.5) can be written as

$$u = e^{-i\phi_3/2} e^{-i\phi_2/2} e^{-i\phi_1/2},$$  \hspace{1cm} (A1)

where $\tau_1, \tau_2, \tau_3$ are the usual Pauli matrices.

We recall that the origin of Eulerian angles $\alpha, \beta, \gamma$ lies in the description of a rigid-body rotation. As shown in Fig. 1, there are two reference systems, the laboratory frame $\Sigma_{lab}$ and the body frame $\Sigma_{body}$. Each frame is defined by a basis of three orthogonal unit vectors: $\hat{x}, \hat{y}, \hat{z}$ for $\Sigma_{lab}$ and $\hat{X}, \hat{Y}, \hat{Z}$ for $\Sigma_{body}$. To go from $\Sigma_{lab}$ to $\Sigma_{body}$, we first rotate an angle $\beta$ along $\hat{Z} = \hat{z}$,

which moves the $y$ axis from $\hat{Y}$ to $\hat{y}$; then an angle $\gamma$ along $\hat{y}$, which rotates the $z$ axis from $\hat{Z}$ to $\hat{z}$; finally an angle $\alpha$ along $\hat{z}$.

Consider now a point $P$ in space, whose coordinates in $\Sigma_{lab}$ and $\Sigma_{body}$ are, respectively, $X, Y, Z$ and $x, y, z$. Let us define

$$R = \frac{1}{2}(\tau_x X + \tau_y Y + \tau_z Z)$$

and

$$r = \frac{1}{2}(\tau_x x + \tau_y y + \tau_z z).$$

From the definition of the Eulerian angles, it follows that

$$R = u ru^t,$$

where $u$ is given by (A1). (Here, $R$ plays the role of $V_i$ in the gauge-field case, and $r$ the role of $A_i$.)

We keep $\Sigma_{lab}$ fixed, and consider the rotation of $\Sigma_{body}$ by changing $a, b, c$. The angular velocity vector is

$$\dot{e}_a + \dot{e}_b + \dot{e}_c,$$  \hspace{1cm} (A2)

where the dot denotes a time derivative. Let us refer to the components of (A2) in $\Sigma_{body}$ and $\Sigma_{lab}$ as $\omega^i$ and $\Omega^i$, respectively. It is useful to define

$$iu^t \Omega = \omega = \frac{1}{2} \Gamma \cdot \omega$$  \hspace{1cm} (A3)

and

$$iu_\mu \Gamma = \Omega = \frac{1}{2} \Gamma \cdot \Omega.$$  \hspace{1cm} (A4)

Then, the matrices $\omega$ and $\Omega$ are related by

$$\Omega = u\omega u^t,$$  \hspace{1cm} (A5)

and the components of $\dot{\omega}$ and $\dot{\Omega}$ are the aforementioned $\omega^i$ and $\Omega^i$, related by

$$\Omega^i = U^{i\mu} \omega^\mu$$  \hspace{1cm} (A6)

with $U^{i\mu}$ given by (4.23). The quantities $\lambda^i$ and $\Lambda^i$, defined by (4.22) and (4.39), are related to $\omega^i$ and $\Omega^i$ by

$$\omega^i = \lambda^i \phi^a$$  \hspace{1cm} (A7)

The Lagrangian $\mathcal{L}$ of a rigid body with no external forces is a function only of $\omega^1, \omega^2, \omega^3$:

$$\mathcal{L} = \mathcal{L} (\omega^1).$$

Through (A2), $\mathcal{L}$ is also a function of $a, b, c$ and $\dot{a}, \dot{b}, \dot{c}$.

A comparison between (A8) and (4.43) gives

$$j^i = \frac{\partial \mathcal{L}}{\partial \omega^i},$$  \hspace{1cm} (A8)

which is simply the component of the angular momentum vector in $\Sigma_{body}$. The same vector viewed in $\Sigma_{lab}$ carries the component $J^i$:

$$\tilde{X} j^1 + \tilde{Y} j^2 + \tilde{Z} j^3 = \tilde{X} \mathcal{L} + \tilde{Y} \mathcal{L} + \tilde{Z} \mathcal{L},$$  \hspace{1cm} (A10)

or $J^i = U^{i\mu} \mathcal{L}^\mu$, as in (4.43).

In the quantum theory,

$$\mathcal{J} = \tilde{X} J^1 + \tilde{Y} J^2 + \tilde{Z} J^3.$$  \hspace{1cm} (A11)
is the rotational operator. Its components in $\Sigma_{\text{lab}}$

satisfy the usual commutation relation

$$[J^i, J^m] = i e^{im} \hat{J}^n.$$  \hfill (A12)

On the other hand, its components the $j^i$'s in $\Sigma_{\text{body}}$ do not. From (A10), we have $j^i = \hat{x} \cdot \hat{J}$ and similar expressions for $j^j$ and $j^k$. Since $\hat{x}$, $\hat{y}$, and $\hat{z}$ are unit vectors fixed in the body frame, under a rotation they transform like $\hat{J}$; all rotate like vectors. Hence, the scalar products $\hat{x} \cdot \hat{J}$, $\hat{y} \cdot \hat{J}$, and $\hat{z} \cdot \hat{J}$ are invariants, and that gives

$$[J^i, J^m] = 0.$$  \hfill (A13)

Likewise, we can show that

$$[j^i, j^m] = -i e^{im} j^n.$$  \hfill (A14)

Equations (4.44)−(4.47) are simply generalizations of these standard commutation rules, but for the gauge field.

It is important to note that in the generalized Laplacian, $\nabla$ (of (4.38)), the angular momentum (i.e., charge) operator that appears naturally is $j^i$. Since $j^i = U^i U^m$, the Gauss theorem (3.27) takes on the form

$$j^i(^) = -U^i U^m \nabla \psi |,$$  \hfill (A15)

which is the alternative expression for (4.60).

In terms of $a$, $b$, and $c$ the components of $\omega$ and $\Omega$ are given by

$$\omega^1 = \hat{a} \sin b - \hat{b} \sin a \cos c,$$

$$\omega^2 = \hat{a} \cos b + \hat{b} \sin a \sin c,$$

$$\omega^3 = \hat{b} \cos c + \hat{c}.$$  \hfill (A16)

The matrices $\lambda$ and $\Lambda$ can be obtained by differentiating these expressions with respect to $\hat{a}$, $\hat{b}$, and $\hat{c}$. Likewise, the components $j^1$ and $J^1$ can be expressed in terms of $p_x = -i/\theta a$, $p_y = -i/\theta b$, and $p_z = -i/\theta c$:

$$j^1 = \sin c p_x - \cos c p_y + \cos a \cos b p_z,$$

$$j^2 = \cos c p_x + \sin c \sin a p_y - \cos a \cos b \sin c p_z,$$

$$j^3 = p_z.$$  \hfill (A17)

By using (A18) and (A19), one can also verify the commutation relations (A12)−(A14) directly.
IS THE PHYSICAL VACUUM A MEDIUM?*

T. D. Lee

Department of Physics
Columbia University
New York, New York 10027

WHAT IS A VACUUM?

In the nineteenth century, to understand how the electromagnetic force, and later the electromagnetic wave, could be transmitted from one place to another, the vacuum was viewed as a medium, called aether. In his note 3075 on experimental research, Faraday wrote:

For my own part, considering the relation of a vacuum to the magnetic force and the general character of magnetic phenomena external to the magnet, I am more inclined to the notion that in the transmission of the force there is such an action, external to the magnet, than that the effects are merely attraction and repulsion at a distance. Such an action may be a function of the aether; for it is not at all unlikely that, if there be an aether, it should have other uses than simply the conveyance of radiations.

However, since at that time the nonrelativistic Newtonian mechanics was the only one available, the vacuum was thought to provide an absolute frame that could be distinguished from other moving frames by measuring the velocity of light. As is well known, this situation led to the downfall of aether and the rise of relativity.

Now we know that the vacuum is Lorentz invariant, which means that just by running around and changing the reference system, we are not going to alter the vacuum. Lorentz invariance is not everything, however. We may still ask: What is this vacuum state? In general, the vacuum is simply the lowest energy state of a Lorentz-invariant theory. Hence, the vacuum can be as complicated as any spin-0 field $\phi(x)$ at the zero 4-momentum limit:

$$\text{vacuum} \sim \phi \text{ at 4-momentum } k_\mu = 0. \tag{1}$$

Like a spin-0 field, perhaps the vacuum state can also carry quantum numbers, such as the isospin $I$, the parity $P$, and the strangeness $S$.

In quantum electrodynamics (QED), the complexity of the vacuum can be calculated in terms of the familiar power series in $\alpha = 1/137$. In the following, we will be particularly interested in the vacuum structure beyond QED. Furthermore, to answer the question in the title of this note, we must inquire whether the properties of the vacuum, like those of any medium, can actually be changed under suitable conditions. If the answer is an affirmative one, it would be physically meaningful to regard the vacuum as a medium. Our analysis will be based primarily on two observations: missing symmetry and quark confinement. These two observations are certainly among the most remarkable phenomena in modern physics.

*Supported in part by a contract with the U.S. Department of Energy.

111
0028-7113/80/0040-0111 $1.75/1 © 1980, NYAS

213
MISSING SYMMETRY

Symmetry principles have played an important part since the beginning of physics. Apart from their intrinsic aesthetic beauty, symmetry considerations have provided an extremely powerful and useful tool in our effort to understand nature. Gradually, they have become the backbone of our theoretic formulation of physical laws. Yet, especially over the past two decades, with few exceptions, most of the symmetries used in physics have been found to be broken. If we consider symmetry quantum numbers, such as the isospin I, the strangeness S, the parity P, . . . , we find

$$\frac{d}{dr} \begin{bmatrix} I \\ S \\ P \\ C \\ \vdots \\ \vdots \end{bmatrix}_\text{matter} \neq 0. \quad (2)$$

Aesthetically, this may appear disturbing. Why should nature abandon perfect symmetry? Physically, this also seems mysterious. What happens to these missing quantum numbers? Where do they go? Can it be that matter alone does not form a closed system? If we also include the vacuum, perhaps symmetry can be restored

$$\frac{d}{dr} \begin{bmatrix} I \\ S \\ P \\ C \\ \vdots \\ \vdots \end{bmatrix}_\text{matter + vacuum} = 0. \quad (3)$$

This is, of course, the basic idea underlying the general heading “spontaneous symmetry breaking.” In such a scheme, one assumes that there exists some spin-0 field $\phi$ that can carry the missing quantum number and whose vacuum expectation value is not zero:

$$\phi_{\text{vac}} = \langle \text{vac} | \phi | \text{vac} \rangle \neq 0. \quad (4)$$

Consequently, the observed asymmetry can be attributed entirely to the state vector of our universe, not to the physical law. On the other hand, unless we have other links that connect matter with the vacuum, how can we be sure that this idea is right and not merely a tautology?

A way out of this dilemma is to realize that in Equation 1, the restriction $k_a = 0$ for the vacuum state is only a mathematical idealization. After all, it is probable that the universe does have a finite radius, and $k_a$ is therefore never strictly zero. In regard to the microscopic system of particle physics, there is little difference between $k_a = 0$ and $k_a$ nearly 0; the latter corresponds to a state that varies only very slowly over a large space-time extension. This means that if the idea expressed by Equation 4 is correct, under suitable conditions, we must be able to produce excitations, or domain structures, in the vacuum. In such an excited state, there exists a volume $\Omega$ whose size
is much greater than the relevant microscopic dimension; inside Ω, we have \( \langle \phi(x) \rangle \neq \phi_{vac} \), but outside Ω, \( \langle \phi(x) \rangle = \phi_{vac} \).

How can we produce such a change in \( \langle \phi(x) \rangle \)? The problem is analogous to the formation of "domain structures" in a ferromagnet. We may draw the analog:

\[ \langle \phi(x) \rangle \leftrightarrow \text{spin} \]

\[ J \rightarrow \text{matter source} \leftrightarrow \text{magnetic field}, \]

as shown in Figure 1.

In the case of a very large ferromagnet, because its spin interacts linearly with the magnetic field, a domain structure can be created by applying an external magnetic field over a large volume. Similarly, by applying over a large volume any matter source \( J \) that has a linear interaction with \( \phi(x) \), one may hope to create\(^2\) a domain structure in \( \langle \phi(x) \rangle \). If the missing symmetry is due to \( \phi_{vac} \neq 0 \), by changing \( \langle \phi(x) \rangle \), we may alter the dynamics of symmetry change inside Ω. Of course, to perform a realistic experiment to reclaim all our missing symmetry is not easy. However, perhaps one should ask whether such domain structures in the form of vacuum excitations have already existed in nature. After all, there do exist numerous sources of matter in this universe. This leads us to discuss our second phenomenon: "quark confinement."

**Quark Model**

There is by now a fairly large body of evidence supporting the validity of the quark\(^3\) model of hadrons. The problem of quark confinement manifests itself in a most
explicit way in the $R$-value measurement in an $e^+e^-$ collision. Let us consider the diagrams for $e^+e^- \rightarrow \mu^+\mu^-$ and the quark–antiquark pair $q\bar{q}$, as shown in Figure 2. The threshold energies for these two reactions are $2m_e$ and $2m_q$, respectively. When the center-of-mass energy $E_{cm}$ is much larger than the threshold energy, and if we neglect the strong interaction between the $q\bar{q}$ pair, the ratio of the cross sections for these two reactions is

$$\frac{\sigma(e^+e^- \rightarrow q\bar{q})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)} = 3Q_q^2,$$

(5)

where $q$ stands for the flavor of the quark, which can be u, d, s, . . . . $Q_q$ is its corresponding charge in units of $e$, and the factor 3 is due to the three colors of each flavor. We define

$$R = \frac{\sum_q \sigma(e^+e^- \rightarrow q\bar{q} \rightarrow \text{hadrons})}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)},$$

where the sum extends over all quarks whose mass $m_q$ is $<\frac{1}{2}E_{cm}$. From Equation 5, we have

$$R \approx 3 \sum_q Q_q^2.$$

The experimental result is consistent with the theoretical expectation, drawn schematically in Figure 3. When $E_{cm}$ is $<2m_e$ but $>1$ GeV, which is assumed to be much larger than $2m_u$, $2m_d$, and $2m_s$, we have

$$R \approx 3 \left[ \left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right)^2 \right] = 2,$$

since the charges for u, d, and s are, respectively, $\frac{2}{3}$, $-\frac{1}{3}$, and $-\frac{1}{3}$. When $E_{cm}$ is between $2m_u$ and $2m_s$, we may write $R = 2 + \Delta R$, where $\Delta R \approx 3 \cdot \left(\frac{2}{3}\right)^2 = \frac{2}{3}$, because the charge of the c quark is $\frac{2}{3}$. When $E_{cm}$ is increased to above $2m_u$, the $R$ value should
increase by another amount $\Delta R = 3 \cdot (\frac{1}{2})^2 - \frac{1}{2}$. The experimental confirmation of these theoretical expectations strongly supports the following conclusions:

Quarks have fractional charges: $Q_q = \frac{2}{3}, -\frac{1}{3}, -\frac{1}{3}, \frac{1}{3},$ and $-\frac{1}{3}$, respectively, for $q = u, d, s, c,$ and $b$ quarks.

Each flavor variety has three colors, and the final states are symmetric with respect to color permutations; that is, the observed final hadron states are color singlets.

Quark “masses” are quite small. Except near resonances, strong interactions between quarks and antiquarks can be neglected (at least in the calculations of total cross sections). The critical question, then, is: Why in the final state do we not “see” these quarks in their free form? This is the well-known problem of quark confinement.

**Quantum Chromodynamics**

In quantum chromodynamics (QCD), there are quark fields $\psi_q$, where $q$, the flavor index, $= u, d, s, \ldots$, and $a$, denoting the color, $= 1, 2, 3$. In addition, we have the gauge fields $V^i$, where $i = 1, 2, \ldots 8$, representing the color index of the gauge-field quanta. The theory is a renormalizable one, so there is no difficulty with respect to divergences in the ultraviolet region. However, there are complications in the infrared region. To give QCD a well-defined meaning, we first contain the whole system within a volume of size $L^3$ to avoid infrared difficulties. Let $g_L$ be the renormalized coupling constant in the long-wavelength limit, momentum $k \sim L^{-1}$. It is possible to prove under rather general assumptions and valid to all orders in (coupling) that

$$g_L \geq g_l \quad \text{if } L > l. \quad (6)$$

This result is connected with the familiar “asymptotic freedom” property of the theory: when $l \to 0$, $g_l$ decreases to 0. Since there is no scale in the theory (assuming that all quarks are of zero mass), this implies that when $l$ increases, $g_l$ must also increase, which leads to Equation 6. The difficulty lies in the infrared limit. When $L \to \infty$, it seems likely that $g_L$ may $\to \infty$, or at least $\geq 1$. Since the true physical system is one with $L = \infty$, we may always be in the ultrastrong coupling limit. In the next section, we shall see how this difficulty can be resolved by regarding the vacuum as a color dielectric medium.

**Color Dielectric Constant**

Let us introduce $\kappa_L$, which is called the color dielectric constant of the vacuum in a volume $L^3$. As a convention, we shall adopt a standard renormalized coupling constant $g$, defined by

$$g = g_l \text{ when } l = \text{ some arbitrarily chosen length, say the proton radius.} \quad (7)$$
The dielectric constant $\kappa_L$ will then be defined as

$$g_L^2 = \frac{g^2}{\kappa_L}.$$  \hfill (8)

Consequently, in accordance with Equation 7,

$$\kappa_f = 1 \quad \text{when } l = \text{proton radius}.$$  \hfill (9)

Equation 6 now implies that

$$\kappa_L < \kappa_f \quad \text{if } L > l.$$  \hfill (10)

To the second order, we have

$$\frac{\kappa_L}{\kappa_f} = \frac{1}{1 + \frac{1}{2\pi} \frac{g^2}{4\pi} \left(11 - \frac{2}{3} n_f\right) \ln \frac{L}{l} + 0(g^4)},$$  \hfill (11)

where $n_f$ is the number of quark flavor varieties (assumed to be less than 17). With the convention of Equation 9, we find for the vacuum in an infinite volume $L = \infty$,

$$\kappa_v < 1,$$  \hfill (12)

which is valid to all orders in $g^2$. The ultrastrong coupling difficulty mentioned before corresponds to

$$\kappa_v = 0,$$  \hfill (13)

or

$$\kappa_v < 1.$$  \hfill (14)

In the former, we call the vacuum in QCD a perfect color dialectric medium, in the latter a nearly perfect color diaelectric medium.

**A HYPOTHETIC PROBLEM IN CLASSICAL ELECTROMAGNETISM**

In QED, our usual convention is to set the dielectric constant of the vacuum state $\kappa_{\text{vac}} = 1$. It is then possible to prove that all physical media have their dielectric constants $\kappa > 1$. This can be seen most easily by using the familiar formula

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P},$$  \hfill (15)

where $\mathbf{D}$ is the displacement vector, $\mathbf{E}$ the electric field, and $\mathbf{P}$ the polarization vector. Since under $\mathbf{E}$, all atoms have their polarization $\mathbf{P}$ in the same direction as $\mathbf{E}$, so as to produce a screening effect, we have $\kappa > 1$.

In this section, we shall consider a hypothetic problem. Let us imagine that in classical electromagnetism, but without the quantum theory of atoms, there could be a medium with a dielectric constant.

$$\kappa = \kappa_{\text{med}} \ll 1,$$  \hfill (16)
that is, this hypothetic medium is antiscreening. Now, suppose we place a small charge
distribution \( \epsilon \) in the medium. As we shall show, no matter how small \( \epsilon \) is, the medium
will develop a hole surrounding the charge. Inside the hole, we have \( \kappa = 1 \), but outside
\( \kappa = \kappa_{\text{med}} \), as shown in Figure 4. To see this, let us assume that such a hole is formed.
Because of the antiscreening nature of the medium, the induced charge on the inner
surface of the hole is of the same sign as \( \epsilon \). Consequently, if we want to reduce the size
of the hole, we must perform work to overcome the repulsion between \( \epsilon \) and the
induced charge. That work is infinite if the hole is to be eliminated in toto. Hence, the
hole will not disappear. (This situation is completely different for a normal medium
whose dielectric constant is \( >1 \). Imagine that a similar hole was created. The
Corresponding induced charge would have a sign opposite to that of \( \epsilon \). The hole would
automatically shrink to 0, resulting in the usual Coulomb field distribution around the
charge. The medium will remain homogeneous if \( \epsilon \) is sufficiently small.) Here, no

\[ \text{OUTSIDE} \quad \kappa = \kappa_{\text{med}} << 1 \]

\[ \kappa = 1 \quad \text{(INSIDE)} \]

**Figure 4.** Effect of a charge in a hypothetic dielectric medium in classical electromagnetism.

matter how small \( \epsilon \) is, when \( \kappa_{\text{med}} \rightarrow 0 \), inhomogeneity in the form of holes must occur in
the medium.

We may estimate the radius \( R \) of such a hole. Let \( D_{\text{out}} \) and \( E_{\text{out}} \) be, respectively, the
normal components of \( D \) and \( E \) outside the hole when \( r = R \). Similarly, let \( D_{\text{in}} \) and \( E_{\text{in}} \)
be the corresponding components inside the hole when \( r = R \). For a spherical hole, we have

\[ D_{\text{in}} = E_{\text{in}} = D_{\text{out}} = \frac{\epsilon}{R^2} \]

and

\[ E_{\text{out}} = \frac{\epsilon}{\kappa_{\text{med}} R^2} \]
where $\epsilon$ is the total charge of this small charge distribution, which is assumed to have a nonzero extension. The electric energy inside the hole is finite and independent of $\kappa_\text{med}$. The electric energy outside the hole is given by the volume integral of $D \cdot E$, which is $\propto \kappa_\text{med}^{-1}$. Therefore, when $\kappa_\text{med} \to 0$, the total electric energy $U_\text{el}$ is

$$U_\text{el} \sim \frac{1}{2} \frac{\epsilon^2}{\kappa_\text{med} R}.$$  \hspace{1cm} (17)

In addition, there is the energy $U_\text{hole}$ needed to create such a hole. The amount $U_\text{hole}$ is a function of $R$. When $R$ is large, $U_\text{hole}$ should be proportional to the volume plus a term proportional to the surface, and so on. We may write

$$U_\text{hole} = \frac{4\pi}{3} R^3 p + 4\pi R^2 s + \cdots,$$ \hspace{1cm} (18)

where $p$, $s$, ... are positive constants. The total energy of the system is given by

$$M = U_\text{el} + U_\text{hole},$$ \hspace{1cm} (19)

which is drawn schematically in Figure 5. Its minimum, as determined by $dM/dR = 0$, gives (in the approximation $U_\text{hole} \approx (4\pi/3) R^3 p$)

$$M = M_0 \sim \frac{4}{3} \left( \frac{\epsilon^2}{2\kappa_\text{med}} \right)^{3/4} (4\pi p)^{1/4},$$

from which we can draw the conclusion that if the total charge $\epsilon \neq 0$, $R \neq 0$. Furthermore, the total energy

$$M_0 \to \infty \text{ when } \kappa_\text{med} \to 0.$$ \hspace{1cm} (20)

It is not difficult to show that the same conclusion can also be reached without the approximation $U_\text{hole} \approx (4\pi/3) R^3 p$.

Next, we replace the single charge distribution by a dipole distribution, that is, two small charge distributions of total charge $\epsilon^+$ and $\epsilon^-$. It is easy to see that when $\kappa_\text{med}$ is sufficiently small, the minimal energy state again requires the formation of a hole around these two charges. As before, inside the hole $\kappa = 1$, and outside $\kappa = \kappa_\text{med}$. When

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure5.png}
\caption{Total energy $M = U_\text{el} + U_\text{hole}$ of the system shown in Figure 4.}
\end{figure}
\[ \kappa_{\text{med}} \to 0, \quad \text{at the surface of the hole, the electric field inside should be parallel to the} \]
\[ \text{surface so that } \mathbf{D} \text{ is } 0 \text{ outside, as shown in Figure 6. Thus, } U_d \text{ remains finite, as does} \]
\[ \text{the total energy } M_{e^+e^-}: \]
\[ M_{e^+e^-} = \text{finite when } \kappa_{\text{med}} \to 0. \quad (21) \]

If we try to separate the two charges \( e^+ \) and \( e^- \) by an infinite distance, because of Equation 20 the work required must also be infinite. This is the analog to "quark confinement" in our hypothetical problem.

**Quark Confinement**

We now return to the corresponding problem in QCD, discussed under QUANTUM CHROMODYNAMICS and COLOR DIELECTRIC CONSTANT. The vacuum is assumed to be a perfect (or nearly perfect) color dielectric medium. By following the same argument given above, we see that whenever there are quarks or antiquarks there must be inhomogeneity in the space surrounding these particles. We may call these either bags,\(^9,10\) or domain structures, or soliton solutions.\(^11\) Inside the "bag," the dielectric constant \( \kappa \) is 1. But outside, \( \kappa = \kappa_{\text{med}} \), which is 0, or \( \ll 1 \). If the total "color" is nonzero, the mass of such a bag would be infinite when \( \kappa_{\text{med}} \to 0 \), in analogy with Equation 20. However, just as in Equation 21, the bag mass remains finite when \( \kappa_{\text{med}} \to 0 \), if inside the bag one has a color singlet, such as

\[ \text{meson: } \bar{q}^a q^a, \quad \text{or baryon: } \epsilon_{abc} q^a q^b q^c, \]

where \( a, b, \) and \( c \) are color indexes that can vary from 1 to 3 (see Figure 7). From this one can derive that the work required to separate the quarks to a large distance \( r \) is approximately proportional to \( r \). The quark confinement is then "explained" by the assumption that the vacuum in QCD is a perfect (or nearly perfect) color dielectric.

Of course, at present it is not known that the mass \( M_{q} \) of a truly free quark is indeed infinite. However, the fact that no free quark has been observed so far in the
final state of any high-energy collision sets a lower bound \( M_q > 5 \text{ GeV} \) for \( q = u, d, \) and \( s \). It can be shown that this lower bound implies

\[
\kappa = 0.013 \left( \frac{g^2}{4\pi} \right), \tag{22}
\]

where \( g \) is defined by Equation 7.

One may wonder in what sense we have bypassed the difficulty of ultrastrong coupling, mentioned under QUANTUM CHROMODYNAMICS. To understand this, let us consider other problems in physics in which interactions are also ultrastrong and yet there are no mathematical difficulties. A good example is the interaction potential \( V(r) \) between two helium atoms. As shown in Figure 8, for \( r < a \), the potential \( V \) is \( \approx \) infinite; for \( r > a \), \( V \) is small. Because the strong potential is of a repulsive nature, the two helium atoms will avoid the strong-interaction region, and therefore this strong interaction does not present any difficulty. The usual method is to replace the strong repulsive part of \( V \) by a hard-sphere potential, which leads to a new boundary value

\[
V(r)
\]

\[
q
\]

\[
r
\]

\[
\text{FIGURE 8. Potential } V \text{ between two helium atoms.}
\]
problem: the distance $r$ between the two helium atoms is restricted to a region $r > a$ in which the potential $V(r)$ is small and therefore can be regarded as a perturbation.

From Figure 7, we see that when $\kappa_\mu \to 0$, that is, $g_\mu^2 = g^2/\kappa_\mu \to \infty$, the quarks are confined to a domain in which $\kappa = 1$. Thus, the ultrastrong region, $g_\mu \to \infty$, exerts a repulsion against the quarks and antiquarks. Just as in the above problem of helium–helium interaction, the particles automatically stay away from the ultrastrong-interaction region. Because of Lorentz invariance, the corresponding boundary-value problem in the present case requires the relativistic soliton solution, as has been given in the literature. When one expands around these soliton solutions, only the coupling $g$ inside the soliton is the relevant parameter, and that resolves the difficulty. (It is only when one attempts to make a simple-minded plane-wave expansion around a homogeneous background that one has to use $g_\mu$ as the expansion parameter. Exactly the same kind of difficulty would appear if one were to force a similar plane-wave expansion in the aforementioned helium–helium problem.)

\begin{figure}[h]
  \centering
  \includegraphics[width=\textwidth]{figure9.png}
  \caption{Superconductivity in QED versus quark confinement in QCD.}
\end{figure}

Because QCD is asymptotically free, the quarks inside the bag behave approximately like free particles, and their effective masses can be relatively small.

Since our theory is relativistic invariant, the velocity of light $c/(\kappa \mu)^{1/2}$ in the vacuum must remain $c$. Hence,

$$\kappa \mu = 1,$$

where $\kappa$ is the color dielectric constant and $\mu$ is the corresponding “magnetic” susceptibility. When $\kappa = \kappa_\mu \to 0$, one must have $\mu = \mu_\omega \to \infty$.

With Equation 23, we can consider $\kappa$ to be a Lorentz-invariant quantity; it plays the same part as the $\phi$-field discussed under WHAT IS A VACUUM? and MISSING SYMMETRY. It may be useful to introduce a phenomenological spin-$0$ field $\phi$ that depends linearly on $\kappa$:

$$\phi \propto 1 - \kappa.$$
All hadrons can be viewed as domain structures in the physical vacuum. Inside the hadron, we have \( \kappa = 1 \), and therefore \( \phi = 0 \). Outside the hadron, \( \kappa = \kappa_v \) and \( \phi = \phi_{\text{vac}} \). In the case that \( \kappa_v = 0 \), Equation 24 becomes simply \( \phi = \phi_{\text{vac}} (1 - \kappa) \).

REMARKS

As mentioned before, in QED there does not exist any dielectric medium at zero frequency; however, there do exist dispersion laws that exhibit dielectric natures at some nonzero frequencies.\(^{13}\) That can lead to physical effects somewhat similar to those discussed above.

Another analog is the comparison between the superconductivity in QED and the quark confinement in QCD. When we switch from QED to QCD, we replace the magnetic field \( H \) by the color electric field \( E \), the superconductor by the QCD vacuum, and the QED vacuum by the interior of the hadron; therefore in FIGURE 9, the inside by the outside and the outside by the inside. In the former, the magnetic field is expelled outward from the superconductor, whereas in the latter, the color electric field is pushed into the bag, which leads to quark confinement. This situation is summarized in TABLE 1.

<table>
<thead>
<tr>
<th>QED Superconductivity</th>
<th>QCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H )</td>
<td>( E )</td>
</tr>
<tr>
<td>( \mu_{\text{inside}} = 0 )</td>
<td>( \kappa_{\text{vacuum}} = 0 )</td>
</tr>
<tr>
<td>( \mu_{\text{vacuum}} = 1 )</td>
<td>( \kappa_{\text{inside}} = 1 )</td>
</tr>
<tr>
<td>Inside</td>
<td>outside</td>
</tr>
<tr>
<td>Outside</td>
<td>inside</td>
</tr>
</tbody>
</table>

Quark confinement is a large-scale phenomenon. Therefore, at least on the phenomenologic level, it should be understandable through a quasiclassical macroscopic theory, much like the London–Landau theory for superfluidity. In this paper, we suggest that the use of the color dielectric constant of the vacuum serves such a purpose.

By viewing the vacuum as a Lorentz-invariant medium, we would expect that many of its parameters, such as the color dielectric constant and the average values of some symmetry numbers can be subjected to changes. We may, by suitably arranging either the matter density or the energy content over a relatively large volume, alter the properties of the vacuum. Thereby, it may lead us to other striking phenomena, hitherto completely unknown to us.

REFERENCES


224
Symmetry Principles

Commentary

In paper [14], the question is raised of whether the observed baryon conservation law is
correlated to a local gauge transformation, similar to that connected to electric charge
conservation. This would lead to a vector field coupling to baryon number in the way that
the electromagnetic field couples to electric charge. The Eotvos experiment on the variation
between chemical elements of the ratio of inertial to gravitational mass shows that any
such coupling must be exceedingly weak, if it exists at all. This raises strong doubts about
the existence of the local gauge transformation for baryon number, and foreshadows the
later suggestion, by Sakharov and others, that baryon number conservation is only an
approximate symmetry.

In paper [20], it is shown that the combined invariance of the strong interactions under
charge conjugation and isospin implies a multiplicative conservation law for systems with
zero baryon number and hypercharge. The conserved quantum number \( G \) is \(-1\) for all
pions, so there cannot be strong transitions in which the pion number changes from odd
to even. This classification of states according to \( G \) proved very useful in later work on
pion-nucleon interactions and on weak interactions. In paper [41], the \( G \) quantum number
is used to analyze annihilations of nucleon and antinucleon into a system containing \( K \),
\( \bar{K} \) and pions. Correlations among the decay products are derived.

By early 1956, it had become clear as the result of analyses by R. Dalitz and others
that a problem existed with respect to the decays of \( K \) mesons. Decays occurred into both
two and three pions, which would not be allowed for a single, spin zero meson, if parity
were conserved in the decay. Following efforts to understand the situation through a variety
of hypotheses, Lee and Yang decided to analyze the evidence for parity conservation in
various particle interactions. The outcome was a historic series of papers, published over
a period of a few months in late 1956 and early 1957. The first result was paper [21], in
which it is shown that whereas there is good evidence for parity \((P)\) conservation in strong
and electromagnetic interactions, there was no such evidence for weak interactions. The
critical point is made that in order to test parity invariance in any interaction, it is necessary
to observe a pseudoscalar correlation. Specific tests to do this are proposed for the pro-
cesses of beta decay, pi-mu-electron decay, and hyperon decay. It is also pointed out that
no evidence exists for charge conjugation \((C)\) or time reversal \((T)\), invariance of weak
interactions.

Evidence was soon provided by experiments, carried out by Wu and collaborators, by
Lederman, Garwin and collaborators, and by Telegdi and collaborators that these pos-
sibilities were very real, and that weak interactions did not satisfy \( P \) or \( C \) invariance.
These early experiments seemed to show that \( CP \) invariance was satisfied. These results
immediately raised the question of what theory correctly describes weak decay processes.
T.D. LEE

Paper [24] describes a two-component theory of massless neutrinos, first discussed by Hermann Weyl, in which the neutrino has only a single spin state, and the antineutrino has opposite spin to the neutrino. It is shown that in this theory, $P$ and $C$ are not conserved, but that $T$ and $CP$ may be. This theory is equivalent to a four-component theory in which the neutrino field always is multiplied by the operator $1 + \gamma_5$. The paper goes on to discuss the implications of this theory for beta decay and pi-mu-electron decay processes, which turned out to be correct. At the time the paper was written, it was thought that the beta decay interaction included a tensor coupling. Later experiments showed that the coupling was instead an axial vector coupling, and it was straightforward to extend the two-component theory to that case.

Paper [25] discusses the implications of the $TCP$ theorem, proved earlier by Pauli and Luders for parity conserving interactions. It shows that $TCP$ invariance implies equality of lifetimes for particle and antiparticle, without the need for assuming $C$ or $CP$ invariance. It analyzes the consequences for experiment of various assumptions about $T$, $C$, and $P$ non-invariance, and shows, for example, that if $TP$ is conserved, then in the absence of final state interactions, a correlation proportional to the pseudoscalar (spin $\cdot$ momentum) cannot be observed. It also discusses the decays of neutral $K$ mesons, and shows that $TCP$ invariance alone implies the existence of two distinct lifetimes for these decays, but that further assumptions, such as $CP$ invariance, are necessary to imply equality of various partial decay rates. This analysis became directly relevant seven years later, when $CP$ non-invariance of neutral $K$-meson decays was discovered experimentally.

Paper [28] analyzes parity non-conservation in hyperon decay. It shows that an angular asymmetry of the decay pion with respect to the production plane would indicate parity non-conservation, and that this asymmetry could be large. Later experiments indeed detected such a large asymmetry in hyperon decays.

Paper [32] is based on lectures Lee delivered at Brookhaven National Laboratory in January 1957. It is a general survey of weak interactions as of that time, following the Wu, Lederman-Garwin, and Telegdi experiments. It was used for a long time afterwards as a working manual for weak interaction experiments and their theoretical analyses.

In 1964 Fitch, Cronin and collaborators made the experimental discovery that the decay of neutral $K$ mesons, $CP$ was not conserved. Later experiments showed that these decays also were not $T$ invariant, although in each case the amount by which the invariance was violated was quite small. In a series of papers, Lee and collaborators explored possible mechanisms for these violations.

In paper [73], it was proposed that the weak interaction Hamiltonian actually is $CP$ invariant, and the observed violation is an environmental effect, due to the existence of a long range vector field present in space, which acts oppositely on $K$ and $\bar{K}$, producing an energy difference between them. As a consequence of this, the amount of $CP$ violation would depend on the energy of the decaying meson. Later experiments showed that this possibility is not realized.

Paper [78] analyzes a variety of phenomenological models of $CP$ violation, in which the violating Hamiltonian has different strengths and different strangeness selection rules. The effect of these Hamiltonians on neutral $K$ meson decays is discussed.

In paper [80], it is hypothesized that the electromagnetic interaction of strongly interacting particles (hadrons) is $C$ violating, and that this is the origin of the observed $CP$ violation in neutral $K$ decays. Various tests of this hypothesis are proposed. Later experiments showed that this hypothesis is not satisfied in nature.
In paper [146], a model is proposed in which $T$ violation, and $CP$ violation occur through spontaneous symmetry breaking in a theory with a $T$ invariant Hamiltonian. In this paper, the model involves a complex scalar field, carrying a conserved quantum number, and interacting with a gauge field. This scalar field can be identified with the Higgs field in some versions of the unified theory of electromagnetic and weak interactions. It is shown that the $T$ violation at low energies should be small in this model. Paper [147] extends the idea of spontaneously broken $T$ violation to the interaction of a single real scalar field, while paper [149] discusses more detailed calculations of $T$ violating effects in these models. The models in which $T$ violation is a spontaneously broken symmetry remain one of the viable approaches to explaining this still mysterious aspect of physics.

In paper [54], the internal symmetries of the strong interactions are studied group theoretically. A criterion that any group must satisfy in order to describe observed properties of these symmetries is proposed.

Paper [93] discusses the relations between $C$, $P$ and $T$ symmetries, and internal particle symmetries that are satisfied in local field theories.
Conservation of Heavy Particles and Generalized Gauge Transformations

T. D. Lee, Columbia University, New York, New York

AND

C. N. Yang, Institute for Advanced Study, Princeton, New Jersey

(Received March 2, 1955)

The possibility of a heavy-particle gauge transformation is discussed.

The conservation laws of nature fall into two distinct categories: those that are related to invariance under space-time displacements and rotations, and those that are not. In the former category there are the conservation laws of momentum, energy, and angular momentum. In the latter category we find the conservation laws of electric charge, of heavy particles, and the approximate conservation laws of isotopic spin, and perhaps others. We notice that the best known within this second category, the conservation of electric charge, is related to invariance under gauge transformations, which expresses the nonmeasurability of the phase of the complex wave function of a charged particle.

We want to ask here whether similar gauge invariances should be related to all conservation laws of the second category. This question has been discussed in connection with the conservation of isotopic spin by Yang and Mills. We wish here to discuss the problem in connection with the conservation of heavy particles.

If we take the conservation of heavy particles to mean invariance under the transformation

\[ \psi_N \rightarrow e^{i\eta} \psi_N, \quad \psi_P \rightarrow e^{i\eta} \psi_P, \]

for the wave function of the heavy particles (neutrons and protons), a general gauge transformation (heavy-particle gauge transformation) is a transformation like (1) with the phase \( \alpha \) an arbitrary function of space-time. Invariance under such a transformation means that the relative phase of the wave function of a heavy particle at two different space-time points is not measurable.

Such a gauge transformation is formally completely identical with the electromagnetic gauge transformation. Invariance under such a transformation therefore necessitates the existence of a neutral vector massless field coupled to all heavy particles. A nucleon would have a "heavy-particle charge" of \( +\eta \) in such a field and an antinucleon would have a "heavy-particle charge" of \(-\eta\). The force between two massive bodies therefore would contain a contribution from the Coulomb-like repulsion between such "heavy-particle charges." The total force including the gravitational attraction is:

\[ \text{Force} = -G(M_1 M_2 / R^3) + \eta^2 (A_1 A_2 / R^2). \]

(2)

Here \( M_1, M_2, A_1, \) and \( A_2 \) are the inertia masses and mass numbers of the two bodies. There should also be a magnetic-dipole-like interaction between individual nuclei because the nucleons are in constant motion in a nucleus. But in a macroscopic object the nuclear spins average out so that (2) is correct unless the two bodies are spinning at high speeds.

Now the packing fraction of various atoms differ so that \( M/A \) varies fractional-wise from substance to substance by \( \sim 10^{-4} \). This means that the observed gravitation mass [which contains a contribution from the \( \eta^2 \) term in (2)] divided by the inertia mass would vary fractional-wise from substance to substance by \( 10^{-4} \eta^2 / G(M_P)^2 \), where \( M_P \) is the mass of the proton.

Very careful measurements by Eotvos and co-workers have shown this variation to be \( <10^{-8} \). Therefore

\[ \eta^2 / G(M_P)^2 < 10^{-8}. \]

It may be remarked that since the packing fraction differs most between hydrogen and, say, carbon, Eotvos' experiment could yield a more sensitive detection of \( \eta^2 \) by a factor of 10 if repeated with a comparison of hydrogen and carbon.

The assumption that leads to the above line of reasoning and the force expression (2) is that the phase factor \( \alpha \) in (1) should be space-time-dependent. It should be noticed that in addition the assumption has also been made that the transformation that generates the conservation of heavy particles is of the specific form (1).

We wish to thank Dr. J. Robert Oppenheimer for an interesting discussion.

Speculations on Heavy Mesons*

T. D. Lee and J. Oker

Department of Physics, Columbia University, New York, New York

(Received August 22, 1955)

From the analysis of Dalitz's it is becoming necessary
to assume that the $K_{e+}$ and $K_{e-}$ are due to at
least two different charged particles. Recent experi-
mental evidence encourages the hypothesis that the
observable lifetimes are exactly the same (at least after
time $>10^{-8}$ sec). As a solution to this possible
dilemma we propose that there exist two heavy mesons,
the $\theta^+$ and the $\tau^-$, where $\theta^+\rightarrow e^++\pi^+$ and
$\tau^-\rightarrow e^-+2\pi^-$. We propose further that the heavier of these two has
a lifetime of $10^{-8}$ sec with a significant branching ratio
for gamma decay to the lighter one. Finally, if we
assign a lifetime of order $10^{-9}$ sec or less to the lighter one,
we achieve a situation where both particles can have
different lifetimes but in most experiments appear
to have exactly the same lifetime. Because of Dalitz's
results we require the $\tau^-\rightarrow e^+$ to be a particle which cannot
decay into two pions because of spin and parity conserva-
tion. However, both $\tau^-$ and $\theta^+$ may contribute to
the $K_{e+}$, $K_{e-}$, and $K_{\nu}$ while the $\tau^-\rightarrow e^+$ may also contribute to
a part of the $K_{e-}$ (unless the spin and parity of the $\tau$
are 0+). We now discuss all the cascade possibilities.

Case I. $\theta$ and $\tau$ both have zero spin (single-gamma
decay forbidden).

$$\theta^+\rightarrow e^++2\gamma,$$

or

$$\tau^-\rightarrow e^+\gamma+2\gamma.$$  

(1)

(2)

With our present knowledge it is not possible to calcu-
late accurately how large a mass difference is needed to
give a two-gamma decay time of $10^{-8}$ sec. Our esti-
mates show that a mass difference as small as 10 Mev
is reasonable if one assumes that the $\theta^+$ and $\tau^-$ are strongly
coupled through nucleon-hyperon intermediate states.

Case II. $\theta$ and $\tau$ do not both have zero spin.—

(a) $M_{\theta}>M_{\tau}$.

$$\theta^+\rightarrow e^++\gamma.$$  

(3)

For magnetic dipole radiation, a mass difference $\leq 100$
kev is required. Electric dipole requires just a few kev.
A larger (and intuitively more reasonable) mass differ-
ence is permissible if we can manufacture a selection rule
forbidding the strong coupling between $\theta$ and $\tau$.
The most convenient selection rule in the framework
of the Gell-Mann scheme is conservation of strangeness,
$S$. Because of the existence of the $\theta^+$, the $\theta$ must have
its usual assignment

$$(S)_{\theta^+}=\pm 1 \quad or \quad I_{\theta}={\frac{1}{2}}.$$  

(4)

The only possibility left for the $\tau$ is

$$(S)_{\tau^+}=\pm 2 \quad or \quad I_{\tau}=0.$$  

(5)

Two important consequences of this assignment are
that the $\theta^+$ should not exist, in contrast to the existence
of $\theta^+$ and $\theta^-$, and that a $\tau^-$ cannot be produced in association
with a single $\Lambda^0$ or $\Sigma$. However, it may be produced by the following reactions:

$$\pi^-+p\rightarrow \Sigma^-+\tau^-.$$  

(6)

$$p+n\rightarrow \Lambda^0+\Sigma^0+\tau^-.$$  

(7)

(b) $M_{\theta}<M_{\tau}$. From the production experiments, the
$K$-mesons must be produced abundantly together with
$\Lambda^0$ and $\Sigma$. Since in this case the $\tau^-$ is to be the lon-
gerived particle, its strangeness must be $\pm 1$. Consequently
in this case the conservation-of-strangeness selection
rule is not possible and the mass difference must not
be greater than $\sim 100$ kev unless the gamma radiation
is of higher order than dipole.

The possible existence of a $\tau^\pm$ with a strangeness
different from the $\theta^\pm$ permits the $\tau^\pm$ to have a lifetime
very much shorter than the $\theta^\pm$, since they would then
not be strongly coupled through intermediate nucleo-
hyperon states. One is now tempted to speculate further
whether the decay of the $\tau^\pm$ may be connected in some
way with the universal Fermi interaction. We have
explored the possibility of regarding the universal Fermi
interaction as a secondary process via a virtual emission of $\tau^\pm$. In order to give the desired scalar and
tensor coupling for the Fermi interaction, the $\tau^\pm$
has to be described by a pseudovector field,$^4$ $\phi_\rho$ ($\rho=1,2,3,4$),
coupled to the fermions by

$$H=g_{\rho\lambda}\phi_\lambda\psi_\rho\psi_\rho-\epsilon_{\rho\eta\lambda\mu}\phi_\lambda\bar{\psi}_\rho\psi_\eta,$$

(8)

where $(A,B)$ represents the $(\rho,n)$, $(\mu,\nu)$, and $(e,\nu)$
groups, and $\epsilon_{\rho\eta\lambda\mu}$ is the isotopic antisymmetric tensor.
The interaction (8) would give the decay of $\tau^\pm$ into $3\nu$, $\mu^+\pi^-$, and $e^+\pi^-$ with comparable rates. If one sets
$g_{\rho\lambda}\sim g_{\lambda\nu}\sim g_{\nu\mu}$, then from the strength of the Fermi
interactions one obtains for the $\tau^\pm$ lifetime $\sim 10^{-16}$ sec.

We wish to thank the theoretical group at the Radiation
Laboratory, University of California for several
communications which have contributed to the
above discussion.

* Supported by the joint program of the U. S. Atomic Energy
Commission and the Office of Naval Research.

1. R. Dalitz, Proceedings of the Fifth Annual Rochester Confer-
ence on High-Energy Physics (Interscience Publishers, Inc.,
New York, 1955), and private communication.

100, 932 (1955)].


4. Conservation of strangeness is mathematically equivalent to
conservation of the $z$ component of isotopic spin, $I_z$. For mesons,
$S_{\tau}=Q$ where $Q$ is the charge.

5. According to the latest analysis of Dalitz, the $K_{e+}$ data
cannot exclude a $(1+)$ particle, although $(0+)$ gives a better fit.

All spin and parity configurations where two-pion decay is also
possible are strongly excluded for spins less than six.
Charge Conjugation, a New Quantum Number $G$, and Selection Rules Concerning a Nucleon-Antinucleon System.

T. D. Lee

Columbia University - New York, N.Y.

C. N. Yang

Institute for Advanced Study - Princeton, N.J.

(ricevuto il 30 Gennaio 1956)

Summary. (*) — A new quantum number is introduced which leads to selection rules concerning transitions between states with heavy particle number $= 0$. They are applied to the system nucleon-antinucleon; it is also shown that they can be extended to include the conservation of strangeness.

(*) Editor's care.

Whenever several conservation laws operate for the same system it is oftentimes possible to obtain new quantum numbers and new selection rules. For the pion-nucleon-antinucleon system in strong interaction one has besides the usual spin parity conservation laws the additional conservation laws of isotopic spin, charge conjugation and heavy particle number. We shall see that these conservation laws together do lead to a new quantum number for all systems with heavy particle number $= 0$. Application of this result to a nucleon-antinucleon system gives the selection rules tabulated in Tables I and II.

We shall be concerned with these operators:

$I_1, I_2, I_3 =$ isotopic spin operators ,

$G = C \exp [i\pi I_3]$

and $N =$ number of heavy particles.
Here $C$ is the charge conjugation operator. The advantage of using the operator $G$ instead of $C$ is that $G$ commutes with $I_1$, $I_2$, and $I_3$, while $C$ does not.

To see this one need only write down in the rest system the explicit form for these operators between the four states \(^1\) with the same $z$ component of spatial spin:

\[
\begin{pmatrix}
    \bar{p} \\
    n \\
    \bar{n} \\
    -\bar{p}
\end{pmatrix}.
\]

They are

\[
I_1 = \begin{pmatrix}
    0 & 1 & 0 & 0 \\
    1 & 0 & 0 & 0 \\
    0 & 0 & 0 & 1 \\
    0 & 0 & 1 & 0
\end{pmatrix}, \quad
I_2 = \begin{pmatrix}
    0 & -i & 0 & 0 \\
    i & 0 & 0 & 0 \\
    0 & 0 & 0 & -i \\
    0 & 0 & i & 0
\end{pmatrix}, \quad
I_3 = \begin{pmatrix}
    1 & 0 & 0 & 0 \\
    0 & -1 & 0 & 0 \\
    0 & 0 & 1 & 0 \\
    0 & 0 & 0 & -1
\end{pmatrix},
\]

\[
G = \begin{pmatrix}
    0 & 0 & 1 & 0 \\
    0 & 0 & 0 & 1 \\
    -1 & 0 & 0 & 0 \\
    0 & -1 & 0 & 0
\end{pmatrix}, \quad
N = \begin{pmatrix}
    1 & 0 & 0 & 0 \\
    0 & 1 & 0 & 0 \\
    0 & 0 & -1 & 0 \\
    0 & 0 & 0 & -1
\end{pmatrix}.
\]

\(^1\) Notice the minus sign in front of the $\bar{p}$ state in the convention here. The definition of $\bar{n}$ and $\bar{p}$ as the antineutron and antiproton states is here so chosen that they are identically related to the neutron and proton states. In other words, charge conjugation is here defined in the same way for the neutron and the proton.
TABLE II. - Selection Rules for $p+p \rightarrow m\pi$ or $\bar{n}+n \rightarrow m\pi$.

<table>
<thead>
<tr>
<th>State</th>
<th>Spin parity</th>
<th>$C$</th>
<th>$I$</th>
<th>$G$</th>
<th>$2\pi^0$</th>
<th>$\pi^+\pi^-$</th>
<th>$3\pi^0$</th>
<th>$\pi^+\pi^- + \pi^0$</th>
<th>$4\pi^0$</th>
<th>$\pi^+\pi^- + 2\pi^0$</th>
<th>$2\pi^+ + 2\pi^-$</th>
<th>$5\pi^0$</th>
<th>$\pi^+\pi^- + 3\pi^0$</th>
<th>$2\pi^+ + 2\pi^- + \pi^0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1S_0$</td>
<td>0$^-$</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$^3S_1$</td>
<td>1$^-$</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>$^1P_1$</td>
<td>1$^+$</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>$^3P_0$</td>
<td>0$^+$</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td></td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$^3P_1$</td>
<td>1$^+$</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$^3P_2$</td>
<td>2$^+$</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$x$ means strictly forbidden and $-$ means forbidden so far as the isotopic spin is a good quantum number.
For a multiple particle system, $I$ and $N$ are additive and $G$ multiplicative. One therefore has in general

\begin{align*}
(I, G) &= (I, N) = 0 , \\
GN + NG &= 0 , \\
G^2 &= (-1)^N .
\end{align*}

Besides $G$ is unitary by definition and $N$ has integral eigenvalues. Solving the relations (1) to (3) one obtains the following:

(1) For a state with a definite value $\neq 0$ for $N$, the quantum numbers are $N$, $I^2$, and $I_3$. The operation of $G$ on the system gives a state with the same $I^2$, $I_3$ but changes the sign of $N$.

(2) For a state with $N=0$ the quantum numbers are $N=0$, $I^2$, $I_3$ and $G = \pm 1$. Besides, all components of the same $I$-multiplet have the same value for $G$. The pions obviously belong to this category, with $I^2 = 1(1+1)$ and $I_3 = \pm 1$ or 0. To find their value for $G$ we notice that a $\pi^0$ has matrix elements connecting it to the $^1S_0$ state of a proton-antiproton system. Such a system has $G = + 1$ (1). Now $\pi^0$ has a total $I$-spin equal to 1. Therefore $\exp[i\pi I_3]$ for $\pi^0$ is $-1$. Hence by definition $G = -1$ for $\pi^0$. Therefore $G = -1$ for all pions, charged or neutral.

The existence of the quantum number $G$ leads to interesting selection rules (2) concerning transitions between states with $N=0$. In particular one sees that an even number of pions cannot by strong interaction go into an odd number of pions. When applied to a system consisting of a nucleon and an antinucleon one obtains the selection rules tabulated in Tables I and II.

The above consideration can easily be extended to include the conservation (3) of strangeness $S$. (S is defined to be $2Q - 2I_3 - N$). The commutation relations between $S$ and the other operators are

\begin{align*}
[N, S] &= [S, I] = 0 , \\
GS + SG &= 0 ,
\end{align*}


and

\[ G^s = (-1)^{N+S}. \]

Solving these commutation relations one obtains results similar to (1) and (2) discussed above, except that \( S \) is now an additional quantum number. \( G \) is a good quantum number = \( \pm 1 \) in this case only if both \( N \) and \( S \) are zero.

---

**RIASSUNTO (*)**

Si introduce un nuovo numero quantico \( G \) che porta a regole di selezione riguardanti le transizioni fra stati con un numero \( N \) di particelle pesanti nullo. Tali regole sono poi applicate al sistema nucleone-antinucleone e viene anche mostrato che esse possono essere estese in modo da comprendere la conservazione della "stranezza" \( S \).

---

(*) A cura della Redazione.
RECENT experimental data indicate closely identical masses\(^1\) and lifetimes\(^2\) of the \(\theta^+ (=K\pi\pi)\) and the \(\tau^+ (=K\pi\pi)\) mesons. On the other hand, analyses\(^3\) of the decay products of \(\tau^+\) strongly suggest on the grounds of angular momentum and parity conservation that the \(\tau^+\) and \(\theta^+\) are not the same particle. This poses a rather puzzling situation that has been extensively discussed.\(^4\)

One way out of the difficulty is to assume that parity is not strictly conserved, so that \(\theta^+\) and \(\tau^+\) are two different decay modes of the same particle, which necessarily has a single mass value and a single lifetime. We wish to analyze this possibility in the present paper against the background of the existing experimental evidence of parity conservation. It will become clear that existing experiments do indicate parity conservation in strong and electromagnetic interactions to a high degree of accuracy, but that for the weak interactions (i.e., decay interactions for the mesons and hyperons, and various Fermi interactions) parity conservation is so far only an extrapolated hypothesis unsupported by experimental evidence. (One might even say that the present \(\theta^+\rightarrow\tau^+\) puzzle may be taken as an indication that parity conservation is violated in weak interactions. This argument is, however, not to be taken seriously because of the paucity of our present knowledge concerning the nature of the strange particles. It supplies rather an incentive for an examination of the question of parity conservation.) To decide unequivocally whether parity is conserved in weak interactions, one must perform an experiment to determine whether weak interactions differentiate the right from the left. Some such possible experiments will be discussed.

\* Work supported in part by the U. S. Atomic Energy Commission.

\* Permanent address: Institute for Advanced Study, Princeton, New Jersey.


\* \(\tau^+\rightarrow\tau^+\) (dimension of system). (1)

The presence of such electric dipole moments would have interesting consequences. For example, if the proton has an electric dipole moment \( \pm e \times (10^{-16}\text{ cm}) \), the perturbation caused by the presence of the neighboring \( 2p \) state of the hydrogen atom would shift the energy of the \( 2s \) state by about 1 MeV/sec. This would be inconsistent with the present theoretical interpretations of the Lamb shift. Another example is found in the electron-neutron interaction. An electric dipole moment for the neutron \( \pm e \times (10^{-18}\text{ cm}) \) is the upper limit allowable by the present experiments.

By far the most accurate measurement of the electric dipole moment was made by Purcell, Ramsey, and Smith. They gave an upper limit for the electric dipole moment of the neutron of \( \pm e \times (5 \times 10^{-30}\text{ cm}) \). This value sets the upper limit for \( f^2 \) as \( f^2 < 3 \times 10^{-18} \), which is also the most accurate verification of the conservation of parity in strong and electromagnetic interactions. We shall see, however, that even this high degree of accuracy is not sufficient to supply an experimental proof of parity conservation in the weak interactions. For such a proof an accuracy of \( f^2 < 10^{-24} \) is necessary.

**QUESTION OF PARITY CONSERVATION IN \( \beta \) DECAY**

At first sight it might appear that the numerous experiments related to \( \beta \) decay would provide a verification that the weak \( \beta \) interaction does conserve parity. We have examined this question in detail and found this to be not so. (See Appendix.) We start by writing down the five usual types of couplings. In addition to these we introduce the five types of couplings that conserve angular momentum but do not conserve parity. It is then apparent that the classification of \( \beta \) decays into allowed transitions, first forbidden, etc., proceeds exactly as usual. (The mixing of parity of the nuclear states would not measurably affect these selection rules. This phenomenon belongs to the discussions of the last section.) The following phenomena are then examined: allowed spectra, unique forbidden spectra, forbidden spectra with allowed shape, \( \beta \)-neutrino correlation, and \( \beta-\gamma \) correlation. It is found that these experiments have no bearing on the question of parity conservation of the \( \beta \)-decay interactions. This comes about because in all of these phenomena no interference terms exist between the parity-conserving and parity-nonconserving interactions. In other words, the calculations always result in terms proportional to \( |C|^2 \) plus terms proportional to \( |C'|^2 \). Here \( C \) and \( C' \) are, respectively, the coupling constants for the usual parity-conserving (a sum of five terms) and the parity-nonconserving interactions (also a sum of five terms). Furthermore, it is well known that without measuring the spin of the neutrino we cannot distinguish the couplings \( C \) from the couplings \( C' \) (provided the mass of the neutrino is zero). The experimental results concerning the above-named phenomena, which constitute the bulk of our present knowledge about \( \beta \) decay, therefore cannot decide the degree of mixing of the \( C' \) type interactions with the usual type.

The reason for the absence of interference terms \( CC' \) is actually quite obvious. Such terms can only occur as a pseudoscalar formed out of the experimentally measured quantities. For example, if three momenta \( p_1, p_2, p_3 \) are measured, the term \( CC'p_1 \cdot (p_2 \times p_3) \) may occur. Or if a momentum \( p \) and a spin \( \sigma \) are measured, the term \( CC'p \cdot \sigma \) may occur. In all the \( \beta \)-decay phenomena mentioned above, no such pseudoscalars can be formed out of the measured quantities.

**POSSIBLE EXPERIMENTAL TESTS OF PARITY CONSERVATION IN \( \beta \) DECAYS**

The above discussion also suggests the kind of experiments that could detect the possible interference between \( C \) and \( C' \) and consequently could establish whether parity conservation is violated in \( \beta \) decay. A relatively simple possibility is to measure the angular distribution of the electrons coming from \( \beta \) decays of oriented nuclei. If \( \theta \) is the angle between the orientation of the parent nucleus and the momentum of the electron, an asymmetry of distribution between \( \theta = 90^\circ \) and \( \theta = 0^\circ \) constitutes an unequivocal proof that parity is not conserved in \( \beta \) decay.

To be more specific, let us consider the allowed \( \beta \) transition of any oriented nucleus, say \( Co^{60} \). The angular distribution of the \( \beta \) radiation is of the form (see Appendix):

\[
I(\theta)d\theta = (\text{constant})(1 + \alpha \cos \theta) \sin \theta d\theta,
\]

where \( \alpha \) is proportional to the interference term \( CC' \). If \( \alpha = 0 \), one would then have a positive proof of parity nonconservation in \( \beta \) decay. The quantity \( \alpha \) can be obtained by measuring the fractional asymmetry between \( \theta < 90^\circ \) and \( \theta > 90^\circ \); i.e.,

\[
\alpha = 2 \left[ \int_0^{90^\circ} I(\theta)d\theta - \int_{90^\circ}^{180^\circ} I(\theta)d\theta \right] / \int_0^{180^\circ} I(\theta)d\theta.
\]

It is noteworthy that in this case the presence of the magnetic field used for orienting the nuclei would automatically cause a spatial separation between the electrons emitted with \( \theta < 90^\circ \) and those with \( \theta > 90^\circ \). Thus, this experiment may prove to be quite feasible.

It appears at first sight that in the study of \( \gamma \)-radiation distribution from \( \beta \)-decay products of oriented nuclei one can form a pseudoscalar from the spin of the oriented nucleus and the \( \gamma \)-ray momentum \( p_3 \). Thus it may seem to offer another possible experimental test of parity conservation. Unfortunately, the nuclear levels have definite parities, and electromagnetic inter-
actions conserve parity. (Any small mixing of parities characterized by $\mathcal{F} < 3 \times 10^{-14}$ would not affect the arguments here.) Consequently the $\gamma$ rays carry away definite parities. Thus the observed probability function must be an even function of $p_r$. This property eliminates the possibility of forming a pseudoscalar quantity. It is therefore not possible to use such experiments as a test of parity conservation.

In $\beta-\gamma-\gamma'\gamma$ triple correlation experiments one can, by some rather similar but more complicated reasoning, prove that a measurement of the three moments cannot supply any information on the question of parity conservation in $\beta$ decay.

In $\beta-\gamma$ correlation experiments the nature of the polarization of the $\gamma$ ray could provide a test. To be more specific, let us consider the polarization state of $\gamma$ rays emitted parallel to the $\beta$ ray. If parity conservation holds for $\beta$ decay, the $\gamma$ ray will be unpolarized. On the other hand, if parity conservation is violated in $\beta$ decay, the $\gamma$ ray will in general be polarized. However, this polarization must be circular in nature and therefore may not lend itself to easy experimental detection. (The usual ways of measuring polarization through Compton effect, photoelectric effect, and photodisintegration of the deuteron are all incapable of detecting circular polarization. This is because circular polarization is specified by an axial vector parallel to the direction of propagation. From the observed momenta in these detection techniques such an axial vector cannot be formed.) For other directions of $\gamma$-ray propagation, elliptical polarization will result if parity is not conserved. This effect will thus be more difficult to detect.

QUESTION OF PARITY CONSERVATION IN MESON AND HYPERON DECAYS

If the weak interactions, such as the $\beta$-decay interactions or the decay interactions of mesons and hyperons, do not conserve parity, parity mixing will occur in all interactions by means of second-order processes. To examine this effect let us consider, for example, the decay of the $\Lambda^0$:

$$\Lambda^0 \rightarrow p + \pi^-.$$  

The assumption that parity is not conserved in this decay implies that the $\Lambda^0$ exists virtually in states of opposite parities. It could therefore possess an electric dipole moment of a magnitude

$$\text{moment} \sim e\mathcal{G} \times \text{dimension of } \Lambda^0,$$  

where $\mathcal{G}$ is the coupling strength of the decay interaction of the $\Lambda^0$. ($\mathcal{G} \lesssim 10^{-22}$.) The electric dipole moment of the $\Lambda^0$ is therefore $\lesssim e \times (10^{-34} \text{ cm})$.

Clearly the proton would have an electric dipole moment of the same order of magnitude. The existence of such a small electric dipole moment is, as we have seen, completely consistent with the present experimental information. Another way of putting this is to observe that by comparing Eq. (3) with Eq. (1), one has

$$\mathcal{F} \sim \mathcal{G}.$$  

Since all the weak interaction including $\beta$ interactions are characterized by coupling strengths $\mathcal{G} < 10^{-12}$, a violation of parity in weak interactions would introduce a parity mixing characterized by an $\mathcal{F} \sim 10^{-24}$. This is outside the present limit of experimental knowledge, as we have discussed before.

If the weak interactions violate parity conservation, parity would be defined and measured in strong and electromagnetic interactions only, just as strangeness is. Furthermore it is important to notice that with the conservation of strangeness, as with every conservation law, there is an element of arbitrariness introduced into the parity of all systems. The parity of all strange particles would be defined only up to a factor of $(\pm 1)^S$, where $S$ is the strangeness. The parity of the $\Lambda^0$ (relative to the nucleons) is therefore a matter of definition. But once this is defined, the parity of other strange particles would be measurable from the strong interactions.

POSSIBLE EXPERIMENTAL TESTS OF PARITY CONSERVATION IN MESON AND HYPERON DECAYS

To have a sensitive unequivocal test of whether parity is conserved in weak interactions, one must decide whether the weak interactions differentiate between the right and the left. This is possible only if one produces interference between states of opposite parities. The mere observation of two decay products of opposite parities originating from a "particle" cannot provide conclusive evidence that parity is not conserved. Such indeed is the state of affairs of the present $\theta - \tau$ puzzle.

As we have discussed before, these interference terms are possible only if the observed quantities can form a pseudoscalar such as $p_1 \cdot (p_2 \times p_3)$. The observation of $\Lambda^0$ decays in association with their production does provide such a possible pseudoscalar and hence a possible test of whether parity is conserved in the $\Lambda^0$ decay interaction. Let us consider the experiment

$$\pi^- \rightarrow p + \Lambda^0 + \nu, \quad \Lambda^0 \rightarrow p + \pi^-.$$  

Let $p_{in}$, $p_A$, and $p_{out}$ be, respectively, the momenta in the laboratory system of the incoming pion, the $\Lambda^0$, and the decay pion. We define a parameter $R$ as the projection of $p_{out}$ in the direction of $p_{in} \times p_A$. The value of $R$ ranges from approximately $-100$ Mev/c to approximately $+100$ Mev/c. Switching from a right-handed convention for vector products (which we use) to a left-handed convention means a switch of the sign of $R$. Parity conservation in the weak decay interaction of $\Lambda^0$ can therefore be experimentally checked by investigating whether $+R$ and $-R$ have equal probabilities of occurrence.
To see more clearly the meaning of the parameter $R$, one transforms $p_\text{out}(-p')$ into the center-of-mass system of $A$. The new vector $p'$ has a constant magnitude $\pm 100$ Mev/c. The frequency distribution of this vector $p'$ can then be plotted on a spherical surface. Taking the $z$ axis for this sphere to be in the direction of $p_\text{out} \times p_{\text{in}}$, one can prove the following two symmetries:

(a) The frequency distribution on the sphere remains unchanged under a rotation through 180° around the $z$ axis. This symmetry follows from parity conservation in the strong reaction producing the $A$'s. It does not depend on the nature of the weak interaction.

(b) If parity is conserved in the decay interaction of $A$, the frequency distribution on the sphere is unchanged under a reflection with respect to the production plane of $A$.

To prove statement (a), one need only consider the invariance of the production process under a reflection with respect to the production plane defined by $p_{\text{out}}$ and $p_{\text{in}}$. This reflection is the resultant of an inversion and a rotation through 180° around the $z$ axis (which is normal to the production plane). The state of polarization of $A'$ is thus invariant under a 180° rotation around the $z$ axis, leading to the stated symmetry.

Statement (b) follows directly from the assumption that the weak interaction as well as the strong interaction conserves parity. A reflection with respect to the production plane must then leave the whole process invariant.

The frequency distribution of $R$ is just the projection of the distribution on the sphere onto the $z$ axis. An asymmetry between $+R$ and $-R$ therefore implies parity nonconservation in $A^0$ decay. However, if the spin of $A'$ is unpolarized, no asymmetry can obtain even if parity is not conserved in $A^0$ decay. To obtain a polarized $A^0$ beam, the experiment is therefore best done at a definite nonforward angle of production of $A^0$ and at a definite incoming energy.

The above discussions apply also to any other strange particle decay if (1) the particle has a nonvanishing spin and (2) it decays into two particles at least one of which has a nonvanishing spin, or it decays into three or more particles. Thus the above considerations can be applied also to the decays of $\Sigma^+$ and $\Sigma^-$ as well as to $K^-\pi^+$, $K^+\pi^-$, and $K^0\pi^-(\equiv \pi^+)$. In the decay processes

$$\pi^-\mu^-\nu, \quad (5)$$

$$\mu^-e^-\nu^-\nu, \quad (6)$$

starting from a $\pi$ meson at rest, one could study the distribution of the angle $\theta$ between the $\mu$-meson momentum and the electron momentum, the latter being in the center-of-mass system of the $\mu$ meson. If parity is conserved in neither (5) nor (6), the distribution will not in general be identical for $\theta$ and $\pi - \theta$. To understand this, consider first the orientation of the muon spin. If (5) violates parity conservation, the muon would be in general polarized in its direction of motion. In the subsequent decay (6), the angular distribution problem with respect to $\theta$ is therefore closely similar to the angular distribution problem of $\beta$ rays from oriented nuclei, which we have discussed before. ( Entirely similar considerations can be applied to $\Sigma^-\rightarrow A^+\pi^-$ and $A^0\rightarrow p^-\pi^0$.)

**REMARKS**

If parity conservation is violated in hyperon decay, the decay products will have mixed parities. This, however, does not affect the arguments of Adair and of Treiman concerning the relationship between the spin of the hyperons and the angular distribution of their decay products in certain special cases.

One may question whether the other conservation laws of physics could also be violated in the weak interactions. Upon examining this question, one finds that the conservations of the number of heavy particles, of electric charge, of energy, and of momentum all appear to be inviolate in the weak interactions. The same cannot be said of the conservation of angular momentum, and of parity. Nor can it be said of the invariance under time reversal. It might appear at first sight that the equality of the life times of $\pi^+$ and of those of $\mu^+$ furnishes proofs of the invariance under charge conjugation of the weak interactions. A closer examination of this problem reveals, however, that this is not so. In fact, the equality of the life times of a charged particle and its charge conjugate against decay through a weak interaction (to the lowest order of the strength of the weak interaction) can be shown to follow from the invariance under proper Lorentz transformations (i.e., Lorentz transformation with neither space nor time inversion). One has therefore at present no experimental proof of the invariance under charge conjugation of the weak interactions. In the present paper, only the question of parity nonconservation is discussed.

---

8 This proof for statement (a) is correct only if $A^0$ exists as a single particle with definite parity in the strong interactions (as discussed in the last section); i.e., if $A^0$ does not exist as two degenerate states $A^0\pi^-$ and $A^0\pi^+$ of opposite parity, as has been suggested by T. D. Lee and C. N. Yang, Phys. Rev. 102, 290 (1956). It is to be emphasized, that if parity is indeed not conserved in the weak interactions, there would be (at present) no necessity to introduce the complication of two degenerate states of opposite parity at all. On the other hand, statement (b) is correct even if $A^0$ exists as two degenerate states $A^0\pi^-$ and $A^0\pi^+$ of opposite parity. To summarize, violation of the symmetry stated in (a) implies the existence of the decay $\pi^-\mu^-\nu$ with a mass difference less than their widths. Violation of the symmetry stated in (b) implies the nonconservation of parity in $A$ decay. See also footnote 12 and T. D. Lee and C. N. Yang, Phys. Rev. (to be published).

4 Also the interference may accidentally be absent if the relative phase between the two particles in the decay product is 90°. This, however, cannot be the case if time-reversal invariance is preserved in the decay process.


11 The existence of $A^0\pi^-$ and $A^0\pi^+$ of opposite parity may affect these relationships. This is similar to the violation of symmetry (a) discussed in footnote 8. See T. D. Lee and C. N. Yang, Phys. Rev. (to be published).
The conservation of parity is usually accepted without questions concerning its possible limit of validity being asked. There is actually no a priori reason why its violation is undesirable. As is well known, its violation implies the existence of a right-left asymmetry. We have seen in the above some possible experimental tests of this asymmetry. These experiments test whether the present elementary particles exhibit asymmetrical behavior with respect to the right and the left. If such asymmetry is indeed found, the question could still be raised whether there could not exist corresponding elementary particles exhibiting opposite asymmetry such that in the broader sense there will still be over-all right-left symmetry. If this is the case, it should be pointed out, there must exist two kinds of protons \( p_R \) and \( p_L \), the right-handed one and the left-handed one. Furthermore, at the present time the protons in the laboratory must be predominantly of one kind in order to produce the supposedly observed asymmetry, and also to give rise to the observed Fermi-Dirac statistical character of the proton. This means that the free oscillation period between them must be longer than the age of the universe. They could therefore both be regarded as stable particles. Furthermore, the numbers of \( p_R \) and \( p_L \) must be separately conserved. However, the interaction between them is not necessarily weak. For example, \( p_R \) and \( p_L \) could interact with the same electromagnetic field and perhaps the same pion field. They could then be separately pair-produced, giving rise to interesting observational possibilities.

In such a picture the supposedly observed right-and-left asymmetry is therefore ascribed to not a basic non-invariance under inversion, but to a cosmologically local preponderance of, say, \( p_R \) over \( p_L \), a situation not unlike that of the preponderance of the positive proton over the negative. Speculations along these lines are extremely interesting, but are quite beyond the scope of this note.

The authors wish to thank M. Goldhaber, J. R. Oppenheimer, J. Steinberger, and C. S. Wu for interesting discussions and comments. They also wish to thank R. Oehme for an interesting communication.

**APPENDIX**

If parity is not conserved in \( \beta \) decay, the most general form of Hamiltonian can be written as

\[
H_{\text{int}} = (\psi^* \gamma^0 \gamma^5 \psi) (C \bar{c} \gamma^5 \gamma^0 \bar{c} + C' \bar{c} \gamma^5 \gamma^0 \bar{c}) + (\psi^* \gamma^0 \gamma^5 \psi) (C \bar{c} \gamma^5 \gamma^0 \bar{c} + C' \bar{c} \gamma^5 \gamma^0 \bar{c}) + \frac{1}{2} (\psi^* \gamma^0 \gamma^5 \psi) (C \bar{c} \gamma^5 \gamma^0 \bar{c}) + (\psi^* \gamma^0 \gamma^5 \psi) (C' \bar{c} \gamma^5 \gamma^0 \bar{c})
\]

\[
\times (-C \bar{c} \gamma^5 \gamma^0 \bar{c} + C' \bar{c} \gamma^5 \gamma^0 \bar{c}) + (\psi^* \gamma^0 \gamma^5 \psi) (C \bar{c} \gamma^5 \gamma^0 \bar{c} + C' \bar{c} \gamma^5 \gamma^0 \bar{c})
\]

where \( \sigma_{\alpha} = -\frac{1}{2} (\gamma_\alpha \gamma_5 - \gamma_5 \gamma_\alpha) \) and \( \gamma_5 = \gamma^0 \gamma^1 \gamma^2 \gamma^3 \). The ten constants \( C \) and \( C' \) are all real if time-reversal invariance is preserved in \( \beta \) decay. This however, will not be assumed in the following.

Calculation with this interaction proceeds exactly as usual. One obtains, e.g., for the energy and angle distribution of the electron in an allowed transition

\[
N(W, \theta) dW \sin \theta d\theta = -\frac{\xi}{4\pi^3} \left( \frac{a \beta}{W} \right) \frac{1}{W} \frac{a \beta}{W} (w - W)^2 \times \left( 1 + \frac{a \beta}{W} \cos \theta + \frac{b}{W} \right) dW \sin \theta d\theta,
\]

where

\[
\xi = \left( |C_\alpha|^2 + |C_\beta|^2 + |C'_\alpha|^2 + |C'_\beta|^2 \right) |M_{\alpha \tau}|^2 + \left( |C_\alpha|^2 + |C_\beta|^2 + |C'_\alpha|^2 + |C'_\beta|^2 \right) |M_{\beta \tau}|^2,
\]

\[
a \xi = \left( |C_\alpha|^2 - |C_\beta|^2 - |C'_\alpha|^2 + |C'_\beta|^2 \right) |M_{\alpha \tau}|^2 - \left( |C_\alpha|^2 - |C_\beta|^2 + |C'_\alpha|^2 - |C'_\beta|^2 \right) |M_{\beta \tau}|^2,
\]

\[
b \xi = \gamma \left[ (C_\alpha^* C_\alpha + C_\beta^* C_\beta) + (C'_\alpha^* C'_\alpha + C'_\beta^* C'_\beta) \right] |M_{\alpha \tau}|^2 + \gamma \left[ (C_\alpha^* C_\alpha + C_\beta^* C_\beta) + (C'_\alpha^* C'_\alpha + C'_\beta^* C'_\beta) \right] |M_{\beta \tau}|^2.
\]

In the above expression all unexplained notations are identical with the standard notations. (See, e.g., the article by Rose.\(^{13}\))

The above expression does not contain any interference terms between the parity-conserving part of the interactions and the parity-nonconserving ones. It is in fact directly obtainable by replacing in the usual expression the quantity \( |C_\alpha|^2 \) by \( |C_\alpha|^2 + |C'_\alpha|^2 \), and \( |C_\beta|^2 \) by \( C_\alpha C_\alpha^* + C_\beta C_\beta^* \), etc. This rule also holds in general, except for the cases where a pseudoscalar can be formed out of the measured quantities, as discussed in the text.

When a pseudoscalar can be formed, for example, in the \( \beta \) decay of oriented nuclei, interference terms would be present, as explicitly displayed in Eq. (2). In an allowed transition \( J \rightarrow J-1 \) (no), the quantity \( \alpha \) is given by

\[
\alpha = \beta(J_J/(J_J-1),
\]

\[
\beta = \Re \left[ C_T C_T^* - C_A C_A^* - C_A^* C_A + C_T^* C_T \right] \frac{Z e^2}{\hbar c} \left( \frac{2}{c \xi + (2b/W)} \right)
\]

where \( M_{\alpha \tau}, \xi, \) and \( b \) are defined in Eqs. (A.3)–(A.5), \( v_\alpha \) is the velocity of the electron, and \( (J_J) \) is the average spin component of the initial nucleus. For an allowed transition \( J \rightarrow J+1 \) (no), \( \alpha \) is given by

\[
\alpha = -\beta(J_J)/(J_J+1).
\]

The effect of the Coulomb field is included in all the above considerations.

Errata

Question of Parity Conservation in Weak Interactions, T. D. Lee and C. N. Yang [Phys. Rev. 104, 254 (1956)]. Equation (A.4) in the Appendix of this paper should be amended to read as follows:

$$
\begin{align*}
\frac{d\xi}{d\tau} &= -\left( |C_2|^2 + |C_1|^2 + |C_0|^2 + |C_3|^2 \right) M_{F.}^2 \\
&\quad + \left( |C_T|^2 + |C_A|^2 + |C_{T'}|^2 + |C_{A'}|^2 \right) M_{G.T.}^2 \\
&\quad + 2 \frac{Z^2}{\hbar c p} \left[ \frac{e}{i} \left( C_0 C_3^* + C_2^* C_{T'} \right) M_{F.}^2 \\
&\quad - \frac{1}{3} \left( C_T C_A^* + C_{T'} C_{A'}^* \right) M_{G.T.}^2 \right].
\end{align*}
$$

(A.4)

This change does not affect the text of the paper, nor does it affect the other parts of the Appendix.

The authors are indebted to Dr. R. B. Curtis and to Dr. M. Morita\(^1\) for pointing out to them the error in the original Eq. (A.4).

Parity Nonconservation and a Two-Component Theory of the Neutrino

T. D. Lee, Columbia University, New York, New York

AND

C. N. Yang, Institute for Advanced Study, Princeton, New Jersey
(Received January 10, 1957; revised manuscript received January 17, 1957)

A two-component theory of the neutrino is discussed. The theory is possible only if parity is not conserved in interactions involving the neutrino. Various experimental implications are analyzed. Some general remarks concerning nonconservation are made.

Recently the question has been raised\(^1,2\) as to whether the weak interactions are invariant under space inversion, charge conjugation, and time reversal. It was pointed out that although these invariances are generally held to be valid for all interactions, experimental proof has so far only extended to cover the strong interactions. (We group here the electromagnetic interactions with the strong interactions.) To test the possible violation of these invariance laws in the weak interactions, a number of experiments were proposed. One of these is to study the angular distribution of the \(\beta\) ray coming from the decay of oriented nuclei. We have been informed by Wu\(^3\) that such an experiment is in progress. The preliminary results indicate a large asymmetry with respect to the spin direction of the oriented nuclei. Since the spin is an axial vector, its observed correlation with the \(\beta\)-ray momentum (which is a polar vector) can be understood only in terms of a violation of the law of space inversion invariance in \(\beta\) decay.

In view of this information and especially in view of the large asymmetry found, we wish to examine here a possible theory of the neutrino different from the conventionally accepted one. In this theory for a given momentum \(\mathbf{p}\) the neutrino has only one spin state, the spin being always parallel to \(\mathbf{p}\). The spin and momentum of the neutrino together therefore automatically define the sense of the screw.

In this theory the mass of the neutrino must be zero, and its wave function need only have two components instead of the usual four. That such a relativistic theory is possible is well known.\(^4\) It was, however, always rejected because of its intrinsic violation of space inversion invariance, a reason which is now no longer valid. (In fact, as we shall see later, in such a theory the violation of space inversion invariance attains a maximum.)

In Sec. 1 we describe this two-component theory of the neutrino. It is then shown in Sec. 2 that this theory is mathematically equivalent to a familiar four-component neutrino formalism for which all parity-conserving and parity-nonconserving Fermi couplings \(C\) and \(C'\) (as defined in the appendix of reference 1) are always related in the following manner: \(C_\alpha = C'_\alpha\), \(C_\tau = C'_\tau\), etc. or \(C_\alpha = -C'_\alpha\), \(C_\tau = -C'_\tau\), etc. Sections 3 to 8 are devoted to the physical consequences of the theory that can be put to experimental test. In the last section some general remarks about nonconservation are made.

I. NEUTRINO FIELD

1. Consider first the Dirac equation for a free spin-\(\frac{1}{2}\) particle with zero mass. Because of the absence of the mass term, one needs only three anticommuting Hermitian matrices. Thus the neutrino can be represented by a spinor function \(\phi\) which has only two components.\(^4\) The Dirac equation for \(\phi\), can be written as \((\hbar = c = 1)\)

\[
\sigma \cdot \mathbf{p} \phi = i \partial \phi / \partial t,
\]

where \(\sigma_1, \sigma_2, \sigma_3\) are the usual \(2 \times 2\) Pauli matrices. The relativistic invariance of this equation for proper Lorentz transformations (i.e., Lorentz transformations without space inversion and time inversion) is well known. In particular, for the space rotations through an angle \(\theta\) around, say, the \(z\) axis, the wave function transforms in the following way:

\[
\phi = \exp(-i \alpha \theta/2) \phi.
\]

The \(\sigma\) matrices are therefore the spin matrices for the neutrino. For a state with a definite momentum \(\mathbf{p}\), the energy and the spin along \(\mathbf{p}\) are given, respectively, by

\[
H = (\sigma \cdot \mathbf{p}),
\]

\[
\sigma = (\sigma \cdot \mathbf{p}) / |\mathbf{p}|.
\]

They are therefore related by

\[
H = |\mathbf{p}| \sigma.
\]

In the \(c\)-number theory, for a given momentum, the particle has therefore two states: a state with positive energy, and with \(\frac{1}{2}\) as the spin component along \(\mathbf{p}\), and a state with negative energy and with \(-\frac{1}{2}\) as the spin component along \(\mathbf{p}\).

It is easy to see that in a hole theory of such particles, the spin of a neutrino (defined to be a particle in the
positive-energy state) is always parallel to its momentum while the spin of an antineutrino (defined to be a hole in the negative-energy state) is always antiparallel to its momentum (i.e., the momentum of the antineutrino). Many of the experimental implications discussed in later sections are direct consequences of this correlation between the spin and the momentum of a neutrino.

We have remarked in the introduction that such a correlation defines automatically the sense of a screw. With the usual (right-handed) conventions which we adopt throughout this paper, the spin and the velocity of the neutrino represent the spiral motion of a right-handed screw while the spin and the velocity of the antineutrino represent the spiral motion of a left-handed screw.

We shall now discuss some general properties of this neutrino field:

(A) In this theory it is clear that the neutrino state and the antineutrino state cannot be the same. A Majorana theory for such a neutrino is therefore impossible.

(B) The mass of the neutrino and the antineutrino in this theory is necessarily zero. This is true for the physical mass even with the inclusion of all interactions. To see this, one need only observe that all the one-particle physical states consisting of one neutrino (or one antineutrino) must belong to a representation of the inhomogeneous proper Lorentz group identical with the representation to which the free neutrino states discussed above belong. For such a representation to exist at all, the mass must be zero.

(C) That the theory does not conserve parity is well known. We see it also in the following way: Under a space inversion $P$, one inverts the momentum of a neutrino but not its spin direction. Since in this theory the two are always parallel, the operator $P$ applied to a neutrino state leads to a nonexistent state. Consequently the theory is not invariant under space inversion.

(D) By the same reasoning one concludes that the theory is also not invariant under charge conjugation $C$ which changes a particle into its antiparticle but does not change its spin direction or momentum.

(E) It is possible, however, for the theory to be invariant under the operation $CP$, as this operation changes a neutrino into an antineutrino and simultaneously reverses its momentum while keeping the spin direction fixed. By the Lüders-Pauli theorem it follows that the theory can be invariant under time reversal $T$.

For the free neutrino field, as described by (1), one can prove that the theory is indeed invariant under time reversal and under $CP$.

2. We shall in this section indicate how one can use the conventional four-component formalism of the neutrino (with violation of parity conservation) and obtain the same results as the present theory.

We start from Eq. (1) and enlarge the matrices by the following definitions (1 represents a $2 \times 2$ unit matrix):

$$\epsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$\psi_r = \begin{pmatrix} \psi \end{pmatrix}, \quad \gamma = \begin{pmatrix} -i \beta \end{pmatrix}, \quad \gamma_4 = \gamma_1 \gamma_2 \gamma_4 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$ (4)

An immediate consequence of these definitions is

$$\gamma_4 \psi_r = -\psi_r.$$ (7a)

The free neutrino part of the Lagrangian is, as usual,

$$L = \psi_r^* \gamma_4 \left( \gamma_5 \frac{\partial}{\partial x} \right) \psi_r,$$ (8)

where $\psi_r^*$ is the Hermitian conjugate of $\psi_r$. The most general interaction Lagrangian not containing derivatives for the process $n \rightarrow p + e + \nu$ is exactly as usual; namely, it is the sum of the usual $S, V, T, A$, and $P$ couplings:

$$+ L_{int} = -H_{int} = \sum \{-2C, (\psi_r^* \partial_0 \psi_r) \psi_r^* \},$$ (10a)

where $i$ runs over $S, V, T, A$, and $P$ and

$$O_S = \gamma_4,$$  \quad $$O_V = \gamma_4 \gamma_5,$$  \quad $$O_T = \frac{1}{2} \psi_r^* (\gamma_5 \gamma_5 - \gamma_5 \gamma_5),$$ (11)

$$O_A = \frac{1}{2} \gamma_5 \gamma_4 \gamma_5,$$  \quad $$O_P = \gamma_4 \gamma_5.$$

It is not difficult to prove that Eqs. (5a) and (7a) are consistent with a relativistic theory even in the presence of the interaction (10a). Another way of proving this is to start from the conventional theory of the neutrino with the interaction Hamiltonian given in (A.1) of reference 1 and observe that when

$$C_S = -C_S', \quad C_V = -C_V', \quad \text{etc.}$$ (12a)

the neutrino field $\psi_r$ always appears in interactions in the combination $(1 - \gamma_5) \psi_r$. In the explicit representa-
tation that we have adopted above, this means that only the first two components of $\psi_0$ contribute to the interaction. All calculations using the conventional theory of the neutrino with the Hamiltonian (A.1) of reference 1 concerning $\beta$ decay therefore give the same result as the present theory if we take the choice of constants (12a).

There exists, however, the possibility that in the decay of the neutron a neutrino is emitted:

$$n \rightarrow \tilde{p} + e^- + \nu.$$  

(9b)

The corresponding general form (not including derivatives of the fields) of the Hamiltonian is

$$H_{\text{int}} = \sum [2C_e(\psi_1^* \psi_0)(\psi_1^* \psi_0^*)],$$

(10b)

where $O_i$ has been defined in Eq. (11). The field $\psi_1^*$ is a four-component spinor defined in terms of the two-component neutrino field $\psi$ by

$$\psi_1^* = \begin{pmatrix} 0 \\ \psi_0 \end{pmatrix}.$$  

(5b)

From Eq. (6), we see that

$$\gamma_\alpha \psi_1^* = + \psi_0^*.$$  

(7b)

It can be shown that (5b) and (7b) are consistent with a relativistic theory even in the presence of interaction (10b). It can also be proved that one can use again the Hamiltonian (A.1) of reference 1 for the conventional theory of the neutrino with the choice of the coupling constants

$$C_{\tilde{e}} = C_e,$$

(12b)

and obtain the same result as the present theory.

The two possible choices (12a) and (12b) depend on whether, in the $\beta$ decay of the neutron, process (9a) or (9b) prevails, i.e., whether a neutrino or an antineutrino is emitted. We shall see in Sec. 3 that experimentally it will be easy to decide which of the two choices is appropriate (if the theory is correct). [We do not consider the possibility here of the simultaneous presence of (9a) and (9b), since the double beta decay process does not seem to be observed experimentally.]

II. EXPERIMENTAL IMPLICATIONS

3. We consider in this section the experiment of the $\beta$ decay of oriented nuclei already discussed in reference 1, and currently being carried out. For the present theory, according to Eqs. (12a) or (12b), Eq. (A.6) reduces to

$$\beta = \frac{\mu \nu / \hbar c}{|C_e|^2 |C_{\tilde{e}}|^2 (2Ze^2/\hbar c \nu) \text{Im}(C_e C_{\tilde{e}}^*)},$$

(13)

The choice of the $\mp$ sign depends on whether

$$n \rightarrow \tilde{p} + e^- + \nu$$

($\mp$ left-handed screw),

or

$$n \rightarrow \tilde{p} + e^- + \nu$$

($\pm$ right-handed screw).

(14)

In writing down (13) the Fierz interference terms has been set equal to zero, which is in conformity with the experimental results, and which implies [see Eq. (A.5) of reference 1]:

$$\text{Real part of } C_e C_{\tilde{e}}^* = 0.$$  

(15)

By measuring the momentum dependence of the asymmetry parameter $\beta$, one can test whether the present theory is correct.

It is interesting to notice that for a positron emitter the asymmetry parameter has the opposite sign. This is a direct consequence of the fact that in positron and electron emission, the neutrino and antineutrino emitted have opposite spindles.

4. An experiment such as the one being carried out by Cowan and collaborators measures the cross section for neutrino absorption, which can be calculated in both the present theory and the usual theory. Now one determines the magnitude of the $\beta$-coupling constants to give the observed lifetimes of nuclei against $\beta$ decay. The calculated value of the cross section turns out then to be twice as great in the present theory as in the usual theory. This follows from the following simple reasoning: The neutrino flux is an experimental quantity independent of the theory. If the neutrinos in a given direction have only one spin state instead of the usual two, by a detailed balancing argument they must have twice the cross section for absorption as the usual ones.

5. In the decay of $\pi^\pm$ mesons at rest, let us consider the component of angular momentum along the direction of $p_n$ the momentum of the $\mu$ meson. The orbital angular momentum contributes nothing to this component. The $\mu$ spin component is therefore completely determined (irrespective of its total spin) by the spin component of the $\nu$ or $\bar{\nu}$. There are then two possibilities:

(A) $$\mu^+ \rightarrow \mu^- + e^+ + \nu_{\mu}$$ (spin along $p_\mu = + \frac{1}{2}$),

$$\mu^- \rightarrow \mu^+ + e^- + \nu_{\mu}$$ (spin along $p_\mu = - \frac{1}{2}$),

(16)

or

(B) $$\mu^+ \rightarrow \mu^- + e^+ + \nu_{\mu}$$ (spin along $p_\mu = - \frac{1}{2}$),

$$\mu^- \rightarrow \mu^+ + e^- + \nu_{\mu}$$ (spin along $p_\mu = + \frac{1}{2}$).

(17)

In each case the $\mu$ mesons with fixed $p_\mu$ form a polarized beam. (It was pointed out in reference 1 that if parity is not conserved in the decay of $\pi$ mesons, the $\mu$ mesons would in general be polarized.) Furthermore, the polarization is now complete (i.e., in a pure state). If this theory of the neutrino is correct, then the $\pi^- - \mu$ decay is a perfect polarizer of the $\mu$ meson, offering a...
natural way to measure the spin and the magnetic moment of the $\mu$ meson. (It turns out that the $\mu^-\gamma$ decay may serve as a good analyzer, as we shall discuss in the next section.)

The choice of the two possibilities (16) and (17) will be further discussed in Sec. 7.

6. For the $\mu^-\gamma$ decay the process can be

\[ \mu^-\rightarrow\nu+\bar{\nu}, \]  

or

\[ \mu^-\rightarrow\mu^-+2\nu, \]  

or

\[ \mu^-\rightarrow\nu+2\bar{\nu}. \]  

Consider process (18) first. The decay coupling can be written with the notations defined in Eq. (11). (We assume no derivative coupling.)

\[ H_{\text{int}} = \sum_{\psi_1, \psi_2} f_1(\psi_1, \Omega_1, \psi_2)(\psi_2, \Omega_1, \psi_2). \]  

(21)

It is easy to see that in the present theory, where $\psi$ satisfies (7a), the $S, T,$ and $P$-type couplings do not exist. We have assumed in writing (21) that the spin of the $\mu$ meson is $\frac{1}{2}$. For a $\mu^-\gamma$ at rest with spin completely polarized, the normalized electron distribution is given by

\[ dN = 2\pi x(3-2x) + \xi\cos^2(1-2x)\Omega, (4\pi)^{-1}, \]

(22)

where $\rho$ = electron momentum, $x = \rho/\text{maximun electron momentum}$, $\theta = \angle$ between electron momentum and the spin direction of the $\mu$, $\Omega$ = solid angle of electron momentum, and

\[ \xi = \left[ |f_1|^2 + |f_2|^2 \right]^{-1} \int f_1 f_2^* + f_2 f_1^* \].

(23)

The mass of the electron is neglected in this calculation. The decay probability per unit time is ($\hbar = c = 1$):

\[ \lambda = M^2 \left[ |f_1|^2 + |f_2|^2 \right] / (3 \times 2 \pi), \]

(24)

where $M$ is the mass of the $\mu$ meson. The spectrum (22) for a nonpolarized $\mu$ meson,

\[ dN = 2\pi x(3-2x)d\Omega, (4\pi)^{-1}, \]

(25)

is characterized by a Michel parameter $\rho = \frac{1}{2}$, which is consistent with known experimental results.

One sees that for not too small values of $\xi$, the spectrum (22) is sensitive to $\cos \theta$, especially in the region of large momentum for the electrons. Therefore the $\mu^-\gamma$ decay may turn out to be a very good analyzer of the $\mu$-meson spin.

An analysis of the so-called universality of the Fermi couplings is easier in this theory because there are fewer coupling constants, and also because $\pi^-\mu^-\gamma$ decay measurements would supply information concerning the parameter $\xi$ of (23).

If process (19) or (20) prevails, the spectrum becomes

\[ dN = 12\pi(1-x)dx(1+y\cos\theta)d\Omega, (4\pi)^{-1}. \]  

(26)

This is characterized by a Michel parameter $\rho = 0$ which is not consistent with experiments. One therefore concludes that (18) is the correct process.

A general theorem concerning the relationship between $\mu^+$ and $\mu^-$ decays will be stated in Sec. 9.

7. If experiments should show that in the decay of the $\pi$ meson, process (16) prevails, and in the $\beta$-decay process (9a) prevails, then one would say that the $\nu$ (the right-handed screw), the $\mu^-$, and the $\nu$ are light particles, and there is a conservation of light particles. If processes (17) and (9b) prevail, one would say that the $\bar{\nu}$ (the left-handed screw), the $\mu^+$, and the $\nu$ are light particles, and there is a conservation of light particles. Similar concepts have been discussed before.

We have already seen in Sec. 3 that the sign of $\beta$ in Eq. (13) determines whether

\[ n \rightarrow p + e + \nu \]  

(9a)

or

\[ n \rightarrow p + e + \bar{\nu} \]  

(9b)

is the process for $\beta$ decay. To decide whether

\[ \pi^+ \rightarrow \mu^+ + \nu, \quad (\mu^+ \text{ spin along } p, \lambda = \frac{1}{2}) \]  

(16)

or

\[ \pi^+ \rightarrow \mu^- + \bar{\nu}, \quad (\mu^- \text{ spin along } p, \lambda = -\frac{1}{2}) \]  

(17)

one will have to determine the spin of $\mu^+$ along its direction of motion.

8. The $\pi^-\mu^-\gamma$ type experiment discussed in Secs. 6 and 7 can be done with the $K_{\mu_2-\mu^-\gamma}$ decays. The analysis is dependent on the spin of $K_{\mu_2}$. If this spin is not zero, the polarization of the $\mu$ meson is not necessarily complete. The degree of polarization can be experimentally found by a comparison of the angular distribution of the electrons in $\pi^-\mu^-\gamma$ decay and in $K_{\mu_2-\mu^-\gamma}$ decay.

Another interesting experiment is to measure the momentum and polarization of the electron emitted in a $\beta$ decay. A polarization of the electron results only if parity is not conserved; a measurement of this polarization is a measurement of a quantity similar to the parameter $\beta$ in Eq. (13). The polarization in such a case will be along the direction of the momentum of the electron. Polarization along other directions can result if the momentum of the recoil nucleus is also determined. Theoretical considerations of such possibilities are being made by Dr. R. R. Lewis.

GENERAL REMARKS

9. Some general remarks concerning the conservation and nonconservation of the parity $P$, the charge conjugation $C$, and the time reversal $T$ will be made in this section. Except for the last paragraph, no assump-


\[ ^{11} \text{See, e.g., Sargent et al., Phys. Rev. 99, 685 (1955).} \]

\[ ^{12} \text{E. J. Konopinski and H. M. Mahmoud, Phys. Rev. 92, 1045 (1953).} \]
tion that the neutrino is a two-component wave is made.

Since the preliminary result of the oriented nucleus experiment that there is a strong asymmetry, Eq. (A.6) of reference 1 shows that not only parity, but also charge conjugation is not conserved in $\beta$ decay. A measurement of the velocity dependence of the asymmetry parameter could supply some information concerning time reversal invariance or noninvariance. If the $\pi^-\mu^-\epsilon$ decay should show any forward-backward asymmetry (as discussed in reference 1, and further analyzed above in Sec. 6 for the two-component neutrino theory), it can be shown from theorem 2 of reference 2 that charge conjugation invariance must be violated in both the $\pi^-\mu$ and $\mu^-\epsilon$ decays.

It is, however, easy to show from the Lüders-Pauli theorem that even if $C$, $T$, and $P$ are all not conserved, a stable particle ($\pi^\pm$ or $\mu^\pm$, or a deuteron, etc.) must have exactly the same mass as its antiparticle.

One can also prove that even if $C$, $T$, and $P$ are all not conserved, the $\epsilon^+$ angular distribution in $\pi^-\mu^-\epsilon^+$ decay is exactly the same as the $\epsilon^-$ angular distribution in $\pi^-\mu^-\epsilon^-$ decay. The only difference in the two cases is that the average spin of $\mu^+$ along $p_\mu$ is the opposite of that of $\mu^-$ along $p_\mu$. (The decays are here assumed to occur in free space from $\pi^\pm$ at rest.)

It is further obvious from the Lüders-Pauli theorem that if time reversal invariance is not violated, the operation $CP$ is conserved. This means that the left-right asymmetry that is found in a laboratory is always exactly opposite to that found in the antilaboratory.

Should it further turn out that the two-component theory of the neutrino described above is correct, one would have a natural understanding of the violation of parity conservation in processes involving the neutrino. An understanding of the $\theta-\tau$ puzzle presents now a problem on a new level because no neutrinos are involved in the decay of $K_{e2}$ and $K_{e3}$. Perhaps this means that a more fundamental theoretical question should be investigated: the origin of all weak interactions. Perhaps the strange particles belong to strange representations of the Lorentz group. Nature seems to make use of simple but odd representations. It is also interesting to note that the massless electromagnetic field is the cause of the breakdown of the conservation of isotopic spin. The similarity to the massless two-component neutrino field that introduces the non-conservation of parity may not be accidental.
Remarks on Possible Noninvariance under Time Reversal and Charge Conjugation

T. D. Lee, Columbia University, New York, New York

AND

REINHARD OEHME AND C. N. YANG, Institute for Advanced Study, Princeton, New Jersey

(Received January 7, 1957)

Interrelations between the nonconservation properties of parity, time reversal, and charge conjugation are discussed. The results are stated in two theorems. The experimental implications for the $\bar{K}-\bar{K}$ complex are discussed in the last section.

In a recent paper the question has been raised as to whether the weak interactions are invariant under a space inversion. It was also pointed out there that similar to the situation for space inversion there exists at present no experimental proof that weak interactions are invariant under charge conjugation. Consequently the absolute invariance under charge conjugation is also an open question.

The present note is devoted to a study of questions concerning the invariance under charge conjugation $C$, and under time reversal $T$ (which is defined to be the Wigner time reversal. It does not switch a particle into its antiparticle; nor does it change the sign of the spatial coordinates).

1. CPT THEOREM

For the discussion of the experimental consequences of possible nonconservation of $P$, $C$, and/or $T$, a theorem which we shall call the CPT theorem proves very important.

To understand the meaning of the theorem one recalls first that the operations $P$ and $C$ in any any many-particle system (with possibilities of creation and annihilation) are represented by unitary operators that operate on the state vectors. The operation $T$, on the other hand, is represented by the operator of complex conjugation multiplied by a unitary operator. In the Schrödinger representation the transformation of a second quantized spin 0 field described by $\psi(r)$ and $\pi(r)$ under these operations can be brought into the following form:

\[ P\psi(r)P^{-1} = \eta_{\pi}\psi(-r), \quad P\pi(r)P^{-1} = \eta_{\pi}^*\pi(-r), \]
\[ C\psi(r)C^{-1} = \eta_{\pi}\bar{\psi}(r), \quad C\pi(r)C^{-1} = \eta_{\pi}^*\bar{\pi}(r), \]
\[ T\psi(r)T^{-1} = \eta_{\pi}\psi(r), \quad T\pi(r)T^{-1} = -\eta_{\pi}^*\pi(r), \]

where $\eta_{\pi}$ means Hermitian conjugate and the phases $\eta_{\pi}$, $\eta_{\pi}^*$, and $\eta_{\pi}$ have absolute values equal to 1. For the spin $\frac{1}{2}$ field $\psi(r)$, the transformations are

\[ P\psi(r)P^{-1} = \eta_{\pi}\gamma_3\psi(-r), \]
\[ C\psi(r)C^{-1} = \eta_{\pi}\gamma_0\psi(r), \]
\[ T\psi(r)T^{-1} = \eta_{\pi}\gamma_1\gamma_2\gamma_3\psi(r), \]

where the $\gamma$ matrices are so chosen that $\gamma_1$, $\gamma_2$, and $\gamma_3$ are real and $\gamma_4$ is pure imaginary. The phases $\eta_{\pi}$, $\eta_{\pi}^*$, and $\eta_{\pi}$ have absolute values equal to unity. The transformation properties of fields of higher spin are similar.

From the CPT theorem one concludes that for any local Hermitian Hamiltonian $H$ which is invariant under proper Lorentz transformations (i.e., Lorentz transformations that involve neither space nor time inversions), there always exists a choice of the phases $\eta_{\pi}$, $\eta_{\pi}$, and $\eta_{\pi}$ for the various fields (usually in more than one way) with the following properties: (a) $H$ commutes with the product of the operators $P$, $C$, and $T$ taken in any order; and (b) if this choice of phases does not make $H$ commute with $P$, then no other choice does, and the theory is not invariant under space inversion. (Of course, if this choice of phases makes $H$ commute with $P$, then the theory is invariant under space inversion.) The same holds for $C$ and $T$.

We shall illustrate this theorem by an example where $H$ is invariant under proper Lorentz transformations. Let $\psi_p, \psi_n, \psi$, and $\psi$ be the fields describing the proton, the neutron, the electron, and the neutrino. The neutrino is assumed to be a non-Majorana particle with a nonvanishing mass. Consider

\[ H = H_{\text{free}} + \int \left( g_1(\psi_p^*\gamma_3\psi_n) (\psi^*_n\gamma_3\psi_p) + g_2(\psi_p^*\gamma_3\bar{\psi}_n) (\psi^*_n\gamma_3\bar{\psi}_p) + g_3(\psi_p^*\gamma_3\bar{\psi}_n) (\psi^*_n\gamma_3\bar{\psi}_p) + \text{Hermitian conjugate} \right) \delta^4(r), \]

\[ = \text{Hermitian conjugate} \delta^4(r), \]
where \( \gamma_a = \gamma_1 \gamma_2 \gamma_3 \gamma_4 \). This example is a special case of an example considered by Pauli.\(^4\)

Writing \( H \) as

\[
H = H(g_1 g_2 g_3 g_4),
\]

one easily proves that

\[
PHP^{-1} = H(g_1 g_2 g_3 g_4, -g_1 g_2 g_3 g_4, -g_1 g_2 g_3 g_4),
\]

\[
CHC^{-1} = H(g_1 g_2 g_3 g_4, -g_1 g_2 g_3 g_4, -g_1 g_2 g_3 g_4)
\]

\[
THT^{-1} = H(g_1 g_2 g_3)
\]

In deriving these formulas, use has been made of the fact that \( TG^{-1} = g_1 g_2 g_3 g_4 \) and \( T\gamma_i T^{-1} = \gamma_i^* \). The phases \( \eta_1, \eta_2, \eta_3, \) and \( \eta_4 \) are products of the respective phases of the four interacting fields. They are given by

\[
\eta_1 = \eta_2^* (p_1) \eta_3 (n_1) \eta_4 (e_1) \eta_4 (n_1),
\]

\[
\eta_2 = \eta_2^* (p_2) \eta_3 (n_2) \eta_4 (e_2) \eta_4 (n_2),
\]

\[
\eta_3 = \eta_2^* (p_3) \eta_3 (n_3) \eta_4 (e_3) \eta_4 (n_3),
\]

\[
\eta_4 = \eta_2^* (p_4) \eta_3 (n_4) \eta_4 (e_4) \eta_4 (n_4).
\]

Using (4), (5), and (6), one can calculate the commutation relation between \( H \) and the six operators \( TCP, TPC, \ldots, PCT \). It is found that with suitable choices of the phases \( \eta \), the Hamiltonian \( H \) commutes with all of the six, as required by the CPT theorem. In fact the conditions on the phases \( \eta \) are simply

\[
\eta_1 = \eta_2 \eta_3 = \eta_4 = \pm 1.
\]

It follows from the CPT theorem that, if one of the three operators \( P, C, \) and \( T \) is not conserved, at least one other must also be not conserved. It is of course also possible that all three are separately not conserved. In the example above, by assigning suitable values to the coupling constants \( g \), one can construct examples for all the five possibilities of conservation or nonconservation of \( P, C, \) and \( T \). These examples are displayed in Table I.

**Table I.** Examples of theories with various possible nonconservation properties.

<table>
<thead>
<tr>
<th>Value of coupling constants</th>
<th>Conserved operators</th>
<th>Nonconserved operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>( g_1 = \text{real}, g_2 = \text{real}, g_3 = g_4 = 0 )</td>
<td>( P, C, T )</td>
<td>( \cdots )</td>
</tr>
<tr>
<td>( g_1 = \text{real}, g_2 = \text{complex}, g_3 = g_4 = 0 )</td>
<td>( P, CT, TC )</td>
<td>( C, T )</td>
</tr>
<tr>
<td>( g_1 = \text{real}, g_2 = \text{imaginary}, g_3 = g_4 = 0 )</td>
<td>( C, PT, TP )</td>
<td>( P, T )</td>
</tr>
<tr>
<td>( g_1 = \text{real}, g_2 = \text{real}, g_3 = g_4 = 0 )</td>
<td>( T, CP, PC )</td>
<td>( C, P )</td>
</tr>
<tr>
<td>( g_1 = \text{real}, g_2 = \text{complex}, g_3 = g_4 = 0 )</td>
<td>( PCT, \text{ and per-} )</td>
<td>( P, C, T )</td>
</tr>
<tr>
<td>mutations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The decays of \( A \) through \( H_1 \) and through \( H_2 \) lead to states \( B_1 \) and \( B_2 \) with opposite parities. They are orthogonal to the order considered, and hence they contribute independently without interference to the decay rate of \( A \). The lifetimes of \( A \) and \( \bar{A} \) are therefore again the same.

A consequence of this theorem has already been mentioned in a previous paper: The identity of the experimental lifetimes of \( \pi^+ \) and of \( \mu^+ \) does not con-

On the other hand, the weak interactions may violate the invariance of \( C, P, \) and \( T \). One can prove the following theorem.

**Theorem 1.**—If a particle \( A \) decays through the interaction \( H_{\text{weak}} \), and if the particle and its antiparticle \( \bar{A} \) do not decay into the same final products (as e.g. when \( A \) is charged), then to the lowest order of \( H_{\text{weak}} \) the lifetimes of \( A \) and \( \bar{A} \) are the same, even if \( H_{\text{weak}} \) is not invariant under charge conjugation.

**Proof.**—Consider the case that particle \( A \) has spin zero. [The proof for the general case follows along the same lines.] Then the final states \( B \) and \( \bar{B} \) in the decays

\[
A \to B, \quad \bar{A} \to \bar{B}
\]

also have spin zero. Using the identity

\[
\langle \psi_1 | \psi_2 \rangle = \langle TP | TP \rangle,
\]

one obtains

\[
\langle B | H_{\text{weak}} | A \rangle = \pm \langle TB | TH_{\text{weak}} T^{-1} | TA \rangle = \pm \langle TB | C^{-1} P^{-1} H_{\text{weak}} PC | TA \rangle,
\]

by the CPT theorem. Consider first the case that \( H_{\text{weak}} \) commutes (or anticommutes) with \( P \). Then

\[
\langle B | H_{\text{weak}} | A \rangle = \pm \langle TB | C^{-1} H_{\text{weak}} C | TA \rangle.
\]

For a spinless system,

\[
|TA\rangle = |A\rangle, \quad |TB\rangle = |B\rangle.
\]

Hence

\[
\langle B | H_{\text{weak}} | A \rangle = \pm (CB | H_{\text{weak}} | CA) = \pm (\bar{B} | H_{\text{weak}} | \bar{A}).
\]

This shows that the lifetimes of \( A \) and \( \bar{A} \) are the same. If \( H_{\text{weak}} \) does not commute with \( P \), we write

\[
H_{\text{weak}} = H_1 + H_2,
\]

where

\[
H_1 = \frac{1}{2}[H_{\text{weak}} + PH_{\text{weak}} P^{-1}],
\]

\[
H_2 = \frac{1}{2}[H_{\text{weak}} - PH_{\text{weak}} P^{-1}].
\]

Then

\[
PH_1 P^{-1} = H_1,
\]

\[
PH_2 P^{-1} = -H_2.
\]
ststitue a proof that charge conjugation invariance holds for the weak interactions.

For a discussion of a case where $A$ and $\bar{A}$ may decay into the same final channels, see Sec. 4.

3. DEPENDENCE OF INTERFERENCE EFFECTS ON CONSERVATION OF C AND T

One would like to ask what are the experimentally detectable manifestations of a weak nonconservation of $P$, $C$, or $T$? For the nonconservation of parity, the answer is clearly to be sought in experiments to differentiate the right-handed screw from the left-handed. Some such experiments have been discussed before.\(^1\)

If parity is indeed not strictly conserved, some of these experiments could also reveal whether $C$ and/or $T$ are or are not conserved. To illustrate this let us consider the experiment of the angular distribution of $\beta$ decay from oriented nuclei. The degree of asymmetry was given in the appendix of reference 1 as proportional to

$$\frac{Z^2}{\hbar c^2} \left[ C_\tau C_\tau^* - C_A C_A^* + \frac{1}{2} (C_A C_\tau^* + C_\tau C_A^*) \right].$$  \(16\)

Applying the arguments of Sec. 1 we recognize that the two terms in (16) are present or absent depending on whether $C$ or $T$ are not conserved. To be more specific: The first term vanishes if $C$ is strictly conserved, the second term vanishes if $T$ is strictly conserved. If this experiment shows any asymmetry, the $P$ dependence and the $Z$ dependence of the asymmetry could therefore reveal whether $C$ and/or $T$ are not conserved. (The existence of any asymmetry rules out the possibility that both $C$ and $T$ are conserved, a conclusion we already drew on general grounds in Sec. 1.)

We notice that if $C$ is strictly conserved, the asymmetry discussed above vanishes in the absence of the Coulomb distortion of the electron wave function. In fact, when $C$ is conserved the asymmetry is directly dependent on the existence of a difference of the Coulomb phase shifts for opposite parities. It turns out that this is a consequence of a general theorem which we state and prove below:

Theorem 2.---If, in addition to the assumptions stated in Sec. 2 concerning $H_{\text{strong}}$ and $H_{\text{weak}}$, we assume that $H$ is strictly under charge conjugation, (i.e., \([H,C]=0\)) and if the decay products in the final state $B$ are free particles, then to the lowest order of $H_{\text{weak}}$ there is no interference between the parity-conserving and the parity-nonconserving parts of $H$ in the decay of $A$, provided the interference is sought for in experiments measuring a term of $\sigma \cdot p$.

Proof.---We again illustrate the proof by considering the case that $A$ is spinless. The general proof follows along the same lines. We perform the decomposition of $H_{\text{weak}}$ as in Eqs. (12)–(15). The final state $B$ consists of two states $B_1$ and $B_2$ of opposite parities reached from $A$ through $H_1$ and $H_2$, respectively. Clearly $B_2$ commutes with $C$, and also, by the $CPT$ theorem, commutes with $CPT$. Hence using identity (9) and Eq. (10) one obtains

$$\langle B_1 | H_1 | A \rangle^* = \langle TB_1 | TH_1 T^{-1} | TA \rangle = \langle B_1 | TH_1 T^{-1} | A \rangle = \langle B_1 | P^{-1} C^{-1} H_1 C P | A \rangle = \langle B_1 | P^{-1} H_1 P | A \rangle = \langle B_1 | H_1 | A \rangle.$$  

Thus $\langle B_1 | H_1 | A \rangle^*$ is real. Similarly one easily proves that $\langle B_2 | H_1 | A \rangle$ is pure imaginary.

In the above the states $B_1$ and $B_2$ are taken as stationary states of $H_{\text{strong}}$ consisting of standing waves. [Otherwise Eq. (10) does not hold.] Transition amplitudes into them have a relative phase factor which is, according to the above, pure imaginary. The observed final states are equal to these amplitudes multiplied by the outgoing part of the stationary states $B_1$ and $B_2$. Such outgoing parts always have real relative amplitudes if the stationary states $B_1$ and $B_2$ represent free particles. The theorem now follows immediately.

Using this theorem, one concludes that if any left-right asymmetry of the form $\sigma \cdot p$ is found, the part of this asymmetry that is independent of the distortion of the final-state wave functions can arise only if charge conjugation symmetry breaks down for the weak interactions. In particular, in decays where there is no strong final-state interactions, as, e.g., in $\pi \rightarrow \mu^+ + \nu$ and $\mu \rightarrow e^- + \nu + \nu$ decays, the detection\(^1\) of parity nonconservation through the observation of $\sigma \cdot p$ becomes impossible if $C$ is strictly conserved.

4. $K^0$, $\bar{K}^0$ DECAY MODES

The existence of the particle $K^0$ and some properties of its decay were predicted\(^4\) and discussed under the assumption that charge conjugation is strictly conserved. We wish to discuss in this section the decay of $K^0$ and $\bar{K}^0$ under the assumption that $C$, $P$, and $T$ are conserved for the strong interactions, but are not necessarily conserved in the weak decay interactions.

In the first place, the conservation of strangeness with respect to the strong interactions still requires that two particles $K^0$ and $\bar{K}^0$ with opposite strangeness exist. To understand their decay processes it is interesting to consider the charge conjugation symmetrical and anti-symmetrical combinations introduced in reference 5 (compare, however, footnote 11):

$$K_1 = \frac{1}{\sqrt{2}} (K^0 + \bar{K}^0), \quad K_2 = \frac{1}{\sqrt{2}} (K^0 - \bar{K}^0).$$  \(17\)

Unlike the situation in reference 5, if $C$ is now not invariant in the decay process, $K_1$ and $K_2$ can decay into the same final states:

$$K_1 \rightarrow \pi^+ + \pi^-, \quad K_1 \rightarrow \pi^+ + \bar{\pi}^-, \quad K_1 \rightarrow e^+ + \pi^-, \quad K_1 \rightarrow e^+ + \bar{\pi}^-,$$

$$K_2 \rightarrow \pi^+ + \pi^- + \pi^+, \quad K_2 \rightarrow e^+ + \pi^++\pi^-, \quad K_2 \rightarrow e^+ + \pi^- + \bar{\pi}^+.$$  \(8\)

\(^1\) M. Gell-Mann and A. Pais, Phys. Rev. 97, 1387 (1955).
Interference effects would therefore set in in these decay processes. The questions that one would like to ask are then: what are the lifetimes of the particles $K^0$ and $\bar{K}^0$? What are the branching ratios into various decay modes?

These questions can be answered by using a Weisskopf-Wigner type of treatment of the time-dependent Schrödinger equation. We write the time-dependent amplitudes of the particles $K^0$ and $\bar{K}^0$ as $a(t)$ and $b(t)$. The various channels of decay are denoted by $j$. $F_j(\omega)e^{-i\omega t}$ represents the amplitude of the decay product in the channel $j$ with the energy $\omega$. [We choose units such that $\hbar = 1$.] The zero of energy is taken to be the rest energy of $K$. The Schrödinger equations are then

$$\frac{da}{dt} = \sum F_{aj}(\omega)a_j(\omega)e^{-i\omega t},$$

$$\frac{db}{dt} = \sum F_{bj}(\omega)b_j(\omega)e^{-i\omega t},$$

$$\frac{dF_j(\omega)}{dt} = e^{i\omega t}[H_{ja}(\omega)a_j + H_{jb}(\omega)b_j],$$

where $H_{ja} = H_{ja}^*$ are the matrix elements. The Weisskopf-Wigner treatment consists of first assuming an exponential time dependence for $a$ and $b$, and then in sums over $\omega$ neglecting the variation of the matrix elements with $\omega$ in the interval $|\omega| \lesssim \text{uncertainty of energy of the original state}$. Using this treatment, one obtains

$$\begin{pmatrix} a \\ b \end{pmatrix} = \psi e^{-i\lambda t}.$$  \hspace{1cm} (21)

The amplitude $\psi$ and the decay constant $\lambda$ are given by the eigenvalue equation

$$\Gamma \psi = \lambda \psi.$$  \hspace{1cm} (22)

$\Gamma$ is a $2 \times 2$ Hermitian matrix with matrix elements given by

$$\Gamma_{11} = \Gamma_{32} = \sum I \Gamma_{ij} = \sum I \Gamma_{hi},$$

$$\Gamma_{12} = \sum I (\Gamma_{ij} \Gamma_{ji}^{*}) e^{i\theta_i} = \Gamma_{21}^{*},$$  \hspace{1cm} (23)

where

$$\Gamma_{ij} = 2\pi |F_{ij}|^2 \text{ (density of states per unit } d\omega)|_{\omega = \omega_j},$$

$$\Gamma_{ii} = 2\pi |H_{ii}|^2 \text{ (density of states per unit } d\omega)|_{\omega = \omega_i},$$  \hspace{1cm} (24)

and

$$e^{i\theta_i} = \text{(phase of } H_{ji} H_{ji}^{-1})|_{\omega = \omega_i}.$$  \hspace{1cm} (25)

In the foregoing derivation, use has been made of Eq. (11) which leads to $H_{ij} = \pm H_{ji}^{*}$, where $j'$ is the charge conjugate channel of $j$ (which may or may not be the same as $j$). It is important to notice that this equation is a consequence of the CPT theorem.

The two eigenvalues $\lambda_+,$ $\lambda_-$ of (22) correspond to the two decay lifetimes. The general solution is a linear superposition of two solutions $\psi_{\pm}$ of the form (21), each of which is characterized by a pure exponential decay. Since the $2 \times 2$ matrix $\Gamma$ is Hermitian, the two solutions represent linear orthogonal combinations of the states $K$ and $\bar{K}$.

In writing down Eqs. (18) and (19) we did not include a slight difference of mass in the form of a mass operation $M$ for the states of the $K$ particle. This restriction can be easily removed by adding to the right-hand sides of (18) and (19) the terms $\frac{1}{2}(M_{11}a + M_{12}b)$ and $\frac{1}{2}(M_{21}a + M_{22}b)$, respectively. The mathematical treatment is very similar to the above simple case except that we have now

$$-d\psi/dt = (\Gamma + iM)\psi,$$  \hspace{1cm} (26)

where $\Gamma$ is the same Hermitian matrix given by Eq. (23). $(iM)$ is an anti-Hermitian matrix representing the effects of the mass shifts. By using Eq. (11) one can show that, similar to Eq. (23), $M$ is a Hermitian matrix with

$$M_{11} = M_{22}.$$  \hspace{1cm} (27)

Equation (26) can now be readily solved. Its eigenstates, defined by

$$(\Gamma + iM)\psi_{\pm} = \lambda_{\pm}\psi_{\pm},$$

are

$$\psi_{\pm} = \begin{pmatrix} \rho \\ \pm q \end{pmatrix} (|\rho|^2 + |q|^2)^{-1},$$  \hspace{1cm} (28)

with the corresponding time constants

$$\lambda_{\pm} = \Gamma_{11} + iM_{11} \pm (\rho q);$$  \hspace{1cm} (29)

where $\rho$ and $q$ are two complex numbers given by

$$\rho = \Gamma_{12} + iM_{12}, \hspace{0.5cm} q = \Gamma_{12} + iM_{11} = \Gamma_{22} + iM_{12}.$$  \hspace{1cm} (30)

If at $t = 0$ a $K$ particle is produced, then at a later time the state function $\psi$ can be expressed in terms of these two eigenstates $\psi_{\pm}$ as

$$\psi(t) = \begin{pmatrix} 1 \\ 2\rho \end{pmatrix} (|\rho|^2 + |q|^2) [\psi_+ e^{-i\gamma_+ t} + \psi_- e^{-i\gamma_- t}].$$  \hspace{1cm} (31)

It is convenient to separate the real and imaginary parts of $\lambda_{\pm}$. Without loss of generality, we may write

$$\lambda_+ = \gamma_+, \hspace{0.5cm} \lambda_- = \gamma_- + 2i\Delta,$$  \hspace{1cm} (32)
where \( \gamma_+ \) and \( \gamma_- \) are two real numbers representing the reciprocal lifetimes of the short-lived ones and the long-lived ones respectively and \( \Delta \) is the mass difference between these two eigenstates. One notices that these two eigenstates \( \psi_+ \) and \( \psi_- \) do not in general represent the states \( K_1 \) and \( \bar{K}_2 \) introduced in (17). In fact they even may not be orthogonal to each other (see footnote 11).

A general discussion of the decay processes is rather involved. We shall make only the following remarks:

(A) The fractional number of \( K \) mesons that decay at time \( t \) after its production is given by
\[
N(t)dt = -d[\psi^\dagger \psi].
\] (33)

Using (26), one easily shows that
\[
\frac{d}{dt}[\psi^\dagger \psi] = \psi^\dagger \Gamma \psi.
\]

By using (28)-(31), Eq. (33) becomes
\[
N(t) = \frac{1}{2}(1+\alpha)^{-1}[\gamma_+ e^{-\gamma_+ t} + \gamma_- e^{-\gamma_- t} + \alpha e^{-\Delta t}] \cos \Delta t - 2 \Delta \sin \Delta t \]
\[
\times \left[ (\gamma_+ + \gamma_-) \cos \Delta t - 2 \Delta \sin \Delta t \right],
\] (34)

where
\[
\alpha = \psi_+ \psi_- = |p|^2 - |q|^2 \frac{|p|^2 + |q|^2}{|p|^2 - |q|^2}^{-1}
\]
(35)
is a real number representing the nonorthogonality of these two eigenstates. The four real numbers \( \gamma_+, \gamma_-, \Delta, \) and \( \alpha \) characterize the decay of the \( K \) particle. They satisfy the inequalities
\[
\gamma_+ \geq 0, \quad \alpha \leq \frac{4 \gamma_+ \gamma_-}{(\gamma_+ + \gamma_-)^2 + 4 \Delta^2},
\] (36)

which follow from the fact that \( \Gamma \) is a positive Hermitian matrix. These conditions also insure that \( N(t) \geq 0 \) for all \( t \).

Experimentally \( N(t) \) is measurable. From \( N(t) \) one can in principle determine all four constants \( \gamma_+, \gamma_-, \Delta, \) and \( \alpha \). Indications from presently existing experiments show that probably \( \gamma_+ / \gamma_- > 100 \). Equation (36) then shows that \( \alpha < 4 \Delta / \gamma_- < 0.04 \).

(B) The above discussion also leads easily to a determination of the branching ratio of the long-lived component (and the short-lived component) into the various decay modes. If charge conjugation is conserved, the long-lived component is an eigenstate of charge conjugation. Consequently its decay into charge conjugate channels such as \( \pi^+ e^- \nu \) and \( \pi^- e^+ \bar{\nu} \) must be equally probable, as is well known. If charge conjugation is not strictly conserved, decays into \( \pi^+ e^- \nu \) and \( \pi^- e^+ \bar{\nu} \) may have different probabilities for the long-lived component.

A more complete discussion of the charge asymmetry of the decay of the long-lived \( K^0 \) will be given in the appendix. We mention here only that Lederman\(^{18}\) has kindly informed us that experimental work in this direction is in progress. It is important to notice that if the experiments should yield a large asymmetry, and a small \( \alpha \) (as mentioned above), Eq. (A7) would impose very strict conditions on the relative magnitudes of the amplitudes \( f_1, g_1, f_2, \) and \( g_2 \). (To see this roughly we need only examine the limiting case \( \alpha = 0 \) discussed below.)

(C) If \( \alpha = 0 \), the two eigenstates are orthogonal.\(^{11}\) Also \( |\psi| = |q| \). [See (35).] This is the case if the mass matrix is negligible. In this case \( \psi_\pm \) are both \( 1:1 \) superpositions of the particle \( K^0 \) and \( \bar{K}^0 \). The fraction of particles decaying in \( dt \), namely \( N(dt) \), becomes the sum of two pure exponentials by (34). Furthermore (A7) shows that the decays of the long-lived component into charge conjugate channels such as \( \pi^+ e^- + \nu \) and \( \pi^- e^+ + \bar{\nu} \) are equally probable.

One of us (Reinhard Oehme) would like to express his gratitude to Professor Robert Oppenheimer for his kind hospitality at the Institute for Advanced Study.

**APPENDIX**

In this appendix we shall show the interrelationship between the parameters \( p, q \) and the branching ratio for the decay of, say, the long-lived component of \( K \) particle into various charge conjugate states.

Consider first the following decay channel of the \( K \) particle
\[
K \rightarrow e^- + \pi^+ + \nu.
\] (A1)

The final product may be in states with either parity \( = +1 \) or parity \( = -1 \). Let us denote the matrix elements for the decay process into these two types of states by \( f_1 \) and \( f_2 \). Similarly, we denote the matrix elements for
\[
K \rightarrow e^+ + \pi^- + \bar{\nu},
\] (A2)

with the final state having parity \( = +1 \) and parity \( = -1 \), by \( g_1 \) and \( g_2 \).

By using the CPT theorem and Eq. (11), the corresponding matrix elements for the decay of \( \bar{K} \),
\[
\bar{K} \rightarrow e^+ + \pi^- + \bar{\nu},
\] (A3)

are related to that of (A1). These elements are \( f_1^* \) and \( -f_2^* \). Similarly the matrix elements for
\[
\bar{K} \rightarrow e^- + \pi^+ + \nu
\] (A4)

are \( g_1^* \) and \( -g_2^* \). Let \( \psi_\pm \) represent the long-lived component \( K \) of the \( K \) particle. The matrix elements for

\(^{18}\) L. Lederman (private communication).

\(^{11}\) We recall that since the strangeness \( S \) is conserved in the strong interaction, the phase \( \psi_\pm \) of a \( K \) particle \((S = +1)\) under charge conjugation is not fixed by the strong interactions. If the weak interaction is not invariant under charge conjugation, the phase \( \psi_\pm \) is defined only up to a factor \( e^{i \theta} \). If \( \psi_\pm \) is orthogonal to \( \psi_- \), there exists however a most convenient choice which makes \( \psi_+ \) identical with the \( K_1, K_2 \) defined in (17).

\(^{8}\) K. Lande et al., Phys. Rev. 103, 1901 (1956).
the decay of $K_+$,

$$K_+ \rightarrow e^- + \pi^+ + \nu,$$  \hspace{1cm} (A5)

into the two different final parity states are proportional to $g_1 + q_1^*$ and $g_2 - q_2^*$, respectively, while the corresponding elements for

$$K_+ \rightarrow e^+ + \pi^- + \bar{\nu}$$  \hspace{1cm} (A6)

are proportional to $g_1 + q_1^*$ and $g_2 - q_2^*$. The branching ratio $r$ for the decay of $K_+$ into $e^- + \pi^+ + \nu$ and $e^+ + \pi^- + \bar{\nu}$ is, therefore,

$$r = \frac{|g_1 + q_1^*|^2 + |g_2 - q_2^*|^2}{|g_1 + q_1|^2 + |g_2 - q_2|^2}. \hspace{1cm} (A7)$$
In addition to the strong and electromagnetic interactions, there exists a large class of interactions characterized by coupling constants of order $10^{-13}$. This class of weak interactions can be divided into two distinct groups. The first group contains processes not involving neutrinos, while the second group consists of processes in which neutrinos participate. Particles involved in the first group also participate in the strong interactions, and therefore each of these particles has a well-defined value of isotopic spin $I$ and its $\pi$ component $I_{\pi}$. The weak interactions in the first group are further characterized by the selection rule $\Delta I_{\pi} = \pm \frac{1}{2}$, which of course is identical with non-conservation of strangeness. The interactions in the second group involve leptons that do not participate in any strong interactions and, in contrast with the particles of the first group, no useful assignment of isotopic spin to these leptons has been found. In spite of this difference, all weak interactions seem to have striking features in common, such as the similarity in the strengths of the coupling constants and the recently observed violation of conservation of parity and charge conjugation. While non-conservation of parity and charge conjugation have been proved only for reactions in the second group, there are strong suspicions that this is true also for the reactions in the first group - I allude, for instance, to the $\tau - \theta$ puzzle.

I shall first review briefly the present experimental status of the violation of conservation laws, together with the immediate theoretical implications. Next I shall develop some further theoretical considerations and speculations, and the particular points of view I shall adopt are fully shared by Prof. Yang.

I. The Experimental Situation

The first experiment demonstrating the non-conservation of parity and charge conjugation was performed by the Columbia and NBS groups, and involved the beta decay of polarized $^{60}\text{Co}$ nuclei, according to the reaction

$$^{60}\text{Co} \ (\text{polarized}) \longrightarrow ^{60}\text{Ni} + e^- + \nu.$$
The alignment of the $^{60}\text{Co}$ nuclei at low temperature was achieved by a magnetic field, and the electrons were found to be emitted preferentially in a direction antiparallel to the nuclear spin. From this asymmetry one can immediately deduce that parity is not conserved in the process, without invoking any detailed theory of beta decay. In fact, the direction of the magnetizing current loop together with the preferential direction of the emitted electrons can be used directly to establish the difference between a right-handed coordinate system and a left-handed coordinate system. Furthermore, from the observed magnitude of the asymmetry one can conclude that charge conjugation as well as parity is not conserved in this reaction. The transition is of pure Gamow-Teller type, so that only tensor coupling is involved. If $C_T$ is the coupling constant for the part of the interaction which commutes with parity, and $C'_T$ is the coupling constant for the part which anticommutes with parity, then the relationship

$$C'_T \approx - C_T$$

would account for the observed asymmetry.

The second confirmation of non-conservation of parity and charge conjugation came from the measurement of the angular distribution of electrons from $\pi^- \rightarrow \mu^- \rightarrow e$ decay, the experiments were done at Columbia and Chicago. The pion decays into a muon, which then decays into an electron:

$$\pi^\pm \rightarrow \mu^\pm + \nu \text{ (or } \bar{\nu})$$

$$\mu^\pm \rightarrow e^\pm + \nu + \bar{\nu}$$

The angular distribution of the electrons with respect to the muon momentum was found to be asymmetrical, the electrons being preferentially emitted in a direction antiparallel to the muon momentum. The angular distribution is, apart from the depolarization effect, the same for both signs of the electric charge.

Following the initial experiments, a large amount of work was
done on the longitudinal polarization of the emitted electrons, on the circular polarization of the associated gamma rays together with beta-gamma correlations, on the beta angular distributions from other nuclei such as $^{60}Co$ and $^{63}Cu$, and on $\pi^-\mu^-\nu$ decay with the muons stopped in many kinds of materials. All results seem to confirm rather conclusively the essential findings of the earlier experiments.

II. The Two-Component Theory of the Neutrino

The quantitative data on the violation of the conservation laws of parity and charge conjugation can be explained in a simple and appealing way by use of the two-component theory of the neutrino. This theory and its applications are by now well known, and I shall make only a few remarks about it.

(a) The two-component theory of the neutrino can be expressed, as a matter of convenience, in terms of the usual four-component Dirac theory of the neutrino, by imposing a subsidiary condition on the neutrino field $\nu$:

$$\gamma_5 \nu = - \nu$$

This ensures that the mass of such particles will be zero.

(b) In the two-component theory, the neutrino (which is the particle of the theory) will always have its spin parallel to its momentum. If we use this structure to define a right-handed coordinate system, then the neutrino always has right-handed spirality and the antineutrino always has left-handed spirality:

![Neutrino: right-handed spirality (ν)](image)

![Anti-neutrino: left-handed spirality (ν̄)](image)

For a given momentum, there exist only these two states.

(c) Let us describe the decay of $^{60}Co$ in terms of these definitions. Since the electrons are seen to be emitted preferentially in a direction antiparallel to the spin of the $^{60}Co$ nucleus, we conclude that $C' = -C$, where $C$ and $C'$ stand for the coupling
constants of the parity conserving and non-conserving parts of the interaction in the conventional four-component theory. If we interpret this result in terms of the two-component theory we are led to the conclusion that the emitted neutrino particle has left-handed spirality, i.e. it actually is an anti-neutrino:

$$C_0^{60} \longrightarrow N_i^{60} + e^- + \bar{\nu}.$$ 

The neutron decay process therefore is

$$n \longrightarrow p^- + e^- + \bar{\nu}.$$

Similarly, when we examine $\pi^- - \mu^- - e$ decay in the light of the two-component theory, we find that both a neutrino and an anti-neutrino must be emitted in the $\mu^-$ decay, not two neutrinos or two anti-neutrinos:

$$\mu^\pm \longrightarrow e^\pm + \nu + \bar{\nu}.$$ 

This can be shown by computing the Michel parameter $\rho$ from the electron spectrum in the two cases. In the former case we get $\rho = 0.75$, while in the latter case we get $\rho = 0$ which is certainly ruled out by the experimental evidence. The best experimental value to date is $\rho = 0.68$. While the value of 0.75 is somewhat higher than the experimental one, the theory nevertheless predicts an electron angular distribution and an energy dependence of that distribution which are in reasonable quantitative agreement with experiment.

(d) Another very interesting conclusion to be drawn from the two-component theory concerns the expected longitudinal polarization of the electrons. The electron-neutrino coupling term (not involving derivatives) has the general form $\psi_e^+ \sigma_i \psi_\nu$, where $\sigma_i$ may be the scalar, tensor, vector, etc. operator. Using the subsidiary condition we may write

$$\psi_e^+ \sigma_i \psi_\nu = \frac{1}{2} \psi_e^+ \sigma_i (1 - \gamma_5) \psi_\nu$$

$$= \frac{1}{2} \psi_e^+ (1 \pm \gamma_5) \sigma_i \psi_\nu \begin{cases} + & \text{for } S, T, P; \\ - & \text{for } V, A \end{cases}$$

One can easily show that this interaction leads to a longitudinal
polarization of the electrons given by

\[ \langle \hat{\sigma} \cdot \hat{p} \rangle_{e^-} = \begin{cases} -\frac{\nu_c}{c} & \text{for } S, T, P \\ +\frac{\nu_c}{c} & \text{for } V, A \end{cases} \]

Thus if the interaction is scalar, tensor, or pseudoscalar, the emitted electron will have left-handed spirality; if the interaction is vector or axial vector, then it will have right-handed spirality. Now we know that beta decay occurs predominantly through scalar and tensor interactions, and consequently we expect the electron to have left-handed spirality. For positrons, of course, the opposite is true. These predictions have been verified experimentally for Gamow-Teller transitions.

(e) Even though the two-component theory describes the experimental phenomena very well, we may ask whether it could also account for the other conceivable reaction

\[ n \rightarrow p^- + e^- + \nu \]

where the spirality is opposite to the observed one. Indeed, the mere use of two-component theories does not exclude such possibilities. However, if both reactions exist at the same time we would have a non-zero rate for double beta decay, and the observed asymmetries would not attain their maximum possible values. In fact, as we shall hear from Prof. Case tomorrow, by using a less restricted Majorana form of the two-component theory we can extend it to get a non-zero neutrino rest mass. Yet experimentally we find that the mass is small, that the rate of the double beta process is also small, and that the asymmetry in the angular distribution is large. We therefore are led to ask whether an underlying principle is behind all this. Indeed, the evidence points very strongly to the law of conservation of leptons. This law and the two-component theory together imply three things:

1. The mass of the neutrino is zero.
2. The rate of the double beta process is zero.
3. Conservation of parity is violated, and the resulting asymmetries can attain their maximum values.

These statements seem to be in good agreement with experiment, and therefore we next turn our attention to the lepton conservation law.
III. The Law of Conservation of Leptons

This law states that whenever a light fermion is emitted, another one is annihilated or an anti-fermion is emitted. We thus may assign a leptonic number to light fermions and anti-fermions, the signs being opposite in the two cases, and are led to the conservation of leptonic number.

By inspecting the beta decay process we find that the leptonic number for the neutrino is the same as that for the electron, and the decay of the muon tells us that $\mu^-$ and electron have the same leptonic number. The lepton conservation leads to several immediate theoretical predictions.

We consider the decay of the negative pion,

$$\pi^- \rightarrow \mu^- + \bar{\nu},$$

and draw a diagram showing the spiralities of the emitted particles:

Here a $\mu^-$ is created, and hence an antineutrino must also come out. Since the latter has left-handed spirality, and since the pion has spin zero, the $\mu^-$ emitted must have left-handed spirality to conserve total angular momentum. Next we draw a similar diagram for the subsequent decay of the muon, $\mu^- \rightarrow e^- + \nu + \bar{\nu}$:

Since the electrons are emitted predominantly in the backward direction with respect to the muon momentum, and since we are merely interested in getting correct signs, we have drawn for simplicity only the high energy electron coming out in the backward direction, with the neutrino pair proceeding in the forward direction. The spiralities in the neutrino pair cancel, and conservation of angular momentum then tells us that the electron has the same spin component as the muon. Thus the $e^-$ emitted in $\mu^-$ decay has right-handed spirality. Both of these predictions may of course be confirmed experimentally.
Let us assume that the above is correct, and look for further implications. The decay electron has right-handed spirality, and if in the interaction the electron and neutrino fields are coupled together, then by the arguments of Section II(d) the coupling must be vector or axial vector. In this case the vector and axial vector interactions are exactly the same, since the extra $\mathcal{V}$'s operating on the $\psi^v$ in each of the $(\xi, \psi^v)$ and $(\nu, \psi^v)$ factors simply produce compensating changes in sign. Thus the coupling contains vector, but not scalar or pseudoscalar or tensor parts. However, an antisymmetrical tensor of second rank is equivalent to a six-vector, and by virtue of the peculiar properties of the two-component neutrino three of its space components precisely cancel three of its time components. Consequently we may either say that there exists no tensor interaction, or that there is a tensor interaction with a non-vanishing coupling constant but what is multiplied by that coupling constant happens to be zero.

IV. The Universal Fermi Interaction

We next turn to the Universal Fermi Interaction, which is an attempt to gain a more unified understanding of certain of the weak interactions. We draw the famous triangle representing the interactions of interest:

![Triangle Diagram](image)

Beta decay information tells us that the interaction between $(\xi, \nu)$ and $(\xi, \psi^v)$ is scalar and tensor, while the two-component neutrino theory plus the law of conservation of leptons implies that the coupling between $(\xi, \psi^v)$ and $(\nu, \psi^v)$ is vector. This means that the Universal Fermi Interaction cannot be realized in the way we have expressed it. If all these coupling types turn out to be experimentally correct, we prefer to think that the similarity in coupling constants cannot be accounted for in terms of such a limited scheme. Rather it is a universal feature of all weak interactions, and not just those involving leptons. Nevertheless, at this moment it is very desirable to recheck even the old beta interactions to see whether the coupling is really scalar, a point to which we shall return later.
We next consider the third leg of the triangle, the muon capture process:

\[ \mu^- + p \rightarrow n + \nu \]

This process used to be hard to analyze because we only observed its rate, but the fact that the emitted neutrino carries away angular momentum can be used to gain more detailed information. The law of conservation of leptons tells us that a neutrino is emitted, not an anti-neutrino. For convenience let us assume that the captured muon is 100% polarized. Then the neutron angular distribution will not be isotropic, but will take the form \( 1 - \cos \theta \) for scalar and vector coupling, and \( 1 + \frac{1}{2} \cos \theta \) for tensor and axial vector coupling, where \( \theta \) is the angle between the spin of the muon and the momentum of the neutron. Hence, by measuring this angular distribution we can get quite definite information on the coupling types for the third leg, and this is particularly important in view of the apparent coupling discrepancy in the other two legs. Further information on the nature of this process can be obtained by observing the bremsstrahlung emitted during the reaction

\[ \mu^- + p \rightarrow n + \nu + \gamma \]

measuring the angular distribution together with the circular polarization of the gamma rays. The conclusions to be drawn from these considerations depend on further experiments. The present data seem fairly well accounted for by the two-component neutrino theory, and we may say that in a modest way we have some understanding of these phenomena. The same cannot be said, however, with regard to the decay of K mesons and hyperons, the topic to which we turn next.

V. The Decay of K-Mesons and Hyperons

(a) We first consider the \( \tau^- - \theta \) puzzle. Once we have observed the non-conservation of parity and charge conjugation in some weak interactions, we are free to conjecture that these violations extend to the entire class. Therefore the \( \tau^- - \theta \) duality can be explained in a simple way, perhaps even the correct way, by just saying that the \( \theta \) particle is identical to the \( \tau \) particle. In the decay of K-mesons into two or three pions we measure at most two independent momenta, going to the center of mass system in the three pion case. If the K-meson has no spin, then the parity conserving and non-conserving modes of decay can never interfere because we cannot
form a pseudoscalar out of two momentum vectors. Consequently the maximum possible evidence that parity is not conserved in K decay is precisely what we have seen; a spinless particle decaying into two or three pions, the masses and lifetimes for the two cases being the same.

Nevertheless, several attempts have been made to base a theory on the idea that only reactions involving neutrinos violate parity conservation, and that reactions not involving neutrinos do not violate parity conservation. Under this assumption the Dalitz-Fabry plot tells us that the Θ particle is not the same as the τ particle. We would then expect to have two particles with different lifetimes, contradictory to what seems to be observed. Thus there is definite experimental evidence against the idea that the K-meson exists in a parity doublet, and in fact there is no theoretical necessity for such a proposal. Assuming spinless particles, we then have in the case of τ and Θ decay the maximum possible information on non-conservation of parity.

(b) We may hope to gain additional information from hyperon decay, since hyperons, being fermions, certainly have spin. Let us consider the production of Λ°:

\[ \pi^- + p \rightarrow \Lambda^0 + K^0. \]

This is a strong interaction which we may regard as a polarizer of the Λ°. We then use the subsequent decay of the Λ°, which is a weak interaction, as an analyzer of this polarization:

\[ \Lambda^0 \rightarrow p + \pi^- . \]

In fact, for this decay to serve as an analyzer at all, it must necessarily violate the conservation law of parity. However, since the polarizer is a strong interaction and since the analyzer involves slow particles, it is possible that neither polarizer nor analyzer are as effective as in the case of the leptons in \( \pi^- \mu^- \nu \) decay, where the polarizer is known to be nearly 100% effective and the analyzer is also extremely effective. To illustrate, let us assume the spins of the \( \Lambda^0 \) and \( K^0 \) to be 1/2 and 0, respectively. If we carry out the production near threshold so that only S and P-waves are involved, and if the production angular distribution has the form \( (1 + \cos \theta)^2 \), then the polarization of the \( \Lambda^0 \) is zero for all angles. Here \( \theta \) is the angle between the incoming pion momentum and the outgoing hyperon momentum in the center of mass.
system. Observation of course does not substantiate that this is the actual angular distribution. It is, however, the most peaked distribution in the forward or backward direction possible for $S$ and $P$-waves, and may not be too far from the actual one. Here we would have a case in which the polarizer is not a very effective one. To obtain definite conclusions on the non-conservation of parity it is imperative to have good knowledge about polarizations. For that reason it will probably be very useful to investigate such production and decay processes in a bubble chamber near threshold energies.

(c) The $\theta_1^0$ and $\theta_2^0$ particles were introduced on the assumption of invariance under charge conjugation. Since this assumption is invalid for at least some of the weak interactions, we shall examine the implications of this violation on their expected behavior. From the strong production reactions we infer that there should exist two states (i.e. two particles) of $\theta^0$ type, with strangeness + 1 and - 1. The existence of two states implies in general the existence of two lifetimes, and we wish to find out how the short-lived and the long-lived particles behave. Let us first assume that charge conjugation is not conserved, but that charge conjugation times parity is conserved. We immediately conclude that the short-lived particle can decay into two pions but the long-lived one cannot. The situation remains the same as under the original assumption, and evidently we cannot see a violation of charge conjugation invariance unless we see a violation of parity. If charge conjugation times parity is not conserved, the long-lived particle can decay into two pions but the $2\pi$ to $3\pi$ branching ratio, based on phase space arguments, would be quite small and the $2\pi$ mode might escape experimental detection. The striking behavior of $\theta_1^0$ and $\theta_2^0$ thus happens to be an insensitive test body for non-conservation of C or CP.

VI. Time Reversal Invariance

The question of whether CP is or is not conserved brings us to the question of invariance under time reversal. According to the famous TCP theorem, if T is invariant then CP is invariant, and if T is not invariant then CP is not invariant. We have seen that $\theta_1^0$ and $\theta_2^0$ do not furnish a sensitive test for T invariance. Fortunately more sensitive tests exist in beta decay, and I shall mention only a few of these.

One method, involving the type of experiment already done, measures the beta angular distribution from polarized nuclei, determining $\vec{F}_T \cdot \vec{F}_e$ for a $\mathcal{J} \rightarrow \mathcal{J}$, $\mathcal{M}_0$ transition. This provides
information about the real part of \( |M_F|/|M_{GT}|(C_s C_T^* + C_T C_s^*) \). The imaginary part of this quantity can be obtained in a beta-gamma cascade by measuring \( \vec{J} \cdot (\vec{p}_e \times \vec{F}_\gamma)(\vec{J} \cdot \vec{F}_\gamma) \) or \( \vec{J} \cdot (\vec{p}_e \times \vec{F}_\gamma) \). A measurement of both the real and the imaginary parts for the same element would yield a great deal of information. Even the measurement of \( \vec{F} \cdot (\vec{p}_e \times \vec{F}_\gamma) \) alone will give much information. The real part of \( |M_F|/|M_{GT}|(C_s C_T^* + C_T C_s^*) \) has been measured by the Leyden and NBS groups for \( C_s C_T^* \), and was found to be zero. The ratio \( |M_F|/|M_{GT}| \) had been determined previously, and neither \( M_F \) nor \( M_{GT} \) was found to vanish. Since \( C_s \) seems to be non-zero, and since certainly \( C_T \) is non-zero, it follows that the quantity in question is purely imaginary. This implies violation of time reversal invariance. But this argument is based on much experimental information obtained at different times, and it seems desirable to check all these results. Of course it would be extremely useful to measure the imaginary part directly, as outlined above. The question of the existence of the scalar interaction needs careful re-examination, and to that end a measurement of the spirality of the emitted positron is suggested. The answers, bearing on the existence of a Universal Fermi Interaction and on time reversal invariance, should be forthcoming in the near future.

VII. The Mach Principle

If it turns out (not to our great surprise, perhaps) that \( T \) is not invariant, what are the implications? This question leads us to a brief discussion of the Mach Principle which states that the laws of physics should not depend upon the choice of any particular geometrical coordinate system. If we believe in this principle then the present information on asymmetries can still be accounted for in two ways, depending on whether \( T \) is invariant or not. If \( T \) is invariant then CP is invariant, and when we perform mirror reflections we must simultaneously change particles to anti-particles to preserve the overall symmetry of space and time. We
nevertheless have lost one symmetry property compared to the earlier situation. If T is not invariant then CP is not invariant, and changing particles into antiparticles does not help. But a firm believer in Mach's Principle might conjecture, for instance, that there exist two different kinds of nucleons. One kind of proton would be a right-handed one (\( p_R \)) and the other would be a left-handed one (\( p_L \)). They have opposite spirality. Our world consists almost 100% of one kind, which accounts for the observed asymmetry. But overall space symmetry can still be maintained by taking into account the \( p_L \) which do not exist in any appreciable quantity in our local cosmological region. Thus P is kept invariant, and from the TCP theorem we know that TC will be invariant, even though C is not invariant. Thus we can still have overall time reversal symmetry by changing particles to anti-particles.

The second possibility is certainly not so simple or attractive as the first.

In conclusion, let me emphasize again the curious nature of all weak interactions. In spite of their great diversity, they display striking similarities in the strengths of coupling constants, in the peculiar selection rule on \( L \), and in the non-conservation of parity and charge conjugation. Perhaps all these are merely different aspects of a single and unifying principle underlying all weak interactions.

Reprinted from
Rochester Conference 1957
Interscience Publisher's, Section VII, pp. 1-12
Possible Detection of Parity Nonconservation in Hyperon Decay

T. D. Lee, J. Steinberger,
Columbia University, New York, New York

AND

G. Feinberg, P. K. Kabir, and C. N. Yang,
Institute for Advanced Study, Princeton, New Jersey

(Received May 2, 1957)

RECENTLY, various experiments established the nonconservation of parity in $\beta$ decay, $\pi$ decay, and $\mu$ decay. The purpose of this note is to emphasize that, in view of these developments, experiments on hyperon production and decay in $(\pi^+ + p)$ collisions of the type done by various groups using bubble chambers, seem now to be especially important for a clarification of the following related questions: (i) whether parity conservation is violated in hyperon decays and (ii) whether parity doublets exist.

A detailed analysis concerning the possible detection of parity doublets exists in the literature. In the following we shall make a phenomenological study of the problem of possible detection of parity nonconservation in hyperon decay under the assumption that there exist no parity doublets for either $K$ mesons or hyperons.

To make the analysis unambiguous and to draw conclusions that are relatively definite, it is necessary that one knows something about the polarization of the hyperons produced. It seems that a good plan is to study hyperon production and decay near threshold.

Production and decay of $\Sigma^-$. For example, let us consider the production of $\Sigma^-$ from $(\pi^+ + p)$ collisions:

$$\pi^+ + p \rightarrow \Sigma^- + K^+$.

(1)

It is perhaps worthwhile to try to do the experiments at laboratory kinetic energies of the pion of, say, 955 Mev and 1 Bev, corresponding to center-of-mass total kinetic energies of the $\Sigma^- + K^+$ system of 30 Mev and 60 Mev. At these energies one hopes that only $s$ and $p$ waves are produced in the $\Sigma^- + K^+$ system.

It is then easy to see that the differential production cross section per unit solid angle $d\Omega$ (in the center-of-mass system of production) of the $\Sigma^-$ produced is given by

$$I(\theta) = |a + b \cos \theta|^2 + |c|^2 \sin \theta,

(2)

where $a$ can be chosen as real and $b$ and $c$ are complex numbers. We use the following notations:

$$p_{in} = \text{momentum of the incoming } \pi^+,

p_{z} = \text{momentum of the } \Sigma^- \text{ produced}, \quad \theta = \text{angle between } p_{in} \text{ and } p_{z}.

(3)

In (3) both $p_{in}$ and $p_{z}$ are measured in the center-of-mass system of production. The polarization of the $\Sigma^-$ produced at the angle $\theta$ is always in the direction of $p_{in} \times p_{z}$ and has the magnitude

$$P(\theta) = |I(\theta)|^{-1} 2 \sin \theta \times \text{Im}[c^*(a + b \cos \theta)],

(4)

where $P(\theta)$ is defined to be the average spin of the $\Sigma^-$ in units of $\hbar^2/2$. In Eq. (4) the assumptions have been made that the spin of $\Sigma^-$ is $\frac{1}{2}$ and that the spin of $K^+$ is 0.

If parity is not conserved in the decay of $\Sigma^-$, the polarization $P(\theta)$ can be measured by using the decay process of $\Sigma^-$,

$$\Sigma^- \rightarrow n + \pi^-,

(5)

as an analyzer. Let $R$ be the projection of the momentum of the decay pion in the direction of $p_{in} \times p_{z}$. The distribution function for $R$ at an angle $\theta$ of production is given by

$$W(\theta, \xi)d\Omega dZ = I(\theta)d\Omega \frac{1}{2} (1 + a \cos \theta) \xi d\xi,

(6)

where

$$\xi = R/(\text{maximum value of } R)|R|(100 \text{ Mev} / c).

In terms of the coefficients $a$, $b$, and $c$, defined in Eq. (2), $W(\theta, \xi)$ can be written as

$$W(\theta, \xi)d\Omega dZ = \left[|a + b \cos \theta|^2 + |c|^2 \sin \theta\right] d\Omega \frac{1}{2} d\xi + \alpha \sin \theta \text{Im}[c^*(a + b \cos \theta)] d\Omega \frac{1}{2} \xi d\xi.

(7)

The existence of a nonvanishing $\alpha$ would constitute an unambiguous proof of parity nonconservation in $\Sigma^-$ decay. In such a case the final state of $(n + \pi^-)$ in process (5) would be a mixture of $s_1$ and $g_1$ states with amplitudes, say, $A$ and $B$ respectively. The asymmetry parameter $\alpha$ is related to these amplitudes by

$$\alpha = 2 \text{ Re}(A^* B) / (|A|^2 + |B|^2).

(8)

If time reversal leaves the invariant decay process of $\Sigma^-$, then

$$\alpha = \pm\frac{2|A| \times |B|}{|A|^2 + |B|^2} \cos(\delta_p - \delta_s),

(9)

where $\delta_p$ and $\delta_s$ are, respectively, the phase shifts of $(n + \pi^-)$ scattering in the $p_1$ and $s_1$ states about 117 Mev in their center-of-mass system. If the decay interaction is invariant under charge conjugation, then

$$\alpha = \pm\frac{2|A| \times |B|}{|A|^2 + |B|^2} \sin(\delta_p - \delta_s).

(10)

The following remarks are useful concerning the measurements of $\alpha$ and $\theta(\theta)$.

1. The polarization $P(\theta)$ may sometimes be very small. E.g., if (1) gives

$$I(\theta) = (1 + \cos \theta)^2, \quad \text{or } I(\theta) = (1 - \cos \theta)^2,

(11)

then $P(\theta) = 0$ identically.
2. At production energies near the threshold, the variations of the quantities \(a, b, c\), and \(\delta\) introduced in Eq. (2), with respect to \(p_2\) are given by

\[
\begin{align*}
    a &= a_0(p_2)^4, \\
    b &= b_0(p_2)^4 \exp(i\delta), \\
    c &= c_0(p_2)^4 \exp(i\delta),
\end{align*}
\]

(12)

where \(a_0, b_0, c_0, \delta\), and \(\chi_0\) are all real constants independent of \(p_2\). Thus by selecting two or three energy values near threshold, it is possible to determine \(a_0, b_0, c_0, \) and \(\chi_0\) from the angular and energy dependence of \(I(\theta)\) alone.

If the energy dependence of the cross section should not be representable by (12), one would have an indication that resonance effects might be important in the \(\pi^- p\) system near the threshold for \(\Sigma^-\) production.

3. If \(\chi_0 \neq 0\), then by comparing the coefficients of the \(\sin \theta\) and \(\sin \cos \theta\) terms in \(W(\theta, \xi)\), the phase \(\chi_0\) can also be determined.

4. From the values of these five real constants, \(a_0, b_0, c_0, \chi_0,\) and \(\chi_0\), the asymmetry parameter \(\alpha\) can then be deduced from \(W(\theta, \xi)\) [Eq. (7)].

5. If

\[|\alpha| > |\sin(\delta - \delta)|,\]

then from Eq. (10) both invariance under charge conjugation and conservation of parity do not hold in the decay of \(\Sigma^-\).

Since the phase shifts in the \(J = \frac{1}{2}\) states are all small, the conclusion is essentially that any appreciable asymmetry with respect to the sign of \(\xi\) in \(W(\theta, \xi)\) is an indication that conservation of parity and invariance under charge conjugation do not hold in the decay of \(\Sigma^-\).

Production and decays of other hyperons. The foregoing analysis can also be applied to the productions and decays of other hyperons. We consider, for definiteness, the following processes concerning \(\Lambda^0\):

\[
\pi^- + p \rightarrow \Lambda^0 + K^0,
\]

(13)

and

\[
\Lambda^0 \rightarrow p + \pi^-.
\]

(14)

All the previous formulas for \(I(\theta), p(\theta), W(\theta, \xi)\), and \(\alpha\) [i.e., Eqs. (2), (4), (6), (7), and (8)] remain unchanged. The only difference is that in Eq. (8) the amplitudes \(A\) and \(B\) of \(s^{-}\) and \(p\)-wave final states in the decay process of \(\Lambda^0\) are now each a mixture of two isotopic spin states. These amplitudes can be written as

\[
\begin{align*}
    A &= (\frac{1}{\sqrt{2}})A_1 + (\frac{1}{\sqrt{2}})A_p, \\
    B &= (\frac{1}{\sqrt{2}})B_1 + (\frac{1}{\sqrt{2}})B_p,
\end{align*}
\]

(15)

where \(A_1, B_1\) are, respectively, the \(s^{-}\) and \(p\)-wave amplitudes for final states with the total isotopic spin value \(I = \frac{1}{2}\), and \(A_p, B_p\) the corresponding amplitudes for states with \(I = \frac{3}{2}\). In place of Eqs. (9) and (10) we have now the following conditions for invariance under time reversal and charge conjugation:

If the decay process is invariant under time reversal, then we can choose

\[
\begin{align*}
    A_1 &= |A_1| e^{i\delta_1}, \\
    A_1 &= \pm |A_1| e^{i\delta_1}, \\
    B_1 &= \pm |B_1| e^{i\delta_1}, \\
    B_1 &= |B_1| e^{i\delta_1}.
\end{align*}
\]

(16)

On the other hand, if the decay process is invariant under charge conjugation operation, then these amplitudes are

\[
\begin{align*}
    A_1 &= |A_1| e^{i\delta_1}, \\
    A_1 &= \pm |A_1| e^{i\delta_1}, \\
    B_1 &= \pm |B_1| e^{i\delta_1}, \\
    B_1 &= |B_1| e^{i\delta_1}.
\end{align*}
\]

(17)

The phase shifts \(\delta_1\) are the usual pion-nucleon scattering phase shifts at 37-Mev total kinetic energy:

\[
\begin{align*}
    \delta_1 &= \text{phase shift for } s \text{ waves, } I = \frac{1}{2}, J = \frac{1}{2}, \\
    \delta_2 &= \text{phase shift for } s \text{ waves, } I = \frac{3}{2}, J = \frac{3}{2}, \\
    \delta_3 &= \text{phase shift for } p \text{ waves, } I = \lambda/2, J = \mu/2.
\end{align*}
\]

All these phase shifts are small at 37-Mev total kinetic energy. Therefore any appreciable asymmetry in \(W(\theta, \xi)\) with respect to the sign of \(\xi\) is an indication that conservation of parity and invariance under charge conjugation do not hold in the decay of \(\Lambda^0\).

A measurement of the branching ratio in the decay processes

\[
\begin{align*}
    \Lambda^0 \rightarrow p + \pi^-, \\
    \Lambda^0 \rightarrow n + \pi^0,
\end{align*}
\]

(18)

(19)

and a measurement of the distribution function \(W(\theta, \xi)\) for process (19) would lead to additional information concerning the amplitudes \(A_1, B_1, A_p,\) and \(B_p\).

* Work supported in part by the U. S. Atomic Energy Commission.


7. In view of the recent experimental developments (references 1, 2, and 3) there appears to be at present no theoretical necessity to introduce the complication of parity doublets. (See footnote 8 in reference 5.)