Some useful properties of a theory of variable mass particles

Daniel M. Greenberger

Department of Physics, City College of the City University of New York, New York, New York 10031
(Received 24 October 1972; final revised manuscript received 11 September 1973)

By a simple extension of the canonical formalism, one can include mass and proper time as dynamical variables in mechanics. Such a theory allows one to treat particles with variable mass and also classically decaying particles. The theory has other properties which offer a fresh approach to classical dynamics. For example, the inertia of a system becomes an active concept, and the rest mass of an interacting particle changes to include its binding energy. Also, the proper time as measured by a clock on a decaying particle runs at a *different* rate from a clock on a stable particle. This effect causes a "decay red shift" to be present in classically decaying systems, over and above other red shift effects. The effect could be used to remove the acausal properties of "pre-acceleration" in radiative reaction, and it is suggested that for rapidly decaying massive objects, like quasars, this effect might be responsible for part of their red shift.

1. INTRODUCTION

Recently an extension of classical mechanics was developed in which particles can have a nonconstant mass[1]. We will show that one can use such a theory to treat classically decaying systems, and also that with it one can include the binding energy of an interacting system in the inertia of the system. In this respect the theory is more closely in accord with the relativistic concept of the unity of energy and mass than is conventional mechanics, where mass is treated as a fixed parameter. In this paper we will restrict ourselves to working out some simple examples in order to indicate the possible uses of this unique feature of the theory, and its promise of offering a fresh approach to mechanics.

The basic concept of this theory is that the proper time of a particle, τ, the time read by a clock moving with the particle, is considered to be an independent degree of freedom. Then the relation between proper time and coordinate time follows as an "equation of motion" for τ. The justification for this procedure comes directly from the equivalence principle, as explained in GI. The formalism of the theory is reviewed in Sec. 2.

In this theory, the "momentum" conjugate to τ is the mass of the particle and in fact one can produce examples to show that there must actually be an uncertainty principle between measurements of mass and proper time (see GII). This is a necessary consequence of the formalism, although the chief motivation for the theory was to provide a classical theory of particle decay. Just as x-dependent potentials produce momentum changes, so τ-dependent potentials produce mass changes in the theory.

Some examples of decaying particles are treated in Sec. 3. However, an extra feature of the theory is that a clock situated on a decaying particle will run at a different rate from one on a stable particle. Thus it is insufficient to know merely the position and velocity of a particle, in order to determine the rate at which proper time passes. One must also know whether it is stable. It is possible to make models of composite particles in both relativity theory and quantum mechanics, which make this result appear quite plausible. But such models are necessarily arbitrary, in a manner similar to classical models of the electron, while the effect follows naturally as a universal phenomenon in this theory.

The τ-dependent interaction that describes decay is essentially the gravitational interaction, appropriately generalized. It is not that gravity "causes" particles to decay, but that if there is *some* interaction causing the decay, then there is also a gravitational τ-dependence induced to cause a change in inertia. This is somewhat analogous to the statement that in electromagnetic theory, if there is some arbitrary nonelectrical force causing a charged particle to accelerate, then there will be an accompanying electromagnetic effect causing it to radiate. And of course it should not be too surprising that changes in inertia should be transmitted via gravity, which after all couples to a system directly through its inertia, although this does represent a new use for gravity.

Even though the theory was developed to describe decay, any interaction at all actually affects the mass of a particle. When a particle interacts with another particle or with an external field, any binding energy must show up as a change in inertia of the system. For example, if two particles are closely bound by a very strong spring, then according to relativity the equality of mass and energy demands that the mass of the composite system must include the binding energy of the spring. And if another particle passes by this system, the gravitational attraction it will feel will be that due to the entire mass, including the binding energy of the spring.

Yet classical physics does not include this effect at all. To the extent that the binding energy is small compared to the rest energy, this is correct to order (v^2/c^2); however, it is certainly conceptually possible to have very strong binding forces between the particles and yet have the system only weakly coupled to the rest of the universe. In such a case classical physics is very poor. (Of course, general relativity must include such effects, but as they are higher order, they are very hard to calculate, while in our theory they are much easier to separate. A tensor form of our theory, which we have not yet explored, should reduce to general relativity for stable particles.)

Unlike the case in classical physics, in our theory the mass of a particle automatically adjusts itself to include the binding energy, so that the mass of an interacting particle is not the same as that of the free particle. As such, the theory provides a much more realistic first approximation in treating gravitational problems than a theory where the mass is a fixed parameter. This inclusion of binding energy is discussed in Sec. 4, and it should make the theory useful in dealing with the classical problem of self energy, though we do not attack that problem here, except to point out that the inclusion of

binding energy, plus the time difference effect between a clock on the particle and one in the laboratory, lead to a possible way of removing the acausal behavior of radiation reaction in classical physics.

In order to gain a better understanding of the mechanism by which binding energy is incorporated in the theory, the formalism is broadened in Sec. 5 to include the case of two interacting particles, rather than merely that of a particle in an external field. Finally, in Sec. 6, a canonical transformation to a new constant mass and linear proper time is introduced, whose main advantage will appear in the discussion of the Schrödinger equation in the companion paper.[2]

Because of the novelty of the "decay red shift" of Sec. 3, which implies that a clock at rest on a decaying particle will run at a slower rate that a clock at rest at the same point, but attached to a stable particle, we include an appendix to discuss the effect more fully. In the Appendix we show that one may construct conventional relativistic models of decaying particles, which provide the particle with an internal structure, and we find that the internal structure produces just such a decay red shift. However, our theory implies that these structure effects persist even when these models are reduced to the point particle limit, in which case they are taken into account by the new degrees of freedom, τ and m. We also show that it is possible to give an equivalence principle argument for the existence of the decay red shift, and we suggest that in rapidly decaying massive systems, such as quasars, a portion of their red shift might be due to this effect.

2. REVIEW OF THE FORMALISM

In this theory, the proper time τ, as read by a clock moving with a particle, and the mass m of the particle, are treated as conjugate dynamical variables, in the same way as the coordinates x_i, and the momenta p_i, of the particle. The relationship between the proper time and coordinate time, $\tau = \tau(t)$, then arises dynamically through the equations of motion, just as does the behavior of $x_i = x_i(t)$. We shall restrict ourselves here to the nonrelativistic case, although the relativistic case was also treated in GI.

The Lagrangian for a single particle in an external field has the form

$$L = L(x_i, v_i, \tau, \dot{\tau}; t), \qquad v_i \equiv \dot{x}_i, \tag{2.1}$$

and the conjugate momenta are defined by

$$p_i = \frac{\partial L}{\partial v_i}, \qquad mc^2 = \frac{\partial L}{\partial \dot{\tau}}. \tag{2.2}$$

(From here on we shall assume that $c \equiv 1$.) The Hamiltonian becomes

$$H(x_i, p_i, \tau, m; t) = p_i v_i + m\dot{\tau} - L \tag{2.3}$$

and Hamilton's equations are

$$\begin{aligned} \frac{\partial H}{\partial p_i} &= v_i, & -\frac{\partial H}{\partial x_i} &= \dot{p}_i, \\ \frac{\partial H}{\partial m} &= \dot{\tau}, & -\frac{\partial H}{\partial \tau} &= \dot{m} \end{aligned} \tag{2.4}$$

If there is only a gravitational potential present, of the form $m\varphi(x, \tau)$, ($L_{grav} \equiv L_0$), then one has the following interesting situation. In the Lagrangian form, the equivalence principle is given by the condition (see GI)

$$L_0 = 0. \tag{2.5}$$

This serves as a guide to the construction of L_0, which we take as

$$\begin{aligned} L_0 &= \alpha(\dot{\tau} - (g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu)^{1/2}) \\ &\approx \alpha(\dot{\tau} - (1 - \tfrac{1}{2}v^2 + \varphi)) \equiv \alpha X, \end{aligned} \tag{2.6}$$

where α is a function of (x, τ, t) (and $\cdot \equiv d/dt$).

If there is also a nongravitational potential V present, then we take

$$L = L_0 - V = \alpha X - V = \alpha(\dot{\tau} - 1 + \tfrac{1}{2}v^2 - \varphi) - V. \tag{2.7}$$

The equations of motion,

$$\frac{dp_i}{dt} - \frac{\partial L}{\partial x_i} = 0, \qquad \frac{dm}{dt} - \frac{\partial L}{\partial \tau} = 0, \tag{2.8}$$

become, from Eq. (2.2), and assuming $X = 0$,

$$p_i = \alpha v_i, \qquad \dot{p}_i = -\alpha\frac{\partial \varphi}{\partial x_i}, \tag{2.9}$$

$$m = \alpha, \qquad \dot{m} = -\alpha\frac{\partial \varphi}{\partial \tau} \equiv -\alpha\varphi' \tag{2.10}$$

(we shall use the convention $' \equiv \partial/\partial\tau$ throughout this paper). The Hamiltonian becomes

$$H = pv + m\dot{\tau} - L = m + \frac{p^2}{2m} + m\varphi + V, \tag{2.11}$$

and the Hamiltonian form of the equations of motion are

$$v_i = \frac{\partial H}{\partial p_i} = \frac{p_i}{m}, \qquad \dot{p}_i = -\frac{\partial H}{\partial x_i} = -m\frac{\partial \varphi}{\partial x_i}, \tag{2.12}$$

$$\begin{aligned} \dot{\tau} &= \frac{\partial H}{\partial m} = 1 - \frac{p^2}{2m^2} + \varphi = 1 - \tfrac{1}{2}v^2 + \varphi, \\ \dot{m} &= -\frac{\partial H}{\partial \tau} = -m\frac{\partial \varphi}{\partial \tau}. \end{aligned} \tag{2.13}$$

The equation for $\dot{\tau}$ is just the equation

$$X = 0. \tag{2.14}$$

It is an interesting mathematical property of the theory that eq. (2.14) follows from the equations of motion in the Hamiltonian formalism, but must be postulated separately in the Lagrangian formalism, where it implies the equivalence principle.

There is one further very important restriction we shall impose on the theory. Rather than allow the potential φ to be arbitrary, we shall assume that the theory is the nonrelativistic limit of the relativistic theory of GI. In that theory φ is the fourth component of a vector potential B_μ, which is subject to a Yang–Mills type of gauge invariance.[3] The invariant coupling is

$$p_\mu \to p_\mu - mB_\mu \tag{2.15}$$

and the gauge condition is

$$\partial_\mu B_\mu - B_\mu B'_\mu = 0. \tag{2.16}$$

In the nonrelativistic limit, $B_\mu \to (\mathbf{0}, i\varphi)$, and this survives as

$$\frac{\partial \varphi}{\partial t} + \varphi\varphi' = 0. \tag{2.17}$$

We shall assume that the arbitrariness of φ is restricted by the gauge condition (2.17). This is a severe nonlinear restriction and it greatly limits the form of the potential, and also strongly affects the possible solutions allowed. We do not know how necessary this restriction is in contributing to the physical significance of the theory. But since we are at present mainly interested in discovering its physical content we feel that any limit on the arbitrariness allowed is helpful, even this stringent one.

One further note relevant to the physical content is that a tensor theory would probably be a more realistic model than a vector theory. However, we have explored the vector theory much further and have the electromagnetic analogy to draw on, and will be able to demonstrate the most interesting effects of the theory within this model. We will examine the tensor theory at a later time.

The gauge condition (2.17) can be solved explicitly. Either there is no t or τ dependence, in which case

$$\varphi = \varphi(x), \tag{2.18}$$

or if there is, it has the following form. Rewrite Eq. (2.17) as

$$-\frac{(\partial\varphi/\partial t)_\tau}{(\partial\varphi/\partial\tau)_t} = (\partial\tau/\partial t)_\varphi = \varphi. \tag{2.19}$$

Then consider the equation as a linear one in the variable τ,

$$\tau = \tau(x, \varphi, t). \tag{2.20}$$

The solution to Eq. (2.19) is then

$$\tau = t\varphi + f(\varphi, x). \tag{2.21}$$

This is the most general solution of the nonlinear gauge restriction on φ, if there is t and τ dependence, and it gives the implicit dependence of φ on t and τ. It also shows that if there is an explicit τ dependence in the problem, which determines the decay of the particle through Eq. (2.10), then there will also be an explicit t dependence.

An alternative expression, equivalent to Eq. (2.21), is given by

$$f(x, \varphi, \tau - t\varphi) = 0. \tag{2.22}$$

3. DECAYING PARTICLES

The advantage of being able to treat particles with varying masses shows up most obviously in the case of a particle which is actually decaying, and losing mass. For conservation of energy, the particle must be coupled to other energy sources so that this energy shows up elsewhere. However, we will be concerned with the case where the particle is considered as isolated, and losing mass under the influence of an external τ-dependent gravitational potential. Then the loss of energy is no more serious than the nonconservation of the momentum of a particle in an external field, where the conservation can always be restored by treating the fully interacting system.

As noted in the introduction, the gravitational potential does not *cause* the particle to decay, but causes the mass of the particle to change as it loses energy. This differs from the standard nonrelativistic decay of a radiating particle, say, in the following way. In the standard theory there is no connection between mass and rest energy, nonrelativistically. So if a particle has rest energy m_0, and an energy of interaction Δ (Δ includes kinetic and potential energy), nonetheless the particle still has inertia m_0. If while accelerating the particle radiates away energy, this process will continue until the particle has radiated away all the energy Δ, and comes to rest. But throughout the entire radiation process, the particle has mass m_0. This means for example, that it has kinetic energy $\frac{1}{2}m_0v^2$, and a gravitational coupling to other particles of Gm_0M/r. This is consistent to order (v^2/c^2) provided the binding energy itself is of order (v^2/c^2).

In our theory however, the mass of the particle will be $m_0 + \Delta$ initially, and as it loses energy, it will also lose mass, until finally the mass will be m_0. It is this change in mass that is brought about by the τ-dependent gravitational field.

The simplest example of a decaying particle is that of a particle of mass $m_0 + \Delta$, at rest, spontaneously decaying to mass m_0 under the influence of an external potential $\varphi(\tau, t)$. There are no position-dependent forces acting and the particle remains at rest. The Hamiltonian for this case is

$$H = m + m\varphi, \tag{3.1}$$

and the relevant gauge condition is

$$\tau = t\varphi + f(\varphi). \tag{3.2}$$

By differentiating with respect to τ, we have

$$1 = t\varphi' + \left(\frac{\partial f}{\partial\varphi}\right)\varphi', \qquad \varphi' = \left(t + \frac{\partial f}{\partial\varphi}\right)^{-1}. \tag{3.3}$$

Taking the total time derivative of Eq. (3.2) gives

$$\dot\tau = \varphi + \left(t + \frac{\partial f}{\partial\varphi}\right)\dot\varphi = \varphi + \dot\varphi/\varphi'. \tag{3.4}$$

The equation of motion for τ, Eq. (2.13), gives

$$\dot\tau = 1 + \varphi, \tag{3.5}$$

which with Eq. (3.4) yields

$$\dot\varphi = \varphi'. \tag{3.6}$$

The mass dependence is determined by Eq. (2.13),

$$\dot m/m = -\varphi' = -\dot\varphi, \tag{3.7}$$

so that

$$m = m_0e^{-\varphi(t)}. \tag{3.8}$$

Thus, by the choice of a function $f(\varphi)$ in Eq. (3.2), one determines the time dependence of φ, $\varphi(t) = \varphi(\tau(t), t)$, and therefore of the mass. Conversely, by choosing a time dependence for the decay, $m = m(t)$, one can use Eq. (3.3) to solve for the explicit from of φ, $\varphi(\tau, t)$.

As a specific example, consider a particle decaying exponentially from mass $m_0 + \Delta$ to mass m_0,

$$m = m_0 + \Delta e^{-\gamma t}. \tag{3.9}$$

We will assume that $\Delta \ll m_0$ and solve only to lowest order in Δ/m_0 for convenience, although this is an unnecessary restriction, of course. Then

$$\frac{\dot{m}}{m} = -\frac{\gamma\Delta e^{-\gamma t}}{m} \approx -\frac{\gamma\Delta e^{-\gamma t}}{m_0} = -\dot{\varphi} \tag{3.10}$$

and

$$\varphi(t) \approx -\frac{\Delta e^{-\gamma t}}{m_0}. \tag{3.11}$$

Note that φ is always < 0.

The fact that the particle ultimately ceases decaying by $t = \infty$ has been used to normalize φ to $\varphi(\infty) = 0$, so that for the stable particle at rest at $t = \infty$, $\dot{\tau} = 1$ from eq. (3.5). To find $\varphi(\tau, t)$, we use Eq. (3.3),

$$\frac{\partial f}{\partial \varphi} = -t + \frac{1}{\dot{\varphi}} = -\left(\frac{1}{\gamma}\right)\ln\left(-m_0\frac{\varphi}{\Delta}\right) - \frac{1}{\gamma\varphi}, \tag{3.12}$$

$$f(\varphi) = -\left(\frac{1}{\gamma}\right)\left[\varphi\ln\left(-\frac{m_0\varphi}{\Delta}\right) - \varphi + \ln(-\varphi) + C\right].$$

The constant is determined from the condition on $\tau(t)$,

$$\tau(0) = 0, \tag{3.13}$$

which implies that the particle's proper time is initially "set" to coincide with the laboratory time. Then, since at $t = 0$, $\varphi = -\Delta/m_0$, Eq. (3.2) gives $f(-\Delta/m_0) = 0$, so that Eq. (3.12) becomes

$$f(\varphi) = -\left(\frac{1}{\gamma}\right)\left[(1+\varphi)\ln\left(-m_0\frac{\varphi}{\Delta}\right) - \left(\varphi + \frac{\Delta}{m_0}\right)\right] \tag{3.14}$$

and this determines $\varphi(\tau, t)$ implicitly from Eq. (3.2).

An important feature of the solution, which is characteristic of all such problems is that at $t = \infty$, $\tau \neq t$. In fact, from Eq. (3.6),

$$\tau = t + \int_0^t \varphi dt \approx t - \left(\frac{\Delta}{\gamma m_0}\right)(1 - e^{-\gamma t}),$$

$$\tau(t \to \infty) \to t - \frac{\Delta}{\gamma m_0}. \tag{3.15}$$

This is a novel feature of the theory, that the rate of change of proper time for a decaying particle is different from that for a stable one. This is true even for a particle at rest, so it should not be confused with the variation of τ from special relativity.

As surprising as this result may be, it is nonetheless possible to give special relativistic and quantum mechanical models of an unstable clock which make it appear quite plausible. However, these models are necessarily arbitrary, as they endow the particle with a structure, like the classical theory of the electron. But in our theory, it appears as an inevitable and natural result, independent of the construction of any specific model, though to lowest order the models we present predict the same result.

For example, one can consider a decaying quantum mechanical system of two states,

$$H_0|1\rangle = m_0|1\rangle, \quad H_0|2\rangle = (m_0 + \Delta)|2\rangle, \tag{3.16}$$

with the normalized wave function

$$|\psi\rangle = (1 - e^{-\gamma t})^{1/2}|1\rangle + e^{-\gamma t/2}|2\rangle, \tag{3.17}$$

so that the average energy agrees with the classical energy.

$$\overline{H} = \langle\psi|H_0|\psi\rangle = (1 - e^{-\gamma t})m_0 + e^{-\gamma t}(m_0 + \Delta)$$
$$= m_0 + \Delta e^{-\gamma t}. \tag{3.18}$$

The total Hamiltonian is of course nondiagonal in this representation. Now one can define a "proper time" for the system ($\hbar \equiv 1$) by the equation

$$\exp(-im_0 t) \equiv \exp\left(-i\int_0^\tau \overline{H}dt\right). \tag{3.19}$$

This equation defines the proper time by equating the corresponding phases of the decaying system and one in the ground state. Then

$$m_0 t = \int_0^\tau (m_0 + \Delta e^{-\gamma t})dt = m_0\tau + \left(\frac{\Delta}{\gamma}\right)(1 - e^{-\gamma t}), \tag{3.20}$$

which is just exactly Eq. (3.15), so that according to the model, the time the particle spends in its excited state effectively slows up its internal clock.

Special relativity gives a similar result. Imagine a block of matter at rest, which is cooling off according to the formula $E = M_0 + \Delta e^{-\gamma t}$. If we define the "proper time" of the block as the average value of the proper time of each of its N molecules, each of rest mass $m_0 = M_0/N$, then

$$d\tau = \left(\frac{1}{N}\right)\Sigma d\tau_i = \left(\frac{1}{N}\right)\Sigma(1 - v_i^2)^{1/2}dt$$
$$\approx \frac{dt}{N}\Sigma\left(1 - \frac{v_i^2}{2}\right) = dt - \frac{dt}{Nm_0}\Sigma\frac{m_0 v_i^2}{2}$$
$$= dt - \frac{dt}{Nm_0}(E - M_0) = dt - \frac{\Delta}{m_0}e^{-\gamma t}dt. \tag{3.21}$$

Again, this gives $\tau - t \to -\Delta/M_0\gamma$, because proper time runs more slowly for an excited particle. While each of these examples is perhaps plausible, though depending on such arbitrary concepts as "average proper time of its molecules," the variable mass theory produces the identical result as a model-independent, universal phenomenon. A more detailed analysis of this effect, the "decay red shift," is contained in the appendix.

Two more points that we note in passing are: First, in our example, where $\Delta \ll m_0$, the total Hamiltonian is conserved to lowest order in Δ/m_0, even though it is explicitly time-dependent by virtue of the time-dependence of $\varphi(\tau, t)$,

$$H = m(1 + \varphi) = m_0\left(1 + \left(\frac{\Delta}{m_0}\right)e^{-\gamma t}\right)\left(1 - \left(\frac{\Delta}{m_0}\right)e^{-\gamma t}\right)$$
$$\approx m_0 + \Theta\left(\left(\frac{\Delta}{m_0}\right)^2\right). \tag{3.22}$$

Second, for the opposite case $\Delta \gg m_0$, the Eq. (3.15) would be replaced by

$$\tau(t \to \infty) \to t - \left[\left(\frac{1}{2\gamma}\right)\left(\ln\frac{\Delta}{m_0}\right)^2 - \left(\frac{m_0}{\Delta\gamma}\right) - \frac{\pi^2}{6\gamma}\right] + \Theta\left(\left(\frac{m_0}{\Delta}\right)^2\right). \tag{3.23}$$

4. BINDING ENERGY AS INERTIA

While the most immediate manifestation of a variable mass occurs in the problem of a decaying particle, yet even a stable particle has a nonconstant mass, in a very real sense. If a stable particle is placed in an external force field there is an energy of interaction between the particle and the field. And according to relativity, this energy should certainly appear as a change in inertia of

the system. In the present theory this effect is in fact present and we shall demonstrate it.

Consider the following simple situation. A particle of free mass m_1 is placed in an external, static nongravitational force field, governed by a potential $V(x)$. Now besides the external force, there will also be present a gravitational inertia-changing potential. This is a new feature of the theory. Since the particle carries inertia, and this inertia is capable of changing, there is necessarily a gravitational coupling present.

There is, of course, one sense in which the theory is incomplete. But all external force theories are incomplete in this same way, so it is not peculiar to this theory. Namely, one must choose the external potential arbitrarily, in order to produce the desired force. In a full relativistic theory of interacting particles, the effective potential acting on a body is determined by the interaction coupling. In the nonrelativistic theory, the potential is essentially produced by magic. Once it is properly chosen, the effects follow. However, any deeper explanation depends on a fully interacting theory. We must content ourselves with producing a potential with the desired properties, without further justification.

The Hamiltonian for the problem is

$$H = m + \frac{P^2}{2m} + m\varphi + V, \tag{4.1}$$

The equations of motion are

$$v = \frac{p}{m}, \tag{4.2a}$$

$$\dot{p} = -m\nabla\varphi - \nabla V, \tag{4.2b}$$

$$\dot{\tau} = 1 + \varphi - \tfrac{1}{2}v^2, \tag{4.2c}$$

$$\dot{m} = -m\varphi'. \tag{4.2d}$$

The potential φ is given by a simple extension of the form of the potential of Sec. 3 [still consistent with the gauge condition, Eq. (2.21)],

$$\tau = t\varphi + f(\varphi) - \tfrac{1}{2}\int_{(P)}^{x} v\cdot dx. \tag{4.3}$$

The last term is to be integrated along the path (P) of the particle, and guarantees that Eq. (4.2c) is identically satisfied. Taking the partial derivative of Eq. (4.3) with respect to τ, we find again that

$$\varphi' = \left(t + \frac{\partial f}{\partial \varphi}\right)^{-1}. \tag{4.4}$$

Taking the partial derivative with respect to x gives

$$0 = \left(\frac{1}{\varphi'}\right)\nabla\varphi - \frac{v}{2}, \qquad \nabla\varphi = \varphi'\frac{v}{2}. \tag{4.5}$$

The total time derivative of Eq. (4.3) is

$$\dot{\tau} = \varphi + t\dot{\varphi} + \left(\frac{\partial f}{\partial \varphi}\right)\dot{\varphi} - \tfrac{1}{2}v^2 = 1 + \varphi - \tfrac{1}{2}v^2$$

$$= \varphi - \tfrac{1}{2}v^2 + \frac{\dot{\varphi}}{\varphi'}, \tag{4.6}$$

from Eq. (4.2c). So we again have, as in Sec. 3,

$$\dot{\varphi} = \varphi' = -\frac{\dot{m}}{m}, \qquad m = m_0 e^{-\varphi(t)} \tag{4.7}$$

The equation of motion, Eq. (4.2b), becomes

$$m\dot{v} + v\dot{m} = m\varphi'\frac{v}{2} - \nabla V, \tag{4.8}$$

which by virtue of Eqs. (4.5) and (4.7) yields

$$\dot{v} - \dot{\varphi}\frac{v}{2} = -\left(\frac{1}{m}\right)\nabla V,$$

$$e^{+\varphi/2}\frac{d}{dt}\left(e^{-\varphi/2}\frac{dx}{dt}\right) = -\left(\frac{e^{\varphi}}{m_0}\right)\nabla V. \tag{4.9}$$

Finally,

$$m_0 e^{-\varphi/2}\frac{d}{dt}\left(e^{-\varphi/2}\frac{dx}{dt}\right) = -\nabla V. \tag{4.10}$$

If we define a new time variable, λ, by

$$\frac{d}{d\lambda} = e^{-\varphi/2}\frac{d}{dt}, \qquad \lambda = \int_0^t dt\, e^{\varphi(t)/2}, \tag{4.11}$$

then the equation of motion reduces to the form

$$m_0\frac{d^2x}{d\lambda^2} = -\nabla V. \tag{4.12}$$

Thus, in terms of the new time coordinate λ, the equation of motion reduces to that of a constant mass moving in the external field, the effect of the inertial potential φ serving to renormalize the mass and scale the time coordinate. But the mass m_0 is *not* the original mass m_1 of the particle. It is the mass *after* the particle has "adjusted" to the potential V.

In the nonrelativistic theory, one can choose the time dependence $m(t)$, by adjusting the function $f(\varphi)$ in Eq. (4.3). After the particle has reached its final mass m_0, the potential φ vanishes. Then

$$d\lambda(t\to\infty)\to dt, \qquad \lambda(t\to\infty)\to t - t_0, \tag{4.13}$$

and

$$H(t)\to m_0 + \tfrac{1}{2}m_0\left(\frac{dx}{d\lambda}\right)^2 + V(x) = m_0 + \epsilon, \tag{4.14}$$

so that the energy asymptotically becomes constant. Furthermore, if one solves for the explicit time dependence of H, one finds

$$\dot{H} = m\frac{\partial\varphi}{\partial t} = -m\varphi\varphi' = \dot{m}\varphi, \tag{4.15}$$

from Eqs. (2.17) and (4.2d), and then Eq. (4.7) for m gives

$$\dot{H} = -m_0 e^{-\varphi}\varphi\dot{\varphi},$$

$$H = m_0 e^{-\varphi}(1+\varphi) + c = m\left(1 - \ln\left(\frac{m}{m_0}\right)\right) + c. \tag{4.16}$$

Since at large times $\varphi\to 0$, and $m\to m_0$, comparison with Eq. (4.14) shows that

$$c = \epsilon. \tag{4.17}$$

The relationship between the initial mass m_1 and the final mass m_0 is determined by choosing the initial value $\varphi(0)$. Nonrelativistically this is arbitrary, and one is not forced to choose m_0 and m_1 to differ by the binding energy. One can only say that this is consistent, rather than necessary. Only in a detailed theory of interacting particles could the relationship be proven.

The subject of binding energy is closely related to that of self energy, and we might expect that the introduction of variable masses can give new insight into the question of the self mass of classical particles. We shall

not pursue this question here, except to note one interesting point. If one assumes that the force $V(x)$ is really of the form $V(x(t))$, that is, that it acts at the position of the particle at time t, then the equation of motion Eq. (4. 2b) will become, instead of Eq. (4. 12),

$$m_0 \frac{d^2x}{d\lambda^2} = -\nabla V(x(t)) = -\nabla V(x(\lambda + t_0)) = F(\lambda + t_0), \tag{4.18}$$

which, for short t_0, becomes

$$m_0 \frac{d^2x}{d\lambda^2} \approx F(\lambda) + t_0 \frac{dF}{d\lambda},$$

$$m_0 \frac{d^2x}{d\lambda^2} - m_0 t_0 \frac{d^3x}{d\lambda^3} = F(\lambda). \tag{4.19}$$

Thus the change from t to λ actually produces an effect equivalent to preacceleration and yields an equation of motion identical to the classical electromagnetic one with radiative reaction, with the identification

$$m_0 t_0 \sim m_0 \left(\frac{\Delta}{m_0 \gamma}\right) = \frac{\Delta}{\gamma} = \left(\frac{2}{3}\right) \frac{e^2}{c^3}. \tag{4.20}$$

Here the parameter γ, which is effectively the level width, would have to be proportional to Δ, the level shift, for electromagnetic radiation. The important point is that the acausal effects due to the third derivative occur in the variable λ, but are completely absent in terms of the laboratory time t, where the equation of motion is Eq. (4. 2b). While the physical significance of such a procedure is not clear, it is one of the more intriguing features of the whole variable mass formalism that there exist new concepts and thus one has the freedom to reinterpret old results of classical physics.

5. THE TWO-BODY PROBLEM

Only a limited insight into the question of binding energy can be gained by considering external forces, for it is only in the mutual effect of interacting particles on each other that the binding mechanism fully reveals itself. However, without tackling the problem in complete generality, we can make some meaningful statements about interacting systems in the lowest approximation.

A possible parametrization of the two-body problem was suggested in GI. For two particles, we have the set of coordinates $q_i = (\mathbf{r}_1, \tau_1, \mathbf{r}_2, \tau_2)$ and the corresponding momenta $\mathbf{p}_i = (\mathbf{p}_1, m_1, \mathbf{p}_2, m_2)$, and the independent variable t. If we introduce center of mass and relative coordinates,

$$\mathbf{R} = \frac{(m_1\mathbf{r}_1 + m_2\mathbf{r}_2)}{(m_1 + m_2)}, \qquad \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2, \tag{5.1}$$

then we have the difficulty that the Poisson brackets do not behave simply; for example,

$$\{\mathbf{R}, \tau_1\} = \sum_i \left(\frac{\partial \mathbf{R}}{\partial q_i}\frac{\partial \tau_1}{\partial p_i} - \frac{\partial \mathbf{R}}{\partial p_i}\frac{\partial \tau_1}{\partial q_i}\right) \neq 0. \tag{5.2}$$

Nonetheless, one can introduce a set of canonical variables satisfying

$$\{q_i, q_j\} = \{p_i, p_j\} = 0, \qquad \{q_i, p_j\} = \delta_{ij}, \tag{5.3}$$

by defining

$$\mathbf{R} = \frac{(m_1\mathbf{r}_1 + m_2\mathbf{r}_2)}{M}, \qquad P = \mathbf{p}_1 + \mathbf{p}_2, \tag{5.4a}$$

$$\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2, \qquad \mathbf{p} = \left(\frac{m_2}{m}\right)\mathbf{p}_1 - \left(\frac{m_1}{m}\right)\mathbf{p}_2, \tag{5.4b}$$

$$T = \frac{(m_1\tau_1 + m_2\tau_2)}{M}, \qquad M = m_1 + m_2 \tag{5.4c}$$

$$\kappa = M(\tau_1 - \tau_2) + \mathbf{P}\cdot\mathbf{r}, \qquad \delta = \frac{(m_1 - m_2)}{2m}. \tag{5.4d}$$

Because of its nonlinearity, the gauge condition would have to be solved from the full field theory to be interpreted correctly. However, if the two particles are considered to be close together, so that

$$|\mathbf{r}| \ll |\mathbf{R}|,$$

and $|\mathbf{r}|$ is also much smaller than the distance between $\mathbf{r}_1$ or $\mathbf{r}_2$ and any other particles around, then one can write an approximate form of the gauge condition. We assume

$$\varphi(\mathbf{r}_1) \approx \varphi(\mathbf{r}_2) \approx \varphi(\mathbf{R}). \tag{5.5}$$

Then the gauge condition, Eq. (4. 3), becomes

$$m_1\tau_1 = m_1 t\varphi + m_1 f(\varphi) - \tfrac{1}{2}m_1 \int \mathbf{v}_1 \cdot d\mathbf{r}_1. \tag{5.6}$$

Similarly,

$$m_2\tau_2 = m_2 t\varphi + m_2 f(\varphi) - \tfrac{1}{2}m_2 \int \mathbf{v}_2 \cdot d\mathbf{r}_2 \tag{5.7}$$

Using the fact that

$$\mathbf{p}_1 \cdot d\mathbf{r}_1 + \mathbf{p}_2 \cdot d\mathbf{r}_2 = \mathbf{P}\cdot d\mathbf{R} + \mathbf{p}\cdot d\mathbf{r}, \tag{5.8}$$

we can add Eqs. (5. 6) and (5. 7) to get

$$MT = Mt\varphi + Mf(\varphi) - \tfrac{1}{2}M \int V \cdot d\mathbf{R} - \tfrac{1}{2}\mu \int v \cdot dr, \tag{5.9}$$

where $\mathbf{V} \equiv \dot{\mathbf{R}}, \mathbf{v} \equiv \dot{\mathbf{r}}$, and $\mu \equiv m_1 m_2/M$. In this approximation, then, φ depends only on T, and not on κ. One would have to assume $\varphi(\mathbf{r}_1) \neq \varphi(\mathbf{r}_2)$ to obtain a κ dependence.

Since there is no κ dependence in the problem, it follows that δ is a constant of the motion. This is equivalent to the statement

$$m_1/m_2 = \text{const} \tag{5.10}$$

which implies that all mass ratios, such as m_1/M and m_2/M, remain constant. The equations of motion for τ_1 and τ_2 will still be given by Eq. (4. 2c), so we may write

$$m_1\dot{\tau}_1 = m_1 + m_1\varphi - m_1 v_1^2/2,$$
$$m_2\dot{\tau}_2 = m_2 + m_2\varphi - m_2 v_2^2/2. \tag{5.11}$$

Adding these equations, dividing by M, and remembering that mass ratios are constants, we obtain

$$\dot{T} = 1 + \varphi - \frac{P^2}{2M^2} - \frac{P^2}{2\mu M}$$
$$= 1 + \varphi - \tfrac{1}{2}V^2 - \left(\frac{\mu}{2M}\right) v^2. \tag{5.12}$$

This is the equation of motion for T.

If the Lagrangian for the problem is taken to be

$$L = m_1(\dot{\tau}_1 - 1 + v_1^2/2) + m_2(\dot{\tau}_2 - 1 + v_2^2/2) - M\varphi - V(r) - U(R), \tag{5.13}$$

where the two particles are bound together by a nongravitational potential $V(r)$ and are acted upon by an external potential U (which we consider as acting only

on the center of mass of the pair), then the Hamiltonian will be

$$H = m_1\dot{\tau}_1 + m_2\dot{\tau}_2 + \mathbf{p}_1\cdot\mathbf{v}_1 + \mathbf{p}_2\cdot\mathbf{v}_2 - L$$
$$= M + \frac{P^2}{2M} + \frac{p^2}{2\mu} + M\varphi + V + U. \tag{5.14}$$

Note that from Eq. (5.9) we have

$$\varphi = \varphi(\mathbf{R}, \mathbf{r}, T, t, \frac{\mu}{M}) \tag{5.15}$$

only. The equations of motion are

$$\dot{M} = -M\frac{\partial\varphi}{\partial T} \equiv -M\varphi', \tag{5.16a}$$

$$\dot{\delta} = 0, \tag{5.16b}$$

$$\dot{T} = 1 - \frac{P^2}{2M^2} - \frac{p^2}{2\mu^2}\left(\frac{\partial\mu}{\partial M}\right) + \varphi, \tag{5.16c}$$

$$\dot{P} = -M\nabla_R\varphi - \nabla_R U, \tag{5.16d}$$

$$\dot{p} = -M\nabla_r\varphi - \nabla_r V. \tag{5.16e}$$

From Eqs. (5.4c) and (5.4d), we have

$$\mu = \tfrac{1}{4}M(1 - 4\delta^2), \quad \frac{\partial\mu}{\partial M} = \frac{\mu}{M}, \tag{5.17}$$

so that Eq. (5.16c) reduces to Eq. (5.12).

From the gauge condition, Eq. (5.9), we find

$$\varphi' \equiv \frac{\partial\varphi}{\partial T} = \left(t + \frac{\partial f}{\partial\varphi}\right)^{-1}, \tag{5.18}$$

$$\nabla_R\varphi = \varphi'\frac{\mathbf{V}}{2}, \tag{5.19}$$

$$\nabla_r\varphi = \left(\frac{\mu}{M}\right)\varphi'\frac{\mathbf{v}}{2}. \tag{5.20}$$

Also, comparing Eq. (5.12) to the time derivative of Eq. (5.9),

$$\dot{T} = \varphi + t\dot{\varphi} + \left(\frac{\partial f}{\partial\varphi}\right)\dot{\varphi} - \tfrac{1}{2}V^2 - \left(\frac{\mu}{2M}\right)v^2$$
$$= \varphi + \frac{\dot{\varphi}}{\varphi'} - \tfrac{1}{2}V^2 - \left(\frac{\mu}{2M}\right)v^2, \tag{5.21}$$

we see that

$$\dot{\varphi} = \varphi', \tag{5.22}$$

as before, which implies from Eq. (5.16a) that

$$\frac{M}{M_0} = \frac{m_1}{m_{10}} = \frac{m_2}{m_{20}} = e^{-\varphi(t)}. \tag{5.23}$$

The equation of motion for P, Eq. (5.16d), becomes

$$(M\mathbf{V})^{\cdot} = -\frac{M\dot{\varphi}\mathbf{V}}{2} - \nabla_R U, \tag{5.24}$$

which is solved exactly as in the previous section, by Eq. (4.8), and becomes

$$M_0\left(\frac{d^2\mathbf{R}}{d\lambda^2}\right) = -\nabla_R U, \quad \lambda = \int_0^t dt\, e^{\varphi/2}. \tag{5.25}$$

The equation for p, Eq. (5.16e), becomes

$$(\mu\mathbf{v})^{\cdot} = -M\left(\frac{\mu}{M}\right)\dot{\varphi}\frac{v}{2} - \nabla_r V(r),$$
$$\mu_0\left(\frac{d^2\mathbf{r}}{d\lambda^2}\right) = -\nabla_r V. \tag{5.26}$$

In these equations it is the approximation $\varphi_1 \sim \varphi_2$, and also $U(r_1, r_2) \sim U(R)$, which eliminates the κ dependence. Since the $M(t)$ dependence in this approximation cannot distinguish between the effect of the two forces V and U, and since

$$H(t \to \infty) \to M_0 + \epsilon_R + \epsilon_r = \text{const}, \tag{5.27}$$

where ϵ_R and ϵ_r are the energies of the two forces separately, it follows that the change in mass is the cumulative effect of the binding energies of both forces separately.

However, this crude approximation is sufficient to show how important the inclusion of binding energy can be. For example, if $V(r)$ is a very strong electromagnetic attraction, whose binding energy contributes a considerable part of the total mass, while $U(R)$ is a very weak coupling of the system to an external field, then the total binding energy is essentially that due to V. So the attraction to U is determined by the actual energy of the system bound by V, not by the free particle rest masses of m_1 and m_2. For the case where U is gravitational this is crucial, as the actual force between the pair of particles and the rest of the universe will be approximately that of one composite particle of the appropriate mass, including binding.

6. A CANONICAL TRANSFORMATION

We shall perform a canonical transformation on our system which will replace τ and m by a new set of variables, which will both be constants of the motion. These variables will be most useful in a discussion of the Schrödinger equation for variable mass particles, but even in the classical case they provide a simple characterization of the solutions of the Hamiltonian.

Consider the Hamiltonian of Sec. 4, of a particle in a nongravitational external field, Eq. (4.1). Look at the variable z, defined as

$$z = z(\varphi, t) = te^{-\varphi} - \int^{\varphi}\left(\frac{\partial f}{\partial\varphi}\right)e^{-\varphi}d\varphi, \tag{6.1}$$

where $f(\varphi)$ is the same function that appears in the gauge condition, Eq. (4.3). The derivatives of z are

$$\left(\frac{\partial z}{\partial\varphi}\right)_t = -\frac{e^{-\varphi}}{\varphi'}, \quad \left(\frac{\partial z}{\partial t}\right)_\varphi = e^{-\varphi}. \tag{6.2}$$

For the significance of z, take $\dot{z} \equiv dz/dt$,

$$\dot{z} = \frac{\partial z}{\partial t} + \left(\frac{\partial z}{\partial\varphi}\right)\dot{\varphi} = e^{-\varphi}\left(1 - \frac{\dot{\varphi}}{\varphi'}\right) = 0 \tag{6.3}$$

along a solution to the equations of motion. Since $z = \text{const}\ (= z_0)$ implies $\varphi' = \dot{\varphi}$, it also implies

$$m = m_0 e^{-\varphi(t)}\Big|_{z=z_0}. \tag{6.4}$$

Another way to see this is to differentiate φ along $z = z_0$,

$$\left(\frac{\partial\varphi}{\partial t}\right)_z = -\left(\frac{\partial z}{\partial t}\right)_\varphi \Big/ \left(\frac{\partial z}{\partial\varphi}\right)_t = \varphi' = \dot{\varphi} \equiv \frac{d\varphi}{dt}, \tag{6.5}$$

which implies that along $z = z_0$, as t varies φ will vary as a solution to the equation of motion, i.e.,

$$\varphi(x, \tau, t)\Big|_{z=z_0} = \varphi(x(t), \tau(t), t) = \varphi(t). \tag{6.6}$$

We can now make a canonical transformation to replace τ by z, so that the variables of the problem change from (x, τ, t) to (x, z, t). This is done by introducing a new set of variables $(\mathbf{x}_1, \mathbf{p}_1, \tau_1, m_1)$ through the generating function[4] $F(\mathbf{x}, \mathbf{p}_1, \tau, m_1)$. Then

$$\begin{aligned} \tau_1 &= \frac{\partial F}{\partial m_1}, \qquad x_1 = \frac{\partial F}{\partial p_1}, \\ m &= \frac{\partial F}{\partial \tau}, \qquad p = \frac{\partial F}{\partial x}, \\ H_1 &= H + \frac{\partial F}{\partial t}. \end{aligned} \tag{6.7}$$

If we choose

$$F = -m_1 z + \mathbf{p}_1 \cdot \mathbf{x}, \tag{6.8}$$

where

$$z = z(\varphi(x, \tau, t), t), \tag{6.9}$$

then

$$\tau_1 = -z, \qquad x_1 = x. \tag{6.10}$$

Also

$$m = \frac{\partial F}{\partial \tau} = \left(\frac{\partial F}{\partial z}\right)\left(\frac{\partial z}{\partial \tau}\right) = -m_1\left(\frac{\partial z}{\partial \varphi}\right)\varphi' = m_1 e^{-\varphi}, \tag{6.11}$$

so that

$$m_1 = m_0. \tag{6.12}$$

The new momentum is determined by

$$\begin{aligned} p &= \frac{\partial F}{\partial x} = p_1 - m_1\left(\frac{\partial z}{\partial \varphi}\right)\nabla\varphi = p_1 + m_1 e^{-\varphi}\frac{\nabla\varphi}{\varphi'}, \\ p_1 &= p - \frac{mv(x)}{2}. \end{aligned} \tag{6.13}$$

The new Hamiltonian becomes

$$\begin{aligned} H_1 &= H - m_0\left(\frac{\partial z}{\partial t}\right)_{x,\tau} = H - m_0\left(\frac{\partial z}{\partial t} + \left(\frac{\partial z}{\partial \varphi}\right)\left(\frac{\partial \varphi}{\partial t}\right)_{x,\tau}\right) \\ &= H - m - m\varphi = \frac{p^2}{2m} + V \\ &= (\tfrac{1}{2}m_1 e^{-\varphi})(p_1 + m_1 e^{-\varphi} v(x)/2)^2 + V(x), \end{aligned} \tag{6.14}$$

where $\varphi = \varphi(x, z, t)$. We can now check the equations of motion. For example,

$$\dot{m}_1 = -\left(\frac{\partial H_1}{\partial z}\right) = -\left[\frac{p^2}{2m} - p\cdot\frac{v}{2}\right]\left(\frac{\partial \varphi}{\partial z}\right) = 0. \tag{6.15}$$

Similarly,

$$\dot{z} = \frac{\partial H}{\partial m_1} = 0. \tag{6.16}$$

If it is desired, one can eliminate the x dependence in the kinetic energy term by another canonical transformation similar to a gauge change in electrodynamics. Since m_0 and z are constant, one can treat them as such and no longer consider them as variables in the problem. Introduce the generating function

$$\overline{F}(x, p_2) = p_2 \cdot x - m_0 \int^x dx e^{-\varphi(x, z_0, t)}\cdot\frac{v(x)}{2}. \tag{6.17}$$

Then

$$x_2 = \frac{\partial \overline{F}}{\partial p_2} = x,$$

$$p_1 = \partial\overline{F}/\partial x = p_2 - m_0 e^{-\varphi} v(x)/2, \tag{6.18}$$

and

$$p_2 = p. \tag{6.19}$$

Along $z = z_0$, we have $\varphi = \varphi(t)$ alone, from Eq. (6.5) so that we can actually remove it from the integral in Eq. (6.17). Then

$$H_2 = H_1 + \frac{\partial \overline{F}}{\partial t} = H_1 + m_0 e^{-\varphi}\dot{\varphi}\int^x dx\cdot v(x)/2. \tag{6.20}$$

Finally, let us look at the simple equation of motion in terms of the variable λ, defined by Eq. (4.11), which clearly can be derived from the Hamiltonian

$$H_3 = \frac{p_3^2}{2m_0} + V(x), \tag{6.21}$$

where

$$p_3 = m_0\frac{dx}{d\lambda}. \tag{6.22}$$

The natural question is how this Hamiltonian is related to the original Hamiltonian H, or its equivalents, H_1 and H_2. The surprising answer is that the transformation from H to H_3 is *not* canonical. This can be seen from the fact that we are seeking a transformation from the variables (x, p) to the variables (x, p_3) where

$$p_3 = \left(\frac{dt}{d\lambda}\right)p = e^{-\varphi/2}p. \tag{6.23}$$

Now

$$\int p_3 \cdot dx = \int e^{-\varphi/2} p\cdot dx \neq \int p\cdot dx, \tag{6.24}$$

so that the transformation does not conserve area in phase space, and so is not canonical. Connected to this situation is the fact that the potential φ is not really a local function, as it depends on the path-dependent integral $\int v\cdot dx$.

APPENDIX: THE DECAY RED SHIFT

The purpose of this appendix is to discuss more fully the decay red shift of Sec. 3. This effect predicts that a clock at rest in an unstable coordinate system (e.g., a decaying particle) will run at a different rate from one at rest at the same point in a stable coordinate system. So the variables of position and velocity are no longer sufficient to determine the rate of a clock–one must also characterize its stability.

This surprising result arises independently of the detailed nature of our theory; *any* theory that attempts to make the mass m, and proper time τ, into independent variables will of necessity produce such an effect. This is because, by its very nature as a dynamical variable, the mass will be changed by a τ-dependent force, in the same was as momentum is changed by ordinary spatial forces. And τ-dependent potentials will effect the rate of clocks just as x-dependent gravitational potentials will.

While these results are inevitable in such a theory, even for point particles, nonetheless it is possible to render them less strange by considering models of decaying systems in conventional relativity. In that case, the structure of the decaying system actually

produces an effect equivalent to the decay red shift, because of the time dilatation effects on the various components of the system. We shall analyze this effect and show that it is of the correct order of magnitude. What our theory proclaims is that this effect persists down to the limit of a point particle, and it is taken into account by the extra degrees of freedom, m and τ.

An analogous situation occurs in the case of the spin of a system. There one may classically discuss finite spinning systems, and take the limit of point particles which have no obvious internal structure and no moment of inertia, and yet one still finds that the spin effects persist and may still yield a completely nonclassical angular momentum. In our case it is the decay effects that persist even down to the point particle limit, giving rise to a nonclassical degree of freedom.

In this appendix we shall first discuss such a conventional model of a decaying system. Next, since we claim that the decay mechanism is produced by an essentially gravitational interaction, albeit acting in quite a new role, it is amusing that we will be able to formulate a model from which the decay red shift is produced by an equivalence principle type of argument. And finally we shall indicate a possible astrophysical significance for the effect.

A. Model of a decaying system

In this section we will use conventional relativity (m and τ are *not* degrees of freedom, and m is constant) in order to describe a model of a decaying system with internal structure. The energy of the system consists of the rest mass of the particles plus their kinetic and potential energies of interaction, which gradually dissipate away.

Consider the following simple example. Two particles of mass m_1 and m_2 are bound by a covariant damped harmonic oscillator potential, their mutual radiation being the source of the damping. The initial energy of the system is $Mc^2 + \epsilon_0$ $(M = m_1 + m_2)$, and the final energy is Mc^2. The Lagrangian for such a system is

$$\mathcal{L} = \int L d\lambda,$$

$$L = (\tfrac{1}{2}m_1 u_1^2 + \tfrac{1}{2}m_2 u_2^2 - \tfrac{1}{2}k(x_1 - x_2)^2)e^{\gamma\lambda}; \tag{A.1}$$

where

$$x_1^\alpha \equiv (\mathbf{x}_1, ict), \quad x_2^\alpha \equiv (\mathbf{x}_2, ict),$$
$$u_1^\alpha = \frac{dx_1^\alpha}{d\lambda}, \quad u_2^\alpha = \frac{dx_2^\alpha}{d\lambda}, \tag{A.2}$$

and k and γ are constants. For the moment we leave the parameter λ unspecified, except that it must be an invariant.

The equations of motion are

$$m_1\frac{du_1^\alpha}{d\lambda} + m_1\gamma u_1^\alpha + k(x_1 - x_2)^\alpha = 0,$$
$$m_2\frac{du_2^\alpha}{d\lambda} + m_2\gamma u_2^\alpha + k(x_2 - x_1)^\alpha = 0. \tag{A.3}$$

One can then introduce the center of mass and relative coordinates,

$$R^\alpha = \left(\frac{m_1}{M}\right)x_1^\alpha + \left(\frac{m_2}{M}\right)x_2^\alpha,$$
$$r^\alpha = x_1^\alpha - x_2^\alpha. \tag{A.4}$$

In these coordinates

$$R^\alpha = (\mathbf{R}, ict), \quad \mathbf{R} = \text{const} \equiv 0,$$
$$r^\alpha = (\mathbf{r}, i0), \tag{A.5}$$

and

$$\mu\frac{d^2r^\alpha}{d\lambda^2} + \mu\gamma\frac{dr^\alpha}{d\lambda} + kr^\alpha = 0, \tag{A.6}$$

where μ is the reduced mass, m_1m_2/M. The solution as a function of λ is

$$\mathbf{r} = \mathbf{r}_0 e^{-\gamma\lambda/2}\cos(\omega_1\lambda + \theta_0),$$
$$\omega_1^2 = \omega_0^2 - \frac{\gamma^2}{4}, \quad \omega_0^2 = \frac{k}{\mu}. \tag{A.7}$$

To complete the model, one must define the parameter λ. For example, one may choose

$$d\lambda_L^2 = -dR^\alpha\frac{dR^\alpha}{c^2}, \tag{A.8}$$

giving

$$\lambda_L = t, \tag{A.9}$$

the laboratory time. In this case the system has a constant frequency ω_1 in the laboratory. One might also take for λ the "center of time" of the two particles,

$$d\lambda_C = \left(\frac{m_1}{M}\right)d\tau_1 + \left(\frac{m_2}{M}\right)d\tau_2. \tag{A.10}$$

In this case, if m_1 and m_2 are equal, then

$$d\lambda_C = d\tau_1 = d\tau_2, \tag{A.11}$$

and the frequency will be constant to an observer sitting on either of the particles.

However, in either case, the relationship between the "center of time" coordinate and the laboratory time will be given by (in the limit $v_1, v_2 \ll c$),

$$d\lambda_C = dt\left[\left(\frac{m_1}{M}\right)\left(1 - \frac{v_1^2}{c^2}\right)^{1/2} + \left(\frac{m_2}{M}\right)\left(1 - \frac{v_2^2}{c^2}\right)^{1/2}\right]$$
$$\approx dt[1 - (\tfrac{1}{2}m_1v_1^2 + \tfrac{1}{2}m_2v_2^2)/Mc^2]. \tag{A.12}$$

If we integrate over a cycle, we have, since the average kinetic energy for an oscillator is half the total energy,

$$d\lambda_C \approx dt[1 - \epsilon/2Mc^2]. \tag{A.13}$$

For a slowly damped oscillator,

$$\epsilon = \epsilon_0 e^{-\gamma t}, \tag{A.14}$$

so that

$$d\lambda_C \approx dt\left[1 - \frac{\epsilon_0 e^{-\gamma t}}{2Mc^2}\right]. \tag{A.15}$$

Integrating over a long time one has

$$\lambda_C - t \underset{t\to\infty}{\longrightarrow} -\frac{\epsilon_0}{2Mc^2\gamma}, \tag{A.16}$$

which is the same order of magnitude as the point particle calculation of our theory, Eq. (3.15), (with the notational changes $\epsilon_0 \to \Delta$, $\gamma \to \lambda$, $c \to 1$).

If the oscillating particles contained an actual clock (such as a precessing spin) one could actually measure the time loss between the laboratory observer, and the

decaying particle system. Also it is clear that the frequency as seen by the laboratory observer is less than that seen by an observer moving with either of the particles (since the period is greater).

It should be evident from this example that the source of the red shift between a laboratory observer and one using the center of time coordinate of the decaying system has nothing to do with the harmonic oscillator nature of the force. Rather it is due to the time dilation effect, because the system has structure, and internal motion, and any model giving the system a structure would produce the same order of magnitude time lag between a clock on the decaying system and one on a stable particle in the laboratory.

By making m and τ into dynamical variables, one is really asserting that a decaying particle has a microscopic structure, and even in the limit where the decaying system reduces down to a point particle, the effects of this structure remain, in the form of the decay red shift. The introduction of m and τ as dynamical variables enables one to take this average effective structure into account, for a point particle, independently of any detailed classical type of model.

B. The equivalence principle and decaying states

In our theory, the mass of a decaying particle changes through a gravitational interaction, though it is the τ dependence rather than the conventional x dependence that causes the decay. Nonetheless, it is possible to introduce a model that illustrates the connection between this gravitational field and the equivalence principle.

Consider the following classical model of a decaying system. The system continuously emits photons of frequency ω_0 in its rest frame. In the laboratory the system has mass m, velocity v, and in time dt it emits dN photons, and loses mass $d\mu(=-dm)$.

We will assume that the photons come off backwards, and the particle accelerates as it loses mass. Momentum conservation in the laboratory frame gives

$$\begin{aligned}(m-d\mu)(v\gamma+d(v\gamma))-\hbar k\,dN &= mv\gamma,\\ -d\mu v\gamma+m\gamma^3dv-dN\hbar k &= 0.\end{aligned} \tag{A.17}$$

Energy conservation gives

$$\begin{aligned}(m-d\mu)c^2(\gamma+d\gamma)+dN\hbar\omega &= mc^2\gamma,\\ m\gamma^3dv &= d\mu v\gamma+dN\hbar k.\end{aligned} \tag{A.18}$$

Combining Eqs. (A. 17) and (A. 18) gives

$$\frac{m\gamma^3dv}{c}=d\mu. \tag{A.19}$$

If one transforms the 4-acceleration, $d^2x^\alpha/d\tau^2$, to the rest frame of the decaying system, one obtains

$$\frac{d^2x^\alpha}{d\tau^2}=(\gamma^2a,i0)\equiv(\alpha,i0), \tag{A.20}$$

where $a=d^2x/dt^2$, and $dt=\gamma d\tau$. So

$$m\alpha d\tau=cd\mu, \tag{A.21}$$

and

$$\frac{dm}{d\tau}=-\frac{d\mu}{d\tau}=-\frac{m\alpha}{c}. \tag{A.22}$$

Comparison with Eq. (3. 7) leads one to introduce the potential (good to lowest order)

$$\varphi=c\int\alpha d\tau,\qquad c\alpha=\frac{\partial\varphi}{\partial\tau}. \tag{A.23}$$

One can also obtain from Eqs. (A. 17) and (A. 18)

$$d\mu c^2=dN\hbar k\left[\frac{(1+\beta)}{(1-\beta)}\right]^{1/2}=dN\hbar k_0, \tag{A.24}$$

where $k_0=\omega_0/c$ is the wave number in the rest frame of the decaying system, rather than the laboratory, and $\beta=v/c$.

To an observer behind the accelerating particle, at rest in the laboratory, $k=k_0[(1-\beta)/(1+\beta)]^{1/2}$, and as the particle accelerates from β to $\beta+d\beta$, we have

$$\begin{aligned}dk&=-k_0\left[\frac{(1-\beta)}{(1+\beta)}\right]^{1/2}\gamma^2d\beta\\ &=-k\gamma^2d\beta=-k\alpha d\tau/c.\end{aligned} \tag{A.25}$$

Thus

$$\frac{\delta\omega}{\omega}=\frac{\delta k}{k}=-\frac{\alpha d\tau}{c}. \tag{A.26}$$

For the period T,

$$\frac{\delta T}{T}=-\frac{\delta\omega}{\omega}=\frac{\alpha d\tau}{c}=\frac{\delta\varphi}{c^2}, \tag{A.27}$$

so that

$$\frac{(T+\delta T)}{T}=\frac{1+\delta\varphi}{c^2}, \tag{A.28}$$

which is consistent with the red shift formula for the decaying particle, Eq. (3. 5).

To an observer at rest on the accelerating system, his acceleration is α. By slowly, continuously varying the direction of emission of the photons, he need not move in a straight line, so that α is a function of his proper time τ, rather than position. The rate at which he decays is

$$\frac{dm}{d\tau}=-\frac{m\alpha}{c}=-\left(\frac{m}{c^2}\right)\frac{\partial\varphi}{\partial\tau}, \tag{A.29}$$

the "equation of motion" for m, and the radiation he has emitted at a time $d\tau$ previously, looks red-shifted to him by an amount given by Eq. (A. 25) or (A. 28). So we see that an observer moving with the decaying system can attribute his acceleration to a τ-dependent gravitational field, which regulates both his decay rate and the red shift.

C. A possible astrophysical consequence

The existence of a decay red shift should certainly be detectable in the laboratory, if it exists. We shall not attempt to discuss this problem here, however, but will merely note that under certain conditions the decay red shift could have a definite astrophysical significance.

In Sec. 5, we showed that there exists a "center of time" coordinate similar to the center of mass coordinate. We also showed that for a strongly interacting system the effect on an outside observer could be approximately characterized by this variable. In other words, in the same way as a strongly interacting system appears to an outside observer as an approximate point source, located at the center of mass of the system, so also the "proper time interaction" is approximately local, and is characterized by the center of time variable.

Thus a large accumulation of particles, like a star, ought to appear to an observer far away as a point mass, characterized by the variables R, the center of mass, and T, the center of time. Even in our example of part A of this appendix, the decay red shift appeared between the center of time and laboratory time variables.

Therefore in a stellar system which we know must be strongly interacting, such as a quasar, it is quite possible that the τ-dependent interaction between the components of the system will appear to a distant observer, to a first approximation, as a single massive particle with R and T as the relevant degrees of freedom. In this case the entire massive decaying object would behave as a simple point source, and one should see a decay red shift, much as in the simple example of Sec. 3, for which we constructed a conventional extended model in part A of this appendix.

Numerically, if as much as half the mass of the quasar were to rapidly radiate away exponentially, then in Eq. (3.9) we would have $\Delta = m_0$. Then

$$\varphi(t) = -\ln\left(1 + \left(\frac{\Delta}{m_0}\right) e^{-\gamma t}\right), \tag{A.30}$$

and for short times $\gamma t \ll 1$,

$$\dot{\tau} = 1 + \varphi \approx 1 - 1 - \ln 2. \tag{A.31}$$

Then, because of the slowing up of proper time in the decaying quasar, there would be an additional "decay red shift"

$$\dot{\tau} = \frac{\lambda}{\lambda'} = \frac{1}{(1+z)} = 1 - \ln 2, \tag{A.32}$$

where z is the red shift parameter $(\lambda' - \lambda)/\lambda$ and equals 2.26 in our example.

Since this shift is of the order of observed red shifts in quasars, it is conceivable that the decay red shift might contribute substantially to the total red shift observed in such objects, so that the cosmological red shift might be smaller than otherwise expected. This in turn would imply that the quasar was closer than predicted by the cosmological red shift alone, and so the quasar would not need to radiate energy at so prodigious a rate. An important difference between the decay and cosmological red shift is that over the lifetime $1/\gamma$ of the decay, the decay red shift would decrease and finally disappear, while the cosmological red shift would scarcely change (and *increase* over cosmological times).

There is of course no independent evidence as yet for the existence of the decay red shift. However, we will close by pointing out two items of information which tend to raise a question as to whether the quasar red shifts are really cosmological, as has been generally accepted, or whether they are in fact caused, at least in part, by dynamical processes peculiar to the quasar phenomenon itself.

First, Arp has reported numerous data supporting the possibility that noncosmological red shifts exist.[5,6] He has discovered evidence in several cases of quasars with very different red shifts that seem to be associated in the same physical cluster. In one case there actually appears to be a filament of matter connecting two quasars with different red shifts. If these cases are confirmed, it would imply the existence of a red shift which is neither cosmological nor gravitational in origin, but rather an intrinsic dynamical feature of the quasar process.

Second, the apparent absence of large quasar red shifts seems to imply a physical cutoff. This has been interpreted cosmologically as showing that astronomers are seeing the "edge of the universe." But in fact, in the absence of corroborative evidence, an equally likely explanation would be that the order of magnitude of quasar red shifts is determined by the dynamics of the quasar mechanism, which somehow provides for a maximum possible red shift.

[1]The motivation and formalism of the theory are explained in D. M. Greenberger, J. Math. Phys. 11, 2329 (1970). Some physical consequences of the theory, especially in regard to the uncertainty principle, are discussed in D. M. Greenberger, J. Math. Phys. 11, 2341 (1970). We shall refer to these papers as GI and GII, respectively.

[2]D. M. Greenberger, J. Math. Phys. 15, 406 (1974).

[3]The gauge invariance of the theory of GI was discussed in D.M. Greenberger, Ann. Phys. (N.Y.) 25, 290 (1963). The classic papers on the subject are C. N. Yang and R. Mills, Phys. Rev. 96, 191 (1954), and R. Utiyama, Phys. Rev. 101, 1597 (1956).

[4]See, for example, H. Goldstein, *Classical Mechanics* (Addison-Wesley, Cambridge, Mass., 1950), Chap. 8. We are using Goldstein's F_2 type of generating function.

[5]H. Arp, Science 174, 1189 (1971). A later report will appear in the proceedings of the Sixth Texas Symposium on Relativistic Astrophysics, New York, December 1972, to be published by the New York Academy of Science.

[6]The suggestion that proper time be considered a dynamical variable was made in a different context by E.A. Schucking and E. L. Spiegel, Comments Astrophys. Space Phys. 2, 121 (1970).

Wavepackets for particles of indefinite mass

Daniel M. Greenberger

Department of Physics, City College of the City University of New York, New York, New York 10031
(Received 24 October 1972; final revised manuscript received 11 September 1973)

Using a formalism that includes particles with indefinite and variable masses, we can create a much richer set of states than in conventional quantum mechanics. For example, in a gravitational field the equivalence principle implies that the motion of a particle is independent of the mass. Thus it is possible to create physical states with large Δm but very small Δv and Δx, so that the trajectory of a particle is well determined, while the $\Delta p \cdot \Delta x$ uncertainty relation is satisfied. We also provide a quantitative description of wavepackets with arbitrary distributions in momentum and mass (or alternatively velocity and mass) and discuss a simple diffraction slit experiment as an example of their physical propagation. The Schrödinger equation is solved for a decaying particle at rest. It is also shown that Galilean invariance plays a natural and cohesive role within the variable mass formalism, unlike the case in conventional quantum mechanics where the existence of multiple mass states would violate Galilean invariance.

1. INTRODUCTION

It has been shown that one can develop a theory of particles with variable mass.[1] With such a theory one can develop a classical description of decaying particles, and the theory also has the property that it includes the binding energy of a particle as part of its inertia. One can also extend the theory to include quantum mechanical effects. The mass m, and proper time τ, of a particle appear as conjugate dynamical variables in the theory, which is a direct extension of the classical canonical formalism, and so one would expect them to obey an uncertainty relation. It was shown in *GII* that this is indeed the case and that the existence of uncertainty relations between p and x, and E and t, necessitate one of the form

$$\Delta m \Delta \tau \geq \hbar \tag{1.1}$$

(we assume throughout this paper that $c \equiv 1$), and several examples of this relation are given there.

In Sec. 2 we indicate how it is possible to create wavepackets consisting of coherent superpositions of different masses, both experimentally and within the variable mass formalism. The motion must be consistent with the relations

$$\Delta p \cdot \Delta x \geq \hbar, \quad p = mv, \tag{1.2}$$

and yet there is still much more freedom available than in the conventional theory, where m is a fixed parameter, and Δp means $m \Delta v$. Specifically it is possible to have a situation where Δm is large and Δv is small, so that

$$\Delta p \approx v \Delta m. \tag{1.3}$$

While this situation does not come about in conventional nonrelativistic quantum theory, it can actually occur in two very important situations, and implies that one can devise an experimental arrangement where both Δx and Δv are small, and yet have Eq. (1.2) satisfied, because Δm is large. Furthermore, if Δx and Δv are both small, the trajectory of the particle is well known, a result which is not possible in conventional quantum theory.

First, such a situation can occur when a particle moves in an external gravitational field. Then the equivalence principle implies that all particles accelerate at the same rate and the lack of knowledge of m does not affect the motion of the particle. However, because

$$d\tau = (g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu)^{1/2}\, dt \sim [1 + \varphi(x) - \tfrac{1}{2} v^2]\, dt, \tag{1.4}$$

where φ is the gravitational potential, it follows that an accurate knowledge of v and x allows one to calculate τ very accurately, where τ is the proper time as read by a clock at rest on the particle. So one can achieve, in a gravitational field, a very accurate knowledge of x and v, and therefore of the trajectory, and of τ, without knowing anything at all about the mass of the particle.

The second such situation occurs when m itself is very large, and also Δm is very large. This is the situation normally encountered in a classical reference frame. By having m large, the laboratory apparatus can absorb momentum transfers from the system being studied without effectively being accelerated itself, and so its trajectory is accurately known. A similar case occurs when a classical particle is subject to a constraint, leaving it free to move in one direction (like a bead on a wire). Then, if the constraint is severe, so that Δx and Δv are small, Δm will be large because of the large uncontrollable energy, and therefore momentum, transfers between the particle and the constraint.

The quantitative description of wavepackets for variable mass particles is carried out in Sec. 3. Of course, there are a wide range of possibilities between the cases where Δp is due solely to Δv, and where it is due solely to Δm, although the extremes are more interesting. In order to point out the physical difference between the extreme cases, in Sec. 4 a slit experiment is analyzed in some detail.

In Sec. 5 we discuss the Schrödinger equation for a decaying particle at rest, and show its relation to the classical problem of decaying particle within the formalism.

In Sec. 6 we analyze the question of Galilean invariance of the theory. In conventional nonrelativistic quantum mechanics, one cannot have a superposition of different masses as it would violate Galilean invariance.[2] One sees this by transforming a free particle into a series of moving reference frames and finally back to the original frame. In the process, the wavefunctions belonging to particles of different masses pick up different phase factors. This would amount to having caused an observable effect through a series of unobservable transformations, and so one imposes on the system a superselection rule to eliminate superpositions of different masses, in order to avoid the problem.

However, it happens that this phase difference acquired by particles of different masses in the transformation to a moving frame is in our theory proportional to the proper time difference between the two frames. While in conventional nonrelativistic theories the proper time

has no place, in our variable mass formulation it plays a vital role. Thus the transformation to a moving reference frame and back is in fact quite observable, in that it leaves a measurable imprint on the system, in the form of the "twin paradox" effect of differential aging between the two systems, and the phase difference between different particles is just a measure of this real phenomenon. So the variable mass theory actually demands this phase difference and is consistent with Galilean invariance. (This phase difference is actually *required* by the uncertainty principle, for it is this phase that is responsible for "washing out" interference patterns caused by different proper times, when an accurate mass measurement is made.) Finally, we use the Galilean invariance of the theory to solve the Schrödinger equation for a free indefinite mass particle in motion.

2. PARTICLE DISTRIBUTIONS WITH INDEFINITE MASS AND VELOCITY

Our formalism enables us to construct wavepackets of indefinite mass. In conventional quantum theory one can do so relativistically, an example being the coherent combination of the K_1 and K_2 particles, whose rest masses differ by about 10^{-5} eV, or about one part in 10^{14}. However, in the conventional nonrelativistic theory there is a total dichotomy between mass and binding energy and one cannot superimpose different mass states.

In our theory this artificial nonrelativistic distinction between rest energy and binding energy is not maintained and so one can easily superimpose different mass wavepackets without destroying Galilean invariance. Furthermore, one can clearly trace the separate contributions to Δp coming from Δm and Δv.

Experimentally it is easy to create a beam of particles with a large Δm. For example, in a high energy baryon–antibaryon collision large numbers of meson of assorted charges and energies are produced. It is possible to sort out these mesons into a definite velocity distribution without knowing their mass. From Newton's law,

$$v(t) - v(0) = \int (F/m)\, dt, \tag{2.1}$$

and so a poor knowledge of m quickly transforms itself into a poor knowledge of v. However, there are two ways around this problem.

The first is through gravitational forces, since then F is proportional to m, and the acceleration will be the same for all particles, whether Δm is small or large. An example of the production of a velocity selected beam is given in Fig. 1a. By performing a time delay experiment, so that the time of passage, t_0, is known, one can select a definite value of v_x and v_y for the center of the beam, regardless of the mass of the particles. The distribution of velocities around the center of the beam, as determined by geometry, will also be independent of the mass of the particles, classically. (Because of the quantum mechanical spread of the packet, after the initial slit, the velocity distribution at the second slit will have some mass dependence, which can be minimized.)

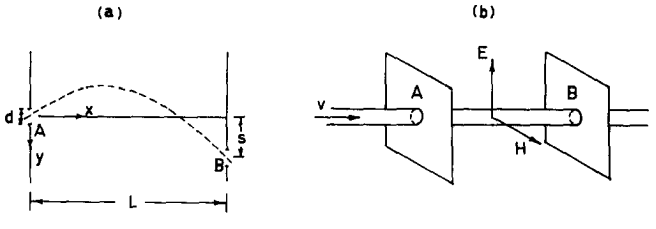

FIG. 1. Experimental velocity selectors. (a) Gravitational: Slit A is opened briefly, and t_0 seconds later slit B is opened briefly. Then at the center of the beam at B, $v_x = L/t_0$, $v_y = s/t_0 + \frac{1}{2} g t_0$, independently of the mass of the particles in the beam. Even at the edges of the beam, the geometrical cutoff will be independent of the mass of the particles (except for quantum mechanical spreading effects, which can be minimized). (b) Nongravitational: The velocity at the center of the beam passing through B is $v = cE/H$, independently of the mass of the particles. However the velocity of particles at the outer edge of the beam at B will be mass dependent.

A second method for selecting a velocity distribution is to have two equal and opposite velocity dependent nongravitational forces, for example those produced by an electric field perpendicular to a magnetic field, both perpendicular to the particle beam, as in Fig. 1b. This procedure selects charged particles of a given velocity. As shown, the beam will be contaminated by neutral particles of all velocities, an effect which could be easily eliminated by a slight effort. The disadvantage of this second technique is that only the center of the beam has the correct velocity, independently of the mass, while near the edges there will be some mass dependence even classically. A gravitational velocity selector is free of this defect.

The uncertainty in mass, besides contributing to an uncertainty in momentum, also must satisfy a relationship of the form of Eq. (1.1). The proper time is the time that would be read on a clock fixed to the particle. The proper time may be uncertain because the particle has an uncertain velocity, or is subject to an uncertain gravitational potential, from Eq. (1.4). This proper time uncertainty represents the inability of an observer fixed in the laboratory to read or calculate the time read by a clock on the particle, and as such is the inability to read, in one reference frame, a clock at rest in another frame–even though an observer moving with the clock in this other frame may be able to read it accurately. As such, it is *not* an uncertainty in time interval in one reference frame, as is Δt. Several examples illustrating the meaning of $\Delta\tau$ are given in GII.

One can obtain a more quantitative measure of the uncertainty in mass, velocity, and momentum. Our formalism is merely an extension of the usual canonical one, the principle difference being that there is an extra degree of freedom, τ. The details of the formalism can be found in GI and GIII. Proceeding analogously to conventional quantum mechanics, we can represent a particle by the wavepacket

$$\psi(x, \tau, t) = (2\pi)^{-2} \int d^3k\, dm\; a(k, m)\, e^{i(k\cdot x - \omega t + m\tau)} \tag{2.2}$$

where nonrelativistically

$$p = \hbar k \equiv k \tag{2.3}$$

(we will assume $\hbar \equiv 1$, unless otherwise noted). In the special case where the distributions in mass and momenta are independent of each other

$$a(k, m) = a_1(k)\, a_2(m), \tag{2.4}$$

there exists a simple relationship between the uncertainties in momentum, mass, and velocity, represented by the wavepacket. In that case, if one defines

$$\mu \equiv 1/m, \qquad v = k\mu, \tag{2.5}$$

then one has

$$\bar{v} = \bar{k}\bar{\mu},$$

$$\overline{(\Delta v)^2}/\bar{v}^2 = \overline{(\Delta k)^2}/\bar{k}^2 + \overline{(\Delta\mu)^2}/\bar{\mu}^2 + \overline{(\Delta k)^2}\,\overline{(\Delta\mu)^2}/\bar{k}^2\bar{\mu}^2, \tag{2.6}$$

where the bar denotes the expectation value.

This result simplifies even further if the distribution is narrow,

$$\overline{(\Delta k)^2} \ll \bar{k}^2, \quad \overline{(\Delta m)^2} \ll \bar{m}^2. \tag{2.7}$$

Then one has

$$\bar{\mu} \approx (1/\bar{m})(1 + \overline{(\Delta m)^2}/\bar{m}^2),$$
$$\overline{(\Delta\mu)^2} \approx \overline{(\Delta m)^2}/\bar{m}^4, \tag{2.8}$$

and

$$\bar{v} \approx (\bar{k}/\bar{m})(1 + \overline{(\Delta m)^2}/\bar{m}^2),$$
$$\overline{(\Delta v)^2}/\bar{v}^2 \approx \overline{(\Delta k)^2}/\bar{k}^2 + \overline{(\Delta m)^2}/\bar{m}^2. \tag{2.9}$$

Of course, if the distributions in m and k are not independent, these conditions can be violated. For example, if $v = v_0$, with $\Delta v = 0$, then $k = mv_0$, and $\Delta k = v_0 \Delta m$. In this case Δk can be finite or infinite, depending on Δm, while $\Delta v = 0$.

3. QUANTITATIVE DISCUSSION OF WAVEPACKETS

We would like to analyze the physical behavior of the wavepacket of Eq. (2.2), containing a distribution in both momentum and mass. The simplest situation occurs when the distribution is peaked about some central value (k_0, m_0) and has a small spread in both k and m, so that

$$\Delta k \ll k_0, \quad \Delta m \ll m_0. \tag{3.1}$$

Then one may use the group velocity approximation, to find the behavior of the packet for $x = f(t), \tau = f(t)$. The canonical equations of motion,

$$\dot{x} = \partial H/\partial p, \quad \dot{\tau} = \partial H/\partial m, \tag{3.2}$$

together with the fact that the distribution $a(k, m)$ of Eq. (2.4) is sharply peaked enable us to use the expansion

$$\omega(k, m) \equiv H(p, m)/\hbar$$
$$= \omega_0(k_0, m_0) + v_0(k - k_0) + \dot{\tau}_0(m - m_0), \tag{3.3}$$

where

$$v_0 \equiv (\partial\omega/\partial k)_0, \quad \dot{\tau}_0 = (\partial\omega/\partial m)_0, \tag{3.4}$$

so that the phase φ of the exponential of Eq. (2.2) becomes

$$\varphi = k_0 \cdot x - \omega_0 t + m_0\tau + (k - k_0)\cdot(x - v_0 t) + (m - m_0)(\tau - \dot{\tau}_0 t). \tag{3.5}$$

Thus, if at $t = 0$, the wavefunction has the form

$$\psi(x, \tau, 0) = e^{i(k_0 x + m_0\tau)}\psi_1(x)\psi_2(\tau), \tag{3.6}$$

where

$$\psi_1(x) = (2\pi)^{-3/2} \int d^3k\, a_1(k)\, e^{i(k-k_0)\cdot x},$$
$$\psi_2(\tau) = (2\pi)^{-1/2} \int dm\, a_2(m)\, e^{i(m-m_0)\tau}, \tag{3.7}$$

then at time t it will have the form

$$\psi(x, \tau, t) = e^{i(k_0 \cdot x - \omega_0 t + m_0\tau)}\psi_1(x - v_0 t)\psi_2(\tau - \dot{\tau}_0 t), \tag{3.8}$$

and it will obviously satisfy the $x - p$ and $m - \tau$ uncertainty relations.

As a special case, for a relativistic free particle, where

$$\omega = (m^2 + k^2)^{1/2}, \tag{3.9}$$

we have

$$v_0 = (\partial\omega/\partial k)_0 = k_0(m_0^2 + k_0^2)^{-1/2},$$
$$k_0 = m_0 v_0 \gamma_0, \quad \gamma_0 = (1 - v_0^2)^{-1/2}, \tag{3.10}$$

and

$$\dot{\tau}_0 = (\partial\omega/\partial m)_0 = m_0(m_0^2 + k_0^2)^{-1/2} = \gamma_0^{-1} = (1 - v_0^2)^{1/2}. \tag{3.11}$$

Similarly, in the nonrelativistic limit

$$\omega = m + k^2/2m, \tag{3.12}$$

and

$$v_0 = (\partial\omega/\partial k)_0 = k_0/m_0,$$
$$\dot{\tau}_0 = (\partial\omega/\partial m)_0 = 1 - k_0^2/2m_0^2 = 1 - v_0^2/2. \tag{3.13}$$

In writing these equations, no provision was made for the restriction $m \geq 0$. It is not clear what role this restriction plays in theoretical physics, including gravitational theory. It is formally equivalent to the restriction on τ that it be analytic in the lower half-plane, but we shall generally not include it unless it seems necessary to do so.

An instructive example, in the opposite extreme limit, is the case where $\Delta m = \infty$. We shall show that the formalism permits the position and velocity to both be precisely known, so that the particle has a perfectly localized trajectory. We have, in this case,

$$v = v_0, \quad \Delta v = \Delta x = \Delta\tau = 0, \quad \Delta k = \Delta m = \infty. \tag{3.14}$$

We assume for convenience a one-dimensional particle with a Dirac δ-function spatial wavefunction. The entire wavepacket has uniform velocity v_0. Under these conditions τ is completely determined, and so

$$\tau = t/\gamma_0 = t(1 - v_0^2)^{1/2}, \tag{3.15}$$

while the mass is totally unknown. At $t = 0$, we have

$$\psi(x, 00) = (2\pi)^{-1} \int dk\, e^{ikx}, \tag{3.16}$$

and so at later times

$$\psi(x, \tau, t) = (2\pi)^{-1} \int dk\, e^{i(kx - \omega t + m\tau)}. \tag{3.17}$$

Under the stated conditions,

$$k = mv_0\gamma_0, \quad \tau = t/\gamma_0,$$
$$\omega = (k^2 + m^2)^{1/2} = k/v_0, \tag{3.18}$$

so that

$$\psi(x, \tau, t) = (2\pi)^{-1} \int dk\, e^{ik(x - t/v_0 + t/v_0\gamma_0^2)} = \delta(x - v_0 t). \tag{3.19}$$

It should be noted that this is the exact wavefunction for all time. Since $\Delta v = 0$, there will be no spreading of the wavepacket, unlike the usual case for particles with a definite mass.

If one wants a packet with a controlled spread in velocity, this can be obtained by using a correlated, rather than independent, spread in k and m. For example, in a distribution of the form

$$a(k, m) = \int dv\; a_1(k)\, a_2(v)\, a_3((k/m) - v) \tag{3.20}$$

the function can be integrated first over k, where then the functions a_1 and a_3 form a convolution giving the dis-

tribution in m and v, or it can be integrated over v first, where it correlates the k and m behavior.

As an example of Eq. (3. 20), we choose a_2 to be a δ function centered about v_0, and a_3 to be slowly varying. We take the case of a relativistic, one-dimensional particle, so that

$$\psi(x,\tau,t) = (2\pi)^{-1}\int dk\,(k^2+m^2)^{-1/2}\,dm\,k^{1/2}\,a_1(k)\,a_3((k/m)-v_0\gamma_0)\exp[kx-(k^2+m^2)^{1/2}t+m\tau] \quad (3.21)$$

where $\gamma_0 = (1-v_0^2)^{-1/2}$. The extra $k^{1/2}$ in the numerator is purely for convenience. Then one can perform the m integral by the method of steepest descent. Denoting the phase by φ

$$\varphi = kx-(k^2+m^2)^{1/2}t+m\tau, \quad (3.22)$$

we must evaluate it at the stationary point $m = m_0$, defined by

$$m_0 = m_0(k,\tau,t), \qquad (\partial\varphi/\partial m)_{m_0} = 0. \quad (3.23)$$

According to this,

$$\tau-(\partial\omega/\partial m)_0 t = \tau - m_0 t(m_0^2+k^2)^{-1/2} = 0,$$
$$m_0 = kf(\tau,t) = k((t^2/\tau^2)-1)^{-1/2}. \quad (3.24)$$

This equation is consistent with Eq. (3. 4). For each value of k, a best value of m, m_0, or equivalently a best velocity is chosen. The range over which different m_0's contribute for a given k is determined by the function a_3.

One can bring this out more clearly by introducing the velocity as an independent variable, $v(\tau,t)$, using Eq. (3. 24),

$$k = m_0 v\gamma, \qquad \gamma \equiv (1-v^2)^{-1/2}, \qquad v\gamma = 1/f(\tau,t), \quad (3.25)$$

so that

$$v(\tau,t) = [1-(\tau^2/t^2)]^{1/2}, \qquad \tau = (1-v^2)^{1/2}t. \quad (3.26)$$

Then the function a_3 in Eq. (3. 21) becomes

$$a_3((k/m)-v_0\gamma_0) = a_3(v\gamma - v_0\gamma_0) \equiv \psi_3(\tau - t/\gamma) \quad (3.27)$$

and it can be taken out of the integral. The phase of Eq. (3. 22) becomes, when expanded about m_0,

$$\varphi(k,\tau,t) = kx-\omega(m_0,k)\,t+m_0\tau-\tfrac{1}{2}(m-m_0)^2(\partial^2\omega/\partial m_0^2)\,t,$$
$$\partial^2\omega/\partial m_0^2 = k^2(m_0^2+k^2)^{-3/2} = v^3/k. \quad (3.28)$$

The Gaussian integral over m then yields

$$\psi = [\psi_3(\tau-t/\gamma)/(2\pi tv)^{1/2}]\int dk\;a_1(k)\,e^{i(kx-\omega(m_0,k)t+m_0\tau)}, \quad (3.29)$$

where $v = v(\tau,t)$ from Eq. (3. 26).

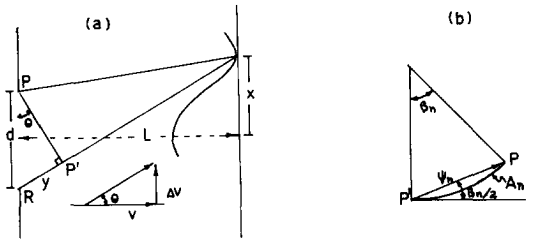

FIG. 2. Diffraction through a slit. (a) For a beam component of mass m_n, a minimum is determined by the condition $y = \lambda_n$. The total wavefunction is $\sum\psi_n$. (b) Determination of the amplitude. If the entire beam component along PR were in phase, the amplitude would be that of A_n. If the phases at P and P' differ by β_n, the amplitude will be that of ψ_n.

Finally, by replacing m_0 by its value from Eq. (3. 25), the wavefunction becomes

$$\psi = \psi_3\,(2\pi tv)^{-1/2}\int dk\;a_1(k)\,e^{ik(x-vt)}$$
$$= (tv)^{-1/2}\psi_3(\tau - t/\gamma_0)\psi_1(x-vt),$$

where ψ_1 is the Fourier transform of a_1 and v is again $v(\tau,t)$. We note that it would not be consistent to take ψ_3 to be a δ function in this example, as then we could not have used the steepest descent technique.

We see that one may construct wavepackets in m and τ to any desired degree of sharpness consistent with the uncertainty principle, in complete analogy to wavepackets in k and x.

4. A DIFFRACTION EXPERIMENT

In order to point up more clearly the distinction between a momentum distribution due to a coherent superposition of different masses, and one due to different velocity states of the same mass, we will analyze a simple diffraction pattern produced on a screen by a beam of particles passing through a slit. If the incident beam is considered to be approximately a plane wave, the resulting diffraction pattern will be as shown in Fig. 2a.

Assume that the incident beam has a unique velocity v, but is a coherent mixture of states of different mass, as described in the previous section. It may have been produced by a high energy collision, and selected for velocity, or it may simply be a coherent combination of many different spin states of a single particle, precessing in a high magnetic field, separated by mass $\Delta m = \mu H$. At any rate, we shall assume that the incident beam has somehow acquired a single nonrelativistic velocity v. The various components of the beam will have masses m_n, momenta $p_n = m_n v$, and wavelength

$$\lambda_n = 1/m_n v. \quad (4.1)$$

If there is a screen placed a distance L from the slit, then for a given component of the beam, of mass m_n, the amplitude of the wave a distance x from the center of the beam will be determined by the different phases present along the line PP' as the beam leaves the slit. If the beam at P' is out of phase with that at P by an angle β_n, and the total amplitude of the in-phase beam PR is given by A_n, as measured along the arc PP' in Fig. 2b, then one can determine the wavefunction ψ_n as in elementary optics,

$$\psi_n = e^{i\beta_n/2}(A_n\sin\beta_n/2)/(\beta_n/2)$$
$$= i(A_n/\beta_n)(1-e^{i\beta_n}), \quad (4.2)$$

where

$$y/\lambda_n = \beta_n/2\pi, \qquad \sin\theta = y/d, \quad (4.3)$$

or equivalently,

$$\beta_n = \kappa_n\sin\theta, \qquad \kappa_n = m_n v d. \quad (4.4)$$

The wavefunction ψ_n can be normalized to unity to determine A_n,

$$\int|\psi_n|^2\,dx = 1, \qquad A_n = A_n(\kappa_n, L). \quad (4.5)$$

For a coherent mixture of many masses, one has

$$\psi = \sum a_n\psi_n. \quad (4.6)$$

In order to be able to carry through the calculation explicitly, we assume for simplicity that there are N equally spaced states, where $N \gg 1$, so that

$$m_n = nm_0, \quad \beta_n = n\beta, \quad \beta = \kappa \sin\theta, \quad \kappa = m_0 v d. \tag{4.7}$$

Then

$$\psi = (i/\beta) \sum_n (a_n A_n/n)(1 - e^{in\beta}). \tag{4.8}$$

Finally, for our particular problem we will choose (purely for calculational convenience)

$$a_n A_n = nae^{2\pi in/N}, \tag{4.9}$$

which makes ψ easy to evaluate,

$$\psi = (ia/\beta) \sum (e^{2\pi in/N} - e^{in(\beta+2\pi/N)}). \tag{4.10}$$

The first terms in Eq. (4.10) sum to 0, and the second terms give

$$|\psi|^2 = \frac{a^2}{\beta^2} \frac{\sin^2(N\beta/2)}{\sin^2(\beta/2 + \pi/N)} \approx \frac{a^2}{\beta^2} \frac{\sin^2(N\beta/2)}{\sin^2(\beta/2)}, \tag{4.11}$$

and therefore the wave pattern on the screen will have minima determined by

$$N\beta/2 \sim l\pi, \quad \sin\theta \sim \theta \sim 2\pi l/N\kappa. \tag{4.12}$$

Of course, only for our special choice of amplitudes, Eq. (4.9), will we have complete cancellation; however, it is characteristic that there will be large fluctuations in the intensity whose minima will be separated by this amount,

$$\Delta\theta \sim 2\pi/N\kappa. \tag{4.13}$$

For incoherent states of different mass, one would have, instead of Eq. (4.6),

$$|\psi|^2 = \sum |\psi_n|^2, \tag{4.14}$$

and this would be governed by the envelope for ψ_1, for which

$$\theta \sim 2\pi/\kappa, \tag{4.15}$$

so that the pattern would be N times as large.

So we see that the great difference between the two cases is that for the coherent superposition of different masses,

$$\Delta v/v \sim \Delta\theta \sim 2\pi/N\kappa, \tag{4.16}$$

a very narrow spread, even though the momentum spread is still governed by the same uncertainty principle, where $\Delta x \sim d$, $\Delta p \sim Nm\Delta v$, for the heavier components and

$$\Delta p\, \Delta x \sim Nm\, \Delta v d \sim Nmv\, \Delta\theta\, d \sim 1. \tag{4.17}$$

The general feature of the calculation is that a coherent mass spread tends to sharpen the velocity distribution of the beam. Thus, for example, one would expect that for a very unstable particle with a large mass spread, like the ρ meson, the diffraction scattering off nucleons would be somewhat narrower in angular spread, than if the ρ were stable.

5. THE SCHRÖDINGER EQUATION FOR A DECAYING PARTICLE

From the wavepacket representation of a particle, Eq. (2.2), it is apparent that the introduction of τ as an independent variable implies that in τ space there exists a mass operator (we temporarily set $\hbar \neq 1$),

$$m_{op} \equiv (\hbar/i)\, \partial/\partial\tau. \tag{5.1}$$

This allows one to immediately write a Schrödinger equation for a free particle,

$$(m_{op}^2 - \hbar^2\nabla^2)\psi = -\hbar^2 \partial^2\psi/\partial t^2. \tag{5.2}$$

In the presence of a gravitational field, in the vector gauge theory of GI and GIII, one makes the replacement

$$\partial_\mu \to (\partial_\mu - B_\mu\, \partial/\partial\tau), \tag{5.3}$$

and Eq. (5.2) becomes (with $p^\mu = ih\partial^\mu$)

$$-\hbar^2(\partial_\mu - B_\mu\, \partial/\partial\tau)(\partial^\mu - B^\mu\, \partial/\partial\tau)\psi = m_{op}^2\psi. \tag{5.4}$$

In the nonrelativistic limit, we have

$$\begin{aligned} &\partial_\mu - B_\mu\, \partial/\partial\tau \to (\partial/\partial t + \varphi\partial/\partial\tau, \nabla), \\ &(m_{op} + p^2/2m_{op})\psi + V(x)\psi = -(\hbar/i)\, \partial\psi/\partial t - \varphi m_{op}\psi. \end{aligned} \tag{5.5}$$

In Eq. (5.5) we have included, besides a gravitational potential $\varphi(x, \tau, t)$, also an external nongravitational potential, $V(x)$. This equation can also be written

$$(m_{op} + \tfrac{1}{2}p^2 + m_{op}\varphi\, m_{op})\psi = -(\hbar/i)\, m_{op}\, \partial\psi/\partial t, \tag{5.6}$$

or

$$[\partial^2/\partial\tau^2 + \tfrac{1}{2}\nabla^2 + (\partial/\partial\tau)(\varphi\partial/\partial\tau) + (i/\hbar)\, V\, \partial/\partial\tau]\psi = -(\partial^2/\partial t\partial\tau)\psi. \tag{5.7}$$

The potential B_μ, or φ, is subject to the gauge condition (see GIII)

$$\partial_\mu B^\mu - B_\mu \partial B^\mu/\partial\tau = 0 \to \partial\varphi/\partial t + \varphi\, \partial\varphi/\partial\tau = 0. \tag{5.8}$$

It was shown in GIII that when a particle decays, it must be coupled to a τ-dependent gravitational field. The gravitational field does not *cause* the decay, but, by coupling to the inertia of the particle, it produces the necessary change in mass.

In this section we will treat the case of a particle decaying from rest under the influence of the potential $\varphi(\tau, t)$, with no x dependence. The classical case is solved in Sec. 3 of GIII. We will assume that φ is given such as to cause a particular $m(t)$ dependence classically, i.e., $\varphi(\tau, t)$ would give rise to a decay

$$m(t) = m_0 e^{-\varphi(\tau(t), t)} = m_0 e^{-\varphi(t)}, \tag{5.9}$$

and we examine what this will imply quantum mechanically. In this case, Eq. (5.7) reduces to

$$(1 + \varphi)(\partial\psi/\partial\tau) = -\partial\psi/\partial t. \tag{5.10}$$

We proceed by making use of the classical solution. First we note that the gauge condition, Eq. (5.8), has the solution

$$\tau = t\varphi + f(\varphi). \tag{5.11}$$

The choice of the function $f(\varphi)$ determines the particular time dependence of the decay, Eq. (5.9). From Eq. (5.11) we see that

$$\varphi' \equiv \partial\varphi/\partial\tau = (\partial f/\partial\varphi + t)^{-1} \tag{5.12}$$

and then we can use this to eliminate the variable τ in favor of φ, transforming from $\psi(\tau, t)$ to $\psi(\varphi, t)$.

In order to accomplish this, we note, using Eq. (5. 8),

$$\left(\frac{\partial\psi}{\partial\tau}\right)_t = \left(\frac{\partial\varphi}{\partial\tau}\right)_t \left(\frac{\partial\psi}{\partial\varphi}\right)_t = \varphi' \frac{\partial\psi}{\partial\varphi},$$
$$\left(\frac{\partial\psi}{\partial t}\right)_\tau = \left(\frac{\partial\varphi}{\partial t}\right)_\tau \left(\frac{\partial\psi}{\partial\varphi}\right)_\tau + \left(\frac{\partial\psi}{\partial t}\right)_\varphi = -\varphi\varphi' \frac{\partial\psi}{\partial\varphi} + \frac{\partial\psi}{\partial t}. \tag{5. 13}$$

Then the Schrödinger equation (5. 10) becomes

$$(1+\varphi)\varphi' \frac{\partial\psi}{\partial\varphi} = -\left(\frac{\partial\psi}{\partial t} - \varphi\varphi' \frac{\partial\psi}{\partial\varphi}\right),$$
$$\varphi' \frac{\partial\psi}{\partial\varphi} = -\frac{\partial\psi}{\partial t}. \tag{5. 14}$$

Finally we change variables again, this time eliminating φ in favor of z, a variable which conveniently characterizes the classical solution, defined by

$$z = z(\varphi, t) = te^{-\varphi} - \int^{\varphi} (\partial f/\partial\varphi) e^{-\varphi} d\varphi. \tag{5. 15}$$

As was shown in GIII, the classical solution is described by

$$z = \text{const} \equiv z_0, \tag{5. 16}$$

and along the line $z = z_0$,

$$\left(\frac{\partial\varphi}{\partial t}\right)_z = \frac{d\varphi}{dt}, \quad \varphi(t) = \int_{z=z_0} \frac{\partial\varphi}{\partial t} dt. \tag{5. 17}$$

Equation (5. 15) gives

$$(\partial z/\partial\varphi)_t = e^{-\varphi}/\varphi', \quad (\partial z/\partial t)_\varphi = e^{-\varphi}, \tag{5. 18}$$

and

$$\left(\frac{\partial\psi}{\partial\varphi}\right)_t = \frac{e^{-\varphi}}{\varphi'} \frac{\partial\psi}{\partial z},$$
$$\left(\frac{\partial\psi}{\partial t}\right)_\varphi = e^{-\varphi} \frac{\partial\psi}{\partial z} + \frac{\partial\psi}{\partial t}. \tag{5. 19}$$

So finally, the Schrödinger equation (5. 14) becomes

$$-e^{-\varphi}(\partial\psi/\partial z) = -(\partial\psi/\partial t + e^{-\varphi}\partial\psi/\partial z), \quad \partial\psi/\partial t = 0. \tag{5. 20}$$

The solution of this equation, of course, is

$$\psi(z, t) = \psi(z), \tag{5. 21}$$

and the physical interpretation of the solution is quite simple. If the wavepacket $\psi(z - z_0)$ is narrow, and peaked around $z = z_0$, it will remain that way throughout time, and the time dependent motion of the system will be given by

$$\psi = \psi(z - z_0) = \psi(te^{-\varphi} - \int^{\varphi} (\partial f/\partial\varphi) e^{-\varphi} d\varphi - z_0), \tag{5. 22}$$

which will be nonzero only in the neighborhood of the region where φ varies as the correct classical function of t (the region where $z \sim z_0$).

The gauge condition plays a strong role in deriving the classical solution, and it in turn stems from the vector relativistic formulation of the theory. It is not at all clear what, if any, physical significance this implies in our nonrelativistic theory, although it greatly reduces the arbitrariness of the potential $\varphi(\tau, t)$.

6. GALILEAN INVARIANCE

In ordinary nonrelativistic quantum mechanics, where the mass of a particle is a fixed parameter, one cannot have a free particle in a superposition of states of different mass. Partly, this is because of the artificial separation, nonrelativistically, of rest energy and binding energy as qualitatively different concepts, and the failure to recognize proper time at all. Relativistically, the restriction does not occur, while nonrelativistically it manifests itself as a superselection rule demanded by the Galilean invariance of the theory.[2]

Basically, the argument against the superposition of different masses runs as follows. If one has a state which is a coherent superposition of two different mass particles, one may make a succession of transformations to coordinate systems moving at different velocities, ultimately returning to the original system. This process results in a new coherent superposition of states, different from the original one. One then argues that by having merely performed a series of coordinate changes on the system, without having affected the system physically, one has yet managed to produce a different physical state. Thus one concludes that such a superposition is ambiguous and excludes it *a priori*.

We shall show that one can reinterpret the argument in a formalism where mass is a dynamical variable and that Galilean invariance poses no obstacle to combining states of different masses in such a formalism. First we review the argument in conventional quantum mechanics. We note that a succession of velocity changes is really a series of accelerations. So we ask the effect of subjecting a rigid Galilean reference frame to an arbitrary acceleration, $\mathbf{r} \to \mathbf{r}_1$,

$$\mathbf{r}_1 = \mathbf{r} - \xi(t). \tag{6. 1}$$

To effect such a transformation, we convert from the coordinates $(\mathbf{r}, t)$ to $(\mathbf{r}_1, t)$,

$$\nabla = \nabla_1, \quad \partial/\partial t)_r = \partial/\partial t)_{r_1} - \dot{\xi}\cdot\nabla_1. \tag{6. 2}$$

Then the Schrödinger equation

$$-(\hbar/2m)\nabla^2\psi = -(\hbar/i)(\partial\psi/\partial t), \tag{6. 3}$$

becomes

$$-(\hbar^2/2m)\nabla_1^2\psi = -(\hbar/i)(\partial\psi/\partial t) - \dot{\xi}\cdot\nabla_1\psi). \tag{6. 4}$$

If one introduces the new wavefunction $u(\mathbf{r}_1, t)$ by a phase change,

$$\psi = u(r_1, t)\exp[i(m/\hbar)(\dot{\xi}\cdot\mathbf{r}_1 + \tfrac{1}{2}\int_0^t \dot{\xi}^2 dt)], \tag{6. 5}$$

then u obeys the equation

$$-\frac{\hbar^2}{2m}\nabla_1^2 u + m\ddot{\xi}\cdot\mathbf{r}_1 u = -\frac{\hbar}{i}\frac{\partial u}{\partial t}. \tag{6. 6}$$

Thus one has made a unitary transformation to a new system where there is an effective gravitational potential,

$$\varphi_{\text{eff.}} = m\mathbf{g}(t)\cdot\mathbf{r}, \quad \mathbf{g}(t) = \ddot{\xi}, \tag{6. 7}$$

where, of course, $\mathbf{g}$ need not be constant in time.

Now if one takes the system around a closed path $\xi(t)$, taking time T, such that

$$\xi(0) = \xi(T) = 0,$$
$$\dot{\xi}(0) = \dot{\xi}(T) = 0, \tag{6. 8}$$
$$\ddot{\xi}(0) = \ddot{\xi}(T) = 0,$$

then one has

$$\psi(T) = u \exp[i(m/2\hbar) \int_0^T \dot{\xi}^2\, dt] \equiv u e^{im\eta(T)}, \quad (6.9)$$

where u is the same solution of the free particle equation, both at $t = 0$ and $t = T$.

Therefore, if we have a superposition of states, u_1 of mass m_1, and u_2 of mass m_2, such that

$$\psi(0) = u_1 + u_2, \quad (6.10)$$

then at $t = T$,

$$\psi(T) = u_1(T) e^{im_1\eta(T)} + u_2(T) e^{im_2\eta(T)}, \quad (6.11)$$

while if we had not made our excursion into an accelerated system, we would have had

$$\psi_0(T) = u_1(T) + u_2(T). \quad (6.12)$$

Thus, by merely transporting ourselves into and back out of an accelerated system, we have altered the relative phase of the states, which would clearly be an observable effect. We thus conclude that such a superposition of states does not occur in nature.

Let us now carry through the same argument for a particle in a dynamic mass theory. The only difference between the two cases will be that now m is an operator, given by Eq. (5. 1), so that if the original wavefunction is $\psi(r, \tau, t)$, then Eq. (6. 9) becomes

$$\begin{aligned}\psi(r, \tau, t) &= \exp(\tfrac{1}{2} \int_0^T \dot{\xi}^2 dt\, \partial/\partial\tau)\, u(r_1, \tau, T) \\ &= u(r, \tau + \tfrac{1}{2} \int_0^T \dot{\xi}^2 dt, T)\end{aligned} \quad (6.13)$$

where, at $t = T$, $\mathbf{r}_1 = \mathbf{r}$. This is to be compared with the state in the original unaccelerated system,

$$\psi_0 = u(r, \tau, T). \quad (6.14)$$

The difference between ψ and ψ_0 is merely an expression of the fact that proper time runs at a different rate in the accelerated and unaccelerated systems. In fact since, relativistically,

$$d\tau = (1 - v^2)^{1/2}\, dt \approx (1 - \tfrac{1}{2} v^2)\, dt, \quad (6.15)$$

it follows that in the laboratory system

$$\tau_0 = T = \tau + \tfrac{1}{2} \int v^2 dt, \quad (6.16)$$

where in our problem, $v = \dot{\xi}$.

And so, far from negating the possibility of a superposition of masses, the formalism produces just that time difference effect which is physically required by relativity, to this order. Specifically, this is due to the fact that proper time enters our formalism explicitly, while the standard formalism does not recognize the concept of proper time as distinct from coordinate time in the nonrelativistic limit.

It should also be pointed out that the phase factor between the two beams is very important in maintaining the uncertainty relation $\Delta m\, \Delta\tau > \hbar$, in much the same way as the phase factor e^{ikx} prevents violation of the $p - x$ uncertainty relation in measurements made on spatially separated beams.

Finally, let us use the Galilean invariance of the theory to convert the solution of the Schrödinger equation for a particle decaying at rest, Eq. (5. 10), into one representing a free particle with velocity v decaying. The Schrödinger equation for the moving particle will be Eq. (5. 7), where there is no external force ($V = 0$).

According to Eq. (6. 5), the solution ψ to Eq. (5. 7) should be represented by

$$\begin{aligned}\psi(r, \tau, t) &= e^{(\mathbf{v}\cdot\mathbf{r} - 1/2\, v^2 t)\partial/\partial\tau}\, u(\tau, t) \\ &= u(\tau + \mathbf{v}\cdot\mathbf{r} - \tfrac{1}{2} v^2 t, t),\end{aligned} \quad (6.17)$$

where $u(\tau, t)$ is a solution of Eq. (5. 10), representing a decaying particle at rest, and

$$\mathbf{r}_1 = \mathbf{r} - \mathbf{v}t. \quad (6.18)$$

This should be a complete solution to the problem. To prove it, note that (with $' \equiv \partial/\partial\tau$)

$$\psi' = u', \quad \nabla\psi = \mathbf{v}u', \quad \frac{\partial\psi}{\partial t} = \frac{\partial u}{\partial t} - \frac{v^2 u'}{2}, \quad (6.19)$$

so that Eq. (5. 7) becomes

$$\begin{aligned}u'' + \frac{v^2}{2} u'' + (\varphi u')' &= -\left(\frac{\partial u'}{\partial t} - \frac{v^2}{2} u''\right), \\ \left(u' + \varphi u' + \frac{\partial u}{\partial t}\right)' &= 0,\end{aligned} \quad (6.20)$$

which is just Eq. (5. 10) for the function u.

[1] D. M. Greenberger, J. Math. Phys. 11, 2329, 2341 (1970), and the paper preceding this one, J. Math. Phys. 15, 395 (1974). We refer to these papers as GI, GII, and GIII, respectively.

[2] The paper that explained the significance of Galilean invariance in quantum mechanics was V. Bargmann, Ann. Math. 59, 1 (1954). See also F. A. Kaempffer, *Concepts in Quantum Mechanics* (Academic, New York, 1965), Appendix 7.

A property of asymptotically flat, static, vacuum space-times*

William Hallidy

California State College, California, Pennsylvania 15419
(Received 24 August 1973)

It is shown that a necessary condition for stationary, asymptotically flat, vacuum space-times to be static is that the Weyl tensor be electric type to asymptotic order r^{-7}, where r is an affine parameter along null geodesics.

I. INTRODUCTION

A space-time is said to be stationary if it possesses a time-like vector field which is a solution of Killing's equations. If this vector field is also hypersurface orthogonal then the space-time is said to be static. As is well known, these geometrical conditions on a space-time may be exploited to pick a preferred coordinate system based on space-like hypersurfaces in which the metric takes an especially simple form.[1]

Newman and Penrose have developed a formalism for general relativity theory in which a preponderant role is played by the Weyl tensor rather than the metric tensor.[2] It is assumed that the reader is familiar with this formalism. Their results are most naturally expressed in a coordinate system based on null hypersurfaces. An asymptotic form of their results was developed in later papers.[3,4] The conditions for a vacuum space-time to be stationary were then expressed as conditions on the coefficients of an asymptotic expansion of the Weyl tensor. Here it is shown that subsidiary conditions are imposed on the asymptotically expanded Weyl tensor by the requirement that the space-time be static. These conditions are shown to require that the Weyl tensor components be electric type to the asymptotic order investigated. This result is in agreement with the long standing assumption that for any asymptotically flat, static, vacuum space-time the Weyl tensor is electric type.[5]

In Sec. II the conditions for a space-time to be stationary are reviewed. The investigation is restricted here to asymptotically flat, vacuum space-times for which the Killing vector is asymptotically time-like. In Sec. III the conditions for these space-times to be static are imposed in the Newman-Penrose formalism and the conditions on the Weyl tensor mentioned above are derived. In Sec. IV some concluding remarks are made.

II. STATIONARY CONDITIONS IN NEWMAN-PENROSE FORMALISM

In this formalism the Weyl tensor components are expressed in terms of five complex functions, ψ_A $(A = 0, \ldots, 3)$, defined on the space-time manifold. In a null coordinate system the metric tensor components take the form[3]

$$g^{0\mu} = \delta^\mu_0, \quad g^{1k} = X^k - (\xi^k\bar{\omega} + \bar{\xi}^k\omega), \qquad g^{11} = 2(u - \omega\bar{\omega}), \quad g^{ke} = -(\xi^k\xi^e + \bar{\xi}^k\bar{\xi}^e), \tag{1}$$

where $k, l = (2, 3)$ and $(x^0, x^1) = (u, r)$ in which u labels outgoing null hypersurfaces and r is an affine parameter along null geodesics on the hypersurfaces. The condition that the space-time be stationary is equivalent to the condition that there exists a null coordinate system in which the Killing vector is given by

$$\eta^\mu \frac{\partial}{\partial X^\mu} = \frac{\partial}{\partial u}.$$

Thus the metric and therefore the Weyl tensor are independent of u. The coordinate system can be further specialized so that if it is assumed that[6] $\psi_0 = \psi_0^0 r^{-5} + \psi_0^1 r^{-6} + O(r^{-7})$, then the Weyl and the metric tensor functions can be written[7]

$$\psi_1 = \psi_1^0 r^{-4} + \bar{\eth}\psi_0^0 r^{-5} + \tfrac{1}{2}\bar{\eth}\psi_0^1 r^{-6} + O(r^{-7}), \tag{2a}$$

$$\psi_2 = \psi_2^0 r^{-3} + \bar{\eth}\psi_1^0 r^{-4} + \tfrac{1}{2}\bar{\eth}^2\psi_0^0 r^{-5} + \tfrac{1}{6}(\bar{\eth}^2\psi_0^1 - 4\psi_1^0\bar{\psi}_1^0) r^{-6} + O(r^{-7}), \tag{2b}$$

$$\psi_3 = \tfrac{1}{2}\bar{\eth}^2\psi_1^0 r^{-4} + \tfrac{1}{6}(\bar{\eth}^3\psi_0^0 - 3\psi_2^0\bar{\psi}_1^0) r^{-5} + \tfrac{1}{24}(\bar{\eth}^3\psi_0^1 - 16\bar{\psi}_1^0\bar{\eth}\psi_1^0 - 3\psi_2^0\eth\bar{\psi}_0^0) r^{-6} + O(r^{-7}), \tag{2c}$$

$$\psi_4 = \tfrac{1}{24}\bar{\eth}^4\psi_0^0 r^{-5} + \tfrac{1}{120}(\bar{\eth}^4\psi_0^1 - 40\bar{\psi}_1^0\bar{\eth}^2\psi_1^0) r^{-6} + O(r^{-7}), \tag{2d}$$

$$U = -1 - \psi_2^0 r^{-1} - \tfrac{1}{6}(\bar{\eth}\psi_1^0 + \eth\bar{\psi}_1^0) r^{-2} - \tfrac{1}{24}(\bar{\eth}^2\psi_0^0 + \eth^2\bar{\psi}_0^0) r^{-3} - \tfrac{1}{120}(\bar{\eth}^2\psi_0^1 + \eth^2\bar{\psi}_0^1 + 10\psi_1^0\bar{\psi}_1^0) r^{-4} + O(r^{-5}), \tag{2e}$$

$$\omega = -\tfrac{1}{2}\psi_1^0 r^{-2} - \tfrac{1}{6}\bar{\eth}\psi_0^0 r^{-3} - \tfrac{1}{24}\bar{\eth}\psi_0^1 r^{-4} + O(r^{-5}), \tag{2f}$$

$$X^k = \tfrac{1}{6}(\xi^{0k}\bar{\psi}_1^0 + \bar{\xi}^{0k}\psi_1^0) r^{-3} + \tfrac{1}{12}(\xi^{0k}\eth\bar{\psi}_0^0 + \bar{\xi}^{0k}\bar{\eth}\psi_0^0) r^{-4} + \tfrac{1}{40}(\xi^{0k}\eth\bar{\psi}_0^1 + \bar{\xi}^{0k}\bar{\eth}\psi_0^1) r^{-5} + O(r^{-6}), \tag{2g}$$

$$\xi^k = \xi^{0k} r^{-1} + \tfrac{1}{6}\bar{\xi}^{0k}\psi_0^0 r^{-4} + \tfrac{1}{12}\bar{\xi}^{0k}\psi_0^1 r^{-5} + O(r^{-6}) \tag{2h}$$

with $\xi^{02} = -i\xi^{03} = P$ and $P = \frac{1}{2}(1 + \zeta\bar{\zeta})$, where $\zeta = -x^2 + ix^3$ and

$$\lim_{r\to\infty} r^2 g^{kl} = -2P^2\delta^{kl}.$$

The metric components may then be written, $g^{0\mu} = \delta^\mu_1$, $g_{1\mu} = \delta^0_\mu$,

$$g^{11} = -2 - 2\psi_2^0 r^{-1} - \tfrac{1}{3}(\bar{\eth}\psi_1^0 + \eth\bar{\psi}_1^0) r^{-2} - \tfrac{1}{12}(\bar{\eth}^2\psi_0^0 + \eth^2\bar{\psi}_0^0) r^{-3} - \tfrac{1}{60}(\bar{\eth}^2\psi_0^1 + \eth^2\bar{\psi}_0^1 + 40\psi_1^0\bar{\psi}_1^0) r^{-4} + O(r^{-5}), \tag{3a}$$

$$g^{1k} = \tfrac{2}{3}(\xi^{0k}\bar{\psi}_1^0 + \bar{\xi}^{0k}\psi_1^0) r^{-3} + \tfrac{1}{4}(\xi^{0k}\eth\bar{\psi}_0^0 + \bar{\xi}^{0k}\bar{\eth}\psi_0^0) r^{-4} + \tfrac{1}{15}(\xi^{0k}\eth\bar{\psi}_0^1 + \bar{\xi}^{0k}\bar{\eth}\psi_0^1) r^{-5} + O(r^{-6}), \tag{3b}$$

$$g^{kl} = -2P^2\delta^{kl} r^{-2} - \tfrac{1}{6}(\psi_0^0\bar{\xi}^{0k}\bar{\xi}^{0l} + \bar{\psi}_0^0\xi^{0k}\xi^{0l}) r^{-5} - \tfrac{1}{12}(\psi_0^1\bar{\xi}^{0k}\bar{\xi}^{0l} + \bar{\psi}_0^1\xi^{0k}\xi^{0l}) r^{-6} + O(r^{-7}), \tag{3c}$$

$$g_{00} = 2 + 2\psi_2^0 r^{-1} + \tfrac{1}{3}(\bar{\eth}\psi_1^0 + \eth\bar{\psi}) r^{-2} + \tfrac{1}{12}(\bar{\eth}^2\psi_0^0 + \eth^2\bar{\psi}_0^0) r^{-3} + \tfrac{1}{180}[3(\bar{\eth}^2\psi_0^1 + \eth^2\bar{\psi}_0^1) - 40\psi_1^0\bar{\psi}_1^0] r^{-4} + O(r^{-5}), \tag{4a}$$

$$g_{0k} = \tfrac{1}{3}P^{-2}(\xi^{0k}\bar{\psi}_1^0 + \bar{\xi}^{0k}\psi_1^0)r^{-1} + \tfrac{1}{8}P^{-2}(\xi^{0k}\eth\bar{\psi}_0^0 + \bar{\xi}^{0k}\bar{\eth}\psi_0^0)r^{-2} + \tfrac{1}{30}P^{-2}(\xi^{0k}\eth\bar{\psi}_0^1 + \bar{\xi}^{0k}\bar{\eth}\psi_0^1)r^{-3} + O(r^{-4}), \quad (4b)$$

$$g_{Kl} = -\tfrac{1}{2}P^{-2}\delta_{kl}r^2 + \tfrac{1}{24}P^{-4}(\psi_0^0\bar{\xi}^{0k}\bar{\xi}^{0l} + \bar{\psi}_0^0\xi^{0k}\xi^{0l})r^{-1} + \tfrac{1}{48}P^{-4}(\psi_0^1\bar{\xi}^{0k}\bar{\xi}^{0l} + \bar{\psi}_0^1\xi^{0k}\xi^{0l})r^{-2} + O(r^{-3}). \quad (4c)$$

The conditions on the Weyl tensor components imposed by the field equations and the condition that the space-time be stationary are given by[4,8]

$$\psi_2^0 = \bar{\psi}_2^0 = \text{const}, \quad \psi_1^0 = \sum_{m=-1}^{1} a_m\, {}_1Y_{1m},$$
$$\psi_0^0 = \sum_{m=-2}^{2} b_m\, {}_2Y_{2m}, \quad \psi_0^1 = \sum_{m=-3}^{3} c_m\, {}_2Y_{3m} + \tfrac{5}{3}[(\psi_1^0)^2 - \tfrac{3}{2}\psi_2^0\psi_0^0], \quad (5)$$

where a_m, b_m, and c_m are constants.

The terms in square brackets in the equation for ψ_0^1 are the Newman-Penrose constants in asymptotically flat, stationary, vacuum space-times.[4]

III. STATIC CONDITIONS IN NEWMAN-PENROSE FORMALISM

The condition that a space-time be static is equivalent to the condition that the Killing vector, η^μ, satisy the following equations[1]:

$$\eta_{[\nu}\eta_{\nu,\sigma]} = 0. \quad (6)$$

As $\eta^\mu = \delta_0^\mu$ in this coordinate system, it follows that $\eta_\mu = g_{\mu 0}$ so that Eq. 6 becomes

$$g_{00,1}g_{02} - g_{02,1}g_{00} - g_{00,2} = 0, \quad (7a)$$

$$g_{00,1}g_{03} - g_{03,1}g_{00} - g_{00,3} = 0, \quad (7b)$$

$$g_{00,2}g_{03} - g_{00,3}g_{02} + (g_{02,3} - g_{03,2})g_{00} = 0, \quad (7c)$$

$$g_{03,1}g_{02} - g_{02,1}g_{03} + g_{02,3} - g_{03,2} = 0. \quad (7d)$$

Using Eqs. 4 to expand Eqs. 7 asymptotically, the coefficient of each separate order of r^{-1} is then equated to zero. The Eq. 7d is the easiest to write out. The coefficients of the r^{-1}, r^{-2}, and r^{-3} terms for this equation yield, respectively,

$$\bar{\eth}\psi_1^0 = \eth\bar{\psi}_1^0,$$
$$\bar{\eth}\,{}^2\psi_0^0 = \eth\,{}^2\bar{\psi}_0^0, \quad (8)$$
$$\bar{\eth}\,{}^2\psi_0^1 = \eth\,{}^2\bar{\psi}_0^1.$$

These are just the conditions that ψ_1^0, ψ_0^0, and ψ_0^1 be electric type.[8] The remaining Eq. (7a, b, c) yield no further conditions. Conditions equivalent to Eqs. 8 appear in the remaining equations but at higher orders in r^{-1}.

Applying the above conditions to the coefficients of each power of the asymptotic expansion of the Weyl tensor component functions given by Eqs. (2a-d) it is easily found that the Weyl tensor is electric type to order r^{-7} in the asymptotic expansion.

As is well known, static space-times are invariant under a discrete transformation which has the form of a time reversal in an appropriate coordinate system.[1] This implies here that the Weyl tensor is asymptotically electric type to order s^{-7}, where s is an affine parameter on incoming null hypersurfaces, and further that the Newman-Penrose constants defined on $\mathcal{I}^-$ are the same as those defined on $\mathcal{I}^+$, as is easily seen by applying the properties of[8] $\eth$ and Eqs. 8 to the constants defined on $\mathcal{I}^-$. The constants defined on $\mathcal{I}^-$ are given by the expression $\frac{1}{4}(\eth\,{}^2\bar{\psi}_1^0)^2 - \frac{1}{16}\psi_2^0(\eth\,{}^4\bar{\psi}_0^0)$ as inspection of Eqs. (2c, d) shows.

IV. CONCLUSIONS

Although it is not possible to conclude from this investigation that the Weyl tensor must be electric type for a space-time to be static, it gives an indication that such an exact result may be true. It is encouraging in this respect to see that the asymptotic result of this paper holds even at an order in r^{-1} for which the nonlinear nature of the field equations is making its effects known through the presence of the Newman-Penrose constants, $(\psi_1^0)^2 - \frac{3}{2}\psi_2^0\psi_0^0$. The basic difficulty in obtaining an exact result is that is has not been possible to write the coefficient of the general term of an asymptotic expansion of the Newman-Penrose equations even for stationary, asymptotically flat, vacuum space-times.

One curious feature of this investigation is the interdependence of Eqs. 7. Apparently, their independence becomes manifest only at higher orders in r^{-1} than were investigated here.[9]

ACKNOWLEDGMENTS

The author would like to thank the support of the members of the relativity group at the University of Hamburg where part of this work was completed and to thank Dr. Allen I. Janis for helpful discussions.

*Part of this work was completed under tenure of a fellowship from the Alexander von Humboldt Foundation.

[1] A. Trautman, *Brandeis Lectures*, (Prentice-Hall, Engelwood Cliffs, N. J., 1964), Vol. 1, p. 130 ff.

[2] E. T. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962); J. Math. Phys. 4, 998 (1963).

[3] E. T. Newman and T. W. J. Unti. J. Math. Phys. 3, 891 (1962).

[4] E. T. Newman and R. Penrose, Proc. Roy. Soc. (London) A305, 175 (1968).

[5] E. T. Newman, private communciation.

[6] Certain differentiability conditions analogous to the asymptotic smoothness conditions of Ref. 2 are also assumed here.

[7] These expressions are in agreement with results obtained by Dr. E. Couch in his unpublished doctoral dissertation (1966). There he obtained the asymptotic expression for ψ_4 only to order r^{-6}.

[8] The functions ${}_sY_{lm}$ are spin weighted eigenfunctions of the differential operator $\eth\bar{\eth}$ defined in the paper by Newman and Penrose, J. Math. Phys. 7, 863 (1966), to which the reader may refer for details.

[9] It has been pointed out to the author by Dr. Allen Janis that only three of Eqs. (7a-d) are independent.

Mayer's theory analysis of repulsive and weak, long range potentials

Alan J. Liebman* and Jean-Yves Parlange

Department of Engineering and Applied Science, Yale University, New Haven, Connecticut
(Received 24 April 1972; revised manuscript received 12 January 1973)

The irreducible integrals defined in the Mayer theory of real gases are derived for one-dimensional systems having hard rod, weak long range, Curie–Weiss and van der Waals potential functions. In two cases, independent knowledge of the exact equation of state of the gas permits the derivation of general star graph degeneracy relationships. These, in turn, are useful for the study of new potential functions; a penetrable repulsive force is shown to produce attractive force behavior in the gas state equation.

I. INTRODUCTION

The exact statistical mechanical properties of systems of many interacting particles are rigorously known for only a few interparticle potentials.[1,2,3,4,5,6] Although the mathematical tools are in principle available, even very simple pair-potentials have defied precise solution. In particular, the Mayer cluster theory[7,8,9,10] has often been introduced as a rigorous theoretical link between gaseous state equations and microscopic forces, but it is difficult to use as a calculational tool. This is the case because the cluster bonding structure and the inseparability of particle coordinates present complex mathematical problems.

For some potential functions of interest, however, it is possible to reduce the mathematical difficulties of the Mayer form to a manageable level. One-dimensional systems can significantly reduce the bonding possibilities. Also the cases of repulsive interactions and of weak, long range forces simplify the Mayer cluster integrands to the degree that coordinate separability is possible over some part of the full integration range. With these two aids, the Mayer formalism can be put to direct calculational use.

In the sections to follow, a variety of one-dimensional gaseous phase interaction potentials will be studied through their behaviour in irreducible cluster integrals. The linkage between irreducible integrals and graph theory[11,12,13,14,15] will be used to provide theorems on star graph degeneracies. Lastly, such theorems will be shown to allow the investigation of new potential functions.

II. THE INFINITELY HARD ROD GAS

A. Cluster integrals from the equation of state

The lattice gas of an infinitely repulsive interaction has been treated by many analytic techniques[16] and therefore provides an apt introduction for the present methods. This gas obeys the interparticle potential

$$U_{ij} = \begin{cases} +\infty, & \text{if rods } i \text{ and } j \text{ of length } m\delta \text{ overlap on at least one site,} \\ 0, & \text{if rods } i \text{ and } j \text{ of length } m\delta \text{ have no common site.} \end{cases} \quad (1)$$

The state equation is obtained once the weighting function for equivalent placements is found. Given a line of length L, having $N = L/\delta$ sites separated by the length δ, the number of ways of arranging n infinitely hard rods of length m on these sites is[17]

$$\binom{n - nm}{n} \quad (2)$$

and the pressure of this gas follows from the use of (2) in the canonical partition function

$$\frac{P}{KT} = \frac{1}{\delta} \ln \frac{1 - m\delta\rho}{1 - (m+1)\delta\rho} \quad (3)$$

where P, ρ, T are the pressure, density, and temperature of the gas, respectively, and K is Boltzmann's constant.

Gaseous equations of state can be expanded in powers of the density ("virial expansion") since the pressure is small and single-valued at low densities. In particular, succeeding terms of a virial expansion are higher order corrections arising from larger clusters of gas molecules. The Mayer theory casts the state equation into the form

$$\frac{P}{\rho KT} = 1 - \sum_{k \geq 1} \frac{k}{k+1} \beta_k \rho^k, \quad (4)$$

where the coefficient of the kth power of the density includes the irreducible integral factor

$$\beta_k = \frac{1}{k!\,V} \int \cdots \int [(\textstyle\sum \Pi)' f_{ij}] d\bar{X}_1 d\bar{X}_2 \cdots d\bar{X}_{k+1} \quad (5)$$

Here, $(\sum \Pi)'$ is the sum over all products of f_{ij} having at least double connectedness and no articulation points. Integration is performed over all possible locations of these particles (position vector $\bar{X}_i$) in space. The Mayer function is defined

$$f_{ij}(r) = e^{-U_{ij}(r)/KT} - 1, \quad (6)$$

where $U_{ij}(r)$ is the two-body configurational energy of the ith and jth particles. For infinitely hard rods, all β_k can be obtained by expanding (3) into the form (4):

$$\beta_k^{(m)} = -\frac{\delta^k}{k} [(m+1)^{k+1} - m^{k+1}]. \quad (7)$$

We observe that the continuum limit of the lattice gas ($m \to \infty$ with $\sigma = m\delta$ a constant) for hard rods is

$$\beta_k^{(\infty)} = -\frac{k+1}{k} \sigma^k. \quad (8)$$

B. Cluster integrals from Mayer definition

For a one-dimensional lattice gas the integrations over $3(k+1)$ coordinates in (5) become sums over $(k+1)$ coordinates. If the star graph representation of the integrand is then introduced, (5) becomes

$$\beta_k = \frac{\delta^k}{k!} \sum_{X_1} \cdots \sum_{X_{k+1}} \left\{ \sum_l \sum_\theta S_{k,l,\theta} (\Pi f)_{k,l,\theta} \right\}. \quad (9)$$

Here l is the number of lines in a star graph of $k+1$ points and is confined to the values $k+1 \leq l \leq k(k+1)/2$. The subscript θ counts the number of topologically different stars having $k+1$ points and l lines; $S_{k,l,\theta}$ is the star "degeneracy." Three problems are presented by (9): the enumeration of star graphs and the determination of their degeneracy values-two mathematical matters—and the summations over f functions—a process that requires the specification of a physical interparticle potential. For perspective, it should be noted that the $S_{k,l,\theta}$ are presently available[18] only for clusters $k \leq 7$. Evaluations of infinite sets of β_k cannot be done therefore by the obvious path.

When the potential function (1) is inserted into (9), the summations over X_i are truncated because a nonzero contribution arises only when *all* the pairs of rods bonded in a graph are within $m+1$ sites of one another. When one of the $k+1$ rods is taken as a fixed reference, the maximal "width" of the cluster with respect to that origin site is prescribed. However, it is *not* true that (9) requires all the $k+1$ rods to be within $m+1$ sites of each other! An overlapping of rods can give nonzero contributions to β_k and, for such a configuration of $k+1$ rods on the line, only some of the star graphs in the complete set will contribute. This is a major difficulty: particular particle configurations choose distinct graph subsets. Such subsets have not been studied systematically.

The configuration shown in Fig. 1a illustrates the graph-specific subset selection with four, labeled, two-unit-long rods ($k=3, m=2$). For $k=3$, there are just three star graphs (Fig. 1b). The 1a configuration will contribute $3(e^{-\infty/KT}-1)^4$ from graph #1, but graphs #2 and #3 do not contribute for that configuration of rods since $f_{ac}=0$ is a factor in their products.

Progress can be made if the collection of rod configurations in (9) is separated into two classes: (i) all rods within an interval of $m+1$ sites, and (ii) at least one rod outside this interval of $m+1$ sites. (For concreteness, each rod of length $m\delta$ can be thought to be located by its left endpoint.) This separation does not abandon any star graph contribution. The configurations now in group (i) share the property that the *entire* set of star graphs of $k+1$ points contributes for *every* configuration in the group. Further, each graph of l bonds has an f function product of value $(-1)^l$ (independent of θ) for each rod configuration.

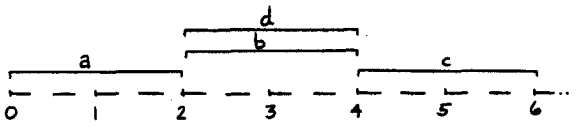

FIG. 1a. A typical overlap configuration of four rods ($k=3, m=2$). The configuration width is 6δ.

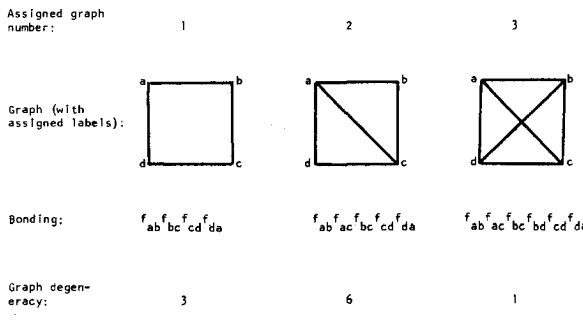

FIG. 1b. The four-point star graphs.

Hence,

$$\beta_k]_{(i)} = \frac{\delta^k}{k!}\sum_{X_1=0}^{m}\cdots\sum_{X_{k+1}=0}^{m}\left\{\sum_l\sum_\theta(-1)^l S_{k,l,\theta}\right\}, \tag{10}$$

where $\beta_k]_{(i)}$ is the contribution to β_k of the rod configurations in group (i). Using the definition,

$$\sum_\theta S_{k,l,\theta} = S_{k,l} \tag{11}$$

and the zeroth alternating moment sum of $S_{k,l}$ [derived at (32)], Eq. (10) can be written

$$\beta_k]_{(i)} = -\frac{\delta^k}{k}\sum_{X_1=0}^{m}\cdots\sum_{X_{k+1}=0}^{m}\{1\}. \tag{12}$$

It is now necessary only to count the number of configurations making up group (i), $\eta_{(i)}$, in order to obtain an exact evaluation of $\beta_k]_{(i)}$.

The number $\eta_{(i)}$ must include all configurations in which $k+1$ labeled rods are within an $m+1$ site interval but must exclude all configurations which are displacements of the $k+1$ rods by some length $\leq m\delta$ ("redundant configurations"). This is accomplished by counting the number of ways of placing $k+1$ labeled particles on r different sites with no site empty and letting r take the values $1 \leq r \leq m+1$. Then

$$\begin{aligned}\eta_{(i)} &= \sum_{r=1}^{m+1}\binom{m}{r-1}\Delta^r 0^{k+1}\\ &= \sum_r\binom{m+1}{r}\Delta^r 0^{k+1} - \sum_r\binom{m}{r}\Delta^r 0^{k+1}\\ &= (m+1)^{k+1} - m^{k+1},\end{aligned} \tag{13}$$

where Δ is the finite difference operator.[12]

Another evaluation of $\eta_{(i)}$ is the following. If the nondisplacement requirement is temporarily neglected, there are $(m+1)^{k+1}$ ways of placing $k+1$ rods on $m+1$ sites. Similarly, the $k+1$ rods could be put on m sites in m^{k+1} ways. This latter number represents the number of configurations of $k+1$ rods for which no rod is required to be on one "chosen" site of the $m+1$ sites. Thus the number of nondisplaced configurations of $k+1$ rods (those with at least one particle on a "chosen" site) is $(m+1)^{k+1} - m^{k+1}$, as in (13).

Combining (9), (12), and (13), we have

$$\begin{aligned}\beta_k = \beta_k]_{(i)} + \beta_k]_{(ii)} &= -\frac{\delta^k}{k}[(m+1)^{k+1} - m^{k+1}]\cdots\\ &\quad + \frac{\delta^k}{k!}\sum_{X_1}{}' \cdots \sum_{X_{k+1}}{}' \quad (\text{at least one } X \geq m+1)\\ &\quad \times\left\{\sum_l\sum_\theta S_{k,l,\theta}(\Pi f)_{k,l,\theta}\right\}.\end{aligned} \tag{14}$$

The comparison of (14) with (7) produces a theorem on hard repulsion star graph subsets:

$$\sum_{X_1}{}' \cdots (\text{at least one } x \geq m+1)\sum_{X_{k+1}}{}'\left\{\sum_l\sum_\theta S_{k,l,\theta}(\Pi f)_{k,l,\theta}\right\} = 0. \tag{15}$$

This equation demonstrates that a special, complex subset of star graphs gives a net zero contribution to the irreducible cluster integrals of a hard rod gas. It illustrates that graph theorems of physical interest can derive from the solution of physical problems. Specifically, when a state equation and the potential function supporting it are known, a graph theorem is available through the Mayer relationship.

A further value of Eq. (15) is the guidance it can give to numerical solutions of virial expansion problems. It is the proof of suggestive observations by Hoover and deRocco[18] and by Ree and Hoover.[19,20] The first authors numerically evaluated hard rod integrals β_2 through β_7 using a "subintegral" method. They found that net contributions to their β_k were generated only by "σ- subintegrals"; that is, only by integrals for which *all* particles were confined within the hard rod length σ.[21] Their σ-subintegrals are equivalent then to the present type (i) configurations which were the only net contributors to β_k. Ree and Hoover introduce "modified star graphs" to calculate virial coefficients more efficiently. They note that just the contribution of the fully bonded star for hard rods entirely accounts for the virial coefficient value and they imply that the sum of contributions of all their other "modified star integrals" is zero. From Eq. (15), therefore, it is seen that the Ree and Hoover modification is tantamount to a transformation of the range of integration of the Mayer integrals for the potential under discussion.

III. WEAK, LONG RANGE INTERACTIONS

The large class of long range potentials allow Mayer formalism usage because the inverse range length is a small parameter whose order can be estimated in the contributions to β_k. Consider the interaction pair-potential with a range of m sites and a strength E_0/m:

$$U_{ij} = \begin{cases} \pm E_0/m, & |X_i - X_j| \leq m\delta, \\ 0\ , & |X_i - X_j| > m\delta, \end{cases} \tag{16}$$

where E_0 is positive and finite and X_i, X_j are the coordinates of the ith and jth particles. The upper sign in (16) corresponds to a repulsive force and the lower one to an attractive force. (This convention will be continued below.) When the parameter m becomes very large, the force range becomes very long and simultaneously the force becomes very weak, but the total energy per particle pair remains constant. For large m, the Mayer functions in the region $|X_i - X_j| \leq m\delta$ can be expanded:

$$f_{ij} = \pm \frac{E_0}{mKT} + \theta\left(\frac{1}{m^2}\right) \tag{17}$$

where $\theta(x)$ is the order of magnitude of x.

Let the summation ranges of (9) be separated into two groups as in Sec II. When all particles are confined within an interval of one force range, the number of configurations is given by (13) and (9) becomes

$$\beta_k^{(m)} = \frac{1}{k!} \sum_{l=k+1}^{(k/2)(k+1)} \sum_{\theta} \left\{ S_{k,l,\theta} \left(e^{\pm \frac{E_0}{mKT}} - 1 \right)^l \right\} \cdots \left[(m+1)^{k+1} - m^{k+1} \right] \cdots + \frac{\delta^k}{k!} \sum_{X_1}{}' \cdots \sum_{X_{k+1}}{}' \text{(at least one } X_i > m+1) \left\{ \sum_l \sum_\theta S_{k,l,\theta} (\Pi f)_{k,l,\theta} \right\}. \tag{18}$$

The net contribution from configurations wider than the force range will not be zero for long range potentials and the exact form of $\beta_k^{(m)}$ for finite m is not extractable. However, the m-dependence of the factors in (18) and the behavior of the expression as $m \to \infty$ can be deduced.

Let the labeling of the star graphs be temporarily disregarded. This will overestimate the true number of configurations of $k+1$ distinguishable particles contributing to (18) but will be a sufficient evaluation. By the double connectedness requirement on star bonding, the maximum width of any cluster is the number of sites in the force range, m, multiplied by the maximum number of force range intervals possible, which is

$$I = \text{largest integer} \leq \frac{k-1}{2} + 2 = \frac{k+3}{2} \tag{19}$$

The quantity I is derived from the configuration in which there are single particles on the end sites of the cluster and two particles on the occupied interior sites. Neglecting star labels and eliminating redundant configurations by fixing one particle to an "origin site", the number of contributing configurations is

$$(2mI)^k = \theta(m^k). \tag{20}$$

But, in either term on the right of (18), the maximum contribution from any one configuration is of order $1/m^{k+1}$, since *every* configuration must have at least $k+1$ bonds and each bond contributes $\theta(1/m)$, according to (17), when m is large. Therefore,

$$\beta_k^{(m)} = \theta(1/m), \quad k \geq 2. \tag{21}$$

The case $k = 1$ is excluded from (21) because it alone is exempt from the double-connectedness property of the star graphs. The integral $\beta_1^{(m)}$ derives from two particles and their one bond. Since no configurations wider than the force range are possible, $\beta_1^{(m)}$ is exactly given by

$$\beta_1^{(m)} = f_{12}[(m+1)^2 - m^2] \underset{m \gg 1}{\longrightarrow} \mp \frac{E_0}{mKT}(2m+1) = \mp \frac{2E_0}{KT}. \tag{22}$$

Clearly when the force range approaches infinity,

$$\beta_1^{(\infty)} = \mp 2E_0/KT, \qquad \beta_k^{(\infty)} = 0, \quad k \geq 2. \tag{23}$$

Entering (23) into (4) gives the equation of state of these potentials:

$$\frac{P}{KT} = \rho \pm (E_0/KT)\rho^2 \tag{24}$$

IV. THE CURIE-WEISS GAS

The repulsive and weak attractive functions just considered allow useful approximations to real systems in combination. The Curie-Weiss model is such a combination.[22] Knowledge of the state equation and potential function of the model suggest there is an opportunity to obtain star graph relationships. The pair potential is

$$U = \begin{cases} +\infty & \text{for 2 particles on the same site,} \\ -2E_0/m & \text{for 2 particles on different sites,} \end{cases} \qquad E_0 > 0, \tag{25}$$

where m is the number of lattice sites over which the attractive force acts, and the state equation

$$\frac{P}{KT} = \frac{1}{\delta} \ln \frac{1}{1-\delta\rho} - \frac{2E_0\delta}{KT}\rho^2 \tag{26}$$

yields irreducible integrals:

$$\beta_1 = -1 + (2E_0/KT), \qquad \beta_k = -1/k, \quad k > 1. \tag{27}$$

These β_k will be valid in the one phase regions.

To derive β_k from the definition (9) with the potential (25), two configuration classes are defined: (i) the single configuration in which all particles are on one site, and (ii) all other configurations. For the class (i) configuration, we have

$$(\Pi f)_{k+1,l,\theta} = (-1)^l \quad \text{for all } k+1, , \tag{28}$$

since $U = +\infty$ for every pair of bonded particles. The contribution to β_k is

$$\beta_k]_{(i)} = \frac{1}{k!} \sum_{l=k+1}^{(k/2)(k+1)} (-1)^l S_{k+1,l} \tag{29}$$

using (11). For the class (ii) configurations,

$$\beta_k]_{(ii)} = \frac{1}{k!} \sum_{X_1}' \cdots \sum_{X_{k+1}}' \left\{ \sum_l \sum_\theta S_{k+1,l,\theta} (\Pi f)_{k+1,l,\theta} \right\}, \tag{30}$$

where $\sum'$ requires that all particles cannot occupy the same site and where the f function follows from (6) and (25). The connectedness of the star graphs requires that every term in the integrand of (30) have at least two factors with $U = -2E_0/m$, that is, every term is at least of order $1/m$ when m is large. In one phase regions, every term in (30) goes to zero when $m \to \infty$. Thus

$$\begin{aligned} \beta_1 &= -1 + (2E_0/KT) \\ \beta_k &= \frac{1}{k!} \sum_l (-1)^l S_{k+1,l}, \quad k > 1. \end{aligned} \tag{31}$$

Equating (27) to (31), the following star graph theorem is demonstrated:

$$\sum_{l=k+1}^{(k/2)(k+1)} (-1)^l S_{k+1,l} = -(k-1)!. \tag{32}$$

This "zeroth alternating moment" of the star degeneracy is used in further theoretical work in Sec. II and VI. Three such moments are known from the work of Riddell and Uhlenbeck[14] but on the basis of a mathematical, not a physical, analysis.

V. THE VAN DER WAALS GAS

A second combination of repulsive and long range attractive forces is realized in the van der Waals model.[23] On a one-dimensional lattice whose sites are spaced δ units apart, allow two characteristic lengths to exist: $\sigma = m\delta$ for a repulsive region and $n\sigma = nm\delta$ for an attractive region of interaction. The potential for a particle pair is

$$U_{ij} = \begin{cases} +\infty, & 0 \leq |X_i - X_j| \leq m\delta, \\ -E_0/nKT, & m\delta < |X_i - X_j| \leq nm\delta, \\ 0, & |X_i - X_j| > nm\delta. \end{cases} \tag{33}$$

When (33) is inserted in (9), the summations over placements of the $k+1$ particles can be separated into three classes:

$$\begin{aligned} \beta_k = \beta_k] \text{ pure repulsion forces} &+ \beta_k] \text{ pure attraction forces} \\ &+ \beta_k] \text{ mixed forces.} \end{aligned} \tag{34}$$

The first term includes configurations in which *every* particle is within m sites of the particles to which it is bonded, for all possible star bondings. Their contribution was calculated in Sec. II:

$$\beta_k] \text{ pure repulsion forces} = -(\delta^k/k)[(m+1)^{k+1} - m^{k+1}]. \tag{35}$$

The second term of (34) includes configurations in which no particle is closer to another particle than $m+1$ sites but no further than nm sites. The derivation of Sec. III then applies, giving

$$\begin{aligned} \beta_1] \text{ pure attraction} &= +\frac{2E_0 m\delta}{KT} + \theta\left(\frac{1}{n}\right) \\ \beta_k] \text{ pure attraction} &= \theta(1/n), \quad k \geq 2. \end{aligned} \tag{36}$$

For the mixed force β_k contribution, the n-dependence of the f product and of the number of configurations are computable. Every configuration contributing to $\beta_k] \cdots$ mixed force can be considered a composite of "islands" of particles. The particles of each island are within m sites of one another while the islands are separated by distances greater than $m\delta$. By the multiconnectedness requirement on the star graphs, every island must have at least two (long range) bondings to particles in other islands.

For an arbitrary configuration of $k+1$ particles let there be ξ islands where, recalling (19),

$$2 \leq \xi \leq I. \tag{37}$$

This means there must be a *minimum* of ξ long range bonds, which contribute a factor of $\theta(n^{-\xi})$ to $\beta_k]$ mixed force for large n, by the expansion of the Mayer function with $U = -E_0/KT$. The number of nonredundant configurations having just ξ islands is of order $n^{\xi-1}$, since every island excepting the "origin" island can move over $2mn$ sites (when the overestimate of unlabeled graphs is made). The contribution, then, to $\beta_k]$ mixed force from all configurations having just ξ islands is of order $1/n$. Since the number of ξ values is finite and independent of n, it is seen that

$$\beta_k] \text{ mixed forces} = \theta(1/n) \tag{38}$$

Collecting, (34), (35), (36), and (37) the irreducible integrals for finite range forces are

$$\begin{aligned} \beta_1 &= -\delta(2m+1) + \frac{2E_0\delta m}{KT} + \theta\left(\frac{1}{n}\right), \\ \beta_k &= -\frac{\delta^k}{k}[(m+1)^{k+1} - m^{k+1}] + \theta\left(\frac{1}{n}\right), \quad k \geq 2. \end{aligned} \tag{39}$$

When the attraction ranges become infinite, $n \to \infty$, and the second equation of (39) reduces to one term. The common van der Walls equation of state follows when the continuum limit of (39) is taken:

$$\frac{P}{KT} = \frac{\rho}{1-\sigma\rho} - \frac{E_0\sigma}{KT}\rho^2. \tag{40}$$

Comparison of the Curie-Weiss and the van der Waals models shows that the van der Waals repulsive force contribution derives from configurations spreading over many lattice sites while the Curie-Weiss law derives from a point (one site) repulsion force. Lengthier arguments by deRocco[24] and Widom[25] lead to an equivalent conclusion. In their terms, the excluded volume is independent of the density for the Curie-Weiss gas but not for the van der Waals gas.

The van der Waals "unphysical region" appears in Eq. (40) because the force range has been allowed to become very large before the β_k were inserted into Eq. (4). A derivation that could predict two phase states would reverse the order of these two steps. Then, however, there would be an infinite number of terms of order $1/n$, and the order of magnitude estimate of β_k used above would be insufficient. Kac et al.[5] succeeded

in carrying out the limits in the proper order for a particular one-dimensional potential having an exponential long range attraction. The proper two phase states were then seen.

VI. A PENETRABLE REPULSION POTENTIAL

The solutions of the preceding sections have in common the property that net contributions to β_k from particle configurations wider than the force range are either zero or of negligible order. Enumeration of the number of contributing wide configurations has been seen to be more intricately connected with the properties of the star graphs than the enumeration of configurations narrower than or equal to the force range. For this reason, it is profitable to consider new models that will allow similar simplification.

An interesting example of this is the system where the binary interaction of two molecules is allowed to be modified by the presence of a third molecule. Consider a cluster with l repulsive bonds. Assume that in small clusters where particles interfere with one another, the Mayer function is not -1 on the average, but $(a_0 - 1)$, where a_0 is very small and represents a slight decrease in the repulsion effect. It is again possible to divide the set of particle configurations into two parts: (i) all particles are within m sites of the "origin" and (ii) at least one particle is further than m sites from the "origin." Introducing the function $f = (a_0 - 1)$ into (9) and using (11) and (13),

$$\beta_k]_{(i)} = \frac{\delta^k}{k!}[(m+1)^{k+1} - m^{k+1}]\left(\sum_l (a_0 - 1)^l S_{k,l}\right). \tag{41}$$

Since $a_0 \ll 1$,

$$(a_0 - 1)^l = (-1)^l + (-1)^{l-1}\binom{l}{1} a_0 + (-1)^{l-2}\binom{l}{2} a_0^2 + \ldots \tag{42}$$

and (41) becomes

$$\beta_k]_{(i)} = \frac{\delta^k}{k!}[(m+1)^{k+1} - m^{k+1}]\left\{\sum_l (-1)^l \binom{l}{0} S_{k,l} \ldots - a_0 \sum_l (-1)^l \binom{l}{1} S_{k,l} + a_0^2 \sum_l (-1)^l \binom{l}{2} S_{k,l} \ldots\right\}. \tag{43}$$

The first term is evaluated using (32); the two succeeding alternating degeneracy moments are known in closed form.[14] If attention is limited to first order in the small quantity a_0, (43) can be written

$$\beta_k]_{(i)} = \left[-\frac{1}{k} + a_0\left(\frac{k+1}{2}\right)\right]\delta^k[(m+1)^{k+1} - m^{k+1}]. \tag{44}$$

The contribution to β_k from the wide configurations $\beta_k]_{(ii)}$ is zero, by Eq. (15). Clearly (44) fails as soon as $k^2 a_0$ is of order one. Implicitly, it has been assumed that the virial expansion of the equation of state converges fast enough so that a β_k for large k not represented by (44) is of no importance. Inserting (44) into (4) the lattice gas equation of state is found:

$$\frac{P}{KT} = \frac{1}{\delta}\ln\left(\frac{1 - m\rho\delta}{1 - (m+1)\rho\delta}\right) \ldots + \frac{\rho^2\delta a_0}{2}\left(\frac{m^2}{(1 - m\rho\delta)^2} - \frac{(m+1)^2}{[1 - (m+1)\rho\delta]^2}\right). \tag{45}$$

In the continuum limit,

$$\frac{P}{KT} = \frac{\rho}{1 - \sigma\rho} - \frac{\sigma a_0 \rho^2}{(1 - \sigma\rho)^2}. \tag{46}$$

The first term recovers the hard rod solution; the second term corrects for the inelasticity of the rods. At low densities the second term is formally similar to the attractive part of the van der Waals equation. That is, the "soft" repulsion force driving (46) can be considered a combination of an infinitely hard short range repulsion and a weak, long range attraction force. Hauge and Hemmer[26] have shown in general that the lowest order state equation for such combinations will take the form

$$P = p_s - a\rho^2 \tag{47}$$

where p_s is the gas pressure from the short range force and a is dependent only on the long range interaction.

VII. CONCLUSIONS

It has been shown that many one-dimensional gas phase solutions can be drawn rigorously from Mayer's irreducible integrals. This is the case even though extremely few exact properties of the integrals are known. In particular, all the physical models presented above derive from a single alternating star degeneracy moment and a theorem following from the hard rod solution. Evaluations of Mayer integrals in problems of physical interest can be used to derive graph theoretical relations. Such mathematical relations, particularly with respect to graph subsets, can be used in turn to solve new physical problems. The characteristics of a graph theorem-yielding physical model are that the equation of state be independently known and that the interparticle potential function be well defined.

*Present Address: Xerox Corporation, Rochester, New York.

[1]E. Ising, Z. Phys. 31, 253 (1925).
[2]L. Onsager, Phys. Rev. 65, 117 (1944).
[3]P. C. Hemmer and E. H. Hauge, Phys. Rev. 133, A1010 (1964).
[4]E. H. Hauge and P. C. Hemmer, Physica 29, 1338 (1963).
[5]M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. 4, 216 (1963).
[6]I. P. Pavlotskii, Sov. Phys.-Doklady 10, 301 (1965).
[7]J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (Wiley, New York 1940), especially Chaps. 13, 14.
[8]J. E. Mayer and S. F. Harrison, J. Chem. Phys. 6, 87 (1938).
[9]M. Born and K. Fuchs, Proc. Roy. Soc. A166, 391 (1938).
[10]J.-Y. Parlange, J. Chem. Phys. 48, 776 (1968).
[11]G. E. Uhlenbeck and G. W. Ford, *Studies in Statistical Mechanics* (North-Holland, Amsterdam, 1962), Vol. I, part B.
[12]J. Riordan, *Introduction to Combinatorial Analysis* (Wiley, New York, 1958), especially Chap. 2 and 5.
[13]F. Harary, *Graph Theory and Theoretical Physics* (Academic, New York, 1967).
[14]R. J. Riddell and G. E. Uhlenbeck, J. Chem. Phys. 21, 2056 (1953).
[15]A. J. Liebman, "Applications of Mayer's Theory to the Study of One-Dimensional Systems," Doctoral dissertation, Yale University, 1970.
[16]L. Tonks, Phys. Rev. 50, 955 (1936).
[17]T. D. Lee and C. N. Yang, Phys. Rev. 87, 410 (1952).
[18]W. G. Hoover and A G. deRocco, J. Chem. Phys. 36, 3141 (1962).
[19]F. H. Ree and W. G. Hoover, J. Chem. Phys. 40, 939 (1964).
[20]F. H. Ree and W. G. Hoover, J. Chem. Phys. 46, 4181 (1967).
[21]If one sums horizontally across subintegral evaluations in Table IX of Ref. 19, one obtains the values $S_{k,l}$; a vertical summation over these numbers with the appropriate f-function factors, $(-1)^l$, corresponds to the introduction of eq. (32) in the present work.
[22]T. L. Hill, *Statistical Mechanics* (McGraw-Hill, New York, 1956). See Chap. 7.
[23]J. D. van der Waals, dissertation, Leiden, 1873.
[24]A. G. deRocco, Phys. Fluids 9, 1348 (1966).
[25]B. Widom, J. Chem. Phys. 39, 2808 (1963).
[26]E. H. Hauge and P. C. Hemmer, J. Chem. Phys. 45, 323 (1966).

Conformal spinor fields in general relativity

M. D. Maia

Universidade de Brasilia, Departmento de Matemática, 70000 Brasilia, Brazil
(Received 20 April 1973)

A spinor field which is covariant under the group of conformal motions in general relativity is defined. These spinor fields extend the concept of isospinors to the cases where conformal motions take place and coincide with Penrose's twistors in the case of flat space–time. It is found that only certain special conformal spinor fields lead to physical entities. The example of Kruskal's conformal spinors is studied.

1. INTRODUCTION

In a previous paper we have introduced the notion of isospinors in general relativity[1]. The isospinors were presented as objects which would be convenient to describe isometric covariance in general relativity. Since isometries are not always present in general relativity the isospinors are not always physically meaningful. To cover this deficiency we turn to the next more general group of symmetries which is the group of conformal motions. Following the same basic ideas we construct a spinor structure which is covariant under the group of conformal motions and consequently are the appropriate vehicle to describe conformal covariance in general relativity. These conformal spinors resemble Penrose's Twistors except by the significant fact that in the case of general relativity they are locally defined.[2,3] However, in special relativity we can establish a one-to-one relation between conformal spinors and twistors.

We use the same notation as in the isospinors paper. Thus a general relativistic space-time is denoted by $R4$. The isometric, minimal embedding space of dimension p and with a metric tensor $\eta_{\mu\nu}$ with r positive and s negative units in the diagonal is denoted by $M(p, r, s)$. Cartesian coordinates in this space are denoted by X^μ, with all Greek indices running from 1 to p, unless explicitly stated to the contrary. Gaussian coordinates in $M(p, r, s)$ based on the hypersurface $R4$ are denoted by x^α and in particular x^i denotes the Gaussian coordinates measured in $R4$, while x^A denotes the Gaussian coordinates measured orthogonally to $R4$ (lower case boldface Latin indices run from 1 to 4 while capital boldface Latin indices run from 5 to p). The embedding is specified when the functions $X^\mu(x^\alpha)$ are given. With the corresponding derivative map, $X^\mu_\alpha = \partial X^\mu/\partial x^\alpha$, we can relate vectors and tensors from the Gaussian to the Cartesian system. The inverse is obtained from $X^\alpha(X^\mu)$ and $x_\mu^{\ \alpha} = \partial x^\alpha/\partial X^\mu$. In the Gaussian system the space-time $R4$ is defined by the equations $x^A = 0$. Any real function $\phi(x^\alpha)$ defined in $M(p, r, s)$ is projected in $R4$ by taking the limit $x^A \to 0$. We use the simplifying notation

$$\phi|_{R4} = \lim_{x^A \to 0} \phi(x^\alpha).$$

If the function ϕ is defined in $R4$ this fact is stated by $\phi(R4)$. It is not always true that $\phi|_{R4} = \phi(R4)$.

Finally, $O(r, s)$ stands for the homogeneous group of isometries in $M(p, r, s)$ while $P(r, s)$ is the inhomogeneous or the "Poincaré" group in the same space. The conformal group will be denoted by $C(r, s)$.

2. THE CONFORMAL GROUP

The group of conformal motions on a flat space $M(p, r, s)$ is generated by the infinitesimal transformations

$$X'^\mu = X^\mu + U^\mu, \tag{2.1}$$

where

$$\pounds_u \eta^{\mu\nu} = U^{(\mu,\nu)} = \frac{1}{p} U^\rho_{\ ,\rho} \eta^{\mu\nu} \tag{2.2}$$

(coma denotes ordinary derivatives). In Gaussian coordinates based on a space-time $R4$, isometrically and minimally embedded in $M(p, r, s)$, the above expressions read

$$\begin{aligned} x'^\alpha &= x^\alpha + \xi^\alpha, \\ \xi^{(\alpha;\beta)} &= \frac{1}{p} \xi^\gamma_{\ ;\gamma} g^{\alpha\beta}, \end{aligned} \tag{2.3}$$

where the semicolon denotes covariant derivative with respect to the metric affine connection of $M(p, r, s)$, and $g^{\alpha\beta}$ are the Gaussian components of the metric tensor of $M(p, r, s)$. The components of the vector ξ^α are related to the components U^μ by $\xi^\alpha = x^\alpha_{\ \mu} U^\mu$. To find the transformations induced in the space-time $R4$, we split (2.3) as

$$x'^i = x^i + \xi^i, \qquad x'^A = x^A + \xi^A \tag{2.4}$$

and

$$\begin{aligned} \xi^{(i;j)} &= \frac{1}{p} \xi^\alpha_{\ ;\alpha} g^{ij}, \\ \xi^{(i;A)} &= \frac{1}{p} \xi^\alpha_{\ ;\alpha} g^{iA}, \\ \xi^{(A;B)} &= \frac{1}{p} \xi^\alpha_{\ ;\alpha} g^{AB}. \end{aligned} \tag{2.5}$$

The first two equations can be joined together and after a contraction with $g_{i\alpha}$ it gives $\xi^i_{\ ;i} = (4/p)\xi^\gamma_{\ ;\gamma}$. Similarly, joining the last two equations and contracting with $g_{A\alpha}$, gives $\xi^A_{\ ;A} = (t/p)\xi^\gamma_{\ ;\gamma}$, where $t = p - 4$. Therefore, (2.5) can be rewritten as

$$\begin{aligned} \xi^{(i;j)} &= \tfrac{1}{4} \xi^k_{\ ;k} g^{ij}, \\ \xi^{(i;A)} &= \tfrac{1}{4} \xi^k_{\ ;k} g^{iA} = (1/t)\, \xi^A_{\ ;A} g^{iA}, \\ \xi^{(A;B)} &= (1/t)\, \xi^C_{\ ;C} g^{AB}. \end{aligned} \tag{2.5'}$$

The space-time projection of (2.4) is

$$x'^i = x^i + \xi^i|_{R4}, \qquad x'^A = \xi^A|_{R4}.$$

Therefore, if we wish to leave unchanged the equations of definition of the hypersurface $R4$, we have to impose the condition

$$\xi^A|_{R4} = 0. \tag{2.6}$$

This condition is also necessary in order that the first equation (2.5') gives the conformal Killing equation in $R4$:

$$\xi^{(i;j)}|_{R4} = \xi^{(i;j)}(R4) = \tfrac{1}{4} \xi^k_{\ ;k}(R4)\, g^{ij}(R4).$$

On the other hand, since $g^{iA}|_{R4} = 0$ (see, e.g., Szekeres[4]),

the second equation (2.5') gives $\xi^{(i;A)}|_{R4} = 0$. This means that the conformal motions in the plane (i-A) give an isometry when projected in $R4$. Finally, the last equation (2.5') describes a conformal motion orthogonal to $R4$. Therefore, the group $C(r, s)$ induce one isometry plus a conformal motion in each space-time $R4$ isometrically embedded in $R4$, provided the condition (2.6) is applied. We denote the resulting group by $C(r, s)|_{R4}$. Conversely, if a conformal motion in the space-time $R4$ embedded in $M(p, r, s)$ is given then $\xi^A = 0$. This condition clearly defines a subgroup of $C(r, s)|_{R4}$ where the transformations orthogonal to $R4$ do not appear.

The embedding of $R4$ in $M(p, r, s)$ is invariant under the group $O(r, s)$, but it is not invariant under $C(r, s)$. Thus if the functions X^μ (x^α) gives the embedding of $R4$ then the new set of functions $X'^\mu(x^\alpha)$ given by (2.1) does not necessarily give the same embedding. In order to avoid this "conformal variance" of the embedding, we replace the space $M(p, r, s)$ by a new space $M'(p', r', s')$ with $p' > p$, such that $C(r, s)$ will be equivalent to $O(r', s')$. This is done with the help of Liouville's theorem[5]: Given a flat space $M(p, r, s)$ $(p > 2)$ a conformal motion in such space can be decomposed as

(a) a transformation of $P(r, s)$ $X'^\mu = \Lambda^\mu{}_\nu X^\nu + \Lambda^\mu$,
(b) a dilatation $X'^\mu = CX^\mu$ $(C = \text{const})$,
(c) an inversion $X'^\mu = X^\mu / \frac{1}{2} X^\rho X_\rho$;

these transformations generate a projective group in a $(p + 1)$ − dimensional projective coordinate space spanned by the coordinates Y^μ, Y^0 defined by $Y^\mu/Y^0 = X^\mu$, $Y^{p+1}/Y^0 = \frac{1}{2} X^\mu X_\mu$, where Y^0 is the homogeneity coordinate and it is equal to 1 for finite points and equal to zero for points at infinity. In this projective space all points of $M(p, r, s)$ are mapped into a quadric surface Q defined by

$$\eta_{\mu\nu} Y^\mu Y^\nu - 2Y^0 Y^{p+1} = 0.$$

The projective transformations induced by (a), (b), and (c) on the projective space are

(a) $Y'^\mu = \Lambda^\mu{}_\nu Y^\mu + Y^0 \Lambda^\mu, \quad Y^{0'} = Y^0,$
$Y'^{p+1} = Y^{p+1} + \eta_{\mu\nu} \Lambda^\mu{}_\rho Y^\rho \Lambda^\nu + \frac{1}{2} Y^0 \eta_{\mu\nu} \Lambda^\mu \Lambda^\nu,$

(b) $Y'^\mu = Y^\mu, \quad Y'^0 = (1/C) Y^0, \quad Y'^{p+1} = C Y^{p+1},$

(c) $Y'^\mu = Y^\mu, \quad Y'^0 = Y^{p+1}, \quad Y'^{p+1} = Y^0.$

These transformations leave the quadric equation invariant. A metric, $(p + 2)$-dimensional coordinate space can be constructed by definition of the new coordinates

$$\begin{aligned} Z^\mu &= Y^\mu, \\ Z^{p+1} &= (1/\sqrt{2})(Y^0 + Y^{p+1}), \\ Z^{p+2} &= (1/\sqrt{2})(Y^0 - Y^{p+1}). \end{aligned} \tag{2.7}$$

In this space the quadric equation corresponds to

$$\eta_{\mu\nu} Z^\mu Z^\nu - (Z^{p+1})^2 + (Z^{p+2})^2 = 0, \tag{2.8}$$

which means that the quadric Q is mapped into the "null cone" of the coordinate metric space $M'(p + 2, r + 1, s + 1)$ spanned by Z^μ, Z^{p+1}, Z^{p+2} and with a metric tensor

$$\eta'_{\mu\nu} = \begin{pmatrix} \eta_{\mu\nu} & \\ & -1+1 \end{pmatrix}. \tag{2.9}$$

Since the points belonging to the space-time are in the quadric Q, these points are all mapped into the null cones of $M'(p + 2, r + 1, s + 1)$. This fact means that the null vectors of the space $M'(p + 2, r + 1, s + 1)$ are more likely to be associated with physical reality than any other.

The transformations of coordinates in $M'(p + 2, r + 1, s + 1)$ induced by the projective group are

(a) $$\begin{aligned} Z'^\mu &= \Lambda^\mu{}_\nu Z^\nu + \frac{1}{\sqrt{2}}(Z^{p+1} + Z^{p+2})\Lambda^\mu, \\ Z'^{p+1} &= Z^{p+1} + \frac{1}{\sqrt{2}} \eta_{\mu\nu} \Lambda^\mu{}_\rho Z^\rho \Lambda^\nu + \frac{\sqrt{2}}{4} \eta_{\mu\nu} \Lambda^\mu \Lambda^\nu Y^0, \\ Z'^{p+2} &= Z^{p+2} - \frac{1}{\sqrt{2}} \eta_{\mu\nu} \Lambda^\mu{}_\rho Z^\rho \Lambda^\nu - \frac{\sqrt{2}}{4} \eta_{\mu\nu} \Lambda^\mu \Lambda^\nu Y^0, \end{aligned}$$

(b) $$\begin{aligned} Z'^\mu &= Z^\mu, \\ Z'^{p+1} &= \tfrac{1}{2}(C + 1/C) Z^{p+1} - \tfrac{1}{2}(C - 1/C) Z^{p+2}, \end{aligned}$$

(c) $$Z'^\mu = Z^\mu, \quad Z'^{p+1} = Z^{p+1}, \quad Z'^{p+2} = -Z^{p+2}.$$

Although the construction of $M'(p + 2, r + 1, s + 1)$ from $M(p, r, s)$ holds globally, the relations to the space-time $R4$ embedded in $M(p, r, s)$ are local. The infinitesimal transformations of the above group are obtained from

$$\Lambda^\mu{}_\nu = \delta^\mu{}_\nu + \epsilon^\mu{}_\nu, \quad \Lambda^\mu = \epsilon^\mu, \quad C = 1 + \epsilon,$$

where $\epsilon^\mu{}_\nu, \epsilon^\mu, \epsilon$ are infinitesimals of the same order and $\epsilon^{\mu\nu} = -\epsilon^{\nu\mu}$.

Neglecting second order infinitesimals, the infinitesimal transformations corresponding to (a), (b), (c) are

(a) $$\begin{aligned} Z'^\mu &= Z^\mu + \epsilon^\mu{}_\nu Z^\nu + \frac{1}{\sqrt{2}}(Z^{p+1} + Z^{p+2})\epsilon^\mu, \\ Z'^{p+1} &= Z^{p+1} + \frac{1}{\sqrt{2}} Z^\mu \epsilon_\mu, \\ Z'^{p+2} &= Z^{p+2} - \frac{1}{\sqrt{2}} Z^\mu \epsilon_\mu, \end{aligned} \tag{2.10}$$

(b) $$\begin{aligned} Z'^\mu &= Z^\mu, \quad Z'^{p+1} = Z^{p+1} - \epsilon Z^{p+2}, \\ Z'^{p+2} &= Z^{p+2} - \epsilon Z^{p+1}, \end{aligned}$$

(c) $$Z'^\mu = Z^\mu, \quad Z'^{p+1} = Z^{p+1}, \quad Z'^{p+2} = -Z^{p+2}.$$

The transformations (a) and (b) can be written in the compact form

$$\left.\begin{aligned} Z'^\mu &= Z^\mu + U^\mu, \\ U^\mu &= \epsilon^\mu{}_\nu Z^\nu, \end{aligned}\right\} \quad \mu, \nu = 1 \cdots p + 2, \tag{2.11}$$

and

$$\epsilon^\mu{}_{p+1} = \epsilon^\mu/\sqrt{2}, \quad \epsilon^\mu{}_{p+2} = \epsilon^\mu/\sqrt{2}, \quad \epsilon^{p+1}{}_{p+2} = -\epsilon,$$
$$\epsilon^{p+1}{}_\mu = \epsilon_\mu/\sqrt{2}, \quad \epsilon^{p+2}{}_\mu = -\epsilon_\mu/\sqrt{2}, \quad \epsilon^{p+2}{}_{p+1} = -\epsilon.$$

Using (2.9) we get $\epsilon^{\mu\nu} = -\epsilon^{\nu\mu}$, $\mu, \nu = 1 \cdots p + 2$. On the other hand, considering that the quadratic form $Z^\mu Z_\mu$ $(\mu = 1 \cdots p + 2)$ is also invariant under the reflections one concludes that the transformations (2.10) plus reflections belong to the improper homogeneous group of isometries of $M'(p + 2, r + 1, s + 1)$, $O(r + 1, s + 1)$. Note that the rotations in the planes (μ, ν), $(\mu, p + 1)$, $(\mu, p + 2)$ correspond to the subgroup (a) $P(r, s)$ of $C(r, s)$. The rotation $(p + 1, p + 2)$ corresponds

to the dilatation (b) and the reflection Z^{p+2} corresponds to the inversion (c).

3. CONFORMAL EMBEDDING BUNDLES

The relation between the coordinates of $M'(p+2, r+1, s+1)$ and those of $M(p, r, s)$ is given by

$$Z^\mu = Y^0 X^\mu, \quad Z^{p+1} = \frac{Y^0}{\sqrt{2}}(1 + \tfrac{1}{2}X^\mu X_\mu), \quad Z^{p+2} = \frac{Y^0}{\sqrt{2}}(1 - \tfrac{1}{2}X^\mu X_\mu). \tag{3.1}$$

Such relation is covariant under $C(r, s)$ and this means that although the isometric embedding space $M(p, r, s)$ is not invariant under the group $C(r, s)$ the new coordinate space $M'(p+2, r+1, s+1)$ is. For that reason we call $M'(p+2, r+1, s+1)$ the conformal embedding space of $R4$. If the number $t = p - 4$ is the embedding class of $R4$, the conformal embedding class will be $t + 2 = p - 2$. The conformal embedding bundle is the manifold

$$\beta = (M'(p+2, r+1, s+1), R4, \pi),$$

where $M'(p+2, r+1, s+1)$ is the fibre, $R4$ the base space and $\pi: M'(p+2, r+1, s+1) \to R4$ the projection map.

As we have seen the homogeneous fiber group of β, $O(r+1, s+1)$, induces the conformal group $C(r, s)$. The group $O(r+1, s+1)|_{R4}$ defined as the subgroup of $O(r+1, s+1)$ under the condition (2.6) induces the group $C(r, s)|_{R4}$ which, as we have seen, induces the group of conformal motions in $R4$. Therefore, $O(r+1, s+1)|_{R4}$ is the subgroup of the fiber group which gives the conformal motions in space-time.

The space-time conformally embedded in $M(p+2, r+1, s+1)$ is defined in the Gaussian system by the set of conditions $x^{\mathbf{A}} = 0$, $Y^0 = 1$. The relations between tensors in $M(p+2, r+1, s+1)$ and in $R4$ are obtained with the use of the Gaussian system of $M(p, r, s)$. Before that let us consider a few relations. From (3.1) we obtain

$$\frac{\partial Z^\mu}{\partial X^\nu} = Y^0 \delta^\mu_\nu, \quad \frac{\partial Z^{p+1}}{\partial X^\nu} = \frac{Y^0}{\sqrt{2}} X_\nu, \quad \frac{\partial Z^{p+2}}{\partial X^\nu} = -\frac{Y^0}{\sqrt{2}} X_\nu \tag{3.2}$$

and from the inverse of (3.1)

$$X^\mu = Z^\mu / Y^0, \quad X^\mu X_\mu = \frac{\sqrt{2}}{Y^0}(Z^{p+1} - Z^{p+2}), \quad Y^0 = \frac{1}{\sqrt{2}}(Z^{p+1} + Z^{p+2}), \tag{3.3}$$

we get

$$\frac{\partial X^\mu}{\partial Z^\nu} = \frac{1}{Y^0}\delta^\mu_\nu. \tag{3.4}$$

However, when Y^0 is an independent parameter,

$$1 = \frac{\partial Z^{p+1}}{\partial X^\rho}\frac{\partial X^\rho}{\partial Z^{p+1}} = \frac{Y^0}{\sqrt{2}} X_\rho \frac{\partial X^\rho}{\partial Z^{p+1}},$$

so that

$$\frac{\partial X^\mu}{\partial Z^{p+1}} = \frac{\sqrt{2}}{Y^0}\frac{X^\mu}{X^\rho X_\rho}. \tag{3.5}$$

Similarly,

$$\frac{\partial X^\mu}{\partial Z^{p+2}} = -\frac{\sqrt{2}}{Y^0}\frac{X^\mu}{X^\rho X_\rho}. \tag{3.6}$$

From (3.3) we also get

$$X_\mu X^\mu{}_{p+1} = \frac{\sqrt{2}}{2Y^0}\left(1 - \frac{\partial Z^{p+2}}{\partial Z^{p+1}}\right), \quad X_\mu X^\mu{}_{p+2} = -\frac{\sqrt{2}}{2Y^0}\left(1 - \frac{\partial Z^{p+1}}{\partial Z^{p+2}}\right)$$

$$\frac{\partial Y^0}{\partial Z^{p+1}} = \sqrt{2}\left(1 + \frac{\partial Z^{p+2}}{\partial Z^{p+1}}\right), \quad \frac{\partial Y^0}{\partial Z^{p+2}} = \sqrt{2}\left(1 + \frac{Z^{p+1}}{Z^{p+2}}\right).$$

But from (3.2) and (3.5), $\partial Z^{p+2}/\partial Z^{p+1} = -1$, so that

$$\frac{\partial Y^0}{\partial Z^{p+1}} = \frac{Y^0}{\partial Z^{p+2}} = 0 \text{ as expected and } X_\mu X^\mu{}_{p+1} = -X_\mu X^\mu{}_{p+2} = \sqrt{2}/Y^0.$$

Now, contraction of (3.2) with $\partial X^\mu/\partial x^\alpha$ gives

$$Z^\mu{}_\alpha = \frac{\partial Z^\mu}{\partial x^\nu}\frac{\partial X^\nu}{\partial x^\alpha} = Y^0 X^\mu{}_\alpha, \quad Z^{p+1}{}_\alpha = \frac{\partial Z^{p+1}}{\partial X^\nu}\frac{\partial X^\nu}{\partial x^\alpha} = \frac{Y^0}{\sqrt{2}} X_\nu X^\nu_\alpha, \quad Z^{p+2}{}_\alpha = \frac{\partial Z^{p+2}}{\partial x^\nu}\frac{\partial X^\nu}{\partial x^\alpha} = -\frac{Y^0}{\sqrt{2}} X_\nu X^\nu{}_\alpha. \tag{3.7}$$

On the other hand, contraction of (3.4), (3.5), and (3.6) with $\partial x^\alpha/\partial X^\mu$ gives

$$\frac{\partial x^\alpha}{\partial Z^\mu} = \frac{\partial x^\alpha}{\partial X^\rho}\frac{\partial X^\rho}{\partial Z^\mu} = \frac{1}{Y^0} x^\alpha_\mu, \quad \frac{\partial x^\alpha}{\partial Z^{p+1}} = \frac{\partial x^\alpha}{\partial X^\rho}\frac{\partial X^\rho}{\partial Z^{p+1}} = \frac{\sqrt{2}}{Y^0}\frac{X^\rho x_\rho{}^\alpha}{X^\mu X_\mu}, \quad \frac{\partial x^\alpha}{\partial Z^{p+2}} = \frac{\partial x^\alpha}{\partial X^\rho}\frac{\partial X^\rho}{\partial Z^{p+2}} = -\frac{\sqrt{2}}{Y^0}\frac{X^\rho x_\rho{}^\alpha}{X^\mu X_\mu}. \tag{3.8}$$

The objects given by (3.7), (3.8) are the elements of the Jacobian matrix of the transformation between the Gaussian and the Z^μ coordinates in $M'(p+2, r+1, s+1)$. Therefore, they relate the components of tensors from one to the other system. Thus, if $U = (U^\mu, U^{p+1}, U^{p+2})$ is a vector in $M'(p+2, r+1, s+1)$, the p Gaussian components of the same vector are

$$\xi^\alpha = \frac{\partial x^\alpha}{\partial Z^\mu}U^\mu + \frac{\partial x^\alpha}{\partial Z^{p+1}}U^{p+1} + \frac{\partial x^\alpha}{\partial Z^{p+2}}U^{p+2}$$

and, in particular, the space-time components are

$$\xi^i\Big|_{R4} = \left(\frac{1}{Y^0}\frac{\partial x^i}{\partial X^\mu}U^\mu + \frac{\sqrt{2}}{Y^0}\frac{X^\rho x_\rho{}^i}{X^\mu X_\mu}(U^{p+1} - U^{p+2})\right)\Big|_{R4}.$$

In the case where U is an infinitesimal generator of a transformation (2.11), then (2.11) and (3.3) gives

$$U^{p+1} - U^{p+2} = (\epsilon^{p+1}_\mu - \epsilon^{p+2}_\mu)Z^\mu + \epsilon^{p+1}{}_{p+2}Z^{p+2} - \epsilon^{p+2}{}_{p+1}Z^{p+1} = \frac{2Y^0}{\sqrt{2}}\epsilon_\mu X^\mu + \epsilon\frac{Y^0}{\sqrt{2}}X^\mu X_\mu,$$

and

$$U^\mu = \epsilon^\mu{}_\nu Z^\nu + \epsilon^\mu{}_{p+1}Z^{p+1} + \epsilon^\mu{}_{p+2}Z^{p+2} = Y^0(\epsilon^\mu{}_\nu X^\mu + \epsilon^\mu),$$

so that

$$\xi^\alpha = x^\alpha{}_\mu X^\nu \epsilon^\mu{}_\nu + \left(\tfrac{1}{2}x^\alpha{}_\mu + \frac{2X^\rho x^\alpha_\rho}{X^\nu X_\nu}X_\mu\right)\epsilon^\mu + X^\rho x_\rho{}^\alpha \epsilon \tag{3.9}$$

from which we obtain the conformal Killing vector field in $R4$, $\xi^i|_{R4}$. Observe that the resulting components are independent of Y^0.

Defining the functions

$$\epsilon^{\alpha\beta} = \frac{\partial x^\alpha}{\partial Z^\mu}\frac{\partial x^\beta}{\partial Z^\nu}\epsilon^{\mu\nu}, \quad \mu,\nu = 1\cdots p+2, \tag{3.10}$$

we can count the various transformations induced in $R4$: $\epsilon^{ij}|_{R4}$, $\epsilon^{iA}|_{R4}$. The transformations $\epsilon^{iA}|_{R4}$ are associated with translations in $R4$, obtained from group contractions. To this end we define the translation parameters $\pi^i = \alpha_A \epsilon^{iA}$ where α_A are appropriate functions of the space-time curvature which vanish in the flat limit. We can write ξ^α in terms of $\epsilon^{\alpha\beta}$:

$$\xi^\alpha = \epsilon^{\alpha\beta}(Z^\mu{}_\beta Z_\mu + Z^{p+1}{}_\beta Z_{p+1} + Z^{p+2}{}_\beta Z_{p+2})$$
$$= \epsilon^{\alpha\beta}\left(Y^0{}^2 X^\mu_\beta X_\mu + \frac{Y^0}{\sqrt{2}} X^\mu{}_\beta X_\mu (Z^{p+1} - Z^{p+2})\right);$$

therefore, the basic condition (2.6) is equivalent to

$$\xi^A|_{R4} = (\epsilon^{AJ} X^\mu{}_J X_\mu + \epsilon^{AB} X^\mu{}_B X_\mu)|_{R4} = 0. \tag{3.11}$$

4. CONFORMAL SPINORS

The construction of conformal spinor fields follows the same process used to construct isospinors. The only difference is that now we use as the algebra space the fibers of the conformal embedding bundles $M'(p+2, r+1, s+1)$ instead of $M(p, r, s)$. The Clifford algebra $\mathcal{C}_{r+1,s+1}$ constructed on $M(p+2, r+2, s+1)$ is generated by $p+2$ elements e_μ, e_{p+1}, e_{p+2} such that

$$e_{(\mu}\ e_{\nu)} = \eta'_{\mu\nu}, \quad \mu,\nu = 1\cdots p+2, \tag{4.1}$$

where $\eta'_{\mu\nu}$ is given by (2.9). The group of automorphisms of such an algebra is isomorphic to $O(r+1, s+1)$ when $p+2$ is even and isomorphic to $SO(r+1, s+1)$ when $p+2$ is odd. To obtain the spinors we use the same Weyl representation as in the case of isospinors which is given by the matrices P_α, Q_α, P_0 $(\alpha = 1\cdots\nu)$ [Ref. 1, expression (IV.2)].

The final representation is obtained by choosing the matrices P_α, Q_α, P_0 to represent the algebraic generators e_μ. By renaming these matrices we get equivalent representations. Having chosen this representation the Lie algebra of the group of automorphisms is given by the generators

$$L_{\mu\nu} = \tfrac{1}{2} e_{[\mu} e_{\nu]}, \quad \mu\nu = 1\cdots p+2, \tag{4.2}$$

which satisfy the commutation relation

$$[L_{\mu\nu} L_{\rho\sigma}] = -i(\eta_{\mu\rho} L_{\nu\sigma} + \eta_{\nu\sigma} L_{\mu\rho} - \eta_{\mu\sigma} L_{\nu\rho} - \eta_{\nu\rho} L_{\mu\sigma}), \tag{4.3}$$

which are the same as those of the Lie algebra of $O(r+1, s+1)$ or $SO(r+1, s+1)$ according to the case. The covering group of each of these groups is generated by the matrices

$$S = 1 + \tfrac{1}{2}\epsilon^{\mu\nu} L_{\mu\nu}. \tag{4.4}$$

As in the case of isospinors these matrices are not always Hermitian and they satisfy a relation of the type

$$M = \pm S^\dagger M S, \tag{4.5}$$

where M is a characteristic of the covering group, and the signs (+) hold for proper transformations while (−) holds for improper transformations. At the end of this paper we give a table with the covering groups of the conformal groups in general relativity, based on the Weyl representation.

In order to relate the algebraic elements of $\mathcal{C}_{r+1,s+1}$ to $R4$ we take the p Gaussian algebraic elements

$$e_\alpha = Z^\mu{}_\alpha e_\mu, \quad \mu = 1\cdots p+2. \tag{4.6}$$

They satisfy

$$e_{(\alpha} e_{\beta)} = Z^\mu{}_\alpha Z^\mu{}_\beta \eta'_{\mu\nu} = g_{\alpha\beta}, \quad \mu,\nu = 1\cdots p+2,$$

and $e_{(i} e_{j)} = g_{ij}(R4)$. On the other hand, we have $e^\alpha = g^{\alpha\beta} e_\beta = g^{\alpha\beta} Z^\mu{}_\beta e_\mu$ $(\mu = 1\cdots p+2)$. But since

$$g^{\alpha\beta} = \frac{\partial x^\alpha}{\partial Z^\mu}\frac{\partial x^\beta}{\partial Z^\nu}\eta'^{\mu\nu}, \quad \mu,\nu = 1\cdots p+2, \tag{4.7}$$

we have

$$e^\alpha = \frac{\partial x^\alpha}{\partial Z^\mu}\frac{\partial x^\beta}{\partial Z^\mu}\frac{\partial Z^\rho}{\partial x^\beta} e_\rho \eta'^{\mu\nu} = \frac{\partial x^\alpha}{\partial Z^\mu} e^\mu, \quad \mu,\nu,\rho = 1\cdots p+2,$$

and

$$e^{(\alpha} e^{\beta)} = \frac{\partial x^\alpha}{\partial Z^\mu}\frac{\partial x^\beta}{\partial Z^\nu} e^{(\mu} e^{\nu)} = g^{\alpha\beta}, \quad \mu,\nu = 1\cdots p+2.$$

On the other hand, the Gaussian components of the Lie algebra operators are

$$L_{\alpha\beta} = Z^\mu{}_\alpha Z^\mu{}_\beta L_{\mu\nu} = \tfrac{1}{2} e_{[\alpha} e_{\beta]}, \quad \mu = 1\cdots p+2, \tag{4.8}$$

which satisfy the commutation relation

$$[L_{\alpha\beta} L_{\gamma\delta}] = -i(g_{\alpha\gamma} L_{\beta\delta} + g_{\beta\delta} L_{\alpha\gamma} - g_{\alpha\delta} L_{\beta\gamma} - g_{\beta\gamma} L_{\alpha\delta}). \tag{4.9}$$

The translation operators can be defined as $\hat{\pi}_i = \alpha^A L_{iA}$, where also α^A vanish in the flat limit.

The commutators between $L_{ij}, L_{AB}, \hat{\pi}_i$ are derived directly from (4.9) and it can easily be seen that in the flat limit they give the Lie algebra of the conformal group in Minkowski space-time. A further step in the representation of the Clifford algebra leads us to the conformal spinors. In short, we consider a left ideal of the Clifford algebra generated by a specific element $\phi_0 \in \mathcal{C}_{r+1,s+1}$. This ideal defines a linear map $\psi(X) = X\phi_0$ for all $X \in \mathcal{C}_{r+1,s+1}$. Restricting this map to the set of inversible elements of the Clifford algebra, we obtain the spinor representation of the Clifford algebra[6]. Under this representation where e_μ is one generator of the algebra, we can decompose $\psi(e_\mu)$ in the product $\phi(e_\mu)\chi(e_\mu)$. Since $e'_\mu = S e_\mu S^{-1}$ then $\psi(e'_\mu) = S e_\mu S^{-1}\phi_0$. We can always choose ϕ_0 so that $S^{-1}\phi_0 S = \phi_0$. Defining $\Delta^{-1} = S^{-1}\phi_0$ and $\bar{\Delta} = \phi_0 S$, then the above condition is equivalent to $\Delta^{-1} = \bar{\Delta}^{-1}$, which indicates the double valuedness of the spinor representation of the group of automorphisms, given by the elements Δ and $\bar{\Delta}$. Taking $\psi(e'_\mu) = S e_\mu \phi_0 S^{-1} = S\psi(e_\mu)S^{-1}$, then for $\psi(e'_\mu) = \phi(e'_\mu)\chi(e'_\mu)$ we get the spinor transformations $\phi(e'_\mu) = S\phi(e_\mu)$ and $\chi(e'_\mu) = \chi(e_\mu)S^{-1}$. The spinors transforming like ϕ are covariant spinors and those transforming like χ are contravariant spinors. When a Weyl representation of the algebra is given, ϕ and χ are column and row matrices, respectively. The spinor space has dimension 2^ν with $\nu = (p+2)/2$ when p is even and $\nu = (p+1)/2$ when p is odd. When p is even we can split the spinors in two equivalent halves called semispinors[7] provided one considers only the group $SO(r+1, s+1)$ acting on the space $M'(p+2, r+1, s+1)$. If the group $O(r+1, s+1)$ is to be considered (that is with reflections included) as it is in the case of conformal groups then semispinors may occur when ν is odd.

Since the isometric embedding classes in general relativity are at least 4 (flat case) and at most 10, the conformal embedding classes vary between 6 and 12 and the cases of ν odd are $P+2 = 6, 7, 10, 11$. However, for $p+2 = 8$ semispinors will also take place thanks to the triality principle of eight dimensional Clifford algebras.[8,9]

The conformal spinors are obtained from this representation of $\mathcal{C}_{r+1,s+1}$ by imposing on the spinor group the same restriction (2.6) applied to the group $O(r+1,$

$s + 1)$. In other words, as the spinor group forms a representation of $O(r + 1, s + 1)$, the conformal spinor group is the corresponding representation of $O(r + 1, s + 1)|_{R4}$. More explicitly, the conformal spinor group is the group generated by the infinitesimal matrices

$$S|_{R4} = 1 + \tfrac{1}{2}\epsilon^{\mu\nu}L_{\mu\nu}|_{R4}. \tag{4.10}$$

It follows from the results in the preceding sections that this group forms a double-valued representation of the conformal group of the space-time $R4$. The spinors defined on $M'(p + 2, r + 1, s + 1)$ which transform under the group generated by (4.10) are called the conformal spinors of $R4$. Since the group of conformal motions in $R4$ is induced by $O(r + 1, s + 1)$, strictly speaking, conformal spinors can only be defined in space-times with even-dimensional embedding space (p even). This is due to the fact that one of the reflections included in $O(r + 1, s + 1)$ is responsible for the inversion of the conformal transformation. Nevertheless, if we consider restricted conformal transformations (without the inversion), then conformal spinors in odd dimension embeddable space-times can also be constructed.

Now we consider the problem of relating conformal spinors to space-time. The vector U in $M'(p + 2, r + 1, s + 1)$ corresponds to the algebraic element $U = U^{\mu}e_{\mu} + U^{p+1}e_{p+1} + U^{p+2}e_{p+2}$. In the spinor representation of the algebra this algebraic element gives a $2^{\nu} \times 2^{\nu}$ matrix or rank two spinor whose elements are denoted by $\psi^{A}{}_{\dot{B}}$ (capital italic spinor indices run from 1 to 2^{ν} or from 1 to $2^{\nu-1}$ in the case of semi-spinors). The dot over one of the indices means that the transformation corresponding to that index is the complex conjugate of the transformations of the undotted index. The relation between $\psi^{A}{}_{\dot{B}}$ and U^{μ} is given by

$$\psi^{A}{}_{\dot{B}} = e_{\mu}{}^{A}{}_{\dot{B}}U^{\mu}, \qquad \mu = 1\cdots p + 2,$$

where $e_{\mu}{}^{A}{}_{\dot{B}}$ are the spin-tensors associated with each generator e_{μ} of the algebra. Conversely, given the rank two spinor $\psi^{A}{}_{\dot{B}}$ we get a vector with components

$$U_{\mu} = e'_{\mu}{}^{A}{}_{\dot{B}}\psi^{\dot{B}}{}_{A}, \qquad \mu = 1\cdots p + 2,$$

where spinor indices are raised and lowered by the two versions of the metric spinor $\epsilon^{AB}, \epsilon_{AB}$. The p Gaussian components of the spin-tensors are defined by

$$e^{A}_{\alpha\dot{B}} = Z^{\mu}{}_{\alpha}e_{\mu}{}^{A}{}_{\dot{B}}, \qquad \mu = 1\cdots p + 2. \tag{4.11}$$

These spin tensors connect spinors to vectors in the Gaussian frame of $M(p, r, s)$:

$$\xi_{\alpha} = e_{\alpha}{}^{A}{}_{\dot{B}}\psi^{\dot{B}}{}_{A}, \qquad \psi^{A}{}_{\dot{B}} = e_{\alpha}{}^{A}{}_{\dot{B}}\xi^{\alpha} \tag{4.12}$$

From which we can obtain the relation between space-time vectors $\xi^{i}(R4)$ and conformal spinors. We note that the space-time as well as $M(p, r, s)$ is entirely contained in the null cone of $M'(p + 2, r + 1, s + 1)$ so that only conformal spinors associated with null vectors of $M'(p + 2, r + 1, s + 1)$ have physical interest.

Let U be a vector in $M'(p + 2, r + 1, s + 1)$, and $\xi^{\alpha} = (\partial x^{\alpha}/\partial Z^{\mu})U^{\mu}$ $(\mu = 1\cdots p + 2)$. Then $U^{\mu}U_{\mu} = \xi^{\alpha}\xi_{\alpha}$ $(\mu = 1\cdots p + 2)$. Thus if ξ^{α} is a null vector defined in space-time, $\xi^{\mathbf{A}} = 0$ and $U^{\mu}U_{\mu} = \xi^{i}\xi_{i} = 0$. Therefore, null vectors in space-time are associated with physical conformal spinors. On the other hand, if $U^{\mu}U_{\mu} = 0$ then $\xi^{i}\xi_{i}|_{R4} = -\xi^{\mathbf{A}}\xi_{\mathbf{A}}|_{R4}$.

Therefore, the conformal spinors associated with U^{μ}, although possibly having a physical character, do not necessarily correspond to a null vector in space-time. This fact seems to have significance in a possible future application of conformal spinors to Penrose's angular momentum quantized physical manifold.[10]

The simplest example of conformal spinors which may occur in relativity is that of the case of Minkowski space-time. In this case the minimal isometric embedding $M(4, 3, 1)$ coincides with the space-time itself. As a consequence the resulting spinor structure is globally defined in this space-time (a peculiarity of this example). The conformal embedding space is $M'(6, 4, 2)$ whose homogeneous group of isometries $O(4, 2)$ induce the conformal group of $M(4, 3, 1)$. Observe that the condition (2.8) here is trivially satisfied so that $O(4, 2)|_{R4} = O(4, 2)$. The Lie algebra of this group is given by $L_{\mu\nu} = \frac{1}{2}e_{[\mu}e_{\nu]}$ where e_{μ} are the generators of the Clifford algebra defined on $M(6, 4, 2)$. Taking the Weyl representation $e_1 = iP_1$, $e_2 = iQ_2$, $e_3 = P_2$, $e_4 = P_3$, $e_5 = P_4$, $e_6 = Q_3$, we obtain the Lie algebra of the covering group $SU(2, 2)$. The elements of the Lie algebra in such representation take the form

$$L_{\mu\nu} = \left(\begin{array}{c|c} L^{(1)}_{\mu\nu} & \\ \hline & L^{(2)}_{\mu\nu} \end{array}\right);$$

since $p + 2$ is even and ν is odd we can consider semi-spinors and the Lie algebra of the covering group are given only by $L^{(1)}_{\mu\nu}$ (or equivalently by $L^{(2)}_{\mu\nu}$). They satisfy the commutation relations

$$[L^{(1)}_{\mu\nu}L^{(1)}_{\rho\sigma}] = -i(\eta'_{\mu\rho}L^{(1)}_{\nu\sigma} + \eta_{\nu\sigma}L^{(1)}_{\mu\rho} - \eta_{\mu\sigma}L^{(1)}_{\nu\rho} - \eta_{\nu\rho}L^{(1)}_{\mu\sigma}),$$

$$[L^{(1)}_{\mu\nu}L^{(1)}_{\rho 5}] = -i(\eta_{\mu\rho}L^{(1)}_{\nu 5} - \eta_{\nu\rho}L^{(1)}_{\mu 5}),$$

$$[L^{(1)}_{\mu\nu}L^{(1)}_{\rho 6}] = -i(\eta_{\mu\rho}L^{(1)}_{\nu 6} - \eta_{\nu\rho}L^{(1)}_{\mu 6}),$$

$$[L^{(1)}_{\mu\nu}L^{(1)}_{56}] = 0, \qquad [L^{(1)}_{\mu 5}L^{(1)}_{\nu 5}] = -[L^{(1)}_{\mu 6}, L^{(1)}_{\nu 6}] = -iL^{(1)}_{\mu\nu},$$

$$[L^{(1)}_{\mu 5}L^{(1)}_{\nu 6}] = -i\eta_{\nu\mu}L^{(1)}_{56}, \qquad [L^{(1)}_{\mu 5}L^{(1)}_{56}] = -iL^{(1)}_{\mu 6},$$

$$[L^{(1)}_{\mu 6}L^{(1)}_{56}] = -iL^{(1)}_{\mu 5}.$$

Defining $\hat{\pi}_{\mu} = \alpha^5 L_{\mu 5} + \alpha^6 L_{\mu 6}$, α^5, α^6 constants, then

$$[L^{(1)}_{\mu\nu}\hat{\pi}_{\rho}] = -i(\eta_{\mu\rho}P_{\nu} - \eta_{\nu\rho}P_{\mu}),$$

$$[\hat{\pi}_{\mu}\hat{\pi}_{\nu}] = 0 \qquad [\hat{\pi}_{\mu}L^{(1)}_{56}] = -i\hat{\pi}_{\mu}$$

which shows that the rotations in the planes $(\mu, 5)$, $(\mu, 6)$ give translations in $M(4, 3, 1)$.

The resulting four component (semi) spinors can have a one-to-one relationship to Penrose's twistors. Let ζ^{i} be the four components of a twistor of valence $[\frac{1}{0}]$, and ϕ^{A} a covariant conformal spinor. We set $\zeta^1 = \phi^1$, $\zeta^2 = \phi^2$, $\zeta^3 = \phi^3$, $\zeta^4 = \phi^4$.

The twistor conjugation is an operation defined by

$$\zeta_1 = \zeta^{*3}, \quad \zeta_2 = \zeta^{*4}, \quad \zeta_3 = \zeta^{*1}, \quad \zeta_4 = \zeta^{*2},$$

where the * means complex conjugation. The result is a twistor of valence $[\frac{1}{0}]$.

In terms of conformal spinors this conjugation corresponds to the operation $\bar{\phi} = \phi^{\dagger}K$ where

$$K = \left(\begin{array}{cc|cc} & & 1 & \\ & & & 1 \\ \hline 1 & & & \\ & 1 & & \end{array}\right).$$

As ϕ is represented by a column matrix, $\bar{\phi}$ is represented by a row matrix. Twistors are classified as right-handed, null, or left-handed according to whether $\bar{\phi}\phi = \phi^{\dagger}K\phi$ is positive, zero, or negative, respectively. This classification may be extended to conformal spinors in general but for each value of $p + 2$ the matrix K has to be redefined.

Let us consider another example of conformal spinors, this time in general relativity. The obvious example would be the case of Schwarzschild space-time. However, the group of conformal motions in this case reduce to ordinary motions, as a consequence of the static condition.[11] This means that the conformal spinors in Schwarzschild space-time should coincide with the isospinors in the same space-time.

However, if we take the maximal analytic extension of this space-time, that is Kruskal's space-time, it ceases to be static in the region interior to Schwarzschild radius and genuine conformal motions take place (taking 2m = 1 this region is characterized by $r < 1$). The isometric embedding class of Kruskal space-time is $M(6, 5, 1)$ and the embedding is given by the functions[12]

$$X^1 = 2(1 - 1/r)^{1/2}\sinh(t/2), \quad X^2 = 2(1 - 1/r)^{1/2}\cosh(t/2),$$

$$X^3 = \int_1^r \left(\frac{r^2 + r + 1}{r^3}\right)^{1/2} dr,$$

$X^4 = r\cos\theta$, $X^5 = r\sin\theta\cos\phi$, $X^6 = r\sin\theta\sin\Phi$, such that $\eta'^{\mu\nu}X^\mu X^\nu|_{R4} = X^\mu X_\mu|_{R4} = f(r)$, where $\eta'_{\mu\nu} = \text{diag}$ $(-+++++)$ and

$$f(r) = 4(1 - 1/r) - \left[\int_1^r \left(\frac{r' + r + 1}{r^3}\right)^{1/2} dr\right]^2 + r^2.$$

We take the Gaussian coordinates to be $x^1 = r$, $x^2 = \theta$, $x^3 = \phi$, $x^4 = t$ and $x^5 = X^2 + a$, $x^6 = X^3 + b$ where a and b are constants chosen so that the space-time $R4$ is defined by $x^5 = x^6 = 0$. From the functions $X^\mu(x^\alpha)$ and $x^\alpha(X^\mu)$ we obtain the elements of the Jacobian matrix: $\partial X^\mu/\partial x^\alpha$, $\partial x^\alpha/\partial X^\mu$. Then as indicated by (3.7), (3.8) we get the elements $\partial Z^\mu/\partial x^\alpha$, $\partial x^\alpha/\partial Z^\mu$. The conformal embedding space is $M(8, 6, 2)$ with a metric tensor $\eta'_{\mu\nu} = \text{diag}\ (-+++++--+)$. The Weyl representation for the Clifford algebra $\mathcal{C}_{6,2}$ is

$$e_1 = iP_1, \quad e_2 = P_2, \quad e_3 = P_5, \quad e_4 = P_4, \quad e_5 = P_3,$$
$$e_6 = Q_4, \quad e_7 = iQ_2, \quad e_8 = Q_3. \tag{4.13}$$

As in the previous example the operators of the Lie algebra of $O(r + 1, s + 1)$ are represented by matrices like

$$L_{\mu\nu} = \tfrac{1}{2}e_{[\mu}e_{\nu]} = \left(\begin{array}{c|c} L^{(1)}_{\mu\nu} & \\ \hline & L^{(2)}_{\mu\nu} \end{array}\right), \quad \mu, \nu = 1\cdots 8$$

and the generators of the spinor group are

$$S = 1 + \tfrac{1}{2}\epsilon^{\mu\nu}L_{\mu\nu}, \quad \mu, \nu = 1\ldots 8,$$

having the characteristic matrix

$$\mu = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix} \otimes \left(\begin{array}{cccc|cccc} 1 & & & & & & & \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & 1 & & & & \\ \hline & & & & -1 & & & \\ & & & & & -1 & & \\ & & & & & & -1 & \\ & & & & & & & -1 \end{array}\right). \tag{4.13'}$$

We denote such spinor group by $SU(1, 1) \otimes SU(4, 4)$. To obtain the conformal spinor group we have in the first place to consider the condition (2.6) or (3.11) which in the case of Kruskal space-time has the coordinate free solution

$$\epsilon^{23} = \epsilon^{24} = \epsilon^{25} = \epsilon^{26} = \epsilon^{34} = \epsilon^{35} = \epsilon^{36} = 0$$
$$\epsilon^2 = \epsilon^3 = \epsilon^4 = \epsilon^5 = \epsilon^6 = 0. \tag{4.14}$$

Furthermore, since we have an eight-dimensional space $M(8, 6, 2)$ we can use the triality principle to obtain eight component semispinors whose spinor group is only $SU(4, 4)$. Therefore, the conformal spinor group of Kruskal's space-time is $SU(4, 4)|_{kr}$ defined by (4.14).

It is interesting to notice the change in the group structure when we compare the conformal groups in Schwarzschild and Kruskal space-times. As the Schwarzschild singularity is shrunk to the origin to produce Kruskal's space-time, genuine conformal motions appear. This seems physically reasonable if we consider that the Schwarzschild solution produces an inertial effect on test particles and these massive particles could not withstand conformal motions. Finally, we remark that the same applications which were mentioned for isospinors also hold for conformal spinors; in particular, the use of conformal spinors for a possible theory of conformally covariant structure of elementary systems in general relativity. These would be obtained by the classification of the continuous unitary irreducible representations of the various conformal spinor groups.

In this respect it is interesting to notice that in the last example the group $SU(3)$ appears as a subgroup of the fiber group of Kraskal's conformal embedding bundle but it does not appear in the fiber group of Schwarzschild's embedding bundle, as if the appearance of group $SU(3)$ in the case were a consequence of the shrinking of Schwarzschild's singularity.

APPENDIX

In the following table we give the various conformal embedding spaces which may occur in general relativity, the homogeneous fiber groups $O(r + 1, s + 1)$, the spinor groups, and some important subgroups. The notation used for the groups is the same as in the isospinor paper.

TABLE I. Conformal embedding spaces, spinor groups and subgroups.

$p + 2$	Conformal embedding	(Spinors) covering group	Main "Little groups"	Important subgroups
6	$M'(6, 4, 2)$	$SU(2, 2)$		
8	$M'(8, 6, 2)$	$SO(4, 4)$		$SU(4)$
8	$M'(8, 5, 3)$	$SL(8, C)$	$SU(2, 2, 2, 2)$	$SU(2) \times SU(2)$
8	$M'(8, 4, 4)$	$SO(2, 2, 2, 2)$		$SU(2) \times SU(2)$
10	$M'(10, 8, 2)$	$SO(8, 8)$		$SU(8)$
10	$M'(10, 7, 3)$	$SL(16\ C)$	$SU(4, 4, 4, 4)$ $SU(2, 2, 2, 2, 2, 2, 2, 2)$	
10	$M'(10, 6, 4)$	$SU(4, 4, 4, 4)$		$SU(4)$
10	$M'(10, 5, 5)$	$SL'(16\ C)$	$SU(2, 2, 2, 2, 2, 2, 2, 2)$	$SU(2) \times SU(2)$
12	$M'(12, 10, 2)$	$SU(16, 16)$		$SU(8)$
12	$M'(12, 11, 3)$	$SL(32, C)$		$SU(8)$
12	$M'(12, 8, 4)$	$SU(8, 8, 8, 8)$		$SU(8)$
12	$M'(12, 7, 5)$	$SL'(32, C)$		$SU(4)$
12	$M'(12, 6, 6)$			$SU(4)$

Note that all spinor groups in the table are semispinors.

$SL'(16\ C)$ differ from $SL(16, C)$ by the fact that $SL'(16\ C)$ does not contain the "little group" $SU(4, 4, 4, 4)$. Also $SL'(32, C)$ and $SL(32, C)$ differ in their "main little groups".

[1] M. D. Maia, J. Math. Phys. 14, 882 (1973).
[2] R. Penrose, J. Math. Phys. 8, 345 (1967).
[3] R. Penrose, Int. J. Theor. Phys. 1, 61 (1968).
[4] P. Szekeres, Nuovo Cimento 43, 1062 (1966).
[5] N. H. Kuiper, Ann. Math. 50, 916 (1949).
[6] C. C. Chevalley, *The Construction and Study of Certain Important Algebras* (Math. Soc. of Japan, Tokyo, 1955).
[7] E. Cartan, *The Theory of Spinors* (M.I.T. Press, Cambridge, Mass., 1966).
[8] C. C. Chevalley, *The Algebraic Theory of Spinors* (Columbia U. P., New York, 1954).
[9] A. Gamba, J. Math. Phys. 8, 775 (1967).
[10] R. Penrose, "An Analysis of the Structure of Space Time." Adam's Prize Essay, 1967.
[11] H. Takeno, *The Theory of Spherically Symmetric Space Times* (Hiroshima University, Hiroshima, 1963).
[12] C. Fronsdal, Phys. Rev. 116, 778 (1959).

The Kerr congruence*

Ezra T. Newman and Jeffrey Winicour

Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260
(Received 16 July 1973)

We present the special relativistic Kerr congruence from two new points of view. The importance of the Kerr congruence stems from its role in the general relativistic rotating body problem. In the special relativistic limit, it is the simplest twisting generalization of the ordinary light cone. We first show that the Kerr congruence can be obtained from the light cone of complex Minkowski space, which induces a family of oblate spheroids in real Minkowski space. We then show that the rigid rotation of one of these spheroids with constant angular velocity also generates the Kerr congruence by shining comoving flashlights normal to the surface. In fact, the oblate spheroid is the unique surface which generates a shear-free twisting null congruence in this manner. This result has direct generalization to the full Kerr geometry.

1. INTRODUCTION

In this paper, we report a curious result which appears to have important bearing on the relativistic theory of angular momentum. Our work originated from considerations of the Kerr metric[1] which describes the general relativistic geometry exterior to a rotating axisymmetric body. Associated with the Kerr metric is a shear-free, twisting congruence of null geodesics. In terms of the tangents k^μ to these null geodesics, the Kerr metric may be expressed in the Kerr-Schild form[2]

$$g_{\mu\nu} = \eta_{\mu\nu} + 2Hk_\mu k_\nu \tag{1.1}$$

where $\eta_{\mu\nu}$ is the Minkowski metric,

$$k^\mu = \left(1, -\frac{(rx + ay)}{r^2 + a^2}, -\frac{(ry - ax)}{r^2 + a^2}, -\frac{z}{r}\right),$$
$$2H = 2mr^3/r^4 + a^2z^2), \tag{1.2}$$

and r is defined by

$$(x^2 + y^2)/(r^2 + a^2) + z^2/r^2 = 1. \tag{1.3}$$

Here we consider the flat-space limit of this congruence obtained by setting the mass of the Kerr geometry equal to zero. This induces a shear-free, twisting, geodesic, null congruence in Minkowski space, referred to as the special relativistic Kerr congruence.[2]

It is well known that shear-free, twist-free, geodesic null congruences in special relativity are generated by the light cones emanating from a time like world line. These congruences form null hypersurfaces whose null rays are orthogonal to families of two-dimensional spheres. The congruences attached to nonaccelerating world lines are the zero twist limit of the Kerr congruence. In keeping with general relativistic nomenclature, we refer to this limiting case as the Schwarzschild congruence. In Sec. 2, we show how the Kerr congruence can be interpreted as a complexified version of the Schwarzschild congruence. From this point of view, the Kerr congruence corresponds to the light cone of an unaccelerated world line in complex Minkowski space.

From the point of view of real Minkowski space, the Kerr congruence, due to its twist, does not form null hypersurfaces. Consequently, the Kerr congruence does not have a simple physical interpretation as does the Schwarzschild congruence which can be generated by flashing light rays normal to the surface of a sphere. However, in Sec. 3 we present our main result that the Kerr congruence can be generated by flashing light rays "kinematically normal" to the surface of a rigidly rotating oblate spheroid (ellipsoid of revolution). In fact, the Kerr congruence is unique in this regard. No other axisymmetric surface rigidly rotating about its axis generates a shear-free congruence.

Because of the invariance of shear under the addition of the $k_\mu k_\nu$ Kerr-Schild term in Eq. (1.1), this result has immediate generalization to the full Kerr geometry. In that case, the spheriodal shape must be interpreted in terms of the Minkowski geometry used as a background geometry. Some general relativistic aspects of this result are given in Sec. 4.

These spheroids also enter in a natural way in treating the light cone with complex vertex. For this reason, we first present the Kerr congruence from the complex point of view to facilitate its later identification as the congruence generated by rotating spheroids.

2. THE COMPLEX POINT OF VIEW

The light cone of (real) Minkowski space, which plays such a dominant role in much of physics, has an interesting generalization to complex Minkowski space.

Complex Minkowski space is a four-dimensional complex manifold endowed with a complex line element and a ten (complex) parameter set of isometries. In natural coordinates (the analytic extension of real Minkowski coordinates) the metric has the form

$$\eta_{\mu\nu} z^\mu z^\nu \equiv \mathbf{z}\cdot\mathbf{z} = \mathbf{x}\cdot\mathbf{x} - \mathbf{y}\cdot\mathbf{y} + 2i\mathbf{x}\cdot\mathbf{y}, \tag{2.1}$$

where $\eta_{\mu\nu} = \text{diag}\,(1, -1, -1, -1)$ and $z^\mu = x^\mu + iy^\mu$. The invariant distance from a fixed point ξ to an arbitrary point $\mathbf{z}$,

$$s^2 = (\mathbf{z} - \xi)\cdot(\mathbf{z} - \xi), \tag{2.2}$$

determines the complex light cone with apex ξ by the condition

$$s^2 = 0. \tag{2.3}$$

(Obviously, in the real case this is the ordinary light cone.) Here we are interested in the intersection of real Minkowski space with the complex cone, i.e., with those $z^\mu = x^\mu$ which satisfy (2.3). We can take, with no loss in generality, the real part of ξ^μ to be zero (by a real Poincaré translation) and the spatial components of the imaginary part of ξ^μ to be along the $x^3 \equiv z$ axis (by a real rotation). ξ^μ then has the form

$$\xi^\mu = i\epsilon\delta^\mu_0 + ia\delta^\mu_3 \tag{2.4}$$

and Eqs. (2.2) and (2.3) become

$$x^2 + y^2 + z^2 - t^2 - a^2 + \epsilon^2 - 2i(za - \epsilon t) = 0,$$

or, since all the variables are real,

$$x^2 + y^2 + z^2 = t^2 + a^2 - \epsilon^2,$$
$$za = \epsilon t. \tag{2.5}$$

Notice that there are no real solutions of (2.5) if $\epsilon^2 > a^2$ and thus the effective range of ϵ is from $-a$ to a. For fixed ϵ (in this range)(2.5) represents an expanding circle which is parallel to the $x-y$ plane. (The two limiting cases $\epsilon = 0$ and $\epsilon = \pm a$ are respectively a circle expanding in the $x-y$ plane and points moving along the positive and negative z axis.)

We thus see that the real manifestation of the complex cone around the fixed point ξ^μ involves only a very limited region of (three-dimensional space. If, however, we "thicken" the complex cone by simultaneously allowing all values of ϵ, we obtain a surface which expands to fill all of space (a hyperboloid in space-time). Explicitly, by eliminating ϵ in (2.5), we obtain the surface

$$(x^2 + y^2)/(t^2 + a^2) + z^2/t^2 = 1, \tag{2.6}$$

which at the limiting value $t = 0$ is a disk, $z = 0$, $x^2 + y^2 \leq a^2$, and which becomes an expanding ellipsoid for increasing t. Eventually, as $t \to \infty$, it approaches a sphere. It is clear that (2.6) represents a time like hypersurface (hyperboloid) in space-time. In the limit $a \to 0$, the hyperboloid becomes the conventional light cone generated by radial light rays.

We now show that the hyperboloid can, in a similar fashion, also be generated by a family of (real) light rays.

The basic idea is to find a congruence of light rays connecting the family of ellipsoids obtained by varying t in Eq. (2.6). For infinitesimal variations, Eq. (2.6) requires that the differentials connecting the ellipsoids satisfy

$$x\,dx + y\,dy + z\,dz(1 + a^2t^{-2}) - z^2a^2t^{-3}\,dt = t\,dt$$

or

$$x\dot{x} + y\dot{y} + z\dot{z}(1 + a^2t^{-2}) - z^2a^2t^{-3} = t, \tag{2.7}$$

wher $(\dot{x}, \dot{y}, \dot{z}) = (dx/dt, dy/dt, dz/dt)$.

The condition that the differentials piece together to form straight lines is $(\ddot{x}, \ddot{y}, \ddot{z}) = 0$. The derivative of Eq. (2.7) then implies

$$\dot{x}^2 + \dot{y}^2 + \dot{z}^2(1 + a^2t^{-2}) - 4z\dot{z}a^2t^{-3} + 3z^2a^2t^{-4} = 1. \tag{2.8}$$

The condition that these straight lines be null is

$$\dot{x}^\mu\dot{x}_\mu = 0, \tag{2.9}$$

where $\dot{x}^\mu = (1, \dot{x}, \dot{y}, \dot{z})$. Equations (2.7)-(2.9) algebraically determine the direction of light rays leaving the ellipsoid corresponding to a given value of t. Their solution is

$$\dot{x}^\mu = \left(1, \frac{-ay \pm tx}{a^2 + t^2}, \frac{ax \pm ty}{a^2 + t^2}, \frac{\pm z}{t}\right), \tag{2.10}$$

with t given by Eq. (2.6). (Due to the reflection symmetry of the ellipsoids there is also another solution $a \to -a$ corresponding to the opposite sense of rotation of the null rays.) All higher derivatives of Eq. (2.10) vanish because of Eq. (2.10) itself so that the light rays so obtained do generate the hyperboloid.

In the limit $a \to 0$, Eq. (2.10) describes the radial null congruence tangent to the light cone $r^2 \equiv x^2 + y^2 + z^2 = t^2$, namely

$$\dot{x}^\mu = (1, x/r, y/r, z/r). \tag{2.11}$$

If the apex of this cone is moved along a straight timelike world line, the vector field (2.11) for each cone gives the Schwarzschild null congruence.

If we move the complex apex (2.4) of the complex cone throughout the complex z° axis, i.e., along the complex world line

$$z^\mu = \phi\delta^\mu_0 + ia\delta^\mu_3, \qquad \phi = \tau + i\epsilon,$$

then the real manifestation of this family of complex cones is the family of hyperboloids

$$x^2 + y^2 + z^2[1 + a^2(t - \tau)^{-2}] = (t - \tau)^2 + a^2,$$

with the associated null vector field (2.10) (with t replaced by $t - \tau$). By comparison with (1.2) and (1.3), we see that (2.10) represents the Kerr congruence when the lower sign is chosen. (The lower sign selects an incoming Kerr congruence and the upper sign an outgoing congruence.)

[This result is a special case of a more general theorem.[3] Just as all (suitably regular) shear and twist-free null geodesic congruences are defined by the light cones from an arbitrary timelike world line, all (suitably regular) shear-free but twisting null congruences are defined by the complex light cones of an arbitrary timelike complex world line, i.e., world line $z^\mu = \xi^\mu(\phi)$ with $\dot{\xi}^\mu\dot{\xi}_\mu = 1$.]

3. THE ROTATING POINT OF VIEW

We now change our point of view and consider the world tube traced out in Minkowski space by the rigid rotation of an axisymmetric 2-surface about its symmetry axis. The world tube is generated by straight lines tangent to the constant timelike unit vector field T^α which pick out the inertial frame in which there is no translational motion. In this frame, let X^α, Y^α, Z^α, T^α be the orthonormal tetrad (signature $---+$) parallel to the x, y, z, t axes with Z^α corresponding to the axis of revolution. That $t = \text{const}$ cross sections of the world tube then have unit normals N^α and unit curvature eigenvectors Φ^α (tangent to the parallels) and Θ^α (tangent to the meridians) which satisfy

$$\Phi^\alpha = (xY^\alpha - yX^\alpha)\rho^{-1},$$
$$\Theta^\alpha = [Z^\alpha + N^\alpha(Z^\beta N_\beta)][1 - (Z^\beta N_\beta)^2]^{-1/2},$$

where $\rho = (x^2 + y^2)^{1/2}$ and

$$\Theta^\alpha T_\alpha = \Phi^\alpha T_\alpha = N^\alpha T_\alpha = N^\alpha\Theta_\alpha = N^\alpha\Phi_\alpha = 0.$$

Here X^α, Y^α, Z^α, and T^α are translational Killing vectors and $\rho\Phi^\alpha$ is a rotational Killing vector. The curvature eigenvalues[4] are

$$k_1 = N_{\alpha;\beta}\Theta^\alpha\Theta^\beta \tag{3.1a}$$

and

$$k_2 = N_{\alpha;\beta}\Phi^\alpha\Phi^\beta. \tag{3.1b}$$

In the course of time, points on the 2-surface trace out world lines with unit four-velocity

$$u^\alpha = T^\alpha\cosh\Lambda + \Phi^\alpha\sinh\Lambda, \tag{3.2}$$

where Λ is determined by the angular velocity ω

$$\cosh\Lambda = (1 - \rho^2\omega^2)^{-1/2}.$$

The local rest frame determined by u^α picks out at each point of the world tube a spatial tangent plane specified by a complex null vector m^α satisfying

$$m^\alpha m_\alpha = m_\alpha u^\alpha = m_\alpha N^\alpha = m_\alpha\bar{m}^\alpha + 1 = 0. \tag{3.3}$$

We choose

$$\sqrt{2}m^\alpha = \Theta^\alpha + i(\Phi^\alpha\cosh\Lambda + T^\alpha\sinh\Lambda). \tag{3.4}$$

We say that a null ray is "kinematically normal" to the 2-surface if its tangent vector l^μ satisfies

$$l^\mu m_\mu = l^\mu l_\mu = 0.$$

The null direction l^μ corresponds to the collimated beam from a small flashlight oriented at each instant ($t = \text{const}$) in the N^α direction but moving with four-velocity u^α. The l^μ direction is not in general orthogonal to the $t = \text{const}$ cross sections of the world tube as is the usual geome trically normal null direction. Analytically we have

$$l^\mu = u^\alpha + N^\alpha. \tag{3.5}$$

On the world tube, the shear of the kinematically normal null rays is

$$\sigma = l_{\alpha;\beta} m^\alpha m^\beta.$$

Because of Eq. (3.5) and the fact that u is constructed from Killing vectors, this reduces to

$$\sigma = N_{\alpha;\beta} m^\alpha m^\beta.$$

If the surface were not rotating, this would lead to the usual relation between the shear and the difference between the two curvature eigenvalues.[5] In the present case, Eqs. (3.1)-(3.4) lead to

$$2\sigma = k_1 - k_2(1 - \rho^2\omega^2)^{-1}. \tag{3.6}$$

If we analytically describe the rotating 2-surface by the revolution of the curve $z = f(\rho)$, then the condition of zero shear obtained from Eq. (3.6) takes the form

$$\frac{f''}{f'(1 + f'^2)} = \frac{1}{\rho(1 - \omega^2\rho^2)}.$$

The most general solution to this equation is a portion of a conic section. Furthermore, the most general solution which generates a smooth compact surface of revolution is an ellipse with equatorial radius A and polar radius B connected by

$$B = A(1 - \omega^2 A^2)^{1/2}. \tag{3.7}$$

Hence, the most general rigidly rotating 2-surface whose kinematically normal null directions are shear-free is an oblate spheroid (ellipsoid of revolution) satisfying Eq. (3.7)

The shear-free null directions leaving the surface of the spheroid remain shear-free when geodetically extended to straight null rays. (This follows from the propagation equations [6,7] for the optical scalars.) It is straightforward to compare the results of this section with those of Sec. 2 to verify that the null congruence l^α [Eq. (3.5)] obtained from a rotating spheroid coincides with $\dot{x}^\mu$ [Eq. (2.11)] and is thus a Kerr congruence. In particular, by referring to Eq. (2.6) the Kerr parameter a can be related to the equatorial radius and angular velocity of the rotating spheroid by

$$a = \omega A^2. \tag{3.8}$$

Since each Kerr congruence corresponds to a fixed value of a, the family of rotating spheroids determined by the full congruence has different angular velocities for different equatorial radii as determined by Eq. (3.8).

4. CONCLUSION

The relationship which we have obtained between the Kerr congruence and rotating spheroids (Kerr spheroids) suggests two possible areas of application.

First, there is the relevance to the general theory of twisting, shear-free congruences which play a basic role in twistor theory, which in turn is the theory of momentum-angular momentum complexes. The Kerr congruence, which we have shown can be represented by unaccelerated, rotating Kerr spheroids, also corresponds to a special analytic function in twistor space. Can more general twistor objects be given physical representation in terms of accelerating and rotating bodies?

Second, the oblateness of the Kerr spheroids suggests a possible connection with the Maclaurin spheroids, which describe the axisymmetric equilibrium configurations of rotating fluids according to Newtonian gravitational theory. However, the oblateness of the Kerr spheroids is due to an aberration effect which is significant only at relativistic velocities. Consequently, there can be no connection with the Maclaurin spheroids except for ultra compact bodies whose rotational deformations are small at nonrelativistic velocities. The quantitative condition for agreement between the Kerr and Maclaurin spheroids follows from the relativistic generalization of the Newtonian equilibrium condition

$$\tfrac{1}{2}\omega^2\rho^2 - V = \text{const on boundary},$$

where V is the Newtonian potential. The relativistic generalization, derived by Boyer,[9] may be put in the form

$$g_{\mu\nu}\xi^\mu\xi^\nu = \text{const on boundary}, \tag{4.1}$$

where ξ^μ is the Killing vector tangent to the hydrodynamical streamlines on the fluid boundary. The requirement that a Kerr spheroid satisfy Eq. (4.1) leads directly to the boundary condition

$$g_{\mu\nu}\xi^\mu\xi^\nu = 0.$$

Hence the hydrodynamical streamlines would necessarily be lightlike on the boundary. The boundary would correspond to the outer Kerr horizon ($r = r_+$). It is unlikely that any physically realistic body could be hydrostatically supported under those conditions.

The curved space generalization of our results does furnish a convenient characterization of the two principle null vectors of the Kerr geometry analogous to (3.5),

$$l^\alpha_\pm = u^\alpha \pm N^\alpha.$$

Here u^α is the unit vector parallel to $T^\alpha + \omega\Phi^\alpha$, where T^α is the time-translational Killing vector, Φ^α is the rotational Killing vector, the angular velocity ω is given by

$$\omega = a/(r^2 + a^2),$$

and the streamlines of u^α lie in the world tubes $r = \text{const}$. having unit normals N^α.

*Supported by the National Science Foundation Grant No. GP-35773.

[1]R. P. Kerr, Phys. Rev. Lett. 11, 237 (1963).
[2]We use the conventions of S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973), p. 161.
[3]R. Lind and E. T. Newman, in preparation.
[4]N. J. Hicks, *Notes on Differential Geometry* (Van Nostrand, New York, 1965).
[5]R. Kantowski, J. Math. Phys. 9, 336 (1968).
[6]R. K. Sachs, Proc. Roy. Soc. (London) **A264**, 309 (1961).
[7]E. T. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962).
[8]R. Penrose, J. Math. Phys. 8, 345 (1967).
[9]R. H. Boyer, Proc. Camb. Phil. Soc. 61, 527 (1965).

Iterative solution of the inverse Sturm–Liouville problem

Victor Barcilon

Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois 60637
(Received 4 September 1973)

An iterative construction of the potential function entering in a regular Sturm-Liouville problem is discussed. The given data is in the form of two spectra associated with distinct boundary conditions at one end point.

1. INTRODUCTION

An inverse problem consists in inferring some properties of a given problem from a complete or partial knowledge of its solution. This definition is rather vague; nevertheless, it points out that the difference between a "direct problem" and an "inverse problem" is, to some extent, arbitrary and its rationale is to be found in the historical development of that particular problem.

Given the canonical Sturm–Liouville problem

$$u'' + \{\lambda - q(x)\}u = 0, \quad x \in (0, 1), \tag{1.1}$$

$$u(0) = u(1) = 0, \tag{1.2}$$

the direct problem consists in finding the eigensolutions $\{\lambda_n, u_n(x)\}_1^\infty$ for a given "potential" $q(x)$. On the other hand, there are many instances of vibrating systems in physics, geophysics, and engineering which reduce to a Sturm–Liouville problem for which the eigenvalues (or normal frequencies) can be measured directly (at least insofar as the lowest ones are concerned). The real problem then consists in deducing $q(x)$. This problem could be referred to as the inverse problem. However, in 1946, Borg[1] has shown that the knowledge of a single spectrum $\{\lambda_n\}_1^\infty$ is *not* sufficient to determine $q(x)$.

Since then, two not altogether equivalent inverse Sturm–Liouville problems have been considered in the literature. In both versions, $q(x)$ is thought out, but these two versions differ insofar as their starting point is concerned.

One approach, which has become mostly associated with the names of Gel'fand and Levitan,[2] uses a spectral function $\rho(\lambda)$ as a starting point. If $u(x, \lambda)$ stands for the solution of (1.1) which satisfies the following conditions:

$$u(0, \lambda) = 0, \tag{1.3}$$

$$u'(0, \lambda) = 1, \tag{1.4}$$

and if we define

$$F(\lambda) = \int_0^1 f(x)u(x, \lambda)dx, \tag{1.5}$$

where $f(x)$ is an arbitrary real $L_2(0, 1)$ function, then the spectral function can be defined by writing Parseval's theorem as follows:

$$\int_0^1 f^2(x)dx = \int_{-\infty}^{\infty} F^2(\lambda)d\rho(\lambda). \tag{1.6}$$

Clearly,

$$\rho(\lambda) = \sum_{\lambda_n < \lambda} \left(\int_0^1 u^2(x, \lambda_n)dx\right)^{-1}, \tag{1.7}$$

and insofar as the regular Sturm–Liouville problem is concerned, the knowledge of $\rho(\lambda)$ is equivalent to the knowledge of the eigenvalues $\{\lambda_n\}_1^\infty$ *and* $\{\int_0^1 u_n^2(x)dx\}_1^\infty$, where the eigenfunctions $\{u_n(x)\}_1^\infty$ have been normalized by means of (1.4).

The other approach to the inverse Sturm–Liouville problem consists in using two spectra as the starting point. In other words, in addition to $\{\lambda_n\}_1^\infty$, which is associated with (1.1) subject to (1.2), we assume that another spectrum, say $\{\mu_n\}_1^\infty$, associated with a different pair of boundary conditions, such as

$$u'(0) = u(1) = 0, \tag{1.8}$$

is also known. Indeed, Borg[1] Marčenko[3] and Levinson[4] have proved that two spectra such as $\{\lambda_n\}_1^\infty$ and $\{\mu_n\}_1^\infty$ are necessary and sufficient to determine $q(x)$. This approach might seem more artificial than the one based on the spectral function, since it required some knowledge about two different eigenvalue problems. However, this is subjective to a certain extent and there are various instances in which the two spectra [and in particular those associated with the canonical boundary conditions (1.2) and (1.8)] arise quite naturally (cf., e.g., Ref. 5). Furthermore, given two spectra, one can easily evaluate $\{\int_0^1 u^2(x, \lambda_n)dx\}_1^\infty$ and cast the problem in terms of the spectral function.[6] Be that as it may, we shall adopt the second approach and assume that we are given two spectra.

The problem of investigating the existence of a function $q(x)$ for a given pair of spectra $\{\lambda_n\}_1^\infty$ and $\{\mu_n\}_1^\infty$ was taken up by Krein.[7,8] His approach, which relies on the theory of continuous fractions and on positive functionals, is extremely ingenious. Given two sequences $\{\lambda_n\}_1^\infty$ and $\{\mu_n\}_1^\infty$ such that

$$\lim_{n\to\infty} (\mu_n/n^2\pi^2) = \lim_{n\to\infty} \lambda_n/n^2\pi^2 = 1 \tag{1.9}$$

and

$$0 < \mu_1 < \lambda_1 < \mu_2 < \cdots, \tag{1.10}$$

Krein first constructs the meromorphic function

$$\Gamma(\lambda) = \prod_{i=1}^{\infty} \frac{(1 - \lambda/\lambda_i)}{(1 - \lambda/\mu_i)} \tag{1.11}$$

and then obtains a partial fraction expansion of $\Gamma(\lambda)$, viz.,

$$\Gamma(\lambda) = \gamma_0 + \sum_1^\infty \frac{m_i}{\mu_i - \lambda}. \tag{1.12}$$

Because of the interlacing of the eigenvalues (1.10), γ_0 and m_i are positive. The related function

$$F(t) = \sum_1^\infty \frac{m_i}{\mu_i} \cos\sqrt{\mu_i}t, \tag{1.13}$$

which is a Hermitian-positive function, is then used to compute the largest value $M(x)$ of r for which $F(t) - r$ is Hermitian-positive over the interval $(0, x)$, where $0 \le x \le 1$. This particular step requires the solution of a Fredholm integral equation with kernel $F(t - s)$. Finally, $q(x)$ is obtained by means of a simple expression containing derivatives of the function $M(2x)$.

Partly because the proof of Krein's existence theorem

does not lend itself to computational purposes, and partly with an eye toward higher-dimensional inverse eigenvalue problems to which Krein's method is not likely to be generalized, I decided to reexamine the actual construction of $q(x)$ from the spectra $\{\lambda_n\}_1^\infty$ and $\{\mu_n\}_1^\infty$. The results presented here are in many ways not completely satisfactory since they are local and apply only for small $\|q\|$. Thus, in some sense, a constructive proof of the existence theorem which can be used for actual computations has yet to be found. In defense of the present analysis, I should say that the approach to be presented is based on an iteration technique which could perhaps be extended to more general $q(x)$ as well as to higher-dimensional problems.

In view of the limited scope of the actual results and for the sake of exposition, I decided to cast the analysis within the framework of the simplest inverse Sturm–Liouville problem, viz., for the canonical boundary conditions (1.1)–(1.8) or, as we shall see, for the so-called symmetric case. Most of these results can be generalized to other boundary conditions, but, as previously mentioned, generalizations in other directions are undoubtedly more relevant.

2. THE CANONICAL INVERSE STURM-LIOUVILLE PROBLEM

We shall be concerned with the following Sturm–Liouville problem:

$$y'' + \{\sigma - q(x)\}y = 0, \quad x \in (-1, 1), \qquad y(\pm 1) = 0 \tag{2.1}$$

for symmetric "potentials" $q(x)$, i.e.,

$$q(-x) = q(x). \tag{2.2}$$

Furthermore, we shall assume throughout our analysis that $q(x)$ has been normalized so that

$$\int_{-1}^{1} q(x)dx = 0. \tag{2.3}$$

The spectrum $\{\sigma_n\}_1^\infty$ determines the symmetric potential $q(x)$ uniquely, and furthermore $q(x)$ exists if the spectrum is such that

$$\lim_{n\to\infty} (\sigma_n/\tfrac{1}{4}n^2\pi^2) = 1. \tag{2.4}$$

Indeed, the above eigenvalue problem is "equivalent" to the following pair of Sturm-Liouville problems:

$$u'' + \{\lambda - q(x)\}u = 0, \quad x \in (0, 1), \qquad u(0) = u(1) = 0, \tag{2.5}$$

and

$$v'' + \{\mu - q(x)\}v = 0, \quad x \in (0, 1), \qquad v'(0) = v(1) = 0, \tag{2.6}$$

and it is quite easy to show that

$$y_{2n}(x) = u_n(x), \quad \sigma_{2n} = \lambda_n, \qquad y_{2n-1}(x) = v_n(x), \quad \sigma_{2n-1} = \mu_n, \qquad n = 1, 2, \cdots. \tag{2.7}$$

Therefore, the determination of $q(x)$ from two spectra $\{\lambda_n\}_1^\infty, \{\mu_n\}_1^\infty$ which are interlaced, viz.,

$$\mu_1 < \lambda_1 < \mu_2 < \lambda_2 < \cdots, \tag{2.8}$$

and which correspond to the eigenvalue problems (2.5), (2.6) is equivalent to the determination of the symmetric $q(x)$ from a knowledge of the spectrum $\{\sigma_n\}_1^\infty$ of (2.1).

Throughout part I (Sec. 3) of the paper, we shall concentrate on the pair of eigenvalue problems (2.5), (2.6). However, in part II (Sec. 4) we shall adopt the point of view that $q(x)$ is symmetric and deal with (2.1).

This changeover will serve to emphasize the separation between part I and II of the paper: For instance, throughout part I, we shall assume that $q \in C^1(0, 1)$ whereas in part II we shall relax this assumption and merely assume that $q(x) \in L_2(-1, 1)$. Although none of the results of part I will be used in part II, the motivation for the iteration technique presented in part II is to be found in part I.

3. PART I

A. A single equation for u^2 and v^2

Throughout this part, we shall focus our attention on eigenvalue problems (2.5) and (2.6) and assume that $q \in C^1(0, 1)$. As a result, both $u(x)$ and $v(x)$ are in $C^3(0, 1)$.

Let us introduce the following notation

$$w = u^2. \tag{3.1}$$

Then

$$w' = 2uu', \tag{3.2}$$

$$w'' = 2u'^2 + 2uu'' = 2u'^2 + 2(q - \lambda)w. \tag{3.3}$$

Multiplying (3.3) by $2w$, we get

$$2ww'' = (2uu')^2 + 4(q - \lambda)w^2,$$

i.e.,

$$ww'' - \tfrac{1}{2}w'^2 - 2qw^2 = -2\lambda w^2. \tag{3.4}$$

Since no mention has been made of boundary conditions, it is clear that v^2 also satisfies (3.4). Therefore, let us modify our original definition (3.1) and define w as a solution of the following nonlinear equation:

$$ww'' - \tfrac{1}{2}w'^2 - 2qw^2 = -\tfrac{1}{2}\nu w^2. \tag{3.5}$$

To complete the specification of w, let us require that

$$w'(0) = w(1) = 0. \tag{3.6}$$

It is obvious from our construction that $\{u_n^2, 4\lambda_n\}$ and $\{v_n^2, 4\mu_n\}$ are eigensolutions of (3.5) and (3.6). We shall presently show that (3.5) and (3.6) admit no other eigensolutions. To that effect let us prove

Lemma 1: The eigenfunctions $w_n(x)$ are nonnegative.

Proof: If $w_n(x)$ is never zero in (0.1), then without loss of generality we can consider that $w_n(x)$ is nonnegative and the lemma is proved.

Let us therefore assume that

$$w_n(x_0) = 0, \quad x_0 \in (0, 1). \tag{3.7}$$

From (3.5) we see that this implies that

$$w_n'(x_0) = 0. \tag{3.8}$$

Thus, in the neighborhood of x_0, we can write

$$w_n(x) = (x - x_0)^{2+\delta}h(x) \tag{3.9}$$

where $h(x_0) \neq 0$. Substituting (3.9) in (3.5), we see that

$$(2 + \delta)(1 + \delta) - \tfrac{1}{2}(2 + \delta)^2 = 0, \tag{3.10}$$

which implies that $\delta = 0$ and hence that all zeros of $w_n(x)$ are double zeros. As a result, without loss of generality we can consider that $w_n(x) \geq 0$.

We can now state

Theorem 1: $\{w_n, \nu_n\}_1^\infty = \{u_n^2, 4\lambda_n\}_1^\infty \cup \{v_n^2, 4\mu_n\}_1^\infty$.

Proof: We have already seen that $\{u_n^2, 4\lambda_n\}_1^\infty$ and $\{v_n^2, 4\mu_n\}_1^\infty$ are eigensolutions of (3.5)–(3.6). Conversely, we shall show that any eigensolution of (3.5)–(3.6) belongs to $\{u_n^2, 4\lambda_n\}_1^\infty \cup \{v_n^2, 4\mu_n\}_1^\infty$.

Since w is nonnegative and has double zeros, we can write

$$w = \varphi^2, \tag{3.11}$$

where φ is in $C^2(0, 1)$. Substituting (3.11) in (3.5), (3.6), we get

$$\begin{aligned} &\varphi^3[\varphi'' + (\tfrac{1}{4}\nu - q(x))\varphi] = 0, \\ &\varphi(0)\varphi'(0) = \varphi(1) = 0. \end{aligned} \tag{3.12}$$

Clearly, either $\varphi = u_n(x)$ and $\nu/4 = \lambda_n$ or $\varphi = v_n(x)$ and $\nu/4 = \mu_n$.

From the preceding discussion, we can write

$$\begin{aligned} &w_{2n}(x) = u_n^2(x), \quad \nu_{2n} = 4\lambda_n \\ &w_{2n-1}(x) = v_n^2(x), \quad \nu_{2n-1} = 4\mu_n \end{aligned} \quad (n = 1, 2, \ldots). \tag{3.13}$$

Thus (3.5)–(3.6) affords us a way of coupling the pair of eigenvalue problems (2.5) and (2.6). That $\{u_n^2\}$ and $\{v_n^2\}$ enter in the solution of the inverse problem can be seen by means of the following heuristic argument. If the potential $q(x)$ were to be changed, say to $q(x) + \delta q(x)$, then, to first order, the nth eigenvalue of (2.5) would change by $\delta\lambda_n$ where

$$\delta\lambda_n = \int_0^1 \delta q u_n^2 dx / \int_0^1 u_n^2 dx. \tag{3.14}$$

Now, the set $\{u_n^2\}_1^\infty$ does not form a basis; this can easily be checked for the simple case $q \equiv 0$. Therefore, we can find a non-zero δq which is orthogonal to all the $\{u_n^2\}_1^\infty$. Thus, two different potentials would correspond to the same eigenvalues. Conversely, if $\{\lambda_n\}$ and $\{\mu_n\}$ are necessary and sufficient to determine $q(x)$ uniquely, then this must imply that $\{u_n^2\}_1^\infty$ together with $\{v_n^2\}_1^\infty$ must constitute a basis for the functions $q(x)$ and $\{\lambda_n\} \cup \{\mu_n\}$ must be related to $\{\int_0^1 q w_n(x) dx\}_1^\infty$.

B. A linear equation for w

In order to investigate the completeness of $\{w_n(x)\}_1^\infty$, it would be much preferable to deal with a linear equation. Such an equation can be obtained by differentiating (3.5):

$$w''' - 4qw' - 2q'w = -\nu w'. \tag{3.15}$$

It is at this stage that the differentiability of $q(x)$ is used. We must now supplement the boundary conditions (3.6) by an additional one. In view of (3.7) and (3.8), it is clear that the new required boundary conditon should be $w'(1) = 0$, and so

$$w'(0) = w(1) = w'(1) = 0. \tag{3.16}$$

It will be convenient to introduce the following three fundamental solutions $W_1(x, s)$, and $W_2(x, s)$, and $W_3(x, s)$, where

$$\mathfrak{L} W_j + s^2 W_j' = 0 \quad (j = 1, 2, 3), \tag{3.17}$$

$$\mathfrak{L} = \frac{d^3}{dx^3} - 4q\frac{d}{dx} - 2q' \tag{3.18}$$

and

$$\begin{aligned} &W_1(0, s) = 1, \quad W_1'(0, s) = 0, \quad W_1''(0, s) = 0, \\ &W_2(0, s) = 0, \quad W_2'(0, s) = 0, \quad W_2''(0, s) = 1, \\ &W_3(1, s) = 0, \quad W_3'(1, s) = 0, \quad W_3''(1, s) = 1. \end{aligned} \tag{3.19}$$

All of the properties of $W_j(x, s)$ which we shall require are derived by classical methods[9]; we shall therefore simply state the various results which are needed.

$W_j(x, s)$ are entire functions of s of order 1. That $W_j(x, s)$ is an entire function of s can be seen by transforming (3.17)–(3.19) into a Volterra equation, viz.,

$$W_j(x, s) = f_j(x) + s\int_{a_j}^x K_j(x, t) W_j(t, s) dt \tag{3.20}$$

and making use of standard results in the theory of linear Volterra equations. The statement regarding the order follows from an investigation of the asymptotic behavior of $W_j(x, s)$ for $|s| \to \infty$.

Lemma 2: As $|s| \to \infty$,

$$\begin{aligned} W_1(x, s) &= 1 + O(e^{|s_i|x}/|s|^3), \\ W_2(x, s) &= (1 - \cos sx)/s^2 + O(e^{|s_i|x}/|s|^3), \\ W_3(x, s) &= \{1 - \cos[s(x-1)]\}/s^2 + O(e^{|s_i|(1-x)}/|s|^3), \end{aligned} \tag{3.21}$$

$$\begin{aligned} W_1'(x, s) &= O(e^{|s_i|x}/|s|^2), \\ W_2'(x, s) &= \frac{\sin sx}{s} + O(e^{|s_i|x}/|s|^2), \\ W_3'(x, s) &= \frac{\sin s(x-1)}{s} + O(e^{|s_i|(1-x)}/|s|^2), \end{aligned} \tag{3.22}$$

$$\begin{aligned} W_1''(x, s) &= O(e^{|s_i|x}/|s|), \\ W_2''(x, s) &= \cos sx + O(e^{|s_i|x}/|s|), \\ W_3''(x, s) &= \cos(x-1) + O(e^{|s_i|(1-x)}/|s|), \end{aligned} \tag{3.23}$$

where $s_i = \operatorname{Im} s$.

Proof: The above formulas are deduced by first deriving the following integral equations for W_j, viz.,

$$W_1 = 1 + \frac{1}{s^2}\int_0^x (4qW_1' + 2q'W_1)[1 - \cos s(x - \xi)] d\xi, \tag{3.24}$$

$$W_2 = \frac{1 - \cos sx}{s^2} + \frac{1}{s^2}\int_0^x (4qW_2' + 2q'W_2) \times [1 - \cos s(x - \xi)] d\xi, \tag{3.25}$$

$$W_3 = \frac{1 - \cos s(x-1)}{s^2} + \frac{1}{s^2}\int_1^x (4qW_3' + 2q'W_3) \times [1 - \cos s(x - \xi)] d\xi, \tag{3.26}$$

and then by applying the same arguments as those used for the second order equations (e.g. Ref. 9, p. 9).

We shall also need to introduce the Wronskian of W_1, W_2, W_3

$$D(s) = \begin{vmatrix} W_1 & W_2 & W_3 \\ W_1' & W_2' & W_3' \\ W_1'' & W_2'' & W_3'' \end{vmatrix} . \tag{3.27}$$

Since $D(s)$ is not a function of x, we can evaluate $D(s)$ by either setting $x = 0$ or $x = 1$ in (3.27). As a result, we see that

$$D(s) = -W_3'(0, s) = W_1(1, s)W_2'(1, s) - W_1'(1, s)W_2(1, s) \tag{3.28}$$

and

$$D(s) = (\sin s)/s + O(e^{|s_i|}/|s|^2). \tag{3.29}$$

If $s = \sqrt{\nu_n}$, $W_3(x, \nu_n^{1/2})$ is an eigenfunction of the original eigenvalue problem and the Wronskian vanishes: Thus, the zeros of the Wronskian are $\{\nu_n^{1/2}\}_1^\infty$. We shall write

$$w_n(x) = W_3(x, \nu_n^{1/2}). \tag{3.30}$$

Finally, the vanishing of the Wronskian for $s = \nu_n^{1/2}$ implies that W_1, W_2, W_3 are linearly dependent, viz.,

$$W_3(x, \nu_n^{1/2}) = A_n W_1(x, \nu_n^{1/2}) + B_n W_2(x, \nu_n^{1/2}). \tag{3.31}$$

C. The adjoint equation

Since (3.15)-(3.16) is not a self-adjoint eigenvalue problem, we shall need to introduce its adjoint, namely,

$$\omega''' - 4q\omega' - 2q'\omega = -\nu\omega', \tag{3.32}$$

$$\omega(0) = \omega''(0) = \omega(1) = 0. \tag{3.33}$$

We recall that the eigenvalues of (3.32)-(3.33) coincide with those of (3.15)-(3.16). It is a simple matter to deduce the following orthogonality relations:

$$\int_0^1 W_n(x)\omega_m'(x)dx = \int_0^1 W_n'(x)\omega_m(x)dx = 0 \quad \text{if } n \neq m. \tag{3.34}$$

Let us also introduce three functions $\Omega_1(x, s)$, $\Omega_2(x, s)$, and $\Omega_3(x, s)$ as follows:

$$\Omega_k(x, s) = \epsilon_{klm} W_l(x, s) W_m'(x, s), \tag{3.35}$$

where ϵ_{klm} is the standard isotropic Cartesian tensor of third order. Differentiating (3.35), we deduce that

$$\Omega_k' = \epsilon_{klm} W_l W_m'', \tag{3.36}$$

$$\Omega_k'' = (4q - s^2)\Omega_k + \epsilon_{klm} W_l' W_m'', \tag{3.37}$$

$$\Omega_k''' = 4q\Omega_k' + 2q'\Omega_k - s^2\Omega_k'. \tag{3.38}$$

Thus

$$\mathcal{L}\Omega_k = -s^2\Omega_k. \tag{3.38'}$$

Furthermore, we should note that

$$\Omega_3(0, s) = 0, \quad \Omega_3'(0, s) = 1, \quad \Omega_3''(0, s) = 0,$$
$$\Omega_3(1, s) = W_1(1, s)W_2'(1, s) - W_1'(1, s)W_2(1, s). \tag{3.39}$$

In view of (3.28), $\Omega_3(x, \nu_n^{1/2})$ is an eigenfunction of the adjoint problem and we shall write

$$\omega_n(x) = \Omega_3(x, \nu_n^{1/2}). \tag{3.40}$$

Finally, for $s = \nu_n^{1/2}$, Ω_1, Ω_2, and Ω_3 are linearly dependent and by means of (3.31), we can show that

$$\Omega_1(x, \nu_n^{1/2}) = -A_n\omega_n(x),$$
$$\Omega_2(x, \nu_n^{1/2}) = -B_n\omega_n(x). \tag{3.41}$$

D. The expansion theorem

We shall investigate the completeness of $\{\omega_n(x)\}_1^\infty$ via the Cauchy integral method. Once again, the method is well known and we shall gloss over most of the details confining our presentation to the key steps.

We focus our attention on the integral

$$I_n = \frac{1}{\pi i} \oint_{C_n} sQ(x, s)\, ds \tag{3.42}$$

where C_n is a circle of radius r_n such that

$$\nu_n^{1/2} < r_n < \nu_{n+1}^{1/2}, \tag{3.43}$$

and the integrand $Q(x, s)$ is defined as follows:

$$Q(x, s) = \frac{\Omega_1'(x, s)}{D(s)} \int_0^x f(\xi)W_1(\xi, s)\, d\xi + \frac{\Omega_2'(x, s)}{D(s)} \int_0^x f(\xi)W_2(\xi, s)\, d\xi - \frac{\Omega_3'(x, s)}{D(s)} \int_x^1 f(\xi)W_3(\xi, s)\, d\xi. \tag{3.44}$$

For our purposes it will be sufficient to assume that $f(x)$ is a differentiable function and

$$\int_0^1 f(x)\, dx = 0 \tag{3.45}$$

Since $Q(x, s)$ is a meromorphic function of s, with simple poles at $s = \nu_m^{1/2}$ $(m = 1, 2, \cdots)$, the computation of I_n can be carried out by means of the calculus of residues, viz.,

$$I_n(x) = \sum_{k=1}^{n} \frac{2\nu_k^{1/2}}{D'(\nu_k^{1/2})} \left\{ \Omega_1'(x, \nu_k^{1/2}) \int_0^x f(\xi)W_1(\xi, \nu_k^{1/2})\, d\xi + \Omega_2'(x, \nu_k^{1/2}) \int_0^x f(\xi)W_2(\xi, \nu_k^{1/2})\, d\xi - \Omega_3'(x, \nu_k^{1/2}) \int_x^1 f(\xi)W_3(\xi, \nu_k^{1/2})\, d\xi \right\} \tag{3.46}$$

or, in view of (3.31) and (3.41),

$$I_n(x) = -\sum_{k=1}^{n} 2\nu_k^{1/2} \frac{\int_0^1 f(\xi)w_k(\xi)\, d\xi}{D'(\nu_k^{1/2})} \omega_k'(x). \tag{3.47}$$

If we define $J_n(x)$ as follows

$$J_n(x) = \sum_{k=1}^{n} f_k \cos k\pi x, \tag{3.48}$$

where

$$f_k = 2\int_0^1 f(x)\cos k\pi x dx,$$

then we can prove (cf., e.g., Ref. 10, p. 303) that $I_n - J_n$ tends to zero as $n \to \infty$ uniformly on $0 \leq x \leq 1$. Since f is continuous, we therefore conclude that

$$f(x) = -\sum_{n=1}^{\infty} \frac{2\nu_n^{1/2} \int_0^1 f(\xi)w_n(\xi)\, d\xi}{D'(\nu_n^{1/2})} \omega_n'(x). \tag{3.49}$$

The above series expansion can be written slightly differently. To that effect we start from the following relation:

$$\int_0^1 \Omega_3(x,s)\mathcal{L}w_n(x)\,dx + \int_0^1 w_n(x)\mathcal{L}\,\Omega_3(x,s)\,dx = \Omega_3 w_n'' - \Omega_3' w_n' + \Omega_3'' w_n\Big|_0^1 \quad (3.50)$$

Using the boundary values for $\Omega_3(x,s)$ and $w_n(x)$, we can see that

$$\int_0^1 \Omega_3(x,s)\mathcal{L}w_n(x)\,dx + \int_0^1 w_n(x)\mathcal{L}\,\Omega_3(x,s)\,dx = D(s), \quad (3.51)$$

or better still,

$$-\nu_n \int_0^1 \Omega_3(x,s)w_n'(x)\,dx - s^2\int_0^1 w_n(x)\Omega_3'(x,s)dx = D(s). \quad (3.52)$$

Therefore,

$$\int_0^1 w_n(x)\Omega_3'(x,s)dx = -D(s)/(s^2-\nu_n), \quad (3.53)$$

and, as $s \to \nu_n^{1/2}$, we see that

$$\int_0^1 w_n(x)\omega_n'(x)dx = -D'(\nu_n^{1/2})/2\nu_n^{1/2}. \quad (3.54)$$

As a result we can state

Theorem 2: If $f(x) \in C^1(0,1)$ and $\int_0^1 f(x)dx = 0$, then

$$f(x) = \sum_{n=1}^{\infty} \left(\int_0^1 f(\xi)w_n(\xi)d\xi \Big/ \int_0^1 \omega_n'(\xi)w_n(\xi)d\xi\right)\omega_n'(x). \quad (3.55)$$

E. An iteration formula

Let us multiply (2.5) and (2.6) by $u_n(x)$ and $v_n(x)$ respectively, and integrate the resulting expressions from 0 to 1. We get

$$\begin{aligned}\lambda_n \int_0^1 u_n^2(x)dx &= \int_0^1 u_n'^2 dx + \int_0^1 q(x)u_n^2(x)dx,\\ \mu_n \int_0^1 v_n^2(x)dx &= \int_0^1 v_n'^2 dx + \int_0^1 q(x)v_n^2(x)dx.\end{aligned} \quad (3.56)$$

Using (3.13), we can combine these two expressions into a single one:

$$\tfrac{1}{4}\nu_n \int_0^1 w_n(x)dx = \tfrac{1}{4}\int_0^1 \frac{w_n'^2(x)}{w_n(x)}\,dx + \int_0^1 q w_n(x)dx. \quad (3.57)$$

Now, let us assume that a sequence of $\{\hat\nu_n\}_1^\infty$ is given and that we want to find the potential $\hat q(x)$ associated with this double spectra. Let us also assume that the] sequence $\{\hat\nu_n\}_1^\infty$ is such as to guarantee that $\hat q(x)$ is differentiable.

If $q^{(0)}(x)$ is an initial guess, then by solving an eigenvalue problem like (3.15), (3.16) we can find the corresponding eigensolutions $\{\nu_n^{(0)}, w_n^{(0)}(x)\}_1^\infty$. Then, we can use (3.58) to get a new, hopefully better, approximation to $\hat q(x)$. This can be accomplished by defining $q^{(1)}(x)$ as follows:

$$\int_0^1 q^{(1)}(x)w_n^{(0)}(x)dx = \tfrac{1}{4}\hat\nu_n \int_0^1 w_n^{(0)}(x)dx - \tfrac{1}{4}\int_0^1 \frac{[w_n^{(0)\prime}(x)]^2}{w_n^{(0)}(x)}\,dx \quad (3.58)$$

or, alternatively,

$$\int_0^1 q^{(1)}(x)w_n^{(0)}(x)dx = \tfrac{1}{4}(\hat\nu_n - \nu_n^{(0)})\int_0^1 w_n^{(0)}(x)dx + \int_0^1 q^{(0)}(x)w_n^{(0)}(x)dx. \quad (3.58')$$

In view of the expansion theorem and the fact that $\hat q$, $q^{(0)}$, $q^{(1)}$ have no mean, we can rewrite (3.58′) as follows:

$$q^{(1)}(x) = \tfrac{1}{4}\sum_{n=1}^{\infty}(\hat\nu_n - \nu_n^{(0)})\frac{\int_0^1 w_n^{(0)}(\xi)d\xi}{\int_0^1 w_n^{(0)}(\xi)\omega_n^{(0)\prime}(\xi)d\xi}\,\omega_n^{(0)\prime}(x) + q^{(0)}(x). \quad (3.59)$$

The cycle is completed when $q^{(1)}$ is substituted in (3.15) which can then be solved for $\{\nu_n^{(1)}, w_n^{(1)}(x)\}_1^\infty$.

In summary, given two appropriate spectra $\{\lambda_n\}_1^\infty$, $\{\hat\mu_n\}_1^\infty$, i.e., a sequence $\{\hat\nu_n\}_1^\infty$, we can construct a sequence of approximations $\{q^{(k)}(x)\}_1^\infty$ as follows:

$$q^{(k)}(x) = q^{(k-1)}(x) + \tfrac{1}{4}\sum_{n=1}^{\infty}(\hat\nu_n - \nu_n^{(k-1)}) \times \frac{\int_0^1 w_n^{(k-1)}(\xi)d\xi}{\int_0^1 w_n^{(k-1)}(\xi)\,(d\omega_n^{(k-1)}/d\xi)d\xi}\,\frac{d\omega_n^{(k-1)}}{dx} \quad (3.60)$$

$$\mathcal{L}_{(k-1)}w_n^{(k-1)} = -\nu_n^{(k-1)}\frac{dw_n^{(k-1)}}{dx}$$

$$\frac{dw_n^{(k-1)}}{dx}\Big|_{x=0} = w_n^{(k-1)}\Big|_{x=1} = \frac{dw_n^{(k-1)}}{dx}\Big|_{x=1} = 0$$

where

$$\mathcal{L}_{(k-1)} = \frac{d^3}{dx^3} - 4q^{(k-1)}\frac{d}{dx} - 2\frac{dq^{(k-1)}}{dx}. \quad (3.61)$$

In part II, we shall investigate the convergence of this iteration scheme, or rather of a comparable one. It should be apparent at this stage that the desired potential $\hat q(x)$ associated with $\{\hat\nu_n\}_1^\infty$ is a fixed point of the operator which maps $q^{(k-1)}(x)$ into $q^{(k)}(x)$.

4. PART II

A. An iteration scheme for the symmetric potential

To proceed with (3.60), we must be careful to add some conditions on the class of $\{\hat\nu_n\}_1^\infty$, so as to guarantee that the desired $\hat q(x)$, if it exists, is differentiable.

We can avoid this problem for the case of a symmetric potential by dealing with the original Sturm-Liouville problem rather than the w equation. We can further simplify the iteration scheme by using a fixed set of eigenfunctions in the formula relating $q^{(k+1)}$ to $q^{(k)}$, for example, those eigenfunctions which correspond to the case $q \equiv 0$.

As a result, let us consider the following iteration scheme:

$$q^{(k)}(x) = q^{(k-1)}(x) - 2\sum_{n=1}^{\infty}(-1)^n[\hat\sigma_n - \sigma_n^{(k-1)}]\cos n\pi x,$$

$$\frac{d^2y_n^{(k-1)}}{dx^2} + \{\sigma_n^{(k-1)} - q^{(k-1)}(x)\}y_n^{(k-1)} = 0, \quad (4.1)$$

$$y_n^{(k-1)}(\pm 1) = 0.$$

Note that if $q^{(k-1)}(x)$ is an even function, then so is $q^{(k)}(x)$. Furthermore, if $q^{(k-1)}(x)$ has no mean over $(-1,1)$, so does $q^{(k)}(x)$.

(4.1) could have been taken as the starting point of our analysis, but the discussion in part I might be helpful in clarifying the genesis of this iteration as well as in attempting future generalizations.

In examining the convergence of the above iteration scheme, it will be useful to write the eigenvalues σ_n of (2.1) in a form reminiscent of their asymptotic expansion[11]:

$$\sigma_n = n^2\pi^2/4 - [(-1)^n/2]q_n + \epsilon_n[q], \quad (4.2)$$

where q_n is the Fourier cosine coefficient of $q(x)$, viz.,

$$q_n = \int_{-1}^{1} q(x)\cos n\pi x dx. \quad (4.3)$$

The eigenvalue problem (2.1) itself can be written thus:

$$y_n(x) = \sin\frac{n\pi}{2}(x+1) + \frac{2}{n\pi}\int_{-1}^{x}\{q(\xi) + \frac{(-1)^n}{2}q_n - \epsilon_n[q]\}y_n(\xi)\cdot\sin\frac{n\pi}{2}(x-\xi)d\xi \quad (4.4)$$

with

$$y_n(1) = 0. \quad (4.5)$$

Finally, given a spectrum $\{\hat{\sigma}_n\}_1^\infty$ satisfying (2.4), we shall define the operator $\mathfrak{A}$ as follows:

$$\mathfrak{A}q \equiv q - 2\sum_1^\infty (-1)^n[\hat{\sigma}_n - \sigma_n^{(q)}]\cos n\pi x \quad (4.6)$$

where $\{\sigma_n^{(q)}\}_1^\infty$ are the eigenvalues of (2.1) associated with the given $L_2(-1,1)$ symmetric, zero-mean function $q(x)$.

If $\hat{\sigma}_n - n^2\pi^2/4$ is in l_2, then $\mathfrak{A}q$ is in $L_2(-1,1)$ whenever q is in $L_2(-1,1)$. This follows from the fact that if (4.2) is looked upon as the asymptotic expansion of σ_n, ϵ_n is $O(1/n)$ (cf., e.g., Ref. 1, p. 11) and q_n is obviously l_2.

Thus $\mathfrak{A}$ maps the complete metric space $L_2(-1,1)$ onto itself and the sequence $\{q^{(k)}(x)\}$ would certainly be convergence if we were able to show that $\mathfrak{A}$ is a contraction mapping, i.e., if given two symmetric, zero-mean, $L_2(-1,1)$ functions $p(x)$ and $q(x)$, there exists a number $\kappa < 1$ such that

$$\|\mathfrak{A}p - \mathfrak{A}q\| < \kappa\|p - q\| \quad (4.7)$$

where $\|\ \|$ stands for the L_2-norm.

We shall see that provided that p and q are in a sufficiently small ball around the origin, $\mathfrak{A}$ is indeed a contraction mapping. This will establish both the existence and uniqueness of $\hat{q}(x)$. However, in view of the restriction on norm of $\hat{q}$, this result is not as general as that of Krein.

B. Some inequalities

In order to establish (4.7), we shall need to derive some inequalities. Let us first note that

$$\mathfrak{A}p - \mathfrak{A}q = 2\sum_1^\infty (-1)^n[\epsilon_n[p] - \epsilon_n[q]]\cos n\pi x \quad (4.8)$$

and so

$$\|\mathfrak{A}p - \mathfrak{A}q\| = 2\left[\sum_1^\infty |\epsilon_n[p] - \epsilon_n[q]|^2\right]^{1/2}. \quad (4.9)$$

Our goal should therefore be to get an upper bound for

$$\delta\epsilon_n = |\epsilon_n[p] - \epsilon_n[q]| \quad (4.10)$$

in terms of

$$\delta q = \|p - q\|. \quad (4.11)$$

Let us first establish

Lemma 3:

$$\int_{-1}^{+1} y_n^{(q)}(x)\sin\frac{n\pi}{2}(x+1)dx \equiv Q_n$$

is bounded away from zero.

Proof: Let us multiply (4.4) by $y_n^{(q)}(x)$ and integrate over $(-1,1)$. Then, with the normalization

$$\|y_n^{(q)}\| = 1 \quad (4.12)$$

we get

$$1 = \int_{-1}^{+1} y_n^{(q)}\sin\frac{n\pi}{2}(x+1)dx + \frac{2}{n\pi}\int_{-1}^{+1} y_n^{(q)}(x)dx\ \int_{-1}^{x}\left\{\frac{(-1)^n}{2}q_n + q(\xi) - \epsilon_n[q]\right\}\cdot y_n^{(q)}(\xi)\sin\frac{n\pi}{2}(x-\xi)d\xi, \quad (4.13)$$

i.e.,

$$1 \le |Q_n| + \frac{2}{n\pi}\left\{|\epsilon_n[q]| + \sqrt{\tfrac{3}{2}}\|q\|\right\}. \quad (4.14)$$

On the other hand, if we evalute (4.4) for $x = 1$ and use the boundary condition (4.5), we see that

$$\epsilon_n[q]\int_{-1}^{+1} y_n(\xi)\sin\frac{n\pi}{2}(1-\xi)d\xi = \int_{-1}^{+1}\left\{q(\xi) + \frac{(-1)^n}{2}q_n\right\}y_n^{(q)}(\xi)\cdot\sin\frac{n\pi}{2}(1-\xi)d\xi \quad (4.15)$$

and hence

$$|\epsilon_n[q]|\,|Q| \le \sqrt{\tfrac{3}{2}}\|q\|. \quad (4.16)$$

Combining (4.16) with (4.14) and solving the quadratic inequality, we see that

$$|Q_n| \ge \tfrac{1}{2} - \sqrt{\tfrac{3}{2}}\|q\|/n\pi + [\tfrac{1}{4} - 3\sqrt{\tfrac{3}{2}}\|q\|/n\pi + (3/2n^2\pi^2)\|q\|^2]^{1/2}. \quad (4.17)$$

In an analogous way, we can show that

$$P_n = \int_{-1}^{+1} y_n^{(p)}(x)\sin\frac{n\pi}{2}(1+x)dx \quad (4.18)$$

is bounded away from zero, and we shall write

$$|P_n + Q_n| \ge a^{-1} > 0. \quad (4.19)$$

Lemma 4:

$$\delta\epsilon_n < M[(\|p\| + \|q\|)/n\pi]\,\delta q, \quad (4.20)$$

where M is a numerical factor.

Proof: Let us return to the integral equations (4.4) corresponding to p and q. Setting $x = 1$ and subtracting them, we get

$$\int_{-1}^{+1}\left\{\frac{(-1)^n}{2}(p_n - q_n) + (p - q)\right\}\left\{y_n^{(p)} + y_n^{(q)}\right\}\cdot\sin\frac{n\pi}{2}(1-\xi)d\xi + \int_{-1}^{+1}\left\{\frac{(-1)^n}{2}(p_n + q_n) + (p + q)\right\}\left\{y_n^{(p)} - y_n^{(q)}\right\}\cdot\sin\frac{n\pi}{2}(1-\xi)d\xi - \{\epsilon_n[p] - \epsilon_n[q]\}\int_{-1}^{+1}\left\{y_n^{(p)} + y_n^{(q)}\right\}\sin\frac{n\pi}{2}(1-\xi)d\xi - \{\epsilon_n[p] + \epsilon_n[q]\}\int_{-1}^{+1}\left\{y_n^{(p)} - y_n^{(q)}\right\}\cdot\sin\frac{n\pi}{2}(1-\xi)d\xi = 0. \quad (4.21)$$

The first term in (4.21) will be written slightly differently in order to obtain the desired inequality. To that effect let us write

$$\int_{-1}^{+1}\left\{\frac{(-1)^n}{2}(p_n-q_n)+(p-q)\right\}\times\left\{y_n^{(p)}+y_n^{(q)}-2\sin\frac{n\pi}{2}(1+\xi)\right\}\sin\frac{n\pi}{2}(1-\xi)d\xi$$
$$=\int_{-1}^{+1}\left\{\frac{(-1)^n}{2}(p_n-q_n)+(p-q)\right\}\left\{y_n^{(p)}+y_n^{(q)}\right\}\sin\frac{n\pi}{2}(1-\xi)d\xi$$
$$-\int_{-1}^{+1}\left\{\frac{(-1)^n}{2}(p_n-q_n)+(p-q)\right\}[\cos n\pi\xi-\cos n\pi]d\xi, \tag{4.22}$$

i.e.,

$$\int_{-1}^{+1}\left\{\frac{(-1)^n}{2}(|p_n|-q_n)+(p-q)\right\}\times\left\{y_n^{(p)}+y_n^{(q)}-2\sin\frac{n\pi}{2}(1+\xi)\right\}\sin\frac{n\pi}{2}(1-\xi)d\xi$$
$$=\int_{-1}^{+1}\left\{\frac{(-1)^n}{2}(p_n-q_n)+(p-q)\right\}\left\{y_n^{(p)}+y_n^{(q)}\right\}\sin\frac{n\pi}{2}(1-\xi)d\xi. \tag{4.23}$$

Substituting (4.23) in (4.21), we get

$$a\delta\epsilon_n\le|\epsilon_n[p]+\epsilon_n[q]|\cdot\delta y_n+\sqrt{\tfrac{3}{2}}\|y_n^{(p)}+y_n^{(q)}-2\sin\tfrac{1}{2}n\pi(1+\xi)\|\cdot\delta q+\sqrt{\tfrac{3}{2}}\|p+q\|\cdot\delta y_n, \tag{4.24}$$

where

$$\delta y_n=\|y_n^{(p)}-y_n^{(q)}\|. \tag{4.25}$$

In view of (4.16) and (4.19) we can rewrite (4.24) thus:

$$\delta\epsilon_n\le\beta\{\|p\|+\|q\|\}\,\delta y_n+\gamma\|y_n^{(p)}+y_n^{(q)}-2\sin\tfrac{1}{2}n\pi(1+\xi)\|\delta q, \tag{4.26}$$

where β and γ are generic constants.

We now need an upper bound for δy_n and for

$$\delta z_n=\|y_n^{(p)}+y_n^{(q)}-2\sin\tfrac{1}{2}n\pi(1+\xi)\|. \tag{4.27}$$

We return to the integral equations (4.4) from which we can deduce that

$$y_n^{(p)}(x)-y_n^{(q)}(x)=\frac{1}{n\pi}\int_{-1}^{x}\left\{\frac{(-1)^n}{2}(p_n-q_n)+(p-q)-\epsilon_n[p]+\epsilon_n[q]\right\}\times\left\{y_n^{(p)}+y_n^{(q)}\right\}\sin\frac{n\pi}{2}(x-\xi)d\xi$$
$$+\frac{1}{n\pi}\int_{-1}^{+1}\left\{\frac{(-1)^n}{2}(p_n+q_n)+(p+q)-\epsilon_n[p]-\epsilon_n[q]\right\}\left\{y_n^{(p)}-y_n^{(q)}\right\}\sin\frac{n\pi}{2}(x-\xi)d\xi. \tag{4.28}$$

As a result

$$\delta y_n\le\frac{\beta}{n\pi}\delta q+\frac{2}{n\pi}\delta\epsilon_n+\frac{\gamma}{n\pi}\{\|q\|+\|p\|\}\delta y_n, \tag{4.29}$$

where β and γ are two generic constants.

Finally, by means of analogous calculations, we can deduce that

$$\delta z_n\le\beta(\|p\|+\|q\|)/n\pi. \tag{4.30}$$

Combining (4.26), (4.30), and (4.31), we get

$$\delta\epsilon_n\le\frac{\beta}{n\pi}\{\|p\|+\|q\|\}\cdot\frac{\gamma\delta q+2\delta\epsilon_n}{1-(\rho/n\pi)\{\|p\|+\|q\|\}}+\theta\frac{\|p\|+\|q\|}{n\pi}\cdot\delta q, \tag{4.31}$$

where again β,γ,θ, and ρ are numerical constants. Solving for $\delta\epsilon_n$, we finally see that

$$\delta\epsilon_n\le M\cdot[(\|p\|+\|q\|)/n\pi]\,\delta q \tag{4.32}$$

where

$$M=\max_n\frac{n\pi+\beta\{\|p\|+\|q\|\}}{n\pi+\gamma\{\|p\|+\|q\|\}}=\max\left(1,\frac{\beta}{\gamma}\right). \tag{4.33}$$

Theorem: Given $\{\hat\sigma_n\}_1^\infty$ such that $[\sum_1^\infty(\hat\sigma_n-n^2\pi^2/4)^2]^{1/2}$ is sufficiently small, there exists a unique $\hat q(x)=\hat q(-x)$ such that

$$y_n''+\{\hat\sigma_n-\hat q(x)\}y_n=0,\qquad y_n(\pm1)=0. \tag{4.34}$$

Proof: Let us define $\hat R$ thus:

$$\hat R=\left[\sum_1^\infty\left(\hat\sigma_n-\frac{n^2\pi^2}{4}\right)^2\right]^{1/2} \tag{4.35}$$

and consider the ball B_R where

$$B_R=\{f(x)\,\big|\,\|f\|\le R\}. \tag{4.36}$$

Then, if p and q are in B_R, (4.9) and (4.33) implies that

$$\|\mathfrak{A}p-\mathfrak{A}q\|\le(4R_M/\sqrt{6})\,\|p-q\|. \tag{4.37}$$

Thus, if R is chosen such that

$$R=\sqrt{6}/4M, \tag{4.38}$$

then the mapping of the ball B_R is contractive.

By picking any even $q^{(0)}(x)\in B_R$, we can generate a Picard sequence $\{q^{(k)}(x)\}_1^\infty$ by means of the iteration

$$q^{(k)}(x)=\mathfrak{A}q^{(k-1)}(x). \tag{4.39}$$

This sequence will necessarily converge toward a unique even function, say $\hat q(x)$, which lies in B_R, i.e.,

$$\|\hat q\|\le R=\sqrt{6}/4M. \tag{4.40}$$

Clearly R and $\hat R$ must be related; hence M must determine $\hat R$. To see this, let us apply (4.33) to the functions $\hat q(x)$ and 0:

$$\hat\sigma_n-\tfrac{1}{4}n^2\pi^2-\tfrac{1}{2}(-1)^n\hat q_n\le M\cdot(R^2/n\pi)=3/8n\pi M. \tag{4.41}$$

As a result

$$\left|\hat\sigma_n-\frac{n^2\pi^2}{4}\right|^2\le\tfrac{1}{4}|\hat q_n|^2+\frac{3}{8n\pi M}|\hat q_n|+\frac{9}{64\pi^2M^2n^2}$$

and so

$$\hat{R}^2 \leq \tfrac{1}{4}\sum |\hat{q}_n|^2 + \frac{3}{8M}\left(\frac{1}{\pi^2}\sum\frac{1}{n^2}\right)^{1/2}\left(\sum|\hat{q}_n|^2\right)^{1/2} + \frac{3}{128M^2}. \quad (4.42)$$

Making use of Parseval's theorem, we deduce that

$$\hat{R} \leq 3\sqrt{6}/16M. \quad (4.43)$$

This upper bound for $\hat{R}$ is very conservative, and the iteration scheme is probably convergent for larger values of $\hat{R}$. Perhaps the single most important cause for this restriction on the existence proof lies in the bounds for the eigenvalues. Indeed, (4.2) together with the bound on $\epsilon_n[q]$ show that we are essentially using the form of the asymptotic estimates of the larger eigenvalues for the *entire* spectrum; this is valid provided that $q(x)$ is in a sufficiently small ball around the origin.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the discussions which I have had on several aspects of this work with Professor P.D. Lax and Professor J. Bona.

[1]G. Borg, Acta Math. **78**, 1 (1946).
[2]I. M. Gel'fand and B. M. Levitan, Izv. Akad. Nauk SSSR Ser. Mat. **15**, 309 (1951) [English transl., Amer. Math. Soc. Transl. **1**, 253 (1955)].
[3]V. A. Marcenko, Dokl. Akad. Nauk SSSR **72**, 457 (1950).
[4]N. Levinson, Mat. Tidsskr. B, 25 (1949).
[5]M. G. Krein, Prikl. Mat. Mekh. **16**, 555 (1952).
[6]B. M. Levitan, Izv. Akad. Nauk SSSR Ser. Mat. **28**, 63 (1964) [English transl., Amer. Math. Soc. Transl. **68**, 1 (1968)].
[7]M. G. Krein, Dokl. Akad. Nauk SSSR **76**, 21 (1951).
[8]M. G. Krein, Dokl. Akad. Nauk SSSR **76**, 345 (1951).
[9]E. C. Titchmarsh: *Eigenfunction Expansions, Part I* (Oxford U. P., Oxford, 1969).
[10]E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).
[11]H. Hochstadt, Comm. Pure Appl. Math. **14**, 749 (1961).

On the problem of diffraction*

Paul B. Bailey

Sandia Laboratories, Albuquerque, New Mexico 87115
(Received 4 June 1973; final revised manuscript received 17 September 1973)

It is shown how the methods of Hadamard, developed for the Cauchy problem in free space, can be extended to include the presence of simple obstacles. The modifications necessary are illustrated by treating the problem of diffraction by a conducting half-plane in two and three dimensions.

1. INTRODUCTION

The problem of diffraction by a planar edge has been studied almost continuously since Sommerfeld[1] first obtained the solution for the problem of plane waves striking a half-plane in 1896. The number of people who have worked on and solved in countless ways the various versions of the problem of diffraction is much too large to be recounted here, although some idea may be obtained by consulting the review article by Bouwkamp[2] or the recent book of Jones.[3] It is all the more surprising then to learn that the complete solution to the half-plane diffraction problem in two dimensions can be expressed in extremely simple form, as was discovered by Friedlander[4] in 1951. The corresponding simple solution for the three dimensional case was subsequently given by Wait[5] in 1957.

The method employed by Friedlander was a modification of that developed by Hadamard[6] for Cauchy's problem in free space. Namely, the use of "the fundamental formula," as Hadamard calls it, for second order partial differential equations (also known as Green's, Gauss', or Stoke's theorem), together with some "elementary solution" of the homogeneous differential equation. We follow the same general procedure but add some additional modifications to those of Friedlander in order to be able to deal with problems in more than two space dimensions.

The particular case of a half-plane in three dimensions is treated in detail, arriving at the solution obtained by Wait[5] by an entirely different method. The reason for going to the trouble of solving the half-plane problem once again is that the method employed here is considerably more general than that of Ref. 5. In particular, it can be used even in the case of a curved edge, which is done very briefly in Sec. 7.

What is new in this paper in the two-dimensional case is the method of determining the "elementary solutions" v^* (the end of Sec. 4) and again w [near the end of Sec. 5, around Eqs. (19)-(26)]. While v^* is already known in the case of reflection from a plane, it would not be known in the case of reflection from a curved surface but could still be found by exactly the same method (even the same formula) described in this paper. In the three-dimensional case, Friedlander's treatment does not apply as it stands. Considerable nontrivial changes must be made to the method employed in Ref. 4 in order to carry through the general idea, and this is done in detail in Sec. 6.

Diffraction by a curved edge was treated in Ref. 7 but that treatment was not wholly satisfactory. The one unsatisfactory feature was the method employed (constructing an infinite sequence of functions) for obtaining the elementary solution w for the diffraction region. But the treatment in this present paper (the end of Sec. 6) culminating in Eq. (56), gives an alternative formula for w which does not require infinite series at all. With this one change in the treatment in Ref. 7, the diffraction problem for a curved edge is complete. The curved edge diffraction problem has not been solved before.

For simplicity of presentation of the main ideas we restrict our attention to only two and three dimensions and to the case of zero initial conditions; but it will be obvious that the same procedure works in spaces of higher dimensions, and even for more general hyperbolic differential equations, as well as for different initial or boundary conditions.

Sections 2 and 3 are devoted to a statement of the problem and of the fundamental formula. In Sec. 4 we review the method of solving the free space and half-space problems in two dimensions as introduction to the solution of the half-plane problem in Sec. 5, which is essentially the same as that of Friedlander.[4] Sections 5 and 6 extend the method to the three-dimensional half-plane problem, while Sec. 7 completes the solution of the problem of diffraction by a curved edge previously treated in Ref. 7.

2. STATEMENT OF THE PROBLEM

The problem is to find the function $u(t,P)$, P a point in 2-space (or 3-space), which satisfies the inhomogeneous wave equation

$$\left(\nabla^2 - \frac{\partial^2}{\partial t^2}\right) u(t,P) = f(t,P), \tag{1}$$

and the boundary condition

$$u(t,P) = U_0(t,P) \quad \text{on the plane,} \tag{2}$$

as well as the zero initial conditions

$$u(O,P) = 0 = \frac{\partial u}{\partial t}\,(O,P), \tag{3}$$

and, finally, the usual "edge condition," that $u(t,P)$ goes to zero as the square root of the distance from the edge. (See Ref. 8.) Although it will not be explicitly referred to again, it is also assumed that f and U_0 are continuous and "compatible":

$$f(O,P) = U_0(O,P) \quad \text{on the plane.}$$

And when we say that u satisfies (1), we mean that u has continuous first and second partial derivatives everywhere except possibly on the plane itself (and on wavefronts), and the second partials satisfy (1).

3. THE FUNDAMENTAL FORMULA

Our main tool—practically our only tool—for obtaining the solution $u(t,P)$ of (1) is the fundamental formula (FF) for the wave operator applied to u and some solution v of the homogeneous wave equation

$$\left(\nabla^2 - \frac{\partial^2}{\partial t^2}\right) v = 0, \tag{4}$$

over an arbitrary region $\mathfrak{D}$; namely

$$\int_{\partial\mathfrak{D}} \left(v\,\frac{du}{d\nu} - u\,\frac{dv}{d\nu}\right) dS + \int_{\mathfrak{D}} fv\,dV = 0, \tag{FF}$$

where $du/d\nu = l_0(\partial u/\partial t) - l_1(\partial u/\partial x_1) - l_2(\partial u/\partial x_2) - l_3(\partial u/\partial x_3)$, l_0, l_1, l_2, l_3 are the direction cosines of the inward normal to the surface $\partial\mathfrak{D}$, the "transversal derivative." The transversal derivative obviously reduces to the outward normal derivative on a time independent surface.

This formula (FF), which is obtained in the usual way[6] by multiplying (1) by v, (4) by u, subtracting, integrating over $\mathfrak{D}$, and applying Stoke's theorem, of course assumes that the functions u and v involved are nice enough in $\mathfrak{D}$ and on the boundary $\partial\mathfrak{D}$. But since we shall often want to use it for functions v and regions $\mathfrak{D}$ such that v is singular on $\partial\mathfrak{D}$, it will be understood that all surface and volume integrals arising are interpreted in the sense of limits of corresponding integrals over approximating interior regions. This is a common practice which need not give rise to any ambiguities so long as all the limits are taken in some consistent fashion.

4. HALF-SPACE PROBLEM IN TWO DIMENSIONS

Let $v(t-t', |PP'|)$ denote the so-called "elementary solution" of (4):

$$v(t-t', |PP'|) \equiv \Gamma^{-1/2}(t-t', |PP'|), \tag{5}$$

where

$$\Gamma(t-t', |PP'|) \equiv (t-t')^2 - |PP'|^2, \tag{6}$$

and $|PP'|$ is the distance between two points P, P' in 2-space.

Recall that this elementary solution v enables one to solve the initial value problem for (1) in the absence of any plane by the simple procedure of merely applying (FF) to u, the desired solution, and v over the region $\mathfrak{D}_i$ bounded by the retrograde light cone. More explicitly, let (t,P) be fixed and apply (FF) to the functions $u(t',P')$, $v(t-t', |PP'|)$ over the region

$$\mathfrak{D}_i = \{(t',P'): 0 < t' < t - |PP'|\}.$$

It turns out, though we will not go through the elementary but slightly tedious details here, that, because of the intimate relation between v and the surface $\Gamma = 0$ (more precisely, the surface $t' = t - \overline{PP'}$), only that portion of the surface near the vertex (t,P) yields a nonzero contribution. This contribution is just 2π times the value of $u(t',P')$ at the vertex (t,P). On the remainder of the boundary of $\mathfrak{D}_i$, which is to say where $t' = 0$, u and $du/d\nu$ each are zero because of the initial conditions (3). Thus (FF) reduces to simply

$$2\pi u(t,P) + \int_{\mathfrak{D}_i} f(t',P')\, v(t-t', |PP'|)\, dV' = 0, \tag{7}$$

which obviously gives the solution u for the free space problem.

In the present case this solution (7) is valid only for such (t,P) that the cone $\Gamma = 0$ does not meet the plane. For larger t there will be reflection from the plane.

One way of obtaining the solution in this case is the well-known method of "images" wherein one notices that the function

$$v^*(t-t', |PP'|) \equiv v(t-t', |P^*P'|),$$

where P^* denotes the image of P in the plane, is a suitable elementary solution of the homogeneous wave equation (4) corresponding to the region $\mathfrak{D}_r$,

$$\mathfrak{D}_r = \{(t',P'): 0 < t' < t - |P^*P'|\}.$$

In this case, one first applies (FF) to $u(t',P')$ and $v(t-t', |PP'|)$ over the region $\mathfrak{D}_i$ (it being understood that both $\mathfrak{D}_i$ and $\mathfrak{D}_r$ are limited to those points P' on the same side of the plane as P), giving

$$-2\pi u(t,P) = \int_{\mathfrak{D}_i} f v\, dV' + \int_{p_i} \left(v\frac{du}{d\nu'} - u\frac{dv}{d\nu'}\right) dS', \tag{8}$$

where p_i denotes the portion of $\partial\mathfrak{D}_i$ consisting of the plane, since the only portions of $\partial\mathfrak{D}_i$ contributing nonzero surface integrals are p_i and the neighborhood of the vertex of $\Gamma = 0$. Second, apply (FF) to $u(t',P')$ and $v^*(t-t', |PP'|)$ over the region $\mathfrak{D}_r$. This time the surface integral over the curved portion of $\partial\mathfrak{D}_r$ (a portion of the cone $t' = t - |P^*P'|$) gives nothing, since the neighborhood of its vertex (t,P^*) is not involved. Hence we get

$$0 = \int_{\mathfrak{D}_r} f v^*\, dV' + \int_{p_r} \left(v^*\frac{du}{d\nu'} - u\frac{dv^*}{d\nu'}\right) dS'. \tag{9}$$

Since obviously

$$v^*(t-t', |PP'|) = v(t-t', |PP'|), \tag{10}$$

$$\frac{dv^*}{d\nu'}(t-t', |PP'|) = -\frac{dv^*}{d\nu'}(t-t', |PP'|), \tag{11}$$

when P' is on the plane p, subtracting (9) from (8) gives

$$-2\pi u(t,P) = \int_{\mathfrak{D}_i} f v\, dV' - \int_{\mathfrak{D}_r} f v^*\, dV' - 2\int_{p_i} u\frac{dv}{d\nu'}\, dS' \tag{12}$$

(where we have used the fact that p_r is the same as p_i). Now substituting U_0 for u in the integral in (12) gives the solution u for the half-space problem.

But suppose for the moment that we did not happen to know the "suitable" v^* for the above argument, and had to try to find it. The properties v^* must possess to qualify it as "the continuation of v corresponding to the region $\mathfrak{D}_r$" are exactly those which we have used above. Namely, v^* should be a solution of (4) which satisfies (10) and (11), and which gives zero surface integral over the curved portion of $\partial\mathfrak{D}_r$ for arbitrary u.

To obtain $v^*(t,P;T,Q)$ from these properties we apply (FF) to $v^*(t',P';T,Q)$ and $v(t-t', |PP'|)$ over the region bounded by the backward light cone from (t,P) and the forward cone from (T,Q) (with $T < t$), together with the plane p. Since (FF) has $f \equiv 0$ in this case we get

$$-2\pi v^*(t,P;T,Q) = \int_p \left(v(t-t', |PP'|)\frac{dv^*}{d\nu'}(t',P';T,Q) - v^*\frac{dv}{d\nu'}\right) dS',$$

since the integrals over the two curved surfaces vanish (except for the part near the vertex) on account of the properties of v^* on the one and v on the other. This can be rewritten, using (10) and (11), as just

$$-2\pi v^*(t,P;T,Q) = -\int_p \Big(v(t-t', |PP'|)\frac{dv}{d\nu'}(t'-T, |P'Q|) + v(t'-T, |P'Q|)\frac{dv}{d\nu'}(t-t', |PP'|)\Big)\, dS', \tag{13}$$

which expresses our supposed unknown v^* in terms of the known v. Here the surface integration is understood to be over those (t',P') on p which are contained between $\mathfrak{D}_i$ and $\mathfrak{D}_r$, of course.

5. HALF-PLANE PROBLEM

In many respects one can treat the diffraction problem in much the same way as the reflection problem was just

treated. There is one important difference, however, which really must be cleared up first. That is, that although the solution v^* corresponding to the reflection wavefront $t' = t + |P^*P'|$ has the property that the integral

$$\int \left(v^*(t,P;t',P')\frac{du}{d\nu'}(t',P') - u\frac{dv^*}{d\nu'}\right)dS'$$

over any portion of the wavefront vanishes, for arbitrary u, just as the corresponding integral with v in place of v^* vanishes over any portion of the direct wavefront $t' = t - |PP'|$ which does not include a neighborhood of the vertex (t,P), the corresponding integral with the solution of (4) corresponding to diffraction must vanish over any portion of the diffraction wavefront which does not include a neighborhood of *the common generator*. That is, the generator which the direct and diffraction wavefronts have in common (or the reflection and diffraction wavefronts have in common, as the case may be). Keeping this difference in mind, we can proceed much as before.

Let q denote the "shadow plane" corresponding to the fixed point (t,P). This is the half-plane determined by (t,P) and the edge of the half-plane separating the region $\mathfrak{D}_i$ from which direct radiation can be received at (t,P) and the region from which only diffracted radiation can be received. Let $\mathfrak{D}_r$ again denote the region from which reflected radiation can be received, and let $\mathfrak{D}_d$ denote the region from which diffracted radiation can be received.

First apply (FF) to the unknown solution $u(t',P')$ and $v(t-t',|PP'|)$ over $\mathfrak{D}_i$

$$-2\pi u(t,P) = \int_{\mathfrak{D}_i} fvdV' + \int_{p_i+q_i}\left(v(t-t',|PP'|\frac{du}{d\nu'}(t',P') - u\frac{dv}{d\nu'}\right)dS', \quad (14)$$

where q_i, of course, denotes the portion of the boundary $\mathfrak{D}_i$ formed by q. [Notice that on q, $du/d\nu' = (1/r')\,\partial u/\partial\theta'$, so that $du/d\nu'$ is not integrable on q unless u goes to zero with r', the distance from the edge.]

Next, let $w(t,P;t',P')$ denote the "solution" of (4), assumed to be unknown for the moment, which is the continuation of v into $\mathfrak{D}_d$. By "continuation" we mean to imply that v in $\mathfrak{D}_i$ and w in $\mathfrak{D}_d$ combine to give a solution of (4) which is appropriate to the total region. This means, among other things, that the total function is continuous and has continuous first partials everywhere, except possibly on p. It follows that w and its normal derivative must have jump discontinuities on q which exactly match those of v in $\mathfrak{D}_i$ [which is why the quotation marks when referring to w as a "solution" of (4)].

Thus if $[w]$ denotes the jump in w in crossing q from the side on which (t,P) lies to the other side, then we must have

$$[w] = v \quad \text{and} \quad \left[\frac{dw}{d\nu'}\right] = 0 \quad \text{on } q \qquad (15)$$

$[dv/d\nu' = 0$ on q, since $(d/d\nu')\,|PP'| = 0$ there].

Next, apply (FF) to $u(t',P')$ and $w(t,P;t',P')$ over the region $\mathfrak{D}_d^+$, which is the portion of $\mathfrak{D}_d$ on the same side of q as (t,P), and over the region $\mathfrak{D}_d^-$, which is the remaining portion of $\mathfrak{D}_d$:

$$0 = \int_{\mathfrak{D}_d^+} fwdV' + \int_{p_d^+q_d^+g^+}\left(w(t,P;t',P')\frac{du}{d\nu'}(t',P') - u\frac{dw}{d\nu'}\right)dS', \quad (16)$$

$$0 = \int_{\mathfrak{D}_d^-} fwdV' + \int_{p_d^-q_d^-g^-}\left(w(t,P;t',P')\frac{du}{d\nu'}(t',P') - u\frac{dw}{d\nu'}\right)dS', \quad (17)$$

where we have used subscripts $+,-$ to denote the side of p or q on which (t,P) lies, and the opposite side, respectively. The notation g^+ or g^- is intended as shorthand for the contribution from the surface integral over the diffraction wavefront in the neighborhood of the common generator, which is the line of intersection of the wavefront and the shadow plane.

Adding (16) and (17), and using (15) and the observation that $d/d\nu'$ has opposite directions on q^+ and q^-, gives

$$0 = \int_{\mathfrak{D}_d} fwdV' + \int_{q_d^+} v(t-t',|PP'|)\frac{du}{d\nu'}(t',P')dS' + \int_{p_d^+p_d^-g^+g^-}\left(w(t,P;t',P')\frac{du}{d\nu'}(t',P') - u\frac{dw}{d\nu'}\right)dS'. \quad (18)$$

Adding (18) to (14) now gives (since $q_d = q_i$)

$$-2\pi u(t,P) = \int_{\mathfrak{D}_i} fvdV' + \int_{\mathfrak{D}_d} fwdV' + \int_{q_i}\left(v(t-t',|PP'|)\frac{du}{d\nu'}(t',P') - u\frac{dv}{d\nu'}\right)dS' + \int_{p_d^+p_d^-g^+g^-}\left(w(t,P;t',P')\frac{du}{d\nu'}(t',P') - u\frac{dw}{d\nu'}\right)dS'. \quad (19)$$

Before proceeding further along these lines with the reflection function v^*, etc., we may as well pause long enough to determine the function $w(t,P;t',P')$, since the phenomenon of diffraction is a direct consequence of the presence of an *edge*, having nothing to do with reflection. That is to say, the function w is determined independently of where the half-plane is oriented about the edge—at least until such time as the half-plane is reached. It follows that the contributions in (19) indicated by g^+ and g^- must cancel each other, for arbitrary u satisfying the edge condition.

To keep clear as to whether it is the function w on the same side of q and p as (t,P), or the other side, let us denote the two functions, temporarily, by w_+, w_-, respectively. First apply (FF) to $w_+(T,Q;t',P')$ and $v(t-t',|PP'|)$ over the region bounded by the backward cone from (t,P) and the forward diffraction cone corresponding to (T,Q), which is assumed to contain (t,P), of course. There being no half-plane present, we get

$$-2\pi w_+(T,Q;t,P) = 2\int_G\left(v(t-t',|PP'|)\frac{dw_+}{d\nu'}(T,Q;t',P') - w_+\frac{dv}{d\nu'}\right)dS', \quad (20)$$

where G denotes that there is a contribution to the surface integral over the diffraction wavefront from the neighborhood of the generator in the shadow plane corresponding to (T,Q). The factor of 2 is to account for the contribution from the surface on each side of this generator.

Next apply (FF) to the same two functions over the region bounded by the backward diffraction cone corresponding to (t,P) and the forward diffraction cone corresponding to (T,Q):

$$0 = 2\int_G\left(v(t-t',|PP'|)\frac{dw_+}{d\nu'}(T,Q;t',P') - w_+\frac{dv}{d\nu'}\right)dS' + 2\int_g\left(v(t-t',|PP'|)\frac{dw_+}{d\nu'}(T,Q;t',P') - w_+\frac{dv}{d\nu'}\right)dS', \quad (21)$$

where g refers to the point (t, P). Combining (20) and (21) gives

$$-2\pi w_+(T,Q;t,P) = -2\int_g \left(v(t-t',|PP'|)\frac{dw_+}{d\nu'}(T,Q;t',P') - w_+\frac{dv}{d\nu'}\right)dS'. \tag{22}$$

Likewise

$$-2\pi w_-(T,Q;t,P) = -2\int_g \left(v(t-t',|PP'|)\frac{dw_-}{d\nu'}(T,Q;t',P') - w_-\frac{dv}{d\nu'}\right)dS'. \tag{23}$$

But the right-hand sides of (22) and (23) must be the negatives of one another, according to the remarks following (19). In other words, $w_- = -w_+$. Taken together with (15) this implies that

$$w_+ = -\tfrac{1}{2}v = -w_- \quad \text{on } q. \tag{24}$$

Hence, finally, from (22),

$$-2\pi w_+(T,Q;t,P) = \int_g \left(v(t-t',|PP'|)\frac{dv}{d\nu'}(t'-T,|QP'|) - v\frac{dv}{d\nu'}\right)dS'. \tag{25}$$

We can now return to Eq. (19), which simplifies slightly to

$$-2\pi u(t,P) = \int_{\mathfrak{D}_i} fv\,dV' + \int_{\mathfrak{D}_d} fw\,dV' + \int_{p_i-p_d^+} \times \left(v(t-t',|PP'|)\frac{du}{d\nu'}(t',P') - u\frac{dv}{d\nu'}\right)dS', \tag{26}$$

where by $p_i - p_d^+$ we mean to indicate that the result of integrating over p_d^+ is to be subtracted from the integration over p_i.

Now apply (FF) to u, v^* over region $\mathfrak{D}_r$, and to u, w^* (where w^* is the appropriate continuation of v^* into $\mathfrak{D}_d$, just as w is the continuation of v) over $\mathfrak{D}_d$, and add the results to give

$$0 = \int_{\mathfrak{D}r} fv^*dV' + \int_{\mathfrak{D}_d} fw^*dV' + \int_{p_i-p_d^+}\left(v^*(t-t',|PP'|)\frac{du}{d\nu'}(t',P') - u\frac{dv^*}{d\nu'}\right)dS'. \tag{27}$$

Finally, subtracting (27) from (26) and using the fact that $u = U_0$ on p, we have

$$-2\pi u(t,P) = \int_{\mathfrak{D}_i} fv\,dV' - \int_{\mathfrak{D}_r} fv^*dV' + \int_{\mathfrak{D}_d} f(w-w^*)dV' - 2\int_{p_i-p_d^+} U_0\frac{dv}{d\nu'}\,dS'. \tag{28}$$

6. HALF-PLANE DIFFRACTION IN THREE DIMENSIONS

The method of obtaining the solution to the half-plane problem in three dimensions (or any number of dimensions, for that matter) will be seen to be very much like that just used in the preceding section for two dimensions, at least in broad outline. There are, however, some rather important differences, due to the fact that the "elementary solution" is now not an ordinary function, as in the two-dimensional case, but is really a generalized function. This fact gives rise to some technical difficulties not present in the previous case, but does not change the problem in any essential way.

Recall that the elementary solution for the wave equation in free space (in three space dimensions) is

$$v \equiv \delta(\overline{t-t'}^2 - R^2), \qquad R = |PP'|, \tag{29}$$

where δ is the Dirac delta function (generalized function), or, equivalently,

$$v \equiv \tfrac{1}{2}R^{-1}\,\delta(t-t'-R). \tag{30}$$

Just as R denotes the shortest distance between two points P and P' in free space, the problem of diffraction by an edge involves the shortest distance between the two points via the edge. Let this minimal distance be denoted by A. Thus if (r,θ,z) denote the cylindrical coordinates of P, relative to the edge, and (r',θ',z') the coordinates of P', then

$$A^2 = (r+r')^2 + (z'-z)^2. \tag{31}$$

According to the solution for the half-plane problem given by Wait,[5] the elementary solution corresponding to the diffraction region is w defined by

$$w \equiv -(1/\pi)\cdot[(A^2-R^2)^{1/2}/(\overline{t-t'}^2 - R^2)(\overline{t-t'}^2 - A^2)^{1/2}]. \tag{32}$$

By letting $\mathfrak{D}_i, \mathfrak{D}_r, \mathfrak{D}_d$ be defined as before, the half-plane problem solution is then (zero initial and boundary conditions)

$$-2\pi u = \int_{\mathfrak{D}_i} fv\,dV - \int_{\mathfrak{D}_r} fv^*dV + \int_{\mathfrak{D}_d} f(w-w^*)\,dV. \tag{33}$$

But suppose we did not happen to have this solution and wanted to find it. Naturally our first step would be to apply (FF) to u and v over the region $\mathfrak{D}_i$, except for the difficulty in interpreting the surface integral

$$\int_{\partial\mathfrak{D}_i}\left(v\frac{du}{d\nu} - u\frac{dv}{d\nu}\right)dS, \tag{34}$$

wherein v is not an ordinary function but a generalized function, whose support is precisely the surface $\partial\mathfrak{D}_i$.

One way of meeting this particular difficulty is that of Hadamard.[6] Namely, change the problem from one in three dimensions to four by adding a dummy fourth variable (his "ascent"), interpret the resulting surface integral in the sense of his "finite part" integrals, and then integrate out the dummy fourth variable (his "descent").

We shall not employ Hadamard's method in the present circumstance, however, because of the serious difficulties which would arise when we come to the diffraction region. [Without going into details, the difficulty we have in mind is due to the fact that w, given by (32) has some of the character of an elementary solution in two dimensions—as appears from the factor $(\overline{t-t'}^2 - A^2)^{-1/2}$—as well as some of the character of three dimensions—as appears from the factor $(\overline{t-t'}^2 - R^2)^{-1}$.]

Another way of interpreting surface integrals such as (34)—and this is the method we shall employ—is that described in Ref. 9 and which we shall refer to as "ln ϵ-part" from here on. Briefly the idea is this:

Take as elementary solution for free space in three dimensions not the generalized function v given by (30), but the ordinary function

$$v \equiv (\overline{t-t'}^2 - R^2)^{-1}. \tag{35}$$

(This is an elementary solution, as is readily verified.) Of course, the surface integrals such as (34) with this v will not make sense in the usual way, being infinite. But we can apply the fundamental formula (FF) to u and v over some approximating region $\mathfrak{D}_i(\epsilon)$, depending upon a parameter $\epsilon > 0$, determine the asymptotic behavior of all the resulting integrals as $\epsilon \downarrow 0$, and retain only those terms of order $\ln\epsilon$. Not only will the equation resulting from this procedure be true, but, as explained in Ref. 9, it will have precisely the properties needed to enable one to solve the Cauchy problem for the wave equation. To signify that only the $\ln\epsilon$ part of an integral is to be retained we shall attach a superscript L to the integral sign.

First apply (FF) to $u(t', P')$, $v(t - t', |PP'|)$ given by (35), over the region $\mathfrak{D}_i$:

$$\int^L_{\mathfrak{D}_i} fv + \int^L_{\partial\mathfrak{D}_i} \left(v\frac{du}{d\nu'} - u\frac{dv}{d\nu'}\right) = 0. \tag{36}$$

The volume integral is easiest, since it can be written

$$\begin{aligned}\int^L f(t', P')[(t-t')^2 - R^2]^{-1}\,dt'\,dV(P') &= \int^L \int f(t - R - \epsilon, P')[\overline{(t-t')^2} - R^2]^{-1}\,dt'\,dV(P') \\ &+ \int^L \int [f(t', P') - f(t - R - \epsilon, P')][\overline{t-t'}^2 - R^2]^{-1}\,dt'\,dV(P').\end{aligned} \tag{37}$$

But the first integral, being equal to

$$\int^L f(t - R - \epsilon, P') \cdot \frac{1}{2R}\left[\ln\frac{t - t' - R}{t - t' + R}\right]^{t'=t-R-\epsilon}_{t'=0} dV(P'),$$

obviously has $\ln\epsilon$ part equal to

$$\int_{0<R<t} f(t - R, P') \cdot \frac{1}{2R}\, dV(P'); \tag{38}$$

and the second integral in (37) converges to a finite limit as $\epsilon \downarrow 0$, hence has $\ln\epsilon$ part equal to zero. In other words the first integral in (36) is just (38), which is an ordinary integral. [Incidentally, (38) is, of course, the same as would be obtained by integrating fv over $\mathfrak{D}_i$ with v given by (30).] Of course the integration region in (38) is bounded still by the "shadow plane," though we have not explicitly indicated this fact.

As for the surface integral in (36), nothing is obtained from the plane $t' = 0$ since u and $\partial u/\partial t$ both vanish there. So there remains only the curved surface $t' = t - R - \epsilon$, say, the half-plane p, and the shadow plane q, which in cylindrical coordinates about the edge has equation $\theta' = \theta + \pi$. Without going through the details here, since they are not difficult in any case, one finds that the only portion of the cone which contributes anything to the $\ln\epsilon$ part is the neighborhood of the vertex; and the contribution here is just 2π times the value of u at the vertex (t, P). Thus (36) can be rewritten

$$-2\pi u(t, P) = \int_{\mathfrak{D}_i} f(t - R, P')\frac{1}{2R}\, dV(P') + \int^L_{q+p}\left(v\frac{du}{d\nu'} - u\frac{dv}{d\nu'}\right),$$

where $\mathfrak{D}_i$ denotes the region $0 < R < t$, bounded by the shadow plane, and where q, p denote the shadow plane and half-plane, respectively.

But on q, $d/d\nu' = (1/r')\,\partial/\partial\theta'$ (in cylindrical coordinates), and clearly $\partial v/\partial\theta' = 0$ since $\partial R/\partial\theta' = 0$ on q. And

$$\int^L_q \frac{1}{r'}\frac{\partial u}{\partial\theta'}[\overline{t-t'}^2 - R^2]^{-1}dt'dr'dz' = \int_q \frac{1}{r'}\frac{\partial u}{\partial\theta'}(t - R, P')\cdot\frac{1}{2R}\, dr'dz'. \tag{39}$$

As for p, the term $v\,du/d\nu'$ gives a contribution exactly as in (38), but the term $u\,dv/d\nu'$ gives (in cylindrical coordinates r', θ', z')

$$\begin{aligned}&\int^L u\cdot 2r\sin\theta\cdot[\overline{t-t'}^2 - R^2]^{-2}dt'\,dr'\,dz' \\ &= \int_p 2r\sin\theta\cdot\left(\frac{1}{4R^2}\frac{\partial u}{\partial t'}(t - R, P') + \frac{1}{8R^3}u(t - R, P')\right) \\ &\quad\times dr'dz'.\end{aligned} \tag{40}$$

Thus, finally, (36) can be written in terms of standard symbols in analysis as

$$\begin{aligned}-2\pi u(t, P) &= \int_{\mathfrak{D}_i} f(t - R, P')\cdot\frac{1}{2R}\, dV(P') + \int_q \frac{1}{r'}\frac{\partial u}{\partial\theta'}(t - R, P')dr'dz' \\ &- \int_p\left[\frac{1}{r'}\frac{\partial u}{\partial\theta'}(t - R, P) - 2r\sin\theta\left(\frac{1}{4R^2}\frac{\partial u}{\partial t'}(t - R, P')\right.\right. \\ &\left.\left. + \frac{1}{8R^3}u(t - R, P')\right)\right]dr'\,dz'.\end{aligned} \tag{41}$$

But let us return now to the much simpler (though less explicit) notation of $\ln\epsilon$ part, so that (41) is condensed to just

$$-2\pi u = \int^L_{\mathfrak{D}_i} fv + \int^L_{p_i+q_i}\left(v\frac{du}{d\nu'} - u\frac{dv}{d\nu'}\right) \tag{42}$$

with p_i, q_i denoting those portions of p, q which constitute the boundaries of $\mathfrak{D}_i$. Applying (FF) to u and v^* (the image of v in the plane p) over $\mathfrak{D}_r$ gives similarly

$$0 = \int^L_{\mathfrak{D}_r} fv^* + \int^L_{p_r+q_r^*}\left(v^*\frac{du}{d\nu'} - u\frac{dv^*}{d\nu'}\right), \tag{43}$$

there being zero contribution from the surface of the cone $t' = t - R^*$, just as in the two-dimensional case, since the integration does not include the neighborhood of the vertex.

Next, apply (FF) to u, w [defined by (32)] over the diffraction region $\mathfrak{D}^+_d$, the "positive" portion of $\mathfrak{D}_d$,

$$0 = \int_{\mathfrak{D}^+_d} fw + \int_{c^+_d+p^+_d+q^+_d}\left(w\frac{du}{d\nu'} - u\frac{dw}{d\nu'}\right) \tag{44}$$

and to the other portion $\mathfrak{D}^-_d$

$$0 = \int_{\mathfrak{D}^-_d} fw + \int_{c^-_d+p^-_d+q^-_d}\left(w\frac{du}{d\nu'} - u\frac{dw}{d\nu'}\right). \tag{45}$$

(The integrals here are ordinary ones.)

Since w is the elementary solution of the homogeneous wave equation which is "appropriate" for the diffraction region $\mathfrak{D}_d$, and u satisfies the edge condition, the only portion of the surface c_d contributing a nonzero amount is the neighborhood of the line of intersection of c_d and q (the generator common to c_d and c_i). And since $w = 0$ on q, and $dw/d\nu'$ has opposite sign on q^+ and q^-, adding (44) and (45) gives

$$0 = \int_{\mathfrak{D}_d} fw + \int_{p^+_d+p^-_d+2c^+_d}\left(w\frac{du}{d\nu'} - u\frac{dw}{d\nu'}\right). \tag{46}$$

Of course, it turns out that the contribution from c^+_d is exactly half the amount obtained in (42) from q^+_i, but opposite in sign, so that these two cancel each other when (42) and (46) are added. Thus

$$-2\pi u = \int^L_{\mathfrak{D}_i} fv + \int_{\mathfrak{D}_d} fw + \int^L_{p_i}\left(v\frac{du}{d\nu'} - u\frac{dv}{d\nu'}\right) + \int_{p^+_d+p^-_d}\left(w\frac{du}{d\nu'} - u\frac{dw}{d\nu'}\right). \tag{47}$$

Similarly the image functions w^*, v^* combine to give the result

$$0 = \int_{\mathfrak{D}_r}^{L} fv^* + \int_{\mathfrak{D}_d} fw^* + \int_{p_r}^{L}\left(v^*\frac{du}{d\nu'} - u\frac{dv^*}{d\nu'}\right) + \int_{p_d^+p_d^-}\left(w^*\frac{du}{d\nu'} - u\frac{dw^*}{d\nu'}\right). \quad (48)$$

Subtracting (48) and (47), finally, gives the solution to our problem

$$-2\pi u = \int_{\mathfrak{D}_i}^{L} fv - \int_{\mathfrak{D}_r}^{L} fv^* + \int_{\mathfrak{D}_d} f(w-w^*) - 2\int_{p_i}^{L} u\frac{dv}{d\nu'} \quad (49)$$

since $v = v^*, w = w^*, dv/d\nu' = -dv^*/d\nu'$ on P, and u is given on p.

But suppose we did not already have the function w and were faced with the problem of finding it. The properties w must have, in addition to being a solution of the homogeneous wave equation, are just those already cited. Namely, $w = 0$ on $q, dw/d\nu'$ has opposite sign on q^+ and q^-, and, for arbitrary functions u which satisfy the edge condition and zero initial conditions,

$$\int_{q_i^+}^{L}\left(v\frac{du}{d\nu'} - u\frac{dv}{d\nu'}\right) + \int_{c_d}\left(w\frac{du}{d\nu'} - u\frac{dw}{d\nu'}\right) = 0. \quad (50)$$

To determine w from these properties, apply (FF) to $w(t'-T, P', Q), v(t-t', |PP'|)$ over the region bounded by the backward cone from (t, P), the forward diffraction "cone" (not really a cone) $t' = T + A(P', Q)$ corresponding to (T, Q), the region being supposed cut along the shadow plane corresponding to (T, Q). Since the orientation of the half-plane is unimportant, as commented on before, we assume the half-plane coincides with the "cut" $q(T, Q)$, for convenience. And since both v and w have singularities, to avoid any possible confusion as to the meaning of $\ln\epsilon$ part integrals we shall explicitly specify the approximating surfaces as

$$c(\epsilon_1): t' = t - \epsilon_1 - R(P, P'), \quad (51)$$

$$c_d(\epsilon_2): t' = T + \epsilon_2 + A(P', Q). \quad (52)$$

With these preliminaries taken care of, we apply (FF) and take the $\ln\epsilon_1$ part, ϵ_2 for the moment being fixed. The only nonzero contribution from c is just 2π times the value of w at the vertex, and there is no contribution at all from q, since the integrals on the two sides of q cancel each other. Hence all we get is

$$-2\pi w(t-T, P, Q) = \int_{c_d(\epsilon_2)}^{L_1}\left(v\frac{dw}{d\nu'} - w\frac{dv}{d\nu'}\right), \quad (53)$$

the integration region being of course bounded by $c(\epsilon_1)$.

It is tempting to try to replace the integral in (53) involving our supposed unknown function w by the shadow plane integral in (50) involving another v in place of w; but unfortunately (50) is not known to hold *except* when the function u appearing there satisfies the edge condition. In fact it turns out that (50) is not true for other u.

However, there is another identity, namely

$$\int_{q_i^+}^{L}\left(v\frac{du}{d\nu'} + u\frac{dv}{d\nu'}\right) = \tfrac{1}{2}\int_{c_d}^{L}\left(v\frac{du}{d\nu'} - u\frac{dv}{d\nu'}\right), \quad (54)$$

which also holds for arbitrary u satisfying the edge condition and zero initial conditions. (See the Appendix for a proof.) Hence we have

$$\int_{c_d}\left(w\frac{du}{d\nu'} - u\frac{dw}{d\nu'}\right) = -\tfrac{1}{2}\int_{c_d}^{L}\left(v\frac{du}{d\nu'} - u\frac{dv}{d\nu'}\right). \quad (55)$$

Initially we know only that this relation holds for functions u satisfying the edge condition, but it is readily extended to hold for other u as well.

Finally, using (55) in (53) gives us

$$-2\pi w(t-T, P, Q) = -\tfrac{1}{2}\int_{c_d(\epsilon_2)}^{L_1, L_2}\left(v(t-t', |PP'|)\frac{dv}{d\nu'}(t'-T, |P'Q|) - v(t'-t, |P'Q|)\frac{dv}{d\nu'}(t-t', |PP'|)\right), \quad (56)$$

where the superscripts L_1, L_2 mean we are to take the $\ln\epsilon_1$ part, with ϵ_2 fixed, and then take the $\ln\epsilon_2$ part of that result.

7. DIFFRACTION BY A CURVED EDGE

The treatment[7] of the problem of diffraction by a planar curved edge had the one unsatisfactory feature that the elementary solution corresponding to the diffraction region—which we have been denoting by w—was constructed in the form of an infinite series. This feature is unsatisfactory for two reasons: For one, the series was not shown to converge; for another, even the computation of the individual terms of the series would be a formidable task.

But we now know how to get w directly without having to resort to infinite series at all. Namely, we use either Eq. (53), if we know the values of w appearing in the surface integral, or use Eq. (56) otherwise. Since these relations were obtained by an argument that did not depend upon the edge being straight (but did require that multiple diffractions do not occur), they still hold.

The reason for mentioning (53) is that it is definitely simpler than (56), and because the values of w on the diffraction wavefront c_d have already been obtained in Ref. 7, Eqs. (18) and (26), even for a curved edge.

With this change in the manner of obtaining w, the solution given in Ref. 7 for diffraction by a curved edge is complete.

APPENDIX

1. Proof of Eq. (50)

Taking cylindrical coordinates (r', θ', z') about the edge, we have

$$\int_{q_i}^{L}\left(v\frac{du}{d\nu'} - u\frac{dv}{d\nu}\right) = \int^{L}\int\int\frac{1}{r'}\frac{\partial u}{\partial\theta'}(t', r', \theta', z')\,[\overline{t-t'}^2 - r^2 - r'^2 + 2rr'\cos\overline{\theta-\theta'} - \overline{z-z'}^2]^{-1}\,dt'\,dr'\,dz'$$

evaluated at $\theta' = \theta + \pi$

$$= \int^{L}\int\int\frac{1}{r'}\frac{\partial u}{\partial\theta'}(t', r', \theta+\pi, z')(\overline{t-t'}^2 - A^2)^{-1} \times dt'\,dr'\,dz',$$

where

$$A^2 = \overline{r+r'}^2 + \overline{z-z'}^2 \quad (A1)$$

$$= \int^{L}\int\int\frac{1}{r'}\frac{\partial u}{\partial\theta'}(t-A-\epsilon, r', \theta+\pi, z')(\overline{t-t'}^2 - A^2)^{-1} \times dt'\,dr'dz' + \int^{L}\int\int\frac{1}{r'}\left(\frac{\partial u}{\partial\theta'}(t', r', \theta+\pi, z') - \frac{\partial u}{\partial\theta'}(t-A-\epsilon, r', \theta+\pi, z')\right) \cdot (\overline{t-t'}^2 - A^2)^{-1}\,dt'\,dr'\,dz',$$

the integration over t' being over $0 < t' < t - A - \epsilon$.

But the last integral clearly converges to a finite limit as $\epsilon \downarrow 0$, so its $\ln\epsilon$ part is zero. The first integral is

$$\frac{1}{2}\int^L\int\frac{1}{r'}\frac{\partial u}{\partial\theta'}(t-A-\epsilon,r',\theta+\pi,z')\ln\frac{\epsilon(T+A)}{(2A+\epsilon)(t-A)}\cdot A^{-1}\,dr'\,dz',$$

which clearly has $1n\epsilon$ part equal to just

$$\frac{1}{2}\int\int\frac{1}{r'}\frac{\partial u}{\partial\theta'}(t-A,r',\theta+\pi,z')\cdot A^{-1}\,dr'dz'. \quad \text{(A2)}$$

As for the other integral in (50), take as approximating surfaces $c_d(\epsilon)$: $t'=t-\epsilon-A$, with A given by (A1). By taking coordinate systems, s,φ,θ' on this surface where s,φ are given by

$$r'+r=s\cos\varphi, \qquad z'-z=s\sin\varphi,$$

the second integral in (50) is

$$\int\int\int\left(\tilde{w}\frac{\partial\tilde{u}}{\partial s}-\tilde{u}\frac{\partial\tilde{w}}{\partial s}\right)s(s\cos\varphi-r)\,ds\,d\varphi\,d\theta',$$

where $\tilde{u}$ denotes

$$\tilde{u}(s,\varphi,\theta')\equiv u(t-s-\epsilon,r',\theta',z') \quad \text{(A3)}$$

$$=\int\int[\tilde{u}\tilde{w}s(s\cos\varphi-r)]_{s=r\sec\varphi}^{s=t-\epsilon}\,d\varphi\,d\theta'-\int\tilde{u}\left(2s(s\cos\varphi-r)\frac{\partial\tilde{w}}{\partial s}+(2s\cos\varphi-r)\tilde{w}\right)\times ds\,d\varphi\,d\theta'. \quad \text{(A4)}$$

The first integral in (A4) vanishes because of the initial condition and edge condition on u. The second integral is equal to

$$-\int\int\int\tilde{u}\tilde{w}\cdot\left\{\frac{2s(2r+s\cos\varphi)}{[4r(s\cos\varphi-r)\cos^2\frac{1}{2}(\theta-\theta')+2s\epsilon+\epsilon^2]}+\frac{\epsilon\cos\varphi-r}{2s+\epsilon}\right\}ds\,d\varphi\,d\theta'$$

$$\doteq\frac{1}{2\pi}\int\int\int\frac{1}{r'}\tilde{u}\,\frac{\lambda^{1/2}\,|\cos\frac{1}{2}(\theta-\theta')|\,d\theta'ds\,d\varphi}{[\cos^2\frac{1}{2}(\theta-\theta')+\lambda]^2}, \qquad \lambda=\frac{s\epsilon}{2rr'}. \quad \text{(A5)}$$

For the θ' integration in (A5), clearly the only portion which has a nonzero limit as $\epsilon\downarrow 0$ is that near $\theta'=\theta\pm\pi$. In this neighborhood we can make the change of variable

$$\theta'=\theta+\pi+2\arcsin x,$$

so that the integral is

$$2\frac{1}{\pi}\cdot\lambda^{1/2}\int\int\int_0^\delta\frac{\tilde{u}}{r'}\cdot\frac{x\,dx}{(x^2+\lambda)^2}$$

$$=\frac{1}{\pi}\lambda^{1/2}\int\int\frac{1}{r'}\left[\frac{\tilde{u}}{x^2+\lambda}\right]_0^\delta ds\,d\varphi-\frac{1}{\pi}\lambda^{1/2}\int\int\int\frac{1}{x^2+\lambda}\cdot\frac{1}{r'}\cdot\frac{1}{dx}\tilde{u}\,dx\,ds\,d\varphi$$

$$\sim\pi^{-1}\int\int\left\{\lambda^{-1/2}\,u(t-s,s\cos\varphi-r,\theta+\pi,z+s\sin\varphi)-\pi\frac{\partial\tilde{u}}{\partial\theta'}\right\}\frac{ds\,d\varphi}{r'}+o(\epsilon) \quad \text{as } \epsilon\downarrow 0,$$

and the finite part of this is clearly

$$\int\frac{1}{r'}\frac{\partial u}{\partial\theta'}(t-s,r',\theta+\pi,z')\,ds\,d\varphi.$$

In the original variables r',z' this is exactly the same as (A2).

2. Proof of Eq. (54)

Taking the same surfaces $c_d(\epsilon)$ and variables (s,φ,θ') as in (i) above, the right-hand side of (54) is

$$\frac{1}{2}\int^L\int\int\left(\tilde{v}\frac{\partial\tilde{u}}{\partial s}-\tilde{u}\frac{\partial\tilde{v}}{\partial s}\right)s(s\cos\varphi-r)\,ds\,d\varphi\,d\theta',$$

where now

$$\tilde{v}=[4r(s\cos\varphi-r)\cos^2\tfrac{1}{2}(\theta-\theta')+2s\epsilon+\epsilon^2]^{-1}$$

$$=-4r\int^L\int\int\tilde{u}\left(\frac{2\epsilon s}{[4r(s\cos\varphi-r)\cos^2\frac{1}{2}(\theta-\theta')+2s\epsilon+\epsilon^2]^2}-\frac{1}{[4r(s\cos\varphi-r)\cos^2\frac{1}{2}(\theta-\theta')+2s\epsilon+\epsilon^2]}\right)ds\,d\theta'\,d\varphi. \quad \text{(A6)}$$

Evidently the only portion of the range of θ' which can contribute a nonzero $\ln\epsilon$ part is that near $\theta'=\theta\pm\pi$. For this range make the same change of variable from θ' to x as before, using now the observation that

$$2\int_0^\delta\tilde{u}\left(\frac{\lambda}{(x^2+\epsilon)^2}-\frac{1/2}{x^2+\lambda}\right)dx=2\int_0^\delta\tilde{u}\frac{d}{dx}\frac{x}{x^2+\lambda}\,dx,$$

$$\lambda=\frac{s}{2rr'}\epsilon, \quad \text{(A7)}$$

$$=\left[\frac{2x\tilde{u}}{x^2+\lambda}\right]_0^\delta-\int_0^\delta\frac{2x}{x^2+\lambda}\cdot\frac{d\tilde{u}}{dx}\,dx$$

$$=\left[\frac{2x\tilde{u}}{x^2+\lambda}-\ln(x^2+\lambda)\frac{d\tilde{u}}{dx}\right]_0^\delta+\int_0^\delta\frac{d^2\tilde{u}}{dx^2}\cdot\ln(x^2+\lambda)\,dx,$$

which evidently has $1n\epsilon$ part equal to just $d\tilde{u}/dx\,|_{x=0}$, or

$$2\cdot\frac{\partial u}{\partial\theta'}(t-s,r',\theta+\pi,z'). \quad \text{(A8)}$$

Now multiplying by $1/4rr'$ to make (A7) comparable to (A6), the result is (A2) again.

3. Verification that (56) is equivalent to (53)

With $w(t'-T,P',Q)$ defined as in (32) and $v(t-t',|PP'|)$ as in (31), the right-hand side of (53) is, approximately (dropping squares and higher powers of ϵ's),

$$J\equiv-\pi^{-1}\iiint\frac{\sqrt{2}\,\epsilon_2^{1/2}\rho^{3/2}\sqrt{r's}\,|\cos\frac{1}{2}(\Theta-\theta')|\,ds\,d\theta'\,d\varphi}{(\Gamma_{\epsilon_2}-2sD_{\epsilon_2})[2\rho r'\cos^2\frac{1}{2}(\Theta-\theta')+s\epsilon_2]^2}, \quad \text{(A9)}$$

where now s,φ are given by

$$\rho+r'=s\cos\varphi, \qquad z'-Z=s\sin\varphi,$$

$\tilde{u}$ means $u(t'=T+s+\epsilon_2)$,

$$\Gamma_{\epsilon_2}\equiv\overline{t-T-\epsilon_2^2}-r^2-\rho^2-\overline{Z-z^2}-2r\rho\cos(\theta-\theta'),$$

$$D_{\epsilon_2}\equiv\overline{t-T-\epsilon_2}-(r\cos\overline{\theta-\theta'}+\rho)\cos\varphi+(Z-z)\sin\varphi.$$

The $1n\epsilon_1$ part of (A9) is obtained by integrating over s, the result being the same as dropping the s integration, replacing $\Gamma_{\epsilon_2}-2sD_{\epsilon_2}$ by $2D_{\epsilon_2}$, and replacing all s's by $\Gamma_{\epsilon_2}/2D_{\epsilon_2}$. The result is

$$\bar{J}\equiv-\rho\pi^{-1}\cdot\iint\frac{\sqrt{\epsilon_2\Gamma_{\epsilon_2}}\cdot\sqrt{2\rho(\Gamma_{\epsilon_2}\cos\varphi-2\rho D_{\epsilon_2})}\,|\cos\frac{1}{2}(\Theta-\theta')|\,d\theta'\,d\varphi}{[2\rho(\Gamma_{\epsilon_2}\cos\varphi-2\rho D_{\epsilon_2})\cos^2\frac{1}{2}(\Theta-\theta')+\epsilon_2\Gamma_{\epsilon_2}]^2}. \quad \text{(A10)}$$

In exactly the same fashion the right-hand side of (56), say,

$$I \equiv \tfrac{1}{2}\rho \iiint \left(\frac{2\epsilon_2 s}{[2\rho r' \cos^2 \tfrac{1}{2}(\Theta - \theta') + s\epsilon_2]^2} - \frac{1}{[2\rho r' \cos^2 \tfrac{1}{2}(\Theta - \theta') + s\epsilon_2]} \right) (\Gamma_{\epsilon_2} - 2sD_{\epsilon_2})^{-1} \, ds d\theta' d\varphi,$$

has $\ln\epsilon_1$ part equal to

$$\tilde{I} \equiv \rho \iint \left(\frac{\Gamma_{\epsilon_2}\epsilon_2}{[2\rho(\Gamma_{\epsilon_2}\cos\varphi - 2\rho D_{\epsilon_2})\cos^2 \tfrac{1}{2}(\Theta - \theta') + \Gamma_{\epsilon_2}\epsilon_2]^2} - \frac{\tfrac{1}{2}}{[2\rho(\Gamma_{\epsilon_2}\cos\varphi - 2\rho D_{\epsilon_2})\cos^2 \tfrac{1}{2}(\Theta - \theta') + \Gamma_{\epsilon_2}\epsilon_2]} \right) d\theta' d\varphi. \tag{A11}$$

We need to show that the $\ln\epsilon_2$ part of (A11) is the same as the finite part of (A10).

Obviously we obtain zero finite part in (A10) and zero $\ln\epsilon_2$ part in (A11) from those portions of the integral for which θ' is bounded away from $\Theta \pm \pi$ and also φ is bounded away from the zeros of $\Gamma_{\epsilon_2}\cos\varphi - 2\rho D_{\epsilon_2}$. Denote by I_1, I_2, I_3 and J_1, J_2, J_3 the remaining integrals wherein, respectively, θ' and φ both are near the critical values, θ' is near $\Theta \pm \pi$ but φ is bounded away from its two critical values, φ is near one of its critical values but θ' is bounded away from $\Theta \pm \pi$.

In I_2, J_2 make the change of variables from θ' to x given by

$$a^{1/2}\Gamma_{\epsilon_2}^{-1/2}\cos\tfrac{1}{2}(\Theta - \theta') = x, \qquad a = 2\rho(\Gamma_{\epsilon_2}\cos\varphi - 2\rho D_{\epsilon_2}) \tag{A12}$$

(note that a here does not involve θ') to get

$$I_2 = \pi^{-1} \iint f(x,\varphi) \left(\frac{\epsilon_2}{(x^2+\epsilon_2)^2} - \frac{\tfrac{1}{2}}{x^2+\epsilon_2} \right) dx d\varphi,$$

$$J_2 = \iint f(x,\varphi) \frac{x\sqrt{\epsilon_2}}{(x^2+\epsilon_2)^2} \, dx d\varphi,$$

where

$$f(x,\varphi) = \rho a^{-1/2}\Gamma_{\epsilon_2}^{1/2} / [\tfrac{1}{2}\Gamma_{\epsilon_2}\sin\tfrac{1}{2}(\Theta - \theta') + 2r\rho \sin(\theta - \theta')\cos\tfrac{1}{2}(\Theta - \theta')].$$

Integration by parts gives

$$I_2 = \frac{1}{2\pi} \int \left(f(x,\varphi) \frac{x}{x^2+\epsilon_2} \right) d\varphi - \frac{1}{2\pi} \iint \frac{\partial f}{\partial x} \cdot \frac{x}{x^2+\epsilon_2} \, dx d\varphi$$

which can now be seen to have $\ln\epsilon_2$ part equal to

$$\frac{1}{4\pi} \int \frac{\partial f}{\partial x}(0,\varphi) \, d\varphi, \tag{A13}$$

while similarly

$$J_2 = \int \left(-\frac{\tfrac{1}{2}\epsilon_2 f}{x^2+\epsilon_2} \right) d\varphi + \frac{1}{2} \iint \frac{\epsilon_2}{x^2+\epsilon_2} \frac{\partial f}{\partial x} \, dx d\varphi,$$

which has finite part equal to (A13).

For I_3, J_3 we make a change of variables from φ to y given by

$$2\rho(\Gamma_{\epsilon_2}\cos\varphi - 2\rho D_{\epsilon_2}) = y^2 \sec^2 \tfrac{1}{2}(\Theta - \theta')\Gamma_{\epsilon_2} \tag{A14}$$

so that

$$I_3 = \pi^{-1} \iint \left(\frac{\epsilon}{(y^2+\epsilon_2)^2} - \frac{\tfrac{1}{2}}{y^2+\epsilon_2} \right) y g(y,\theta') \, dy d\theta',$$

$$J_3 = \iint \frac{\sqrt{\epsilon_2}\, y^2}{(y^2+\epsilon_2)^2} \, g(y,\theta') \, dy d\theta',$$

$$g(y,\theta') = \sec^2 \tfrac{1}{2}(\Theta - \theta') \Big/ \frac{d}{d\varphi} (\Gamma_{\epsilon_2}\cos\varphi - 2\rho D_{\epsilon_2}).$$

One finds readily that the finite part of I_3 and the $\ln\epsilon_2$ part of J_3 are each equal to

$$\frac{1}{4\pi} \int g(0,\theta') \, d\theta'.$$

Finally, in I_1 and J_1 make the change of variables from θ', φ to x, y given by

$$2\rho(\Gamma_{\epsilon_2}\cos\varphi - 2\rho D_{\epsilon_2}) = y^2, \tag{A15}$$

$$|\cos\tfrac{1}{2}(\Theta - \theta')| = x\Gamma_{\epsilon_2}^{1/2}, \tag{A16}$$

so that

$$I_1 = \pi^{-1} \iint \left(\frac{\epsilon_2}{(x^2y^2+\epsilon_2)^2} - \frac{\tfrac{1}{2}}{(x^2y^2+\epsilon_2)} \right) y h(x,y) \, dx dy,$$

$$J_1 = \iint \frac{\sqrt{\epsilon_2}\, xy^2}{(x^2y^2+\epsilon_2)^2} \, h(x,y) \, dx dy,$$

where

$$h(x,y) = \Gamma_{\epsilon_2}^2 \left(\frac{d}{d\varphi}(\Gamma_{\epsilon_2}\cos\varphi - 2\rho D_{\epsilon_2}) \times [\tfrac{1}{2}\Gamma_{\epsilon_2}\sin\tfrac{1}{2}(\Theta - \theta') + r\rho \sin(\theta - \theta')\cos\tfrac{1}{2}(\Theta - \theta')] \right)^{-1}.$$

The computation is a little more complicated in this case than in the previous cases but one gets, integrating over $x \in (0,\delta)$ and $y \in (0,\gamma)$, the finite part of J_1 and the $\ln\epsilon_2$ part of I_1 are each equal to

$$-\frac{1}{4\pi\delta} h(\delta,0) + \frac{1}{4\pi} \frac{\partial h}{\partial x}(0,\gamma) \ln(\delta\gamma).$$

This completes the verification.

*This work was supported by the United States Atomic Energy Commission.

[1]A Sommerfeld, Math. Ann. 47, 317 (1896).
[2]C. J. Bouwkamp, Rept. Progr. Phys. 17, 35 (1954).
[3]D. S. Jones, *The Theory of Electromagnetism* (MacMillan, New York, 1964).
[4]F. G. Friedlander, Quart. J. Mech. Appl. Math. 4, 344 (1951).
[5]J. R. Wait, Can. J. Phys. 35, 693 (1957).
[6]J. Hadamard, *Lectures on Cauchy's Problem* (Yale U. P., New Haven, 1923).
[7]P. B. Bailey and G. E. Barr, J. Math. Phys. 12, 2080 (1971).
[8]F. G. Friedlander, Proc. Camb. Phil. Soc. 45, 395 (1949).
[9]P. B. Bailey, SIAM J. Math. Anal. 2, 500 (1971).

An approach to the study of quantum systems*

F. T. Hioe

Institute for Fundamental Studies, Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

(Received 4 September 1973)

A perturbation method for quantum mechanical problems is presented in which the successive terms are obtained by algebraic recurrence relations without the use of any diagrams. Two examples, one an anharmonic oscillator with quartic self-interaction, and one a spin with quadratic self-interaction, are given to illustrate the use of this recurrence relation perturbation method.

1. INTRODUCTION

The study of a quantum system in equilibrium has centered mainly in determining the energies corresponding to the Hamiltonian of the system and in determining the thermodynamic partition function of the system. The basic quantum mechanical operators encountered are well known: the boson creation and annihilation operators $a^\dagger$ and a, and the spin operators S^z, S^+, and S^-, and the basis functions most commonly used for these operators are the $|n\rangle$ for the boson operators and the $|S,m\rangle$ for the spin operators of total spin $[S(S+1)]^{1/2}$, where $m=-S$, $-S+1,\ldots,S$. In general, the energies and the partition function of a quantum system can only be determined approximately by perturbation methods and the most widely used of these up to now is probably the diagram method.

More recently, the following two developments have provided the motivation for the work reported in this paper:

(1) The author[1,2] calculated exactly the partition functions and derived the phase transition properties of the Dicke and some generalized Dicke models of superradiance in the thermodynamic limit by employing the Glauber's coherent states[3] $|\alpha\rangle$ for the photon field. An interesting point to note from these calculations is that the coherent states $|\alpha\rangle$ are a convenient basis for the calculation of the thermodynamic properties of the Dicke model although they appear to lack any direct physical meaning. It was also noted by the author[2] that the thermodynamic and phase transition properties of the BCS model of superconductivity can be very simply derived by using the Bloch coherent states[4,5,6] $|\mu\rangle$, which are the analog of the Glauber's coherent states for the photon field, for the spin system. The concept of coherent states[7] was originally developed in connection with the so-called phase space methods which go as far back as Wigner's[8] original work on the Wigner distribution function. Following the work on the Dicke and the BCS models, the author was prompted to try to look for new representations for the Hamiltonian of an arbitrary quantum system which might prove to be more convenient to use than the traditionally used representations.

(2) In another development, Bender and Wu[9] recently published a series of papers dealing with the energy expansions of a quantum anharmonic oscillator. One of the methods they used was a difference equation method which made no use of any diagrams. Independently, Helleman and Montroll[10] also developed a recurrence relation method to study the classical anharmonic oscillator. A more general recurrence relation method was also used earlier by Schweber.[11] However, a recurrence relation method has never been developed, to the best of the author's knowledge, for the study of systems involving (one or more) spins.

The contents of this paper consist of the following:

(i) In Sec. 2, we deduce, by means of the coherent state representation, the following representation of the boson operators:

$$a_i^\dagger \to z_i, \quad a_i \to \frac{\partial}{\partial z_i} \tag{1.1}$$

where z's are arbitrary complex variables (independent for different i). Representation (1.1) was studied by Bargmann[12] and others.[13,14] In terms of this representation of the boson operators, the energy equation

$$H(a_i^\dagger, a_i)|E\rangle = E|E\rangle, \tag{1.2}$$

where $H(a_i^\dagger, a_i)$ denotes $H(a_1^\dagger, a_1, a_2^\dagger, a_2, \cdots)$, is transformed into the form of

$$H\left(z_i, \frac{\partial}{\partial z_i}\right) f(z_1, z_2, \cdots) = E f(z_1, z_2, \cdots), \tag{1.3}$$

where $f(z_1, z_2, \cdots)$ is a power series in $z_1, z_2, \cdots$.

(ii) In Sec. 3, we deduce, also by the coherent state representation, the following representation of the spin operators:

$$\begin{aligned} S_i^+ &\to z_i \frac{\partial}{\partial \zeta_i}, \\ S_i^- &\to \zeta_i \frac{\partial}{\partial z_i}, \\ S_i^z &\to \tfrac{1}{2}\left(z_i \frac{\partial}{\partial z_i} - \zeta_i \frac{\partial}{\partial \zeta_i}\right), \end{aligned} \tag{1.4}$$

where z's and ζ's are arbitrary complex variables. Representation (1.4) was used by Bargmann[15] and others.[14,16] In terms of this representation of the spin operators, the energy equation

$$H(S_i^z, S_i^+, S_i^-)|E\rangle = E|E\rangle \tag{1.5}$$

is transformed into the form of

$$\begin{aligned} &H\left(\tfrac{1}{2}\left(z_i \frac{\partial}{\partial z_i} - \zeta_i \frac{\partial}{\partial \zeta_i}\right), z_i \frac{\partial}{\partial \zeta_i}, \zeta_i \frac{\partial}{\partial z_i}\right) \\ &\quad \times f(\zeta_1, z_1, \zeta_2, z_2, \cdots)\Big|_{\zeta_1, \zeta_2, \cdots = 1} \\ &= E f(\zeta_1, z_1, \zeta_2, z_2, \cdots)\big|_{\zeta_1, \zeta_2, \cdots = 1}, \end{aligned} \tag{1.6}$$

where $f(\xi_1, z_1, \xi_2, z_2, \cdots)$ is of the form

$$\prod_i \{\xi_i^{2S-K} z_i^K \sum_p c_p \xi_i^{-p} z_i^p\}. \tag{1.7}$$

(iii) In Sec. 4, we consider a general eigenvalue equation of the form

$$\left\{H_0\left(z_i, \frac{\partial}{\partial z_i}\right) + \lambda H_1\left(z_i, \frac{\partial}{\partial z_i}\right)\right\} f(z_1, z_2, \cdots) = Ef(z_1, z_2, \cdots) \tag{1.8}$$

and describe a recurrence relation method by which the eigenvalues $E(\lambda)$ can be expanded in a power series of λ.

(iv) In Sec. 5, the recurrence relation method is applied to two specific examples.

(v) Some concluding remarks are given in Sec. 6.

2. SYSTEMS OF BOSONS

Let us first consider the case of a single mode oscillator. The creation and annihilation operators, $a^\dagger$ and a, as is well known, satisfy the commutation relation

$$[a, a^\dagger] = 1 \tag{2.1}$$

and the state vectors $|n\rangle$ acted upon by a and $a^\dagger$ are given by

$$a|n\rangle = n^{1/2}|n-1\rangle \tag{2.2}$$

and

$$a^\dagger|n\rangle = (n+1)^{1/2}|n+1\rangle. \tag{2.3}$$

Another convenient set of state vectors for representing the state of the system is the set of Glauber's coherent states[3] $|\alpha\rangle$ which are defined as the eigenstates of the annihilation operators, namely

$$a|\alpha\rangle = \alpha|\alpha\rangle \tag{2.4}$$

and

$$\langle\alpha|a^\dagger = \alpha^*\langle\alpha|. \tag{2.5}$$

In terms of the state vectors $|n\rangle$, the coherent states $|\alpha\rangle$ are given by

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_n \frac{\alpha^n}{(n!)^{1/2}} |n\rangle \tag{2.6}$$

and

$$\langle\alpha| = e^{-|\alpha|^2/2} \sum_n \frac{(\alpha^*)^n}{(n!)^{1/2}} \langle n|. \tag{2.7}$$

The states $|\alpha\rangle$, while not orthogonal, are complete, namely,

$$\int \frac{d^2\alpha}{\pi} |\alpha\rangle\langle\alpha| = 1 \tag{2.8}$$

where $d^2\alpha$ means $d(\mathrm{Re}\alpha)d(\mathrm{Im}\alpha)$. The scalar product of two coherent states is given by

$$\langle\alpha|\beta\rangle = \exp(\alpha^*\beta - \tfrac{1}{2}|\alpha|^2 - \tfrac{1}{2}|\beta|^2). \tag{2.9}$$

Thus an arbitrary state $|g\rangle$ given by, say,

$$|g\rangle = \sum_n c_n |n\rangle \tag{2.10}$$

can be expressed in terms of the states $|\alpha\rangle$ by

$$|g\rangle = \int \frac{d^2\alpha}{\pi} e^{-|\alpha|^2/2} g(\alpha^*) |\alpha\rangle, \tag{2.11}$$

where

$$g(\alpha^*) = \sum_n \frac{c_n(\alpha^*)^n}{(n!)^{1/2}}. \tag{2.12}$$

The above properties of the coherent states [Eqs. (2.4) to (2.12)] were given by Glauber.[3]

Now consider a Hamiltonian $H(a^\dagger, a)$ in which the operators a and $a^\dagger$ are normally ordered, namely, let $H(a^\dagger, a)$ be of the form

$$H(a^\dagger, a) = \sum_m \sum_n H_{mn} (a^\dagger)^m a^n, \tag{2.13}$$

and consider the eigenvalue equation

$$H(a^\dagger, a)|E\rangle = E|E\rangle. \tag{2.14}$$

The state $|E\rangle$ which, in terms of the state vectors $|n\rangle$, is given by

$$|E\rangle = \sum_{n=0}^{\infty} c_n |n\rangle$$

can be expressed in terms of the set of Glauber's coherent states by

$$|E\rangle = \int \frac{d^2\beta}{\pi} e^{-|\beta|^2/2} f(\beta^*) |\beta\rangle \tag{2.15}$$

where

$$f(\beta^*) = \sum_{n=0}^{\infty} \frac{c_n(\beta^*)^n}{(n!)^{1/2}}. \tag{2.15'}$$

Multiplying both sides of Eq. (2.14) by an arbitrary state $\langle\alpha|$, we obtain on the left-hand side

$$\langle\alpha|H|E\rangle = \int \frac{d^2\beta}{\pi} e^{-|\beta|^2/2} H(\alpha^*, \beta) f(\beta^*) e^{\alpha^*\beta - |\alpha|^2/2 - |\beta|^2/2} \tag{2.16}$$

by noting the relations (2.4), (2.5), and (2.9). Therefore,

$$\langle\alpha|H|E\rangle = e^{-|\alpha|^2/2} \sum_m \sum_n H_{mn} (\alpha^*)^m \int \frac{d^2\beta}{\pi} e^{\alpha^*\beta - |\beta|^2} \beta^n f(\beta^*) \tag{2.17}$$

$$= e^{-|\alpha|^2/2} \sum_m \sum_n H_{mn} (\alpha^*)^m \left(\frac{\partial}{\partial\alpha^*}\right)^n f(\alpha^*) \tag{2.18}$$

$$= e^{-|\alpha|^2/2} H\left(\alpha^*, \frac{\partial}{\partial\alpha^*}\right) f(\alpha^*), \tag{2.19}$$

where (2.18) was obtained from (2.17) by using the formula

$$\int \frac{d^2\beta}{\pi} e^{\alpha^*\beta - |\beta|^2} \beta^n f(\beta^*) = \left(\frac{\partial}{\partial\alpha^*}\right)^n f(\alpha^*). \tag{2.20}$$

On the right-hand side, we obtain

$$E\langle\alpha|E\rangle = E \int \frac{d^2\beta}{\pi} e^{-|\beta|^2/2} f(\beta^*) e^{\alpha^*\beta - |\alpha|^2/2 - |\beta|^2/2}$$

$$= e^{-|\alpha|^2/2} E f(\alpha^*). \tag{2.21}$$

From (2.19) and (2.21), we have

$$H\left(\alpha^*, \frac{\partial}{\partial\alpha^*}\right) f(\alpha^*) = Ef(\alpha^*)$$

or, since α^* is an arbitrary complex variable, we write

$$H\left(z, \frac{\partial}{\partial z}\right) f(z) = Ef(z). \tag{2.22}$$

This suggests a representation of the boson creation and annihilation operators given by

$$a^\dagger \to z$$

and

$$a \to \frac{\partial}{\partial z} \tag{2.23}$$

which, as can be verified, satisfies the commutation relation $[a, a^\dagger]=1$. We shall refer to (2.23) as the Bargmann representation (or realization) of the boson operators. In the transformation of Eq. (2.14) into Eq. (2.22), the eigenvalues of the equation are kept unchanged. The eigenfunctions, on the other hand, are expressed in a different form from the original eigenfunctions $|E\rangle$. This representation $f(z)$ of the eigenfunctions has the form of a power series in z and cannot be directly associated with any physical interpretation. The energies, of course, are the physically most significant quantities. Whether or not the corresponding representation of the eigenfunctions has any direct physical meaning should not worry us or prevent us from looking for the most appropriate form of the eigenvalue equation for the determination of the energies of the system. To get an idea of what advantage would be gained by working with the Bargmann representation, let us compare it with the Schrödinger representation for the case of a simple harmonic oscillator for which the energy levels are given by $E_K=\frac{1}{2}+K$, $K=0,1,2,\ldots$. In the Schrödinger representation, the position and momentum variables are represented by $x \to x$ and $p \to -i\partial/\partial x$ and the energy equation for the oscillator is

$$\tfrac{1}{2}\left(-\frac{d^2}{dx^2}+x^2\right)u(x)=Eu(x) \tag{2.24}$$

with the eigenfunction $u_K(x)$ corresponding to the Kth energy level given by

$$u_K(x)=e^{-x^2/2}H_K(x), \tag{2.25}$$

$H_K(x)$ being the Kth Hermite polynomial. In the Bargmann representation, the energy equation is given by

$$\left(z\frac{d}{dz}+\tfrac{1}{2}\right)f(z)=Ef(z) \tag{2.26}$$

with the eigenfunction $f_K(z)$ corresponding to the Kth energy level given by

$$f_K(z)=z^K. \tag{2.27}$$

It is seen that the eigenfunctions (2.27) are very much simpler than (2.25).[11,12] Indeed one of the reasons that the occupation-number representation is simpler to use than the Schrödinger representation is that one does not have to deal with eigenfunctions of the form (2.25) which are somewhat cumbersome, even though generally speaking the operators in the occupation-number representation are more abstract than the differential operators. The remarkable simplicity of the eigenfunctions (2.27) suggests that the Bargmann representation (2.23) might be more convenient to use in the study of some quantum system of bosons than would the occupation-number representation.

The generalization to the case of more than one mode is straightforward. Instead of functions of one variable, we have functions of several independent variables $z_1, z_2, \ldots$, where the subscripts refer to modes with frequencies $\nu_1, \nu_2, \ldots$, respectively.

3. SYSTEMS OF SPINS

Consider a single spin whose x, y, and z components are S^x, S^y, and S^z, respectively. Let the total angular momentum of the spin be $[S(S+1)]^{1/2}$, i.e., let the eigenvalues of S^z be given by $-S, -S+1, \ldots, S$. The S^+ and S^- operators are defined as usual by

$$S^+=S^x+iS^y \tag{3.1}$$

and

$$S^-=S^x-iS^y. \tag{3.2}$$

The basic commutation relations are

$$[S_i^z, S_j^\pm]=\pm\delta_{ij}S_i^\pm, \tag{3.3}$$

$$[S_i^+, S_j^-]=2\delta_{ij}S_i^z. \tag{3.4}$$

The state vectors $|S,p\rangle$ acted upon by S^+, S^-, or S^z are given by

$$S^+|S,p\rangle=\{(p+1)(2S-p)\}^{1/2}|S,p+1\rangle, \tag{3.5}$$

$$S^-|S,p\rangle=\{p(2S-p+1)\}^{1/2}|S,p-1\rangle, \tag{3.6}$$

$$S^z|S,p\rangle=(S-p)|S,p\rangle. \tag{3.7}$$

There exist spin states analogous to the Glauber's coherent states for the bosons. They are referred to as the Bloch coherent states[4,5] $|\mu\rangle$ defined by

$$|\mu\rangle=\frac{1}{(1+|\mu|^2)^S}\sum_{p=0}^{2S}\left(\frac{(2S)!}{p!(2S-p)!}\right)^{1/2}\mu^p|p\rangle \tag{3.8}$$

and

$$\langle\mu|=\frac{1}{(1+|\mu|^2)^S}\sum_{p=0}^{2S}\left(\frac{(2S)!}{p!(2S-p)!}\right)^{1/2}(\mu^*)^p\langle p| \tag{3.9}$$

where, for simplicity of notation, we have denoted the state vector $|S,p\rangle$ by $|p\rangle$. The various properties of the coherent states $|\mu\rangle$ were given by Radcliffe[4] and can be summarized as follows:

(i) The states $|\mu\rangle$, while not orthogonal, are complete, namely,

$$\frac{2S+1}{\pi}\int\frac{d^2\mu}{(1+|\mu|^2)^2}|\mu\rangle\langle\mu|=1 \tag{3.10}$$

where $d^2\mu$ means $d(\mathrm{Re}\,\mu)d(\mathrm{Im}\,\mu)$.

(ii) The scalar product of two coherent states is given by

$$\langle\lambda|\mu\rangle=\frac{(1+\lambda^*\mu)^{2S}}{(1+|\lambda|^2)^S(1+|\mu|^2)^S}. \tag{3.11}$$

The completeness relation enables us to express an arbitrary state $|E\rangle$ given, say, by

$$|E\rangle=\sum_{p=0}^{2S}c_p|p\rangle \tag{3.12}$$

in terms of the coherent states $|\mu\rangle$ by

$$|E\rangle=\frac{2S+1}{\pi}\int\frac{d^2\mu}{(1+|\mu|^2)^{S+2}}f(\mu^*)|\mu\rangle \tag{3.13}$$

where

$$f(\mu^*)=\sum_{p=0}^{2S}c_p(\mu^*)^p. \tag{3.14}$$

Now consider a Hamiltonian $H(S^z, S^+, S^-)$ in which the operators S^+ and S^- are normally ordered, and consider the eigenvalue equation

$$H(S^z, S^+, S^-)\,|E\rangle = E\,|E\rangle\,. \tag{3.15}$$

Multiply both sides of Eq. (3.15) by an arbitrary coherent state $\langle\lambda|$. On the right-hand side, we get

$$E\langle\lambda|E\rangle = [1/(1+|\lambda|^2)^S]Ef(\lambda^*) \tag{3.16}$$

by making use of the formula

$$\frac{2S+1}{\pi}\int \frac{d^2\mu}{(1+|\mu|^2)^{2S+2}}(\mu^*)^k\mu^m = \delta_{km}\frac{m!(2S-m)!}{(2S)!}\,. \tag{3.17}$$

On the left-hand side, a typical term is

$$\frac{2S+1}{\pi}\int \frac{d^2\mu}{(1+|\mu|^2)^{S+2}} f(\mu^*)\langle\lambda|(S^+)^m(S^-)^k\,|\mu\rangle\,. \tag{3.18}$$

But,

$$\begin{aligned}(S^-)^k|\mu\rangle &= \frac{1}{(1+|\mu|^2)^S}\sum_p\left\{\frac{(2S)!}{p!(2S-p)!}\,p(p-1)\cdots(p-k+1)\right.\\ &\quad\left.\times(2S-p+1)(2S-p+2)\cdots(2S-p+k)\right\}^{1/2}\mu^p|p-k\rangle\\ &= \frac{1}{(1+|\mu|^2)^S}\sum_q\left\{\frac{(2S)!}{q!(2S-q)!}\right\}^{1/2}\\ &\quad\times(2S-q)(2S-q-1)\cdots(2S-q-k+1)\mu^{q+k}|q\rangle\end{aligned} \tag{3.19}$$

and

$$\begin{aligned}\langle\lambda|(S^+)^m &= \frac{1}{(1+|\lambda|^2)^S}\sum_p\left\{\frac{(2S)!}{p!(2S-p)!}\right.\\ &\quad\times p(p-1)\cdots(p-m+1)(2S-p+1)(2S-p+2)\\ &\quad\left.\cdots(2S-p+m)\right\}^{1/2}(\lambda^*)^p\langle p-m|\\ &= \frac{1}{(1+|\lambda|^2)^S}\sum_q\left\{\frac{(2S)!}{q!(2S-q)!}\right\}^{1/2}(2S-q\rangle\\ &\quad\times(2S-q-1)\cdots(2S-q-m+1)(\lambda^*)^{q+m}\langle q|\,.\end{aligned} \tag{3.20}$$

Hence,

$$\begin{aligned}\langle\lambda|(S^+)^m(S^-)^k|\mu\rangle &= \frac{1}{(1+|\lambda|^2)^S(1+|\mu|^2)^S}\sum_n\\ &\quad\times\frac{(2S)!}{(n-k)!(2S-n+k)!}(2S-n+k)\\ &\quad\times(2S-n+k-1)\cdots(2S-n+k-m+1)\\ &\quad\times(2S-n+k)(2S-n+k-1)\\ &\quad\cdots(2S-n+1)(\lambda^*)^{n+m-k}\mu^n\end{aligned} \tag{3.21}$$

and

$$\begin{aligned}&\frac{2S+1}{\pi}\int\frac{d^2\mu}{(1+|\mu|^2)^2}\,f(\mu^*)\,\langle\lambda|(S^+)^m(S^-)^k|\mu\rangle\\ &= \frac{1}{(1+|\lambda|^2)^S}\sum_n c_n n(n-1)\cdots(n-k+1)(2S-n+k)\\ &\quad\times(2S-n+k-1)\cdots(2S-n+k-m+1)(\lambda^*)^{n+m-k}\\ &= \frac{1}{(1+|\lambda|^2)^S}\sum_n c_n\left(\lambda^*\frac{\partial}{\partial\zeta}\right)^m\left(\zeta\frac{\partial}{\partial\lambda^*}\right)^k\zeta^{2S-n}(\lambda^*)^n\Big|_{\zeta=1}\,.\end{aligned} \tag{3.22}$$

Equation (3.22) suggests a representation of the spin operators S^+ and S^- given by

$$S^+ \to z\frac{\partial}{\partial\zeta} \quad\text{and}\quad S^- \to \zeta\frac{\partial}{\partial z}\,. \tag{3.23a}$$

Considering in a similar way $\langle\lambda|(S^z)^m|\mu\rangle$, or by simply using the commutation relation (3.4), gives the corresponding representation for S^z:

$$S^z \to \tfrac{1}{2}\left(z\frac{\partial}{\partial z} - \zeta\frac{\partial}{\partial\zeta}\right)\,. \tag{3.23b}$$

It is easy to verify that the representation (3.23a), (3.23b) satisfies the commutation relations (3.3) and (3.4). We shall refer to (3.23a), (3.23b) as the Bargmann representation of the spin operators.

It follows that if we consider a function $f(\zeta, z)$ given by

$$f(\zeta, z) = \sum_n c_n\zeta^{2S-n}z^n\,, \tag{3.24}$$

the energy equation (3.15) is transformed into

$$H\left(\tfrac{1}{2}\left(z\frac{\partial}{\partial z} - \zeta\frac{\partial}{\partial\zeta}\right),\ z\frac{\partial}{\partial\zeta},\ \zeta\frac{\partial}{\partial z}\right)f(\zeta,z)\Big|_{\zeta=1} = Ef(\zeta,z)\Big|_{\zeta=1}\,. \tag{3.25}$$

The variable z is the main variable while ζ is what we shall call the companion variable as it is set equal to 1 after being operated on.

The advantage of the above representation over the traditional spin representation is again plainly clear. In the Bargmann representation, the eigenfunctions are simple power series of the form (3.24). In the traditional representation

$$\begin{aligned}S^x &\to y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\,,\\ S^y &\to z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}\,,\\ \text{and}\qquad S^z &\to x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\,,\end{aligned} \tag{3.26}$$

the eigenfunctions involve, in general, the product of some exponential with the associated Legendre polynomials which are somewhat cumbersome.

The eigenfunctions of S^z, namely, the solutions of

$$\tfrac{1}{2}\left(z\frac{\partial}{\partial z} - \zeta\frac{\partial}{\partial\zeta}\right)f(\zeta,z)\Big|_{\zeta=1} = Ef(\zeta,z)\Big|_{\zeta=1}\,, \tag{3.27}$$

are easily seen to be given by

$$f_K(\zeta,z) = \zeta^{2S-K}z^K \tag{3.28}$$

with

$$E_K = -S+K,\quad K=0,1,2,\ldots\,. \tag{3.29}$$

The generalization to the case of more than one spin is again simply done by affixing the subscripts $i=1,2,\ldots$ which refer to the spins, to the variables ζ and z.

4. PERTURBATION THEORY USING RECURRENCE RELATIONS

In this section, we present a perturbation theory using a recursive method. As pointed out in the Introduction, the recurrence relation method was used by Bender and Wu[9] and Schweber[11] in their study of harmonic and anharmonic oscillators. We shall describe the method in a more general setting here so that the generality as well as the limitation of the method will become clearer. We start with a given eigenvalue equation in the form of a differential equation such that the operator is a function of z_i and $\partial/\partial z_i$, $i=1,2,\ldots$, where z_i's are arbitrary complex variables, and the eigenfunctions are functions of z_i. Consider first the single variable case. Let the eigenvalue equation be

$$H\left(z,\frac{\partial}{\partial z}\right)\Phi(z)=E\Phi(z)\,. \tag{4.1}$$

Suppose that H can be divided into H_0, the unperturbed part, and λH_1, the perturbing part; i.e., if

$$H\left(z,\frac{\partial}{\partial z}\right)=H_0\left(z,\frac{\partial}{\partial z}\right)+\lambda H_1\left(z,\frac{\partial}{\partial z}\right), \tag{4.2}$$

then

$$\left\{H_0\left(z,\frac{\partial}{\partial z}\right)+\lambda H_1\left(z,\frac{\partial}{\partial z}\right)\right\}\Phi(\lambda,z)=E(\lambda)\Phi(\lambda,z)\,. \tag{4.3}$$

Let

$$E_n^K(\lambda)=\sum_{n=0}^{\infty}A_n^K\lambda^n \tag{4.4}$$

and

$$\Phi^K(\lambda,z)=\sum_{n=0}^{\infty}B_n^K(z)\lambda^n\,, \tag{4.5}$$

where K denotes, in the case H represents a Hamiltonian, the energy level. The superscript K if often omitted for convenience in the following discussions. Substituting (4.4) and (4.5) into (4.3), we get

$$\sum_{n=0}^{\infty}\lambda^n\left\{H_0\left(z,\frac{\partial}{\partial z}\right)B_n(z)+H_1\left(z,\frac{\partial}{\partial z}\right)B_{n-1}(z)\right\}$$
$$=\sum_{n=0}^{\infty}\lambda^n\sum_{p=0}^{n}A_pB_{n-p}(z)\,. \tag{4.6}$$

Let us write

$$\Phi(\lambda,z)=B_0(z)\sum_{n=0}^{\infty}\beta_n(z)\lambda^n \tag{4.6'}$$

so that $\beta_0(z)=1$. We now consider only the case that $B_0(z)$ is of the form z^K which, as indicated in the preceding sections, is the form of the eigenfunctions of the unperturbed Hamiltonian normally considered. If H_0 and H_1 are given by

$$H_0\left(z,\frac{\partial}{\partial z}\right)=\sum_{i=0}^{I}\sum_{j=0}^{J}f_{ij}z^i\left(\frac{\partial}{\partial z}\right)^j, \tag{4.7}$$

$$H_1\left(z,\frac{\partial}{\partial z}\right)=\sum_{p=0}^{P}\sum_{q=0}^{Q}g_{pq}z^p\left(\frac{\partial}{\partial z}\right)^q, \tag{4.8}$$

where I and J are the highest powers of z and $\partial/\partial z$ in H_0 for which $f_{ij}\neq 0$, and P and Q are the highest powers of z and $\partial/\partial z$ in H_1 for which $g_{pq}\neq 0$. Then $H_0(z,\partial/\partial z)B_0(z)\times\beta_n(z)$ and $H_1(z,\partial/\partial z)B_0(z)\beta_{n-1}(z)$ can be written in the form of

$$H_0\left(z,\frac{\partial}{\partial z}\right)B_0(z)\beta_n(z)=\beta_n(z)H_0\left(z,\frac{\partial}{\partial z}\right)B_0(z)+z^K\sum_{r=1}^{J}f_r(z)\beta_n^{(r)}(z), \tag{4.9}$$

where $\beta_n^{(r)}$ denotes the rth derivative of $\beta_n(z)$ with respect to z, and

$$H_1\left(z,\frac{\partial}{\partial z}\right)B_0(z)\beta_{n-1}(z)=z^K\sum_{r=0}^{Q}g_r(z)\beta_{n-1}^{(r)}(z)\,, \tag{4.10}$$

$f_r(z)$ and $g_r(z)$ being a finite linear combination of powers, positive or negative, of z. Substituting (4.9) and (4.10) into (4.6) and noting that

$$H_0\left(z,\frac{\partial}{\partial z}\right)B_0(z)=A_0B_0(z)\,, \tag{4.11}$$

we get

$$\sum_{n=1}^{\infty}\lambda^n\left\{\sum_{r=1}^{J}f_r(z)\beta_n^{(r)}(z)+\sum_{r=0}^{Q}g_r(z)\beta_{n-1}^{(r)}(z)\right\}=\sum_{n=1}^{\infty}\lambda^n\sum_{p=0}^{n-1}A_{n-p}\beta_p(z)\,. \tag{4.12}$$

If $K=0$, $f_r(z)$ and $g_r(z)$ contain no term involving negative powers of z, and the highest positive powers of z in $f_r(z)$ and $g_r(z)$ do not exceed I and P, respectively. If $K\neq 0$, $f_r(z)$ and $g_r(z)$ in general contain terms involving negative as well as positive powers of z. It is easy to show that the highest positive powers of z in $f_r(z)$ do not exceed $I+r$, the highest positive powers of z in $g_r(z)$ do not exceed $P+r$, the highest negative powers of z in $f_r(z)$ do not exceed $-J+r$ [i.e., the lowest powers of z in $f_r(z)$ are not smaller than $-J+r$] and the highest negative powers of z in $g_r(z)$ do not exceed $-Q+r$.

Now comes the crucial step in the recurrence relation method. Remembering that $\beta_0(z)=1$, we consider three cases separately:

Case 1: The powers of z in all $g_r(z)$ are all nonnegative (i.e., zero or positive). This is the case for (a) $K=0$ or (b) $K\neq 0$, $P>0$, $Q=0$ in (4.8). We assume, for $n\geq 1$,

$$\beta_n(z)=\sum_{j=1}^{Pn}b_{n,j}z^j\,. \tag{4.13}$$

Then, if the highest powers of z in $f_r(z)$ do not exceed r and the lowest powers of z in $f_r(z)$ are not smaller than -1, a substitution of (4.13) into (4.12) gives, on comparing the coefficients of z^j, $j=0,1,2,\ldots,Pn$, $Pn+1$ equations involving $Pn+1$ unknowns, $A_n, b_{n,1}, b_{n,2},\ldots, b_{n,Pn}$ which can therefore be determined uniquely in terms of the A's and b's of orders $<n$. The conditions on the highest and lowest powers of z in $f_r(z)$ are seen to be necessary because if, for example, the highest power of z in one of the $f_r(z)$ is greater than r, then we would have more equations than the number of unknowns which in general would give inconsistent results, and similarly if the lowest power of z in one or more of the $f_r(z)$ is smaller than -1.

Case 2: The powers of z in all $g_r(z)$ are all nonpositive (i.e., zero or negative). This is the case for $K\neq 0$, $P=0$, $Q>0$ in (4.8). In this case, we assume, for $n\geq 1$,

$$\beta_n(z)=\sum_{j=-Qn}^{-1}b_{n,j}z^j\,. \tag{4.14}$$

Then, assuming that $f_r(z)$ contains one and only one term

$c_r z^r$, where c_r is some constant, a substitution of (4.14) into (4.12) gives, on comparing the coefficients of z^j, $j=0,-1,-2,\ldots,-Qn$, $Qn+1$ equations involving $Qn+1$ unknowns $A_n, b_{n,-1}, b_{n,-2},\ldots,b_{n,-Qn}$ which can then be determined recursively in terms of the A's and b's of orders $<n$. The condition on the form of $f_r(z)$ is again necessary to avoid inconsistency.

Case 3: In the more general case in which $K\neq 0$, P and Q are both >0 in (4.8), there are positive as well as negative powers of z in the g_r's. In this case, we assume, for $n\geq 1$,

$$\beta_n(z)=\sum_{j=-Qn}^{Pn}{}' b_{n,j}z^j, \tag{4.15}$$

where the prime in the summation indicates that the term $j=0$ is excluded. Then assuming that $f_r(z)=c_r z^r$, c_r being some constant, a substitution of (4.15) into (4.12) gives, on comparing the coefficients of z^j, $j=-Qn,-Qn+1,\ldots,-1,0,1,\ldots,Pn$, $(P+Q)n+1$ equations involving $(P+Q)n+1$ unknowns, $b_{n,-Qn},\ldots,b_{n,-1}$, $A_n, b_{n,1},\ldots,b_{n,Pn}$ which can therefore be determined recursively in terms of the A's and b's of orders $<n$. The condition that $f_r(z)$ must be of the form $c_r z^r$ is equivalent to the condition that the unperturbed part $H_0(z,\partial/\partial z)$ must be of the form

$$H_0\left(z,\frac{\partial}{\partial z}\right)=\sum_i c_i z^i\left(\frac{\partial}{\partial z}\right)^i. \tag{4.16}$$

As can be seen from the preceding sections, the unperturbed Hamiltonians normally considered are happily precisely of this form in the Bargmann representation.

In the case of a spin, we have an additional variable ξ beside the variable z, but the analysis can be carried out in a very similar manner. We have H_0 and H_1 as functions of ξ, $\partial/\partial\xi$, z, and $\partial/\partial z$. Let

$$\Phi^K(\lambda,\xi,z)=\sum_{n=0}^{\infty}B_n^K(\xi,z)\lambda^n$$

and write

$$B_n^K(\xi,z)=B_0^K(\xi,z)\beta_n^K(\xi,z),$$

where $B_0^K(\xi,z)$ is $\xi^{2S-K}z^K$ and $\beta_0^K(\xi,z)=1$. Instead of Eq. (4.12), we have

$$\begin{aligned}\sum_{n=1}^{\infty}\lambda^n\Big\{&\sum_{r_1,r_2}f_{r_1,r_2}(\xi,z)\beta_n^{(r_1,r_2)}(\xi,z)+\sum_{r_1,r_2}g_{r_1,r_2}(\xi,z)\\&\times\beta_{n-1}^{(r_1,r_2)}(\xi,z)\Big\}\Big|_{\xi=1}\\&=\sum_{n=1}^{\infty}\lambda^n\sum_{p=0}^{n-1}A_{n-p}\beta_p(\xi,z)\Big|_{\xi=1},\end{aligned} \tag{4.17}$$

where

$$\beta^{(r_1,r_2)}(\xi,z)=\frac{\partial^{r_1+r_2}}{\partial\xi^{r_1}\partial z^{r_2}}\beta(\xi,z).$$

We assume, for $n\geq 1$,

$$\beta_n(\xi,z)=\sum_j{}' b_{n,j}\xi^{-j}z^j, \tag{4.18}$$

where the prime in the summation indicates that the term $j=0$ is excluded. In trying to determine the length of the power series representing $\beta_n(\xi,z)$, the companion variable ξ is ignored and only the highest and lowest powers of z in $g_{r_1,r_2}(\xi,z)$ are considered as in the previous case when z is the only variable. The conditions which must be placed on the forms of $f_{r_1,r_2}(\xi,z)$ in order that the recurrence relations give consistent results are also similarly obtained by considering the highest and lowest powers possible for z in $f_{r_1,r_2}(\xi,z)$. For example, if H_0 and H_1 are given by

$$H_0\left(\xi,\frac{\partial}{\partial\xi},z,\frac{\partial}{\partial z}\right)=\sum_{i=0}^{I}\sum_{j=0}^{J}f_{ij}\left(z\frac{\partial}{\partial z}\right)^i\left(\xi\frac{\partial}{\partial\xi}\right)^j, \tag{4.19}$$

$$H_1\left(\xi,\frac{\partial}{\partial\xi},z,\frac{\partial}{\partial z}\right)=\sum_{p=0}^{P}\sum_{q=0}^{Q}g_{pq}\left(z\frac{\partial}{\partial\xi}\right)^p\left(\xi\frac{\partial}{\partial z}\right)^q, \tag{4.20}$$

where I, J, P, Q are the highest powers of those factors in (4.19) and (4.20) for which f_{ij} and g_{pq} are $\neq 0$, then following the previous analysis, we have the following results:

Case 1: For (a) $K=0$ or (b) $K\neq 0$, $P>0$, $Q=0$ in (4.20), assume, for $n\geq 1$,

$$\beta_n(\xi,z)=\sum_{j=1}^{Pn}b_{n,j}\xi^{-j}z^j. \tag{4.21}$$

Case 2: For $K\neq 0$, $P=0$, $Q>0$ in (4.20), assume, for $n\geq 1$,

$$\beta_n(\xi,z)=\sum_{j=-Qn}^{-1}b_{n,j}\xi^{-j}z^j. \tag{4.22}$$

Case 3: For $K\neq 0$, P and Q are both >0 in (4.20), assume, for $n\geq 1$,

$$\beta_n(\xi,z)=\sum_{j=-Qn}^{Pn}{}' b_{n,j}\xi^{-j}z^j. \tag{4.23}$$

It is easy to show that the recurrence relations obtained by substituting (4.21), (4.22), or (4.23) into (4.17) work for any H_0 given in the form (4.19).

Once the principles and the limitations involved in the recurrence relation method are understood, the generalization to the case involving more than one variable is straightforward. Thus for the case of a system of boson oscillators with frequencies $\nu_1,\nu_2,\ldots,\nu_N$, we assume

$$\begin{aligned}\Phi^{\{K\}}(\lambda,z_1,z_2,\ldots,z_N)=B_0^{\{K\}}(z_1,z_2,\ldots,z_N)\sum_{n=0}^{\infty}\\\times\beta_n^{\{K\}}(z_1,z_2,\ldots,z_N)\lambda^n\end{aligned} \tag{4.24}$$

where $\{K\}$ denotes the set $\{K_1,K_2,\ldots,K_N\}$, with

$$\begin{aligned}&B_0^{\{K\}}(z_1,z_2,\ldots,z_N)=z_1^{K_1}z_2^{K_2}\cdots z_N^{K_N},\\&\beta_0^{\{K\}}(z_1,z_2,\ldots,z_N)=1,\\&\text{and}\\&\beta_n^{\{K\}}(z_1,z_2,\ldots,z_N)=\sum_{j_1,j_2,\ldots,j_N}{}'\\&\qquad\times b_{n,j_1,j_2,\ldots,j_N}^{\{K\}}z_1^{j_1}z_2^{j_2}\cdots z_N^{j_N}\end{aligned} \tag{4.25}$$

for $n\geq 1$. Similarly, for a system of N spins, we assume

$$\begin{aligned}\Phi^{\{K\}}(\lambda,\xi_1,z_1,\ldots,\xi_N,z_N)=B_0^{\{K\}}(\xi_1,z_1,\ldots,\xi_N,z_N)\\\times\sum_{n=0}^{\infty}\beta_n^{\{K\}}(\xi_1,z_1,\ldots,\xi_N,z_N)\lambda^n\end{aligned} \tag{4.26}$$

with

$$B_0^{\{K\}}(\xi_1,z_1,\ldots,\xi_N,z_N)=\xi_1^{2S-K_1}z_1^{K_1}\cdots\xi_N^{2S-K_N}z_N^{K_N},$$

$$\beta_n^{[K]}(\zeta_1, z_1, \ldots, \zeta_N, z_N) = \sum_{j_1, j_2, \ldots, j_N}{}' b^{[K]}_{n, j_1, j_2, \ldots, j_N} \times \zeta_1^{-j_1} z_1^{j_1} \cdots \zeta_N^{-j_N} z_N^{j_N} \quad (4.27)$$

for $n \geq 1$, and $\beta_0(\zeta_1, z_1, \ldots, \zeta_N, z_N) = 1$.

We are reminded that the energies $E^K(\lambda)$ and hence the coefficients A_n^K are of the greatest interest. The variables z's and ζ's on which the energies of course do not depend are merely convenient "tools" which we have selected for our purpose, and the set of numbers b's may be thought of merely as some numbers which need to be calculated at the intermediate stage of our calculation in order to arrive at the numbers A_n^K.

5. TWO SIMPLE EXAMPLES

In this section, we illustrate the recurrence relation method as well as the use of the Bargmann representation described in the preceding sections by two simple examples.

Example 1: Consider an anharmonic oscillator with quartic self-interaction. The Hamiltonian H is assumed to be (in the units $h=1$, $\nu=1$)

$$H = a^\dagger a + \tfrac{1}{2} + \tfrac{1}{4}\lambda(a^\dagger + a)^4 . \quad (5.1)$$

In the Bargmann representation, the energy equation is

$$\left[z\frac{d}{dz} + \tfrac{1}{2} + \tfrac{1}{4}\lambda\left(z + \frac{d}{dz}\right)^4\right] f(z) = Ef(z) . \quad (5.2)$$

The operators are assumed to be normally (Wick) ordered such that $(z + d/dz)^4$ means

$$\left(z + \frac{d}{dz}\right)^4 = z^4 + 4z^3\frac{d}{dz} + 6z^2\frac{d^2}{dz^2} + 4z\frac{d^3}{dz^3} + \frac{d^4}{dz^4} . \quad (5.3)$$

Let

$$E^K(\lambda) = \sum_{n=0}^{\infty} A_n^K \lambda^n \quad (5.4)$$

where

$$A_0^K = \tfrac{1}{2} + K , \quad (5.5)$$

and let

$$f_K(z) = B_0^K(z) \sum_{n=0}^{\infty} \beta_n^K(z)\lambda^n \quad (5.6)$$

where

$$B_0^K(z) = z^K , \quad (5.7)$$

$$\beta_0^K(z) = 1 , \quad (5.8)$$

and

$$\beta_n^K(z) = \sum_{j=-4n}^{4n}{}' b_{n,j}^K z^j \text{ for } n \geq 1 , \quad (5.9)$$

the prime in the summation indicating the exclusion of the term $j=0$. Substituting (5.4) and (5.6) into (5.2), we obtain (omitting the superscript K in the A's and b's for convenience),

$$A_0 = \tfrac{1}{2} + K ,$$

$$A_1 = \tfrac{3}{2}K(K-1) ,$$

and

$$b_{1,-3} = b_{1,-1} = b_{1,1} = b_{1,3} = 0 ,$$

$$b_{1,-4} = -\tfrac{1}{4}K(K-1)(K-2)(K-3) ,$$

$$b_{1,-2} = -K(K-1)(K-2) ,$$

$$b_{1,2} = -K ,$$

$$b_{1,4} = -\tfrac{1}{4} \quad (5.10)$$

and, for $n \geq 2$,

$$A_n = \tfrac{1}{4}\{b_{n-1,-4} + 4(K-2)b_{n-1,-2} + [4K(K-1)(K-2) + 24K(K-1) + 24K]b_{n-1,2} + [K(K-1)(K-2)(K-3) + 16K(K-1)(K-2) + 72K(K-1) + 96K + 24]b_{n-1,4}\} \quad (5.11)$$

and

$$jb_{n,j} + \tfrac{1}{4}\{b_{n-1,j-4} + 4(K+j-2)b_{n-1,j-2} + 6[K(K-1) + 2Kj + j(j-1)]b_{n-1,j} + 4[K(K-1)(K-2) + 3K(K-1)(j+2) + 3K(j+2)(j+1) + (j+2)(j+1)j]b_{n-1,j+2} + [K(K-1)(K-2)(K-3) + 4K(K-1)(K-2)(j+4) + 6K(K-1)(j+4)(j+3) + 4K(j+4)(j+3)(j+2) + (j+4)(j+3)(j+2)(j+1)] \times b_{n-1,j+4}\} = \sum_{p=1}^{n-1} A_{n-p} b_{p,j} . \quad (5.12)$$

The recurrence relations (5.11) and (5.12) together with (5.10) enable us to determine the coefficients A's and b's recursively up to any order. Thus, for $K=0$, $E^0(\lambda)$, the perturbed energy from the ground state, is given by

$$E^0(\lambda) = \tfrac{1}{2} - \tfrac{3}{8}\lambda^2 + \tfrac{27}{16}\lambda^3 + \cdots \quad (5.13)$$

which agrees with the result obtained by the diagram method or by the recurrence relation method using the Schrödinger representation given by Bender and Wu.[9]

Example 2: Consider a spin S having total spin $[S(S+1)]^{1/2}$ with quadratic self-interaction. The Hamiltonian is assumed to be

$$H = S^z + \lambda(S^+ + S^-)^2 , \quad (5.14)$$

where the operators S^+ and S^- are assumed to be normally ordered so that $(S^+ + S^-)^2$ means

$$(S^+ + S^-)^2 = (S^+)^2 + 2S^+S^- + (S^-)^2 . \quad (5.15)$$

In the Bargmann representation, the energy equation is given by

$$\left\{\tfrac{1}{2}\left(z\frac{\partial}{\partial z} - \zeta\frac{\partial}{\partial \zeta}\right) + \lambda\left(z^2\frac{\partial^2}{\partial \zeta^2} + 2z\frac{\partial}{\partial \zeta}\zeta\frac{\partial}{\partial z} + \zeta^2\frac{\partial^2}{\partial z^2}\right)\right\} \times f(\zeta, z)\big|_{\zeta=1} = Ef(\zeta, z)\big|_{\zeta=1} . \quad (5.16)$$

Let

$$E^K(\lambda) = \sum_{n=0}^{\infty} A_n^K \lambda^n \quad (5.17)$$

with

$$A_0^K = -S + K \quad (5.18)$$

and let

$$f_K(\zeta, z) = B_0^K(\zeta, z) \sum_{n=0}^{\infty} \beta_n^K(\zeta, z)\lambda^n \quad (5.19)$$

with

$$B_0^K(\zeta,z)=\zeta^{2S-K}z^K, \tag{5.20}$$

$$\beta_0^K(\zeta,z)=1, \tag{5.21}$$

and

$$\beta_n^K(\zeta,z)=\sum_{j=-2n}^{2n}{}' b_{n,j}^K\zeta^{-j}z^j \quad \text{for } n\geq 1, \tag{5.22}$$

where the prime in the summation indicates that the term $j=0$ is excluded. Substituting (5.17) and (5.19) into (5.16), we obtain

$$\begin{aligned} &A_0=-S+K,\\ &A_1=2K(2S-K+1),\\ &\text{and}\\ &b_{1,-1}=b_{1,1}=0,\\ &b_{1,-2}=\tfrac{1}{2}K(K-1),\\ &b_{1,2}=-\tfrac{1}{2}(2S-K)(2S-K-1), \end{aligned} \tag{5.23}$$

and, for $n\geq 2$,

$$A_n=(2S-K+1)(2S-K+2)b_{n-1,-2}+(K+1)(K+2)b_{n-1,2} \tag{5.24}$$

and

$$\begin{aligned} jb_{n,j}&+\{(2S-K)(2S-K-1)+2(2S-K)(-j+2)\\ &+(-j+2)(-j+1)\}b_{n-1,j-2}+\{2K(2S-K+1)-2Kj\\ &+2(2S-K+1)j-2j^2\}b_{n-1,j}+\{K(K-1)+2K(j+2)\\ &+(j+2)(j+1)\}b_{n-1,j+2}\\ &=\sum_{p=1}^{n-1}A_{n-p}b_{p,j}. \end{aligned} \tag{5.25}$$

The recurrence relations (5.24) and (5.25) together with (5.23) enable us to determine the coefficients A's and b's recursively up to any order. Thus, we find

$$\begin{aligned} A_0&=-S+K,\\ A_1&=2K(2S-K+1),\\ A_2&=\tfrac{1}{2}(2S-K+1)(2S-K+2)K(K-1)\\ &\quad-\tfrac{1}{2}(K+1)(K+2)(2S-K)(2S-K-1),\\ A_3&=-(2S-K+1)(2S-K+2)K(K-1)(2S-2K+3)\\ &\quad+(K+1)(K+2)(2S-K)(2S-K-1)(2S-2K-1),\\ A_4&=\tfrac{1}{16}(2S-K+1)(2S-K+2)K(K-1)\{(2S-K+3)\\ &\quad\times(2S-K+4)(K-2)(K-3)+32(2S-2K+3)^2\\ &\quad-2(2S-K+1)(2S-K+2)K(K-1)\\ &\quad+2(K+1)(K+2)(2S-K)(2S-K-1)\}\\ &\quad-\tfrac{1}{16}(K+1)(K+2)(2S-K)(2S-K-1)\{(K+3)(K+4)\\ &\quad\times(2S-K-2)(2S-K-3)+32(2S-2K-1)^2\\ &\quad+2(2S-K+1)(2S-K+2)K(K-1)\\ &\quad-2(K+1)(K+2)(2S-K)(2S-K-1)\}. \end{aligned} \tag{5.26}$$

For $K=0$, $E^0(\lambda)$, the perturbed energy from the ground state, is given by

$$\begin{aligned} E^0(\lambda)=&-S-2S(2S-1)\lambda^2+4S(2S-1)^2\lambda^3\\ &-2S(2S-1)(20S^2-30S+13)\lambda^4+\cdots. \end{aligned} \tag{5.27}$$

6. SUMMARY AND CONCLUSIONS

The Bargmann analytic function representation for the boson and spin operators has been shown useful and convenient for the study of some quantum systems. The representation enables us to formulate the recurrence relation method in perturbation theory in a general and systematic manner. The recurrence relation method offers an excellent alternative to the traditional diagram methods. We hope to apply the recurrence relation method to the study of a number of quantum systems involving one and several particles and we shall present the results in future publications.

ACKNOWLEDGMENT

I am very grateful to Professor Elliott Montroll for his interest and support.

*Research partially supported by Advanced Research Projects Agency and monitored by ONR under Contract No. N00014-67-0398-0005.

[1]F. T. Hioe, Phys. Rev. A8, 1440 (1973).
[2]F. T. Hioe, unpublished.
[3]R. Glauber, Phys. Rev. 131, 2766 (1963).
[4]J. M. Radcliffe, J. Phys. A (London) 4, 313 (1971).
[5]F. T. Arecchi *et al.*, Phys. Rev. A6, 2211 (1972).
[6]A. O. Barut and L. Girardello, Commun. Math. Phys. 21, 41 (1971).
[7]For more recent articles on the subject, see for example K. E. Cahill and R. J. Glauber, Phys. Rev. 177, 1857 (1969); G. S. Agarwal and E. Wolf, Phys. Rev. D 2, 2161 (1970); J. Kutzner, Phys. Lett. 41A, 475 (1972).
[8]E. Wigner, Phys. Rev. 40, 749 (1932).
[9]C. M. Bender and T. T. Wu, Phys. Rev. 184, 1231 (1969); Phys. Rev. D 7, 1620 (1973).
[10]R. H. G. Helleman and E. W. Montroll, preprint.
[11]S. Schweber, Ann. Phys. 41, 205 (1967).
[12]V. Bargmann, Comm. Pure Appl. Math. 20, 1 (1967).
[13]V. Fock, Z. Phys. 49, 339 (1928).
[14]B. Simon, J. Math. Phys. 12, 140 (1971).
[15]V. Bargmann, Rev. Mod. Phys. 34, 829 (1962).
[16]J. Schwinger, "On Angular Momentum," U. S Atomic Energy Commission, NYO-3071 (1952); reprinted in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. Van Dam (Academic, New York, 1965).

Note on gauge invariance and conservation laws for a class of nonlinear partial differential equations

Yoiti Watanabe

Department of Physics, College of General Education, Osaka University, Toyonaka, Osaka 560, Japan
(Received 13 December 1971; final manuscript received 6 September 1972)

Relations between gauge invariance and the conservation laws are discussed relative to the Korteweg–deVries and related equations. The technique used here is conventional in field theory—the canonical formalism related to the invariant variational problem. The function of constants of motion, when represented in terms of canonical variables, is to generate infinitesimal partial gauge transformations of any order, under which the variational problem is invariant. From this point of view, one example is presented together with the general theory.

1. INTRODUCTION

The equation derived from the invariant variation problem

$$\delta \int dt \bar{L}[\phi]=0, \tag{1}$$

$$\bar{L}[\phi]=\int dx\left[\frac{d\phi}{dx}\frac{d\phi}{dt}+\frac{1}{3}\left(\frac{d\phi}{dx}\right)^3-\left(\frac{d^2\phi}{dx^2}\right)^2\right] \tag{2}$$

is known as the Korteweg–deVries equation, which yields a number of conservation laws. Whitham investigated this kind of variational problem,[2] and Gardner later discussed this problem in the Hamiltonian form related to this Lagrangian for a limited case.[3]

The main advantages of the theory of invariant variation, however, exist in helping to demonstrate the relation between invariance and conservation laws.[4] For example, the Lagrangian (2) is Galilean-invariant. Invariance of this kind was discussed by Kruskal and his group.[5] Another invariance to be pointed out here is a kind of gauge invariance. The quantity $\phi(x,t)$ in (2) is unphysical but a "potential." That is, only its gradient, namely "strength,"

$$u=d\phi/dx \tag{3}$$

is "physical." This circumstance implies the gauge invariance of the variational problem just as in the Maxwell electromagnetic theory.[6]

The purpose of the work is to show the relation of the gauge invariance to such conservation laws contained in the KdV equation in the canonical frame.[7] The potential ϕ and the strength $d\phi/dx$ are treated as field quantities, here. Except for ϕ the quantities are assumed to vanish on the boundary of the space region, together with their independent small variations $\delta\phi$, $\delta(d\phi/dx)$, ... etc.

2. LAGRANGIAN AND HAMILTONIAN

The independent variational problem

$$\delta \int \bar{L}[\phi]dt=0, \tag{4}$$

$$\bar{L}[\phi]=\int dx\left[\frac{d\phi}{dx}\frac{d\phi}{dt}+\frac{1}{3}\left(\frac{d\phi}{dx}\right)^3+\frac{\alpha^2}{36}\left(\frac{d\phi}{dx}\right)^4-\left(\frac{d^2\phi}{dx^2}\right)^2\right], \tag{5}$$

with respect to $\phi(x,t)$ under fixed boundary values, leads to

$$\frac{d}{dt}\left(\frac{d\phi}{dx}\right)+\left[\frac{d\phi}{dx}+\frac{1}{6}\alpha^2\left(\frac{d\phi}{dx}\right)^2\right]\frac{d^2\phi}{dx^2}+\frac{d^4\phi}{dx^4}=0, \tag{6}$$

which is taken from the work of Miura and his group.[8] (This equation tends to the KdV equation as $\alpha\to 0$.) The variational problem (4) and the equation are both invariant under the gauge transformation

$$\phi\to\phi(x,t)+\Lambda(t), \tag{7}$$

where $\Lambda(t)$ stands for an arbitrary function of t, independent of x. In fact, (7) adds on the right-hand side of (5) a term, dependent only on the boundary values of ϕ, irrelevant for the variation problem.

Because of the linear dependence of $\bar{L}[\phi]$ on $d\phi/dt$, the Legendre transformation $d\phi/dt$ to p (the canonical momentum conjugate to ϕ) is not unique. The situation yields many possible Hamiltonians. They are formally different from each other, and have different functions in the framework of canonical theory, although they give equal values of "energy" under a certain compatibility condition as we shall see. A choice for the Hamiltonian is

$$\bar{H}[p,\phi]=\int dx H, \tag{8}$$

$$\bar{H}=-\left(\frac{1}{2}p^2+\frac{\alpha^2}{18}p^3+\frac{d^2p}{dx^2}\right)\frac{d\phi}{dx}+\frac{1}{6}p^3+\frac{\alpha^2}{36}p^4\ , \tag{9}$$

(see Appendix I), this leads to the proper forms of canonical field equations,

$$\frac{dp}{dt}=-\frac{\delta\bar{H}}{\delta\phi}=-\frac{d}{dx}\left(\frac{1}{2}p^2+\frac{\alpha^2}{18}p^3+\frac{d^2p}{dx^2}\right), \tag{10}$$

$$\frac{d\phi}{dt}=\frac{\delta\bar{H}}{\delta p}=-\left(p+\frac{\alpha^2}{6}p^2\right)\frac{d\phi}{dx}-\frac{d^3\phi}{dx^3}+\frac{1}{2}p^2+\frac{\alpha^2}{9}p^3. \tag{11}$$

Equation (10) is essentially identified with the KdV equation at $\alpha=0$, while Eq. (11), after x-differentiation, gives

$$\frac{d}{dt}\left(p-\frac{d\phi}{dx}\right)+\frac{d}{dx}\left[\left(p+\frac{\alpha^2}{6}p^2+\frac{d^2}{dx^2}\right)\left(p-\frac{d\phi}{dx}\right)\right]=0, \tag{12}$$

with the aid of (10). Thus, the quantity $[p-(d\phi/dx)]$ itself is a conserved density of Eqs. (10), (11), but, if we set at some time, (e.g., $t=0$)

$$p-d\phi/dx=0, \tag{13}$$

then (13) holds forever. It seems that the choice of the Hamiltonian is different from that of Gardner. The functions of these Hamiltonians are different, while they have equal values of the "energy" under the condition (13).

3. POISSON BRACKET

The poisson bracket, defined by

$$[\bar{F},\bar{G}]=\int dx\left(\frac{\delta\bar{F}}{\delta p(x)}\cdot\frac{\delta\bar{G}}{\delta\phi(x)}-\frac{\delta\bar{G}}{\delta p(x)}\cdot\frac{\delta\bar{F}}{\delta\phi(x)}\right) \tag{14}$$

leads to

$$[p(x,t),\phi(x',t)]=\delta(x-x'),$$
$$[p(x,t),p(x',t)]=[\phi(x,t),\ \phi(x',t)]=0. \tag{15}$$

Also, we have

$$\frac{dp}{dt}=-[p(x,t),\bar{H}],$$
$$\frac{d\phi}{dt}=-[\phi(x,t),\bar{H}], \tag{16}$$

the canonical equations in the P.B. form, just as in the usual analytical dynamics. For any well-defined functional $\bar{G}[p,\phi]$,

$$\frac{d\bar{G}}{dt}=-[\bar{G},\bar{H}] \tag{17}$$

gives the time development of the dynamical variable $\bar{G}$, and implies that $\bar{G}$ is a constant of motion if $\bar{G}$ "commutes" with $\bar{H}$. Of course, other properties of the Poisson bracket such as Jacobi's identity also hold in our case.

A dynamical variable $\bar{F}[p,\phi]$ is replaced by an appropriate functional $\bar{G}[p,\phi]$ according to

$$\delta\bar{F}=\epsilon[\bar{F},\bar{G}], \tag{18}$$

where ϵ stands for an infinitesimal parameter of continuous transformation (see Appendix II). Equation (17) implies that $\bar{F}[p,\phi]$ is invariant under the $\bar{G}$-transformation, if $\bar{F}$ commutes with $\bar{G}$, and vice versa. The functional containing many linearly independent parameters α_n,

$$\bar{G}[p,\phi]=\sum\alpha_n\bar{G}_n[p,\phi] \tag{19}$$

plays a similar rôle for a certain displacement of the variables. In particular, any $\bar{G}_n[p,\phi]$ makes a "partial" displacement on the variables, if all $\bar{G}_n$'s are compatible:

$$[\bar{G}_m,\bar{G}_n]=0,\quad \text{for all } n,\ m \tag{20}$$

and then $[\bar{G}_n,H]=0$, if $[\bar{G},\bar{H}]=0$. In other words, $\bar{H}$ is also invariant under the partial displacement.

4. GAUGE TRANSFORMATION

The functional

$$\epsilon\lambda(t)\bar{p}=\epsilon\lambda(t)\int dx\,p(x,t) \tag{21}$$

generates the infinitesimal gauge transformation on $\phi(x,t)$:

$$\delta\phi=\epsilon\lambda(t)[\phi(x,t),\bar{p}]=\epsilon\lambda(t). \tag{22}$$

The Hamiltonian (8) is evidently gauge-invariant. This invariance implies

$$[\bar{p},\bar{H}]=0, \tag{23}$$

i.e., $\bar{p}$ is a constant of motion.

Now, consider a generator

$$\bar{\Phi}[p,\phi']=\int dx\left(p+i\alpha\frac{dp}{dx}+\frac{1}{6}\alpha^2p^2\right)\phi'. \tag{24}$$

The old pair (p,ϕ) turns into the new pair (p',ϕ') by the time-independent, finite canonical transformation generated by $\bar{\Phi}$;

$$p'(x,t)=p(x,t)+i\alpha\frac{d}{dx}p(x,t)+\frac{1}{6}\alpha^2p(x,t)^2,$$
$$\phi(x,t)=\left(1-i\alpha\frac{d}{dx}+\frac{1}{3}\alpha^2p(x,t)\right)\phi'(x,t). \tag{25}$$

The new momentum p' satisfies the proper KdV equation,[8] i.e., $\alpha=0$. Thus, we can regard this canonical transformation as the one to make α vanish. In terms of the new variables, consequently, the Hamiltonian is written as

$$\bar{H}'[p',\phi']=\int dxH',$$
$$H'=-\left(\frac{1}{2}p'^2+\frac{d^2p'}{dx^2}\right)\frac{d\phi'}{dx}+\frac{1}{6}p'^3. \tag{26}$$

The new momentum $\bar{p}'$ can be considered as the generator of an infinitesimal gauge transformation on the new ϕ', since P.B. relations do not change under canonical transformations. Moreover, time-independent canonical transformations do not change the canonical equations. This fact yields

$$\frac{d\bar{p}}{dt}=0=[\bar{p},\bar{H}'[p',\phi']]. \tag{27}$$

The old momentum p has been formally found in terms of p' stepwisely

$$p=p'-i\alpha\frac{dp'}{dx}+\sum\alpha^nT_n\left(p',\frac{dp'}{dx},\ldots\right). \tag{28}$$

The densities $T_n(p',dp'/dx,\ldots)$ are given as linearly independent polynomials (see Ref. 8). Thus, (27) and (28) lead to

$$\frac{d\bar{T}_n}{dt}=[\bar{T}_n,\bar{H}']=0. \tag{29}$$

It is essential that all of the $T_n[p']$ are the functionals of p' only, not dependent on ϕ'. It implies the compatibility relation

$$[\bar{T}_n[p'],\bar{T}_m[p']]=0\quad\text{for all } n,\ m \tag{30}$$

to be true. Since $\bar{p}$ generates an infinitesimal gauge transformation on ϕ, $\bar{T}_n[p']$ generates the infinitesimal partial gauge transformation of order n on $\phi(\phi',p')$. The Hamiltonian is, therefore, invariant under partial gauge transformations of any order. This invariance thus yields many constants of motion of the KdV equation.

The parameter α, running over all real nonnegative values, gives a class of equations, any element of which turns into another by means of a canonical transformation. This class is hereafter called "KdV class" for short, since the class contains the KdV equation as its element $\alpha=0$.

5. GENERAL REMARKS

The Lagrangian functional of general form

$$\bar{L}[\phi]=\int dx\,\frac{d\phi}{dx}\frac{d\phi}{dt}-\bar{F}\left[\frac{d\phi}{dx}\right], \tag{31}$$[10]

where

$$\bar{F}\left[\frac{d\phi}{dx}\right]=\int dx\,F\left(\frac{d\phi}{dx},\ \frac{d^2\phi}{dx^2},\ \ldots\right), \tag{32}$$

contains (2) and (5) as special cases. The variation problem of (31) is certainly invariant under the transformation (7); linear dependence of (31) on $d\phi/dt$ leads to nonuniqueness of the Legendre transformation, and consequently to many possibilities for the Hamiltonian formalism. As is seen in Appendix I, however, the Hamiltonian

$$\bar{H}[p,\phi]=\bar{F}[p]-\frac{1}{2}\int dx\,\frac{\delta\bar{F}}{\delta p}\left(p-\frac{d\phi}{dx}\right) \tag{33}$$

is a reasonable choice, and (33) leads to canonical equations

$$\frac{dp}{dt}=-\frac{\delta\bar{H}}{\delta\phi}=\frac{1}{2}\frac{d}{dx}\frac{\delta\bar{F}}{\delta p}, \tag{34}$$

$$\frac{d\phi}{dt}=\frac{\delta\bar{H}}{\delta p}=-\frac{\delta^2\bar{F}}{\delta p^2}\left(p-\frac{d\phi}{dx}\right)+\frac{1}{2}\frac{\delta\bar{F}}{\delta p}. \tag{35}$$[11]

Equation (35) is equivalent to Eq. (34) only under the condition $p=d\phi/dx$, which should be given as an initial condition. On the other hand, Eq. (34) is formally similar to Gardner's representation of the KdV equation [cf. Ref. 3, Eq. (5) and Eq. (31)]. However, the concrete functional form of $\bar{F}[p]$ now remains free. Thus, the gauge invariant variation problem with respect to (31), in general, yields the Gardner-type equation (34), which naturally implies $[\bar{H},\bar{p}]=0$.

Next, consider a one-parametric canonical transformation by the generator

$$\bar{\Phi}[p,\phi']=\int dx\,\phi' S\left(p,\ \frac{dp}{dx},\ \ldots;\alpha\right), \tag{36}$$

where α stands for the parameter of the transformation, and

$$p=S\left(p,\ \frac{dp}{dx},\ \ldots;\ 0\right) \tag{37}$$

is assumed for simplicity. Evidently this transformation

$$p'=\frac{\delta\bar{\Phi}}{\delta\phi'}=S\left(p,\ \frac{dp}{dx},\ \ldots;\ \alpha\right) \tag{38}$$

preserves the gauge properties. In terms of new canonical variables (p',ϕ'), therefore, another Gardner-type canonical equation

$$\frac{dp'}{dt}=\frac{1}{2}\frac{d}{dx}\frac{\delta\bar{F}'}{\delta p'} \tag{39}$$

results, where the identity

$$\bar{F}'[p']=\bar{F}[p] \tag{40}$$

defines a new $\bar{F}'[p']$. So, Eqs. (34) and (39) imply, respectively,

$$\frac{d\bar{p}}{dt}=0,\qquad \frac{d\bar{p}'}{dt}=0. \tag{41}$$

On the other hand, p can be expressed in terms of p' in the formal power series in α:

$$p=p'+\sum\alpha^n A_n\left(p',\ \frac{dp'}{dx},\ \ldots\right) \tag{42}$$

by solving (38). It is concluded that

$$\frac{d}{dt}A_n[p']=0 \tag{43}$$

holds for all n, because (34), (39), and (41) must be correct for all values of α. In other words, every A_n is a conserved density of Eq. (39).

The compatibility condition

$$[\bar{A}_n,\bar{A}_m]=0$$

is also a triviality, and the variables $\bar{p}$ and $\bar{p}'$ gives rise to the infinitesimal gauge transformations on ϕ and ϕ', respectively. So, every $\bar{A}_n[p']$ gives rise to a partial gauge transformation on ϕ'.

6. ANOTHER EXAMPLE

The KdV class is one example. We now want to show another. Let us look for the properties of a class of nonlinear equations,

$$\frac{du}{dt}-\exp\left(2i\alpha\frac{du}{dx}\right)\left\{\frac{du}{dx}+2i\alpha\left[(1-u)\frac{d^2u}{dx^2}-\left(\frac{du}{dx}\right)^2\right]-\alpha^2\left(u\frac{du}{dx}\frac{d^2u}{dx^2}+\frac{d^3u}{dx^3}\right)+4i\alpha^3\left(\frac{d^2u}{dx^2}\right)^2\right\}=0, \tag{44}$$

where α runs over all real values. Equation (44) is of course an artificial one, and appears somewhat complicated, but it is employed as a simple example in investigating properties similar to those found in the KdV class. Equation (44) is derived from the invariant variation problem of

$$\bar{L}[\phi]=\int dx\left[\frac{d\phi}{dx}\cdot\frac{d\phi}{dt}-\frac{d\phi}{dx}\exp\left(2i\alpha\frac{d^2\phi}{dx^2}\right)\right], \tag{45}$$

where $u=d\phi/dx$. Consequently, by putting

$$\bar{F}[p]=\int dx\,p^2\exp\left(2i\alpha\frac{dp}{dx}\right) \tag{46}$$

in the Gardner-type equation (34), we get Eq. (44) (there u must be replaced by p).

Next, consider the canonical transformation

$$p'=\delta\bar{\Phi}/\delta\phi'=p\exp\left(i\alpha\frac{dp}{dx}\right), \tag{47}$$

$$\phi=\delta\bar{\Phi}/\delta p, \tag{48}$$

generated from

$$\bar{\Phi}[p,\phi']=\int dx\phi' p\exp\left(i\alpha\frac{dp}{dx}\right). \tag{49}$$

The second expression (48) is no longer given in detail, because it is less important. After the transformation (47) $\bar{F}[p]$ turns into

$$\bar{F}[p']=\int dx\,p'^2, \tag{50}$$

and the equation turns into

$$\frac{dp'}{dt}=\frac{dp'}{dx} \tag{51}$$

which implies

$$\frac{d\bar{p}'}{dt}=0 \tag{52}$$

for all real values of α. The old momentum p can be expressed in terms of the new momentum p' in the formal power series in α:

$$p=p'+\alpha A_1+\alpha^2 A_2+\cdots, \tag{53}$$

where

$$A_1=-i\frac{dp'}{dx}\frac{d^2p'}{dx^2},$$

$$A_2=\frac{1}{2}p'\left(\frac{dp'}{dx}\right)^2+\frac{1}{2}\frac{d^2}{dx^2}\left[p'\left(\frac{dp'}{dx}\right)^2\right],\cdots. \tag{54}$$

The conservation property of A_1 is trivial; i.e., $d\bar{A}_1/dt=0$ holds without Eq. (51), while that of A_2 is essential; $d\bar{A}_2/dt=0$ holds if and only if p' is a solution of Eq. (51). Thus, we have two explicit examples to show the relation between gauge invariance and the conservation laws of a class of nonlinear equations; one the KdV class, the other that shown above.

7. CONCLUDING REMARKS

The conclusion of this work is that *the existence of a number of conserved densities originates in the gauge invariance of the variational problem.*

The gauge transformation (7) contains an arbitrary function $\Lambda(t)$. According to Noether, if any transformation, under which the variational problem remains invariant, possesses r parameters, then the Euler equations have r conserved densities. An arbitrary function, as in our case, behaves as an infinite number of parameters, and correspondingly results in the derivation of many conserved densities.

It is possible to construct more examples arbitrarily by employing appropirate specific $\bar{F}[p]$. Moreover, similar situations are expected when the field variable ϕ has a number of components. That may motivate wider examinations for classes of nonlinear differential equation.

APPENDIX A: MANY TYPES OF HAMILTONIAN

Two propositions from classical variation calculus are mainly relied on in the following arguments:

(i) The variational problem remains unaltered after the results of the Euler equations are partially introduced into the original problem.

(ii) The relations among variables, if any, must be considered as subsidiary conditions, and can be taken into the calculation by means of indefinite multipliers.

Let us change $d\phi/dt$ to $\dot{\phi}$ in (31), and consider $d\phi/dt=\dot{\phi}$ as a subsidiary condition on the independent variables ϕ and $\dot{\phi}$. By using a multiplier p, the variational problem turns into

$$\delta\int dt\left\{\int dx\left(\frac{d\phi}{dx}\dot{\phi}\right)-\bar{F}\left[\frac{d\phi}{dx}\right]-\int dx\,p\left(\dot{\phi}-\frac{d\phi}{dt}\right)\right\}=0, \tag{A1}$$

which is performed with respect to ϕ, $\dot{\phi}$, and p independently. One of the Euler equations derived from (A1)

$$p=\frac{\delta\bar{L}}{\delta\dot{\phi}}=\frac{d\phi}{dx}, \tag{A2}$$

usually regarded as the Legendre transformation, $(\phi,\dot{\phi})$ to (ϕ,p), does not contain $\dot{\phi}$ in this case. So, the equation, an imposed relation between p and ϕ, must be considered as a new subsidiary condition.

The lack of relation p to $\dot{\phi}$ leads to a multiplicity of the related Hamiltonians. One of the tentative choices among them is

$$\bar{H}[p,\phi,v]=\int dx\frac{\delta\bar{L}}{\delta\dot{\phi}}\dot{\phi}-\bar{L}\left[\frac{d\phi}{dx}\right]+\int dx\,v\left(p-\frac{d\phi}{dx}\right)$$

$$=\bar{F}\left[\frac{d\phi}{dx}\right]+\int dx\,v\left(p-\frac{d\phi}{dx}\right), \tag{A3}$$

where v stand for the indefinite multiplier. This Hamiltonian yields $p=d\phi/dx$ as one of the Euler equations, which may be put into $\bar{F}$. Another tentative Hamiltonian is, then

$$\bar{H}[p,\phi,v]=\bar{F}[p]+\int dx\,v\left(p-\frac{d\phi}{dx}\right), \tag{A3}$$

the last term of which is necessary to derive that condition.

In the next step, what we have to do is to eliminate v so as to get the proper form of $\bar{H}$, which must be alein to the additional variable other than the canonical variables (p,ϕ). The Euler equations from (A3) are

$$\frac{dp}{dt}=-\frac{\delta\bar{H}}{\delta\phi}=-\frac{dv}{dx},$$

$$\frac{d\phi}{dt}=\frac{\delta\bar{H}}{\delta p}=\frac{\delta\bar{F}}{\delta p}+v,$$

$$p=d\phi/dx, \tag{A4}$$

which yield

$$v=-\frac{1}{2}\frac{\delta\bar{F}}{\delta p}. \tag{A5}$$

By putting (A5) into (A3), we obtain

$$\bar{H}[p,\phi]=\bar{F}[p]-\frac{1}{2}\int\frac{\delta\bar{F}}{\delta p}\left(p-\frac{d\phi}{dx}\right)dx, \tag{A6}$$

which leads to the canonical equation

$$\frac{dp}{dt}=\frac{1}{2}\frac{d}{dx}\frac{\delta\bar{F}}{\delta p},$$

$$\frac{d\phi}{dt}=-\frac{1}{2}\left[\frac{\delta^2\bar{F}}{\delta p^2}\left(p-\frac{d\phi}{dx}\right)\right]+\frac{1}{2}\frac{\delta\bar{F}}{\delta p}. \tag{A7}$$

The first equation has obviously the general Gardner-type, while the second together with the first is changed into

$$\frac{d}{dt}\left(p-\frac{d\phi}{dx}\right)-\frac{1}{2}\frac{d}{dx}\left[\frac{\delta^2\bar{F}}{\delta p^2}\left(p-\frac{d\phi}{dx}\right)\right]=0, \tag{A8}$$

where by the notation $\delta^2\bar{F}/\delta p^2$ we understand an operator on the immediate right member: e.g., if $F=\frac{1}{2}(dp/dx)^2$, $\delta^2\bar{F}/\delta p^2$ functions as the differential operator d^2/dx^2. Equation (A8) implies that $[p-(d\phi/dx)]$ itself is a conserved density; if it is set equal to zero at an initial moment, then it vanishes forever. This statement is

never a triviality, although $p = d\phi/dx$ is partially used in going from the Lagrangian to the Hamiltonian formalism, since p and ϕ have to be treated as independent variables in the canonical framework.

Thus, by (A6) we understand the terminal choice of Hamiltonian; which must be alien to v, and yields canonical equations equivalent to the original equation if $p - (d\phi/dx) = 0$ is given as an initial condition.

APPENDIX B. CANONICAL TRANSFORMATION

A canonical transformation $(p, \phi) \rightleftharpoons (p', \phi')$ must be obtained from a functional $\overline{W}[p, \phi']$ by

$$p' = \frac{\delta \overline{W}}{\delta \phi'(x, t)}, \qquad \phi = \frac{\delta \overline{W}}{\delta p(x, t)}. \tag{B1}$$

Clearly the transformation (B1) does not alter the Poisson bracket (14); and also it does not change the canonical equations of motion if $\overline{W}$ does not explicitly contain t.

In particular,

$$\overline{W}[p, \phi'] = \int dx\, p\phi' + \epsilon \overline{G}[p. \phi'], \quad \epsilon\text{: infinitesimal} \tag{B2}$$

gives rise to the infinitesimal displacement on p and ϕ;

$$\delta\phi = \phi' - \phi = \epsilon[\phi, \overline{G}[p, \phi]], \qquad \delta p = p' - p = \epsilon[p, \overline{G}[p. \phi]], \tag{B3}$$

up to the first order in ϵ. Note that ϕ' in the last members is replaced by ϕ. That is, a functional $\overline{G}[p, \phi]$ plays the rôle of the infinitesimal displacement operator on p and ϕ in the sense of P.B.

[1]Throughout this article, the differential operators with respect to space–time are denoted by d/dx, d/dt instead of $\partial/\partial x$, $\partial/\partial t$, while ones with respect to the "quantities" and their derivatives are denoted by $(\partial/\partial\phi)$ $(\partial/\partial\phi_x)$...etc. The functional defined by the "density" $G(\phi, d\phi/dx, d^2\phi/dx^2, ...)$ is $\bar{G}[\phi] = \int dx\ G(\phi, d\phi/dx, d^2\phi/dx^2 ...)$. In particular, a point function can be regarded as a functional when it is represented in terms of Dirac's delta function as $\phi(x) = \int dx'\phi(x')\delta(x' - x)$, $d\phi/dx = -\int dx'\phi(x')\delta'(x' - x)$, where $\delta'(x)$ stands for $d\delta/dx$.

[2]G. B. Whitham, Proc. R. Soc. A **283,** 238 (1965).
[3]C. S. Gardner, J. Math. Phys. **12,** 1548 (1971).
[4]E. Noether, Nachr. Akad. Wiss. Goett. Math.-Phys. 2, 235 (1918).
[5]M. D. Kruskal, R. M. Miura, C. S. Gardner, and N. J. Zabusky, J. Math. Phys. **11,** 952 (1970).
[6]See, for example, W. Heisenberg and W. Pauli, Z. Phys. **56,** 1 (1927).
[7]See, for example, P. A. M. Dirac, Can. J. Phys. **2,** 129 (1951).
[8]R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9,** 1204 (1968).
[9]The functional derivatives are defined by $\delta\bar{G}/\delta\phi = (\partial G/\partial\phi) - (d/dx)(\partial G/\partial\phi_x) + (d^2/dx^2)(\partial G/\partial\phi_{xx}) - ...$ (see Ref. 6).
[10]In general, the Lagrangian may be in the form $G(d\phi/dx, d^2\phi/dx^2, ...)\ (d\phi/dt) - F[d\phi/dx]$, which gives a gauge-invariant variational problem, but this general form can be modified to (33) by means of an appropriate canonical transformation.
[11]$\delta^2\overline{F}/\delta p^2$ in (35) must be regarded as an operator on the term immediately following it (see Appendix A).

Two-particle approximation for the quantum third virial coefficient: Comparison of two approaches*

R. E. Johnson, N. K. Pope, and A. J. Barrett

Department of Mathematics, Royal Military College of Canada, Kingston, Ontario, Canada
(Received 5 November 1973)

Two distinct approaches have been followed to produce an approximate expression for the third virial coefficient of a quantum system in terms of the pertinent two-particle quantities (standard T matrix elements and bound state eigenvalues and eigenstates): One approach is based on the T *approximation* for the self-energy in the theory of temperature Green functions, and the other uses the diagrams which represent the various terms in the multiple-scattering expansion of the three-particle T operator. It is shown that the result given by the former method is reproduced exactly by the contribution of a small number of selected diagrams. The Green function method requires the evaluation of the Kadanoff and Baym generalized T matrix to one order in the activity higher than its familiar low-density limit. The common result can be written entirely in terms of the lowest order self-energy and should give a reasonable approximation to the third virial coefficient for systems with short-range forces.

1. INTRODUCTION

The purpose of this paper is to study and compare two methods by which the third virial coefficient of a quantum gas can be approximated by expressions involving the two-particle scattering amplitude or T matrix.

The first method is the well-known T *approximation* according to the Green function formalism introduced by Martin and Schwinger.[1] We follow the development of Kadanoff and Baym[2] which involves a generalized two-particle T matrix including many-body effects. This procedure gives a closed expression for the third virial coefficient in terms of the standard two-particle T matrices. This result is developed in Sec. 2.

The second method involves the use of formal operator techniques[3] and diagram expansions[4] for the particle density. This approach yields an infinite set of terms which contain two-particle T matrices; the total contribution of these terms must, of course, be equivalent to that of the three-particle T matrix. We present the results for certain selected diagrams in Sec. 3.

In Sec. 4 we show that the contribution of a small number of diagrams is identical to the result obtained by the Green function method. In this manner we obtain an expression for the third virial coefficient of a quantum gas whose origin is understood and which is relatively simple but contains all of the contributions which have been explicitly calculated in terms of T matrices previously[5,6,7]; this is discussed in Sec. 5. For a concise summary of the history of the theory of the quantum-mechanical third virial coefficient the introduction in the paper by Larsen and Mascheroni[8] is recommended.

We consider a dilute quantum gas with spinless particles of mass m which interact through two-body forces. The Hamiltonian of the system will be written

$$H=H_0+H_1=\sum_i K_i+\sum_{i<j} V_{ij}, \tag{1.1}$$

where K represents the one-particle kinetic energy operator and V the two-particle interaction operator. The two-body system with Hamiltonian $H_0^{(2)}+V$ may have bound states.

For a system having inverse temperature β and chemical potential μ we define $\lambda=(2\pi\beta/m)^{1/2}$ and the activity or fugacity $z=e^{\beta\mu}$. (We set $\hbar=1$ throughout.) For a dilute gas $\beta\mu\ll -1$, and the density of the system can be written

$$n=n_1z+n_2z^2+n_3z^3+\cdots, \tag{1.2}$$

where $n_1=\lambda^{-3}$. The second and third virial coefficients of the system are then given by

$$B_2=-\tfrac{1}{2}\lambda^6 n_2 \tag{1.3}$$

and

$$B_3=4B_2^2-\tfrac{2}{3}\lambda^9 n_3, \tag{1.4}$$

respectively. Since B_2 is well known (and will be rederived in the course of our discussion), the problem of computing B_3 is equivalent to that of obtaining the density coefficient n_3.

The standard two-particle T operator is defined by

$$T_0(\omega)=V+V(\omega-H_0^{(2)})^{-1}T_0(\omega), \tag{1.5}$$

where ω is a complex variable; n_2 can be expressed simply in terms of matrix elements of T_0, and in this paper we develop an approximate expression for n_3 in terms of such matrix elements. (In fact, we will show that n_3 can be written concisely in terms of the lowest order self-energy function.)

2. GREEN FUNCTIONS AND THE T APPROXIMATION FOR THE DENSITY

A. Review of the theory

The Green functions for our many-body system are defined by

$$G_n(1\cdots n,1'\cdots n')=(-i)^n\,\mathrm{Tr}\{e^{-\beta(H-\mu\hat{N})}T[\psi(1)\cdots\psi(n)\psi^\dagger(n')\cdots\psi^\dagger(1')]\}/\mathrm{Tr}\,e^{-\beta(H-\mu\hat{N})}, \tag{2.1}$$

where $n=1,2,3,\ldots$, $\hat{N}$ is the number operator, $\psi(1)$ and $\psi^\dagger(1)$ are either Bose or Fermi field operators in the Heisenberg representation (1 stands for $\mathbf{r}_1,t_1$), and T is the Wick time-ordering operator. In statistical mechanics one considers imaginary "times" satisfying $0\leq t\leq -i\beta$.

The one-particle Green function G is related to G_2 by the equation

$$\left(i\frac{\partial}{\partial t_1}+\frac{\nabla_1^2}{2m}\right)G(1,1')=\delta(1-1')+\int_0^{-i\beta}dt_2 \times\int d\mathbf{r}_2\Sigma(1,2)G(2,1'), \tag{2.2}$$

where

$$\Sigma(1,2)G(2,1')=\pm iV(\mathbf{r}_1-\mathbf{r}_2)G_2(12,1'2^+)\big|_{t_2=t_1} \tag{2.3}$$

defines the *self-energy* Σ. In addition, G satisfies the boundary condition

$$G(\mathbf{r}_1t_1-i\beta,1')=\pm e^{-\beta\mu}G(\mathbf{r}_1t_1,1'), \tag{2.4}$$

the convention that the upper sign pertain to bosons and the lower to fermions being adhered to throughout this paper. If we construct the Fourier transform with respect to space variables and the Fourier series with respect to (imaginary) times of (2.2) and incorporate condition (2.4), the corresponding Fourier coefficients are related by

$$G(\mathbf{p},\omega_p)=[\omega_p-E_p-\Sigma(\mathbf{p},\omega_p)]^{-1}, \tag{2.5}$$

where $E_p=p^2/2m$, $(p=|\mathbf{p}|)$, and $\omega_p=i\pi\beta^{-1}p+\mu$ ($p=0,\pm2,\pm4,\ldots$ for bosons or $\pm1,\pm3,\pm5,\ldots$ for fermions). We also define

$$G^{(0)}(\mathbf{p},\omega_p)=[\omega_p-E_p]^{-1}. \tag{2.6}$$

The particle density of our system is defined by

$$n=\mathrm{Tr}[e^{-\beta(H-\mu\hat{N})}\psi^\dagger(\mathbf{r},t)\psi(\mathbf{r},t)]/\mathrm{Tr}\,e^{-\beta(H-\mu\hat{N})}, \tag{2.7}$$

which is independent of $\mathbf{r}$ and t and can be put in the form

$$n=\mp\beta^{-1}\lim_{\tau\to0^+}\int\frac{d\mathbf{p}}{(2\pi)^3}\sum_p e^{\tau\omega_p}G(\mathbf{p},\omega_p), \tag{2.8}$$

where τ is real. From (2.5) and (2.3) it is clear that to obtain n a knowledge of Σ and, consequently, of G_2 is required.

The two-particle Green function $G_2(12,1'2')$ describes the propagation of two particles added to the medium at $1'$ and $2'$ and removed at 1 and 2. In the *ladder approximation*[2] for G_2 the only processes considered are those in which the two particles propagate independently (but not freely), except that they may interact with each other any number of times, just once at any particular time. The physical idea is that the collision time is short compared to the time between collisions; this is appropriate for a dilute system with short range forces. This approximation is described by the integral equation

$$G_2(12,1'2')=G(11')G(22')\pm G(12')G(21') +i\int d3d4\,G(13)G(24)V(3-4)G_2(34,1'2'). \tag{2.9}$$

In connection with (2.3) the *ladder approximation* for G_2 is equivalent to the T *approximation* for Σ whose Fourier coefficient is given by[9]

$$\Sigma(\mathbf{p},\omega_p)=\mp\Omega^{-1}\beta^{-1}\lim_{\tau\to0^+}\int\frac{d\mathbf{p}_2}{(2\pi)^3} \times\sum_{p_2}e^{\tau\omega_{p_2}}(\mathbf{pp}_2|\,T(\omega_p+\omega_{p_2})\,|\mathbf{pp}_2)_sG(\mathbf{p}_2,\omega_{p_2}), \tag{2.10}$$

where $|\mathbf{pp}_2)_s=|\mathbf{pp}_2)\pm|\mathbf{p}_2\mathbf{p})$, Ω is the volume of the system ($\lim\Omega\to\infty$ implied), and the (generalized) T matrix must satisfy

$$(\mathbf{p}_1\mathbf{p}_2|\,T(\omega_\nu)\,|\mathbf{k}_1\mathbf{k}_2)=(\mathbf{p}_1\mathbf{p}_2|\,V\,|\mathbf{k}_1\mathbf{k}_2) +\int\frac{d\mathbf{k}_1'\,d\mathbf{k}_2'}{(2\pi)^6}(\mathbf{p}_1\mathbf{p}_2|\,T(\omega_\nu)\,|\mathbf{k}_1'\mathbf{k}_2')Q(\mathbf{k}_1'\mathbf{k}_2',\omega_\nu)(\mathbf{k}_1'\mathbf{k}_2'|\,V\,|\mathbf{k}_1\mathbf{k}_2), \tag{2.11}$$

where

$$Q(\mathbf{k}_1'\mathbf{k}_2',\omega_\nu)=-\beta^{-1}\sum_{p'}G(\mathbf{k}_1',\omega_{p'})G(\mathbf{k}_2',\omega_\nu-\omega_{p'}). \tag{2.12}$$

[In (2.10), (2.11), and (2.12) ω_p, ω_{p_2}, and $\omega_{p'}$ carry the appropriate statistics while ω_ν must have boson character.) Note that the T matrix defined here is more general than that which was used in nuclear matter calculations by Puff.[10]

The T matrix defined by (2.11) is assumed to possess an expansion in powers of the activity of the form

$$T=T_0+T_1z+T_2z^2+\cdots. \tag{2.13}$$

The first two terms in this expansion are required in this paper: T_0 corresponds to the standard two particle T matrix [i.e., matrix elements of (1.5)] and T_1 can be expressed in terms of T_0; these points are discussed in Appendix A.

B. Calculation of density coefficients

The Fourier coefficients $G(\mathbf{p},\omega_p)$, $\Sigma(\mathbf{p},\omega_p)$, and $(\mathbf{p}_1\mathbf{p}_2|\,T(\omega_p)\,|\mathbf{k}_1\mathbf{k}_2)$ have been defined by (2.5), (2.10), (2.11), and (2.12): When $p\neq0$ these coefficients may be continued to define the functions $G(\mathbf{p},\omega)$, $\Sigma(\mathbf{p},\omega)$, and $(\mathbf{p}_1\mathbf{p}_2|\,T(\omega)\,|\mathbf{k}_1\mathbf{k}_2)$ which are analytic when ω is not purely real. Branch cuts along the entire real axis are introduced for these functions. These three functions all have the property $F(\omega^*)=F^*(\omega)$ and, therefore, $\mathrm{Re}F(\omega)$ is continuous across the real axis while $\mathrm{Im}F(\omega)$ is not: For each of these functions we will denote $\lim\epsilon\to0$ $\mathrm{Re}F(x+i\epsilon)$ where x is real by $\mathrm{Re}F(x)$, and the symbols x^+, x^- will mean $x+i\epsilon$ and $x-i\epsilon$, respectively, where the limit $\epsilon\to0^+$ is to be taken after the integration is performed. (See Appendix B.) Note that in the case of Bose statistics we encounter the Fourier coefficients $G(\mathbf{p},\omega_0)$ and $\Sigma(\mathbf{p},\omega_0)$: These are connected to the corresponding functions by $G(\mathbf{p},\omega_0)=\mathrm{Re}G(\mathbf{p},\mu)$ and $\Sigma(\mathbf{p},\omega_0)=\mathrm{Re}\Sigma(\mathbf{p},\mu)$.

Let us define

$$S(\omega)=(e^{\beta(\omega-\mu)}\mp1)^{-1}, \tag{2.14}$$

a function having simple poles at $\omega=\omega_p$, each with residue $\pm\beta^{-1}$. By considering

$$\lim_{\tau\to0^+}\int_\Gamma d\omega\,e^{\tau\omega}S(\omega)G(\mathbf{p},\omega) \tag{2.15}$$

where Γ is described in Fig. 1 and assuming that $|\omega G(\mathbf{p},\omega)|$ is bounded as $|\omega|\to\infty$, it can be shown that (2.8) becomes

$$n=-\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}\,\frac{\mathrm{Im}G(\mathbf{p},x^+)}{z^{-1}e^{\beta x}+1} \tag{2.16}$$

for Fermi statistics and

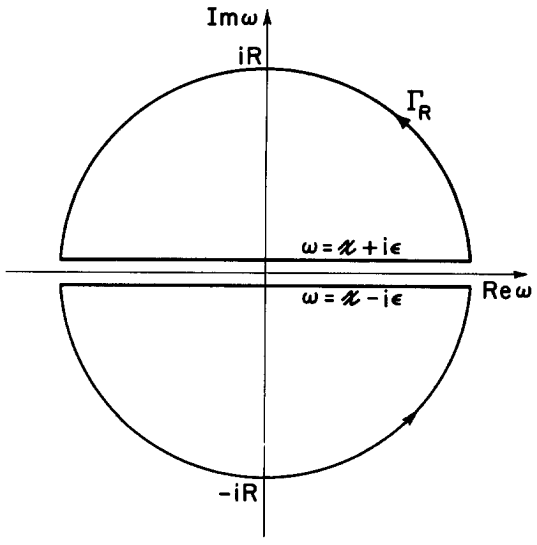

FIG. 1. The integration contour Γ referred to in the text is Γ_R in the limit $R\to\infty$ and $\epsilon\to 0$.

$$n=-\beta^{-1}\int\frac{d\mathbf{p}}{(2\pi)^3}G(\mathbf{p},\omega_0)-\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}\mathrm{Im}\left(\frac{G(\mathbf{p},x^+)}{z^{-1}e^{\beta x^+}-1}\right)$$
$$=-\int\frac{d\mathbf{p}}{(2\pi)^3}\mathrm{P.V.}\int_{-\infty}^{\infty}\frac{dx}{\pi}\frac{\mathrm{Im}G(\mathbf{p},x^+)}{z^{-1}e^{\beta x}-1} \qquad (2.17)$$

for Bose statistics (P.V. means principal value of the following integral with respect to $x=\mu$). These results can be combined in the form

$$n=-\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}\mathrm{Re}S(x^+)\mathrm{Im}G(\mathbf{p},x^+). \qquad (2.18)$$

The self-energy function will be shown to have an activity expansion of the form

$$\Sigma(\mathbf{p},\omega)=z\Sigma_1(\mathbf{p},\omega)+z^2\Sigma_2(\mathbf{p},\omega)+\cdots. \qquad (2.19)$$

Combination of (2.5) with (2.19) and substitution into (2.18) gives the following results for the density coefficients:

$$n_1=-\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x}\mathrm{Im}G^{(0)}(\mathbf{p},x^+)=\lambda^{-3}, \qquad (2.20)$$

$$n_2=-\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}(\pm e^{-2\beta x}\mathrm{Im}G^{(0)}(\mathbf{p},x^+)$$
$$+e^{-\beta x}\mathrm{Im}\{[G^{(0)}(\mathbf{p},x^+)]^2\Sigma_1(\mathbf{p},x^+)\}), \qquad (2.21)$$

$$n_3=-\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}(e^{-3\beta x}\mathrm{Im}G^{(0)}(\mathbf{p},x^+)$$
$$\pm e^{-2\beta x}\mathrm{Im}\{[G^{(0)}(\mathbf{p},x^+)]^2\Sigma_1(\mathbf{p},x^+)\}$$
$$+e^{-\beta x}\mathrm{Im}\{[G^{(0)}(\mathbf{p},x^+)]^2\Sigma_2(\mathbf{p},x^+)\}$$
$$+e^{-\beta x}\mathrm{Im}\{[G^{(0)}(\mathbf{p},x^+)]^3[\Sigma_1(\mathbf{p},x^+)]^2\}). \qquad (2.22)$$

The next step is to determine Σ_1 and Σ_2 according to the *T approximation*, (2.11) and (2.12). The summation in (2.10) can be performed by considering the integral

$$\lim_{\tau\to 0^+}\int_\gamma d\omega\, e^{\tau\omega}S(\omega)(\mathbf{pp}_2|T(\omega+\omega_p)|\mathbf{pp}_2)_s G(\mathbf{p}_2,\omega),$$

where γ is described in Fig. 2 and ω_p is fixed with $p\neq 0$. Since $(\mathbf{pp}_2|T(\omega+\omega_p)|\mathbf{pp}_2)_s$ is bounded as $|\omega|\to\infty$ and since $0\leq\tau\leq\beta$ we have, as $|\omega|\to\infty$,

$$|\omega\, e^{\tau\omega}S(\omega)(\mathbf{pp}_2|T(\omega+\omega_p)|\mathbf{pp}_2)_s G(\mathbf{p}_2,\omega)|\to 0.$$

The T matrix, considered as a function of ω, is analytic everywhere except on the line $\omega=x-\omega_p$; $G(\mathbf{p}_2,\omega)$ is analytic everywhere except on the real axis; the poles of $S(\omega)$ have already been mentioned. [The poles of $S(\omega)$ on the real axis in the Bose case and on the line $\omega=x-\omega_p$ in either case demand special attention, but there is no difficulty.] From these considerations we obtain

$$\Sigma(\mathbf{p},\omega_p)=-\Omega^{-1}\int\frac{d\mathbf{p}_2}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}[\mathrm{Re}S(x^+)(\mathbf{pp}_2|T(x+\omega_p)|\mathbf{pp}_2)_s$$
$$\times\mathrm{Im}G(\mathbf{p}_2,x^+)+\mathrm{Re}S(x^+-\omega_p)G(\mathbf{p}_2,x-\omega_p)\,\mathrm{Im}(\mathbf{pp}_2|T(x^+)|\mathbf{pp}_2)_s]. \qquad (2.23)$$

[See Appendix B for the meaning of $\mathrm{Re}S(x^+)$ and $\mathrm{Re}S(x^+-\omega_p)$.] Noting that $S(x-\omega_p)$ is of higher order than $S(x)$ by one power of z, we obtain for the desired self-energy coefficients

$$\Sigma_1(\mathbf{p},\omega_p)=-\Omega^{-1}\int\frac{d\mathbf{p}_2}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x}(\mathbf{pp}_2|T_0(x+\omega_p)|\mathbf{pp}_2)_s$$
$$\times\,\mathrm{Im}G^{(0)}(\mathbf{p}_2,x^+), \qquad (2.24)$$

$$\Sigma_2(\mathbf{p},\omega_p)=-\Omega^{-1}\int\frac{d\mathbf{p}_2}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}[\pm e^{-2\beta x}(\mathbf{pp}_2|T_0(x+\omega_p)|\mathbf{pp}_2)_s$$
$$\times\,\mathrm{Im}G^{(0)}(\mathbf{p}_2,x^+)$$
$$+e^{-\beta x}(\mathbf{pp}_2|T_0(x+\omega_p)|\mathbf{pp}_2)_s\,\mathrm{Im}\{[G^{(0)}(\mathbf{p}_2,x^+)]^2\Sigma_1(\mathbf{p}_2,x^+)\}$$
$$+e^{-\beta x}(\mathbf{pp}_2|T_1(x+\omega_p)|\mathbf{pp}_2)_s\,\mathrm{Im}G^{(0)}(\mathbf{p}_2,x^+)$$
$$\pm e^{-\beta x}G^{(0)}(\mathbf{p}_2,x-\omega_p)\,\mathrm{Im}(\mathbf{pp}_2|T_0(x^+)|\mathbf{pp}_2)_s]. \qquad (2.25)$$

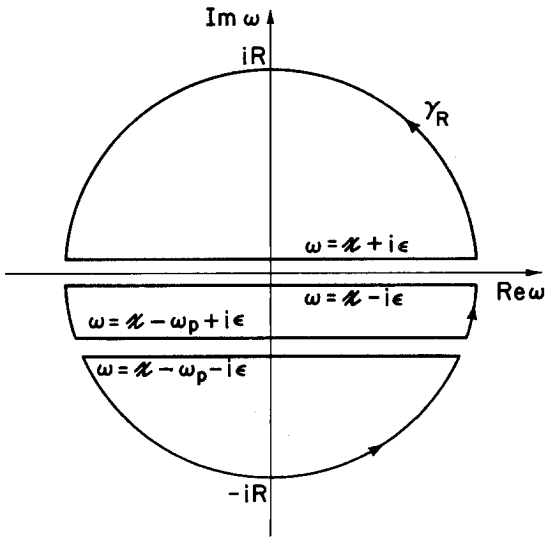

FIG. 2. The integration contour γ referred to in the text is γ_R in the limit $R\to\infty$ and $\epsilon\to 0$ where $\omega_p-\mu\neq 0$.

Now we can use Eq. (A11) to express T_1 in terms of T_0 and perform certain integrations to obtain

$$\Sigma_1(\mathbf{p},x^+)=\Omega^{-1}\int\frac{d\mathbf{p}_2}{(2\pi)^3}e^{-\beta E_{p_2}}(\mathbf{pp}_2|\,T_0(x^+ + E_{p_2})|\mathbf{pp}_2)_s, \tag{2.26}$$

$$\begin{aligned}\Sigma_2(\mathbf{p},x^+)=&\pm\Omega^{-1}\int\frac{d\mathbf{p}_2}{(2\pi)^3}e^{-2\beta E_{p_2}}(\mathbf{pp}_2|\,T_0(x^+ + E_{p_2})|\mathbf{pp}_2)_s\\
&-\Omega^{-1}\int\frac{d\mathbf{p}_2}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx'}{\pi}e^{-\beta x'}(\mathbf{pp}_2|\,T_0(x'+x^+)|\mathbf{pp}_2)_s\\
&\times\mathrm{Im}\left(\frac{\Sigma_1(\mathbf{p}_2,x'^+)}{(x'^+ - E_{p_2})^2}\right)\\
&\pm\Omega^{-1}\int\frac{d\mathbf{p}_2d\mathbf{k}_1d\mathbf{k}_2}{(2\pi)^9}e^{-\beta E_{p_2}}(e^{-\beta E_{k_1}}+e^{-\beta E_{k_2}})\\
&\times(\mathbf{pp}_2|\,T_0(x^+ + E_{p_2})|\mathbf{k}_1\mathbf{k}_2)\\
&\times(\mathbf{k}_1\mathbf{k}_2|\,T_0(x^+ + E_{p_2})|\mathbf{pp}_2)_s(x^+ + E_{p_2}-E_{k_1}-E_{k_2})^{-1}\\
&+\Omega^{-1}\int\frac{d\mathbf{p}_2d\mathbf{k}_1d\mathbf{k}_2}{(2\pi)^9}e^{-\beta E_{p_2}}\Sigma_1(\mathbf{k}_1,x^+ + E_{p_2}-E_{k_2})\\
&\times(\mathbf{pp}_2|\,T_0(x^+ + E_{p_2})|\mathbf{k}_1\mathbf{k}_2)\\
&\times(\mathbf{k}_1\mathbf{k}_2|\,T_0(x^+ + E_{p_2})|\mathbf{pp}_2)_s(x^+ + E_{p_2}-E_{k_1}-E_{k_2})^{-2}\\
&-\Omega^{-1}\int\frac{d\mathbf{p}_2d\mathbf{k}_1d\mathbf{k}_2}{(2\pi)^9}e^{-\beta E_{p_2}}(\mathbf{pp}_2|\,T_0(x^+ + E_{p_2})|\mathbf{k}_1\mathbf{k}_2)\\
&\times(\mathbf{k}_1\mathbf{k}_2|\,T_0(x^+ + E_{p_2})|\mathbf{pp}_2)_s\\
&\times\int_{-\infty}^{\infty}\frac{dx'}{\pi}(x^+ - x' - E_{k_1})^{-1}\,\mathrm{Im}\left(\frac{\Sigma_1(\mathbf{k}_2,x'^+ + E_{p_2})}{(x'^+ + E_{p_2}-E_{k_2})^2}\right)\\
&\mp\Omega^{-1}\int\frac{d\mathbf{p}_2}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx'}{\pi}e^{-\beta x'}\frac{\mathrm{Im}(\mathbf{pp}_2|\,T_0(x'^+)|\mathbf{pp}_2)}{x'-x^+-E_{p_2}}.\end{aligned} \tag{2.27}$$

It turns out that the density coefficients n_2 and n_3, given by (2.21) and (2.22), can be written entirely in terms of Σ_1; this requires the use of Eq. (A14). The results are

$$n_2=\pm 2^{-3/2}\lambda^{-3}-\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x}\,\mathrm{Im}\left(\frac{\Sigma_1(\mathbf{p},x^+)}{(x^+ - E_p)^2}\right), \tag{2.28}$$

$$\begin{aligned}n_3^{(GF)}=&3^{-3/2}\lambda^{-3}\mp\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-2\beta x}\,\mathrm{Im}\left(\frac{\Sigma_1(\mathbf{p},x^+)}{(x^+ - E_p)^2}\right)\\
&\mp\int\frac{d\mathbf{p}}{(2\pi)^3}e^{-\beta E_p}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x}\,\mathrm{Im}\left(\frac{\Sigma_1(\mathbf{p},x^+)}{(x^+ - E_p)^2}\right)\\
&+\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x}\,\mathrm{Im}\left(\frac{\Sigma_1(\mathbf{p},x^+)}{(x^+ - E_p)^2}\right)\int_{-\infty}^{\infty}\frac{dx'}{\pi}e^{-\beta x'}\\
&\times\mathrm{Im}\left(\frac{\Sigma_1(\mathbf{p},x'^+ + x)}{(x'^+)^2}\right)\\
&\pm 2\int\frac{d\mathbf{p}}{(2\pi)^3}e^{-\beta E_p}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x}\,\mathrm{Im}\left(\frac{(\partial/\partial x)\Sigma_1(\mathbf{p},x^+)}{x^+ - E_p}\right)\\
&+\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x}\,\mathrm{Im}\left(\frac{\Sigma_1(\mathbf{p},x^+)(\partial/\partial x)\Sigma_1(\mathbf{p},x^+)}{(x^+ - E_p)^2}\right)\\
&-\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x}\,\mathrm{Im}\left(\frac{\Sigma_1(\mathbf{p},x^+)}{(x^+ - E_p)^2}\right)\int_{-\infty}^{\infty}\frac{dx'}{\pi}e^{-\beta x'}\\
&\times\mathrm{Im}\left(\frac{(\partial/\partial x')\Sigma_1(\mathbf{p},x'^+ + x)}{x'^+}\right)\\
&\mp\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x}\int_{-\infty}^{\infty}\frac{dx'}{\pi}e^{-\beta x'}\,\mathrm{Im}\Sigma_1(\mathbf{p},x'^+)\\
&\times\mathrm{Im}\left(\frac{1}{(x^+ - E_p)^2(x^+ - x')}\right)\\
&-\int\frac{d\mathbf{p}}{(2\pi)^3}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x}\,\mathrm{Im}\left(\frac{[\Sigma_1(\mathbf{p},x^+)]^2}{(x^+ - E_p)^3}\right).\end{aligned} \tag{2.29}$$

The latter is the result of the *T approximation* for n_3. One can replace Σ_1 in (2.29) by T_0 according to (2.26), but it is convenient to postpone this step until the summary in Sec. 5.

3. SELECTED DIAGRAMS TECHNIQUE FOR THE DENSITY IN THE MULTIPLE-SCATTERING THEORY

The use of formal operator techniques in quantum statistical mechanics, specifically the Watson[11] multiple-scattering formalism, was suggested by Reiner[3]; he also discussed the relation between this approach and the binary collision method developed by Lee and Yang.[12] Certain advantages of the former method compared to the latter were discussed recently by Gibson.[13]

The diagram techniques described in the review by Bloch[4] are appropriate for calculating the density in powers of the activity: The diagrams themselves are not essential, of course, but they serve as a guide in accounting for the various contributions. The density is the expectation of the operator $\hat{N}/\Omega$ in the grand canonical ensemble for our system:

$$n=Z_{gr}^{-1}\Omega^{-1}\mathrm{Tr}\{e^{-\beta(H-\mu\hat{N})}\hat{N}\}, \tag{3.1}$$

where $Z_{gr}=\mathrm{Tr}\,e^{-\beta(H-\mu\hat{N})}$ is the grand partition function, and $\lim\Omega\to\infty$ is implied; the trace is to be evaluated with respect to symmetric or antisymmetric states as required. By introducing complete orthonormal sets of states containing N particles, we may write

$$\begin{aligned}n=&Z_{gr}^{-1}\Omega^{-1}\sum_{N=1}^{\infty}\frac{z^N}{(N-1)!}\sum_{m_1\cdots m_N}\sum_P\epsilon^P\\
&\times(m_1\cdots m_N|\,e^{-\beta H^{(N)}}P|m_1\cdots m_N),\end{aligned} \tag{3.2}$$

where $H^{(N)}$ is the Hamiltonian for the N-particle system, P is the permutation operator, and $\epsilon^P=1$ for

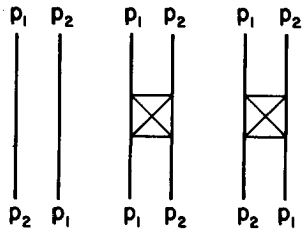

FIG. 3. Linked diagrams for n_2.

bosons, ± 1 for fermions according to whether P is an even or odd permutation.

The required matrix elements of $e^{-\beta H^{(N)}}$ will be determined using the operator relations[11]

$$e^{-\beta H^{(N)}} = \int_C \frac{d\omega}{2\pi i} e^{-\beta\omega} \frac{1}{\omega - H^{(N)}} \tag{3.3}$$

and

$$\frac{1}{\omega - H^{(N)}} = \sum_{n=0}^{\infty} \frac{1}{\omega - H_0^{(N)}} \left(H_1^{(N)} \frac{1}{\omega - H_0^{(N)}} \right)^n, \tag{3.4}$$

where the path of integration C in (3.3) surrounds the real axis of the ω plane in the positive sense. For $N=2$, (3.4) becomes simply

$$(\omega - H^{(2)})^{-1} = (\omega - H_0^{(2)})^{-1} + (\omega - H_0^{(2)})^{-1} T_0(\omega)(\omega - H_0^{(2)})^{-1}, \tag{3.5}$$

where $T_0(\omega)$ is the standard two-particle T operator defined by (1.5). Similarly, one could introduce the three-particle T operator,

$$T^{(3)}(\omega) = (V_{12} + V_{23} + V_{13})[1 + (\omega - H_0^{(3)})^{-1} T^{(3)}(\omega)], \tag{3.6}$$

but that is not our intention.

The diagrams which represent the various matrix elements in (3.2) can be divided into two types, *linked* and *unlinked*. It is well known that the z-dependence of Z_{gr}^{-1} can be provided for by including only the contributions from *linked* diagrams.[4] We choose to use the momentum representation,[9,14] such that the product state $|\mathbf{p}_1 \cdots \mathbf{p}_N)$ is an eigenstate of $H_0^{(N)}$ with eigenvalue $E_{\mathbf{p}_1} + \cdots + E_{\mathbf{p}_N}$. Now, using (3.3), (3.4), and (3.5) in (3.2), we obtain

$$n_2 = \Omega^{-1} \int \frac{d\mathbf{p}_1 d\mathbf{p}_2}{(2\pi)^6} \int_C \frac{d\omega}{2\pi i} e^{-\beta\omega} \left((\mathbf{p}_1\mathbf{p}_2| \frac{1}{\omega - H_0^{(2)}} + \frac{1}{\omega - H_0^{(2)}} \times T_0(\omega) \frac{1}{\omega - H_0^{(2)}} |\mathbf{p}_1\mathbf{p}_2)_s \right)_L, \tag{3.7}$$

$$n_3 = \frac{1}{2}\Omega^{-1} \int \frac{d\mathbf{p}_1 d\mathbf{p}_2 d\mathbf{p}_3}{(2\pi)^9} \sum_P \epsilon^P \int_C \frac{d\omega}{2\pi i} e^{-\beta\omega} \left\{ (\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3| \frac{1}{\omega - H_0^{(3)}} \times \sum_{n=0}^{\infty} \left[(V_{12} + V_{13} + V_{23}) \frac{1}{\omega - H_0^{(3)}} \right]^n P |\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3) \right\}_L, \tag{3.8}$$

where the subscript L means that only matrix elements corresponding to *linked* diagrams are to be included and the subscript s is defined following (2.10).

Let us define the operators

$$T_0^{(12)}(\omega) = V_{12} + V_{12}(\omega - H_0^{(2)}(1,2))^{-1} T_0^{(12)}(\omega), \tag{3.9}$$

and $T_0^{(23)}(\omega)$ and $T_0^{(31)}(\omega)$ in analogous fashion, where $H_0^{(2)}(1,2) = K_1 + K_2$, etc. Now if we assume that the infinite series in (3.8) for a particular matrix element is absolutely convergent, then it is permissible to rearrange its terms in the form

$$(\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3| (\omega - H_0^{(3)})^{-1} + (\omega - H_0^{(3)})^{-1} [T_0^{(12)}(\omega - K_3) + T_0^{(23)}(\omega - K_1) + T_0^{(31)}(\omega - K_2) + T_0^{(12)}(\omega - K_3)(\omega - H_0^{(3)})^{-1} T_0^{(23)}(\omega - K_1) + \cdots](\omega - H_0^{(3)})^{-1} |\mathbf{p}_1'\mathbf{p}_2'\mathbf{p}_3'), \tag{3.10}$$

where there are three terms involving one T (as shown), six terms involving two T's (one of which is shown), twelve terms involving three T's and so on. Note that in any term the same T may not occur twice in succession.

The matrix elements of the operator (3.9) are simply related to the usual two-particle T_0 matrices: thus, for example,

$$(\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3| T_0^{(12)}(\omega - K_3) |\mathbf{p}_1'\mathbf{p}_2'\mathbf{p}_3') = (2\pi)^3 \delta(\mathbf{p}_3 - \mathbf{p}_3')(\mathbf{p}_1\mathbf{p}_2| T_0(\omega - E_{\mathbf{p}_3}) |\mathbf{p}_1'\mathbf{p}_2'). \tag{3.11}$$

This means that after introducing intermediate states as required each term in (3.10) can be expressed in terms of T_0 matrices. Since (3.10) represents a typical matrix element of (3.8) it follows that n_3 can be written as an infinite series whose terms involve T_0 matrices and energy denominators: these terms can be represented by diagrams.[4] We intend to select a finite number of such diagrams to obtain an approximation for n_3.

The diagrams for the density consist of vertical lines representing the free-particle propagator and crossed rectangles (called T_0-insertions), representing the T_0 operator. Each line carries a factor of z, and each point on a line represents a "time" or, strictly speaking, an inverse temperature playing the role of imaginary time. Since T_0 contains the interaction V to all orders there must not be two successive T_0-insertions connecting the same two lines. The diagrams must be labeled at top and bottom by the momenta of the desired matrix element, and each internal line segment must also be assigned a momentum label. For a particular order N one requires all possible diagrams having N lines. For $N=2$ the number of diagrams is finite (see Fig. 3); for $N=3$ the number is infinite and any finite selection of diagrams gives an approximation for n_3.

The contribution of a given diagram is given by $(2\pi i\Omega)^{-1} \int_C d\omega e^{-\beta\omega} F(\omega)$ where $F(\omega)$ is composed of

(1) A factor $(\omega - \sum_{i=1}^{N} E_{k_i})^{-1}$ for each space between two successive T_0-insertions, for each space between the bottom of the diagram and the T_0-insertion immediately above or the top and the T_0-insertion immediately below, or for the space between the top and bottom of the diagram if it contains no T_0-insertions, where E_{k_i} corresponds to the line segment with label $\mathbf{k}_i$;

(2) A factor $(\mathbf{k}_i\mathbf{k}_j| T_0(\omega - \hat{E}) |\mathbf{k}_i'\mathbf{k}_j')$ for each T_0-insertion where the momenta refer to the line segments entering and leaving the rectangle, and $\hat{E}$ equals the sum of the energies associated with the points at the same "time" on each line segment not touching the rectangle;

(3) A factor $(2\pi)^3 \delta(\mathbf{p}_i - \mathbf{p}_i')$ for each line extending

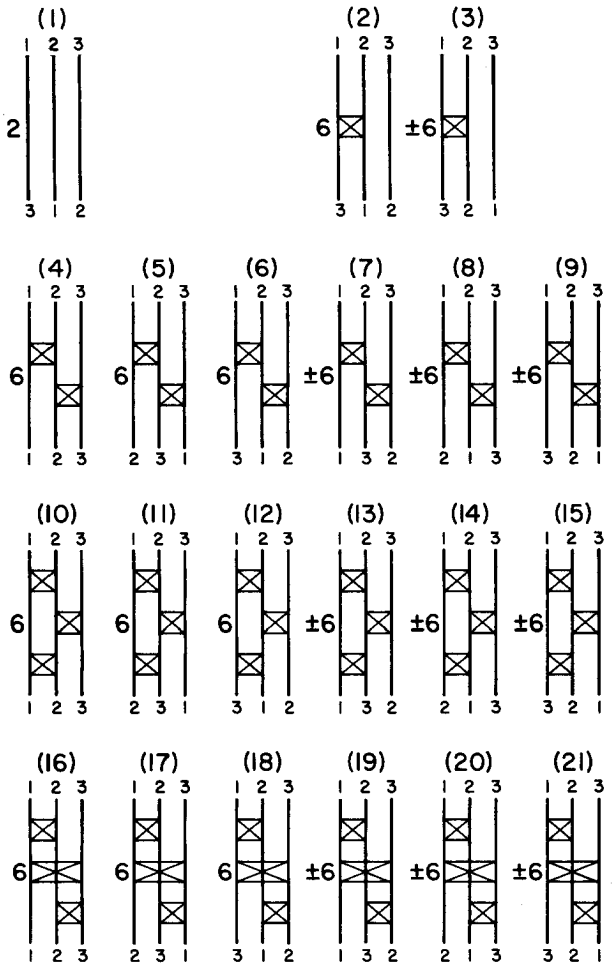

FIG. 4. Linked diagrams for n_3 containing three or fewer T_0-insertions. The number to the left of a diagram indicates the number of completely labeled diagrams which are equivalent to it. (See Bloch.[4])

from the bottom to the top of the diagram without touching a rectangle, (except that for a diagram containing no rectangles the contribution of one of the lines is just Ω);

(4) Integration over all momentum labels is required. Property (A13) of the T_0 matrices permits the complex integrals to be replaced by real integrals.

The diagrams in Fig. 3 give

$$n_2 = \pm 2^{-3/2}\lambda^{-3} - \Omega^{-1}\int \frac{dp_1 dp_2}{(2\pi)^6}\int_{-\infty}^{\infty}\frac{ax}{\pi}e^{-\beta x}\,\mathrm{Im}\left(\frac{(p_1p_2|\,T_0(x^+)\,|p_1p_2)_s}{(x^+ - E_{p_1} - E_{p_2})^2}\right). \tag{3.12}$$

In Fig. 4 we present the diagrams for n_3 which contain three or fewer T_0-insertions. The number to the left of a diagram is the number of diagrams which can be shown by simple symmetry arguments to be equivalent to it. Each diagram is labeled by a number appearing above it, and the corresponding contributions to n_3 are given by

$$n_3^{(1)} = \int \frac{dp}{(2\pi)^3}e^{-3\beta E_p}, \tag{3.13a}$$

$$n_3^{(2)} + n_3^{(3)} = \mp 3\Omega^{-1}\int \frac{dp_1 dp_2}{(2\pi)^6}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x} \times \mathrm{Im}\left(\frac{(p_1p_2|\,T_0(x^+ - E_{p_2})\,|p_1p_2)_s}{(x^+ - E_{p_1} - 2E_{p_2})^2}\right), \tag{3.13b}$$

$$n_3^{(4)} + n_3^{(5)} + n_3^{(7)} + n_3^{(8)} = -3\Omega^{-1}\int \frac{dp_1 dp_2 dp_3 dk}{(2\pi)^{12}}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x} \times \mathrm{Im}\left(\frac{(p_1p_2|\,T_0(x^+ - E_{p_3})\,|p_1k)_s(kp_3|\,T_0(x^+ - E_{p_1})\,|p_2p_3)_s}{(x^+ - E_{p_1} - E_{p_2} - E_{p_3})^2(x^+ - E_{p_1} - E_k - E_{p_3})}\right), \tag{3.13c}$$

$$n_3^{(6)} + n_3^{(9)} = -3\Omega^{-1}\int \frac{dp_1 dp_2 dp_3 dk}{(2\pi)^{12}}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x} \times \mathrm{Im}\left(\frac{(p_1p_2|\,T_0(x^+ - E_{p_3})\,|p_3k)(kp_3|\,T_0(x^+ - E_{p_3})\,|p_1p_2)_s}{(x^+ - E_{p_1} - E_{p_2} - E_{p_3})^2(x^+ - E_k - 2E_{p_3})}\right), \tag{3.13d}$$

$$n_3^{(10)} + n_3^{(14)} + n_3^{(18)} + n_3^{(21)} = -3\Omega^{-1}\int \frac{dp_1 dp_2 dp_3 dk_1 dk_2 dk_3}{(2\pi)^{18}} \times \int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x} \times \mathrm{Im}\left(\frac{(p_1p_2|\,T_0(x^+ - E_{p_3})\,|k_1k_2)(k_2p_3|\,T_0(x^+ - E_{k_1})\,|k_3p_3)_s}{(x^+ - E_{p_1} - E_{p_2} - E_{p_3})^2(x^+ - E_{k_1} - E_{k_2} - E_{p_3})} \times \frac{(k_1k_3|\,T_0(x^+ - E_{p_3})\,|p_1p_2)_s}{(x^+ - E_{k_1} - E_{k_3} - E_{p_3})}\right), \tag{3.13e}$$

$$n_3^{(11)} + n_3^{(12)} + n_3^{(13)} + n_3^{(15)} + n_3^{(16)} + n_3^{(17)} + n_3^{(19)} + n_3^{(20)} = -3\Omega^{-1}\int \frac{dp_1 dp_2 dp_3 dk_1 dk_2 dk_3}{(2\pi)^{18}}\int_{-\infty}^{\infty}\frac{dx}{\pi}e^{-\beta x} \times \mathrm{Im}\left(\frac{(p_1p_2|\,T_0(x^+ - E_{p_3})\,|k_1k_2)_s(k_2p_3|\,T_0(x^+ - E_{k_1})\,|k_3p_1)_s}{(x^+ - E_{p_1} - E_{p_2} - E_{p_3})^2(x^+ - E_{k_1} - E_{k_2} - E_{p_3})} \times \frac{(k_1k_3|\,T_0(x^+ - E_{p_1})\,|p_2p_3)_s}{(x^+ - E_{k_1} - E_{k_3} - E_{p_3})}\right) \tag{3.13f}$$

These results can be simplified by using the T_0 matrix properties (A15) for momentum conservation and (A14) for the derivative. Furthermore, it can be shown that the results (3.13a) to (3.13e) can be expressed in terms of

$$\Sigma_1(p, x^+) = \Omega^{-1}\int \frac{dp'}{(2\pi)^3}e^{-\beta E'_p}(pp'\,|\,T_0(x^+ + E'_p)\,|pp')_s, \tag{3.14}$$

which is precisely the self-energy Σ_1 derived by the Green function analysis of Sec. 2. It can be seen that (3.13f) as well as the contributions from all of the diagrams for n_3 which are not included in Fig. 4 cannot be written in terms of Σ_1 (note that Σ_1 involves the diagonal T_0 matrix). The results for (3.13a) to (3.13e) are

$$n_3^{(SD)} = 3^{-3/2}\lambda^{-3} \mp 3\int \frac{dp}{(2\pi)^3} e^{-\beta E_p} \int_{-\infty}^{\infty} \frac{dx}{\pi} e^{-\beta x} \mathrm{Im}\left(\frac{\Sigma_1(p,x^+)}{(x^+ - E_p)^2}\right)$$

$$- 3\int \frac{dp}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{dx}{\pi} e^{-\beta x} \mathrm{Im}\left(\frac{[\Sigma_1(p,x^+)]^2}{(x^+ - E_p)^3}\right)$$

$$\pm 3\int \frac{dp}{(2\pi)^3} e^{-\beta E_p} \int_{-\infty}^{\infty} \frac{dx}{\pi} e^{-\beta x} \mathrm{Im}\left(\frac{(\partial/\partial x)\Sigma_1(p,x^+)}{x^+ - E_p}\right)$$

$$+ \frac{3}{2}\int \frac{dp}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{dx}{\pi} e^{-\beta x} \mathrm{Im}\left(\frac{(\partial/\partial x)[\Sigma_1(p,x^+)]^2}{(x^+ - E_p)^2}\right). \tag{3.15}$$

This will be our *selected diagram* approximation to n_3.

4. COMPARISON OF THE TWO METHODS

In Secs. 2 and 3 we have developed two approximations for the density coefficient n_3 which is needed to compute the third virial coefficient. We now demonstrate that these two results, (2.29) and (3.15), are identical.

Using integration by parts,[15] (3.15) becomes

$$n_3^{(SD)} = 3^{-3/2}\lambda^{-3} \pm 3\beta\int \frac{dp}{(2\pi)^3} e^{-\beta E_p} \int_{-\infty}^{\infty} \frac{dx}{\pi} e^{-\beta x} \mathrm{Im}\left(\frac{\Sigma_1(p,x^+)}{x^+ - E_p}\right)$$

$$+ \frac{3}{2}\beta\int \frac{dp}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{dx}{\pi} e^{-\beta x} \mathrm{Im}\left(\frac{[\Sigma_1(p,x^+)]^2}{(x^+ - E_p)^2}\right). \tag{4.1}$$

According to Appendix B the second term on the right can be written as

$$\mp 3\beta\int \frac{dp}{(2\pi)^3} e^{-2\beta E_p} \mathrm{Re}\,\Sigma_1(p,E_p) \pm 3\beta\int \frac{dp}{(2\pi)^3} e^{-\beta E_p}$$

$$\times \int_{-\infty}^{\infty} \frac{dx}{\pi} e^{-\beta x} \mathrm{Im}\Sigma_1(p,x^+)\mathrm{Re}(x^+ - E_p)^{-1}, \tag{4.2}$$

and, if each of the four terms in (2.29) which involve Σ_1 to first order is expanded in similar fashion, it can be seen that the sum of these terms is the same.

Similarly, we compare the third term on the right of (4.1) and the sum of the four terms in (2.29) which contain Σ_1 to second order. The difference between these quantities is

$$\beta\int \frac{dp}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{dx}{\pi} e^{-\beta x} \mathrm{Im}\left(\frac{[\Sigma_1(p,x^+)]^2}{(x^+ - E_p)^2}\right) + \beta\int \frac{dp}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{dx}{\pi} e^{-\beta x}$$

$$\times \mathrm{Im}\left(\frac{\Sigma_1(p,x^+)}{(x^+ - E_p)^2}\right) \int_{-\infty}^{\infty} \frac{dx'}{\pi} e^{-\beta x'} \mathrm{Im}\left(\frac{\Sigma_1(p,x'^+ + x)}{x'^+}\right). \tag{4.3}$$

Using Appendix B one can show that (4.3) is equal to

$$\beta\int \frac{dp}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{dx}{\pi} \left[e^{-\beta x} \mathrm{Im}\Sigma_1(p,x^+)\mathrm{Re}\left(\frac{\Sigma_1(p,x^+)}{(x^+ - E_p)^2}\right)\right.$$

$$\left. + \mathrm{Im}\left(\frac{\Sigma_1(p,x^+)}{(x^+ - E_p)^2}\right) \int_{-\infty}^{\infty} \frac{dx'}{\pi} e^{-\beta x'} \mathrm{Im}\Sigma_1(p,x^+)\mathrm{Re}(x'^+ - x)^{-1}\right]$$

$$= -\beta\int \frac{dp}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{dx}{\pi} e^{-\beta x} \mathrm{Im}\Sigma_1(p,x^+)$$

$$\times \int_{-\infty}^{\infty} \frac{dx'}{\pi} \mathrm{Im}\left(\frac{\Sigma_1(p,x'^+)}{(x'^+ - x)(x'^+ - E_p)^2}\right); \tag{4.4}$$

this equals zero because the final integral does. This can be demonstrated by considering

$$\int_\Gamma d\omega \Sigma_1(p,\omega)(\omega - x)^{-1}(\omega - E_p)^{-2} \tag{4.5}$$

where the contour Γ is described in Fig. 1. Since x and E_p are real and since $\Sigma_1(p,\omega)$ is analytic in any region not including the real axis this integral must equal zero. Since $|\Sigma_1(p,\omega)|$ is bounded as $|\omega| \to \infty$, (4.5) must also equal the last integral in (4.4). This completes the proof that $n_3^{(SD)} = n_3^{(GF)}$.

5. DISCUSSION

It has been established that the particular selection of diagrams made in Sec. 3 gives the same formula for n_3 as the *T approximation* of Green function theory. The interaction enters this formula via the self-energy function Σ_1 only; thus, the *T approximation* must correspond to the set of diagrams in which the T_0 matrices are diagonal or are reducible to diagonal form by means of (A14). This requirement is satisfied by all diagrams having two or fewer T_0-insertions but by no diagrams with four or more. Of the diagrams having three T_0-insertions (see Fig. 4), certain ones are admitted and the others are excluded. The favored diagrams (10), (14), (18), and (21) share the property that the line labeled 3 enters and leaves the same insertion: this suggests that particles 1 and 2 form a highly correlated pair. These diagrams can be interpreted as representing the three successive events: a scattering between a pair of particles, a scattering between one particle of the pair and a third particle, and a scattering involving the original pair. The rejected diagrams do not possess this feature.

Having confirmed that the two approximations presented in this paper give the same result for n_3, we think that the formula, thus obtained, deserves further consideration. The result can be expressed in the form

$$n_3 \approx 3^{-3/2}\lambda^{-3} \mp 3\beta\int \frac{dp}{(2\pi)^3} e^{-2\beta E_p} \mathrm{Re}\Sigma_1(p,E_p)$$

$$\pm 3\beta\int \frac{dp}{(2\pi)^3} e^{-\beta E_p} \int_{-\infty}^{\infty} \frac{dx}{\pi} e^{-\beta x}$$

$$\times \mathrm{Im}\Sigma_1(p,x^+)\mathrm{Re}(x^+ - E_p)^{-1}$$

$$+ \tfrac{3}{2}\beta^2\int \frac{dp}{(2\pi)^3} e^{-\beta E_p}\mathrm{Re}[\Sigma_1(p,E_p^+)]^2$$

$$- \tfrac{3}{2}\beta\int \frac{dp}{(2\pi)^3} e^{-\beta E_p} \frac{\partial}{\partial E_p}\mathrm{Re}[\Sigma_1(p,E_p^+)]^2$$

$$+ 3\beta\int \frac{dp}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{dx}{\pi} e^{-\beta x}\mathrm{Re}\Sigma_1(p,x)$$

$$\times \mathrm{Im}\Sigma_1(p,x^+)\mathrm{Re}(x^+ - E_p)^{-2}. \tag{5.1}$$

The last two terms in this formula can be combined by using the fact that

$$\int_{-\infty}^{\infty} dx \, \mathrm{Im}\left(\frac{[\Sigma_1(\mathbf{p},x^+)]^2}{(x^+ - E_p)^2}\right) = 0; \tag{5.2}$$

this result can be obtained by an argument similar to that used to show that (4.5) equals zero. Thus, the sum of the last two terms in (5.1) is equal to

$$3\beta \frac{d\mathbf{p}}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{dx}{\pi} (e^{-\beta x} - e^{-\beta E_p}) \mathrm{Re}\Sigma_1(\mathbf{p},x) \mathrm{Im}\Sigma_1(\mathbf{p},x^+) \mathrm{Re}(x^+ - E_p)^{-2}. \tag{5.3}$$

To see how n_3 depends on the T_0 matrices we must use (3.14) which gives Σ_1 in terms of T_0 and then perform the x integration by using expression (A21) for Im T_0; the reduced T_0 matrix elements (A15) are introduced, and the center-of-mass momentum integration may be performed. Thus, we obtain

$$\begin{aligned}
n_3 \approx{}& 3^{-3/2}\lambda^{-3} \mp \frac{8\lambda^{-3}\beta}{3^{1/2}} \int \frac{d\mathbf{p}}{(2\pi)^3} e^{-4\beta p^2/3m} \mathrm{Re}(\mathbf{p}|T_0(\mathbf{p}^2/m)|\mathbf{p})_s \\
&\pm \frac{8\lambda^{-3}\beta}{3^{1/2}} \mathrm{P.V.} \int \frac{d\mathbf{p}d\mathbf{k}}{(2\pi)^6} \\
&\times \frac{e^{-\beta p^2/3m} e^{-\beta k^2/m} (\mathbf{p}|T_0(\mathbf{k}^2/m^+)|\mathbf{k})(\mathbf{k}|T_0(\mathbf{k}^2/m^-)|\mathbf{p})_s}{k^2/m - p^2/m} \\
&\mp \frac{8\lambda^{-3}\beta}{3^{1/2}} \sum_B \int \frac{d\mathbf{p}}{(2\pi)^3} e^{-\beta p^2/3m} e^{-\beta\epsilon_B} (\epsilon_B - p^2/m)(\mathbf{p}|B)(B|\mathbf{p})_s \\
&+ \frac{32\lambda^{-3}\beta^2}{3^{1/2}} \int \frac{d\mathbf{p}d\mathbf{k}}{(2\pi)^6} e^{-(4\beta/3m)(p^2+k^2+\mathbf{p}\cdot\mathbf{k})} \\
&\times \mathrm{Re}[(\mathbf{p}|T_0(p^2/m^+)|\mathbf{p})_s (\mathbf{k}|T_0(k^2/m^+)|\mathbf{k})_s] \\
&- \frac{64\lambda^{-3}\beta}{3^{1/2}} \int \frac{d\mathbf{p}d\mathbf{k}}{(2\pi)^6} e^{-(4\beta/3m)(p^2+k^2+\mathbf{p}\cdot\mathbf{k})} \mathrm{P.V.} \\
&\int \frac{d\mathbf{l}}{(2\pi)^3} \frac{e^{-(\beta/m)(l^2-k^2)} - 1}{(l^2/m - k^2/m)^2} \\
&\times \mathrm{Re}(\mathbf{p}|T_0(p^2/m + l^2/m - k^2/m)|\mathbf{p})_s \\
&\times (\mathbf{k}|T_0(l^2/m^+)|\mathbf{l})(\mathbf{l}|T_0(l^2/m^-)|\mathbf{k})_s \\
&- \frac{64\lambda^{-3}\beta}{3^{1/2}} \sum_B \int \frac{d\mathbf{p}d\mathbf{k}}{(2\pi)^6} e^{-(4\beta/3m)(p^2+k^2+\mathbf{p}\cdot\mathbf{k})} (e^{-\beta(\epsilon_B - k^2/m)} - 1) \\
&\times \mathrm{Re}(\mathbf{p}|T_0(p^2/m + \epsilon_B - k^2/m)|\mathbf{p})_s (\mathbf{k}|B)(B|\mathbf{k})_s,
\end{aligned} \tag{5.4}$$

where the bound states $|B)$ and energies ϵ_B are defined by (A17) and $|\mathbf{p})_s = |\mathbf{p}) \pm |-\mathbf{p})$. To use this formula the two-particle problem must be solved to obtain the bound state energies and wave functions and the T_0 matrices [on- and half-on-shell T_0 matrices will suffice except for the last two terms in (5.4)]. For low temperatures the leading contribution is the fifth term which involves forward (and backward for exchange) scattering amplitudes only.

The exact result for n_2 can be obtained from either (2.28) or (3.12) and written in a form analogous to the above expression for n_3.[16] For reference we give it here:

$$\begin{aligned}
n_2 ={}& \pm 2^{-3/2}\lambda^{-3} - 2^{3/2}\lambda^{-3}\beta \int \frac{d\mathbf{p}}{(2\pi)^3} e^{-\beta p^2/m} \mathrm{Re}(\mathbf{p}|T_0(\mathbf{p}^2/m)|\mathbf{p})_s \\
&+ 2^{3/2}\lambda^{-3} \mathrm{P.V.} \int \frac{d\mathbf{p}d\mathbf{k}}{(2\pi)^6} \frac{e^{-\beta k^2/m} - e^{-\beta p^2/m}}{(k^2/m - p^2/m)^2} \\
&\times (\mathbf{p}|T_0(k^2/m^+)|\mathbf{k})(\mathbf{k}|T_0(k^2/m^-)|\mathbf{p})_s \\
&+ 2^{3/2}\lambda^{-3} \sum_B e^{-\beta\epsilon_B} - 2^{3/2}\lambda^{-3} \sum_B \int \frac{d\mathbf{p}}{(2\pi)^3} e^{-\beta p^2/m} (\mathbf{p}|B)(B|\mathbf{p})_s.
\end{aligned} \tag{5.5}$$

Our result for the third virial coefficient is obtained by combining (5.4) and (5.5) according to

$$B_3 = \lambda^{12} n_2^2 - \tfrac{2}{3}\lambda^9 n_3. \tag{5.6}$$

Certain parts of our results have been derived previously by other authors. Reiner[5] obtained part of the first order in Σ_1 contribution to n_3, namely, the second term in our Eq. (3.15) [his method was analogous to ours of Sec. 3, but he did not use the formula (A14) for the derivative of the T_0 matrix]; his other terms involve three-particle T matrices. Baumgartl[6] obtained the third term in our Eq. (3.15); he concentrated on terms of second and third order in Re T. Dashen and Ma,[7] in their Eq. (4.23), have obtained the fourth and fifth terms in our result (5.1); they did not attempt to classify the various contributions to B_3 systematically.

We believe the value of our result for B_3 to lie in its generality (as regards the two-particle interaction and statistical and bound states effects), its relative simlicity [as displayed by equation (4.1)], and the fact that it combines the Green function and multiple-scattering theory.

APPENDIX A

Here we present the pertinent details concerning the generalized T matrix as well as some properties of the familiar T_0 matrix. The generalized T matrix is defined by

$$\begin{aligned}
(\mathbf{p}_1\mathbf{p}_2|T(\omega_\nu)|\mathbf{k}_1\mathbf{k}_2) ={}& (\mathbf{p}_1\mathbf{p}_2|V|\mathbf{k}_1\mathbf{k}_2) \\
&+ \int \frac{d\mathbf{l}_1 d\mathbf{l}_2}{(2\pi)^6} (\mathbf{p}_1\mathbf{p}_2|T(\omega_\nu)|\mathbf{l}_1\mathbf{l}_2) Q(\mathbf{l}_1\mathbf{l}_2,\omega_\nu)(\mathbf{l}_1\mathbf{l}_2|V|\mathbf{k}_1\mathbf{k}_2),
\end{aligned} \tag{A1}$$

where $\omega_\nu = i2\pi\beta^{-1}\nu + 2\mu$ $(\nu = 0, \pm 1, \pm 2, \cdots)$, and

$$Q(\mathbf{l}_1\mathbf{l}_2,\omega_\nu) = -\beta^{-1} \sum_n G(\mathbf{l}_1,\omega_n) G(\mathbf{l}_2,\omega_\nu - \omega_n), \tag{A2}$$

and ω_n has either Bose or Fermi character as required.

If it is assumed that both T and Q have activity expansions,

$$T = T_0 + T_1 z \cdots, \tag{A3}$$

$$Q = Q^{(0)} + Q^{(1)} z + \cdots, \tag{A4}$$

it follows that

$$T_0 = V + T_0 Q^{(0)} V, \tag{A5}$$

$$T_1 = T_1 Q^{(0)} V + T_0 Q^{(1)} V$$
$$= T_1 Q^{(0)} V + T_0 Q^{(1)} T_0 - T_0 Q^{(1)} T_0 Q^{(0)} V$$
$$= T_0 Q^{(1)} T_0. \tag{A6}$$

The summation in (A2) can be performed by considering

$$\int_{\gamma'} d\omega\, S(\omega) G(l_1, \omega) G(l_2, \omega_\nu - \omega),$$

where the contour γ' consists of a large circle with branch cuts along the lines $\omega = x$ and $\omega = -x + \omega_\nu$ where $\nu \neq 0$. Thus,

$$Q(l_1 l_2, \omega_\nu) = \mp \int_{-\infty}^{\infty} \frac{dx}{\pi} \mathrm{Re}S(x^+) \mathrm{Im}G(l_1, x^+) G(l_2, \omega_\nu - x) \tag{A7}$$
$$\pm \int_{-\infty}^{\infty} \frac{dx}{\pi} \mathrm{Re}S(\omega_\nu - x^+) G(l_1, \omega_\nu - x) \mathrm{Im}G(l_2, x^+).$$

Noting that $S(\omega_\nu - x) = \mp (1 \mp z e^{-\beta x})^{-1}$ we can compute

$$Q^{(0)}(l_1 l_2, \omega_\nu) = (\omega_\nu - E_{l_1} - E_{l_2})^{-1}, \tag{A8}$$

$$Q^{(1)}(l_1 l_2, \omega_\nu) = \pm (e^{-\beta E_{l_1}} + e^{-\beta E_{l_2}})(\omega_\nu - E_{l_1} - E_{l_2})^{-1}$$
$$+ \int_{-\infty}^{\infty} \frac{dx}{\pi} \frac{\Sigma_1(l_1, \omega_\nu - E_{l_2})}{(\omega_\nu - E_{l_1} - E_{l_2})^2}$$
$$- \int_{-\infty}^{\infty} \frac{dx}{\pi} (\omega_\nu - E_{l_1} - x)^{-1} \mathrm{Im}\left(\frac{\Sigma_1(l_2, x^+)}{(x^+ - E_{l_2})^2}\right). \tag{A9}$$

Combination of these results with (A5) and (A6) gives (when ω is not purely real),

$$(p_1 p_2 | T_0(\omega) | k_1 k_2) = (p_1 p_2 | V | k_1 k_2) + \int \frac{dl_1 dl_2}{(2\pi)^6} \frac{(p_1 p_2 | T_0(\omega) | l_1 l_2)(l_1 l_2 | V | k_1 k_2)}{\omega - E_{l_1} - E_{l_2}} \tag{A10}$$

which is the familiar definition of the standard T matrix, and

$$(p_1 p_2 | T_1(\omega) | k_1 k_2) =$$
$$\pm \int \frac{dl_1 dl_2}{(2\pi)^6} (e^{-\beta E_{l_1}} + e^{-\beta E_{l_2}}) \frac{(p_1 p_2 | T_0(\omega) | l_1 l_2)(l_1 l_2 | T_0(\omega) | k_1 k_2)}{\omega - E_{l_1} - E_{l_2}}$$
$$+ \int \frac{dl_1 dl_2}{(2\pi)^6} \frac{\Sigma_1(l_1, \omega - E_{l_2})(p_1 p_2 | T_0(\omega) | l_1 l_2)(l_1 l_2 | T_0(\omega) | k_1 k_2)}{(\omega - E_{l_1} - E_{l_2})^2}$$
$$- \int \frac{dl_1 dl_2}{(2\pi)^6} (p_1 p_2 | T_0(\omega) | l_1 l_2)(l_1 l_2 | T_0(\omega) | k_1 k_2)$$
$$\times \int_{-\infty}^{\infty} \frac{dx'}{\pi} (\omega - x' - E_{l_1})^{-1} \mathrm{Im}\left(\frac{\Sigma_1(l_2, x'^+)}{(x'^+ - E_{l_2})^2}\right). \tag{A11}$$

Certain special properties of T_0 are used in this paper. If it is assumed that V is invariant under the product of parity and time reversal, then $(p_1 p_2 | V | k_1 k_2) = (k_1 k_2 | V | p_1 p_2)$ implies

$$(p_1 p_2 | T_0(\omega) | k_1 k_2) = (k_1 k_2 | T_0(\omega) | p_1 p_2). \tag{A12}$$

By definition (1.5) it follows that $[T_0(\omega)]^\dagger = T_0(\omega^*)$, and this implies

$$(p_1 p_2 | T_0(x^+) | k_1 k_2)^* = (p_1 p_2 | T_0(x^-) | k_1 k_2). \tag{A13}$$

A formula for the derivative is provided by[6]

$$\frac{\partial}{\partial \omega} (p_1 p_2 | T_0(\omega) | k_1 k_2) = -\int \frac{dl_1 dl_2}{(2\pi)^6} \frac{(p_1 p_2 | T_0(\omega) | l_1 l_2)(l_1 l_2 | T_0(\omega) | k_1 k_2)}{(\omega - E_{l_1} - E_{l_2})^2}, \tag{A14}$$

which follows immediately from the definition (1.5).

The center-of-mass momentum may be separated from these matrix elements according to

$$(p_1 p_2 | T_0(\omega) | k_1 k_2) = (2\pi)^3 \delta(p_1 + p_2 - k_1 - k_2)(p | T_0(P, \omega) | k), \tag{A15}$$

where $P = p_1 + p_2$, $p = (p_1 - p_2)/2$, and $k = (k_1 - k_2)/2$. If $\omega = P^2/4m + \xi$, where ξ is independent of P, then $(p | T_0(P, \omega) | k)$ is independent of P. In particular, if $\omega = P^2/4m + p^2/m \pm i\epsilon$ and $k^2 = p^2$ we write $(p | T_0(p^2/m^\pm) | k)$ which is an *on-shell* T matrix; if $\omega = P^2/4m + p^2/m \pm i\epsilon$ or $\omega = P^2/4m + k^2/m \pm i\epsilon$, but $k^2 \neq p^2$, then $(p | T_0(p^2/m^\pm) | k)$ or $(p | T_0(k^2/m^\pm) | k)$ are *half-on-shell* T matrices.

An alternative expression for T_0 is

$$T_0(\omega) = V + V(\omega - H^{(2)})^{-1} V. \tag{A16}$$

By denoting the eigenstates and eigenvalues of $H^{(2)}$ by $|\alpha\rangle$ and ϵ_α, we obtain

$$\mathrm{Im}(p_1 p_2 | T_0(x^+) | k_1 k_2) = -\pi \sum_\alpha \delta(x - \epsilon_\alpha)(p_1 p_2 | V | \alpha\rangle\langle\alpha | V | k_1 k_2). \tag{A17}$$

The two-particle Hamiltonian may possess a continuous eigenvalue spectrum with eigenstates $|\alpha'\rangle$ and a discrete spectrum with eigenstates $|\alpha_B\rangle$. Thus, we write

$$H^{(2)} |\alpha'\rangle = (P^2/4m + p'^2/m) | P) | p'), \tag{A18}$$

$$H^{(2)} |\alpha_B\rangle = (P^2/4m + \epsilon_B) | P) | B). \tag{A19}$$

Under the assumption that the continuous spectrum of $H^{(2)}$ and the complete spectrum of $H_0^{(2)}$ are identical, we obtain[17]

$$V |\alpha'\rangle = T_0(P^2/4m + p'^2/m^+) | P) | p'). \tag{A20}$$

Then (A15) becomes[18]

$$\mathrm{Im}(p | T_0(P, x^+) | k) = -\pi \int \frac{dp'}{(2\pi)^3} \delta(x - P^2/4m - p'^2/m)$$

$$\times (\mathbf{p}|T_0(p'^2/m^+)|\mathbf{p}')(\mathbf{p}'|T_0(p'^2/m^-)|\mathbf{k})$$

$$-\pi\sum_B \delta(x - P^2/4m - \epsilon_B)(\epsilon_B - p^2/m)(\epsilon_B - k^2/m)(\mathbf{p}|B)(B|\mathbf{k}), \tag{A21}$$

the first term corresponding to the *right-hand cut* with branch point at $P^2/4m$ and the second to the bound-state poles.

APPENDIX B

The meaning of certain integrals which appear frequently in this paper are given here:

$$\int_{-\infty}^{\infty} dx F(x)\mathrm{Im}(x^+ - a)^{-n} = -\pi F^{(n-1)}(a)/(n-1)!, \quad n = 1, 2, \ldots, \tag{B1}$$

$$\int_{-\infty}^{\infty} dx F(x)\mathrm{Re}(x^+ - a)^{-n} = \begin{cases} \mathrm{P.V.}\int_{-\infty}^{\infty} dx F(x)(x-a)^{-n}, & n = 1, 3, \ldots, \\ \mathrm{P.V.}\int_{-\infty}^{\infty} dx[F(x) - F(a)](x-a)^{-n}, & n = 2, 4, \ldots, \end{cases} \tag{B2}$$

where P.V. denotes the principal value of the integral at a. Certain integrals appearing in Sec. 2 and Appendix A are given here:

$$\int_{-\infty}^{\infty} dx F(x)\mathrm{Re}S(x^+) = \begin{cases} \mathrm{P.V.}\int_{-\infty}^{\infty} dx F(x)(z^{-1}e^{\beta x} - 1)^{-1} & \text{(bosons)}, \\ \int_{-\infty}^{\infty} dx F(x)(z^{-1}e^{\beta x} + 1)^{-1} & \text{(fermions)}, \end{cases} \tag{B3}$$

$$\int_{-\infty}^{\infty} dx F(x)\mathrm{Re}S(x^+ - \omega_p) = \pm\,\mathrm{P.V.}\int_{-\infty}^{\infty} dx F(x)(z^{-2}e^{\beta x} - 1)^{-1}, \tag{B4}$$

where P.V. means principal value at μ in (B3) and at 2μ in (B4).

*Research supported in part by the Defence Research Board of Canada, Grant No. 9510-30.

[1]P.C. Martin and J. Schwinger, Phys. Rev. 115, 1342 (1959).
[2]L.P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Benjamin, New York, 1962).
[3]A.S. Reiner, Physica 26, 700 (1960).
[4]C. Bloch, in *Studies in Statistical Mechanics* (North-Holland, Amsterdam, 1965), Vol. III.
[5]A.S. Reiner, Phys. Rev. 151, 170 (1966).
[6]B.J. Baumgartl, Z. Physik 198, 148 (1967).
[7]R. Dashen and S. Ma, J. Math. Phys. 12, 689 (1971).
[8]S.Y. Larsen and P.L. Mascheroni, Phys. Rev. A 2, 1018 (1970).
[9]The state $|\mathbf{p})$ is taken to be an eigenstate of kinetic energy K with eigenvalue E_p and normalization given by $(\mathbf{p}|\mathbf{p}') = (2\pi)^3\delta(\mathbf{p} - \mathbf{p}')$: this implies that $(\mathbf{p}|\mathbf{p})$ should be replaced by Ω. The N-particle state $|\mathbf{p}_1 \cdots \mathbf{p}_N)$ is a product of such states. Note that $(\mathbf{p}_1\mathbf{p}_2|\,T(\omega)\,|\mathbf{p}_1'\mathbf{p}_2')$ has the dimension energy × volume-squared.
[10]R.D. Puff, Ann. Phys. (N.Y.) 13, 317 (1961).
[11]K.M. Watson, Phys. Rev. 89, 577 (1953); 103, 489 (1956).
[12]T.D. Lee and C.N. Yang, Phys. Rev. 113, 1165 (1959).
[13]W.G. Gibson, Phys. Rev. A 6, 2469 (1972).
[14]R. Dashen and S. Ma [J. Math. Phys. 11, 1136 (1970)] show how to express cluster coefficients in terms of on-shell T matrices only using an angular momentum representation. They point out that the momentum representation may be used provided the off-shell quantities are treated with care: this is the spirit of our approach.
[15]Integration by parts is valid here because $\Sigma_1(p, x^+)$ is differentiable for all x, and the boundary term vanishes due to the properties of $\mathrm{Im}(x^+ - E_p)^{-n}$ in Appendix B and of $\mathrm{Im}\Sigma_1(\mathbf{p}, x^+)$ in Appendix A, (A21).
[16]Our expression for n_2 depends on the use of the fact that $\int_{-\infty}^{\infty} dx\,\mathrm{Im}[\Sigma_1(\mathbf{p}, x^+)(x^+ - E_p)^{-2}] = 0$ to handle the derivative of the T_0 matrix which appears when (2.28) or (3.12) is expanded. The result differs in form from that of Baumgartl (Ref. 6).
[17]P. Roman, *Advanced Quantum Theory* (Addison-Wesley, Reading, Mass., 1965).
[18]C. Lovelace, in *Strong Interactions and High Energy Physics*, edited by R.G. Moorhouse (Oliver and Boyd, London, 1964).

Applications of the concept of strength of a system of partial differential equations

K. H. Mariwalla

Matscience, The Institute of Mathematical Sciences, Madras, India
(Received 12 August 1971; revised manuscript received 16 April 1972)

The concept of "strength" of a system of field equations was introduced by Einstein, and is of such generality that one can compare vastly different systems of field equations. We review here this concept in arbitrary number of dimensions and apply it to some of the well-known equations of physics. We calculate the strength, in arbitrary dimensions, of massless Klein–Gordon equations, Maxwell equations (in both potential and field formulation), and Einstein equations. We also determine the strength of massless Dirac equation and Weyl's neutrino equation for the case of four dimensions. It turns out that the strength for all these equations is identical for space–time dimensionality of four. Other possible applications of this concept are indicated.

Einstein in the formulation of his unified field theory was faced with the problem of showing that the system of field equations he had developed was rich enough (in number of solutions) to describe a variety of situations in electromagnetic and gravitational theories without being too vague. The first such attempt he made was when he published his "Generalized Theory of Gravitation" as Appendix II to his Princeton Lectures on Relativity.[1] The unsatisfactory nature of Einstein's analysis was pointed out to him by E. Strauss, and Einstein therefore gave a revised discussion in the fifth edition of his book.[2] In a further edition[3] published by Princeton University Press (1953), Einstein completely revised his approach and introduced the concept of "strength" of a system of field equations. Here he obtains the result that potential and field formulations of Maxwell equations have different strengths[4] for the dimension four. He apparently realized the paradoxical nature of this result. Consequently, in the following (and the last revised) edition[5] of this book he rewrote this portion completely and presented the concept of strength in a very neat and concise form. The potential formulation of Maxwell's equations is, however, not discussed here. Lucid as Einstein's discussion has been, one would expect a plethora of papers applying this concept to other problems in physics. Actually, there is only one such discussion in the literature[6] generalizing Einstein's discussion to an arbitrary number of dimensions d (Einstein had confined himself to the case of four dimensions). Unfortunately, however, this paper contains errors which give a completely wrong picture, of the method of approach. Considering also some of the new applications that we have in mind we give here a complete review of this method and apply it to some well known massless wave equations. Plan of the paper is the following. We first consider simple cases in lower dimensions to give credibility to the terms we introduce and thus arrive at a heuristic definition of the concept of strength of a system of field equations. Next we consider some of the well-known equations of physics in arbitrary number of dimensions.

If we have the first order differential equation

$$\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial t}\right)\phi(x,t)=0,$$

we know its solutions have the form $\phi(x,t)=F(x+t)$; for the second order equation

$$\left(\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial t^2}\right)\phi(x,t)=0,$$

the solutions are somewhat more general of the type $\phi(x,t)=F_1(x+t)+F_2(x-t)$. We say that the first system determines the solutions more strongly than the second system. This statement generalizes to the d-dimensional analogs of these equations ($\partial_\mu=\partial/\partial x_\mu$):

$$\sum_{\mu=1}^{d}\partial_\mu\phi(x_1,\ x_2,\ \ldots,\ x_d)=0, \tag{1}$$

$$\sum_{\mu=1}^{d-1}(\partial_\mu^2-\partial_d^2)\phi(x_1,x_2,\ldots,x_d)=0. \tag{2}$$

We are interested in obtaining a numerical characterization of the concept of "strength" that would enable us to compare different systems of equations as in the above examples. Suppose we have an analytic field function of d variable. We can expand it in Taylor series and the totality of its coefficients describe the field completely. Let us consider the terms of the nth order of differentiation in the Taylor expansion; these are of the form

$$\partial_{\mu_1}\partial_{\mu_2}\cdots\partial_{\mu_n}\phi,\quad \mu_j=1,\ 2,\ \cdots,\ d$$

and the total number of such coefficients is

$$\left[\begin{matrix}d\\n\end{matrix}\right]\equiv\frac{(d+n-1)!}{(d-1)!N!}\equiv\binom{d+n-1}{n}.$$

If the function satisfies some field equations or constraints, these give several relations between the nth order coefficients, thereby reducing the number of coefficients left free to be assigned arbitrary values. If we denote the number of coefficients thus left free by Z_n, then it is clear that from our viewpoint the quantity of importance is the fraction of the total number of nth order coefficients left free, i.e., $Z_n/\left[\begin{matrix}d\\n\end{matrix}\right]$. For instance, if the field equation satisfied by our function is (1), then $(n-1)$-fold differentiation of it gives $\left[\begin{matrix}d\\n-1\end{matrix}\right]$ relations

$$\sum_{\mu=1}^{d}\partial_\mu\partial_{\mu_1}\partial_{\mu_2}\cdots\partial_{\mu_{n-1}}\phi=0$$

between the nth order coefficients. In particular for $d=2$, these relations are

$$\frac{\partial\phi}{\partial t}=\frac{-\partial\phi}{\partial x},\quad \frac{\partial^2\phi}{\partial t^2}=\frac{-\partial^2\phi}{\partial x\,\partial t}=\frac{\partial^2\phi}{\partial x^2},\quad \text{etc.},$$

and when substituted back into the Taylor series yield the solutions

$$\phi(x,t)=\phi(x_0,t_0)+\frac{\partial\phi}{\partial x}\Big|_0(x-t)+\cdots=F(x-t).$$

The number of nth order coefficients remaining free are

$$Z_n^{(1)}=\begin{bmatrix}d\\n\end{bmatrix}-\begin{bmatrix}d\\n-1\end{bmatrix}=\begin{bmatrix}d\\n\end{bmatrix}\left\{1-\left(1-\frac{d-1}{n}\right)^{-1}\right\}.$$

If on the other hand the field equation is Eq. (2), then the relations between the nth order coefficients are obtained by $(n-2)$-fold differentiation of (2); these are $\left[\begin{smallmatrix}d\\n-2\end{smallmatrix}\right]$ in number, and therefore

$$Z_n^{(2)}=\begin{bmatrix}d\\n\end{bmatrix}\left(1-\frac{n(n-1)}{(d+n-1)(d+n-2)}\right).$$

In order to compare these, it is clear that we must expand these for large n, in powers of $1/n$, otherwise the result would strongly depend on n; such expansion to the first order in $1/n$ is

$$Z_{n\,\text{large}}^{(1)}\approx\begin{bmatrix}d\\n\end{bmatrix}\left(0+\frac{d-1}{n}\right), \tag{1a}$$

$$Z_{n\,\text{large}}^{(2)}\approx\begin{bmatrix}d\\n\end{bmatrix}\left(0+2\frac{d-1}{n}\right). \tag{2a}$$

Since there is only one field variable involved in our analysis, the vanishing of the first (constant) term indicates that no function of d variables $(x_1, x_2, \cdots, x_d)$, which satisfies the field equations is left free (i.e., without restriction). Therefore, this term, in general, gives the number of functions of d variables left free. The comparison of the coefficients of $1/n$ shows that this coefficient for the system (2) is twice as large as for the system (1); since the system (2) is known to be weaker (as we saw from the explicit example in the two dimensional case) we conclude that this coefficient gives a measure of "strength" of the system: the larger this coefficient, the weaker is the system.

Now let us indicate a procedure for determining strength of a general system of field equations in arbitrary dimensions. Let the number of field variables be k; then, the number of nth-order coefficients are

$$N_n=k\begin{bmatrix}d\\n\end{bmatrix}.$$

Frequently equations arise which are not homogeneous in the order of differentiation (i.e., the field equations contain derivatives of different orders of a given variable). In that case one must consider these derivatives as independent variables together with the subsidiary conditions relating them. In the general case, therefore, one has several field variables $k_1, k_2\cdots(k=\Sigma_j k_j)$ satisfying field equations of orders $r_1, r_2, \cdots$ together with $P=\Sigma_\alpha p_\alpha$ identities, each of order l_α. These various numbers k, r, l, p can be expressed in terms of dimension d. Thus the number of conditions M_n (due to the field equations together with the constraints) on the nth order coefficients has the form

$$M_n(k_i, r_\alpha; p_\alpha, l_\alpha; d)=M_n(d).$$

The number of nth order coefficients remaining free is then given by

$$Z_n=N_n-M_n=\begin{bmatrix}d\\n\end{bmatrix}\left\{k-M_n\Big/\begin{bmatrix}d\\n\end{bmatrix}\right\}.$$

If we expand the expression in the parenthesis for large n making use of the expansion formulas

$$\begin{bmatrix}d\\n\pm r\end{bmatrix}\approx\begin{bmatrix}d\\n\end{bmatrix}\left(1\pm r\frac{d-1}{n}\cdots\right), \tag{3}$$

we get

$$Z_n=\begin{bmatrix}d\\n\end{bmatrix}\left(z_0+\frac{z_1}{n}+\frac{z_2}{n^2}+\cdots\right).$$

It is clear that Z_n is always ≥ 0 for all n, for an absolutely compatible system. The constant z_0 gives the number of functions of d variables $(x_1, x_2\cdots x_d)$ left free; it is clearly desirable that this number be zero for any physically reasonable system. The coefficient of $1/n$, z_1, is of special interest and gives measure of strength of the system of field equations under consideration.

Larger the value of z_1, weaker is the system. We call z_1, the coefficient of freedom. We now proceed to illustrate the calculation of "strength" for some of the well-known field equations.

MAXWELL EQUATIONS IN TERMS OF FIELD STRENGTHS

Field strengths of the electromagnetic field are given as components of an antisymmetric tensor of the second rank, which we denote by $F_{\mu\nu}$. Since the number of components of an antisymmetric tensor of the kth rank is $\binom{d}{k}$, the number of nth order coefficients in the expansion of $F_{\mu\nu}$ is

$$N_n=\binom{d}{2}\begin{bmatrix}d\\n\end{bmatrix}=\frac{1}{2}d(d-1)\begin{bmatrix}d\\n\end{bmatrix}. \tag{4}$$

The field equations in the absence of sources, are

$$U^\mu\equiv\partial_\nu F^{\mu\nu}=0, \tag{5}$$

$$h_{\mu\nu\sigma}\equiv\partial_\sigma F_{\mu\nu}+\partial_\mu F_{\nu\sigma}+\partial_\nu F_{\sigma\mu}=0. \tag{6}$$

Both these are of the first order. Since $h_{\mu\nu\sigma}$ is a completely antisymmetric tensor of the third rank it represents $\binom{d}{3}$ conditions, whereas the Eqs. (5) are d in number. The number of conditions on the nth order coefficients that the field equations impose therefore are

$$\left[d+\binom{d}{3}\right]\begin{bmatrix}d\\n-1\end{bmatrix} \tag{7}$$

and may be obtained by $(n-1)$-fold differentiation of the field equations. These are to be subtracted from (4). All these conditions are not independent because of the existence of identities. One such identity is

$$\partial_\mu\partial_\nu F^{\mu\nu}=0, \tag{8}$$

which holds in virtue of the antisymmetry of $F_{\mu\nu}$, and gives

$$\begin{bmatrix}d\\n-2\end{bmatrix} \tag{9}$$

relations between the conditions (on nth order coefficients) due to the field equations and are to be subtracted from them [i.e., from Eq. (7)]. If we were working in 3-dimensions these equations are all we need as $h_{\mu\nu\sigma}$ represents just one independent component. For $d\geq 4$, the cyclic divergence of $h_{\mu\nu\sigma}$

$$J_{\mu\nu\sigma\lambda}\equiv h_{\mu\nu\sigma,\lambda}-h_{\nu\sigma\lambda,\mu}+h_{\sigma\lambda\mu,\nu}-h_{\lambda\mu\nu,\sigma} \tag{10}$$

also vanishes, on account of which not all the conditions (6) are independent. Equations (10) furnish

$$\binom{d}{4}\left[\begin{matrix} d \\ n-2 \end{matrix}\right] \tag{11}$$

relations to be subtracted from (7). For $d \geq 5$ not all the relations (10) are independent as the cyclic divergence of $J_{\mu\nu\sigma\lambda}$ also vanishes, so that we get

$$\binom{d}{5}\left[\begin{matrix} d \\ n-3 \end{matrix}\right]$$

relations to be substracted from (11). Enumerating this telescopic series of identities we get for an arbitrary dimension $k=d$, the total number of

$$\left[\begin{matrix} d \\ n-2 \end{matrix}\right] + \sum_{r=2}^{d=k} \binom{d}{2+r} (-1)^r \left[\begin{matrix} d \\ n-r \end{matrix}\right] \tag{12}$$

relations between the conditions on the nth order coefficients due to field equations, and are to be subtracted from (7) to obtain the number of independent conditions on the nth order coefficients. Collecting terms (Eqs. 4, 7 and 12) we obtain

$$Z_n = \binom{d}{2}\left[\begin{matrix} d \\ n \end{matrix}\right] - \left\{\left[d + \binom{d}{3}\right]\left[\begin{matrix} d \\ n-1 \end{matrix}\right] + \left[\begin{matrix} d \\ n-2 \end{matrix}\right] - \sum_{r=2}^{d} (-1)^r \binom{d}{2+r}\left[\begin{matrix} d \\ n-r \end{matrix}\right]\right\}. \tag{13}$$

If we expand this for large n, we get on using formula (3) and rearranging terms

$$Z_n \approx \left[\begin{matrix} d \\ n \end{matrix}\right]\left\{\sum_{r=0}^{d} (-1)^r \binom{d}{r} + \frac{d-1}{n}\left[2(d-2) + \sum_{r=0}^{d} (-1)^r (r-2)\binom{d}{r}\right]\right\}.$$

Since

$$r\binom{d}{r} = d\binom{d-1}{r-1}, \quad \sum_{r=0}^{k} (-1)^r \binom{d}{r} = (-1)^h \binom{d-1}{h},$$

and the latter vanishes for $d \leq h$, we finally get

$$z_0 = 0, \quad z_1 = 2(d-1)(d-2). \tag{14}$$

This result agrees with that of Einstein[5] for $d=4$, but does not agree with the work of Penney.[6,7] We now turn to discuss the potential formulation.

MAXWELL EQUATIONS IN TERMS OF POTENTIALS

In terms of the potentials A^μ, the field equations of Maxwell take the form

$$j^\mu \equiv (\partial^\nu \partial_\mu \eta^{\mu\alpha} - \partial^\mu \partial^\alpha) A_\alpha = 0. \tag{15}$$

The number of components of A^μ is d, and the number of nth order coefficients in its expansion is $d\binom{d}{n}$.

Actually, only a part of these nth order coefficients serve to characterize essentially different potentials due to the freedom of the gauge:

$$A^\mu \to A'^\mu = A^\mu + \partial^\mu \phi,$$

where ϕ is an arbitrary (but differentiable, etc.!) function of x^μ. The A^μ and A'^μ represent in fact the same electromagnetic potentials. If the above equation is differentiated n times with respect to x, one notices that all the $(n+1)$th derivatives of the function ϕ with respect to x enter into the nth order coefficients of the expansion of A'^μ; i.e., there appear (in the nth order terms of differentiation) $\left[\begin{matrix} d \\ n+1 \end{matrix}\right]$ numbers that have no part in the characterization of the potentials. Therefore, in any theory involving this type of freedom of the gauge, one must subtract this number from the total number of nth order coefficients to take account of the gauge invariance.[8] Thus, the "relevant" coefficients of the nth order are only

$$d\left[\begin{matrix} d \\ n \end{matrix}\right] - \left[\begin{matrix} d \\ n+1 \end{matrix}\right]$$

in number.

The number of conditions on these nth order coefficients due to the fields equations are $d\left[\begin{matrix} d \\ n-2 \end{matrix}\right]$. Between them there are $\left[\begin{matrix} d \\ n-3 \end{matrix}\right]$ relations, due to the 3rd order identity

$$\partial_\mu j^\mu = 0, \tag{16}$$

so that the number of independent conditions on the nth order coefficients are $d\left[\begin{matrix} d \\ n-2 \end{matrix}\right] - \left[\begin{matrix} d \\ n-3 \end{matrix}\right]$. Collecting terms, we get for the number of nth order coefficients remaining free

$$Z_n = d\left[\begin{matrix} d \\ n \end{matrix}\right] - \left[\begin{matrix} d \\ n+1 \end{matrix}\right] - d\left[\begin{matrix} d \\ n-2 \end{matrix}\right] + \left[\begin{matrix} d \\ n-3 \end{matrix}\right].$$

On expanding for large n we obtain

$$z_0 = 0, \quad z_1 = 2(d-1)(d-2) \tag{17}$$

in complete agreement with the strength as obtained for the "field strength" formulation.

In obtaining the value of the "strength" in the above, we have made no use of any gauge condition, as it is only an unnecessary complicating factor and does not add anything to the clarity of the discussion. For instance, if we introduce the Lorentz condition as a gauge condition, the system of equations we get are

$$V^\mu \equiv \partial^\nu \partial_\nu A^\mu = 0, \quad \partial_\mu A^\mu = 0, \tag{18}$$

and the identity

$$\partial_\mu V^\mu = 0. \tag{19}$$

Comparing it with the system (15,16) we see that the second of the Eqs. (18)—the Lorentz condition—is an additional field equation, so that the number of independent conditions on the nth order coefficients is increased by $\left[\begin{matrix} d \\ n-1 \end{matrix}\right]$ (since this equation is of the first order). On the other hand, because of the Lorentz condition there is a reduction in the arbitrariness of the potentials as the function ϕ is now no longer completely arbitrary but has to satisfy the field equation

$$\Box^2 \phi = \partial_\mu (\partial^\mu \phi) = 0.$$

Since this is of the second order, the number of terms that appear in the nth order of differentiation of A^μ but play no part in characterization of the potentials is reduced by $\left[\begin{matrix} d \\ n+1-2 \end{matrix}\right]$. Therefore, the number of "relevant" coefficients of nth order are

$$d\left[\begin{matrix} d \\ n \end{matrix}\right] - \left[\begin{matrix} d \\ n+1 \end{matrix}\right] + \left[\begin{matrix} d \\ n-1 \end{matrix}\right].$$

In the final analysis there is no net change in the expression for Z_n and hence also in the values of z_0 and z_1. The value of the strength is thus a gauge invariant quantity as it should be.[8] It is, however, important to note that from a group theoretical viewpoint the system (18,19) is not "equivalent" to the system (15,16) as the latter system is invariant under the pseudogroup of

conformal coordinate changes,[9] whereas the former does not possess this symmetry.

EINSTEIN EQUATIONS OF GENERAL RELATIVITY

The field equations for the source-free gravitational field are

$$R_{ik}=0, \tag{20}$$

where R_{ik} are components of the symmetric Ricci tensor which is obtained as a contraction of the Riemann tensor,

$$R^h_{ijk}=\Gamma^h_{lj}\Gamma^l_{ik}-\Gamma^l_{ij}\Gamma^h_{lk}+\partial_j\Gamma^h_{ik}-\partial_k\Gamma^h_{ij},$$

where Γ's are the Christoffel affinities. We notice that the Riemann tensor contains terms in both zero*th* and first order of differentiation in the affinities. Because of the various symmetries (which we need not go into here), the Riemann tensor has only $\frac{1}{12}d^2(d^2-1)$ independent components. In addition, it satisfies two differential identities[10] known after Bianchi and Veblen; both these identities are of the first order of differentiation in the Riemann tensor and reduce for the contracted tensor to the set of d equations (Einstein identity)

$$(R^i_k-\tfrac{1}{2}R\delta^i_k)_{;i}=0. \tag{21}$$

At this point we notice that in addition to the affinities, there enter the field equations and the identities, the components of the metric tensor. In fact, metric tensor is essential for raising and lowering indices and for contraction. The property of the metric tensor that makes it so useful is the differential condition

$$g_{\mu\nu;\xi}\equiv g_{\mu\nu,\xi}-g_{\alpha\nu}\Gamma^\alpha_{\mu\xi}-g_{\mu\beta}\Gamma^\beta_{\xi\nu}=0. \tag{22}$$

One may consider this condition as a field equation for the $g_{\mu\nu}$; it may be considered as an equation giving the affinities in terms of the first order derivatives of the metric tensor components. On account of this the affinities are to be considered as of the first order of differentiation and field equations as of the second order of differentiation. The Einstein identity (21) is then of the third order of differentiation. It is, further more, clear that g's and Γ's are to be considered as independent variables, since the field equations are not homogeneous in the order of differentiation (they are nonlinear).

Since affinities are symmetric with respect to the interchange of two of its indices, there are $\frac{1}{2}d^2(d+1)$ independent components. The metric tensor has on account of its symmetry $\frac{1}{2}d(d+1)$ components. Therefore, the total number of nth order terms in the expansion of the fields g and Γ is

$$\frac{1}{2}d(d+1)\begin{bmatrix}d\\n\end{bmatrix}+\frac{1}{2}d^2(d+1)\begin{bmatrix}d\\n-1\end{bmatrix},$$

where we have taken account of the fact that Γ are to be taken as of the first order of differentiation. Because of the general covariance of the theory, a situation analogous to that of Maxwell's equations in potential formulation arises here. In that case, all potentials differing only in "gauge" (i.e., connected by transformations of the gauge group) were considered equivalent, as the field equations were unaffected under transformations of the gauge group. In the present case, general covariance means that field equations are invariant under the groupoid of coordinate transformations. This means that if $g_{ij}(x)$ are representatives of the components of the metric tensor then so are all the other representatives related to $g_{ij}(x)$ by a coordinate transformation

$$g'_{lm}(x')=\frac{\partial x^i}{\partial x'^l}\frac{\partial x^j}{\partial x'^m}g_{ij}(x),$$

so that g and g' are to be considered as equivalent. If we make a Taylor series expansion of g', it's nth order terms would contain $(n+1)$-th order derivatives of the d-functions x^i, which have therefore no role in characterization of the field. Thus, from the number of nth order coefficients of g_{ij} one must subtract $d\begin{bmatrix}d\\n+1\end{bmatrix}$. Let us now consider the number of relations between the nth-order coefficients due to the field equations. First of the field equations (20) are $\frac{1}{2}d(d+1)$ in number and are of the second order. These therefore furnish $\frac{1}{2}d(d+1)\begin{bmatrix}d\\n-2\end{bmatrix}$ relations between the nth order coefficients. The second set of equations (22) similarly gives $\frac{1}{2}d^2(d+1)\begin{bmatrix}d\\n-1\end{bmatrix}$ relations. All of these relations are not independent; between these there are $d\begin{bmatrix}d\\n-3\end{bmatrix}$ identities due to the d Einstein identities (21) of the third order. Collecting terms we get for the number of nth order terms remaining free:

$$Z_n=\binom{d}{2}\begin{bmatrix}d\\n\end{bmatrix}+d\binom{d}{2}\begin{bmatrix}d\\n-1\end{bmatrix}-d\begin{bmatrix}d\\n+1\end{bmatrix}-\left\{\binom{d}{2}\begin{bmatrix}d\\n-2\end{bmatrix}+d\binom{d}{2}\begin{bmatrix}d\\n-1\end{bmatrix}-d\begin{bmatrix}d\\n-3\end{bmatrix}\right\}. \tag{23}$$

In this expression we see that the terms arising from considering Γ's as independent variables cancel the terms arising from the field equations (22). It is tempting therefore to altogether disregard the role of the Γ's and consequently ignore also the field equations (22). In fact Penney, who obtains the same result as above, actually ignores these contributions, which however because of their accidental cancellation make no difference in the final result. However, as outlined in detail, above, such enumeration is in principle incorrect.

If we expand (23) for large n, we get

$$z_0=0,\quad z_1=d(d-1)(d-3). \tag{24}$$

We note paranthetically that our considerations strictly exclude $d\leq 3$, since for $d\leq 3$ the field equations (20) imply that space is flat.[11]

WEYL'S NEUTRINO EQUATION

Written in the form of a wave equation, the Weyl's free field neutrino equation is [12,13]

$$\boldsymbol{\sigma}\cdot\nabla\psi=-\frac{\partial}{\partial t}\psi, \tag{25}$$

where ψ is a two component spinor, and $\boldsymbol{\sigma}=(\sigma_1,\ \sigma_2,\ \sigma_3)$ are the Pauli matrices[12] which satisfy

$$[\sigma_i,\sigma_j]_+=2\delta_{ij},\quad \sigma_i\sigma_j=i\sigma_k. \tag{26}$$

It is important to consider explicitly the properties of σ's in detail for two reasons. It clarifies the status of this equation as a representation of the Lorentz group and yields the relevant identities. It enables one to determine the number of independent components of ψ correctly. The spinor field ψ has 2 complex components

and is arbitrary up to a phase factor. Because of this circumstance and properties of the Pauli matrices, ψ and $\sigma_2\psi^*$ are to be considered as the same representation

$$\psi \approx \sigma_2\psi^*. \tag{27}$$

These are two equations in two complex variables and imply that ψ has only two independent real variables. Eliminating σ we find that the two real components must satisfy the second order equation

$$(\nabla^2 - \partial_t^2)\psi = 0. \tag{28}$$

This equation being of the type of Eq. (2), its strength is given by

$$z_0 = 0, \quad z_1 = 12. \tag{29}$$

Now let us consider the massless Dirac equation[13] in four dimensions. The field equation for the four component object (bispinor) ψ is

$$\gamma_\mu \partial_\mu \psi = 0, \quad \mu = 1, 2, 3, 4, \tag{30}$$

where the matrices[12] $\{\gamma\}$ consist of five anticommuting matrices satisfying

$$\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu},$$

such that the fifth matrix γ_5, as also 10 other matrices including the identity, may be obtained as some products of the other four; these sixteen matrices form an algebra (Clifford algebra).[12] Because γ_5 anticommutes with the four γ_μ in (30), the fields ψ and their transform by γ_5 are in fact the same representation:

$$\psi \approx \exp(i\alpha\gamma_5)\psi, \quad \alpha \text{ real}. \tag{31}$$

This reduces the number of independent components of ψ to two complex components. Also, there exists a matrix[14] C_1, such that $C_1\psi^*$ and ψ satisfy the same equation (alternately,[15] one may make appeal to the existence of a Majorana representation)

$$C_1\psi^* \approx \psi. \tag{32}$$

The Eqs. (31) and (32) together imply that there are just two real independent components which on elimination of γ_μ satisfy Eqs. (28) so that its "strength" is given by (29). This result is in agreement with the well-known equivalence[15,13] of these two formulations of the neutrino equations.

DISCUSSION

In the above we have discussed and obtained the values of the strength of the following systems of partial difference equations of physical interests. Scalar wave equation, the Maxwell equations in the potential and "field strength" formulations and Einstein equations of general relativity, all for arbitrary dimensional space; Weyl neutrino equation and massless Dirac equations in four dimensions. We find that, as expected, the Maxwell equations in the potential and "field strength" formulation have the same strength for all dimensionalities. The calculation in the literature[6] of the strength for the Maxwell equations in field formulation in arbitrary number of dimensions is found to be incorrect. If we compare the strengths for the various equations of physical interest we find that the space-time dimensionality *four* stands out as a preferred dimension for which the strengths of the following zero mass wave equations are identical: complex Klein—Gordon field, Maxwell field, Einstein's gravitational field, neutrino field. The strength parameters are $z_0 = 0$, $z_1 = 12$.

Considering the generality of the method of comparing strengths of two systems of field equations we feel that other possible applications are indicated. For instance, recently there has been a great proliferation of various kinds of wave equations in relativistic quantum theory. The absolute conditions that a set of wave equations should satisfy are that they pose a well-defined Cauchy problem[16] and furnish a rich enough set of solutions so as to provide an integrable representation[17] of the Lorentz group. For the latter condition to be satisfied the following seems essential: (a) Since action of the Lorentz group on various momentum hyperboloids is transitive, any wave equation that leads to space-like momenta[18] is to be *a priori* excluded. It is not enough to just exclude states with space-like momenta; (b) The wave equation should satisfy restrictions due to the two Casimir invariants.[19] The operators of the Poincaré algebra, as also the sum of the squares of these operators, should be self-adjoint on the domain defined by the solutions of the wave equation. (c) This domain should be invariant under the Lorentz group; a related requirement is stability under a scalar perturbation.

A situation can arise when all these conditions are satisfied by more than one wave equation for the same mass and spin. It would then be useful to have a test, in the absence of any other compelling physical requirement, to enable one to single out a preferred system of equations. We thus see that the concept of strength of a system of field equations is precisely of the type necessary for this purpose. Considering the great generality of this concept it can be applied to compare systems that differ both in the number of fields involved as well as in kind. Thus, for example, we may require that all physically significant zero mass equations have the same strength, 12. This is in fact the case for the equations we considered above. In a subsequent paper we plan to extend these considerations to other types of equations.

[1] A. Einstein, *Meaning of Relativity* (Methuen, London, 1950), 4th ed.
[2] A. Einstein, *Meaning of Relativity* (Methuen, London, 1951), 5th ed.; this was also published by Princeton U. P., Princeton, N. J., 1950, as its 3rd ed.
[3] A. Einstein, *Meaning of Relativity* (Princeton U. P., Princeton, N. J., 1953), 4th ed. This edition has no analog in the Methuen series. I am grateful to the referee for pointing out the existence of this work. This edition unfortunately is not available outside the area served directly by the Princeton University Press (e.g., India), where Methuen has the distribution rights.
[4] I am indebted to the referee for this information.
[5] A. Einstein, *Meaning of Relativity* (Methuen, London, 1956), 6th ed.; also Princeton U. P., Princeton, N. J., 1955, 5th ed.
[6] R. Penney, J. Math. Phys. **6**, 1607 (1965). The errors in this paper first came to my attention when the referee criticized the disparity between the strength for Maxwell equations in potential formulation as given here and that calculated by Penney for the field formulation of these equations. If turns out that Penney's generalization of Einstein's calculation of strength of Maxwell equations is valid only for $d \leq 4$. His presentation of the calculation of the strength of Einstein equation is also misleading and his discussion of the strength of Weyl and massless Dirac equation is wrong.

[7]The reason for this disagreement is that Penney has disregarded the additional cyclic identities that arise in higher dimensions. See also the discussion at the end of this paper.

[8]That "strength" for the case of potential formulation has to be a gauge invariant quantity is already clear from Einstein's discussion of the case of "coordinate gauge freedom" in general relativity. In fact calculating the strength of any system one must take into account all the relevant symmetries; therein lies the value of this concept.

[9]The conformal invariance of the system (15, 16) follows from the conformal invariance of the system (5,6). See also K. H. Mariwalla, Lett. Nuov. Cim. **4**, 295 (1972).

[10]L. P. Eisenhart, *Non-Riemannian Geometry* (Am. Math. Soc. Coll. Publs., New York, 1927), Vol. VIII.

[11]For a (2+1) space–time, even though R^{h}_{ijk} vanish, the space–time need not be *globally* flat. For remarks in this connection, see J. A. Wheeler in *Relativity, Groups and Topolgy* (Les Houches, Paris, 1963), edited by C. Dewitt and B. Dewitt (Gordon and Breach, New York, 1964), pp. 519–20, Problem 56.

[12]See, e.g., A. Ramakrishnan, *Elementary Particles and Cosmic Rays* (Pergamon, London, 1962). For a very lucid discussion of Pauli matrices their generalization to higher dimensions and to generalized Clifford algebras, see Alladi Ramakrishnan, *L-Matrix Theory or Grammar of Dirac Matrices* (Tata McGraw-Hill, Bombay, New Delhi, India 1972), and J. Math. Anal. Appl. **20**, 9 (1967).

[13]B. Kursunoglu, *Modern Quantum Theory* (Freeman, London, 1962).

[14]K. H. Mariwalla, Rev. Mod. Phys. **34**, 215 (1962).

[15]C. Ryan and S. Okubo, Nuovo Cimento Suppl. **2**, 234 (1964).

[16]The Cauchy problem for a partial differential operator $D = \partial_1^m + \Sigma_{r=1}^{m} p_r(\partial_2, \partial_3, ..., \partial_d)\partial_1^{m-r}$, where p_r are polynomials, is to determine ψ such that $D\psi = \varphi(x_1, x_2, ..., x_d)$ and $\partial_1^j \psi|_{x_1=0} = \chi_j(x_2, x_3, ..., x_d)$, $j = 0,1,...,m-1$, where ψ, ϕ, χ are functions of a specified type. Associated with the Cauchy initial value problem is an Abelian group which transforms the Cauchy data from $x=0$ to another x.

[17]For instance in the *formal* Lie algebra $[q, p] = i$, $p = i\partial/\partial q$, if $q = (-\infty, +\infty)$, it integrates to a group; but that is not the case if the range of q is in any way restricted. See also E. Nelson, Ann. Math. **70**, 572 (1959); M. Born and P. Jordan, *Elementare Quantenmechanik* (Berlin, 1930), p. 128.

[18]For instance, the so-called Weinberg equation is unacceptable from this viewpoint. We note parenthetically that already from the point of view of conformal invariance the Weinberg equation exhibits anomalous behavior. See, e.g., M. Flato *et al.*, Ann. Phys. (N.Y.) **61**, 93 (1970).

[19]Not all wave equations in the literature satisfy this condition.

Occupation statistics from exact recursion relations for occupation by dumbbells of a $2 \times N$ array

M. A. Rashid

Institute of Mathematics University of Islamabad, Islamabad, Pakistan
(Received 7 July 1972; final manuscript received 3 October 1972)

Starting from an exact recursion relation of the form $A(q, N) = \Sigma_{i,j} A(q - i, N - j)C_{ij}$, we present a recursion method for calculating the various moments $\mu_N(m) = \Sigma_{q=0}^{N} q^m A(q, N)$. These results are applied to obtain the occupation statistics for occupation of a $2 \times N$ array by parallel and not necessarily parallel dumbbells.

INTRODUCTION

Study of the regular space lattices is useful in understanding a variety of physical phenomenon. In particular, the problem of occupation of a two-dimensional lattice space by dumbbells appears in the theory of adsorption of diatomic molecules.[1]

Exact solutions for such problems are usually known only for the one-dimensional lattices.[2] Recently McQuistan *et al.* have attacked the problem of obtaining the occupation statistics for the placement of dumbbells on a $2 \times N$ array, a dumbbell occupying adjacent sites. They have given exact recursion relations for the two situations—when all the dumbbells are oriented parallel to the long (N) dimension of the array, or are placed unrestrictedly.[3,4] The first problem is simpler, and they have been able to calculate the pseudovariance, using an exact solution of the recursion relation and some approximation techniques.

The purpose of the present paper is to show that we can obtain these results from the recursion relations and the initial conditions only *without obtaining a general solution*. This greatly increase the domain of validity of the method. In particular, we consider the problem of finding the number of arrangements of dumbbells on a $2 \times N$ array and the corresponding moments and occupation statistics for the two situations referred to above.

NORMALIZATION

Let us start with a recursion relation

$$A(q,N) = \sum_{i=0}^{I} \sum_{j=0}^{J} C_{ij} A(q-i, N-j), \tag{1}$$

where C_{ij} are constants and $A(q,N)$ are defined only for nonnegative values of q and N. Also $A(q,N) = 0$ whenever $q > N$. Define

$$C_j = \sum_{i=0}^{I} C_{ij} \tag{2}$$

and

$$\Delta_N = \sum_{q=0}^{N} A(q,N). \tag{3}$$

In this section, we determine the normalization Δ_N for our distribution.

From (1), (2), and (3)

$$\Delta_N = \sum_{q=0}^{N} A(q,N) = \sum_{j=0}^{J} C_j \Delta_{N-j}. \tag{4}$$

Equation (4) is a recursion relation satisfied by Δ_N.

Now multiply Eq. (4) by x^N and sum over N from $N = J$ to ∞. This gives

$$G(x) = \sum_{N=0}^{\infty} x^N \Delta_N = \left(\sum_{p=0}^{J-1} x^p \Delta_p - \sum_{j=0}^{J-1} C_j x^j \sum_{p=0}^{J-j-1} x^p \Delta_p \right) \Big/ \left(1 - \sum_{j=0}^{J} C_j x^j \right). \tag{5}$$

In the above equation, we have expressed $G(x)$, the generating function for Δ_N, as a quotient of two polynomials in x, with the numerator having a lesser degree than the denominator. Note that to compute $G(x)$, we need the J initial values Δ_p $(0 \leq p \leq J-1)$.

We can, therefore, express $G(x)$ in terms of partial fractions. Suppose that the roots of the equation

$$1 - \sum_{j=0}^{J} C_j x^j = 0 \tag{6}$$

are all unequal. Then Eq. (5) can be rewritten as

$$G(x) = \sum_{N=0}^{\infty} x^N \Delta_N = \sum_{j=0}^{J} \frac{k_j}{1 - S_j x} = \sum_{N=0}^{\infty} \left(\sum_{j=0}^{J} k_j S_j^N \right) x^N, \tag{7}$$

i.e.,

$$\Delta_N = \sum_{j=0}^{J} k_j S_j^N. \tag{8}$$

If $1/S = 1/S_1$ is the root with the smallest absolute value, and $k_1 = k$, we have

$$\lim_{N \to \infty} \Delta_N = k S^N, \tag{9}$$

where

$$k = \lim_{Sx \to 1} (1 - Sx) G(x). \tag{10}$$

Equations (9) and (10) are valid whenever the root with the smallest absolute value is nondegenerate. For the case when this root is degenerate, a modification of these equations can be obtained easily. We do not present this modification since our examples do not require it.

HIGHER MOMENTS

Define[5]

$$\mu_N(m) = \frac{1}{\Delta_N} \sum_{q=0}^{N} q^m A(q,N) \tag{11}$$

$$\simeq \sum_{i=0}^{m} \mu_i(m) N^i \quad \text{for large } N. \tag{12}$$

In Eqs. (11), (12), $m = 0, 1, 2, \ldots$ and

$$\mu_0(m) = \delta_{m0}. \tag{13}$$

Now multiply the recursion relation in Eq. (1) by q^m and sum both sides from 0 to N to obtain

$$\mu_N(m)=\frac{1}{\Delta_N}\sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{q=0}^{N} q^m C_{ij}A(q-i,\ N-j)$$

$$=\frac{1}{\Delta_N}\sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{k=0}^{m} C_{ij}(m_k)i^{m-k}\Delta_{N-j}\mu_{N-j}(k). \tag{14}$$

In the last step, we have manipulated the q-summation and used Eq. (11).

Finally we make use of Eqs. (9) and (12) to arrive at

$$\sum_{r=0}^{m}\mu_r(m)N^r=\sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{k=0}^{m}\sum_{l=0}^{k}\sum_{r=0}^{l}$$
$$\times[C_{ij}(m_k)i^{m-k}S^{-j}\mu_l(k)x(l_r)N^r(-j)^{l-r}]. \tag{15}$$

Comparing the coefficients of N^r on both sides, we obtain

$$\mu_r(m)=\sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{k=0}^{m}\sum_{l=0}^{k} C_{ij}(m_k)i^{m-k}S^{-j}\mu_l(k)(l_r)(-j)^{l-r}, \tag{16}$$

for $r=0,\ 1,2,\ldots,m$.

Out of these $m+1$ equations for each m, only m are interesting. For when $r=m$, Eq. (16) simplifies to

$$\sum_{j=0}^{J}C_jS^{-j}=1, \tag{17}$$

which is trivially satisfied since S^{-1} is a root of Eq. (6). Thus we have again found that S^{-1} satisfies a certain equation. However, in the last section on normalization, we were able to see that S^{-1} is a root with the smallest absolute value. In that section, we were also able to determine the normalization constant k. The remaining m equations are linear equations in the m quantities $\mu_r(m)$ $(1\leq r\leq m)$ with coefficients depending upon the lower order moments $\mu_s(m)$ $(s<r)$. Equation (16) is therefore the recursion scheme we were seeking.

Remark: For $r=m-1$, Eq. (16) takes the form

$$\mu_m(m)/\mu_{m-1}(m-1)=\sum_{i=0}^{J}\sum_{j=0}^{J}iC_{ij}S^{-j}\Big/\sum_{i=0}^{I}\sum_{j=0}^{J}jC_{ij}S^{-j}, \tag{18}$$

which shows that $\mu_m(m)/\mu_{m-1}(m-1)$ is a constant independent of m. Since $\mu_0(0)=1$, we find

$$\mu_m(m)=(\mu(1))^m, \tag{19}$$

where, for simplicity, we have written

$$\mu_1(1)=\mu(1). \tag{20}$$

Thus we can immediately determine the coefficient of N^m in $\mu_N(m)$. This is the content of the central limit theorem proved in Ref. 4. We have, however, presented a technique by which coefficients of other powers of N in the various moments, could also be determined.

Remark: Since we will also calculate the pseudo-variance in our examples, we give the equation which determines $\mu_1(2)$ below.[6]

$$\mu_1(2)=\sum_{i=0}^{I}\sum_{j=0}^{J}C_{ij}S^{-j}[j^2\mu_2(2)-2ij\mu(1)+i^2]\Big/\sum_{i=0}^{I}\sum_{j=0}^{J}jC_{ij}S^{-j}. \tag{21}$$

EXAMPLES

Occupation of 2 × N array by parallel dumbbells

If $A_p(q,N)$ is the number of ways of arranging q parallel dumbbells in a $2\times N$ array, it is known that $A_p(q,N)$ satisfies the recursion relation[4]

$$A_p(q,N)=A_p(q,N-1)+A_p(q-1,N-1)+A_p(q-1,N-2)$$
$$+A_p(q-2,N-2)-A_p(q-3,N-3). \tag{22}$$

Thus, in this example, the coefficients C_{ij} and C_j are

$$C_{00}=0,\quad C_{10}=0,\quad C_{20}=0,\quad C_{30}=0,$$
$$C_{01}=1,\quad C_{11}=1,\quad C_{21}=0,\quad C_{31}=0,$$
$$C_{12}=1,\quad C_{22}=1,\quad C_{32}=0,$$
$$C_{33}=-1,$$
$$C_0=0,\quad C_1=2,\quad C_2=2,\quad C_3=-1.$$

Here $J=3$, and the initial values of Δ_N are

$$\Delta_0=1,\quad \Delta_1=1,\quad \Delta_2=4.$$

Thus

$$G(x)=(1-x)/(1-2x-2x^2+x^3) \tag{23}$$

or[7]

$$\Delta_N\simeq[(3+\sqrt{5})/10][(3+\sqrt{5})/2]^N, \tag{24}$$

i.e.,

$$k=(3+\sqrt{5})/10\ \text{ and }\ S=(3+\sqrt{5})/2. \tag{25}$$

Also

$$\mu(1)=(S^2+3S-3)/(2S^2+4S-3)=1-1/\sqrt{5} \tag{26}$$

and

$$\mu_2(2)=(1-1/\sqrt{5})^2. \tag{27}$$

Finally

$$\mu_1(2)=[\mu_2(2)(2S^2+8S-9)-2\mu(1)(S^2+6S-9)$$
$$+(S^2+5S-9)]/(2S^2+4S-3), \tag{28}$$
$$=2/5\sqrt{5}.$$

Occupation of 2 × N array by dumbbells (not necessarily parallel)

If $A(q,N)$ is the number of possible arrangements of q dumbbells, the recursion relation is[3]

$$A(q,N)=A(q,N-1)+2A(q-1,N-1)+A(q-1,N-2)$$
$$-A(q-3,\ N-3). \tag{29}$$

Thus the coefficients C_{ij}, C_j in this example are

$$C_{00}=0,\quad C_{10}=0,\quad C_{20}=0,\quad C_{30}=0,$$
$$C_{01}=1,\quad C_{11}=2,\quad C_{21}=0,\quad C_{31}=0,$$
$$C_{12}=1,\quad C_{22}=0,\quad C_{32}=0,$$
$$C_{23}=0,\quad C_{33}=-1,$$
$$C_0=0,\quad C_1=3,\quad C_2=1,\quad C_3=-1.$$

The initial values of Δ_N are

$$\Delta_0=1,\quad \Delta_1=2,\quad \Delta_2=7,$$

and $J=3$.

Thus

$$G(x)=(1-x)/(1-3x-x^2+x^3) \tag{30}$$

or

$$\Delta_N \simeq 0.655\ (3.214)^N, \tag{31}$$

i.e.,

$$S = 3.214, \quad k = 0.655. \tag{32}$$

Here

$$\mu(1) = (2S^2 + S - 3)/(2S^2 + 2S - 3) \simeq 0.606, \tag{33}$$

$$\mu_2(2) = 0.367, \tag{34}$$

and

$$\mu_1(2) = [\mu_2(2)(3S^2 + 4S - 9) - 2\mu(1)(2S^2 + 2S - 9) + (2S^2 + S - 9)]/(3S^2 + 2S - 3) \simeq 0.155. \tag{35}$$

CONCLUSION

Starting from a recursion relation, we have presented a technique for calculating higher moments and have illustrated our method by two examples. For the first example, alternate method based on explicit solution of the recursion relation is also available, though the calculation of moments higher than the second will become formidable. The second problem is such where the solution to the recursion relation is not known so far. We have been able to demonstrate the usefulness of our method by explicitly calculating the moment $\mu_1(2)$ which is related to the pseudovariance by $\sigma^2 = \mu_1(2)^N$. From the results that we have given, we can express the occupation numbers $A(q, N)$ as

$$A(q, N) \simeq [\Delta_N / \sqrt{2\pi\mu_1(2)N}] \times \exp\{-[q - \mu_1(2)N]^2 / 2\mu_1(2)N\}. \tag{36}$$

[1] J. K. Roberts, Proc. R. Soc. Lond. A **161**, 141 (1937).
[2] R. B. McQuistan, J. Math. Phys. **10**, 2205 (1969).
[3] R. B. McQuistan and S. J. Lichtman, J. Math. Phys. **11**, 3095 (1970).
[4] R. B. McQuistan, S. J. Lichtman, and L. P. Levine, J. Math. Phys. **13**, 242 (1972).

[5] This is where we have gone beyond the work in Ref. 4.

[6] The pseudovariance is defined as $\mu_N(2) - [\mu_N(1)]^2$. For large N, using Eq. (15), it is just $\mu_1(2)N$.

[7] In fact from Eq. (23), one immediately obtains $\Delta_N = [3 + \sqrt{5})/10][(3 + \sqrt{5})/2]^N + [(3 - \sqrt{5})/10]\,[(3 - \sqrt{5})/2]^N + (2/5)(-1)^N$, which for large N becomes $[(3 + \sqrt{5})/10][(3 + \sqrt{5})/2]^N$. We can rewrite this answer as $(1/5)[(1 + \sqrt{5})/2]^{2N+2}$, where $(1 + \sqrt{5})/2 = 1.618$ is the golden mean. In Appendix B of Ref. 4, this normalization is calculated using another argument. However, the authors have made minor errors. Using their notation, $\Sigma_q X(q, 1 \times 2N)$ is not negligible. Also $\Sigma_q A(q, 1 \times N) = (1/\sqrt{5})\{[(1 + \sqrt{5})/2]^{N+1} - [(1 - \sqrt{5})/2]^{N+1}\}$ and $\Sigma_q X(q, 1 \times 2N) = [\Sigma_q A(q, 1 \times (N-1))]^2$. These finally give the same answer as derived by us using the generating function method.

Spacelike representations of the inhomogeneous Lorentz group in a Lorentz basis

B. Radhakrishnan and N. Mukunda

Tata Institute of Fundamental Research, Bombay, India
(Received 8 August 1972)

The problem of introducing a Lorentz basis in a spacelike representation of the inhomogeneous Lorentz group in the spinless case, is solved by the methods of integral geometry. A new description of the representation associated with isotropic lines, which is peculiar to the spacelike case, is also given.

INTRODUCTION

The unitary irreducible representations (UIR's) of the inhomogeneous Lorentz group (IHLG), which were constructed by Wigner in 1939, may be analyzed from several different points of view.[1] One point of view may be described as follows. The IHLG has a semidirect product form, namely of the homogeneous Lorentz group (HLG), which is semisimple, by the four-parameter group of space–time translations, which is Abelian. As was originally done by Wigner, one can therefore set up the UIR's in such a way that the operators corresponding to the space–time translations are all diagonal. In physical terms, the infinitesimal generators of space–time translations are the operators of energy and momentum, and it is sought to diagonalize simultaneously these four operators. One then obtains the UIR's of the IHLG in a momentum basis, and can then analyze the various possible descriptions that can be given of the "spin" or helicity variable and the transformations that connect these descriptions. This development is very well known.[2] A totally different point of view from which to analyze the UIR's places the emphasis on the HLG rather than the space–time translation group; this leads to the UIR's of the IHLG in a Lorentz basis. The vectors of the Lorentz basis belong to definite UIR's of the HLG; and in this basis the unitary representation of the HLG contained in any given UIR of the IHLG is explicitly displayed as a direct sum (integral) of UIR's of the HLG.

The nontrivial UIR's of the IHLG form three distinct families, namely those with timelike, lightlike, and spacelike energy momenta. (By trivial we mean those UIR's of the IHLG in which the operators of energy and momentum vanish, and which are therefore just UIR's of the HLG). Analyses of the timelike and lightlike UIR's according to the second point of view described above have been given by several authors.[3] In this connection it is worth remarking that it is of interest to see not only how a Lorentz basis may be introduced, but also how the energy–momentum operators act in this basis. The mathematical methods employed in these treatments would not appear to be particularly unfamiliar to an experienced quantum mechanic. However the decomposition of a spacelike UIR of the IHLG into UIR's of the HLG does not appear to be as easily achieved with these same methods; at any rate some characteristic new complications peculiar to these UIR's show up.[4]

The simplest spacelike UIR of the IHLG is that corresponding to vanishing helicity; in every other spacelike UIR the helicity variable can assume an infinite number of distinct values. In this paper, we present a solution to the problem of decomposing the simplest spacelike UIR of the IHLG into UIR's of the HLG, making use of the methods of integral geometry developed by Gel'fand *et al.*[5] As a matter of fact, in a sense to be made precise later on, the work of Gel'fand *et al.* includes the solution to exactly half of the problem that we shall set ourselves, and our work constitutes an extension of their methods to a full solution of our problem.

Apart from its intrinsic interest, we were led to the present problem because it has unexpected connections with the Clebsch–Gordan problem for the UIR's of the three-dimensional Lorentz group. This connection will be explained elsewhere. In the meantime, Bassetto and Toller arrived at the same problem during their study of multiperipheral models, and they too have made use of the methods of integral geometry to obtain a solution.[6] However their solution and ours have, formally, quite different appearances. In particular, we have developed our solution in such a way that at every stage the specialization needed to recover the formulas of Gel'fand *et al.* is completely obvious. Such is not the case with the final formula of Bassetto and Toller. In addition, we have given a more detailed description of the representation of the HLG associated with isotropic lines.

The material of this paper is arranged as follows. Section 1 contains a statement of the problem we are interested in, and a resumé of the material from Ref. 5 that is relevant here. This resumé has been included because the methods used may be unfamiliar to a physicist, and because we will just be extending the work of Ref. 5. In Sec. 2 we give the derivation of the generalized inversion formula which is at the basis of the present method of reduction of group representations. Section 3 contains an analysis of the representation of the HLG on isotropic lines using essentially the approach of Ref. 5; while in Sec. 4 we relate this representation to another one obtained by the group expansion method from a representation of the Euclidean group $E(3)$. In the concluding Sec. 5 we have given a clear statement of our results, and suggested some interesting questions worth pursuing.

I. STATEMENT OF THE PROBLEM, RESUME OF KNOWN RESULTS

A general element of the IHLG may be denoted by (Λ, a^μ), and is made up of a proper orthochronous homogeneous Lorentz transformation Λ and a space–time translation a^μ.[7] The 4×4 real matrix $\|\Lambda^\mu{}_\nu\|$ corresponding to Λ obeys

$$\Lambda^{\mu}{}_{\nu}\Lambda_{\mu\sigma}=g_{\nu\sigma} \tag{1.1}$$

in addition to the conditions imposed by being a proper orthochronous transformation; while the components a^{μ} are four arbitrary real numbers. The action of the IHLG on the space–time coordinates x^{μ}, and the group composition law for the IHLG, are summarized by

$$(\Lambda, a^{\mu}): x^{\mu} \to \Lambda^{\mu}{}_{\nu}x^{\nu}+a^{\mu};$$
$$(\Lambda'^{\mu}{}_{\nu}, a'^{\mu})(\Lambda^{\mu}{}_{\nu}, a^{\mu})=(\Lambda'^{\mu}{}_{\sigma}\Lambda^{\sigma}{}_{\nu}, a'^{\mu}+\Lambda'^{\mu}{}_{\sigma}a^{\sigma}). \tag{1.2}$$

In a UIR of the IHLG, the generators of homogeneous Lorentz transformations, $J_{\mu\nu}$ and of translations, P_{μ} are both Hermitian and obey the following commutation rules:

$$-i[J_{\mu\nu}, J_{\lambda\rho}]=g_{\nu\lambda}J_{\mu\rho}-g_{\mu\lambda}J_{\nu\rho}+g_{\nu\rho}J_{\lambda\mu}-g_{\mu\rho}J_{\lambda\nu},$$
$$-i[J_{\mu\nu}, P_{\lambda}]=g_{\nu\lambda}P_{\mu}-g_{\mu\lambda}P_{\nu},$$
$$[P_{\mu}, P_{\lambda}]=0. \tag{1.3}$$

The set of six operators $J_{\mu\nu}$ can be divided into two subsets of three operators each:

$$J_k=\tfrac{1}{2}\epsilon_{klm}J_{lm}, \quad K_k=J_{k0}. \tag{1.4}$$

J_k generate space rotations, comprising the group $R(3)$, and the K_k give rise to pure Lorentz transformations.

The spacelike UIR of the IHLG corresponding to vanishing helicity can be described thus. In momentum space, that is the space of the eigenvalues p^{μ} of the operators P^{μ}, define the spacelike hypersurface Γ (single-sheeted hyperboloid) by

$$\Gamma: p^2\equiv(p^0)^2-|\mathbf{p}|^2=-1. \tag{1.5}$$

We shall write (dp) for the Lorentz-invariant volume element on Γ. We can now set up the Hilbert space $\mathcal{H}$ for the UIR: It consists of all "wavefunctions" $f(p)$ defined on Γ for which

$$\|f\|^2=\int_{\Gamma}(dp)|f(p)|^2\equiv\int\frac{d^3p}{|p^0|}\sum_{p^0=\pm\omega(\mathbf{p})}|f(p)|^2<\infty,$$
$$\omega(\mathbf{p})=[|\mathbf{p}|^2-1]^{1/2}\geq 0, \quad |\mathbf{p}|\geq 1. \tag{1.6}$$

(In three-dimensional momentum space, $\mathbf{p}$ is restricted to the exterior of the unit sphere). In $\mathcal{H}$, the required UIR acts on wavefunctions $f(p)$ in the following manner:

$$(\Lambda, a)\to T_{(\Lambda,a)}: (T_{(\Lambda,a)}f)(p)=\exp(ipa)f(\Lambda^{-1\mu}{}_{\nu}p^{\nu}). \tag{1.7}$$

Restricting ourselves to the subgroup of homogeneous Lorentz transformations, we have in $\mathcal{H}$ a unitary representation of the HLG provided by the operators $T_{\Lambda}\equiv T_{(\Lambda,0)}$. Our problem is to decompose this unitary representation of the HLG into UIR's.

From the manner in which the above representation of the HLG has been defined, we can immediately draw certain conclusions which restrict the UIR's of the HLG that can possibly occur. Recall that the (single-valued) UIR's of the HLG can be labelled in the form $\{j_0, \rho\}$; j_0 is a nonnegative integer or zero, and specifies the "smallest" representation of the subgroup $R(3)$ that is present (the least value of the angular momentum), while ρ is any real number.[8] Except for the fact that $\{0, \rho\}$ and $\{0, -\rho\}$ are the same UIR, different pairs $\{j_0, \rho\}$ correspond to inequivalent UIR's. Two Casimir invariants may be formed out of the generators J_k, K_k of the HLG; their forms, and the values they assume in a UIR, are

$$C_1=K_kK_k-J_kJ_k=1+\rho^2-j_0^2,$$
$$C_2=K_kJ_k=\rho j_0. \tag{1.8}$$

Now it is easily seen that the representation of the HLG defined in the space $\mathcal{H}$ has the generators

$$J_{\mu\nu}=i\left(p_{\mu}\frac{\partial}{\partial p^{\nu}}-p_{\nu}\frac{\partial}{\partial p^{\mu}}\right); \tag{1.9}$$

it is immaterial whether the restriction of p to Γ is made before or after the application of the J's to a function $f(p)$. With these expressions for $J_{\mu\nu}$ it follows that

$$K_kJ_k=0. \tag{1.10}$$

Therefore, on comparing with Eq. (1.8), it is clear that only UIR's of the HLG of the two types $\{0, \rho\}$, $\{j_0, 0\}$ can be present in $\mathcal{H}$.

Since on the single-sheeted hyperboloid Γ there is no Lorentz-invariant distinction between positive and negative values of p^0, we can define a reflection operation $\mathbb{R}$ in this way:

$$(\mathbb{R}f)(p^{\mu})=f(-p^{\mu}), \quad \mathbb{R}^2=1. \tag{1.11}$$

The unitary operator $\mathbb{R}$ commutes with the transformations T_{Λ} representing the HLG, but of course not with the space–time translations. It therefore can be used to split our problem into two independent parts: the eigenspaces $\mathcal{H}_+$ and $\mathcal{H}_-$ of $\mathbb{R}$, consisting of functions $f(p)$, respectively, even and odd under $p^{\mu}\to -p^{\mu}$, are individually invariant under the HLG, so the two corresponding representations of the HLG can be separately decomposed into UIR's. The problem of decomposing the representation appearing in $\mathcal{H}_+$ into irreducibles is precisely what has been solved in Ref. 5. We shall now give a brief resumé of this treatment.

Let ξ^{μ} stand for a variable lightlike vector on the positive light cone, i.e., fulfilling $\xi^2=0$, $\xi^0>0$. With any given function $f_+(p)$ belonging to $\mathcal{H}_+$ we associate a function $h_+(\xi)$ on the positive light cone according to[9]

$$h_+(\xi)=2\int(dp)\delta(\xi\cdot p-1)f_+(p). \tag{1.12}$$

The values that a function of ξ^{μ} obtained in this manner may assume for different points on the positive light cone are not independent of one another; $h_+(\xi)$ obeys the linear constraints

$$\int(d\xi)h_+(\xi)[\delta(p\cdot\xi-t)+\delta(p\cdot\xi+t)]$$
$$=\int(d\xi)h_+(\xi)[\delta(p\cdot\xi-t^{-1})+\delta(p\cdot\xi+t^{-1})], \tag{1.13}$$

where p is any point on Γ, t any positive real number, and $(d\xi)=d^3\xi/\xi^0$ the Lorentz-invariant volume element on the positive light cone.[10] One may now ask whether it is possible to recover $f_+(p)$ given a function $h_+(\xi)$ which is known to obey Eq. (1.13), i.e., whether Eq. (1.12) can be inverted. The answer is that this is not possible; in addition to $h_+(\xi)$ we need to know the values of the integrals of f_+ along all possible "isotropic lines" in Γ before we can obtain an inversion formula for $f_+(p)$. An isotropic line l is a one-parameter set of points in

Γ of the form $b^\mu + t\xi^\mu$ with the parameter t in the range $-\infty < t < \infty$, where b^μ is any point on Γ and ξ^μ is a lightlike vector such that $b \cdot \xi = 0$. (Given any point on Γ, one can find isotropic lines passing through it). By suitable choice of t it is evident that we can arrange to have $b^0 = 0$, $\xi^0 = 1$; we may then say that the isotropic line l is given in standard form. Thus, isotropic lines correspond one-to-one to pairs of unit vectors $(\hat{\xi}, \hat{b})$ in three-dimensional space, fulfilling $\hat{\xi} \cdot \hat{b} = 0$. In fact the line determined by the pair $(\hat{\xi}, \hat{b})$ comprises the points $(t, \hat{b} + t\hat{\xi})$ of Γ. With $f_+(p)$ we now associate a function on the isotropic lines according to

$$\varphi_+(\hat{\xi}, \hat{b}) = \int_{-\infty}^{\infty} dt\, f_+(t, \hat{b} + t\hat{\xi}). \tag{1.14}$$

Once again, the values that a function φ_+ obtained in this manner may assume for different "values" of its argument are not independent of one another. We have the set of linear constraint equations

$$\int d\Omega(\hat{\xi}) \varphi_+[\hat{\xi}, R(\alpha, \hat{\xi})\hat{b}]\delta(p^0 - \mathbf{p} \cdot \hat{\xi}) = \int d\Omega(\hat{\xi}) \varphi_+[\hat{\xi}, R(\pi - \alpha, \hat{\xi})\hat{b}]\delta(p^0 - \mathbf{p} \cdot \hat{\xi}). \tag{1.15}$$

Here, $d\Omega(\hat{\xi})$ is the rotation-invariant integration measure over the direction of the unit vector $\hat{\xi}$; p^μ is any point on Γ; α is any angle in the range $0 \leq \alpha \leq \pi$; $R(\alpha, \hat{\xi})$ denotes a right-handed rotation of amount α about the axis $\hat{\xi}$; and $\hat{b}$ is determined by $\hat{\xi}$ and p^μ according to $\hat{b} = \mathbf{p} - p^0\hat{\xi}$. Notice that for a given point p on Γ, and for each direction $\hat{\xi}$ that is then permitted by the delta function in Eq. (1.15), $\hat{b}$ is such that the isotropic line determined by the pair $(\hat{\xi}, \hat{b})$ passes through p; indeed we have $p^\mu = (p^0, \hat{b} + p^0\hat{\xi})$. Now it is possible to invert the pair of Eqs. (1.12, 14) and express $f_+(p)$ in terms of $h_+(\xi)$, $\varphi_+(\hat{\xi}, \hat{b})$:

$$f_+(p) = -(4\pi)^{-2} \int (d\xi) h_+(\xi) \times [\delta''(p \cdot \xi - 1) + \delta''(p \cdot \xi + 1)] - (2\pi)^{-2} \int d\Omega(\hat{\xi}) \delta(p^0 - \mathbf{p} \cdot \hat{\xi}) \times \int_0^\pi d\alpha \cot^2\alpha \varphi_+(\hat{\xi}, R(\alpha, \hat{\xi})\hat{b}). \tag{1.16}$$

δ'' denotes the second derivative of the delta function with respect to its argument, and once again the unit vector $\hat{b}$ inside the integral is chosen according to $\hat{b} = \mathbf{p} - p^0\hat{\xi}$.[11]

The association of the pair $\{h_+(\xi), \varphi_+(\hat{\xi}, \hat{b})\}$ with each wavefunction $f_+(p)$ in $\mathcal{H}_+$, and the inversion formula (1.16) expressing the latter in terms of the former, complete the first step in decomposing the representation of the HLG appearing in $\mathcal{H}_+$ into UIR's. The action of Lorentz transformations on the functions $f_+(p)$ can be translated into suitable actions on the functions $h_+(\xi)$ and $\varphi_+(\hat{\xi}, \hat{b})$. In the case of $h_+(\xi)$, the result is obvious. As for $\varphi_+(\hat{\xi}, \hat{b})$: evidently a homogeneous Lorentz transformation Λ^{-1} applied to the set of points on Γ constituting an isotropic line $l = (\hat{\xi}, \hat{b})$ results in the set of points $\Lambda^{-1}(t, \hat{b} + t\hat{\xi})$, $-\infty < t < \infty$, which forms another isotropic line, $l' = \Lambda^{-1}l$ say. If $\Lambda \in R(3)$, we get a description of l' in standard form by allowing Λ^{-1} to act directly on $\hat{\xi}$ and $\hat{b}$; otherwise a redefinition of t is needed to get such a description. Let us write $\Lambda^{-1}(\hat{\xi}, \hat{b})$ for the standard description of $l' = \Lambda^{-1}l$. Then the change in the function $f_+(p)$ caused by the action of T_Λ,

$$f_+(p^\mu) \to f_+(\Lambda^{-1\mu}{}_\nu p^\nu) \tag{1.17}$$

is accompanied by the following changes in $h_+(\xi)$ and $\varphi_+(\hat{\xi}, \hat{b})$:

$$h_+(\xi^\mu) \to h_+(\Lambda^{-1\mu}{}_\nu \xi^\nu), \tag{1.18a}$$

$$\varphi_+(\hat{\xi}, \hat{b}) \to [\Lambda_0{}^0 + \Lambda_j{}^0 \xi_j]^{-1} \varphi_+[\Lambda^{-1}(\hat{\xi}, \hat{b})]. \tag{1.18b}$$

Thus the originally given unitary representation of the HLG in the space $\mathcal{H}_+$ has been expressed as the direct sum of two other representations: one acts according to (1.18a) on a class of functions that are defined on the positive light cone and obey the constraints (1.13); the other acts according to (1.18b) on a class of functions that are defined on the isotropic lines and obey the constraints (1.15). These two representations must now be decomposed into irreducibles, and then the original scalar product in $\mathcal{H}_+$ must be suitably reexpressed.

The decomposition of the representation associated with $h_+(\xi)$ is achieved as follows: with $h_+(\xi)$ we associate a function of ξ and a real variable ρ by

$$F_+(\xi; \rho) = \int_0^\infty dt\, t^{-i\rho/2} h_+(t\xi), \quad -\infty < \rho < \infty. \tag{1.19}$$

Then for fixed ρ, $F_+(\xi; \rho)$ is homogeneous in ξ of degree $-1 + i\rho/2$:

$$F_+(a\xi; \rho) = a^{-1+i\rho/2} F_+(\xi; \rho), \quad a > 0. \tag{1.20}$$

So we may think of $F_+(\xi; \rho)$ as essentially a function of $\hat{\xi}$, i.e., as a function on the unit sphere in three-space. It then turns out that when $h_+(\xi)$ transforms according to Eq. (1.18a) under the HLG, $F_+(\xi; \rho)$ transforms according to the UIR $\{0, \rho\}$ of this group. The equations that relate $F_+(\xi; \rho)$ directly to $f_+(p)$, express $h_+(\xi)$ in terms of $F_+(\xi; \rho)$, and give the constraints (1.13) in terms of $F_+(\xi; \rho)$ are these:

$$F_+(\xi; \rho) = 2 \int (dp) \theta(\xi \cdot p) [\xi \cdot p]^{-1+i\rho/2} f_+(p) = \int (dp) |\xi \cdot p|^{-1+i\rho/2} f_+(p); \tag{1.21a}$$

$$h_+(\xi) = (4\pi)^{-1} \int_{-\infty}^{\infty} d\rho\, F_+(\xi; \rho); \tag{1.21b}$$

$$\int (d\xi) F_+(\xi; \rho)[\delta(p \cdot \xi - 1) + \delta(p \cdot \xi + 1)] = \int (d\xi) F_+(\xi; -\rho)[\delta(p \cdot \xi - 1) + \delta(p \cdot \xi + 1)] \quad \text{for all } \rho, \ p \text{ on } \Gamma. \tag{1.21c}$$

The last equation above is compatible with the irreducible transformation properties of the $F_+(\xi; \rho)$ stated above because $\{0, \rho\}$ and $\{0, -\rho\}$ are equivalent UIR's of the HLG; it essentially says that knowledge of $F_+(\xi; \rho)$ for all positive ρ suffices to determine this function for all ρ. Making use of these equations, the first term on the right-hand side of the inversion formula (1.16) can be written in terms of $F_+(\xi; \rho)$ for $\rho \geq 0$, and it has the form

$$\tfrac{1}{2}(4\pi)^{-3} \int_0^\infty \rho^2 d\rho \int (d\xi) F_+(\xi; \rho)[\delta(p \cdot \xi - 1) + \delta(p \cdot \xi + 1)]. \tag{1.22}$$

The decomposition of the representation associated with $\varphi_+(\hat{\xi}, \hat{b})$ is achieved as follows: With $\varphi_+(\hat{\xi}, \hat{b})$ we associate a function of $(\hat{\xi}, \hat{b})$ and an even integer $2n$ by

$$F_+(\hat{\xi}, \hat{b}; 2n) = \int_0^\pi d\alpha \exp(-2in\alpha) \varphi_+[\hat{\xi}, R(\alpha, \hat{\xi})\hat{b}], \quad n = 0, \pm 1, \pm 2, \cdots. \tag{1.23}$$

Then for each n, the behavior of $F_+(\hat{\xi}, \hat{b}; 2n)$ under rotation of $\hat{b}$ is given by

$$F_+(\hat{\xi}, R(\beta, \hat{\xi})\hat{b}; 2n) = \exp(2in\beta) F_+(\hat{\xi}, \hat{b}; 2n). \tag{1.24}$$

This is the analogue to Eq. (1.20); so if for each $\hat{\xi}$ we make some choice of $\hat{b}$ according to some convention,

we can again think of $F_+(\hat{\xi}, \hat{b}; 2n)$ as just dependent on $\hat{\xi}$, i.e., as a function on the unit sphere in three-space. It then turns out that when $\varphi_+(\hat{\xi}, \hat{b})$ is transformed according to Eq. (1.18b) under the HLG, $F_+(\hat{\xi}, \hat{b}; 2n)$ transforms according to the UIR $\{2|n|, 0\}$ of this group. The replacements for Eqs. (1.21) are, respectively,

$$F_+(\hat{\xi}, \hat{b}; 2n) = \int (dp) \exp[-2in\alpha(\hat{\xi}, \hat{b}; p)]\delta(p^0 - \mathbf{p}\cdot\hat{\xi}) f_+(p), \tag{1.25a}$$

$$\varphi_+(\hat{\xi}, \hat{b}) = \frac{1}{\pi}\sum_{n=-\infty}^{\infty} F_+(\hat{\xi}, \hat{b}; 2n), \tag{1.25b}$$

$$\int d\Omega(\hat{\xi}) F_+(\hat{\xi}, \hat{b}; 2n)\delta(p^0 - \mathbf{p}\cdot\hat{\xi}) = \int d\Omega(\hat{\xi}) F_+(\hat{\xi}, \hat{b}; -2n)\delta(p^0 - \mathbf{p}\cdot\hat{\xi}) \quad \text{for all } n,\ p \text{ on } \Gamma. \tag{1.25c}$$

In the first of the three equations above, the dependence of α on $\hat{\xi}$, $\hat{b}$, and p is to be determined from

$$\mathbf{p} - p_0\hat{\xi} = R(\alpha(\hat{\xi}, \hat{b}; p), \hat{\xi})\hat{b}. \tag{1.26}$$

Equation (1.25c) is clearly compatible with the stated transformation properties of $F_+(\hat{\xi}, \hat{b}; 2n)$, and it essentially says that this series of functions for $n \geq 0$ already determines those for $n < 0$. Making use of these equations, the second term on the right-hand side of the inversion formula (1.16) can be written in terms of $F_+(\hat{\xi}, \hat{b}; 2n)$ for $n > 0$ (the term with $n = 0$ drops out!); adding this expression to what appears in (1.22), we get

$$f_+(p) = \tfrac{1}{2}(4\pi)^{-3}\int_0^\infty \rho^2 d\rho \int (d\xi) F_+(\xi; \rho)[\delta(p\cdot\xi - 1) + \delta(p\cdot\xi + 1)] + 4\pi^{-2}\sum_{n=1}^{\infty} n \int d\Omega(\hat{\xi}) F_+(\hat{\xi}, \hat{b}; 2n) \times \delta(p^0 - \mathbf{p}\cdot\hat{\xi}),$$
$$\hat{b} = \mathbf{p} - p^0\hat{\xi}. \tag{1.27}$$

Finally, it is shown in Ref. 5 that the original expression for the norm of a wave function in $\mathcal{H}_+$ takes this form:

$$\|f_+\|^2 \equiv \int (dp)|f_+(p)|^2 = \tfrac{1}{2}(4\pi)^{-3}\int_0^\infty \rho^2 d\rho \int d\Omega(\hat{\xi})|F_+(\hat{\xi}; \rho)|^2 + 4\pi^{-2}\sum_{n=1}^{\infty} n \int d\Omega(\hat{\xi})|F_+(\hat{\xi}, \ldots; 2n)|^2. \tag{1.28}$$

From all of this we conclude that in the subspace $\mathcal{H}_+$ of $\mathcal{H}$ the set of UIR's $\{0, \rho\}$ of the HLG appear in the form of a direct integral from $\rho = 0$ to $\rho = \infty$, the multiplicity of appearance of each of these UIR's being one; and in addition there is a discrete direct sum of the UIR's $\{2n, 0\}$ for $n = 1, 2, \cdots$, each appearing once.

In this account of the treatment given in Ref. 5 of the representation of the HLG appearing on $\mathcal{H}_+$, we have retained all those important equations for which we shall find parallels when analyzing the representation on $\mathcal{H}_-$. As mentioned in the Introduction, our general treatment will be such that the equations valid for $\mathcal{H}_+$ alone will at all stages be completely obvious.

II. DERIVATION OF GENERALIZED INVERSION FORMULA

We shall now obtain an inversion formula that generalizes Eq. (1.16) to all functions $f(p) \in \mathcal{H}$. Our method will be an extension of that employed in Ref. 5. To begin with, it is necessary to restrict ourselves to functions $f(p)$ that are infinitely often differentiable and have compact support on Γ; however since these niceties are very well explained in Ref. 5, we will not repeatedly refer to them.

Given then a function $f(p)$ on Γ, let us set up two independent functions on the positive light cone by

$$h_a(\xi) = \int (dp)\delta(\eta_a\xi\cdot p - 1), \quad a = 1, 2 \qquad \eta_1 = +1,\ \eta_2 = -1. \tag{2.1}$$

We will obtain a formula that gives $f(p)$ in terms of $h_1(\xi)$, $h_2(\xi)$ and the integrals of f along isotropic lines. To this end, let us form four expressions $K_{ab}(p'; \mu; f)$, where p' is a point on Γ and μ a complex variable, by

$$K_{ab}(p'; \mu; f) = \int (d\xi) h_a(\xi)(\eta_b p'\cdot\xi - 1)_+^\mu, \quad a \text{ and } b = 1, 2. \tag{2.2}$$

[Here, for real t we define $t_\pm^\mu$ to be equal to $\theta(\pm t)|t|^\mu$.] The generalized function appearing in the integrand can be rewritten using the formula[12]

$$t_+^\mu = (1/2i\sin\mu\pi)[\exp(i\mu\pi)(t - i0)^\mu - \exp(-i\mu\pi)(t + i0)^\mu] \tag{2.3}$$

and then K_{ab} takes the form

$$K_{ab}(p'; \mu; f) = \int (dp) G_{ab}(p, p'; \mu) f(p),$$
$$G_{ab}(p; p'; \mu) \equiv (1/2i\sin\mu\pi)\int (d\xi)\delta(\eta_a\xi\cdot p - 1) \times [\exp(i\mu\pi)(\eta_b p'\cdot\xi - 1 - i0)^\mu - \exp(-i\mu\pi)(\eta_b p'\cdot\xi - 1 + i0)^\mu]. \tag{2.4}$$

By making use of the delta function present in the integrand in G_{ab}, and writing $\mp i0$ as $\mp ic\cdot\xi$, where c is an infinitesimal positive timelike vector, the functions G_{ab} can all be expressed in terms of the standard function $J(p, z; \mu)$ introduced in Ref. 5, p. 310:

$$G_{ab}(p, p'; \mu) \equiv (1/2i\sin\mu\pi)[\exp(i\mu\pi)J(\eta_a p, \eta_b p' - \eta_a p - ic; \mu) - \exp(-i\mu\pi)J(\eta_a p, \eta_b p' - \eta_a p + ic; \mu)],$$
$$J(p, z; \mu) = \int (d\xi)\delta(p\cdot\xi - 1)(z\cdot\xi)^\mu. \tag{2.5}$$

[Note that $J(p, z; \mu)$ is defined for p on Γ, z arbitrary; we need to take Imz = positive or negative timelike and infinitesimal.]

Now $J(p, z; \mu)$ is an explicitly Lorentz-invariant function, being unchanged if both p and z are subjected to one and the same proper orthochronous homogeneous Lorentz transformation. Consequently, $G_{ab}(p, p'; \mu)$ is a Lorentz-invariant function of p and p'. Let us work out the values of these functions in the particular situation where p has the form $p = (0, 0, 0, 1)$, while p' remains an arbitrary point on Γ. One then needs the values of $J(p, z; \mu)$ and $J(-p, z; \mu)$ for various choices of Rez but with Imz always either a positive or a negative timelike infinitesimal vector. If we write $z = q \pm i0$, respectively, in these two cases, for $J(p, q \pm i0; \mu)$ Ref. 5, p. 329 gives the following values:

$$J(p, q + i0; \mu) = -[2\pi/(\mu + 1)]\exp(i\mu\pi/2)(-P \mp i0)^{-1/2} \times \{(-P \mp i0)^{1/2} - iq^3\}^{\mu+1} \text{ acc. as } q^0 \gtrless 0;$$
$$J(p, q - i0; \mu) = -[2\pi/(\mu + 1)]\exp(-i\mu\pi/2)(-P \pm i0)^{-1/2}$$

$$\times\{(-P\pm i0)^{1/2}+iq^3\}^{\mu+1} \text{ acc. as } q^0\gtrless 0;$$
$$P=(q^0)^2-(q^1)^2-(q^2)^2,$$
$$t^\mu=|t|^\mu\exp(i\mu\arg t),\quad -\pi<\arg t<\pi. \tag{2.6}$$

The values of $J(-p,q\pm i0;\mu)$ follow from the observation that $J(p,z;\mu)$ is generally unchanged if p and z are subjected to a common Lorentz transformation; since with $p=(0,0,0,1)$, p is taken into $-p$ by a rotation of 180° about the second axis, all that happens is that q^3 changes sign in the above expressions. In using these expressions, one can also avail of the formulas

$$(-P\pm i0)^{1/2}=P_-^{1/2}\pm iP_+^{1/2},\quad (-P\pm i0)^{-1/2}=P_-^{-1/2}\mp iP_+^{-1/2}. \tag{2.7}$$

One then finds for the set of four functions $G_{ab}(p,p';\mu)$ if

$$P(p')\equiv(p'^0)^2-(p'^1)^2-(p'^2)^2>0,$$

then

$$\begin{aligned}
G_{11}(p,p';\mu)&=-[2\pi/(\mu+1)]P^{\mu/2}\theta(p'^0)\theta(p'^3-1)\\
&\quad\times[1+(p'^3-1)/P^{1/2}]^{\mu+1};\\
G_{12}(p,p';\mu)&=-[2\pi/(\mu+1)]P^{\mu/2}\theta(-p'^0)\theta(-p'^3-1)\\
&\quad\times[1-(p'^3+1)/P^{1/2}]^{\mu+1};\\
G_{21}(p,p';\mu)&=-[2\pi/(\mu+1)]P^{\mu/2}\theta(p'^0)\theta(-p'^3-1)\\
&\quad\times[1-(p'^3-1)/P^{1/2}]^{\mu+1};\\
G_{22}(p,p';\mu)&=-[2\pi/(\mu+1)]P^{\mu/2}\theta(-p'^0)\theta(p'^3-1)\\
&\quad\times[1+(p'^3-1)/P^{1/2}]^{\mu+1}.
\end{aligned} \tag{2.8}$$

In all these expressions, $P=P(p')$. On the other hand, if $P(p')<0$, we have

$$\begin{aligned}
G_{11}(p,p';\mu)&=G_{22}(p,p';\mu)\\
&=[\pi i(-P)^{\mu/2}/(\mu+1)\sin\mu\pi]\\
&\quad\times\{\exp(i\mu\pi/2)[1-i(1-p'^3)/(-P)^{1/2}]^{\mu+1}\\
&\quad-\exp(-i\mu\pi/2)[1+i(1-p'^3)/(-P)^{1/2}]^{\mu+1}\},\\
G_{12}(p,p';\mu)&=G_{21}(p,p';\mu)\\
&=[\pi i(-P)^{\mu/2}/(\mu+1)\sin\mu\pi]\\
&\quad\times\{\exp(i\mu\pi/2)[1-i(1+p'^3)/(-P)^{1/2}]^{\mu+1}\\
&\quad-\exp(-i\mu\pi/2)[1+i(1+p'^3)/(-P)^{1/2}]^{\mu+1}\}.
\end{aligned} \tag{2.9}$$

By studying the results in Eq. (2.8), we can try to form a linear combination of $G_{ab}(p,p';\mu)$ which does discriminate between the regions $p'^3>1$ and $p'^3<-1$, but does not discriminate between positive and negative p'^0. Let us choose the combination $G=G_{11}+G_{22}$; with $p=(0,0,0,1)$ we get for G

$$\begin{aligned}
G(p,p';\mu)&=-\frac{2\pi}{\mu+1}P^{\mu/2}\left(1+\frac{(p'^3-1)}{P^{1/2}}\right)^{\mu+1} &&\text{if } P>0,\ p'^3>1\\
&=0 &&\text{if } P>0,\ p'^3<-1\\
&=\frac{2\pi i(-P)^{\mu/2}}{(\mu+1)\sin\mu\pi}\left[\exp(i\mu\pi/2)\left(1-i\frac{1-p'^3}{(-P)^{1/2}}\right)^{\mu+1}\right.\\
&\quad\left.-\exp(-i\mu\pi/2)\left(1+i\frac{1-p'^3}{(-P)^{1/2}}\right)^{\mu+1}\right] &&\text{if } P<0.
\end{aligned} \tag{2.10}$$

Another combination we could have considered is $G'=G_{12}+G_{21}$ but this is easily derivable from G [at least when $p=(0,0,0,1)$]. In any case, it will suffice to work with G alone hereafter. Going over now to an arbitrary point p on Γ, not necessarily $p=(0,0,0,1)$ Eq. (2.10) can be written in invariant form thus:

$$\begin{aligned}
G(p,p';\mu)&=-\frac{2\pi}{\mu+1}[(p\cdot p')^2-1]^{\mu/2}\\
&\quad\times\left(1-\frac{p\cdot p'+1}{\sqrt{(p\cdot p')^2-1}}\right)^{\mu+1} &&\text{if } p\cdot p'<-1,\\
&=0 &&\text{if } p\cdot p'>1,\\
&=\frac{2\pi i}{(\mu+1)\sin\mu\pi}[1-(p\cdot p')^2]^{\mu/2}\\
&\quad\times\left[\exp(i\mu\pi/2)\left(1-i\frac{1+p\cdot p'}{\sqrt{1-(p\cdot p')^2}}\right)^{\mu+1}\right.\\
&\quad\left.-\exp(-i\mu\pi/2)\left(1+i\frac{1+p\cdot p'}{\sqrt{1-(p\cdot p')^2}}\right)^{\mu+1}\right]\\
&&&\text{if } (p\cdot p')^2<1.
\end{aligned} \tag{2.11}$$

We can also go back to Eqs. (2.2), (2.4), and consider that linear combination that involves just the function G; this then gives

$$\begin{aligned}
&\int(dp)G(p,p';\mu)f(p)\\
&=(1/2i\sin\mu\pi)\\
&\quad\times\{\int(d\xi)[\exp(i\mu\pi)(p'\cdot\xi-1-i0)^\mu\\
&\quad-\exp(-i\mu\pi)(p'\cdot\xi-1+i0)^\mu]h_1(\xi)\\
&\quad+\int(d\xi)[\exp(i\mu\pi)(-p'\cdot\xi-1-i0)^\mu\\
&\quad-\exp(-i\mu\pi)(-p'\cdot\xi-1+i0)^\mu]h_2(\xi)\}.
\end{aligned} \tag{2.12}$$

As in Ref. 5, it is by examining the residues of both sides at the simple pole that exists at $\mu=-3$ that we shall derive our inversion formula. Knowing now the function $G(p,p';\mu)$ for all p, p', and since Eq. (2.12) is "manifestly Lorentz invariant," let us first take p' in this equation to be the point $(0,0,0,1)$ on Γ, and develop the left-hand side:

$$\begin{aligned}
&\int(dp)G(p,p';\mu)f(p)=I_1+I_2,\\
&I_1=\int_{p^3>1}(dp)G(p,p';\mu)f(p)\\
&\quad=\int_{p^3>1}\frac{dp^0\,dp^1\,dp^2}{p^3}\left(-\frac{2\pi}{\mu+1}\right)\\
&\quad\times[(p^0)^2-(p^1)^2-(p^2)^2]_+^{\mu/2}\left(1+\frac{p^3-1}{\sqrt{(p^3)^2-1}}\right)^{\mu+1}f(p),\\
&I_2=\int_{(p^3)^2<1}(dp)G(p,p';\mu)f(p).
\end{aligned} \tag{2.13}$$

The residue of the term I_1 at the simple pole at $\mu=-3$ may be computed using the result[13]

$$\operatorname*{res}_{\mu=-3}[(p^0)^2-(p^1)^2-(p^2)^2]_+^{\mu/2}=-4\pi\delta(p^0)\delta(p^1)\delta(p^2). \tag{2.14}$$

On the other hand, in the case of the term I_2, the pole

at $\mu=-3$ arises from the explicit factor $(\sin\mu\pi)^{-1}$ present in $G(p',p;\mu)$ for $(p\cdot p')^2<1$; the factor $((p^1)^2+(p^2)^2-(p^0)^2)_+^{\mu/2}$ has no pole at this value of μ.[14] so from (2.13) we obtain

$$\operatorname*{res}_{\mu=-3}\int(dp)G(p,p';\mu)f(p)=-4\pi^2 f(0,0,0,1)+\operatorname*{res}_{\mu=-3}\int_{(p^3)^2<1}(dp)G(p,p';\mu)f(p). \tag{2.15}$$

It is straightforward algebra to check that in the case $(p\cdot p')^2<1$, for quite general p', we have

$$\operatorname*{res}_{\mu=-3}G(p,p';\mu)\big|_{(p\cdot p')^2<1}=-p\cdot p'(p\cdot p'-1)[1-(p\cdot p')^2]^{-3/2}; \tag{2.16}$$

we must remember later that the exponent $-3/2$ here is the limit of $\mu/2$ as $\mu\to-3$. Thus the calculation of the residue of the left-hand side in Eq. (2.12) for $\mu=-3$ is complete. The calculation for the right-hand side can be done using the general result[15]

$$\operatorname*{res}_{\mu=-3}(1/2i\sin\mu\pi)[\exp(i\mu\pi)(t-i0)^\mu-\exp(-i\mu\pi)(t+i0)^\mu]=\tfrac{1}{2}\delta''(t). \tag{2.17}$$

Putting together Eqs. (2.12)–(2.16), we get in the first place an expression for the value of f at the special point $p'=(0,0,0,1)$ on Γ, but because of the manifest Lorentz invariance we can immediately generalize this to an arbitrary point p' on Γ. Interchanging also the variables p and p', we get

$$f(p)=-(8\pi^2)^{-1}\int(d\xi)[h_1(\xi)\delta''(p\cdot\xi-1)+h_2(\xi)\delta''(p\cdot\xi+1)]+(2\pi)^{-2}\int_{(p'\cdot p)^2<1}(dp')p\cdot p'(1-p\cdot p')[1-(p\cdot p')^2]^{-3/2}f(p'). \tag{2.18}$$

Let us at this stage change over the pair of functions $h_{1,2}(\xi)$ to the linear combinations

$$h_\pm(\xi)=h_1(\xi)\pm h_2(\xi). \tag{2.19}$$

The point is that if $f(p)$ were split also into an even part and an odd part under $p\to-p$, corresponding to its components in $\mathcal{H}_+$ and $\mathcal{H}_-$ respectively, then h_+ depends on f_+ alone and h_- on f_- alone:

$$f(p)=f_+(p)+f_-(p),\quad f_\pm(-p)=\pm f_\pm(p),$$
$$h_\pm(\xi)=2\int\delta(\xi\cdot p-1)f_\pm(p)(dp). \tag{2.20}$$

This relation between $h_+(\xi)$ and $f_+(p)$ coincides exactly with that in Eq. (1.12). We then put (2.18) into the form

$$f(p)=-(4\pi)^{-2}\int(d\xi)\{h_+(\xi)[\delta''(p\cdot\xi-1)+\delta''(p\cdot\xi+1)]+h_-(\xi)[\delta''(p\cdot\xi-1)-\delta''(p\cdot\xi+1)]\}+(2\pi)^{-2}\int_{(p'\cdot p)^2<1}(dp')p\cdot p'(1-p\cdot p')\times[1-(p\cdot p')^2]^{-3/2}f(p'). \tag{2.21}$$

This is the desired generalization of the inversion theorem obtained in Ref. 5 and quoted in Eq. (1.16). The part involving the functions $h_\pm(\xi)$ is already in a neat form separating the contributions to $f_+(p)$ and $f_-(p)$ explicitly. The second term involving integration over the region $(p\cdot p')^2<1$ in p' will be dealt with in detail in the next section; there we will express it in terms of the integrals of $f(p)$ over isotropic lines, and show how this too splits explicitly into contributions to $f_+(p)$ and $f_-(p)$. [Actually the even and odd parts in p in the second term of (2.21) can already be separated.]

The originally given unitary representation of the HLG in $\mathcal{H}$ is now expressed as the sum of three representations: one acting on the functions $h_+(\xi)$, another on the functions $h_-(\xi)$, and a third on the functions of isotropic lines. This last can evidently be again expressed as the sum of two other representations, acting on functions of isotropic lines arising from functions of p respectively even or odd in p. These details are reserved for the following section. In the rest of this section, we consider the representations acting on $h_\pm(\xi)$.

First we establish the fact that the functions $h_-(\xi)$ obey a family of constraints analogous to Eq. (1.13) for $h_+(\xi)$. $h_-(\xi)$ arises from $f(p)$ thus:

$$h_-(\xi)=\int(dp)[\delta(p\cdot\xi-1)-\delta(p\cdot\xi+1)]f(p). \tag{2.22}$$

The contribution made by $h_-(\xi)$ to $f(p)$ in the inversion formula (2.21) suggests that we define a function $H_-(p;t)$ for t a real positive number in this way:

$$H_-(p;t)=\int(d\xi)h_-(\xi)[\delta(p\cdot\xi-t)-\delta(p\cdot\xi+t)]. \tag{2.23}$$

Then apart from the factor $-(4\pi)^{-2}$, the term involving $h_-(\xi)$ in Eq. (2.21) is just the second derivative of $H_-(p;t)$ with respect to t, at $t=1\cdot H_-(p;t)$ is explicitly odd in p. Unlike the analogous function defined in Ref. 5 and related to $h_+(\xi)$, the definition given above for H_- involves no divergent quantities, and no regularization of the type used in Ref. 5 is needed to make it well defined; this happens because the *difference* of two delta functions occurs in (2.23). $H_-(p;t)$ can be related to $f(p')$ via a kernel $K_-(p,p';t)$:

$$H_-(p;t)=\int(dp')K_-(p,p';t)f(p'),$$
$$K_-(p,p';t)=\int(d\xi)[\delta(p\cdot\xi-t)-\delta(p\cdot\xi+t)][\delta(p'\cdot\xi-1)-\delta(p'\cdot\xi+1)]. \tag{2.24}$$

We can now express K_- in terms of the invariant function J used earlier. The steps are

$$\begin{aligned}K_-(p,p';t)&=\int(d\xi)[\delta(p\cdot\xi-t)-\delta(p\cdot\xi+t)][\delta(p'\cdot\xi-1)-\delta(p'\cdot\xi+1)],\\&=t\int(d\xi)[\delta(p\cdot\xi-1)-\delta(p\cdot\xi+1)][\delta(tp'\cdot\xi-1)-\delta(tp'\cdot\xi+1)],\\&=\frac{t}{\pi}\operatorname{Im}\int(d\xi)\delta(p\cdot\xi-1)\{[(tp'-p-i0)\cdot\xi]^{-1}-[(tp'+p-i0)\cdot\xi]^{-1}\}\\&\quad-\frac{t}{\pi}\operatorname{Im}\int(d\xi)\delta(p\cdot\xi+1)\{[(tp'+p-i0)\cdot\xi]^{-1}-[(tp'-p-i0)\cdot\xi]^{-1}\},\\&=\frac{t}{\pi}\operatorname{Im}\lim_{\mu\to-1}[J(p,tp'-p-i0;\mu)-J(p,tp'+p-i0;\mu)-J(-p,tp'+p-i0;\mu)+J(-p,tp'-p-i0;\mu)].\end{aligned} \tag{2.25}$$

Since K_- is a manifestly Lorentz invariant function of p and p' let us evaluate it in the case $p=(0,0,0,1)$, p' arbitrary. Using Eqs. (2.6), (2.7), one finds some delicate cancellations taking place as the limit $\mu\to-1$

is taken, and one gets the following values for various configurations of p':

$p=(0,0,0,1)$:

$$\begin{aligned} K_-(p,p';t) &= 2\pi P^{-1/2}[\theta(1-t(P^{1/2}+p'^3)) \\ &\quad +\theta(1-t^{-1}(P^{1/2}+p'^3))-1] \quad \text{if } p'^3>1; \\ &= 2\pi P^{-1/2}[\theta(1+t(P^{1/2}+p'^3)) \\ &\quad +\theta(1+t^{-1}(P^{1/2}+p'^3))-1] \quad \text{if } p'^3<-1; \\ &= 2(-P)^{-1/2}\ln[(t+t^{-1}+2p'^3)/(t+t^{-1}-2p'^3)] \\ &\quad \text{if } |p'^3|<1. \end{aligned} \tag{2.26}$$

Here $P\equiv P(p')=(p'^0)^2-(p'^1)^2-(p'^2)^2$. Generalizing to arbitrary p, we have the symmetry relation

$$K_-(p,p';t)=K_-(p,p';t^{-1}); \tag{2.27}$$

written in terms of $h_-(\xi)$, this relation is

$$\begin{aligned} &\int (d\xi)h_-(\xi)[\delta(p\cdot\xi-t)-\delta(p\cdot\xi+t)] \\ &\quad =\int (d\xi)h_-(\xi)[\delta(p\cdot\xi-t^{-1})-\delta(p\cdot\xi+t^{-1})]. \end{aligned} \tag{2.28}$$

This replaces Eq. (1.13).[16]

Under the transformations of the HLG, $h_-(\xi)$ transforms in the same way as $h_+(\xi)$, i.e., via Eq. (1.18a):

$$h_-(\xi^\mu)\to h_-(\Lambda^{-1}{}^\mu{}_\nu\,\xi^\nu). \tag{2.29}$$

The decomposition of this representation into irreducibles is therefore achieved in exactly the same way as in the case of $h_+(\xi)$. If for each real ρ we define $F_-(\xi;\rho)$ by

$$F_-(\xi;\rho)=\int_0^\infty dt\, t^{-i\rho/2}h_-(t\xi), \quad -\infty<\rho<\infty, \tag{2.30}$$

then we will have a homogeneity property for F_- and can recover h_- from F_-:

$$F_-(a\xi;\rho)=a^{-1+i\rho/2}F_-(\xi;\rho), \quad a>0, \tag{2.31a}$$

$$h_-(\xi)=(4\pi)^{-1}\int_{-\infty}^{\infty}d\rho F_-(\xi;\rho). \tag{2.31b}$$

Following the arguments of Ref. 5, we conclude that when h_- transforms via Eq. (2.29), $F_-(\xi;\rho)$ transforms by the UIR $\{0,\rho\}$ of the HLG. The replacements for Eqs. (1.21a,c) are

$$F_-(\xi;\rho)=\int (dp)\,|\xi\cdot p|^{-1+i\rho/2}[\theta(\xi\cdot p)-\theta(-\xi\cdot p)]f(p), \tag{2.32a}$$

$$\begin{aligned} &\int (d\xi)F_-(\xi;\rho)[\delta(p\cdot\xi-1)-\delta(p\cdot\xi+1)] \\ &\quad =\int (d\xi)F_-(\xi;-\rho)[\delta(p\cdot\xi-1)-\delta(p\cdot\xi+1)], \\ &\qquad \text{for all } \rho,\ p \text{ on } \Gamma. \end{aligned} \tag{2.32b}$$

The characteristic differences as compared to the previous case must be noted. Again the constraints on $F_-(\xi;\rho)$ are compatible with their known irreducible transformation laws. Finally, we can express the first term in the general inversion formula (2.21) in terms of $F_+(\xi;\rho)$ for $\rho>0$, to get:

$$\begin{aligned} f(p) &= \tfrac{1}{2}(4\pi)^{-3}\int_0^\infty \rho^2\,d\rho\int (d\xi)\{F_+(\xi;\rho)[\delta(p\cdot\xi-1) \\ &\quad +\delta(p\cdot\xi+1)]+F_-(\xi;\rho)[\delta(p\cdot\xi-1)-\delta(p\cdot\xi+1)]\} \\ &\quad +(2\pi)^{-2}\int_{(p\cdot p')^2<1}(dp')p\cdot p'\,(1-p\cdot p') \\ &\quad \times(1-(p\cdot p')^2)^{-3/2}f(p'). \end{aligned} \tag{2.33}$$

In $\mathcal{K}_-$, as in $\mathcal{K}_+$, we find that the UIR's $\{0,\rho\}$ for $\rho>0$ appear, once each, in the form of a direct integral. We turn now to the representation associated with isotropic lines.

III. ANALYSIS OF REPRESENTATION ON ISOTROPIC LINES

In dealing with the second integral in Eq. (2.33), we face essentially this problem: given a function $g(p)$ on Γ, express

$$W(p;g)\equiv\int_{(p'\cdot p)^2<1}(dp')g(p'), \quad p\in\Gamma, \tag{3.1}$$

in terms of integrals of g along isotropic lines. [A possible dependence of g on the p occurring in $W(p;g)$ is suppressed.] We do not assume that g is an even function on Γ, and so we shall go through the analysis in spite of some overlap with Ref. 5.

First let us consider the case

$$p=(\sinh\xi,0,0,\cosh\xi); \tag{3.2}$$

writing S and C for $\sinh\xi$ and $\cosh\xi$, we deal with

$$W(\mathrm{S},0,0,\mathrm{C})=\int_{(p'^0\mathrm{S}-p'^3\mathrm{C})^2<1}(dp')g(p'). \tag{3.3}$$

To get a neater integration domain, change variables from p' on Γ to q on Γ by

$$p'^0=q^0\mathrm{C}+q^3\mathrm{S}, \quad p'_\perp=q_\perp, \quad p'^3=q^3\mathrm{C}+q^0\mathrm{S}. \tag{3.4}$$

$\perp$ refers to the projection in the 1—2 plane. Then,

$$\begin{aligned} W(\mathrm{S},0,0,\mathrm{C};g) &= \int_{(q^3)^2<1}(dq)g(q^0\mathrm{C}+q^3\mathrm{S},q_\perp,q^3\mathrm{C}+q^0\mathrm{S}), \\ &= \int_{-1}^{+1}dq^3\int_{q_1^2\geq 1-q_3^2}\frac{dq^1dq^2}{\omega(\mathbf{q})}\sum_{q^0=\pm\omega(\mathbf{q})}g(q^0\mathrm{C} \\ &\quad +q^3\mathrm{S},q_\perp,q^3\mathrm{C}+q^0\mathrm{S}), \\ \omega(\mathbf{q}) &= +(|\mathbf{q}|^2-1)^{1/2}. \end{aligned} \tag{3.5}$$

For fixed q^3, we have to integrate in the q_1-q_2 plane over the region outside the circle of radius $(1-q_3^2)^{1/2}$ centered at the origin, and this must be converted into integration along isotropic lines. So let us parametrize q^μ for fixed q^3 in this way:

$$\begin{aligned} &q^0\mathrm{C}+q^3\mathrm{S}=t, \quad q^3\mathrm{C}+q^0\mathrm{S}=b^3+t\xi^3, \\ &q_\perp=b_\perp+t\xi_\perp, \end{aligned}$$

i.e.,

$$\begin{aligned} &q^0=(\mathrm{C}-\xi^3\mathrm{S})t-b^3\mathrm{S}, \quad q^3=(\xi^3\mathrm{C}-\mathrm{S})t+b^3\mathrm{C}, \\ &q_\perp=b_\perp+t\xi_\perp. \end{aligned} \tag{3.6}$$

If for all t we want q to lie on Γ and q_3^2 to be bounded by unity, we find these conditions:

$$\xi_\perp^2+\xi_3^2=b_\perp^2+b_3^2=1, \quad \xi\cdot\mathrm{b}=0, \quad \xi^3=\mathrm{S}/\mathrm{C}, \quad b^3=q^3/\mathrm{C}. \tag{3.7}$$

In particular, as expected, $\hat{\xi}$ and $\hat{b}$ are unit vectors orthogonal to one another. Further, from these equations we find that as $\xi_\perp$ and $b_\perp$ vary with $q_\perp$ over a plane of constant q^3, they enclose an angle γ that does not vary with $q_\perp$ but depends only on q^3 (and ξ):

$$\begin{aligned} &\cos\gamma=(\xi_\perp\cdot b_\perp)/|\xi_\perp||b_\perp|=-\mathrm{S}q^3/[\mathrm{C}^2-(q^3)^2]^{1/2}, \\ &\qquad 0\leq\gamma\leq\pi. \end{aligned} \tag{3.8}$$

Therefore, if $\xi_\perp$ makes an angle ϕ with the q^1 axis in the

1–2 plane, $b_\perp$ must make an angle $(\phi-\gamma)$ or $(\phi+\gamma)$. We have then two alternative ways in which new integration variables t, ϕ may be introduced in place of q_1-q_2, concisely expressed by:

$$\text{(A)}\quad q^1+iq^2=\exp(i\phi)(t+[C^2-(q^3)^2]^{1/2}\exp(-i\gamma))/C,$$

$$\text{(B)}\quad q^1+iq^2=\exp(i\phi)(t+[C^2-(q^3)^2]^{1/2}\exp(i\gamma))/C. \tag{3.9}$$

We shall agree to use (A), but the existence of the alternative (B) will lead to a symmetry relation. The vectors $\hat{\xi}$ and $\hat{b}$ corresponding to (A) have components

$$\hat{\xi}(\phi)=C^{-1}(\cos\phi,\sin\phi,S),$$

$$\hat{b}(q^3,\phi)=C^{-1}([C^2-(q^3)^2]^{1/2}\cos(\phi-\gamma),[C^2-(q^3)^2]^{1/2}\times\sin(\phi-\gamma),q^3). \tag{3.10}$$

We also easily check that

$$\int_{q_1^2\geq 1-q_3^2}\frac{dq^1dq^2}{\omega(\mathbf{q})}\sum_{q^0=\pm\omega(\mathbf{q})}\cdots=\frac{1}{C}\int_{-\infty}^{\infty}dt\int_0^{2\pi}d\phi\cdots \tag{3.11}$$

so Eq. (3.5) takes the form

$$W(S,0,0,C;g)=\int_{-1}^{1}dq^3\frac{1}{C}\int_{-\infty}^{\infty}dt\int_0^{2\pi}d\phi\, g(t,\hat{b}(q^3,\phi)+t\hat{\xi}(\phi)). \tag{3.12}$$

Let us now deal with the q^3 integration. Since $\hat{\xi}$ has no dependence on q^3, $\hat{b}(q^3,\phi)$ is orthogonal to $\hat{\xi}(\phi)$ for all q^3. So $\hat{b}(q^3,\phi)$ arises from $\hat{b}(1,\phi)$ via a right-handed rotation of some amount α about $\hat{\xi}(\phi)$. $\cos\alpha$ is easily determined:

$$\cos\alpha=\hat{b}(1,\phi)\cdot\hat{b}(q^3,\phi)=q^3. \tag{3.13}$$

It can be checked that the range $0\leq\alpha\leq\pi$ as q^3 goes from -1 to $+1$ corresponds to picking alternative (A) in (3.9). So (3.12) is transformed into

$$W(S,0,0,C;g)=\int_0^{\pi}\sin\alpha\,d\alpha\frac{1}{C}\int_{-\infty}^{\infty}dt\int_0^{2\pi}d\phi\times g(t,t\hat{\xi}(\phi)+R(\alpha,\hat{\xi}(\phi))\hat{b}(1,\phi)),$$

$$\hat{b}(1,\phi)=\mathbf{p}-p^0\hat{\xi}(\phi). \tag{3.14}$$

As a last step, let us express the ϕ integration in terms of an integration over all directions of $\hat{\xi}$ and a suitable delta function:

$$\int_0^{2\pi}d\phi\cdots=C\int d\Omega(\hat{\xi})\delta(p^0-\mathbf{p}\cdot\hat{\xi})\cdots. \tag{3.15}$$

With this, we may free ourselves from the special configuration (3.2) for p, and write the general formula

$$W(p;g)=\int_0^{\pi}\sin\alpha\,d\alpha\int d\Omega(\hat{\xi})\int_{-\infty}^{\infty}dt\,\delta(p^0-\mathbf{p}\cdot\hat{\xi})\times g(t,t\hat{\xi}+R(\alpha,\hat{\xi})\hat{b}), \tag{3.16}$$

$$\hat{b}=\mathbf{p}-p^0\hat{\xi}.$$

The fact that the q_1-q_2 integration could be converted into a $t-\phi$ integration using either choice (A) or choice (B) in (3.9) can be seen to lead to the symmetry relation

$$\int d\Omega(\hat{\xi})\int_{-\infty}^{\infty}dt\,\delta(p^0-\mathbf{p}\cdot\hat{\xi})g(t,t\hat{\xi}+R(\alpha,\hat{\xi})\hat{b})=\int d\Omega(\hat{\xi})\int_{-\infty}^{\infty}dt\,\delta(p^0-\mathbf{p}\cdot\hat{\xi})g(t,t\hat{\xi}+R(2\pi-\alpha,\hat{\xi})\hat{b})$$

$$\text{for all } \alpha,\ p \text{ on } \Gamma,\ \hat{b}=\mathbf{p}-p^0\hat{\xi}. \tag{3.17}$$

(The symmetry relation used in Ref. 5 is obtained by combining this more general one with the evenness of g.)

Let us now use formula (3.16) to handle the second integral in (2.33). Define, as an extension of (1.14), the function $\varphi(\hat{\xi},\hat{b})$ on the isotropic lines by

$$\varphi(\hat{\xi},\hat{b})=\int_{-\infty}^{\infty}dt\,f(t,\hat{b}+t\hat{\xi}). \tag{3.18}$$

Then (2.33) becomes

$$f(p)=\cdots-(2\pi)^{-2}\int_0^{\pi}d\alpha\sin\alpha\frac{\cos\alpha(\cos\alpha+1)}{\sin^3\alpha}\times\int d\Omega(\hat{\xi})\delta(p^0-\mathbf{p}\cdot\hat{\xi})\varphi(\hat{\xi},R(\alpha,\hat{\xi})\hat{b}),$$

$$\hat{b}=\mathbf{p}-p^0\hat{\xi}. \tag{3.19}$$

The dots stand for the terms involving $h_\pm(\xi)$, which we have omitted writing again. Therewith we have completed the proof of our generalization of the inversion formula of Ref. 5 from even functions on Γ to all functions on Γ. The constraints on φ following from Eq. (3.17) are

$$\int d\Omega(\hat{\xi})\delta(p^0-\mathbf{p}\cdot\hat{\xi})\varphi(\hat{\xi},R(\alpha,\hat{\xi})\hat{b})=\int d\Omega(\hat{\xi})\delta(p^0-\mathbf{p}\cdot\hat{\xi})\varphi(\hat{\xi},R(2\pi-\alpha,\hat{\xi})\hat{b}),$$

$$\hat{b}=\mathbf{p}-p^0\hat{\xi},\quad \text{for all } \alpha,\ p \text{ on } \Gamma. \tag{3.20}$$

This generalizes Eq. (1.15) to all $f(p)$. Making use of it, one can also obtain

$$f(-p)=\cdots-(2\pi)^{-2}\int_0^{\pi}d\alpha\sin\alpha\frac{\cos\alpha(\cos\alpha-1)}{\sin^3\alpha}\times\int d\Omega(\hat{\xi})\delta(p^0-\mathbf{p}\cdot\hat{\xi})\varphi(\hat{\xi},R(\alpha,\hat{\xi})\hat{b}),$$

$$\hat{b}=\mathbf{p}-p^0\hat{\xi}. \tag{3.21}$$

We have used here the same value for $\hat{b}$ as is needed in Eq. (3.19), not its negative, so that on comparing Eqs. (3.19) and (3.21), it is easy to see what terms in the general inversion formula contribute to $f_+(p)$, and what terms to $f_-(p)$.

For the further analysis and decomposition of the representation of the HLG on functions φ on the isotropic lines, we shall use an approach slightly different from, but of course equivalent in essence to, that of Ref. 5. The notation we will use is suggested by the structure of the regular representation of $R(3)$; this is convenient because $\varphi(\hat{\xi},\hat{b})$ can be thought of as a function on $R(3)$, obeying suitable constraints. Let $e_1^{(0)}$, $e_2^{(0)}$, $e_3^{(0)}$ be three unit vectors in the directions of the three axes in $\mathbf{p}$-space. Given an acceptable pair of vectors $\hat{\xi}$ and $\hat{b}$, the triplet $\hat{b}$, $\hat{\xi}\times\hat{b}$, $\hat{\xi}$ always defines some right-handed coordinate system, obtained from the fixed $e_j^{(0)}$ by a unique element of $R(3)$. In this way, we have a well-defined correspondence between pairs $(\hat{\xi},\hat{b})$ and elements in $R(3)$; we agree to call $\hat{\xi}$ and $\hat{b}$ the unit vectors e_3, e_1 respectively, of a variable triad, this triad being related to the fixed one $e_j^{(0)}$ by a variable element in $R(3)$. A function $\varphi(\hat{\xi},\hat{b})$ becomes a function on $R(3)$. Referring to elements of $R(3)$ by the matrices $A(\boldsymbol{\alpha})$, the triad $e_j(\boldsymbol{\alpha})$ is defined in Appendix A by

$$e_j(\boldsymbol{\alpha})=A_{jk}(\boldsymbol{\alpha})e_k^{(0)}. \tag{3.22}$$

At the same time let us use for $\varphi(\hat{\xi},\hat{b})$ the bra–ket notation:

$$\varphi(\hat{\xi},\hat{b})=\langle\hat{\xi},\hat{b}\,|\varphi\rangle=\langle e_j|\varphi\rangle. \tag{3.23}$$

We can now express the constraints (3.20) on the allowed vectors $|\varphi\rangle$ using the operators of the left and right regular representations of $R(3)$. Since the α rotation in Eq. (3.20) is about the variable e_3 axis, this dependence is naturally expressed via the left regular representation operators:

$$\varphi(\hat{\xi}, R(\alpha, \hat{\xi})\hat{b}) = \langle \hat{\xi}, R(\alpha, \hat{\xi})\hat{b} | \varphi \rangle = \langle A_{jk}(0, 0, -\alpha) e_k | \varphi \rangle$$
$$= \langle e_k | \exp(-i\alpha J_3^L) | \varphi \rangle. \quad (3.24)$$

Therefore, the constraints on $|\varphi\rangle$ can be written as

$$\int d\Omega(\hat{\xi}) \delta(p^0 - \mathbf{p}\cdot\hat{\xi}) \langle \hat{\xi}, \hat{b} | [\exp(-i\alpha J_3^L) - \exp(i\alpha J_3^L)] | \varphi \rangle = 0,$$
$$\hat{b} = \mathbf{p} - p^0\hat{\xi}, \quad \text{for all } \alpha,\ p \text{ on } \Gamma. \quad (3.25)$$

Now these constraints are all preserved under the action of the HLG; this suggests that the operators representing the elements of the HLG will all commute with J_3^L. That this is the case for the $R(3)$ subgroup of the HLG can be easily established; indeed the corresponding operators give just the right regular representation of $R(3)$! Under a general transformation $\Lambda \in$ HLG, the representation on the functions $\varphi(\hat{\xi}, \hat{b})$ is given again by Eqs. (1.17), (1.18):

$$f(p^\mu) \to f'(p^\mu) = f(\Lambda^{-1\mu}{}_\nu p^\nu),$$
$$\varphi(\hat{\xi}, \hat{b}) \to \varphi'(\hat{\xi}, \hat{b}) = [\Lambda_0{}^0 + \Lambda_j{}^0 \xi_j]^{-1} \varphi(\Lambda^{-1}(\hat{\xi}, \hat{b})). \quad (3.26)$$

If Λ is the element $A(\boldsymbol{\beta})$ in $R(3)$, then, resolving $\hat{\xi}$ and $\hat{b}$ into their components along $e_k^{(0)}$, we have

$$\varphi'(\hat{\xi}_k, \hat{b}_k) = \int_{-\infty}^{\infty} dt f'(t, \hat{b}_k + t\hat{\xi}_k) = \varphi(A_{jk}(\boldsymbol{\beta})\hat{\xi}_j, A_{jk}(\boldsymbol{\beta})\hat{b}_j).$$

Reverting to vector notation, this reads [see Eq. (A.7)]

$$\langle e_k | \varphi' \rangle = \langle \hat{\xi}, \hat{b} | \varphi' \rangle = \langle R(\boldsymbol{\beta})\hat{\xi}, R(\boldsymbol{\beta})\hat{b} | \varphi \rangle$$
$$= \langle R(\boldsymbol{\beta}) e_k | \varphi \rangle = \langle e_k | \exp(i\boldsymbol{\beta}\cdot\mathbf{J}^R) | \varphi \rangle,$$
$$\text{i.e.,}\quad |\varphi'\rangle = \exp(i\boldsymbol{\beta}\cdot\mathbf{J}^R)|\varphi\rangle. \quad (3.27)$$

Later on we shall explicitly compute the generators for pure Lorentz transformations in this representation of the HLG on functions $\varphi(\hat{\xi}, \hat{b})$ and show that they too commute with J_3^L.

To decompose this representation of the HLG into irreducibles, we must therefore break up each $|\varphi\rangle$ into its projections in the various eigenspaces of J_3^L. We shall define, for an allowed $|\varphi\rangle$,

$$|\varphi; n\rangle = \tfrac{1}{2}\int_0^{2\pi} d\alpha \exp[-i\alpha(J_3^L + n)] |\varphi\rangle, \quad n = 0, \pm 1, \pm 2, \cdots \quad (3.28)$$

$|\varphi\rangle$ may be recovered via

$$|\varphi\rangle = \pi^{-1} \sum_{n=-\infty}^{\infty} |\varphi; n\rangle \quad (3.29)$$

and $|\varphi; n\rangle$ is an eigenvector of J_3^L with eigenvalue $-n$. In conformity with Eq. (1.23), we shall write the wavefunctions corresponding to $|\varphi; n\rangle$ and their properties in this way[17]

$$F(\hat{\xi}, \hat{b}; n) = \langle \hat{\xi}, \hat{b} | \varphi; n \rangle = \tfrac{1}{2}\int_0^{2\pi} d\alpha \exp(-i\alpha n)\varphi(\hat{\xi}, R(\alpha, \hat{\xi})\hat{b}),$$
$$F(\hat{\xi}, R(\beta, \hat{\xi})\hat{b}; n) = \exp(in\beta) F(\hat{\xi}, \hat{b}; n),$$
$$\varphi(\hat{\xi}, \hat{b}) = \pi^{-1} \sum_{n=-\infty}^{\infty} F(\hat{\xi}, \hat{b}; n). \quad (3.30)$$

(The second equation above corresponds to $|\varphi; n\rangle$ being an eigenvector of J_3^L.) Analogous to Eq. (1.25a), we now have

$$F(\hat{\xi}, \hat{b}; n) = \tfrac{1}{2}\int (dp) \exp[-in\alpha(\hat{\xi}, \hat{b}; p)] \delta(p^0 - \mathbf{p}\cdot\hat{\xi}) f(p),$$
$$\mathbf{p} - p^0\hat{\xi} = R(\alpha(\hat{\xi}, \hat{b}; p), \hat{\xi})\hat{b}. \quad (3.31)$$

The constraint relations on φ are translated into this form:

$$\int d\Omega(\hat{\xi}) \delta(p^0 - \mathbf{p}\cdot\hat{\xi}) \langle \hat{\xi}, \hat{b} | \{ |\varphi; n\rangle - |\varphi; -n\rangle \} = 0,$$
$$\hat{b} = \mathbf{p} - p^0\hat{\xi}, \quad \text{for all } n,\ p \text{ on } \Gamma. \quad (3.32)$$

As will be clear from Appendix B, these relations state that $|\varphi; n\rangle$ determines $|\varphi; -n\rangle$ for each n, so only the vectors $|\varphi, n\rangle$ for $n \geq 0$, or corresponding wavefunctions $F(\hat{\xi}, \hat{b}; n)$ for $n \geq 0$ are truly independent.

We saw above that in the representation of the HLG on the space of functions $\varphi(\hat{\xi}, \hat{b})$, the $R(3)$ subgroup of the HLG is realized via the right regular representation of $R(3)$ (but, of course, the space is reduced by a factor 2 due to the constraints). From the structure of the regular representation of $R(3)$ we know that in the eigenspace of J_3^L with eigenvalue $-n$, the representation of $R(3)$ generated by the operators J_k^R is the direct sum of infinitely many finite dimensional UIR's of $R(3)$, with each of the "j values" $|n|$, $|n|+1$, $|n|+2, \cdots, \infty$ occurring once. We know that each eigenspace of J_3^L is invariant under the HLG, given the law of transformation (3.26); on the other hand, we saw in Sec. 1 that the full space $\mathcal{H}$ can contain only two types of UIR's of the HLG, namely $\{j_0, 0\}$ and $\{0, \rho\}$. From all of this it follows that when $\varphi(\hat{\xi}, \hat{b})$ is transformed via Eq. (3.26) under the HLG, the functions $F(\hat{\xi}, \hat{b}; n)$ for fixed n transform via the UIR $\{|n|, 0\}$ of this group. Taking into account the constraints on φ, we have the result that the representation of the HLG on the functions $\varphi(\hat{\xi}, \hat{b})$ is the direct sum of the UIR's $\{n, 0\}$ for $n = 1, 2, 3, \cdots$, each appearing once. (As pointed out in Sec. 1, the term $n = 0$ is absent!). Clearly, the even n values belong to $\mathcal{H}_+$, the odd ones to $\mathcal{H}_-$.

We can express now the isotropic line contribution in the general inversion formula in terms of the various n values. The relevant expression is in Eq. (3.19) and can be developed as follows:

$$(2\pi)^{-2} \int_0^{\pi} d\alpha \sin\alpha \frac{\cos\alpha(\cos\alpha + 1)}{\sin^3\alpha} \int d\Omega(\hat{\xi}) \delta(p^0 - \mathbf{p}\cdot\hat{\xi})$$
$$\times \varphi(\hat{\xi}, R(\alpha, \hat{\xi})\hat{b}) = (2\pi)^{-2} \int_0^{\pi} d\alpha \frac{\cos\alpha(\cos\alpha + 1)}{\sin^2\alpha}$$
$$\times \int d\Omega(\hat{\xi}) \delta(p^0 - \mathbf{p}\cdot\hat{\xi}) \cdot \pi^{-1} \sum_{n=-\infty}^{\infty}$$
$$\times \exp(in\alpha) F(\hat{\xi}, \hat{b}; n)$$
$$= 4(2\pi)^{-3} \sum_{n=1}^{\infty} C_n \int d\Omega(\hat{\xi}) \delta(p^0 - \mathbf{p}\cdot\hat{\xi})$$
$$\times F(\hat{\xi}, \hat{b}; n),$$
$$C_n = (1 + (-1)^n) \int_0^{\pi/2} d\alpha \cos n\alpha \frac{\cos^2\alpha}{\sin^2\alpha} + (1 - (-1)^n)$$
$$\times \int_0^{\pi/2} d\alpha \cos n\alpha \frac{\cos\alpha}{\sin^2\alpha}. \quad (3.33)$$

These integrals diverge, and as remarked immediately after Eq. (2.16), they must be understood as the limits,

as $\lambda \to -2$, of similar integrals in which $\sin^{-2}\alpha$ is replaced by $\sin^{\lambda}\alpha$. Using standard formulas, we get uniformly for all $n \geq 1$

$$C_n = -4\pi n. \tag{3.34}$$

The final expansion formula for a function $f(p)$ in terms of the associated functions $F_+(\xi;\rho)$, $F_-(\xi,\rho)$, and $F(\hat{\xi},\hat{b};n)$, where the first two transform via the UIR $\{0,o\}$ of the HLG and the third via the UIR $\{n,0\}$, is obtained by putting together Eqs. (2.33), (3.19), (3.33), and (3.34):

$$\begin{aligned} f(p) &= \tfrac{1}{2}(4\pi)^{-3}\int_0^\infty \rho^2 d\rho \int (d\xi)\{F_+(\xi;\rho)[\delta(p\cdot\xi-1) \\ &\quad + \delta(p\cdot\xi+1)] + F_-(\xi;\rho)[\delta(p\cdot\xi-1)-\delta(p\cdot\xi+1)]\} \\ &\quad + 2\pi^{-2}\sum_{n=1}^{\infty} n\int d\Omega(\hat{\xi})\delta(p^0-\mathbf{p}\cdot\hat{\xi})F(\hat{\xi},\hat{b};n), \\ \hat{b} &= \mathbf{p} - p^0\hat{\xi}. \end{aligned} \tag{3.35}$$

The last item in this section is to compute the forms of the generators of pure Lorentz transformations for the representation of the HLG acting on $\varphi(\hat{\xi},\hat{b})$, and check that they do commute with J_3^L. It suffices to get the form of the operator K_3. Let us now introduce Euler angles ψ, θ, ϕ to describe the orientation of the triad defined by $\hat{\xi}$ and $\hat{b}$ relative to $e_j^{(0)}$; i.e., for the components of $\hat{\xi}$ and $\hat{b}$ we have (see Appendix A)

$$\begin{aligned} \hat{\xi} &= (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \\ \hat{b} &= (\cos\psi\cos\theta\cos\phi - \sin\psi\sin\phi, \cos\psi\cos\theta\sin\phi \\ &\quad + \sin\psi\cos\phi, -\cos\psi\sin\theta). \end{aligned} \tag{3.36}$$

The expressions for the two commuting sets of $R(3)$ generators J_k^L, J_k^R are given in Appendix A. If we now consider an infinitesimal pure Lorentz transformation in the 3 direction, whose effect on a function $f(p)$ is the following:

$$f(p) \to f'(p) = [(1-i\lambda K_3)f](p) = f(p^0-\lambda p^3, p_\perp, p^3-\lambda p^0), \quad |\lambda| \ll 1, \tag{3.37}$$

then one finds after simple algebra that the change in $\varphi(\hat{\xi},\hat{b})$ is

$$\begin{aligned} &\varphi(\hat{\xi},\hat{b}) \to [(1-i\lambda K_3)\varphi](\hat{\xi},\hat{b}) \approx (1+\lambda\xi^3)\varphi(\hat{\xi}+\delta\hat{\xi},\hat{b}+\delta\hat{b}), \\ &\delta\xi_\perp = \lambda\xi^3\xi_\perp, \quad \delta\xi^3 = \lambda(\xi_3^2-1), \\ &\delta b_\perp = \lambda b^3\xi_\perp, \quad \delta b^3 = \lambda b^3\xi^3. \end{aligned} \tag{3.38}$$

By using Eq. (3.36), the associated changes in ψ, θ, and ϕ turn out to be

$$\delta\psi = \delta\phi = 0, \quad \delta\theta = \lambda\sin\theta. \tag{3.39}$$

So we may write (3.38) in the form

$$\varphi(\psi,\theta,\phi) \to [(1-i\lambda K_3)\phi](\psi,\theta,\phi) = (1+\lambda\cos\theta)\varphi(\psi,\theta+\lambda\sin\theta,\phi)$$

and this gives immediately the form of K_3 as far as the action on $\varphi(\hat{\xi},\hat{b})$ is concerned:

$$K_3 = i\left(\cos\theta + \sin\theta\frac{\partial}{\partial\theta}\right). \tag{3.40}$$

We see explicitly that K_3 commutes with both J_3^L and J_3^R; the latter of course follows from the commutation rule, in Eq. (1.3). Since the components $K_{1,2}$ can be gotten by commuting K_3 with $J_{2,1}^R$, we have shown that J_3^L commutes with all the HLG generators in the representation on functions on isotropic lines.

IV. ALTERNATIVE DESCRIPTION OF REPRESENTATION ON ISOTROPIC LINES

When one compares the timelike, lightlike, and spacelike UIR's of the IHLG (all with vanishing helicity) with one another, from the point of view of introducing a Lorentz basis the really distinguishing feature of the spacelike case is the occurrence of the representation associated with functions on isotropic lines. Of course the reasons for this are very beautifully explained in Ref. 5; nevertheless it is of some value to give an alternative way of looking at this representation, especially as it helps us understand the structure of its generators.

It is a well known fact that if one is given a unitary representation of the three-dimensional Euclidean group $E(3)$, with generators J_k of rotations and P_k of translations obeying the commutation rules

$$\begin{aligned} &[J_k,J_l] = i\epsilon_{klm}J_m, \quad [J_k,P_l] = i\epsilon_{klm}P_m, \\ &[P_k,P_l] = 0, \end{aligned} \tag{4.1}$$

and if one then defines three operators K_k in terms of J_k and P_k by

$$K_k = (\rho+i)P_k - \epsilon_{klm}P_lJ_m, \tag{4.2}$$

then J_k and K_k obey the commutation rules corresponding to the HLG and so generate a unitary representation of this group.[18] In this construction, ρ is any real number. The UIR's of $E(3)$ can be labelled (j_0,p): j_0 is an integer [positive or zero and denotes the "lowest" UIR of $R(3)$ present], while p is any real number.[19] The Casimir invariants of $E(3)$ and their values in (j_0,p) are

$$P_kP_k = p^2, \quad P_kJ_k = pj_0. \tag{4.3}$$

If we carry out the construction (4.2) using the generators of the UIR $\{j_0,p\}$ of $E(3)$, we end up with the UIR $\{j_0,\rho p\}$ of the HLG. As with the HLG, here also two different pairs (j_0,p), (j_0',p') denote inequivalent UIR's of $E(3)$ except for the equivalence of $(0,p)$ and $(0,-p)$.

There is a very natural way to construct a unitary reducible representation of $E(3)$ in the space of the regular representation of $R(3)$, and it can be briefly explained as follows.[20] The space consists of all square integrable functions on $R(3)$; using Euler angles ψ, θ, ϕ to parameterize $R(3)$, we deal with functions $f(\psi,\theta,\phi)$ obeying

$$\int_0^{2\pi} d\psi \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta\, d\theta\, |f(\psi,\theta,\phi)|^2 < \infty. \tag{4.4}$$

A continuous basis of ket vectors $|\psi\theta\phi\rangle$ may be introduced with appropriate orthonormality relations. The matrix of $R(3)$ corresponding to given angles ψ, θ, ϕ is given in Appendix A, and is written $A(\psi,\theta,\phi)$. If we take for the $R(3)$ subgroup of $E(3)$ the right regular representation generated by the operators J_k^R given in Appendix A, and if for P_k we define the action on $|\psi\theta\phi\rangle$ by

$$P_k|\psi\theta\phi\rangle = \sum_j p_j A_{jk}(\psi\theta\phi)|\psi\theta\phi\rangle, \tag{4.5}$$

then J_k^R and P_k together give rise to a unitary representation of $E(3)$. Here p_k is any triplet of real numbers and characterizes the representation. Let us now choose

p_k to have the form $(0,0,1)$: Then the operators P_k, which are in any case diagonal in the basis $|\psi\theta\phi\rangle$ and so just operators of multiplication, reduce to

$$P_1=\sin\theta\cos\phi,\quad P_2=\sin\theta\sin\phi,\quad P_3=\cos\theta. \tag{4.6}$$

Now, the generators J_k^R and P_k all commute with the operator J_3^L from the left regular representation of $R(3)$. We can break up the full space into the eigenspaces of J_3^L for various eigenvalues $0, \pm1, \pm2, \cdots$. In the eigenspace on which $J_3^L=n\geq 0$, we then find that we have the UIR $(n,1)$ of $E(3)$; while in that on which $J_3^L=n<0$, we have the UIR $(|n|,-1)$. In particular the UIR's of $E(3)$ on the subspaces with $J_3^L=n$ and $J_3^L=-n$ are *not* equivalent.

Starting with this direct sum of UIR's of $E(3)$, which may be symbolically written

$$\sum_{n=0}^{\infty}\oplus(n,1)\oplus\sum_{n=1}^{\infty}\oplus(n,-1),$$

let us apply the construction (4.2) *for* $\rho=0$. Clearly, the subspace on which the UIR $(n,1)$ of $E(3)$ operates now supports the UIR $\{n,0\}$ of the HLG; and in the subspace carrying the UIR $(n,-1)$ of $E(3)$ too we now get the UIR $\{n,0\}$ of the HLG. We have thus the generators for a unitary representation of the HLG, in the space of the regular representation of $R(3)$; and this consists of the direct sum of UIR's

$$\sum_{n=0}^{\infty}\oplus\{n,0\}\oplus\sum_{n=1}^{\infty}\oplus\{n,0\}$$

of the HLG. Notice now that the two subspaces corresponding to $J_3^L=n$ and $J_3^L=-n$ both support the *same* UIR $\{n,0\}$ of the HLG; this has happened because we chose $\rho=0$; otherwise one subspace would support the UIR $\{n,\rho\}$, the other $\{n,-\rho\}$. (Of course, all the HLG generators commute with J_3^L). The generators of this rather special unitary representation of the HLG are easily obtained, using Eq. (4.2) for $\rho=0$, Eq. (4.6) and J_k^R as given in Appendix A. K_3, for example, turns out to be

$$K_3=iP_3+P_2J_1^R-P_1J_2^R=i\left(\cos\theta+\sin\theta\frac{\partial}{\partial\theta}\right). \tag{4.7}$$

This coincides exactly with the result of Sec. 3! Thus with this construction we understand in a new light the forms of the generators for the representation of the HLG analyzed in the last section.

To reduce the present representation to the previous one, two things have to be done and they will be described qualitatively. The operator $\exp(i\pi J_2^L)$ interchanges the eigenspaces corresponding to $J_3^L=n$ and to $J_3^L=-n$ with one another. It is easily checked to also switch the signs of P_k and K_k, i.e.,

$$\exp(i\pi J_2^L)(P_k \text{ or } K_k)\exp(-i\pi J_2^L)=(-P_k \text{ or } -K_k). \tag{4.8}$$

[K_k defined in Eq. (4.2) using P_k and J_k^R]. On the other hand, within a UIR of the HLG of the form $\{j_0,0\}$, there always exists an operator C that commutes with the $R(3)$ generators and anticommutes with the K_k's. Such an operator can then be set up on each eigenspace of J_3^L, and so on the entire space of the regular representation of $R(3)$. This C commutes with J_3^L, and of course with J_k^R. The product $C\exp(i\pi J_2^L)$ then has these properties: it anticommutes with J_3^L but commutes with all the generators of the HLG representation that has been set up on the space of the regular representation of $R(3)$. The subspace consisting of functions on $R(3)$ which are eigenfunctions of $C\exp(i\pi J_2^L)$ with eigenvalue $+1$ is then a Lorentz invariant subspace, and is (apart from a modification in the scalar product) just the space of functions on the isotropic lines obeying the symmetry conditions of Eq. (3.20). In this subspace, each of the UIR's $\{n,0\}$ for $n=0,1,2,\cdots$ occurs just once. (A more detailed analysis of these symmetry relations is given in Appendix B.) The next step to be taken is to essentially modify the scalar product given in Eq. (4.4) by including the operator (J_3^L) as a metric operator. This is to accomodate the factor n that appears in the discrete summation in Eq. (3.35). By these steps, then, the representation of the HLG analyzed in Sec. 3 may be produced from the one set up in this section starting from a representation of $E(3)$.

V. CONCLUSION

In Sec. 1 we defined the "spinless" spacelike UIR of the IHLG, on the Hilbert space $\mathcal{H}$ of all square integrable functions $f(p)$ on the single-sheeted hyperboloid Γ. By a simple extension of the techniques of Ref. 5, we have shown that this contains the following UIR's of the HLG: each of the UIR's $\{0,\rho\}$ for $0<\rho<\infty$ in the form of a direct integral, with multiplicity two; each of the UIR's $\{n,0\}$ for $n=1,2,\cdots$ in the form of a direct sum, with multiplicity one. Half of the former, and those of the latter with n even, belong to the subspace $\mathcal{H}_+$ of even functions on Γ; the rest belong to the subspace $\mathcal{H}_-$ of odd functions. The components of $f(p)$ belonging to the subspaces of these UIR's are given by $F_+(\xi;\rho)$, $F_-(\xi;\rho)$, and $F(\hat{\xi},\hat{b};n)$; these are expressed in terms of f in Eqs. (1.21a), (2.32a), and (3.31), and f can be reconstructed using Eq. (3.35). The original scalar product in $\mathcal{H}$ can be reexpressed in terms of $F_\pm(\xi;\rho)$, $F(\hat{\xi},\hat{b};n)$, exactly as in Ref. 5; the result reads

$$\begin{aligned}\int_\Gamma (dp)\,|f(p)|^2=&\tfrac{1}{2}(4\pi)^{-3}\int_0^\infty \rho^2\,d\rho\int d\Omega(\hat{\xi})[\,|F_+(\hat{\xi};\rho)|^2\\&+|F_-(\hat{\xi};\rho)|^2]\\&+4\pi^{-2}\sum_{n=1}^{\infty}n\int d\Omega(\hat{\xi})\,|F(\hat{\xi},\hat{b};n)|^2.\end{aligned} \tag{5.1}$$

It is interesting to recall the results in the cases of the "spinless" timelike and lightlike UIR's of the IHLG (both with positive energy), because one can then see the changes that occur in the reduction under the HLG. The spinless timelike UIR contains each of the UIR's $\{0,\rho\}$ of the HLG *once*, of course as a direct integral from $\rho=0$ to $\rho=\infty$. When we examine the "spinless" lightlike case, the spectrum just doubles: Each of the UIR's $\{0,\rho\}$ for $0<\rho<\infty$ appears *twice*. Finally in the case analyzed in this paper we get all the UIR's present in the lightlike case, as often, plus the infinite discrete sum of the UIR's $\{n,0\}$. This last is the most striking feature of the spacelike case.

The problem of setting up Lorentz bases in general (not spinless) timelike and lightlike UIR's of the IHLG has been solved by several authors. It would be most interesting to see how the results of the present paper change when a spacelike UIR with nonzero helicity is

examined. Equally interesting, even in the spinless case, is to understand how the energy–momentum operators P_μ act in the Lorentz basis, especially how they "connect" the continuous basis vectors corresponding to $\{0,\rho\}$ to the discrete ones corresponding to $\{n,0\}$.[21] We hope to analyze these problems elsewhere.

APPENDIX A: REGULAR REPRESENTATION OF $R(3)$

The structure of the regular representation of $R(3)$ is of course well known. Our purpose here is just to express it in a slightly unusual form, and to collect some useful expressions.

Elements of $R(3)$ can be identified with real orthogonal unimodular matrices in three dimensions. Using axis-angle variables, for every triplet of real numbers $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ we define the 3×3 matrix $A(\boldsymbol{\alpha})$ by giving its matrix elements

$$A_{jk}(\boldsymbol{\alpha}) = \delta_{jk}\cos\alpha + \alpha_j\alpha_k(1-\cos\alpha)/\alpha^2 - \epsilon_{jkl}\alpha_l \sin\alpha/\alpha, \quad \alpha^2 = \alpha_k\alpha_k. \tag{A1}$$

By giving all directions to $\boldsymbol{\alpha}$, and allowing the range $0 \leq \alpha \leq \pi$, all elements of $R(3)$ are obtained. Some useful relations are

$$A(\boldsymbol{\alpha})^{-1} = A(-\boldsymbol{\alpha}),$$
$$A(\boldsymbol{\alpha})A(\boldsymbol{\beta})A(-\boldsymbol{\alpha}) = A(\boldsymbol{\beta}'), \quad \beta_j' = A_{jk}(\boldsymbol{\alpha})\beta_k. \tag{A2}$$

Another useful geometrical relation is needed. We will think of some fixed triad of unit vectors $e_j^{(0)}$ as given once for all, and will write e_j for a variable triad; both are right-handed. By $R(\boldsymbol{\omega})$ we shall mean the geometrical operation of applying a right-handed rotation by the angle $|\boldsymbol{\omega}|$ about the axis $\omega_j e_j^{(0)}$; $\boldsymbol{\omega}$ is just a triplet of real numbers, and $R(\boldsymbol{\omega})$ is an operator to be applied to vectors. (Sometimes in the text, the magnitude and direction of $\boldsymbol{\omega}$ are separately given). Between the operator $R(\boldsymbol{\omega})$, the matrix $A(\boldsymbol{\alpha})$ and any triad e_j, we then have the useful relation

$$R(\omega)e_j = A_{kj}(\boldsymbol{\alpha})e_k \text{ if } \omega_j e_j^{(0)} = \alpha_j e_j. \tag{A3}$$

The regular representation of $R(3)$ acts in the space of all square integrable functions $f(\boldsymbol{\alpha})$ on $R(3)$, the scalar product being given by

$$(f,f) = \int_0^\pi (1-\cos\alpha)\,d\alpha \int d\Omega(\hat{\alpha})\,|f(\boldsymbol{\alpha})|^2 < \infty. \tag{A4}$$

For the corresponding continuous basis of ket vectors, we will write $|A(\boldsymbol{\alpha})\rangle\rangle$. The operators for the left and right regular representations will be $U(A(\boldsymbol{\alpha}))$, $V(A(\boldsymbol{\alpha}))$, respectively; their generators are J_k^L, J_k^R. These two representation operators act according to

$$U(A(\boldsymbol{\alpha}))\,|A(\boldsymbol{\beta})\rangle\rangle = \exp(-i\boldsymbol{\alpha}\cdot\mathbf{J}^L)\,|A(\boldsymbol{\beta})\rangle\rangle = |A(\boldsymbol{\alpha})A(\boldsymbol{\beta})\rangle\rangle,$$
$$V(A(\boldsymbol{\alpha}))\,|A(\boldsymbol{\beta})\rangle\rangle = \exp(-i\boldsymbol{\alpha}\cdot\mathbf{J}^R)\,|A(\boldsymbol{\beta})\rangle\rangle = |A(\boldsymbol{\beta})A(-\boldsymbol{\alpha})\rangle\rangle. \tag{A5}$$

Now an alternative notation for these same vectors can be developed as follows: For each $\boldsymbol{\alpha}$, let the triad $e_j(\boldsymbol{\alpha})$ be obtained from the fixed triad by using the matrix $A(\boldsymbol{\alpha})$,

$$e_j(\boldsymbol{\alpha}) = A_{jk}(\boldsymbol{\alpha})e_k^{(0)}. \tag{A6}$$

Then we have a one-one correspondence between right-handed triads e_j and elements $A(\boldsymbol{\alpha})$ in $R(3)$, the group element being determined by the relation of the triad e_j to the fixed one $e_j^{(0)}$. The basis kets for the regular representation of $R(3)$ can now be written as $|e_j\rangle\rangle$, the "range" of e_j being all right-handed triads in 3-space, and always being measured relative to $e_j^{(0)}$. By combining Eqs. (A2), (A3), (A5), and (A6), we can express the U's and V's in the new notation in this way:

$$|A_{jk}(\boldsymbol{\alpha})e_k\rangle\rangle = \exp(-i\boldsymbol{\alpha}\cdot\mathbf{J}^L)\,|e_j\rangle\rangle, \tag{A7a}$$
$$|R(\boldsymbol{\alpha})e_j\rangle\rangle = \exp(-i\boldsymbol{\alpha}\cdot\mathbf{J}^R)\,|e_j\rangle\rangle. \tag{A7b}$$

These are very simple equations. It is possible to express the left-hand side of (A7a) in terms of $\mathbf{J}^R$, and that of (A7b) in terms of $\mathbf{J}^L$; but the resulting equations are cumbersome and unnatural. In using this notation for handling functions on isotropic lines, we identified $\hat{\xi}$ with e_3 and $\hat{b}$ with e_1.

If we use Euler angles ψ, θ, ϕ as parameters for $R(3)$, we use in place of the matrix $A(\boldsymbol{\alpha})$ the matrix $A(\psi,\theta,\phi)$

$$A(\psi,\theta,\phi) = \begin{pmatrix} \cos\psi\cos\theta\cos\phi - \sin\psi\sin\phi; & \cos\psi\cos\theta\sin\phi + \sin\psi\cos\phi; & -\cos\psi\sin\theta \\ -\sin\psi\cos\theta\cos\phi - \cos\psi\sin\phi; & -\sin\psi\cos\theta\sin\phi + \cos\psi\cos\phi; & \sin\psi\sin\theta \\ \sin\theta\cos\phi; & \sin\theta\sin\phi; & \cos\theta \end{pmatrix}. \tag{A8}$$

Then, J^R and J^L are given as partial differential operators as follows:

$$J_1^R = i\left(\cot\theta\cos\phi\frac{\partial}{\partial\phi} + \sin\phi\frac{\partial}{\partial\theta} - \frac{\cos\phi}{\sin\theta}\frac{\partial}{\partial\psi}\right),$$
$$J_2^R = i\left(\cot\theta\sin\phi\frac{\partial}{\partial\phi} - \cos\phi\frac{\partial}{\partial\theta} - \frac{\sin\phi}{\sin\theta}\frac{\partial}{\partial\psi}\right),$$
$$J_3^R = -i\frac{\partial}{\partial\phi}, \tag{A9a}$$
$$J_1^L = i\left(\cot\theta\cos\psi\frac{\partial}{\partial\psi} + \sin\psi\frac{\partial}{\partial\theta} - \frac{\cos\psi}{\sin\theta}\frac{\partial}{\partial\phi}\right),$$
$$J_2^L = i\left(-\cot\theta\sin\psi\frac{\partial}{\partial\psi} + \cos\psi\frac{\partial}{\partial\theta} + \frac{\sin\psi}{\sin\theta}\frac{\partial}{\partial\phi}\right),$$
$$J_3^L = i\frac{\partial}{\partial\psi}. \tag{A9b}$$

These expressions are used in Sec. 4.

APPENDIX B: SYMMETRY RELATIONS ON THE FUNCTIONS $\phi(\xi, b)$

The functions $\varphi(\hat{\xi},\hat{b})$ on isotropic lines that arise from functions $f(p)$ on Γ via Eq. (3.18) are subject to the continuous family of constraints given in Eq. (3.20). While these constraints on φ are by their origin Lorentz-

invariant, they are somewhat difficult to grasp for two reasons: (a) the continuous variables p and α appear in them and in fact enumerate them; (2) they are nonlocal in φ in the sense that they do not relate φ at one "point" to φ at another "point" but involve integrals of φ. Their meaning would be more transparant if we could find an alternative description of φ, i.e., a "wavefunction" for $|\varphi\rangle$ in some other basis, in which p and α drop out in the statement of the constraints, and at the same time this statement becomes local. We show here how this may be achieved.

The basic condition on $|\varphi\rangle$ is this: For each p on Γ, the expression

$$\int d\Omega(\hat{\xi})\delta(p^0-\mathbf{p}\cdot\hat{\xi})\langle\hat{\xi},\hat{b}\,|\exp(i\alpha J_3^L)|\,\varphi\rangle, \quad \hat{b}=\mathbf{p}-p^0\hat{\xi}, \tag{B1}$$

should be an even function of α. Let us write here,

$$p=(\sinh\zeta,\hat{p}\cosh\zeta), \tag{B2}$$

and let $\hat{p}$ arise from the vector $e_3^{(0)}$ by the rotation $R(\mathbf{n})$, $\mathbf{n}$ dependent on $\hat{p}$. If then we change variables in (B1) from $\hat{\xi}$ to $\hat{\xi}'$ by $\hat{\xi}=R(\mathbf{n})\hat{\xi}'$, we have the statement

$$\int d\Omega(\hat{\xi}')\delta(\sinh\zeta-\hat{\xi}'^3\cosh\zeta)\langle\hat{\xi}',\hat{b}'\,|\exp(i\mathbf{n}\cdot\mathbf{J}^R)\exp(i\alpha J_3^L)|\,\varphi\rangle$$

should be an even function of α, for all choices of ζ and $\mathbf{n}$. Here $\hat{b}'=e_3^{(0)}\cosh\zeta-\hat{\xi}'\sinh\zeta$. With $|\varphi\rangle$, $\exp(i\mathbf{n}\cdot\mathbf{J}^R)|\varphi\rangle$ will also obey the constraints: dropping primes on $\hat{\xi}$ and $\hat{b}$ we can then say that the allowed vectors $|\varphi\rangle$ are all those that satisfy

$$\int d\Omega(\hat{\xi})\delta(\xi^3-\tanh\zeta)\langle\hat{\xi},\hat{b}\,|\exp(i\alpha J_3^L)|\,\varphi\rangle=\text{even function of } \alpha \text{ for all } \zeta,$$

$$\hat{b}=e_3^{(0)}\cosh\zeta-\hat{\xi}\sinh\zeta, \tag{B3}$$

and all transforms of $|\varphi\rangle$ by the right regular representation of $R(3)$. Let us now use Euler angles for the description of $\hat{\xi}$ and $\hat{b}$, so that the spherical polar angles of $\hat{\xi}$ are θ, ϕ; let us also write $\tanh\zeta=\cos\chi$, $0\leq\chi\leq\pi$. Then it is easily seen that (B3) takes the form

$$\int_0^{2\pi}d\phi\langle e_k^{(0)}\,|\exp(-i\chi J_2^R)\exp(i\phi J_3^R)\exp(i\alpha J_3^L)|\,\varphi\rangle=\text{even in } \alpha \text{ for all } \chi. \tag{B4}$$

We now introduce the basis in which the regular representation of $R(3)$ appears fully reduced. The elements of this basis can be written $|jmn\rangle$ and their basic properties are as follows:

$$|jmn\rangle=\frac{(2j+1)^{1/2}}{8\pi^2}\int_0^{2\pi}d\psi\int_0^{2\pi}d\phi\times\int_0^{\pi}\sin\theta\,d\theta D^j_{mn}(\psi,\theta,\phi)\,|A(\psi,\theta,\phi)\rangle,$$

$$|A(\psi,\theta,\phi)\rangle=\sum_{jmn}(2j+1)^{1/2}D^j_{mn}(\psi,\theta,\phi)^*\,|jmn\rangle,$$

$$\exp(-i\phi J_3^L)\exp(-i\theta J_2^L)\exp(-i\psi J_3^L)\,|jmn\rangle=\sum_{m'}D^j_{mm'}(\psi,\theta,\phi)\,|jm'n\rangle,$$

$$\exp(i\psi J_3^R)\exp(i\theta J_2^R)\exp(i\phi J_3^R)\,|jmn\rangle=\sum_{n'}D^j_{n'n}(\psi,\theta,\phi)\,|jmn'\rangle,$$

$$\langle j'm'n'\,|jmn\rangle=\delta_{j'j}\delta_{m'm}\delta_{n'n}, \quad j=0,1,2,\cdots,$$

$$|m|,\ |n|\leq j. \tag{B5}$$

The functions $D^j_{mn}(\psi,\theta,\phi)$ are familiar from angular momentum theory, and the main properties of these that we need are

$$D^j_{mn}(\psi,\theta,\phi)=\exp(im\psi+in\phi)d^j_{mn}(\theta),$$

$$d^j_{-m,-n}(\theta)=(-1)^{m-n}d^j_{mn}(\theta). \tag{B6}$$

The state $|e_k^{(0)}\rangle$ corresponds to taking $\psi=\theta=\phi=0$, and we can use for it

$$|e_k^{(0)}\rangle=\sum_{jm}(2j+1)^{1/2}\,|jmm\rangle. \tag{B7}$$

Then (B4) reads

$$\sum_{jmm'}(2j+1)^{1/2}\langle jmm\,|\exp(-i\chi J_2^R)|\,jmm'\rangle\int_0^{2\pi}d\phi\times\exp(im'\phi+im\alpha)\langle jmm'\,|\,\varphi\rangle=\text{even in } \alpha \text{ for all } \chi,$$

i.e.,

$$\sum_{jm}(2j+1)^{1/2}d^j_{0m}(\chi)\exp(im\alpha)\langle jm0\,|\,\varphi\rangle=\text{even in } \alpha \text{ for all } \chi. \tag{B8}$$

Using (B6) we can make the coefficient of $\exp(im\alpha)$ in this relation odd in m; alternatively, we can write (B8) as

$$\sum_{jm}(2j+1)^{1/2}d^j_{0m}(\chi)\exp(im\alpha)\{\langle jm0\,|\,\varphi\rangle-(-1)^m\times\langle j,-m,0\,|\,\varphi\rangle\}=0 \text{ for all } \alpha \text{ and } \chi. \tag{B9}$$

From the α dependence we get

$$\sum_j(2j+1)^{1/2}d^j_{0m}(\chi)\{\langle jm0\,|\,\varphi\rangle-(-1)^m\langle j,-m,0\,|\,\varphi\rangle\}=0 \text{ for all } m,\chi. \tag{B10}$$

For fixed m, the $d^j_{0m}(\chi)$ for various j constitute a complete linearly independent set of functions of χ, so that we have the simple constraints

$$\langle jm0\,|\,\varphi\rangle=(-1)^m\langle j,-m,0\,|\,\varphi\rangle \text{ for all } j \text{ and } m. \tag{B11}$$

Applying to $|\varphi\rangle$ a general element of the right regular representation of $R(3)$, we have the final statement of the constraints in the form:

$$\langle jmn\,|\,\varphi\rangle\equiv\varphi^{(m)}_{jn},$$

$$\varphi^{(-m)}_{jn}=(-1)^m\varphi^{(m)}_{jn}. \tag{B12}$$

We can look upon the sequence of numbers $\varphi^{(m)}_{jn}$ as constituting the "wavefunction" of $|\varphi\rangle$ in the basis $\langle jmn|$. The constraints then appear in a "local" form here, and also are not dependent on parameters like a variable point on Γ. m is the Lorentz-invariant index on φ, and is the eigenvalue of $-J_3^L$.

A similar treatment can no doubt be given to put the symmetry relations on $h_\pm(\xi)$, namely Eqs. (1.13) and (2.28), into a simple form; but we will omit the details here.

[1]E. P. Wigner, Ann. Math. **40**, 149 (1939).
[2]For a textbook treatment, see, for example, A. D. Martin and T. D. Spearman, *Elementary Particle Theory* (North-Holland, Amsterdam, 1970).
[3]I. S. Shapiro, Sov. Phys.-Dokl. **1**, 91 (1956); Chou Kuang-Chao and L. G. Zastavenko, Sov. Phys.-JETP **8**, 990 (1959); N. Mukunda, J.

Math. Phys. **9**, 532 (1968); A. Chakrabarti, M. Levy-Nahas, and R. Seneor, J. Math. Phys. **9**, 1274 (1968); A. Chakrabarti, J. Math. Phys. **12**, 1822 (1971).

[4]This problem has been considered in A. Chakrabarti, Ref. 3; however, this paper does not contain a definite statement saying what UIR's of the HLG are present in a given spacelike UIR of the IHLG.

[5]I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5, especially Chaps. V. and VI.

[6]A. Bassetto and M. Toller, "Harmonic Analysis on the One-Sheet Hyperboloid and Multiperipheral Inclusive Distributions," TH. 1499-CERN (May 1972).

[7]Improper transformations like space reflection and time reversal will not be considered in this paper. Greek indices run from 0 to 3, Latin ones from 1 to 3; the metric tensor $g_{\mu\nu}$ with $g_{00} = +1$, is used for raising and lowering indices. The invariant scalar product of two 4-vectors p^μ, q^μ is written $p \cdot q$.

[8]I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Macmillan, New York, 1963). In our work, we do not need to work with the Supplementary Series of UIR's of the HLG.

[9]Strictly speaking, the following analysis is carried out working with functions $f_+(p)$ which are infinitely differentiable and have compact support on Γ; later by the process of completion results valid for all f_+ in $\mathscr{H}_+$ are obtained.

[10]As it stands, Eq. (1.13) may contain divergent quantities, so it has to be suitably regularized; this is done in Ref. 5.

[11]The second term in Eq. (1.16) may diverge near $\alpha = 0$, so it has to be regularized.

[12]I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1, p. 60.

[13]Reference 12, p. 258.

[14]Reference 12, p. 257.

[15]Reference 12, p. 60.

[16]Given the convergence of the defining integral (2.24) for $K_-(p, p', t)$, one could arrive at the symmetry relation (2.27) more easily by using this fact: for any two points p, p' on Γ, a homogeneous Lorentz transformation Λ can be found that will do one of two things, either interchange p and p', or take p into $-p'$ and p' into $-p$.

[17]If $f(p)$ is an even function of p, these equations become exactly the same as (1.23), (1.24), (1.25b).

[18]A. Sankaranarayanan, Nuovo Cimento **38**, 1441 (1965); M. Y. Han, Nuovo Cimento **42B**, 367 (1966); J. Rosen and P. Roman, J. Math. Phys. **7**, 2072 (1966). See also N. Mukunda, J. Math. Phys. **10**, 897 (1969), especially Appendix A.

[19]W. Pauli, " Continuous Groups in Quantum Mechanics," in *Ergebnisse der exakten Naturwissenschaften* (Springer-Verlag, Berlin, 1965), Vol. 37.

[20]The construction is due to C. Goebel, Phys. Rev. Lett. **16**, 1130 (1966); also in *Noncompact Groups in Particle Physics*, edited by Y. Chow (Benjamin, New York, 1966). it is also summarized in N. Mukunda (Ref. 18), Appendix B. Note, however, that, in comparison with the latter, we have switched the roles of the left and right regular representations.

[21]A somewhat similar problem, and its solution, are discussed in N. Mukunda, J. Math. Phys. **9**, 50 (1968).

On the relaxation to quantum-statistical equilibrium of the Wigner-Weisskopf atom in a one-dimensional radiation field. VI. Influence of the coupling function on the dynamics

Janilla Jyy-wen Yang*

Department of Chemistry, University of Notre Dame, Notre Dame, Indiana 46556

Russell Davidson

Department of Physics, University of British Columbia, Vancouver 8, Canada

John J. Kozak

Department of Chemistry and Radiation Laboratory, University of Notre Dame, Notre Dame, Indiana 46556

(Received 26 March 1973)

An investigation is made of the effect on the dynamics of spontaneous emission of the Wigner–Weisskopf atom in an infinite-system limit of various choices of the coupling function or form factor describing the atom's interaction with the spectrum of the radiation field. This is carried out both for the exact solution to the problem of spontaneous emission, obtained in the earlier papers in the series, and for some approximate solutions, also previously considered, in particular one based on the Schrödinger equation of the problem and one based on the weak-coupling Prigogine–Résibois master equation. The details of the form factor are found, by numerical computation of the solutions, to be critical in determining the nonexponential parts of the solutions, and these parts are seen to be capable in some cases of dominating the exponential parts, which are given only by the values of the form factor near the resonance energy. The approximate solutions discussed are found to vary widely in their worth, and one, which yields the exact solution for the Wigner–Weisskopf problem, is singled out as being of probable use in the statistical-mechanical description of more complicated systems.

I. INTRODUCTION

This study is a continuation of previous work of the authors, (Refs. 1–5, hereafter referred to as I–V, respectively) especially that presented in V. There a study was undertaken of the properties of an exact solution of the problem of spontaneous emission of a two-level atom in interaction with a one-dimensional radiation field, in the limit where the size of the system becomes infinite. This exact solution had been obtained earlier in III, and for a certain choice of coupling function it was found in V that the solution could be written down in closed form, and the resulting expression studied numerically. This exact solution was then compared with various weak-coupling approximations, and it was concluded that in the limit of weak coupling, an approximation based on the Schrödinger equation was acceptable, whereas for all values of the coupling constant in the given choice of coupling function, the predictions of the Prigogine–Résibois master equation were unacceptable, except when the model was ergodic in a certain sense.

To illustrate more clearly the objective of the present paper, the main features of the model studied in I–V will now be reviewed. The Hamiltonian is

$$H=\epsilon_1\alpha\alpha^*+\epsilon_2\alpha^*\alpha+\sum_\lambda\left[\tfrac{1}{2}\hbar\omega_\lambda(a_\lambda^*a_\lambda+1)\right]+\sum_\lambda\left(h_\lambda^*\,\alpha^*\,a_\lambda+h_\lambda\,\alpha\,a_\lambda^*\right), \tag{1}$$

in which ϵ_1, ϵ_2 are, respectively, the energies of the ground state $|1\rangle$ and excited state $|2\rangle$ of the two-level atom. The operators α and a_λ are defined by the following equations:

$$\alpha=|1\rangle\langle 2|,$$

$$\alpha^*=|2\rangle\langle 1|,$$

$$\langle n_\lambda|a_\lambda|m_\lambda\rangle=[2(n_\lambda+1)]^{1/2}\delta^{\mathrm{kr}}(m_\lambda-n_\lambda-1)=\langle m_\lambda|a_\lambda^*|n_\lambda\rangle.$$

The state $|n_\lambda\rangle$ is one with $n_\lambda(=0,1,2,\cdots)$ photons in the λth mode, and $\delta^{\mathrm{kr}}(\cdots)$ denotes the Kronecker delta. The states of the system, between which matrix elements of the Hamiltonian, Eq. (1), are to be taken, are given by

$$|i;\{n_\lambda\}\rangle=|i\rangle\prod_\lambda|n_\lambda\rangle,$$

with $i=1,2$ and $n_\lambda=0,1,2,\cdots$. But, when the problem of spontaneous emission is considered, that is, when the initial state of the system is taken as

$$|2;\{0\}\rangle \tag{2}$$

then the only states accessible under the evolution given by Eq. (1) are

$$|2;\{0\}\rangle,\quad |1;0,0,\cdots,0,1,0,\cdots 0,0\rangle,$$

where, if the atom is de-excited, there can be only one photon present. These states will be written as $|/\!/\rangle$ and $|\lambda\rangle$, respectively; λ ranges over all modes of the field, the properties of which are to be determined by the size and dimensionality of the system. If the zero of energy is chosen to be the state

$$|1;\{0\}\rangle$$

—a state not accessible given the initial condition (2)—then the Hamiltonian can be conveniently rewritten as

$$H=\hbar E|/\!/\rangle\langle/\!/|+\sum_\lambda\hbar\omega_\lambda|\lambda\rangle\langle\lambda|+\sum_\lambda\left[h_\lambda\sqrt{2}\,|\lambda\rangle\langle/\!/|+h_\lambda^*\sqrt{2}\,|/\!/\rangle\langle\lambda|\right], \tag{3}$$

where $\hbar E=\epsilon_2-\epsilon_1$ is the energy separating the two levels of the atom, $\hbar\omega_\lambda$ is the energy of a photon in the λth mode, and h_λ is a (possibly complex) coupling function to be specified below. As shown explicitly in III, if the system is in the state $|/\!/\rangle$ at $t=0$, then the probability

amplitude for finding the state at time t is exactly given by [see Eq. (III. 12)]

$$\langle N|\Psi(t)\rangle = -\frac{1}{2\pi i}\int_C dz\, \exp(-izt)\left(z - E - \sum_\lambda \frac{2|h_\lambda|^2}{\hbar^2(z-\omega_\lambda)}\right)^{-1}, \tag{4}$$

where C is a contour above the real axis and parallel to it. This result was obtained by solving directly the time-dependent Schrödinger equation for the problem, and it is to be noted that Eq. (4) is in the form of an inverse Laplace transform integral.

These remarks bring us to the main point of the present paper. In I, it was shown that in order to make the model defined by the Hamiltonian, Eq. (1), resemble as closely as possible the situation of an electron in an atom interacting with a (one-dimensional) field of electromagnetic radiation, then the following specification for the coupling function, h_λ, was necessary:

$$|h_\lambda|^2 = \alpha\hbar^2c^2|k_\lambda|/L, \tag{5}$$

where L is the length of the system, c is the velocity of light, and α is a dimensionless coupling constant which corresponds for a one-dimensional system to the fine-structure constant of quantum electrodynamics. It was noted in I that this choice led to the ultraviolet divergence in the calculation of the probability $|\langle N|\Psi(t)\rangle|^2$ from the Prigogine—Résibois master equation. There, it was pointed out that this divergence could be avoided in several ways, the simplest being to cut off the divergent part of the offending integral [Eq. (I. 41)] at a finite upper limit. This possibility was not pursued in I for several reasons, not the least of which was that the use of this approximation would have necessitated the introduction of an additional parameter, namely, a (rather arbitrary) bound on the frequency spectrum of the radiation field. Instead, in I and II, the coupling function itself was modified by removing its dependence on the wave vector, $|k_\lambda|$. Specifically, the factor $|k_\lambda|$ was replaced by its resonant value, E/c, so that the coupling function assumed the form

$$|h_\lambda|^2 = \alpha\hbar^2cE/L. \tag{6}$$

This was enough to ensure the convergence of the expressions derived in I and II as approximate solutions to the Prigogine—Résibois master equation for the diagonal elements of the density matrix. A motivation for the choice, Eq. (6), was the observation that this procedure is effectively what is done in the Born approximation of time-dependent perturbation theory in quantum mechanics for such problems. It was interesting, then, that in going to a representation of the problem based on the time-dependent Schrödinger equation in III, the choice of $|h_\lambda|^2$ adopted in I and II, namely Eq. (6), caused the summation in the expression for the wave function, Eq. (4), to diverge, given the usual choice of frequency spectrum

$$\omega_\lambda = |2\pi cn_\lambda/L|, \quad n_\lambda \text{ an integer}.$$

Accordingly, in that study and in IV and V as well, to avoid the necessity of introducing an upper bound on the frequency spectrum, the following expedient was used: $|h_\lambda|^2$ was made proportional to a negative power $-p$ of ω_λ in such a way that, for $\omega_\lambda = E$, it assumed the value given by Eq. (6). This led to

$$|h_\lambda|^2 = \alpha\hbar^2c^{1-p}E^{1+p}/L|k_\lambda|^p \tag{7}$$

and here it is necessary to specify further that $0 < p < 1$, in order to avoid the infrared divergence in the summation Eq. (4). In III, IV, and V, a full study, both analytical and numerical, was made of the consequences of Eq. (4) with the $|h_\lambda|^2$ of Eq. (7), for systems both finite and infinite in extent. It was seen (in V) that still more difficulties arose, connected with the ergodic properties of the system. It turned out to be necessary to make $|h_\lambda|^2$ go to zero with $|k_\lambda|$ in such a way that the quantity

$$\sum_\lambda (2|h_\lambda|^2/\hbar^2\omega_\lambda)$$

remained finite if the "ergodic" property

$$|\langle N|\Psi(t)\rangle| \to 0 \quad \text{as } t \to \infty$$

was to hold even for arbitrarily small coupling. A possible choice of $|h_\lambda|^2$ that would achieve this was given [Eq. (V-32)]:

$$|h_\lambda|^2 = \hbar^2\alpha c^2|k_\lambda|/L \quad \text{for } k_\lambda \leq E/c$$
$$= \hbar^2\alpha c^{1/2}E^{3/2}/L|k_\lambda|^{1/2} \quad \text{for } k_\lambda \geq E/c.$$

A priori, there is no means of knowing what the effect is of making different choices of $|h_\lambda|^2$ on the detailed dynamics of the system. It would seem to be of some interest, then, to investigate this problem in some generality, in order to see which of the conclusions drawn in I—V regarding the worth of various approximation schemes are true irrespective of the exact form of $|h_\lambda|^2$, as well as to see the influence of this form on the exact solution. For dimensional reasons, we may take as a general expression for $|h_\lambda|^2$:

$$|h_\lambda|^2 = (\alpha\hbar^2Ec/L)f(c|k_\lambda|/E), \tag{8}$$

where we impose $f(1) = 1$ so that $|h_\lambda|^2$ shall always have the same value at resonance, where in order to avoid divergences we require

$$f(x) = O(x) \quad \text{as } x \to \infty,$$
$$xf(x) = O(1) \quad \text{as } x \to 0^+,$$

and where we *may* also ask that

$$f(x) \to 0 \quad \text{as } x \to 0^+, \tag{9}$$

for the ergodic property. We shall be especially interested in the cases

$$f(x) = x^p,$$

where, if p is positive, an upper cutoff is employed to evaluate the summation in Eq. (4), and, if p is negative (but always $p > -1$) Eq. (9) is not satisfied. With the choice $p = 1$, Eq. (8) reduces to Eq. (5); with $p = 0$, it reduces to Eq. (6).

We shall also be interested in a choice of form factor that satisfies Eq. (9). Although that given above [see Eq. (V. 32)], for which

$$f(x) = x \quad (x \leq 1),$$
$$f(x) = x^{-1/2} \quad (x \geq 1)$$

is a possible such choice, it has been thought preferable to use instead

$$f(x)=4x/(1+x)^2$$

which has several advantages of simplicity for the purposes of the calculations of the remainder of the paper, as well as having a distinctly different analytic form from those of the other coupling functions considered.

In Sec. II, some of the details of the computation of Eq. (4) will be given, with special emphasis on the case $f(x)=x^p$ (with cutoff). The matter of weak-coupling approximations will be brought up in Sec. III; Sec. IV contains a description of the numerical work performed to elucidate the arguments of the paper, and in Sec. V these arguments are discussed and conclusions are drawn.

II. THE FORMAL SOLUTION

We proceed from the exact solution for the probability amplitude for finding the system in the state $|N\rangle$ at time t, given that the system was in this state at time $t=0$. This exact solution was obtained in III and is given by Eq. (4) of the preceding section. As noted in that section, we shall specify the coupling function to be given by Eq. (8).

Our calculation will be performed in the limit of an infinite system, for which summations over modes λ can be replaced by integrals over the wavenumber k, with the following correspondences:

$$(2\pi/L)\sum_\lambda \to \int_{-\infty}^{+\infty},$$
$$\omega_\lambda = c|k_\lambda|.$$

In this limit, then, the summation appearing in Eq. (4) becomes

$$\sum_\lambda \frac{2|\hbar_\lambda|^2}{\hbar^2(z-\omega_\lambda)} = \frac{2\alpha E c}{L}\sum_\lambda \frac{f(c|k_\lambda|/E)}{z-c|k_\lambda|},$$
$$\to \frac{L}{2\pi}\,\frac{2\alpha E c}{L}\int_{-\infty}^{+\infty} d|k_\lambda|\,\frac{f(c|k_\lambda|/E)}{z-c|k_\lambda|},$$
$$= \frac{2\alpha E}{\pi}\int_0^\infty d\omega \frac{f(\omega/E)}{z-\omega}. \tag{10}$$

Equation (4) now becomes

$$\langle N|\Psi(t)\rangle = -\frac{1}{2\pi i}\int_c \frac{dz\, e^{-izt}}{z-E+(2\alpha E/\pi)\int_0^\infty d\omega[f(\omega/E)/(\omega-z)]} \tag{11}$$

At this stage it is convenient to introduce the dimensionless variables used in I–V and necessary for the proper taking of the weak-coupling limit:

$$\tau = \alpha E t,$$
$$\xi = z/\alpha E.$$

Equation (11) then becomes

$$\langle N|\Psi(\tau)\rangle = -\frac{1}{2\pi i}\int_c \frac{d\xi\, e^{-i\xi\tau}}{\xi-\alpha^{-1}+(2/\pi)\int_0^\infty [f(\alpha\xi')d\xi'/\xi'-\xi]}. \tag{12}$$

We shall first consider the evaluation of the integral in the denominator of Eq. (12) for the case

$$f(x)=x^p, \quad p=1,2,3.$$

with a cutoff at $\xi'=R/\alpha$. It is an elementary result that

$$\int_0^{R/\alpha}\frac{(\alpha\xi')^p\,d\xi'}{\xi'-\xi} = \sum_{n=0}^{p-1}\frac{(\alpha\xi)^n}{p-n}R^{p-n}+(\alpha\xi)^p\log\left(1-\frac{R}{\alpha\xi}\right), \tag{13}$$

where the logarithm, as a function of ξ, is defined on a plane cut between zero and αR, and is purely real for real ξ outside the interval $[0, \alpha R]$. In order to demonstrate more clearly the steps in the evaluation of Eq. (12), we consider for the moment only the case $p=1$; the procedure for $p=2, 3, \cdots$ is exactly similar, and only final results will be given in these cases. The denominator of the integrand in Eq. (12) is then

$$\xi-\frac{1}{\alpha}+\frac{2R}{\pi}+\frac{2\alpha\xi}{\pi}\log\left(1-\frac{R}{\alpha\xi}\right). \tag{14}$$

This expression can be separated into real and imaginary parts, so as to determine the nature and location of its zeros. Writing

$$\xi = x+iy,$$

we obtain

$$\mathrm{Re(den)} = x-\frac{1}{\alpha}+\frac{2R}{\pi}+\frac{2\alpha x}{\pi}\log\left|1-\frac{R}{\alpha\xi}\right|-\frac{2\alpha y}{\pi}\arg\left(1-\frac{R}{\alpha\xi}\right),$$

$$\mathrm{Im(den)} = y+\frac{2\alpha y}{\pi}\log\left|1-\frac{R}{\alpha\xi}\right|+\frac{2\alpha x}{\pi}\arg\left(1-\frac{R}{\alpha\xi}\right).$$

One can see that simultaneous vanishing of both Re(den) and Im(den) can take place only if $y=0$ and $x<0$ or $x>R/\alpha$ (this condition corresponds to requiring hermiticity of the Hamiltonian); further, if $y=0$, one finds that

$$\mathrm{Re(den)} = x-\frac{1}{\alpha}+\frac{2R}{\pi}+\frac{2\alpha x}{\pi}\log\left|1-\frac{R}{\alpha x}\right|,$$
$$\mathrm{Im(den)} = 0.$$

Since

$$\lim_{x\to\pm\infty}\mathrm{Re(den)} = \pm\infty,$$

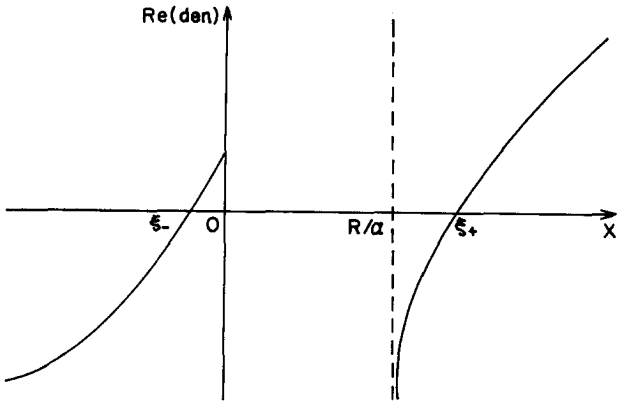

FIG. 1. A sketch of Re(den) versus x. The ξ_+ and ξ_- refer to the simple poles on the positive and negative x axis, respectively.

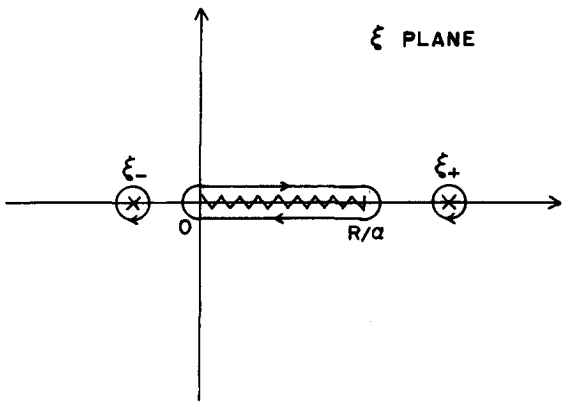

FIG. 2. The contour used in the evaluation of Eq. (12).

$$\lim_{x\to R/\alpha^+} \mathrm{Re(den)} = -\infty,$$

$$\lim_{x\to 0} \mathrm{Re(den)} = (2R/\pi) - \alpha^{-1},$$

the behavior of Re(den) as a function of x is (approximately) as shown in Fig. 1 if $2R/\pi > \alpha^{-1}$ (large cutoff or strong coupling). There are accordingly at most two zeros of the denominator, at $\xi_+ (> R/\alpha)$ and at $\xi_- (< 0)$, although the latter does not appear—for a given cutoff—if the coupling is sufficiently weak [see Eq. (V. 29)]. The zero ξ_+ always occurs, and prevents the appearance of the ergodic property even for small α with this kind of cutoff $|h_\lambda|^2$. Given this behavior, then, and choosing the contour shown in Fig. 2 for the evaluation by Cauchy's theorem of Eq. (12), we have that

$$\langle N|\Psi(\tau)\rangle = +\sum_{\xi_+,\xi_-} \text{Residue} + \frac{2\alpha}{\pi}\int_0^{R/\alpha} \frac{dx\, x e^{-i\tau x}}{[x-\alpha^{-1}+(2R/\pi)+(2\alpha x/\pi)\log|1-R/\alpha x|]^2+4\alpha^2x^2}$$

since the contributions from the large semicircle and the two small circles obviously vanish. The residues are determined in the usual manner:

$$\text{Residue}_\pm = \exp(-i\xi_\pm\tau)/D_1(\xi_\pm),$$

where

$$D_1(\xi) \equiv 1 + \frac{2\alpha}{\pi}\log\left|1-\frac{R}{\alpha\xi}\right| + \frac{2\alpha R}{\pi(\alpha\xi - R)}$$

is just the derivative of the denominator (14). Then for the "one-pole" case, $\alpha < \pi/2R$, the result is

$$\begin{aligned}\langle N|\Psi(\tau)\rangle &= \frac{\cos(\tau\xi_+)}{D_1(\xi_+)} + \frac{2\alpha}{\pi}\int_0^{R/\alpha} \frac{x\cos(x\tau)\,dx}{[x-\alpha^{-1}+(2R/\pi)+(2\alpha x/\pi)\log|1-R/\alpha x|]^2+4\alpha^2x^2} \\ &\quad - i\left(\frac{\sin(\tau\xi_+)}{D_1(\xi_+)} + \frac{2\alpha}{\pi}\int_0^{R/\alpha} \frac{x\sin(x\tau)dx}{[x-\alpha^{-1}+(2R/\pi)+(2\alpha x/\pi)\log|1-R/\alpha x|]^2+4\alpha^2x^2}\right).\end{aligned} \tag{15}$$

Accordingly, the (diagonal) element of the density matrix giving the probability of the state $|N\rangle$ as a function of time is

$$\begin{aligned}\rho(\tau) &\equiv |\langle N|\Psi(\tau)\rangle|^2 \\ &= \left(\frac{\cos(\tau\xi_+)}{D_1(\xi_+)} + \frac{2\alpha}{\pi}\int_0^{R/\alpha} \frac{x\cos(x\tau)\,dx}{[x-\alpha^{-1}+(2R/\pi)+(2\alpha x/\pi)\log|1-R/\alpha x|]^2+4\alpha^2x^2}\right)^2 \\ &\quad + \left(\frac{\sin(\tau\xi_+)}{D_1(\xi_+)} + \frac{2\alpha}{\pi}\int_0^{R/\alpha} \frac{x\sin(x\tau)\,dx}{[x-\alpha^{-1}+(2R/\pi)+(2\alpha x/\pi)\log|1-R/\alpha x|]^2+4\alpha^2x^2}\right)^2\end{aligned} \tag{16}$$

The corresponding results for $p = 2, 3$ for the case that a single pole contributes are given by the following expressions:
For $p = 2$

$$\begin{aligned}\rho(\tau) &= \left(\frac{\cos(\tau\xi_+)}{D_1(\xi_+)} + \frac{2\alpha^2}{\pi}\int_0^{R/\alpha} \frac{x^2\cos(x\tau)\,dx}{[x-\alpha^{-1}+(R^2/\pi)+(2\alpha Rx/\pi)+(2\alpha^2x^2/\pi)\log|1-R/\alpha x|]^2+4\alpha^4x^4}\right)^2 \\ &\quad + \left(\frac{\sin(\tau\xi_+)}{D_1(\xi_+)} + \frac{2\alpha^2}{\pi}\int_0^{R/\alpha} \frac{x^2\sin(x\tau)\,dx}{[x-\alpha^{-1}+(R^2/\pi)+(2\alpha Rx/\pi)+(2\alpha^2x^2/\pi)\log|1-R/\alpha x|]^2+4\alpha^4x^4}\right)^2,\end{aligned}$$

where

$$D_1(\xi_+) = 1 + \frac{2\alpha R}{\pi} + \frac{4\alpha^2\xi}{\pi}\log\left|1-\frac{R}{\alpha\xi}\right| + \frac{2\alpha^2 R\xi}{\pi(\alpha\xi - R)}.$$

For $p = 3$

$$\begin{aligned}\rho(\tau) &= \left(\frac{\cos(\tau\xi_+)}{D_1(\xi_+)} + \frac{2\alpha^3}{\pi}\int_0^{R/\alpha} \frac{x^3\cos(x\tau)\,dx}{[x-\alpha^{-1}+(2R^3/3\pi)+(\alpha R^2x/\pi)+(2\alpha^2Rx^2/\pi)+(2\alpha^3x^3/\pi)\log|1-R/\alpha x|]^2+4\alpha^6x^6}\right)^2 \\ &\quad + \left(\frac{\sin(\tau\xi_+)}{D_1(\xi_+)} + \frac{2\alpha^3}{\pi}\int_0^{R/\alpha} \frac{x^3\sin(x\tau)\,dx}{[x-\alpha^{-1}+(2R^3/3\pi)+(\alpha R^2x/\pi)+(2\alpha^2Rx^2/\pi)+(2\alpha^3x^3/\pi)\log|1-R/\alpha x|]^2+4\alpha^6x^6}\right)^2,\end{aligned} \tag{17}$$

where

$$D_1(\xi_+)=1+\frac{\alpha R^2}{\pi}+\frac{4\alpha^2R\xi}{\pi}+\frac{6\alpha^3\xi^2}{\pi}\log\left|1-\frac{R}{\alpha\xi}\right|+\frac{2\alpha^3\xi^2R}{\pi(\alpha\xi-R)}.$$

We remark that the "two-pole" results are similar in structure to the "one-pole" results, except that there exist in the expression for $\rho(\tau)$ additional terms which spring from the residue at ξ_-. Thus, for example, the density matrix for $p=1$ for the case that two poles contribute is given by

$$\rho(\tau)=\left(\frac{\cos(\tau\xi_-)}{D_1(\xi_-)}+\frac{\cos(\tau\xi_+)}{D_1(\xi_+)}+\frac{2\alpha}{\pi}\int_0^{R/\alpha}\frac{x\cos(x\tau)\,dx}{[x-\alpha^{-1}+(2R/\pi)+(2\alpha x/\pi)\log|1-R/\alpha x|]^2+4\alpha^2x^2}\right)^2$$
$$+\left(\frac{\sin(\tau\xi_-)}{D_1(\xi_-)}+\frac{\sin(\tau\xi_+)}{D_1(\xi_+)}+\frac{2\alpha}{\pi}\int_0^{R/\alpha}\frac{x\sin(x\tau)\,dx}{[x-\alpha^{-1}+(2R/\pi)+(2\alpha x/\pi)\log|1-R/\alpha x|]^2+4\alpha^2x^2}\right)^2. \tag{18}$$

The cases $p=2, 3$ have similar generalizations.

Finally, for purposes of a later discussion, we include below the expression obtained for the density matrix $f(\tau)$ derived assuming the form factor

$$f(x)=4x/(1+x)^2 \quad (\alpha<\pi/8).$$

The exact solution for the time evolution of $\rho(\tau)$ corresponding to this choice is

$$\rho(\tau)=\left(\frac{8}{\pi}\int_0^\infty\frac{\cos(\xi\tau)\,\alpha\xi(\alpha\xi+1)^2\,d\xi}{(\alpha\xi+1)^4[\xi-\alpha^{-1}-G(\xi)]^2+64(\alpha\xi)^2}\right)^2$$
$$+\left(\frac{8}{\pi}\int_0^\infty\frac{\sin(\xi\tau)\,\alpha\xi(\alpha\xi+1)^2\,d\xi}{(\alpha\xi+1)^4[\xi-\alpha^{-1}-G(\xi)]^2+64(\alpha\xi)^2}\right)^2, \tag{19}$$

where

$$G(\xi)=(8/\pi)[(\alpha\xi+1)^2]^{-1}[\alpha\xi\log(\alpha\xi)-(\alpha\xi+1)].$$

III. USE OF A WEAK-COUPLING APPROXIMATION

As a matter of practical necessity, it is often necessary to consider weak-coupling approximations to a given formal result. The validity of a weak-coupling approximation to the Schrödinger equation and to the Prigogine—Résibois master equation was studied extensively in IV and V for systems finite in extent and infinite in extent, and for the particular choice of coupling function, Fig. (7). In this section, we investigate the validity of a weak-coupling approximation to these two equations, but with the dynamics governed by the choice, Eq. (8).

A. The weak-coupling quantum-mechanical solution

Given the remarks in IV and V, one might reasonably expect that a weak-coupling approximation to the time-dependent Schrödinger equation would be in fair accord with the results obtained in the previous section. To investigate whether this is indeed the case, given a different choice of coupling function, we develop a weak-coupling approximation to Eq. (4).

We make use of the operator identity

$$(z-H/\hbar)^{-1}=(z-H_0/\hbar)^{-1}+(z-H_0/\hbar)^{-1}(V/\hbar)(z-H/\hbar)^{-1}, \tag{20}$$

where we make the identifications

$$H_0=\epsilon_1\alpha\alpha^*+\epsilon_2\alpha^*\alpha+\sum_\lambda[\tfrac{1}{2}\hbar\omega_\lambda(a_\lambda^*a_\lambda+1)],$$
$$V=\sum_\lambda(h_\lambda^*\alpha^*a_\lambda+h_\lambda\alpha a_\lambda^*).$$

Here, V represents the interaction part of the Hamiltonian, Eq. (1). By repeated application of this operator identity, one can develop an iterative expansion of Eq. (20) which leads to the result [see Eq. (III. 25)]

$$\langle\mathcal{N}|\Psi(t)\rangle=(2\pi i)^{-1}\int_c dz$$
$$\times\exp(-izt)\left(\sum_{n=0}^\infty(z-E)^{-1}\langle\mathcal{N}|[(V/\hbar)(z-H_0/\hbar)^{-1}]^n|\mathcal{N}\rangle\right). \tag{21}$$

The terms corresponding to $n=0$ and $n=1$ vanish, so we take as our weak-coupling approximation the term corresponding to $n=2$ in the above expansion. The result is

$$\langle\mathcal{N}|\Psi(t)\rangle=-(2\pi i)^{-1}\int_c dz\exp(-izt)(z-E)^{-1}[1+(z-E)^{-1}\Sigma(z)],$$

where

$$\Sigma(z)=\sum_\lambda[2|h_\lambda|^2/\hbar^2(z-\omega_\lambda)]$$

whence, on introducing dimensionless variables, one obtains

$$\langle\mathcal{N}|\Psi(\tau)\rangle=-(2\pi i)^{-1}\int_c d\xi$$
$$\times\exp(-i\xi\tau)(\xi-\alpha^{-1})^{-1}[1+\sigma(\xi)/(\xi-\alpha^{-1})],$$
$$\sigma(\xi)=\Sigma(\alpha E\xi)/\alpha E.$$

Just as in III, it is necessary to extract a factor $\exp(-iEt)=\exp(-i\tau/\alpha)$ if this expression is to be meaningful in the weak-coupling limit $\alpha\to 0$. Then, we have

$$\exp(iEt)\langle\mathcal{N}|\Psi(\tau)\rangle=1-(2\pi i)^{-1}\int_c d\xi\exp(-i\xi\tau)[\sigma(\xi+\alpha^{-1})/\xi^2]. \tag{22}$$

For an infinite system we may put here

$$\sigma(\xi+\alpha^{-1})=(2/\pi)\int_0^\infty d\xi'\,[f(\alpha\xi')/(\xi+\alpha^{-1}-\xi')], \tag{23}$$

from which it is seen that $\sigma(\xi+\alpha^{-1})$ is analytic everywhere as a function of ξ except for cuts along those sections of the real interval $[-\alpha^{-1},\infty]$ where $f(\alpha\xi+1)\neq 0$. In the evaluation of the contour integral in Eq. (22), it is thus necessary to take account of these cuts and the double pole at $\xi=0$. The residue at $\xi=0$ does not provide an exponential time-dependence as one would wish, but rather a linear dependence. As has been seen in III [see Eq. (III. 27)], this can be properly accounted for by noticing that the expansion in Eq. (21) amounts to treating τ as well as α as a small quantity, and then simply replacing the term of the form $1+A\tau$ by $\exp(A\tau)$.

To evaluate Eq. (22) the contour of Fig. 3 is used, the integral of Eq. (23) being taken as an analytic function of ξ on the cut-plane of the figure. Then there results

$$\exp(iEt)\langle N|\Psi(\tau)\rangle = 1+\sigma'(\alpha^{-1})-i\tau\sigma(\alpha^{-1})-(2\pi i)^{-1}\int_{\mathcal{C}} d\xi \times[\exp(-i\xi\tau)/\xi^2]\sigma(\xi+\alpha^{-1})$$

which we replace by

$$[1+\sigma'(\alpha^{-1})]\exp[-i\tau\sigma(\alpha^{-1})]-(2\pi i)^{-1}\int_{\mathcal{C}} d\xi \times[\exp(-i\xi\tau)/\xi^2]\sigma(\xi+\alpha^{-1}). \quad (24)$$

(Here, $\sigma'(\alpha^{-1})$ is considered as a small quantity, since it is proportional to α.)

We have seen that for $f(x)=x^p$, $p=1,2,3,\cdots$, we obtain [Eq. (13)]

$$\sigma(\xi)=-(2/\pi)\left(\sum_{n=0}^{p-1}[(\alpha\xi)^n/(p-n)]R^{p-n}+(\alpha\xi)^p\log(R/\alpha\xi-1)+\pi i(\alpha\xi)^p\right) \quad (25)$$

with the usual cutoff, so that Eq. (24) becomes

$$\exp(iEt)\langle N|\Psi(\tau)\rangle=(1+\alpha A_p)\exp(-2\tau)\exp(iB_p\tau) - (2\pi i)^{-1}\int_{\mathcal{C}_1} d\xi\,[\exp(-i\xi\tau)/\xi^2](-2/\pi)\,2\pi i\,(\alpha\xi+1)^p, \quad (26)$$

where

$$A_p=-(2/\pi)\left(\sum_{n=1}^{p-1}[nR^{p-n}/(p-n)]+p\log(R-1)-[R/(R-1)]+\pi ip\right), \quad (26')$$

$$B_p=(2/\pi)\left(\sum_{n=0}^{p-1}[R^{p-n}/(p-n)]+\log(R-1)\right). \quad (26'')$$

In Eq. (26), $\mathcal{C}_1$ is the contour of Fig. 4 along which one has to integrate the discontinuity of $\sigma(\xi+\alpha^{-1})$. The contour $\mathcal{C}_1$ is an obvious deformation, appropriate to this choice of f, of the cut shown in Fig. 3. [For other choices of the coupling function f, exactly similar analyses can be performed.] The explicit expression for $\rho(\tau)$ corresponding to the choice $f(x)=x^p$ is

$$\rho(\tau)=\Big\{(1+\alpha\,\mathrm{Re}\,A_p)\exp(-2\tau)\cos(B_p\tau)+2\alpha p \times\exp(-2\tau)\sin(B_p\tau)+(2/\pi)\int_\alpha^{\alpha/(R-1)} d\xi \times\cos(\tau/\xi)[(\xi+\alpha)/\xi]^p+(2\alpha/\pi)\int_0^\pi d\theta \times\exp[-\tau(R-1)\sin\theta/\alpha] \times\Big[\cos\theta/(R-1)\Big(\cos[\tau(R-1)\cos\theta/\alpha]\sum_{n=0}^{p}\binom{p}{n}(-1)^n(R-1)^n \sin n\theta + \sin[\tau(R-1)\cos\theta/\alpha]\sum_{n=0}^{p}\binom{p}{n}(-1)^n(R-1)^n\cos n\theta\Big) - [\sin\theta/(R-1)]\Big(\cos[\tau(R-1)\cos\theta/\alpha]\sum_{n=0}^{p}\binom{p}{n}(-1)^n(R-1)^n \cos n\theta - \sin[\tau(R-1)\cos\theta/\alpha]\sum_{n=0}^{p}\binom{p}{n}(-1)^n(R-1)^n\sin n\theta\Big)\Big]\Big\}^2 + \Big\{(1+\alpha\,\mathrm{Re}\,A_p)\exp(-2\tau)\sin(B_p\tau)-2\alpha p\exp(-2\tau)\cos(B_p\tau) + (2/\pi)\int_\alpha^{\alpha/(R-1)} d\xi\,\sin(\tau/\xi)[(\xi+\alpha)/\xi]^p-(2\alpha/\pi)\int_0^\pi d\theta \times\exp(-\tau(R-1)\sin\theta/\alpha) \times[\cos\theta/(R-1)]\Big(\cos[\tau(R-1)\cos\theta/\alpha]\sum_{n=0}^{p}\binom{p}{n}(-1)^n(R-1)^n \times\cos n\theta - \sin[\tau(R-1)\cos\theta/\alpha]\sum_{n=0}^{p}\binom{p}{n}(-1)^n(R-1)^n\sin n\theta\Big) + [\sin\theta/(R-1)]\Big(\cos[\tau(R-1)\cos\theta/\alpha]\sum_{n=0}^{p}\binom{p}{n}(-1)^n(R-1)^n \times\sin n\theta + \sin[\tau(R-1)\cos\theta/\alpha]\sum_{n=0}^{p}\binom{p}{n}(-1)^n(R-1)^n\cos n\theta\Big)\Big]\Big\}^2. \quad (27)$$

We may remark here that the weak-coupling approximation used in III and based on Eq. (12)—that which was shown in IV and V to compare well for weak-coupling with the exact solution for a finite system—gives for an infinite system a purely exponential result [see Eq. (III. 23)]. This approximation consisted in writing Eq. (12) as

$$\exp(iEt)\langle N|\Psi(\tau)\rangle=-\frac{1}{2\pi i}\int_c\frac{d\xi\exp(-i\xi\tau)}{\xi-\sigma(\xi+\alpha^{-1})}$$

and then setting $\alpha=0$ in the expression for $\sigma(\xi+\alpha^{-1})$. One can, of course, retain terms to first (or any desired) order in α in this, but since $\sigma(\xi+\alpha^{-1})$ is not an entire function, all nonexponential contributions to $\exp(iEt)\langle N|\Psi(\tau)\rangle$ will still be neglected in this way. However, it is easy to see that the purely exponential contribution can be obtained to any desired accuracy by this method. In fact, retention of the term proportional to α in the expansion of $\sigma(\xi+\alpha^{-1})$ yields not only the coefficient [for example, $(1+\alpha A_p)$ in Eq. (26)] of the exponential to order α, but also the correction to that order in the exponent itself. This latter was lost in our earlier procedure because τ there appeared as a small quantity as well as α.

Finally, to complete this study, we include for comparison the weak-coupling solution to Eq. (4) corresponding to the choice of form factor

$$f(x)=4x/(1+x)^2 \quad (\alpha<\pi/8).$$

Here we find that the time evolution of $\rho(\tau)$ is given by

$$\rho(\tau)=\left[\left(1+\frac{4\alpha}{\pi}\right)\exp(-2\tau)\cos\left(\frac{4\tau}{\pi}\right)+\frac{8\alpha}{\pi}\int_0^{\pi/2}\right.$$

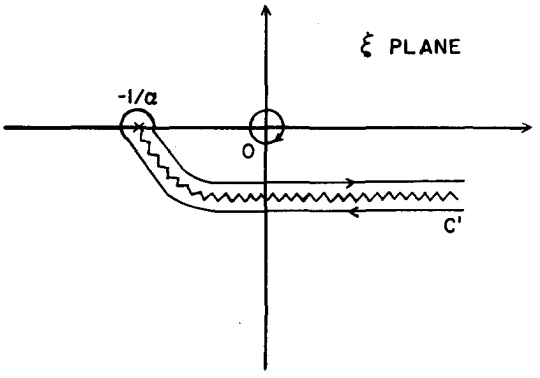

FIG. 3. The contour used in the evaluation of Eq. (22).

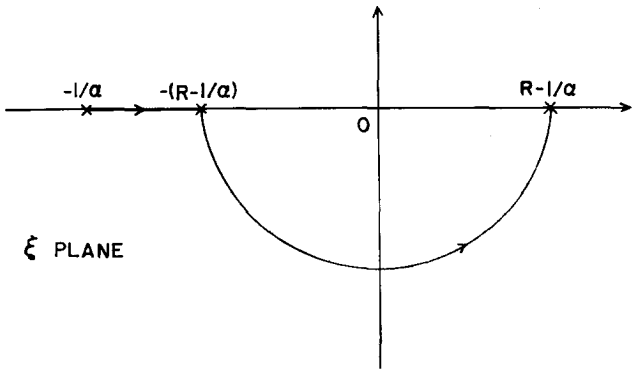

FIG. 4. The contour used in the evaluation of Eq. (26).

$$\frac{d\theta \exp(-\tau\sin\theta/\alpha)}{(5-4\cos\theta)^2}$$

$$\times[\cos(\tau\cos\theta/\alpha)(-8\sin\theta+5\sin2\theta-\sin3\theta)$$

$$-\sin(\tau\cos\theta/\alpha)(4-8\cos\theta+5\cos2\theta-\cos3\theta)]$$

$$-\frac{8\alpha^3}{\pi}\int_0^\alpha \frac{d\xi \exp(-\tau/\xi)\xi}{(4\xi^2+\alpha^2)^2}\Bigg]^2$$

$$+\left[\left(1+\frac{4\alpha}{\pi}\right)\exp(-2\tau)\sin\left(\frac{4\tau}{\pi}\right)+\frac{8\alpha}{\pi}\int_0^{\tau/2}\frac{d\theta \exp(-\tau\sin\theta/\alpha)}{(5-4\cos\theta)^2}\right.$$

$$\times[\cos(\tau\cos\theta/\alpha)(4-8\cos\theta+5\cos2\theta-\cos3\theta)$$

$$+\sin(\tau\cos\theta/\alpha)(-8\sin\theta+5\sin2\theta-\sin3\theta)]$$

$$\left.+\frac{8\alpha^3}{\pi}\int_0^\alpha \frac{d\xi\, c^{-\tau/\xi}\xi^2(4\xi^2+3\alpha^2)}{(4\xi^2+\alpha^2)^2}\right]^2. \tag{28}$$

B. The Prigogine–Resibois master equation

The time evolution of a system with many degrees of freedom can also be described using the Prigogine–Résibois master equation.[6] We consider here a weak-coupling solution to this master equation, subject to the same approximations as introduced in previous sections, and for the general choice of coupling function, Eq. (8). Since the methods for treating the Wigner–Weisskopf problem using the Prigogine–Résibois master equation have been laid down in I, we can proceed at once to Eq. (I. 30), and write

$$\rho_0(N,t)=-(2\pi i)^{-1}\int_c dz \exp(-izt)\, z^{-1}[1+S(z)]^{-1},$$

where

$$S(z)=-(4/\hbar^2)\sum_\lambda\{|h_\lambda|^2/[z^2-(w_\lambda-E)^2]\} \tag{29}$$

and $\rho_0(N,t)$ is the probability at time t of finding the system in the state $|N\rangle$. With the usual choice of dimensionless variables, one obtains the result

$$\rho_0(N,\tau)=-(2\pi i)^{-1}\int_c d\xi \exp(-i\xi\tau)\,\xi^{-1}[1+S(\alpha E\xi)].$$

If Eq. (8) is substituted into Eq. (29) and the definition of $\sigma(\xi)$, Eq. (23), recalled, it is seen that

$$S(\alpha E\xi)=-\xi^{-1}[\sigma(\xi+\alpha^{-1})-\sigma(-\xi+\alpha^{-1})],$$

so that

$$\rho_0(N,\tau)=-(2\pi i)^{-1}\int_c d\xi \times\exp(-i\xi\tau)[\xi-\sigma(\xi+\alpha^{-1})+\sigma(-\xi+\alpha^{-1})]^{-1}. \tag{30}$$

As in previous sections, we shall consider the case $f(x)=x^p$ explicitly in order to demonstrate the evaluation of Eq. (30). Firstly, from Eq. (25), we have

$$\xi-\sigma(\xi+\alpha^{-1})+\sigma(-\xi+\alpha^{-1})$$

$$=\xi+\frac{2}{\pi}\left[\sum_{n=0}^{p-1}\frac{(\alpha\xi+1)^n-(-\alpha\xi+1)^n}{p-n}R^{p-n}\right.$$

$$+(\alpha\xi+1)^p\log\left(\frac{R}{\alpha\xi+1}-1\right)$$

$$\left.-(-\alpha\xi+1)^p\log\left(\frac{R}{-\alpha\xi+1}-1\right)+\pi i(\alpha\xi+1)^p+\pi i(-\alpha\xi+1)^p\right],$$

where the log functions are defined on the cut-plane of Fig. 5. It is seen at once that this expression has a zero on the lower imaginary axis, at a point

$$\xi_0=-4i+\frac{16i\alpha}{\pi}\left(\sum_{n=1}^{p-1}\frac{nR^{p-n}}{p-n}+p\log(R-1)-\frac{R}{R-1}\right)+O(\alpha^2)$$

$$=-4i(1+2\alpha\,\mathrm{Re}\,A_p)+O(\alpha^2)$$

with A_p given by Eq. (26a). The residue of the integrand in Eq. (30) at this point is

$$\exp(-i\xi_0\tau)\left(1-\frac{d}{d\xi}[\sigma(\xi+1/\alpha)-\sigma(-\xi+1/\alpha)]\Big|_{\xi=\xi_0}\right)^{-1}$$

$$=\exp(-i\xi_0\tau)[1+2\alpha\,\mathrm{Re}\,A_p+O(\alpha^2)]^{-1}$$

$$=[1-2\alpha\,\mathrm{Re}\,A_p+O(\alpha^2)]\exp\{-4\tau[1+2\alpha\,\mathrm{Re}\,A_p+O(\alpha^2)]\}.$$

The coefficient here can be seen to correspond to this order in α with that of the square of the modulus of the exponential part of $\exp(iEt)\langle N|\Psi(\tau)\rangle$ as given by Eq. (26). The exponent itself, however, does not agree in the term proportional to α with the "true" value of the exponent as obtained by doubling the imaginary part of the zero of $\xi-\sigma(\xi+\alpha^{-1})$. This is analogous to the result of Eq. (26), where no term proportional to α appears in the exponent. For the rest of $\rho_0(N,\tau)$ as given by Eq. (30), it is necessary to compute the contributions arising from the discontinuities of the integrand along the cuts of Fig. 5. In the following analysis, we have chosen $(1<)\ R\leq 2$, since this is the choice made in the subsequent numerical work on Eq. (30). The case $R>2$ can be treated in the same way. It is readily seen that the contributions to $\rho_0(N,\tau)$ from the cuts are

$$-(2\pi i)^{-1}\int_{-\infty}^{-1/\alpha} d\xi \exp(-i\xi\tau)\{[\xi+G(\xi)]^{-1}[\xi+G(\xi)$$

$$+4i(\alpha\xi+1)^p+4i(-\alpha\xi+1)^p]^{-1}\},$$

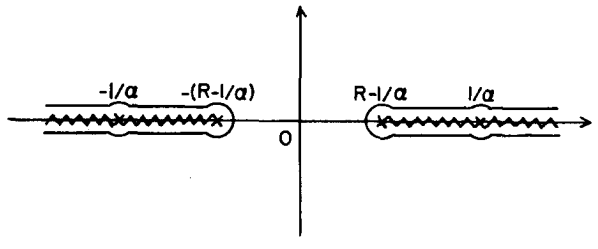

FIG. 5. The contour used in the evaluation of Eq. (30).

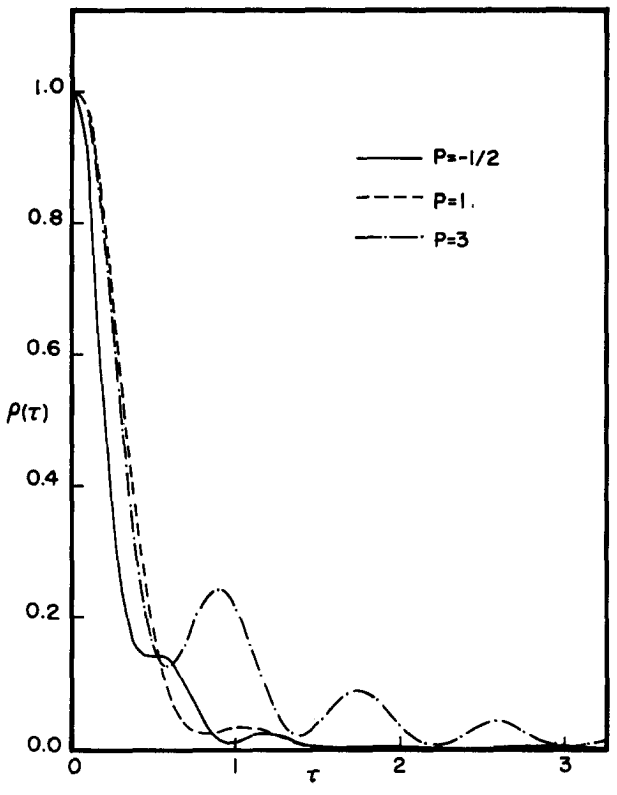

FIG. 6. A plot of $\rho(\tau)$ versus τ determined from the Schrödinger equation. The calculations were performed on Eqs. (16) and (17) with no weak-coupling approximation for the choice of parameters, $\alpha=0.1$ and $R=1.5$.

$$-(2\pi i)^{-1}\int_{-1/\alpha}^{-(R-1)/\alpha} d\xi\,\exp(-i\xi\tau)\{[\xi+G(\xi)+2i(\alpha\xi+1)^p]^{-1}[\xi+G(\xi)+2i(\alpha\xi+1)^p+4i(-\alpha\xi+1)^p]^{-1}\},$$

$$-(2\pi i)^{-1}\int_{(R-1)/\alpha}^{1/\alpha} d\xi\,\exp(-i\xi\tau)\{[\xi+G(\xi)+2i(-\alpha\xi+1)^p]^{-1}[\xi+G(\xi)+4i(\alpha\xi+1)^p+2i(-\alpha\xi+1)^p]^{-1}\},$$

$$-(2\pi i)^{-1}\int_{1/\alpha}^{\infty} d\xi\,\exp(-i\xi\tau)\{[[\xi+G(\xi)]^{-1}[\xi+G(\xi)+4i(\alpha\xi+1)^p+4i(-\alpha\xi+1)^p]^{-1}\}, \tag{31}$$

where $G(\xi)$ is the function, real for real ξ

$$G(\xi)=\frac{2}{\pi}\left(\sum_{n=0}^{p-1}\frac{(\alpha\xi+1)^n-(-\alpha\xi+1)^n}{p-n}R^{p-n}+(\alpha\xi+1)^p\log\left|\frac{R}{\alpha\xi+1}-1\right|-(-\alpha\xi+1)^p\log\left|\frac{R}{-\alpha\xi+1}-1\right|\right). \tag{32}$$

It is seen that $G(\xi)=-G(-\xi)$ for real ξ. If one makes the change of variables $\xi=\pm\xi^{-1}$, then the expression (31) becomes

$$C(\tau)\equiv\frac{1}{\pi}\int_0^{\alpha} d\xi\left(\frac{\sin(\tau/\xi)}{\xi[1+\xi G(\xi^{-1})]}-\frac{[1+\xi G(\xi^{-1})]\sin(\tau/\xi)+4\xi^{1-p}[(\xi-\alpha)^p+(\xi+\alpha)^p]\cos(\tau/\xi)}{\xi\{[1+\xi G(\xi^{-1})]^2+16\xi^{2-2p}[(\xi-\alpha)^p+(\xi+\alpha)^p]^2\}}\right)$$

$$+\frac{1}{\pi}\int_{\alpha}^{\alpha/(R-1)} d\xi\left(\frac{[1+G(\xi^{-1})]\sin(\tau/\xi)+2\xi^{1-p}(\xi-\alpha)^p\cos(\tau/\xi)}{\xi\{[1+\xi G(\xi^{-1})]^2+4\xi^{2-2p}(\xi-\alpha)^{2p}\}}\right.$$

$$\left.-\frac{[1+\xi G(\xi^{-1})]\sin(\tau/\xi)+2\xi^{1-p}[(\xi-\alpha)^p+2(\xi+\alpha)^p]\cos(\tau/\xi)}{\xi\{[1+\xi G(\xi^{-1})]^2+4\xi^{2-2p}[(\xi-\alpha)^p+2(\xi+\alpha)^p]^2\}}\right). \tag{33}$$

The final result is then

$$\rho_0(\mathcal{N},\tau)=(1-2\alpha\,\mathrm{Re}A_p)\exp[-4\tau(1+2\alpha\,\mathrm{Re}A_p)]+C(\tau), \tag{34}$$

for this approximation.

IV. NUMERICAL RESULTS

The expressions derived in the preceding sections for $\rho(\tau)=|\langle\mathcal{N}|\Psi(\tau)\rangle|^2$ or $\rho(\tau)=\rho_0(\mathcal{N},\tau)$ have been computed numerically for various choices of the coupling parameter α and of the coupling function (form factor) f of Eq. (8). The first series of calculations was based on Eq. (12) for the exact solution [with use of the material following Eq. (12) in Sec. II]. Figures 6–9 display graphically the results for $f(x)=x^p$ ($p=1,2$) with cutoffs $R=1.5$ and 2, and for "weak-coupling," $\alpha=0.1$, and "strong-coupling," $\alpha=0.8$. The weak-coupling cases display very little of the nonergodicity which, as has been seen, is due to the occurrence of the poles $\xi_\pm$ of the denominator of the integrand of Eq. (12), and which always must arise with the form factors in question. In fact, all the nonexponential contributions to $\rho(\tau)$ can be seen to be relatively unimportant for $\alpha=0.1$, although the case $p=3$, especially with the larger ($R=2$) cutoff, already shows quite noticeable nonexponential effects.

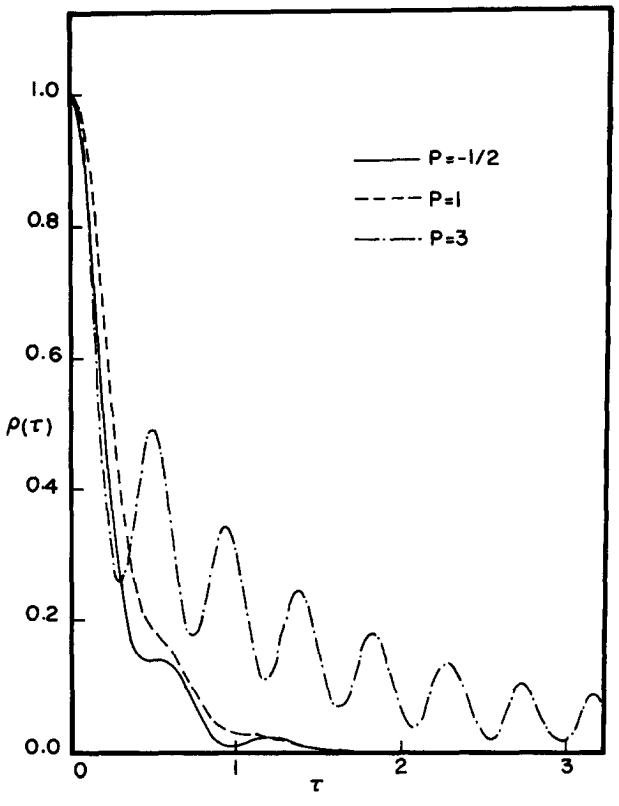

FIG. 7. A plot of $\rho(\tau)$ versus τ determined from the Schrödinger equation. The calculations were performed on Eqs. (16) and (17) with no weak-coupling approximation for the choice of parameters, $\alpha=0.1$ and $R=2.0$.

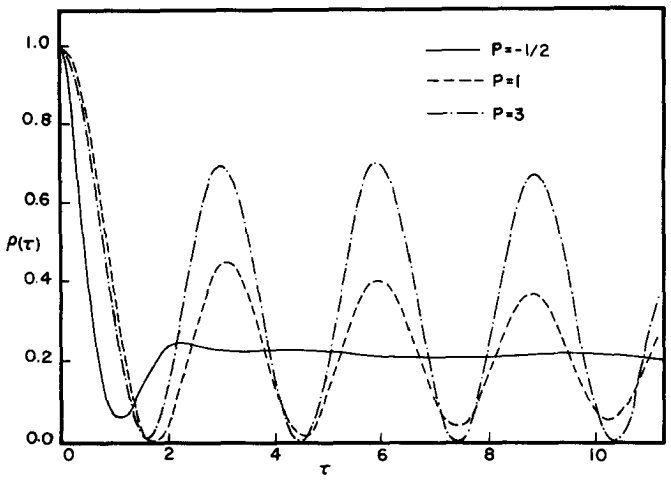

FIG. 8. A plot of $\rho(\tau)$ versus τ determined from the Schrödinger equation. The calculations were performed on Eqs. (16) and (17) with no weak-coupling approximation for the choice of parameters, $\alpha = 0.8$ and $R = 1.5$.

These are much magnified for $\alpha = 0.8$, to the extent that (see Fig. 9) decay has become very slow, with large oscillations the predominant feature. The results obtained in V, for the form factor $p = -1/2$ (no cutoff), are also included in Figs. 6—9 for further comparison. The oscillations in these solutions, although of the same nature as those for $p = 1, 3$, are considerably less marked, especially for the larger choice of α. The case $p = 2$ has not been shown in the figures, since it turns out on calculation to be entirely intermediate to the cases $p = 1$ and $p = 3$. Lastly in this series, $\rho(\tau)$ has been plotted in Fig. 10 for

$$f(x) = 4x/(1+x)^2$$

and $\alpha = 0.1$. This case is actually ergodic, that is $\rho(\tau) \to 0$ as $\tau \to \infty$. Here the decay is not strictly exponential, but all other contributions have become minor, and there is no visible sign of oscillation.

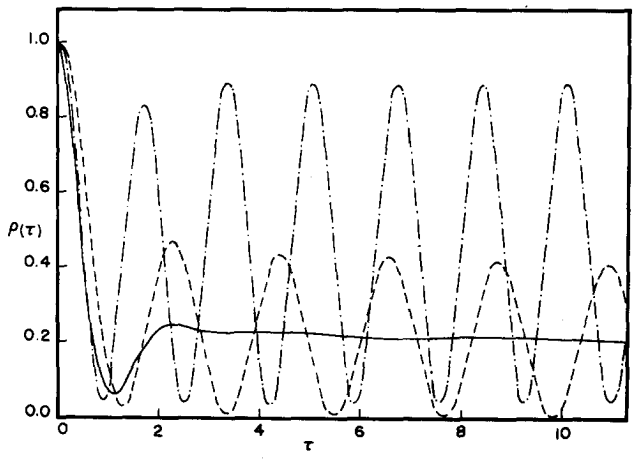

FIG. 9. A plot of $\rho(\tau)$ versus τ determined from the Schrödinger equation. The calculations were performed on Eq. (18) and the corresponding "two-pole" generalization for $p = 3$ with no weak-coupling approximation and for the choice of parameters, $\alpha = 0.8$ and $R = 2.0$.

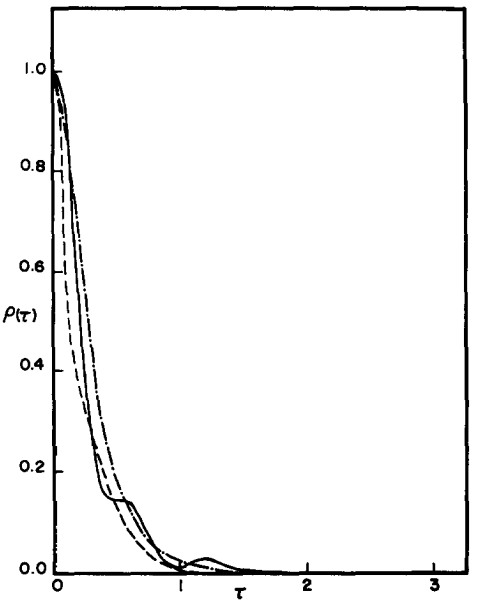

FIG. 10. The Schrödinger exact solution and the associated weak-coupling approximation for the intermediate form factor, $f(x) = 4x/(1+x)^2$ for $\alpha = 0.1$ The curve - - - - gives the time evolution of $\rho(\tau)$ determined using Eq. (19), the curve - · - · - gives the time evolution determined using Eq. (28), and the curve —— gives, for comparison, the exact dynamics corresponding to the choice $p = -1/2$.

The next series of calculations was based on Eq. (24) for the weak-coupling quantum-mechanical solution. The results are given in Figs. 11 and 12, for the same choices of the coupling function as in Figs. 6—9, but only for $\alpha = 0.1$, since $\alpha = 0.8$ is too large for weak coupling to be a sensible approximation. These solutions

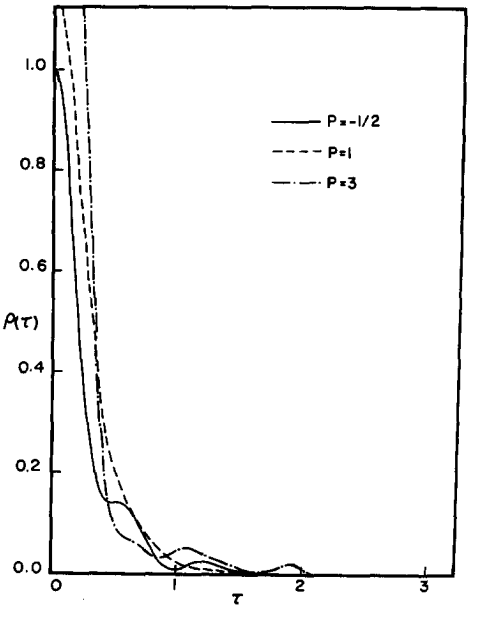

FIG. 11. The Schrödinger weak-coupling approximation for the form factor x^p. The calculations were performed on Eq. (27) for the particular choices $p = 1, 3$ with $\alpha = 0.1$ and $R = 1.5$. We include as well the curve corresponding to the exact solution for $p = -1/2$.

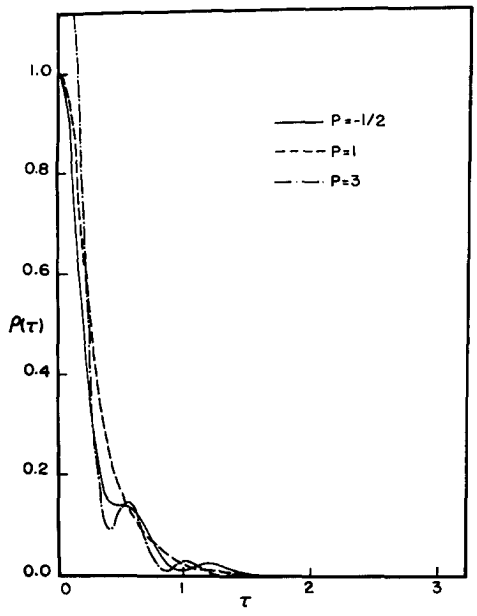

FIG. 12. The Schrödinger weak-coupling approximation for the form factor x^p. The calculations were performed on Eq. (27) for the particular choices $p=1,3$ with $\alpha=0.1$ and $R=2.0$. We include as well the curve corresponding to the exact solution for $p=-1/2$.

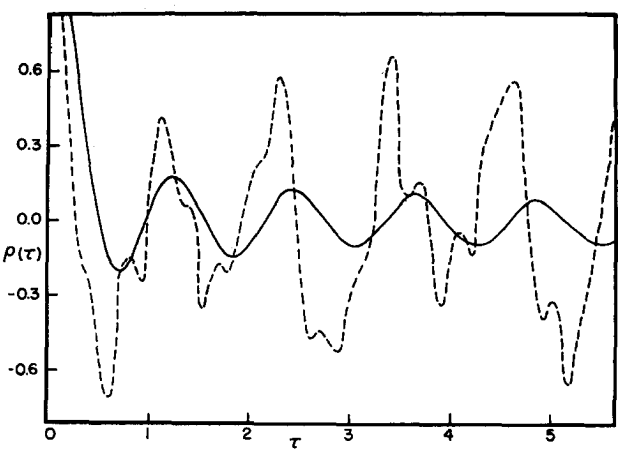

FIG. 13. The Prigogine—Résibois weak-coupling approximation for the form factor x^p. The calculations were performed on Eq. (34) for the particular choices $p=1$ (full line) and $p=3$ (dotted line) with $\alpha=0.1$ and $R=1.5$.

are of course ergodic by construction, and further, as seen clearly in the figures, $\rho(0)$ need not in general be unity. These are not the only discrepancies from the exact results however, for it appears that the oscillations in the tail of the decay are much reduced in this approximation scheme. For the case $p=-1/2$, the results from V are again plotted for comparison; again the oscillations are less important than for $p=1,3$ and are reduced relative to the exact solution. Figure 10 shows the result for

$$f(x)=4x/(1+x)^2$$

and, as in the exact solution, no oscillations can be seen.

The last set of calculations is based on Eq. (30) and the subsequent material of Sec. III giving the solution to the weak-coupling Prigogine—Résibois master equation. The results appear in Figs. 13 and 14, again for $f(x)=x^p$ $p=1,3$, $R=1.5,2$, and $\alpha=0.1$. Here, the solutions are no longer confined between zero and unity, as noted in III. In fact, they bear little other than gross qualitative resemblance to either the exact solutions or the quantum-mechanical weak-coupling ones. It was seen in V that this is true also in the case $p=-1/2$, and so the results for that case have not been reproduced here.

V. DISCUSSION

There have been two main points at issue in this paper. The first was the extent to which the choice of the coupling function f affected the dynamics of spontaneous emission for our model, and the other was the worth of the two weak-coupling schemes discussed with various choices of f. It should be mentioned here that the cutoff parameter R that has been used has been thought of as intrinsic in the definition of f, and in no way as a physical parameter. Although in the corresponding problem in solid-state theory—spontaneous emission of phonons[7]—a physical (Debye) cutoff exists to bound the excitation spectrum, such matters have not concerned us here; what was of interest was to determine how f, as opposed to its value at resonance, affected the model's predictions. The choices of f with cutoff are clearly not continuous functions over the excitation spectrum. This has not impeded the calculations for $\rho(\tau)$, since f appears only in integrands (in the infinite-system limit). In fact, if one f is given, another differing from it only on a set of measure zero will yield the same results; that is, the value of f at an isolated point (even at resonance) does not matter, as decay proceeds only via the interaction with a continuum of states. On the other hand, intervals of the spectrum where f *vanishes* influence the dynamics greatly: it has been seen that this aspect of the cutoff form factors gives rise to isolated modes and the related nonergodicity, and also to branch cuts in the integrands involved in calculating $\rho(\tau)$, with consequent significant nonexponential contributions to

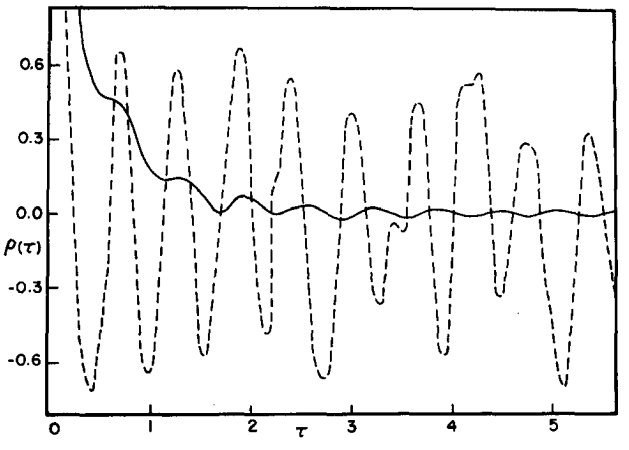

FIG. 14. The Prigogine—Resibois weak-coupling approximation for the form factor x^p. The calculations were performed on Eq. (34) for the particular choices $p=1$ (full line) and $p=3$ (dotted line) with $\alpha=0.1$ and $R=2.0$.

$\rho(\tau)$. If it were physically reasonable (which it is not in quantum mechanics) to consider an excitation spectrum extending from $-\infty$ to $+\infty$ and a form factor f not vanishing over all this range, then not only would $\rho(\tau)$ be ergodic, but it would be possible for it to be purely exponential for certain choices of f.

The numerical calculations have shown that, although of course the exponential part of $\rho(\tau)$ depends, to lowest order in α, only on the value of f at resonance (or rather in the neighborhood of resonance), the nonexponential parts depend critically on the details of f, and may well be so large as to dominate the exponential part completely. But the better behaved f is, in the sense of not vanishing on the excitation spectrum, being continuous or even analytic, and obeying Eq. (9) at $x=0$, the less substantial the nonexponential parts become. In fact, with

$$f(x)=4x/(1+x)^2$$

they could hardly be detected graphically. It seens permissible to draw the conclusion that, if in a physical situation only exponential behavior is observed, great caution must be exercised in describing the system by a model involving a cutoff or any other irregularity in the interaction form factor.

The quantum-mechanical weak-coupling approximation, Eq. (24), appears to be reasonable in a variety of cases. Over the whole range of f's treated in this paper, it has been seen that the exponential part of the approximation is in very good accord with that of the exact solution [in fact, as noted at the end of Sec. IIIA, by taking sufficiently high powers of α in $\sigma(\xi+\alpha^{-1})$ one can have arbitrary accuracy], and that the nonexponential parts behave in much the same way as in the exact solution, but are consistently underestimated. Again clearly those f's yielding small nonexponential parts give the best results.

The solutions obtained from the Prigogine—Résibois master equation are quite different. By virtue of the fact that Eq. (34) is at best a weak-coupling equation, arbitrary accuracy in the exponential part is not available. The nonexponential parts are very badly given by this equation: Not only do they yield values of $\rho(\tau)$ outside $[0,1]$, but the resemblance they bear to the exact-solution results, which had been seen in IV and V to be only qualitative, seems to disappear entirely with the less well-behaved f's considered here.

Another possibility exists for a weak-coupling approximation that may in general be superior to both of those considered in this paper. In III, it was pointed out that for our model, the *exact* solution was given by the following equation [see Eq. (III. 15)]:

$$\frac{\partial}{\partial t}\langle\mathcal{N}|\Psi(\tau)\rangle=-iE\langle\mathcal{N}|\Psi(\tau)\rangle+\int_0^t d\tau\, C(\tau)\langle\mathcal{N}|\Psi(t-\tau)\rangle, \quad (35)$$

where $C(\tau)$ is the inverse Laplace transform of the expression

$$\psi(z)=(i\hbar)^{-1}\sum_{n=1}^{\infty}\langle\mathcal{N}|H_1[(\hbar z-H_0)^{-1}H_1]^n|\mathcal{N}\rangle_{\mathrm{irr}}$$

the suffix "irr" meaning that when the matrix element is expanded using the completeness relation

$$|\mathcal{N}\rangle\langle\mathcal{N}|+\sum_{\lambda}|\lambda\rangle\langle\lambda|=1$$

the term $|\mathcal{N}\rangle\langle\mathcal{N}|$ is to be omitted. Equation (35) was deduced in III from the Schrödinger equation as an exact result, provided that the series defining $\psi(z)$ converges, by a method entirely analogous to that given in I to derive the Prigogine—Résibois generalized master equation. That equation, too, is an exact result if convergent when taken to all orders, and it has just been seen how poor is the lowest-order approximation to it. On the other hand, Eq. (35) gives the exact Eq. (4) as its lowest-order solution with the Hamiltonian Eq. (3), all other orders vanishing (see Sec. II of III). Thus for this one model at least the perturbative scheme defined by Eq. (35) is the best of the possibilities found so far.

The authors would like to suggest that the situation described above may be a rather general state of affairs. Master equations, Markovian or not, provide extremely useful approaches to the description of weak-coupled systems, but it is evident that they do not afford a sensible way to study how approximate solutions arise from the exact dynamics of a system, simply because their lowest orders give the details of an exact solution in a spurious manner. On the other hand, especially since even so crude a result as Eq. (24) yields a recognizable approximation to the time evolution for our model with quite ill-behaved form factors, it seems permissible to conclude that Eq. (35), which is a better perturbative scheme than Eq. (24) for the same Schrödinger equation, will often give a good description of the dynamics associated with a Hamiltonian not permitting an exact solution to be found in closed form. Clearly there are Hamiltonians for which the approximations given by Eq. (35) will be as pathological as those here derived from the master equation but, for some systems at least, by means of Eq. (35) a proper understanding should be available of the detailed dynamics and the relevance of its various aspects to the problems connected with the taking of the thermodynamic limit, the appearance of irreversibility or ergodicity for certain observables and so forth to an extent much greater than that provided by a master equation. More complicated systems involving the interaction of atoms and radiation than the simple Wigner—Weisskopf model of this series of papers are at present being studied by the authors from this point of view in the hope of resolving some of the yet outstanding conceptual difficulties of nonequilibrium statistical mechanics.

ACKNOWLEDGMENTS

The authors have benefited from funds provided by the United States Atomic Energy Commission through contract AT(11-1)-38. One of us (R. D.) would like to thank the National Research Council of Canada for support.

*This work is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of Notre Dame.

[1]R. Davidson and J.J. Kozak, J. Math. Phys. 11, 189 (1970).
[2]R. Davidson and J.J. Kozak, J. Math. Phys. 11, 1420 (1970).
[3]R. Davidson and J.J. Kozak, J. Math. Phys. 12, 903 (1971).
[4]R. Davidson and J.J. Kozak, J. Math. Phys. 14, 414 (1973).
[5]R. Davidson and J.J. Kozak, J. Math. Phys. 14, 423 (1973).
[6]I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience, New York, 1962); P. Resibois, "A Perturbative Approach to Irreversible Statistical Mechanics," in *Physics of Many-Particle Systems* (Gordon and Breach, New York, 1966), Vol. I.
[7]H.D. Kutz, R. Davidson, and J.J. Kozak, Phys. Rev. B 4, 1471 (1971).

On the existence of solutions to S matrix models

J. V. Greenman

Research Institute for Theoretical Physics, University of Helsinki, Siltavuorenpenger 20B Helsinki 17, Finland

*Department of Mathematics, University of Essex, Colchester, Essex, England**

(Received 15 September 1972; final revised manuscript received 19 March 1973)

We analyze the charged scalar Low equation and show that, in this case at least, the breakdown of the standard fixed point theorems occurs because of their inherent weakness rather than through some topological change in the structure of the family of solutions. We also show that unless we reformulate the equation the Schauder principle only handles the basic solution, that solution without CDD poles. We discuss methods, alternative to the N/D method, to handle the other solutions.

1. INTRODUCTION

The question of the feasibility of the S matrix program has been studied in some detail.[1-6] The results, however, have been somewhat disappointing,[3,6] in that existence of solutions of the various proposed models has been proved only for weak coupling. In particular, Warnock[1] and McDaniel and Warnock[2,3] have proved existence of solutions to the πN static Low equation[7] only for coupling constant less than about 0.003 compared to the πN coupling constant value of about 0.2.

It is the purpose of the paper to gain some insight into why the standard fixed point theorems[6,8,9] fail at such low coupling constants. We apply the theorems to a soluble version of the Low equation, the charged scalar model, and compare the results with the exact solutions.[10] We conclude, in this model at least, that the theorems fail not because of any topological change in the structure of the family of solutions but through inherent weakness of the theorems. We also show that the Schauder theorem gives no improvement on the contraction mapping theorem, there being only one fixed point in each case. This fixed point corresponds to the basic solution, that solution without CDD poles.[11] The best bound obtained is about 0.2 whereas the basic solution is well defined up to $\lambda=1$, when a second sheet pole enters the physical sheet. The basic solution then and thereafter ceases to be a solution.

To gain access to solutions other than the basic solution we reformulate the equation to incorporate the singularity structure of solutions with CDD poles thus excluding the basic solution. The contraction mapping principle then generates an iterative solution for any assumed singularity structure compatible with unitarity. This is again only possible for low values of the coupling constant. To move along the trajectories of these solutions in function space with increasing λ one must iterate the contraction mapping iteration[6] or use the more powerful Fréchet derivative method.[6] What our reformulation has accomplished is the vital first step of constructing a solution with CDD pole, gaining the necessary access to the solution trajectory.

We collect the fixed point theorems for convenience in the Appendix.

2. EXISTENCE OF SOLUTIONS WITHOUT CDD POLES

The Low equation has the form

$$f_\alpha(\omega)=P_\alpha(\omega)+\frac{1}{\pi}\int_1^\infty d\omega'\rho(\omega')\left(\frac{|f_\alpha(\omega')|^2}{\omega'-\omega-i\epsilon}+\sum_\beta A_{\alpha\beta}\frac{|f_\beta(\omega')|^2}{\omega'+\omega}\right),$$

$$\alpha=1,\ 2,\ \ldots,\ n, \tag{2.1}$$

where $P_\alpha(\omega)$ is the sum of bound state poles, $\rho(\omega)$ is the product of the centrifugal barrier factor and a cutoff, and $A_{\alpha\beta}$ are the elements of the crossing matrix A. The charged scalar model is defined by the identification $n=2$, $\lambda_\alpha=\lambda$, $P_\alpha(\omega)=\lambda\omega^{-1}$, $A=\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$, $\rho(\omega)=(\omega^2-1)^{1/2}$. The equation expressed in terms of the variable $t=\omega^{-1}$ is then

$$f_+(t)=\lambda+\pi^{-1}\int_{-1}^1 du(1-u^2)^{1/2}\left[|f_+(u)|^2/(t-u-i\epsilon)\right]\equiv N(f_+), \tag{2.2}$$

where $f_+(t)=t^{-1}f_1(\omega)$ and $f_2(\omega)=-f_1(-\omega)$. Equation (2.2) can also be expressed as an equation for the discontinuity function $\sigma=|f_+|^2$:

$$\sigma=f_+f_-, \tag{2.3a}$$

$$f_+=\lambda+\pi^{-1}\int_{-1}^1 du(1-u^2)^{1/2}\left[\sigma(u)/(t-u-i\epsilon)\right], \tag{2.3b}$$

$$f_-=\lambda+\pi^{-1}\int_{-1}^1 du(1-u^2)^{1/2}\left[\sigma(u)/(t-u+i\epsilon)\right]. \tag{2.3c}$$

We first determine what information is provided by the Schauder fixed point theorem if we take the closed convex set K of that theorem to be the set of functions defined by the constraints;

$$|f(t)|\leq A,\quad |t|\leq 1, \tag{2.4a}$$

$$|f(t_1)-f(t_2)|\leq B|t_1-t_2|^\mu,\quad -1\leq t_1,t_2\leq 1,\quad 0<\mu<1. \tag{2.4b}$$

The theorem applies if the operator N of Eq. (2.2) is completely continuous and maps K into itself. The first property has been proved by Warnock[1] using the method of Pogorzelski.[9] A range of values of for which K is mapped into itself is given by the following constraints:

$$|\lambda|+M_1A^2+M_2AB\leq A, \tag{2.5a}$$

$$M_3A^2+M_4AB\leq B, \tag{2.5b}$$

where M_i are known constants. M_1, M_2, M_3 are easily found to have values 1, $4(\pi\mu)^{-1}$, 2.4, respectively. The fourth constant, M_4, is the hardest to calculate since it is a product of the intricate analysis of the Privalov lemma.[12,13] Its value, about 17, is large compared to the other constants of (2.5). An upper bound on $|\lambda|$ for

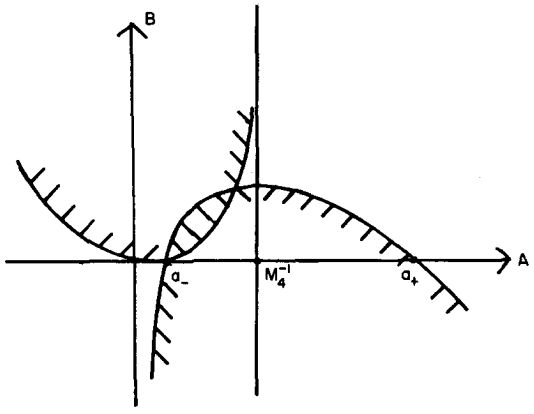

FIG. 1. The region in (A, B) space defined by the constraints (2.5).

which the constraints (2.5) are consistent can be found immediately by noting that the boundary of the region (2.5b) is a hyperbola with asymptote $A = M_4^{-1}$ and the boundary of (2.5a) is a hyperbola with asymptote $A = 0$ that intersects the A axis at the two points (Fig. 1):

$$A = a_\pm = [1 \pm (1 - 4M_1|\lambda|)^{1/2}](2M_1)^{-1}. \tag{2.6}$$

An upper bound is obtained by equating a_- and M_4^{-1}. In fact this is a good approximation to the least upper bound on $|\lambda|$ since the boundary of (2.5b) has a sharp corner in the first quadrant of the A-B plane. The best bound turns out to be $|\lambda| = 0.05$.

We can check whether the theorem is an inherently weak theorem since we know the exact solutions. In fact the general solution of (2.2) is[10]

$$f_*(t) = \{[R(t) - t] - i(1 - t^2)^{1/2}\}^{-1} \tag{2.7}$$

or in terms of the discontinuity function

$$\sigma(t) = \{[R(t) - t]^2 + (1 - t^2)\}^{-1}, \tag{2.8}$$

where

$$R(t) = \lambda^{-1} + \sum_i [a_i/(b_i - t)]. \tag{2.9}$$

$A_i = a_i b_i^{-2}$, $(B_i = b_i^{-1})$ is the residue (position) of the ith CDD pole in the variable ω. For convenience we label the discontinuity function solution with n CDD poles with parameters $\{(A_i, B_i)\}$, $i = 1, \ldots, n$ by $\sigma_n(\{(A_i, B_i)\})$. $\sigma_n(\{(A_i, B_i)\})$ is the discontinuity solution only for certain values of the parameters $\{(A_i, B_i)\}$.

In fact, one can prove the following results:

(i) $\sigma_n(\{(A_i, B_i)\})$ is not a solution of more than one B_i satisfies the constraint $|B_i| < 1$.

(ii) If there is one B_i such that $|B_i| < 1$ then $\sigma_n(\{(A_i, B_i)\})$ ceases to be a solution for sufficiently small λ since a second sheet pole eventually comes onto the physical sheet. $\sigma_n(\{(A_i, B_i)\})$ then ceases to be a solution of the equation.[10]

(iii) For all solutions a second sheet pole comes onto the physical sheet for sufficiently large values of $|\lambda|$ then ceasing to be a solution.

It may appear at first sight that the solutions (2.8) form a continuum in norm, but this is in fact not so. The convergence of the solutions $\sigma_n(\{(A_i, B_i)\})$, $n > 1$, as λ tends to zero is nonuniform. In fact, there is a confluence on the real axis of each of the resonance pole pairs of $\sigma_n(\{(A_i, B_i)\})$ as λ tends to zero. $\sigma_n(\{(A_i, B_i)\})$ is of order λ almost everywhere, the exceptions being in a neighborhood of width of order λ near each CDD pole position. In each of these neighborhoods there is a point, determined by the condition $R(t) = 0$, where $\sigma_n = 1$. Hence, defining

$$A(\sigma_n, \lambda) = \sup_{t \in [-1,1]} |\sigma_n(\lambda, t)|$$

for a discontinuity function σ_n with n CDD poles,

$$\lim_{\lambda \to 0} A(\sigma_n, \lambda) \geq 1 \text{ for } n > 1 \text{ and } \lim_{\lambda \to 0} A(\sigma_0, \lambda) = 0.$$

Thus there is separation in absolute value between the lowest member of the family of solutions and all the others, the ranking being in terms of subfamilies defined by the number of CDD poles. In Fig. 2 we plot the lower boundary L of the curves $A(\sigma_n, \lambda)$ for $n > 1$ and the curve $A(\sigma_0, \lambda)$. In addition, we plot the region R defined by the Schauder analysis. It is clear from this figure that the Schauder analysis gives us information only about the existence of σ_0, the basic solution. It is not surprising therefore that the bound on $|\lambda|$ is little better than the value one obtains by applying the contraction mapping principle to Eq. (2.2). It is also clear that the bound on $|\lambda|$ has nothing to do with any structural change in the family of solutions. The only type of structural change that can occur is when a second sheet pole of a solution moves onto the physical sheet. This occurs to the basic solution at $\lambda = 1$.

There have been two proposals[6] in the literature for improving the rather weak bound on $|\lambda|$ for σ_0. The first is to use the contraction mapping principle to explicitly construct solutions for λ up to the bound for the ball centered at the origin of the function space. One then uses the solution at the maximum bound as the location of the center of the ball to which one applies the contraction mapping principle, in the expectation that one can move along the trajectory $A(\sigma_0, \lambda)$ in these successive movements. The process eventually breaks

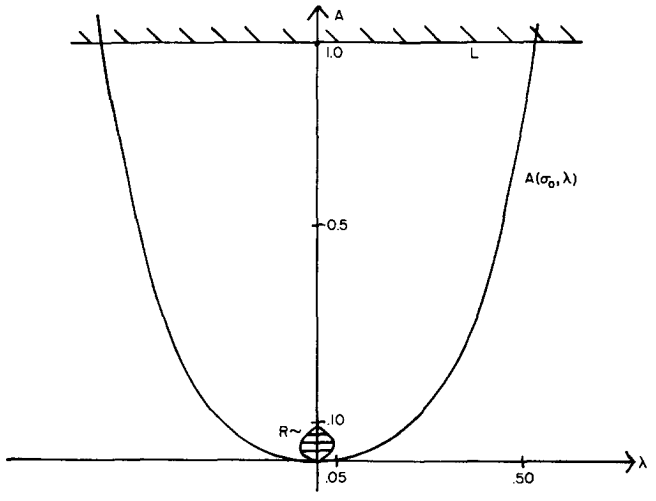

FIG. 2. The region R in (A, λ) space defined by the constraints (2.5).

down because of the weakness of the contraction mapping principle, long before the solution σ_0 has unbounded norm at $\lambda=1$. In fact, the maximum norm obtained by this method is about $\lambda=0.08$. More powerful methods, such as the Fréchet derivative method, would work throughout the trajectory since no bifurcation occurs.[6]

The second proposal is to reformulate the equation. Two possibilities that have been discussed[2,3] are the inverse amplitude equation and the N/D equation. Both give improved bounds but the improvement is not dramatic for the charged scalar equation. The immediate cause for the low bounds on $|\lambda|$ is the fact that the parameter M_4 in expression (2.5) is large. M_4 arises from the Privalov bounding procedure for Hölder-continuous functions. The space of Hölder-continuous functions is larger than we require since we know that the solutions σ are infinitely differentiable. It seems worthwhile, therefore, to examine the possibility of embedding Eq. (2.3) in spaces of analytic functions.[14]

The space of functions continuous on and analytic in the closed disk D_R of radius $R>1$ centered at the origin of the t Argand plane is a complete Banach space S_1 if endowed with the norm

$$\|\sigma\| = \sup_{|z|\leq R} |\sigma(z)|. \tag{2.10}$$

This space is isomorphic to a subspace S_2 of the Banach space of continuous functions defined on the boundary C_R of D_R, endowed with the usual sup norm. This subspace is defined by the constraints

$$\int_0^{2\pi} f(\theta)\exp(in\theta)\,d\theta = 0, \quad n=1,\ 2,\ \ldots\ . \tag{2.11}$$

The isomorphism is the identification of an analytic function with its boundary values. The constraints (2.11) lead to the property

$$\int_{C_R} \frac{dz' f(z')}{z'-z} = 0, \quad z \text{ outside } C_R, \tag{2.12}$$

where without risk of confusion we have written $f(Re^{i\theta}) = f(\theta)$. Relation (2.12) is obtained from the constraints (2.11) on multiplying the nth constraint by $z^{-(n+1)}$ and summing. Relation (2.12) is the necessary and sufficient condition[12] that

$$\lim_{\substack{z\to z_0 \\ z_0\in C_R}} \frac{1}{2\pi i}\int_{C_R} \frac{dz' f(z')}{z'-z} = f(z_0), \quad z \text{ within } C_R. \tag{2.13}$$

The proof is as follows:

$$\int_{C_R} \frac{dz' f(z')}{z'-z} = \int_{C_R} \frac{dz' f(z')}{z'-z} - \int_{C_R} \frac{dz' f(z')}{z'-R^2\bar{z}^{-1}}, \quad z \text{ within } C_R,$$

$$= i\int_0^{2\pi} d\theta f(\theta) P(R,\theta;r,\phi), \tag{2.14}$$

where $z=re^{i\phi}$, $r<R$ and

$$P(R,\theta;r,\phi) = (R^2-r^2)/[R^2+r^2-2Rr\cos(\theta-\phi)]. \tag{2.15}$$

Hence

$$\frac{1}{2\pi i}\int_{C_R} \frac{dz' f(z')}{z'-z} = f(\phi)\frac{1}{2\pi}\int_0^{2\pi} d\theta P(R,\theta;r,\phi) + \frac{1}{2\pi} \times \int_0^{2\pi} d\theta P(R,\theta;r,\phi)[f(\theta)-f(\phi)]. \tag{2.16}$$

The first integral in the limit has value 2π and the second by the usual ϵ, δ analysis has value zero. The operator N with action

$$(N\sigma)(z) = f_+^2(z,\sigma) - 2i\rho(z) f_+(z,\sigma)\frac{1}{2\pi}\int_0^{2\pi} \frac{Re^{i\theta}\sigma(Re^{i\theta})}{Re^{i\theta}-z}\,d\theta, \tag{2.17}$$

where

$$f_+(z,\sigma) = \lambda + \frac{1}{\pi}\int_{-1}^{1} \frac{du\rho(u)}{z-u}\cdot\frac{1}{2\pi}\int_0^{2\pi} \frac{Re^{i\theta}\sigma(Re^{i\theta})}{Re^{i\theta}-z}\,d\theta \tag{2.18}$$

maps S_2 into S_1. This is clear except possibly for the points ± 1. One can show, however, by straightforward distortion of the contour techniques[15] that if one takes the function $N\sigma$ around either end point one ends up with the original value. In the limit $z\to Re^{i\phi}$ we obtain an equation on S_2:

$$\sigma(Re^{i\phi}) = (N\sigma)(Re^{i\phi}) = f_+^2(Re^{i\phi},\sigma) - 2i\rho(Re^{i\phi}) f_+(Re^{i\phi},\sigma)\sigma(Re^{i\phi}). \tag{2.19}$$

Equation (2.19) is equivalent to Eq. (2.3) since if we define

$$\sigma(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{Re^{i\theta}\sigma(Re^{i\theta})}{Re^{i\theta}-z}\,d\theta \tag{2.20}$$

then $(N\sigma)(z)=\sigma(z)$. Evaluating expression (2.17) with z on the real segment one obtains Eq. (2.3).

We cannot use the Schauder principle on the operator N of Eq. (2.19) since N is not completely continuous. We have to use the contraction mapping principle or a slightly stronger variant, Theorem 3 of the Appendix. It is clear that the operator N_1 corresponding to the first term of the right-hand side of Eq. (2.19) is completely continuous. We will find bounds on $|\lambda|$ to ensure that $N_2 \equiv N - N_1$ is a contraction mapping on the ball $B_M \equiv \{\sigma \mid \|\sigma\| \leq M\}$ and that if $\sigma_1, \sigma_2 \in B_M$ then so does $N_1\sigma_1 + N_2\sigma_2$. Theorem 3 then guarantees us at least one fixed point in B_M. The bounding procedure is straightforward using the maximum modulus principle. The bounds are in fact

$$|\lambda|^2 + M^2(a^2+2c) + M(2|\lambda|(a+b)-1) \leq 0 \tag{2.21}$$

and

$$2(|\lambda| + 2M_c) < 1, \tag{2.22}$$

where

$$a = \sup_{z\in C_R} \pi^{-1}\int_{-1}^{1} du\rho(u)|z-u|^{-1}, \quad b = (R^2+1)^{1/2},$$

$$c = \sup_{z\in C_R} [\rho(z)/\pi]\int_{-1}^{1} du\rho(u)|z-u|^{-1}.$$

The region (2.21) is denoted in Fig. 3 by the single shaded region. The parameters N_m, λ_m are given by

$$N_m = (a^2+2c)^{-1}, \quad \lambda_m = \tfrac{1}{2}((a+b)-(a^2+2c)^{1/2})(b^2+2ab-2c)^{-1}. \tag{2.23}$$

The contraction mapping constraint (2.22) denoted by the straight lines in Fig. 3 is not binding. The best result is obtained for $R=1.5$ and the bound on $|\lambda|$ is 0.15. If R gets smaller (larger) a and c (b) gets larger. There is a trade off position about $R=1.5$.

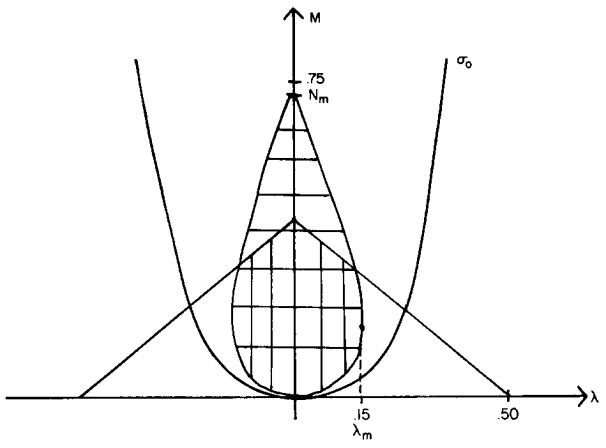

FIG. 3. The region in (M, λ) space defined by the constraints (2.21), (2.22).

One can get improved bounds if one takes different regions of analyticity about the line segment $[-1,1]$. For example if we take an ellipse with foci ± 1 and major and minor axes 1.32, 0.86, respectively, then the bound is improved to 0.18. Thus the analytic space technique certainly improves the $|\lambda|$ bound but not dramatically so. The amount of hard analysis required is however drastically reduced.

Similar techniques can be applied to the Eq. (2.2) for $f_+(t)$. We consider the two functions $f_+(z), f_-(z)$ analytic in the unit disk whose values on the line segment $[-1,1]$ are the functions $f_\pm(t)$. $f_\pm(z)$ are related by the complex conjugation relation $f_+(\bar{z}) = \overline{f_-(z)}$. Because of the analyticity assumption we can distort the integration contour of (2.2) into the upper boundary C_{1+} of the unit disk:

$$f_+(z) = \lambda + \frac{1}{\pi}\int_{C_{1+}} \frac{du\rho(u)}{z-u} f_+(u) f_-(u). \tag{2.24}$$

If we evaluate this expression on the boundary of the unit disc we obtain coupled equations for $\tilde{f}_\pm(\theta)$, where $\tilde{f}_\pm(\theta) \equiv f_\pm(e^{i\theta})$. We embed this equation in the space S_2 with $R=1$. We can use the relation (2.13) to rewrite Eq. (2.24) as

$$f_+(z) = \lambda - \frac{1}{\pi}\int_{C_1} \frac{du\rho(u)}{z-u} f_+(u) f_-(u) + \frac{1}{\pi}\int_{C_{1-}} \frac{du\rho(u)}{z-u} f_+(u) f_-(u)$$

$$= \lambda + 2i\rho(z) f_+(z) f_-(z) + \frac{1}{\pi}\int_{C_{1-}} \frac{du\rho(u)}{z-u} f_+(u) f_-(u), \tag{2.25}$$

where C_{1-} is the lower half of the unit circle C_1 traced in the anticlockwise direction. Hence we can use (2.24) for evaluating $f_+(z)$ on C_{1-} and (2.25) on C_{1+}. With this procedure the integrals involved are not singular other than at the end points.

The nonlinear operator implied by the relations (2.24) and (2.25) maps the ball $B_M = \{f \mid \|f\| \le M\}$ into itself if

$$|\lambda| + (\tilde{a} + 2\tilde{b})M^2 \le M, \tag{2.26}$$

where

$$\tilde{a} = \sup_{z\in C_{1+}} \frac{1}{\pi}\int_{C_{1-}} \frac{du\rho(u)}{|z-u|} = 2^{1/2}, \quad \tilde{b}^2 = \sup_{z\in C_{1+}} |\rho| = 2^{1/2}.$$

The maximum bound on $|\lambda|$ is $\frac{1}{4}(a+2b)^{-1} = 0.06$. The contraction mapping bound is again not binding. This bound is only marginally better than the Hölder analysis bound. It is inferior to the σ analysis principally because the $\tilde{a}$ integral has end point singularities. The region (2.26) is depicted in Fig. 4.

None of the techniques discussed so far give information about solutions other than the basic solution σ_0. In the next section we discuss methods for gaining access to these other solutions.

3. EXISTENCE OF SOLUTIONS WITH CDD POLES

Using spaces of analytic functions it is easy to incorporate simple singularity structure. Suppose, for example, that $f_+(z)$ has one simple pole in the upper half of the unit disk. To be precise suppose $\tilde{f}_+(z) = f_+(z) - [g/(z-a)]$ is analytic in the unit disk, $z=a$ lying in that disk. In distorting the contour of Eq. (2.2) into C_{1+} one picks up the residue from the pole obtaining an equation for $\tilde{f}_+(z)$:

$$\tilde{f}_+(z) = \lambda + \frac{1}{\pi}\int_{C_{1+}} \frac{du\rho(u)}{z-u}\left[\left(\frac{a-\bar{a}}{u-a}\right)\left(\frac{1}{2i\rho(\bar{a})} + \tilde{f}_+(\bar{a})\right) + \tilde{f}_+(u)\right]$$

$$\times\left[\left(\frac{a-\bar{a}}{u-\bar{a}}\right)\left(\frac{1}{2i\rho(a)} - \tilde{f}_-(a)\right) + \tilde{f}_-(u)\right] \tag{3.1}$$

and, for consistency, an equation for g:

$$g = (a-\bar{a})\{[2i\rho(\bar{a})]^{-1} + \tilde{f}_+(\bar{a})\}. \tag{3.2}$$

Evaluating expression (3.1) on the unit circle and at points $z=a$, $\bar{a}$ we obtain coupled equations for $\tilde{f}_+(e^{i\theta})$, $\tilde{f}_+(a)$, $\tilde{f}_+(\bar{a})$ using the fact that $\tilde{f}_+(e^{i\theta}) = \tilde{f}_-(e^{-i\theta})$.

The equivalent equation for the analytic part of the discontinuity $\tilde{\sigma}(z) \equiv \sigma(z) - [G/(z-a)] - [\bar{G}/(z-\bar{a})]$ is

$$\tilde{\sigma} = f_+^2 - 2if_+\rho\tilde{\sigma} - 2iG\left(\frac{\rho(z)f_-(z) - \rho(a)f_-(a)}{z-a}\right)$$

$$- 2i\bar{G}\left(\frac{\rho(z)f_+(z) - \rho(\bar{a})f_+(\bar{a})}{z-\bar{a}}\right), \tag{3.3}$$

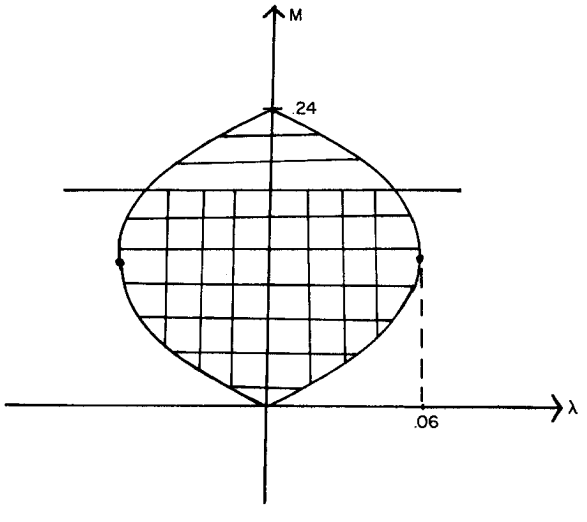

FIG. 4. The region in (M, λ) space defined by the constraint (2.26).

where f_+ is given by (2.3b) and G is given by

$$G=\left[\left(I_2(\bar{a},\bar{a})\frac{1}{2i\rho(a)}-I_2(a,\bar{a})\frac{1}{2i\rho(\bar{a})}\right)+[\bar{\lambda}I_2(\bar{a},\bar{a})-\lambda I_2(a,\bar{a})]\right.$$
$$\left.+[I_1(a,\tilde{\sigma})I_2(\bar{a},\bar{a})-I_1(\bar{a},\tilde{\sigma})I_2(a,\bar{a})]\right]\Delta^{-1}. \quad (3.4)$$

I_1, I_2, Δ are given by

$$I_2(\bar{a},\bar{a})=\frac{1}{\pi}\int_{-1}^{1}\frac{du\rho(u)}{(\bar{a}-u)(u-\bar{a})}, \quad (3.5a)$$

$$I_1(\bar{a},\tilde{\sigma})=\frac{1}{\pi}\int_{-1}^{1}\frac{du\rho(u)\tilde{\sigma}(u)}{\bar{a}-u}, \quad (3.5b)$$

$$\Delta=I_2(\bar{a},a)I_2(a,\bar{a})-I_2(\bar{a},\bar{a})I_2(a,a). \quad (3.5c)$$

One can straightforwardly show that the Eqs. (3.1), (3.3) are equivalent to the original Low equation if we use Hölder continuous functions or functions continuous on and analytic in appropriate disks.

The equation set for σ is somewhat difficult to analyze since the undistorted contour lies close to the poles. We therefore consider the set (3.1), (3.2) analyzing it with respect to the space $S_2\times\mathbb{C}$, the Cartesian product of the special subspace of functions continuous on the boundary of the unit disk and the space of complex numbers. In order to simplify the analysis we will take the pole to move with λ. In fact, we will take $a=0.25+0.5i\lambda$. The reason is that as λ tends to zero the complex conjugate pole pairs of the discontinuity function coalesce on the real axis. Hence with our choice of trajectory we will be sure of a solution at $\lambda=0$, namely the trivial solution $\tilde{f}_\pm(z)=0$. To facilitate evaluation of $\tilde{f}_+(z)$ on C_{1_-} we use the device used to obtain (2.25) from (2.24). In fact,

$$\tilde{f}_+(z)=\frac{1}{\pi}\int_{C_{1_-}}\frac{du\rho(u)}{z-u}\left(\frac{g}{u-a}+\tilde{f}_+(u)\right)\left(\frac{\bar{g}}{u-\bar{a}}+\tilde{f}_-(u)\right)+\lambda-E(z), \quad (3.6)$$

where

$$E(z)=2ig\bar{g}\left(\frac{-\rho(z)}{(z-a)(z-\bar{a})}+\frac{\rho(a)}{(a-\bar{a})(z-a)}+\frac{\rho(z)}{(a-\bar{a})(z-\bar{a})}\right)$$
$$-2i\rho(z)\tilde{f}_+(z)\tilde{f}_-(z)+2ig\left(\frac{\rho(a)\tilde{f}_-(a)-\rho(z)\tilde{f}_-(z)}{z-a}\right)$$
$$+2i\bar{g}\left(\frac{\rho(\bar{a})\tilde{f}_+(\bar{a})-\rho(z)\tilde{f}_+(z)}{z-\bar{a}}\right). \quad (3.7)$$

The ball

$$B_{MN}=\{(\tilde{f}_+(e^{i\theta}),\tilde{f}_+(\bar{a}))\,\big|\,\ |\tilde{f}_+(e^{i\theta})|\le M,\ \tilde{f}_+(\bar{a})\le N\} \quad (3.8)$$

is mapped itself if

$$|\lambda|+|\lambda|^2(N+0.52)(5.84M+5.16N+4.6N|\lambda|)$$
$$+|\lambda|(N+0.52)(9.61M+3.35M|\lambda|+5.16N+4.6N|\lambda|)$$
$$+|\lambda|^2(N+0.52)^2\,8.60+4.24M^2\le M \quad (3.9a)$$

and

$$|\lambda|+(1.51-0.33|\lambda|^2)M^2+|\lambda|^2(N+0.52)^2\,2.68$$
$$+M(N+0.52)|\lambda|(3.56+3.16|\lambda|)\le N. \quad (3.9b)$$

These regions are depicted in Fig. 5 for $\lambda=0.03$. There is a region, the double shaded region, compatible with constraints (3.9) and the straight line contraction mapping constraint. The best bound is $\lambda=0.032$. There is little difference in results between this and the Hölder space analysis of Eq. (3.1). This is probably because in the analytic space reformulation the equations become much more complicated [(3.6), (3.7)], and our very crude bounding methods are at a disadvantage.

ACKNOWLEDGMENTS

I would like to thank Dr. K. Kajantie for hospitality at the Theoretical Physics Institute in Helsinki. I would also like to thank the referee for his helpful comments on the first version of this paper.

APPENDIX

Theorem 1 (*The Contraction Mapping Principle*): If N maps a complete metric space K into itself and satisfies the condition

$$\delta(Nf_1,Nf_2)\le\beta(f_1,f_2),\quad \beta<1,$$

for all $f_1,f_2\in K$ and with β independent of f_1,f_2, then the equation $f=Nf$ has a unique solution in K that can be obtained by iteration from any point in K.

Theorem 2 (*The Schauder Principle*): Let K be a convex closed subset of a Banach space. If N is a continuous mapping of K into itself such that $N(K)$ is compact then N has at least one fixed point in K.

Theorem 3: Let N_1,N_2 be operators defined on a bounded closed convex set K of a Banach space S and suppose

(i) $N_1f+N_2g\in K$ for every pair $f,\ g\in K$,

(ii) N_2 is a contraction mapping on K,

(iii) N_1 is completely continuous on K,

then there exists at least one fixed point $h\in K$ satisfying

$$h=N_1h+N_2h.$$

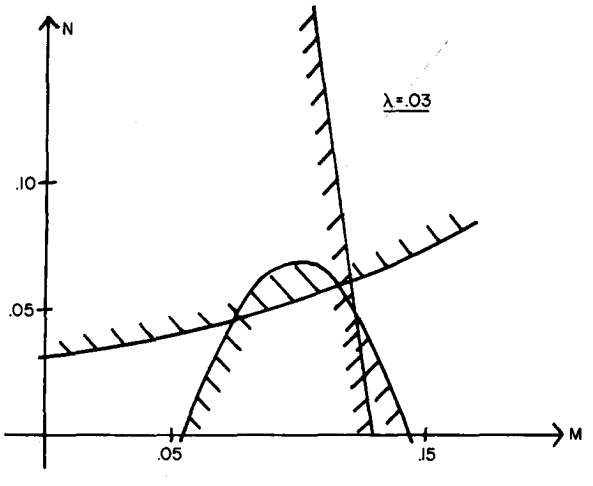

FIG. 5. The region in (N,M) space defined by the constraints (3.9) for $\lambda=0.03$.

*Permanent address.
[1]R. L. Warnock, Phys. Rev. 170, 1323 (1968).
[2]H. McDaniel and R. L. Warnock, Phys. Rev. 180, 1433 (1969).
[3]H. McDaniel and R. L. Warnock, Nuovo Cimento A 64, 905 (1969).
[4]D. Atkinson, Nucl. Phys. 7, 375 (1968); D. Atkinson, Nucl. Phys. 8, 377 (1968); D. Atkinson, Nucl. Phys. 13, 415 (1969).
[5]J. Kupsch, Nucl. Phys. 11, 573 (1969); J. Kupsch. Nucl. Phys. 12, 155 (1969); J. Kupsch, Nuovo Cimento A 66, 202 (1970).
[6]R. L. Warnock, *Lectures in Theoretical Physics*, edited by K. T. Mahanthappa *et al.* (Gordon and Breach, New York, 1969), Vol. XI-D.
[7]F. Low, Phys. Rev. 97, 1392 (1955); G. F. Chew and F. Low, Phys. Rev. 101, 1570 (1956); S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row Petersen, Evanston, Ill. 1961).
[8]M. A. Krasnosel'skii, *Topological Methods in the Theory of Nonlinear Integral Equations* (Pergamon, Oxford, 1964); L. V. Kantorivich and G. P. Akilov, *Functional Analysis in Normed Spaces* (Pergamon, Oxford, 1964).
[9]W. Pogorzelski, *Integral Equations and their Applications* (Pergamon, Oxford, 1966), Vol. I.
[10]K. Huang and F. E. Low, J. Math. Phys. 4, 795 (1965).
[11]L. Castillejo, R. H. Dalitz, and F. J. Dyson, Phys. Rev. 101, 453 (1956).
[12]N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, Holland, 1953).
[13]F. D. Gakhov, *Boundary Value Problems* (Pergamon, Oxford, 1966).
[14]K. Hoffmann, *Banach Spaces of Analytic Functions* (Prentice-Hall, Englewood Cliffs, N.J., 1962).
[15]R. J. Eden *et al.*, *The Analytic S-Matrix* (Cambridge U.P., Cambridge, England, 1965).

Energy transfer in radiation chemistry. I. Dynamics of the electron–oscillators system

Henry Douglas Kutz* and John J. Kozak

Department of Chemistry and Radiation Laboratory, University of Notre Dame, Notre Dame, Indiana 46556
(Received 24 October 1972)

Single-mode excitation events in radiation chemistry are investigated by considering a simple model: an electron (characterized by mass m and momentum $\hbar\mathbf{k}$) in interaction with a collection of harmonic oscillators. The Hamiltonian corresponding to this model is identified, and then used within the framework of a master equation constructed from the time-dependent Schrödinger equation. The usual long-wavelength approximation is employed, which, in the infinite-system limit, necessitates an upper bound on the frequency spectrum of the oscillator bath. Once these constraints have been imposed, however, the time evolution of the system is determined exactly (i.e., no "weak-coupling" approximation) in two limiting regimes—the case of a low-energy incident electron and that of a high-energy incident electron. Our results are compared with those obtained using a weak-coupling approximation, first within the framework of the master equation cited above and then in conjunction with the Prigogine–Résibois master equation, and it is pointed out that for the system under study, the use of a weak-coupling approximation in both equations leads to unphysical results. Our results are also compared with those obtained earlier by Van Hove and co-workers on the electron–random scatterers model, and an attempt is made to understand the role of the approximations introduced in both models. Finally, suggestions for future work are given; in particular, we discuss the generalization of our approach to deal with multiple-mode excitation events in radiation chemistry.

1. INTRODUCTION

A problem of contemporary interest in radiation chemistry is the study of the deposition of energy in condensed media from energetic charged particles, and the possible localization of this energy to produce a chemical effect.[1] Burton[2] has noted that these two events are not necessarily identical and that the chemical event may not even occur in the same vicinity in which the energy was deposited initially. Several authors[3] have addressed themselves to the problem of identifying the separate effects associated with these two events, focusing attention on such aspects as energy migration, tracks and spurs, time scales, mechanisms, and so on. Oftentimes, in the study of these phenomena, a particular event is idealized by posing a simple model, and then, subsequently, the mathematical consequences of adopting the model are interpreted physically. The present contribution is similar in spirit to this type of study in that, recognizing the complexity of the general problem of energy transfer in radiation chemistry, we consider here only those dynamical events associated with the initial deposition of energy from the charged particle to the condensed medium (as opposed to those effects associated with later chemical events) and, secondly, in that we introduce a model which, hopefully, preserves the essential features of the underlying physical situation. In fact, the present study represents a departure from earlier work not so much in the conceptualization of the problem as in the particular approach used to investigate the consequences of adopting a model Hamiltonian.

We begin by recognizing that the dynamics governed by the Hamiltonian for a quantum-statistical system possessing many degrees of freedom can often be usefully described within the framework of a generalized master equation. Accordingly, in our approach we consider a model consisting of an electron in interaction with a collection of harmonic oscillators, and then for the associated Hamiltonian we determine the time evolution of the system using a master-equation approach. We restrict the class of problems discussed in this paper by considering only single-mode excitation events; multiple-mode excitation events are discussed only qualitatively, and a mathematical study of this more complicated situation will be deferred to a later paper. It is worth noting that the study of the electron-oscillators problem via a master-equation approach goes right back to the classic work of Van Hove and coworkers,[4–7] and although these authors obtained numerical solutions only to a certain "truncated" version of the electron-oscillators problem (i.e., the electron—random scatterers problem), until quite recently this was one of the few models for which a full study, both analytical *and* numerical, had been carried through starting from a generalized master equation. Indeed, one of the auxiliary objectives of the present study is to go beyond the results presented in Refs. 4—7, so that the time-dependent behavior of the electron-oscillators system can be characterized, both qualitatively and quantitatively, and the results compared with similar studies on other simple models—the Wigner—Weisskopf atom in interaction with a radiation field (Refs. 8—12, hereafter referred to as I—V respectively), and an effective spin in interaction with the phonon modes of a lattice (Ref. 13, hereafter referred to as VI). A very practical objective of this kind of comparative study is that it might be possible to determine the consequences, both analytical and numerical, of adopting specific approximations. In turn, it may be possible to ascertain for a given class of Hamiltonians which characteristics of the solution arise solely from features inherent in the representation itself and which are due to approximations introduced in order to make the problem solvable. This kind of question has received comparatively little attention, and it is hoped that the present study will cast some light on this very difficult problem.

The mathematical methods we use in this paper are drawn almost entirely from earlier work by the authors

in collaboration with Davidson.[8–13] After the Hamiltonian has been introduced in Sec. 2, a quantum-mechanical master equation is constructed, following the procedure in III, and then a formal solution of the electron-oscillators problem is obtained; this solution takes the form of a contour integral, and it is this integral which forms the basis of our study in the remaining sections of the paper. It should be emphasized at the very outset that our formal results are obtained assuming (1) a particular wave vector dependence describing the electron-oscillators interaction term and (2) an upper bound on the frequency spectrum of the lattice, this in order to avoid the ultraviolet divergence at a certain stage in the calculation. Explicit analytical results are then reported in Sec. 4 for two limiting situations, the case of a low-energy incident electron and the case of a high-energy incident electron. Our numerical results are presented in Sec. 5 along with a discussion and our conclusions.

We remark that the final section attempts to deal with the apparent similarities and differences between our numerical results and those obtained in the study of the electron—random scatterers model; as pointed out in Ref. 5, the latter may be viewed as a contraction of the electron-oscillators system in that the energy of the oscillators is supressed. It is noted that the evolution of both systems is nonexponential in the time, and the implications of this observation as regards an earlier remark of Zwanzig[14] are brought out. That certain of the solutions exhibit a nonzero asymptotic limit is also noted, and correlations with earlier work on the Wigner—Weisskopf atom are indicated. Finally, on the basis of two calculations presented in Appendices A and B (the first dealing with the weak-coupling solution to the Schrödinger master equation and the second with a weak-coupling solution to the Prigogine—Résibois master equation), some concluding comments on the inapplicability of a weak-coupling approximation in treating the electron-oscillators problem are offered.

2. THE HAMILTONIAN

The model considered in this paper consists of an electron characterized by mass m and momentum $\hbar\mathbf{k}$ in interaction with a collection of harmonic oscillators. Using the language of second quantization,[15] we write the Hamiltonian for the free electron as

$$H_{\mathrm{elec}}=\sum_{k}\epsilon(k)\,\alpha_k^*\,\alpha_k, \tag{1}$$

where

$$\epsilon(k)=\hbar^2|\mathbf{k}|^2/2m. \tag{2}$$

The electron creation and annihilation operators satisfy the fermion commutation relation

$$\alpha_k\alpha_k^*+\alpha_k^*\alpha_k=1. \tag{3}$$

The Hamiltonian for the bath of oscillators is given by[15]

$$H_{\mathrm{bath}}=\sum_{q_\lambda}\hbar\omega(q_\lambda)(a_{q_\lambda}^*\,a_{q_\lambda}+\tfrac{1}{2}), \tag{4}$$

where q_λ labels the vibrational modes; the creation and annihilation operators here satisfy the boson commutation relation

$$a_{q_\lambda}a_{q_\lambda}^*-a_{q_\lambda}^*\,a_{q_\lambda}=1. \tag{5}$$

Finally, the interaction between the electron and the oscillator bath is written as[6,16]

$$V_{\mathrm{int.}}=i(8\pi^3/\Omega)^{1/2}\gamma\sum_{k,q_\lambda}\sqrt{\omega(q_\lambda)}\,(a_{q_\lambda}^*\,\alpha_{k-q_\lambda}^*\,\alpha_k-a_{q_\lambda}\,\alpha_k^*\alpha_{k-q_\lambda}), \tag{6}$$

where Ω is the volume of the system and γ is a coupling constant. Since it is desirable to deal with a coupling constant which is dimensionless, we shall eventually replace γ by a dimensionless coupling constant, F; in particular, following Fröhlich,[17] we introduce a characteristic density n and energy ζ such that

$$F=6\pi^2n\gamma^2/\zeta\hbar. \tag{7}$$

Although n and ζ will not be specified explicitly until later, we note that in the Fröhlich theory, where one considers a conduction electron in interaction with the vibrational modes of a lattice, n represents the number of ions per unit volume and ζ denotes the Fermi energy.

With the above remarks in mind, the full Hamiltonian corresponding to our model is given by

$$H=\sum_{k}\epsilon(k)\,\alpha_k^*\,\alpha_k+\sum_{q_\lambda}\hbar\omega(q_\lambda)(a_{q_\lambda}^*\,a_{q_\lambda}+\tfrac{1}{2}) + i(8\pi^3/\Omega)^{1/2}\gamma\sum_{k,q_\lambda}\sqrt{\omega(q_\lambda)}\,(a_{q_\lambda}^*\,\alpha_{k-q_\lambda}^*\,\alpha_k-a_{q_\lambda}\,\alpha_k^*\,\alpha_{k-q_\lambda}). \tag{8}$$

3. THE FORMAL SOLUTION

The time evolution of the electron-oscillators system, governed by the Hamiltonian of Eq. (8), will now be studied within the framework of the time-dependent Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\,|\Psi(t)\rangle=H|\Psi(t)\rangle, \tag{9}$$

using an approach suggested in III. We begin by noting that Eq. (9) has the formal solution

$$|\Psi(t)\rangle=e^{-iHt/\hbar}\,|\Psi(0)\rangle.$$

The state vector for the electron-oscillators system is written as $|\mathbf{k},\mathbf{q}_\lambda\rangle$, where, in this ket, $\mathbf{k}$ signifies an electron with momentum $\hbar\mathbf{k}$ and $\mathbf{q}_\lambda$ signifies an excitation in the q_λth vibrational mode of the lattice. We choose the initial state of the system to be

$$|N\rangle=|\mathbf{k},0\rangle, \tag{10}$$

for which the electron has momentum $\hbar\mathbf{k}$ with $\mathbf{k}$ fixed and all vibrational modes of the lattice are de-excited; i.e., the oscillator bath is in the ground vibrational state. Given this initial state and the Hamiltonian of Eq. (8), the interaction of the electron with the oscillator field may result in the new state

$$|\lambda\rangle=|\mathbf{k}-\mathbf{q}_\lambda,\mathbf{q}_\lambda\rangle,$$

wherein one mode of the oscillator field has been excited. In what follows, we choose the zero point of energy as that state for which the kinetic energy of the electron is zero and all vibrational modes of the oscillator field are de-excited, a state which is inaccessible given the initial condition (10) and the Hamiltonian of Eq. (8).

The Hilbert space is spanned by the states $|N\rangle$ and

$|\lambda\rangle$. The corresponding completeness relation is

$$|N\rangle\langle N| + \sum_{\lambda} |\lambda\rangle\langle\lambda| = 1. \tag{11}$$

The properties of the electron and oscillator, creation and annihilation operators are defined by the following equations:

$$a^*_{q_\lambda}|\mathbf{k}, 0\rangle = (8\pi^3/\Omega)^{1/2}|\mathbf{k}, \mathbf{q}_\lambda\rangle,$$
$$a_{q_\lambda}|\mathbf{k}, \mathbf{q}_\lambda\rangle = (8\pi^3/\Omega)^{1/2}|\mathbf{k}, 0\rangle,$$
$$\alpha^*_k|0, \mathbf{q}_\lambda\rangle = (8\pi^3/\Omega)^{1/2}|\mathbf{k}, \mathbf{q}_\lambda\rangle,$$
$$\alpha_k|\mathbf{k}, \mathbf{q}_\lambda\rangle = (8\pi^3/\Omega)^{1/2}|0, \mathbf{q}_\lambda\rangle.$$

When the Hamiltonian acts on the state $|N\rangle$, the result is

$$H|N\rangle = \epsilon(k)|N\rangle + i\gamma(8\pi^3/\Omega)\sum_{q'_\lambda}\sqrt{\omega(q'_\lambda)}\,|\mathbf{k}-\mathbf{q}'_\lambda, \mathbf{q}'_\lambda\rangle. \tag{12}$$

If we consider only single-mode excitation events, then the action of the Hamiltonian on the state $|\lambda\rangle$ yields the result:

$$H|\lambda\rangle = -i\gamma\sqrt{\omega(q_\lambda)}\,|N\rangle + [\epsilon(k-q_\lambda) + \hbar\omega(q_\lambda)]\,|\lambda\rangle. \tag{13}$$

The state vectors obey the following normalization condition:

$$\langle\lambda|\lambda'\rangle = \delta(\mathbf{q}_\lambda - \mathbf{q}'_\lambda), \tag{14a}$$
$$\langle\lambda|N\rangle = \langle N|\lambda\rangle = 0, \tag{14b}$$
$$\langle N|N\rangle = 1. \tag{14c}$$

Having specified the various operators and their properties, we now express the Schrödinger equation in terms of the resolvent operator $(z - H/\hbar)^{-1}$:

$$|\Psi(t)\rangle = -\frac{1}{2\pi i}\int dz\, e^{-izt}(z - H/\hbar)^{-1}\,|\Psi(0)\rangle, \tag{15}$$

where C is a contour parallel to and slightly above the real axis. Making the identification

$$|\Psi(0)\rangle = |N\rangle,$$

we then introduce the definition

$$|S(z, N)\rangle = (z - H/\hbar)^{-1}\,|N\rangle, \tag{16}$$

for which there exists the expansion

$$|S(z, N)\rangle = S_N|N\rangle + \sum_{\lambda} S_\lambda|\lambda\rangle. \tag{17}$$

Inverting the first of these two equations and combining it with the second yields

$$S_N(z - H/\hbar)|N\rangle + \sum_{\lambda} S_\lambda(z - H/\hbar)|\lambda\rangle = |N\rangle. \tag{18}$$

Taking the inner product of this expression with $\langle N|$ and also with $\langle\lambda|$ allows one to solve explicitly for S_N, thus obtaining

$$S_N = \left[z - \frac{\epsilon(k)}{\hbar} - \frac{\gamma^2}{\hbar^2}\left(\frac{8\pi^3}{\Omega}\right)\sum_{\lambda}\frac{\omega(q_\lambda)}{z - (1/\hbar)[\epsilon(k-q_\lambda) + \hbar\omega(q_\lambda)]}\right]^{-1}. \tag{19}$$

With this result, one can then write down the formally exact expression for the wavefunction of the system corresponding to the state $|N\rangle$; it is

$$\langle N|\Psi(t)\rangle = -\frac{1}{2\pi i}\int_C dz\, e^{-izt} \times\left[z - \frac{\epsilon(k)}{\hbar} - \frac{\gamma^2}{\hbar^2}\left(\frac{8\pi^3}{\Omega}\right)\times\sum_{\lambda}\frac{\omega(q_\lambda)}{z - (1/\hbar)[\epsilon(k-q_\lambda) + \hbar\omega(q_\lambda)]}\right]^{-1}. \tag{20}$$

The associated probability that the system be found in the state $|N\rangle$ at time t is then given by

$$\rho(t) = |\langle N|\Psi(t)\rangle|^2,$$

say, in conformity with the notation of the density-matrix formalism.

In the limit of an infinite system, summations over modes λ can be replaced by integrals over the wave vector $\mathbf{q}$, that is,

$$(2\pi)^3/\Omega\sum_{\lambda} \to \int d\mathbf{q}_\lambda.$$

In this limit, then, Eq. (20) reads

$$\langle N|\Psi(t)\rangle = -\frac{1}{2\pi i}\int_C dz\, e^{-izt} \times\left(z - \frac{\epsilon(k)}{\hbar} - \frac{\gamma^2}{\hbar^2}\int\frac{\omega(q_\lambda)dq_\lambda}{z - (1/\hbar)[\epsilon(k-q_\lambda) + \hbar\omega(q_\lambda)]}\right)^{-1}. \tag{21}$$

Before one can proceed with the solution of this contour integral, it is necessary to specify first the wave vector dependence of $\omega(q_\lambda)$. Here we shall use a well-known approximation,[16] namely, that in the long-wavelength limit, this dependence is given by

$$\omega(q_\lambda) = v|q_\lambda|, \tag{22}$$

where v is the velocity of sound. Given the structure of the integral appearing in the denominator of Eq. (21), it is seen that with this specification of q_λ – dependence, an upper bound μ must be established on the wave vector spectrum so as to avoid the well-known ultraviolet divergence. With these two assumptions, then, Eq. (21) reads

$$\langle N|\Psi(t)\rangle = -\frac{1}{2\pi i}\int_C dz\, e^{-izt} \times\left(z - \frac{\epsilon(k)}{\hbar} - \frac{4\pi\gamma^2 v}{\hbar^2}\int_0^{\mu}\frac{q_\lambda^3\, dq_\lambda}{z - (1/\hbar)\epsilon(k-q_\lambda) - vq_\lambda}\right)^{-1}, \tag{23}$$

where we have used

$$\int d\mathbf{q}_\lambda = \int 4\pi q_\lambda^2\, dq_\lambda$$

and imposed the bound μ on the spectrum.

We note that the formal solution of Eq. (23) can be carried one step further by dealing explicitly with the μ integral appearing in the denominator of the contour integral. The result is

$$\langle N|\Psi(t)\rangle = -\frac{1}{2\pi i}\int_C dz\, e^{-izt} \times\left(z - \frac{\epsilon(k)}{\hbar} + \frac{4\pi\gamma^2 v}{\hbar^2}I(z)\right)^{-1}, \tag{24}$$

where

$$I(z) = m\hbar\mu + 2m\hbar\mu(2k - 2m\hbar v) + \frac{m\hbar\mu}{\frac{1}{2}(q_+ - q_-)}\left[q_+^3\ln\left(\frac{\mu - q_+}{-q_+}\right) - q_-^3\ln\left(\frac{\mu - q_-}{q_-}\right)\right] \tag{25a}$$

and

$$q_\pm = k - m\hbar v \pm [m\hbar v(m\hbar v - k + 2z/v)]^{1/2}. \quad (25b)$$

Although the complicated structure of the integrand in the contour integral, Eq. (24), leads to something of an impasse in the analysis at this point, one can determine the explicit solution of Eq. (23) in two different regimes of energy transfer, and this task will be taken up in the following section.

4. REGIMES OF ENERGY TRANSFER

A. Low-energy incident electron

Here we assume that the range of kinetic energies of the incident electron (fixed initially at some constant value) is bounded from above by the upper bound on the energy spectrum of the lattice; that is, we assume that $\epsilon(k) \leq \hbar v\mu$. Upon energy transfer, it is then assumed that the final kinetic energy of the electron is negligible relative to the consequent single-mode excitation of the initially de-excited lattice. In particular, we require that $\epsilon(k - q_\lambda) \ll \hbar\omega(q_\lambda)$, a simplification which permits us to replace Eq. (13) by the slightly different expression

$$H|\lambda\rangle = -i\gamma\sqrt{\omega(q_\lambda)}\,|\mathcal{N}\rangle + \hbar\omega(q_\lambda)|\lambda\rangle. \quad (26)$$

Physically, the requirement $\epsilon(k - q_\lambda) \ll \hbar\omega(q_\lambda)$ approximates the complete "stopping" of the incident electron.

To study this case, we follow the same procedure as was outlined in the previous section, but proceed here from Eqs. (12) and (26) [as opposed to Eqs. (12) and (13)]. One then obtains the following expression for the probability amplitude:

$$\langle\mathcal{N}|\Psi(t)\rangle = -\frac{1}{2\pi i}\int_C dz\, e^{-izt} \times\left(z - \frac{\epsilon(k)}{\hbar} - \frac{4\pi v\gamma^2}{\hbar^2}\int_0^\mu \frac{q_\lambda^3\, dq_\lambda}{z - vq_\lambda}\right)^{-1}. \quad (27)$$

The μ integral appearing in the denominator of Eq. (27) can be handled explicitly, with the result

$$\langle\mathcal{N}|\Psi(t)\rangle = -\frac{1}{2\pi i}\int_C dz\, e^{-izt} \times\left\{z - \frac{\epsilon(k)}{\hbar} + \frac{4\pi\gamma^2}{\hbar^2}\left[\frac{\mu^3}{3} + \frac{\mu^2}{2}\left(\frac{z}{v}\right) + \mu\left(\frac{z}{v}\right)^2 + \left(\frac{z}{v}\right)^3 \ln\left(\frac{v\mu - z}{-z}\right)\right]\right\}^{-1} \quad (28)$$

It is convenient at this point to introduce the dimensionless variables

$$R = \hbar v\mu/\epsilon(k), \quad (29)$$

$$\tau = F\epsilon(k)tR/\hbar, \quad (30)$$

$$\xi = z\hbar/RF\,\epsilon(k), \quad (31)$$

where F is the dimensionless interaction parameter of Fröhlich, Eq. (7). Then, with the additional definition

$$B = 2\epsilon(k)^2 R^2\zeta/3\pi n\hbar^3 v^3, \quad (32)$$

the expression for the contour integral, Eq. (28), may be written

$$\langle\mathcal{N}|\Psi(\tau)\rangle = -\frac{1}{2\pi i} \times\int_C \frac{(1/B)\,d\xi\, e^{-i\xi\tau}}{F^3\xi^3\ln(1 - 1/F\xi) + F^2\xi^2 + (\frac{1}{2}F + 1/B)\xi + (\frac{1}{3} - 1/RFB)}, \quad (33)$$

a form in which the analytic structure of the integral is more clearly displayed. To evaluate this contour integral (via Cauchy's theorem), we separate the denominator into real and imaginary parts so as to determine the location of the poles. We write

$$\xi = x + iy$$

and obtain

$$\begin{aligned}\mathrm{Re\,(den)} &= F^3(x^3 - 3xy^2)\ln|1 - 1/F\xi| \\ &\quad - F^3(3x^2y - y^3)\arg(1 - 1/F\xi) \\ &\quad + F^2(x^2 - y^2) + (1/B + F/2)x + (\tfrac{1}{3} - 1/BRF), \end{aligned} \quad (34a)$$

$$\begin{aligned}\mathrm{Im\,(den)} &= F^3(3x^2y - y^3)\ln|1 - 1/F\xi| \\ &\quad + F^3(x^3 - 3xy^2)\arg(1 - 1/F\xi) \\ &\quad + F^2\,2xy + (1/B + F/2)y. \end{aligned} \quad (34b)$$

So that the logarithm be a single-valued function, we choose as the range of its argument, $-\pi < \arg \leq +\pi$. Given that the variable ξ occurs in the argument as $(1 - 1/F\xi)$, the choice $-\pi < \arg \leq +\pi$ corresponds to choosing a branch cut in the ξ-plane from the origin to the point $1/F$. One observes that the argument vanishes only when $y = 0$ and $\{x < 0$ or $x > 1/F\}$. With this specification, we see that

$$\mathrm{Re(den)} = F^3x^3\ln|1 - 1/Fx| + F^2x^2 + (1/B + F/2)x + (\tfrac{1}{3} - 1/RFB), \quad (35)$$

$$\mathrm{Im(den)} = 0. \quad (36)$$

To gain a preliminary understanding of the behavior of Re(den) in the limit of large x, we make use of the well-known approximation[18]

$$\ln(1 \pm \eta\cdot 10^{-n}) = \pm\eta\cdot 10^{-n} - \tfrac{1}{2}\eta^2\cdot 10^{-2n} \quad (n > 10)$$

(stated here in terms of η, an arbitrary real variable) to study the behavior of $\ln(1 - 1/Fx)$ as $x \to \pm\infty$, and find that

$$\lim_{x\to\pm\infty} \mathrm{Re(den)} = \pm\infty. \quad (37a)$$

Also,

$$\lim_{x\to(1/F)_+} \mathrm{Re(den)} = -\infty, \quad (37b)$$

$$\lim_{x\to 0_-} \mathrm{Re(den)} = \tfrac{1}{3} - 1/RFB, \quad (37c)$$

the last result following since[18]

$$\lim_{\eta\to 0} \eta^\alpha \ln\eta = 0 \quad \text{for } \alpha > 0.$$

Taking into account the analytic structure of the expression for Re(den) and making use of the limits, Eqs. (37), one can sketch the approximate behavior of Re(den) vs x, and this is displayed in Fig. 1. Upon examining the behavior at $x = 0$, we notice that if

$$\tfrac{1}{3} > 1/RFB,$$

then the intercept on the Re(den) axis lies on the positive segment of that axis. As a consequence, the function Re(den) is expected to cross the x axis at two points, once on the negative x axis (at a point we shall refer to as ξ_-) and once for $\xi > 1/F$ on the positive x axis (at a

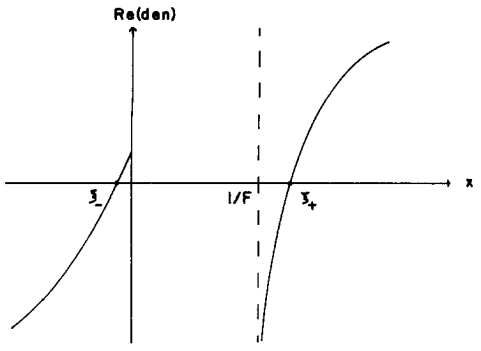

FIG. 1. A sketch of Re(den) vs x. The ξ_+ and ξ_- refer to the simple poles on the positive and negative x axis, respectively.

point we shall denote as ξ_+). Indeed, a full numerical study of Eq. (35) reveals that for $\frac{1}{3} > 1/RFB$ these are the only two points in the ξ plane for which the denominator of Eq. (33) vanishes. Conversely, if

$$\tfrac{1}{3} < 1/RFB,$$

then the intercept on the Re(den) axis will be on the negative segment of that axis, and consequently the denominator of Eq. (33) will vanish only once, in particular for $\xi > 1/F$ on the positive x axis. Given these remarks, it is anticipated that the solution of the contour integral, Eq. (33), and hence the time evolution of the system will depend crucially on the coupling constant F, the incident kinetic energy of the electron, and the cutoff on the wave vector spectrum.

Choosing the contour shown in Fig. 2, and noting that the contribution to the integral from the large semicircle and the two small circles vanishes [Jordan's lemma applied to the integrand of Eq. (33)], we have

$$\langle \mathcal{N} | \Psi(\tau) \rangle = \frac{1}{B} \sum_{\xi_+,\xi_-} \text{Res} + \frac{F^3}{B} \int_0^{1/F} \frac{x^3 e^{-ix\tau}\, dx}{\text{Re}(x)^2 + \pi^2 F^6 x^6}, \tag{38}$$

where $\text{Re}(x)$ is given by Eq. (35) and, as indicated above, ξ_- and ξ_+ designate the simple poles on the negative and positive x axis, respectively. To determine the residues at the poles, we expand the denominator in a Taylor series about the poles:

$$\text{denominator} = D_0 + D_1(\xi - \xi_i) + \tfrac{1}{2} D_2(\xi - \xi_i)^2 + \ldots,$$

where D_1 is the first derivative, D_2 is the second derivative, and so on; the coefficient D_0 must be zero since the denominator vanishes at $\xi = \xi_i$. Numerical studies reveal that D_1 is nonzero in the neighborhood of the poles, so that the residue at either pole is

$$\text{Res} = e^{-i\xi_i\tau}/D_1(\xi_i).$$

The final expression for the wavefunction of the system in the state $|\mathcal{N}\rangle$ for the case of one pole $\xi_+ (\frac{1}{3} < 1/RFB)$ is, therefore,

$$\langle \mathcal{N} | \Psi(\tau) \rangle = \frac{1}{B} \left(\frac{e^{-i\xi_+\tau}}{D_1(\xi_+)} + F^3 \int_0^{1/F} \frac{x^2 e^{-ix\tau}\, dx}{\text{Re}(x)^2 + \pi^2 F^6 x^6} \right). \tag{39}$$

Hence, the diagonal element of the density matrix for the electron-oscillators system for the case that a single pole contributes is

$$\rho(\tau) = \frac{1}{B^2} \left(\frac{1}{D_1(\xi_+)^2} + 2F^3 \frac{\cos(\xi_+\tau)}{D_1(\xi_+)} C(\tau) + 2F^3 \frac{\sin(\xi_+\tau)}{D_1(\xi_+)} S(\tau) + F^6 [C(\tau)^2 + S(\tau)^2] \right), \tag{40}$$

where

$$S(\tau) = \int_0^{1/F} \frac{x^3 \sin(x\tau)\, dx}{\text{Re}(x)^2 + \pi^2 F^6 x^6}, \tag{41}$$

$$C(\tau) = \int_0^{1/F} \frac{x^3 \cos(x\tau)\, dx}{\text{Re}(x)^2 + \pi^2 F^6 x^6}. \tag{42}$$

The wavefunction for the system in the state $|\mathcal{N}\rangle$ for the case that both poles, ξ_+ and ξ_-, contribute is

$$\langle \mathcal{N} | \Psi(\tau) \rangle = \frac{1}{B} \left(\frac{e^{-i\xi_+\tau}}{D_1(\xi_+)} + \frac{e^{-i\xi_-\tau}}{D_1(\xi_-)} + F^3 \int_0^{1/F} \frac{x^3 e^{-ix\tau}\, dx}{\text{Re}(x)^2 + \pi^2 F^2 x^6} \right). \tag{43}$$

The diagonal element of the density matrix for the electron-oscillator system is then given by

$$\rho(\tau) = \frac{1}{B^2} \left[\frac{1}{D_1(\xi_+)^2} + \frac{1}{D_1(\xi_-)^2} + \frac{2\cos(\xi_+ - \xi_-)\tau}{D_1(\xi_+) D_1(\xi_-)} + 2F^3 C(\tau) \left(\frac{\cos(\xi_+\tau)}{D_1(\xi_+)} + \frac{\cos(\xi_-\tau)}{D_1(\xi_-)} \right) + 2F^3 S(\tau) \left(\frac{\sin(\xi_+\tau)}{D_1(\xi_+)} + \frac{\sin(\xi_-\tau)}{D_1(\xi_-)} \right) + F^6 [C^2(\tau) + S^2(\tau)] \right], \tag{44}$$

with $\text{Re}(x)$ given by Eq. (35). These, then, are the expressions for the evolution of the system as determined from the time-dependent Schrödinger equation, subject to (1) the long-wavelength approximation, (2) the use of an upper bound on the frequency spectrum, and, specifically for this case, (3) the restriction of the initial kinetic energy of the electron to low energies. The third assumption will be changed in the following section, where we will consider the case of a high-energy incident electron, but, before proceeding, we indicate briefly the method of solving Eqs. (40) and (44) numerically. First, one determines the number and location of the poles and then evaluates the corresponding first derivatives. Next, the $S(\tau)$ and $C(\tau)$ are computed, and, finally, the expression for $\rho(\tau)$ is studied as a function of τ. The results of our numerical studies will be reported in Sec. 5.

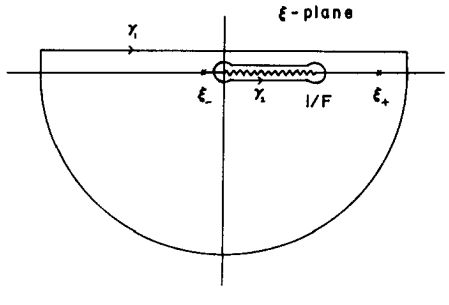

FIG. 2. The contour used in evaluating the expressions for $\langle \mathcal{N} | \Psi(t) \rangle$.

B. High-energy incident electron

Here one considers the limiting situation wherein the kinetic energy of the incident electron (again, fixed at some constant value) is assumed to be much larger than the energy corresponding to any single excitation of the initially de-excited lattice; that is, we assume that $\epsilon(k) \gg \hbar v\mu$. We then require that, upon energy transfer, although the electron may deposit a quantum of energy $\hbar\omega_\lambda$, where $\hbar\omega_\lambda \ll \epsilon(k)$, the overall change in the kinetic energy of the electron is negligible relative to its incident kinetic energy. The consideration of this limiting situation is equivalent to the statement that $\epsilon(k - q_\lambda) \simeq \epsilon(k)$, so, in terms of our formulation of the problem, Eq. (13) is replaced by the modified expression

$$H|\lambda\rangle = -i\gamma\sqrt{\omega(q_\lambda)}\,|N\rangle + [\epsilon(k) + \hbar\omega(q_\lambda)]\,|\lambda\rangle. \tag{45}$$

Again, following the same procedure as was utilized in Section 3, but proceeding now from Eqs. (12) and (45), the following expression is obtained for the inverse Laplace transform of the probability amplitude:

$$\langle N|\Psi(t)\rangle = \frac{1}{2\pi i}\int_C dz\, e^{-izt} \times \left(z - \frac{\epsilon(k)}{\hbar} - \frac{\gamma^2 4\pi v}{\hbar^2}\int_0^\mu \frac{q_\lambda^3\, dq_\lambda}{z - (1/\hbar)[\epsilon(k) + \hbar v q_\lambda]}\right)^{-1}. \tag{46}$$

It is convenient at this point to extract the phase $e^{-i\epsilon(k)t/\hbar}$ from the right-hand side, thus obtaining

$$\langle N|\Psi(t)\rangle e^{+i\epsilon(k)t/\hbar} = -\frac{1}{2\pi i}\int_C dz\, e^{-izt} \times \left(z - \frac{\gamma^2 4\pi v}{\hbar^2}\int_0^\mu \frac{q_\lambda^3\, dq_\lambda}{z - vq_\lambda}\right)^{-1}. \tag{47}$$

Treating the μ integral explicitly leads to the result

$$\langle N|\Psi(t)\rangle e^{+i\epsilon(k)t/\hbar} = -\frac{1}{2\pi i}\int_C dz\, e^{-izt} \times \left\{z + \frac{\gamma^2 4\pi}{\hbar^2}\left[\frac{\mu^3}{3} + \frac{z\mu^2}{2v} + \frac{z^2\mu}{v^2} + \frac{z^3}{v^3}\ln\left(1 - \frac{v\mu}{z}\right)\right]\right\}^{-1}. \tag{48}$$

We now transform to the dimensionless variables R, τ, and ξ introduced earlier, and replace γ^2 by the corresponding expression involving F. The result is

$$\langle N|\Psi(\tau)\rangle e^{i\tau/FR} = -\frac{1}{2\pi i}\int_C \frac{(1/B)\,d\xi\, e^{-i\xi\tau}}{F^3\xi^3\ln(1 - 1/F\xi) + F^2\xi^2 + (1/B + F/2)\xi + \frac{1}{3}}. \tag{49}$$

It is seen that this contour integral is similar in structure to Eq. (33), but with one important difference; again referring to Fig. 1, we see here that the intercept on the Re(den) axis is always positive, and hence the denominator will always vanish once for ξ on the negative real axis and once for $\xi > 1/F$ on the positive real axis. Accordingly, the diagonal element of the density matrix will have the same formal structure as the "two-pole" result, Eq. (44), but with the Re(x) given by

$$\mathrm{Re}(x) = F^3x^3\ln|1 - 1/Fx| + F^2x^2 + (1/B + F/2)x + \tfrac{1}{3}. \tag{50}$$

Our numerical study of this case will be presented in the following section, along with a general discussion and our conclusions.

5. DISCUSSION

In this paper we have introduced a simple model to study the deposition of energy in a condensed medium from an energetic charged particle. Our approach is different from previous work in radiation chemistry in that we determine the dynamics of the electron-oscillators system starting from a master equation constructed from the time-dependent Schrödinger equation. It was found that, provided one imposed a frequency cutoff to avoid the (ultraviolet) divergence associated with the adoption of the long-wavelength approximation, an explicit analysis could be carried through in two, quite different, energy regimes: the case of a low-energy incident electron and that of a high-energy incident electron. It is worth emphasizing that once the long-wavelength and cutoff approximations have been introduced, and the energy regime specified, the subsequent analysis is carried through without any restriction on the coupling between the electron and the oscillator bath; that is, a weak-coupling approximation is not introduced in solving the resulting master equation. Indeed, it is most instructive to compare the results obtained with and without the use of a weak-coupling approximation, and this comparison will be made later on in this section.

Whereas our over-all objective in this work is the application of the master-equation approach to problems of experimental interest in radiation chemistry, it is our belief that before specific property calculations are undertaken, the dynamical behavior of the electron-oscillators system must be thoroughly characterized, both qualitatively and quantitatively. Accordingly, in the present paper, we report only a few, representative calculations which serve to display the time evolution of $\rho(\tau)$, as determined by the master equation introduced in Sec. 3, when the electron-oscillators system is subjected to a certain, well-defined set of constraints (that is, particular values of F, R, and B).

We consider first the numerical results obtained in the low-energy regime. Here we had required the range of possible incident kinetic energies of the electron to be bounded from above by $\hbar v\mu$; that is, we specified $\epsilon(k) \leq \hbar v\mu$. Furthermore, upon energy transfer, it was assumed that the final kinetic energy of the electron was negligible relative to a single excitation of the initially de-excited field; that is, we required that $\epsilon(k - q_\lambda) \ll \hbar\omega(q_\lambda)$. To facilitate our discussion of this case, it is useful to point out that the dimensionless parameter R, defined by Eq. (29), is just the ratio

$$R = \frac{\text{upper bound on the energy spectrum of the lattice}}{\text{incident kinetic energy of the electron}}.$$

Clearly, then, for the low-energy case, the range $R \geq 1$ is the regime of interest, and for definiteness we consider the choice $R = 2.0$; this value of R corresponds to an incident kinetic energy of the electron equal to one-half the upper bound on the energy spectrum of the lattice. The next parameter which must be specified is the

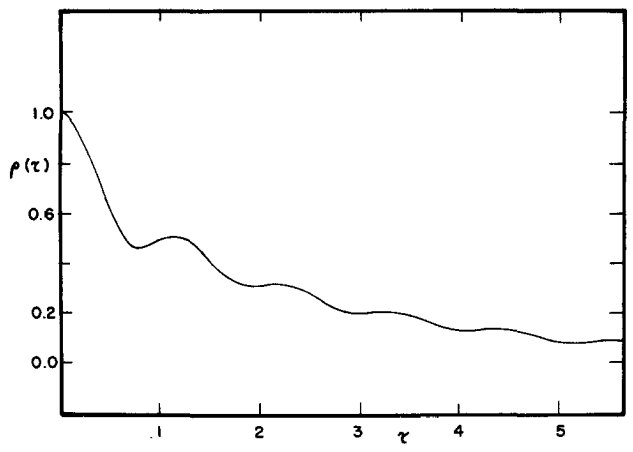

FIG. 3. A plot of $\rho(\tau)$ vs τ for the case of a low-energy incident electron, for the choice of parameters $F=0.1$, $R=2.0$, and $B=1.0$.

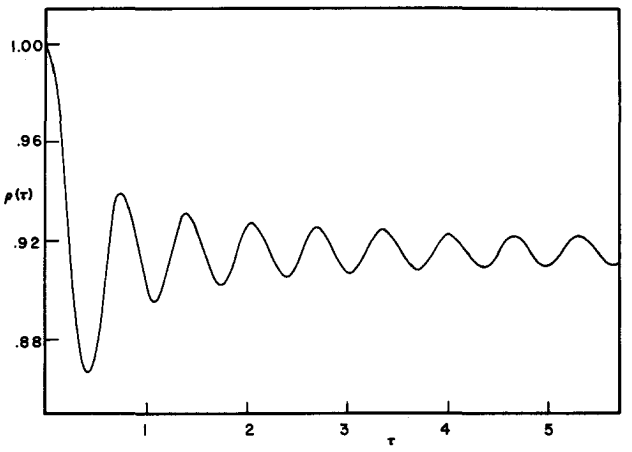

FIG. 4. A plot of $\rho(\tau)$ vs τ for the case of a high-energy incident electron, for the choice of parameters $F=0.1$ and $B=1.0$.

dimensionless coupling constant F, and here, in order to permit a meaningful comparison with the two weak-coupling theories mentioned earlier, we adopt the value $F=0.1$ as our initial choice for the value of the coupling between the electron and the oscillator bath. Finally, the parameter B, although devoid of any obvious physical interpretation, must be chosen so as to be consistent with the particular values selected for F and R. In addition, we must identify the characteristic density n and energy ζ used in the definition, Eq. (7). Given the nature of the electron-oscillators model, the density n is conveniently chosen to be the density of modes, that is, the number of modes of the oscillator bath per unit volume available for energy deposition via interaction with an energetic electron; this density will be proportional to $(2\pi)^{-3}(\frac{4}{3}\pi\mu^3)$. The characteristic energy parameter ζ can be chosen to be the quantum-mechanical energy for a single particle of mass m and momentum $\hbar\mathbf{k}$ in a volume Ω, and for definiteness we assume that $\zeta=0.1\epsilon(k)$. With these specifications, the magnitude of the parameter B is of the order of unity, and whereas we have a certain flexibility in assigning a value to B, in our subsequent calculations we shall set $B=1.0$.

The time evolution of $\rho(\tau)$, as determined by the solution to the master equation for the set of values $F=0.1$, $R=2.0$, and $B=1.0$, is displayed in Fig. 3. An examination of this figure reveals that the electron-oscillators system relaxes, eventually, in such a way that $\rho(\tau)\to 0$ as $\tau\to\infty$; this property of the solution is equivalent to the statement that in the asymptotic limit the electron has zero probability of being found in the initial state $|N\rangle$. The associated decrease in the kinetic energy of the electron approximates the "stopping" of a charged particle passing through a condensed medium. Of some considerable interest is the observation that the essentially exponential decay is overlaid with nonexponential oscillations; calculations show that these oscillations tend to become less rapid as the parameter R approaches unity, while, at the same time, the individual amplitudes of these oscillations tend to increase. (See Fig. 6, discussed later.) With particular reference to the role of the cutoff in determining the characteristics of these nonexponential contributions, this behavior is reminiscent of that found in VI, where a weak-coupling solution to the Prigogine–Résibois master equation was obtained for the Wigner–Weisskopf atom in a three-dimensional radiation field and for a single, effective spin in interaction with the vibrational modes of a three-dimensional lattice; in this respect, the cutoff parameter seems to play the same role in all three models.

Turning our attention to the case of a high-energy incident electron, we recall that the range of possible kinetic energies of the electron was assumed to be bounded from below by $\hbar v\mu$; that is, we required that $\epsilon(k)>\hbar v\mu$. This assumption, and the further requirement that the change in the kinetic energy of the electron upon interaction is negligible relative to its initial kinetic energy, effectively remove the R-dependence from the final ex-

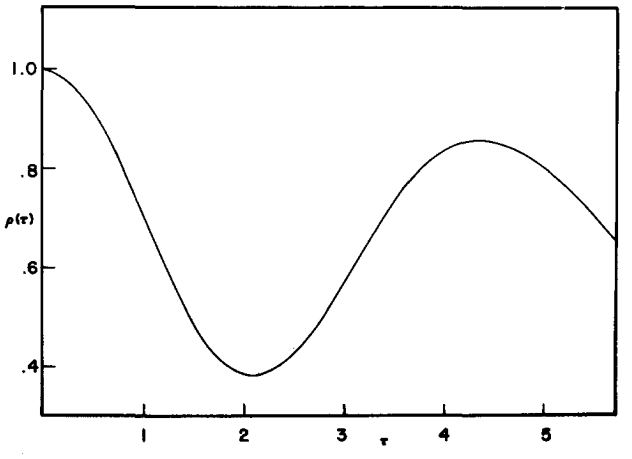

FIG. 5. A plot of $\rho(\tau)$ vs τ for the case of a high-energy incident electron, for the choice of parameters $F=1.0$ and $B=1.0$.

pression for $\rho(\tau)$. [See Eq. (46).] Hence, we need specify only the parameters F and B in our calculations here, and for purposes of comparison we choose the values $F=0.1$ and $B=1.0$. The resulting behavior of $\rho(\tau)$ is graphed in Fig. 4, where it is seen that the probability that the electron be in the state $|\Lambda\rangle$ at time τ is close to unity, which implies that the electron is essentially unaffected as it migrates through the lattice. Since our study was performed without imposing a weak-coupling approximation, we can also consider the case wherein the incident electron, although possessing a very high energy, couples very strongly to the phonon modes of the lattice. This situation is studied by fixing $B=1.0$, but allowing the coupling parameter F to increase by an order of magnitude; that is, we set $F=1.0$. The result is shown in Fig. 5, where it is seen that the probability that the electron be in the state $|\Lambda\rangle$ at time τ is characterized by slowly varying oscillations with large amplitude. This behavior, associated with an increase in the magnitude of the coupling constant, is again reminiscent of the behavior found in the two models studied in VI using the Prigogine–Rêsibois master equation.

For single-mode excitation processes, we notice that in the high-energy case, $\rho(\tau)$ does not approach zero as $\tau\to\infty$. Of course, physically, we expect that as the electron migrates through the lattice, it will deposit energy in a succession of modes, eventually losing all of its kinetic energy with the consequence that $\rho(\tau)\to 0$ as $\tau\to\infty$. Now, to generalize our approach to deal with a sequence of de-excitation events using the Hamiltonian, Eq. (8), one must retain the explicit q_λ-dependence in the term $\epsilon(k-q_\lambda)$. The immediate consequence of this generalization is that a whole hierarchy of coupled equations arises. To illustrate this point for the simplest possible case, suppose that two modes of the lattice are successively excited by interaction with the energetic electron; as regards the history of the electron, it passes from an incident energy $\epsilon(k)$ to $\epsilon(k-q_\lambda)$ to $\epsilon(k-q_\lambda-q_{\lambda'})$, while, at the same tiem, the lattice is excited successively from $|0\rangle$ to $|q_\lambda\rangle$ to $|q_\lambda, q_{\lambda'}\rangle$. Thus, even for a three-state problem one has two coupled (singular) equations which would have to be solved simultaneously, and for obvious reasons this problem is even more difficult to solve exactly than the one discussed in this paper. Nonetheless, as will be shown in a subsequent paper where we take up the problem of multiple-mode excitation events, progress can be made if one adopts a rationale for decoupling the hierarchy of equations.

It should be remarked that in the low-energy regime, the assumption $\epsilon(k)\ll\hbar\omega(q_\lambda)$ also has the effect of removing the q_λ-dependence from the term $\epsilon(k-q_\lambda)$, but here we find that $\rho(\tau)\to 0$ as $\tau\to\infty$ as expected, since in this regime $\epsilon(k)<\hbar v\mu$, and the electron can transfer its kinetic energy via a single-mode excitation event. To pursue this point further, one can consider the borderline case wherein one sets $\epsilon(k)=\hbar v\mu$ and then works out the dynamics using the low-energy formulation. As is seen from Fig. 6, removal of the q_λ-dependence in the regime $\epsilon(k)=\hbar v\mu$ $(R=1.0)$ leads to results which are intermediate between those displayed in Fig. 3 (the case $\epsilon(k)<\hbar v\mu$) and Fig. 4 (the case $\epsilon(k)>\hbar v\mu$). Physically, these results suggest that in the intermediate-energy regime (an electron characterized by an incident kinetic energy $\epsilon(k)\cong\hbar v\mu$) a charged particle tends not to be "stopped" via any single-mode de-excitation event. It should also be mentioned that nonzero asymptotic limits (as displayed, for example, in Figs. 4 and 5) are linked to the question of ergodicity, as pointed out in V; we shall not discuss this point here, since the use of the approximation $\epsilon(k-q_\lambda)\cong\epsilon(k)$ in the high-energy case would make the comparison between the results found here and the results found in V somewhat tentative—despite the fact that strong qualitative similarities do exist between the results found in the two studies.

The next point to be taken up concerns the comparison of our results with those obtained by Van Hove and coworkers.[4–7] As stated in the Introduction, Van Hove developed a master equation for the system of an electron interacting with a collection of static, random scatterers. Three different expressions were derived describing the behavior of this system for weak, intermediate, and strong coupling between the electron and the scatterers. Their numerical studies showed the time evolution of the system to be characterized by damped oscillations in the limit of strong coupling and by an essentially monotonic decay in the limit of weak coupling. Now, it is stated in Ref. 5 that the master equation for the electron-scatterers system can be regarded as a contracted form of the master equation for the electron-oscillators system, but one for which the energy of the oscillators has been suppressed. It is most interesting, therefore, to compare the evolution of the electron-oscillators system as determined via our analysis with the evolution characterizing the electron-scatterers system, and, in particular, with the numerical results reported in Ref. 7 for various coupling regimes. The chief similarity in the behavior of the two systems is the emergence of nonexponential behavior as the coupling between the electron and the system under study increases. As

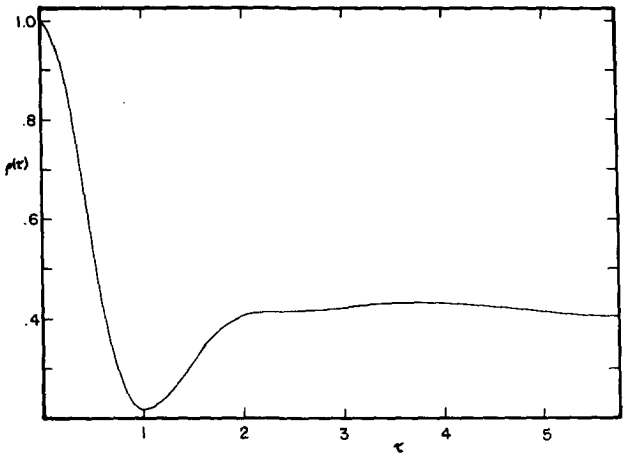

FIG. 6. A plot of $\rho(\tau)$ vs τ for the case of an intermediate-energy incident electron, for the choice of parameters $F=0.1$, $R=1.0$, and $B=1.0$.

pointed out by Zwanzig,[14] exponential decay in time may not be universal, and may, in fact, be hidden behind some other, more complicated type of time dependence. Indeed, other studies on simple models treated via a master-equation approach (e.g., the Wigner–Weisskopf atom in a radiation field and a single, effective spin in interaction with the phonon modes of a lattice) reveal this same property, and hence one is led to the tentative conclusion that perhaps this behavior is general enough to be of importance in energy-transfer problems in radiation chemistry. In particular, it would be most interesting to examine the experimental consequences of this predicted nonexponential relaxation in the specific problem of electrons scattered from thin films, a problem currently under experimental investigation by W. Hamill.[19]

The significant difference which emerges from the numerical work on the two systems is that the time evolution of the scatterers system is characterized by $\rho(\tau)\to 0$ as $\tau\to\infty$ for all couplings, whereas in our study this behavior is found only in the limit of weak coupling in the low-energy regime. As mentioned earlier in our discussion of the electron-oscillators system, the removal of the q_λ-dependence from the term $\epsilon(k-q_\lambda)$ is responsible for our $\rho(\tau)$ not exhibiting this behavior in the asymptotic time limit. On the other hand, Van Hove, in his solution of the electron-scatterers problem, assumed that his transition probability $W_{ll'}(k,k')$ satisfied the condition

$$W_{ll'}(k,k')=\begin{cases}W, & k,k'\le a,\\ 0, & \text{otherwise},\end{cases} \tag{51}$$

where W and a are arbitrary positive constants; it is noted in Ref. 5 that this approximation is necessary in order to obtain an explicit solution to the scatterers problem. Now, the statement, Eq. (51), is equivalent to the assumption that the functional dependence of the transition probability on the wave vector may be totally suppressed, and in one sense this is equivalent to the procedure whereby the q_λ-dependence in the term $\epsilon(k-q_\lambda)$ is removed in our treatment. Moreover, the assumption, Eq. (48), is equivalent to establishing an upper bound on the energy spectrum, in that the restriction $\mathbf{k},\mathbf{k}'\le a$ is imposed, and so here again a similarity with our treatment, whereby a bound is imposed on the energy spectrum of the lattice, can be noted. Therefore, at first sight, it would seem that the kinds of approximations introduced in solving the electron-scatterers problem are somewhat similar to those employed in our treatment of the electron-oscillators problem, and this suggests that the difference in behavior exhibited by the two models may be due to an even more fundamental difference in the formulation of the two problems. Here we remark that in Van Hove's treatment of the electron-scatterers problem, a significant number of excited vibrational modes are considered. On the other hand, in our study the states considered are those in which either no vibrational mode or (after deposition) only one vibrational mode of the lattice is excited. Since the number of vibrational modes excited may be thought of as a measure of the temperature of the system, perhaps one can conclude from this that the temperature of the bath plays an essential role in determining the behavior $\rho(\tau)\to 0$ as $\tau\to\infty$. This point will be pursued further in our subsequent study.

The final point to be made regarding our analysis concerns the two weak-coupling calculations presented in Appendices A and B. In Appendix A, the evolution of the electron-oscillators system is determined from the Schrödinger equation for the case of a high-energy incident electron by using the operator expansion

$$\left(z-\frac{H}{\hbar}\right)^{-1}=\sum_{n=0}^{\infty}\left(z-\frac{H_0}{\hbar}\right)^{-1}\left[\frac{V}{\hbar}\left(z-\frac{H_0}{\hbar}\right)^{-1}\right]^n, \tag{52}$$

where

$$H_0=\sum_k \epsilon(k)\alpha_k^*\alpha_k+\sum_{q_\lambda}\hbar\omega(q_\lambda)\left[a_{q_\lambda}^* a_{q_\lambda}+\tfrac{1}{2}\right], \tag{53}$$

and where V is the interaction part of the Hamiltonian, defined by Eq. (6). The weak-coupling approximation here corresponds to retaining only the first nonzero term in the above expansion. The behavior predicted by the solution, Eq. (A11), is clearly undesirable, due to the presence of secular terms $O(\tau^2)$ in this expression for the diagonal element of the density matrix. In particular, the term $\tau^2B^2/9$ causes the density-matrix element to grow without bound as τ increases, even in the limit that the coupling parameter approaches zero. To remedy this, one might try to retain higher-order terms in the operator expansion, but this does not seem to result in an improvement. It is possible to understand why this perturbation expansion fails by comparing Eqs. (19) and (A4). One sees that the perturbation expansion is equivalent to $(z-\epsilon(k)/\hbar)^{-1}$ times the geometric series expansion for $(1-f(z))^{-1}$, where

$$f(z)=\frac{\gamma^2}{\hbar^2}\left(\frac{8\pi^3}{\Omega}\right)\frac{1}{(z-\epsilon(k)/\hbar)}\sum_\lambda\frac{\omega(q_\lambda)}{z-(1/\hbar)[\epsilon(k-q_\lambda)+\hbar\omega(q_\lambda)]},$$

and that the condition on the radius of convergence for such series, $|f(z)|<1$, is not satisfied. Thus, for this problem we conclude that the use of a weak-coupling approximation generated by using the operator expansion, Eq. (49), is unacceptable.

In Appendix B, the evolution of the electron-oscillators system is studied by constructing a weak-coupling solution to the Prigogine–Rêsibois master equation, and it is found that the result is identical to the equation obtained in our earlier study of spin-lattice relaxation.[13] The evolution is characterized by an exponential component plus nonexponential contributions consisting of damped oscillations. The behavior of the system depends on the values chosen for the coupling parameter F and the parameter B. This dependence is described in detail in VI, where the results of several numerical studies are reported. As is the case for spontaneous photon emission and spontaneous phonon emission—phenomena previously studied via a weak-coupling solution to the Prigogine–Rêsibois master equation—negative values for the overall density-matrix element appear in the electron-oscillators problem. Comments concerning this undesirable feature are presented in I and also apply here.

The significance of these two weak-coupling studies as regards the present work can now be understood. It might have been hoped that a weak-coupling approxi-

mation, properly introduced within the framework of the Schrödinger master equation or the Prigogine–Rêsibois master equation, might have allowed a "shortcut" in the mathematical analysis of the problem that, in turn, would have permitted the subsequent consideration of more complicated dynamical events. In the former case, however, the use of a weak-coupling approximation leads to the occurrence of secular terms which cause the diagonal element of the density matrix to grow without bound with increasing time, whereas the use of this approximation in the Prigogine–Rêsibois master equation results in negative contributions to the density matrix, which defy the physical interpretation of a density matrix. In other words, for this class of problems a weak-coupling approximation seems to fail even in the simplest case of an effective two-state relaxation process.

In conclusion, then, we have considered in this paper a model of an electron in interaction with the phonon modes of a lattice, with particular emphasis on the "stopping" of the electron via single-mode de-excitation events. The formal solution was obtained, Eq. (24), but the mathematical difficulties associated with quadratic terms in $\epsilon(k-q_\lambda)$ led to the consideration of the problem in two limiting energy regimes (the case of a high-energy incident electron and that of a low-energy incident electron), the mathematical consequence of which was that the q_λ-dependence in the term $\epsilon(k-q_\lambda)$ could be removed. Once the energy regime was specified, the subsequent analysis was carried through exactly; that is, without imposing a weak-coupling approximation. The consequences of our study as regards predicted physical effects would seem to demand a careful study, and this shall be presented in a subsequent paper, where we will include as well a detailed comparison with the predictions of other theories of energy transfer. For the most part, however, the principal objective of the present paper has been achieved: to construct a mathematical framework, starting from first principles, within which the energy-transfer problem in radiation chemistry could be discussed. This has been carried out by starting from a well-defined Hamiltonian, using an approach based on a master equation (in this case, one derived from the time-dependent Schrödinger equation) and employing only those approximations whose mathematical consequences could be determined. It is in this spirit that our further studies will be pursued.

ACKNOWLEDGMENT

The authors have benefited from funds provided by the United States Atomic Energy Commission through contract AT(11-1)-38.

APPENDIX A: THE WEAK-COUPLING QUANTUM-MECHANICAL SOLUTION

It is possible to obtain a weak-coupling description for the time evolution of the electron-oscillators system by making use of the operator identity

$$\left(z-\frac{H}{\hbar}\right)^{-1}=\left(z-\frac{H_0}{\hbar}\right)^{-1}+\left(z-\frac{H_0}{\hbar}\right)^{-1}\frac{V}{\hbar}\left(z-\frac{H}{\hbar}\right)^{-1}, \tag{A1}$$

where H_0 is the unperturbed part of the Hamiltonian, Eq. (53), and V is the interaction part of the Hamiltonian, defined by Eq. (6). There exists an iterative expansion for Eq. (16) which can be derived by repeated application of this identity; it is

$$S_{\mathcal{N}}=\sum_{n=0}^{\infty}\left(z-\frac{\epsilon(k)}{\hbar}\right)^{-1}\langle\mathcal{N}|\left[\frac{V}{\hbar}\left(z-\frac{H_0}{\hbar}\right)^{-1}\right]^n|\mathcal{N}\rangle. \tag{A2}$$

The terms corresponding to $n=0$ and $n=1$ vanish, so we take as the weak-coupling solution the term corresponding to $n=2$ in the expansion

$$S_{\mathcal{N}}=\left(z-\frac{\epsilon(k)}{\hbar}\right)^{-1}\left\{1+\langle\mathcal{N}|\left[\frac{V}{\hbar}\left(z-\frac{H_0}{\hbar}\right)^{-1}\right]^2|\mathcal{N}\rangle\right\}. \tag{A3}$$

Repeated application of the completeness relation affords the result

$$S_{\mathcal{N}}=\left(z-\frac{\epsilon(k)}{\hbar}\right)^{-1}\times\left(1+\frac{(8\pi^3/\Omega)\gamma^2}{\hbar^2(z-\epsilon(k)/\hbar)}\sum_\lambda\frac{\omega(q_\lambda)}{z-\epsilon(k-q_\lambda)/\hbar-\omega(q_\lambda)}\right). \tag{A4}$$

We transpose from γ^2 to F and obtain an expression for the wavefunction for the system in the state $|\mathcal{N}\rangle$:

$$\langle\mathcal{N}|\Psi(t)\rangle=-\frac{1}{2\pi i}\int_C\frac{dz\,e^{-izt}}{(z-\epsilon(k)/\hbar)}\times\left(1+\frac{4\pi\zeta F}{3n\Omega\hbar(z-\epsilon(k)/\hbar)}\sum_\lambda\frac{\omega(q_\lambda)}{z-\epsilon(k-q_\lambda)/\hbar-\omega(q_\lambda)}\right). \tag{A5}$$

It is convenient at this point to introduce the same assumptions that were invoked earlier (the long-wavelength and cutoff approximations), and for definiteness we consider the case of a high-energy incident electron. The result is

$$\langle\mathcal{N}|\Psi(t)\rangle=-\frac{1}{2\pi i}\int_C\frac{dz\,e^{-izt}}{(z-\epsilon(k)/\hbar)}\times\left(1+\frac{2\zeta Fv}{3n\pi\hbar(z-\epsilon(k)/\hbar)}\int_0^\mu\frac{q_\lambda^{\,3}\,dq_\lambda}{z-\epsilon(k)/\hbar-vq_\lambda}\right). \tag{A6}$$

We extract the phase factor $e^{-i\epsilon(k)t/\hbar}$ by performing the transformation $z-\epsilon(k)/\hbar\to z$, and obtain

$$\langle\mathcal{N}|\Psi(t)\rangle e^{+i\epsilon(k)t/\hbar}=-\frac{1}{2\pi i}\int_C\frac{dz\,e^{-izt}}{z}\times\left(1+\frac{2\zeta Fv}{3n\pi\hbar z}\int_0^\mu\frac{q_\lambda^{\,3}\,dq_\lambda}{z-vq_\lambda}\right). \tag{A7}$$

It is also convenient to introduce the dimensionless variables employed earlier, so that after performing the integration of the μ-integral, one obtains

$$\langle\mathcal{N}|\Psi(\tau)\rangle e^{+i\tau/FR}=-\frac{1}{2\pi i}\int_C d\xi\,e^{-i\xi\tau}\times\left\{\frac{1}{\xi}-\frac{B}{\xi^2}\left[\frac{1}{3}+\frac{F\xi}{2}+F^2\xi^2+F^3\xi^3\ln\left(1-\frac{1}{F\xi}\right)\right]\right\}. \tag{A8}$$

This result readily reduces to

$$\langle N|\Psi(\tau)\rangle e^{+i\tau/FR}$$

$$=1+\frac{i\tau B}{3}-\frac{FB}{2}+F^3B\int_C d\xi\, e^{-i\xi\tau}\,\xi\ln\left(1-\frac{1}{F\xi}\right). \quad \text{(A9)}$$

The contour illustrated in Fig. 2 is suitable for evaluating the integral appearing in this expression. One finds that

$$\langle N|\Psi(\tau)\rangle e^{+i\tau/FR}$$

$$=1+\tfrac{1}{3}i\tau B-\tfrac{1}{2}FB+F^3B[e^{-i\tau/F}(1/\tau^2+i/\tau F)-1/\tau^2], \quad \text{(A10)}$$

and the corresponding diagonal element of the density matrix is given by

$$\rho(\tau)=1-FB+\tfrac{1}{4}F^2B^2+\tfrac{1}{3}2F^2B^2\cos(\tau/F)+\tfrac{1}{9}\tau^2B^2$$
$$+(1/\tau)(2-\tfrac{5}{3}FB)F^2B\sin(\tau/F)$$
$$+(1/\tau^2)\{(F^4B^2-2F^3B)[1-\cos(\tau/F)]$$
$$-2F^6B^2\cos(\tau/F)+F^4B^2\sin^2(\tau/F)\}$$
$$-(1/\tau^3)[2F^5B^2\sin(\tau/F)]$$
$$+(1/\tau^4)[2-\cos^2(\tau/F)]F^6B^2. \quad \text{(A11)}$$

This result is the weak-coupling, quantum-mechanical solution for the electron-oscillators system determined using a master-equation approach, and it is noted at once that the expression contains secular terms $O(\tau^2)$.

APPENDIX B: WEAK-COUPLING SOLUTION TO THE PRIGOGINE-RESIBOIS MASTER EQUATION

An alternate method for studying the dynamics of a system possessing many degrees of freedom is to use the master equation derived by Prigogine and Rêsibois[20]; this master equation is obtained from the Liouville—von Neumann equation for the density matrix ρ. As it happens, the weak-coupling solution to the Prigogine—Rêsibois master equation for the electron-oscillators system can be determined quite easily, since many of the formal results obtained in our earlier work, in particular I and VI, can be taken over directly and used within the context of the present calculation; the reader is referred to these papers for a complete statement of the formal results, notation, etc. In order that the electron-oscillators model may correspond as closely as possible to the two models studied in I and VI, it is necessary to extract the factor $\sqrt{2}$ from the bath operators, so that

$$H_{\rm bath}=\sum_\lambda \tfrac{1}{2}\hbar\omega(q_\lambda)[a^*_{q_\lambda}a_{q_\lambda}+1] \quad \text{(B1)}$$

and

$$\langle n_\lambda|a_{q_\lambda}|m_\lambda\rangle=\sqrt{2(n_\lambda+1)}\,\delta(m_\lambda-n_\lambda-1). \quad \text{(B2)}$$

It is also convenient to write the interaction term of the Hamiltonian in the equivalent form

$$V_{\rm int}=\sum_\lambda [h^*_{q_\lambda}\alpha^*_{k-q_\lambda}\alpha_k a^*_{q_\lambda}+h_{q_\lambda}\alpha^*_k\alpha_{k-q_\lambda}a_{q_\lambda}], \quad \text{(B3)}$$

where

$$|h_{q_\lambda}|^2=2(8\pi^3/\Omega)\gamma^2\,\omega(q_\lambda). \quad \text{(B4)}$$

Notice that the summation over electron states has been suppressed, in accordance with the specification of the initial condition $|N\rangle$; this choice of initial condition also results in the vanishing of the destruction operator.

To determine the time evolution of the system, it is necessary to evaluate the matrix element $\langle\nu|R^o(z)|\nu\rangle$:

$$\langle\nu|R^o(z)|\nu\rangle=-i[(1/\hbar)\langle N+\tfrac{1}{2}\nu|H_0|N+\tfrac{1}{2}\nu\rangle$$
$$-(1/\hbar)\langle N-\tfrac{1}{2}\nu|H_0|N-\tfrac{1}{2}\nu\rangle-z]^{-1}$$
$$=-i((1/\hbar)\{\epsilon(k)[\delta(N_p+\tfrac{1}{2}\nu_p-k)-\delta(N_p-\tfrac{1}{2}\nu_p-k)]$$
$$+\sum_\lambda\hbar\omega_\lambda\nu_\lambda\}-z)^{-1}$$
$$\equiv-i(E(N_p+\tfrac{1}{2}\nu_p)-E(N_p-\tfrac{1}{2}\nu_p)+\sum_\lambda\omega_\lambda\nu_\lambda-z)^{-1}. \quad \text{(B5)}$$

In addition, it is necessary to evaluate $V_{\nu-\nu'}(N)$; here one finds

$$V_{\nu-\nu'}(N)=\langle N+\tfrac{1}{2}(\nu-\nu')|V|N-\tfrac{1}{2}(\nu-\nu')\rangle$$
$$=\sum_\lambda\delta(N_p+\tfrac{1}{2}(\nu_p-\nu_{p'})-n_{k-q_\lambda})\,\delta(N_p-\tfrac{1}{2}(\nu_p-\nu_{p'})-n_k)$$
$$\times h^*_{q_\lambda}\sqrt{2N_\lambda+1}\,\delta(\nu_\lambda-\nu'_\lambda-1)$$
$$+\sum_\lambda\delta(N_p+\tfrac{1}{2}(\nu_p-\nu_{p'})-n_k)\,\delta(N_p-\tfrac{1}{2}(\nu_p-\nu_{p'})-n_{k-q_\lambda})$$
$$\times h_{q_\lambda}\sqrt{2N_\lambda+1}\,\delta(\nu_{\lambda'}-\nu_\lambda-1). \quad \text{(B6)}$$

This equation may be written in the equivalent form

$$V_{\nu-\nu'}(N)=\sum_\lambda\delta(N_p-\tfrac{1}{2}(n_{k-q_\lambda}+n_k))\,\delta(\nu_p-\nu_{p'}+n_k-n_{k-q_\lambda})$$
$$\times h^*_{q_\lambda}\sqrt{2N_\lambda+1}\,\delta(\nu_\lambda-\nu_{\lambda'}-1)$$
$$+\sum_\lambda\delta(N_p-\tfrac{1}{2}(n_{k-q_\lambda}+n_k))\,\delta(\nu_p-\nu_{p'}-n_k+n_{k-q_\lambda})$$
$$\times h_{q_\lambda}\sqrt{2N_\lambda+1}\,\delta(\nu_{\lambda'}-\nu_\lambda-1). \quad \text{(B7)}$$

It is most instructive to compare Eq. (B7) with Eq. (33) of I. It is readily apparent, upon comparison of the initial and final states chosen for the electron-oscillators system with the corresponding states chosen for the model discussed in I, that Eq. (B7) is formally equivalent to Eq. (I-33). One also sees that the expression for $\langle\nu|R^o(z)|\nu\rangle$ for the electron-oscillators system, Eq. (B5), is equivalent to the corresponding expression, Eq. (I-32), for the model of I. Consequently, the expression for $\psi(z)$, the Laplace transform of the collision operator, for the electron-oscillators system is formally equivalent to Eq. (I-34). While the two expressions are similar in structure, however, the explicit dependence of $|h_\lambda|^2$ on the wavevector is different for the two models considered. Nonetheless, this kind of formal similarity can be exploited, as was pointed out in VI, where, for one of the models considered in that paper, namely, an effective spin in interaction with the vibrational modes of a lattice, the corresponding expression for was also formally equivalent to the $\psi(z)$ of I. Now, the wave vector dependence of $|h_\lambda|^2$ for the spin-lattice model is linear in k. As it happens, if one adopts the long-wavelength approximation, the *same* dependence of $|h_\lambda|^2$ on the wave vector k is found for the electron-oscillators system. These arguments serve to demonstrate that the

probability that the electron-oscillators system be in the state $|N\rangle$ is exactly equivalent to the probability that the spin-lattice system be in its initial state, the latter probability being given explicitly by Eq. (VI-44). Given this correspondence, we can write down at once the weak-coupling solution to the Prigogine—Résibois master equation for the electron-oscillators system. In particular, the diagonal element of the density matrix which describes the probability that the system be in the state $|N\rangle$ at time τ is given by

$$\rho(\tau)=\frac{\exp[-4+(8R^2+32R+48\ln\mathcal{R})\sigma/\pi]\tau}{1+(2\sigma R^2+8\sigma R-4\sigma+12\sigma\ln\mathcal{R})/\pi}$$
$$-\frac{1}{2\pi}\int_0^{\sigma/\mathcal{R}}\frac{d\xi[6a(\xi)\cos(\tau/\xi)+2q(\xi)\sin(\tau/\xi)]}{\xi^2[9a(\xi)^2+q(\xi)^2]}$$
$$-\frac{1}{2\pi}\int_0^{\sigma/\mathcal{R}}\frac{d\xi[2a(\xi)\cos(\tau/\xi)-2q(\xi)\sin(\tau/\xi)]}{\xi^2[a(\xi)^2+q(\xi)^2]}$$
$$-\frac{1}{2\pi}\int_0^{\sigma/\mathcal{R}}\frac{d\xi\{2[b(\xi)-3a(\xi)]\cos(\tau/\xi)-2q(\xi)\sin(\tau/\xi)\}}{\xi^2\{[b(\xi)-3a(\xi)]^2+q(\xi)^2\}}$$
$$-\frac{1}{2\pi}\int_0^{\sigma/\mathcal{R}}\frac{d\xi\{2[b(\xi)+a(\xi)]\cos(\tau/\xi)+2q(\xi)\sin(\tau/\xi)\}}{\xi^2\{[b(\xi)+a(\xi)]^2+q(\xi)^2\}}. \tag{B8}$$

Here

$$a(\xi)=\frac{6\sigma^2}{\xi^2}+2,$$
$$b(\xi)=\frac{6\sigma}{\xi}-2\left(\frac{\sigma}{\xi}\right)^3,$$
$$q(\xi)=\frac{1}{\xi}\left\{1+\frac{2\sigma R^2}{\pi}+\frac{8\sigma R}{\pi}\right\}+\frac{b(\xi)}{\pi}\ln\left|\frac{\sigma^2-\xi^2R^2}{\sigma^2-\xi^2}\right|$$
$$+\frac{a(\xi)}{\pi}\ln\left|\frac{\sigma-\xi}{\sigma+\xi}\right|.$$

Also,

$$R=\hbar v\mu/\hbar E,$$
$$\mathcal{R}=(v\mu-E)/E,$$

and

$$\hbar E=\epsilon(k)-\epsilon(k-q_\lambda),$$

where in these results we have set

$$\sigma=\frac{4\zeta\hbar EF}{3v^3n}$$

Extensive numerical calculations performed on Eq. (B8) (reported in VI) reveal that $\rho(\tau)$ assumes negative values over certain ranges of τ, even in the limit of very weak coupling.

*This work is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of Notre Dame.

[1]M. Burton, K. Funabashi, R.R. Hentz, P.K. Ludwig, J.L. Magee and A. Mozumder, "Energy Transfer in Radiation Chemistry," in *Transfer and Storage of Energy by Molecules*, edited by G.M. Burnett and A.M. North (Interscience, New York, 1969), Vol. I.

[2]See, for example, M. Burton, Discuss. Faraday Soc. 36, 7 (1963).

[3]See Ref. 1, and the references cited therein.

[4]L. Van Hove, Physica 23, 441 (1957).

[5]L. Van Hove and E. Verboven, Physica 27, 418 (1961).

[6]A. Janner, L. Van Hove, and E. Verboven, Physica 28, 1341 (1962).

[7]E. Verboven and L. Buyst, Physica 29, 653 (1963).

[8]R. Davidson and J.J. Kozak, J. Math. Phys. 11, 189 (1970).

[9]R. Davidson and J.J. Kozak, J. Math. Phys. 11, 1420 (1970).

[10]R. Davidson and J.J. Kozak, J. Math. Phys. 12, 903 (1971).

[11]R. Davidson and J.J. Kozak, J. Math. Phys. 14, 414 (1973).

[12]R. Davidson and J.J. Kozak, J. Math. Phys. 14, 423 (1973).

[13]H.D. Kutz, R. Davidson, and J.J. Kozak, Phys. Rev. B 4, 1471 (1971).

[14]R. Zwanzig, "Statistical Mechanics of Irreversibility," in *Lectures in Theoretical Physics* (Boulder, 1960), edited by W.E. Brittin, B.W. Downs, and J. Downs (Interscience, New York, 1961), Vol. III.

[15]G. Baym, *Lectures on Quantum Mechanics* (Benjamin, New York, 1969).

[16]J.M. Ziman, *Electrons and Phonons* (Oxford U. P., Oxford, 1960).

[17]H. Frohlich, Proc. Roy. Soc. (London) A 215, 291 (1952).

[18]M. Abramowitz and I.A. Stegum, Eds., *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1964).

[19]P.B. Merkel and W.H. Hamill, J. Chem. Phys. 55, 1409 (1971).

[20]I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience, New York, 1962); P. Resibois, "A Perturbative Approach to Irreversible Statistical Mechanics," in *Physics of Many-Particle Systems* (Gordon and Breach, New York, 1966), Vol. I.

Errata: Ether flow through a drainhole: A particle model in general relativity [J. Math. Phys. 14, 104 (1973)]

Homer G. Ellis

Department of Mathematics, University of Colorado, Boulder, Colorado 80302
(Received 3 January 1974)

Page	Column	Line	Should read:
105	2	31	The coframe system $\{\omega^\mu\}$ ...
105	2	32	... $\omega^1 = d\rho - f(\rho)dt$,
106	1	4	... $(r'/r)f[(\omega^2 e_0)e_2 + ...$
106	2	14	... Sec. V it will be
106	2	18	$\left(\frac{d\rho}{dt} - f(\rho)\right)^2 + ...$
107	1	24	... token $\nabla(f^2/2)$ ...
107	2	18	... $\partial/\partial t = \partial/\partial T$.
111	2	31	... $(\sin\vartheta)^2\,\dot{\varphi}^2$. ...
112	1	45	... Eq. (52) implies that $\dot{\rho}$ is bounded. ...
113	1	42	... $(\rho^2 + a^2)$. ...
114	1	21	(i) $-1 \leqslant 2E \leqslant -1 + ...$
116	2	35	... basis $\{(\partial/\partial x^\mu)(P)\}$ of $\mathcal{T}^P$ ...
116	2	46	$de_\mu = \omega_\mu^{\ \kappa} \otimes e_\kappa$ ...
117	1	43	... Preuss. Akad. Wiss. Phys.-Math. Kl. 7 ...
118	2	4	... issue.

Erratum: The evaluation of lattice sums. II. Number theoretic approach [J. Math. Phys. 14, 701 (1973)]

M. Lawrence Glasser

Battelle Memorial Institute, Columbus, Ohio 43201
(Received 11 January 1974)

Equation (10) should read:

$$\sum_{p,q=1}^{\infty} \frac{p^4 + 26p^2q^2 + 25q^4}{(p^4 - 6p^2q^2 + 25q^4)^2} = \frac{\pi^2}{3}\left(G - \frac{13}{50}\frac{\zeta(4)}{\zeta(2)}\right).$$

The right-hand side of Eq. (34) should read:

$$2^s(1-2^s)\zeta(s)\beta(s) + (2^{2s}-1)\zeta(2s) - 2^{2s}[A^2(s) - B^2(s) - \beta(2s-1)]$$