Covariant theory of the quantized electromagnetic field with only physical photons

1970 J. Phys. A: Gen. Phys. 3 357

(http://iopscience.iop.org/0022-3689/3/4/006)

J. Phys. A: Gen. Phys., 1970, Vol. 3. Printed in Great Britain

Covariant theory of the quantized electromagnetic field with only physical photons

P. M. MATHEWS
Division of Theoretical Physics, University of Madras, Madras 25, India
MS. received 13th February 1970

Abstract. A manifestly covariant procedure for quantization of the electromagnetic potentials A^{μ} is presented, wherein only two types of photons with space-like polarization vectors find a place. The non-appearance of unphysical photons corresponding to the longitudinal and scalar photons of other formulations is due to our insistence from the beginning that the Lorentz condition be satisfied (as an operator equation) by the A^{μ}, and it makes the introduction of any indefinite metric unnecessary. The wave functions of the physical photons, which appear in the plane wave expansion of the A^{μ}, are definable *in principle* only to within arbitrary gauge terms. Consequently, expressions for the commutator between the field components and for the photon propagator contain indeterminate terms which serve merely to ensure that A^{μ} is coupled only to conserved currents; they do not contribute to physical matrix elements. Our theory therefore leads to the same results in quantum electrodynamics as earlier formulations, but in a manifestly Lorentz covariant and gauge invariant fashion and without recourse to concepts like the indefinite metric or any other *ad hoc* prescriptions.

1. Introduction

It is a strange fact that though the quantization of the electromagnetic field was carried out soon after the formulation of the quantum theory itself (Dirac 1927), and subsequent developments in the field of electrodynamics have led to dazzlingly successful results, the basic theory of the quantized electromagnetic field has remained fundamentally unsatisfactory in several respects. One of these, with which the present paper is concerned, is the quantization procedure itself. Since Maxwell's equations for electromagnetism, expressible in the form

$$F^{\mu\nu} = \partial^{\mu}A^{\nu} - \partial^{\nu}A^{\mu} \tag{1a}$$

$$\partial_{\mu}F^{\mu\nu} \equiv \partial_{\mu}\partial^{\mu}A^{\nu} - \partial^{\nu}\partial_{\mu}A^{\mu} = 0 \tag{1b}$$

are manifestly covariant under Lorentz transformation and also invariant under the gauge transformation

$$A^{\mu} \rightarrow A^{\mu} + \partial^{\mu}\Lambda \tag{2}$$

where Λ is an arbitrary function, it would be natural to expect the procedure for quantizing the field A^{μ} to be (*a*) manifestly covariant, (*b*) applicable in an arbitrary gauge and (*c*) such that Maxwell's equations (1) continue to be valid, now as equations for the *operators* A^{μ}. However, it appeared right from the beginning that this expectation was doomed to remain unrealized, for the different requirements seemed to impose conflicting conditions on the field variables. It is well known, for instance, that the manifestly covariant quantum rule†

$$[A^{\mu}(x),\, A^{\nu}(y)] = -\mathrm{i}g^{\mu\nu}D(x-y) \tag{3}$$

† Notation: $x = \{x^{\mu}\} \equiv (x^0, x^1, x^2, x^3)$; $\partial_{\mu} = \partial/\partial x^{\mu}$. The metric $g^{\mu\nu}$ is diagonal, with signature $+ - - -$. Units such that $\hbar = c = 1$ are assumed.

obtained by independently quantizing the four components A^μ which are taken to satisfy

$$\Box A^\mu \equiv \partial_\nu \partial^\nu A^\mu = 0 \tag{4}$$

is inconsistent with the Lorentz condition

$$\partial_\mu A^\mu = 0 \tag{5}$$

needed to supplement (4) in order that Maxwell's equation (1*b*) be obeyed. To get around this difficulty, Fermi (1932) proposed a compromise whereby equation (5) would not be insisted upon as an operator, but would be replaced by a supplementary condition which serves to define a class of 'physical' states in which alone Maxwell's equations would hold. The Gupta–Bleuler method (Gupta 1950, Bleuler 1950) which is now presented as standard in many textbooks on quantum field theory† is the culmination of this line of thinking. The essential features of this method which are of interest to us here are the following.

(i) The use of an indefinite metric in the definition of scalar products between states of the quantized field. This is necessitated by the difficulty in physical interpretation arising from the negative sign $(-g^{00})$ on the right-hand side of (3) when $\mu = \nu = 0$. The sign difficulty, in turn, is a consequence of quantizing all the four components A^μ independently of each other, ignoring the interconnection between them which is implied by Maxwell's equations.

(ii) Validity of Maxwell's equations only in a very weak sense, as relations between expectation values of the $F^{\mu\nu}$ in states Ψ picked out by the supplementary condition‡

$$(\partial_\mu A^\mu)^+ \Psi = 0$$

which replaces the operator equation (5). The abandonment of Maxwell's equations in operator form and the introduction of a supplementary condition on states is forced, as already noted, by the incompatibility of (5) with the covariant commutation rules (3).

(iii) The existence of states of vanishing norm (permitted by the indefinite metric) which obey the supplementary condition. The admixture of such states with states of positive norm leads to gauge transformations on *expectation* values of the A^μ.

Although for practical calculations the Gupta–Bleuler formalism has been quite satisfactory, it must be considered fundamentally unsatisfactory in principle that Maxwell's equations for the *observable* fields $\boldsymbol{E}$ and $\boldsymbol{H}$, based on direct experimental evidence, are weakened (for the sake of covariance in the commutation rules of the *unobservable* fields A^μ!). Further, the use of the indefinite metric and the concomitant appearance of ghost states of zero norm are sufficiently unpalatable that many authors prefer to employ a different approach in which the covariance is not manifest but physical interpretation, according to the conventional concepts of quantum theory, is possible. In this approach, employed originally by Dirac (1927) in the quantization of the pure radiation field, a *non-covariant* decomposition of the potentials into transverse, longitudinal and scalar parts is made, the scalar part is expressed in terms of the charged matter current with which the electromagnetic field interacts, and the longitudinal part is eliminated by choosing the Coulomb

† See, for example, Heitler (1954), Jauch and Rohrlich (1955), Bogoliubov and Shirkov (1959), Schweber (1961).

‡ $(\partial_\mu A^\mu)^+$ means the positive energy part of $\partial_\mu A^\mu$.

gauge.† This leaves only the two transverse degrees of freedom to be independently quantized in the usual fashion. One of the most unappealing features of the resulting theory is the appearance of a non-covariant term in the transverse-photon propagator, whose effects are to be cancelled ultimately by a contribution from an instantaneous Coulomb interaction. Further, by working in a specific gauge (whose characterization is by itself non-covariant), one has also lost explicit gauge invariance. Of course one verifies in the end that the *final results* are Lorentz covariant and gauge invariant.

During the course of four decades that have elapsed since Dirac's original work, approaches other than the above two widely used ones have been proposed, for example, one based on the use of potentials defined gauge-invariantly (Mandelstam 1962, Rhorlich and Strocchi 1965), but not with any conspicuous success. Efforts in recent years aimed at deriving manifestly covariant schemes free of the less palatable features of the Fermi–Gupta–Bleuler method have produced a number of *ad hoc* prescriptions—like the introduction of auxiliary fields (Nakanishi 1967, Yokoyama 1968)—and others relying on rather drastic steps like dissociating the free-field theory from any limiting form of the theory with interaction, denying thereby the possibility of having an interaction picture (Just 1965). These have been resorted to because it was clear that a straightforward modification of the commutation rule (3) to take account of the Lorentz condition by subtracting from $g^{\mu\nu}$ a term proportional to $k^\mu k^\nu$ (in the momentum representation) would lead to meaningless results (unlike the case of the massive vector field), because of the zero length of the four-vector k^μ. What had not been generally realized, however, is that the A^μ, which define the electric and magnetic fields through equations (1), need not transform strictly as the components of a vector and that the conditions on the commutator arising from its supposed tensor character therefore become correspondingly less stringent. Once this fact is fully appreciated, the foundations of the widely held belief that potentials of the Maxwell fields cannot be covariantly quantized (without the aid of indefinite metric, etc.) get seriously eroded. In fact the author has already set up a covariant, gauge-invariant quantization procedure wherein the Maxwell equations are taken care of (in a covariant fashion) right at the beginning, before quantization, and no supplementary conditions, indefinite metric etc., are required. The salient points of the new approach have been given in outline in a recent paper (Mathews 1969 a). In this paper we present the theory in greater detail, including a discussion of certain features which will undoubtedly appear strange at first sight (e.g. terms in the commutator whose magnitude is indeterminate in principle). The latter are intimately related to questions of gauge invariance and covariance, which are important in themselves and are also of direct relevance in understanding the nature of the Hilbert space of photon states. An analysis of these will form an essential part of our discussion.

2. Quantization

In our approach to the quantization problem it is a fundamental requirement that the electromagnetic potential must satisfy Maxwell's equations (1) or

† For accounts of this type of approach, see, for instance, Schiff (1955), Bjorken and Drell (1965) and Weinberg (1965). The decomposition of the A^μ suggested by Schwinger (1948) is also essentially of the same kind, though the lack of covariance is here masked by the introduction of an extraneous time-like unit vector. A very instructive comparison of calculations using the non-covariant and Gupta–Bleuler methods may be found in the book by Heitler (1954).

equivalently,† *both* the Klein–Gordon equation (4) and the Lorentz condition (5). Therefore we can decompose the real field $A^\mu(x)$ in the usual manner into

$$A^\mu(x) = (2\pi)^{-3/2}\int \frac{\mathrm{d}^3k}{(2k^0)^{1/2}}\{A^\mu(\boldsymbol{k})\,\mathrm{e}^{-\mathrm{i}k.x}+A^{\mu *}(\boldsymbol{k})\,\mathrm{e}^{\mathrm{i}k.x}\} \tag{6}$$

with

$$k\,.\,x = k^0x^0-\boldsymbol{k}\,.\,\boldsymbol{x} \qquad \text{and} \qquad k^0 = +|\boldsymbol{k}|.$$

On account of our insistence on equation (5), the Fourier coefficients $A^\mu(k)$ must satisfy

$$k_\mu\, A^\mu(\boldsymbol{k}) = 0. \tag{7}$$

Now, k^μ being a light-like four-vector, it is known (Wigner 1939) that the space orthogonal to it can be spanned by a basis consisting of the vector k^μ itself and two space-like vectors, $u_1{}^\mu$ and $u_2{}^\mu$, all mutually orthogonal. It follows then that $A^\mu(\boldsymbol{k})$, which by (7) is orthogonal to k^μ, can be written as

$$A^\mu(\boldsymbol{k}) = \sum_{r=1}^{3} a_r(\boldsymbol{k})\, u_r{}^\mu(\boldsymbol{k}) \tag{8}$$

with

$$u_3{}^\mu = k^\mu, \qquad u_3{}^\mu u_{3\mu} \equiv k^\mu k_\mu = 0 \tag{9a}$$

$$u_i{}^\mu u_{j\mu} = -\delta_{ij} \qquad (i,j = 1,2) \tag{9b}$$

and

$$u_3{}^\mu u_{i\mu} \equiv k^\mu u_{i\mu} = 0. \tag{9c}$$

Thus the *four* quantities $A^\mu(k)$ are expressed in terms of *three* amplitudes $a_r(k)$ of which only the first *two* contribute to observable quantities (since the term in $u_3{}^\mu \equiv k^\mu$ in (8) drops out when $F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu$ is formed). It is easy, for instance, to verify that the energy momentum vector

$$P^\mu = \int (F^{\mu\rho}F_\rho{}^0+\tfrac{1}{4}F^{\rho\sigma}F_{\rho\sigma}g^{\mu 0})\,\mathrm{d}^3x \tag{10}$$

which may be obtained from the gauge-invariant Lagrangian density

$$\mathscr{L} = -\tfrac{1}{4}F^{\rho\sigma}F_{\rho\sigma} \tag{11}$$

by the application of Noether's theorem, takes the form

$$P^\mu = -\tfrac{1}{2}\int \mathrm{d}^3k\, k^\mu[A^\rho(\boldsymbol{k})A_\rho{}^*(\boldsymbol{k})+A_\rho{}^*(\boldsymbol{k})A^\rho(\boldsymbol{k})] \tag{12}$$

which reduces to

$$P^\mu = \tfrac{1}{2}\int \mathrm{d}^3k\, k^\mu \sum_{i=1}^{2}(a_i a_i{}^* + a_i{}^* a_i) \tag{13}$$

† Actually the general solution of (1) contains, in addition to the general solution of the simultaneous equations (4) and (5), a term of the form $\partial^\mu\chi$ which does not contribute to $F^{\mu\nu}$ and hence to any observable quantities. Since $\partial^\mu\chi$, by definition, does not satisfy (5), its Fourier components are associated with four vectors k^μ of non-zero length. Consequently χ must remain as a c number function even in the quantized theory, since if its Fourier coefficients were elevated to operator status, it would amount to introducing 'photons' of various masses $k^2 \neq 0$. Such a c number part in A^μ would have no role to play in the theory and can indeed be eliminated by a c number gauge transformation which takes one into the Lorentz gauge. Gauge transformations *within* the Lorentz gauge, however, are of a fundamentally different nature, as will be seen below.

on introducing (8) and using equations (9). It is noteworthy that P^μ here depends only on two amplitudes† and that H is positive definite as it stands, though this is really not surprising since we have already taken account of Maxwell's equations in full. The quantum conditions required for particle interpretation are clearly

$$\begin{aligned} [a_i(\boldsymbol{k}), a_j^*(\boldsymbol{k}')] &= \delta_{ij}\,\delta(\boldsymbol{k}-\boldsymbol{k}') \\ [a_i(\boldsymbol{k}), a_j(\boldsymbol{k}')] &= [a_i^*(\boldsymbol{k}), a_j^*(\boldsymbol{k}')] = 0. \qquad (i,j = 1,2) \end{aligned} \tag{14}$$

At this point it appears that the amplitude a_3 does not have any particular role to play, and one might be tempted to treat it as a mere c number and get rid of it from A^μ by a c number gauge transformation. However, such a procedure would not be invariant because the distinction between a_1 and a_2 on the one hand and a_3 on the other is not an invariant one—this can be seen by changing to a basis u_r' related to the u_r by

$$u_1 = u_1' + f u_3', \qquad u_2 = u_2' + g u_3', \qquad u_3 = h u_3' \tag{15}$$

where f, g, h are arbitrary constants. The u_r' possess the same orthonormality properties‡ (9) as the u_r. On substituting (15) in (8) we see that the change to the new basis induces the following transformation on the expansion coefficients:

$$a_1 \to a_1' = a_1, \qquad a_2 \to a_2' = a_2, \qquad a_3 \to a_3' = f a_1 + g a_2 + h a_3. \tag{16}$$

It is clear that even if a_3 were chosen to be a c number a_3' would not be one. However, it is important to recognize that a_1, a_2 (and their Hermitian conjugates) are still the only independent operators in the theory. The operator nature of a_3 arises solely from possible admixture§ of a_1 and a_2 as in (16). A consequence of this, which has no parallel in the quantum theory of any other known system, is that all commutators involving a_3 or a_3^* are indeterminate *in principle*. It may be verified that it is impossible to keep the value of such commutators invariant under the transformation (16), for all f, g, and h.

Since we can neither eliminate a_3 from the theory in an invariant way, nor assign definite values to its commutators, the presence of undetermined terms in the commutator of the potentials A^μ is inescapable. We have, in fact,

$$[A^\mu(\boldsymbol{k}), A^{\nu *}(\boldsymbol{k}')] = \{u_1^\mu u_1^\nu + u_2^\mu u_2^\nu + k^\mu f^\nu + f^{\mu *} k^\nu\}\,\delta(\boldsymbol{k}-\boldsymbol{k}') \tag{17}$$

where

$$\delta(\boldsymbol{k}-\boldsymbol{k}') f^\nu \equiv [a_3(\boldsymbol{k}), a_1^*(\boldsymbol{k}')\,u_1^\nu + a_2^*(\boldsymbol{k}')\,u_2^\nu + \tfrac{1}{2} a_3^*(\boldsymbol{k}')\,u_3^\nu] \tag{18}$$

† The absence of any term containing a_3 in (13) is one of the important consequences of the vanishing norm of u_3. The presence in A^μ of a part of vanishing norm is responsible for practically all the difficulties peculiar to the massless vector field, in the quantization procedure as well as in physical interpretation. It may be remarked here that, while the massive vector field also can be resolved in exactly the same form as in (8), all three u_r are of non-vanishing norm there and so the P^μ contain all the three a_r. See Mathews (1969 b).

‡ That this is so despite the complete arbitrariness of f, g, and h, is another consequence of the zero norm of u_3. This may be contrasted with the usual situation where the preservation of orthonormality requires the coefficients characterizing the basis transformation to be constrained by unitarity conditions.

§ In view of this we may (after eliminating any c number part in a_3 through a c number gauge transformation) write a_3 simply as $f a_1 + g a_2$, where the values of the coefficients f and g are arbitrary and indeterminate. Such a representation will be found particularly useful in discussing the Hilbert space of photon states. See below.

is indeterminate. Nevertheless it follows from (9*c*) and (18) that

$$k_\mu f^\mu = 0 \tag{19}$$

and therefore it is trivial to verify explicitly that the commutation relation (17) is consistent with the Lorentz condition. The validity of Maxwell's equations as operator equations, on which we have insisted from the beginning, is thus ensured, and there is no need for supplementary conditions. The price paid for this is the occurrence of indeterminate terms in the commutator which cannot be objected to on principle since A^μ itself is unobservable† and undefined to the extent of arbitrary gauge terms. It must be proved of course that no practical difficulties arise from the presence of these terms, and we shall come to this aspect presently, but first we rewrite (17) in an alternative form which brings the familiar $g^{\mu\nu}$ into the commutator.

Let us introduce a fourth vector $u_0{}^\mu$ which, together with the $u_r{}^\mu$, forms a basis for the whole four-dimensional vector space. The scalar product of u_0 with u_3 must be necessarily non-zero‡ and we take it to be unity. Without loss of generality we can take u_0 to be light-like and orthogonal to u_1 and u_2. Thus

$$u_0{}^\mu u_{0\mu} = 0, \qquad u_0{}^\mu u_{i\mu} = 0 \quad (i = 1, 2), \qquad u_0{}^\mu u_{3\mu} = 1. \tag{20}$$

The tensor $g^{\mu\nu}$ can be expressed in terms of the u_α ($\alpha = 0, 1, 2, 3$) as

$$g^{\mu\nu} = u_0{}^\mu u_3{}^\nu - u_1{}^\mu u_1{}^\nu - u_2{}^\mu u_2{}^\nu + u_3{}^\mu u_0{}^\nu. \tag{21}$$

That the expression (21) does act as the index-raising operator can be verified with the aid of (9) and (20). Using (21) in (17) we now rewrite the commutator as

$$[A^\mu(\boldsymbol{k}), A^{\nu *}(\boldsymbol{k}')] = \{-g^{\mu\nu} + u_0{}^\mu k^\nu + k^\mu u_0{}^\nu + k^\mu f^\nu + f^{\mu *} k^\nu\}\delta(\boldsymbol{k}-\boldsymbol{k}'). \tag{22}$$

This expression differs from the familiar $-g^{\mu\nu}\,\delta(\boldsymbol{k}-\boldsymbol{k}')$ of the Fermi–Gupta–Bleuler theory by terms which are all proportional to k^μ or k^ν, and of course a similar situation holds in the case of the propagator. However, it has already been proved in connection with the non-covariant (Coulomb gauge) approach that such terms do not make any contribution to the complete amplitude§ for any physical process. Consequently, our theory does not change any of the standard results. This is indeed eminently satisfactory.

3. Covariance

Returning now to the question of covariance, we observe that though we have had to prescribe quantum rules for only two truly independent pairs of operators a_i, a_i^* ($i = 1, 2$), exactly as in the non-covariant Coulomb gauge approach, we have nevertheless not had to abandon covariance in order to achieve this. The resolution of $A^\mu(\boldsymbol{k})$ at each point in momentum space, which gave us the amplitudes to be quantized, was done in terms of a basis $u_r(\boldsymbol{k})$ which was defined in a *manifestly covariant* fashion

† It may be recalled here that the unobservability of A^μ has had to be invoked in some other treatments too, e.g. in the Coulomb gauge approach where the commutator of $A^\mu(x)$ and $A^\nu(y)$ at space-like separations $(x-y)$ turns out to be non-vanishing. See, for example, Schiff (1955).

‡ If u_α ($\alpha = 0, 1, 2, 3$) formed a basis for the four-vector space, and u_3 were orthogonal to all four of these, then it would be orthogonal to the whole space and must necessarily vanish.

§ The contribution at a given vertex is not necessarily zero, but when the different possible positions (relative to other vertices) at which a given photon line can be attached to a given electron line are all taken into account, the total contribution vanishes. See, for instance, Bjorken and Drell (1965, chap. 17).

through the Lorentz condition $k_\mu u_i^\mu = 0$. Despite this, however, the fact remains (paradoxical as it may seem) that the expression (22) for the commutator is not Lorentz invariant in the conventional sense, i.e. the right-hand member in (22), considered as a function of $\boldsymbol{k}$, does not form an *invariant* tensor field as would normally be expected, because the factors $u_i^\mu (i = 1, 2)$ contained in the f^μ and u_0^μ are not invariant vector fields. The source of this strange phenomenon is of course to be traced again to the vanishing norm of the vector k^μ. In the case of the massive vector field ($k^2 - m^2 \neq 0$) an identical procedure to what we have advocated in this paper (Mathews 1969 b) leads to an *invariant* form

$$[A^\mu(\boldsymbol{k}), A^{\nu *}(\boldsymbol{k}')] = -\left\{g^{\mu\nu} - \frac{k^\mu k^\nu}{m^2}\right\} \delta(\boldsymbol{k}-\boldsymbol{k}') \tag{23}$$

for the commutator, the expression in curly brackets being simply the projection operator to the three-dimensional space defined by $k_\mu A^\mu = 0$. When $k^2 = 0$, on the other hand, k^μ itself belongs to this three-dimensional space and *no* fourth vector orthogonal to this space exists, so that there is no possibility of obtaining the relevant projection operator by removing from the unit operator ($g^{\mu\nu}$) a projection operator to a complementary (one-dimensional) subspace as in (23). In fact there is not even an invariant distinction between the 'Lorentz subspace' spanned by the u_r^μ ($r = 1, 2, 3$) and the 'complementary subspace' defined by u_0^μ, for a basis change of the type (15) within the former must be accompanied by a change of u_0 also, such that

$$u_0 = \frac{1}{h}[u_0' + f u_1' + g u_2' + \tfrac{1}{2}(f^2+g^2)u_3'] \tag{24}$$

in order that the conditions (20) be maintained. The significant point to note is the admixture of the u_r' ($r = 1, 2, 3$) with the 'complementary' vector u_0'. The lack of Lorentz invariance of the commutator as a whole is directly due to the failure of the Lorentz condition to achieve a separation of the four-dimensional space into two invariant complementary spaces (just as the presence of indeterminate terms in the commutator was caused by the presence of a distinction between u_3 on the one hand and u_1 and u_2 on the other, which is not an invariant one).

It is important to observe that the departure from normal expectations in the transformation character of the right-hand side of (22) is intimately linked to the fact that the A^μ themselves do not transform 'normally'. Of particular relevance in this connection is the work of Strocchi (1967) which shows from very general considerations that if the A^μ are required (*a*) to satisfy Maxwell's equations and (*b*) to transform strictly as the components of a four-vector then the electric and magnetic fields must vanish, so that the resulting theory is trivial.† Since we do insist on Maxwell's equations, we have to relax the transformation law of the A^μ to the extent of allowing

† Essentially the same result is contained in the earlier work of Wightman and Gårding (1964). It has been known for some time that, when one works in some special gauge (e.g. one in which the vector potential is transverse), the effect of a change in Lorentz frame on the A^μ is a combination of the usual linear homogeneous vector transformation, together with a gauge transformation, and that the gauge function is in general an operator function (see, for example, Bjorken and Drell 1965, chap. 14, Weinberg 1965). What had not been clear is that the occurrence of the extra 'gauge' term in the transformation law is not just a consequence of using an explicitly non-covariant approach assuming some special gauge, but that it is necessitated by the very requirement that Maxwell's equations be satisfied.

extra 'gauge' terms in the transformed field.† The lack of strict Lorentz invariance of the commutator, despite the manifest covariance of the quantization procedure, is a manifestation of this relaxation. Nevertheless, as we have already seen, all physical results are completely covariant because the ambiguous terms in the commutator (those other than $-g^{\mu\nu}$ in (22)) do not contribute to the total amplitude for any such process.

4. The Hilbert space

Finally, we would like to discuss briefly the definition of the Hilbert space of photon states. We have already mentioned that the only operators in the theory are a_1, a_2 and their Hermitian conjugates, and that, while a_3 cannot be invariantly defined to be a c number, its operator properties are nevertheless entirely due to arbitrary admixtures of a_1 and a_2. Another way of expressing this is to rewrite (8) as

$$A^{\mu}(\boldsymbol{k}) = a_1\{u_1{}^{\mu}+fu_3{}^{\mu}\}+a_2\{u_2{}^{\mu}+gu_3{}^{\mu}\} \tag{25}$$

with f and g arbitrary. One might even leave out the explicit terms in $u_3{}^{\mu}$ provided $u_1{}^{\mu}$ and $u_2{}^{\mu}$ are understood to be *not* specific four-vectors orthogonal to k^{μ}, but four-vectors which are defined only to within an additive term which is an arbitrary multiple of $u_3{}^{\mu} \equiv k^{\mu}$.

Let us now define the vacuum state $|0\rangle$ by

$$a_1\,|0\rangle = a_2\,|0\rangle = 0, \qquad \langle 0|0\rangle = 1 \tag{26}$$

and from it, build up by repeated application with operators $a_i{}^*(\boldsymbol{k})$ $(i = 1, 2)$, states

$$\prod_{r=1}^{m} a_1{}^*(\boldsymbol{k}_r) \prod_{s=1}^{n} a_2{}^*(\boldsymbol{k}_s)\;|0\rangle \tag{27}$$

with m photons of 'type 1' and n photons of 'type 2'. Every one of these states, for any m and n, has positive norm, as can easily be verified by using (26) and (14). The linear space spanned by these states in thus endowed with a positive definite metric and so is a Hilbert space, exactly as in the case of massive fields. And the operators $a_1{}^*(\boldsymbol{k})$ and $a_2{}^*(\boldsymbol{k})$ are creation operators of photons of two types, with momentum $\boldsymbol{k}$ and energy $|\boldsymbol{k}|$, as can be seen by considering the effect of P^{μ}, equation (13), on the states (27).

However, there is one outstanding respect in which the present situation differs from theories of massive fields. It is that the *wave function* of the particle (photon) created by $a_1{}^*$ (or $a_2{}^*$) is *not* uniquely determined. It is determined only modulo an arbitrary multiple of k^{μ} (e.g. $u_1{}^{\mu}+fk^{\mu}$ with f arbitrary) and we have a whole equivalence class of wave functions to describe a specific one-photon state,‡ the difference

† Observe that in view of the nature of the commutator in (22), the two-point Wightman function $\langle 0|A^{\mu}(x)\;A^{\nu}(y)|0\rangle$ in our theory will *not* have the form $g^{\mu\nu}D_1(x-y)+\partial^{\mu}\partial^{\nu}D_2(x-y)$ which it would have been constrained to have, according to Strocchi (1967), if the A^{μ} transformed strictly as a four-vector.

‡ Moses (1966) has suggested a quantization procedure similar to ours but with the fundamental difference that he uses a specific set of four-component functions of k in the place of our wave functions $u_{\alpha}{}^{\mu}$. In doing so, the fact that a_3 cannot be invariantly defined to be a c number, and the consequent arbitrariness in the wave functions created by $a_1{}^*$ and $a_2{}^*$ (which is the essence of the difference between the photon field and other fields), are completely missed. Moreover, the use of complicated explicit forms for the u_{α} masks the essential simplicity of the quantization procedure and makes the covariance anything but manifest.

between the various members of the class being a multiple of k^{μ} (or in coordinate space language, the gradient of a scalar function). The essential manner in which the gauge freedom enters into the quantum theory of the electromagnetic potentials is thus vividly displayed.

The fact that a whole equivalence class of wave functions, rather than a single unique wave function, may have to be used in the description of a state of a massless particle has been recognized in the literature (Shaw 1965) but has certainly not been very widely appreciated. It may be useful, therefore, to point out here how this circumstance arises. It is well known from Wigner's work on the irreducible representations of the Poincaré group (Wigner 1939) that the zero-mass irreducible representations are labelled by an additional quantum number, helicity (which takes the place of the spin quantum number of massive particles) and are one-dimensional; a massless particle state with definite helicity transforms into itself under all transformations of the Poincaré group. However, one does not try to represent this state by a single-component wave function because the effect of Lorentz transformation on such a wave function is expressed through a phase factor with a complicated dependence on the transformation itself and on the momentum value of the state being transformed† (i.e. it is a non-local transformation in coordinate space). One employs instead a wave function which transforms locally, according to some irreducible representation $D(k, l)$—or a direct sum of such representations—of the *homogeneous* Lorentz group. Since the maximum helicity contained in $D(k, l)$ is $k+l$, to describe a massless particle of helicity λ one needs irreducible representations $D(k, l)$ with $k+l \geqslant \lambda$, and when λ is non-zero this means that k and/or l must be non-zero, so that the dimension $(2k+1)(2l+1)$ of the wave function is greater than unity. Then λ is not the only helicity contained in the wave function. Nevertheless *if* the helicity λ projection of this wave function transforms into itself under all Lorentz transformations, then one gets a one-to-one correspondence between the states and wave functions of a massless particle of helicity λ. Now, there is a general theorem, proved by Weinberg (1964 b) by applying 'little group' considerations to quantized massless fields and by the present author (Mathews 1969 c) from elementary quantum mechanics, which states that if, for the description of a massless particle, wave functions transforming according to $D(k, l)$ are used the *only invariant helicity* is $(l-k)$, i.e. the projection to the helicity value $(l-k)$ is the only one which does not get mixed up with other helicity projections under Lorentz transformations. If one insists nevertheless on using $D(k, l)$ to describe some helicity $\lambda \neq (l-k)$, then the admixture of other helicity projections which results when Lorentz transformations are performed has to be considered to be in the nature of a gauge transformation (and not as the introduction of particle states with physical helicity values different from λ). A whole class of wave functions, all of which have the same content of the helicity λ projection, but differ in the admixture of the other helicity parts, is to be identified then with a single physical state of helicity λ.

In the photon case, which is what we are concerned with here, the admissible helicity values are $+1$ and -1. Unique wave functions to represent these helicity states can be obtained if we use wave functions transforming according to $D(0, 1) \oplus D(1, 0)$ to describe the photon, the $D(0, 1)$ and $D(1, 0)$ parts giving us *invariant* helicities $+1$ and -1 respectively by the theorem quoted above. This is precisely what we have when a description in terms of the antisymmetric tensor field $F^{\mu\nu}$ is employed; the transformation of $F^{\mu\nu}$ is equivalent to $D(0, 1) \oplus D(1, 0)$.

† See, for instance, Weinberg (1964 a, appendix A).

However, when the four-potential A^μ transforming as $D(\frac{1}{2}, \frac{1}{2})$ is employed instead, the only invariant helicity is zero, which is uninteresting. The physically interesting helicities $+1$ and -1 get admixtures of the zero-helicity state (with wave function proportional to k^μ) under Lorentz transformations, and this is why one has an *equivalence class* of wave functions (the members of which differ only by gauge transformations corresponding to arbitrary admixtures of the helicity zero part) to describe a state of given helicity $+1$ or -1, when the potentials A^μ are employed.

5. Conclusions

Our considerations above have been confined to the free Maxwell field but, as is well known, this is all that is needed to make calculations on processes involving matter–field interactions, via perturbation theory in the interaction picture. We have demonstrated that, in the description of the free electromagnetic field, one needs to talk of only two types of photons with space-like polarization vectors, covariantly defined (which can be made 'transverse' in the three-dimensional sense, if need be, by choosing the arbitrary constants f and g in (25) in such a way as to make the time-like component of the wave function vanish). There is no need to introduce even the notion of 'longitudinal' and 'scalar' photons which have no physical existence. It is on account of this that we are able to have a Hilbert space (with a positive definite metric) for photons, unlike in the Gupta–Bleuler formalism. At the same time, compared with the Coulomb gauge approach (in which also only two types of photons are explicitly introduced), we have the advantage of maintaining explicit gauge-invariance and manifest covariance.

Acknowledgments

The major part of this work was done while the author was visiting at the International Centre for Theoretical Physics, Trieste. It is a pleasure to acknowledge our gratitude to Professors Abdus Salam and P. Budini and to the International Atomic Energy Agency, for hospitality at Trieste.

Note Added in Proof. It may be noted that in the presence of interactions the A^μ in the Heisenberg picture will have, besides the 'intrinsic' part with space-like polarization vectors considered above, also an 'induced' part which is just a function of the current which interacts with the field. The strong resemblance to the Coulomb gauge approach (apart from covariance) is obvious, and provides a heuristic argument that the present theory would reproduce the conventional results. A detailed treatment of the interacting case will be presented separately.

References

Bjorken, J. D., and Drell, S. D., 1965, *Relativistic Quantum Fields* (New York: McGraw-Hill).
Bleuler, K., 1950, *Helv. Phys. Acta*, **23**, 567–86.
Bogoliubov, N. N., and Shirkov, D. V., 1959, *Introduction to the Theory of Quantised Fields* (New York: Interscience).
Dirac, P. A. M., 1927, *Proc. R. Soc.* A, **114**, 243–65.
Fermi, E., 1932, *Rev. mod. Phys.*, **4**, 87–132.
Gupta, S. N., 1950, *Proc. Phys. Soc.* A, **63**, 681–91.
Heitler, W., 1954, *The Quantum Theory of Radiation* (Oxford: Clarendon Press).
Jauch, J. M., and Rohrlich, F., 1955, *The Theory of Photons and Electrons* (Reading, Mass.: Addison-Wesley).
Just, K., 1965, *Nuovo Cim.*, **38**, 400–9.
Mandelstam, S., 1962, *Ann. Phys., N.Y.*, **19**, 1–24.
Mathews, P. M., 1969 a, *Lett. Nuovo Cim.*, **2**, 739.

—— 1969 b, to be published in *J. math. Phys.*
—— 1969 c, to be published in *J. math. phys. Sci., Madras.*
MOSES, H. E., 1966, *Nuovo Cim.*, **42** A, 757–81.
NAKANISHI, N., 1967, *Prog. theor. Phys.*, **38,** 881–91.
ROHRLICH, F., and STROCCHI, F., 1965, *Phys. Rev.*, **139,** B476–82.
SCHIFF, L. I., 1955, *Quantum Mechanics* (New York: McGraw-Hill).
SCHWEBER, S. S., 1961, *An Introduction to Relativistic Quantum Field Theory* (New York: Harper and Row).
SCHWINGER, J., 1948, *Phys. Rev.*, **74,** 1439–61.
SHAW, R., 1965, *Nuovo Cim.*, **37,** 1086–99.
STROCCHI, F., 1967, *Phys. Rev.*, **162,** 1429–38.
WEINBERG, S., 1964 a, *Phys. Rev.*, **135,** B1049–56.
—— 1964 b, *Phys. Rev.*, **134,** B882–96.
—— 1965, *Phys. Rev.*, **138,** B988–1002.
WIGHTMAN, A. S., and GÅRDING, L., 1964, *Ark. Fys.*, **28,** 129–84.
WIGNER, E. P., 1939, *Ann. Math.*, **40,** 149–204.
YOKOYAMA, K., 1968, *Prog. theor. Phys.*, **40,** 160–6.