PHYSICAL REVIEW VOLUME 75, NUMBER 11 JUNE 1, 1949

The S Matrix in Quantum Electrodynamics

F. J. DYSON
Institute for Advanced Study, Princeton, New Jersey
(Received February 24, 1949)

The covariant quantum electrodynamics of Tomonaga, Schwinger, and Feynman is used as the basis for a general treatment of scattering problems involving electrons, positrons, and photons. Scattering processes, including the creation and annihilation of particles, are completely described by the S matrix of Heisenberg. It is shown that the elements of this matrix can be calculated, by a consistent use of perturbation theory, to any desired order in the fine-structure constant. Detailed rules are given for carrying out such calculations, and it is shown that divergences arising from higher order radiative corrections can be removed from the S matrix by a consistent use of the ideas of mass and charge renormalization.

Not considered in this paper are the problems of extending the treatment to include bound-state phenomena, and of proving the convergence of the theory as the order of perturbation itself tends to infinity.

I. INTRODUCTION

IN a previous paper[1] (to be referred to in what follows as I) the radiation theory of Tomonaga[2] and Schwinger[3] was applied in detail to the problem of the radiative corrections to the motion of a single electron in a given external field. It was shown that the rules of calculation for corrections of this kind were identical with those which had been derived by Feynman[4] from his own radiation theory. For the one-electron problem the radiative corrections were fully described by an operator H_T (Eq. (20) of I) which appeared as the "effective potential" acting upon the electron, after the interactions of the electron with its own self-field had been eliminated by a contact transformation. The difference between the Schwinger and Feynman theories lay only in the choice of a particular representation in which the matrix elements of H_T were calculated (Section V of I).

The present paper deals with the relation between the Schwinger and Feynman theories when the restriction to one-electron problems is removed. In these more general circumstances the two theories appear as complementary rather than identical. The Feynman method is essentially a set of rules for the calculation of the elements of the Heisenberg S matrix corresponding to any physical process, and can be applied with directness to all kinds of scattering problems.[5] The Schwinger method evaluates radiative corrections by exhibiting them as extra terms appearing in the Schrödinger equation of a system of particles and is suited especially to bound-state problems. In spite of the difference of principle, the two methods in practice involve the calculation of closely related expressions; moreover, the theory underlying them is in all cases the same. The systematic technique of Feynman, the exposition of which occupied the second half of I and occupies the major part of the present paper, is therefore now available for the evaluation not only of the S matrix but also of most of the operators occurring in the Schwinger theory.

The prominent part which the S matrix plays in this paper is due to its practical usefulness as the connecting link between the Feynman technique of calculation and the Hamiltonian formulation of quantum electrodynamics. This practical usefulness remains, whether or not one follows Heisenberg in believing that the S matrix may eventually replace the Hamiltonian altogether. It is still an unanswered question, whether the finiteness of the S matrix automatically implies the finiteness of all observable quantities, such as bound-state energy levels, optical transition probabilities, etc., occurring in electrodynamics. An affirmative answer to the question is in no way essential to the arguments of this paper. Even if a finite S matrix does not of itself imply finiteness of other observable quantities, it is probable that all such quantities will be finite; to verify this, it will be necessary to repeat the analysis of the present paper, keeping all the time closer to the original Schwinger theory than

[1] F. J. Dyson, Phys. Rev. **75**, 486 (1949).

[2] Sin-Itiro Tomonaga, Prog. Theor. Phys. **1**, 27 (1946); Koba, Tati, and Tomonaga, Prog. Theor. Phys. **2**, 101 (1947) and **2**, 198 (1947); S. Kanesawa and S. Tomonaga, Prog. Theor. Phys. **3**, 1 (1948) and **3**, 101 (1948); Sin-Itiro Tomonaga, Phys. Rev. **74**, 224 (1948); Ito, Koba, and Tomonaga, Prog. Theor. Phys. **3**, 276 (1948); Z. Koba and S. Tomonaga, Prog. Theor. Phys. **3**, 290 (1948).

[3] Julian Schwinger, Phys. Rev. **73**, 416 (1948); **74**, 1439 (1948); **75**, 651 (1949).

[4] Richard P. Feynman, Phys. Rev. **74**, 1430 (1948).

[5] The idea of using standard electrodynamics as a starting point for an explicit calculation of the S matrix has been previously developed by E. C. G. Stueckelberg, Helv. Phys. Acta, **14**, 51 (1941); **17**, 3 (1944); **18**, 195 (1945); **19**, 242 (1946); Nature, **153**, 143 (1944); Phys. Soc. Cambridge Conference Report, 199 (1947); E. C. G. Stueckelberg and D. Rivier, Phys. Rev. **74**, 218 (1948). Stueckelberg anticipated several features of the Feynman theory, in particular the use of the function D_F (in Stueckelberg's notation D^C) to represent retarded (i.e., causally transmitted) electromagnetic interactions. For a review of the earlier part of this work, see Gregor Wentzel, Rev. Mod. Phys. **19**, 1 (1947). The use of mass renormalization in scattering problems is due to H. W. Lewis, Phys. Rev. **73**, 173 (1948).

has here been possible. There is no reason for attributing a more fundamental significance to the S matrix than to other observable quantities, nor was it Heisenberg's intention to do so. In the last section of this paper, tentative suggestions are made for a synthesis of the Hamiltonian and Heisenberg philosophies.

II. THE FEYNMAN THEORY AS AN S MATRIX THEORY

The S matrix was originally defined by Heisenberg in terms of the stationary solutions of a scattering problem. A typical stationary solution is represented by a time-independent wave function Ψ' which has a part representing ingoing waves which are asymptotically of the form Ψ_1', and a part representing outgoing waves which are asymptotically of the form Ψ_2'. The S matrix is the transformation operator S with the property that

$$\Psi_2' = S\Psi_1' \tag{1}$$

for every stationary state Ψ'.

In Section III of I an operator $U(\infty)$ was defined and stated to be identical with the S matrix. Since $U(\infty)$ was defined in terms of time-dependent wave functions, a little care is needed in making the identification. In fact, the equation

$$\Psi_2 = U(\infty)\Psi_1 \tag{2}$$

held, where Ψ_1 and Ψ_2 were the asymptotic forms of the ingoing and outgoing parts of a wave function Ψ in the Ψ-representation of I (the "interaction representation" of Schwinger[3]). Now the time-independent wave function Ψ' corresponds to a time-dependent wave function

$$\exp[(-i/\hbar)Et]\Psi'$$

in the Schrödinger representation, where E is the total energy of the state; and this corresponds to a wave function in the interaction representation

$$\Psi = \exp[(+i/\hbar)t(H_0 - E)]\Psi', \tag{3}$$

where H_0 is the total free particle Hamiltonian. However, the asymptotic parts of the wave function Ψ', both ingoing and outgoing, represent freely traveling particles of total energy E, and are therefore eigenfunctions of H_0 with eigenvalue E. This implies, in virtue of (3), that the asymptotic parts Ψ_1 and Ψ_2 of Ψ are actually time-independent and equal, respectively, to Ψ_1' and Ψ_2'. Thus (1) and (2) are identical, and $U(\infty)$ is indeed the S matrix. Incidentally, $U(\infty)$ is also the "invariant collision operator" defined by Schwinger.[3]

There is a series expansion of $U(\infty)$ analogous to (32) of I, namely,

$$U(\infty) = \sum_{n=0}^{\infty} \left(\frac{-i}{\hbar c}\right)^n \frac{1}{n!} \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_n \times P(H_1(x_1), \cdots, H_1(x_n)). \tag{4}$$

Here the P notation is as defined in Section V of I, and

$$H_1(x) = H^I(x) + H^e(x) \tag{5}$$

is the sum of the interaction energies of the electron field with the photon field and with the external potentials. The Feynman radiation theory provides a set of rules for the calculation of matrix elements of (4), between states composed of any number of ingoing and outgoing free particles. Also, quantities contained in (4) are the only ones with which the Feynman rules can deal directly. The Feynman theory is thus correctly characterized as an S matrix theory.

One particular way to analyze $U(\infty)$ is to use (5) to expand (4) in a series of terms of ascending order in H^e. Substitution from (5) into (4) gives

$$U(\infty) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left(\frac{-i}{\hbar c}\right)^{m+n} \frac{1}{m!n!} \int_{-\infty}^{\infty} dx_1 \cdots \times \int_{-\infty}^{\infty} dx_{m+n} P(H^e(x_1), \cdots, H^e(x_m), \times H^I(x_{m+1}), \cdots, H^I(x_{m+n})). \tag{6}$$

In this double series, the term of zero order in H^e is $S(\infty)$, given by (32) of I. The term of first order is

$$U_1 = (-i/hc) \int_{-\infty}^{\infty} H_F(x)dx, \tag{7}$$

where H_F is given by (31) of I. Clearly, $S(\infty)$ is the S matrix representing scattering of electrons and photons by each other in the absence of an external potential; U_1 is the S matrix representing the additional scattering produced by an external potential, when the external potential is treated in the first Born approximation; higher terms of the series (6) would correspond to treating the external potential in the second or higher Born approximation. The operator H_F played a prominent part in I, where it was in no way connected with a Born approximation; however, it was there introduced in a somewhat unnatural manner, and its physical meaning is made clearer by its appearance in (7). In fact, H_F may be defined by the statement that

$$(-i/\hbar)(\delta t)(\delta\omega)H_F(x)$$

is the contribution to the S matrix that would be produced by an external potential of strength H^e, acting for a small duration δt and over a small volume $\delta\omega$ in the neighborhood of the space-time point x.

The remainder of this section will be occupied with a statement of the Feynman rules for evaluating $U(\infty)$. Proofs will not be given, because the rules are only trivial generalizations of the rules

which were given in I for the evaluation of matrix elements of H_F corresponding to one-electron transitions.

In evaluating $U(\infty)$ we shall not make any distinction between the external and radiative parts of the electromagnetic field; this is physically reasonable since it is to some extent a matter of convention how much of the field in a given situation is to be regarded as "external." The interaction energy occurring in (4) is then

$$H_1(x) = -ieA_\mu(x)\bar{\psi}(x)\gamma_\mu\psi(x) - \delta mc^2\bar{\psi}(x)\psi(x), \quad (8)$$

where A_μ is the total electromagnetic field, and the term in δm is included in order to allow for the fact that the interaction representation is defined in terms of the total mass of an electron including its "electromagnetic mass" δm (see Section IV of I). The first step in the evaluation of $U(\infty)$ is to substitute from (8) into (4), writing out in full the suffixes of the operators $\bar{\psi}_\alpha$, ψ_β which are concealed in the matrix product notation of (8). After such a substitution, (4) becomes

$$U(\infty) = \sum_{n=0}^{\infty} J_n, \quad (9)$$

where J_n is an n-fold integral with an integrand which is a polynomial in $\bar{\psi}_\alpha$, ψ_β and A_μ operators.

The most general matrix element of J_n is obtained by allowing some of the $\bar{\psi}_\alpha$, ψ_β and A_μ operators to annihilate particles in the initial state, some to create particles in the final state, while others are associated in pairs to perform a successive creation and annihilation of intermediate particles. The operators which are not associated in pairs, and which are available for the real creation and annihilation of particles, are called "free"; a particular type of matrix element of J_n is specified by enumerating which of the operators in the integrand are to be free and which are to be associated in pairs. As described more fully in Section VII of I, each type of matrix element of J_n is uniquely represented by a "graph" G consisting of n points (bearing the labels $x_1, \cdots, x_n$) and various lines terminating at these points.

The relation between a type of matrix element of J_n and its graph G is as follows. For every associated pair of operators $(\bar{\psi}(x), \psi(y))$, there is a directed line (electron line) joining x to y in G. For every associated pair of operators $(A(x), A(y))$, there is an undirected line (photon line) joining x and y in G. For every free operator $\bar{\psi}(x)$, there is a directed line in G leading from x to the edge of the diagram. For every free operator $\psi(x)$, there is a directed line in G leading to x from the edge of the diagram. For every free operator $A(x)$, there is an undirected line in G leading from x to the edge of the diagram. Finally, for a particular type of matrix element of J_n it is specified that at each point x_i either the part of $H_1(x_i)$ containing $A_\mu(x_i)$ or the part containing δm is operating; correspondingly, at each vertex x_i of G there are either two electron lines (one ingoing and one outgoing) and one photon line, or else two electron lines only. Lines joining one point to itself are always forbidden.

In every graph G, the electron lines form a finite number m of open polygonal arcs with ends at the edge of the diagram, and perhaps in addition a number l of closed polygonal loops. The corresponding type of matrix element of J_n has m free operators $\bar{\psi}$ and m free operators ψ; the two end segments of any one open arc correspond to two free operators, one $\bar{\psi}$ and one ψ, which will be called a "free pair." The matrix elements of J_n are now to be calculated by means of an operator $J(G)$, which is defined for each graph G of n vertices, and which is obtained from J_n by making the following five alterations.

First, at each point x_i, $H_1(x_i)$ is to be replaced by either the first or the second term on the right of (8), as indicated by the presence or absence of a photon line at the vertex x_i of G. Second, for every electron line joining a vertex x to a vertex y in G, two operators $\bar{\psi}_\alpha(x)$ and $\psi_\beta(y)$ in J_n, regardless of their positions, are to be replaced by the function

$$\tfrac{1}{2}S_{F\beta\alpha}(x-y), \quad (10)$$

as defined by (44) and (45) of I. Third, for every photon line joining two vertices x and y of G, two operators $A_\mu(x)$ and $A_\nu(y)$ in J_n, regardless of their positions, are to be replaced by the function

$$\tfrac{1}{2}\hbar c\delta_{\mu\nu}D_F(x-y), \quad (11)$$

defined by (42) of I. Fourth, all free operators in J_n are to be left unaltered, but the ordering by the P notation is to be dropped, and the order of the free $\bar{\psi}$ and ψ operators is to be arranged so that the two members of each free pair stand consecutively and in the order $\bar{\psi}\psi$; the order of the free pairs among themselves, and of all free A_μ operators, is left arbitrary. Fifth, the whole expression J_n is to be multiplied by

$$(-1)^{n-l-m}. \quad (12)$$

The Feynman rules for the evaluation of $U(\infty)$ are essentially contained in the above definition of the operators $J(G)$. To each value of n correspond only a finite number of graphs G, and all possible matrix elements of $U(\infty)$ are obtained by substituting into (9) for each J_n the sum of all the corresponding $J(G)$. It is necessary only to specify how the matrix element of a given $J(G)$ corresponding to a given scattering process may be written down.

The matrix element of $J(G)$ for a given process may be obtained, broadly speaking, by replacing

each free operator in $J(G)$ by the wave function of the particle which it is supposed to create or annihilate. More specifically, each free $\bar{\psi}$ operator may either create an electron in the final state or annihilate a positron in the initial state, and the reverse processes are performed by a free ψ operator. Therefore, for a transition from a state involving A electrons and B positrons to a state involving C electrons and D positrons, only operators $J(G)$ containing $(A+D)=(B+C)$ free pairs contribute matrix elements. For each such $J(G)$, the $(A+D)$ free ψ operators are to be replaced in all possible combinations by the A initial electron wave functions and the D final positron wave functions, and the $(B+C)$ free $\bar{\psi}$ operators are to be similarly replaced by the initial positron and final electron wave functions, and the results of all such replacements added together, taking account of the antisymmetry of the total wave functions of the system in the individual particle wave functions. In the case of the free A_μ operators, the situation is rather different, since each such operator may either create a photon in the final state, or annihilate a photon in the initial state, or represent merely the external potential. Therefore, for a transition from a state with A photons to a state with B photons, any $J(G)$ with not less than $(A+B)$ free A_μ operators may give a matrix element. If the number of free A_μ operators in $J(G)$ is $(A+B+C)$, these operators are to be replaced in all possible combinations by the $(A+B)$ suitably normalized potentials corresponding to the initial and final photon states, and by the external potential taken C times, and the results of all such replacements added together, taking account now of the symmetry of the total wave functions in the individual photon states.

In practice cases are seldom likely to arise of scattering problems in which more than two similar particles are involved. The replacement of the free operators in $J(G)$ by wave functions can usually be carried out by inspection, and the enumeration of matrix-elements of $U(\infty)$ is practically complete as soon as the operators $J(G)$ have been written down.

The above rules for the calculation of $U(\infty)$ describe the state of affairs before any attempt has been made to identify and remove the various divergent parts of the expressions. In particular, contributions are included from all graphs G, even those which yield nothing but self-energy effects. For this reason, the rules here formulated are superficially different from those given for the one-electron problem in Section IX of I, which described the state of affairs after many divergencies had been removed. Needless to say, the rules are not complete until instructions have been supplied for the removal of all infinite quantities from the theory; in Sections V–VII of this paper it will be shown how the formal structure of the S matrix makes such a complete removal of infinities appear attainable.

Another essential limitation is introduced into the S matrix theory by the use of the expansion (4). All quantities discussed in this paper are expansions of this kind, in which it is assumed that not only the radiation interaction but also the external potential is small enough to be treated as a perturbation. It is well known that an expansion in powers of the external potential does not give a satisfactory approximation, either in problems involving bound states or in scattering problems at low energies. In particular, whenever a scattering problem allows the possibility of one of the incident particles being captured into a bound state, the capture process will not be represented in $U(\infty)$, since the initial and final states for processes described by $U(\infty)$ are always free-particle states. It is the expansion in powers of the external potential which breaks down when such a capture process is possible. Therefore it must be emphasized that the perturbation theory of this paper is applicable only to a restricted class of problems, and that in other situations the Schwinger theory will have to be used in its original form.

III. THE S MATRIX IN MOMENTUM SPACE

Both for practical applications to specific problems, and for general theoretical discussion, it is convenient to express the S matrix $U(\infty)$ in terms of momentum variables. For this purpose, it is enough to consider an expression which will be denoted by M, and which is a typical example of the units out of which all matrix elements of $U(\infty)$ are built up. A particular integer n and a particular graph G of n vertices being supposed fixed, the operator $J(G)$ is constructed as in the previous section, and M is defined as the number obtained by substituting for each of the free operators in $J(G)$ one particular free-particle wave function. More specifically, for each free operator $\psi(x)$ in $J(G)$ there is substituted

$$\psi(k)e^{ik_\mu x_\mu}, \tag{13}$$

where k_μ is some constant 4 vector representing the momentum and energy of an electron, or minus the momentum and energy of a positron, and where $\psi(k)$ is a constant spinor. For each free operator $\bar{\psi}(x)$ there is substituted

$$\bar{\psi}(k')e^{-ik_\mu' x_\mu}, \tag{14}$$

where $\bar{\psi}(k')$ is again a constant spinor. For each free operator $A_\mu(x)$ there is substituted

$$A_\mu(k'')e^{ik_\mu'' x_\mu}, \tag{15}$$

where $A_\mu(k'')$ is a constant 4 vector which may

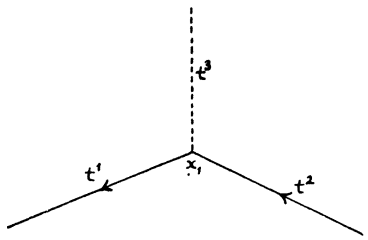

FIG. 1.

represent the polarization vector of a quantum whose momentum-energy 4-vector is either plus or minus k_μ''; alternatively, $A_\mu(k'')$ may represent the Fourier component of the external potential with a particular wave number and frequency specified by the 4 vector k''. There is no loss of generality in splitting up the external potential into Fourier components of the form (15). When the substitutions (13), (14), (15) are made in $J(G)$, the expression M which is obtained is still an n-fold integral over the whole of space-time, and in addition depends parametrically upon E constant 4 vectors in momentum-space, where E is the number of free operators in $J(G)$.

The graph G will contain E external lines, i.e., lines with one end at a vertex and the other end at the edge of the diagram. To each of these external lines corresponds one constant 4 vector, which may be denoted by $k_\mu{}^i$, $i=1, \cdots, E$, and one constant spinor or polarization vector appearing in M, either $\psi(k^i)$ or $\bar{\psi}(k^i)$ or $A_\mu(k^i)$.

Suppose that G contains F internal lines, i.e., lines with both ends at vertices. To each of these lines corresponds a D_F or an S_F function in M, as specified by (11) or (10). These functions have been expressed by Feynman as 4-dimensional Fourier integrals of very simple form, namely

$$D_F(x)=\frac{1}{4\pi^3}\int e^{-ip_\mu x_\mu}\delta_+(p^2)dp, \tag{16}$$

$$S_F(x)=\frac{1}{4\pi^3}\int e^{-ip_\mu x_\mu}[+ip_\mu\gamma_\mu-\kappa_0] \times\delta_+(p^2+\kappa_0^2)dp, \tag{17}$$

where κ_0 is the electron reciprocal Compton wavelength,

$$p^2=p_\mu p_\mu=p_1{}^2+p_2{}^2+p_3{}^2-p_0{}^2, \tag{18}$$

and the δ_+ function is defined by

$$\delta_+(a)=\tfrac{1}{2}\delta(a)+\frac{1}{2\pi ia}=\frac{1}{2\pi}\int_0^\infty e^{-iaz}dz. \tag{19}$$

Substituting from (16) and (17) into M will introduce an F-fold integral over momentum space. Corresponding to each internal line of G, there will appear in M a 4 vector variable of integration, which may be denoted by $p_\mu{}^i$, $i=1, \cdots, F$. However, after this substitution is made, the space-time variables $x_1, \cdots, x_n$ occur in M only in the exponential factors, and the integration over these variables can be performed. The result of the integration over x_j is to give

$$(2\pi)^4\delta(q_j), \tag{20}$$

where the δ represents a simple 4-dimensional Dirac δ-function, and q_j is a 4 vector formed by taking an algebraic sum of the k^i and p^i 4 vectors corresponding to those lines of G which meet at x_j. The factor (20) in the integrand of M expresses the conservation of energy and momentum in the interaction occurring at the point x_j.

The transformation of M into terms of momentum variables is now complete. To summarize the results, M now appears as an F-fold integral over the variable 4 vectors $p_\mu{}^i$ in momentum space. In the integrand there appear, besides numerical factors;

(i) a constant spinor or polarization-vector, $\psi(k^i)$ or $\bar{\psi}(k^i)$ or $A_\mu(k^i)$, corresponding to each external line of G;

(ii) a factor

$$D_F(p^i)=\delta_+((p^i)^2) \tag{21}$$

corresponding to each internal photon line of G;

(iii) a factor

$$S_F(p^i)=[+ip_\mu{}^i\gamma_\mu-\kappa_0]\delta_+((p^i)^2+\kappa_0^2) \tag{22}$$

corresponding to each internal electron line of G;

(iv) a factor

$$\delta(q_j) \tag{23}$$

corresponding to each vertex of G;

(v) a γ_μ operator, surviving from Eq. (8), corresponding to each vertex of G at which there is a photon line.

The important feature of the above analysis is that all the constituents of M are now localized and associated with individual lines and vertices in the graph G. It therefore becomes possible in an unambiguous manner to speak of "adding" or "subtracting" certain groups of factors in M, when G is modified by the addition or subtraction of certain lines and vertices. As an example of this method of analysis, we shall briefly discuss the treatment in the S matrix formalism of the "Lamb shift" and associated phenomena.

Suppose that a graph G, of any degree of complication, has a vertex x_1 at which two electron lines and a photon line meet. These three lines may be either internal or external, and the momentum 4 vectors associated with them in M may be either p^i or k^i; these 4 vectors are denoted by t^1, t^2, t^3 as indicated in Fig. 1. The factors in the integrand of

M arising from the vertex x_1 are

$$-ie\gamma_\mu(2\pi)^4\delta(t^1-t^2-t^3), \tag{24}$$

the two spinor indices of the γ_μ being available for matrix multiplication on both sides with the factors in M arising from the two electron lines at x_1.

Now suppose that G' is a graph identical with G, except that in the neighborhood of x_1 it is modified by the addition of two new vertices and three new lines, as indicated in Fig. 2. With the three new lines, which are all internal, are associated three 4 vector variables p^1, p^2, p^3, which occur as variables of integration in the expression M' formed from G' as M is from G. It can be proved, in view of Eqs. (21), (22), (23), that M' may be obtained from M simply by replacing the factor (24) in M by the expression

$$\begin{aligned}-\frac{ie^3}{\hbar c}(2\pi)^3\int\int\int dp^1dp^2dp^3\\ \delta(t^1-p^2+p^3)\delta(p^2-p^1-t^3)\delta(p^1-p^3-t^2)\\ \gamma_\lambda(+ip_\rho{}^2\gamma_\rho-\kappa_0)\gamma_\mu(+ip_\sigma{}^1\gamma_\sigma-\kappa_0)\gamma_\lambda\\ \delta_+((p^2)^2+\kappa_0{}^2)\delta_+((p^1)^2+\kappa_0{}^2)\delta_+((p^3)^2).\end{aligned} \tag{25}$$

(The factorial coefficients appearing in (4) are just compensated by the fact that the two new vertices of G' may be labelled x_i, x_j in $(n+1)(n+2)$ ways, where n is the number of vertices in G.) In (25), two of the 4-dimensional δ-functions can be eliminated at once by integration over p^1 and p^2, and the third then reduces to the δ-function occurring in (24). Therefore M' can be obtained from M by replacing the operator γ_μ in (24) by an operator

$$\begin{aligned}L_\mu=L_\mu(t^1,t^2)\\ =2\alpha\int dp[\gamma_\lambda(+i(p_\rho+t_\rho{}^1)\gamma_\rho-\kappa_0)\gamma_\mu\\ \times(+i(p_\sigma+t_\sigma{}^2)\gamma_\sigma-\kappa_0)\gamma_\lambda]\\ \times\delta_+((p+t^1)^2+\kappa_0{}^2)\\ \times\delta_+((p+t^2)^2+\kappa_0{}^2)\delta_+(p^2).\end{aligned} \tag{26}$$

Here α is the fine-structure constant, $(e^2/4\pi\hbar c)$ in Heaviside units. The operator L_μ can without great difficulty be calculated explicitly as a function of the 4 vectors t^1 and t^2, by methods developed by Feynman.

In the special case when Fig. 1 represents the graph G in its entirety, M is a matrix element for the scattering of a single electron by an external potential. Figure 2 then represents G' in its entirety, and M' is a second-order radiative correction to the scattering of the electron. In this case then the operator L_μ gives rise to what may be called "Lamb shift and associated phenomena." However, the above analysis applies equally to an expression M which may occur anywhere among the matrix elements of $U(\infty)$, and may represent any physical process whatever involving electrons, positrons and photons. There will always appear in $U(\infty)$, together with M, terms M' representing second-order radiative corrections to the same process; one term M' arises from each vertex of G at which a photon line ends; and M' is always to be obtained from M by substituting for an operator γ_μ the same operator L_μ. Furthermore, some higher radiative corrections to M will be obtained by substituting L_μ for γ_μ independently at two or more of the vertices of G.

By a "vertex part" of any graph will be meant a connected part of the graph, consisting of vertices and internal lines only, which touches precisely two electron lines and one photon line belonging to the remainder of the graph. The central triangle of Fig. 2 is an example of such a part. In other words, a vertex part of a graph is a part which can be substituted for the single vertex of Fig. 1 and give a physically meaningful result. Now the argument, by which the replacement of Fig. 1 by Fig. 2 was shown to be equivalent to the replacement of γ_μ by L_μ, can be used also when a more complicated vertex part is substituted for the vertex in Fig. 1. If G is any graph with a vertex x_1 as shown in Fig. 1, and G' is obtained from G by substituting for x_1 any vertex part V, and if M and M' are elements of $U(\infty)$ formed analogously from G and G', then M' can be obtained from M by replacing an operator γ_μ by an operator

$$\Lambda_\mu=\Lambda_\mu(V,t^1,t^2), \tag{27}$$

dependent only on V and the 4 vectors t^1, t^2 and independent of G.

To summarize the results of this section, it has been shown that the S matrix formalism allows a wide variety of higher order radiative processes to be calculated in the form of operators in momentum space. Such operators appear as radiative corrections to the fundamental interaction between the photon and electron-positron fields, and need only to be calculated once to be applicable to the various special problems of electrodynamics.

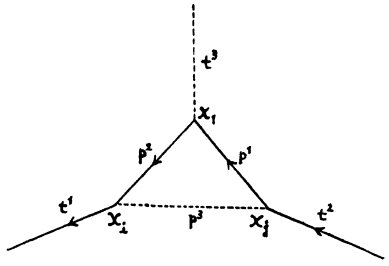

FIG. 2.

IV. FURTHER REDUCTION OF THE S MATRIX

It was shown in Section VII of I that, for the one-electron processes there considered, only connected graphs needed to be taken into account. In constructing the S matrix in general, this is no longer the case; disconnected graphs give matrix elements of $U(\infty)$ representing two or more collision processes occurring simultaneously among separate groups of particles, and such processes have physical reality. It is only permissable to omit a disconnected graph when one of its connected components is entirely lacking in external lines; such a component without external lines will give rise only to a constant multiplicative phase factor in every matrix element of $U(\infty)$ and is therefore devoid of physical significance.

On the other hand, the treatment in Section VII of I of graphs with "self-energy parts" applies almost without change to the general S matrix formalism. A "self-energy part" of a graph is a connected part, consisting of vertices and internal lines only, which can be inserted into the middle of a single line of a graph G so as to give a meaningful graph G'. In Fig. 3 is shown an example of such an insertion made in one of the lines of Fig. 1. Let M and M' be expressions derived in the manner of the previous section from the graphs G and G' of which parts are shown in Figs. 1 and 3. Suppose for definiteness that the line labelled t^1 is an internal line of G; then according to (22) it will contribute a factor $S_F(t^1)$ in M. By an argument similar to that leading to (26), it can be shown that M' may now be obtained from M by replacing $S_F(t^1)$ by

$$S_F(t^1)N(t^1)S_F(t^1)$$
$$= S_F(t^1)2\alpha\int dp[\gamma_\lambda(+i\gamma_\rho(p_\rho+t_\rho^1)-\kappa_0)\gamma_\lambda]$$
$$\times\delta_+((p+t^1)^2+\kappa_0^2)\delta_+(p^2)S_F(t^1). \quad (28)$$

In the same way, if G' were obtained from G by inserting in the t^1 line any self-energy part W, then M' would be obtained from M by replacing $S_F(t^1)$ by

$$S_F(t^1)\Sigma(W, t^1)S_F(t^1), \quad (29)$$

where Σ is an operator dependent only on W and t^1 and not on G. Moreover, if the t^1 line were an external line of G, then M' would be obtained from M by replacing a factor $\bar{\psi}(t^1)$ by

$$\bar{\psi}(t^1)\Sigma(W, t^1)S_F(t^1). \quad (30)$$

As a special case, W may consist of a single point; then at this point it is the term in δm of the interaction (8) which is operating, and Σ reduces to a constant,

$$\Sigma(W, t^1) = -2\pi i(\delta mc/\hbar) = -2\pi i\delta\kappa_0. \quad (31)$$

The operator $N(t^1)$ in (28) describes in a general way the second-order contribution to the electron self-energy and to the phenomenon called "vacuum polarization of the second kind" in Section VIII of I. The self-energy contribution is supposed to be cancelled by (31); the constant $\delta\kappa_0$ being a power series in α, the linear term only is required to cancel the self-energy part of (28), and the higher terms are available for the cancellation of self-energy effects from operators $\Sigma(W, t^1)$ of higher order. The S matrix formalism makes clear the important fact that, since the operators $\Sigma(W, t^1)$ are universal operators independent of the graph G, the electron self-energy effects will be formally cancelled by a constant $\delta\kappa_0$ independent of the physical situation in which the effects occur.

When a self-energy part W' is inserted into a photon line of a graph G, for example the line labelled t^3 in Fig. 1, then the modification produced in M may be again described by a function $\Pi(W', t^3)$ independent of G. Specifically, if the t^3 line in G is internal, M' is obtained from M by replacing a factor $D_F(t^3)$ by

$$D_F(t^3)\Pi(W', t^3)D_F(t^3). \quad (32)$$

If the t^3 line is external, the replacement is of a factor $A_\mu(t^3)$ by

$$A_\mu(t^3)\Pi(W', t^3)D_F(t^3). \quad (33)$$

In addition to terms of the form (33), there will appear terms such as

$$A_\nu(t^3)t_\nu^3 t_\mu^3\Pi'(W', t^3)D_F(t^3); \quad (34)$$

these however are zero in consequence of the gauge condition satisfied by A_ν. Similar terms in $t_\nu^3 t_\mu^3$ will also appear with the expression (32); in this case the extra terms can be shown to vanish in consequence of the equation of conservation of charge satisfied by the electron-positron field. The functions $\Pi(W', t^3)$ describe the phenomenon of photon self-energy and the "vacuum polarization of the first kind" of Section VIII of I. Following Schwinger, one does not explicitly subtract away the divergent photon self-energy effects from the $\Pi(W', t^3)$, but one asserts that these effects are zero as a consequence of the gauge invariance of electrodynamics.

In Section VII of I, it was shown how self-energy parts could be systematically eliminated from all

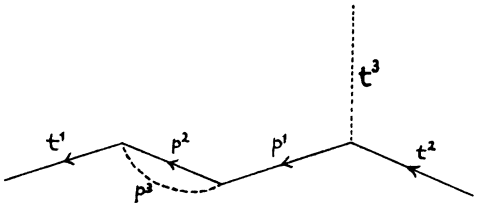

FIG. 3.

graphs, and their effects described by suitably modifying the functions D_F and S_F. The analysis was carried out in configuration space, and was confined to the one-electron problem. We are now in a position to extend this method to the whole S matrix formalism, working in momentum space, and furthermore to eliminate not only self-energy parts but also the "vertex parts" defined in the last section.

Every graph G has a uniquely defined "skeleton" G_0, which is the graph obtained by omitting all self-energy parts and vertex parts from G. A graph which is its own skeleton is called "irreducible;" all of its vertices will be of the kind depicted in Fig. 1. From every irreducible G_0, the G of which it is the skeleton can be built by inserting pieces in all possible ways at all vertices and lines of G_0; these G form a well-defined class Γ. The term "proper vertex part" must here be introduced, denoting a vertex part which is not divisible into two pieces joined only by a single line; thus a vertex part which is not proper is essentially redundant, being a proper vertex part plus one or more self-energy parts. The graphs of Γ are then accurately enumerated by inserting at some or all of the vertices of G_0 a proper vertex part, and in some or all of the lines a self-energy part, these insertions being made independently in all possible combinations.

Suppose that M is a constituent of a matrix element of $U(\infty)$, obtained from G_0 as described in Section III. Then every graph G in Γ will yield additional contributions to the same matrix element of $U(\infty)$; the sum of all such contributions, including M, is denoted by M_S. As a result of the analysis leading to (27), (29), and (32), and in view of the statistical independence of the insertions made at the different vertices and lines of G_0, the sum M_S will be obtained from M by the following substitutions. For every internal electron line of G_0, a factor $S_F(p^i)$ of M is replaced by

$$S_F'(p^i)=S_F(p^i)+S_F(p^i)\Sigma(p^i)S_F(p^i), \quad (35)$$

where $\Sigma(p^i)$ is the sum of the $\Sigma(W, p^i)$ over all electron self-energy parts W. For every internal photon line, a factor $D_F(p^i)$ is replaced by

$$D_F'(p^i)=D_F(p^i)+D_F(p^i)\Pi(p^i)D_F(p^i), \quad (36)$$

where $\Pi(p^i)$ is the sum of the $\Pi(W', p^i)$ over all photon self-energy parts W'. For every external line, a factor $\psi(k^i)$ or $\bar{\psi}(k^i)$ or $A_\mu(k^i)$ is replaced by

$$\begin{aligned}\psi'(k^i)&=S_F(k^i)\Sigma(k^i)\psi(k^i)+\psi(k^i),\\ \bar{\psi}'(k^i)&=\bar{\psi}(k^i)\Sigma(k^i)S_F(k^i)+\bar{\psi}(k^i),\\ A_\mu'(k^i)&=A_\mu(k^i)\Pi(k^i)D_F(k^i)+A_\mu(k^i),\end{aligned} \quad (37)$$

respectively. For every vertex of G_0, where the incident lines carry momentum variables as shown in Fig. 1, an operator γ_μ is replaced by

$$\Gamma_\mu(t^1, t^2)=\gamma_\mu+\Lambda_\mu(t^1, t^2), \quad (38)$$

where $\Lambda_\mu(t^1, t^2)$ is the sum of the $\Lambda_\mu(V, t^1, t^2)$ over all proper vertex parts V. The matrix elements of $U(\infty)$ will be correctly calculated, if one includes contributions only from irreducible graphs, after making in each contribution the replacements (35), (36), (37), (38).

To calculate the operators Λ_μ, Σ and Π, it is necessary to write down explicitly the integrals in momentum space, examples being (26) and (28), corresponding to every self-energy part W or proper vertex part V. When considering effects of higher order than the second, the parts W and V will themselves often be reducible, containing in their interior self-energy and vertex parts. It will again be convenient to omit such reducible V and W, and to include their effects by making the substitutions (35)–(38) in the integrals corresponding to irreducible V and W. In this way one obtains in general not explicit formulas, but integral equations, for Λ_μ, Σ and Π. For example,

$$\Lambda_\mu=\alpha I_\mu(\Lambda, \Sigma, \Pi) \quad (39)$$

where I_μ is an integral in which Λ_μ, Σ and Π occur explicitly. Fortunately, the appearance of α on the right side of (39) makes it easy to solve such equations by a process of successive substitution, the forms of Λ_μ, Σ, and Π being obtained correct to order α^n when values correct to order α^{n-1} are substituted into the integrals.

The functions D_F' and S_F' of (35) and (36) are the Fourier transforms of the corresponding functions in I. The interpretation of these functions in Section VIII of I can be extended in an obvious way to include the operator Γ_μ. Since $\bar{\psi}\gamma_\mu\psi$ is the charge-current 4 vector of an electron without radiative corrections, $\bar{\psi}\Gamma_\mu\psi$ may be interpreted as an "effective current" carried by an electron, including the effects of exchange interactions between the electron and the electron-positron field around it.

An additional reduction in the number of graphs effectively contributing to $U(\infty)$ is obtained from a theorem of Furry.[6] The theorem was shown by Feynman to be an elegant consequence of his theory. In any graph G, a "closed loop" is a closed electron polygon, at the vertices of which a number p of photon lines originate; the loop is called odd or even according to the parity of p. If G contains a closed loop, then there will be another graph $\bar{G}$ also contributing to $U(\infty)$, obtained from G by reversing the sense of the electron lines in the loop. Now if M and $\bar{M}$ are contributions from G and $\bar{G}$, $\bar{M}$ is derived from M by interchanging the roles of electron and positron states in each of the interactions at the vertices of the loop; such an interchange is called "charge conjugation." It was shown by Schwinger that his theory is invariant under charge

[6] Wendell H. Furry, Phys. Rev. 51, 125 (1937).

conjugation, provided that the sign of e is at the same time reversed (this is the well-known charge symmetry of the Dirac hole theory). It is clear from (8) that the constant e appears once in M for each of the p loop vertices at which there is a photon line; at the remaining vertices only the constant δm is involved, and δm is an even function of e. Therefore the principle of charge-symmetry implies

$$\bar{M}=(-1)^p M. \tag{40}$$

Taking p odd in (40) gives Furry's theorem; all contributions to $U(\infty)$ from graphs with one or more odd closed loops vanish identically.

By an "odd part" of a graph is meant any part, consisting only of vertices and internal lines, which touches no electron lines, and only an odd number of photon lines, belonging to the rest of the graph. The simplest type of odd part which can occur is a single odd closed loop. Conversely, it is easy to see that every odd part must include within itself at least one odd closed loop. Therefore, Furry's theorem allows all graphs with odd parts to be omitted from consideration in calculating $U(\infty)$.

V. INVESTIGATION OF DIVERGENCES IN THE S MATRIX

The δ_+ function defined by (19) has the property that, if b is real and $f(a)$ is any function analytic in the neighborhood of b, then

$$\int f(a)\delta_+(a-b)da=(1/2\pi i)\int f(a)(1/(a-b))da, \tag{41}$$

where the first integral is along a stretch of the real axis including b, and the second integral is along the same stretch of the real axis but with a small detour into the complex plane passing underneath b. In the matrix elements of $U(\infty)$ there appear integrals of the form

$$\int dpF(p)\delta_+(p_1^2+p_2^2+p_3^2-p_0^2+c^2), \tag{42}$$

integrated over all real values of p_1, p_2, p_3, p_0. By (41), one may write (42) in the form

$$\frac{1}{2\pi i}\int dp\frac{F(p)}{(p_1^2+p_2^2+p_3^2-p_0^2+c^2)}, \tag{43}$$

in which it is understood that the integration is along the real axis for the variables p_1, p_2, p_3, and for p_0 is along the real axis with two small detours, one passing above the point $+(p_1^2+p_2^2+p_3^2+c^2)^{\frac{1}{2}}$, and one passing below the point $-(p_1^2+p_2^2+p_3^2+c^2)^{\frac{1}{2}}$. To equate (42) with (43) is certainly correct, when $F(p)$ is analytic at the critical values of p_0. In practice one has to deal with integrals (42) in which $F(p)$ itself involves δ_+ functions (see for example (26) and (28)); in these cases it is legitimate to replace each δ_+ function by a reciprocal, making a separate detour in the p_0 integration for each pole in the integrand, provided that no two poles coincide. Thus every constituent part M of $U(\infty)$ can be written as an integral of a rational algebraic function of momentum variables, by using instead of (21) and (22)

$$D_F(p^i)=\frac{1}{2\pi i(p^i)^2}, \tag{44}$$

$$S_F(p^i)=\frac{(ip_\mu{}^i\gamma_\mu-\kappa_0)}{2\pi i((p^i)^2+\kappa_0^2)}. \tag{45}$$

This representation of D_F and S_F as rational functions in momentum-space has been developed and extensively used by Feynman (unpublished).

There may appear in M infinities of three distinct kinds. These are (i) singularities caused by the coincidence of two or more poles of the integrand, (ii) divergences at small momenta caused by a factor (44) in the integrand, (iii) divergences at large momenta due to insufficiently rapid decrease of the whole integrand at infinity.

In this paper no attempt will be made to explore the singularities of type (i). Such singularities occur, for example, when a many-particle scattering process may for special values of the particle momenta be divided into independent processes involving separate groups of particles. It is probable that all singularities of type (i) have a similarly clear physical meaning; these singularities have long been known in the form of vanishing energy denominators in ordinary perturbation theory, and have never caused any serious trouble.

A divergence of type (ii) is the so-called "infra-red catastrophe," and is well known to be caused by the failure of an expansion in powers of α to describe correctly the radiation of low momentum quanta. It would presumably be possible to eliminate this divergence from the theory by a suitable adaptation of the standard Bloch-Nordsieck[7] treatment; we shall not do this here. From a practical point of view, one may avoid the difficulty by arbitrarily writing instead of (44)

$$D_F(p^i)=\frac{1}{2\pi i((p^i)^2+\lambda^2)}, \tag{46}$$

where λ is some non-zero momentum, smaller than any of the quantum momenta which are significant in the particular process under discussion.[8]

[7] F. Bloch and A. Nordsieck, Phys. Rev. **52**, 54 (1937).

[8] The device of introducing λ in order to avoid infra-red divergences must be used with circumspection. Schwinger (unpublished) has shown that a long standing discrepancy between two alternative calculations of the Lamb shift was due to careless use of λ in one of them.

It is the divergences of type (iii) which have always been the main obstacle to the construction of a consistent quantum electrodynamics, and which it is the purpose of the present theory to eliminate. In the following pages, attention will be confined to type (iii) divergences; when the word "convergent" is used, the proviso "except for possible singularities of types (i) and (ii)" should always be understood.

A divergent M is called "primitive" if, whenever one of the momentum 4 vectors in its integrand is held fixed, the integration over the remaining variables is convergent. Correspondingly, a primitive divergent graph is a connected graph G giving rise to divergent M, but such that, if any internal line is removed and replaced by two external lines, the modified G gives convergent M. To analyze the divergences of the theory, it is sufficient to enumerate the primitive divergent M and G and to examine their properties.

Let G be a primitive divergent graph, with n vertices, E external and F internal lines. A corresponding M will be an integral over F variable p^i of a product of F factors (44) and (45) and n factors (23). Since G is connected, the δ-functions (23) in the integrand enable $(n-1)$ of the variables p^i to be expressed in terms of the remaining $(F-n+1)$ p^i and the constants k^i, leaving one δ-function involving the k^i only and expressing conservation of momentum and energy for the whole system. An example of such integration over the δ-functions was the derivation of (26) from (25). After this, the integrations in M may be arranged as follows; the fourth components of the $(F-n+1)$ independent p^i are written

$$p_4{}^i = ip_0{}^i = i\alpha\pi_0{}^i, \tag{47}$$

and the integration over α is performed first; subsequently, integration is carried out over the $3(F-n+1)$ independent $p_1{}^i$, $p_2{}^i$, $p_3{}^i$, and over the $(F-n)$ ratios of the $\pi_0{}^i$. M then has the form

$$M = \int dp_1{}^i dp_2{}^i dp_3{}^i d\pi_0{}^i \int_{-\infty}^{\infty} R\alpha^{F-n} d\alpha, \tag{48}$$

where R is a rational function of α, the denominator of which is a product of F factors

$$(p_1{}^i)^2+(p_2{}^i)^2+(p_3{}^i)^2+\mu^2-(\alpha\pi_0{}^i+c^i)^2. \tag{49}$$

Here the constants $\pi_0{}^i$, c^i are defined by the condition that

$$p_4{}^j = ip_0{}^j = i(\alpha\pi_0{}^j+c^j), \quad j=1, 2, \cdots, F. \tag{50}$$

Thus the c^i corresponding to the $(F-n+1)$ independent p^i are zero by (47), and the remainder are linear combinations of the k^i; also $(n-1)$ of the $\pi_0{}^i$ are linear combinations of the independent $\pi_0{}^i$ defined in (47). In view of (43), we take the integration variables in (48) to be real variables, with the exception of α which is to be integrated along a contour C deviating from the real axis at each of the $2F$ poles of R. As a general rule, C will detour above the real axis for $\alpha>0$, and below it for $\alpha<0$; the reverse will only occur at certain of the poles corresponding to denominators (49) for which

$$(p_1{}^i)^2+(p_2{}^i)^2+(p_3{}^i)^2+\mu^2 \leqslant (c^i)^2. \tag{51}$$

Such poles will be called "displaced." The integration over α alone will always be absolutely convergent. Therefore the contour C may be rotated in a counter-clockwise direction until it lies along the imaginary axis, and the value of M will be unchanged except for residues at the displaced poles.

Regarded as a function of the parameters k^i describing the incoming and outgoing particles, M will have a complicated behavior; the behavior will change abruptly whenever one of the c^i has a critical value for which (51) begins to be soluble for $p_1{}^i$, $p_2{}^i$, $p_3{}^i$, and a new displaced pole comes into existence. This behavior is explained by observing that displaced poles appear whenever there is sufficient energy available for one of the virtual particles involved in M to be actually emitted as a real particle. It is to be expected that the behavior of M should change when the process described by M begins to be in competition with other real processes. It is a feature of standard perturbation theory, that when a process A involves an intermediate state I which is variable over a continuous range, and in this range occurs a state II which is the final state of a competing process, then the matrix element for A involves an integral over I which has a singularity at the position II. In standard perturbation theory, this improper integral is always to be evaluated as a Cauchy principal value, and does not introduce any real divergence into the matrix element. In the theory of the present paper, the displaced poles give rise to similar improper integrals; these come under the heading of singularities of type (i) and will not be discussed further.

If $p_1{}^i$, $p_2{}^i$, $p_3{}^i$ satisfying (51) are held fixed, then the value of $p_4{}^i$ at the corresponding displaced pole is fixed by (50). The contribution to M from the displaced pole is just the expression obtained by holding the 4-vector p^i fixed in the original integral M, apart from bounded factors; since M is primitive divergent, this expression is convergent. The total contribution to M from the i'th displaced pole is the integral of this expression over the finite sphere (51) and is therefore finite. Strictly speaking, this argument requires not only the convergence of the expression, but uniform convergence in a finite region; however, it will be seen that the convergent integrals in this theory are convergent for large

momenta by virtue of a sufficient preponderance of large denominators, and convergence produced in this way will always be uniform in a finite region.

M is thus, apart from finite parts, equal to the integral M' obtained by replacing α by $i\alpha$ in (48) and (49). Alternatively, M' is obtained from the original integral M by substituting for each $p_0{}^i$

$$ip_4{}^i+(1-i)c^i, \tag{52}$$

and then treating the $4(F-n+1)$ independent $p_\mu{}^i$, $\mu=1, 2, 3, 4$, as ordinary real variables. In M' the denominators of the integrand take the form

$$(p_1{}^i)^2+(p_2{}^i)^2+(p_3{}^i)^2+\mu^2+(p_4{}^i-(1+i)c^i)^2, \tag{53}$$

and are uniformly large for large values of $p_\mu{}^i$. The convergence of M' can now be estimated simply by counting powers of $p_\mu{}^i$ in numerator and denominator of the integrand. Since M' is known to converge whenever one of the p^i is held fixed and integration is carried out over the others, the convergence of the whole expression is assured provided that

$$K=2F-F_e-4[F-n+1]\geqslant 1. \tag{54}$$

Here $2F$ is the degree of the denominator, and F_e that of the numerator, which is by (44) and (45) equal to the number of internal electron lines in G. Let E_e and E_p be the numbers of external electron and photon lines in G, and let n_s be the number of vertices without photon lines incident. It follows from the structure of G that

$$2F=3n-n_s-E_e-E_p,$$
$$F_e=n-\tfrac{1}{2}E_e,$$

and so the convergence condition (52) is

$$K=\tfrac{3}{2}E_e+E_p+n_s-4\geqslant 1. \tag{55}$$

This gives the vital information that the only possible primitive divergent graphs are those with $E_e=2$, $E_p=0, 1$, and with $E_e=0$, $E_p=1, 2, 3, 4$. Further, the cases $E_e=0$, $E_p=1, 3$, do not arise, since these give graphs with odd parts which were shown to be harmless in Section IV. It should be observed that the course of the argument has been "if E_e and E_p do not have certain small values, then the integral M is convergent at infinity;" there is no objection to changing the order of integrations in M as was done in (48), since the argument requires that this be done only in cases when M is, in fact, absolutely convergent.

The possible primitive divergent graphs that have been found are all of a kind familiar to physicists. The case $E_e=2$, $E_p=0$ describes self-energy effects of a single electron; $E_e=0$, $E_p=2$ self-energy effects of a single photon; $E_e=2$, $E_p=1$ the scattering of a single electron in an electromagnetic field; and $E_e=0$, $E_p=4$ the "scattering of light by light" or the mutual scattering of two photons. Further, (55) shows that the divergence will never be more than logarithmic in the third and fourth cases, more than linear in the first, or more than quadratic in the second. Thus it appears that, however far quantum electrodynamics is developed in the discussion of many-particle interactions and higher order phenomena, no essentially new kinds of divergence will be encountered. This gives strong support to the view that "subtraction physics," of the kind used by Schwinger and Feynman, will be enough to make quantum electrodynamics into a consistent theory.

VI. SEPARATION OF DIVERGENCES IN THE S MATRIX

First it will be shown that the "scattering of light by light" does not in fact introduce any divergence into the theory. The possible primitive divergent M in the case $E_e=0$, $E_p=4$ will be of the form

$$\delta(k^1+k^2+k^3+k^4)A_\lambda(k^1)A_\mu(k^2)A_\nu(k^3)A_\rho(k^4)I_{\lambda\mu\nu\rho}, \tag{56}$$

where $I_{\lambda\mu\nu\rho}$ is an integral of the type

$$\int R_{\lambda\mu\nu\rho}(k^1, k^2, k^3, k^4, p^i)dp^i, \tag{57}$$

at most logarithmically divergent, and R is a certain rational function of the constant k^i and the variable p^i. In any physical situation where, for example, the $A(k)$ are the potentials corresponding to particular incident and outgoing photons, there will appear in $U(\infty)$ a matrix element which is the sum of (56) and the 23 similar expressions obtained by permuting the suffixes of $I_{\lambda\mu\nu\rho}$ in all possible ways. It may therefore be supposed that at the start $R_{\lambda\mu\nu\rho}$ has been symmetrized by summation over all permutations of suffixes; (56) is then a sum of contributions from 24 or fewer (according to the degree of symmetry existing) graphs G.

If, under the sign of integration in (57), the value $R(0)$ of R for $k^1=k^2=k^3=k^4=0$ is subtracted from R, the integrand acquires one extra power of $|p_\mu{}^i|^{-1}$ for large $|p_\mu{}^i|$, and the integral becomes absolutely convergent at infinity. Therefore

$$I_{\lambda\mu\nu\rho}=I_{\lambda\mu\nu\rho}(0)+J_{\lambda\mu\nu\rho}, \tag{58}$$

where $I(0)$ is a possibly divergent integral independent of the k^i, and J is a convergent integral vanishing when all k^i's are zero. To interpret this result physically, it is convenient to write (56) again in terms of space-time variables; this gives

$$M=\int I_{\lambda\mu\nu\rho}(0)A_\lambda(x)A_\mu(x)A_\nu(x)A_\rho(x)dx+N, \tag{59}$$

where N is a convergent expression involving derivatives of the $A(x)$ with respect to space and

time. Now the first term in (59) is physically inadmissable; it is not gauge-invariant, and implies for example a scattering of light by an electric field depending on the absolute magnitude of the scalar potential, which has no physical meaning. Therefore $I(0)$ must vanish identically, and the whole expression (56) is convergent.

The fact that the scattering of light by light is finite in the lowest order in which it occurs has long been known.[9] It has also been verified by Feynman by direct calculation, using his own theory as described in this paper. The graphs which give rise to the lowest order scattering are shown in Fig. 4. It is found that the divergent parts of the corresponding M exactly cancel when the three contributions are added, or, what comes to the same thing, when the function $R_{\lambda\mu\nu\rho}$ is symmetrized. It is probable that the absence of divergence in the scattering of light by light is in all cases due to a similar cancellation, and it should not be difficult to prove this by calculation and thus avoid making an appeal to gauge-invariance.

The three remaining types of primitive divergent M are, in fact, divergent. However, these are just the expressions which have been studied in Sections III and IV and shown to be completely described by the operators Λ_μ, Σ, and Π. More specifically, when $E_e=2$, $E_p=0$, M will be of the form

$$\bar{\psi}(k^1)\Sigma(W, k^1)\psi(k^1), \tag{60}$$

where W is some electron self-energy part of a graph. When $E_e=0$, $E_p=2$, M will be

$$A_\mu(k^1)\Pi(W', k^1)A_\mu(k^1), \tag{61}$$

with W' some photon self-energy part. When $E_e=2$, $E_p=1$, M will be

$$\bar{\psi}(k^1)\Lambda_\mu(V, k^1, k^2)\psi(k^2)A_\mu(k^1-k^2), \tag{62}$$

with V some vertex part. Therefore, if some means can be found for isolating and removing the divergent parts from Λ_μ, Σ, and Π, the "irreducible" graphs defined in Section IV will not introduce any fresh divergences into the theory, and the rules of Section IV will lead to a divergence-free S matrix.

In considering Λ_μ, Σ, and Π in Section IV it was found convenient to divide vertex and self-energy parts themselves into the categories reducible and irreducible. An irreducible self-energy part W is required not only to have no vertex and self-energy parts inside itself; it is also required to be "proper," that is to say, it is not to be divisible into two pieces joined by a single line. In Section IV it was shown that to avoid redundancy the operator Λ_μ should be defined as a sum over proper vertex parts V only. By the same argument, in order to make (35), (36), (37) correct, it is essential to define Σ and Π as sums over both proper and improper self-energy parts. However, it is possible to define S_F' and D_F' in terms of proper self-energy parts only, at the cost of replacing the explicit definitions (35), (36) by implicit definitions. Let $\Sigma^*(p^i)$ be defined as the sum of the $\Sigma(W, p^i)$ over proper electron self-energy parts W, and let $\Pi^*(p^i)$ be defined similarly. Every W is either proper, or else it is a proper W joined by a single electron line to another self-energy part which may be proper or improper. Therefore, using (35), S_F' may be expressed in the two equivalent forms

$$\begin{aligned} S_F'(p^i) &= S_F(p^i)+S_F(p^i)\Sigma^*(p^i)S_F'(p^i) \\ &= S_F(p^i)+S_F'(p^i)\Sigma^*(p^i)S_F(p^i). \end{aligned} \tag{63}$$

Similarly,

$$\begin{aligned} D_F'(p^i) &= D_F(p^i)+D_F(p^i)\Pi^*(p^i)D_F'(p^i) \\ &= D_F(p^i)+D_F'(p^i)\Pi^*(p^i)D_F(p^i). \end{aligned} \tag{64}$$

It is sometimes convenient to work with the Σ and Π in the starred form, and sometimes in the unstarred form.

Consider the contribution $\Sigma(W, t^1)$ to the operator Σ^*, arising from an electron self-energy part W. It is supposed that W is irreducible, and the effects of possible insertions of self-energy and vertex parts inside W are for the time being neglected. Also it is supposed that W is not a single point, of which the contribution is given by (31). Then W has an even number $2l$ of vertices, at each of which a photon line is incident; and $\Sigma(W, t^1)$ will be of the form

$$e^{2l}\int R(t^1, p^i)dp^i, \tag{65}$$

where R is a certain rational function of the t^1 and p^i, and the integral is at most linearly divergent. The integrand in (65) is now written in the form

$$\begin{aligned} R(t^1, p^i) &= R(0, p^i) \\ &\quad + t_\mu^1\left(\frac{\partial R}{\partial t_\mu^1}(0, p^i)\right)+R_c(t^1, p^i), \end{aligned} \tag{66}$$

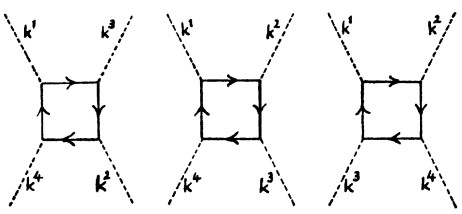

FIG. 4.

[9] H. Euler and B. Kockel, Naturwiss. **23**, 246 (1935); H. Euler, Ann. d. Phys. **26**, 398 (1936). In these early calculations of the scattering of light by light, the theory used is the Heisenberg electrodynamics, in which certain singularities are eliminated at the start by a procedure involving non-diagonal elements of the Dirac density matrix. In Feynman's calculation, on the other hand, a finite result is obtained without subtractions of any kind.

and for large values of the $|p_\mu{}^i|$ the remainder term R_c will tend to zero more rapidly by two powers of $|p_\mu{}^i|$ than R. Therefore, in complete analogy with (58),

$$\Sigma(W, t^1)=e^{2l}[A+B_\mu t_\mu{}^1+\Sigma_c(W, t^1)], \quad (67)$$

where A and B_μ are constant divergent operators, and $\Sigma_c(W, t^1)$ is defined by a covariant and absolutely convergent integral. $\Sigma_c(W, t^1)$ must, on grounds of covariance, be of the form

$$R_1((t^1)^2)+R_2((t^1)^2)t_\mu{}^1\gamma_\mu \quad (68)$$

with R_1 and R_2 particular functions of $(t^1)^2$; for the same reason, B_μ must be of the form $B\gamma_\mu$ with B a certain divergent integral. Now if t^1 happens to be the momentum-energy 4 vector of a free electron,

$$(t^1)^2=-\kappa_0{}^2, \quad t_\mu{}^1\gamma_\mu=i\kappa_0. \quad (69)$$

It is convenient to write

$$\Sigma_c(W, t^1)=A'+B'(t_\mu{}^1\gamma_\mu-i\kappa_0)+(t_\mu{}^1\gamma_\mu-i\kappa_0)S(W, t^1), \quad (70)$$

where $S(W, t^1)$ is zero for t^1 satisfying (69), and to include the first two terms in the constants A and B of (67); since all terms in (70) are finite, the separation of $S(W, t^1)$ is without ambiguity. Thus an equation of the form (67) is obtained, with

$$\Sigma_c(W, t^1)=(t_\mu{}^1\gamma_\mu-i\kappa_0)S(W, t^1). \quad (71)$$

Summing (67) over all irreducible W and including (31), gives for the operator Σ^*,

$$\Sigma^*(t^1)=A-2\pi i\delta\kappa_0+B(t_\mu{}^1\gamma_\mu-i\kappa_0)+(t_\mu{}^1\gamma_\mu-i\kappa_0)S_c(t^1). \quad (72)$$

Hence by (63) and (45)

$$S_F'(t^1)=(A-2\pi i\delta\kappa_0)S_F(t^1)S_F'(t^1)+\frac{1}{2\pi}BS_F'(t^1)+S_F(t^1)+\frac{1}{2\pi}S_c(t^1)S_F'(t^1). \quad (73)$$

In (72) and (73), A and B are infinite constants, and S_c a divergence-free operator which is zero when (69) holds; A, B, and S_c are power series in e starting with a term in e^2. In (72) and (73), however, effects of higher order corrections to the $\Sigma(W, t^1)$ themselves are not yet included.

A similar separation of divergent parts may be made for the $\Pi(W', t^1)$, when W' is an irreducible photon self-energy part. The integral (65) may now be quadratically divergent, and so it is necessary to use instead of (66)

$$R(t^1, p^i)=R(0, p^i)+t_\mu{}^1\left(\frac{\partial R}{\partial t_\mu{}^1}(0, p^i)\right)+\tfrac{1}{2}t_\mu{}^1t_\nu{}^1\left(\frac{\partial^2 R}{\partial t_\mu{}^1\partial t_\nu{}^1}(0, p^i)\right)+R_c(t^1, p^i),$$

and derive instead of (67)

$$\Pi(W', t^1)=e^{2l}[A+B_\mu t_\mu{}^1+C_{\mu\nu}t_\mu{}^1t_\nu{}^1+\Pi_c(W', t^1)]. \quad (74)$$

The A, B_μ, $C_{\mu\nu}$ are absolute constant numbers (not Dirac operators) and therefore covariance requires that $B_\mu=0$, $C_{\mu\nu}=C\delta_{\mu\nu}$. $\Pi_c(W', t^1)$ is defined by an absolutely convergent integral, and will be an invariant function of $(t^1)^2$ of a form

$$\Pi_c(W', t^1)=(t^1)^2D(W', t^1), \quad (75)$$

where $D(W', t^1)$ is zero for t^1 satisfying

$$(t^1)^2=0 \quad (76)$$

instead of (69). Summing (74) over all irreducible W's will give

$$\Pi^*(t^1)=A'+C(t^1)^2+(t^1)^2D_c(t^1), \quad (77)$$

and hence by (64) and (44)

$$D_F'(t^1)=A'D_F(t^1)D_F'(t^1)+\frac{1}{2\pi i}CD_F'(t^1)+D_F(t^1)+\frac{1}{2\pi i}D_c(t^1)D_F'(t^1). \quad (78)$$

In (77) and (78), D_c is zero for t^1 satisfying (76), and is divergence free.

The constant A' in (77) is the quadratically divergent photon self-energy. It will give rise to matrix elements in $U(\infty)$ of the form

$$M=A'\int A_\mu(x)A_\mu(x)dx, \quad (79)$$

which are non-gauge invariant and inadmissable. Such matrix elements must be eliminated from the theory, as the first term of (59) was eliminated, by the statement that A' is zero. The verification of this statement, by direct calculation of the lowest order contribution to A', has been given by Schwinger.[3,10]

The separation of the divergent part of Λ_μ again follows the lines laid down for Σ^*. Since the integral analogous to (65) is now only logarithmically divergent, no derivative term is required in (66), and the analog of (67) is

$$\Lambda_\mu(V, t^1, t^2)=e^{2l}[L_\mu+\Lambda_{\mu c}(V, t^1, t^2)], \quad (80)$$

where L_μ is a constant divergent operator, and $\Lambda_{\mu c}$ is convergent and zero for $t^1=t^2=0$. In (80), L_μ can only be of the form $L\gamma_\mu$. Also, if $t^1=t^2$ and t^1 satisfies (69), $\Lambda_{\mu c}$ will reduce to a finite multiple of γ_μ which can be included in the term $L\gamma_\mu$. Therefore it may be supposed that $\Lambda_{\mu c}$ in (80) is zero not for $t^1=t^2=0$ but for $t^1=t^2$ satisfying (69). The meaning of this

[10] Gregor Wentzel, Phys. Rev. **74**, 1070 (1948), presents the case against Schwinger's treatment of the photon self-energy.

physically is that $\Lambda_{\mu c}$ now gives zero contribution to the energy of a single electron in a constant electromagnetic potential, so that the whole measured static charge on an electron is included in the term $L\gamma_\mu$. Summing (80) over all irreducible vertex parts V, and using (38),

$$\Lambda_\mu(t^1, t^2) = L\gamma_\mu + \Lambda_{\mu c}(t^1, t^2), \tag{81}$$

$$\Gamma_\mu(t^1, t^2) = (1+L)\gamma_\mu + \Lambda_{\mu c}(t^1, t^2). \tag{82}$$

In (81) and (82), effects of higher order corrections to the $\Lambda_\mu(V, t^1, t^2)$ are again not yet included. Formally, (82) differs from (73) and (78) in not containing the unknown operator Γ_μ on both sides of the equation.

VII. REMOVAL OF DIVERGENCES FROM THE S MATRIX

The task remaining is to complete the formulas (73), (78), and (82), which show how the infinite parts can be separated from the operators Γ_μ, S_F', and D_F', and to include the corrections introduced into these operators by the radiative reactions which they themselves describe. In other words, we have to include radiative corrections to radiative corrections, and renormalizations of renormalizations, and so on *ad infinitum*. This task is not so formidable as it appears.

First, we observe that Λ_μ, Σ^*, and Π^* are defined by integral equations of the form (39), which will be referred to in the following pages as "the integral equations." More specifically, consider the contribution $\Lambda_\mu(V, t^1, t^2)$ to Λ_μ represented by (80), arising from a vertex part V with $(2l+1)$ vertices, l photon lines, and $2l$ electron lines. This contribution is defined by an integral analogous to (65), with an integrand which is a product of $(2l+1)$ operators γ_μ, l functions D_F, and $2l$ operators S_F. The exact $\Lambda_\mu(V, t^1, t^2)$ is to be obtained by replacing these factors, respectively, by Γ_μ, D_F', S_F', as described in Section IV. Now suppose that S_F' in the integrand is represented, to order e^{2n} say, by the sum of S_F and of a finite number of finite products of S_F with absolutely convergent operators $S(\bar{W}, t^1)$ such as appear in (71); similarly, let D_F' be represented by D_F plus a finite sum of finite products of D_F with functions $D(\bar{W}', t^1)$ appearing in (75); and let Γ_μ be represented by the sum of γ_μ and of a finite set of $\Lambda_{\mu c}(\bar{V}, t^1, t^2)$ from (80). Then the integral $\Lambda_\mu(V, t^1, t^2)$ will be determined to order e^{2n+2l}; and since the operators $S(\bar{W}, t^1)$, $D(\bar{W}', t^1)$, $\Lambda_{\mu c}(\bar{V}, t^1, t^2)$ always have a sufficiency of denominators for convergence, the theory of Section V can be applied to prove that this $\Lambda_\mu(V, t^1, t^2)$ will not be more than logarithmically divergent. Therefore the new $\Lambda_\mu(V, t^1, t^2)$ can be again separated into the form (80). The sum of these $\Lambda_\mu(V, t^1, t^2)$ will then be a $\Lambda_\mu(t^1, t^2)$ of the form (81), with constant L and convergent operator $\Lambda_{\mu c}$ determined to order e^{2n+2}. Thus (82) provides a new expression for Γ_μ, determined to order e^{2n+2}.

FIG. 5.

The above procedure describes the general method for separating out the finite part from the contribution to Γ_μ arising from a reducible vertex part V_R. First, V_R is broken down into an irreducible vertex part V plus various inserted parts $\bar{W}$, $\bar{W}'$, $\bar{V}$; the contribution to Γ_μ from V_R is an integral $M(V_R)$ which is not only divergent as a whole, but also diverges when integrated over the variables belonging to one of the insertions $\bar{W}$, $\bar{W}'$, $\bar{V}$, the remaining variables being held fixed. The divergences are to be removed from $M(V_R)$ in succession, beginning with those arising from the inserted parts, and ending with those arising from V itself. This successive removal of divergences is a well-defined procedure, because any two of the insertions made in V are either completely non-overlapping or else arranged so that one is completely contained in the other.

In calculating the contribution to Σ^* or Π^* from reducible self-energy parts, additional complications arise. There is in fact only one irreducible photon self-energy part, the one denoted by W' in Fig. 5; and there is, besides the self-energy part consisting of a single point, just one irreducible electron self-energy part, denoted by W in Fig. 5. All other self-energy parts may be obtained by making various insertions in W or W'. However, reducible self-energy parts are to be enumerated by inserting vertex parts at only one, and not both, of the vertices of W or W'; otherwise the same self-energy part would appear more than once in the enumeration. And the contribution $M(W_R)$ to Σ^* arising from a reducible part W_R will be, in general, an integral which involves simultaneously divergences corresponding to each of the ways in which W_R might have been built up by insertions of vertex parts at either or both vertices of W. This complication arises because, in the special case when two vertex parts are both contained in a self-energy part and each contains one end-vertex of the self-energy part (and in no other case), it is possible for the two vertex parts to overlap without either being completely contained in the other.

The finite part of $M(W_R)$ is to be separated out as follows. In a unique way, W_R is obtained from W by inserting a vertex part $\bar{V}_a$ at a, and self-energy parts $\bar{W}_a$ and $\bar{W}_a'$ in the two lines of W. From $M(W_R)$ there are subtracted all divergences arising from $\bar{V}_a$, $\bar{W}_a$, $\bar{W}_a'$; let the remainder after this sub-

traction be $M'(W_R)$. Next, W_R is considered as built up from W by inserting some vertex part $\bar{V}_b$ at b, and self-energy parts $\bar{W}_b$ and $\bar{W}_b'$ in the two lines of W. The integral $M'(W_R)$ will still contain divergences arising from $\bar{V}_b$ (but none from $\bar{W}_b$ and $\bar{W}_b'$), and these divergences are to be subtracted, leaving a remainder $M''(W_R)$. The finite part of $M''(W_R)$ can finally be separated by applying to the whole integral the method of Section VI, which gives for $M''(W_R)$ an expression of the form (67), with Σ_c given by (71). Therefore the finite part of $M(W_R)$ is a well-determined quantity, and is an operator of the form (71).

The behavior of the higher order contributions to Σ^* and Π^* having now been qualitatively explained, we may describe the precise rules for the calculation of Σ^* and Π^* by the same kind of inductive scheme as was given for Λ_μ in the second paragraph of this Section. Apart from the constant term $(-2\pi i\delta\kappa_0)$, Σ^* is just the contribution $\Sigma(W, t^1)$ from the W of Fig. 5; and $\Sigma(W, t^1)$ is represented by an integral of the form (65) with $l=1$. The integrand in (65) was a product of two operators γ_μ, one operator D_F, and one operator S_F. The exact $\Sigma(W, t^1)$ is to be obtained by replacing D_F by D_F', S_F by S_F', and one only of the factors γ_μ by Γ_μ, say the γ_μ corresponding to the vertex a of W. Suppose that S_F' in the integrand is represented, to order e^{2n}, by the sum of S_F and of a finite number of finite products of S_F with operators $S(\bar{W}, t^1)$ such as appear in (71); and suppose that D_F' and Γ_μ are similarly represented. Then $\Sigma(W, t^1)$ will be determined to order e^{2n+2}. The new $\Sigma(W, t^1)$ will be a sum of integrals like the $M'(W_R)$ of the previous paragraph, still containing divergences arising from vertex parts at the vertex b of W, in addition to divergences arising from the graph W_R as a whole. When all these divergences are dropped, we have a $\Sigma_c(W, t^1)$ which is finite; substituting this $\Sigma_c(W, t^1)$ for Σ^* in (63) gives an S_F' which is also finite and determined to the order e^{2n+2}.

The above procedures start from given S_F', D_F' and Γ_μ represented to order e^{2n} by, respectively, S_F plus S_F multiplied by a finite sum of products of $S(\bar{W}, t^1)$, D_F plus D_F multiplied by a finite sum of products of $D(\bar{W}', t^1)$, and γ_μ plus a finite sum of $\Lambda_{\mu c}(\bar{V}, t^1, t^2)$. From these there are obtained new expressions for S_F', D_F', Γ_μ. In the new expressions there appear new convergent operators $S(W, t^1)$, $D(W', t^1)$, $\Lambda_{\mu c}(V, t^1, t^2)$, determined to order e^{2n+2}; in the divergent terms which are separated out and dropped from the new expressions, there appear divergent coefficients A, B, C, L, such as occur in (73), (78), (82), also now determined to order e^{2n+2}. After the dropping of the divergent terms, the new Γ_μ by (82) is a sum of γ_μ and a finite set of $\Lambda_{\mu c}(V, t^1, t^2)$; the new S_F' by (73) is S_F plus S_F multiplied by a finite sum of products of $S(W, t^1)$; and the new D_F' by (78) is D_F plus D_F multiplied by a finite sum of products of $D(W', t^1)$. That is to say, the new Γ_μ, S_F', D_F' can be substituted back into the integrals of the form (65), and so a third set of operators Γ_μ, S_F', D_F' is obtained, determined to order e^{2n+4}, and again with finite and divergent parts separated. In this way, always dropping the divergent terms before substituting back into the integral equations, the finite parts of Γ_μ, S_F', D_F', may be calculated by a process of successive approximation, starting with the zero-order values γ_μ, S_F, D_F. After n substitutions, the finite parts of Γ_μ, S_F', D_F' will be determined to order e^{2n}.

It is necessary finally to justify the dropping of the divergent terms. This will be done by showing that the "true" Γ_μ, S_F', D_F', which are obtained if the divergent terms are not dropped, are only numerical multiples of those obtained by dropping divergences, and that the numerical multiples can themselves be eliminated from the theory by a consistent use of the ideas of mass and charge renormalization. Let $\Gamma_{\mu 1}(e)$, $S_{F1}'(e)$, $D_{F1}'(e)$ be the operators obtained by the process of substitution dropping divergent terms; these operators are power series in e with finite operator coefficients (to avoid raising the question of the convergence of these power-series, all quantities are supposed defined only up to some finite order e^{2N}). Then we shall show that the true operators Γ_μ, S_F', D_F' are of the form

$$\Gamma_\mu = Z_1^{-1}\Gamma_{\mu 1}(e_1), \tag{83}$$

$$S_F' = Z_2 S_{F1}'(e_1), \tag{84}$$

$$D_F' = Z_3 D_{F1}'(e_1), \tag{85}$$

where Z_1, Z_2, Z_3 are constants to be determined, and e_1 is given by

$$e_1 = Z_1^{-1}Z_2 Z_3^{\frac{1}{2}} e. \tag{86}$$

This e_1 will turn out to be the "true" electronic charge. It has to be proved that the result of substituting (83), (84), (85) into the integral equations defining Γ_μ, S_F', D_F', is to reproduce these expressions exactly, when Z_1, Z_2, Z_3, and $\delta\kappa_0$ are suitably chosen.

Concerning the $\Gamma_{\mu 1}(e)$, $S_{F1}'(e)$, $D_{F1}'(e)$, it is known that, when these operators are substituted into the integral equations, they reproduce themselves with the addition of certain divergent terms. The additional divergent terms consist partly of the terms involving A, B, C, L, which are displayed in (73), (78), (82), and partly of terms arising (in the case of S_F' and D_F' only) from the peculiar behavior of the vertices b, b' in Fig. 5. The terms arising from b and b' have been discussed earlier; they may be called for brevity b-divergences. Originally, of course, there is no asymmetry between the divergences arising in Σ^* from vertex parts inserted at

the two ends a and b of W; we have manufactured an asymmetry by including the divergences arising at a in the coefficient Z_1^{-1} of (83), while at b the operator γ_μ has not been replaced by Γ_μ and so the b divergences have not been so absorbed. It is thus to be expected that the effect of the b divergences, like that of the a divergences, will be merely to multiply all contributions to Σ^* by the constant Z_1^{-1}. Similarly, we expect that divergences at b' will multiply Π^* by the constant Z_1^{-1}. It can be shown, by a detailed argument too long to be given here, that these expectations are justified. (The interested reader is recommended to see for himself, by considering contributions to Σ^* arising from various self-energy parts, how it is that the finite terms of a given order are always reappearing in higher order multiplied by the same divergent coefficients.) Therefore, the complete expressions obtained by substituting $\Gamma_{\mu 1}(e)$, $S_{F1}'(e)$, $D_{F1}'(e)$, into the integral equations defining Λ_μ, Σ^*, Π^*, are

$$\Lambda_{\mu 1}(e)=\Lambda_{\mu c}(e)+L(e)\gamma_\mu, \tag{87}$$

$$S_F\Sigma_1^*(e)=-2\pi i\delta\kappa_0 S_F + Z_1^{-1}\left(A(e)S_F+\frac{1}{2\pi}B(e)+\frac{1}{2\pi}S_c(e)\right), \tag{88}$$

$$D_F\Pi_1^*(e)=Z_1^{-1}\left(\frac{1}{2\pi i}C(e)+\frac{1}{2\pi i}D_c(e)\right). \tag{89}$$

Here $A(e)$, $B(e)$, $C(e)$, $L(e)$ are well-defined power series in e, with coefficients which diverge never more strongly than as a power of a logarithm. The finite operators $\Lambda_{\mu c}(e)$, $S_c(e)$, $D_c(e)$, will, when all divergent terms are dropped, lead back to the $\Gamma_{\mu 1}(e)$, $S_{F1}'(e)$, $D_{F1}'(e)$, from which the substitution started; thus, according to (38), (63), (64),

$$\Gamma_{\mu 1}(e)=\gamma_\mu+\Lambda_{\mu c}(e), \tag{87'}$$

$$S_{F1}'(e)=S_F+\frac{1}{2\pi}S_c(e)S_{F1}'(e), \tag{88'}$$

$$D_{F1}'(e)=D_F+\frac{1}{2\pi i}D_c(e)D_{F1}'(e). \tag{89'}$$

Equations (87)–(89), (87')–(89'), describe precisely the way in which the $\Gamma_{\mu 1}(e)$, $S_{F1}'(e)$, $D_{F1}'(e)$, when substituted into the integral equations, reproduce themselves with the addition of divergent terms. And from these results it is easy to deduce the self-reproducing property of the operators (83)–(85), when substituted into the same equations.

Consider for example the effect of substituting from (83)–(85) into the term $\Sigma(W, t^1)$, given by (65) with $l=1$. The integrand of (65) is a product of one factor Γ_μ, one γ_μ, one S_F', and one D_F'. Therefore the substitution gives

$$Z_1^{-1}Z_2Z_3\Sigma_0(W), \tag{90}$$

where $\Sigma_0(W)$ is the expression (65) obtained by substituting $\Gamma_{\mu 1}(e_1)$, $S_{F1}'(e_1)$, $D_{F1}'(e_1)$, without the Z factors. Now the Z factors in (90) combine with the e^2 of (65) to give

$$Z_1Z_2^{-1}e_1^2,$$

and the remaining factor of $\Sigma_0(W)$ is explicitly a function of e_1 and not of e. Therefore (90) is

$$Z_1Z_2^{-1}\Sigma_1(W, e_1),$$

where $\Sigma_1(W, e)$ is the expression obtained by substituting the operators $\Gamma_{\mu 1}(e)$, $S_{F1}'(e)$, $D_{F1}'(e)$ into $\Sigma(W, t^1)$. Thus the $\Sigma^*(t^1)$, obtained by substituting from (83)–(85) into (65), is identical with the result of substituting the operators $\Gamma_{\mu 1}(e)$, $S_{F1}'(e)$, $D_{F1}'(e)$, and afterwards changing e to e_1 and multiplying the whole expression (except for the constant term in $\delta\kappa_0$) by $Z_1Z_2^{-1}$. More exactly, using (88), one can say that the Σ^* obtained by substituting from (83)–(85) is given by

$$S_F\Sigma^*=-2\pi i\delta\kappa_0 S_F + Z_2^{-1}\left(A(e_1)S_F+\frac{1}{2\pi}B(e_1)+\frac{1}{2\pi}S_c(e_1)\right). \tag{91}$$

Further, the S_F' obtained by substituting from (83)–(85) into the integral equations is given by (91) and

$$S_F'=S_F+S_F\Sigma^*S_F'. \tag{92}$$

It is now easy to verify, using (88'), that S_F' given by (91) and (92) will be identical with (84), provided that

$$Z_2=1+\frac{1}{2\pi}B(e_1), \tag{93}$$

$$\delta\kappa_0=\frac{1}{2\pi i}Z_2^{-1}A(e_1). \tag{94}$$

In a similar way, the D_F' obtained by substituting from (83)–(85) into the integral equations can be related with the $\Pi_1^*(e)$ of (89). This D_F' will be identical with (85) provided that

$$Z_3=1+\frac{1}{2\pi i}C(e_1). \tag{95}$$

Finally, the Γ_μ obtained by substituting from (83)–(85) can be shown to be

$$\Gamma_\mu=\gamma_\mu+Z_1^{-1}\Lambda_{\mu 1}(e_1),$$

with $\Lambda_{\mu 1}(e)$ given by (87). Using (87'), this Γ_μ will

be identical with (83) provided that

$$Z_1=1-L(e_1). \tag{96}$$

Therefore, if Z_1, Z_2, Z_3, $\delta\kappa_0$ are defined by (96), (93), (95), (94), it is established that (83)–(85) give the correct forms of the operators Γ_μ, S_F', D_F', including all the effects of the radiative corrections which these operators introduce into themselves and into each other. The exact Eqs. (83)–(85) give a much simpler separation of the infinite from the finite parts of these operators than the approximate equations (73), (78), (82).

Consider now the result of using the exact operators (83)–(85) in calculating a constituent M of $U(\infty)$, where M is constructed from a certain irreducible graph G_0 according to the rules of Section IV. G_0 will have, say, F_e internal and E_e external electron lines, F_p internal and E_p external photon lines, and

$$n=F_e+\tfrac{1}{2}E_e=2F_p+E_p \tag{97}$$

vertices. In M there will be $\frac{1}{2}E_e$ factors $\psi'(k^i)$, $\frac{1}{2}E_e$ factors $\bar{\psi}'(k^i)$ and E_p factors $A_\mu'(k^i)$ given by (37). In $\psi'(k^i)$, k^i is the momentum-energy 4 vector of an electron, which satisfies (69), and the $S_c(k^i)$ in (73) are zero at every stage of the inductive definition of $S_{F1}'(e)$. Therefore (84), (35), (37) give in turn

$$\begin{aligned} &S_F'(k^i)=Z_2S_F(k^i),\\ &\Sigma(k^i)=2\pi(Z_2-1)(k_\mu{}^i\gamma_\mu-i\kappa_0),\\ &\psi'(k^i)=\psi(k^i)+2\pi(Z_2-1)S_F(k^i)(k_\mu{}^i\gamma_\mu-i\kappa_0)\psi(k^i). \end{aligned} \tag{98}$$

The expression (98) is indeterminate, since $(k_\mu{}^i\gamma_\mu-i\kappa_0)$ operating on $\psi(k^i)$ gives zero, while operating on $S_F(k^i)$ it gives the constant $(1/2\pi)$. Thus, according to the order in which the factors are evaluated, (98) will give for $\psi'(k^i)$ either the value $\psi(k^i)$ or the value $Z_2\psi(k^i)$. Similarly, $\bar{\psi}'(k^i)$ is indeterminate between $\bar{\psi}(k^i)$ and $Z_2\bar{\psi}(k^i)$, and, excluding for the moment $A_\mu(k^i)$ which are Fourier components of the external potential, $A_\mu'(k^i)$ is indeterminate between $A_\mu(k^i)$ and $Z_3A_\mu(k^i)$. In any case, considerations of covariance show that the $\psi'(k^i)$, $\bar{\psi}'(k^i)$, $A_\mu'(k^i)$ are numerical multiples of the $\psi(k^i)$, $\bar{\psi}(k^i)$, $A_\mu(k^i)$; thus the indeterminacy lies only in a constant factor multiplying the whole expression M.

There cannot be any indeterminacy in the magnitude of the matrix elements of $U(\infty)$, so long as this operator is restricted to be unitary. The indeterminacy in fact lies only in the normalization of the electron and photon wave functions $\psi(k^i)$, $\bar{\psi}(k^i)$, $A_\mu(k^i)$, which may or may not be regarded as altered by the continual interactions of these particles with the vacuum-fields around them. It can be shown that, if the wave functions are everywhere normalized in the usual way, the apparent indeterminacy is removed, and one must take

$$\begin{aligned} \psi'(k^i)&=Z_2{}^{\frac{1}{2}}\psi(k^i),\\ \bar{\psi}'(k^i)&=Z_2{}^{\frac{1}{2}}\bar{\psi}(k^i),\\ A_\mu'(k^i)&=Z_3{}^{\frac{1}{2}}A_\mu(k^i). \end{aligned} \tag{99}$$

It will be seen that (99) gives just the geometric mean of the two alternative values of $\psi'(k^i)$ obtained from (98).

When $A_\mu(k^i)$ is a Fourier component of the external potential, then in general $(k^i)^2\neq 0$, and $A_\mu'(k^i)$ is not indeterminate but is given by (37) and (85) in the form

$$A_\mu'(k^i)=2\pi iZ_3D_{F1}'(e_1)(k^i)^2A_\mu(k^i). \tag{100}$$

However, the unit in which external potentials are measured is defined by the dynamical effects which the potentials produce on known charges; and these dynamical effects are just the matrix elements of $U(\infty)$ in which (100) appears. Therefore the factor Z_3 in (100) has no physical significance, and will be changed when A_μ is measured in practical units. The correct constant which appears when practical units are used is $Z_3{}^{\frac{1}{2}}$; this is because the photon potentials A_μ in (99) were normalized in terms of practical units; and (100) should reduce to (99) when $(k^i)^2\to 0$, if the external A_μ and the photon A_μ are measured in the same units. Therefore the correct formula for A_μ', covering the cases both of photon and of external potentials, is

$$\left.\begin{aligned} A_\mu'(k^i)&=2\pi iZ_3{}^{\frac{1}{2}}D_{F1}'(e_1)(k^i)^2A_\mu(k^i), \quad (k^i)^2\neq 0,\\ A_\mu'(k^i)&=Z_3{}^{\frac{1}{2}}A_\mu(k^i), \quad (k^i)^2=0. \end{aligned}\right\} \tag{101}$$

In M there will appear F_e factors S_F', F_p factors D_F', and n factors Γ_μ, in addition to the factors of the type (99), (101). Hence by (97) the Z factors will occur in M only as the constant multiplier

$$Z_1{}^{-n}Z_2{}^nZ_3{}^{\frac{1}{2}n}.$$

By (86), this multiplier is exactly sufficient to convert the factor e^n, remaining in M from the original interaction (8), into a factor $e_1{}^n$. Thereby, both e and Z factors disappear from M, leaving only their combination e_1 in the operators $\Gamma_{\mu 1}(e_1)$, $S_{F1}'(e_1)$, $D_{F1}'(e_1)$, and in the factor $e_1{}^n$. If now e_1 is identified with the finite observed electronic charge, there no longer appear any divergent expressions in M. And since M is a completely general constituent of $U(\infty)$, the elimination of divergences from the S matrix is accomplished.

It hardly needs to be pointed out that the arguments of this section have involved extensive manipulations of infinite quantities. These manipulations have only a formal validity, and must be justified *a posteriori* by the fact that they ultimately lead to a clear separation of finite from infinite expressions. Such an *a posteriori* justification of dubious manipulations is an inevitable feature of

any theory which aims to extract meaningful results from not completely consistent premises.

We conclude with two disconnected remarks. First, it is probable that $Z_1=Z_2$ identically, though this has been proved so far only up to the order e^2. If this conjecture is correct, then all charge-renormalization effects arise according to (86) from the coefficient Z_3 alone, and the arguments of this paper can be somewhat simplified. Second, Eqs. (88′), (89′), which define the fundamental operators S_{F1}', D_{F1}', may be solved for these operators. Thus

$$S_{F1}'(e)=\left[1-\frac{1}{2\pi}S_c(e)\right]^{-1}S_F, \tag{88''}$$

$$D_{F1}'(e)=\left[1-\frac{1}{2\pi i}D_c(e)\right]^{-1}D_F. \tag{89''}$$

In electrodynamics, the S_c and D_c are small radiative corrections, and it will always be legitimate and convenient to expand (88″) and (89″) by the binomial theorem. If, however, the methods of the present paper are to be applied to meson fields, with coupling constants which are not small, then it will be desirable not to expand these expressions; in this way one may hope to escape partially from the limitations which the use of weak-coupling approximations imposes on the theory.

VIII. SUMMARY OF RESULTS

The results of the preceding sections divide themselves into two groups. On the one hand, there is a set of rules by which the element of the S matrix corresponding to any given scattering process may be calculated, without mentioning the divergent expressions occurring in the theory. On the other hand, there is the specification of the divergent expressions, and the interpretation of these expressions as mass and charge renormalization factors.

The first group of results may be summarized as follows. Given a particular scattering problem, with specified initial and final states, the corresponding matrix element of $U(\infty)$ is a sum of contributions from various graphs G as described in Section II. A particular contribution M from a particular G is to be written down as an integral over momentum variables according to the rules of Section III; the integrand is a product of factors $\psi(k^i)$, $\bar{\psi}(k^i)$, $A_\mu(k^i)$, $S_F(p^i)$, $D_F(p^i)$, $\delta(q_i)$, γ_μ, the factors corresponding in a prescribed way to the lines and vertices of G. According to Section IV, contributions M are only to be admitted from irreducible G; the effects of reducible graphs are included by replacing in M the factors ψ, $\bar{\psi}$, A_μ, S_F, D_F, γ_μ, by the corresponding expressions (37), (35), (36), (38). These replacements are then shown in Section VII to be equivalent to the following: each factor S_F in M is replaced by $S_{F1}'(e)$, each factor D_F by $D_{F1}'(e)$, each factor γ_μ by $\Gamma_{\mu 1}(e)$, each factor A_μ when it represents an external potential is replaced by

$$A_{\mu 1}(k^i)=2\pi i D_{F1}'(e)(k^i)^2 A_\mu(k^i), \tag{102}$$

factors ψ, $\bar{\psi}$, A_μ representing particle wave-functions are left unchanged, and finally e wherever it occurs in M is replaced by e_1. The definition of M is completed by the specification of $S_{F1}'(e)$, $D_{F1}'(e)$, $\Gamma_{\mu 1}(e)$; it is in the calculation of these operators that the main difficulty of the theory lies. The method of obtaining these operators is the process of successive substitution and integration explained in the first part of Section VII; the operators so calculated are divergence-free, the divergent parts at every stage of the calculation being explicitly dropped after being separated from the finite parts by the method of Section VI.

The above rules determine each contribution M to $U(\infty)$ as a divergence-free expression, which is a function of the observed mass m and the observed charge e_1 of the electron, both of which quantities are taken to have their empirical values. The divergent parts of the theory are irrelevant to the calculation of $U(\infty)$, being absorbed into the unobservable constants δm and e occurring in (8). A place where some ambiguity might appear in M is in the calculation of the operators $S_{F1}'(e)$, $D_{F1}'(e)$, $\Gamma_{\mu 1}(e)$, when the method of Section VI is used to separate out the finite parts $S(W, t^1)$, $D(W', t^1)$, $\Lambda_{\mu c}(V, t^1, t^2)$, from the expressions (67), (74), (80). Even in this place the rules of Section VI give unambiguous directions for making the separation; only there is a question whether some alternative directions might be equally reasonable. For example, it is possible to separate out a finite part from $\Sigma(W, t^1)$ according to (67), and not to make the further step of using (70) to separate out a finite part $S(W, t^1)$ which vanishes when (69) holds. Actually it is easy to verify that such an alternative procedure will not change the value of M, but will only make its evaluation more complicated; it will lead to an expression for M in which one (infinite) part of the mass and charge renormalizations is absorbed into the constants δm and e, while other finite mass and charge renormalizations are left explicitly in the formulas. It is just these finite renormalization effects which the second step in the separation of $S(W, t^1)$ and $\Lambda_{\mu c}(V, t^1, t^2)$ is designed to avoid. Therefore it may be concluded that the rules of calculation of $U(\infty)$ are not only divergence-free but unambiguous.

As anyone acquainted with the history of the Lamb shift[11] knows, the utmost care is required

[11] H. A. Bethe, *Electromagnetic Shift of Energy Levels*, Report to Solvay Conference, Brussels (1948).

before it can be said that any particular rule of calculation is unambiguous. The rules given in this paper are unambiguous, in the sense that each quantity to be calculated is an integral in momentum-space which is absolutely convergent at infinity; such an integral has always a well-defined value. However, the rules would not be unambiguous if it were allowed to split the integrand into several parts and to evaluate the integral by integrating the parts separately and then adding the results; ambiguities would arise if ever the partial integrals were not absolutely convergent. A splitting of the integrals into conditionally convergent parts may seem unnatural in the context of the present paper, but occurs in a natural way when calculations are based upon a perturbation theory in which electron and positron states are considered separately from each other. The absolute convergence of the integrals in the present theory is essentially connected with the fact that the electron and positron parts of the electron-positron field are never separated; this finds its algebraic expression in the statement that the quadratic denominator in (45) is never to be separated into partial fractions. Therefore the absence of ambiguity in the rules of calculation of $U(\infty)$ is achieved by introducing into the theory what is really a new physical hypothesis, namely that the electron-positron field always acts as a unit and not as a combination of two separate fields. A similar hypothesis is made for the electromagnetic field, namely that this field also acts as a unit and not as a sum of one part representing photon emission and another part representing photon absorption.

Finally, it must be said that the proof of the finiteness and unambiguity of $U(\infty)$ given in this paper makes no pretence of being complete and rigorous. It is most desirable that these general arguments should as soon as possible be supplemented by an explicit calculation of at least one fourth-order radiative effect, to make sure that no unforeseen difficulties arise in that order.

The second group of results of the theory is the identification of δm and e by (94) and (86). Although these two equations are strictly meaningless, both sides being infinite, yet it is a satisfactory feature of the theory that it determines the unobservable constants δm and e formally as power series in the observable e_1, and not vice versa. There is thus no objection in principle to identifying e_1 with the observed electronic charge and writing

$$(e_1^2/4\pi\hbar c)=\alpha=1/137. \tag{103}$$

The constants appearing in (8) are then, by (94) and (86),

$$\delta m=m(A_1\alpha+A_2\alpha^2+\cdots), \tag{104}$$

$$e=e_1(1+B_1\alpha+B_2\alpha^2+\cdots), \tag{105}$$

where the A_i and B_i are logarithmically divergent numerical coefficients, independent of m and e_1.

IX. DISCUSSION OF FURTHER OUTLOOK

The surprising feature of the S matrix theory, as outlined in this paper, is its success in avoiding difficulties. Starting from the methods of Tomonaga, Schwinger and Feynman, and using no new ideas or techniques, one arrives at an S matrix from which the well-known divergences seem to have conspired to eliminate themselves. This automatic disappearance of divergences is an empirical fact, which must be given due weight in considering the future prospects of electrodynamics. Paradoxically opposed to the finiteness of the S matrix is the second fact, that the whole theory is built upon a Hamiltonian formalism with an interaction-function (8) which is infinite and therefore physically meaningless.

The arguments of this paper have been essentially mathematical in character, being concerned with the consequences of a particular mathematical formalism. In attempting to assess their significance for the future, one must pass from the language of mathematics to the language of physics. One must assume provisionally that the mathematical formalism corresponds to something existing in nature, and then enquire to what extent the paradoxical results of the formalism can be reconciled with such an assumption. In accordance with this program, we interpret the contrast between the divergent Hamiltonian formalism and the finite S matrix as a contrast between two pictures of the world, seen by two observers having a different choice of measuring equipment at their disposal. The first picture is of a collection of quantized fields with localizable interactions, and is seen by a fictitious observer whose apparatus has no atomic structure and whose measurements are limited in accuracy only by the existence of the fundamental constants c and h. This observer is able to make with complete freedom on a sub-microscopic scale the kind of observations which Bohr and Rosenfeld[12] employ in a more restricted domain in their classic discussion of the measurability of field-quantities; and he will be referred to in what follows as the "ideal" observer. The second picture is of a collection of observable quantities (in the terminology of Heisenberg), and is the picture seen by a real observer, whose apparatus consists of atoms and elementary particles and whose measurements are limited in accuracy not only by c and h but also by other constants such as α and m. The real observer

[12] N. Bohr and L. Rosenfeld, Kgl. Dansk. Vid. Sels. Math.-Phys. Medd. 12, No. 8 (1933). A second paper by Bohr and Rosenfeld is to be published later, and is abstracted in a booklet by A. Pais, *Developments in the Theory of the Electron* (Princeton University Press, Princeton, 1948).

makes spectroscopic observations, and performs experiments involving bombardments of atomic systems with various types of mutually interacting subatomic projectiles, but to the best of our knowledge he cannot measure the strength of a single field undisturbed by the interaction of that field with others. The ideal observer, utilizing his apparatus in the manner described in the analysis of the Hamiltonian formalism by Bohr and Rosenfeld,[12] makes measurements of precisely this last kind, and it is in terms of such measurements that the commutation-relations of the fields are interpreted. The interaction-function (8) will presumably always remain unobservable to the real observer, who is able to determine positions of particles only with limited accuracy, and who must always obtain finite results from his measurements. The ideal observer, however, using non-atomic apparatus whose location in space and time is known with infinite precision, is imagined to be able to disentangle a single field from its interactions with others, and to measure the interaction (8). In conformity with the Heisenberg uncertainty principle, it can perhaps be considered a physical consequence of the infinitely precise knowledge of location allowed to the ideal observer, that the value obtained by him when he measures (8) is infinite.

If the above analysis is correct, the divergences of electrodynamics are directly attributable to the fact that the Hamiltonian formalism is based upon an idealized conception of measurability. The paradoxical feature of the present situation does not then lie in the mere coexistence of a finite S matrix with an infinite interaction-function. The empirically found correlation, between expressions which are unobservable to a real observer and expressions which are infinite, is a physically intelligible and acceptable feature of the theory. The paradox is the fact that it is necessary in the present paper to start from the infinite expressions in order to deduce the finite ones. Accordingly, what is to be looked for in a future theory is not so much a modification of the present theory which will make all infinite quantities finite, but rather a turning-round of the theory so that the finite quantities shall become primary and the infinite quantities secondary.

One may expect that in the future a consistent formulation of electrodynamics will be possible, itself free from infinities and involving only the physical constants m and e_1, and such that a Hamiltonian formalism with interaction (8), with divergent coefficients δm and e, may in suitably idealized circumstances be deduced from it. The Hamiltonian formalism should appear as a limiting form of a description of the world as seen by a certain type of observer, the limit being approached more and more closely as the precision of measurement allowed to the observer tends to infinity.

The nature of a future theory is not a profitable subject for theoretical speculation. The future theory will be built, first of all upon the results of future experiments, and secondly upon an understanding of the interrelations between electrodynamics and mesonic and nucleonic phenomena. The purpose of the foregoing remarks is merely to point out that there is now no longer, as there has seemed to be in the past, a compelling necessity for a future theory to abandon some essential features of the present electrodynamics. The present electrodynamics is certainly incomplete, but is no longer certainly incorrect.

In conclusion, the author would like to express his profound indebtedness to Professor Feynman for many of the ideas upon which this paper is built, to Professor Oppenheimer for valuable discussions, and to the Commonwealth Fund of New York for financial support.