Cosmology and action-at-a-distance electrodynamics

F. Hoyle

102 Admirals Walk, West Cliff Road, West Cliff, Bournemouth, Dorset BH2 5HF, United Kingdom

J.V. Narlikar

Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411 007, India

This article reviews the developments in the electrodynamics of direct interparticle action, emphasizing the achievements in quantum as well as classical electrodynamics. It is shown that the application of the Wheeler-Feynman absorber theory of radiation places stringent requirements on the asymptotic future and past light cones of the universe. All Friedmann cosmologies fail to meet these requirements, but the steady-state and the quasi-steady-state models have the right kind of structure to make the theory work. Further, it is shown that the working theory is free from the problems of divergence that trouble the classical and quantum field theory. In particular, no renormalization is needed: The bare mass and bare charge of an electron are finite. A few ideas relating to the response of the universe to a local microscopic experiment are presented as well as on possible clues to the outstanding issues of foundations of quantum theory.

CONTENTS

I. HISTORICAL BACKGROUND

A. From Newton to Gauss

The foundations of theoretical physics were laid by Isaac Newton's book *Philosophiae Naturalis Principia Mathematica* published in the mid-1680s. The laws of gravitation and dynamics described therein successfully demonstrated how to explain the various dynamical phenomena ranging from the motions of terrestrial projectiles to the orbits of planets. They also established an important principle: that with suitable initial conditions the subsequent behavior of a dynamical system can be completely determined provided the forces acting on it are known. Until the advent of quantum mechanics in the early part of this century, this deterministic view prevailed.

The next addition to fundamental physics came a century later with the discovery of the electrical force. The law of electrical attraction and/or repulsion between unlike and/or like electrical charges as stated by Coulomb was strikingly similar to the inverse square law of gravitation. For a comparison we state Newton's and Coulomb's laws in familiar notation:

 0034-6861/95/67(1)/113(43)/$13.45

$$F_N = -\frac{Gm_1m_2}{r^2}, \tag{1.1}$$

$$F_C = \frac{Ke_1e_2}{r^2}. \tag{1.2}$$

[The constant K can be taken as unity by a suitable choice of units as we shall do hereafter.]

It is possible that Coulomb may have been inspired to think in terms of an inverse square law because of the successes of the law of gravitation. However, the experiments in electrostatics clearly pointed to such a law. Also, in spite of their superficial similarity there was one fundamental difference between the two laws, a difference that led to their subsequent development along different routes. In gravitation there is always attraction whereas in electrostatics the presence of positive and negative charges allows both repulsion and attraction to be present. [Note also that for like charges the rule is of *repulsion* as opposed to *attraction* in gravitation.]

The commonality between the two laws, however, extends beyond the functional (inverse square) form to a deeper level in that they both assume *instantaneous* action at a distance. So far as gravitation was concerned there was no apparent conflict with any observation because of this assumption. In electrodynamics the situation turned out to be different. It became clear as a result of several experiments on rapidly moving charges that the Coulomb law was not sufficient to describe all the observed details. On March 19, 1845 Gauss in a letter to Weber summarized the difficulty in these words (Gauss, 1867):

> *... I would doubtless have published my researches long since were it not that at the time I gave them up I had failed to find what I regarded as the keystone, Nil actum reputans si quid superesset agendum: namely, the derivation of the additional forces—to be added to the interaction of electrical charges at rest, when they are both in motion—from an action which is propagated not instantaneously but in time as is the case with light...*

Thus, in a sense Gauss had anticipated the future work of Maxwell but did not get down to the actual description of delayed action at a distance with the speed of light playing the key role. In the postspecial relativity era one could express the above requirement that *the action at a distance should be a relativistically invariant concept.* Evidently, with its effect traveling at infinite speed the Newton-Coulomb action at a distance was not consistent with relativity.

B. The formula for delayed action

The problem posed by Gauss was partially solved in the early part of this century by Schwarzschild (1903), Tetrode (1922), and Fokker (1929a, 1929b, 1932). We restate below the Fokker formula for delayed action at a distance in a notation that will be useful for describing the subsequent developments.

We will use the four-dimensional spacetime notation that became common after special relativity. Thus ($i = 0, 1, 2, 3$) will denote the four spacetime coordinates with $x^\circ = ct$ the timelike coordinate and x^μ ($\mu = 1, 2, 3$) the three spacelike ones. Here c is the speed of light which occasionally will be set equal to unity to simplify writing. The same will apply to the Planck symbol $\hbar$ which will be needed in our discussions of quantum electrodynamics. In general the Latin indices shall take four values 0,1,2,3; while the Greek indices will take three values 1,2,3. The summation convention shall be assumed. In special relativity the line element is given by

$$ds^2 = \eta_{ik}dx^i dx^k, \tag{1.3}$$

where the metric tensor $\eta_{ik} = \mathrm{diag}(1, -1, -1, -1)$. In general relativity, the metric tensor will be denoted by g_{ik}. The line element will continue to have the signature of Eq. (1.3) even in the latter case where the metric tensor may not be diagonal.

We will also need the Dirac delta function $\delta(x)$ which has the properties

$$\delta(x) = 0, \qquad \text{for } x \neq 0; \qquad \int_{-\infty}^{\infty} \delta(x)dx = 1. \tag{1.4}$$

This satisfies the identity

$$\eta^{ik}\delta(s^2_{XA})_{,ik} = \Box_X \delta(s^2_{XA}) = -4\pi\delta_4(X, A) \tag{1.5}$$

where δ_4 is the four-dimensional delta function for spacetime points $X \equiv (x^i)$ and $A \equiv (a^i)$ and s^2_{XA} is the square of the interval between them as computed by (1.3). [A suffix i following the comma denotes differentiation with respect to the coordinate x^i.]

In Eq. (1.5) $\Box$ is the wave operator and $\delta(s^2_{XA})$ is its *Green's function.* This identity is valid in the flat spacetime of special relativity and needs to be generalized to the curved spacetime of general relativity which we shall introduce in the following section. For the present we will work within the framework of special relativity.

Having stated our notation we now write the Fokker action formula which describes the interaction between electric charges labeled $a, b, c, \ldots$, etc. as follows:

$$J = -\sum_a \int m_a da - \sum_a \sum_{<b} \int\int e_a e_b \delta(s^2_{AB})\eta_{ik}da^i db^k. \tag{1.6}$$

In the above expression m_a is the mass of particle a and e_a its electric charge. da is the element of proper time of particle a. The first term is the usual inertial term while the second term is the electrodynamic interaction term. In the latter, the delta function ensures that the typical points A and B on the worldlines of a and b interact if and only if they are connectible by a null ray. This is

another way of saying that the interaction between A and B propagates with the speed of light. Thus conceptually at least the program envisioned by Gauss seems to have been achieved. [See Fig. 1.1.] But how does it work in practice?

C. The problems of action-at-a-distance electrodynamics

The formula (1.6) looks quite different from the field theory action which is usually stated in the form

$$J = -\sum_a \int m_a da - \frac{1}{16\pi}\int F_{ik}F^{ik}d^4x - \sum_a \int e_a A_i da^i. \tag{1.7}$$

In the above the particles $a, b, c, \ldots$ are not interacting directly with one another; they do so through the medium of a field F_{ik} which is defined in terms of a four-potential A_i by

$$F_{ik} = A_{k,i} - A_{i,k}. \tag{1.8}$$

The field has its own uncountably infinite degrees of freedom which are called into play in describing phenomena like radiation. What is the corresponding picture in the action at a distance defined by the Fokker formula?

To see the correspondence with the field picture the following definitions of *direct particle potentials* and *direct particle fields* are useful:

$$A_i^{(b)}(X) = e_b \int \delta(s_{XB}^2)\eta_{ik}db^k,$$

$$F_{ik}^{(b)} = A_{k,i}^{(b)} - A_{i,k}^{(b)}. \tag{1.9}$$

Thus we have a field and a potential associated with each particle and these identically satisfy the following relations:

$$A^{(b)i}_{\ ;i} = 0, \tag{1.10}$$

$$\Box A_i^{(b)}(X) = -F_i^{(b)\ k}{}_{,k} = 4\pi J_i^{(b)}(X), \tag{1.11}$$

$$J_i^{(b)}(X) = e_b \int \delta_4(X,B)\eta_{ik}db^k. \tag{1.12}$$

Superficially these look similar to the gauge condition, the field equations, and the wave equation of the Maxwell field theory. However, these are identities in view of the definitions (1.9). In fact these "fields" do not have degrees of freedom of their own: they are functionals of particle paths. For this reason it is misleading to call them fields. We shall refer to them as *direct particle fields.* In terms of these direct particle fields the variation of the worldline of a typical particle a gives us the analog of the Maxwell-Lorentz equations of motion:

$$m_a \frac{d^2a^i}{da^2} = e_a \sum_{b\neq a} F^i{}_k{}^{(b)} \frac{da^k}{da}. \tag{1.13}$$

Notice that the particle a is acted on by all *other* particles $b \neq a$, i.e., there is no self-action. This absence of self-action was in fact evident from the Fokker formula which has in the second term the summation excluding self-action.

This formulation therefore satisfies the requirement of relativistic invariance and seems to resemble the Maxwellian field theory which is already known as a successful theory of electrodynamics. There are, however, several questions that this formulation has to answer before it can be accepted as a working theory. We list them below.

(1) The complete time symmetry of the formulation tells us that the electromagnetic interaction proceeds not only forward in time but also—in equal strength, it proceeds backwards in time. Figure 1 illustrates this result. Thus there is a manifest violation of causality. How can such a theory explain causal and unidirectional phenomena like radiation?

(2) With no degrees of freedom vested in direct particle fields, will the theory be able to account for all electrodynamic observations?

(3) How is the theory described in curved spacetime? How does it interact with spacetime geometry? This question assumes significance when we recall that the electromagnetic energy momentum tensor in Einstein's field equations depends entirely on *fields* in Maxwell's theory and that there is no corresponding field term in the present theory.

(4) The bulk of the effects of electrodynamics fall within the quantum domain. Can the action-at-a-distance formulation be quantized? Recall again that in the usual formulation it is the field that is quantized and here we have no field.

(5) Finally, at a deeper level, we may ask whether this

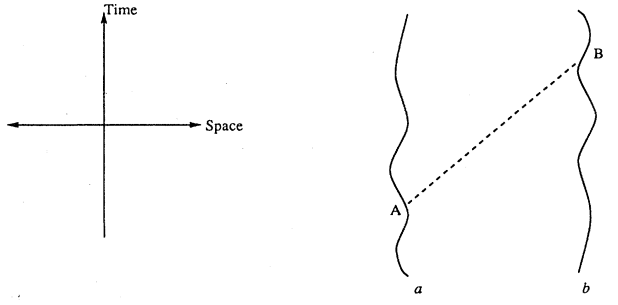

FIG. 1. Typical points A (on the worldline of charge a) and B (on the worldline of charge b) interact if the dotted line connecting them is a null ray. However, the interaction can be both forward (A to B) and backward (B to A) in time.

new formulation fares better than the standard field theory.

These challenges have been addressed by various workers over a span of several decades. In this review we will summarize the progress in light of the above questions. We begin with the seminal work of Wheeler and Feynman first reported in this journal nearly fifty years ago.

II. THE ABSORBER THEORY OF RADIATION

A. The problems of classical field theory

The issues raised above *vis a vis* action at a distance can be better appreciated against the background of the problems faced by the classical field theory of Maxwell. We itemize them below although they happen to be interrelated.

1. Explanation of causality

The wave equation satisfied by the four-potential A_i in the Maxwell theory is similar to the relation (1.11) except that in this case it is a genuine equation rather than an identity:

$$\Box A_i = 4\pi J_i. \tag{2.1}$$

Here the right-hand side is the current density four-vector. In terms of our direct particle definition (1.12) it is the sum of all such four vectors.

In solving any problem in field theory involving the above equation, it is common practice to choose those solutions of Eq. (2.1) that are consistent with the principle of causality. The most fundamental problem is the one referred to by Gauss (1867), viz. that of the accelerated electric charge. It is well known that the wave equation (2.1) has two independent basic solutions, one having support on the future light cone (the so-called *retarded solution*) and the other having support on the past light cone (the *advanced solution*). Symbolically we will denote these solutions by $A_i^{(\mathrm{ret})}$ and $A_i^{(\mathrm{adv})}$, respectively, for the potentials and by $F_{ik}^{(\mathrm{ret})}$ and $F_{ik}^{(\mathrm{adv})}$ for the corresponding fields.

Now in the problem of the accelerated charge, it is customary to select the retarded solution to describe the physical situation. The advanced solution is rejected on the grounds of causality. Thus it is argued that it is physically realistic to have the charge radiating electromagnetic waves which travel outwards from it and reach a distant point at a *later* instant; and the retarded solution describes this situation. The advanced solution describing waves *converging from infinity* onto the source charge and crossing a distant point *before* they reach the source is manifestly unrealistic. Hence the retarded solution is the reasonable one.

While this procedure is entirely consistent with physical reality, at a deeper level it is incomplete; for it does not take us any further towards understanding why the principle of causality should operate. Expressed in a somewhat different form, the phenomenon of radiation by the accelerated electric charge is a unidirectional one in terms of time whereas the basic Maxwell equations are time symmetric. The question therefore is, *why do we have an electrodynamic arrow of time?* Field theory does not offer any answer. It stops at providing a scenario consistent with causality. The choice of the retarded solution is imposed *ad hoc* rather than deduced.

2. Radiation damping

As a result of the choice of the retarded solution and the phenomenon of radiation by the accelerated charge, the charge loses energy and its motion is damped. It is possible to compute the damping force on the charge by using the law of conservation of energy and momentum. In the notation of the preceding section, the equation of motion of a typical charge a is modified from the Maxwell-Lorentz form to the following:

$$m_a \frac{d^2 a^i}{da^2} = e_a F^i{}_k \frac{da^k}{da} + \frac{4e_a}{3} g_{lk} \left(\frac{d^3 a^i}{da^3} \frac{da^l}{da} - \frac{d^3 a^l}{da^3} \frac{da^i}{da} \right) \frac{da^k}{da} \quad , \quad c = 1. \tag{2.2}$$

The $F^i{}_k$ term here denotes the *external field* acting on the charge. The extra term on the right-hand side is the damping force. Notice that it has not been deduced from the basic field theory action whose Lagrangian only gives the Lorentz force. It has been put in from the requirement of energy loss by radiation. For example, if we had chosen a time-symmetric solution, i.e., a solution with half the advanced plus half the retarded fields then there would be no emission of radiation and no damping.

In a highly perceptive discussion of the problem Dirac (1938b) had provided a new modus operandi for the computation of the force of radiative damping. His prescription was as follows. To the field $F^i{}_k$ used in computing the Lorentz force in Eq. (2.2) add an extra field

$$R^{(a)i}{}_k \equiv \frac{1}{2} \left\{ F^{(a)\ \mathrm{ret}\ i}{}_k - F^{(a)\ \mathrm{adv}\ i}{}_k \right\}. \tag{2.3}$$

Here $R^{(a)i}{}_k$ is evaluated at the electric charge a. Although both the advanced and retarded fields due to the motion of a diverge on the worldline of a, their difference is finite and as shown by Dirac, its force on the charge is exactly equal to the extra term in Eq. (2.2). Thus the

motion of an electric charge a is given by the modified equations:

$$m_a \frac{d^2 a^i}{da^2} = e_a \left\{ F^i{}_k + R^{(a)i}{}_k \right\} \frac{da^k}{da} \tag{2.4}$$

with the second term apparently arising from the charge itself.

The Dirac prescription despite its elegance was somewhat mystifying, however, in that it brought in the advanced solution that had been discarded as unphysical. Dirac sought to relate its presence to another outstanding problem of field theory, namely the problem of infinite self-action. We will consider it next.

3. The paradox of self-action

Dirac (1938b) highlighted the problem with the help of an idealized situation. Imagine an electric charge a at rest and under the action of no forces until it is hit by a hammer. The hit is thus an impulsive force which sets the charge in motion. What happens to the charge thereafter when it finds itself once again under no external forces?

There are two possible solutions for describing the motion of the charge. The first solution has the charge moving with a uniform velocity that it acquired as a result of the hit. The second solution is more peculiar and describes the charge moving with a momentum that increases exponentially with time, and according to the full relativistic treatment given by Dirac its velocity rapidly approaches the speed of light. This happens because of the self-action force introduced in Eq. (2.4).

Although the first solution appears reasonable, it is the second that matches the prescribed initial conditions. Under the circumstances Dirac reexamined the initial conditions and argued that they need to be altered if the first solution is to apply. The new situation has the charge moving from rest at infinite past and attaining the final velocity just before being hit; a velocity it maintains thereafter.

The crucial mathematical point to appreciate here is that, with the self-action included, the differential equation of motion is of third rather than second order. Thus, after an impulsive force the acceleration rather than the velocity changes discontinuously. Physically, however, the new situation seems acausal, for the charge accelerates *in anticipation* of the hit in such a way that it builds up the right velocity just before being hit by the hammer.

This acausal behavior of the charge can be rationalized by pointing out that the self-action force as computed by Dirac's method does include the advanced field. In practical terms the duration of acausality is of the order $e_a^2/m_a c^3$ which is not only very small but also is small by the factor 1/137 compared to the Compton time scale associated with the charge (assuming that it is an electron or a proton).

The above discussion (see also Hoyle and Narlikar, 1993 for details) offers us a choice between two unphysical solutions. In one we have causality but infinite self-energy while in the other the motions are finite but acausal. Not surprisingly, it was believed that the problem of self-force of the charge would not be solved except by recourse to quantum theory.

This hope has not been fully realized. Quantum field theory does alleviate the self-energy problem but cannot surmount it without introducing the renormalization program. We shall consider the quantum problem in Sec. IV. For the present we will confine ourselves to the classical electrodynamics.

These comments therefore underscore the fact that there are conceptual problems with the classical field theory, and thus provide further motivation for looking at the alternative offered by action at a distance.

B. The Wheeler-Feynman approach

Fifty years ago, Wheeler and Feynman (1945) addressed the above issues in an attempt to revive the action-at-a-distance formulation as derived by Schwarzschild, Tetrode, and Fokker (see references in the previous section). The central themes of their argument were that an action-at-a-distance theory was necessarily nonlocal and that the apparent acausality in its results arose from inadequate attention being paid to the interaction of a typical charge a with all the other charges in the universe, even if they happen to be located far away.

1. A simple illustrative example

To illustrate how the distant charges influence a local experiment we will repeat briefly the simple derivation given by Wheeler and Feynman in their above-mentioned paper.

We assume the universe to be static, Euclidean, with a uniform number density of charges e and with the line element of special relativity as given in Eq. (1.3). Let the charge a be located near the origin O of spherical polar coordinates (r, θ, ϕ) and suppose that its motion there is Fourier analyzed with a typical component of the acceleration given by

$$\mathbf{u}_o e^{-i\omega t}. \tag{2.5}$$

To simplify the picture further, Wheeler and Feynman assumed the local region around the charge a to be empty, in the form of a spherical cavity centered at $r = 0$, and extending as far as $r = R$ and the universe beyond having N charges per unit volume.

In vacuum the full retarded electric field of the charge a at a point P located at a large distance r from it would be given by

$$E_\theta = u_o \frac{e}{r} \sin\theta \exp\left[i\omega(r - t)\right] \tag{2.6}$$

in the direction of increasing θ where θ is the angle made by the direction OP with that of the acceleration vector $\mathbf{u}_o$. We have taken $c = 1$.

This result, however, needs to be modified to include the refraction effect at the boundary $r = R$ of the cavity and the phase change due to the refractive index $n - ik$ of the medium beyond. The latter is related to the effect the basic field produces on the motion of a typical charge at P. Thus, we modify (2.6) to

$$E_\theta = \frac{2eu_o \sin\theta}{r(1+n-ik)} \exp\left[i\omega\{r - t + (n-ik-1)(r-R)\}\right], \tag{2.7}$$

and use the field E_θ to compute the acceleration of the charge at P. This is given in the direction of E_θ by

$$\frac{e}{m} p(\omega) E_\theta \tag{2.8}$$

where $p(\omega)$ is a frequency-dependent function, determined in terms of the refractive index by the formula

$$(n-ik)^2 = 1 - \frac{4\pi N e^2}{m\omega^2} p(\omega). \tag{2.9}$$

The crucial step in the Wheeler-Feynman theory was to recognize that in the action-at-a-distance formulation the motion of the particle at P will generate a reaction which will arrive at a backwards in time, i.e., at the instant that the original retarded field left it. This reaction is the half advanced field of the particle at P. Further, to study the electrodynamics in the vicinity of a we must evaluate such responses from all particles lying on the future light cone of a.

The half advanced electric field at a due to the source acceleration at P as given by Eq. (2.6) when resolved in the direction of the acceleration of a then becomes

$$E_\theta \frac{e}{m} p(\omega) \left[\frac{e}{2r} \sin\theta\right] \exp(-i\omega r). \tag{2.10}$$

The net response of all such particles along the future light cone of a is given by the integral

$$\begin{aligned} R &= \int_{r=R}^{\infty} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \frac{e^2}{2mr} p(\omega) \sin\theta\, e^{-i\omega r} E_\theta N r^2 \\ &\quad \times \sin\theta\, dr d\theta d\phi \\ &= -\frac{2}{3} i\omega e u_o e^{-iwt}. \end{aligned} \tag{2.11}$$

The responses normal to the acceleration vector cancel out and so we may use Eq. (2.11) to sum over all frequencies and arrive at the result

$$\mathbf{R} = \frac{2e}{3} \dddot{\mathbf{a}}. \tag{2.12}$$

This is the field that the charge a itself would experience because of its action at a distance with the rest of the universe. Multiplying this field by the charge gives us the standard formula for the radiative damping force

$$\mathbf{R}e = \frac{2e^2}{3} \dddot{\mathbf{a}}. \tag{2.13}$$

It can be verified that this is the nonrelativistic version of the Dirac term in Eq. (2.2).

If instead of calculating the sum of responses at the location of a we had calculated it at an arbitrary point in its neighboring region, we would have found that the field is Dirac's extra field (2.3),

$$\frac{1}{2}\left\{F^{(a)\text{ret}} - F^{(a)\text{adv}}\right\}. \tag{2.14}$$

This calculation is slightly more involved and may be found in the work of Wheeler and Feynman (1945). Using this result, they built up a self-consistent picture of action at a distance in the following way.

In the above calculation the net field emanating from charge a is the full retarded field. How is it made up? It is made up of two components as given below:

$$\begin{aligned} F^{(a)\text{ret}} &= \frac{1}{2}\left\{F^{(a)\text{ret}} + F^{(a)\text{adv}}\right\} \\ &\quad + \frac{1}{2}\left\{F^{(a)\text{ret}} - F^{(a)\text{adv}}\right\}. \end{aligned} \tag{2.15}$$

The first term on the right-hand side is the basic time-symmetric field of charge a while the second term, as we just saw, represents the response of the universe. The calculation is thus self-consistent since it was the full retarded field that was used in computing the response.

We therefore see that Dirac's mysterious prescription receives a natural derivation in the action at a distance framework. We also see that the radiative reaction is not a self-force but is the combined reaction of the universe to the motion of the charge a. Further, even though our theory is time symmetric, we seem to have arrived at an explanation of why retarded solutions operate in practice: it is not an *ad hoc* choice required by causality but forced on us by the way the universe responds.

It might be argued that the elegant result obtained above may be due to our oversimplified choice of parameters describing the universe. The universe is not homogeneous. It may consist of charged particles of various masses (e.g., electrons and protons). The cavity imagined around the charge a may not be spherical. Is the result sensitive to these issues?

Wheeler and Feynman demonstrated that these issues are not important. The crucial issue is that of *complete absorption.* The integral in Eq. (2.11) must then converge to the value it has in Eq. (2.11). This is ensured by the presence of a sufficient number of particles to the future of a that can absorb the disturbance coming out from a and react to it. The condition may be stated thus: *the universe must be a perfect absorber of all electromagnetic fields emanating from within.* The self-consistency

argument implied by Eq. (2.15) will not work if the universe is an imperfect absorber. We will now state this requirement mathematically and use it to give a general derivation of the above result.

2. The general result

Let us consider the universe as static and with Euclidean geometry. The electric charges in it are moving arbitrarily and we will denote by $F^{(a)\text{ret}}$ and $F^{(a)\text{adv}}$ the retarded and advanced fields of a typical charge a. We remind the reader that the fields referred to here are direct particle fields and hence do not have extra degrees of freedom of their own. Thus the retarded and/or advanced field implied here is well defined with respect to the light cones future and/or past of the corresponding particle.

We now state the property of perfect absorption as implied by Wheeler and Feynman (1945) as follows: When any arbitrary electric charge a is accelerated, all electromagnetic fields arising from its motion—directly or through its interaction with other charges—should tend to zero sufficiently rapidly at great distances from a. If we confine our attention to only such fields, then the above condition means

$$\sum_b \frac{1}{2}\left[F^{(b)\text{ret}} + F^{(b)\text{adv}}\right] \sim o\left(\frac{1}{r}\right) \text{ as } r \to \infty. \quad (2.16)$$

In vacuum, a radiative field falls asymptotically as r^{-1} and the more rapid fall implied by Eq. (2.16) indicates perfect absorption. The proof given by Wheeler and Feynman that in a perfectly absorbing universe only retarded interactions survive is as follows.

Since in (2.16) we have a combination of incoming and outgoing waves, for the relation to hold at all times we need the two types of waves to vanish asymptotically *separately*. Thus Eq. (2.16) implies two relations:

$$\sum_b \frac{1}{2}F^{(b)\text{ret}} \sim o\left(\frac{1}{r}\right), \quad (2.17)$$

$$\sum_b \frac{1}{2}F^{(b)\text{adv}} \sim o\left(\frac{1}{r}\right) \quad \text{as } r \to \infty,$$

and hence also

$$\sum_b \frac{1}{2}\left[F^{(b)\text{ret}} - F^{(b)\text{adv}}\right] \sim o\left(\frac{1}{r}\right) \quad \text{as } r \to \infty. \quad (2.18)$$

However, unlike Eq. (2.16) the above combination represents a sourceless field and hence a solution of the homogeneous wave equation. As such, its faster than r^{-1} behavior at infinity implies that it must vanish identically everywhere. Hence

$$\sum_b \frac{1}{2}\left[F^{(b)\text{ret}} - F^{(b)\text{adv}}\right] \equiv 0. \quad (2.19)$$

The field acting on charge a therefore becomes

$$\sum_{b\neq a} \frac{1}{2}\left[F^{(b)\text{ret}} + F^{(b)\text{adv}}\right]$$

$$= \sum_{b\neq a} F^{(b)\text{ret}} + \frac{1}{2}\left[F^{(a)\text{ret}} - F^{(a)\text{adv}}\right]. \quad (2.20)$$

The first term on the right-hand side represents the retarded field of all other charges $b \neq a$ acting together on a while the second term is the Dirac radiative reaction.

We therefore arrive at the general version of the result derived in the simple example considered earlier, thus highlighting the role of perfect absorption by the universe. For this reason, Wheeler and Feynman called this theory the *absorber theory of radiation.*

3. Enter cosmology

The apparent resolution of the causality problem in action at a distance was, however, not quite complete in its logical framework as Wheeler and Feynman themselves pointed out. We can see the problem in the following way. In the above general argument interchange the words advanced and retarded to find that the chain of reasoning still goes through with (2.20) replaced by

$$\sum_{b\neq a} \frac{1}{2}\left[F^{(b)\text{ret}} + F^{(b)\text{adv}}\right]$$

$$= \sum_{b\neq a} F^{(b)\text{adv}} + \frac{1}{2}\left[F^{(a)\text{adv}} - F^{(a)\text{ret}}\right]. \quad (2.21)$$

This means that the charge a is acted on by the advanced fields of all other charges and a radiative reaction that is the exact opposite of that given by Dirac.

There is nothing to prevent us from using Eq. (2.21) instead of Eq. (2.20), but in practice it would be very awkward. For example, in the simple example discussed earlier, we had assumed that the absorber particle was at rest before being hit by the retarded wave from the source. This is reflected in the first term on the right-hand side of Eq. (2.20) which is uncorrelated with the motion of a. In Eq. (2.21), on the other hand, the first term is highly correlated with the motion of a and hence if we took those correlations into account we would recover Eq. (2.20).

On the other hand, we could use Eq. (2.21) to describe a new situation in which the universe admits only the advanced solutions. The simple example of Sec. II.B.1 would then have a counterpart in which the acceleration of a generates advanced, i.e., incoming waves which hit the absorber particles *before* reaching the source charge. A typical absorber particle must move in such a way that

after the incoming wave has hit it, it comes to rest.

Wheeler and Feynman argued that while *a priori* there is nothing to prevent us from imagining a universe with initial conditions set up in the above fashion, in the thermodynamic context such artificial initial conditions would appear highly unlikely. In fact, this distinction between Eqs. (2.20) and (2.21) was, according to them, dictated by thermodynamic time asymmetry. The time asymmetry in electromagnetic radiation arises from asymmetrical initial conditions that favor Eq. (2.20) over Eq. (2.21), i.e., from the time asymmetry in thermodynamics.

It turns out, however, that this recourse to thermodynamics is unnecessary. The crucial consideration that breaks the time symmetry of the action at a distance theory comes from cosmology. This was first pointed out by Hogarth (1962) who argued that, if due note is taken of the cosmological fact that the universe is expanding, then the symmetry between the two situations leading to Eqs. (2.20) and (2.21) is broken. For, if we examine the proof of the general result of Wheeler and Feynman given above, we notice that to prove the consistency of retarded solutions we require perfect absorption in the future, and likewise we need perfect absorption in the past for the consistency of the advanced solutions.

The early observations of Hubble (1929) based on the redshifts of the nearby galaxies and clusters have since been extended to galaxies considerably farther away and the picture of the expanding universe has come to be generally accepted. Most cosmological models today are based on this concept. Thus the assumption of a static universe by Wheeler and Feynman was unrealistic. Hogarth's argument can be reworded in the following way to underscore the crucial role of cosmology in action-at-a-distance electrodynamics.

Suppose that we have a universe that has future and past absorbers operating at different efficiencies which we shall denote by factors f and p. Thus $f = 1$ denotes a perfect future absorber while $p = 1$ a perfect past absorber. Let such a universe lead to a net self-consistent solution of the form

$$F_{\text{total}} = AF^{(\text{ret})} + BF^{(\text{adv})} \tag{2.22}$$

where A and B are constants. Now, we have seen that a full retarded solution gives the Dirac radiative reaction in a perfectly absorbing universe. With an absorber of efficiency f the field $AF^{(\text{ret})}$ will therefore generate a radiative reaction Af times the Dirac value. Similarly, the field $BF^{(\text{adv})}$ will generate a reaction $-Bp$ times the Dirac value. For self-consistency therefore, the net radiative reaction $(Af - Bp)$ times the Dirac value added to the basic elementary field of the charge a should give us the net field assumed in Eq. (2.22):

$$F_{\text{total}} = \frac{1}{2}\left[F^{(\text{ret})} + F^{(\text{adv})}\right] + \frac{1}{2}(Af - Bp)\left[F^{(\text{ret})} - F^{(\text{adv})}\right]. \tag{2.23}$$

Equating the coefficients of the advanced and retarded fields in Eqs. (2.22) and (2.23) separately, we determine the coefficients A and B as

$$A = \frac{1-p}{2-f-p}, \qquad B = \frac{1-f}{2-f-p}. \tag{2.24}$$

Notice that if $f = 1$ we get the full retarded field as the self-consistent answer so long as $p \neq 1$. Similarly, for $p = 1$ and $f \neq 1$ we get the full advanced field as the self-consistent answer. Only for the case $p = f = 1$ do we run into an ambiguous situation. This last was precisely the case encountered by Wheeler and Feynman. Hogarth on the other hand showed that most familiar cosmological models lead to unambiguous results.

Before we can examine the cosmological implications in detail, however, we have to prepare the groundwork for describing action at a distance in curved spacetime, since cosmology uses that framework.

C. Cosmological considerations

1. Action at a distance in curved spacetime

As discussed above we will first develop the general framework for describing the action-at-a-distance electrodynamics in Riemannian spacetime and then apply it to some of the standard models of the universe. Such a framework was first given by Hoyle and Narlikar (1964a) in their attempts to follow up Hogarth's lead in a more comprehensive manner.

Thus instead of the Minkowski line element of Eq. (1.3) we have

$$ds^2 = g_{ik}dx^i dx^k \tag{2.25}$$

and the question arises, in what way can we generalize the $\delta(s^2_{AB})$ type of interaction to the above spacetime. Although the square of the interval s^2_{AB} between two world points A, B along the geodesic joining them (assuming it to exist and to be unique) is definable in a Riemannian spacetime, any operations of calculus on it are extremely intricate and do not lead us to Maxwell-like equations. The correct procedure lies in the generalization of the wave equation (1.5) to curved spacetime.

Synge (1960) had developed the necessary basic framework which was subsequently used by Dewitt and Brehme (1961) for defining the Green's functions of the wave equation in a Riemannian spacetime. In formulating action at a distance these Green's functions play the basic role of the above delta function. We will use the notation of Hoyle and Narlikar (1963) in what follows.

Accordingly we will rewrite the Fokker action (1.6) in curved spacetime in the following form:

$$S = -\sum_a \int m_a da - \sum_a \sum_{b<} 4\pi e_a e_b \times \int\int G_{i_A i_B} da^{i_A} db^{i_B}. \tag{2.26}$$

Here the first term is a straightforward generalization but the second one needs some explanation. The Green's function $G_{i_A i_B}$ takes the place of the flat spacetime term $\delta(s^2_{AB})\eta_{ik}$ and has the following properties:

$$G_{i_A i_B} = G_{i_B i_A}, \tag{2.27}$$

$$\Box_X G_{i_X i_B} + R_{i_X}{}^{l_X} G_{l_X i_B} = [-\bar{g}(X,B)]^{-1/2}\bar{g}_{i_X i_B}\delta_4(X,B). \tag{2.28}$$

Notice first that we have attached suffixes to the tensor indices to indicate the spacetime point at which they operate. This is necessary since the property of tensor covariance is a local one in curved spacetime. Instead of functions of one spacetime point common in field theory, here we are forced to use quantities that are invariant or covariant at two points where they are defined.

The relation (2.27) indicates the property of symmetry between the two points A, B at which $G_{i_A i_B}$ acts as a vector. It is this property that ensures time symmetry of the action defined above. The relation (2.28) is the identity satisfied by the Green's function. The two-point vector on the right-hand side is the so-called *parallel propagator* introduced by Synge (1960) to describe the parallel propagation of a vector along the geodesic joining two points.

For detailed properties of $G_{i_A i_B}$ see Dewitt and Brehme (1961) and Hoyle and Narlikar (1963, 1964a). For example, the Green's function has the following structure

$$G_{i_A i_B} = p_{i_A i_B}\delta(s^2_{AB}) + q_{i_A i_B}\theta(s^2_{AB}) \tag{2.29}$$

where $p_{i_A i_B}$ and $q_{i_A i_B}$ are two-point functions and θ is the Heaviside function. The latter part of the Green's function vanishes in flat spacetime. We therefore see a connection with the $\delta(s^2_{AB})$ term of the Fokker action. Here we briefly run through the formalism just to illustrate how the action at a distance is describable analogously to the flat spacetime version. Thus the direct particle potential and field are defined by

$$A^{(b)}_{i_X} = 4\pi e_b \int \bar{G}_{i_X i_B} db^{i_B}, \qquad F^{(b)}_{i_X k_X} = A^{(b)}_{k_X;i_X} - A^{(b)}_{i_X;k_X}, \tag{2.30}$$

and, in view of Eq. (2.28) the following relations also hold:

$$A^{(b)_i}{}_{;i} \equiv 0, \qquad F^{(b)_{ik}}{}_{;k} = -4\pi J^{(b)_i}, \tag{2.31}$$

where the current vector is defined as a straightforward curved space analog of Eq. (1.12). (Where there is no ambiguity the suffix on a tensor index is dropped.) The Lorentz force equation is likewise a generalization of Eq. (1.13) which need not be explicitly stated.

The action so formulated answers question (3) raised in Sec. I.C, even in relation to the effect of direct particle field on the spacetime geometry. For a variation of the metric tensor alters the spacetime in which the Green's function $G_{i_A i_B}$ is defined. As a result the Green's function also changes and hence the action. It can be shown that this variation leads to an energy momentum tensor defined in terms of direct particle fields that resembles the energy momentum tensor of the Maxwell field theory (Hoyle and Narlikar, 1964a).

Earlier Wheeler and Feynman (1949) had speculated whether the action-at-a-distance theory would produce such a gravitational effect. From purely flat spacetime arguments they had arrived at two possible forms of the energy tensor:

$$T^{ik}_{\text{Frenkel}} = \frac{1}{4\pi}\left\{\frac{1}{2}g^{ik}\sum_a\sum_{b\neq a} F^{(a)}_{lm}F^{(b)lm} - \sum_a\sum_{b\neq a}\left[F^{(a)il}F^{(b)k}{}_l + F^{(b)il}F^{(a)k}{}_l\right]\right\} \tag{2.32}$$

and

$$T^{ik}_{\text{canonical}} = \frac{1}{4\pi}\left\{\frac{1}{4}g^{ik}\sum_a\sum_{b\neq a}\left[F^{(a)\text{adv}}_{lm}F^{(b)\text{ret }lm} + F^{(b)\text{adv}}_{lm}F^{(a)\text{ret }lm}\right] - \sum_a\sum_{b\neq a}\left[F^{(a)\text{adv }il}F^{(b)\text{ret }k}{}_l + F^{(b)\text{adv }il}F^{(a)\text{ret }k}{}_l\right]\right\}. \tag{2.33}$$

The first one they called the *Frenkel tensor* and the second the *canonical tensor*. They had concluded

> *...From the standpoint of pure electrodynamics it is not possible to choose between the two tensors. The difference is of course significant for the general theory of relativity, where energy has associated with it a gravitational mass. So far we have not attempted to discriminate between the two possibilities by way of this higher standard...*

It was subsequently shown by Narlikar (1974) that it is the canonical tensor that arises from the above varia-

tional procedure.

We now leave these formal aspects of action at a distance in curved spacetime since we shall need them only marginally. Our aim has been to demonstrate that with the above framework it is legitimate to talk of an absorber theory in the curved spacetime of an expanding universe.

2. Cosmological models

For completeness it is now necessary to describe the cosmological models in which in the following subsection we discuss the absorber theory of radiation. Although there are several cosmological approaches we will restrict our attention to those normally described within the metric theories of gravitation. Again, we will limit our discussion to those aspects that we shall need for this article. The reader may refer to standard texts in cosmology for details. Here we will use the notation of Narlikar (1993a).

To begin with, we will consider only those models that require the universe to be homogeneous and isotropic on a large enough scale. Such models are described by the Robertson-Walker line element which in standard notation is

$$ds^2 = dt^2 - Q^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right]. \quad (2.34)$$

Here (r, θ, ϕ) are the comoving coordinates of a typical galaxy which ideally is presumed to be at rest in the above expanding cosmological frame. The time coordinate t is called the *cosmic time* and such a global time coordinate can be defined because of the large scale symmetry (homogeneity and isotropy) assumed for the spacelike sections t = constant. These symmetry arguments allow the spacelike sections to have three types of geometry, all with *constant curvature*. The parameter k denotes the type of curvature: thus for $k = 0$ we have flat Euclidean sections, for $k = +1$ we have closed sections while for $k = -1$ we have open hyperbolic sections. All these sections have an overall scale that varies with epoch.

The scale factor itself is given by $Q(t)$. Observations extending over several billion years along the past light cone indicate that the universe has been expanding, i.e., the function $Q(t)$ has been increasing with time over that timespan. Different cosmological models have, however, specified different functional forms for the scale factor. Thus in some cases the universe expands from a singular (pointlike) beginning, the so-called *big bang*, and either expands forever or contracts back to a pointlike singularity (the *big crunch*). There are also models which are nonsingular.

For our purpose we do not need the specific details of gravity theories that lead to these models. As we shall see in the next subsection, we need the geometrical quantities $Q(t)$ and k of the Robertson-Walker model and the way the density of matter $\rho(t)$ falls off at asymptotic past and future. Table I gives these details for some well-known models.

Despite differences in their geometrical details these models share certain common features which we now highlight. First, the redshift. If the source of light being observed now, at epoch t_o, shows the wavelength of a certain spectral line to be λ, then the wavelength of that line at the epoch t_e of emission was λ_e where

$$1 + z \equiv \frac{\lambda}{\lambda_e} = \frac{Q(t_o)}{Q(t_e)}. \quad (2.35)$$

The parameter z is called the redshift. For $z > 0$, the line has shifted towards the red end of the spectrum. This is

TABLE I. Some important cosmological models.

Model	k	$Q(t)$	$\rho(t)$	Reference
Einstein–de Sitter	0	$t^{2/3}$	$\propto Q^{-3}$	Einstein and de Sitter (1932)
Closed Friedmann	+1	$A(1-\cos\psi)$, A = constant $t = B(\psi - \sin\psi)$	$\propto Q^{-3}$	Friedmann (1922, 1924)
Open Friedmann	−1	$B(\cosh\psi - 1)$, B = constant $t = B(\sinh\psi - \psi)$	$\propto Q^{-3}$	Friedmann (1922, 1924)
Steady-state	0	$\exp Ht$	constant	Bondi and Gold (1948), Hoyle (1948)
Quasi-steady-state	0	$(1 + \alpha\cos\beta t)\exp At$ A, α, β constant	$(1 + \alpha\cos\beta t)^{-3}$	Hoyle, Burbidge, and Narlikar (1993)
Brans-Dicke	0	t^A, $A = (2\omega + 2)/(3\omega + 4)$	$\propto Q^{-3}$	Brans and Dicke (1961)
Dirac	0	$t^{1/3}$	$\propto Q^{-3}$	Dirac (1938a)

invariably the case and so we conclude that observations of all discrete extragalactic sources show that the length scale of the universe has increased since the time light left a typical source. Thus in an expanding universe, light traveling towards the future (as in a retarded solution) is *redshifted*, while that traveling towards the past (as in an advanced solution) is *blueshifted*.

A second feature of expanding models different from flat spacetime is the epoch dependence of density. As seen from Table I the density in general behaves as a function of the epoch and thus the past absorber is physically different from the future absorber. Hence, unless one carries out explicit computations one cannot decide how one absorber will respond based on the knowledge of how the other does.

Thus it is clear that when discussing the interaction with the future absorber we are dealing with low energy waves while in the case of the past absorber the interacting waves are of high energy. Likewise, except in the case of the steady-state and the quasi-steady-state models, and the closed Friedmann model, the future absorber has low density and the past absorber high density of absorbing matter. These issues will be relevant to our discussion of the absorber theory in these models.

3. Conformal transformations

A fortunate circumstance simplifies the discussion of the absorber theory in the above cosmological models. This arises because (a) these cosmological models are conformally flat and (b) the electrodynamic equations are conformally invariant. More specifically, a conformal transformation relates two metrics that are defined on the same spacetime manifold. We write

$$g_{ik} = \Omega^2 \bar{g}_{ik} \tag{2.36}$$

where the conformal function $\Omega(x^i)$ is a twice differentiable function with values in the range $0 < \Omega < \infty$.

If we can find a conformal function and a set of coordinates such that $\bar{g}_{ik} = \eta_{ik}$ then the spacetime described by the metric is said to be *conformally flat*. It was shown by Infeld and Schild (1945) that the Robertson-Walker model is conformally flat. The following series of transformations are needed to explicitly demonstrate this result:

$$k = 0: \quad \tau = \int_0^t \frac{du}{Q(u)}, \quad \rho = r.$$

$$k = +1: \quad T = \int_0^t \frac{du}{Q(u)}, \quad r = \sin R, \quad \xi = \frac{1}{2}(T + R),$$
$$\eta = \frac{1}{2}(T - R), \quad \tau = \frac{1}{2}(\tan \xi + \tan \eta),$$
$$\rho = \frac{1}{2}(\tan \xi - \tan \eta).$$

$k = -1$: Same as above with hyperbolic functions replacing trigonometric ones.

$$ds^2_{R-W} = \Omega^2[d\tau^2 - d\rho^2 - \rho^2(d\theta^2 + \sin^2\theta d\phi^2)]. \tag{2.37}$$

We will give below the explicit examples of the Einstein–de Sitter model and the steady-state model.

Einstein–de Sitter model:

$$Q(t) = \left(\frac{t}{t_0}\right)^{2/3}, \quad t_o = \text{constant},$$
$$\tau = 3t_o^{2/3}t^{1/3}, \quad \tau_o = 3t_o$$
$$\Omega(\tau) = \left(\frac{\tau}{\tau_o}\right)^2, \qquad 0 < \tau < \infty. \tag{2.38}$$

We may identify t_0, τ_0 as the time coordinates of the present epoch.

Steady-state model: Here we will assume, without loss of generality $t = 0$, $\tau = 0$ to denote the present epoch. We then have

$$Q(t) = e^{Ht}, \qquad H = \text{constant}, \qquad \tau = \frac{1}{H}\left\{1 - e^{-Ht}\right\}$$

$$\Omega(\tau) = \frac{1}{1 - H\tau} \quad , \qquad -\infty < \tau < H^{-1}. \tag{2.39}$$

In the latter case note that the time axis on the τ scale ends at $\tau = H^{-1}$. This happens because there is an event horizon to the future of any fundamental observer. This fact will turn out to have very significant implications for quantum electrodynamics.

D. Response of the expanding universe

With the inputs brought by cosmology it is now worth taking a second look at the absorber theory of radiation. There is, however, one subtle issue that Feynman [1963, but also see Mr. X in *The Nature of Time*, edited by T. Gold (Cornell, 1967)] had pointed out with regard to Hogarth's treatment of the problem that needs to be highlighted. For this, we go back to the general treatment of Sec. II.B.2.

As we pointed out, the condition for perfect absorption in the future demands [cf. Eq. (2.16)] that as $r \to \infty$

$$\sum_b \frac{1}{2}\left[F^{(b)\text{ret}} + F^{(b)\text{adv}}\right] \to 0 \quad \text{faster than} \quad 1/r.$$

In practice this is ensured by the absorptive part of the refractive index $n - ik$, i.e., by the parameter k. Hogarth had used the phenomenon of collisional damping to calculate k. Further, when he discussed the condition of perfect absorption in the past, he had used the same formula for k, but with its sign reversed. Feynman's criticism was that this sign reversal brought in thermodynamics that Hogarth was seeking to avoid, for the phenomenon of collisional damping is a collective phenomenon that assumes the second law of thermodynamics and asymmetrical initial conditions. Thus, the claim that cosmology and not

thermodynamics determined the unidirectionality of electromagnetic radiation was vitiated.

To get around Feynman's criticism, Hoyle and Narlikar (1963) proceeded in a different way: they used the radiation reaction on the charge to determine the damping parameter k. Their approach involved first choosing a particular combination of advanced and retarded solutions as the final solution and then testing it for self-consistency. Let us say that the pure retarded solution is to be so tested. Then, given that all charges interact finally through retarded waves, the radiation reaction is as given by Dirac [cf. Eq. (2.20)]. This reaction gives a force of damping that, in the nonrelativistic limit, leads to the following equation of motion for a typical absorber particle acted on by an external electric field $\mathbf{E}$:

$$m\ddot{\mathbf{r}} = e\mathbf{E} + \frac{2e^2}{3}\dddot{\mathbf{r}}. \tag{2.40}$$

Here e is the charge and m the mass of the particle.

If we Fourier analyze with ω the angular frequency of a typical field component, then it is easy to see that the refractive index $n - ik$ of the absorber medium is given by the equation

$$(n-ik)^2 = 1 - \frac{4\pi N e^2}{m\omega^2}\left\{1 + \frac{2ie^2\omega}{3m}\right\}^{-1}. \tag{2.41}$$

Notice that in deriving the above equation and using it for calculating the imaginary part of the refractive index we have not gone beyond electrodynamics; certainly not to thermodynamics. Further, if we were testing the self-consistency of the advanced solutions we would likewise use Eq. (2.40) with the sign of the radiative reaction term reversed. This would change Eq. (2.41) to

$$(n-ik)^2 = 1 - \frac{4\pi N e^2}{m\omega^2}\left\{1 - \frac{2ie^2\omega}{3m}\right\}^{-1}. \tag{2.42}$$

This reversal of sign has no relationship to thermodynamics (as was the case with Hogarth's use of collisional damping), but follows logically from electrodynamics. Further, the presence of the imaginary part in the refractive index arising from the radiative reaction term tells us that we are not dealing with a pure scattering phenomenon.

The next step in the argument is from cosmology. Because a typical wave from the source undergoes a spectral shift while traveling into the past or the future, we have to take into account its changed frequency at the time of its interaction with the absorber particle.

To study this effect we will consider two explicit examples from Table I. First consider the Einstein–de Sitter model whose geometrical details were given in Eq. (2.38). We rewrite its line element in the Robertson-Walker form as

$$ds^2 = dt^2 - \left(\frac{t}{t_o}\right)^{4/3}[dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)]. \tag{2.43}$$

In manifestly conformally flat form this is

$$ds^2 = \left(\frac{\tau}{\tau_o}\right)^4[d\tau^2 - dr^2 - r^2(d\theta^2 + \sin^2\theta d\phi^2)]. \tag{2.44}$$

Let us first test for the consistency of retarded solutions.

Suppose a general disturbance leaves the source at $r = 0$ at $t = t_0, \tau = \tau_o$, and travels along the future light cone. We consider a typical Fourier component of angular frequency ω of the electric field emanating from the accelerated source. At the source we have already adjusted the conformal factor of Eq. (2.44) to be unity. Thus the frequency ω measured on the t scale is the same as that on the τ scale. Since the electric field is conformally invariant it is convenient to work with the line element of Eq. (2.44) as we can take over the flat spacetime solution intact in these coordinates. However, as we found in the general treatment of Wheeler and Feynman, we need not go into specific details of the electric field but need only verify that it does indeed fall off faster than $1/r$ at large distances.

The flat spacetime expression tells us that without the interaction with the absorber, the field falls off as $1/r$. The absorber introduces a frequency-dependent factor

$$\xi = \exp\left(+\int k dr\right) = \exp(-I), \text{ say}, \tag{2.45}$$

of further damping. [Here $\omega(r)$ is the frequency of the wave at the radial coordinate r.] It is the asymptotic behavior of this factor that decides whether the future absorber is perfect or not. If the integral in the exponent of Eq. (2.45) diverges, the absorption is complete; otherwise it is incomplete.

Our task therefore is reduced to computing the asymptotic form of the parameter k in the refractive index. To calculate this first note that although in the flat spacetime solution the frequency of the field does not change, the τ scale does not measure the proper frequency in the cosmological rest frame. Thus ω is not the proper frequency with which the typical absorber particle at coordinate r interacts. The proper frequency is measured on the t scale and the value of the conformal factor at the absorber particle will determine it in terms of the constant frequency on the τ scale. The proper frequency is therefore given by

$$\omega(r) = \omega\left[1 + \frac{r}{\tau_o}\right]^{-2}. \tag{2.46}$$

A comparison with the Eq. (2.35) will tell us that the above is a restatement of the phenomenon of cosmological redshift. The absorber particle encounters a lower frequency than what was sent out by the source. The determination of k from the formula (2.41) requires us also to know the asymptotic behavior of the number density N. A reference to Table II tells us that, since the density

TABLE II. Consistency of advanced and/or retarded solutions.

Model	Future absorber	Past absorber	Outcome
Einstein–de Sitter	imperfect	perfect	advanced
Closed Friedmann	perfect	perfect	ambiguous
Open Friedmann	imperfect	perfect	advanced
Steady-state	perfect	imperfect	retarded
Quasi-steady-state	perfect	imperfect	retarded
Brans-Dicke	imperfect	perfect	advanced
Dirac	imperfect	perfect	advanced

ρ is proportional to N,

$$N(r) = N(0)\left[1 + \frac{r}{\tau_o}\right]^{-6}. \tag{2.47}$$

Using these formulas in Eq. (2.41) we find that the asymptotic value of k is given by

$$k \cong -\alpha\omega^2 \quad , \quad \alpha = \text{constant} \tag{2.48}$$

and the integral for absorption as given in Eq. (2.45) is

$$I \sim \int^{\infty} \alpha\omega^2 dr \sim \int^{\infty} \left(1 + \frac{r}{\tau_o}\right)^{-4} dr < \infty. \tag{2.49}$$

Note that this integral converges, thus indicating that the absorption is imperfect. *It follows therefore that in the Einstein–de Sitter cosmology, the retarded solution is not consistent.*

What about the advanced solution? We similarly consider the above formulas in the asymptotic limit of very large blueshifts. The situation at high energies is, however, not so clearcut. If we assume that interaction cross sections saturate as $\omega \to \infty$, then it can be shown [cf. Hoyle and Narlikar, 1963] that as $\omega \to \infty$,

$$1 - (n - ik)^2 = \frac{4\pi N e^2}{m\omega^2}\left[1 + O\left(\frac{1}{\omega}\right)\right]. \tag{2.50}$$

Using the $N \propto \omega^3$ dependence and keeping in mind the fact that τ is bounded below at $\tau = 0$, the integral for absorption in the past becomes

$$I \propto \int_0^{\tau_o} \frac{dr}{\tau_o - r} = +\infty, \tag{2.51}$$

i.e., it diverges. Thus here we have the past absorber perfect and the future absorber imperfect, a situation leading to the advanced solutions being self-consistent.

We therefore have a cosmology that does distinguish between the past and the future absorbers, the main point made by Hogarth. The final outcome, however, is the opposite of what is found in real life. Let us now examine another case from Table I, the steady-state model.

Using Eq. (2.39) it is easy to see that a retarded wave emitted at $t = 0$ by a source at $r = 0$ arrives at the absorber particle at the coordinate r at $\tau = r$ and therefore the frequency ω at the source is redshifted to $\omega(r) = \omega(1 - Hr)$. The number density of absorber particles per unit proper volume, however, remains constant at $N = N_0$, say. Again, evaluating the parameter k in the low frequency limit we finally get the absorption integral of Eq. (2.45) as

$$I = -\int_0^{H^{-1}} k dr \sim \int_0^{H^{-1}} \frac{dr}{1 - Hr} = \infty. \tag{2.52}$$

This integral clearly diverges, thus ensuring perfect absorption in the future.

For the past absorber, a similar calculation gives the blueshifted frequency of the advanced wave at the absorber particle with coordinate r to be $\omega(r) = \omega(1 + Hr)$. Thus again we are dealing with high frequency waves at the asymptotic past infinity. However, the number density is still constant and hence from Eq. (2.50) the limiting value of the constant $-k$ is $\propto \omega(r)^{-3}$. So the absorption integral of Eq. (2.45) becomes

$$I = -\int^{\infty} k dr \propto \int^{\infty} (1 + Hr)^{-3} dr < \infty. \tag{2.53}$$

This integral converges, indicating imperfect absorption.

This is another example of the past and future absorbers behaving differently. In this case, however, we do get the right answer, viz. that only the retarded solution is self-consistent.

The calculations with regard to the response of the past absorber as given above carry a caveat. The determination of the refractive index for very high energy waves cannot really be carried out entirely classically. Quan-

tum effects cannot be ignored. Nevertheless, as stated earlier, if the quantum cross sections converge at high energy (as they must do) the conclusions drawn here will stand.

With regard to the models mentioned in Table I we find a variety of answers to the above type of calculation. The results are summarized in Table II. For the cosmological models with the curvature parameter $= 1$ or -1 the calculation is more involved and was carried out by Roe (1969). Davies (1972b) has also examined a whole class of cosmological models with somewhat different refractive indices. His conclusions in general are similar to that of Table II. He has, however, questioned if the trapping of redshifted waves of frequencies below the plasma frequency in the future absorber of the steady-state universe can be interpreted as absorption. The point, however, is that whatever the physical process it will eventually ensure absorption of all waves of progressively decreasing frequencies as they travel through a future absorber of constant density and infinite extent. Thus the condition (2.16) is satisfied.

Davies (1973) has also pointed out that absorption will take place by macroscopic objects at all wavelengths; i.e., a galaxy will absorb photons at a rate proportional to the photon density ($\propto R^{-3}$) and hence if R increases slower than $t^{1/3}$ as in the Dirac model the time-integrated photon absorption diverges, giving perfect absorption. In the Dirac model with $G \propto 1/t$ the black hole radius tends to zero and so if all matter in galaxies, etc., ultimately ends in black holes the universe would not be opaque. In this sense the Dirac model in Table II is a borderline case.

In some cases in Table II the models give both the advanced and retarded solutions as self-consistent. We call such a case ambiguous since there the cosmological time asymmetry is not able to resolve the issue and we are not better off than Wheeler and Feynman working within the static universe.

Can we use collisional damping to settle these issues as Hogarth had attempted to do? In a self-consistent picture the following must hold. If retarded solutions are to be justified, the future absorber must be perfect and the past absorber imperfect. Now a particular cosmological model may have a perfect future absorber using collisional damping; but how do we judge the efficacy of the past absorber? Being of thermodynamic origin the nature of the phenomenon along the past light cone cannot be determined unambiguously. Thus we cannot settle the issue without an *extra assumption* about the thermodynamic arrow of time. Hence, if we wish to work entirely within the framework of electrodynamics and cosmology we have to avoid the usage of collisional damping as the means of absorption. Once the consistency of retarded solutions is established, however, we can use the above process to compute any actual damping.

Where there is the correct (i.e., retarded) solution emerging clearly we are not only better off *vis a vis* Wheeler and Feynman, but we are also better off compared to the classical field theory, because in such models we are able to link the local electrodynamic time asymmetry to the cosmological one and are thus able to demonstrate that the choice of retarded solutions is not *ad hoc* but forced by the universe. The analysis given here therefore answers the first of the questions raised at the end of Sec. I.C. This gain is very important in rehabilitating action at a distance as a viable classical theory.

From Table II it is clear that cosmologies with ongoing creation of matter deliver the right answer, because in their cases the future absorber has sufficient absorbing matter to be perfect while the past absorber is rarefied enough (for high frequency waves) to be imperfect. Thus, if a workable action-at-a-distance theory is to be the decisive criterion, these theories have to be preferred to those others (like the big-bang models of Friedmann) which give wrong or ambiguous answers. Considering, however, that this verdict on cosmology is the exact opposite to the current beliefs in the validity of big-bang models more has to be said in justification of action at a distance as the correct approach to electrodynamics.

In this context, the most important issues are raised in questions 4 and 5 at the end of Sec. I.C. Can the action-at-a-distance formulation be developed at the quantum level and does it throw more light on issues which trouble the field theory, e.g., the problem of self-action? We will review the progress on these fronts next.

III. QUANTUM ELECTRODYNAMICS—NONRELATIVISTIC PROCESS

A. The path-integral approach to quantum mechanics

1. Introduction

We have so far proceeded along classical lines. We have shown that the direct-particle approach to electromagnetism works at least as well as the Maxwell field approach in explaining all the classical phenomena of the interaction of charges, and that it links with cosmology in an interesting way. The choice of retarded potentials is not an *ad hoc* choice, but is dictated by the universe at large. Moreover, the unbounded motions of charges moving under the self-force do not arise in this theory.

Although success in the classical domain is necessary for any physical theory, it is not sufficient. Nature, as we understand it today, is quantum in character. In electrodynamics, quantum theory has unearthed a vast collection of phenomena outside the concepts of classical physics. These have been explained with remarkable success by the quantized version of Maxwell's theory, although there have been conceptual and mathematical stumbling blocks, too. Can the direct-particle theory do as well here, if not better?

At first sight, an attempt to extend the classical direct-particle theory to include quantum phenomena seems unlikely to succeed. In Maxwell's theory we have fields to quantize. The degrees of freedom of these fields result in packets of energy called "photons," which play such an

important part in quantum electrodynamics. We have no analogous degrees of freedom in direct-particle fields. Thus photons do not appear to exist in the latter theory. The only degrees of freedom are those vested in the particles. Can we get all the conventional quantum electrodynamics by a first quantization alone? If not, the theory fails. If we can, however, the theory must be regarded as the superior theory, because it reproduces all observations under fewer degrees of freedom.

Other difficulties can be anticipated concerning particles and antiparticles. In classical theory all worldlines are endless and timelike. In relativistic quantum electrodynamics, the worldlines can go forward and backward in time. What happens to the "identity" of a worldline under these circumstances? How does the rule of no self-interaction operate under such conditions?

These are some of the problems which arise when we undertake to quantize the Fokker theory. We shall proceed by stages in solving them, beginning with the simpler nonrelativistic picture and ending with fully relativistic interactions of electrons and positrons.

2. Path amplitudes

Suppose a physical system has action S. This is defined in terms of "paths" Γ that the system can follow in coordinate space. Classical physics tells us that not all Γ are permissible. In general, there is a unique path Γ from a given point P_1 in coordinate space to a given point P_2. Writing Γ_0 for this path, Γ_0 is given by the principle of stationary action

$$\delta S = 0 \quad \text{for} \quad \Gamma = \Gamma_0. \tag{3.1}$$

Earlier, we have used this mysterious prescription, and have noted its remarkable success in classical physics. In the thirties Dirac (1935) suggested the interesting idea, later developed quantitatively by Feynman, that (3.1) is the consequence of a more general principle operating in quantum mechanics. In quantum theory the system permits any of the paths Γ from P_1 to P_2, but each path has a definite probability amplitude proportional to

$$\exp\,(iS/\hbar), \tag{3.2}$$

where $\hbar$ is Planck's constant. All amplitudes add, giving a total amplitude for the system that goes from P_1 to P_2. In the classical limit $\hbar \to 0$, and (3.2) oscillates wildly as we move from path to path—with the exception of Γ_0 where (3.1) holds. Paths in the neighborhood of Γ_0 make a significant contribution in this limit, whereas the contributions from other paths average out to zero. Hence the classical principle of stationary action.

Feynman carried these ideas further by introducing the concept of the path integral. He defined the nonrelativistic quantum-mechanical propagator $K(P_2;P_1)$ for the system to go from P_1 to P_2 by

$$K(P_2;P_1) = \sum_{\Gamma}(\text{constant})\,\exp\,(iS/\hbar), \qquad t_2 > t_1,$$
$$= 0, \qquad t_2 < t_1. \tag{3.3}$$

In (3.3) the action S is computed for each path Γ, and the sum is over all Γ from P_1 to P_2. The constant is a normalization constant. If, as is usual, the paths form a continuum, the sum is replaced by an integral

$$K(P_2;P_1) = \int \exp[iS(\Gamma)/\hbar]\mathcal{D}\Gamma, \qquad t_2 < t_1. \tag{3.4}$$

The integral is over the continuum of paths and is more complicated than the Riemann or the Lebesgue integral, which are summed over sets of points. Only limited progress has been made toward giving a rigorous mathematical foundation to this concept. Feynman was able, however, to obtain all required physical answers by various subtle devices. We shall draw heavily on these techniques. For details see Feynman and Hibbs (1965). The constant in (3.3) can be absorbed in the measure of $\mathcal{D}\Gamma$.

Suppose P_1 represents the spacetime point (x_1,t_1) and P_2 the point (x_2,t_2) in the motion of a particle. Let $t_2 > t_1$. Any path Γ from P_1 to P_2 will pass through some intermediate point P having time coordinate t. Let the spatial position of P be x. Since the action functional is additive over paths, we can write

$$S(\Gamma_{P_2P_1}) = S(\Gamma_{P_2P}) + S(\Gamma_{PP_1}), \tag{3.5}$$

where the path $\Gamma_{P_2P_1}$ from P_1 to P_2 is made up of the segment Γ_{PP_1} from P_1 to P and the segment Γ_{P_2P} from P to P_2. Hence

$$\exp\,[iS(\Gamma_{P_2P_1})/\hbar] = \exp\,[iS(\Gamma_{P_2P})\hbar]\exp\,[iS(\Gamma_{PP_1})/\hbar]. \tag{3.6}$$

The sum over all paths from P_1 to P_2 can be obtained by summing all paths Γ_{PP_1} and Γ_{P_2P} and integrating over the spatial coordinates of P. From (3.6) together with an appropriate measure for the path integral (3.4), we get

$$K(x_2,t_2;x_1,t_1) = \int K(x_2,t_2;x,t)K(x,t;x_1,t_1)d^3x. \tag{3.7}$$

Together, (3.4) and (3.7) suggest an alternative way of defining the path amplitude. Suppose we divide $\Gamma_{P_2P_1}$ into a large number of small segments with intermediate points Q_r $(r = 1,\ldots,N-1)$. We can define $Q_0 = P_1$ and $Q_N = P_2$. Over each segment (Q_{r-1},Q_r) we may imagine S to change very slowly. We define the amplitude for such a segment to be proportional to $K(Q_r;Q_{r-1})$. The amplitude along the entire path is then given by the product

$$\lim_{N\to\infty}\prod_{r=1}^{N} A_r K(Q_r;Q_{r-1}), \tag{3.8}$$

where A_r is a constant of proportionality with the dimensions of spatial volume. Summing over all paths gives

$$\int \prod_{r=1}^{N} A_r K(Q_r; Q_{r-1}) \mathcal{D}\Gamma = \int \cdots \int K(P_2; Q_{N-1}) K(Q_{N-1}; Q_{N-2}) \cdots K(Q_2; Q_1) K(Q_1; P_1) d^3Q_{N-1} \cdots d^3Q_1$$
$$= K(P_2; P_1) \tag{3.9}$$

by (3.7), the integrations of d^3Q_r being over the time sections $t = t_r; r = 1, \ldots, N-1$.

The constants A_r are again absorbed in the measure of Γ. Because of (3.7), the definition (3.8) leads to the same function $K(P_2; P_1)$ as before.

Sometimes we know K but not S. Then (3.8) is useful to define a path amplitude. The original definition (3.2) is the more direct one, however. We shall use (3.8) in relativistic path-integral theory.

3. The wave function

Suppose that, instead of knowing that the particle is at (x_1, t_1), we only know the probability amplitude $\psi(x_1, t_1)$ for it to be at various spatial positions x_1 on the time section $t = t_1$. We then ask, what is the probability amplitude $\psi(x_2, t_2)$ of finding the particle at (x_2, t_2)? This is obtained from the weighted mean of $K(x_2, t_2; x_1, t_1)$,

$$\psi(x_2, t_2) = \int K(x_2, t_2; x_1, t_1) \psi(x_1, t_1) d^3x_1. \tag{3.10}$$

As $t_2 \to t_1, \psi(t_2) \to \psi(t_1)$. This implies

$$\lim_{t_2 \to t_1} K(x_2, t_2; x_1, t_1) = \delta_3(x_2 - x_1). \tag{3.11}$$

The function ψ is the usual Schrödinger wave function. Does it satisfy the Schrödinger equation? The answer is "yes," and we illustrate this by a simple example. For a particle moving in a potential field V, we have

$$S = \int \left(\frac{1}{2} m \dot{x}^2 - V \right) dt. \tag{3.12}$$

If we substitute this in (3.4) and use (3.10), we find that ψ satisfies the differential equation

$$-\frac{\hbar^2}{2m} \nabla_2^2 \psi + V\psi = i\hbar \frac{\partial \psi}{\partial t_2}. \tag{3.13}$$

This is the one-dimensional Schrödinger equation. The same argument can readily be extended to obtain the three-dimensional Schrödinger equation.

In view of (3.11) and the fact that $K = 0$ for $t_2 < t_1$ (no propagation backward in time) (3.13) implies that K satisfies the equation

$$[(\partial/\partial t_2) - (i\hbar/2m)\nabla_2^2 + (i/\hbar)V] K(x_2, t_2; x_1, t_1)$$
$$= \delta_3(x_2 - x_1) \delta(t_2 - t_1). \tag{3.14}$$

4. Transition probability

Let the particle be in an initial state $\phi_i(x, t_1)$. We wish to determine the probability that it is in state $\phi_f(x, t_2)$ for $t_2 > t_1$. Using (3.10), the amplitude is given by

$$\langle \phi_f | \phi_i \rangle = \int \int \phi_f^*(x_2, t_2) K(x_2, t_2; x_1, t_1) \phi_i(x_1, t_1) d^3x_2 d^3x_1$$
$$= \int \int \int \phi_f^*(x_2, t_2) \exp\left[iS(\Gamma_{21})/\hbar \right] \phi_i(x_1, t_1) \mathcal{D}\Gamma_{21} d^3x_2 d^3x_1, \tag{3.15}$$

where the asterisk denotes complex conjugate.

The transition probability from ϕ_i to ϕ_f is given by the square of the modulus of (3.15), i.e., by

$$P(\phi_i \to \phi_f) = |\langle \phi_f | \phi_i \rangle|^2$$
$$= \int \int \int \int \int \int \phi_f^*(x_2, t_2) \exp\left[iS(\Gamma_{21})/\hbar \right] \phi_i(x_1, t_1) \phi_i^*(x_1', t_1) \exp\left[-iS(\Gamma_{21}'/\hbar) \right] \phi_f(x_2', t_2)$$
$$\times \mathcal{D}\Gamma_{21} \mathcal{D}\Gamma_{21}' d^3x_1 d^3x_2 d^3x_1' d^3x_2'. \tag{3.16}$$

We shall often refer to Γ_{21}' as a conjugate path, implying that the action along Γ_{21}' is multiplied by $-i$ instead of $+i$.

5. Perturbation theory

We now consider the problem of perturbation expansion from the path-integral point of view. Suppose we know the behavior of a system with action S_0. Let the system be disturbed by a potential $V(x, t)$ over the time interval (t_1, t_2). The action during this interval is therefore

$$S = S_0 - \int_{t_1}^{t_2} V(x, t) dt. \tag{3.17}$$

We now wish to determine $K_V(x_2,t_2;x_1,t_1)$ in terms of the unperturbed propagator $K_0(x_2,t_2;x_1,t_1)$. We have

$$K_0(x_2,t_2;x_1,t_1)=\int \exp\left[\frac{iS_0(\Gamma_{21})}{\hbar}\right]\mathcal{D}\Gamma_{21}, \quad (3.18)$$

$$K_V(x_2,t_2;x_1,t_1)=\int \exp\left[\frac{iS_0(\Gamma_{21})}{\hbar}-\frac{i}{\hbar}\int_{\Gamma_{21}} V dt\right]\mathcal{D}\Gamma_{21}. \quad (3.19)$$

Expanding the exponential involving the integral over V, then

$$K_V(x_2,t_2;x_1,t_1)=\int \exp\left[\frac{iS_0(\Gamma_{21})}{\hbar}\right]\left[1-(i/\hbar)\int_{\Gamma_{21}} V dt-(1/2\hbar^2)\left(\int_{\Gamma_{21}} V dt\right)^2+\cdots\right]\mathcal{D}\Gamma_{21}. \quad (3.20)$$

The path integral for the unity term in this expansion gives K_0. We consider the next term

$$-\int \exp\left[\frac{iS_0(\Gamma_{21})}{\hbar}\right]\frac{i}{\hbar}\int_{\Gamma_{21}} V dt\mathcal{D}\Gamma_{21}. \quad (3.21)$$

The integral $\int_{\Gamma_{21}} V dt$ is over a specific path Γ_{21}, given by a function $x(t)$. Suppose we take a particular instant t of time and take all paths which pass through $x(t)$ on their way from P_1 to P_2. If we sum over these paths alone, we will get, as in the analysis leading to (3.7), the product

$$K_0(x_2,t_2;x,t)K_0(x,t;x_1,t_1). \quad (3.22)$$

If the integral over V were absent, we would simply have obtained (3.7). But now we have to weight (3.22) with $-iV(x,t)/\hbar$. Then we have to sum over all x, to include all paths from 1 to 2. Finally we perform the time integral to get the entire contribution:

$$-\frac{i}{\hbar}\int_{t_1}^{t_2} K_0(x_2,t_2;x,t)V(x,t)K_0(x,t;x_1,t_1)d^3x dt. \quad (3.23)$$

The physical meaning of this operation is illustrated in Fig. 2. If we consider a given t and x, we imagine the system to proceed undisturbed from (x_1,t_1) to (x,t). Then it is scattered by $V(x,t)$, after which it proceeds undisturbed from (x,t) to (x_2,t_2). This scattering could occur anywhere within the spacetime slab $t_1 \le t \le t_2$. Hence the integration in (3.23). The four-dimensional volume element, $d^3x dt = d\tau$, say, can actually be taken over the whole of spacetime, since none of the K_0 functions involves propagation backward in time. The integrand is therefore zero whenever the point x,t, falls outside the slab $t_1 \le t \le t_2$.

Higher-order terms in the expansion give multiple scattering processes. Thus (3.19) is the closed form of the usual infinite perturbation series:

$$K_V(x_2,t_2;x_1,t_1)=K_0(x_2,t_2;x_1,t_1)-\frac{i}{\hbar}\int K_0(x_2,t_2;x,t)V(x,t)K_0(x,t;x_1,t_1)d\tau + \left(\frac{i}{\hbar}\right)^2\int\int K_0(x_2,t_2;x_3,t_3)V(x_3,t_3)K_0(x_3,t_3;x_4,t_4)V(x_4,t_4)K_0(x_4,t_4;x_1,t_1)d\tau_3 d\tau_4+\cdots. \quad (3.24)$$

6. Transition element

Classically we are use to continuous changes of dynamical variables. In quantum mechanics, transitions are in general discontinuous. Can we give meaning to "velocity" or "acceleration" in discontinuous transitions? In the path-integral formulation, we can indeed give a meaning to such terms, by means of the concept of transition elements. Suppose ϕ_i and ϕ_f are the initial and final states of a system described by action S. We have already seen that the probability amplitude can be defined by the path integral

$$\langle\phi_f|\phi_i\rangle=\int\int\int \phi_f^* \exp(iS/\hbar)\phi_i\mathcal{D}\Gamma d^3x_1 d^3x_2. \quad (3.25)$$

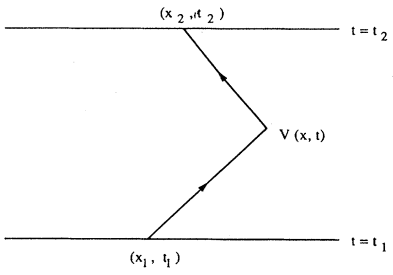

FIG. 2. The potential $V(x,t)$ scatters the system from (x_1,t_1) towards (x_2,t_2). The scattering can be along any set of paths weighted by their respective probability amplitude. This represents the first-order perturbation.

Suppose we now have a functional $F[\Gamma]$ of a path Γ. Then the transition element of F is given by

$$\langle\phi_f|F|\phi_i\rangle = \int\int\int \phi_f^* \exp\left[\frac{iS(\Gamma)}{\hbar}\right] \times F[\Gamma]\phi_i \mathcal{D}\Gamma d^3x_1 d^3x_2. \tag{3.26}$$

This definition means that $\langle\phi_f|F|\phi_i\rangle$ is a certain kind of average of $F[\Gamma]$ over all paths suitably weighted by the initial and final wave functions. Classically, we could have calculated $F[\Gamma]$ exactly, since we know that $\Gamma = \Gamma_0$ is the unique solution. In quantum mechanics, we cannot make such a definitive statement.

Equation (3.26) is easy to apply to simple cases. For instance, for a free particle of mass m, the transition element of velocity is given by

$$\langle\phi_f|\dot{x}|\phi_i\rangle = \int -\frac{i\hbar}{m}\phi_f^* \nabla_x \phi_i d^3x. \tag{3.27}$$

As seen in this example, a transition element need not be real even if the original dynamical variable is real.

7. Influence functional

We come now to the concept most useful for the quantum development of direct-particle theories. Suppose we have one quantum-mechanical system in interaction with another, and suppose we are interested only in the detailed behavior of the first system regardless of what happens to the second. We may speak of the second system as the "external environment" of the first. To fix ideas, let the first system be described by a coordinate q and the second by Q. The combined action for the two systems is taken to be of the form

$$S = S_0[q(t)] + S_E[Q(t)] + S_I[q(t), Q(t)], \tag{3.28}$$

where S_0 represents the action of the first system alone, S_E is the action of the second system alone, and S_I represents the interaction of the two systems. Let $\phi_i(q_i, t_i)$ be the initial state and $\phi_f(q_f, t_f)$ the final state of the first system. Using (3.16), the probability of transition $\phi_i \to \phi_f$ is given by

$$P(\phi_i \to \phi_f) = \int\int\int\int\int\int \phi_f^*(q_f,t_f)\phi_i^*(q_i',t_i)\phi_i(q_i,t_i)\phi_f(q_f',t_f) \times \exp\left(\frac{i}{\hbar}\{S_0[q(t)] - S_0[q'(t)]\}\right) F[q,q']\mathcal{D}q\mathcal{D}q' dq_f dq_f' dq_i dq_i', \tag{3.29}$$

where

$$F[q(t), q'(t)] = \sum_f \left\langle \psi_f \middle| \exp\left\{\frac{i}{\hbar}S_I[q(t), Q(t)]\right\} \middle| \psi_i \right\rangle \left\langle \psi_f \middle| \exp\left\{\frac{i}{\hbar}S_I[q'(t), Q'(t)]\right\} \middle| \psi_i \right\rangle^*. \tag{3.30}$$

Here ψ_i is the initial state of the second system. We have summed over all final states ψ_f, since we are not interested in how the environment ends up.

$F[q(t), q'(t)]$ is called the *influence functional* and represents the "force" exerted by the environment on the first system.

Feynman and Hibbs (1965) discuss influence functionals and their applications in some detail; so we shall not discuss them further. We expect the universe to act as the external environment in the quantum version of the Wheeler-Feynman theory. It is our aim to determine the quantum analog of the response of the universe that earlier we calculated classically. This response will appear in the form of an influence functional.

B. Absorption and stimulated emission

While a first quantization of particles suffices to determine transition probabilities with respect to a prescribed potential function, as in A.4 and A.5, transitions in a radiation field are usually considered to require a quantization of the field itself—i.e., a resolution into photons rather than classical wave theory. Then to match the quantization of the field it is usual to adopt a second quantization of the particles. However, we shall show here that none of this is necessary, all the standard radiation results being obtainable without any more quantization than that given in Sec. A being needed.

In a simplified notation the probability $P(m \to n)$ given by (3.16) of transition in the time interval $0 \le t \le T$ from the state m to an orthogonal state n in a specified (unquantized) external field is given by

$$P(m \to n) = \int\int\int\int \phi_n^*(\mathbf{a}_f)\phi_n(\mathbf{a}_f')J\phi_m(\mathbf{a}_i) \times \phi_m^*(\mathbf{a}_i')d^3\mathbf{a}_f d^3\mathbf{a}_f' d^3\mathbf{a}_i d^3\mathbf{a}_i', \tag{3.31}$$

where ϕ_m, ϕ_n are the wave functions for the states, and J is the double path integral

$$J = \int\int \exp\left[\frac{i}{\hbar}\{S[\mathbf{a}(t)] - S[\mathbf{a}'(t)]\}\right] \mathcal{D}^3\mathbf{a}(t)\mathcal{D}^3\mathbf{a}'(t). \tag{3.32}$$

Here the path $\mathbf{a}(t)$ "begins" at $\mathbf{r} = \mathbf{a}_i$, $t = 0$ and "ends" at $\mathbf{r} = \mathbf{a}_f$, $t = T$, while the path $\mathbf{a}'(t)$ begins at $\mathbf{r} = \mathbf{a}_i'$, $t = 0$ and ends at $\mathbf{r} = \mathbf{a}_f'$, $t = T$. The action $S[\mathbf{a}(t)]$ is a

functional of the path $\mathbf{a}(t)$ and is defined by

$$S[\mathbf{a}(t)] = \int_0^T L(\mathbf{a}, t)dt, \tag{3.33}$$

and similarly for $S[\mathbf{a}'(t)]$.

We are concerned with an electronic transition and with a situation in which

$$L = \frac{1}{2}m\dot{\mathbf{a}}^2 + eV(\mathbf{a}, t) - e\dot{\mathbf{a}} \cdot \mathbf{A}(\mathbf{a}, t), \tag{3.34}$$

where e, m are the electronic charge and mass, the velocity of light is taken as unity, the nonrelativistic kinetic energy is used, and $V(\mathbf{a}, t), \mathbf{A}(\mathbf{a}, t)$ are the potentials of the specific field. In a simple one-electron atomic problem the external field is made up of the electrostatic field within the atom together with the field incident on the atom. Provided the latter has no electrostatic component it yields the vector potential $\mathbf{A}$, which in the Coulomb gauge satisfies

$$\text{div}\mathbf{A} = 0. \tag{3.35}$$

The external field $\mathbf{A}$ is then wholly transverse. With this division of V and $\mathbf{A}$, we write

$$S[\mathbf{a}(t)] = S_0[\mathbf{a}(t)] - e\int_0^T \dot{\mathbf{a}} \cdot \mathbf{A}dt, \tag{3.36}$$

so that $S_0[\mathbf{a}(t)]$ is the action for the atomic field alone.

Next we regard $\mathbf{A}$ as small enough for

$$\exp\left[-\frac{ie}{\hbar}\int_0^T \dot{\mathbf{a}} \cdot \mathbf{A}dt\right] = 1 - \frac{ie}{\hbar}\int_0^T \dot{\mathbf{a}} \cdot \mathbf{A}dt + \cdots \tag{3.37}$$

to be rapidly convergent. The dominant term in (3.31) then involves the product of the first-order term in (3.36) with the first-order term in the complex conjugate of (3.36). After a reduction that is straightforward, except perhaps for the calculation of the transition element of the velocity (Feynman and Hibbs, 1965, p. 184) one obtains

$$P(m \to n) = \frac{e^2}{\hbar^2}\left(\frac{E_n - E_m}{\hbar}\right)^2 \left|\int_0^T \exp\frac{i(E_n - E_m)t}{\hbar} \times \int \phi_n^*(\mathbf{a})\mathbf{a} \cdot \mathbf{A}\phi_m(\mathbf{a})d^3\mathbf{a}dt\right|^2. \tag{3.38}$$

To express this result in a more familiar form expand the incident field in a Fourier series. For this step consider the atom as situated inside a cube with side of unit length, the latter chosen to be very large compared with any wavelength of importance in the transition problem. Write

$$\mathbf{A}(\mathbf{a}, t) = \sqrt{4\pi}\sum_{\mathbf{k}}\{\mathbf{c}_k \exp[i(\mathbf{k} \cdot \mathbf{a} - kt)] + \mathbf{c}_\mathbf{k}^* \exp[-i(\mathbf{k} \cdot \mathbf{a} - kt)]\}, \tag{3.39}$$

noting that $\mathbf{k} = 2\pi(n_1, n_2, n_3)$ where n_1, n_2, n_3 are integers, and that $\mathbf{k} \cdot \mathbf{c}_\mathbf{k} = 0$ because of the Coulomb gauge. Inserting (3.39) in (3.38) leads to a double sum, $\sum_\mathbf{k}\sum_{\mathbf{k}'}$ say. Waves with $\mathbf{k} \neq \mathbf{k}'$ are regarded as coming from different oscillators and so may be regarded as being in random phase with respect to each other. After averaging $\sum_\mathbf{k}\sum_{\mathbf{k}'}$ thus reduces to $\sum_\mathbf{k}$. Moreover, T is very large; even for allowed transitions there are a very large number of atomic oscillations in time T so that

$$\left|\int_0^T \exp\left[\frac{i}{\hbar}(E_n - E_m \pm \hbar k)t\right]dt\right|^2 \simeq 2\pi T\delta\left(k \pm \frac{E_n - E_m}{\hbar}\right). \tag{3.40}$$

When $E_n > E_m$ the important terms involve the minus sign in the delta function, and (3.38) leads to

$$(4\pi e^2/\hbar)^2\left(\frac{E_n - E_m}{\hbar}\right)^2 \sum_\mathbf{k}\left|\mathbf{c}_\mathbf{k} \cdot \int \phi_n^*(\mathbf{a})\mathbf{a}e^{i\mathbf{k}\cdot\mathbf{a}}\phi_m(\mathbf{a})d^3\mathbf{a}\right|^2 \left|\int_0^T \exp\left[\frac{i}{\hbar}(E_n - E_m - \hbar k)t\right]dt\right|^2, \tag{3.41}$$

where we have deferred using (3.40) until after the sum has been converted to an integral. This is the case of absorption.

When $E_m > E_n$ we have the case of stimulated emission, and the corresponding result is

$$(4\pi e^2/\hbar^2)\left(\frac{E_m - E_n}{\hbar}\right)^2 \sum_\mathbf{k}\left|\mathbf{c}_\mathbf{k}^* \cdot \int \phi_n^*(\mathbf{a})\mathbf{a}e^{-i\mathbf{k}\cdot\mathbf{a}}\phi_m(\mathbf{a})d^3\mathbf{a}\right|^2 \left|\int_0^T \exp\left[\frac{i}{\hbar}(E_n - E_m + \hbar k)t\right]dt\right|^2. \tag{3.42}$$

It is not hard to see that (3.42) is the same as (3.41). Although $\mathbf{c}_\mathbf{k}^* e^{-i\mathbf{k}\cdot\mathbf{a}}$ has replaced $\mathbf{c}_k e^{i\mathbf{k}\cdot\mathbf{a}}$ the roles of the states m, n have been switched; ϕ_m is the state of lower energy in (3.41) whereas ϕ_n is the wave function of the lower state in (3.42). This confirms that the transition probability for stimulated emission in an external field is equal to that for absorption.

It is usual to express the transition probability in terms

of the intensity of the applied field. In order to do this it is necessary either to approximate (3.41) or to average the transition probability with respect to the orientation of the atom. We adopt the latter procedure. We define

$$\mathbf{a}_{mn}(\mathbf{k}) = \int \phi_n^*(\mathbf{a})\mathbf{a}e^{i\mathbf{k}\cdot\mathbf{a}}\phi_m(\mathbf{a})d^3\mathbf{a}, \tag{3.43}$$

and let $\alpha_{\mathbf{k}}^{(1)}, \alpha_{\mathbf{k}}^{(2)}$ be unit vectors which together with $\mathbf{k}/k$ form an orthogonal triad. Because of $\mathbf{c}_{\mathbf{k}} \cdot \mathbf{k} = 0$,

$$\mathbf{c}_{\mathbf{k}} = \sum_{j=1,2} [\mathbf{c}_{\mathbf{k}} \cdot \mathbf{a}_{\mathbf{k}}^{(j)}]\alpha_{\mathbf{k}}^{(j)} \tag{3.44}$$

and

$$|\mathbf{c}_{\mathbf{k}} \cdot \mathbf{a}_{mn}(\mathbf{k})|^2 = \left| \sum_{j=1,2} (\mathbf{c}_{\mathbf{k}} \cdot \alpha_{\mathbf{k}}^{(j)}) \int \phi_n^*(\mathbf{a})\alpha_{\mathbf{k}}^{(j)} \cdot \mathbf{a}e^{i\mathbf{k}\cdot\mathbf{a}}\phi_m(\mathbf{a})d^3\mathbf{a} \right|^2. \tag{3.45}$$

We have to average (3.45) with respect to orientation.

It is not hard to show that the sum of

$$\int \phi_n^*(\mathbf{a})\alpha_{\mathbf{k}}^{(1)} \cdot \mathbf{a}e^{i\mathbf{k}\cdot\mathbf{a}}\phi_m(\mathbf{a})d^3\mathbf{a} \cdot \int \phi_n(\mathbf{a})\alpha_{\mathbf{k}}^{(2)} \cdot \mathbf{a}e^{-i\mathbf{k}\cdot\mathbf{a}}\phi_m^*(\mathbf{a})d^3\mathbf{a} \tag{3.46}$$

and its complex conjugate averages to zero, and that (3.45) is proportional to $|\mathbf{c}_{\mathbf{k}}|^2$ and can be written in the form $\frac{1}{3}|\mathbf{c}_{\mathbf{k}}|^2|\mathbf{a}_{mn}(k)|^2$, where $|\mathbf{a}_{mn}(k)|^2$ depends on the magnitude but not the direction of $\mathbf{k}$, and is equal to $|\mathbf{a}_{mn}|^2$ when the factor $\exp(i\mathbf{k}\cdot\mathbf{a})$ in (3.45) is approximated by unity. Hence the average value of (3.41) is

$$\frac{4\pi e^2}{3\hbar^2}\left(\frac{E_n - E_m}{\hbar}\right)^2 \sum_{\mathbf{k}} |\mathbf{c}_{\mathbf{k}}|^2 \cdot |a_{mn}(k)|^2 \left| \int_0^T \exp\left[\frac{i}{\hbar}(E_m - E_n + \hbar k)t\right] dt \right|^2. \tag{3.47}$$

There are $d^3\mathbf{k}/(2\pi)^3$ terms of this series in the element $d^3\mathbf{k}$ of $\mathbf{k}$ space. Defining $\overline{|\mathbf{c}_{\mathbf{k}}|^2}$ by

$$\overline{|\mathbf{c}_{\mathbf{k}}|^2}\frac{d^3\mathbf{k}}{(2\pi)^3} = \sum_{d^3\mathbf{k}} |\mathbf{c}_{\mathbf{k}}|^2, \tag{3.48}$$

we now write (3.47) in the integral form

$$\frac{e^2}{6\pi^2\hbar^2}\left(\frac{E_n - E_m}{\hbar}\right)^2 \int \overline{|\mathbf{c}_{\mathbf{k}}|^2} \cdot |\mathbf{a}_{mn}(k)|^2$$

$$\times \left| \int_0^T \exp\left[\frac{i}{\hbar}(E_n - E_m - \hbar k)t\right] dt \right|^2 \cdot d^3\mathbf{k}. \tag{3.49}$$

To define the intensity $I(\mathbf{k})$ per unit solid angle we note first that

$$\frac{1}{4\pi}\mathbf{E} \times \mathbf{H} = 2\sum_{\mathbf{k}} |\mathbf{c}_{\mathbf{k}}|^2 k\mathbf{k} = \frac{1}{4\pi^3}\int \overline{|\mathbf{c}_{\mathbf{k}}|^2} k\mathbf{k}d^3\mathbf{k} \tag{3.50}$$

and that the contribution of solid angle $d\Omega$ to (3.50) is

$$\frac{d\Omega}{4\pi^3} \cdot \frac{\mathbf{k}}{k}\int \overline{|\mathbf{c}_{\mathbf{k}}|^2} k^4 dk, \tag{3.51}$$

where $\mathbf{k}$ lies in $d\Omega$. The intensity $I(\mathbf{k})$ is now defined by equating (3.51) to

$$d\Omega \cdot \frac{\mathbf{k}}{k}\int I(\mathbf{k})dk.$$

For this definition to hold irrespective of $\overline{|\mathbf{c}_{\mathbf{k}}|^2}$ we must have

$$I(\mathbf{k}) = \frac{1}{4\pi^3}k^4\overline{|\mathbf{c}_{\mathbf{k}}|^2}. \tag{3.52}$$

Eliminating $\overline{|\mathbf{c}_{\mathbf{k}}|^2}$ between (3.49) and (3.52), and using (3.40), one easily obtains

$$\frac{4\pi^2 e^2}{3\hbar^2}\int d\Omega \int_0^\infty I(\mathbf{k})|\mathbf{a}_{mn}(k)|^2 \delta\left(k - \frac{|E_n - E_m|}{\hbar}\right) dk \tag{3.53}$$

for the transition per unit time. The result (3.53) applies both to absorption and stimulated emission, and is the usual relation between the intensity and the transition probability. We can introduce separate intensities $I^{(j)}(\mathbf{k})$ for the "polarization" directions $\alpha_{\mathbf{k}}^{(j)}$ by defining

$$I^{(j)}(\mathbf{k}) = \frac{k^4}{4\pi^3}\overline{|\mathbf{c}_{\mathbf{k}} \cdot \alpha_{\mathbf{k}}^{(j)}|^2}; \qquad j = 1, 2. \tag{3.54}$$

Evidently

$$I(\mathbf{k}) = \sum_{j=1,2} I^{(j)}(\mathbf{k}). \tag{3.55}$$

We end the present section by introducing the concept of opacity. Suppose there are $n(k)dk$ atoms per unit volume in state m, where $E_m < E_n$ and $(E_n - E_m)/\hbar$ lies between k and $k + dk$. Define a mean value of $|\mathbf{a}_{mn}(k)|^2$ by

$$V\overline{|\mathbf{a}_{mn}(k)|^2} \cdot n(k)dk = \sum_V |\mathbf{a}_{mn}(k)|^2, \tag{3.56}$$

the summation being taken through the small volume V for all atoms with $(E_n - E_m)/\hbar$ between k and $k + dk$. Both $\overline{|\mathbf{a}_{mn}(k)|^2}$ and $n(k)$ can be functions of position as well as of k.

Suppose radiation of frequency k travels along a three-dimensional spatial path Γ connecting two points P_1 and P_2. Then the opacity difference between these points is $\int_\Gamma d\tau$, where $d\tau$ is the difference for an element of the path. To obtain $d\tau$, let s be the three-dimensional length along Γ and let $\mathbf{u}(s)$ be the unit tangent vector at s. The opacity differential $d\tau(k)/ds$ for frequency k is defined in terms of absorption by the equation

$$I(k\mathbf{u})d\Omega dk\frac{d\tau(k)}{ds} = \frac{4\pi^2e^2}{3\hbar^2}d\Omega I(k\mathbf{u})\overline{|\mathbf{a}_{mn}(k)|^2}\hbar k n(k)dk, \tag{3.57}$$

and the opacity difference is

$$\int_{P_1}^{P_2} d\tau(k) = \frac{4\pi^2e^2}{3\hbar} \cdot k \int_{P_1}^{P_2} \overline{|\mathbf{a}_{mn}(k)|^2} n(k)ds. \tag{3.58}$$

Radiation traveling from P_1 and P_2 is reduced in intensity by the factor $\exp[-\int_{P_1}^{P_2} d\tau]$, it being supposed that enough atoms are involved for the absorption probability to be averaged. It may be noted that the factor $\hbar k$ appears on the right-hand side of (3.57) because of $E_n - E_m \cong \hbar k$, not because of field quantization.

C. Spontaneous emission

We come now to a crux in the discussion of this article, which unless it can be overcome would end the entire development. Whereas in Maxwell's theory the field has independent degrees of freedom that, after quantization, generate spontaneous emission from the excited states of atoms, here we have no such resource. How then is spontaneous emission to be understood and calculated?

To fix ideas, let us consider the motion of particle a in the time-interval $0 \le t \le T$ (see Fig. 3), and denote the displacement of a by $\mathbf{a}(t)$. Let b be a typical absorber particle whose world line is intersected in intervals $\mathbf{\Delta}_-$ and $\mathbf{\Delta}_+$, respectively, by the past and future light cones from the initial point $[\mathbf{a}(0), 0]$ and the final point $[\mathbf{a}(T), T]$ on the path $\mathbf{a}(t)$ of a. From what has been said above, the induced transitions of a arise from its interaction with the retarded field of b, i.e., from the portion $\mathbf{\Delta}_-$ of the worldline of b. The action governing induced transitions is therefore

$$-e_a \int_0^T \mathbf{A}_{\text{ret}}^{(b)}(\mathbf{a}) \cdot \dot{\mathbf{a}}dt, \tag{3.59}$$

where $\mathbf{A}_{\text{ret}}^{(b)}$ is the full retarded 3 potential from b. To calculate spontaneous transitions, we need, on the other hand, the transitions of b induced by the full retarded field of a. The action governing this is

$$-e_b \int_{\mathbf{\Delta}_+} \mathbf{A}_{\text{ret}}^{(a)}(\mathbf{b}) \cdot \dot{\mathbf{b}}dt, \tag{3.60}$$

with a similar notation. As in the classical case, we are looking for a self-consistent cycle of argument in which the net field is the retarded one. Throughout this calculation, we shall work in a conformally flat cosmological model with a perfect future absorber.

Our aim is to calculate the influence functional governing the motion of a and arising from the whole universe. Since we may assume the different absorber particles to act independently, the influence functional has the form

$$F[\mathbf{a}(t), \mathbf{a}'(t)] = \prod_{b \neq a} F^{(b)}[\mathbf{a}(t), \mathbf{a}'(t)], \tag{3.61}$$

where $F[\mathbf{a}(t), \mathbf{a}'(t)]$ is the influence functional exerted by a typical particle b.

To calculate $F^{(b)}[\mathbf{a}(t), \mathbf{a}'(t)]$ we consider all transitions produced by (3.60) in the absorber particle b. If $\psi_i(\mathbf{b})$ is the initial wave function and $\psi_f(\mathbf{b})$ any final wave function of b, we have from (3.30)

$$F^{(b)}[\mathbf{a}(t), \mathbf{a}'(t)] = \sum_f \int\int\int\int \psi_f^*(\mathbf{b}_f)\psi_f(\mathbf{b}'_f)J^{(b)}\psi_i(\mathbf{b}_i)\psi_i^*(\mathbf{b}'_i)d^3\mathbf{b}_i d^3\mathbf{b}_f d^3\mathbf{b}'_i d^3\mathbf{b}'_f, \tag{3.62}$$

where

$$J^{(b)} = \int\int \exp\frac{i}{\hbar}\{S_E[\mathbf{b}(t)] - S_E[\mathbf{b}'(t)] + S_I[\mathbf{a}(t), \mathbf{b}(t)] - S_I[\mathbf{a}'(t), \mathbf{b}'(t)]\} \cdot \mathcal{D}\mathbf{b}\mathcal{D}\mathbf{b}'. \tag{3.63}$$

With the interaction governed by (3.59), we write

$$S_I[\mathbf{a}(t), \mathbf{b}(t)] = -e_b \int_{\mathbf{\Delta}_+} \mathbf{A}_{\text{ret}}^{(a)}(\mathbf{b}) \cdot \dot{\mathbf{b}}dt \tag{3.64}$$

in (3.63), with a similar expression for $S_I[\mathbf{a}'(t), \mathbf{b}'(t)]$.

Because of the cooling produced by the expansion of the universe, b is usually in its ground state. We shall take ψ_i to be the ground state. We can expand the part of the exponential in (3.63) containing the interaction terms only. The unity term in the expansion corresponds to $\psi_f \to \psi_i$. Thus we have

$$F^{(b)}[\mathbf{a}, \mathbf{a}'] = 1 + \sum_{f \neq i} \int\int\int\int\int\int \psi_f^*(\mathbf{b}_f)\psi_f(\mathbf{b}'_f)\psi_i^*(\mathbf{b}'_i)\psi_i(\mathbf{b}_i)(1/\hbar^2)\left[e_b^2 \int_{\Delta_+} \mathbf{A}_{\text{ret}}^{(a)}(\mathbf{b}) \cdot \dot{\mathbf{b}} dt \int_{\Delta'_+} \mathbf{A}_{\text{ret}}^{(a')}(\mathbf{b}') \cdot \dot{\mathbf{b}}' dt\right]$$
$$\times \exp\left((i/\hbar)\{S_E[\mathbf{b}(t)] - S_E[\mathbf{b}'(t)]\}\right) \mathcal{D}\mathbf{b}' \mathcal{D}\mathbf{b} d^3\mathbf{b}_f d^3\mathbf{b}_i d^3\mathbf{b}'_i d^3\mathbf{b}'_f$$
$$+ \text{ terms in } \left[\int_{\Delta_+}\right]^2, \left[\int_{\Delta'_+}\right]^2 + \cdots, \tag{3.65}$$

where $\mathbf{A}_{\text{ret}}^{(a)}$ is calculated for the path $\mathbf{a}(t)$, $\mathbf{A}_{\text{ret}}^{(a')}$ for the conjugate path $\mathbf{a}'(t)$.

Let E_f and E_i be the energies of states ψ_f and ψ_i, respectively. Then first-order perturbation theory shows that the contribution to the above expression from the transition $\psi_i \to \psi_f$ is

$$\frac{e_b^2}{\hbar^2}\left(\frac{E_f - E_i}{\hbar}\right)^2 M[\mathbf{a}(t)]M^*[\mathbf{a}'(t)], \tag{3.66}$$

where

$$M[\mathbf{a}(t)] = \int_{\Delta_+} \exp\left[\frac{i(E_f - E_i)t}{\hbar}\right] dt \int \psi_f^*(\mathbf{b}) \times \mathbf{A}_{\text{ret}}^{(a)}(\mathbf{b}) \cdot \mathbf{b}\psi_i(\mathbf{b}) d^3\mathbf{b}. \tag{3.67}$$

To calculate (3.67), we proceed as follows. Take an origin near a to measure the displacement $\mathbf{a}(t)$. Similarly, let $\mathbf{b}(t)$ measure the displacement from an origin near b. Let the origin near b have a relative displacement $\mathbf{R}$ with respect to the origin near a. Then the vector from $\mathbf{a}(t)$ to $\mathbf{b}(t)$ is given by

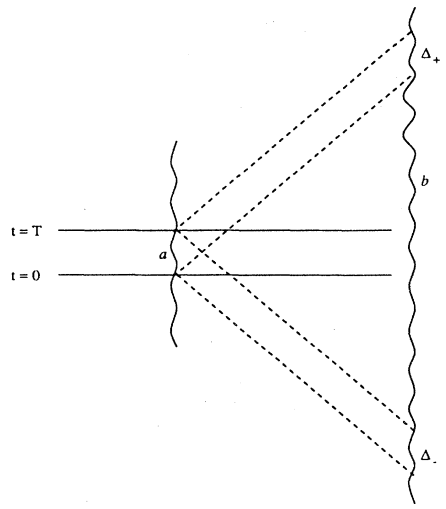

FIG. 3. A schematic diagram of interaction between local particle a and absorber particle b.

$$\mathbf{r} = \mathbf{R} + \mathbf{b} - \mathbf{a}. \tag{3.68}$$

For $|\mathbf{R}| \gg |\mathbf{b}|, |\mathbf{a}|$, we get

$$r = |\mathbf{r}| \cong R + \frac{1}{R}\mathbf{R} \cdot (\mathbf{b} - \mathbf{a}). \tag{3.69}$$

In the absence of any dispersion, we would have, in the Coulomb gauge,

$$\mathbf{A}_{\text{ret}}^{(a)}(\mathbf{b}) = \frac{e_a}{r - \mathbf{r} \cdot \dot{\mathbf{a}}} \sum_{j=1,2} [\alpha^{(j)} \cdot \dot{\mathbf{a}}]\alpha^{(j)}, \tag{3.70}$$

where $\alpha^{(j)}$ are two unit vectors which form a mutually orthogonal triad with $\mathbf{r}/|\mathbf{r}|$.

The quantity $r - \mathbf{r} \cdot \dot{\mathbf{a}}$ varies only slightly with $\mathbf{b}$, and it is sufficiently accurate to replace it in (3.70) by $R - \mathbf{R} \cdot \dot{\mathbf{a}}$,

$$\mathbf{A}_{\text{ret}}^{(a)}(\mathbf{b}) = \frac{e_a}{R - \mathbf{R} \cdot \dot{\mathbf{a}}} \sum_{j=1,2} [\alpha^{(j)} \cdot \dot{\mathbf{a}}]\alpha^{(j)}. \tag{3.71}$$

We Fourier analyze (3.71) to get

$$\mathbf{A}_{\text{ret}}^{(a)}(\mathbf{b}) = \sum_{l=-\infty}^{\infty} \sum_{j=1,2} [\alpha^{(j)} \cdot \mathbf{A}_l]\alpha^j \exp\left(-\frac{2\pi i l t}{T'}\right), \tag{3.72}$$

where

$$\mathbf{A}_l = \frac{e_a}{RT'} \int \frac{\dot{\mathbf{a}}}{1 - \dot{\mathbf{a}} \cdot \mathbf{R}/R} \exp\left(\frac{2\pi i l t'}{T}\right) dt'. \tag{3.73}$$

The range T' of t' is given by

$$t' = t + R + \frac{1}{R}\mathbf{R} \cdot (\mathbf{b} - \mathbf{a}), \qquad 0 \leq t \leq T. \tag{3.74}$$

Since $\mathbf{b}, \mathbf{R}$ are not to be regarded as varying with t, we have

$$\frac{dt'}{dt} = 1 - \frac{\dot{\mathbf{a}} \cdot \mathbf{R}}{R}, \tag{3.75}$$

so that changing the variable from t' to t in (3.73) gives

$$\mathbf{A}_l = \frac{e_a}{RT'} \exp\left[\frac{2\pi i l}{T'}\left(R + \frac{\mathbf{R}\cdot\mathbf{b}}{R}\right)\right] \times \int_0^T \dot{\mathbf{a}} \exp\left[\frac{2\pi i l}{T'}\left(t - \frac{\mathbf{a}\cdot\mathbf{R}}{R}\right)\right] dt, \tag{3.76}$$

with

$$T' = T - \frac{\mathbf{R}}{R}[\mathbf{a}(T) - \mathbf{a}(0)]. \tag{3.77}$$

The effect of dispersion in the cosmological medium is to introduce both a phase change and damping into (3.76), modifying it to

$$\mathbf{A}_l = \frac{e_a}{RT'} \exp\left\{-\frac{1}{2}\tau_l + i\left[\frac{2\pi l}{T'}\left(R + \frac{\mathbf{R}\cdot\mathbf{b}}{R}\right) + \chi_l\right]\right\} \int_0^T \dot{\mathbf{a}} \exp\left[\frac{2\pi i l}{T'}\left(t - \frac{\mathbf{a}\cdot\mathbf{R}}{R}\right)\right] dt, \tag{3.78}$$

the phase change being expressed by χ_l and the damping by τ_l.

We now substitute (3.72) into (3.67). We shall make use of the fact (to be shown later) that χ_l is a very large phase angle, and use this to wipe out any cross products of $\mathbf{A}_l, \mathbf{A}_{l'}$, which arise in (3.66), except those with $l' = -l$, giving

$$(e_a^2 e_b^2/3R^2T'^2)[(E_f - E_i)^2/\hbar^4] \sum_{l=-\infty}^{\infty} \left|\mathbf{b}_{if}(2\pi l/T')\right|^2 e^{-\tau_l} \left| \int_0^{T'} \exp\left[\frac{it'}{\hbar}\left(E_f - E_i - \frac{2\pi l\hbar}{T'}\right)\right] dt' \right|^2$$

$$\times \sum_{j=1,2} \int_0^T \alpha^{(j)} \cdot \dot{\mathbf{a}} \exp\left[(2\pi l i/T')\left(t - \frac{\mathbf{a}\cdot\mathbf{R}}{R}\right)\right] dt \int_0^T \alpha^{(j)} \cdot \dot{\mathbf{a}}' \exp\left[2\pi i l/T'\left(\frac{\mathbf{a}'\cdot\mathbf{R}}{R} - t\right)\right] dt, \tag{3.79}$$

in which we have already averaged with respect to all orientations of b to remove product terms in the components of $\alpha^{(1)}$ and $\alpha^{(2)}$, and where $\mathbf{b}_{if}(\frac{2\pi l}{T'})$ is the matrix element of $\mathbf{b}$ exp $[i\frac{2\pi l}{T'}\frac{\mathbf{R}\cdot\mathbf{b}}{R}]$ with respect to ψ_i and ψ_f. We have also ignored variations of $\mathbf{b}$ over $\boldsymbol{\Delta}+$ in comparison with T'. Expression (3.86) can be simplified further by the results

$$\left| \int_0^{T'} \exp\left[\frac{it'}{\hbar}\left(E_f - E_i - \frac{2\pi\hbar l}{T'}\right)\right] dt' \right|^2 \simeq 2\pi T'\hbar\delta\left(E_f - E_i - \frac{2\pi l\hbar}{T'}\right), \tag{3.80}$$

$$\sum_{i=-\infty}^{\infty} \frac{\sin^2\alpha}{(\pi l - \alpha)^2} = 1. \tag{3.81}$$

Writing

$$\hbar k = E_f - E_i, \quad \mathbf{k} = \frac{k\mathbf{R}}{R}, \tag{3.82}$$

and using (3.80) and (3.81), we reduce (3.79) to the form, say,

$$(e_a^2 e_b^2 k^2/3R^2\hbar^2) e^{-\tau(k)} |\mathbf{b}_{if}(k)^2|^2 \sum_{j=1,2} \int_0^T \alpha^{(j)} \cdot \dot{\mathbf{a}} e^{-i\mathbf{k}\cdot\mathbf{a} + ikt} dt \int_0^T \alpha^{(j)} \cdot \dot{\mathbf{a}}' e^{i\mathbf{k}\cdot\mathbf{a}' - ikt} dt = X. \tag{3.83}$$

Suppose that at a distance R in a particular solid angle $d\Omega$, there are $n(k)dk$ particles per unit volume with states f and i such that $(E_f - E_i)/\hbar$ lies in the range k to $k + dk$. Writing $\overline{|\mathbf{b}_{if}(k)|^2}$ for the average of $|\mathbf{b}_{if}(k)|^2$ for all systems satisfying this requirement, the contribution to $F[\mathbf{a}, \mathbf{a}']$ from all absorbers between R and $R + dR$ and in $d\Omega$ is

$$[1 + X]^{n(k)dkR^2dRd\Omega}. \tag{3.84}$$

Since X is very small and the index is large, we can rewrite (3.84) in the form

$$\exp\left[X \cdot n(k)dkR^2dRd\Omega\right]. \tag{3.85}$$

The function $\tau(k)$ in (3.83) is just the optical depth of the absorbing medium at frequency k. It is not hard to show that

$$\frac{d\tau(k)}{dR} = \frac{4\pi^2}{3}\frac{e_b^2 k}{\hbar}\overline{|\mathbf{b}_{if}(k)|^2} n(k), \tag{3.86}$$

the damping expressed by $\tau(k)$ being due to induced upward transitions in the absorber. Thus (3.86) follows from first-order perturbation theory. Remembering that absorber particles contribute as a product, as in (3.61), we next integrate the exponent of (3.85) with respect to R and with respect to k. Using (3.86), and letting $\tau \to \infty$ as $R \to \infty$, we obtain

$$\exp\left[(e_a^2/4\pi^2\hbar)d\Omega\int_0^\infty kdk\sum_{j=1,2}\int_0^T(\alpha^{(j)}\cdot\dot{\mathbf{a}})e^{-i\mathbf{k}\cdot\mathbf{a}+ikt}dt\int_0^T(\alpha^{(j)}\cdot\dot{\mathbf{a}}')e^{i\mathbf{k}\cdot\mathbf{a}'-ikt}dt\right]. \tag{3.87}$$

Integrating finally with respect to Ω, we obtain

$$F[\mathbf{a}(t),\mathbf{a}'(t)]=\exp\left[(e^2/4\pi^2\hbar)\int d\Omega\int_0^\infty kdk\sum_{j=1,2}\int_0^T(\alpha_{\mathbf{k}}^{(j)}\cdot\dot{\mathbf{a}})e^{-i\mathbf{k}\cdot\mathbf{a}+ikt}dt\cdot\int_0^T(\alpha_{\mathbf{k}}^{(j)}\cdot\dot{\mathbf{a}}')e^{i\mathbf{k}\cdot\mathbf{a}'-ikt}dt\right]. \tag{3.88}$$

The subscript $\mathbf{k}$ has been added to $\alpha^{(j)}$ since we are now integrating with respect to Ω, and the vectors $\alpha^{(j)}$ change as $d\Omega$ changes—$\mathbf{k}$ is a vector in $d\Omega$.

The last part of the calculation is similar to the procedure that led to (3.38). Expand the exponential in (3.88) and retain only the first-order term in $e^2/\hbar$. Previously we had

$$\frac{e^2}{\hbar^2}\int_0^T\dot{\mathbf{a}}\cdot\mathbf{A}dt\int_0^T\dot{\mathbf{a}}'\cdot\mathbf{A}dt, \tag{3.89}$$

in which $\mathbf{A}$ was a specified field. Now we have the exponent of (3.88). Noting that if $\mathbf{A}$ in (3.89) had not been a real field

$$\frac{e^2}{\hbar^2}\int_0^T\dot{\mathbf{a}}\cdot\mathbf{A}dt\int_0^T\dot{\mathbf{a}}'\cdot\mathbf{A}^*dt \tag{3.90}$$

would still have led to (3.38), we see that provided we replace $e^2/\hbar^2$ by

$$\frac{e^2}{4\pi^2\hbar}\int d\Omega\int_0^\infty kdk\sum_{j=1,2},$$

and provided we write $\mathbf{A}=\alpha_{\mathbf{k}}^{(j)}e^{-i\mathbf{k}\cdot\mathbf{a}+ikt}$, the present case is the same as the previous one. We obtain

$$P(m\to n)=\frac{e^2}{4\pi^2\hbar}\left(\frac{E_m-E_n}{\hbar}\right)^2\int d\Omega\int_0^\infty kdk\sum_{j=1,2}\left|\int\phi_n^*(\mathbf{a})\mathbf{a}\cdot\alpha_{\mathbf{k}}^{(j)}e^{i\mathbf{k}\cdot\mathbf{a}}\phi_m(\mathbf{a})d^3\mathbf{a}\right|^2$$
$$\times\left|\int_0^T\exp\left[\frac{i}{\hbar}(E_n-E_m+\hbar k)t\right]dt\right|^2. \tag{3.91}$$

Using (3.40), we see that $E_m>E_n$ is necessary to obtain a nonzero result, and that the spontaneous emission probability per unit time is

$$(e^2/2\pi\hbar)[(E_m-E_n)/\hbar]^3\int d\Omega\int_0^\infty dk\sum_{k=1,2}\left|\int\phi_n^*(\mathbf{a})\mathbf{a}\cdot\alpha_{\mathbf{k}}^{(j)}e^{-i\mathbf{k}\cdot\mathbf{a}}\phi_m(\mathbf{a})d^3\mathbf{a}\right|^2\delta\left(k-\frac{E_m-E_n}{\hbar}\right) \tag{3.92}$$

in agreement with the usual expression.

Averaging (3.92) with respect to solid angle and using the same notation as in (3.53) gives

$$\frac{e^2}{3\pi\hbar}\left(\frac{E_m-E_n}{\hbar}\right)^3\int_0^\infty dk|\mathbf{a}_{mn}(k)|^2\delta\left(k-\frac{E_m-E_n}{\hbar}\right)\times\int d\Omega. \tag{3.93}$$

Now equate the contribution from $d\Omega$ to (3.53) to $\bar{q}(\mathbf{k})$ times the contribution from $d\Omega$ to (3.93). This gives the following definition of $\bar{q}(\mathbf{k})$

$$\bar{q}(\mathbf{k})=\frac{1}{\hbar}\left(\frac{2\pi}{k}\right)^3\frac{1}{2}\sum_{j=1,2}I^{(j)}(\mathbf{k}), \tag{3.94}$$

where $\hbar k=E_m-E_n$ and $I(\mathbf{k})$ is separated into the two polarizations defined in (3.55). In the usual quantum electrodynamics (3.94) is the relation between the field intensity and the average number of quanta per vacuum oscillator in the frequency range k to $k+dk$. Although quanta do not appear explicitly in the present theory, it is interesting that we obtain the usual formulas by taking the spontaneous transition rate as a reference standard. It follows that, if $I(\mathbf{k})$ were to have the value appropriate to a thermodynamic radiation field at temperature T, $\bar{q}(\mathbf{k})$ would follow Planck's law,

$$\bar{q}(\mathbf{k})\equiv\bar{q}(k)=\frac{1}{\exp(\hbar k/T)-1} \tag{3.95}$$

in which the temperature scale has been chosen so that the Boltzmann constant is unity.

The delta function in (3.92) gives an asymmetry between emission and absorption. Spontaneous transitions are downward because we have taken the absorber particles as being in their ground levels, $E_i\le E_f$ for all f. We see therefore that the asymmetry of spontaneous emission arises from the assumption of a cold universe. We shall return to this point at the end of this section.

D. The complete influence functional and the level shift formula

The complete influence functional was given by (3.65), of which the two terms in the last line remain to be discussed:

$$-(1/2\hbar^2)\int\int\int S_I^2[\mathbf{a}(t),\mathbf{b}(t)]\exp\{iS_0[\mathbf{b}(t)]/\hbar\}\psi_i^*(\mathbf{b}_f)\psi_i(\mathbf{b}_i)d^3\mathbf{b}_f d^3\mathbf{b}_i\mathcal{D}^3\mathbf{b}(t),$$

$$-(1/2\hbar^2)\int\int\int S_I^2[\mathbf{a}'(t),\mathbf{b}'(t)]\exp\{-iS_0[\mathbf{b}'(t)]/\hbar\}\psi_i(\mathbf{b}'_f)\psi_i^*(\mathbf{b}'_i)d^3\mathbf{b}'_f d^3\mathbf{b}'_i\mathcal{D}^3\mathbf{b}'(t). \tag{3.96}$$

It will be sufficient to work out the first of these terms, since the second can then be written down by inspection.

Inserting (3.72) for $\mathbf{A}_{\rm ret}^{(a)}$ in the expression (3.64) for S_I leads to

$$-(e^2/2\hbar^2)\sum_{l=-\infty}^{\infty}\int\int\int\int_0^{T'}(\mathbf{A}_l\cdot\dot{\mathbf{b}})e^{-2\pi ilt'/T}dt'$$

$$\times\int_0^{T'}(\mathbf{A}_l^*\cdot\dot{\mathbf{b}})e^{2\pi ilt'/T}dt'\exp\{iS_0[\mathbf{b}(t)]/\hbar\}\psi_i^*(\mathbf{b}_f)\psi_i(\mathbf{b}_i)d^3\mathbf{b}_f d^3\mathbf{b}_i\mathcal{D}^3\mathbf{b}(t). \tag{3.97}$$

Using ordinary perturbation methods for second-order transitions, (3.97) can be reduced to

$$-(e^2/\hbar^2)\sum_{l=-\infty}^{\infty}\sum_g[(E_i-E_g)/\hbar]^2\left|\int\psi_i^*(\mathbf{b})(\mathbf{b}\cdot\mathbf{A}_l)\psi_g(\mathbf{b})d^3\mathbf{b}\right|^2\int_0^{T'}\exp\left(\frac{i}{\hbar}[E_i-E_g-(2\pi l\hbar/T')]t'\right)dt'$$

$$\times\int_0^{t'}\exp\left(\frac{i}{\hbar}[E_0-E_i+(2\pi l\hbar/T')]\tilde{t}'\right)d\tilde{t}', \tag{3.98}$$

where the summation with respect to g is over all intermediate states of b.

Now insert (3.78) for $\mathbf{A}_l$ in (3.98), to give

$$(e^2/3\hbar^2R^2T'^2)\sum_g\left(\frac{E_g-E_i}{\hbar}\right)^2\sum_{l=-\infty}^{\infty}|\mathbf{b}_{ig}(2\pi l/T')|^2e^{-\tau_l}\int_0^{T'}\exp\left[(i/\hbar)\left(E_i-E_g-\frac{2\pi\hbar l}{T'}\right)t'\right]dt'$$

$$\times\int_0^{t'}\exp\left[(i/\hbar)\left(E_g-E_i+2\pi\hbar l/T'\right)\tilde{t}'\right]d\tilde{t}'\sum_{j=1,2}\int_0^T\alpha^{(j)}\cdot\dot{\mathbf{a}}\exp\left[(2\pi il/T')\left(t-\frac{\mathbf{a}\cdot\mathbf{R}}{R}\right)\right]dt$$

$$\times\int_0^T\alpha^{(j)}\cdot\dot{\mathbf{a}}\exp\left[-(2\pi il/T')\left(\tilde{t}-\frac{\mathbf{a}\cdot\mathbf{R}}{R}\right)\right]d\tilde{t} \tag{3.99}$$

after averaging with respect to the orientation of b.

Although (3.99) appears complicated we shall find simplifications. For fixed ψ_g, the main contributions come from $2\pi/\hbar/T'\cong E_i-E_g$. Only when this condition is satisfied can the integrals in the second line of (3.99) yield a contribution that behaves as T'^2. Defining

$$\mathbf{k}=\frac{E_i-E_g}{\hbar}\cdot\frac{\mathbf{R}}{R}, \tag{3.100}$$

we can therefore write

$$-\frac{e^2}{3\hbar^2R^2T'^2}\sum_g k^2|\mathbf{b}_{ig}(k)|^2\cdot e^{-\tau(k)}\int_0^{T'}e^{-ikt'}dt'\int_0^{t'}e^{ik\tilde{t}'}d\tilde{t}'$$

$$\times\sum_{j=1,2}\int_0^T\alpha^{(j)}\cdot\dot{\mathbf{a}}e^{-i\mathbf{k}\cdot\mathbf{a}}dt\int_0^T\alpha^{(j)}\cdot\dot{\mathbf{a}}e^{i\mathbf{k}\cdot\mathbf{a}}d\tilde{t}\sum_{l=-\infty}^{\infty}\exp\left[\frac{2\pi il}{T'}(t-\tilde{t}-t'+\tilde{t}')\right]. \tag{3.101}$$

Furthermore

$$\sum_{l=-\infty}^{\infty}\exp\left[\frac{2\pi il}{T'}(t-\tilde{t}-t'+\tilde{t}')\right]=T'\delta(t-\tilde{t}-t'+\tilde{t}'). \tag{3.102}$$

We must therefore have $t \geq \tilde{t}$ since $t' \geq \tilde{t}'$, and (3.101) becomes

$$-\frac{e^2}{3\hbar^2 R^2}\sum_g k^2 e^{-\tau(k)}|\mathbf{b}_{ig}(k)|^2 \sum_{j=1,2}\int_0^T \alpha^{(j)}\cdot\dot{\mathbf{a}}e^{-i\mathbf{k}\cdot\mathbf{a}-ikt}dt\int_0^t \alpha^{(j)}\cdot\dot{\mathbf{a}}e^{i\mathbf{k}\cdot\mathbf{a}+ik\tilde{t}}d\tilde{t}. \tag{3.103}$$

This contribution to $F^{(b)}$ must be added to (3.88).

As before it is sufficient to consider a single state g, since all states were automatically included in the discussion that followed (3.88). Indeed the summation with respect to all absorbers proceeds exactly as before, and in place of (3.87) we now have

$$\begin{aligned}F[\mathbf{a}(t),\mathbf{a}'(t)] = \exp\Bigg[\frac{e^2}{4\pi^2\hbar}\int d\Omega\int_0^\infty kdk\sum_{j=1,2}\Bigg\{&\int_0^T(\alpha_{\mathbf{k}}^{(j)}\cdot\dot{\mathbf{a}})e^{-i\mathbf{k}\cdot\mathbf{a}+ikt}dt\int_0^T(\alpha_{\mathbf{k}}^{(j)}\cdot\dot{\mathbf{a}}')e^{i\mathbf{k}\cdot\dot{\mathbf{a}}'-ikt'}dt'\\ &-\int_0^T(\alpha_{\mathbf{k}}^{(j)}\cdot\dot{\mathbf{a}})e^{-i\mathbf{k}\cdot\mathbf{a}-ikt}dt\int_0^t(\alpha_{\mathbf{k}}^{(j)}\cdot\dot{\mathbf{a}})e^{i\mathbf{k}\cdot\dot{\mathbf{a}}+ik\tilde{t}}d\tilde{t}\\ &-\int_0^T(\alpha_{\mathbf{k}}^{(j)}\cdot\dot{\mathbf{a}}')e^{-i\mathbf{k}\cdot\mathbf{a}'+ikt}dt\int_0^t(\alpha_{\mathbf{k}}^{(j)}\cdot\dot{\mathbf{a}}')e^{i\mathbf{k}\cdot\dot{\mathbf{a}}'-ik\tilde{t}}d\tilde{t}\Bigg\}\Bigg],\end{aligned} \tag{3.104}$$

in which we have also included the second term of (3.96).

The expression (3.104) is an influence functional and it obeys the general rules discussed by Feynman and Hibbs [1965, Eq. (99), p. 348]. If we identify the paths, $\mathbf{a}(t) \equiv \mathbf{a}'(t)$, the exponent in (3.104) vanishes. The paths do not act on themselves via the response of the Universe.

The new terms in (3.104) have no effect on the calculation of $P(m \to n), m \neq n$, but they are necessary to obtain $P(m \to m)$. We now show that

$$P(m \to m) = 1 - \sum_{n\neq m} P(m \to n). \tag{3.105}$$

$P(m \to m)$ is given by writing m for n in (3.31). Expanding the exponential in (3.104) to first order we have

$$\begin{aligned}P(m \to m) = 1 &- (e^2/4\pi^2\hbar)\int\int\int \phi_m^*(\mathbf{a}_f)\exp\left\{\frac{i}{\hbar}S_0[\mathbf{a}(t)]\right\}\phi_m(\mathbf{a}_i)\int d\Omega\int_0^\infty kdk\\ &\times\sum_{j=1,2}\int_0^T(\dot{\mathbf{a}}\cdot\alpha_{\mathbf{k}}^{(j)})e^{-i\mathbf{k}\cdot\mathbf{a}-ikt}dt\int_0^t(\dot{\mathbf{a}}\cdot\alpha_{\mathbf{k}}^{(j)})e^{i\mathbf{k}\cdot\mathbf{a}+ik\tilde{t}}d\tilde{t}\mathcal{D}^3\mathbf{a}(t)d^3\mathbf{a}_f d^3\mathbf{a}_i\\ &-(e^2/4\pi^2\hbar)\int\int\int \phi_m(\mathbf{a}'_f)\exp\left\{-\frac{i}{\hbar}S_0[\mathbf{a}'(t)]\right\}\phi_m^*(\mathbf{a}'_i)\int d\Omega\int_0^\infty kdk\\ &\times\sum_{j=1,2}\int_0^T(\dot{\mathbf{a}}'\cdot\alpha_{\mathbf{k}}^{(j)})e^{-i\mathbf{k}\cdot\mathbf{a}'+ikt}dt\int_0^t(\dot{\mathbf{a}}'\cdot\alpha_{\mathbf{k}}^{(j)})e^{i\mathbf{k}\cdot\mathbf{a}'-ik\tilde{t}}d\tilde{t}\mathcal{D}^3\mathbf{a}'(t)d^3\mathbf{a}'_f d^3\mathbf{a}'_i.\end{aligned} \tag{3.106}$$

The term involving $\mathbf{a}(t)$ separates from the term involving $\mathbf{a}'(t)$. The path integrals are not hard to evaluate. We obtain

$$\begin{aligned}P(m \to m) = 1 &- (e^2/4\pi^2\hbar)\sum_{j=1,2}\int d\Omega\int_0^\infty kdk\sum_n[(E_n-E_m)/\hbar]^2\left|\int\phi_m^*(\mathbf{a})\alpha_{\mathbf{k}}^{(j)}\cdot\mathbf{a}e^{-i\mathbf{k}\cdot\mathbf{a}}\phi_n(\mathbf{a})d^3\mathbf{a}\right|^2\\ &\times\int_0^T dt\int_0^t\left[\exp\left\{\frac{i}{\hbar}(E_m-E_n-\hbar k)(t-\tilde{t})\right\}+\exp\left\{\frac{i}{\hbar}(E_m-E_n-\hbar k)(\tilde{t}-t)\right\}\right]d\tilde{t}.\end{aligned} \tag{3.107}$$

The term in $\exp\{i/\hbar(E_m-E_n-\hbar k)(t-\tilde{t})\}$ in (3.107) comes from the quadratic term in the path $\mathbf{a}(t)$ in $F[\mathbf{a}(t),\mathbf{a}'(t)]$, and the term in $\exp[i/\hbar(E_m-E_n-\hbar k)(\tilde{t}-t)]$ comes from the quadratic term in $\mathbf{a}'(t)$.

The last two integrals of (3.107) give

$$2\int_0^T\frac{\sin[(E_m-E_n)/\hbar-k]t}{(E_m-E_n)/\hbar-k}dt \cong 2\pi T\delta\left(k-\frac{E_m-E_n}{\hbar}\right). \tag{3.108}$$

Hence (3.107) is just $1-\sum_{E_n<E_m}P(m\to n)$. Since $P(m\to n)$ is zero for $E_n > E_m$ we obtain (3.105). Probability is therefore conserved.

The system a is by hypothesis in the state m at $t=0$. The effect of the response of the universe is to change the

amplitude for the system to be in the state m at time T from $\phi_m \exp(-iE_mT/\hbar)$ to

$$\phi_m \exp\left[-\frac{i}{\hbar}(E_m+\Delta E_m)-\frac{1}{2}\gamma\right]T \cong \left(1-\frac{\gamma T}{2}-\frac{i\Delta E_m T}{\hbar}\right)\phi_m e^{iE_mT/\hbar} \tag{3.109}$$

for T not too large. The probability of the system being in the state m at time T is therefore

$$\begin{aligned} P(m\to m) &\cong \left(1-\frac{\gamma T}{2}-\frac{i\Delta E_m T}{\hbar}\right)\left(1-\frac{\gamma T}{2}+\frac{i\Delta E_m T}{\hbar}\right) \\ &= 1-\left(\frac{\gamma}{2}+\frac{i\Delta E_m}{\hbar}\right)T-\left(\frac{\gamma}{2}-\frac{i\Delta E_m}{\hbar}\right)T+O(\Delta E_m^2). \end{aligned} \tag{3.110}$$

Suppose we identify the second and third terms on the right-hand side of (3.110) with the second and third terms on the right-hand side of (3.106). Then

$$\begin{aligned} \left(\frac{1}{2}\gamma+i\frac{\Delta E_m}{\hbar}\right)T = \frac{e^2}{4\pi^2\hbar}\int\int\int \phi_m^*(\mathbf{a}_f)\phi_m(\mathbf{a}_i)\exp\left\{\frac{i}{\hbar}S_0[\mathbf{a}(t)]\right\}\int d\Omega\int_0^\infty kdk \\ \times\sum_{j=1,2}\int_0^T \dot{\mathbf{a}}\cdot\alpha_{\mathbf{k}}^{(j)}e^{-i\mathbf{k}\cdot\mathbf{a}-ikt}dt\int_0^t \dot{\mathbf{a}}\cdot\alpha_{\mathbf{k}}^{(j)}e^{i\mathbf{k}\cdot\mathbf{a}+ik\tilde{t}}d\tilde{t}\mathcal{D}^3\mathbf{a}(t)d^3\mathbf{a}_f d^3\mathbf{a}_i. \end{aligned} \tag{3.111}$$

Care is needed in evaluating the path integral because a "crossover" term arises at $t=\tilde{t}$. This term yields

$$\frac{ie^2T}{\pi m}\int_0^\infty kdk, \tag{3.112}$$

while the main term in the reduction is

$$\begin{aligned} &(e^2/4\pi^2\hbar)\sum_{j=1,2}\int d\Omega\int_0^\infty kdk\sum_n[(E_n-E_m)/\hbar]^2\left|\int\phi_n^*(\mathbf{a})\mathbf{a}\cdot\alpha_{\mathbf{k}}^{(j)}e^{-i\mathbf{k}\cdot\mathbf{a}}\phi_m(\mathbf{a})d^3\mathbf{a}\right|^2 \\ &\quad\times\int_0^T dt\int_0^t \exp\left\{\frac{i}{\hbar}(E_m-E_n-\hbar k)(t-\tilde{t})\right\}d\tilde{t} \\ &\quad=\frac{1}{2}\sum_{E_n<E_m}P(m\to n)+(ie^2/4\pi^2\hbar)\sum_{j=1,2}\int d\Omega\int_0^\infty kdk \\ &\quad\times\sum_n[(E_n-E_m)/\hbar]^2\left|\int\phi_n^*(\mathbf{a})\mathbf{a}\cdot\alpha_{\mathbf{k}}^{(j)}e^{-i\mathbf{k}\cdot\mathbf{a}}\phi_m(\mathbf{a})d^3\mathbf{a}\right|^2\int_0^T dt\int_0^t \sin[(E_m-E_n-\hbar k)(t-\tilde{t})/\hbar]d\tilde{t}. \end{aligned} \tag{3.113}$$

Collecting terms

$$\begin{aligned} \frac{\gamma}{2}+\frac{i}{\hbar}\Delta E_m = \frac{1}{2T}\sum_{E_n<E_m}P(m\to n)+\frac{ie^2}{\pi m}\int_0^\infty kdk+\frac{ie^2}{4\pi^2}\sum_n\left(\frac{E_m-E_n}{\hbar}\right)^2\sum_{j=1,2}PP\int_0^\infty\frac{kdk}{E_m-E_n-\hbar k}\int d\Omega \\ \times\left|\int\phi_n^*(\mathbf{a})\alpha_{\mathbf{k}}^{(j)}\cdot\mathbf{a}e^{-i\mathbf{k}\cdot\mathbf{a}}\phi_m(\mathbf{a})d^3\mathbf{a}\right|^2. \end{aligned} \tag{3.114}$$

The summation in the last term is not restricted to $E_n<E_m$. The divergent term on the right-hand side does not, however, survive in the expression (3.110).

E. The radiation cutoff at the absorber

The cosmological features of the response of the universe was discussed in Sec. II.D, where Table II gave results for the classical electromagnetic theory. Of the cosmological models discussed in the literature only the steady-state and quasi-steady-state theories met the requirements of the absorber theory, and the same may be anticipated in the present quantum version of the absorber theory. There may well be other cosmological models still not examined in detail that also meet the same requirements. Indeed any model with a proper density of matter that does not fall below some nonzero lower limit along the future light cone is expected to satisfy the required response condition. Here we consider a further feature of the absorber theory using the steady-state model as the simplest example of such a class of models.

The work of Secs. III.B–III.D was in flat spacetime. This work can be taken over to cosmological spaces of the Robertson-Walker type by making a conformal transformation of the latter, as in Sec. II.C.3. There is, however, the slight notational problem that in Secs. III.B–III.D the time was denoted by t, whereas in II.C.3 the time t was in the cosmological space and τ in the flat conformal space. To retain the explicit formulas of III.B–III.D we therefore invert t and τ as they appeared in II.C.3. For an observer at $t = 0, r = 0$ the line element of the steady-state model then takes the form

$$ds^2 = (1 - Ht)^{-2}[dt^2 - dr^2 - r^2(d\theta^2 + \sin^2\theta d\phi^2)], \tag{3.115}$$

where $H \simeq 3.10^{-18}$ sec^{-1} is a constant of the theory.

In the spontaneous transition $E_m \to E_n$ of the local system, the important frequencies involved in the influence functional (3.104) are those in the neighborhood of

$$k = (E_m - E_n)/\hbar. \tag{3.116}$$

This frequency has to be matched by the value of $(E_f - E_i)/\hbar$ for the absorber transition $\psi_i \to \psi_f$. If, however, we wish to consider E_i, E_f, with respect to proper time at the absorber, it is necessary to take account of the redshift effect of the expansion of the universe. Thus when a wave, starting with frequency k at the source, reaches an absorber at coordinate r, the proper frequency has become

$$\omega = (1 - Hr)k, \tag{3.117}$$

and E_i, E_f, with respect to local proper time are related to k by

$$E_f - E_i = \hbar k(1 - Hr). \tag{3.118}$$

If we displace the absorber by a proper distance dl away from the source, the frequency in tune with the wave changes by $d\omega$, where

$$d\omega = -Hkdr, \qquad dr = (1 - Hr)dl. \tag{3.119}$$

The second relation in (3.119) follows from the conformal transformation. Hence we get

$$\frac{d\omega}{\omega} = -Hdl. \tag{3.120}$$

Suppose the absorbers are effective over a range $[\omega_{\min}, \omega_{\max}]$ of frequencies. Then if $k > \omega_{\max}$, the range of l over which the absorbers can be in tune with k is

$$l = \int dl = H^{-1}\int_{\omega_{\min}}^{\omega_{\max}} \frac{d\omega}{\omega} = H^{-1}\ln\frac{\omega_{\max}}{\omega_{\min}}. \tag{3.121}$$

For complete absorption, we need $l \to \infty$. Hence we need a process with $\omega_{\min} \to 0$. Collisional absorption provides such a process, and will be used from hereon, although any process with $\omega_{\min} \to 0$ will suffice equally well.

At the low frequencies at which the main collisional absorption occurs it is sufficient to use classical considerations, according to which the alteration of a wave of frequency ω over a length dl caused by collisions with frequency ν is

$$\exp\left[-\frac{2\pi\nu Ne^2}{m\omega^2}dl\right], \tag{3.122}$$

where N is the number density of electrons, e the electronic charge, and m the electronic mass. Using (3.117) and (3.119), the damping produced over the coordinate range $0 \le r \le R$ is

$$\exp\left[-\frac{2\pi\nu Ne^2}{mk^2}\int_0^R \frac{dr}{(1 - Hr)^3}\right]. \tag{3.123}$$

For large R, appreciable damping occurs when

$$(1 - Hr)^2 \simeq \frac{2\pi\nu Ne^2}{mk^2H}, \tag{3.124}$$

which corresponds to an effective proper frequency

$$\omega_{\text{eff}}^2 \simeq \frac{2\nu Ne^2}{mH}. \tag{3.125}$$

For ionized hydrogen, ν at ω_{eff} is given by

$$\nu = 2\pi Nv\left(\frac{e^2}{mv^2}\right)^2 \ln\left(\frac{mv^2}{\hbar\omega_{\text{eff}}}\right), \tag{3.126}$$

where v is a typical electron veleocity. Taking $H^{-1} \sim 3.10^{17}$ sec, $N \sim 10^{-9}$ cm^{-3}, $v = 1/300$ (of velocity of light), and substituting for $e, m, \hbar$ in (3.125) and (3.126) we can solve for ω_{eff} and ν,

$$\omega_{\text{eff}} \simeq 80 \text{ sec}^{-1}, \quad \nu = 1.3 \times 10^{-14} \text{ sec}^{-1}. \tag{3.127}$$

Thus a wave with frequency greater than 10^2 sec^{-1} is first redshifted to $\sim 10^2$ sec^{-1} and then absorbed. The effective absorption takes place over a proper distance of the order of 10^{28} cm. A wave with frequency less than 10^2 sec^{-1} is absorbed without having to be redshifted.

From (3.127) we note that the dimensionless parameter

$$\frac{4\pi Ne^2}{m\omega_{\text{eff}}^2} \simeq 5.10^{-4}, \tag{3.128}$$

while the real part n of the refractive index given by $1 - 2\pi Ne^2/m\omega_{\text{eff}}^2$ is a little different from unity, showing that ω_{eff} is appreciably larger than the plasma frequency $(4\pi Ne^2/m)^{1/2}$, showing also that the wave is effectively

absorbed well before the redshift causes the frequency to fall so low that the real part of the refractive index becomes negative, with the wave no longer able to propagate.

From this deduction two important inferences can be made. One explains the situation at (3.79) where random phasing was used to average cross products of A_l and $A_{l'}, l \neq -l'$, to zero. Consider two waves with slightly different values of the initial frequency, k and $k + dk$. From (3.124), absorption occurs at r and $r + dr$ with

$$Hdr = \omega_{\text{eff}} dk/k^2. \tag{3.129}$$

From the line element (3.115) the proper distance associated with this coordinate displacement is $(1 - Ht)^{-1}dr$, and with $r = t$ along a light ray, this is the same as $(1 - Hr)^{-1}dr$. Using (3.129) and $\omega_{\text{eff}} = (1 - Hr)k$ this proper distance is

$$H^{-1}dk/k \simeq 3 \times 10^{17} dk/k \text{ sec } \simeq 10^{28} dk/k \text{ cm}. \tag{3.130}$$

For $k \simeq 10^{15}$ sec^{-1} at optical frequencies this is a large distance even for dk as small as 1 sec^{-1}. The proper distance between the particles absorbing the wave of frequency k and that of frequency $k + dk$ is very large compared with the interparticle distance. The responses to k and to $k + dk$ are therefore uncorrelated in phase and the cancelling of waves with $l' = -l$ at (3.79) is explained. The same issue arises in the usual version of quantum mechanics. How is random phasing between the degrees of freedom of the quantized field to be understood? So far as we are aware it is assumed and not understood.

The second inference that will assume importance in Sec. V is that a frequency cutoff now emerges cosmologically (see Hoyle and Narlikar, 1993).

In the above discussion we wrote that emission at $r = 0$ occurs for $t \simeq 0$, not $t = 0$, since some finite interval of t is needed for emission to take place. The duration of emission for spontaneous emission was $0 \leq t \leq T$ with $T \gg k^{-1}$, this condition being necessary to establish the delta-function property at (3.40). The early oscillations emitted immediately after $t = 0$ are absorbed by particles at $t = r$ with

$$r = H^{-1}(1 - \omega_{\text{eff}}/k), \tag{3.131}$$

while the later oscillations emitted at $t = T$ are absorbed at $t = r$ with

$$r = T + H^{-1}(1 - \omega_{\text{eff}}/k). \tag{3.132}$$

Since, however, the cosmological model has an horizon cutoff at $t = H^{-1}$ we cannot have $t = r$ with r given by (3.132) unless

$$k < \omega_{\text{eff}} H^{-1}/T. \tag{3.133}$$

For $\omega_{\text{eff}} = 80$ sec^{-1}, $H^{-1} = 3.10^{17}$ sec, $T = 10^{-12}$ sec as an example, the cutoff on k is at 2.4×10^{31} sec^{-1}. Above this limit the response condition fails. It does so for the same value of k for every observer in the cosmological Hubble flow, provided every observer examines the same radiation process with the same T.

We have thus deduced the existence of a very high frequency cutoff. It has no practical relevance for the calculation of transition probabilities. But at a later stage, when we come to the radiative correction process in relativistic quantum electrodynamics, it will be of critical importance to the logical development of the theory. It avoids the infinities of quantum electrodynamics.

IV. RELATIVISTIC QUANTUM ELECTRODYNAMICS

A. Introduction

The work of the preceding section has amply demonstrated that the Wheeler-Feynman absorber theory of radiation can be extended into the quantum domain. The explanation of the phenomenon of spontaneous transition was hitherto considered to demand a quantum field theory. The degrees of freedom vested in the electromagnetic field make the quantum vacuum nontrivial and therefore the atomic electron is supposed to jump down the energy ladder even in the absence of the external field, because it interacts with the field vacuum.

In the action-at-a-distance picture, the role of the vacuum is taken over by the response of the Universe. The calculation of the transition probability in the previous section has shown that provided we live in the right type of universe (perfect future absorber and imperfect past absorber) the answer comes out right. Thus we could argue that a quantum field theory is sufficient but not necessary for understanding spontaneous transition. Can action at a distance replace quantum field theory altogether just as it can demonstrably replace the classical field theory?

The answer is not immediately obvious. First of all, spontaneous transition represents but the tip of the iceberg of phenomena coming under the purview of quantum field theory. What about phenomena like Compton scattering, pair creation and annihilation, vacuum polarization, self-energy effects like the Lamb shift, etc.? Unless the full gamut of quantum electrodynamic results are described by the action at a distance we cannot look upon it as a viable alternative to field theory.

In this section we will attempt to review the progress made in the above direction. This necessitates, however, the development of a formalism for path integrals for relativistic particles. First we will discuss the motion of a single relativistic fermion, i.e., a Dirac particle in terms of Feynman's path integral formalism suitably extended. We will then discuss a system containing many fermions interacting via the electromagnetic action at a distance *a la* Fokker. This will bring in the response of the universe and the quantum analog of the Dirac formula [cf. Eq.

(2.3)] for radiative reaction. We will then consider the issues relating to self-action and renormalization.

B. The motion of a Dirac particle

1. The nonrelativistic propagator

We first summarize the path integral approach from Sec. III in a form that will be adaptable to the relativistic case.

Consider the motion of an electric charge a with mass m_a and charge e_a moving freely and nonrelativistically from a spacetime point 1 to a spacetime point 2. Let us assume that the spacetime coordinates of 1 and 2 are respectively $(\mathbf{a}_1, t_1)$ and $(\mathbf{a}_2, t_2)$. In Newtonian mechanics this particle would move along a definite path Γ_c in spacetime connecting point 1 to point 2. This is the path of zero acceleration. In quantum mechanics, as discussed in the last section, there is no such unique path but a whole range of paths Γ all starting at 1 and ending at 2. The overall motion of the particle from 1 to 2 is described by a propagator $K[2;1]$ that is obtained by summing the probability amplitudes along all the paths according to the formula:

$$K[2;1] = \int \exp\{iJ[\Gamma]\}\mathcal{D}\Gamma \tag{4.1}$$

where $J[\Gamma]$ is the classical action computed for path Γ. The path integral can be evaluated and the answer is

$$K[2;1] = \left[\frac{m_a}{2\pi i(t_2 - t_1)}\right]^{3/2} \times \exp\left\{\frac{im_a(\mathbf{a}_2 - \mathbf{a}_1)^2}{2(t_2 - t_1)}\right\}\theta(t_2 - t_1). \tag{4.2}$$

Here θ is the Heaviside function. [The reader is reminded that we have taken $\hbar = 1, c = 1$.] The propagator $K[2;1]$ satisfies the well-known Schrödinger equation

$$\left[\frac{\partial}{\partial t_2} + \frac{1}{2im_a}\nabla_2^2\right]K[2;1] = \delta_4(2;1). \tag{4.3}$$

Instead of using (4.1) we could proceed in the following way. Along the typical path Γ mark points X_i, $i = 0, 1, 2, \ldots, N$, with the end points 1 and 2 corresponding to X_0 and X_N, respectively. With N sufficiently large, we can consider a typical segment X_iX_{i+1} as infinitesimal. Then define $P(\Gamma)$ as the product

$$P(\Gamma) = \prod_{i=1}^{N} A_i^{-1}K[X_i; X_{i-1}], \tag{4.4}$$

where the A_i are normalizing constants. Proceeding to a limit as $N \to \infty$, we can recover the expression (4.1). In a sense Eq. (4.1) is the inverse of Eq. (4.4). For details we refer the reader to Feynman's original paper (1949) or the classic book by Feynman and Hibbs (1965).

Another useful result relates the particle propagator to the complete set $\{u_n\}$ of normalized stationary eigensolutions of the homogeneous Schrödinger equation

$$K[2;1] = \sum_n u_n(2)\bar{u}_n(1)\theta(t_2 - t_1). \tag{4.5}$$

Note that the propagator is time asymmetric in the sense that it assigns a zero probability amplitude for motion backwards in time while the full set of eigensolutions is used to describe the amplitude for forward propagation. In the discussion of the relativistic motion to be considered next, this aspect undergoes a serious modification.

2. The relativistic free particle

The classical action for a relativistic particle of rest mass m_a is given by

$$J = -\int m_a da. \tag{4.6}$$

This action, however, does not describe a fermion like an electron because it contains no information on the intrinsic spin. Rather than look for a classical action containing this information we will follow the alternative procedure of Eq. (4.4), for we already know that the wave equation generalizing Eq. (4.3) is the Dirac equation which for the propagator $K[2;1]$ becomes

$$(\not{\nabla}_2 + im_a)K_0[2;1] = \delta_4(2,1). \tag{4.7}$$

For reasons to be made clear shortly, we have distinguished the propagator by a suffix 0. In analogy with the nonrelativistic limit, $K_0[2;1]$ is expected to satisfy the temporal condition

$$K_0[2;1] = 0 \quad \text{for } t_2 < t_1. \tag{4.8}$$

However, here we run into the well-known problem of negative energy states. If we express the solution of Eq. (4.7) in terms of an expression like (4.5) we find that the complete set $\{u_n\}$ has to include negative energy solutions also. Dirac had sought to get around the problem via the "hole theory," thus effectively converting it into a many-particle problem.

Feynman (1949) got around the problem of negative energy states by redefining the propagator solution of Eq. (4.7) in the following way:

$$K_+[2;1] = \begin{cases} \sum\limits_{E_n>0} u_n(2)\bar{u}_n(1), & t_2 > t_1, \\ \sum\limits_{E_n<0} u_n(2)\bar{u}_n(1), & t_2 < t_1. \end{cases} \tag{4.9}$$

Notice that the propagator $K_+[2;1]$ allows propagation

backwards as well as forward in time. The positive energy states ($E_n > 0$) contribute to forward propagation while the negative energy states ($E_n < 0$) contribute to backward propagation. In standard language we say that a negative energy electron going backward in time corresponds to a positive energy positron going forward in time. We shall refer to $K_+[2;1]$ as the Feynman propagator.

Let us now see how all this affects our definition of the probability amplitude if we seek to generalize Eq. (4.4) with the Feynman propagator K_+ replacing K_0. One difference is immediately apparent. In the nonrelativistic case a path that turned backwards in time would automatically have zero probability amplitude. Not so any more. We could now have paths as shown in Fig. 4 with nonzero amplitudes.

How do we interpret these paths? In principle they tell us that a forward (in time) moving electron can be scattered to go backwards in time and vice versa. This with Feynman's reinterpretation corresponds to the phenomena of pair annihilation or creation which can happen with or without any external electromagnetic disturbance. We will next consider how these cases are to be looked at in the action-at-a-distance picture.

C. Motion in an external potential

1. The perturbation expansion

Given the Feynman propagator for a free particle a, we next ask for a quantum-mechanical description of a charge e_a moving under the external electromagnetic potential A_i of other electric charges. Instead of the propagator of Eq. (4.9) we now have another denoted by $K_+^A[2;1]$ which satisfies the equation

$$[\nabla_2 + ie_a \not{A}(2) + im_a]K_+^A[2;1] = \delta_4(2,1). \tag{4.10}$$

Both in Eqs. (4.7) and (4.10) a suitable limiting process is used to define the derivatives of the K_0 or K_+ propagators.

The result proved by Feynman, using second quantization of the particle wave function [cf. Feynman 1949, Appendix], was that we can ignore the hole theory in an amplitude calculation provided we use $K_+^A[2;1]$ and multiply the amplitude by C_v, the amplitude for a vacuum to remain a vacuum. C_v is given by

$$C_v = \exp(-L), \quad L = \sum_{n\geq 2} L^{(n)}, \tag{4.11}$$

where $L^{(n)}$ is the amplitude for the occurrence of a closed loop in which A_i acts n times:

$$L^{(n)} = \frac{(-ie_a)^n}{n}\int\cdots\int T_r\{K_+[n;1]\not{A}(1)K_+[1;2]\not{A}(2) \times\cdots K[n-1;n]\not{A}(n)\}d\tau_1\cdots d\tau_n. \tag{4.12}$$

It can be shown via Furry's theorem that $L^{(n)}$ vanishes for odd n [cf. Bjorken and Drell, 1965, for example].

Can we use the above result in the action-at-a-distance framework? Not directly, since Feynman used field theory to arrive at it, i.e., he had to use quantization of the fermion fields. However, it is possible to rederive the result without recourse to second quantization, as was shown by Hoyle and Narlikar (1971). We briefly describe this work.

Like the propagator K_0 also defines another K_0^- by the relation

$$K_0^-[2;1] = -\theta(t_1 - t_2)\sum_n u_n(2)\bar{u}_n(1). \tag{4.13}$$

Thus the $K_0^-[2;1]$ propagator describes motion backwards in time within the past light cone at point 1. To have a more symmetric notation, we will denote $K_0[2;1]$ of Eq. (4.7) by $K_0^+[2;1]$. Both $K_0^\pm$ satisfy the inhomogeneous Dirac equation (4.7). $K_0^+[2;1]$ is nonzero along the future light cone at point 1.

Corresponding to these two propagators, we also distinguish between two types of paths, Γ_{21}^+ going forward in time and Γ_{21}^- going backwards in time in going from point 1 to point 2. Now use formula (4.4) to define the amplitude $P(\Gamma_{21}^+)$ along Γ_{21}^+, while using a similar expression but with Γ_{21}^+ replaced by Γ_{21}^- to define the amplitude along Γ_{21}^-. The path integrals corresponding to these definitions ensure that

$$K_0^\pm[2;1] = \int P(\Gamma_{21}^\pm)\mathcal{D}\Gamma_{21}^\pm. \tag{4.14}$$

Now suppose that we have a free particle in the four-dimensional spacetime slab $t_1 \leq t \leq t_2$ and assume that the amplitude for the particle to reach $t = t_1$ from its previous history is given by $\psi_+(1)$. Likewise at $t = t_2$ we denote the amplitude for the particle to come from the future $t > t_2$ by $\psi_-(2)$. Thus the probability amplitude for the particle to be at an interior point 3 is given by $\psi(3)$ which is the sum of amplitudes along all Γ_{31}^+ type paths coming from points 1 on the time section $t = t_1$ weighted by $\psi(1)$ together with amplitudes along all Γ_{32}^- paths from points 2 on the time section $t = t_2$ to the

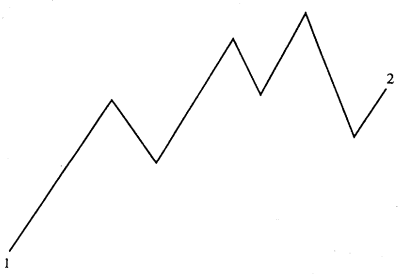

FIG. 4. A path from 1 to 2 could have reversals in time with each segment (backward or forward) having a nonzero probability amplitude.

point 3 again weighted by $\psi(2)$:

$$\psi(3) = \int K_0^+[3;1]\gamma_4\psi_+(1)d^3\mathbf{x}_1 - \int K_0^-[3;2]\gamma_4\psi_-(2)d^3\mathbf{x}_2. \quad (4.15)$$

We impose as a boundary condition that $\psi_+(1)$ is made of positive energy solutions while $\psi_-(2)$ is made of negative energy solutions. Then Eq. (4.15) is equivalent to the relation

$$\psi(3) = \int K_+[3;1]\gamma_4\psi_+(1)d^3\mathbf{x}_1 - \int K_+[3;2]\gamma_4\psi_-(2)d^3\mathbf{x}_2. \quad (4.16)$$

Thus the Feynman propagator serves the convenient role of book keeping of positive and negative energy states, of how the former travel forward in time along paths Γ^+ and the latter backwards in time along paths Γ^-.

Consider next the external potential A_i acting on the particle which is such that A_i vanishes outside the slab. In that case we may have scattering of paths backwards and forward in time due to the potential. Here again the K_+^A propagator helps in the book-keeping process. Only, we need to invoke the perturbation expansion to keep track of how many times the particle has been scattered.

To begin with, the amplitude for a path Γ_{21}^+ or Γ_{12}^- in the slab is defined by

$$\begin{aligned} P^A(\Gamma_{21}^+) &= P(\Gamma_{21}^+)\exp\left[-ie\int_{\Gamma_{21}^+} A_i da^i\right], \\ P^A(\Gamma_{12}^-) &= P(\Gamma_{12}^-)\exp\left[-ie\int_{\Gamma_{12}^-} A_i da^i\right], \end{aligned} \quad (4.17)$$

with the understanding that point 1 is on $t = t_1$ and 2 on $t = t_2$. Paths within the slab need not be monotonic, however, with respect to t. Suppose we have a path from 1 to 2 with $2n$ reversals. Denoting intermediate points by i, sections i to $i+1$ are monotonic, and the amplitude for such a path is given by

$$P^A(\Gamma_{21}) = \prod_i P^A(\Gamma_{i,i-1}^{\pm}) \quad (4.18)$$

where the plus sign holds for the forward going sections and the minus sign for the backward going sections.

We return to a point that was taken for granted in the discussion so far. In the absence of an external potential A_i there would be no reversals, with the forward going paths continuing forward and likewise for the backward going paths. (Later we will reexamine this assumption in the light of the response of the Universe; but for the time being we will continue with it.) In the presence of A_i, however, reversals can occur, with paths starting at t_1 ending also at t_1. In this case we get a ψ_- at $t = t_1$ even though there may be no ψ_- at $t = t_2$. Similarly, we could have $\psi_+ \neq 0$ at t_2 originating solely from ψ_- at t_2. We therefore wish to know, what are ψ_+ on the time section $t = t_2$ and ψ_- on $t = t_1$ given a ψ_+ on $t = t_1$ and $\psi_- = 0$ on $t = t_2$?

Formally, we have

$$\psi_+^A(2) = \int\int P^A(\Gamma_{21})\gamma_4\psi_+(1)\mathcal{D}\Gamma_{21}d^3\mathbf{x}_1 \quad (4.19)$$

which includes paths with reversals. This expression can be calculated using the definitions given above and the standard path integral evaluation procedure [cf. Feynman and Hibbs, 1965]. The calculation is given in Hoyle and Narlikar (1971) and we simply quote the result

$$\psi_+^A(2) = \int K_+^A[2;1]\gamma_4\psi_+(1)d^3\mathbf{x}_1, \quad (4.20)$$

where

$$\begin{aligned} K_+^A[2;1] = K_+[2;1] &- ie\int K_+[2;3]\,\not{A}\,(3)K_+[3;1]d\tau_3 \\ &+(-ie)^2\int\int K_+[2;4]\,\not{A}\,(4)K_+[4;3] \\ &\times \not{A}\,(3)K_+[3;1]d\tau_3 d\tau_4 + \cdots. \end{aligned} \quad (4.21)$$

A similar analysis gives us $\psi_-^A(1')$:

$$\psi_-^A(1') = \int K_+^A[1';1]\gamma_4\psi_+(1)d^3\mathbf{x}_1, \quad (4.22)$$

where $K_+^A[1';1]$ is given by replacing point 2 in Eq. (4.21) by point $1'$ on the time section $t = t_1$.

Equation (4.21) is the perturbation expansion and is equivalent to Eq. (4.10). What about the factor C_v which appeared in the Feynman calculation? Does it have an analog in the particle sans field framework described here? We will consider this question next.

2. Vacuum loops

The answer to the above questions is provided by including the as yet ignored feature of antisymmetrization of the wave function. The basic concept of indistinguishability of identical particles in quantum mechanics intervenes in the above analysis in the following way.

In Fig. 5(a) we have two particles going forward in time: the path Γ_{31}^+ describes the motion of one particle from point 1 to point 3 while path Γ_{24}^+ describes the motion of the other particle from point 4 to point 2. However, when we want the amplitude for there to be a particle at 3 and at 2, given that each has a particle at the earlier epochs at points 1 and 4, the answer must take note of the exclusion principle, and is obtained by subtracting from the amplitude for Fig. 5(a) the amplitude for Fig. 5(b), wherein we have interchanged the final states with the particle at point 1 going to point 2 and

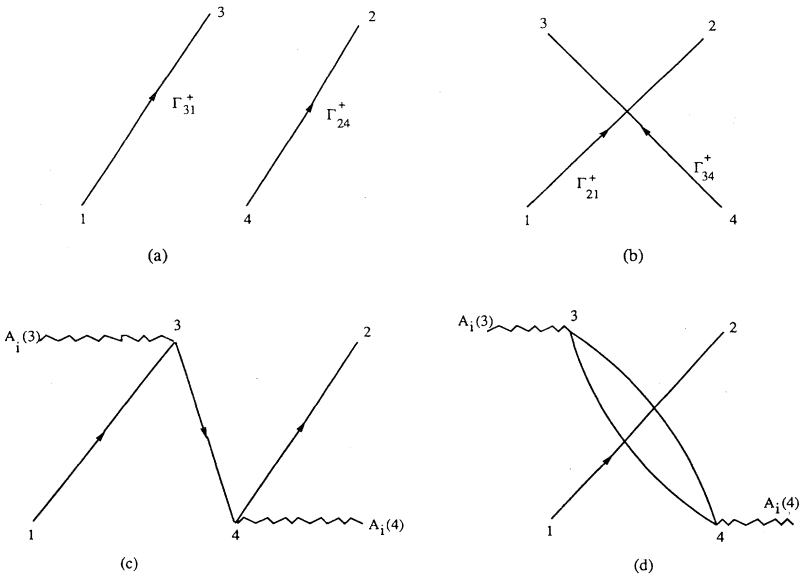

FIG. 5. Closed loops and antisymmetrization. (a) and (b) illustrate the result of antisymmetrization, which when applied to a double scattering process, shows how a closed loop arises in (d) from antisymmetrization of (c).

that at 4 going to 3.

What has been stated just now for a pair of particles also has relevance to the perturbation problem discussed above. In Fig. 5(c) we have a particle starting at point 1 and reaching a point 2 after being scattered twice by the potential A_i at points 3 and 4. Notice that the configuration of the Γ^+ paths in this figure is the same as in Fig. 5(a) and so our antisymmetrization criterion leads us to include also the amplitude for the path configuration of Fig. 5(d) that corresponds to that of Fig. 5(b). This last diagram, however, describes a particle moving unperturbed from point 1 to point 2, along with a closed loop which has double scattering at points 3 and 4.

For a quantitative description of this idea see Hoyle and Narlikar (1971). The result can be stated as follows. If to zeroth order the particle wave function at point 1 is $v_+(1)$ then to that order it is $v_+(2)$ at point 2. [The plus suffix denotes that it is made up of positive energy solutions.] Then corresponding to Fig. 5(d), we have to add to the amplitude the term

$$-\frac{1}{2}(-ie)^2 \int\int T_r\{K_+[3;4]\, \not{A}\,(4) K_+[4;3]\, \not{A}\,(3)\} \times d\tau_3 d\tau_4 v_+(2) \quad (4.23)$$

besides the second-order term from the diagram 5(c) obtained earlier from the perturbation expansion (4.21). Notice that apart from the unity term, the coefficient of $v_+(2)$ in Eq. (4.23) is the lowest-order term in the expansion of C_v. In fact, if we proceed further, using higher orders in the perturbation expansion following the antisymmetrization rule, we will recover the full factor C_v.

Hoyle and Narlikar (1971) have also shown that the quantity L in Eq. (4.11) is equal to the path integral over the loop amplitudes defined as follows:

$$L = \int P(\Gamma^0) \exp\left[-ie \oint_{\Gamma^0} A_i dl^i\right] \mathcal{D}\Gamma^0. \quad (4.24)$$

Here Γ^0 is a typical closed loop and the probability amplitude $P(\Gamma^0)$ is defined for a free loop by the same prescription as that of Eq. (4.4), but as before, the infinitesimal sections of it are propagated by $K_0^{\pm}$ rather than by K_+. Likewise the exponential factor in Eq. (4.24) denotes the influence of the external potential A_i on the loop Γ^0, following the same formula as Eq. (4.17).

The expression (4.12) can now be obtained from Eq. (4.24) by expanding the exponential phase factor and using a perturbation expansion. We leave it to the reader to convince himself that this indeed is the case. In particular, he can verify that even though the propagators $K_0^{\pm}$ were used in the definitions of the probability amplitude along a path or a loop, the final answer contains the K_+ propagators which keep the correct accounting.

D. Many particle interactions and the quantum response of the Universe

1. The problem of many particles

We begin by restating the results obtained so far in a slightly different notation. We will denote our spacetime slab by $0 \le t \le T$ rather than by $t_1 \le t \le t_2$; and instead of denoting the wave functions by suffixes $\pm$ we will use the notation of "in" and "out." Thus ψ_{in} denotes what we earlier called ψ_+ on $t = 0$ and ψ_- on $t = T$, while ψ_{out} denotes what we earlier called ψ_+ on the time section $t = T$ and ψ_- on the time section $t = 0$. We will denote by S the surface of the slab and n^i will denote the unit outward normal to dS, the typical surface element.

To describe action at a distance we need to generalize Eq. (4.22) to a system of many particles $a, b, c, \ldots$. We will proceed step by step. First suppose that these particles are all in an external potential A_i but do not interact with one another. Then their wave function will follow the formula

$$\Psi_{\text{out}}[a',b',\ldots] = \int\int\cdots\int K_+^A[a';a]\,\not{a}_a\,K_+^A[b';b]\,\not{a}_b \times\cdots\Psi_{\text{in}}[a,b,\ldots]dS_a dS_b\cdots . \quad (4.25)$$

Here the Ψ function has four spinorial components for each particle and the propagators act independently on these particles. From previous work we have

$$K_+^A[a';a] = \int P(\Gamma_{a'a})\exp\left(-ie_a\int_{\Gamma_{a'a}} A_i da^i\right)\mathcal{D}\Gamma_{a'a}, \quad (4.26)$$

where the typical path for particle a starts from point a and ends on point a' on the surface S. Thus the multiparticle propagator is simply the product of the individual propagators

$$K[a',b',\ldots;a,b,\ldots] = \int\int\cdots\int P(\Gamma_{a'a})P(\Gamma_{b'b})\cdots\exp\left(-ie_a\int_{\Gamma_{a'a}} A_i da^i - ie_b\int_{\Gamma_{b'b}} A_i db^i\cdots\right)\mathcal{D}\Gamma_{a'a}\mathcal{D}\Gamma_{b'b}\cdots . \quad (4.27)$$

Stated in this form the transition to the interactive multiparticle system is natural. The formula (4.27) is generalized to

$$K[a',b',\ldots;a,b,\ldots] = \int\int\cdots\int P(\Gamma_{a'a})P(\Gamma_{b'b})\cdots\exp\left[-ie_a\int_{\Gamma_{a'a}} A_i da^i - \cdots\right]\exp(iR)\mathcal{D}\Gamma_{a'a}\mathcal{D}\Gamma_{b'b}\cdots, \quad (4.28)$$

where the extra factor $\exp iR$ in the path integral is none other than that given by the Fokker formula for interparticle action. We have

$$R = -\sum_{a<}\sum_{b} e_a e_b \int_{\Gamma_{a'a}}\int_{\Gamma_{b'b}} \delta(s_{AB}^2)\eta_{ik}da^i db^k. \quad (4.29)$$

We now show that all the well-known results for quantum electrodynamics must follow from the formula (4.28) provided, as in the classical theory due note is taken of the response of the Universe.

Before coming to grips with this fundamental problem we make one comment on the wave functions Ψ_{in} and Ψ_{out}. The exclusion principle requires the wave functions to be antisymmetric with respect to the interchange of any two particles, and if Ψ_{in} is antisymmetric then so is Ψ_{out}. But what does this mean in the present path integral approach? Note that the exclusion principle prevents two paths $\Gamma_{a'a}, \Gamma_{b'b}$ from crossing if $a \neq b$; more specifically, paths with common points make a zero contribution to the amplitude. Paths for the same particle can, however, cross since they are alternatives for the propagation of the particle. Thus we can tell from this property whether two paths belong to the same particle or different ones. Note also that the vacuum loops arose when we considered the antisymmetry property for a single particle.

2. The influence functional

We have already seen in the previous section how the quantum response of the universe appears in the form of an influence functional in the problem of transition of an atomic electron. We now look for a corresponding expression in the full relativistic problem of interacting electric charges.

To begin with, we note that in Eq. (4.28) it is the paths outside the slab, i.e., those satisfying the temporal conditions $t < 0$ or $t > T$ that contribute to the potentials A_i. The discussions of preceding sections tell us how to deal with the external potentials arising from the past portions, i.e., $t < 0$. The future portions contribute because of the $\delta(s^2)$ interaction. Without loss of generality we can take

$$A^i = A^i_{t<0} + A^i_{t>T}, \quad A^i_{t<0} = 0. \quad (4.30)$$

Then the future $A^i_{t>T}$ interactions in Eq. (4.28) are the so-called *response of the Universe.*

Turning now to the interaction term R in Eq. (4.28), although we have written it in the classical fashion, we will later show that technically it should include the hitherto excluded self-action terms $a = b$. For the time being we will continue with the classical expression (4.29) and exclude these terms.

Thus, we have

$$-\sum_{a<}\sum_{b} e_a e_b \int\int \delta(s_{AB}^2)\eta_{ik}da^i db^k = -\frac{1}{2}\sum_a e_a\int A_{(a)i}da^i \quad (4.31)$$

where

$$A_{(a)i}(X) = \sum_{b\neq a} e_b\int\delta(s_{XB}^2)\eta_{ik}db^k = \frac{1}{2}[A^{\text{ret}}_{(a)i}(X) + A^{\text{adv}}_{(a)i}(X)]. \quad (4.32)$$

Here we have separated the potentials into their advanced and retarded components as per our earlier discussion of classical direct-particle electrodynamics. Therefore, the classical electrodynamic action is written in the form

$$-\sum_a e_a \int A_{it>T} da^i - \frac{1}{2}\sum_a e_a \int [A^{\rm ret}_{(a)i} + A^{\rm adv}_{(a)i}] da^i. \tag{4.33}$$

(We have not considered the inertial part of the action here, which is of course assumed to be present.) If the Universe is a perfect future absorber and an imperfect past absorber then the classical response is such that

$$A^R_i(X)_{t>T} = \frac{1}{2}\sum_a [A^{(a)}_i(X)^{\rm ret} - A^{(a)}_i(X)^{\rm adv}]. \tag{4.34}$$

Hence the classical action becomes

$$-\sum_a e_a \int \{A^{\rm ret}_{(a)i} + \tfrac{1}{2}[A^{(a)\rm ret}_i - A^{(a)\rm adv}_i]\} da^i. \tag{4.35}$$

The first term in the above sum is the total retarded potential of all other particles $b \neq a$ while the second term is the Dirac radiation reaction formula. What is the corresponding quantum version of this result?

This is where we refer back to the discussion of Sec. III. There we saw that the apparently local behavior of a quantum system actually involves the response of the Universe via an influence functional which arises when we take into account how the absorber reacts back (via advanced potentials) on the local system. The influence functional enters into any probability calculation in the path integral approach (cf. Feynman and Hibbs, 1965) whenever the effects of external variables on the local system are integrated out. It is a double integral over paths and conjugate paths.

In Sec. III we saw how the conjugate paths arise in the calculation of probability for spontaneous transition of the atomic electron, involving the response of the Universe, when the effects of the individual absorber particles are integrated out. The calculation requires paths $\mathbf{a}(t), \mathbf{b}(t), \ldots$ starting from points $a, b, \ldots$, etc. in addition to the conjugate paths $\mathbf{a}'(t), \mathbf{b}'(t), \ldots$ which start from points $a^*, b^*, \ldots$ on S. But both the paths and conjugate paths end at the same points $a', b', \ldots$, respectively. As we saw in Sec. III we end up with a transition probability instead of a transition amplitude. Experiments, however, are concerned with the measurements of the former only and so the theory does not suffer from any incompleteness on this count. In fact, we saw that the infinite term cancels out in the full probability calculation of Eq. (3.110).

The paths and conjugate paths together permit the separation of positive and negative frequencies with the paths giving positive frequencies and the conjugate paths the negative frequencies. In the explicit example of Sec. III we saw how the quantum transitions in the future absorber lead to this distinction. We generalize the concept here.

We define the positive and negative frequency components of the advanced and retarded potentials by

$$A^i(X)^{\rm ret}_{\pm} = \sum_b e_b \int \frac{\delta_\pm(t_X - t_B - |\mathbf{x} - \mathbf{x}_B|)}{2|\mathbf{x} - \mathbf{x}_B|} db^i, \tag{4.36}$$

$$A^i(X)^{\rm adv}_{\pm} = \sum_b e_b \int \frac{\delta_\pm(t_X - t_B + |\mathbf{x} - \mathbf{x}_B|)}{2|\mathbf{x} - \mathbf{x}_B|} db^i, \tag{4.37}$$

where the $\delta_\pm$ functions have the usual meaning

$$\delta_\pm(x) = \frac{1}{\pi}\int_0^\infty e^{\mp iwx} dw = \delta(x) \mp \frac{i}{\pi}\frac{\mathcal{P}}{x}. \tag{4.38}$$

It can be shown that although the individual expressions defined in Eqs. (4.36) and (4.37) are not vectors, the differences

$$[A^{i\ \rm ret}_{+} - A^{i\ \rm adv}_{+}]\ ,\ [A^{i\ \rm ret}_{-} - A^{i\ \rm adv}_{-}]$$

transform as vectors. These combinations have no place in classical electrodynamics, but they arise in quantum electrodynamics in a natural way.

The quantum response corresponding to the classical expression (4.34) is then given by

$$A^R_i(X)_{t>T} = \frac{1}{2}\sum_b \{[A^{(b)}_i(X)^{\rm ret}_{+} - A^{(b)}_i(X)^{\rm adv}_{+}] + [A^{\prime(b)}_i(X)^{\rm ret}_{-} - A^{\prime(b)}_i(X)^{\rm adv}_{-}]\}. \tag{4.39}$$

As stated above, the paths carry positive frequencies and the conjugate paths carry negative frequencies. If we imagine a coalescence of paths and conjugate paths, Eq. (4.39) collapses into Eq. (4.34). To obtain an equivalent condition for $A'^R_i(X)_{t>T}$ we interchange paths with conjugate paths and positive with negative frequencies, and we find that

$$A'^R_i(X)_{t>T} = A^R_i(X)_{t>T}. \tag{4.40}$$

Typically, in the calculation of the influence functional we have two pairs of paths and conjugate paths $(\mathbf{a}, \mathbf{a}')$ and $(\mathbf{b}, \mathbf{b}')$ for pairs of interacting particles a, b. We therefore have four possible combinations. Consider the combination $(\mathbf{a}, \mathbf{b})$ which has two terms. The first is

$$-e_a e_b \int\int \delta(s^2_{AB}) \eta_{ik} da^i db^k$$

from the time symmetric interparticle action, and the second

$$-\frac{1}{2}e_a e_b\left\{\int_{t_A>t_B}\int[\delta_+(s^2_{AB})-\delta_-(s^2_{AB})]\eta_{ik}da^i db^k+\int_{t_B>t_A}\int[\delta_+(s^2_{BA})-\delta_-(s^2_{BA})]\eta_{ik}da^i db^k\right\}$$

arises from A_i. The time inequalities in the integrals reflect the advanced and/or retarded nature of the potential components. A little book-keeping exercise gives (cf. Hoyle and Narlikar, 1974) the sum of the above two contributions as

$$-e_a e_b\int\int\delta_+(s^2_{AB})\eta_{ik}da^i db^k. \tag{4.41}$$

Likewise, the combination $(\mathbf{a}',\mathbf{b}')$ gives

$$+e_a e_b\int\int\delta_-(s^2_{A'B'})\eta_{ik}da'^i db'^k. \tag{4.42}$$

The remaining two combinations $(\mathbf{a}',\mathbf{b})$, $(\mathbf{a},\mathbf{b}')$ combine to give

$$\begin{aligned}e_a e_b\Bigg\{&\int_{t_{A'}>t_B}\int\delta_+(s^2_{A'B})\eta_{ik}da'^i db^k-\int_{t_B>t_{A'}}\int\delta_-(s^2_{BA'})\eta_{ik}da'^i db^k\\&+\int_{t_{B'}>t_A}\int\delta_+(s^2_{B'A})\eta_{ik}da^i db'^k-\int_{t_A>t_{B'}}\int\delta_-(s^2_{AB'})\eta_{ik}da^i db'^k\Bigg\}.\end{aligned} \tag{4.43}$$

We now have the influence functional in the form that generalizes Eq. (3.104) of the previous section:

$$\begin{aligned}F[\mathbf{a},\mathbf{b};\mathbf{a}',\mathbf{b}']=\exp\Bigg[&(e_a e_b/4\pi^2)\int d\Omega\int KdK\Bigg\{\int\int\exp[-iK|t_A-t_B|+i\mathbf{k}\cdot(\mathbf{x}_B-\mathbf{x}_A)]\eta_{ik}da^i db^k\\&+\int\int\exp[iK|t_{A'}-t_{B'}|+i\mathbf{k}\cdot(\mathbf{x}_{A'}-\mathbf{x}_{B'})]\eta_{ik}da'^i db'^k-\int\int\exp[iK(t_A-t_{B'})\\&+i\mathbf{k}\cdot(\mathbf{x}_{B'}-\mathbf{x}_A)]\eta_{ik}da^i db'^k-\int\int\exp[ik(t_B-t_{A'})+i\mathbf{k}\cdot(\mathbf{x}_{A'}-\mathbf{x}_B)]\eta_{ik}da'^i db^k\Bigg\}\Bigg].\end{aligned} \tag{4.44}$$

We interpret the various terms in $F[\mathbf{a},\mathbf{b};\mathbf{a}',\mathbf{b}']$ as follows. For $t_A>t_B$, the positive terms in the curly brackets contribute to a downward transition of particle b and an upward transition of particle a, and vice versa for $t_B>t_A$. The negative terms contribute to downward transitions of both a and b. Thus, (4.44) describes absorption and stimulated emission. In general we may consider these phenomena as energy exchanges between particles of a pair and between the pair and the surroundings. Since the coefficient in front of all terms is the same, the probabilities for these processes are also the same.

3. Self-action

We now come to an issue that distinguishes the quantum treatment of direct interparticle action from its classical treatment. Suppose self-action were included in the classical theory. This would mean the addition of an integral of the following form for each typical charge a in the action formula (1.6):

$$-\frac{1}{2}e_a^2\int\int\delta(s^2_{A\tilde{A}})\eta_{ik}da^i d\tilde{a}^k. \tag{4.45}$$

Here both points A and $\tilde{A}$ lie on the same path $a(t)$. Since all classical paths are timelike, we have $\delta(s^2_{A\tilde{A}})=0$ for $A\neq\tilde{A}$. That is, for two distinct points on the path the segment $A\tilde{A}$ is timelike and so the delta function vanishes. Hence the term adds nothing to the action except in the case $A=\tilde{A}$. It turns out, however, that this addition leads to the notorious infinities of electrodynamics.

In quantum electrodynamics, the situation is different since the paths here are made up of both Γ^+ and Γ^- segments, i.e., they can turn backwards in time. Thus we can find distinct points A and $\tilde{A}$ on the same path $a(t)$ such that $s^2_{A\tilde{A}}=0$. In other words segments on which these points lie do interact. Can we therefore retain the rule that there is no self-interaction of a typical path? This would mean that two segments with points on one connectible to points on the other by null rays still do not interact if they belong to the same path.

Such a rule would be difficult to implement in practice since it would require knowing beforehand the full history of the path or paths to which the segments belong to decide whether they interact or not. In any case the phenomenon of positronium annihilation tells us that the two segments do interact. Thus it seems necessary to add

the self-action term (4.45) to the action and limit the lack of self-interaction to the proviso $A \neq \tilde{A}$.

Including the self-action therefore, we need to consider the influence functional for the case $a = b$. Hoyle and Narlikar (1971) have shown that in such a situation the essential contributions are from the combinations $(\mathbf{a}, \mathbf{a}')$, $(\mathbf{a}, \mathbf{a})$, and $(\mathbf{a}', \mathbf{a}')$. The resulting expression is

$$F[\mathbf{a}, \mathbf{a}'] = \exp\left\{ ie_a^2 \left[\int_{t_{A'} > t_A} \int \delta_+(s^2_{A'A}) \eta_{ik} da'^i da^k - \int_{t_A > t_{A'}} \int \delta_-(s^2_{AA'}) \eta_{ik} da'^i da^k \right.\right.$$
$$\left.\left. - \frac{1}{2} \int\int \delta_+(s^2_{A\tilde{A}}) \eta_{ik} da^i d\tilde{a}^k + \frac{1}{2} \int\int \delta_-(s^2_{A'\tilde{A}'}) \eta_{ik} da'^i d\tilde{a}'^k \right] \right\}, \tag{4.46}$$

or, in terms of Fourier integrals it becomes

$$F[\mathbf{a}, \mathbf{a}'] = \exp\left[(e_a^2/4\pi^2) \int d\Omega \int_0^\infty K dK \left\{ - \int\int \exp iK(t_A - t_{A'}) + i\mathbf{k} \cdot (\mathbf{x}_{A'} - \mathbf{x}_A) \eta_{ik} da^i da'^k \right.\right.$$
$$+ \int_{t_A > t_{\tilde{A}}} \int \exp iK(t_{\tilde{A}} - t_A) + i\mathbf{k} \cdot (\mathbf{x}_{\tilde{A}} - \mathbf{x}_A) \eta_{ik} da^i d\tilde{a}^k$$
$$\left.\left. + \int_{t_{A'} > t_{\tilde{A}}} \int \exp[iK(t'_{A'} - t_{\tilde{A}'}) + i\mathbf{k} \cdot (\mathbf{x}_{\tilde{A}'} - \mathbf{x}_{A'})] \eta_{ik} da'^i d\tilde{a}'^k \right\} \right]. \tag{4.47}$$

This is the same result as that obtained in Sec. III. The first term in the curly bracket gives the spontaneous transitions while the second and third terms contribute the radiative correction effects. In principle a factor of this form should exist for all particles $a, b, c, \ldots$ of the system. We will consider the radiative corrections in the next section.

4. Interaction with vacuum loops

The expression (4.28) is still incomplete because we have not included the interactions of paths with loops. As discussed in IV.C.2. the loops arise in the theory by the requirement of antisymmetrization and thus will influence any phenomenon of interaction between "real" particles.

This generalization is straightforward, given the earlier discussion of the influence functional for paths of particles. As shown by Hoyle and Narlikar (1971) the loops by themselves do not affect the probability calculation but the interactive term between a loop and a particle path does. The influence functional for a loop-particle interaction is given by

$$F[\mathbf{a}, \mathbf{l}; \mathbf{a}', \mathbf{l}'] = \exp\left[iee_a \left\{ \int_{t_{A'} > t_L} \int \delta_+(s^2_{A'L}) \eta_{ik} da'^i dl^k - \int_{t_L > t_{A'}} \int \delta_-(s^2_{LA'}) \eta_{ik} da'^i dl^k \right.\right.$$
$$+ \int_{t_{L'} > t_A} \int \delta_+(s^2_{L'A}) \eta_{ik} da^i dl'^k - \int_{t_A > t_{L'}} \int \delta_-(s^2_{AL'}) \eta_{ik} da^i dl'^k$$
$$\left.\left. - \int\int \delta_+(s^2_{AL}) \eta_{ik} da^i dl^k + \int\int \delta_-(s^2_{A'L'}) \eta_{ik} da'^i dl'^k \right\} \right]. \tag{4.48}$$

As we shall see, this type of interaction produces the effects normally ascribed to "vacuum polarization" in field theory.

With this result the action-at-a-distance quantum electrodynamics can be said to have reached the same level of attainment as the conventional quantum field theoretic electrodynamics. Although we have used (throughout this section) the language of flat spacetime, we have done so because electrodynamics is conformally invariant and our cosmological models are conformally flat. Thus our arguments work in an expanding universe with the correct past and future boundary conditions, i.e., with the response required by the condition (4.34). However, it is cosmology that brings out the real differences between the two approaches when we consider the so-called radiative corrections and the renormalization program. We turn to these issues next.

V. COSMOLOGICAL RESPONSE: SOME IMPLICATIONS

A. Radiative corrections

1. The electron self-energy correction

We now examine the so-called self-energy correction due to the radiative processes in action-at-a-distance electrodynamics. Recall that the classical self-energy problem is solved in this theory by the use of advanced reaction from the rest of the Universe. The problem arises in quantum field theory from the ultraviolet divergence, i.e., from the degrees of freedom of the electromagnetic field of very high frequency. In the action-at-a-distance version, the self-energy problem in principle appears from the identification of the two legs $A, \tilde{A}$, of the delta function in the interaction

$$-\frac{1}{2}e_a^2 \int\int \delta_+(s^2_{A\tilde{A}})\eta_{ik}da^i d\tilde{a}^k \tag{5.1}$$

describing the action of a charge a on itself. Note that the delta function δ_+ instead of δ appears in Eq. (5.1) after we include the response of the Universe.

Any computation of a quantum electrodynamic cross section using Eq. (5.1) leads to a divergent result if one uses it literally as given. However, action at a distance requires a lower cutoff of the kind

$$|X^i_A - X^i_{\tilde{A}}| \geq |\epsilon^i| \tag{5.2}$$

with the cutoff vector ϵ^i having length ϵ small compared to the Compton wavelength of the mass m_a. Hoyle and Narlikar (1971) had conjectured that if a more complete theory includes classical gravity then a natural cutoff would be the Schwarzschild radius of the charge, $2Gm_a/c^2$. Later Padmanabhan (1985) showed that in a quantum gravity context the cutoff is the Planck length $(G\hbar/c^3)^{1/2}$.

Neither of these cutoffs, however, reflect the global nature of the problem, i.e., the fact that any local quantum measurement is subject to the interference of the response of the universe. A clue to this type of cutoff was provided by our earlier discussion of Sec. III.E. There we found that because of the event horizon in the future absorber the response is limited to frequencies up to those satisfying the inequality

$$k < \omega_{\text{eff}}/HT, \tag{5.3}$$

where T is the time duration of the measurement. This was the limit in the spontaneous transition problem. However, as we saw in Sec. IV the limit will appear in the more general influence functional calculated in formulas (4.44) and (4.47). This limit to high frequencies in the momentum space translates to a lower limit in the configuration space. Identifying the lower limit ϵ in formula (5.2) with that in Eq. (5.3) we get

$$\epsilon \sim k^{-1}. \tag{5.4}$$

With a finite cutoff the calculation of the bare and observed masses of the electric charge can be performed using the usual methods of the renormalization program. (See Hoyle and Narlikar, 1971, for details.)

The cutoff on k at the high frequency and given by (5.3) works out to $\sim 10^{31}$ sec^{-1} for the atomic and cosmological parameters described above. This cutoff may vary from one microscopic process to another; it also is linked with the properties of the cosmic absorber. However, the reasoning given above tells us that *for every microscopic process in electrodynamics a cutoff exists.*

The purely local approach to QED demands Lorentz invariance in every operation that may be performed. Our method, on the other hand, picks out a specific local reference frame, viz. the so-called *cosmological rest frame*, to define the response of the Universe. Thus Lorentz invariance is manifestly not present, although one can *use* the Lorentz transformation to describe any process of QED in a frame different from the cosmological rest frame.

Choose the cosmological rest frame in which, by the arguments of the preceding section, all Fourier integrals in the computations of the influence functional have a high frequency cutoff at $k_{\max}$, say. With $c = 1$, $\hbar = 1$, this cutoff corresponds to a restriction in time coordinate

$$|t_P - t_Q| \geq k^{-1}_{\max}. \tag{5.5}$$

We will shortly specify $k_{\max}$; for the time being $k^{-1}_{\max}$ remains a small quantity akin to ϵ used in Eq. (5.2).

Using the derivation of Hoyle and Narlikar [1971: see their Eq. (146)] therein we find that *in the cosmological rest frame* the observed mass m_{obs} is related to the theoretical mass m_{th} of the electron by the relation

$$m_{\text{obs}} = m_{\text{th}}\left\{1 + \frac{3e^2}{2\pi}\ln\left(\frac{k_{\max}}{m_{\text{th}}}\right)\right\}. \tag{5.6}$$

Notice that a similar formula comes from quantum field theory but there the cutoff is a purely abstract quantity and so no numerical significance is attached to the mass difference

$$\Delta m = m_{\text{obs}} - m_{\text{th}}. \tag{5.7}$$

In the present theory $k_{\max}$ is related to physical parameters and as a result it is possible to estimate it and Δm, which we proceed to do now.

The upper limit on k is given by (5.3) in which we have $\omega_{\text{eff}} \sim 80$ sec^{-1}, $H^{-1} \sim 3 \times 10^{17}$ sec, and T a time scale large compared to the characteristic time for the process—in this case the free motion of the electron. We may take $T \sim \hbar/mc^2 \sim 10^{-21}$ sec. Thus we have

$$k_{\max} \sim 10^{40} \text{ sec}^{-1}. \tag{5.8}$$

Using these values, (5.6) and (5.7) give in dimensionless form

$$\frac{\Delta m}{m} \sim \frac{3e^2}{2\pi\hbar c}\ln(10^{19}) \sim 0.15. \tag{5.9}$$

In fact, with $T \sim \hbar/mc^2$, (5.9) expressed in symbols is

$$\frac{\Delta m}{m} \sim \frac{3\alpha}{2\pi}\ln\left(\frac{\omega_{\text{eff}}}{H}\right), \tag{5.10}$$

where α is the fine-structure constant.

Two comments are needed to elaborate the above conclusion. First, (5.10) shows clearly the cosmological input to the correction term at which no purely local attempt to resolve the divergence problem will arrive. Second, the correction has been obtained in the cosmological rest frame, and so the statement is not strictly Lorentz invariant. This in our view is an unavoidable conclusion echoing the first comment that only a global theory can lead to the resolution of the divergence problem. In this context we recall Dirac's intuitive perception when he wrote:

> *"With a cut-off we eliminate at once all difficulties about divergent integrals which have been plaguing theoretical physics for decades. These difficulties arise only because people want to have strict Lorentz invariance in an imperfect theory. In doing so they are aiming for something which may very well be impossible"* [– Dirac, 1969].

We would agree with the above sentiment with one modification: replace the adjective "imperfect" by "incomplete" to underscore the one crucial element missed out in field theory, viz. the response of the Universe.

2. Charge renormalization

The "renormalization" of electric charge occurs in the present theory through the interaction of charged closed loops as intermediaries between the interaction of any two charges. Suppose an external potential B_i scatters a particle from state u_1 to u_2 in the presence of a loop. The loop-path influence functional then leads to the amplitude

$$-(-ie^2)(-ie^2)\int\int\int \bar{u}_2(1)\gamma_i u_1(1)\delta_+(s_{12}^2) \times \mathrm{Tr}[\gamma^i K_+(2;3)\, \not{B}\,(3) K_+(3;2)]d\tau_1 d\tau_2 d\tau_3 \tag{5.11}$$

with the integrations with respect to τ_1, τ_2, and τ_3 being over the slab $0 \le t \le T$. The scattering is the same as that produced by a potential

$$A^i(1) = ie^2\int\int \delta_+(s_{12}^2)\mathrm{Tr}[\gamma^i K_+(2;3)\, \not{B}\,(3) \times K_+(3;2)]d\tau_2 d\tau_3, \tag{5.12}$$

which in turn corresponds to the current

$$j^i(1) = \frac{1}{4\pi}\Box_1 A^i(1) = ie^2\int \mathrm{Tr}[\gamma^i K_+(1;3)\, \not{B}\,(3) K_+(3;1)]d\tau_3. \tag{5.13}$$

The evaluation of $j^i(1)$ is given in detail by Hoyle and Narlikar (1971) and we simply quote the result which is cutoff dependent:

$$j_i(1) = \frac{2e^2}{3\pi}\ln(m\epsilon)J_i(1), \tag{5.14}$$

where m is the mass of the charge. The ϵ here arises again from the lower limit on the separation of the two legs of the delta function and its interpretation in terms of the upper limit on frequencies of the absorber response is as given by Eqs. (5.3) and (5.4). Thus in terms of the cosmological parameters of the absorber we get the charge modification Δe as

$$\frac{\Delta e}{e} = \frac{2e^2}{3\pi\hbar c}\ln\left(\frac{\omega_{\text{eff}}}{H}\right). \tag{5.15}$$

The result is that closed loops effectively lower the theoretical (or "bare") value of the electric charge by some 0.04 fraction of its original value.

We may briefly comment on the relationship (5.4) further as follows. For A and $\tilde{A}$ separated by smaller than ϵ the corresponding frequencies in the influence functional are too high for the absorber to react and influence the local experiment. This results in slight modifications of the "bare" values of mass and charge of a typical particle, for any experiment measures these values not for isolated (i.e., bare) charges but for charges in continuous interaction with the universe.

The renormalization program in the quantum field theory of charged particles has merit in that it gives an unambiguous way of handling infinite integrals which are only logarithmically divergent. It has been felt that the actual values of these integrals do not contribute to observable quantities and as such the success of the program is judged by how the residuals left after removing the infinite integrals pass the observational tests. However, as Dirac observed:

> *"... this so called good theory (QED) ... involves neglecting infinities, neglecting them in an arbitrary way. This is not sensible mathematics. Sensible mathematics involves neglecting a quantity when it is small—not neglecting it just because it is infinitely great and you do not want it"* [– Dirac, 1978].

The proposed remedy in the present approach solves this outstanding difficulty—at a price that the theoretical physicist trained at viewing the problem in a purely local way will find it difficult to appreciate. Yet, the merit of the solution presented here should induce him to take

into account the missing link, namely the response of the Universe. It is this link that forces us to consider cosmological boundary conditions for seemingly local problems.

B. Response calculation using the S-matrix formulation

The path integral approach of Feynman provides a natural way of quantizing the Wheeler-Feynman action-at-a-distance electrodynamics, as is clear from the discussion of Secs. III and IV. Nevertheless the path integral approach, especially for relativistic particles, is a relatively unfamiliar one and one would like to see some formal contact with the methods of quantum field theory. This was provided by Davies in a series of papers (see, for example, Davies, 1970, 1971, and 1972a). We give a brief account of this work below.

In the first of these papers Davies discussed the analog of the work of Sec. III, i.e., the formulation of the quantum response of the Universe that leads to the spontaneous transition of the atomic electron. Starting with the Lagrangian for the typical ath particle

$$L_a = -e_a \sum_{b\neq a} e_b \int \delta(s_{AB}^2)\eta_{ik} da^i da^k \tag{5.16}$$

and a direct particle potential due to source b

$$A_i^{(b)}(X) = e_b \int \delta(s_{XB}^2)\eta_{ik} db^k, \tag{5.17}$$

Davies showed how to construct an S-matrix perturbation expansion

$$S = \sum_n \frac{(-1)^n}{n!} \int \cdots \int P[\mathcal{L}(X_1)\mathcal{L}(X_2)\cdots \times \mathcal{L}(X_n)] d\tau_1 \cdots d\tau_n. \tag{5.18}$$

Here P denotes a time-ordered product as usual, but the $\mathcal{L}(X)$ function cannot be the formal product of the current and a potential operator since the potential in Eq. (5.17) is not a vector field in its own right. Instead it is possible to use for $\mathcal{L}$ the basic first-order interaction

$$\mathcal{L}(X) = -\int j^{(a)i}(X)\delta(s_{XX'}^2) j^{(b)k}(X')\eta_{ik} d\tau d\tau'. \tag{5.19}$$

Davies showed that if one uses the Wheeler-Feynman theory and the perfect absorber condition developed in Secs. II and III above, then one can recover the usual quantum rules for direct particle fields as defined by Eq. (1.9):

$$\langle 0|A_i(X)A_k(X')|0\rangle = -iD^+(X,X')\eta_{ik}, \tag{5.20}$$

$$\langle 0|P[A_i(X)A_k(X')]|0\rangle = -iD_F(X,X')\eta_{ik}. \tag{5.21}$$

Note that the D^+ and D_F functions used here have their usual definitions in field theory but here they have not been used to formulate the quantum rules of $A_i(X)$ treated as operators. Instead they were deduced by Davies from the application of the response of the Universe.

In subsequent papers Davies (1971, 1972a) generalized the formulation to give an exact expression rather than a perturbation expansion for S. In particular, in the first of the two papers he showed that

$$\begin{aligned} S &= P \exp\left\{ -i \int \mathcal{L}(X) d^4x \right\} \\ &= P \exp\left\{ \frac{1}{2} i \sum_a \sum_b \int\int j_i^{(a)}(X) D_F(X,X') \right. \\ &\quad \left. \times j_k^{(b)}(X') \eta^{ik} d\tau d\tau' \right\} \end{aligned} \tag{5.22}$$

gives a complete description of quantum electrodynamics for processes which involve only virtual photons. Note that in the summation over all particle pairs in Eq. (5.22) the self-action is also included as was done by Hoyle and Narlikar (1971). The expression (5.22) is analogous to that derived by Feynman (1950) by eliminating all photons from the description of electrodynamics.

In the second paper Davies discussed the role of real photons. The question is whether there are any real photons at all. As Feynman put it all photons if observed over a sufficiently long time scale are virtual. Thus if the universe is a light tight box that absorbs all photons emitted within it then the above formalism would describe all processes that apparently involve the emission and absorption of "real" photons. The condition is basically the same as that obtained by Hoyle and Narlikar, viz. that the universe should have a future absorber that absorbs completely all photons of positive frequencies.

C. Experimental search for advanced potentials

Assuming the above approach to electrodynamics to be valid, it follows that the nature of the accepted cosmological model should be consistent with the local experiments of electrodynamics. In particular, if the cosmological response is not such as to give pure retarded solutions, then it may be possible to detect advanced effects. There have been attempts to look for small advanced effects in local radiation experiments, although their interpretation itself is shrouded in controversy.

As we saw in earlier sections, no standard big bang cosmology satisfies the absorber condition to give unambiguous pure retarded solutions. It follows therefore that if one of these cosmologies is right then the pure retarded solution is untenable. Can it be that the incompleteness of future absorption would show itself through the presence of small fractions of advanced effects in local experiments?

Partridge (1973) attempted to detect such an effect in the radiation of a microwave source as it alternately radiated into free space and a local absorber. Partridge argued that advanced potentials lead to power gain rather than power loss in the source. Hence if a tiny fraction

of radiation is via advanced potentials, the power drain from the source would be less than in the pure retarded case.

Partridge set up an arrangement in which radiation was blocked by a local absorber in one direction and was allowed to move freely in another. The argument was that the local absorber will ensure pure retarded effects whereas the radiation into free space would travel long distances and through an incompletely absorbing universe. A switching arrangement allowed these possibilities alternately. Within the accuracy of the experiment (estimated at 1 part in 10^8) there was no difference in the two cases. Thus Partridge claimed to have found no evidence of advanced effects.

Subsequently, Heron and Pegg (1974) argued that Partridge's use of a static absorber would inevitably lead to a null result. Instead what they proposed was an experiment with a time asymmetric *chopper absorber* to alter the boundary conditions. This would allow them to alter the ratio of advanced to retarded components, leading to a possible detection of the former.

However, Davies (1975) has criticized both the above approaches on the grounds that with proper inclusion of thermodynamics, attempts like these are bound to give null results. The objection raised by Davies goes in fact deeper than the specific issues relating to the proposed experiments. Davies has argued that one cannot bypass thermodynamics as proposed by Hogarth, Hoyle, and Narlikar and that ultimately the thermodynamic asymmetry like that in the Boltzmann H theorem will have to be included in any realistic discussion of electrodynamic time asymmetry. In other words, Davies was reverting to the explanation of time asymmetry given by Wheeler and Feynman (1945) referred to in Sec. III.

While this could be a possible line of argument it misses the entire spirit of the action-at-a-distance theory. First, it postulates *ad hoc* asymmetrical initial conditions which are basic to the H theorem. Second, once one decides to work within the action-at-a-distance framework the nonlocality of the problem forces one to take cognizance of the large scale structure of the universe, and the cosmological considerations of Secs. III–V become relevant and unavoidable. The self-consistent mixture of advanced and retarded potentials is determined by including the response of the universe. Finally, rather than use the statistical laws of thermodynamics as fundamental laws, attempts should be made to understand them as a consequence of other more fundamental arrows of time like electrodynamics and cosmology.

Another aspect of the Partridge-type experiment relates to the deeper question of a relationship of thermodynamic and electrodynamic arrows to the expanding and contracting phases of a time-symmetric universe, such as the Friedmann model with $k = +1$. Do these time arrows reverse when the universe contracts? This question, so long as one sticks to the Wheeler-Feynman electrodynamics, is not uniquely answered, as we saw in Sec. II.C. Recently Gell-Mann and Hartle (1991) have discussed this problem in a different way. They investigate a way in which the rules of quantum mechanics might be adapted to impose a time symmetry on the boundary conditions. Thus when the universe enters the contracting phase, these microscopic degrees of freedom of the universe conspire to reverse the time-asymmetric processes. In a reanalysis of the Partridge experiment Davies and Twamley (1993) argue that its null result goes against the Gell-Mann Hartle model, but suggest that a more stringent test would be to repeat the Partridge experiment with a laser rather than microwave antenna. This is because the universe is apparently transparent out to great distances at the GHz frequencies and to include its absorptive effects along the future light cone in a more significant way much higher frequencies should be used.

VI. CONCLUSION

The above remarks motivate an important extension of the action-at-a-distance concept to all basic interactions of physics, so that the thermodynamics of macroscopic systems can be understood as a consequence of the largest scale time asymmetry, viz. the expansion of the universe. Thus one needs to go deeper into the controversial issue (see Gold, 1968, for a discussion) as to whether in a contracting universe thermodynamics goes in the reverse direction. There are other compelling reasons for seeking such an extension, since, so far as electrodynamics is concerned, action at a distance has now demonstrated the following advantages over the field theoretic description:

(1) The choice of retarded solutions is not *ad hoc* as in field theory but dictated by the time asymmetry of the universe.

(2) There is no paradox involving infinities due to self-action in the classical theory.

(3) The theory is able to account for all classical as well as quantum electrodynamics without the extra degrees of freedom vested in free fields, i.e., it is more economical in its postulates.

(4) The cosmological boundary conditions provide the cutoff at high frequencies and thereby eliminate the divergences that normally plague quantum electrodynamics.

(5) The concept of the response of the universe provides a powerful tool for limiting viable cosmological models.

These advantages are sufficient to motivate the generalization of action at a distance to other areas of physics.

There is one further hint of the possible role of the response of the universe in local phenomena, a role that takes us beyond electrodynamics. The discussions of Secs. III–V tell us that it is not proper to talk of a probability amplitude for a local microscopic system. The correct description of the physical behavior of the system follows from the probability calculation that includes the response of the universe. Thus one is dealing with a "square of the amplitude" type of expression rather than the amplitude itself.

This may explain the mystery that surrounds such epistemological issues like the *collapse of the wave function.* What is missing from the usual discussion of the problem is the response of the universe. The wave-function collapse represents the final course of action taken by the system consistent with the response of the universe. We suggest this idea as a way of understanding many other conceptual issues of quantum mechanics. It may well be that the real nonlocal "hidden variables" are contained in the response of the universe. For a detailed discussion of this idea see Hoyle (1982) and Narlikar (1993b).

What has been the progress towards extending the action-at-a-distance formulation to other interactions? In the late sixties Narlikar (1968) showed how to construct an action-at-a-distance counterpart for a field theory of arbitrary spin having a quadratic Lagrangian and linear field equations. Such a formulation will naturally have an "absorber theory" similar to the Wheeler-Feynman theory. Earlier Narlikar (1962) had discussed an absorber theory involving neutrinos on lines similar to virtual photons assuming that the neutrinos travel with the speed of light and mediate in weak interactions.

Hoyle and Narlikar (1964b, 1966, and 1974) have shown how to obtain a theory of inertia and gravity by a natural extension to scalar conformally invariant theory. The theory reduces in the many particle approximation to general relativity with the additional demonstration that the sign of the gravitational constant has to be positive. The theory is Machian in origin in the sense that it relates the inertia of matter to the large scale structure of the universe.

More recently, Hoyle, Burbidge, and Narlikar (1995) have further generalized the formulation to describe the creation of matter, including the deduction of the cosmological constant. Here the creation is through the basic unit of Planck mass $(\hbar c/G)^{1/2}$ which subsequently decays through a series of high energy physics interactions to baryonic matter—a process yet to be determined by the particle physicists. This may very well involve a "grand unification" but through action at a distance instead of fields.

Perhaps the stiffest resistance to the concept of action at a distance would come today not from microphysics but from cosmology. As we found earlier, all the popular big bang models fail to meet the appropriate temporal boundary conditions while the steady-state model satisfies them. However, the latter model has several difficulties in explaining the observed large scale features of the universe. (So does the big bang idea; but it has gained acceptance because of the belief that only it can provide an explanation for the microwave background and the abundances of light nuclei, which had baffled the steady-state model.)

Recently, however, Hoyle, Burbidge, and Narlikar (1993, 1994a, 1994b) have produced the so-called "quasi-steady-state cosmology" (QSSC) that appears to circumvent the difficulties faced by the steady-state model. This model combines some features of the big bang model with some of the steady-state model. It is thus able to explain observed features like the microwave background, abundances of light nuclei, the redshift magnitude relation for galaxies, radio source counts, angular size redshift relation, etc., that are normally claimed as successes of the big bang cosmology. It also explains features which the big bang cosmology finds hard to accommodate like the age distribution of galaxies, baryonic dark matter, relationship to high energy astrophysics, and above all the explanation of the primary creation of matter within a framework that respects the law of conservation of matter and energy.

Although not claimed as "the cosmology" by its authors, the QSSC therefore is indicative of the kind of cosmology that may be required to accommodate the growing list of extragalactic phenomena being discovered. For the present work, it has the merit of having the right kind of cosmological response. We therefore end this article with the hope that in the growing interaction between fundamental physics and cosmology the action-at-a-distance approach may have a lot to offer to both the disciplines.

REFERENCES

Bjorken, J.D., and S.D. Drell, 1965, *Relativistic Quantum Fields* (McGraw Hill, New York).

Bondi, H., and T. Gold, 1948, Mon. Not. R. Astron. Soc. **108**, 252.

Brans, C., and R.H. Dicke, 1961, Phys. Rev. **124**, 125.

Davies, P.C.W., 1970, Proc. Camb. Phil. Soc. **68**, 751.

Davies, P.C.W., 1971, J. Phys. A **4**, 836.

Davies, P.C.W., 1972a, J. Phys. A **5**, 1025.

Davies, P.C.W., 1972b, J. Phys. A **5**, 1722.

Davies, P.C.W., 1973, Mon. Not. R. Astron. Soc. **161**, 1.

Davies, P.C.W., 1975, J. Phys. A **8**, 272.

Davies, P.C.W., and J. Twamley, 1993, Class. Quantum Gravity **10**, 931.

DeWitt, B.S., and R.W. Brehme, 1961, Ann. Phys. (USA) **9**, 220.

Dirac, P.A.M., 1935, *The Principles of Quantum Mechanics* (Clarendon, Oxford).

Dirac, P.A.M., 1938a, Proc. R. Soc. A **165**, 199.

Dirac, P.A.M., 1938b, Proc. R. Soc. A **167**, 148.

Dirac, P.A.M., 1969, in *Fundamental Interactions at High Energy*, edited by T. Gudehus, G. Kaiser, and A. Perlmutter (Gordon and Breach, New York), p. 1.

Dirac, P.A.M., 1978, in *Directions in Physics* (Wiley, New York), p. 36.

Einstein, A., and W. deSitter, 1932, Proc. Natl. Acad. Sci. **18**, 213.

Feynman, R.P., 1949, Phys. Rev. **76**, 749.

Feynman, R.P., 1950, Phys. Rev. **80**, 440.

Feynman, R.P., 1963, private communication.

Feynman, R.P., and A.R. Hibbs, 1965, *Quantum Mechanics and Path Integrals* (McGraw Hill, New York).

Fokker, A.D., 1929a, Z. Phys. **58**, 386.

Fokker, A.D., 1929b, Physica **9**, 33.

Fokker, A.D., 1932, Physica **12**, 145.

Friedman, A., 1922, Z. Phys. **10**, 377.

Friedman, A., 1924, Z. Phys. **21**, 326.
Gauss, C.F., 1867, Werke **5**, 629.
Gell-Mann, M., and J.B. Hartle, 1991, Santa Barbara Report No. UCSBTH–91–31.
Gold, T., 1968, *The Nature of Time* (Cornell University, Ithaca).
Heron, M.L., and D.T. Pegg, 1974, J. Phys. A **7**, 1965.
Hogarth, J.E., 1962, Proc. R. Soc. London A **314**, 529.
Hoyle, F., 1948, Mon. Not. R. Astron. Soc. **108**, 372.
Hoyle, F., 1982, Annu. Rev. Astron. Astrophys. **20**, 1.
Hoyle, F., G. Burbidge, and J.V. Narlikar, 1993, Astrophys. J. **410**, 437.
Hoyle, F., G. Burbidge, and J.V. Narlikar, 1994a, Mon. Not. R. Astron. Soc. **267**, 1007.
Hoyle, F., G. Burbidge, and J.V. Narlikar, 1994b, Astron. Astrophys. **289**, 729.
Hoyle, F., G. Burbidge, and J.V. Narlikar, 1995, Proc. R. Soc. A (in press).
Hoyle, F., and J.V. Narlikar, 1963, Proc. R. Soc. London A **277**, 1.
Hoyle, F., and J.V. Narlikar, 1964a, Proc. R. Soc. London A **282**, 184.
Hoyle, F., and J.V. Narlikar, 1964b, Proc. R. Soc. London A **282**, 191.
Hoyle, F., and J.V. Narlikar, 1966, Proc. R. Soc. London A **294**, 138.
Hoyle, F., and J.V. Narlikar, 1969, Ann. Phys. (NY) **54**, 207.
Hoyle, F., and J.V. Narlikar, 1971, Ann. Phys. (NY) **62**, 44.
Hoyle, F., and J.V. Narlikar, 1974, *Action at a Distance in Physics and Cosmology* (Freeman, New York).
Hoyle, F., and J.V. Narlikar, 1993, Proc. R. Soc. London A **442**, 469.
Hubble, E.P., 1929, Proc. Nat. Acad. Sci. **15**, 168.
Infeld, L., and A. Schild, 1945, Phys. Rev. **68**, 250.
Narlikar, J.V., 1962, Proc. R. Soc. London A **270**, 553.
Narlikar, J.V., 1968, Proc. Camb. Phil. Soc. **64**, 1071.
Narlikar, J.V., 1974, J. Phys. A **7**, 1274.
Narlikar, J.V., 1993a, *Introduction to Cosmology* (Cambridge University, Cambridge, England).
Narlikar, J.V., 1993b, in *Philosophy of Science: Perspectives from Natural and Social Sciences*, edited by J.V. Narlikar, L. Banga, and C. Gupta (Indian Institute of Advanced Study, Shimla), p. 69.
Padmanabhan, T., 1985, *G.R.G. Journal* **17**, 215.
Partridge, R.B., 1973, Nature **244**, 263.
Roe, P.E., 1969, *Mon. Not. R. Astron. Soc.* **144**, 219.
Schwarzschild, K., 1903, Göttinger Nachrichten **128**, 132.
Synge, J.L., 1960, *Relativity: The General Theory* (North-Holland, Amsterdam).
Tetrode, H., 1922, Z. Phys. **10**, 317.
Wheeler, J.A., and R.P. Feynman, 1945, Rev. Mod. Phys. **17**, 157.
Wheeler, J.A., and R.P. Feynman, 1949, Rev. Mod. Phys. **21**, 425.