UNIVERSAL
LIBRARY

PROCEEDINGS OF
SYMPOSIA IN APPLIED MATHEMATICS
VOLUME II

ELECTROMAGNETIC THEORY

PUBLISHED BY THE
AMERICAN MATHEMATICAL SOCIETY
531 WEST 116TH STREET, NEW YORK CITY
1950

PROCEEDINGS OF THE SECOND SYMPOSIUM IN APPLIED MATHEMATICS OF THE AMERICAN MATHEMATICAL SOCIETY

Held at the Massachusetts Institute of Technology
July 29–31, 1948

COSPONSORED BY

THE AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
THE AMERICAN INSTITUTE OF PHYSICS
THE INSTITUTE OF RADIO ENGINEERS

TABLE OF CONTENTS

THE NEW QUANTUM ELECTRODYNAMICS

BY

HERMAN FESHBACH

1. **Introduction.** In this report we shall attempt to summarize the recent developments in this field. We should remain free to imagine that the final form of the theory has not yet been delineated. In the present discussion we shall rely almost exclusively upon the unpublished work of Julian Schwinger [17; 18; 19]. Similar methods have been used by French and Weisskopf [3] and Tomanaga [22; 5] and Feynman.

Recent experiments by Lamb and Retherford [7], and by Nelson, Nafe, and Rabi [11], which have since been substantiated by several workers [10; 12; 6; 4], have forced the consideration of the problems raised by the point theory of the electron. In electrodynamics it is usual to use perturbation theory; the expansion in question being in powers of the fine structure constant $\alpha = e^2/\hbar c$ (e = electronic charge, $\hbar$ = Planck's constant/2π; c = velocity of light). Before the present theory it was possible to compute results from theory only up to the first non-vanishing term in the power series in α. The next term in the series diverged. Yet excellent agreement with experiment was obtained using only this first term. Arguments [13; 8] have been devised to show that these higher order terms should in actuality be small, but these arguments were not a straightforward consequence of the theory. The recent experimental work has given an accurate measurement of these higher order terms, terms which are in the same order as the divergent terms. The achievement of the theory to be discussed today is that it permits the calculation of the next order in a non-ambiguous manner. The agreement with experiment is remarkably close.

In order to appreciate the points involved, it is useful to first discuss the divergence and the consequent difficulties, the attempts of the older theories to circumvent them. It is, of course, impossible to include in such a survey all attempted solutions. We shall limit ourselves to those suggestions which are particularly relevant to the theory we shall discuss today. In addition, we shall give a simple method of computing the result of the Lamb-Retherford experiment, so as to indicate the important physical concepts involved.

2. **History of problem.** The theory of the electron when formulated by Lorentz [9] was immediately confronted with the question of the stability of the electron. A local cohesive force was introduced to hold it together, but then it was found that the electron had an additional inertia. If, for example, the electron were of radius a, this inertia is e^2/a. In that event, it is clear that the theory of a point electron becomes impossible for the inertia (we shall also refer to this quantity as *self-energy* in the following) diverges as $a \to 0$.

It is possible to perform a simple non-relativistic calculation which will show the origin of this effect. Consider a rigid electron of radius a, with center at $\mathbf{r}'$. This will generate a scalar potential φ in virtue of the equation

$$\Pi^2\varphi = -\rho(r - r'), \int \rho(r - r')\, d\tau = e, \qquad (1)$$
$$\Pi^2 = \nabla^2 - \partial^2/\partial x_0^2, \qquad x_0 = ct,$$

where ρ is the density function describing the distribution of charge about $\mathbf{r}'$; we shall assume $\rho = \rho(\,|\,r - r'\,|\,)$. The equation of motion of the center of mass neglecting magnetic forces is

$$m\ddot{r} + e\int \rho(r'' - r')\nabla\varphi(r'')\, d\tau'' = 0. \qquad (2)$$

For simplicity assume that the time dependence is simple harmonic, with angular frequency ω. Then

$$\varphi(r'') = \frac{1}{4\pi}\int \frac{e^{ikR}}{R}\rho(r''' - r')\, d\tau''',$$
$$R = |\,r''' - r''\,|, \qquad k = \omega/c.$$

If the electron has a dimension a, for small a

$$\varphi(r'') \xrightarrow[ka\to 0]{} \frac{1}{4\pi}\left\{\int \frac{\rho(r''' - r')}{R}\, d\tau''' + ik\int \rho(r''' - r')\, d\tau''' - \frac{k^2}{2}\int R\rho(r''' - r')\, d\tau''' - ik^3\int R^2\rho(r''' - r')\, d\tau''' + \cdots\right\}. \qquad (3)$$

We now take (3) and introduce it into (2). The first term in (2), being independent of the motion of the electron, gives simply the repulsive force tending to blow the electron up; this we shall assume is balanced out by some cohesive force of short range. The next term, being a constant (see (1)), vanishes when the gradient operator is applied. Finally, the last two terms behave as ω^2 and $i\omega^3$ respectively. Therefore when introduced into (2) they will behave as reactive and resistive terms respectively. The reactance is masslike and is

$$X = \frac{2}{3}\iint \frac{\rho(r'' - r')\rho(r' - r''')}{|\,r'' - r'''\,|}\, d\tau''\, d\tau'''. \qquad (4a)$$

The resistive term is

$$R = \frac{2}{3}\frac{e^2}{c^3}\omega^3 \qquad (4b)$$

and is structure independent. Higher order terms yield zero as $a \to 0$ but are structure dependent. The equation of motion for a *free* electron becomes

$$(m + X)\ddot{r} - 2e^2/3c^3\,\dddot{r} = 0. \tag{5}$$

It seemed natural at the time for Lorentz to incorporate X into m and to say that $(m + X)$ is the *actual experimental mass* observed for a free electron. This renormalization of mass is precisely the method used in the new electrodynamics to absorb the self-energy term appearing there. In order to keep this electromagnetic mass *finite*, a definite radius a for the electron had to be chosen. Since it is difficult to define a radius so that it is meaningful from the point of view of the Lorentz transformation, it was not considered a satisfactory way of modifying the theory. One solution of this dilemma involves assuming that a satisfactory definition of a exists so that X is known, calling $m + X$ the experimental mass, and finally pointing out that the structure dependent terms will be important for only higher order effects. Indeed the first term which does appear, the resistive term, is structure independent.

It was Poincaré [15] who first pointed out that the requirements of the special theory did not permit the entire mass of the electron to be electromagnetic in origin. His solution of the problem required the introduction of another field, short range in character, which would be attractive overcoming the instability of the electron due to the Coulomb repulsion of the charge. Such a theory may be made Lorentz-invariant. It has the unpleasant (but perfectly possible) feature of introducing an unobserved field.

Poincaré's procedure has been reconsidered again in recent years by Stueckelberg [21] and Bopp [2] in classical theory and by Pais [14], Podolsky[1] et al., and Sakata [16] in quantum theory. These, at least in the order to which their consequences have been carried, seem to yield consistent results.

It will be noted that the Poincaré philosophy differs from that of Lorentz in that the former attempts to specify the electron structure giving a specific model, whereas Lorentz does not specify a model but rather attempts to discover those properties which are independent of structure. Poincaré modifies the Coulomb law of force at short distances between the charge; Lorentz says it must be modified but does not say how, leaving it rather for future experiment to determine.

Finally we shall mention, in passing, attempts to define an electron radius in a Lorentz-invariant way. This may be done giving the electron some extension in space-time. Unfortunately it leads to rather difficult equations of motion. As a consequence it has been difficult to quantize these theories in a satisfactory manner.

The introduction of quantum theory introduced grave new difficulties in the theory of the point electron; indeed the divergences become more severe. The main source of these troubles stems from the fact that in quantum field theory,

---

[1] See, for example, a review article, Boris Podolsky and Philip Schwed, Reviews of Modern Physics vol. 20 (1948) p. 40.

the vacuum instead of being simple "emptiness" becomes a complex "emptiness". This is due to *fluctuation* phenomena [23; 24; 25] which occur both for the radiation field and for the matter field. We shall discuss the effects of the radiation field first, for we shall be able to understand the quantum infinity occurring here by means of a relatively simple argument. Consider, for example, the electric field in a vacuum. Of course its average value is zero

$$E_{av} = 0.$$

However from quantum mechanics it follows that the average of E^2 is not zero

$$\langle E^2 \rangle_{av} \neq 0.$$

These fluctuations are part and parcel of quantum mechanics as formulated at present. It is responsible for the "spontaneous" emission part of the radiation transition probabilities as first discovered by Einstein. Stated in another way, in quantum mechanics an electron in a Bohr orbit does not radiate, the orbit is stationary. However, the fluctuating vacuum electromagnetic fields can induce transitions down to a lower state, which, since they occur in vacuo, are called spontaneous emissions.

Consider now the behavior of a free electron in a vacuum, which for simplicity is taken to be at rest. But is it at rest? In the presence of the fluctuating vacuum fields it will perform a sort of Brownian motion. Its average displacement and position will be zero; however the average of the velocity squared is not zero so that it acquires energy because of these fields. This energy is called the fluctuation energy W_f. We shall now compute it semi-classically.

We may analyze the vacuum field into its harmonic components of angular frequency ω. We shall compute the fluctuation energy $W_f(\omega)$ for each such component and obtain the total fluctuation energy by integrating over the frequency spectrum $n(\omega)\,d\omega$ of the vacuum:

$$W_f = \int_0^\infty W_f(\omega) n(\omega)\, d\omega. \tag{6}$$

The energy acquired by the particle in virtue of the vacuum fluctuations is

$$W_f(\omega) = \frac{e}{c} \langle v \cdot A \rangle_{av} \tag{7}$$

where A is the vector potential associated with the vacuum field. To compute v we must solve the electron equation of motion:

$$m\dot{v} = eE_0 \cos \omega t \qquad \text{or} \qquad v = \frac{eE_0}{m\omega} \cos \omega t$$

so that

$$W_f(\omega) = \frac{e^2}{2mc^2} \langle A_0^2 \rangle_{av} \tag{8}$$

where A_0 is the amplitude of A. Equation (8) is precisely the equation obtained by Weisskopf in his paper [25] of 1939. To obtain the frequency dependence of the integrand of (6) we need the frequency dependence of A_0. Assuming 1 photon per unit volume V (the final result will be independent of V), and using $E_0^2/4\pi = \hbar\omega/2V$ one finds that $A \sim 1/\omega^{1/2}$. Using $n(\omega) = V\omega^2/\pi^2c^3$ we finally obtain

$$\frac{W_f}{me^2} = \frac{1}{\pi}\frac{e^2}{\hbar c}\left(\frac{\hbar}{mc^2}\right)^2 \int_0^\infty \omega \, d\omega, \tag{9}$$

a quadratic divergence. The classical divergence of equation (4a) is linear in frequency space.

In addition to the fluctuations of the vacuum electromagnetic field, the vacuum also contains fluctuating charge fields. Although the average charge density $\rho_{AV} = 0$, the mean square fluctuation is not. These charge fluctuations manifest themselves experimentally in the phenomenon of pair production. Charge fluctuations are closely connected with electromagnetic field fluctuations. A simple picture of the vacuum which makes this connection more intuitive, envisages the vacuum as containing positron-electron dipoles. The charge per unit volume is zero. An incident light wave will be absorbed by the dipole, and if it contains enough energy ($>2mc^2$) will break the dipole apart producing a pair. The vacuum electric field will also polarize the dipole producing consequently a fluctuation in the charge density. We see that in this picture charge fluctuations are a consequence of field fluctuations.

Vacuum charge fluctuations give rise to two phenomena of interest in the context of this paper. The first suggests that a photon, because of the fact that it polarizes the vacuum by polarizing the dipoles it contains, may possess an inertia, giving rise to the problem of the photon self-energy. From the covariance properties of electromagnetic fields, it may however be shown that this inertia must be zero. A quantum calculation [20] gives an ambiguous answer, again because of the presence of a divergence. As evaluated in the reference [20] the photon self-energy is finite, a result as disturbing as the linear and quadratic divergences reported above for the self-energy of the electron.

The self-energy of the electron is also affected by the fluctuating vacuum charge. Introducing an electron into the vacuum will polarize the vacuum. The electron will accumulate positive charge around it and will push a corresponding amount of negative charge away. This has the net effect of spreading the charge of the electron out over a volume whose radius is $(\hbar/mc)$. The self-energy of the electron arising from the electrostatic field no longer diverges as e^2/a but much more slowly [23; 24; 25]

$$\frac{1}{4\pi}\frac{e^2}{\hbar c}\ln\frac{\hbar}{mca}. \tag{10}$$

The fluctuation self-energy of the Einstein-Bose charged particle is not affected by the vacuum charge. However, in the case of a Fermi-Dirac particle obeying the Pauli-exclusion principle, the fluctuation energy can be shown to vanish [23; 24; 25], though as we shall see this is not an unambiguous result. The essence of this calculation consists in noting that the vacuum charge also possesses a fluctuation energy W_{fl}^{vac} . The fluctuation self-energy of an electron should then be taken to be

$$W_{fl}\text{ (electron + vacuum)} - W_{fl}\text{ (vacuum)} = W_{fl}. \tag{11}$$

Both terms of this difference are infinite but the difference, if taken properly, will not diverge as rapidly. The term W_{fl} (electron + vacuum) contains the fluctuation of the electron plus the fluctuations of the vacuum electrons. However, not all the fluctuations of the vacuum electrons are possible in virtue of the Pauli exclusion principle. Indeed, one may readily verify that those vacuum fluctuations which are not allowed equal, term by term, the possible fluctuations of the electron so that W_{fl} (electron + vacuum) $\simeq$ W_{fl} (vacuum). However, since we are subtracting two infinities, the difference in (11) may not be unique, it may not transform properly when passing to another Lorentz reference system, and so on. Assuming, however, that the subtraction may somehow be justified, the self-energy of the electron still diverges in virtue of equation (10).

To summarize, the theory of the interaction of matter and radiation leads in both classical and quantum theories to infinities in the self-energy of the electron which in their dependence on e are of the order of $e^2/\hbar c$. We should emphasize that quantum theory brings in difficulties which are of a different character from those of classical theory. This has the consequence that any modifications of the description of matter interacting with radiation are pertinent in quantum theory only.

Let us now turn to the Lamb-Retherford experiment. Here we are concerned with the energy of the hydrogen atom, or more specifically with the energy levels of an electron moving in the Coulomb field of a proton. Of course similar results are obtained if the proton is replaced by any atomic nucleus. The Dirac theory of the hydrogen atom (neglecting the magnetic moment of the proton) gives a degeneracy for the first excited state, a degeneracy between $2s_{1/2}$ and $2p_{1/2}$. The $2s_{1/2}$ level is metastable, since any transitions to the ground state $1s_{1/2}$ is forbidden. The purpose of the experiment was to measure the energy difference between the $2s_{1/2}$ and $2p_{1/2}$ level and to see if they were really degenerate. It was found that they were not degenerate, that actually the $2s_{1/2}$ level was higher than the $2p_{1/2}$ level by 1065 megacycles per second on a frequency scale.

This deviation from quantum theory was interpreted as being due to the different energy of interaction of the electrons in the $2s_{1/2}$ and $2p_{1/2}$ states with the vacuum radiation field. Only now we have the difficulty that the interaction energies for both of these states is infinite as noted above, so that their difference is not unambiguous. The recent developments in quantum electrodynamics have revolved about the question.

It is possible to derive, by means of the semi-classical method used earlier, an expression for the energy difference of these two levels and so obtain a clue to the relevant physics.

The problem is that of computing the effect of radiation on a bound state. We know that this is infinite. However, the important physical quantity is *the difference between the interaction energy of the particle when bound and the particle when free.* As a simple case, consider the case of a harmonic oscillator which under the influence of the vacuum electric field:

$$m\ddot{r} + m\omega_0^2 r = eE_0 \cos \omega t.$$

Proceeding as in equations (7) and (8), we obtain

$$W_{fl}^{\text{bound}}(\omega) = (e^2E_0^2)/2m(\omega^2 - \omega_0^2)$$

so that

$$(\Delta W)_{fl} = W_{fl}^{\text{bound}} - W_{fl}^{\text{free}} = \frac{1}{\pi}\frac{e^2}{\hbar c}\left(\frac{\hbar}{mc^2}\right)^2 \int_0^\infty \omega^2 \left[\frac{1}{\omega^2 - \omega_0^2} - \frac{1}{\omega^2}\right] d\omega.$$

We take the principal value of the integral. It diverges logarithmically at the upper limit. However, we omitted several effects, such as the recoil of the electron upon absorbing the vacuum photons, the effects of retardation, and finally the relativistic variation of mass with velocity. All these combine to cut the integral off effectively at $\omega_{\max} \simeq mc^2/\hbar$, so that

$$\text{(12)} \qquad \Delta W_{fl}(\omega_0) = \frac{1}{\pi}\alpha\left(\frac{\hbar\omega_0}{mc^2}\right)^2\left(\ln\frac{\omega_{\max}}{\omega_0}\right)mc^2.$$

The results for the harmonic oscillator may be used to obtain W_{fl} for any potential V following a suggestion due to Kramers. The transitions in any atom may be considered to be equivalent to oscillators with the frequency $\omega_{ik} = (E_k - E_i)/h$ and of strength $f_{ik} = (2m/3\hbar) \mid (\psi_i, x\psi_k) \mid^2$ for each coordinate. Writing then $\Delta W_{fl} = \sum W_{fl}(\omega_{ik})f_{ik}$ one obtains

$$\text{(13)} \qquad \frac{\Delta W_{fl}}{mc^2} = \frac{\alpha}{3\pi}\left(\frac{\hbar}{mc^2}\right)\langle\nabla^2 V\rangle_{av}\left\langle\ln\frac{mc^2}{E_i - E_k}\right\rangle_{av}$$

where $\langle\nabla^2 V\rangle_{av}$ is the average value of $\nabla^2 V$ in the state being considered, while the $\langle\ln mc^2/(E_i - E_k)\rangle_{av}$ is the average value with respect to the sum over k. Formula (13) was first obtained by Bethe [1]. When applied to the Coulomb potential expression, (13) yields excellent agreement with the results of the Lamb-Retherford experiment. A procedure similar to that outlined above and leading more directly to equation (13) has been given by Welton [26]. We note again that the procedure used involving as it does the term by term subtraction of two infinities cannot be regarded as rigorous and thus requires further justification.

3. Formulation of the problem. We have noted in §2 a number of infinities and a number of places at which these infinities have been cancelled against each other to obtain a finite and therefore interpretable result. We have emphasized the general ambiguity of such a procedure. There are however some restrictions which limit the scope of the ambiguity; indeed in the case of the Lamb-Retherford experiment, one may obtain a result which is unambiguous to well within the accuracy of the experiment, so that the principal part of the effect can be calculated as a consequence of the theory.

From the special theory of relativity we have the general requirement of *covariance*, that is, the predictions of quantum electrodynamics must transform properly when evaluated in different Lorentz frames. Since we are dealing with a divergent theory in which we must be prepared to subtract or rather absorb some infinities, it is essential that the covariance be transparent at each step so that the terms remaining after the infinities are removed will have obvious transformation properties.

From the relations between the electromagnetic potentials and the corresponding fields, we obtain the general requirement of *gauge-invariance*. For this purpose, it is most useful to arrange the theory so that its predictions do not depend in any way upon the particular gauge employed. This is accomplished in Schwinger's [17; 18; 19] formulation by making a covariant separation of the electromagnetic field into its longitudinal and transverse parts and finally eliminating the longitudinal part to obtain a covariant description of the Coulomb interaction between charges. We shall now proceed to describe a covariant-gauge invariant formulation of the interaction of matter with radiation.

4. **A covariant formulation of the equations of motion of quantum mechanics.**[2] We are interested in setting up the equations of motion of two systems, a matter field and a radiation field which interact. The Hamiltonian operator in the absence of interaction, H_0, is the sum of the Hamiltonians for the matter field and radiation field in the absence of interaction. The interaction operator H_{12} is:

$$H_{12} = -\frac{1}{c}\int \mathfrak{H}_\mu(r)A_\mu(r)\,d\tau \tag{14}$$

where $\mathfrak{H}_\mu$ is the current density and A_μ is the four-vector electromagnetic potential, $d\tau$ is the volume element in three-space.

It is clear that the usual Schroedinger equation of motion is not in covariant form:

$$(H_0 + H_{12})\Psi = i\hbar\,\frac{\partial\Psi}{\partial t}\,. \tag{15}$$

Here Ψ is a state vector in Hilbert space. It is now necessary to find a covariant equation of motion which reduces to (15) when a special surface in four-space, t = constant, is employed. A hint as to how this may be done is suggested by the fact that the interaction energy density H_{12} is an invariant. To focus our attention on this term, we use a moving set of axes in Hilbert space, as determined by the

[2] We shall use here the formalism developed by Schwinger [17; 18; 19] and Tomanaga et al. [22; 5].

operator H_0. The motion of the state vectors with respect to this moving frame will then be completely determined by H_{12}. This transformation to the moving set of axes is performed by means of the contact transformation

(16) $$\Psi = e^{-iH_0t/\hbar}\Phi$$

yielding the equation of motion for Φ:

$$e^{iH_0t/\hbar}H_{12}e^{-iH_0t/\hbar}\Phi = i\hbar\frac{\partial\Phi}{\partial\tau}.$$

Under this transformation each operator, which in the Schroedinger representation was independent of the time, will be replaced by an operator with an explicit time dependence:

(17) $$\mathfrak{H}_\mu(r) \rightarrow e^{iH_0t/\hbar}\mathfrak{H}_\mu(r)e^{-iH_0t/\hbar} = \mathfrak{H}_\mu(x)$$

where by $\mathfrak{H}_\mu(x)$ we shall mean $\mathfrak{H}_\mu(r, t)$. Moreover under the transformation each operator will obey its *free* (no interaction) *field equation* of motion so that the time dependence of $\mathfrak{H}_\mu$ for example will be determined by the Dirac equation in the absence of radiation. These equations of motion are summarized below.

Electromagnetic field:

$$\Pi^2 A_\mu = 0, \qquad F_{\mu\nu} = \frac{\partial A_\mu}{\partial x_\nu} - \frac{\partial A_\nu}{\partial x_\mu}$$

are the fields.

In addition we have the Lorentz condition[3]

(18) $$\frac{\partial A_\mu}{\partial x_\mu}\Phi = 0.$$

As a consequence we have the free field Maxwell equation as a supplementary condition

(19) $$\left(\frac{\partial F_{\mu\nu}}{\partial x_\nu}\right)\Phi = 0.$$

The effect of charge is contained in the equations of motion of Φ. A_μ satisfies commutation relations

(20) $$A_\mu(x)A_\nu(x') - A_\nu(x')A_\mu(x) = [A_\mu(x), A_\nu(x')]_- = \frac{\hbar c}{i}\delta_{\mu\nu}D(x - x')$$

where $\delta_{\mu\nu}$ is the Kronecker and $D(x)$ is defined by

(21) $$\Pi^2 D = 0, \qquad D(r, 0) = 0, \qquad \left(\frac{\partial D}{\partial t}\right)_{t=0} = \delta(r).$$

The function D vanishes everywhere but on the surface of the light cone (with apex) at $r, t = 0$.

---

[3] See discussion at beginning of §5 for limitations on (18).

One of the great advantages of the representation (16) used here is the simplicity of the resulting commutation relations.

Electron field:

$$\left(\gamma_\mu \frac{\partial}{\partial x_\mu} + \kappa\right)\psi = 0 \tag{22}$$

where ψ is a four element matrix and γ_μ satisfies $\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2\delta_{\mu\nu}$. κ is a constant proportional to the mass of the electron. The adjoint of ψ is written $\bar{\psi}$ and satisfies $\bar{\psi}(\gamma_\mu\partial/\partial x_\mu - \kappa) = 0$. The current density is

$$\mathfrak{H}_\mu = -\frac{iec}{2}[\bar{\psi}\gamma_\mu\psi - \psi\tilde{\gamma}_\mu\bar{\psi}]. \tag{23}$$

$\tilde{\gamma}_\mu$ is the transpose of γ_μ. The commutation rules are

$$\psi_\alpha(x)\bar{\psi}_\beta(x') + \bar{\psi}_\beta(x')\psi_\alpha(x) = [\psi_\alpha(x), \bar{\psi}_\beta(x')]_+ = i_\alpha\delta_\beta(x - x'). \tag{24}$$

Here the matrix

$$\delta(x) = (\gamma_\mu\, \partial/\partial x_\mu - \kappa)\Delta(x)$$

where

$$(\Pi^2 - \kappa^2)\Delta = 0, \qquad \Delta(r, 0) = 0, \qquad \left(\frac{\partial\Delta}{\partial t}\right)_{t=0} = \delta(r). \tag{25}$$

We now return to the equation of motion for Φ. Writing $e^{iH_0t/\hbar}H_{12}e^{-iH_0t/\hbar} = -(1/c)\int\mathfrak{H}_\mu(x)A_\mu(x)\,d\tau$ as $\int\mathcal{H}(x)\,d\tau$ the equation of motion reads

$$\left[\int \mathcal{H}(x)\,d\tau\right]\Phi = i\hbar\frac{\partial\Phi}{\partial t}. \tag{26}$$

Replacing, for the purposes of clarity, the integral by a Riemann sum we have

$$[\sum_i \mathcal{H}(r_i, t)\Delta\tau_i]\Phi = i\hbar\frac{\Delta\Phi}{\Delta t}$$

or

$$[\sum_i \mathcal{H}(r_i, t)\Delta\omega_i]\Phi = i\hbar c\Delta\Phi \tag{27}$$

where $\Delta\omega_i$ is a volume element in four-space located at (r_i, t). It is now possible to state the proper generalization of the equation of motion of Φ by asking how does Φ change on going from one spacelike surface σ to another, σ'. Suppose the change from σ to σ' occurs at a particular point (r, t). Then because of the fact that σ is a spacelike surface, and since

$$[\mathcal{H}(x), \mathcal{H}(x')]_- = 0 \qquad \text{on a spacelike surface} \tag{28}$$

it is possible to say that the only term in (27) which will be effective is the one located at (r, t). Therefore $i\hbar c\Delta\Phi = \mathcal{H}(x)\Delta\omega\Phi$ or

$$i\hbar c\frac{\delta\Phi}{\delta[\sigma]} = \mathcal{H}(x)\Phi \tag{29}$$

where $\delta\Phi/\delta[\sigma]$ is the variational derivative of Φ at x, where σ is a spacelike surface passing through x. The state vector Φ, which in (26) was a function of t, a particular spacelike surface, is now a function of the parameters describing the possible σ surfaces considered by (29), or better still it is a functional dependent upon the details of the particular surface σ considered. From the derivation it is clear that (29) is equivalent to (26) if and only if (28) is satisfied. Equation (29) is a covariant description of the equation of motion of Φ; it does not refer to any particular Lorentz frame.

The covariance exhibited by (29) is not enough to assure covariance for the entire theory. In addition, it is necessary to demonstrate the covariance of the procedures used and their resulting equations. The calculation will proceed by a series of contact transformations of the type

$$\Phi' = e^{-iS}\Phi. \tag{30}$$

Covariance requires that S be an invariant, a condition which is met in electromagnetic theory. However it should be noted that because of the presence of infinite self-energy terms discussed in §2, S in the electromagnetic case is in general singular and the contact transformation (30) non-existent. However, we shall assume that at some future time, the theory will be modified so that it is indeed finite. This may be accomplished by (1) some relativistic cut-off which is equivalent to giving the electron a finite extension in 4-space, or (2) by the use of the f field which would not change the point character of the electron.

The large variety of the possible methods which can be used leads to a non-uniqueness of the final results. There is no way in which we can decide a priori which cut-off method is correct. The proper method will presumably be suggested by experiment.

In method (1), the resultant operators obtained by transformation (30) would no longer be point interactions and it would be impossible to go from the transformed (29) back to the original time dependent equation (26). Thus we see that if it is desired to set up a theory for an electron with structure one should not expect to be able to describe it with the usual Schroedinger equation, which is no surprise. However, by a procedure similar to Lorentz described in section (2), it will be possible to absorb the troublesome terms leaving terms which remain finite when the cut-off is dropped. Indeed because of the ambiguity introduced by the cut-off, it is convenient to take this limit, so that we may discuss those effects which are not sensitive to electron structure. Upon doing this we should then be back to the theory of a point-electron. We may then demand that these final interaction terms should be point functions and moreover should satisfy commutation rule (28). In order then to be certain that the procedures used did not disturb covariance, it is necessary to check these points. For the f-field theories, it is easy to show these requirements are rigorous.

5. **Schwinger's covariant elimination of the longitudinal component of the electromagnetic field.** As was mentioned earlier, the predictions of the theory

must be gauge invariant, which in turn requires a separation of the electromagnetic field by some covariant procedure into its longitudinal and transverse components and the subsequent elimination of the former.

Whatever the choice of gauge, A_μ, in order that it transform as a four vector, must satisfy the supplementary condition (18) $(\partial A_\mu/\partial x_\mu)\Phi = 0$. The elimination of the longitudinal field consists in first writing A_μ in terms of a transverse plus longitudinal field, thus making the choice of gauge explicit, and secondly rearranging the interaction energy density so that it does not depend on the choice of gauge and no longer involves the variables occurring in the revamped supplementary condition.

We must first generalize the supplementary condition (18) $(\partial A_\mu/\partial x_\mu)\Phi[\sigma(x)] = 0$

$$\Omega[\sigma', x]\Phi[\sigma'] = 0 \tag{31}$$

where $\sigma' = \sigma(x')$, x is not necessarily on σ; Ω is an operator satisfying the initial condition

$$\Omega[\sigma(x), x] = \partial A_\mu/\partial x_\mu \tag{32}$$

so that (31) reduces to (18) when x is on σ. The operator Ω must satisfy the equation of motion for any dynamical variable in order that (31) be consistent with the equations of motion of Φ, (29). Therefore

$$i\hbar c\,\frac{\delta\Omega[\sigma', x]}{\delta[\sigma']} + [\Omega(\sigma', x), \mathcal{H}(x')]_- = 0. \tag{33}$$

Noting that $\partial A_\mu/\partial x_\mu$ does not satisfy (33) suggests writing the form

$$\Omega = \frac{\partial A_\mu}{\partial x_\mu} + W,\ (W, \mathcal{H})_- = 0.$$

Then from (33), W satisfies

$$i\hbar c\,\frac{\delta W}{\delta[\sigma']} + \left[\frac{\partial A_\mu}{\partial x_\mu}, \mathcal{H}(x')\right]_- = 0.$$

The commutator may be evaluated using (20) and the equation for W integrated to obtain:

$$\Omega[\sigma', x] = \frac{\partial A_\mu}{\partial x_\mu} - \frac{i}{c}\int_{\sigma'} D(x - x')\mathfrak{H}_\mu(x')\,d\sigma_\mu' \tag{34}$$

where $d\sigma_\mu'$ is a surface element on surface σ. We note, in virtue of the properties of A_μ and D, that

$$\Pi^2\Omega = 0. \tag{35}$$

Employing (33) and (35), we can assert that if (31) holds for a given σ', that it will hold subsequently.

The covariant elimination of the static field is the relativistic generalization of the usual procedure which we shall therefore first outline. In a special Lorentz frame, it is customary to write

$$A_k = -\frac{\partial \Lambda'}{\partial x_k} + B_k, \qquad \frac{\partial B_k}{\partial x_k} = 0,$$

$$A_4 = -\frac{\partial \Lambda}{\partial x_4}, \quad \Pi^2 \Lambda = 0 = \Pi^2 \Lambda'.$$

Then $\partial A_\mu/\partial x_\mu = \partial^2(\Lambda' - \Lambda)/\partial x_4^2$ so that the Lorentz condition is satisfied if $(\Lambda' - \Lambda)\Phi = 0$ and $(\partial(\Lambda' - \Lambda)/\partial x_4)\Phi = 0$. The interaction density is modified from $-(1/c)\ (\mathfrak{H}_\mu A_\mu)$ to

$$\mathfrak{H}_\mu A_\mu = \mathfrak{H}_k B_k - \frac{\partial}{\partial x_\mu}(\mathfrak{H}_\mu \Lambda') + \mathfrak{H}_4 \frac{\partial(\Lambda - \Lambda')}{\partial x_4}.$$

In virtue of the Lorentz condition, the last term vanishes, when applied to Φ, while the next to the last term may be removed by a contact transformation. Thus in the final formulation the choice of gauge, as given through Λ and Λ', no longer affects the interaction energy density; the supplementary condition no longer contains the variables present in the interaction energy density. Thus the longitudinal field A_4 has been eliminated, the electrostatic energy $\mathfrak{H}_4 A_4$ now appearing in the $E^2 + H^2/8\pi$ term of the energy. The covariant generalization of this procedure in which x is no longer required to be on σ' will lead to an explicit covariant formulation of the Coulomb interaction.

The generalization will involve a vector $\mathfrak{A}_\mu$ playing the same role as B_k, and thus satisfying

(36) $$\partial \mathfrak{A}_\mu/\partial x_\mu = 0, \qquad \eta_\mu \mathfrak{A}_\mu = 0.$$

The vector η is a constant timelike vector, $\eta_\mu^2 = -1$. In the special case discussed above $\eta_1 = \eta_2 = \eta_3 = 0$, $\eta_4 = 1$. The decomposition of A_μ is then

(37) $$A_\mu = \eta_\mu \eta_\nu \frac{\partial \Lambda}{\partial x_\nu} - \left(\frac{\partial}{\partial x_\mu} + \eta_\mu \eta_\nu\right)\Lambda' + \mathfrak{A}_\mu.$$

The first of these terms is the component of a gradient of a scalar in the timelike direction and is thus the generalization of $A_4 = -\partial\Lambda/\partial x_4$. The second term is the space component of the gradient of Λ' analogous to $\partial\Lambda'/\partial x_k$, in the equation relating to A_k and B_k.

From the commutation rules satisfied by A_μ, it is possible to set up a consistent set of commutation rules for the new field variables, Λ, Λ', $\mathfrak{A}_\mu$:

$$[\Lambda(x), \Lambda'(x')] = [\Lambda(x), \mathfrak{A}_\mu(x')] = [\Lambda'(x), \mathfrak{A}_\mu(x')]_- = 0,$$

$$[\mathfrak{A}_\mu(x), \mathfrak{A}_\nu(x')]_- = iW_{\mu\nu}(x - x'),$$

(38) $$W_{\mu\nu}(x) = \hbar c\left\{\delta_{\mu\nu} D(x) - \left[\frac{\partial}{\partial x_\mu}\frac{\partial}{\partial x_\nu} + \left(\eta_\mu \frac{\partial}{\partial x_\nu} + \eta_\nu \frac{\partial}{\partial x_\mu}\right)\eta_\lambda \frac{\partial}{\partial x_\lambda}\right]\mathfrak{D}(x)\right\},$$

$$\left(\eta_\mu \frac{\partial}{\partial x_\mu}\right)^2 \mathfrak{D}(x) = D(x).$$

We see that Λ, Λ', and $\mathfrak{A}_\mu$ are independent fields at the cost of a complex commutation relation for $\mathfrak{A}_\mu$. The supplementary condition implies

$$[\Lambda - \Lambda' - \frac{i}{c}\int \mathfrak{D}(x - x')\mathfrak{H}_\mu(x')\, d\sigma'_\mu]\Phi[\sigma] = 0. \tag{39}$$

The operator $\mathcal{H}(x)$ is

$$\mathcal{H}(x) = -\frac{1}{c}\mathfrak{H}_\mu \mathfrak{A}_\mu + \frac{\partial}{\partial x_\mu}\left(\frac{1}{c}\mathfrak{H}_\mu \Lambda'\right) - \frac{1}{c}\eta_\mu \mathfrak{H}_\mu \eta_\nu \frac{\partial}{\partial x_\nu}(\Lambda - \Lambda'). \tag{40}$$

We again make a contact transformation:

$$\Phi[\sigma] = e^{-iG}\Theta[\sigma], \qquad G = \frac{i}{\hbar c^2}\int \mathfrak{H}_\mu \Lambda'\, d\sigma'_\mu.$$

The operator Λ' transforms to

$$e^{iG}\Lambda' e^{-iG} = \Lambda' - \frac{1}{c}\int_{\sigma'} \mathfrak{D}(x - x')\mathfrak{H}_\mu(x')\, d\sigma'_\mu$$

so that the *supplementary condition becomes*

$$(\Lambda - \Lambda')\Theta[\sigma] = 0. \tag{41}$$

$\mathcal{H}(x)$ now transforms to a new operator $H(x)$ so that $H(x)\Theta = i\hbar c\delta\Theta/\delta[\sigma]$

$$\begin{aligned} H(x) = -\frac{1}{c}\mathfrak{H}_\mu \mathfrak{A}_\mu - \int_\sigma &\left[\frac{1}{2}\frac{\partial \mathfrak{D}(x - x')}{\partial x_\mu} + \eta_\mu \eta_\nu \frac{\partial \mathfrak{D}(x - x')}{\partial x_\nu}\right] \\ &\cdot \left[\frac{1}{c^2}\mathfrak{H}_\mu(x)\mathfrak{H}_\lambda(x')\right] d\sigma'_\lambda - \frac{1}{c}\eta_\mu \mathfrak{H}_\mu \eta_\nu \frac{\partial}{\partial x_\nu}(\Lambda - \Lambda'). \end{aligned} \tag{42}$$

When applied to a state vector Θ, the last term may be made to vanish in virtue of (41) and the equations $\Pi^2\Lambda = \Pi^2\Lambda' = 0$. The first term is the interaction with the transverse field $\mathfrak{A}_\mu$, while the second term is the covariant expression for the Coulomb interaction. In this term a dependence on η_μ remains. Since η_μ is an arbitrary timelike vector, it is necessary for gauge invariance of the theory that the final results be independent of any particular choice for η_μ.

6. **Applications of the new formulation.** In the preceding two sections we have discussed the procedural innovations which have been introduced in order to render the theory Lorentz and gauge-invariant. We must now discuss the manner in which the relevant expressions for a given physical process are computed with particular reference to the self-energy of the photon and electron and finally the Lamb-Retherford shift.

The method of calculation is that of perturbation theory carried out by a series of contact transformations. We begin by considering the operator $H(x)$. With respect to the process in question, some part of $H(x)$ may contribute a term directly without involving any higher transitions through intermediate states. The higher order transitions may be included by transforming, via a contact transformation, the remainder of $H(x)$ into the next order. The transformed operator is now examined and again the terms giving direct contributions are separated off, and the remainder transformed on to the next order, and so on. The main advantage of this technique over customary perturbation theory lies in the possibility of being able to express each of these contact transformations in terms of integrals over space-time. A close correspondence to the results of classical theory is also established. Thus an examination of the behavior of the transformation and of perturbation theory under a Lorentz transformation becomes possible. We may recall from §3 that one of the requirements on the contact transformation is that it be an invariant.

We shall illustrate this procedure by outlining the calculation of the self-energy of the photon. We thus have only one photon present, whose transverse vector potential $\mathcal{A}_\mu = a_\mu e^{ik_\lambda x_\lambda}$ where $a_\mu k_\mu = 0$ from $\partial \mathcal{A}_\mu / \partial x_\mu = 0$ and $k_\mu^2 = 0$ from $\Pi^2 \mathcal{A}_\mu = 0$. The proper equation of motion is

$$H(x)\Theta = i\hbar c \frac{\delta\Theta}{\delta[\sigma]} \tag{43}$$

where $H(x)$ is given by (42). From (42) we see that $\mathcal{A}_\mu$ enters in H only linearly and thus cannot contribute to the photon self-energy which must contain terms which are at least bilinear in $\mathcal{A}_\mu$. We therefore introduce a transformation S

$$e^{iS}\Theta = \Theta'.$$

The equation of motion satisfied by Θ' is

$$e^{iS} H e^{-iS} \Theta' = i\hbar c e^{iS} \frac{\delta}{\delta[\sigma]} [e^{-iS} \Theta'].$$

H is proportional to the electric charge e. Expanding to second order in e on both sides:

$$[H + i(S, H)]_{-}\Theta' = i\hbar c \frac{\delta\Theta'}{\delta[\sigma]} + i\hbar c \left[-i \frac{\delta S}{\delta\sigma} + \frac{1}{2}\left(S, \frac{\delta S}{\delta\sigma}\right)\right]\Theta'.$$

We eliminate H by choosing S so that

$$\hbar c \frac{\delta S}{\delta\sigma} = H(x)$$

or

$$S = \frac{1}{\hbar c} \int^{\sigma} H(x')\, d\omega \tag{44}$$

so that Θ' obeys the equation

$$\frac{i}{2} (S, H)\Theta' = i\hbar c \frac{\delta\Theta'}{\delta[\sigma]}. \tag{45}$$

The commutator (S, H) does contain terms which are bilinear in $\mathcal{A}_\mu$ for

$$(S, H) = \frac{1}{\hbar c}\int^{\sigma} [H(x'), H(x)]_{-}\, d\omega'.$$

The terms in the integrand bilinear in $\mathcal{A}_\mu$ are $[\mathfrak{S}_\mu(x')\mathcal{A}_\mu(x'), \mathfrak{S}_\nu(x)\mathcal{A}_\nu(x)]_{-}$ which in turn equals

$$\mathfrak{S}_\mu(x')\mathfrak{S}_\nu(x)(\mathcal{A}_\mu(x'), \mathcal{A}_\nu(x))_{-} + (\mathfrak{S}_\mu(x'), \mathfrak{S}_\nu(x))_{-}\mathcal{A}_\nu(x)\mathcal{A}_\mu(x').$$

The first term, in virtue of the commutation rules between $\mathcal{A}_\mu$ and $\mathcal{A}_\nu$, is independent of $\mathcal{A}_\mu$ so that only the second term will contribute. Thus the relevant term for the self-energy of the photon is

$$(46) \qquad \frac{i}{2\hbar c}\int^{\sigma} \mathcal{A}_\nu(x)\mathcal{A}_\mu(x')\, \langle\, [\mathfrak{S}_\mu(x'), \mathfrak{S}_\nu(x)]_{-}\, \rangle_{\text{vac}}\, d\omega'$$

where $\langle[\mathfrak{S}_\mu(x'), \mathfrak{S}_\nu(x)]_{-}\rangle_{\text{vac}}$ is the expectation value for the vacuum (no charged particles present). The appearance of this multiplying factor demonstrates that the photon self-energy arises from the vacuum fluctuations of the charge and current densities. (Of course $\langle\mathfrak{S}_\mu\rangle_{\text{vac}} = 0$.) The remaining terms of $(S, H)_{-}$ which we have not considered include (1) terms corresponding to the electron self-energy, (2) corrections to the Coulomb interaction between charged particles (bilinear in $\mathfrak{S}_\mu$), (3) terms corresponding to the production of quanta upon electron-electron scattering (trilinear in $\mathfrak{S}_\mu$ and linear in $\mathcal{A}_\mu$). The terms $\mathcal{A}_\mu\mathcal{A}_\nu[(\mathfrak{S}_\mu, \mathfrak{S}_\nu)_{-} - \langle(\mathfrak{S}_\mu, \mathfrak{S}_\nu)_{-}\rangle_{\text{vac}}]$ involve such processes as Compton scattering. Of course our results are correct to only second order in e.

We would become involved in too many details if we should now proceed to evaluate (46) explicitly. It is sufficient to note that the integrals involved are divergent with the consequence that the result is ambiguous. The use of a procedure which is clearly Lorentz and gauge invariant is not sufficient unless these features are maintained even in the evaluation of the integrals. An important criterion for testing the validity of an integration procedure requires that the equation of continuity of charge and current be satisfied at each stage of the calculation. To apply this criterion to the present problem, it is necessary to notice that

$$\langle\mathfrak{S}_\mu\rangle_{\text{induced}} = \frac{i}{\hbar c^2}\int^{\sigma} \langle(\mathfrak{S}_\mu(x), \mathfrak{S}_\nu(x'))_{-}\,\rangle_{\text{vac}}\, \mathcal{A}_\mu(x')\, d\omega'$$

is just the current induced by the vacuum fluctuations of charge, that is, the current induced by the "polarization of the vacuum". In the absence of any real charge or current $\langle\mathfrak{S}_\mu\rangle$ induced must be zero for conservation of charge. It is thus not surprising that the self-energy of a photon turns out to be zero when we insist upon this condition. However, it must be emphasized at this point that calculations yielding any other result, such as a finite value for the self-energy of the photon, have violated the principle of conservation of charge. The general procedure described in this review has the merit of making this point obvious. Moreover, it now sets up general integration rules and procedures for evaluating certain ambiguous integrals, which for consistency must be maintained whenever they arise again.

A similar discussion may be made for the electron self-energy which arises from another part of $(S, H)_-$. It may be demonstrated that the electron self-energy operator is proportional to $(\bar{\psi}\psi + \psi\bar{\psi})$ and thus transform under a Lorentz transformation as an addition to the electron mass. This addition may be performed explicitly by means of a contact transformation though it is sometimes just as convenient to carry it along in the interaction energy. The experimental mass of the electron will then be a sum of the initial mechanical mass and the electromagnetic mass just described. However, as expected, the multiplicative constant is a divergent integral. Here the assumption (or philosophy) is made that the present theory is incorrect at higher energies, that a correct theory would effectively introduce an invariant cut-off for this divergent integral without disturbing the foundations, for example, commutation rules, which we have employed. In addition it must be assumed that, since we are using perturbation theory, the electromagnetic mass must be a small fraction of the total mass. On the other hand, the cut-off must occur in an energy region which is very much above the regions for which the theory has already been experimentally verified. It is easy to meet both of these conditions for the Dirac electron because the self-energy only diverges logarithmically.

We turn now to the Lamb-Retherford experiment in which the interest is focussed on the energy of an electron moving in external field of force. The modification introduced into our considerations above is that H in (43) is not replaced by $H + H^{\text{ext}}$ where

$$H^{\text{ext}} = -\frac{1}{c}\mathfrak{H}_\mu\mathfrak{A}_\mu^{\text{ext}}$$

where $\mathfrak{A}_\mu^{\text{ext}}$ is the external field, and is of course not an operator. The calculation then proceeds in the same manner as that outlined for the self-energy of the photon; H^{ext} is of course not transformed away according to the general scheme outlined in this section. The relevant terms in the interaction energy become

$$H_{\text{ext}} \rightarrow H^{\text{ext}} + i(S, H^{\text{ext}})_- + [i(T, H^{\text{ext}})_- - 2^{-1}(S, (S, H^{\text{ext}})_-)_-$$

where S was defined in (44) and

$$T = \frac{1}{\hbar c}\int^{\sigma}\frac{i}{2}(S, H)\ d\omega' .$$

The second term is the term which gives rise to radiation processes such as occur in X-ray production. It also affects the electron energy to the same order as the third term but we have not carried out this transformation explicitly. The principal term is the third term. It may be shown to satisfy the criteria outlined in §3. There is, however, still one infinity which must be properly disposed of. An external field polarizes the vacuum. In the Lamb-Retherford experiment the Coulomb field due to the proton nucleus will polarize the vacuum, and the charge on the proton would be substantially changed. The experimentally observed charge is just the sum of the actual charge plus the induced charge. In the calculation this turns up in a rather simple fashion for a part of the third term comes out to be just proportional to H_{ext} and therefore may be considered to change the interaction constant (proportional to the charge) to the experimental value. Again this change involves a divergent integral and the same sort of discussion that occurred in the photon and electron self-energy would have to be repeated here. Suffice it to say that once this term is removed by renormalization of the charge, the calculation of the Lamb-Retherford shift may be performed. It is substantially independent of any of the cut-off procedures discussed above.

7. **Discussion.** The methods and calculations outlined in the previous section give rise to a number of questions, some of which are now under investigation. For example:

(a) Will the renormalization of charge and mass suffice in higher order effects to maintain finiteness of the theory? There may be high energy phenomena which are sensitive to the cut-off but low energy phenomena should be relatively independent.

(b) What are the general conditions on a field theory permitting the application of the program of sections (3), (4), and (5)? For example, will it be possible to apply these procedures to meson theories and to the theory of nuclear forces?

(c) Can these methods be applied to the situation where perturbation methods no longer apply, that is, to so-called strong coupling theories?

The theory as discussed here keeps the usual interaction energy density of matter and radiation intact. It thus maintains as close a connection with phenomena which have already been explored and understood as is possible. This, however, results in a very complex picture of the vacuum which cannot be considered as very satisfying. The theory of the future may perhaps lead to a simpler vacuum picture.

Finally, one might ask how essential is the relativistic formulation and the particular use of invariant contact transformations? This question has been answered for the Lamb-Retherford effect by a calculation carried out by French and Weisskopf, using the more customary techniques. They obtain the same results proving the older "subtraction" theories were indeed properly covariant.

8. **Acknowledgments.** The writer is very greatly indebted for many clarifying discussions with both Professors Schwinger and Weisskopf. It has been a great privilege to witness the development of their attack on this problem. He has also greatly benefited from a joint work carried on in this field with Dr. H. Goldstein, and from a series of discussions with Dr. H. S. Snyder and from the discussions carried on at the Pocono conference on theoretical physics.

Bibliography

1. H. A. Bethe, Physical Review vol. 72 (1947) p. 339.
2. F. Bopp, Annalen der Physik vol. 38 (1940) p. 345.

3. B. French and V. F. Weisskopf, in press.
4. H. M. Foley and P. Dusch, Physical Review vol. 72 (1947) p. 256.
5. I. Koba, T. Tati, and S. Tomonaga, Progress of Theoretical Physics vol. 2 (1947) p. 101; vol. 2 (1947) p. 198.
6. P. Kusch and H. M. Foley, Physical Review vol. 72 (1947) p. 256.
7. W. E. Lamb, Jr., and R. C. Retherford, Physical Review vol. 72 (1947) p. 241.
8. L. Landau, Journal of Physics, USSR. vol. 2 (1940) p. 483.
9. H. A. Lorentz, *Theory of electrons.*
10. J. E. Mack and N. Austern, Physical Review vol. 72 (1947) p. 972.
11. J. E. Nafe, E. B. Nelson, and I. I. Rabi, Physical Review vol. 71 (1947) p. 914.
12. D. E. Nagle, H. S. Julian, and J. R. Zacharias, Physical Review vol. 72 (1947) p. 971.
13. J. R. Oppenheimer, H. Snyder, and R. Serber, Physical Review vol. 57 (1946) p. 75.
14. A. Pais, Verhandelingen K. Akademia van Wetenschappen, Eerste sectie, vol. 19 (1947).
15. H. Poincaré, *Le mecanique nouvelle*, Paris, Gauthier-Villars, 1924.
16. S. Sakata, Progress of Theoretical Physics vol. 2 (1947) p. 30.
17. J. Schwinger, *Notes on a series of lectures given at the Pocono Conference on Theoretical Physics*, April, 1948.
18. ———, *Quantum electrodynamics*, mimeographed notes, June, 1948.
19. ———, Physical Review vol. 73 (1948) p. 315.
20. Robert Serber, Physical Review vol. 49 (1936) p. 545.
21. E. C. G. Stückelberg, Helvetian Physical Acta vol. 14 (1941) p. 51.
22. S. Tomonaga, Progress of Theoretical Physics vol. 1 (1946) p. 27.
23. V. F. Weisskopf, Zeitschrift für Physik vol. 89 (1934) p. 27; vol. 89 (1934) p. 817.
24. ———, Det. K. Danski Videnskabernes Selskab vol. 14 (1936) p. 1.
25. ———, Physical Review vol. 56 (1939) p. 72.
26. T. A. Welton, in press.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
CAMBRIDGE, MASS.

A METHOD OF ANALYTICAL CONTINUATION IN THE EIGENVALUE AND SCATTERING PROBLEMS OF QUANTUM THEORY[1]

BY

EUGENE FEENBERG

The formal infinite series for the probability amplitudes in scattering problems are transformed by (1) a repetitive procedure which separates out roundabout terms from all orders beyond the second and combines them to produce a common factor multiplying all orders, (2) a procedure of summation to a closed form (essentially an analytical continuation) replacing the above common factor by a generalized complex energy denominator, and (3) unlimited repetition of 1 and 2. Procedure 2 is based on the generalized energy quantity

$$\mathcal{E}_{gh\cdots np} = E_p + \sum_{q\neq gh\cdots np} \frac{V_{pq}V_{qp}}{(q)} + \sum_{qr\neq gh\cdots np} \frac{V_{pq}V_{qr}V_{rp}}{(q)(r)} + \cdots$$

and the formal identity

$$[E - \mathcal{E}_{gh\cdots np}]^{-1} = \frac{1}{(p)}\left[1 + \sum_{q\neq gh\cdots n} \frac{V_{pq}V_{qp}}{(p)(q)} + \sum_{qr\neq gh\cdots n} \frac{V_{pq}V_{qr}V_{rp}}{(q)(p)(r)} + \cdots\right]$$

employing the notation $(x) = E - E_x$.

A repetitive term is defined as one in which duplications occur among the quantum numbers labeling the matrix elements associated in products. As a result of the transformation all such terms are removed from the explicit products of matrix elements and relegated to relatively harmless positions in the generalized energy denominators.

A closely related transformation is applicable to eigenvalue problems.

---

[1] Published in Physical Review vol. 74 (1948) pp. 206–210 under the title *A note on perturbation theory*. See also *Theory of scattering processes*, Physical Review vol. 74 (1948) pp. 664–669.

WASHINGTON UNIVERSITY,
ST. LOUIS, MO.

ELECTROMAGNETISM WITHOUT METRIC

BY

J. L. SYNGE

1. **Introduction.** Einstein's theory of gravitation (and indeed the whole of Riemannian geometry) is based on the study of a *symmetric* covariant tensor g_{mn}. If the curvature tensor R_{mnrs} vanishes, and only then, it is possible to choose throughout space-time a particularly simple system of coordinates, viz. that system for which the matrix g_{mn} becomes diagonal with elements ± 1. In relativity these elements are $(1, 1, 1, -1)$ or $(-1, -1, -1, 1)$. Such coordinates may be called *canonical*; transformations of canonical coordinates form the Lorentz group.

If the historical order of development had been reversed and the general theory of relativity had preceded the special theory, we might have learned to attach a direct physical meaning to the gravitational potentials g_{mn} before rigid bodies or clocks had intruded themselves as essential elements of the theory. Instead of regarding the rectangular Cartesian coordinates and time of the special theory of relativity as the results of measurement by such ideal pieces of apparatus, we would simply say that they formed a canonical system of coordinates in the absence of gravitational field. In fact, it is possible that an intelligent race of beings might have been led to the concepts of rigid bodies and clocks, not through direct observation of physical objects approximating conditions of rigidity and isochronism, but through a mathematical curiosity regarding particularly simple coordinate systems in those particular gravitational fields in which the curvature tensor vanished.

Electromagnetic fields appear in relativity in the form of *skew-symmetric* covariant tensors. The outstanding problem of unified field theory is to weld together into a homogeneous mathematical structure the *symmetric* tensor of gravity and the *skew-symmetric* tensor of electromagnetism. Now Einstein in 1916 had available a wide knowledge of the theory of the symmetric tensor; this theory was not developed ad hoc; it represented the work of geometers stretching back through many decades. The theory of skew-symmetric tensors was then, and is still, much less fully developed. Can we hope to work out a successful general electromagnetic theory without a knowledge of skew-symmetric tensors comparable to our knowledge of symmetric tensors? It seems hardly likely.

This paper is therefore devoted to the study of skew-symmetric tensors. To bring out the intrinsic properties of the tensors it seems best to study them per se rather than against a background of Riemannian geometry. This means that a metric tensor g_{mn} is *not* given.

A contribution to the geometry of a single skew-symmetric tensor field has been made by H. C. Lee [4]. However, for our purposes, the geometry of a single skew-symmetric tensor is not sufficiently rich. We shall suppose that *two* skew-symmetric tensors (P_{mn} and Q_{mn}) are given. This is done to take care of the twelve components of the vectors $\boldsymbol{E}$, $\boldsymbol{D}$, $\boldsymbol{H}$, $\boldsymbol{B}$, of electromagnetic theory.

The name *Maxwellian geometry* is suggested as a title for the theory of a pair of skew-symmetric covariant tensors in four dimensions. From the standpoint of numbers of components, Maxwellian geometry is not much more complicated than the corresponding Riemannian geometry; in Maxwellian geometry we have twelve components, in Riemannian geometry ten. Since skew-symmetric tensors are sensitive with regard to dimensionality (particularly the change from even to odd), it would seem at this stage wiser to stick to four dimensions than to attempt a general treatment in N dimensions.

To plan an attack of Maxwellian geometry, we turn to Riemannian geometry for suggestions. Actually Riemannian geometry has developed through generalizations, and it took two thousand years to do it. For our purposes we consider Riemannian geometry as starting with an assigned symmetric tensor g_{mn}, free of conditions save those of analytic smoothness and nonsingularity (det $g_{mn} \neq 0$). On this basis we develop the general aspects of Riemannian geometry. Then we impose on g_{mn} certain invariant partial differential equations, and derive the properties of Einstein spaces, spaces of constant curvature, and flat spaces. In the last case we are particularly interested in the existence of special coordinate systems (Cartesian), because their use greatly simplifies the formalism in a flat space.

To carry out such a plan in Maxwellian geometry, we should first study the geometry of a pair of skew-symmetric tensors without imposing any particular conditions on them. This would give us the analogue of general Riemannian geometry. We would then proceed to impose invariant conditions and study the resulting special geometries, seeking special coordinate systems to simplify the formalism.

The invariant conditions imposed may be algebraic or differential. In Riemannian geometry the usual algebraic conditions concern the signature of the form $g_{mn}\, dx^m\, dx^n$, in particular those conditions that make it positive definite or give it the signature appropriate to relativity. Such conditions are in the nature of inequalities; the corresponding equations would describe singular cases, which do not seem to have been extensively discussed, and which are perhaps not of much interest.

In Maxwellian geometry we can at once write down three invariant algebraic conditions which may be imposed if we wish to restrict the geometry. They are

(1.1) $$\epsilon^{mnrs} P_{mn} P_{rs} = 0,$$

(1.2) $$\epsilon^{mnrs} Q_{mn} Q_{rs} = 0,$$

(1.3) $$\epsilon^{mnrs} P_{mn} Q_{rs} = 0.$$

Here ϵ^{mnrs} is the usual four-dimensional permutation symbol, taking the value zero unless the suffixes are distinct, and the values $+1$ or -1 according as they form an even or odd permutation of 1234. The expressions written above are relative invariants, and so the equations hold in all coordinate systems if they hold in one. The left-hand sides of (1.1) and (1.2) are actually the square roots of the determinants of P_{mn} and Q_{mn}.

As regards differential equations which we may impose on the two tensors, the left-hand sides of the following equations are well known tensors:

(1.4) $$P_{mn,r} + P_{nr,m} + P_{rm,n} = 0,$$ $$Q_{mn,r} + Q_{nr,m} + Q_{rm,n} = 0.$$

Here the comma denotes partial differentiation. These may be written in the equivalent form

(1.5) $$\epsilon^{mnrs} P_{mn,r} = 0, \qquad \epsilon^{mnrs} Q_{mn,r} = 0.$$

Since operation on either of these with $\partial/\partial x^s$ gives an identity, we are to regard (1.4) as containing only six independent conditions on the two tensors.

In §5 we shall meet other invariant partial differential equations.

To get suggestions from the physical side, let us make an "identification" as usually made in the discussion of electromagnetism in the special theory of relativity:

(1.6)
$$\begin{aligned}
(\ P_{14}, \quad P_{24}, \quad P_{34}) &= \boldsymbol{E} = \text{electric vector},\\
(\ P_{23}, \quad P_{31}, \quad P_{12}) &= \boldsymbol{B} = \text{magnetic induction},\\
(-Q_{14}, -Q_{24}, -Q_{34}) &= \boldsymbol{H} = \text{magnetic vector},\\
(\ Q_{23}, \quad Q_{31}, \quad Q_{12}) &= \boldsymbol{D} = \text{electric displacement}.
\end{aligned}$$

Then the invariant relations (1.1), (1.2), and (1.3) correspond respectively to

(1.7) $$\boldsymbol{E}\cdot\boldsymbol{B} = 0, \qquad \boldsymbol{H}\cdot\boldsymbol{D} = 0, \qquad \boldsymbol{E}\cdot\boldsymbol{D} - \boldsymbol{B}\cdot\boldsymbol{H} = 0.$$

We recognize that (1.1), (1.2), (1.3) correspond to rather special types of electromagnetic fields, and we shall therefore not impose any of these three conditions, since they would restrict too much the scope of the theory. Instead we consider the invariant equation

(1.8) $$\epsilon^{mnrs} P_{mn} P_{rs} + \epsilon^{mnrs} Q_{mn} Q_{rs} = 0.$$

This corresponds to

(1.9) $$\boldsymbol{E}\cdot\boldsymbol{B} - \boldsymbol{H}\cdot\boldsymbol{D} = 0.$$

We certainly expect to have this relation satisfied in vacuo, and we shall in fact restrict ourselves to tensor pairs satisfying (1.8); the full significance of this equation will appear later.

If we substitute from (1.6) in (1.4) we get the usual Maxwell equations:

$$(1.10)\qquad \begin{aligned} \nabla\cdot \boldsymbol{B} &= 0, & \partial \boldsymbol{B}/\partial x^4 &= -\nabla \times \boldsymbol{E}, \\ \nabla\cdot \boldsymbol{D} &= 0, & \partial \boldsymbol{D}/\partial x^4 &= \nabla \times \boldsymbol{H}. \end{aligned}$$

At this point we could complete our non-metrical electromagnetism at once if we could add to (1.10) the "structural equations"

$$(1.11)\qquad \boldsymbol{D} = \boldsymbol{E}, \qquad \boldsymbol{B} = \boldsymbol{H}.$$

Translated back by (1.6), these equations read

$$(1.12)\qquad \begin{aligned} Q_{23} &= P_{14}, & Q_{14} &= -P_{23}, \\ Q_{31} &= P_{24}, & Q_{24} &= -P_{31}, \\ Q_{12} &= P_{34}, & Q_{34} &= -P_{12}. \end{aligned}$$

But these equations are not invariant under general coordinate transformations; therein lies the crux of our problem.

The idea of dealing with electromagnetism without metric occurred to van Dantzig [2; 3] some time ago. However, his approach was somewhat different. He did not limit the data in space-time to the pair of electromagnetic tensors only. He introduced a "structural tensor" M^{rs}_{nm}, skew-symmetric in the upper and in the lower suffixes so that there are 36 components. Then (1.12) are replaced by the structural equations

$$(1.13)\qquad Q_{mn} = M^{rs}_{mn} P_{rs}.$$

Equations (1.4) and (1.13) yield 12 equations for the 12 components of P_{mn} and Q_{mn}. At first sight that seems to be as it should be—12 equations for 12 unknowns—until we recall that in the general theory of relativity we have only *six* equations for *ten* unknowns (g_{mn}), on account of the Bianchi identities (four in number) which link together the field equations ($R_{mn} = 0$). This we must expect. The four-fold freedom introduced by the Bianchi identities corresponds to the four-fold freedom which we have in the choice of coordinate system. *In fact, any field theory, expressed in a form independent of coordinate system, should display this four-fold freedom.* It appears therefore that van Dantzig's theory is over-determined.

In the present stage of the present theory the question of determination or over-determination cannot be raised, because we have been only feeling our way in Maxwellian geometry, and have not committed ourselves yet, except to the single equation (1.8). We shall reserve judgment with regard to the admission of (1.4)—the Maxwell equations.

It is not the purpose of the present paper to present a systematic treatment of

Maxwellian geometry. Our interest will center on the structural equations (1.12). We ask: *Granted that these equations are not invariant under general transformation of coordinates, is it possible to choose coordinates such that these equations are true throughout a domain of space-time?* This question is analogous to, but mathematically very different from, the question of the possibility of reducing a symmetric tensor g_{mn} to a diagonal matrix $(1, 1, 1, -1)$.

I am very much indebted to Mr. G. H. F. Gardner for stimulating discussions on the material of this paper. He has prepared for publication a paper on the algebraic problem of the existence of canonical coordinates at a point, and consequently this algebraic problem is not discussed in the present paper (see Theorem III). I have also to thank my colleague, Professor E. G. Olds, for valuable assistance in connection with §2.

In discussions at the Symposium Professor A. H. Taub pointed out the significance of conformal transformations in connection with §6. As a result of his kind assistance, that section has been completely revised and greatly improved.

2. **Notation and definitions.** Small Latin suffixes take the range 1, 2, 3, 4, with the usual convention for summation with respect to repeated suffixes. The summation convention does not apply to capital suffixes, and different capital suffixes in an equation indicate different numbers from the range 1, 2, 3, 4.

Coordinates are denoted by x^r or $x^{r'}$ or $x^{r''}$, the primes being attached to the suffix in accordance with the practice introduced by Schouten. Real coordinates are used throughout except in §6. The matrix of a transformation is written

$$\frac{\partial x^r}{\partial x^{s'}} = X^r_{s'}, \qquad \frac{\partial x^{r'}}{\partial x^s} = X^{r'}_s, \tag{2.1}$$

so that we have

$$X^r_{s'} X^{s'}_t = \delta^r_t, \qquad X^r_{s'} X^{t'}_r = \delta^t_s. \tag{2.2}$$

For the Jacobians we write

$$X' = \det X^{r'}_s, \qquad X_{\prime} = \det X^r_{s'}, \qquad X' X_{\prime} = 1. \tag{2.3}$$

The following notation is due to Gardner:

$$\begin{aligned}(\epsilon PQ) &= \frac{1}{8}\,\epsilon^{mnrs} P_{mn} Q_{rs} = \frac{1}{2}\,(P_{23}Q_{14} + P_{31}Q_{24} + P_{12}Q_{34} \\ &\qquad + P_{14}Q_{23} + P_{24}Q_{31} + P_{34}Q_{12}), \\ (\epsilon PP) &= \frac{1}{8}\,\epsilon^{mnrs} P_{mn} P_{rs} = P_{23}P_{14} + P_{31}P_{24} + P_{12}P_{34}.\end{aligned} \tag{2.4}$$

We note that

$$\det P_{mn} = (\epsilon PP)^2. \tag{2.5}$$

We define the *dual* T^{*mn} of any skew-symmetric covariant tensor T_{mn} by the formula

$$T^{*mn} = \frac{1}{2}\epsilon^{mnrs}T_{rs}, \tag{2.6}$$

or, explicitly,

$$\begin{aligned} T^{*23} &= T_{14}, & T^{*14} &= T_{23}, \\ T^{*31} &= T_{24}, & T^{*24} &= T_{31}, \\ T^{*12} &= T_{34}, & T^{*34} &= T_{12}. \end{aligned} \tag{2.7}$$

T^{*mn} is a relative tensor transforming according to

$$T^{*m'n'} = T^{*rs}X_r^{m'}X_s^{n'}X, . \tag{2.8}$$

We note that the relations (2.7) are somewhat similar to (1.12), but they are not the same on account of the minus signs in (1.12).

Let us examine (1.12) more carefully, because those relations play a basic role in what we have to say. We must remember that we start with a four-dimensional manifold of space-time, and a quite general coordinate system. In choosing a set of four numbers x^r to specify an event, there is no a priori reason to suppose that three of those numbers are in any sense "space-like" and the fourth "time-like". In fact, at this stage no meaning can be attached to the words "space-like" and "time-like". Thus we are not to expect to find one of the coordinates, say x^4, playing a privileged role relative to the other three. But in (1.12) x^4 *does* play a privileged role; we note that as we come down either column the number 4 is held fixed and there is a cyclic permutation of the other three numbers. That is one "objection" to (1.12). Another is that the relations are not symmetric in P and Q. Let us see how these objections are to be met.

Consider the whole set of equations

$$P_{MN} + Q_{RS} = 0, \tag{2.9}$$

which include the six equations (1.12). We shall adopt the symbol $[MNRS]$ to denote the equation (2.9). Since there are 24 permutations of 1234, there are 24 equations of the form (2.9). However, since (2.9) is equivalent to

$$P_{NM} + Q_{SR} = 0,$$

we may write symbolically

$$[MNRS] = [NMSR]. \tag{2.10}$$

With this understanding, we have in (2.9) just twelve distinct equations. We shall now discuss how these twelve equations may be split in four different ways into two sets of six equations, (1.12) being one of those sets of six equations.

From the four numbers 1234, let us pick out one which we shall call a *base.* Let us start with a base D, and let us denote the three other numbers by ABC. These may be written in six different ways as follows:

$$\begin{matrix} ABC & BCA & CAB \\ BAC & CBA & ACB. \end{matrix} \tag{2.11}$$

Writing the base behind and in front, we generate from the first line of (2.11) the six symbols

$$[ABCD] \quad [BCAD] \quad [CABD] \quad [DABC] \quad [DBCA] \quad [DCAB], \tag{2.12}$$

each of which stands for an equation of the type (2.9). Similarly, from the second line of (2.11) we generate

$$[BACD] \quad [CBAD] \quad [ACBD] \quad [DBAC] \quad [DCBA] \quad [DACB]. \tag{2.13}$$

We have displayed in (2.12) and (2.13) twelve of the 24 permutations of 1234: but actually on account of (2.10) the whole 24 permutations are to be regarded as present in (2.12) and (2.13).

We now introduce the following symbols:

$$\begin{aligned} [ABC{:}D] &= \text{the set (2.12) and those derived from them by (2.10);} \\ [BAC{:}D] &= \text{the set (2.13) and those derived from them by (2.10).} \end{aligned} \tag{2.14}$$

Obviously

$$\begin{aligned} [ABC{:}D] &= [BCA{:}D] = [CAB{:}D], \\ [BAC{:}D] &= [CBA{:}D] = [ACB{:}D]. \end{aligned} \tag{2.15}$$

By (2.14) we have

$$[123{:}4] = \{[1234] \quad [2314] \quad [3124] \quad [4123] \quad [4231] \quad [4312]\}, \tag{2.16}$$

which is precisely the set of equations (1.12). Further [213:4] represents the set of equations

$$\begin{aligned} P_{23} &= Q_{14}, & P_{14} &= -Q_{23}, \\ P_{31} &= Q_{24}, & P_{24} &= -Q_{31}, \\ P_{12} &= Q_{34}, & P_{34} &= -Q_{12}, \end{aligned} \tag{2.17}$$

which is (1.12) with P and Q interchanged.

The objections raised earlier are overcome if, instead of taking (1.12) as the fundamental relation, we take as fundamental all the sets of equations represented by $[ABC{:}D]$ and $[BAC{:}D]$. These we shall now write in another form, making use of Fig. 1, which shows six points marked 1, 2, 3, 4, 5, 6. It does not appear possible to make a geometrical representation having the full symmetry implicit in our problem, but it is perhaps helpful to regard 1, 2, 3, 4 as the corners of a square in a horizontal plane and the line 56 as vertical, the whole forming a regular octahedron of which however only eight edges are filled in. In what follows we shall use capital Latin letters for 1234 and capital Greek letters for 56, different letters in any one equation standing for different numbers.

The labels in Fig. 1 are attached as follows. First we attach the label $[\Theta D]$ to the line joining the points Θ and D, for example [51], to the line joining 5 and 1. Now let us start with $D = 4$, although we might as well start with any other value of D. With 4 as base, there are just two 4-index symbols, [123:4] and [213:4]. One of these we attach to [54] and the other to [64]. The choice

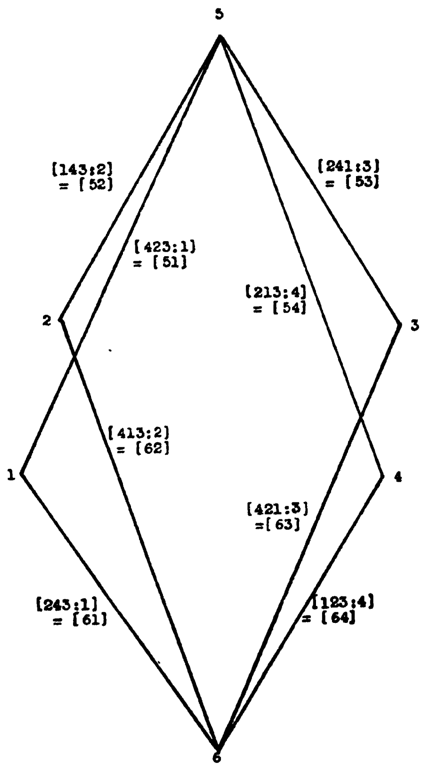


FIG. 1

here is arbitrary, but once it has been made, the rest of the labels are fixed. In the diagram we happen to have chosen

$$[54] = [213:4], \qquad [64] = [123:4].$$

The rest of the labelling is carried out according to the rules

$$(2.18) \qquad \begin{aligned} (AB)[\Theta A] &= [\Phi B], \\ (BC)[\Theta A] &= [\Phi A], \end{aligned}$$

where (AB) and (BC) are substitution symbols applied to the 4-index symbols which correspond to the 2-index symbols. Thus, for example, let us put $A = 4$,

$B = 1, \Phi = 6$ in the first of (2.18). According to our rules we have $\Theta = 5$, and so we get

$$[\Phi B] = [61] = (AB)[\Theta A] = (41)[54] = (41)[213:4] = [243:1].$$

Thus $[61] = [243:1]$, as marked on the diagram. The other labels are found similarly, and it is easy to verify that the rules (2.18) hold without contradiction. We note that there are just two ways of labelling Fig. 1, of which of course only one is shown.

It is clear that [64] represents the equations (1.12), and [54] represents (2.17). We shall now prove the following theorem:

THEOREM I. *Any one of the sets of equations* $[\Theta A]$ *may be transformed into any other by a coordinate transformation consisting of interchange of axes, the signs of the Jacobians being as shown:*

$$(2.19)\qquad \begin{aligned} [\Theta A] &\to [\Phi A] \text{ with } X' = -1, \\ [\Theta A] &\to [\Phi B] \text{ with } X' = -1, \\ [\Theta A] &\to [\Theta B] \text{ with } X' = +1. \end{aligned}$$

Apply the transformation

$$(2.20)\qquad x^{A'} = x^A, \quad x^{B'} = x^C, \quad x^{C'} = x^B, \quad x^{D'} = x^D \qquad (X' = -1).$$

This transforms $[\Theta A]$ into $(BC)[\Theta A] = [\Phi A]$ by (2.18). Thus the first line of (2.19) is established.

Apply the transformation

$$(2.21)\qquad x^{A'} = x^B, \quad x^{B'} = x^A, \quad x^{C'} = x^C, \quad x^{D'} = x^D \qquad (X' = -1).$$

This transforms $[\Theta A]$ into $(AB)[\Theta A] = [\Phi B]$ by (2.18). Thus the second line of (2.18) is established.

Now follow this transformation with

$$(2.22)\qquad x^{A''} = x^{A'}, \quad x^{B''} = x^{B'}, \quad x^{C''} = x^{D'}, \quad x^{D''} = x^{C'} \qquad (X''_{\prime} = -1).$$

This transforms $[\Phi B]$ into $(CD)[\Phi B] = [\Theta B]$ by (2.18), and so the third line of (2.19) is established, since $X'' = X''_{\prime}X' = +1$. Actually in this last case we have used

$$(2.23)\qquad (CD)(AB)[\Theta A] = [\Theta B].$$

Another operation we might mention is that of reversing an axis, denoted by R_A for the x^A axis. It is easily verified that we have

$$(2.24)\qquad R_A[\Theta A] = [\Phi A], \qquad R_B[\Theta A] = [\Phi A].$$

We see that the eight sets of equations $[\Theta A]$ ($\Theta = 5, 6$; $A = 1, 2, 3, 4$) are equivalent to one another in the sense that any one of these sets may be transformed into any other by interchange of axes. Nevertheless complete equivalence has not yet been established, because the Jacobian of the transformation is $+1$

in some cases and -1 in others. We note from (2.19) that the Jacobian is $+1$ if the auxiliary point Θ $(= 5, 6)$ is held fixed, and -1 if it is changed. Referring to Fig. 1, this means that the upper lines transform into one another with Jacobian $+1$, and so do the lower lines, but a transformation from an upper line to a lower line, or vice versa, has Jacobian -1. To bring out this point in our symbolism, we might attach a third label $+1$ to each upper line and -1 to each lower line. Then the Jacobian of the transformation is given by multiplying these labels $(1 \times 1 = 1, -1 \times -1 = 1, 1 \times -1 = -1, -1 \times 1 = -1)$.

However, if we do not restrict ourselves to such simple transformations, and there is no reason why we should, we may establish the following theorem:

THEOREM II. *It is possible to pass from any one of the sets of equations* $[\Theta A]$ *to any other by means of a transformation with Jacobian* $+1$ *or with Jacobian* -1, *at our choice.*

It is immaterial which of the set $[\Theta A]$ we start with; let us accept [64]. That means that we assume that the tensors satisfy [64], that is, the equations (1.12). We shall keep the base coordinate x^4 unchanged and apply to the other three a rotation, that is, an orthogonal transformation with Jacobian $+1$. With respect to such a transformation the following triads behave like vectors:

$$(Q_{23}, Q_{31}, Q_{12}), \qquad (P_{14}, P_{24}, P_{34}), \qquad (Q_{14}, Q_{24}, Q_{34}), \qquad (P_{23}, P_{31}, P_{12}).$$

Moreover the equations (1.12) remain satisfied. These equations tell us that the first two of the above four vectors have a common line, and so also do the last two. Let us rotate the axes $x^1x^2x^3$ so as to make the new x^1 axis perpendicular to these two lines, so that we have

$$Q_{23} = P_{14} = Q_{14} = P_{23} = 0, \tag{2.25}$$

and we have the following equations unchanged from (1.12):

$$\begin{aligned} Q_{31} &= P_{24}, & Q_{24} &= -P_{31}, \\ Q_{12} &= P_{34}, & Q_{34} &= -P_{12}. \end{aligned} \tag{2.26}$$

Now apply the substitution (14)(23), which consists of two interchanges of axes, a transformation with Jacobian $+1$. This leaves (2.25) unchanged, and changes (2.26) into

$$\begin{aligned} Q_{24} &= P_{31}, & Q_{31} &= -P_{24}, \\ Q_{43} &= P_{21}, & Q_{21} &= -P_{43}. \end{aligned} \tag{2.27}$$

But (2.25) and (2.27) together make up the set [213:4], that is, [54]. This means that we have passed from [64] to [54] by a transformation with Jacobian $+1$. We have in fact passed from a lower line of Fig. 1 to an upper line by a transformation with Jacobian $+1$, and so the truth of Theorem II is evident.

Thus we see that all the sets of equations represented by $[\Theta A]$ are fully equivalent, in the sense that we can pass from any one set to any other by a transformation with Jacobian of whichever sign we please, and indeed equal to ± 1. We therefore proceed to make the following definitions:

DEFINITION. *Two skew-symmetric tensors P_{mn} and Q_{mn} are said to be Maxwellian complements of one another in a domain D of space-time if there exists a system of coordinates in D for which* (1.12), *or equivalently any one of the sets of equations* [ΘA], *hold at each point of D. Such a system of coordinates is said to be canonical in D.*

DEFINITION. *Two skew-symmetric tensors P_{mn} and Q_{mn} are said to be Maxwellian complements of one another at a point M of space-time if there exists a system of coordinates such that* (1.12), *or equivalently any one of the sets of equations* [ΘA], *hold at M. Such a system of coordinates is said to be canonical at M.*

The reader should note the following rather subtle point about these two definitions: two tensors may be Maxwellian complements *at each point* of a domain D without being Maxwellian complements in D. It is a question of special systems of coordinates, one for each point, versus one system of coordinates for the whole domain D. This distinction is very important because it involves the idea of integrability.

We shall close this section by writing the equations [ΘA] in a compact notation due to Gardner. We introduce four diagonal matrices with the following diagonal elements:

$$\text{(2.28)} \qquad \eta_{(1)}^{rs} = (-1, 1, 1, 1), \qquad \eta_{(2)}^{rs} = (1, -1, 1, 1), \qquad \eta_{(3)}^{rs} = (1, 1, -1, 1), \qquad \eta_{(4)}^{rs} = (1, 1, 1, -1).$$

Now write the two sets of equations

$$\text{(2.29)} \qquad \eta_{(A)}^{mr}\eta_{(A)}^{ns}Q_{rs} = \frac{1}{2}\epsilon^{mnrs}P_{rs} \qquad (A = 1, 2, 3, 4),$$

and

$$\text{(2.30)} \qquad \eta_{(A)}^{mr}\eta_{(A)}^{ns}P_{rs} = \frac{1}{2}\epsilon^{mnrs}Q_{rs} \qquad (A = 1, 2, 3, 4).$$

If we put $A = 4$ in (2.29), we get (1.12); if we put $A = 4$ in (2.30), we get (2.17). In fact, in the other notation, (2.29) is [$6A$] and (2.30) is [$5A$].

We may add a remark about the geometrical significance of canonical coordinates. Let λ^r, μ^r be any two linearly independent contravariant vectors at a point M of space-time. They define an elementary V_2 at M. The expression $P_{mn}\lambda^m\mu^n$ is an invariant, and so is $Q_{mn}\lambda^m\mu^n$. The ratio

$$\text{(2.31)} \qquad R = \frac{P_{mn}\lambda^m\mu^n}{Q_{mn}\lambda^m\mu^n}$$

is therefore an invariant under coordinate transformations. It has however another invariance, namely, invariance with respect to the choice of the vectors λ^r, μ^r in V_2; in fact, R is invariant under the transformation

$$\lambda^r = a\bar{\lambda}^r + b\bar{\mu}^r, \qquad \mu^r = c\bar{\lambda}^r + d\bar{\mu}^r, \tag{2.32}$$

where a, b, c, d are any four scalars. R is therefore an invariant defined by the elementary V_2 and the tensor pair (P, Q).

We can form six invariants (2.31) by taking for the elementary V_2's those defined by pairs of coordinate lines. These invariants may be written $R_{(AB)} = R_{(BA)}$, and we have

$$R_{(23)} = P_{23}/Q_{23}, \cdots, \cdots, \qquad R_{(14)} = P_{14}/Q_{14}, \cdots, \cdots. \tag{2.33}$$

Consider now the invariants $R_{(AB)}R_{(CD)}$. There are three of these, namely,

$$\begin{aligned} R_{(23)}R_{(14)} &= P_{23}P_{14}/Q_{23}Q_{14}, \\ R_{(31)}R_{(24)} &= P_{31}P_{24}/Q_{31}Q_{24}, \\ R_{(12)}R_{(34)} &= P_{12}P_{34}/Q_{12}Q_{34}. \end{aligned} \tag{2.34}$$

It is clear that if P_{mn} and Q_{mn} are Maxwellian complements and the coordinate system canonical at M, then

$$R_{(AB)}R_{(CD)} = -1. \tag{2.35}$$

Conversely, it can be shown that if (2.35) hold at M, then there exists a system of coordinates canonical at M.

3. **The condition $(\epsilon PP) + (\epsilon QQ) = 0$ and its insufficiency for the existence of canonical coordinates in a domain D.** Throughout we shall assume that the tensors P_{mn} and Q_{mn} are linearly independent; or in other words, the only pair of quantities a, b satisfying $aP_{mn} + bQ_{mn} = 0$ are $a = 0, b = 0$.

THEOREM III. *Two skew-symmetric tensors P_{mn} and Q_{mn} are Maxwellian complements at a point M if, and only if, they satisfy*

$$(\epsilon PP) + (\epsilon QQ) = 0 \tag{3.1}$$

at M, in the notation of (2.4).

The necessity of this condition is obvious if we substitute from (1.12) in (3.1). That the condition is also sufficient has been proved by G. H. F. Gardner in a paper prepared for publication. We shall accept the theorem here without further comment.

We shall prove the following theorem:

THEOREM IV. *The condition* (3.1) *is insufficient to insure that P_{mn} and Q_{mn} are Maxwellian complements in a domain D. Equivalently, two tensors may be Maxwellian complements at each point of a domain D without being Maxwellian complements in D.*

We shall use a reductio ad absurdum. Let us assume that Theorem IV is false; that is, we accept

HYPOTHESIS A. *The condition* (3.1) *is sufficient to insure that* P_{mn} *and* Q_{mn} *are Maxwellian complements in* D.

Let P_{mn} and Q_{mn} be any Maxwellian complements in D and let x^r be canonical coordinates such that (1.12) hold throughout D. Let $\bar{P}$ and $\bar{Q}$ be another pair of skew-symmetric tensors also satisfying (3.1), and let us define skew-symmetric tensors p and q by

$$\bar{P}_{mn} = P_{mn} + p_{mn}, \qquad \bar{Q}_{mn} = Q_{mn} + q_{mn}. \tag{3.2}$$

We shall take p and q to be infinitesimal and so linearize the work. Thus, rejecting product terms in p and q, we have

$$\begin{aligned}(\epsilon\bar{P}\bar{P}) &= [\epsilon(P + p)(P + p)] \\ &= (\epsilon PP) + 2(\epsilon Pp),\end{aligned} \tag{3.3}$$

and a similar equation in Q and q. Since (3.1) is satisfied by both (P, Q) and $(\bar{P}, \bar{Q})$, we have then

$$(\epsilon Pp) + (\epsilon Qq) = 0. \tag{3.4}$$

But P and Q satisfy (1.12); substitute then for Q in terms of P and get from (3.4) the equation

$$\sum(P_{23}B_{14} + P_{14}B_{23}) = 0, \tag{3.5}$$

where the summation sign indicates summation over cyclic permutations of 123, and B is defined by

$$B_{23} = p_{23} + q_{14}, \qquad B_{14} = p_{14} - q_{23}, \tag{3.6}$$

and four other relations obtained by cyclic permutation of 123.

Under Hypothesis A the tensors $\bar{P}$ and $\bar{Q}$ are Maxwellian complements, but there is no reason to suppose that the coordinates x^r which we have been using are canonical coordinates for $\bar{P}$, $\bar{Q}$ as well as for P, Q. However, since these two tensor pairs differ only infinitesimally, there will exist a transformation $x^r \to x^{r'}$, differing only infinitesimally from the identity, which carries us from the canonical coordinates x^r of (P, Q) to the canonical coordinates $x^{r'}$ of $(\bar{P}, \bar{Q})$. Let us write the matrix of this transformation

$$X^r_{s'} = \delta^r_s + Y^r_{s'}, \tag{3.7}$$

where the Y's are infinitesimal functions of position, satisfying the integrability conditions

$$Y^r_{s',t'} = Y^r_{t',s'}, \tag{3.8}$$

where the second subscript indicates partial differentiation. Under this transformation we have

$$\bar{P}_{m'n'} = \bar{P}_{rs}X^r_{m'}X^s_{n'}, \qquad \bar{Q}_{m'n'} = \bar{Q}_{rs}X^r_{m'}X^s_{n'}, \tag{3.9}$$

or, by (3.2) and (3.7), to the first order of infinitesimals,

$$(3.10)\qquad \begin{aligned} \tilde{P}_{m'n'} &= P_{mn} + p_{mn} + P_{rn}Y^r_{m'} + P_{mr}Y^r_{n'}, \\ \tilde{Q}_{m'n'} &= Q_{mn} + q_{mn} + Q_{rn}Y^r_{m'} + Q_{mr}Y^r_{n'}. \end{aligned}$$

But since $x^{r'}$ are canonical coordinates for $(\tilde{P}, \tilde{Q})$, we have by (1.12)

$$(3.11)\qquad \tilde{Q}_{2'3'} = \tilde{P}_{1'4'}, \qquad \tilde{Q}_{1'4'} = -\tilde{P}_{2'3'},$$

and similar relations. When we substitute in these equations from (3.10), we get relations connecting (P, p, q, Y), since we can express Q in terms of P from (1.12). We have also (3.5) connecting (P, p, q).

Let us examine these relations. Substitution from (3.11) in (3.10) gives, when we use (1.12),

$$(3.12)\qquad \begin{aligned} P_{r4}Y^r_{1'} - Q_{r3}Y^r_{2'} - Q_{2r}Y^r_{3'} + P_{1r}Y^r_{4'} &= -B_{14}, \\ Q_{r4}Y^r_{1'} + P_{r3}Y^r_{2'} + P_{2r}Y^r_{3'} + Q_{1r}Y^r_{4'} &= -B_{23}, \end{aligned}$$

together with four other equations obtained from these by cyclic permutation of 123. On carrying out the summations with respect to r, and again making use of (1.12), we get

$$(3.13)\qquad \begin{aligned} &P_{14}(Y^1_{1'} - Y^2_{2'} - Y^3_{3'} + Y^4_{4'}) \\ &\quad + P_{24}(Y^2_{1'} + Y^1_{2'}) + P_{34}(Y^3_{1'} + Y^1_{3'}) \\ &\quad + P_{12}(Y^2_{4'} - Y^4_{2'}) + P_{13}(Y^3_{4'} - Y^4_{3'}) = -B_{14}, \\ &P_{23}(Y^1_{1'} - Y^2_{2'} - Y^3_{3'} + Y^4_{4'}) \\ &\quad + P_{31}(Y^2_{1'} + Y^1_{2'}) + P_{12}(Y^3_{1'} + Y^1_{3'}) \\ &\quad + P_{24}(Y^3_{4'} - Y^4_{3'}) + P_{34}(Y^4_{2'} - Y^2_{4'}) = B_{23}, \end{aligned}$$

together with four more equations obtained by cyclic permutation of 123.

For the purpose of establishing the contradiction, we may take throughout the domain D all the components of P_{mn} zero except $P_{23} = 1$, the burden of making x^r canonical for the pair (P, Q) being placed on Q_{mn}. Then the condition (3.5) on B_{mn} implies only

$$(3.14)\qquad B_{14} = 0,$$

the rest of the functions B_{mn} being arbitrary infinitesimals throughout D. On substituting these special values in (3.13), the first is satisfied identically, and the remainder read

$$(3.15)\qquad \begin{aligned} Y^3_{4'} - Y^4_{3'} &= -B_{24}, \\ Y^2_{4'} - Y^4_{2'} &= B_{34}, \\ Y^1_{1'} - Y^2_{2'} - Y^3_{3'} + Y^4_{4'} &= B_{23}, \\ Y^1_{2'} + Y^2_{1'} &= B_{31}, \\ Y^1_{3'} + Y^3_{1'} &= B_{12}. \end{aligned}$$

If we differentiate the first of (3.15) with respect to $x^{1'}$, or equivalently with respect to x^1 on account of the nearly identical character of the transformation (3.7), we get

$$Y^3_{4',1'} - Y^4_{3',1'} = -B_{34,1}\,. \tag{3.16}$$

Similarly, from the last of (3.15),

$$Y^1_{3',4'} + Y^3_{1',4'} = B_{12,4}\,. \tag{3.17}$$

Subtracting (3.16) from (3.17) and using (3.8), we get

$$(Y^1_{4'} + Y^4_{1'})_{,3'} = B_{24,1} + B_{12,4}\,. \tag{3.18}$$

Similarly, from the second and fourth of (3.15), we get

$$(Y^1_{4'} + Y^4_{1'})_{,2'} = B_{31,4} - B_{34,1}\,. \tag{3.19}$$

From (3.18) and (3.19), we obtain

$$B_{24,12} + B_{12,42} = B_{31,43} - B_{34,13}\,. \tag{3.20}$$

But we have already seen that these B's may be chosen arbitrarily in D; in fact we can choose them so that these equations are *not* satisfied, for example, $B_{12} = x^2x^4$, with the other B's zero. Thus Hypothesis A has led us to a contradiction, (3.20) satisfied and not satisfied at the same time. Thus Hypothesis A is false, and so Theorem IV is true. The result is that we must look beyond (3.1) for a sufficient condition for the existence of canonical coordinates in D.

The following question is interesting, but I have not been able to answer it: Are the conditions

$$\begin{aligned} &(\epsilon PP) + (\epsilon QQ) = 0,\\ &P_{mn,r} + P_{nr,m} + P_{rm,n} = 0,\\ &Q_{mn,r} + Q_{nr,m} + Q_{rm,n} = 0 \end{aligned} \tag{3.21}$$

sufficient to insure that P_{mn} and Q_{mn} are Maxwellian complements in a domain D?

4. **Some integrability conditions.** For a set of total differential equations in N-space:

$$T^{(\mu)}_n\, dx^n = 0 \qquad (\mu = 1, 2, 3, \cdots, M), \tag{4.1}$$

necessary and sufficient conditions of integrability are

$$(T^{(\mu)}_{m,n} - T^{(\mu)}_{n,m})\, dx^m \delta x^n = 0 \qquad (\mu = 1, 2, \cdots, M), \tag{4.2}$$

for all dx, δx satisfying (4.1). This result is due to Frobenius [1].

Consider now the set of four total differential equations in space-time

$$p_{rs}\, dx^s = 0, \tag{4.3}$$

where p_{rs} is a skew-symmetric tensor field. These equations are of course inconsistent unless det p_{rs} vanishes, or equivalently

$$(4.4) \qquad (\epsilon pp) = 0.$$

We shall suppose this condition satisfied in a domain D of space-time. Then only two of (4.3) are independent, since the rank of skew-symmetric matrix is always even [5]. Then the set (4.3) is equivalent to two equations, say to

$$(4.5) \qquad p_{1s}\, dx^s = 0, \qquad p_{2s}\, dx^s = 0.$$

By (4.2), necessary and sufficient conditions for the integrability of (4.5) are

$$(4.6) \qquad \begin{aligned} (p_{1r,s} - p_{1s,r})\, dx^r \delta x^s &= 0, \\ (p_{2r,s} - p_{2s,r})\, dx^r \delta x^s &= 0, \end{aligned}$$

for all dx, δx satisfying (4.5). Since (4.5) define an elementary 2-space, we have to consider only two independent vectors, and we shall take the following, which obviously satisfy (4.5):

$$(4.7) \qquad \begin{aligned} dx^1 &= p_{23}, & dx^2 &= p_{31}, & dx^3 &= p_{12}, & dx^4 &= 0, \\ \delta x^1 &= p_{24}, & \delta x^2 &= p_{41}, & \delta x^3 &= 0, & \delta x^4 &= p_{12}. \end{aligned}$$

Then the first of (4.6) reads explicitly

$$(4.8) \qquad \begin{aligned} -(p_{12,3} - p_{13,2})p_{12}p_{41} + p_{13,1}p_{12}p_{24} \\ -p_{12,1}(p_{23}p_{41} - p_{31}p_{24}) - p_{14,1}p_{23}p_{12} \\ + (p_{12,4} - p_{14,2})p_{31}p_{12} \\ + (p_{13,4} - p_{14,3})p_{12}^2 = 0. \end{aligned}$$

By (4.4) we have

$$(4.9) \qquad p_{23}p_{14} + p_{31}p_{24} + p_{12}p_{34} = 0,$$

so that

$$(4.10) \qquad p_{23}p_{41} - p_{31}p_{24} = p_{12}p_{34}.$$

If we substitute this in the third term in (4.8), we see that p_{12} is a factor in each term. We naturally assume that p_{rs} has at least one nonzero component. We may assume $p_{12} \neq 0$, because we can always make it so by a coordinate transformation. Then, dividing (4.8) by p_{12}, we obtain

$$(4.11) \qquad \begin{aligned} (p_{12,3} + p_{31,2})p_{14} - p_{31,1}p_{24} - p_{12,1}p_{34} - p_{14,1}p_{23} \\ + (p_{12,4} + p_{41,2})p_{31} + (p_{13,4} + p_{41,3})p_{12} = 0. \end{aligned}$$

By adding terms inside the parentheses and subtracting them outside, we convert the above equation into

$$
\begin{aligned}
&(p_{12,3} + p_{23,1} + p_{31,2})p_{14} \\
&+ (p_{12,4} + p_{24,1} + p_{41,2})p_{31} \\
&+ (p_{13,4} + p_{34,1} + p_{41,3})p_{12} \\
&- (p_{23}p_{14} + p_{31}p_{24} + p_{12}p_{34})_{,1} = 0.
\end{aligned} \tag{4.12}
$$

The last term vanishes on account of (4.9), and the resulting equation may be written, in the notation of (2.7),

$$
p_{12}p^{*2n}{}_{,n} + p_{13}p^{*3n}{}_{,n} + p_{14}p^{*4n}{}_{,n} = 0. \tag{4.13}
$$

We have derived this equation from the first of (4.6). A second equation follows similarly from the second of (4.6), and these two equations are necessary and sufficient for the integrability of (4.3), or equivalently (4.5). These two conditions of integrability are, however, two of a set of four symmetric equations. Of these four, any two must be deducible from the other two. We may state the following result:

THEOREM V. *Necessary and sufficient conditions for the integrability of* (4.3), *subject to* (4.4), *are*

$$
p_{rm}p^{*mn}{}_{,n} = 0; \tag{4.14}
$$

only two of these equations are independent.

The integrability conditions (4.14) may also be written

$$
p_{rm}\epsilon^{mnab}p_{ab,n} = 0. \tag{4.15}
$$

We note that

$$
(\epsilon pp) = 0, \qquad p^{*mn}{}_{,n} = 0, \tag{4.16}
$$

or equivalently

$$
(\epsilon pp) = 0, \qquad p_{mn,r} + p_{nr,m} + p_{rm,n} = 0 \tag{4.17}
$$

are sufficient for the integrability of $p_{rs}\,dx^s = 0$.

5. Sufficient conditions for the existence of canonical coordinates in a domain D. Let P_{mn} and Q_{mn} be two skew-symmetric tensors in a domain D of space-time, satisfying

$$
(\epsilon PP) + (\epsilon QQ) = 0 \tag{5.1}
$$

at each point of D. We have already seen in §3 that (5.1) alone is not a sufficient condition for the existence of canonical coordinates in D. We shall add other conditions to (5.1).

Consider the determinantal equation

$$\det (P_{mn} - \lambda Q_{mn}) = 0. \tag{5.2}$$

This skew-symmetric determinant of the fourth order is necessarily a perfect square, and we may write (5.2) in the equivalent form

$$[\epsilon(P - \lambda Q)(P - \lambda Q)] = 0. \tag{5.3}$$

This expands into

$$(\epsilon PP) - 2(\epsilon PQ)\lambda + (\epsilon QQ)\lambda^2 = 0. \tag{5.4}$$

In view of (5.1), this equation always has real roots, and in general their product is -1. However, there are certain singular cases to consider, so we make the following classification:

Case A: $(\epsilon PP) \neq 0$, $(\epsilon QQ) \neq 0$: roots real with product -1.

Case B: $(\epsilon PP) = (\epsilon QQ) = 0$, $(\epsilon PQ) \neq 0$: the only root of (5.4) is $\lambda = 0$.

Case C: $(\epsilon PP) = (\epsilon QQ) = (\epsilon PQ) = 0$: equation (5.4) is satisfied by every value of λ.

Let μ be a root of (5.4); then $\mu \neq 0$ in Case A and $\mu = 0$ in Case B; we shall choose $\mu = 0$ in Case C. Let us define p and q by

$$p_{mn} = P_{mn} - \mu Q_{mn}, \qquad q_{mn} = \mu P_{mn} + Q_{mn}. \tag{5.5}$$

Then it is clear from (5.3) that

$$(\epsilon pp) = 0, \qquad (\epsilon qq) = 0. \tag{5.6}$$

Consider now the total differential equations

$$p_{rs}\, dx^s = 0, \qquad q_{rs}\, dx^s = 0. \tag{5.7}$$

In general, these equations are not integrable. Let us however subject p and q to the Maxwell conditions

$$p^{*mn}{}_{,n} = 0, \qquad q^{*mn}{}_{,n} = 0, \tag{5.8}$$

so that, as we saw in (4.16), the equations (5.7) are integrable. The conditions (5.8) are equivalent to

$$\begin{aligned} (P^{*mn} - \mu Q^{*mn})_{,n} &= 0, \\ (\mu P^{*mn} + Q^{*mn})_{,n} &= 0. \end{aligned} \tag{5.9}$$

Under these conditions of integrability, there exist four functions F_r of the coordinates such that if $dF_1 = dF_2 = 0$, then the first of (5.7) is satisfied, and if $dF_3 = dF_4 = 0$, then the second of (5.7) is satisfied.

On calculating (ϵpq) from (5.5) and using the fact that μ satisfies (5.4), we get

$$2\mu[(\epsilon PP)^2 + (\epsilon PQ)^2] = (\epsilon PP)(\epsilon pq). \tag{5.10}$$

Thus $(\epsilon pq) \neq 0$ in Case A. Obviously $(\epsilon pq) \neq 0$ in Case B also. By our choice of $\mu = 0$ we have further $(\epsilon pq) = (\epsilon PQ) = 0$ in Case C.

Gardner has shown that $(\epsilon pq) = 0$ is a necessary and sufficient condition for the intersection of the two elementary 2-spaces (5.7). Thus in Cases A and B these two 2-spaces do not intersect. This means that there is no displacement satisfying the four equations

(5.11) $$F_{r,s}\, dx^s = 0.$$

Hence the Jacobian det $F_{r,s} \neq 0$, and so the transformation $x^{r'} = F_r(x)$ is not singular.

Let us complete the argument for Cases A and B, omitting the singular Case C. Since (5.7) are invariant equations, we see that, if we put primes on the suffixes, the first of (5.7) is satisfied provided $dx^{1'} = dx^{2'} = 0$; hence

(5.12) $$p_{r'3'}\, dx^{3'} + p_{r'4'}\, dx^{4'} = 0$$

for arbitrary $dx^{3'}$, $dx^{4'}$, and so

(5.13) $$p_{r'3'} = p_{r'4'} = 0.$$

Similarly,

(5.14) $$q_{r'1'} = q_{r'2'} = 0.$$

In fact, throughout D the only surviving components of $p_{m'n'}$ and $q_{m'n'}$ are $p_{1'2'}$ and $q_{3'4'}$. By the invariant conditions (5.8) we have

(5.15) $$p_{1'2',3'} = 0, \qquad p_{1'2',4'} = 0, \qquad q_{3'4',1'} = 0, \qquad q_{3'4',2'} = 0,$$

so that

(5.16) $$p_{1'2'} = p_{1'2'}(x^{1'}, x^{2'}), \qquad q_{3'4'} = q_{3'4'}(x^{3'}, x^{4'}).$$

Neither of these functions can vanish since we are considering only the case $(\epsilon pq) \neq 0$.

A transformation

(5.17) $$\begin{aligned} x^{1''} &= f_1(x^{1'}, x^{2'}), \qquad & x^{2''} &= f_2(x^{1'}, x^{2'}), \\ x^{3''} &= f_3(x^{3'}, x^{4'}), \qquad & x^{4''} &= f_4(x^{3'}, x^{4'}) \end{aligned}$$

will not revive any components which we have caused to vanish, and the two surviving components will satisfy

(5.18) $$\begin{aligned} p_{1'2'} &= p_{1''2''}(X^{1''}_{1'}\, X^{2''}_{2'} - X^{2''}_{1'}\, X^{1''}_{2'}), \\ q_{3'4'} &= q_{3''4''}(X^{3''}_{3'}\, X^{4''}_{4'} - X^{4''}_{3'} X^{3''}_{4'}). \end{aligned}$$

Thus we shall have

(5.19) $$q_{3''4''} = -p_{1''2''} = 1$$

provided the transformation (5.17) satisfies the two conditions

(5.20) $$X^{1''}_{1'} X^{2''}_{2'} - X^{2''}_{1'} X^{1''}_{2'} = -p_{1'2'}, \qquad X^{3''}_{3'} X^{4''}_{4'} - X^{4''}_{3'} X^{3''}_{4'} = q_{3'4'}.$$

These equations can be solved with considerable arbitrariness. For example, we can take $f_2 = x^{2'}, f_4 = x^{4'}$, so that they become

$$X^{1''}_{1'} = -p_{1'2'}, \qquad X^{3''}_{3'} = q_{3'4'}, \tag{5.21}$$

which are immediately integrable. We have then a coordinate system $x^{r''}$ in which all components of p and q vanish except $p_{1''2''}$ and $q_{3''4''}$, and these components satisfy

$$q_{3''4''} = -p_{1''2''} \tag{5.22}$$

throughout D. Then, by (5.5), we have throughout D

$$\begin{aligned} P_{2''3''} - \mu Q_{2''3''} &= 0, & \mu P_{2''3''} + Q_{2''3''} &= 0,\\ P_{3''1''} - \mu Q_{3''1''} &= 0, & \mu P_{3''1''} + Q_{3''1''} &= 0,\\ P_{1''4''} - \mu Q_{1''4''} &= 0, & \mu P_{1''4''} + Q_{1''4''} &= 0,\\ P_{2''4''} - \mu Q_{2''4''} &= 0, & \mu P_{2''4''} + Q_{2''4''} &= 0, \end{aligned} \tag{5.23}$$

and also

$$\begin{gathered} P_{3''4''} - \mu Q_{3''4''} = 0, \qquad \mu P_{1''2''} + Q_{1''2''} = 0,\\ \mu P_{3''4''} + Q_{3''4''} = -(P_{1''2''} - \mu Q_{1''2''}). \end{gathered} \tag{5.24}$$

From (5.23) it follows that all the components of P and Q vanish except

$$P_{1''2''}, \qquad P_{3''4''}, \qquad Q_{1''2''}, \qquad Q_{3''4''}. \tag{5.25}$$

If we substitute from the first line of (5.24) in the second line, we get

$$(1 + \mu^2)Q_{3''4''} = -(1 + \mu^2)P_{1''2''}, \tag{5.26}$$

and hence

$$Q_{3''4''} = -P_{1''2''}, \qquad Q_{1''2''} = P_{3''4''}. \tag{5.27}$$

Thus we have a system of canonical coordinates in D. It follows from Theorem II that the transformation from the original coordinates to the canonical coordinates may be made to have a positive or negative Jacobian at our choice. So we have the result:

THEOREM VI. *Omitting the singular Case* C, *the conditions* (5.1) *and* (5.9) *are sufficient to insure that* P_{mn} *and* Q_{mn} *are Maxwellian complements in a domain* D, *and the transformation to canonical coordinates may be made with a positive or negative Jacobian at our choice.*

We note that (5.9) are equivalent to

$$\begin{gathered} (P_{mn} - \mu Q_{mn})_{,r} + (P_{nr} - \mu Q_{nr})_{,m} + (P_{rm} - \mu Q_{rm})_{,n} = 0,\\ (\mu P_{mn} + Q_{mn})_{,r} + (\mu P_{nr} + Q_{nr})_{,m} + (\mu P_{rm} + Q_{rm})_{,n} = 0. \end{gathered} \tag{5.28}$$

If $\mu = 0$, (5.28) become the usual Maxwellian equations.

6. The general transformation preserving canonical form. Let D be a region of space-time and x^r a coordinate system. Consider all tensor pairs (P_{mn}, Q_{mn}) for which x^r are canonical coordinates in D. We ask: *What is the group of transformations of canonical coordinates?* In other words: *What transformations $x^r \rightarrow x^{r'}$ preserve canonical relationships?*

Hitherto we have used real coordinates, but now it is convenient to introduce Minkowskian coordinates with x^4 a pure imaginary, so that the relations (1.12) read

(6.1)
$$Q_{mn} = 2^{-1} i \epsilon^{mnrs} P_{rs}.$$

It is well known that for an orthogonal transformation covariant and contravariant tensors are indistinguishable, and the permutation symbol is a tensor, provided the transformation has positive Jacobian. Hence (6.1) is preserved under such an orthogonal transformation. It is also preserved under equal expansions of all four coordinates, that is, $x^{r'} = kx^r$ where k is a constant. But these obvious results are special cases of a general theorem, as we shall now see.

A transformation $x^r \rightarrow x^{r'}$ is *conformal* if

(6.2)
$$dx^r\, dx^r = dx^{r'}\, dx^{r'} \cdot F,$$

where F is a function of position. In the notation of (2.1), necessary and sufficient conditions for a conformal transformation are

(6.3)
$$X^r_{m'} X^r_{n'} = F\delta_{mn}.$$

This gives

(6.4)
$$X^2_{\prime} = F^4.$$

Equation (6.3) also gives

(6.5)
$$X^s_{m'} = FX^{m'}_s.$$

Consider, for *any* pair of skew-symmetric tensors P_{mn} and Q_{mn}, the expression

(6.6)
$$S_{mn} = Q_{mn} - 2^{-1} i \epsilon^{mnrs} P_{rs}.$$

Under *any* transformation we have

(6.7)
$$\begin{aligned} S_{m'n'} &= Q_{m'n'} - 2^{-1} i \epsilon^{m'n'r's'} P_{r's'} \\ &= Q_{ab} X^a_{m'} X^b_{n'} - 2^{-1} i \epsilon^{m'n'r's'} P_{ab} X^a_{r'} X^b_{s'}. \end{aligned}$$

But

(6.8)
$$\epsilon^{m'n'r's'} = X_{\prime} \epsilon^{cduv} X^{m'}_c X^{n'}_d X^{r'}_u X^{s'}_v,$$

and so (6.7) gives

(6.9)
$$\begin{aligned} S_{m'n'} &= Q_{ab} X^a_{m'} X^b_{n'} - 2^{-1} i X_{\prime} \epsilon^{cduv} X^{m'}_c X^{n'}_d X^{r'}_u X^{s'}_v P_{ab} X^a_{r'} X^b_{s'} \\ &= Q_{ab} X^a_{m'} X^b_{n'} - 2^{-1} i X_{\prime} \epsilon^{cduv} P_{uv} X^{m'}_c X^{n'}_d. \end{aligned}$$

So far the transformation is quite general. Let us now take it to be conformal.

Then, by (6.5), (6.9) may be written

$$S_{m'n'} = Q_{ab}X^a_{m'}X^b_{n'} - 2^{-1}iX_{\prime}F^{-2}\epsilon^{cduv}P_{uv}X^c_{m'}X^d_{n'} . \tag{6.10}$$

If the Jacobian of the transformation is positive, (6.4) tells us that $X_{\prime} = F^2$, and (6.10) may be written

$$S_{m'n'} = S_{ab}X^a_{m'}X^b_{n'} . \tag{6.11}$$

Thus we have the following result:

THEOREM VII. *If P_{mn} and Q_{mn} are any two skew-symmetric tensors, the skew-symmetric quantities S_{mn} defined by* (6.6) *in Minkowskian coordinates are components of a covariant tensor with respect to conformal transformations with positive Jacobian.*

Now $S_{mn} = 0$ expresses the fact that the coordinates are canonical. Hence it is clear that canonical character is preserved under conformal transformations with positive Jacobian. The positive character of the Jacobian here is not important. If we follow such a transformation by the reversal of one coordinate, the whole transformation is conformal and its Jacobian is negative, while the effect of the reversal of the axis is merely to change from one of the sets of canonical coordinates discussed in §2 to another of those sets. Hence we may state this result:

THEOREM VIII. *Canonical form is preserved under any conformal transformation, the Jacobian having either sign.*

We have seen that conformal character is *sufficient* to preserve canonical form. Let us now look at the question the other way round, and seek the most general transformation preserving canonical form. We assume that P_{mn} and Q_{mn} satisfy (6.1) in the Minkowskian coordinates x^r. Then (6.9) gives for any transformation

$$S_{m'n'} = Q_{ab}(X^a_{m'}X^b_{n'} - X_{\prime}X^{m'}_a X^{n'}_b). \tag{6.12}$$

If canonical form is to be preserved no matter what Q_{ab} may be, it is necessary that the matrix of the transformation satisfy

$$X^a_{m'}X^b_{n'} - X^b_{m'}X^a_{n'} = X_{\prime}(X^{m'}_a X^{n'}_b - X^{m'}_b X^{n'}_a). \tag{6.13}$$

If we multiply this by $X^{m'}_c$ and introduce the notation

$$(ab) = X^{m'}_a X^{m'}_b = (ba), \tag{6.14}$$

we get

$$\delta^a_c X^b_{n'} - \delta^b_c X^a_{n'} = X_{\prime}[(ac)X^{n'}_b - (bc)X^{n'}_a]. \tag{6.15}$$

If we put

(6.16) $$a = A, \qquad b = B, \qquad c = A,$$

we get

(6.17) $$X^B_{n'} = X_{\prime}[(AA)X^{n'}_B - (AB)X^{n'}_A],$$

and multiplication by $X^{n'}_B$ (remember that there is no summation for capitals) gives

(6.18) $$1 = X_{\prime}[(AA)(BB) - (AB)^2].$$

Multiplication of (6.17) by $X^{n'}_C$ gives (since we consider only transformations with $X_{\prime} \neq 0$)

(6.19) $$0 = (AA)(BC) - (AB)(AC),$$

or, with interchange of A and B,

(6.20) $$0 = (BB)(AC) - (AB)(BC).$$

Now multiply (6.18) by (BC) and use (6.19); this gives

(6.21) $$\begin{aligned}(BC) &= X_{\prime}[(BB)(AB)(AC) - (AB)^2(BC)] \\ &= X_{\prime}(AB)[(BB)(AC) - (AB)(BC)] \\ &= 0\end{aligned}$$

by (6.20). It follows then from (6.18) that $(AA) = (BB)$, and if we define F by $F = X_{\prime}(AA)$, then F is a function of position, independent of A. Then (6.17) reads

(6.22) $$X^b_{n'} = FX^{n'}_b,$$

and hence

(6.23) $$X^b_{m'}X^b_{n'} = F\delta_{mn},$$

as in (6.3), showing that the transformation is conformal. Hence we have the following result:

THEOREM IX. *The most general transformation which preserves canonical form for all skew-symmetric tensor-pairs already in canonical form is a conformal transformation.*

Since the equation $dx^r\, dx^r = 0$ is invariant under conformal transformation, Theorem IX connects the concept of the null-cone with the concept of canonical coordinates. The result is, however, not as interesting as may appear at first sight. Suppose we are given a pair of skew-symmetric tensors which are Maxwellian complements. This means that there exists at least one canonical coordinate system for this tensor-pair. We know, by Theorem VIII, that any conformal transformation yields new canonical coordinates for this tensor-pair. But we have no reason to suppose that canonical form for this tensor-pair is preserved *only* under conformal transformations. This statement does not conflict with Theorem IX, which concerns the preservation of canonical form for a whole class of tensor-pairs and not just one such pair; it will be remembered that after (6.12) we assumed Q_{ab} arbitrary.

An interesting question is thus raised at this point: *What group of transformations preserves canonical form for a given tensor-pair?*

We see from (6.12) that such a transformation must satisfy

$$Q_{ab}(X^a_{m'}X^b_{n'} - X_{\prime}X^{m'}_a X^{n'}_b) = 0, \tag{6.24}$$

which may also be written in the equivalent form

$$Q_{ab}[\delta^a_c \delta^b_d - X_{\prime}(ac)(bd)] = 0 \tag{6.25}$$

in the notation of (6.14). We have here six linear homogeneous equations in the six independent components of Q_{ab}, and so the vanishing of the determinant of these equations is a necessary condition on the transformation. This condition is of course independent of the particular tensor Q_{ab}, and it might be of interest to study those transformations for which the aforesaid determinant vanishes. But of course the vanishing of this determinant merely ensures the existence of some tensor-pairs which remain canonical under the transformation; it does not ensure the preservation of canonical form for a given tensor-pair. The condition for that is the consistency of (6.24) or (6.25), and we shall here make no further attempt to analyze this complicated situation.

We may, however, push the argument further for infinitesimal transformations. Let us write

$$X^a_{m'} = \delta^a_m + Y^a_{m'}, \qquad X^{m'}_a = \delta^m_a + Y^{m'}_a, \tag{6.26}$$

where the Y's are infinitesimal. It follows from (2.2) that

$$Y^a_{b'} + Y^{a'}_b = 0, \qquad X_{\prime} = 1 + Y^c_{c'}. \tag{6.27}$$

If we substitute from (6.26) in (6.24), the finite part cancels and the principal part of the equation reads

$$Q_{am}e_{an} - Q_{an}e_{am} + 2^{-1}e_{aa}Q_{mn} = 0, \tag{6.28}$$

where

$$e_{ab} = Y^b_{a'} + Y^a_{b'} = e_{ba}. \tag{6.29}$$

We may regard the transformation as an infinitesimal strain of space-time, and e_{ab} as the components of the strain tensor. So we may state the following result:

THEOREM X. *The most general infinitesimal transformation which preserves the canonical form of a given tensor-pair is an infinitesimal strain satisfying* (6.28).

As a check on (6.28), we note that for an infinitesimal conformal transformation we have

$$e_{AA} = e_{BB}, \qquad e_{AB} = 0. \tag{6.30}$$

Now (6.28) is obviously satisfied if $m = n$. If $m = M, n = N$, the left-hand side reads, by (6.30),

$$Q_{aM}e_{aN} - Q_{aN}e_{aM} + 2^{-1}e_{aa}Q_{MN}$$
$$= e_{AA}(Q_{NM} - Q_{MN}) + 2e_{AA}Q_{MN} \tag{6.31}$$
$$= 0.$$

Thus (6.28) is satisfied by any infinitesimal conformal transformation.

If we define

$$f_{mn} = e_{mn} - 4^{-1}\delta_{mn}e_{cc} = f_{nm}, \tag{6.32}$$

(6.28) becomes

$$Q_{ma}f_{an} + f_{ma}Q_{an} = 0, \tag{6.33}$$

or, in matrix notation,

$$Qf + fQ = 0. \tag{6.34}$$

Thus f is a symmetric matrix which anticommutes with the skew-symmetric matrix Q. However, it must be remembered that the solution of (6.28) is not merely an algebraic problem. The strain components must satisfy the well known compatibility equations [6]

$$e_{mn,rs} + e_{rs,mn} - e_{mr,ns} - e_{ns,mr} = 0. \tag{6.35}$$

7. **Modified Maxwell equations in vacuo.** The physical hypothesis offered below is highly tentative, but there seems to be no harm in making a suggestion, which may assist in the development of electromagnetic theory even though it should be shown to be physically untenable in its present form. In this spirit the following hypothesis is made. Real coordinates are used in this section.

Modified Maxwellian theory in vacuo: In space-time in vacuo there is no metric and an electromagnetic field is specified by two skew-symmetric vectors, P_{mn} and Q_{mn}. They are subject to the invariant algebraic relation

$$(\epsilon PP) + (\epsilon QQ) = 0, \tag{7.1}$$

and the singular case where $(\epsilon PP) = (\epsilon QQ) = (\epsilon PQ) = 0$ is excluded. (This exclusion may not be essential, but is made here because the argument has not been completed in this singular case.) An invariant function of position in space-time μ is defined as either root of the quadratic equation

$$(\epsilon PP) - 2\lambda(\epsilon PQ) + \lambda^2(\epsilon QQ) = 0. \tag{7.2}$$

The two tensors are further subject to the following invariant modified Maxwell equations:

$$(P^{*rn} - Q^{*rn})_{,n} = 0, \qquad (P^{*rn} + Q^{*rn})_{,n} = 0, \tag{7.3}$$

or the equivalent form (5.28).

In the above formulation there are 12 dependent variables (components of the field) and 7 independent equations for them. Thus there is a 5-fold freedom instead of the 4-fold freedom which we expect in a theory formulated for general coordinates. This suggests two possibilities: (a) some other condition, at present unknown, should be imposed, (b) partial indeterminacy is an essential feature of the electromagnetic field in vacuo.

The conditions imposed above insure that P and Q are Maxwellian complements. If we use general coordinates, this fact does not lead to expressions for the components of Q in terms of those of P. For simplicity we may however choose canonical coordinates for which (1.12) hold. For such coordinates the equation for μ becomes

$$\sum P_{23}P_{14} + \mu(\sum P_{23}^2 - \sum P_{14}^2) - \mu^2 \sum P_{23}P_{14} = 0, \tag{7.4}$$

and the differential equations (7.3) or (5.28) read explicitly

$$\begin{aligned}
(P_{23} - \mu P_{14})_{,1} + (P_{31} - \mu P_{24})_{,2} + (P_{12} - \mu P_{34})_{,3} &= 0,\\
(P_{23} - \mu P_{14})_{,4} + (P_{34} + \mu P_{12})_{,2} + (P_{42} - \mu P_{31})_{,3} &= 0,\\
(P_{31} - \mu P_{24})_{,4} + (P_{14} + \mu P_{23})_{,3} + (P_{43} - \mu P_{12})_{,1} &= 0,\\
(P_{12} - \mu P_{34})_{,4} + (P_{24} + \mu P_{31})_{,1} + (P_{41} - \mu P_{23})_{,2} &= 0,\\
(\mu P_{23} + P_{14})_{,1} + (\mu P_{31} + P_{24})_{,2} + (\mu P_{12} + P_{34})_{,3} &= 0,\\
(\mu P_{23} + P_{14})_{,4} + (\mu P_{34} - P_{12})_{,2} + (\mu P_{42} + P_{31})_{,3} &= 0,\\
(\mu P_{31} + P_{24})_{,4} + (\mu P_{14} - P_{23})_{,3} + (\mu P_{43} + P_{12})_{,1} &= 0,\\
(\mu P_{12} + P_{34})_{,4} + (\mu P_{24} - P_{31})_{,1} + (\mu P_{41} + P_{23})_{,2} &= 0.
\end{aligned} \tag{7.5}$$

In (7.4) and (7.5) we have seven independent equations for μ and the six components of P.

Solutions of these equations may be obtained by the following device. Let us go back to classical physics, and choose a Maxwellian field $(\boldsymbol{e}, \boldsymbol{h})$ in vacuo, using rectangular Cartesians (x^1, x^2, x^3) and time (x^4) such that the velocity of light is unity. Thus $(\boldsymbol{e}, \boldsymbol{h})$ satisfy

$$\begin{aligned}
\nabla \cdot \boldsymbol{h} = 0, \qquad \partial \boldsymbol{h}/\partial x^4 &= -\nabla \times \boldsymbol{e},\\
\nabla \cdot \boldsymbol{e} = 0, \qquad \partial \boldsymbol{e}/\partial x^4 &= \nabla \times \boldsymbol{h}.
\end{aligned} \tag{7.6}$$

Let us however further subject the field to the condition

$$\boldsymbol{e} \cdot \boldsymbol{h} = 0; \tag{7.7}$$

this condition is satisfied by many well known fields.

We now write $\boldsymbol{b} = \boldsymbol{h}$ and $\boldsymbol{d} = \boldsymbol{e}$, and define p_{mn} and q_{mn} by relations as in (1.6), but with small letters. We have then

$$\begin{aligned}
q_{23} &= p_{14}, \qquad q_{14} = -p_{23},\\
q_{31} &= p_{24}, \qquad q_{24} = -p_{31},\\
q_{12} &= p_{34}, \qquad q_{34} = -p_{12}.
\end{aligned} \tag{7.8}$$

Then by (7.7) and (7.6) we have

$$(\epsilon pp) = 0, \qquad (\epsilon qq) = 0, \tag{7.9}$$

$$p^{*mn}{}_{,n} = 0, \qquad q^{*mn}{}_{,n} = 0, \tag{7.10}$$

and also

$$(\epsilon pq) = 2^{-1}(e^2 - h^2). \tag{7.11}$$

To avoid the singular case we shall assume $e^2 - h^2 \neq 0$.

We now choose *any* function f of the space-time coordinates (it may be zero), and define P_{mn} and Q_{mn} by the equations

$$\begin{aligned} P_{mn} &= (p_{mn} + fq_{mn})(1 + f^2)^{-1}, \\ Q_{mn} &= (-fp_{mn} + q_{mn})(1 + f^2)^{-1}. \end{aligned} \tag{7.12}$$

Then by (7.9)

$$(\epsilon PP) = 2f(\epsilon pq)(1 + f^2)^{-2} = -(\epsilon QQ) \tag{7.13}$$

so that (7.1) is satisfied, and we have also

$$(\epsilon PQ) = (\epsilon pq)(1 - f^2)(1 + f^2)^{-2}. \tag{7.14}$$

Since $(\epsilon pq) \neq 0$ by hypothesis, the singular case $(\epsilon PP) = (\epsilon QQ) = (\epsilon PQ) = 0$ is avoided. Further from (7.13) and (7.14) we have

$$(\epsilon PP) - 2f(\epsilon PQ) + f^2(\epsilon QQ) = 0, \tag{7.15}$$

so that f satisfies (7.2) and therefore may be identified with the μ defined by P and Q.

From (7.12) we have

$$\begin{aligned} p_{mn} &= P_{mn} - fQ_{mn}, \\ q_{mn} &= fP_{mn} + Q_{mn}. \end{aligned} \tag{7.16}$$

Since $f = \mu$, it follows from (7.10) that P and Q satisfy (7.3). Thus starting from a Maxwellian field $(\boldsymbol{e}, \boldsymbol{h})$ satisfying $\boldsymbol{e}\cdot\boldsymbol{h} = 0$, $e^2 - h^2 \neq 0$, we have generated a field P, Q belonging to the modified electromagnetic theory. Using (1.6) and (7.12), we express the vectors of this field as follows:

$$\begin{aligned} \boldsymbol{E} = \boldsymbol{D} &= (\boldsymbol{e} - f\boldsymbol{h})(1 + f^2)^{-1}, \\ \boldsymbol{B} = \boldsymbol{H} &= (\boldsymbol{h} + f\boldsymbol{e})(1 + f^2)^{-1}. \end{aligned} \tag{7.17}$$

If we choose $f = 0$, the modified field of course coincides with the Maxwellian field $(\boldsymbol{e}, \boldsymbol{h})$.

Let us conclude with a simple example. Let $(\boldsymbol{e}, \boldsymbol{h})$ be the Coulomb field

$$\boldsymbol{e} = kr^{-3}\boldsymbol{r}, \qquad \boldsymbol{h} = 0, \tag{7.18}$$

so that $\boldsymbol{e}\cdot\boldsymbol{h} = 0$ as required. Then (7.17) gives

$$E = kr^{-3}\boldsymbol{r}\,(1 + f^2)^{-1}, \quad H = kr^{-3}\boldsymbol{r}f(1 + f^2)^{-1}, \tag{7.19}$$

f being an arbitrary function of position and time. Let us choose $f = br^{-2}$, where b is a constant. Then the field approximates to a Coulomb field for large values of r, and for small r we have approximately

$$E = kb^{-2}r\boldsymbol{r}, \qquad H = kb^{-1}r^{-1}\boldsymbol{r}. \tag{7.20}$$

Thus as we approach the origin, E tends to zero, the magnitude of H tends to kb^{-1} and its direction is radial.

Bibliography

1. E. Cartan, *Leçons sur les invariants integraux*, Paris, Hermann, 1922.
2. D. van Dantzig, Proc. Cambridge Philos. Soc. vol 30 (1934) p. 99.
3. ———, Proceedings K. Akademie van Wetenschappen te Amsterdam vol. 37 (1934) pp. 521–531, 644–652, 825–836; vol. 39 (1936) pp. 126–131.
4. H. C. Lee, *A kind of even-dimensional geometry and its application to exterior calculus*, Amer. J. Math. vol. 65 (1943) pp. 433–438.
5. C. C. MacDuffee, *The theory of matrices*, New York, 1946, p. 12.
6. I. S. Sokolnikoff, *Mathematical theory of elasticity*, New York, 1946, p. 27.

Carnegie Institute of Technology,
Pittsburgh, Pa.

DISCONTINUITY IN ELECTROMAGNETISM

BY

W. H. WATSON

The work which I am about to present grew out of the belief that quantum mechanics may be regarded as a means of representing discontinuous motion statistically. This approach to atomic physics is at first glance far removed from the current pursuit of formalism that eschews looking on electrons and nuclei with anything like the naive concreteness of the biologist's thought. In theoretical physics, it hardly seems to matter that we are supposed to be talking about *things* actually existing in the world and that these things subserve their organization into the higher forms of existence in this, the only world that we know.

Let me explain what I mean by discontinuous motion. The continuous motion of a particle may be represented geometrically by showing the functional relation between its coordinates and the time, or it may be shown kinematically for example by drawing a chalk mark on the blackboard and erasing it continuously so that the same size of mark is seen all the time. In discontinuous motion the successive processes of creation and annihilation are discrete. A mark is made and rubbed out before the next is made. There are intervals of time corresponding to which it has no sense to ask where the object is, for the answer is "nowhere". The world track is a discrete series of events, and there is time for the processes of creation and annihilation, whereas in the limit of continuous motion there is no time for these processes [1].

What modifications are required in classical electromagnetic theory to permit such discontinuous motion by a singularity of the field? At once we encounter the conservation of electric charge. For, if an electric charge moves by successive annihilation and creation, conservation is not possible. We have to modify the basic equations to accommodate a set of quantities which behave as potential electric charge and current density. The simplest form that we can give to our equations is obtained by introducing a scalar function whose (negative) space gradient is proportional to the potential current density and whose time derivative is proportional to the potential charge density. Let us call this scalar N, then the proposed equations read [2, 3]

$$(1)\qquad \begin{aligned} \operatorname{curl} \boldsymbol{E} &= -\dot{\boldsymbol{B}}, & \operatorname{curl} \boldsymbol{H} - \dot{\boldsymbol{D}} &= \boldsymbol{j} - (1/\mu_0) \operatorname{grad} N, \\ \operatorname{div} \boldsymbol{B} &= 0, & \operatorname{div} \boldsymbol{D} - K_0 \dot{N} &= \rho, \\ \boldsymbol{B} &= \mu_0 \boldsymbol{H}, & \boldsymbol{D} &= K_0 \boldsymbol{E}. \end{aligned}$$

Obviously, when N = constant we have Maxwell's theory. Introduce scalar potential φ, and vector potential $\boldsymbol{A}$ in the usual way and require the preservation of

$$(2)\qquad \Box \boldsymbol{A} = -\mu_0 \boldsymbol{j}, \qquad \Box \varphi = -\rho/K_0 .$$

Then we find

(3) $$\square N = \mu_0(\operatorname{div} j + \dot{\rho}),$$

(4) $$\operatorname{div} A + (1/c^2)\dot{\varphi} = -N.$$

The diagonal components N of the electromagnetic field cartesian tensor are therefore produced by the creation or destruction of electricity.

The stress-energy-momentum tensor for the field obeying these new equations may be set up and the corresponding equation of motion of a charge q in the field introduced: thus

(5) $$\frac{dP}{ds} = q\left[N\frac{dr}{ds} + E\frac{dt}{ds} + \frac{dr}{ds} \times B\right],$$

from which we can deduce in obvious notation

(6) $$\frac{dm}{ds} = qN \text{ where } m^2c^2 = \frac{W^2}{c^2} - P^2.$$

This field component therefore would cause the particle to change its mass. Actually it was after considering the possibility of such a field that I was led to set up the electromagnetic equations already given [2].

As a step towards the treatment of a moving singularity we consider first what is the field associated with the sudden creation of a point-charge q at the event (x_1, y_1, z_1, t_1), it being further predicated that the charge continues to remain at rest thereafter. Since this creation involves a point-instant singularity (equation (3)), we use Fourier transforms to represent the field in (k_1, k_2, k_3, ck_0)-space, writing F to denote passage from (x, y, z, t)-space to the transform. From equation (3)

(7) $$F\frac{1}{\mu_0}\square N = \iiiint \left(\frac{\partial\rho}{\partial t} + \operatorname{div} j\right) e^{-i(k\cdot r - k_0ct)}\, dxdydzdt$$

integrated over all space and time. We assume that $\partial\rho/\partial t + \operatorname{div} j$ is a singular distribution vanishing everywhere except at (r_1, t_1) and tending to infinity there in such a way that

(8) $$\iiiint \left(\frac{\partial\rho}{\partial t} + \operatorname{div} j\right) dxdydzdt = q.$$

Then

(9) $$F\frac{1}{\mu_0}\square N = qe^{-i(k\cdot r - k_0ct_1)} \equiv qu_1.$$

From (9) we deduce

(10) $$FN = \frac{\mu_0 q u_1}{k_0^2 - k^2}.$$

By like reasoning we find for the scalar potential φ,

$$F\varphi = \frac{-qu_1}{K_0(k_0^2 - k^2)} \int_0^\infty e^{ik_0ct}\, dt. \tag{11}$$

Introduce the half δ-function

$$\delta_+(ck_0) = \frac{1}{2\pi} \int_0^\infty e^{ik_0ct}\, dt \tag{12}$$

understanding Im $(k_0) > 0$, then

$$F\varphi = -\frac{2\pi qu_1}{K_0(k_0^2 - k^2)}\, \delta_+(ck_0). \tag{13}$$

Clearly the destruction of charge would differ from the foregoing only in respect of substituting for the δ_+-function the complementary

$$\delta_-(ck_0) = \frac{1}{2\pi} \int_{-\infty}^0 e^{ik_0ct}\, dt. \tag{14}$$

with $\mathrm{Im}(k_0) < 0$.

We are of course interested in the spatial distributions of φ and N. Let $R = |\,r - r_1\,|$; then, by inversion, φ is easily seen to be given by

$$\varphi = -\frac{q}{8\pi^2 K_0 R} \int_{-\infty}^{\infty} \frac{e^{ik_0(R - c(t-t_1))}}{ik_0}\, dk_0 \tag{15}$$

with $\mathrm{Im}(k_0) > 0$. The integral has the value 0 or -2π according as in the exponent the multiplier of $k_0 >$ or < 0. Consequently

$$\begin{aligned} \varphi &= \frac{q}{4\pi K_0 R} \quad \text{if} \quad R < c(t - t_1), \\ \varphi &= 0 \qquad\quad \text{if} \quad R > c(t - t_1). \end{aligned} \tag{16}$$

The form (13) for φ therefore represents a spherical wave of discontinuity between the region in which the usual electrostatic field due to the point charge at the point r_1 is set up and the outer region in which the field is not yet established. The discontinuity entails infinite electric force.

The Fourier inversion for N leads to the result

$$N = -\frac{\mu_0 qc}{4\pi R}\, \delta(R - c(t - t_1)), \tag{17}$$

which represents a pulse travelling outwards from the place when the charge is created. The calculation is as follows

$$N = \frac{\mu_0 qc}{(2\pi)^4} \iiiint \frac{e^{i\{k.(r-r_1) - k_0c(t-t_1)\}}}{k_0^2 - k^2}\, dk\, dk_0 \tag{18}$$

where $d\mathbf{k} = -k^2\, dk\, d\mu\, d\psi$, $\mu = \cos\theta$, $\mathbf{k}\cdot(\mathbf{r} - \mathbf{r}_1) = kR\mu$; θ is the angle between k and the axis $(\mathbf{r} - \mathbf{r}_1)$, while ψ is the longitude measured about this axis.

Integrating with respect to μ and ψ, we have

$$N = \frac{\mu_0 qc}{(2\pi)^3 R} \int_{-\infty}^{\infty} e^{-ik_0 c(t-t_1)}\, dk_0 \int_{-\infty}^{\infty} \frac{e^{ikR}\, k}{i(k_0^2 - k^2)}\, dk. \tag{19}$$

The integral with respect to k may be written

$$\frac{1}{2i} \int_{-\infty}^{\infty} \left\{ \frac{e^{ikR}}{k_0 - k} - \frac{e^{ikR}}{k_0 + k} \right\} dk. \tag{20}$$

Now we require $\mathrm{Im}(k_0) > 0$ and since this infinite integral may be replaced by the contour integral along the real axis of k and around the semicircle at infinity in the positive half of the k-plane, only the first fraction contributes to the answer

$$-\pi e^{ik_0 R}. \tag{21}$$

Hence

$$\begin{aligned} N &= -\frac{\pi \mu_0 qc}{(2\pi)^3 R} \int_{-\infty}^{\infty} e^{ik_0(R - c(t-t_1))}\, dk_0 \\ &= -\frac{\mu_0 qc}{4\pi R}\, \delta\{R - c(t - t_1)\}. \end{aligned}$$

If k_0 had been required to be real we should have proceeded by writing the integral (20) as

$$-\frac{1}{2i}\,[e^{ik_0 R} + e^{-ik_0 R}] \int_{-\infty}^{\infty} \frac{e^{izR}}{z}\, dz \tag{22}$$

in which the infinite integral would be taken as

$$\lim_{A \to \infty} P \int_{-A}^{A} \frac{e^{izR}}{z}\, dz.$$

Since $R > 0$, its value is πi, which leads to

$$N = -\frac{\mu_0 qc}{4\pi R}\cdot\frac{1}{2}\{\delta(R - c(t - t_1)) + \delta(R + c(t - t_1))\}. \tag{23}$$

This combination of an incoming and outgoing wave does not however correspond to the distribution of φ which grows outwards from the singularity at (r_1, t_1).

On account of the presence of improper integrals the analysis tends to conceal the distinction between incoming and outgoing waves. The equation (10) for instance can represent either, depending on the restrictions imposed on the imaginary parts of the complex variables k and k_0 to give meaning to the improper integrals involved. Thus quite clear physical distinctions are disguised in the niceties of the formalism. This thought might well be pondered on by formalists.

It appears then that we can distinguish different types of creation and annihilation processes, involving combinations of incoming and outgoing waves. We may, for instance, imagine the process of creation to begin on the sphere $R = b$, from which there spreads outwards a wave establishing the potential due to the charge and also a wave travelling inward which will be reflected at the pole $R = 0$. If this reflected wave were to travel outwards to infinity the field would be annulled after the passage of the wave and the charge destroyed. Consequently if the distant field of the charge is to persist, we must imagine a second creation process to cancel the outgoing wave and to initiate a new ingoing wave from the sphere $R = b$. This is equivalent to reflexion of the first outgoing wave inside the sphere due to the instantaneous appearance of charge on its

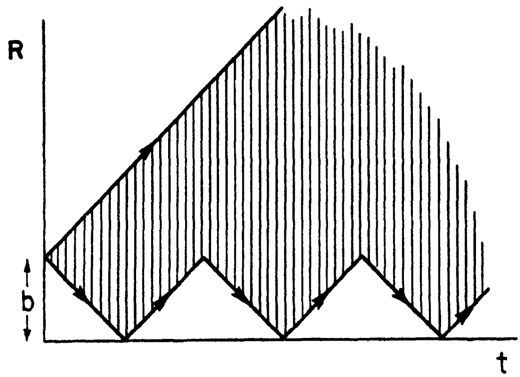


FIG. 1
A process of field creation initiated on the sphere $R = b$

surface. The new ingoing wave is in turn reflected at the pole with like consequences indicated in Fig. 1. The usual electrostatic field of the charge is established for those values of R and t in the hatched part of the figure. For $R < b$, the field is established and annihilated periodically.

When the front of discontinuity of potential is confined in this way the distribution of N is given by repeating the pattern

$$N = \frac{\mu_0 qc}{4\pi R}\frac{1}{2}\{\delta(R - ct) - \delta(R + ct - 2b)\} \quad \text{for} \quad R \leqq b, \tag{24}$$

N vanishing elsewhere, $0 < t < 2b/c$, and superposing the patterns for successive periods.

Let us consider the Fourier transform of (24).

$$\frac{F8\pi N}{\mu_0 q} = \frac{2\pi}{k}\left[\frac{e^{i(k_0-k)b} - 1}{k_0 - k} - \frac{e^{i(k_0+k)b} - 1}{k_0 + k}\right] + \frac{2\pi}{k}e^{2ik_0b}\left[\frac{e^{-i(k_0+k)b} - 1}{k_0 + k} - \frac{e^{-i(k_0-k)b} - 1}{k_0 - k}\right]. \tag{25}$$

From this

$$F\,\frac{8\pi(k_0^2 - k^2)N}{\mu_0 q} = -\frac{2\pi}{k}\,[2ik_0 e^{ik_0 b}\sin kb + 2k - ke^{ik_0 b}\cos kb] \tag{26}$$
$$-\frac{2\pi}{k}\,[2ik_0 e^{ik_0 b}\sin kb - 2ke^{2ik_0 b} + ke^{ik_0 b}\cos kb]$$

$$= \frac{8\pi ik_0}{k}\,e^{ik_0 b}\sin kb - 4\pi(1 - e^{2ik_0 b}). \tag{27}$$

This equation shows that at time $t = b/c$ there exists a charge distribution whose Fourier transform is

$$F\rho = \frac{q\sin kb}{k}\,e^{ik_0 b}.$$

The corresponding density in $(xyzt)$-space is

$$\rho = \frac{q}{2\pi R}\cdot\delta(R - b)\delta(ct - b) \tag{28}$$

leading to the time integral of charge $2bq/c$ appearing on the sphere $R = b$ at $t = b/c$. Thus the charge is switched on and off at each reflection of the wave at $R = b$. The second term on the right of (27) merely reflects the fact that the disturbance of the field began at $t = 0$ and disappeared at $t = 2b/c$.

It is not difficult to deduce that the corresponding solution for the potential φ is the usual electrostatic value $q/4\pi K_0 R$ for $R < ct < R + 2b$, while for times outside the interval, it vanishes. By repetition the enduring field can be built up.

It seems a natural step to attempt to extend the representation to the creation of a moving charge by transformation of coordinates corresponding to continuous motion of the charge after its creation. In place of the wave of scalar potential only, we should then have both vector and scalar potential waves. The effect of transforming coordinates would be to produce the classical field of the moving charge on one side of the discontinuity separating the region of space where the field has been established from that where it has not. The N pulse, however, arises only from the creation, hence it is dissociated from the subsequent motion if continuous. But if the motion is a series of appearances by successive creations and annihilations, N waves do arise from the subsequent motion.

Let us consider discontinous motion of an erratic type. We shall think of each creation and annihilation process as the establishing and elimination of the static field of the charge singularity. To produce the magnetic field of a moving charge, we require the equivalent of electric current, namely, the creation of electric moment. When a positive charge q at P is annihilated and an equal positive charge is created at Q, the field of the electric dipole $q\cdot PQ$ is the adjustment that must be radiated outwards and superposed on the original static field of q at P, to give the field due to q at Q. That is, the discontinuous motion from P to Q is equivalent to the creation of an electric dipole.

Although the creation of an electric doublet requires no radiation of N component (since electric charge is conserved in the process), the creation of a dipole of finite extent does cause the radiation of waves of creation and annihilation which do not cancel exactly because centered on two distinct points. Thus evidence of annihilation and creation of the charge is radiated to infinity. In addition, the creation of electric moment requires the radiation of transverse electric and magnetic fields. We have therefore raised in acute form the classical difficulty associated with the stationary motion of an electron in an atom, viz. that it ought to radiate. Hence we must find some way of containing the disturbance of the electromagnetic field so that the distant field is physically appropriate for the system of which the charge singularity forms a part.

Now we have already encountered the problem of containing N waves associated with the static charge. The wave constituting the internal field boundary is reflected when charge is switched on and off. When the pole moves erratically we may, to avoid misfit, introduce the Lorentz transformation characterized by the average velocity of the pole and derive the form of the field from the static case. The wave front will then be deformed and its pole translated so that after reflexion the wave will converge on the new pole. The Lorentz transformation will then have to supply the magnetic field we have attributed to the creation of electric moment. However, in a series of such displacements of the pole the Lorentz transformation will in general have to be different for each. Such a series of transformations applied discontinuously introduces misfits of higher order in our space-time and electromagnetic description. To remove these we should have to use the appropriate transformation of general relativity. That is, in order to transform the regular pulsation which we have designated as the static case with fixed pole, into a process without kinematic misfit corresponding to the erratic discontinuous motion of the pole, we require the corresponding fluctuating metric to distort the propagation of the internal boundary of the field.

Whichever metric we choose for the representation—either Euclidean or non-Euclidean—we have at some stage to impose conditions that lead to discontinuities in the electromagnetic field. It is these that integrate the erratic discontinuous motion of the singularity and it is the possibility of discontinuities that subserves the organization of the successive appearances of the singularity into a coherent motion as a part of a larger structure.

In one respect, however, discontinuity can be mitigated. The formulation of the static case involves discontinuity on the null cone through the position of the singularity. This discontinuity may be given structure in a finer representation and correspondingly the δ-functions are modified. In a previous paper a model was exhibited in which the suddenness of the jump of potential on the null cone was characterized by the brief time interval σ/c (of the order e^2/mc^3 for the electron). The δ-function may be modified by some such device as the following: For the unit step function write $\lim_{\sigma\to 0}$ $(1 + \tanh(x/\sigma))2$ and do not pass to the limit $\sigma \to 0$; the δ-function is replaced by the derivative $(1/2\sigma)\,\mathrm{sech}^2\,(x/\sigma)$ or to sufficient approximation by $(2/\sigma)e^{-2|x|/\sigma}$. Similarly we replace $\delta_+(k)$ by

$$\frac{1}{2\pi}\int_0^\infty \frac{1}{2}\left(\tanh\frac{x}{\sigma}+1\right)e^{ikx}\,dx$$

which to the same degree of approximation is

$$\delta_+(k) = -\frac{1}{2\pi}\left\{\frac{1}{ik} - \frac{1}{2/\sigma - ik}\right\}.$$

In so doing we introduce a new pole on the imaginary k-axis, as the means of

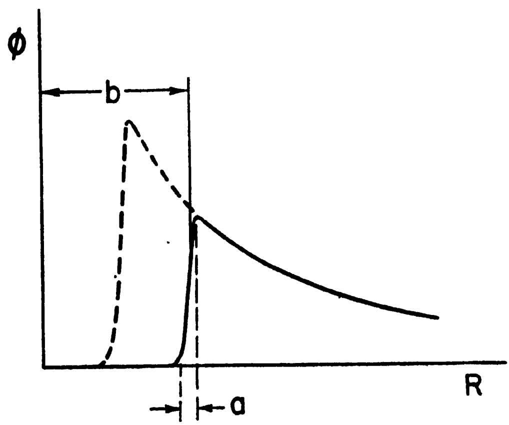


FIG. 2

The electric potential distribution, its pulsating internal boundary being shown for two different times

representing the time scale of the pulse of N (or jump in φ). By substituting the revised δ_+ function in the expression (13) for φ we obtain

$$\varphi = 0 \quad \text{if } R > c(t - t_1),$$

$$\varphi = \frac{q}{4\pi K_0 R}\left[1 - e^{2(R-c(t-t_1))/\sigma}\right] \quad \text{if} \quad R < c(t - t_1).$$

Such a potential distribution implies strong electric force varying exponentially with separation from the locus of the quasi discontinuity.

Our study of the imagined discontinuous motion of an electric charge has led to the following:

(i) modification of Maxwell's equations to permit processes which we can call creation and annihilation of charge;

(ii) to secure the appropriate distant field we assert that outgoing waves are reflected; according to our equations these reflexions are caused by the sudden appearance of singular distributions which prevent the self-destruction of the field;

(iii) different distributions are required to characterize different systems of discontinuous motions;

(iv) a kinematic interpretation of the lengths e^2/mc^2 and h/mc is possible; the former gives the scale of the "thickness" of the wavefront constituting the internal boundary of the electron's field, the latter gives the size of the hole in the field caused by the waxing and waning boundary. These lengths are indicated by a and b respectively in Fig. 2.

All of the foregoing representation is intended as a kinematic one to replace the field forms characteristic of singularities in classical electricity. Since we have in mind discontinuous motion, classical conceptions associated with the differential calculus are inappropriate and further since we contemplate erratic motion, the treatment of dynamics must be statistical. In any case our representation is much too fine grained in space-time to allow any restriction of the spectrum of the process in momentum and energy. We have in fact to represent probabilities of creation and annihilation and to discover how a distribution will change or preserve itself and how these probabilities are influenced by what we designate classically as a field of force. Here one is guided by the successful inventions of quantum mechanics. Once we recover forms such as Schrodinger's equation satisfied by a complex function defined by the real probabilities just mentioned, ordinary dynamical laws for the system follow by Fourier transformation, aided by the usual hypothesis as to the role of h. It is intended to present the details of this in a separate paper. The important point for electromagnetism is this: Electromagnetic forms are being used in atomic physics. By insisting on continuity except at punctual singularities we are making physics unreasonable. By boldly imagining extended discontinuities, or at least loci of rapid field change, we can see the possibility of representing the fields of atomic particles as existing, however transiently, in the world. The problem of the continued existence of an electron is then placed on the same basis as that of an atom. Lots may be going on within a region of space inaccessible to our measurements but accessible to our imagination.

Bibliography

1. W. H. Watson, Canadian Journal of Research vol. A23 (1945) pp. 39–46.
2. ———, Canadian Journal of Research vol. A18 (1940) pp. 1–21.
3. ———, Canadian Journal of Research vol. A23 (1945) pp. 33–38.

National Research Council of Canada,
Chalk River, Ontario.

THE FACTORIZATION METHOD AND ITS APPLICATION TO DIFFERENTIAL EQUATIONS IN THEORETICAL PHYSICS

BY

LEOPOLD INFELD

Introduction. I should like to report on a new technique of treating eigenvalue problems as they most frequently appear in Maxwell's theory with imposed boundary conditions, and in quantum-mechanics.

Let me say a few words about a subject often discussed: the analogy between Maxwell's and Dirac's equations. Both are linear systems of equations; each of them contains partial derivatives of the first order. Both Maxwell's and Dirac's equations are Lorentz-invariant. I may remark in passing that in the case of Maxwell's equations, the linearity may be an over-simplification that leads to difficulties with infinite self-energies. Yet, if we consider only regular solutions, as we shall, we may ignore this difficulty.

Historically, both the Maxwell and Dirac equations were preceded by a scalar theory. Especially in the case of Dirac's theory the preceding one was the Schroedinger theory, which is still applied to a wide range of quantum mechanical problems. The scalar theory leads to one partial differential equation of the second order, containing the Laplacian or the D'Alembertian. As both these names indicate, the study of this scalar equation and its solution is an important chapter of the mathematics of the nineteenth century. It led to the potential theory, to Legendre, Laguerre, Jacobi, Tschebyscheff, and Hermite polynomials, to Bessel functions, all of which form a chapter in mathematical physics which was completed when in theoretical physics the scalar field theories changed to vector, tensor, or spinor theories.

Thus the technique of solving Maxwell's or Dirac's system of equations is modeled upon the scalar theory. This is especially evident in the case of Maxwell's equations. There the usual procedure is to introduce the vector-potential and then obtain four equations of the type studied by the scalar theory. If you think about an application of Maxwell's equations to a wave guide with rectangular or circular crossection, or to an antenna, you see how the boundary conditions finally lead us to a set of ordinary differential equations of the second order.

In many respects, the situation is even simpler in wave mechanics. There the boundary conditions are organically connected with the differential equation itself and they usually mean uniqueness or finiteness or integrability.

Thus, both in electromagnetic and in quantum theory, we are led to equations of the type:

$$\frac{d^2y}{dx^2} + r(x, m)y + \lambda y = 0. \tag{1}$$

Here the only given function is

$$r(x, m),$$

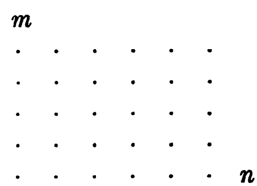
which characterizes the particular problem. We shall assume m to be a non-negative integer

$$m = 0, 1, 2, 3, \cdots. \tag{2}$$

It is well to recall how and why this m enters our equation. We represented the function of many variables by the product of many functions, each depending on one variable only. The constant m appears because of the process of separation and its integer values are due to the boundary conditions imposed upon a function by which y was multiplied. In all the typical cases λ has to be discrete because of the imposed boundary condition, and the solution of (1) leads to the eigenvalues of λ:

$$\lambda_0, \lambda_1, \lambda_2, \cdots, \lambda_n, \cdots. \tag{3}$$

Thus the typical eigenvalue problem can be represented by a lattice of points in a (n, m) plane:



For every line connecting the lattice points and parallel to the m-axis there exists a λ_n. For every point on the lattice there exists a function $y_{n(x)}^{m}$ satisfying some boundary conditions.

All I have done so far is to recapitulate a problem common to every theoretical physicist, almost a cornerstone in mathematical investigations of physical problems.

General characterization of the factorization method. Whereas the classical method consists in finding polynomials belonging to y_n^m and directly solving the differential equations, the factorization method outflanks and unifies the historical development. It either goes back to the original (Maxwell's and Dirac's) equations, or replaces the equations of the second order by a proper set of equations of the first order. While using the factorization method, we are not interested in the explicit solution, as we are in the classical method, but rather in the structure of the differential equation itself. We are led to a system of equations of the first order and we can quickly answer many reasonable questions without knowing the explicit solutions; it is, essentially, an operational method.

The method originated from Dirac's treatment of the harmonic oscillator. Some eight years ago, it was extended by Schroedinger to other differential equations that appear in wave-mechanics. Work on it was and is being done by myself and my students.

There are both a theoretical and a technical aspect of the factorization method. As far as the theoretical part is concerned, many things are still unclear to me. Therefore, I shall concentrate mostly on the technical side. There is not the slightest doubt that in many cases, these methods form a great simplification if compared with the old methods. The fact that we know everything about the solution itself without using the power series is convenient, but not important enough because the polynomials *are* well known. But the problems with which we are confronted in wave-mechanics are not those of finding solutions but of using them. To give some specific examples: the tedious problem of normalization of the Schroedinger functions does not even appear in the factorization method. The more complicated problem of normalizing the Dirac solutions was solved by the use of the factorization method in a brief "letter to the editor" by C. C. Lin. One of the basic problems in wave-mechanics is to calculate intensities of the hydrogen spectral lines. These radial integrals are very troublesome to compute and Gordon's paper on this subject is some twenty-five pages long. This problem can be solved literally in a few lines by the proper use of the operational methods.

I should like here to answer one more question: What is the range of validity of this method? Let us concentrate our attention on wave-mechanics. There we find some "pure cases", by which I mean those that can be solved rigorously without the use of any perturbation procedure. *All* these pure cases can be solved quickly in a unifying way by the factorization method. More than that is possible. Because of an investigation by A. C. Stevenson and (along different lines) by myself we know all the types of differential equations to which this method can be applied. Strangely enough there are eleven types! Each of them has a physical image in Maxwell's theory, or quantum theory, or both.

Yet "purity", though a desired phenomenon, is a rare one, and as science and its technique develops, the number of non-pure cases built around pure cases constantly increases. It is therefore gratifying to know that a perturbation procedure can be built around the pure cases, and the factorization method properly extended. I do not know whether all the cases treated in quantum theory by the perturbation method are accessible to the factorization method. But again in some cases, as in the Stark effect, the method leads us more quickly than any other to the proper solution.

Special characterization of the factorization method. We start from the equation

(A) $$\frac{d^2 y^m}{dx^2} + r(x, m)y^m + \lambda y^m = 0.$$

We assume that (A) can be written in *either* of two forms:

$$(^{+}H^{m+1})\,(^{-}H^{m+1})y^m = (\lambda - L(m+1))y^m,$$

(B) $$(^{-}H^{m}\;)\,(^{+}H^{m}\;)y^m = (\lambda - L(m))y^m,$$

$$^{\pm}H^m = k(x, m) \pm \frac{d}{dx}.$$

Let us explain the notation: ^{-}H and ^{+}H (H "up" and H "down") are linear operators; $L(m)$ is a function of m alone, and not of x. The second equation in (B) is obtained from the first by interchanging the "up" and "down" operators and by replacing $(m + 1)$ by (m), both in the H's and L's, but not in the y's. It is essential to understand that *both* the first and second (B) equations mean *the same* as (A). We say that (A) can be factorized if (B) can be written down. If we are faced with (A) we can answer the following questions: Can (A) be factorized? What is the $k(x, m)$ and $L(m)$ corresponding to a given $r(x, m)$, if the factorization is possible? Without going into details, we shall just state that both questions can be answered by the use of a little algebra and that we have

$$r(x, m) \leftrightarrow k(x, m); L(m),$$

if factorization is possible. As a matter of fact, as we know all the types that can be factorized, the problem of finding k and L is that of looking up a table.

The next step is a transition from (B) to a set of equations of the first order. Omitting some mathematical subtleties we say (though it is not absolutely correct) that (B) is equivalent to

(C) $$\begin{aligned}(^{-}H^{m+1})y^m &= (\lambda - L(m+1))^{1/2}y^{m+1},\\ (^{+}H^{m+1})y^{m+1} &= (\lambda - L(m+1))^{1/2}y^m.\end{aligned}$$

The proof of the equivalence between (B) and (C), and its restrictions, are easily found if the content of (B) and (C) is understood from the point of view of the factorization method.

Equations (C) describe a manufacturing process. The first equation tells us how to find y^{m+1} if y^m is known, how to move *up* a ladder, increasing the m's. The second equation tells us how to move *down* a ladder, how to manufacture y^{m-1}, y^{m-2}, etc., if y^m is known. Thus if we return to our (n, m) lattice, (C) tells us how to move up and down along a line parallel to the m axis. Thus we know everything we need if we know one *key* function, for one m, and for a given λ. At the first glance the ladders seem to go to infinity since the manufacturing process can be prolonged indefinitely. Yet this is not true, at least for certain values of λ. Let us take for λ the value

(4) $$\lambda_n = L(n+1)$$

where n is a non-negative integer:

$$n = 0, 1, 2, 3, \cdots .$$

Then in the process of manufacturing new functions, the first (C) equation becomes:

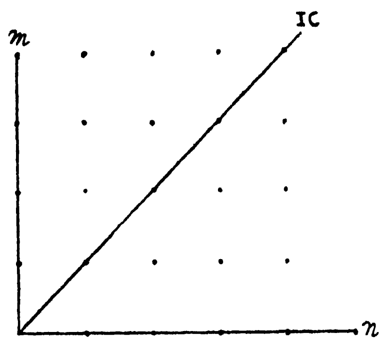
$$(^{-}H^{n+1})y_n^n = 0. \tag{5}$$

But this equation can be integrated immediately since it is a simple differential equation of the first order, with the solution:

$$y_n^n \sim \exp\left\{\int^x k(x, n + 1)\, dx\right\}. \tag{6}$$

Thus the manufacturing process is stopped. We can satisfy (C) by taking $\lambda = \lambda_n$ and $y^{n+1} = y^{n+2} = \cdots = 0$. By looking at the equations (C), we can write down mechanically the *eigenvalues* (4) for λ_n ($n = 0, 1, 2, \cdots$), the *key* function (6), which is the *top* of the ladder; from there we can go down by the second equation in (C). Thus we know everything, if we know (C).

Let us clarify the situation thus described on the (n, m) lattice:



We know the value of λ_n on each lattice line parallel to the m-axis. The quarter (n, m) plane is bisected by the IC (Iron Curtain) line. On this line, that is for $n = m$, we have the key solutions. There are no solutions belonging to our ladder above IC. The manufacturing process allows us to find solutions to every lattice point below the IC line. Its explicit form is of little importance, because what we need in applications is not the functions but their manufacturing process.

Let us now go back for a moment to (B) and we shall see why (B) is essentially equivalent to (C). What does the first (B) equation tell us? It tells us: go one step up the ladder, then one step down the ladder and you will arrive at the function from which you started. Similarly the second equation in (B) tells us: go one step down, then one step up the ladder and you will arrive at the function from which you started.

Special examples. We shall now illustrate our method on six examples; they are collected in a table:

No.	Interval	$r(x, m)$	$k(x, m)$	$L(n+1) = \lambda_n$	$\sim y_n^n$
1.	$(-\infty, +\infty)$	$-x^2 - 2m$	$-x$	$2n + 1$	$\exp\left(-\frac{x^2}{2}\right)$
2.	$(0, \pi)$	$\frac{1/4 - m^2}{\sin^2 x} + 1/4$	$\left(m - \frac{1}{2}\right) \cot x$	$n(n + 1)$	$(\sin x)^{n+1/2}$
3.	$(0, \infty)$	$\frac{2}{x} - \frac{m(m+1)}{x^2}$	$\frac{m}{x} - \frac{1}{m}$	$-\frac{1}{(n+1)^2}$	$x^{n+1}\exp\left(-\frac{x}{n+1}\right)$
4.	$(0, \pi)$	$2\nu \cot x - \frac{m(m+1)}{\sin^2 x}$	$m\cot x - \frac{\nu}{m}$	$n(n+2) - \frac{\nu^2}{(n+1)^2}$	$(\sin x)^{n+1}\exp\left(\frac{-x\nu}{n+1}\right)$
5.	$(0, \pi)$	$\frac{\frac{1}{4}-(\alpha+m)^2}{\sin^2 x} + \frac{\frac{1}{4}-(\beta+m)^2}{\cos^2 x}$	$(\alpha-1/2+m)\cot x - (\beta-1/2+m)\tan x$	$(\alpha + \beta + 2n + 1)^2$	$(\sin x)^{n+\alpha+1/2} \times (\cos x)^{n+\beta+1/2}$
6.	$(0, +\infty)$	$-\frac{(m^2-\frac{1}{4})}{x^2}$	$\frac{m-\frac{1}{2}}{x}$	$L(m) = 0$	—

The first example is that of a harmonic oscillator. The second example is that of associated harmonics (densities), the third example refers to one of the most important equations in quantum theory: the radial equation for an electron in a Coulomb field.

The fourth equation is little known and it is perhaps worth while to say a few words about it. It is one of those types that have an interesting image in wave mechanics and it was discussed for the first time by Schroedinger.

In the case of hydrogen and positive energy we have a continuous spectrum as $\lambda - L(m + 1)$ is always positive and therefore there is no top to our ladders. But, as is well known, we can change the continuous spectrum into a discontinuous spectrum by enclosing the hydrogen atom in a sphere. Mathematically a neater procedure is gained by assuming the hydrogen atom in a finite (Einstein) universe. A straightforward procedure gives us for the potential energy

$$\text{(7)} \qquad \cot x \text{ instead of } 1/x$$

and only the radial equation is different from the old Schroedinger problem Indeed the fourth example *is* the radial equation of a hydrogen atom in an Einstein universe. Going back to the energy levels, we have, in the usual notation:

$$\text{(8)} \qquad E_n = \frac{n(n + 2)\hbar^2}{2\mu R^2} - Z^2 \frac{e^4 \mu}{2\hbar^2(n + 1)^2},$$

where the only symbol needing explanation is R, the radius of the universe. For small (n/R) we regain Bohr's energy levels and for great (n/R) we have a very dense but discrete spectrum. In principle (though this remark must not be taken too seriously) we could deduce the radius of the universe from the spectral lines of hydrogen.

The example just discussed is interesting for another reason, too. Although the recursion formulas are valid whether the ladders are finite or not, the full method, including finding the key functions, can be used only for discrete λ's. Our example shows that by changing our problem into another, physically equivalent, problem, we may be able to assure the full use of the factorization methods.

The fifth example gives us the Jacobian polynomial (densities). Spherical harmonics, Weyl's spherical harmonics with spin, and Tschebyscheff polynomials are particular cases of these solutions.

The fifth example is an important but degenerate case. It gives us the Bessel functions (densities). But $L(m)$ here turns out to be zero, therefore we have infinite ladders. All that we gain in this case are the well known recursion formulas.

General remarks. 1. To all points of the lattice below the IC there exists, at least in the examples discussed here, a *good* solution satisfying the integrability condition, that is $\int y^2\,dx$ is finite. Usually we wish the integral to be equal to "one" which involves us in a normalization procedure. I wish to show on this one particular example of normalization how the factorization method works.

Multiply the first (C) equation on the left hand by y^{m+1} and integrate over the interval. Then, integrating by parts, shift the ${}^{-}H$ operator to y^{m+1}, changing it into ${}^{+}H$. If the functions vanish at the end of the interval, you obtain

$$\int (y^{m+1})^2\,dx = \int (y^m)^2\,dx.$$

Thus the manufacturing process preserves the normalization! All that is needed is to normalize the simple key functions!

2. To every point on the lattice, we found in our examples one *good* solution. But, of course, there must be two independent solutions for every point below and above the IC line. It is easy, by our method, to find the one *bad* solution below IC and the two *bad* solutions above IC. Thus we have good and bad ladders. They do not mix. For only one of them do we obtain integrable functions.

3. I should like to say a few words about the approximation procedure that allows us to go beyond the pure cases. The theory is very different from the usual perturbation theory, since now each pure case originates its class of modified equations that can be solved by the perturbation method.

Let us assume

$$r = \overset{(0)}{r} + \epsilon \overset{(1)}{r} \qquad (\epsilon \text{ small parameter})$$

and the existence of

$$k = \overset{(0)}{k} + \epsilon \overset{(1)}{k}, \qquad L = \overset{(0)}{L} + \epsilon \overset{(1)}{L}$$

so that the *approximate* factorization is possible. That is, (B) is equivalent to (A), not rigorously now, but only if we neglect expressions of the order of ϵ^2. Then we can proceed the same way as before and obtain (C), the manufacturing process, the

$$\lambda = \overset{(0)}{\lambda} + \epsilon \overset{(1)}{\lambda}$$

and the key functions.

We shall illustrate the procedure on an example of an equation closely related to that appearing in the Stark effect.

Let

$$(9) \qquad r = \frac{2a}{x} - \frac{m(m+1)}{x^2} + \epsilon \overset{(1)}{\left(\frac{2a}{x} + \frac{m(m+1)}{a} + 2x\right)}.$$

Then:

$$k = \frac{m}{x} - \frac{a}{m} + \epsilon \overset{(1)}{\left(\frac{mx}{a} - \frac{a}{m}\right)}, \qquad L(m) = -\frac{a^2}{m^2} + \epsilon \overset{(1)}{\left(-\frac{3m^2}{a} - \frac{2aa}{m^2}\right)},$$

$$(10) \qquad \lambda = -\frac{a^2}{(n+1)^2} + \epsilon \left(-\frac{3(n+1)^2}{a} - \frac{\overset{(1)}{2aa}}{(n+1)^2}\right),$$

$$y^n = \overset{(0)}{y^n} \exp \epsilon \left(\frac{(n+1)x^2}{2a} - \frac{\overset{(1)}{ax}}{(n+1)}\right) = \exp\left\{\int k(x, n+1)\, dx\right\}$$

and the last equations represent the solution of our equation. Similarly we can find the solution of an equation up to the next approximation, though the procedure then becomes more complicated.

4. I have stressed here rather the technical than the principal part of the problem of factorization. There is no doubt in my mind that the method is an improvement over old methods. It does not vary from one type of differential equation to another. It deals, not with the explicit solutions but with the structure of the differential equations themselves. Yet I believe that the method hides some deeper problems of symmetry that probably lie in the region of group theory. I should be happy if my report were to induce some mathematicians to clarify the questions connected with the use of the factorization method.

Bibliography

1. E. Schroedinger, Proceedings of the Royal Inst. Academy vol. 46 A (1940) p. 9; ibid. (1941) p. 183; ibid. vol. 47 A (1941) p. 53.

2. L. Infeld, Physical Review vol. 59 (1941), p. 737; ibid. vol. 72 (1947) p. 1125. Transactions of the Canadian Royal Society vol. 36 (1942) p. 7.

3. C. C. Lin, Physical Review vol. 59 (1941) p. 841.

4. T. E. Hull and L. Infeld, Physical Review vol. 74 (1948) p. 905.

University of Toronto,
Toronto, Ontario.

NONLINEAR ELECTRICAL NETWORKS

BY

R. J. DUFFIN

This paper discusses electrical networks which, though nonlinear, nevertheless have a simple qualitative behavior.[1] Networks of this class shall be termed *reliable*. The qualitative behavior of reliable networks may be summarized in the following statement: *Reliable networks share the uniqueness properties of linear networks.*

1. Direct current networks. A direct current network is a collection of resistors and batteries, interconnected. The network is linear if the resistors obey Ohm's law, $p = Ri$. Here p is the potential drop across the resistor, i is the current through the resistor, and R is a positive constant called the resistance.

The network problem is to determine the currents flowing in the various resistors. To express equations which determine the current flow, it is customary to choose a maximal set of independent circuits of the network. If these circuits are maximal and independent, then any conservative flow of current may be uniquely expressed as a superposition of cyclic currents $i_1, i_2, \cdots, i_n$ flowing in these n closed circuits. Let $e_1, e_2, \cdots, e_n$ be the sum of the electromotive forces of the batteries in these circuits. The sum of the potential drops across the resistors in the first circuit is equated to e_1, and so on. This yields n equations for the cyclic currents. By use of vector notation these equations may be expressed as the following single equation:

$$(1) \qquad e = Ri.$$

Here e and i are vectors with n components as indicated above, and R is a symmetric matrix.

If u and v are vectors, let their scalar product $\sum_1^n u_j v_j$ be indicated by (u, v). Then the power supplied by the batteries is given by (e, i). This power is dissipated in the resistors in the form of heat. The rate of emanation of heat in a resistor is its resistance multiplied by the square of the current through the resistor. Thus the power supplied can not be negative; therefore, relation (1) gives the inequality $(Ri, i) \geqq 0$. In other words, R is a semi-definite matrix. In an actual physical network none of the wires can have zero resistance, so it is clear that R is actually positive-definite. A positive-definite matrix is nonsingular; therefore the currents are uniquely determined in a linear network.

[1] Some of the material in this paper has been presented in greater detail in the Bulletin of the American Mathematical Society, vol. 52 (1946) pp. 833-838; vol. 53 (1947) pp. 963-971; and vol. 54 (1948) pp. 119–127 and vol. 55 (1949) pp. 119–129.

If the resistors obey a nonlinear law relating potential drop to current flow, there arises a set of n simultaneous nonlinear equations. It is not altogether evident to the intuition how many stable states of current could exist; at least, it was not evident to mine when I encountered such problems for the first time. For the purpose of amusement I was attempting to construct a network with several states. Rectifiers and Ohmic resistors were combined by cut and try methods. But cut and try as I might, the ammeter readings were disappointingly unique. The negative result of this amateur experiment suggested the concept of reliable networks.

A large class of resistors used in engineering is such that the differential resistance lies between positive limits. Thus the linear law $p = Ri$ is replaced by $p = V(i)$, so if $dp = V'di$, then

$$b \leqq V' \leqq c \tag{2}$$

for positive constants b and c. Such resistors I have termed *quasi-linear*. Examples of quasi-linear resistors are selenium, copper oxide, silicon carbide (called "thyrite"), and thermionic rectifiers. It is to be shown now that networks with quasi-linear resistors are reliable.

Consider a linear network with resistance matrix R in which the Ohmic resistors are replaced by quasi-linear resistors. Then equation (1) is replaced by

$$e = V(i) \tag{3}$$

where $V(i)$ is a vector function. Then $de = V'di$ where V' is the matrix of differential coefficients. This last relation may be thought of as defining a "differential network." In the linear case the differential network is defined by $de = R\,di$. Comparing the heat dissipated by these two differential networks gives the inequality

$$b(Rx, x) \leqq (V'x, x) \leqq c(Rx, x). \tag{4}$$

Here $x = di$ is an arbitrary vector and b and c are positive constants. If the linear resistors were all of unit resistance and the quasi-linear resistors satisfied (2) with uniform bounds b and c, then the same b and c are valid in (4). A vector function $V(i)$ which satisfies (4) I have termed a *quasi-linear replacement* of a linear vector function Ri.

If R is positive-definite, it is clear from (4) that V' is positive-definite. Thus the determinant of V' (the Jacobian) never vanishes. By the implicit function theorem of the differential calculus this implies that i is a function of e. In other words, the state of currents is uniquely determined by the electromotive forces.

2. **Variable current networks.** In a variable current network the batteries are replaced by generators whose electromotive forces are functions of the time. Account must now be taken of inductance and capacitance. First consider the case of a single circuit. An inductor of inductance L, a resistor of resistance R, and a capacitor of capacitance $1/S$ are connected in series with a generator of

electromotive force e. Equating the potential drops to the electromotive force yields the following differential equation:

$$L\,di/dt + Ri + S\int i\,dt = e. \tag{5}$$

Energy considerations demand that L and S be non-negative.

In the general linear network with inductors, resistors, and capacitors, the vector network equation may be put in the form (5) provided i and e are given the same vectorial interpretation as before. Because the inductors and capacitors enter the geometry of the circuits in the same way as resistors, it follows that L and S are symmetric semi-definite matrices.

If the linear resistors are replaced by quasi-linear resistors, a reliable network again results. The network equation becomes:

$$L\,di/dt + V(i) + S\int i\,dt = e. \tag{6}$$

Variable current linear networks have the following uniqueness property: A given electromotive force may give rise to more than one state of current, but as time goes on the transient currents die out and there is a unique relation between the electromotive force and the current state. In other words, equation (5) has a unique asymptotic solution. This property will now be demonstrated for equation (6).

Suppose that i' and i^* are two solutions of (6) for a given e which may be an arbitrary function of time. Let $x = i' - i^*$. Then $V(i') - V(i^*) = V'_m x$ where $V'_m = \int_0^1 V'(i^* + \theta x)\,d\theta$. Integrating (4) with respect to θ shows that V'_m satisfies (4). Subtracting the equations satisfied by i' and i^* gives

$$L\,dx/dt + V'_m x + S\int x\,dt = 0.$$

Taking the scalar product with x gives

$$(L\,dx/dt, x) + (V'_m x, x) + \left(S\int x\,dt, x\right) = 0.$$

Integrating from 0 to t gives

$$(Lx, x)/2 + \int_0^t (V'_m x, x)\,dt + \left(S\int x\,dt, \int x\,dt\right)\Big/2 = A.$$

Here A is a constant of integration. The terms on the left are non-negative, so they must be uniformly bounded for all t by the constant A. By inequality (4), $b\int_0^t (Rx, x)\,dt \leq A$. The integrand (Rx, x) is non-negative, so on the average it must approach zero as t approaches infinity. If R is positive-definite, this implies that x approaches zero in mean. In other words, all solutions approach one another as time goes on, and there is a unique asymptotic solution.

In the case in which the electromotive force is periodic, the asymptotic solution of a linear network is also periodic. The same result can be demonstrated for reliable networks. The method of proof employs the notion of Hilbert space of periodic functions. The network equation is considered to be a transformation from the Hilbert space of electric current to the Hilbert space of electromotive force. A well known lemma on linear transformations of Hilbert space states that closure implies completeness. A similar lemma was developed here for a class of nonlinear transformations of Hilbert space. This lemma guaranteed that the network transformation had a unique inverse; that is, equation (6) has a periodic solution.

3. **Nonlinear inductors.** Many of the inductors used in engineering are coils wound on iron cores. The permeability of iron is not constant; hence, the inductance is not constant. In most applications this nonlinearity is an undesirable feature which designers try to minimize. There are, however, important applications of iron core inductors whose operation depends essentially on the nonlinearity, such as flux gate magnetometers, magnetic voltage regulators, frequency multipliers, and magnetic amplifiers.

It is to be shown now that networks composed of iron core inductors and Ohmic resistors are reliable. The definition of the inductance of an inductor is the flux linked with the circuit when unit current flows. If ϕ is the flux, then $\phi = Li$. In a nonlinear case this relation is replaced by $\phi = F(i)$. It is a well known experimental fact that neglecting hysteresis, the differential permeability of iron lies between positive limits. Thus, $b \leqq F' \leqq c$. This is the same relation as that stated for quasi-linear resistors. The vector network equation now becomes:

$$d/dtF(i) + Ri = e. \tag{7}$$

Except for the operator d/dt, the inductors enter the network equations in the same way as the resistors. Hence the vector function $F(i)$ is a quasi-linear replacement of Li.

Integrating equation (7) with respect to time gives

$$F(i) + R\int i\,dt = g. \tag{8}$$

Here g is essentially the integral of the electromotive force vector e. The equation for a network containing quasi-linear resistors and linear capacitors but no inductors is

$$V(i) + S\int i\,dt = e. \tag{9}$$

Equations (8) and (9) are abstractly identical, F corresponding to V, R corresponding to S, and g corresponding to e. The results stated previously for networks with quasi-linear resistors therefore apply to these networks with iron core inductors having negligible hysteresis. Thus a given electromotive force determines a unique current flow after the transient currents have died out. If e is periodic and has the average value zero, then g is also periodic, so a unique periodic flow of current can subsist.

To shorten the exposition, networks with transformers have not been discussed. However, the mathematical formulation of such networks introduces no change, so the results of this section and of the previous section remain valid for networks with transformers.

4. **Reliable operators.** The main characteristic of the mathematics of this century has been the development of unified treatments of seemingly unrelated problems. For example, linear problems involve linear operators, and the treatment of linear operators has been neatly unified by such postulational theories as Hilbert space and Banach space. It would be desirable to have a unified theory for nonlinear problems, but very little has been accomplished in this direction. The trouble with nonlinearity is that it is too vast a territory. Before it can be conquered by unified theories it seems necessary to find some way of subdividing it.

The consideration of reliable networks suggests that a useful subdivision of nonlinearity to study would be operators having a unique inverse. Such operators might be termed *reliable*. Linear and nonlinear reliable operators are topologically equivalent; that is, they are both one-to-one transformations of some space on some other space. It is to be expected, therefore, that reliable operators can be handled by methods somewhat similar to those used for linear operators. As a consequence, reliable operators should not be too unpleasant to handle.

These remarks seem rather obvious; nevertheless, most of the literature on nonlinearity is concerned with operators that do not have a unique inverse. Thus the student is first introduced to nonlinearity in the study of algebraic equations, where the main emphasis is on the concept that an equation of the nth degree has n roots. In applied mathematics a typical example of this preoccupation with unreliable systems is the theory of the Duffing equation. This equation describes the forced oscillations of a mass connected to a nonlinear spring. It has been shown that a fantastic variety of states may result from a given impressed force.

In the theory of linear operators the demonstration of linearity is in most cases trivial. By contrast a central problem in a theory of reliable operators would be the demonstration of reliability. Any well-developed linear theory should be fair game for a reliable generalization. For example, it appears that potential theory has such a nonlinear generalization.

Carnegie Institute of Technology,
Pittsburgh, Pa.

RAY THEORY VS. NORMAL MODE THEORY IN WAVE PROPAGATION PROBLEMS

BY

C. L. PEKERIS

1. Introduction. In the problem of propagation of radiation emanating, say, from a point source it is usually possible to express the field at a distant point either as a sum of contributions from various rays or from normal modes. In the immediate vicinity of the source the number of rays necessary to determine the field is small, because the path of the rays due to multiple reflection (when the medium is bounded vertically) is longer than that of the direct ray. On the other hand, the number of modes required to describe the field at close ranges is large. At great distances the situation is reversed, the normal mode solution being as a rule more rapidly convergent than the ray solution. There are cases when only one type of solution exists, such as the normal mode solution for points beyond the horizon, when the question of a choice of representation does not arise. A still outstanding problem in propagation theory is to provide an efficient method of determining the field at intermediate ranges where both the number of rays required and the number of normal modes is large.

The aim of this investigation is to establish the mathematical relationship between the ray and normal mode representations. In order not to encumber the analysis with matters which are not germane to the central topic, we shall base our discussion on the simple case of propagation from a point source situated at $z = h$ in a uniform medium bounded by perfectly reflecting horizontal surfaces at $z = 0$ and $z = H$. The Hertzian potential Ψ is expressed in §2 by the integrals (11) or (12). These integrals are then transformed in §3 into the ray solution (15) on the one hand, and the normal mode solution (16) on the other hand. The next sections are devoted to transformations of one type of solution into another. In §4 it is shown that in the limit of $H \to \infty$ the normal mode solution can be summed to give the direct ray and the ray reflected from the surface $z = 0$. In §5 it is shown that the ray solution is the Poisson transform[1] of the normal mode solution.

The medium with perfectly reflecting boundaries was chosen because for it both the ray representation and the normal mode representation are *exact*. In the case of more general types of media the ray representation is only approximate, while the normal mode solution may need to be appended by a "branch line" integral.[2]

[1] E. C. Titchmarsh, *Introduction to the theory of Fourier integrals*, Oxford University Press, 1937, p. 61.

[2] C. L. Pekeris, *Theory of propagation of explosive sound in shallow water*, Memoir 27 Geological Society of America, New York, 1948.

2. **Formal solution.** We consider a point source situated at $z = h$ in a uniform medium bounded at $z = 0$ and $z = H$ by perfectly reflecting horizontal boundaries. The field can be derived from a Hertzian potential Ψ satisfying the wave equation

$$\nabla^2\Psi + k_0^2\Psi = 0, \qquad k_0 = 2\pi/\lambda, \tag{1}$$

and the boundary conditions

$$\Psi = 0 \quad \text{at} \quad z = 0 \quad \text{and} \quad z = H; \tag{2}$$

$$\Psi \to (1/R)e^{-ik_0R}, \tag{3}$$

for small distances R from the source. Here a factor $e^{i\omega t}$ is implied throughout. Since the boundaries are horizontal, we attempt to synthesize Ψ from elementary solutions of the form

$$\psi = J_0(kr)F(z)G(k). \tag{4}$$

From (1) it follows that

$$(d^2F/dz^2) + \beta^2F = 0, \tag{5}$$

$$\beta = (k_0^2 - k^2)^{1/2}, \qquad k_0 > k,$$
$$= -i(k^2 - k_0^2)^{1/2}, \qquad k > k_0.^{3} \tag{6}$$

The appropriate solutions of (5) are

$$F_1 = A \sin \beta z, \qquad 0 < z < h, \tag{7}$$

$$F_2 = B \sin \beta(H - z), \qquad h < z < H. \tag{8}$$

At $z = h$ we must have[4]

$$F_1 = F_2, \qquad \dot{F}_1 - \dot{F}_2 = 2k \tag{9}$$

provided we integrate with respect to k from 0 to ∞. We thus obtain

$$A = 2k \sin \beta(H - h)/(\beta \sin \beta H), \qquad B = 2k \sin \beta h/(\beta \sin \beta H), \tag{10}$$

$$\Psi_1 = 2 \int_0^\infty J_0(kr)k\, dk \sin \beta z \sin \beta(H - h)/(\beta \sin \beta H), \qquad 0 < z < h, \tag{11}$$

$$\Psi_2 = 2 \int_0^\infty J_0(kr)k\, dk \sin \beta h \sin \beta(H - z)/(\beta \sin \beta H), \qquad h < z < H. \tag{12}$$

3. **The ray solution and the normal mode solution.** The ray solution can be written down directly, because the reflection coefficient of both boundaries is -1. The effect of the boundaries can be exactly simulated by replacing them

[3] Since the integrands in (11) and (12) are even functions of β, the sign in front of i is immaterial.

[4] See Appendix in paper cited in footnote 2.

with a system of images strung along a vertical through the source. These consist first of the dipole formed by the source and its negative image at $z = -h$ plus other dipoles spaced at a distance $2H$ apart, with neighboring dipoles having opposite polarity. The ray solution can also be derived from (11) or (12). Taking Ψ_1, for example, we make the following expansion in the integrand,

$$\text{(13)} \quad \begin{aligned} 2 \sin \beta z \sin \beta(H - h)/(\sin \beta H) &= -i(1 + e^{-2i\beta H} + e^{-4i\beta H} + \cdots) \\ &\times [e^{i\beta(z-h)} + e^{-i\beta(z-h+2H)}) - e^{-i\beta(z+h)} - e^{i\beta(z+h-2H)}] \end{aligned}$$

and use the relation

$$\text{(14)} \quad \int_0^\infty J_0(kr)k\, dke^{-i\beta z}/\beta = ig(R^2); \qquad g(R^2) \equiv (1/R)e^{-ik_0R},$$

obtaining

$$\text{(15)} \quad \begin{aligned} \Psi_1 = {} & \{g[r^2 + (h - z)^2] - g[r^2 + (h + z)^2]\} \\ & + \sum_{n=1}^{\infty} \{g[r^2 + (h - z + 2nH)^2] + g[r^2 + (z - h + 2nH)^2] \\ & - g[r^2 + (z + h + 2nH)^2] - g[r^2 + (-z - h + 2nH)^2]\}. \end{aligned}$$

The identical result would have been obtained had we used Ψ_2. The first two terms represent the direct ray and the ray reflected from the boundary $z = 0$. Each quadruplet of terms in the following sum represents the four rays which suffer n reflections from the plane $z = 0$.

The normal mode solution is obtained by evaluating the integral (11) or (12) in terms of the residues of the integrand. This can be accomplished by making, in the first instance, a transformation of the integral over $J_0(kr)$ from 0 to ∞ into an integral over $H_0^{(2)}(kr)$ taken along a different path of integration in the complex k-plane. The details of this step will be found elsewhere,[4] and we shall give here merely the final result:

$$\text{(16)} \quad \Psi = -(2i\pi/H) \sum_{n=1}^{\infty} H_0^{(2)}(k_n r) \sin (n\pi h/H) \sin (n\pi z/H),$$

$$\text{(17)} \quad k_n = [k_0^2 - (n^2\pi^2/H^2)]^{1/2}, \qquad k_0 > (n\pi/H),$$

$$\text{(18)} \quad k_n = -i[(n^2\pi^2/H^2) - k_0^2]^{1/2}, \qquad k_0 < (n\pi/H)^{\cdot}$$

(16) expresses the Hertzian potential Ψ as a sum of normal modes. These are of different nature depending on whether condition (17) or (18) holds. In the former case the amplitude of the mode is weakened with distance r only through the cylindrical divergence, because

$$\text{(19)} \quad H_0^{(2)}(k_n r) \to (2/\pi k_n r)^{1/2} \exp [(i\pi/4) - ik_n r)].$$

On the other hand, when $k_0 < (n\pi/H)$ and k_n is imaginary, the factor $\exp(-ik_n r)$ gives an exponential attenuation with range. Thus when $\lambda > 2H$, all modes suffer exponential attenuation with range. For $H < \lambda < 2H$, the first mode is free, but all higher modes are attenuated, and so on. For moderate and large ranges (in terms of H), the number n of modes required in the representation (16) is less than $2H/\lambda$. The normal mode solution has poor convergence when the wavelength is a small fraction of the thickness H of the medium.

4. **Summation of the normal mode solution in the limit $H \to \infty$.** For a given wavelength λ the number of modes necessary to sum is about $(2H/\lambda)$, which can become very large with increasing H. On the other hand, we know that in the limit of $H \to \infty$ the solution reduces exactly to the direct ray and the ray reflected from the surface $z = 0$. We shall now perform a summation of (16) in this limit and prove that this is actually the case. With $x = (n\pi/H)$, we can write that

$$-\frac{2i\pi}{H}\sum_{n=1}^{\infty} H_0^{(2)}(k_n r)\sin(n\pi h/H)\sin(n\pi z/H)$$

$$\simeq -i2\int_0^\infty dx H_0^{(2)}[r(k_0^2 - x^2)^{1/2}]\sin(hx)\sin(zx) \tag{20}$$

$$= [r^2 + (z-h)^2]^{-1/2}\exp\{-ik_0[r^2 + (z-h)^2]^{1/2}\} - [r^2 + (z+h)^2]^{-1/2}\cdot$$

$$\cdot\exp\{-ik_0[r^2 + (z+h)^2]^{1/2}\}.$$

In the above derivation use was made of the relation

$$\int_0^\infty H_0^{(2)}[r(k_0^2 - x^2)^{1/2}]\cos(zx)\,dx = i[r^2 + z^2]^{-1/2}\exp\{-ik_0[r^2 + z^2]^{1/2}\}. \tag{21}$$

5. **Transformation of the normal mode solution into the ray solution by a Poisson summation.** We now turn to the principal theme of this investigation, which is to establish the mathematical relationship between the two series (15) and (16), of rays and modes respectively, by which the solution for the Hertzian potential Ψ can be represented. We shall show that the two series are related by a Poisson transform:[1]

$$\beta^{1/2}\left[(1/2)F_c(0) + \sum_{n=1}^{\infty} F_c(n\beta)\right] = \alpha^{1/2}[(1/2)f(0) + \sum_{n=1}^{\infty} f(n\alpha)]; \tag{22}$$

$$F_c(x) \equiv (2/\pi)^{1/2}\int_0^\infty f(t)\cos(xt)\,dt; \qquad \alpha\beta = 2\pi. \tag{23}$$

Poisson's relation (22) affords a transformation of one series into another, the new terms being cosine transforms of the old. In our case

$$f(n) = -i(2\pi/H)H_0^{(2)}[r(k_0^2 - n^2\pi^2/H^2)^{1/2}]\sin(n\pi z/H)\sin(n\pi h/H), \tag{24}$$

and with $u = (t\pi/H)$, $v = (xH/\pi)$, and g as defined in (14),

$$F_e(x) = -2i(2/\pi)^{1/2} \int_0^\infty \cos [(xH/\pi)u] H_0^{(2)}[r(k_0^2 - u^2)^{1/2}] \sin (uh) \sin (uz)\, du$$

$$(25) \qquad = (2\pi)^{-1/2}\{g[r^2 + (v + z - h)^2] + g[r^2 + (v - z + h)^2] - g[r^2 + (v + z + h)^2] - g[r^2 + (v - z - h)^2]\},$$

$$(26) \qquad (1/2)F_e(0) = (2\pi)^{-1/2}\{g[r^2 + (z - h)^2] - g[r^2 + (z + h)^2]\}.$$

Here again use was made of (21). Now letting $\alpha = 1$, $\beta = 2\pi$ in (22), we obtain precisely the ray solution in (15).

6. **Discussion.** We have chosen a particularly simple propagation problem and have proven for it that the ray solution and normal mode solution stand in the relationship of a Poisson transform. It would be of interest to establish whether this relationship holds in more complicated propagation problems. The difficulty that can be foreseen is a practical one, namely the carrying out of the Fourier transform in (23), which in our case was facilitated by relation (21). Another line to follow up would be to use the Euler-Maclaurin expansion[5] to sum either the normal mode solution or the ray solution. It should also be emphasized that in the general case it may happen that neither the ray series nor the normal mode series are exact solutions.

THE WEIZMANN INSTITUTE,
REHOVOTH, ISRAEL.

[5] Whittaker and Watson, *Modern analysis*, Cambridge University Press, 1935, p. 127.

SYSTEMS OF WIENER-HOPF INTEGRAL EQUATIONS AND THEIR APPLICATION TO SOME BOUNDARY VALUE PROBLEMS IN ELECTROMAGNETIC THEORY

BY

ALBERT E. HEINS

When one recalls the number of mixed boundary value problems for which a satisfactory solution may be offered, it is indeed gratifying to know that there is one class of such problems for which a great deal may be said. The mixed boundary value problem is a product of the development of classical physics. Such problems arise in the solution of linear partial differential equations subject to linear boundary conditions. Instead of assigning a single boundary condition along a coordinate surface, we are required to assign two boundary conditions along such a surface. For example we may find that the function itself is assigned along a part of this surface, and a linear combination of the normal derivative and the function along the remainder. Such boundary value problems fall outside of the scope of the elementary classical methods. They are best formulated as integral equations. We shall limit ourselves here to regions bounded by coordinate surfaces or parts of coordinate surfaces.

More specifically we shall be interested in finding the solution of

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + k^2\phi = 0 \qquad (k > 0) \tag{1}$$

for such regions as Fig. 1 or Fig. 2, where equation (1) is to be solved subject to the fact that either (i) $\phi = 0$ on the solid parallel lines in Figs. 1 or 2 or (ii) $\partial\phi/\partial y = 0$ on the same lines in the same figures. We recall now that the solution of equation (1) for Figs. 1 or 2 requires that we assign conditions at infinity. For example in Fig. 1 we must decide what modes are propagating in the ducts for $x >> 0$ and $x << 0$. Similarly for Fig. 2, we not only assign conditions on $\phi(x, y)$ for $x >> 0$ in the ducts but also conditions on $\phi(x, y)$ for $r = (x^2 + y^2)^{1/2} >> 0$ outside of the ducts. These conditions plus the boundary conditions (i) or (ii) on the solid lines supply us with the necessary information to formulate these boundary value problems.

Let us now examine the geometry of Figs. 1 or 2 in more detail. Both figures may be viewed as nontrivial generalizations of the classical diffraction problem treated by Sommerfeld [10] at the turn of the century. Sommerfeld considered, amongst other things, the diffraction of a plane wave incident upon a semi-infinite perfectly conducting, "zero thickness", barrier (see Fig. 3). If the exciting plane wave has only one component of the electric field (or magnetic field) perpendicular to the plane of the paper, then the problem we have described above may be formulated as an integral equation of the Wiener-Hopf [8] type. That is, this problem is a scalar electromagnetic problem in the sense that the solution of equation (1) subject to the boundary condition (i) and appropriate conditions at infinity, depends on one field quantity $\phi(x, y)$. This function $\phi(x, y)$ can be identified with E_z and the y derivative of $\phi(x, y)$ is proportional to the x component of the magnetic field. The boundary condition (ii) corresponds to the case in which there is only a single component of the magnetic field perpendicular to the plane of the paper. The geometry for Fig. 3 is nothing more than the semi-infinite straight line, while the difficulties in Figs. 1 and 2 arise from a similar geometry.[1] The boundary value problem for Fig. 3, using equation (1) and the boundary condition (i), plus the fact that the

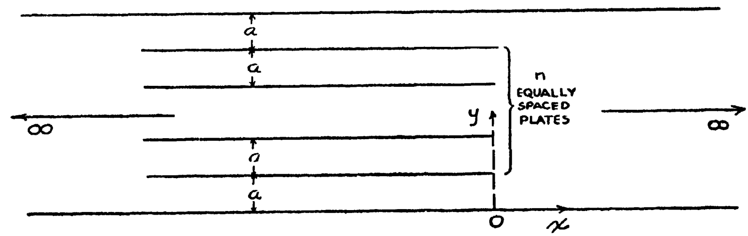


FIG. 1

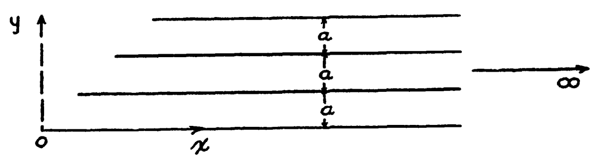


FIG. 2

plate has been excited by a plane wave, may be formulated as an integral equation of a special type, as we have already remarked. Indeed, for $x > 0$, we have

$$(2) \qquad 0 = e^{ikx\cos\theta} + \frac{i}{4}\int_0^\infty H_0^{(1)}[k\,|x - x'|]\,I(x')\,dx'$$

where the symbol $H_0^{(1)}$ is the usual one for the Hankel function of the first kind while $I(x')$ is the discontinuity of H_x across a plate, that is, essentially the surface current density on the plate. We recognize that equation (2) is a nonhomogeneous integral equation of the Wiener-Hopf type because of its limits of integration and the $x - x'$ character of the kernel. This problem distinguishes itself from those in Figs. 1 and 2 by one important point. Here we can formulate the boundary value problem as a single integral equation and as such, the original Wiener-Hopf technique carries over. For Figs. 1 and 2, such is not the case, since now we are required to find the current densities on several semi-infinite barriers. In these cases then, we get a system of such integral equations. We shall still employ the boundary conditions of the types (i) or (ii), that is, the orientation of the exciting electric field or the magnetic field is perpendicular to the plane of the paper. The

[1] The variable z does not enter into our calculations and all field quantities are considered independent of it.

technique for the single integral equation has been applied to numerous problems by Carlson [1; 5], Copson [2], Feshbach [6], Heins [3; 4], Levine [7], and Schwinger [9]. Applications are not limited by these authors to electromagnetic field theory problems, but to studies of boundary value problems in acoustics, elasticity, and hydromechanics.

We have already remarked that the integral equations for Figs. 1, 2, or 3 may be formulated in terms of unknown current densities on the solid semi-infinite barriers in these figures. They might equally well have been formulated in terms of the tangential electric field on the geometric extension of these barriers and we should still get a Wiener-Hopf type of integral equation. For Fig. 1, for example, such a formulation would lead a system of n integral equations of the Wiener-Hopf type. Once we specified the polarization (which is in turn

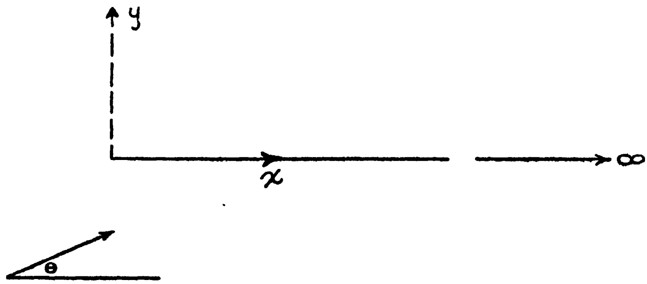


FIG. 3

equivalent to stating the boundary conditions (i) or (ii)) and also the wave-length range, that is, the interval in which ak falls, then the system may be formulated [9].

The calculation of the asymptotic form of the tangential component of the electric field on the surfaces $y = \sigma a$, $\sigma = 1, \cdots, n$, $x > 0$, enables us to proceed with the calculation of the Fourier transform of the system. Of course, we are required to know that all transforms have a common strip of regularity parallel to the real axis of a complex w plane. To illustrate what we should get, let $E_i(x)$ be the tangential component of the electric field on the line $y = ia$, $i = 1, \cdots, n$. Let $K_{ij}(x - x')$ be certain linear combinations of the Green's functions used in formulating the integral equations. Finally let $F_j(x)$ be the form of the propagating modes in the jth duct. Then we find

$$\sum_{j=1}^{n} \int_0^\infty K_{ij}(x - x')E_i(x')\, dx' + F_j(x) = 0$$

for $x > 0$. This is a system of Wiener-Hopf integral equations for the functions $E_i(x)$.

We now let $k(w)$ be the matrix of the Fourier transforms of the kernels $K_{ij}(x)$, an $n \times n$ matrix, and further write

$$e_m(w) = \int_0^\infty E_m(x')e^{-iwx'}\, dx', \qquad m = 1, \cdots, n,$$

$$d_m(w) = \int_0^\infty D_m(x')e^{-iwx'}\,dx', \qquad m = 1, \cdots, n,$$

$$f_m(w) = \int_0^\infty F_m(x')e^{-iwx'}\,dx', \qquad m = 1, \cdots, n.$$

Here $E_m(x)$ is the tangential component of the electric field along the line $y = ma$, $x > 0$. $D_m(x) = 0$, $x > 0$, is the extending function of the mth integral equation for negative arguments, and we define $F_m(x)$ and $E_m(x)$ to vanish identically for negative x. Let

$$\tilde{d}(w) = (d_1, \cdots, d_n), \qquad \tilde{e}(w) = (e_1, \cdots, e_n), \qquad \tilde{f}(w) = (f_1, \cdots, f_n)$$

be horizontal matrices. The symbol $\tilde{}$ is to be understood as that of transposition. Then the Fourier transform of the system is

$$k(w)e(w) = f(w) + d(w). \tag{3}$$

Here $k(w)$ and $f(w)$ are known matrices. $k(w)$ and its determinant are regular in some strip parallel to the real axis of the w-plane. $f(w)$ and $e(w)$ are regular in some lower half-plane, while $d(w)$ is regular in some upper half-plane. All of these matrices are regular in a common strip. The matrix $k(w)$ enjoys the special property that it is symmetric about both diagonals for Figs. 1 and 2.

We now ask whether we can factor the matrix $k(w)$ to the product $k_+(w)\,k_-(w)$ where $k_+(w)$ and $k_-(w)$ are matrices which are regular in the appropriate upper and lower half-planes. Further, from the point of view of taking inverses we require that det $k_+(w)$ and det $k_-(w)$ are also regular in the appropriate half-planes. This can be done with the aid of Sylvester's Theorem [11]. That is, we can write

$$k(w) = e^{l(w)}$$

and decompose the matrix $l(w)$ additively into two terms $l_+(w)$ and $l_-(w)$ where the subscripts $+$ and $-$ carry their customary connotation. In this case $l_+(w)$ and $l_-(w)$ commute because of the special form of $l(w)$, if we are concerned with Fig. 1. Hence equation (3) becomes

$$e^{l_-(w)}e(w) = e^{-l_+(w)}f(w) + e^{-l_+(w)}d(w). \tag{4}$$

The left side of the matrix equation (4) is regular in some lower half-plane while the term $e^{-l_+(w)}\,d(w)$ is regular in some upper half-plane. The middle term is still regular in a strip common to these half-planes. Because of the simplicity of $f(w)$ in this case, it is possible to decompose this term

$$e^{-l_+(w)}f(w)$$

further so that one term will be regular in the lower half plane while the other will be regular in the upper half-plane and we shall still have the common strip of regularity. This completes the decomposition of equation (3). We have

$$e^{l_-(w)}e(w) + h(w) = e^{-l_+(w)}d(w) + g(w) \tag{5}$$

where

$$g(w) - h(w) = e^{-l_+(w)}f(w)$$

and $g(w)$ is regular in the upper half-plane, while $h(w)$ is regular in the lower half-plane described above.

Since every element of the matrix equation

$$e^{l_-(w)}e(w) + h(w) \tag{6}$$

is regular in some lower half-plane, while every element of the matrix equation

$$e^{-l_+(w)}d(w) + g(w) \tag{7}$$

is regular in some upper half-plane, and both half-planes have a common strip of analyticity, every element on the left side of (5) is the analytic continuation of

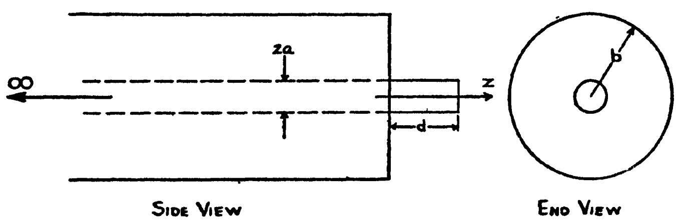


Fig. 4

the corresponding element on the right side of equation (5). It follows that every element of (6) and (7) is regular everywhere, that is,

$$e^{l_-(w)}e(w) + h(w) = j(w)$$

and

$$e^{-l_+(w)}d(w) + g(w) = j(w)$$

where $j(w)$ is a vertical matrix of integral functions. $j(w)$ can be determined by studying the asymptotic forms of $l_-(w)$, $l_+(w)$, $e(w)$, $d(w)$, $h(w)$, and $g(w)$. For the problem in Fig. 1, $j(w)$ is identically zero. This then enables us to find the matrix $e(w)$. Given $e(w)$, it is a fairly simple calculation to determine the relation between the amplitudes of the reflected and transmitted waves for $x > 0$ and $x < 0$ and hence we can reduce the problem to one in transmission line networks. The details of this problem will appear in a subsequent publication.

In order to write $k(w)$, an $n \times n$ matrix which is regular in some strip of the complex w-plane, in the form $e^{l(w)}$ where $l(w)$ is a matrix with similar properties, we are required to know the characteristic values of $k(w)$. In the decomposition of $l(w)$ into $l_+(w) + l_-(w)$ we are required to have the further assurance that $l_+(w)$ and $l_-(w)$ commute. There are boundary value problems in electromagnetic theory for which the above method of factoring matrices is no longer applicable. As an example of this case we cite the problem of radiation from a semi-infinite coaxial line (see Fig. 4).[2] The radius of the inner conductor is a, while that of the outer one is b. The outer conductor is open at $z = 0$ while the inner one is open at $z = d$. The coaxial region sustains the principal mode. In this case the 2×2 matrix of the Fourier transforms of the kernels is symmetric about the main diagonal, and the factoring technique we have described is not applicable because of the non-commutability of the matrices l_+ and l_-. Further investigation is being carried on in this direction.

Bibliography

1. J. F. Carlson and A. E. Heins, *The reflection of an electromagnetic plane wave by an infinite set of plates*, II, Quarterly of Applied Mathematics vol. 5 (1947) pp. 82–88.

2. E. T. Copson, *On an integral equation arising in the theory of diffraction*, Oxford Quarterly of Mathematics vol. 17 (1946) pp. 19–34.

3. A. Heins, *The radiation and transmission properties of a pair of semi-infinite parallel plates*, I, Quarterly of Applied Mathematics vol. 6 (1948) pp. 157–166.

4. ———, *The radiation and transmission properties of a pair of semi-infinite parallel plates*, II, Quarterly of Applied Mathematics vol. 6 (1948) pp. 157–166.

5. A. E. Heins and J. F. Carlson, *The reflection of an electromagnetic plane wave by an infinite set of plates*, II, Quarterly of Applied Mathematics vol. 5 (1947) pp. 82–88.

6. A. E. Heins and H. Feshbach, *The coupling of two acoustical ducts*, Journal of Mathematics and Physics vol. 26 (1948) pp. 143–155.

7. H. Levine and J. S. Schwinger, *On the radiation of sound from an unflanged circular pipe*, Physical Review vol. 73 (1948) pp. 383–406.

8. R. E. A. C. Paley and N. Wiener, *The Fourier transforms in the complex domain*, Amer. Math. Soc. Colloquium Publications, vol. 19, 1934, chap. 4.

9. J. S. Schwinger, *Theory of guided waves*, Radiation Laboratory Publication, forthcoming.

10. A. Sommerfeld, *Mathematische Theorie der Diffraction*, Math. Ann. vol. 47 (1896) pp. 317–374.

11. J. H. M. Wedderburn, *Lectures on matrices*, Amer. Math. Soc. Colloquium Publications, vol. 17, 1934.

Carnegie Institute of Technology,
Pittsburgh, Pa.

[2] Fig. 2 falls in this category also.

ORBITS OF CHARGED PARTICLES IN CONSTANT FIELDS[1]

BY

A. H. TAUB

Solutions of the relativistic equations of motion of a charged particle in a constant electromagnetic field are obtained and classified. The discussion is based on the observation that the four-dimensional anti-symmetric tensor, $f_{\sigma\tau}$, which describes the external field determines one parameter group of Lorentz matrices L of which it is the infinitesimal generator. It is a consequence of the equations of motion that the four-dimensional velocity vector at any point of the orbit is related to its initial value by means of this group of Lorentz transformations.

Results obtained by O. Veblen, J. W. Givens, and the author give a complete classification of these Lorentz transformations in terms of invariants formed from $f_{\sigma\tau}$ as well as closed expressions for L. These results are used to give a classification of the orbits and to obtain various properties of the orbits.

The methods used may be applied to the orbit of a charged particle in the field of a plane wave.

INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

[1] Published in Physical Review vol. 73 (1948) pp. 786–798.

REFLECTIONS FROM BENDS AND CORNERS IN ELECTROMAGNETIC WAVE GUIDES[1]

BY

S. O. RICE

When an electromagnetic wave traveling along a wave guide strikes an irregularity, a reflected wave is set up. In some communication systems it is important to make these reflections as small as possible. Corners and bends in the guide constitute one source of troublesome reflections and accordingly any theory regarding them is of practical interest.

The present paper is concerned with (1) the reflection from a simple corner which changes the direction of a rectangular wave guide by some assigned angle and (2) the reflection from a sectoral horn which may be regarded as a sort of generalized corner since the side walls are bent in opposite directions. The method used goes back, in its essentials, to the nineteenth century work of E. J. Routh. He studied the vibrations of membranes of complicated shape by transforming them, conformally, into rectangles of non-uniform density. By the same device a bent wave guide may be transformed into a straight wave guide in which the propagation constant is not uniform. The magnitude of the reflection coefficient in the transformed guide is equal to that in the original guide.

The problem of obtaining the reflection coefficient for the simple corner may be reduced to the problem of solving an integral equation. When the angle of bend is small the solution may be obtained by successive substitutions and the reflection coefficient expressed approximately as an infinite series.

This series, representing the reflection caused by a discontinuous change of direction, turns out to be related to the series giving the reflection coefficient for a gentle circular bend of the same angle. In fact, if the radius of curvature of the latter is held constant while the angle of bend is made small, the series for the circular bend reduces to that for the corner.

Even when the angle of the corner is not small, numerical results obtained by only two successive substitutions appear to be reasonably accurate. Associated with this is the fact that the reflection coefficient is a stationary value of a certain form associated with the integral equation. This is somewhat analogous to results obtained by Schwinger.

When one attempts to study the reflection from a sectoral horn by setting up the corresponding integral equation, difficulties are encountered. An alternative and more general method, which consists of dealing with an infinite set of ordinary differential equations, may be used.

BELL TELEPHONE LABORATORIES,
NEW YORK, N. Y.

[1] Published in Bell System Technical Journal vol. 28 (1949) pp. 104–156.

WAVE PROPAGATION IN ELECTROMAGNETIC HORNS

BY

A. F. STEVENSON

Exact solutions for the propagation of electromagnetic waves in conical or sectoral horns are known, but there appears to have been no general discussion for horns of any shape. In this paper, approximate solutions for the propagation of electromagnetic waves in a perfectly conducting horn of arbitrary section and flare are obtained, it being assumed that the cross-section of the horn varies slowly and that the solution for a tube of the same *constant* cross-section is known.

These approximate solutions are of the E- or H-wave type, and their general characteristics are discussed. There is in general an attenuation of the wave and a gradual change in the phase velocity as we travel down the horn.

The particular cases of horns of circular and rectangular section with various flares are discussed in detail.

UNIVERSITY OF TORONTO,
TORONTO, ONTARIO

PROBLEMS RELATED TO MEASURING THE FIELD STRENGTH OF HIGH FREQUENCY ELECTROMAGNETIC FIELDS[1]

BY

ROHN TRUELL

An electron beam is directed parallel to a uniform steady magnetic field and parallel to the midplane of two plane parallel plates with a high frequency electromagnetic field having the electric component at right angles to the steady magnetic field. Analysis of this case produces a relation among the variables which is particularly simple when the steady magnetic field H is adjusted so that $\omega = eH/mc \equiv \omega_c$ where ω is the frequency of the electromagnetic field. One use of the relation mentioned is that of measuring the field strength of the high frequency field. With the geometry described the end effects of the fringe fields on the electron paths may be appreciable, and, under some circumstances, the alternative geometry of a rectangular cavity may be used in place of the plane parallel plates. Analysis of this case leads to the conclusion that when $\omega = \pm\pi v_s/z_0 \pm eH/mc$ and when $\omega_c\tau = n\pi$ (n even) (where v_s is the velocity of the electron entering the cavity and z_0 is the length of the cavity in the direction in which the electron beam is introduced, τ is the transit time of the electron through the cavity) there again results a special solution of the equations of motion which, for the purposes of field measurement, differs from the solution to the parallel plate case only by a numerical factor.

BROWN UNIVERSITY,
PROVIDENCE, R. I.

[1] Proceedings of the Institute of Radio Engineers vol. 37 (1949) pp. 1144–1147.

ABERRATION CORRECTION WITH ELECTRON MIRRORS[1]

BY

E. G. RAMBERG

The fulfillment of the Laplace equation by the refracting fields of electron optics severely restricts the possibility of aberration correction in electron-optical systems. Thus third-order spherical and axial chromatic aberration can be made to vanish only by introducing conductors or space charge in the central region of the lens system, by the employment of high-frequency fields, by departure from axial symmetry as suggested by Scherzer, or by the employment of electron mirrors. The last approach appears to be the easiest to realize in practice. This renders it attractive in spite of the fact that special precautions must be taken to prevent undue interference of the source and the specimen with the observation of the latter.

From the point of view of practical utilization it is advantageous that the mirror field be restricted to a relatively short region of the optic axis. This requirement is fulfilled by the family of axial potential distributions $\Phi = C -$ tanh (sinh z), whose focusing conditions and spherical and chromatic aberration have been investigated in some detail. For this purpose the paraxial ray equations are solved in parametric form and the corresponding aberration integrals evaluated. The possibility, in principle, of compensating the spherical aberration of typical magnetic electron microscope objectives in this manner is demonstrated. It is found, however, that the required mirror dimensions are too small for practical realization.

WASHINGTON UNIVERSITY,
ST. LOUIS, MO.

[1] To be published in the Journal of Applied Physics.

DISTRIBUTION PROBLEMS IN THE THEORY OF RANDOM NOISE[1]

BY

MARK KAC

If a receiver consists of an intermediate frequency amplifier (IF), a rectifier, and a linear filter, the output noise is of the form

$$(1) \qquad z(t) = \int_0^\infty K(u)V(x)(t-u))\,du = \int_{-\infty}^t K(t-\tau)V(x(\tau))\,d\tau,$$

where $V(x)$ describes the action of the rectifier, $K(u)$ is the response function of the linear filter, and $x(t)$ a stationary Gaussian process whose correlation function $\rho(t)$ is determined by the IF spectrum. The distributions of quantities (1) are thus of considerable interest in the theory of noise.

For a quadratic rectifier ($V(x) = x^2$) the present author and A. J. F. Siegert proved (Journal of Applied Physics vol. 18 (1947) pp. 383–397) that the characteristic function of the distribution of (1) is $(D(i\zeta))^{-1/2}$, where $D(\lambda)$ is the Fredholm determinant of the integral equation

$$(2) \qquad \int_0^\infty \rho(s-t)K(t)f(t)\,dt = \lambda f(s).$$

For other rectifiers some special results can be obtained. For instance, let $x(t)$ be a Markoffian process (i.e., the IF is an R-L circuit) and let $K(u) = \exp(-\gamma u)$, $u > 0$, $K(u) = 0$, $u < 0$ (i.e., the linear filter is an R-L circuit too). Denote by $P(0, 0 \mid z, x; t)dzdx$ the conditional probability that if $x(0) = 0$ and $z(0) = 0$ we shall have at time t, $z < z(t) < z + dz$ and $x < x(t) < x + dx$.

It can then be shown (by a heuristic argument) that P must be that fundamental solution of

$$(3) \qquad \frac{\partial P}{\partial t} = D\frac{\partial^2 P}{\partial x^2} + \beta\frac{\partial(xP)}{\partial x} + (\gamma z - V(x))\frac{\partial P}{\partial z},$$

which becomes $\delta(x)\delta(z)$ as $t \to 0$. The constants D and β are those in the differential equation

$$(4) \qquad \frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2} + \beta\frac{\partial(xp)}{\partial x}$$

satisfied by the transition probabilities $p(x_1 \mid x; t)$ of the IF output $x(t)$.

Equations analogous to (3) (but in more variables) can be derived for the

[1] The results discussed above were mostly obtained in collaboration with A. J. F. Siegert and will form a part of a joint article, now under preparation for the Review of Modern Physics.

more general case when both IF and the linear filter are circuits with lumped elements.

A derivation of (3) can be based on the observation that $\{z(t), x(t)\}$ form a two-dimensional Markoffian process and that consequently P satisfies an appropriate Smoluchowski (Chapman-Kolmogoroff) integral equation. Equation (3) is then the Fokker-Planck equation which can be derived from the integral equation in a known way (Kolmogoroff, Feller). This derivation presupposes certain differentiability conditions on P which in our case of *explicitly* defined processes must be justified. This is one of the difficulties in rigorizing the procedure.

In the simplest case when $x(t)$ ($x(0) = 0$) is a Wiener process

$$z(t) = \int_0^t V(x(\tau))\, d\tau, \tag{5}$$

$V(x) > 0$ and satisfies some other mild conditions, it can be shown rigorously that if $\sigma(\alpha; t)$ is the distribution function of $z(t)$, then

$$\int_0^\infty e^{-st}\left\{\int_0^\infty e^{-n\alpha}\, d_\alpha \sigma(\alpha; t)\right\} dt = \int_{-\infty}^\infty \psi(x)\, dx, \tag{6}$$

where $\psi(x)$ is the Green's function of the differential equation

$$(1/2)\psi'' - (s + uV(x))\psi = 0, \tag{7}$$

satisfying the conditions $\psi(x) \to 0$, as $x \to \pm\infty$, $|\psi'(x)| < A$, $\psi'(-0) - \psi'(+0) = 2$. Equation (7) is obtained by Laplace transforming (with respect to z and t) the equation

$$\frac{\partial P}{\partial t} = \frac{1}{2}\frac{\partial^2 P}{\partial x^2} - V(x)\frac{\partial P}{\partial z} \tag{8}$$

which is a special case ($\beta = 0, \gamma = 0, D = 1/2$) of (3). Although (7), with $x(t)$ a Wiener process, does not correspond to any realistic noise problem, it is of some theoretical interest. Various limit theorems (including the P. Lévy arc sin law) can, for instance, be derived in this way.

The analogy between (7) and Schroedinger's equation is noteworthy.

CORNELL UNIVERSITY,
ITHACA, N. Y.

ENTROPY AND INFORMATION

BY

NORBERT WIENER

It has already been ascertained by Shannon and Wiener that amount of information is a quantity corresponding to negative entropy. Wiener has already suggested in his book *Cybernetics* that this is relevant to the action of a possible Maxwell demon. The Maxwell demon is able to lower the entropy of the mechanical output compared with the mechanical input because the demon itself possesses a negative entropy of information. This negative entropy must give information of the momentum and position of particles approaching the gateway operated by the demon in such a way that when the particle collides with the gateway, the gateway is either open or locked shut, so that no absorption of energy appears. This involves a mode of transmission more rapid than the motion of the particles to be separated, and a mode which is probably light. This light itself has a negative entropy which decreases whenever the frequency of the light is lowered by a Compton effect or similar cause.

Thus the Maxwell demon gives at least one way for comparing entropy of light with mechanical entropy. Since the information of the Maxwell demon includes both position and momentum it can be only imperfect information in quantum theory. This whole point of view suggests that one place to look for Maxwell demons may be in the phenomena of photosynthesis. It also suggests that absorption and emission of light may be an essential part of enzyme action in other forms of catalysis. Since there has been at least a suggestion that enzyme action and similar biological catalytic phenomena operate across an inert membrane, this light may give at least a partial explanation of the phenomena observed. At present the absence of an adequate entropic theory of radiation and in general of an adequate quantum theory of radiation able to take care of the particle-like as well as the wave-like aspects of light make a more precise verification of the opinions suggested here an impossibility.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
CAMBRIDGE, MASS.

THE STATISTICAL THEORY OF MESSAGE TRANSMISSION[1]

BY

Y. W. LEE

This paper presents a heuristic treatment of the statistical transmission theory of N. Wiener in a general form for application to physical problems. In contrast with present-day communication engineering techniques where a message is inadequately considered to be either a steady-state or a transient phenomenon, it is here represented by a stationary time series. The transmission problem treated in this paper is expressed as the design of a linear system to operate in cascade with an existing system for the extraction of a desired message from an additive mixture of message and disturbance with the least mean-square error. The desired message may be the original message, or its derivative or its integral, with a lead (prediction) or a lag in each case. It may take still other forms provided that it is coherent with the input. The formulation of the problem leads to an integral equation of the Wiener-Hopf type the solution of which is effected by a factorization of the input spectrum. Expressions for the irremovable error caused by the overlapping of message and disturbance spectra, and for the removable error which depends upon the amount of lag allowed, are derived. This theory shows that the basic data for design are the autocorrelation function of the input and the crosscorrelation function of the input and the desired output.

MASSACHUSETTS INSTITUTE OF TECHOLOGY,
CAMBRIDGE, MASS.

[1] To be published as a technical report by the Research Laboratory of Electronics, Massachusetts Institute of Technology.

TRANSIENT RESPONSE AND THE CENTRAL LIMIT THEOREM OF PROBABILITY

BY

HENRY WALLMAN

In many recent communication methods, such as radar, television, and pulse-modulation, and in instrumentation for particle-physics, it is essential to transmit discontinuous signals, typically step-functions of time, with as much fidelity as possible. Often this requires that the step-function response be a monotonic time-function, or, in the language of engineering, be "free of overshoot". For technical reasons it is usually necessary for the original step-function signal to pass through a considerable number of electrical networks and amplifiers, each of which may be required to be free of overshoot. It is consequently a matter of considerable interest to know the limiting behavior of the step-function response of cascades of networks each free of overshoot.

It is pointed out in this paper that the problem is solved by suitable interpretation of Laplace's central limit theorem of probability. Suppose one sets up the following "dictionary" between the languages of engineering and probability:

Engineering	Probability
E_1. Step-function response, free of overshoot	P_1. Distribution function
E_2. System function	P_2. Characteristic function
E_3. Time	P_3. Independent variable of distribution function
E_4. Frequency	P_4. Independent variable of characteristic function

(The key to this dictionary is the fact that just as E_2 is the Laplace transform of E_1 so is P_2 the Laplace transform of P_1.)

Then the central limit theorem yields the result that "The step-function response of cascaded networks, each free of overshoot, tends to the error-function integral". This has important practical consequences, among them the fact that if a network is free of overshoot its time-of-response inevitably increases rapidly upon cascading, namely as the square-root of the number of cascaded networks.

The underlying mathematical situation is the following: both the central limit theorem of probability and the theorem on cascaded step-function responses are consequences of the same theorem on iterated convolutions of bounded monotonic functions.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
CAMBRIDGE, MASS.