Conservation of Isotopic Spin and Isotopic Gauge Invariance*

C. N. Yang † and R. L. Mills

Brookhaven National Laboratory, Upton, New York

(Received June 28, 1954)

It is pointed out that the usual principle of invariance under isotopic spin rotation is not consistant with the concept of localized fields. The possibility is explored of having invariance under local isotopic spin rotations. This leads to formulating a principle of isotopic gauge invariance and the existence of a **b** field which has the same relation to the isotopic spin that the electromagnetic field has to the electric charge. The **b** field satisfies nonlinear differential equations. The quanta of the **b** field are particles with spin unity, isotopic spin unity, and electric charge $\pm e$ or zero.

INTRODUCTION

THE conservation of isotopic spin is a much discussed concept in recent years. Historically an isotopic spin parameter was first introduced by Heisenberg[1] in 1932 to describe the two charge states (namely neutron and proton) of a nucleon. The idea that the neutron and proton correspond to two states of the same particle was suggested at that time by the fact that their masses are nearly equal, and that the light

stable even nuclei contain equal numbers of them. Then in 1937 Breit, Condon, and Present pointed out the approximate equality of $p-p$ and $n-p$ interactions in the ^1S state.[2] It seemed natural to assume that this equality holds also in the other states available to both the $n-p$ and $p-p$ systems. Under such an assumption one arrives at the concept of a total isotopic spin[3] which is conserved in nucleon-nucleon interactions. Experi-

[1] W. Heisenberg, Z. Physik **77**, 1 (1932).

[2] Breit, Condon, and Present, Phys. Rev. **50**, 825 (1936). J. Schwinger pointed out that the small difference may be attributed to magnetic interactions [Phys. Rev. **78**, 135 (1950)].

[3] The total isotopic spin **T** was first introduced by E. Wigner, Phys. Rev. **51**, 106 (1937); B. Cassen and E. U. Condon, Phys. Rev. **50**, 846 (1936).

ments in recent years[4] on the energy levels of light nuclei strongly suggest that this assumption is indeed correct, An implication of this is that all strong interactions such as the pion-nucleon interaction, must also satisfy the same conservation law. This and the knowledge that there are three charge states of the pion, and that pions can be coupled to the nucleon field *singly*, lead to the conclusion that pions have isotopic spin unity. A direct verification of this conclusion was found in the experiment of Hildebrand[5] which compares the differential cross section of the process $n+p \rightarrow \pi^0+d$ with that of the previously measured process $p+p \rightarrow \pi^++d$.

The conservation of isotopic spin is identical with the requirement of invariance of all interactions under isotopic spin rotation. This means that when electromagnetic interactions can be neglected, as we shall hereafter assume to be the case, the orientation of the isotopic spin is of no physical significance. The differentiation between a neutron and a proton is then a purely arbitrary process. As usually conceived, however, this arbitrariness is subject to the following limitation: once one chooses what to call a proton, what a neutron, at one space-time point, one is then not free to make any choices at other space-time points.

It seems that this is not consistent with the localized field concept that underlies the usual physical theories. In the present paper we wish to explore the possibility of requiring all interactions to be invariant under *independent* rotations of the isotopic spin at all space-time points, so that the relative orientation of the isotopic spin at two space-time points becomes a physically meaningless quantity (the electromagnetic field being neglected).

We wish to point out that an entirely similar situation arises with respect to the ordinary gauge invariance of a charged field which is described by a complex wave function ψ. A change of gauge[6] means a change of phase factor $\psi \rightarrow \psi'$, $\psi' = (\exp i\alpha)\psi$, a change that is devoid of any physical consequences. Since ψ may depend on x, y, z, and t, the relative phase factor of ψ at two different space-time points is therefore completely arbitrary. In other words, the arbitrariness in choosing the phase factor is local in character.

We define *isotopic gauge* as an arbitrary way of choosing the orientation of the isotopic spin axes at all space-time points, in analogy with the electromagnetic gauge which represents an arbitrary way of choosing the complex phase factor of a charged field at all space-time points. We then propose that all physical processes (not involving the electromagnetic field) be invariant under an isotopic gauge transformation, $\psi \rightarrow \psi'$, $\psi' = S^{-1}\psi$, where S represents a space-time dependent isotopic spin rotation.

To preserve invariance one notices that in electro-

dynamics it is necessary to counteract the variation of α with x, y, z, and t by introducing the electromagnetic field A_μ which changes under a gauge transformation as

$$A_\mu' = A_\mu + \frac{1}{e}\frac{\partial\alpha}{\partial x_\mu}.$$

In an entirely similar manner we introduce a B field in the case of the isotopic gauge transformation to counteract the dependence of S on x, y, z, and t. It will be seen that this natural generalization allows for very little arbitrariness. The field equations satisfied by the twelve independent components of the B field, which we shall call the **b** field, and their interaction with any field having an isotopic spin are essentially fixed, in much the same way that the free electromagnetic field and its interaction with charged fields are essentially determined by the requirement of gauge invariance.

In the following two sections we put down the mathematical formulation of the idea of isotopic gauge invariance discussed above. We then proceed to the quantization of the field equations for the **b** field. In the last section the properties of the quanta of the **b** field are discussed.

ISOTOPIC GAUGE TRANSFORMATION

Let ψ be a two-component wave function describing a field with isotopic spin $\frac{1}{2}$. Under an isotopic gauge transformation it transforms by

$$\psi = S\psi', \tag{1}$$

where S is a 2×2 unitary matrix with determinant unity. In accordance with the discussion in the previous section, we require, in analogy with the electromagnetic case, that all derivatives of ψ appear in the following combination:

$$(\partial_\mu - i\epsilon B_\mu)\psi.$$

B_μ are 2×2 matrices such that[7] for $\mu = 1$, 2, and 3, B_μ is Hermitian and B_4 is anti-Hermitian. Invariance requires that

$$S(\partial_\mu - i\epsilon B_\mu')\psi' = (\partial_\mu - i\epsilon B_\mu)\psi. \tag{2}$$

Combining (1) and (2), we obtain the isotopic gauge transformation on B_μ:

$$B_\mu' = S^{-1}B_\mu S + \frac{i}{\epsilon}S^{-1}\frac{\partial S}{\partial x_\mu}. \tag{3}$$

The last term is similar to the gradiant term in the gauge transformation of electromagnetic potentials. In analogy to the procedure of obtaining gauge invariant field strengths in the electromagnetic case, we

[4] T. Lauritsen, Ann. Rev. Nuclear Sci. **1**, 67 (1952); D. R. Inglis, Revs. Modern Phys. **25**, 390 (1953).
[5] R. H. Hildebrand, Phys. Rev. **89**, 1090 (1953).
[6] W. Pauli, Revs. Modern Phys. **13**, 203 (1941).

[7] We use the conventions $\hbar = c = 1$, and $x_4 = it$. Bold-face type refers to vectors in isotopic space, not in space-time.

define now

$$F_{\mu\nu}=\frac{\partial B_\mu}{\partial x_\nu}-\frac{\partial B_\nu}{\partial x_\mu}+i\epsilon(B_\mu B_\nu-B_\nu B_\mu). \qquad (4)$$

One easily shows from (3) that

$$F_{\mu\nu}'=S^{-1}F_{\mu\nu}S \qquad (5)$$

under an isotopic gauge transformation.‡ Other simple functions of B than (4) do not lead to such a simple transformation property.

The above lines of thought can be applied to any field ψ with arbitrary isotopic spin. One need only use other representations S of rotations in three-dimensional space. It is reasonable to assume that different fields with the same total isotopic spin, hence belonging to the same representation S, interact with the same matrix field B_μ. (This is analogous to the fact that the electromagnetic field interacts in the same way with any charged particle, regardless of the nature of the particle. If different fields interact with different and independent B fields, there would be more conservation laws than simply the conservation of total isotopic spin.) To find a more explicit form for the B fields and to relate the B_μ's corresponding to different representations S, we proceed as follows.

Equation (3) is valid for any S and its corresponding B_μ. Now the matrix $S^{-1}\partial S/\partial x_\mu$ appearing in (3) is a linear combination of the isotopic spin "angular momentum" matrices T^i ($i=1, 2, 3$) corresponding to the isotopic spin of the ψ field we are considering. So B_μ itself must also contain a linear combination of the matrices T^i. But any part of B_μ in addition to this, $\bar{B}_\mu$, say, is a scalar or tensor combination of the T's, and must transform by the homogeneous part of (3), $\bar{B}_\mu'=S^{-1}\bar{B}_\mu S$. Such a field is extraneous; it was allowed by the very general form we assumed for the B field, but is irrelevant to the question of isotopic gauge. Thus the relevant part of the B field is of the form

$$B_\mu=2\mathbf{b}_\mu\cdot\mathbf{T}. \qquad (6)$$

(Bold-face letters denote three-component vectors in isotopic space.) To relate the $\mathbf{b}_\mu$'s corresponding to different representations S we now consider the product representation $S=S^{(a)}S^{(b)}$. The B field for the combination transforms, according to (3), by

$$B_\mu'=[S^{(b)}]^{-1}[S^{(a)}]^{-1}BS^{(a)}S^{(b)}$$
$$+\frac{i}{\epsilon}[S^{(a)}]^{-1}\frac{\partial S^{(a)}}{\partial x_\mu}+\frac{i}{\epsilon}[S^{(b)}]^{-1}\frac{\partial S^{(b)}}{\partial x_\mu}.$$

‡ *Note added in proof.*—It may appear that B_μ could be introduced as an auxiliary quantity to accomplish invariance, but need not be regarded as a field variable by itself. It is to be emphasized that such a procedure violates the principle of invariance. Every quantity that is not a pure numeral (like 2, or M, or any definite representation of the γ matrices) should be regarded as a dynamical variable, and should be varied in the Lagrangian to yield an equation of motion. Thus the quantities B_μ must be regarded as independent fields.

But the sum of $B_\mu^{(a)}$ and $B_\mu^{(b)}$, the B fields corresponding to $S^{(a)}$ and $S^{(b)}$, transforms in exactly the same way, so that

$$B_\mu=B_\mu^{(a)}+B_\mu^{(b)}$$

(plus possible terms which transform homogeneously, and hence are irrelevant and will not be included). Decomposing $S^{(a)}S^{(b)}$ into irreducible representations, we see that the twelve-component field $\mathbf{b}_\mu$ in Eq. (6) is the same for all representations.

To obtain the interaction between any field ψ of arbitrary isotopic spin with the $\mathbf{b}$ field one therefore simply replaces the gradiant of ψ by

$$(\partial_\mu-2i\epsilon\mathbf{b}_\mu\cdot\mathbf{T})\psi, \qquad (7)$$

where T^i ($i=1, 2, 3$), as defined above, are the isotopic spin "angular momentum" matrices for the field ψ.

We remark that the nine components of $\mathbf{b}_\mu$, $\mu=1, 2, 3$ are real and the three of $\mathbf{b}_4$ are pure imaginary. The isotopic-gauge covariant field quantities $F_{\mu\nu}$ are expressible in terms of $\mathbf{b}_\mu$:

$$F_{\mu\nu}=2\mathbf{f}_{\mu\nu}\cdot\mathbf{T}, \qquad (8)$$

where

$$\mathbf{f}_{\mu\nu}=\frac{\partial\mathbf{b}_\mu}{\partial x_\nu}-\frac{\partial\mathbf{b}_\nu}{\partial x_\mu}-2\epsilon\mathbf{b}_\mu\times\mathbf{b}_\nu. \qquad (9)$$

$\mathbf{f}_{\mu\nu}$ transforms like a vector under an isotopic gauge transformation. Obviously the same $\mathbf{f}_{\mu\nu}$ interact with all fields ψ irrespective of the representation S that ψ belongs to.

The corresponding transformation of $\mathbf{b}_\mu$ is cumbersome. One need, however, study only the infinitesimal isotopic gauge transformations,

$$S=1-2i\mathbf{T}\cdot\delta\omega.$$

Then

$$\mathbf{b}_\mu'=\mathbf{b}_\mu+2\mathbf{b}_\mu\times\delta\omega+\frac{1}{\epsilon}\frac{\partial}{\partial x_\mu}\delta\omega. \qquad (10)$$

FIELD EQUATIONS

To write down the field equations for the $\mathbf{b}$ field we clearly only want to use isotopic gauge invariant quantities. In analogy with the electromagnetic case we therefore write down the following Lagrangian density:[8]

$$-\tfrac{1}{4}\mathbf{f}_{\mu\nu}\cdot\mathbf{f}_{\mu\nu}.$$

Since the inclusion of a field with isotopic spin $\tfrac{1}{2}$ is illustrative, and does not complicate matters very much, we shall use the following total Lagrangian density:

$$\mathcal{L}=-\tfrac{1}{4}\mathbf{f}_{\mu\nu}\cdot\mathbf{f}_{\mu\nu}-\bar{\psi}\gamma_\mu(\partial_\mu-i\epsilon\boldsymbol{\tau}\cdot\mathbf{b}_\mu)\psi-m\bar{\psi}\psi. \qquad (11)$$

One obtains from this the following equations of motion:

$$\partial\mathbf{f}_{\mu\nu}/\partial x_\nu+2\epsilon(\mathbf{b}_\nu\times\mathbf{f}_{\mu\nu})+\mathbf{J}_\mu=0,$$
$$\gamma_\mu(\partial_\mu-i\epsilon\boldsymbol{\tau}\cdot\mathbf{b}_\mu)\psi+m\psi=0, \qquad (12)$$

[8] Repeated indices are summed over, except where explicitly stated otherwise. Latin indices are summed from 1 to 3, Greek ones from 1 to 4.

where

$$\mathbf{J}_\mu = i\epsilon\bar{\psi}\gamma_\mu\boldsymbol{\tau}\psi. \qquad (13)$$

The divergence of $\mathbf{J}_\mu$ does not vanish. Instead it can easily be shown from (13) that

$$\partial\mathbf{J}_\mu/\partial x_\mu = -2\epsilon\mathbf{b}_\mu\times\mathbf{J}_\mu. \qquad (14)$$

If we define, however,

$$\mathfrak{J}_\mu = \mathbf{J}_\mu + 2\epsilon\mathbf{b}_\nu\times\mathbf{f}_{\mu\nu}, \qquad (15)$$

then (12) leads to the equation of continuity,

$$\partial\mathfrak{J}_\mu/\partial x_\mu = 0. \qquad (16)$$

$\mathfrak{J}_{1,2,3}$ and $\mathfrak{J}_4$ are respectively the isotopic spin current density and isotopic spin density of the system. The equation of continuity guarantees that the total isotopic spin

$$\mathbf{T} = \int \mathfrak{J}_4 d^3x$$

is independent of time and independent of a Lorentz transformation. It is important to notice that $\mathfrak{J}_\mu$, like $\mathbf{b}_\mu$, does not transform exactly like vectors under isotopic space rotations. But the total isotopic spin,

$$\mathbf{T} = -\int \frac{\partial\mathbf{f}_{4i}}{\partial x_i}d^3x,$$

is the integral of the divergence of $\mathbf{f}_{4i}$, which transforms like a true vector under isotopic spin space rotations. Hence, under a general isotopic gauge transformation, if $S\to S_0$ on an infinitely large sphere, $\mathbf{T}$ would transform like an isotopic spin vector.

Equation (15) shows that the isotopic spin arises both from the spin-$\frac{1}{2}$ field ($\mathbf{J}_\mu$) and from the $\mathbf{b}_\mu$ field itself. Inasmuch as the isotopic spin is the source of the $\mathbf{b}$ field, this fact makes the field equations for the $\mathbf{b}$ field nonlinear, even in the absence of the spin-$\frac{1}{2}$ field. This is different from the case of the electromagnetic field, which is itself chargeless, and consequently satisfies linear equations in the absence of a charged field.

The Hamiltonian derived from (11) is easily demonstrated to be positive definite in the absence of the field of isotopic spin $\frac{1}{2}$. The demonstration is completely identical with the similar one in electrodynamics.

We must complete the set of equations of motion (12) and (13) by the supplementary condition,

$$\partial\mathbf{b}_\mu/\partial x_\mu = 0, \qquad (17)$$

which serves to eliminate the scalar part of the field in $\mathbf{b}_\mu$. This clearly imposes a condition on the possible isotopic gauge transformations. That is, the infinitesimal isotopic gauge transformation $S = 1 - i\boldsymbol{\tau}\cdot\delta\boldsymbol{\omega}$ must satisfy the following condition:

$$2\mathbf{b}_\mu\times\frac{\partial}{\partial x_\mu}\delta\boldsymbol{\omega} + \frac{1}{\epsilon}\frac{\partial^2}{\partial x_\mu{}^2}\delta\boldsymbol{\omega} = 0. \qquad (18)$$

This is the analog of the equation $\partial^2\alpha/\partial x_\mu{}^2 = 0$ that must be satisfied by the gauge transformation $A_\mu' = A_\mu + e^{-1}(\partial\alpha/\partial x_\mu)$ of the electromagnetic field.

QUANTIZATION

To quantize, it is not convenient to use the isotopic gauge invariant Lagrangian density (11). This is quite similar to the corresponding situation in electrodynamics and we adopt the customary procedure of using a Lagrangian density which is not obviously gauge invariant:

$$\mathcal{L} = -\frac{1}{2}\frac{\partial\mathbf{b}_\mu}{\partial x_\nu}\cdot\frac{\partial\mathbf{b}_\mu}{\partial x_\nu} + 2\epsilon(\mathbf{b}_\mu\times\mathbf{b}_\nu)\frac{\partial\mathbf{b}_\mu}{\partial x_\nu}$$
$$-\epsilon^2(\mathbf{b}_\mu\times\mathbf{b}_\nu)^2 + \mathbf{J}_\mu\cdot\mathbf{b}_\mu - \bar{\psi}(\gamma_\mu\partial_\mu + m)\psi. \qquad (19)$$

The equations of motion that result from this Lagrangian density can be easily shown to imply that

$$\frac{\partial^2}{\partial x_\nu{}^2}\mathbf{a} + 2\epsilon\mathbf{b}_\nu\times\frac{\partial}{\partial x_\nu}\mathbf{a} = 0,$$

where

$$\mathbf{a} = \partial\mathbf{b}_\mu/\partial x_\mu.$$

Thus if, consistent with (17), we put on one space-like surface $\mathbf{a} = 0$ together with $\partial\mathbf{a}/\partial t = 0$, it follows that $\mathbf{a} = 0$ at all times. Using this supplementary condition one can easily prove that the field equations resulting from the Lagrangian densities (19) and (11) are identical.

One can follow the canonical method of quantization with the Lagrangian density (19). Defining

$$\boldsymbol{\Pi}_\mu = -\partial\mathbf{b}_\mu/\partial x_4 + 2\epsilon(\mathbf{b}_\mu\times\mathbf{b}_4),$$

one obtains the equal-time commutation rule

$$[b_\mu{}^i(x), \Pi_\nu{}^j(x')]_{t=t'} = -\delta_{ij}\delta_{\mu\nu}\delta^3(x-x'), \qquad (20)$$

where $b_\mu{}^i$, $i = 1, 2, 3$, are the three components of $\mathbf{b}_\mu$. The relativistic invariance of these commutation rules follows from the general proof for canonical methods of quantization given by Heisenberg and Pauli.[9]

The Hamiltonian derived from (19) is identical with the one from (11), in virtue of the supplementary condition. Its density is

$$H = H_0 + H_{\text{int}},$$

$$H_0 = -\frac{1}{2}\boldsymbol{\Pi}_\mu\cdot\boldsymbol{\Pi}_\mu + \frac{1}{2}\frac{\partial\mathbf{b}_\mu}{\partial x_j}\cdot\frac{\partial\mathbf{b}_\mu}{\partial x_j} + \bar{\psi}(\gamma_j\partial_j + m)\psi, \qquad (21)$$

$$H_{\text{int}} = 2\epsilon(\mathbf{b}_i\times\mathbf{b}_4)\cdot\boldsymbol{\Pi}_i - 2\epsilon(\mathbf{b}_\mu\times\mathbf{b}_j)\cdot(\partial\mathbf{b}_\mu/\partial x_j)$$
$$+\epsilon^2(\mathbf{b}_i\times\mathbf{b}_j)^2 - \mathbf{J}_\mu\cdot\mathbf{b}_\mu.$$

The quantized form of the supplementary condition is the same as in quantum electrodynamics.

[9] W. Heisenberg and W. Pauli, Z. Physik 56, 1 (1929).

PROPERTIES OF THE b QUANTA

The quanta of the b field clearly have spin unity and isotopic spin unity. We know their electric charge too because all the interactions that we proposed must satisfy the law of conservation of electric charge, which is exact. The two states of the nucleon, namely proton and neutron, differ by charge unity. Since they can transform into each other through the emission or absorption of a b quantum, the latter must have three charge states with charges $\pm e$ and 0. Any measurement of electric charges of course involves the electromagnetic field, which necessarily introduces a preferential direction in isotopic space at all space-time points. Choosing the isotopic gauge such that this preferential direction is along the z axis in isotopic space, one sees that for the nucleons

$$Q=\text{electric charge}=e(\tfrac{1}{2}+\epsilon^{-1}T^z),$$

and for the b quanta

$$Q=(e/\epsilon)T^z.$$

The interaction (7) then fixes the electric charge up to an additive constant for all fields with any isotopic spin:

$$Q=e(\epsilon^{-1}T^z+R).\tag{22}$$

The constants R for two charge conjugate fields must be equal but have opposite signs.[10]

FIG. 1. Elementary vertices for b fields and nucleon fields. Dotted lines refer to b field, solid lines with arrow refer to nucleon field.

We next come to the question of the mass of the b quantum, to which we do not have a satisfactory answer. One may argue that without a nucleon field the Lagrangian would contain no quantity of the dimension of a mass, and that therefore the mass of the b quantum in such a case is zero. This argument is however subject to the criticism that, like all field theories, the b field is beset with divergences, and dimensional arguments are not satisfactory.

One may of course try to apply to the b field the methods for handling infinities developed for quantum electrodynamics. Dyson's approach[11] is best suited for the present case. One first transforms into the interaction representation in which the state vector Ψ

[10] See M. Gell-Mann, Phys. Rev. **92**, 833 (1953).
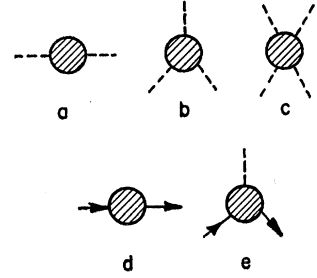[11] F. J. Dyson, Phys. Rev. **75**, 486, 1736 (1949).

FIG. 2. Primitive divergences.

satisfies

$$i\partial\Psi/\partial t=H_{\text{int}}\Psi,$$

where H_{int} was defined in Eq. (21). The matrix elements of the scattering matrix are then formulated in terms of contributions from Feynman diagrams. These diagrams have three elementary types of vertices illustrated in Fig. 1, instead of only one type as in quantum electrodynamics. The "primitive divergences" are still finite in number and are listed in Fig. 2. Of these, the one labeled a is the one that effects the propagation function of the b quantum, and whose singularity determines the mass of the b quantum. In electrodynamics, by the requirement of electric charge conservation,[12] it is argued that the mass of the photon vanishes. Corresponding arguments in the b field case do not exist[13] even though the conservation of isotopic spin still holds. We have therefore not been able to conclude anything about the mass of the b quantum.

A conclusion about the mass of the b quantum is of course very important in deciding whether the proposal of the existence of the b field is consistent with experimental information. For example, it is inconsistent with present experiments to have their mass less than that of the pions, because among other reasons they would then be created abundantly at high energies and the charged ones should live long enough to be seen. If they have a mass greater than that of the pions, on the other hand, they would have a short lifetime (say, less than 10^{-20} sec) for decay into pions and photons and would so far have escaped detection.

[12] J. Schwinger, Phys. Rev. **76**, 790 (1949).
[13] In electrodynamics one can formally prove that $G_{\mu\nu}k_\nu=0$, where $G_{\mu\nu}$ is defined by Schwinger's Eq. (A12). ($G_{\mu\nu}A_\nu$ is the current generated through virtual processes by the arbitrary external field A_ν.) No corresponding proof has been found for the present case. This is due to the fact that in electrodynamics the conservation of charge is a consequence of the equation of motion of the electron field alone, quite independently of the electromagnetic field itself. In the present case the b field carries an isotopic spin and destroys such general conservation laws.