Bifurcation properties of Dicke Hamiltonians*

R. Gilmore

Institute de Physique Théorique, Université de Louvain, Louvain-la-Neuve, B—1348, Belgium

C. M. Bowden

Quantum Physics, Physical Sciences Directorate, Redstone Arsenal, Alabama 35809
(Received 22 March 1976)

A variation in the coupled order parameter treatment of Dicke Hamiltonians in thermodynamic equilibrium is presented. The Hamiltonian is linearized by introducing disposable c-number parameters. These parameters are chosen to minimize the resulting free energy. This requirement leads to a system of coupled nonlinear equations whose bifurcation properties are studied. The solution branches are labelled by the inertia of the free energy stability matrix. We prove that the parameters on the solution branch which provide the global minimum free energy also produce a linearized Hamiltonian thermodynamically equivalent to the original Hamiltonian provided only a finite number of field modes are present. This method is used to discuss the bifurcation and stability properties of the Dicke Hamiltonian with A^2 and counterrotating terms. We also discuss why the phase transition disappears in the presence of external currents or fields. We show how an internal gauge destroying mechanism may lead to the persistence of the phase transition even in the presence of external coupling. The method is used to discuss the phase transitions and multiplicity of ordered state phases in multilevel molecular systems. We also present a simple method for determining whether an external source will or will not destroy a second order phase transition and discuss the conditions under which such models may exhibit first order phase transitions.

1. INTRODUCTION

Recent interest in the equilibrium statistical mechanics of Dicke Hamiltonians[1] has been stimulated by the proof by Hepp and Lieb,[2] of the existence of a second order phase transition for sufficiently large values of the coupling constant λ.

The presence and location of the phase transition is an example of the bifurcation[3] of a nontrival solution of a particular nonlinear equation from its trivial solution. In a bifurcation analysis of Dicke Hamiltonians, $\beta=1/kT$ plays the role of the bifurcation parameter and the gap equations, which determine the critical temperatures, are simply the bifurcation equations.[3]

In the present work, we treat Dicke Hamiltonians by a variation of the coupled order parameter method.[4,5] This intrinsically nonlinear treatment emphasizes the bifurcation properties of nonlinear equations associated with specific model Hamiltonians. This method involves an attempt to find a linear Hamiltonian which is thermodynamically equivalent to the original Hamiltonian. This process is carried out in two steps:

1. The Hamiltonian is linearized by introducing unknown disposable c-number parameters. The intensive free energy, or free energy per particle, is computed for the linearized Hamiltonian, and the disposable c-number parameters are chosen to minimize the free energy. The disposable parameters obey a system of coupled nonlinear equations which always possess one solution, called the thermal or disordered branch. For large enough values of the coupling constants, other solutions, called ordered branches, may be possible below certain critical temperatures. These branches may arise either through bifurcation from the disordered branch or some other ordered branch, or otherwise. The branches are characterized by the inertia of the free energy stability matrix (FESM). Bifurcations and turn-arounds are characterized by the change in sign of at least one eigenvalue of this real symmetric matrix.

2. It is then necessary to determine whether any of the linear Hamiltonians associated with solutions of the nonlinear equations is thermodynamically equivalent to the original Hamiltonian. This can be done by determining the convergence properties of a perturbation series,[4,5] or by a direct estimate of the intensive free energy of the original Hamiltonian. On the globally stable branch, the disposable parameters have a natural interpretation as order parameters for the system.

In Sec. 2 we illustrate step 1 of this method using the original Dicke Hamiltonian.[1] In Sec. 3 this method is applied to the Dicke Hamiltonian "dressed" by counterrotating and A^2 terms.[4,6] In Secs. 4 and 5 we discuss the effect of external sources—classical fields or classical currents—on the bifurcation properties[7] of the Hamiltonian treated in Secs. 2 and 3.

In Sec. 6 we prove that the linear Hamiltonain associated with the global minimum solution of the coupled nonlinear equations is in fact thermodynamically equivalent to the original Hamiltonian, provided there are only a finite number of field modes present. These results are extended in Sec. 7 to a qualitative discussion of the ordered states of multilevel molecular systems interacting with a finite number of modes of the radiation field. In Sec. 8 we give a physical interpretation to the density operators which arise in connection with the Hamiltonians discussed in Secs. 2—5.

2. DICKE HAMILTONIAN

The Dicke Hamiltonian

$$H_{\mathrm{D}}=\omega a^{\dagger}a+\epsilon\sum_{j=1}^{N}\tfrac{1}{2}\,\sigma_j^z+(\lambda/\sqrt{N})\sum_{j=1}^{N}\left(a^{\dagger}\sigma_j^- + a\sigma_j^+\right) \tag{2.1}$$

is linearized by expanding the operators $a^{\#}$, $\sigma_j^{\pm}$ in the interaction term about disposable c-number parameters $\mu\sqrt{N}$ and ν as follows:

$$a=\mu\sqrt{N}+(a-\mu\sqrt{N}),\quad \sigma_j^-=\nu+(\sigma_j^--\nu). \tag{2.2}$$

The factor $\sqrt{N}$ has been introduced for convenience. Then

$$H_D=H_L+H_{BL},$$

where the bilinear term is

$$H_{BL}=(\lambda/\sqrt{N})\sum_{j=1}^{N}\{(a^\dagger-\mu^*\sqrt{N})(\sigma_j^--\nu)+\text{h.c.}\}. \tag{2.3}$$

The linear Hamiltonian is of the form

$$H_L=H_0+H_1+H_2, \tag{2.4}$$

$$H_0=-\lambda\sum_{j=1}^{N}(\mu^*\nu+\mu\nu^*),$$

$$H_1=\omega a^\dagger a+(\lambda/\sqrt{N})\left(a^\dagger\sum_{j=1}^{N}\nu+a\sum_{j=1}^{N}\nu^*\right),$$

$$H_2=\sum_{j=1}^{N}\{\epsilon\tfrac{1}{2}\sigma_j^z+\lambda\mu^*\sigma_j^-+\lambda\mu\sigma_j^+\}. \tag{2.5}$$

The terms H_0, H_1, H_2 commute. As a result, the free energy F_L associated with H_L is the sum of the free energies associated with H_0, H_1, H_2. These free energies are simple to compute since they are analytic continuations of characters of representations of the Lie groups H(4) and SU(2):

$$F_0=H_0,\quad F_1=-(\lambda^2/\omega N\left|\sum_{j=1}^{N}\nu\right|^2+(1/\beta)\ln(1-e^{-\beta\omega}),$$

$$F_2=-(1/\beta)N\ln 2\cosh\beta\theta, \tag{2.6}$$

$$\theta^2=(\epsilon/2)^2+\lambda^2\mu^*\mu. \tag{2.7}$$

The intensive linear free energy is

$$F_L/N=-\lambda(\mu^*\nu+\mu\nu^*)-(\lambda^2/\omega)\nu^*\nu$$
$$\times-(1/\beta)\ln 2\cosh\beta\theta+(1/N\beta)\ln(1-e^{-\beta\omega}). \tag{2.8}$$

Next, F_L/N is minimized by appropriate choice of the parameters μ, ν. A necessary condition is the vanishing of the first derivatives,

$$\frac{\partial}{\partial\nu^*}(F_L/N)=-\lambda\mu-(\lambda^2/\omega)\nu=0, \tag{2.9a}$$

$$\frac{\partial}{\partial\mu^*}(F_L/N)=-\lambda\nu-(\lambda^2\mu/2\theta)\tanh\beta\theta=0. \tag{2.9b}$$

Similar equations relating μ^* and ν^* are easily obtained. The coupled nonlinear equations (2.9) may be treated by eliminating either μ or ν:

$$[\omega-(\lambda^2/2\theta)\tanh\beta\theta]\mu=0. \tag{2.10}$$

This nonlinear equation always has one solution, $\mu=0$, called the disordered branch. A nontrivial solution $\mu\neq0$, $\nu\neq0$ is possible if the implicit equation[8]

$$\omega=(\lambda^2/2\theta)\tanh\beta\theta \tag{2.11}$$

can be solved. This is only possible if $\lambda^2/\epsilon\omega\geq1$. In this case, a nontrivial solution bifurcates from the trivial solution at a critical temperature determined by the bifurcation equation

$$\omega=(\lambda^2/\epsilon)\tanh\tfrac{1}{2}\beta_c\epsilon. \tag{2.12}$$

For $T<T_c$, $\mu\neq0$ is determined uniquely up to a phase factor by (2.11).

The coupled equations (2.9) are necessary but not sufficient to determine the minimum value of F_L/N. It is also necessary to examine the free energy stability matrix in a Cartesian coordinate system. Since μ and ν are not independent, it is useful to express F_L/N as a function of μ,μ^* or ν,ν^*. Using (2.9a) to eliminate ν,ν^*, and writing $\mu=x+iy=x_1+ix_2$, we have in the thermodynamic limit

$$f(x,y)=\lim_{N\to\infty}F_L/N=\omega(x^2+y^2)-(1/\beta)\ln 2\cosh\beta\theta. \tag{2.13}$$

The free energy stability matrix $f_{ij}=\partial^2 f/\partial x_i\partial x_j$ may then be evaluated on each branch of the nonlinear equations (2.9). The inertia of this matrix then characterizes the stability properties of the various branches.

On the thermal disordered branch, the inertia is $(++)$ for $T>T_c$, (00) for $T=T_c$, and $(--)$ for $T<T_c$. On the ordered branch, it is $(+0)$ for $T<T_c$. Since $\partial^2f/\partial x_i\partial x_j$ is positive definite for $T>T_c$ and positive semidefinite[9] for $T<T_c$, there is a local and global minimum on the disordered branch above the critical temperature and a nonlocal minimum on the ordered branch below the critical temperature, respectively.

For $T<T_c$, the potential $f(x,y)$ has the form of Fig. (63) of Ref. 10, rotated around the symmetry axis. The free energy assumes its minimum value on the circle whose radius $|\mu|$ is determined by (2.11) but whose azimuth is undetermined. This SO(2) gauge invariance is responsible for the fact that the free energy stability matrix is not positive definite. Evaluated on the minimum circle, the radial eigenvalue is positive and the azimuthal eigenvalue is zero.

3. A^2 AND COUNTERROTATING TERMS

Next, we consider the Dicke Hamiltonian (2.1) modified by the inclusion of the A^2 and counterrotating terms.[4,6] This "dressed" Dicke Hamiltonian is

$$H_{DD}=\tfrac{1}{2}\omega(a^\dagger a+aa^\dagger)+\kappa(a^\dagger+a)^2+\epsilon\sum_{j=1}^{N}\tfrac{1}{2}\sigma_j^z$$
$$+(\lambda/\sqrt{N}\sum_{j=1}^{N}(a^\dagger\sigma_j^-+a\sigma_j^++r^*a^\dagger\sigma_j^++ra\sigma_j^-). \tag{3.1}$$

By making a canonical transformation

$$\begin{bmatrix} b\\ b^\dagger\end{bmatrix}=\begin{bmatrix}\cosh\gamma & \exp(i\varphi)\sinh\gamma\\ \exp(-i\varphi)\sinh\gamma & \cosh\gamma\end{bmatrix}\begin{bmatrix}a\\ a^\dagger\end{bmatrix}, \tag{3.2}$$

it is possible to eliminate the double frequency terms a^2, $a^{\dagger2}$ or the counterrotating terms $a^\dagger\sigma_j^+$, $a\sigma_j^-$. In Table I we list the values of $\tanh\gamma$ which cause this elimination, as well as the relevant parameters of the resulting Hamiltonian. These calculations have been carried out assuming r real, $-1<r<+1$.

It is clear from Table I that for

$$s\equiv1+4\,\frac{\kappa}{\omega}-\left(\frac{1+r}{1-r}\right)^2=0 \tag{3.3}$$

the Hamiltonian (3.1) can be reduced to (2.1) with renormalized parameters. As a result, we should expect

TABLE I. Renormalized parameters for H_{DD} in (3.1) under a canonical transformation (3.2) which eliminates either the A^2 or the counterrotating terms. In this table, $f=1+4(\kappa/\omega)$.

$\tanh\gamma$	ω'	κ	λ'	r'
$-\dfrac{f^{1/2}-1}{f^{1/2}+1}$	$f^{1/2}\omega$	0	$\frac{\lambda}{2}\{(1+r)f^{-1/4}+(1-r)f^{+1/4}\}$	$\dfrac{1-\left(\frac{1-r}{1+r}\right)f^{1/2}}{1+\left(\frac{1-r}{1+r}\right)f^{1/2}}$
$-r$	$\left(\frac{1+r}{1-r}\right)\omega$	$\kappa\left(\frac{1-r}{1+r}\right)-\frac{\omega r}{1-r^2}$	$\lambda(1-r^2)^{1/2}$	0

the surface $s(\omega,\kappa,r)=0$ to play the role of a separatrix for the branches of the nonlinear order parameter equations associated with (3.1).

Following the procedure described in Sec. 2, it is possible to determine H_L, F_L/N, and the coupled order parameter equations associated with (3.1). These equations are

$$\begin{bmatrix}\omega+2\kappa & 2\kappa\\ 2\kappa & \omega+2\kappa\end{bmatrix}\begin{bmatrix}\mu\\ \mu^*\end{bmatrix}+\lambda\begin{bmatrix}1 & r^*\\ r & 1\end{bmatrix}\begin{bmatrix}\nu\\ \nu^*\end{bmatrix}=0, \tag{3.4a}$$

$$\epsilon\begin{bmatrix}\nu\\ \nu^*\end{bmatrix}+\frac{\lambda\epsilon}{2\theta}\tanh\beta\theta\begin{bmatrix}1 & r^*\\ r & 1\end{bmatrix}\begin{bmatrix}\mu\\ \mu^*\end{bmatrix}=0, \tag{3.4b}$$

$$\theta^2=(\epsilon/2)^2+\lambda^2|\mu+r^*\mu^*|^2. \tag{3.5}$$

For simplicity, we assume r is real. These equations may be diagonalized by performing a similarity transformation with $S=(I_2-i\sigma_y)/\sqrt{2}$. Eliminating ν,ν^* from the resulting equations results in the matrix equation

$$\left\{\begin{bmatrix}\omega+4\kappa & \\ & \omega\end{bmatrix}-\frac{\lambda^2}{2\theta}\tanh\beta\theta\begin{bmatrix}1+r & \\ & 1-r\end{bmatrix}^2\right\}\begin{bmatrix}x\\ y\end{bmatrix}=0, \tag{3.6}$$

$$\theta^2=(\epsilon/2)^2+\lambda^2(1+r)^2x^2+\lambda^2(1-r)^2y^2, \tag{3.7}$$

where $\mu=x+iy$.

Equations (3.6) always possess the trivial solution $x=0$, $y=0$. This is the disordered branch.

Nontrivial solutions $(x\neq 0, y=0)$ or $(x=0, y\neq 0)$ are also possible if $(\omega+4\kappa)<\lambda^2(1+r)^2/\epsilon$ or $\omega<\lambda^2(1-r)^2/\epsilon$. When such solutions exist, they are called the real and imaginary branches, respectively. On the real branch, $x\neq 0$ is determined uniquely up to sign by

$$1=\frac{\lambda^2(1+r)^2}{2\theta_R(\omega+4\kappa)}\tanh\beta\theta_R \tag{3.8}$$

$$\theta_R^2=(\epsilon/2)^2+\lambda^2(1+r)^2x^2. \tag{3.9}$$

The real branch bifurcates from the disordered branch at a critical temperature T_R determined by the gap equation

$$1=\frac{\lambda^2(1+r)^2}{\epsilon(\omega+4\kappa)}\tanh\tfrac{1}{2}\beta_R\epsilon. \tag{3.10}$$

The imaginary branch bifurcates from the disordered branch at T_I determined from

$$1=\frac{\lambda^2(1-r)^2}{\epsilon\omega}\tanh\tfrac{1}{2}\beta_I\epsilon. \tag{3.11}$$

For $T<T_I$, $y\neq 0$ is determined uniquely up to sign by

$$1=\frac{\lambda^2(1-r)^2}{2\theta_I\omega}\tanh\beta\theta_I, \tag{3.12}$$

$$\theta_I^2=(\epsilon/2)^2+\lambda^2(1-r)^2y^2. \tag{3.13}$$

If there is a secondary bifurcation[3] from either of the primary branches onto a secondary branch $(x\neq 0,\ y\neq 0)$, then both diagonal matrix elements in (3.6) must simultaneously vanish. This is possible only if $s=0$ in (3.3). However, this is precisely the condition for the Hamiltonian (3.1) to be equivalent to the original Dicke Hamiltonian (2.1) under a canonical transformation. Since the Dicke Hamiltonian (2.1) exhibits a doubly degenerate bifurcation at the critical temperature T_c, there is no secondary bifurcation from either of the ordered branches of (3.6). This is consistent with the results of Ref. 5. Since both Lie algebras SU(2) and h(4) have maximal roots of level 1,[11] there may be several primary branches but no secondary branches.

The intensive free energy determined from the linearized form of H_{DD} is, in the thermodynamic limit

$$f(x,y)=(\omega+4\kappa)x^2+\omega y^2-(1/\beta)\ln 2\cosh\beta\theta. \tag{3.14}$$

The free energy stability matrix can be computed from (3.14) and the inertia evaluated on each of the branches. The results are

1. Disordered branch:

$$\left[\operatorname{sgn}\left(1-\frac{\lambda^2(1+r)^2}{\epsilon(\omega+4\kappa)}\tanh\tfrac{1}{2}\beta\epsilon\right),\ \operatorname{sgn}\left(1-\frac{\lambda^2(1-r)^2}{\epsilon\omega}\tanh\tfrac{1}{2}\beta\epsilon\right)\right].$$

2. Real branch, if it exists:

$[+1,\ \operatorname{sgn}(-s)]$.

3. Imaginary branch, if it exists:

$[\operatorname{sgn}(+s),\ +1]$.

The inertia of the free energy stability matrix is constant along each of the ordered branches. On the disordered branch, one eigenvalue changes sign at each bifurcation point, even in the degenerate case $s=0$.

If neither of the two equations (3.8) and (3.12) can be solved, there is only the trivial solution to (3.6), and the inertia is $(++)$ on the disordered branch. If only one of the two equations can be solved, the inertia is $(++)$ on the nonzero solution branch, $(++)$ on the disordered branch for $T>T_c$, and $(+-)$ for $T<T_c$. On both branches it is $(+0)$ at the bifurcation point $T=T_c$.

TABLE II. Intertia of the free energy stability matrix obtained from $f(x,y)$ in (3.14) for different values of s in (3.3). For $s<0$, the real branch is the high temperature branch. For $s>0$, the imaginary branch bifurcates at the higher temperature and is stable. On the separatrix $s=0$, the free energy is invariant under the gauge group U(1) and the FESM is positive semidefinite.

	$s<0$	$s=0$	$s>0$
Real branch	(++)	(+ 0)	(+ −)
Imaginary branch	(+ −)	(+ 0)	(++)

If both (3.8) and (3.12) can be solved, the situation is more complicated, depending on the sign of the difference s in (3.3). The inertia is (+ +) on the disordered branch above both bifurcations, (− −) below both, and (+ −) between. One eigenvalue changes sign at each bifurcation. The inertia is (+ +) on the high temperature ordered branch and (+ −) on the low temperature ordered branch. As $s\to 0$, the bifurcation points coalesce, and the discussion reduces to that of Sec. 2. The results are summarized in Table II. These results are illustrated in Fig. 1 for $s<0$.

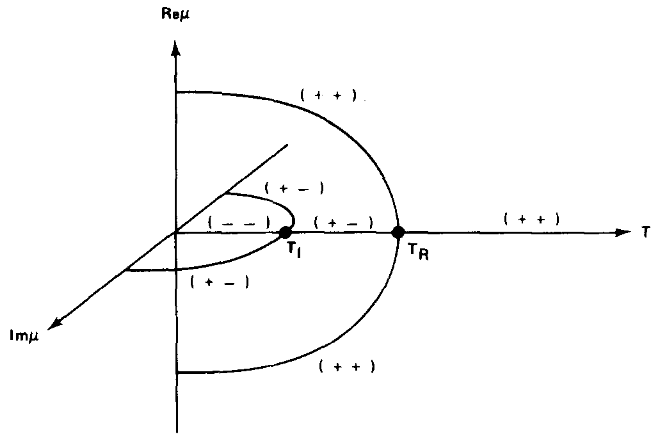

FIG. 1. Solutions of the nonlinear equation (3.6). For $\lambda^2(1+r)^2/\epsilon(\omega+4\kappa)>\lambda^2(1-r)^2/\epsilon\omega>1$, $s<0$, the real branch bifurcates at a higher temperature than the imaginary branch and is globally stable. Each branch is labelled by the inertia of the free energy stability matrix. The two critical points are shown by dots.

4. EXTERNAL SOURCES

We now consider the changes brought about by the addition of classical external sources. The Hamiltonian we consider is

$$H=H_D+H_{\mathrm{ext}}. \tag{4.1}$$

A classical current will produce a coupling to the electromagnetic field of the form

$$H_{\mathrm{C1.C.}}=h\sqrt{N}\,a^\dagger+h^*\sqrt{N}\,a. \tag{4.2}$$

A classical field will produce a coupling to the atomic system of the form

$$H_{\mathrm{C1.F.}}=\sum_{j=1}^{N}\left(\lambda'\sigma_j^+ +\lambda'^*\sigma_j^-\right). \tag{4.3}$$

The Hamiltonian $H_{\mathrm{DD}}+H_{\mathrm{C1.C.}}$ is equivalent to $H_{\mathrm{DD}}+H_{\mathrm{C1.F.}}$ under a canonical transformation provided the parameters h in (4.2), λ' in (4.3), and ω, κ, r in (3.1) are related by

$$\begin{pmatrix}\lambda'\\ \lambda'^*\end{pmatrix}=-\lambda\begin{pmatrix}1 & r^*\\ r & 1\end{pmatrix}\begin{pmatrix}\omega+2\kappa & 2\kappa\\ 2\kappa & \omega+2\kappa\end{pmatrix}^{-1}\begin{pmatrix}h\\ h^*\end{pmatrix}. \tag{4.4}$$

For this reason, it is sufficient to study the effects of only classical external currents (4.2) in this and the following section.

The coupled nonlinear equations describing the linearized form of (4.1) = (2.1) + (4.2) are

$$\omega\mu+\lambda\nu=-h, \tag{4.5a}$$

$$\nu+(\lambda/2\theta)\tanh\beta\theta=0, \tag{4.5b}$$

where θ is given by (2.7). Eliminating ν leads to the inhomogeneous equation

$$[\omega-(\lambda^2/2\theta)\tanh\beta\theta]\mu=-h. \tag{4.6}$$

Since $\mu\neq 0$, this equation can be rewritten

$$(\lambda^2/2\theta)\tanh\beta\theta=\omega+h/\mu. \tag{4.7}$$

This equation always has one solution, for which μ and $-h$ have the same phase, and for which $\omega\mu/h\leq -1$. For $\lambda^2/\epsilon\omega$ sufficiently large, two additional solutions are possible, but these no longer bifurcate from the thermal branch. The presence of the inhomogeneous term in (4.5a) "unhinges" the bifurcation, as can be seen by inspecting (4.7). The solution branches for the inhomogeneous nonlinear equation (4.6) are shown in Fig. 2.

The intensive free energy obtained from the linearized form of (4.1) is

$$f(\mu,\mu^*)=\omega\mu^*\mu+\mu^*h+\mu h^*-(1/\beta)\ln 2\cosh\beta\theta.$$

This differs from (2.13) by only linear terms. As a result, the matrix elements of the FESM obtained from (2.13) and (4.8) are identical. However, these matrices must be evaluated on the solutions of (2.10) and (4.6), respectively. The results are shown in Fig. 2.

The inclusion of external terms destroys the gauge invariance of the model. As a result, zero is no longer

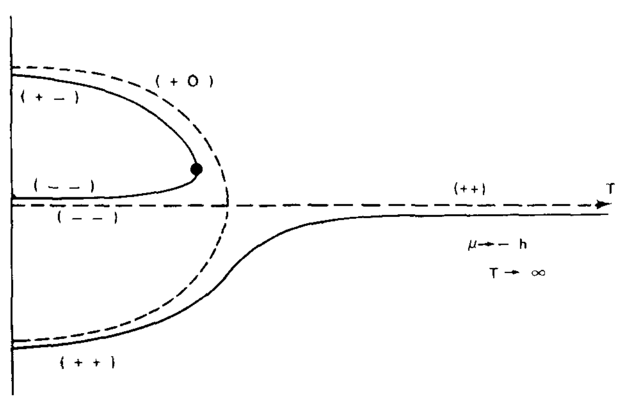

FIG. 2. In the presence of external sources, additional low temperature solutions to (4.6) do not bifurcate from the thermal branch. Solutions to the corresponding homogeneous equation are shown by a dotted line. When one of the eigenvalues of the FESM passes through zero (dot), there is a turn-around on the low temperature branch.

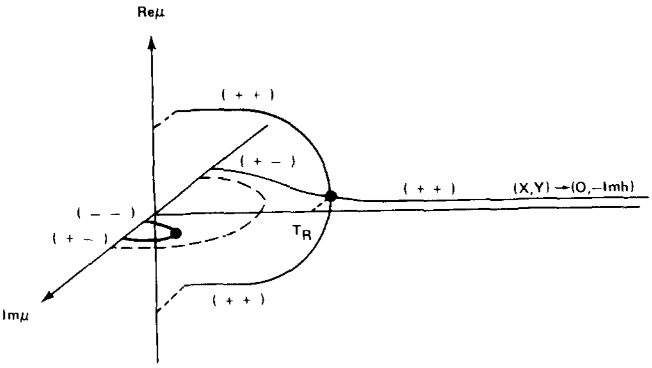

FIG. 3. Solutions of the nonlinear equations (5.2) with $\mathrm{Re}h=0$. Bifurcations of ordered branches from the disordered branch are still possible in the presence of external fields provided there is an internal mechanism to destroy gauge invariance. The two critical points (dots) lead to bifurcation, since (5.2a) is homogeneous, and to turn around, since (5.2b) is inhomogeneous.

an eigenvalue of the FESM. There is no bifurcation, and no phase transition.[12] The change in sign of one root of the stability matrix is associated with a vertical tangent.

In general,[5] a sign change in the inertia of the stability matrix is associated with bifurcation (homogeneous equation, cf. Fig. 1) or with turn around of a branch (inhomogeneous equations, cf. Fig. 2).

5. RIGID HAMILTONIANS WITH EXTERNAL COUPLING

Finally, we consider the Hamiltonian (3.1) in the presence of external sources (4.2)

$$H=H_{\mathrm{DD}}+H_{\mathrm{C1.C}}. \tag{5.1}$$

The nonlinear equations arising from the coupled order parameter treatment differ from (3.4) only by the addition of the matrix col (h,h^*) to the left-hand side of (3.4a). After diagonalization, the nonlinear equations for $\mu=x+iy$ are

$$\left\{(\omega+4\kappa)-\frac{\lambda^2(1+r)^2}{2\theta}\tanh\beta\theta\right\}x=-\mathrm{Re}h, \tag{5.2a}$$

$$\left\{\omega-\frac{\lambda^2(1-r)^2}{2\theta}\tanh\beta\theta\right\}y=-\mathrm{Im}h, \tag{5.2b}$$

where $h=\mathrm{Re}h+i\mathrm{Im}h$ and θ is given by (3.7). The intensive free energy obtained from the linearized form of (5.1) differs from (3.14) by linear terms,

$$f(x,y)=(\omega+4\kappa)x^2+\omega y^2+2x\mathrm{Re}h + 2y\mathrm{Im}h-(1/\beta)\ln 2\cosh\beta\theta. \tag{5.3}$$

Equations (5.2) always possess one soluction (disordered branch) with asymptotic limits $\lim_{T\to\infty}\mu=-h$, and for which $\omega x/\mathrm{Re}h\leqslant -1$, $\omega y/\mathrm{Im}h\leqslant -1$. The presence of the inhomogeneous term $h\neq 0$ destroys at least one bifurcation but need not destroy both. If $\mathrm{Re}h=0$, (5.2a) may possess a nonzero bifurcating solution which we call loosely the "real ordered branch." See Fig. 3. Similarly, if $\mathrm{Im}h=0$, (5.2b) may possess a nontrivial bifurcating solution called the "imagninary ordered branch."

For T sufficiently large, the FESM has inertia $(++)$ on the disordered branch. The inertia along this branch can change only at a bifurcation, since there are no turn arounds. If there are no bifurcations from the disordered branch, there will be no phase transition. If an ordered branch does bifurcate from the disordered branch, the inertia on the disordered branch is $(+-)$ for $T<T_c$ and is $(++)$ on the ordered branch. There is then a phase transition.

We illustrate a situation now for which $h\neq 0$ but there is a phase transition. We choose $\mathrm{Re}h=0$, $s<0$ (3.3). The real ordered branch obeys a homogeneous equation, and is in fact the nonzero solution to (3.8). On this branch, the value of y is constant and uniquely determined by the relation

$$\left\{\omega-(\omega+4\kappa)\left(\frac{1-r}{1+r}\right)^2\right\}y=-\mathrm{Im}h. \tag{5.4}$$

The real branch bifurcates from the disordered branch at a critical temperature determined by the gap equation

$$(\omega+4\kappa)-\frac{\lambda^2(1+r)^2}{\epsilon_{\mathrm{eff}}}\tanh\tfrac{1}{2}\beta_R\epsilon_{\mathrm{ff}}=0, \tag{5.5}$$

$$\left(\frac{\epsilon_{\mathrm{eff}}}{2}\right)^2=\left(\frac{\epsilon}{2}\right)^2+\lambda^2(1-r)^2(\mathrm{Im}h)^2 \times\omega^2\left[1-\left(1+4\frac{\kappa}{\omega}\right)\left(\frac{1-r}{1+r}\right)^2\right]^2. \tag{5.6}$$

For $s=0$, there is no bifurcation (Sec. 4), and for $s>0$, if there is a bifurcation, it occurs on one of the disconnected imaginary solutions and is never globally stable.

6. THERMODYNAMIC LIMIT

In Secs. 2—5 we have been concerned entirely with the problem of determining linearized Hamiltonians which may be thermodynamically equivalent to the original Hamiltonian. In this section we prove that the linearized Hamiltonian associated with the global minimum solution to the coupled nonlinear order parameter equations is in fact thermodynamically equivalent to the original Hamiltonian.

To do this, it will be convenient to derive a useful technical result. It is necessary to compute limits of the form

$$i(\beta)=\lim_{N\to\infty}-\frac{1}{N\beta}\ln I_N(\beta), \tag{6.1}$$

$$I_N(\beta)=\int g(x)\{e^{-\beta\varphi(x)}\}^N dx. \tag{6.2}$$

In this integral, we assume $g(x)$, $\varphi(x)$ are analytic functions of κ real variables, $x=(x_1,\ldots,x_\kappa)$, the integral extends over $\mathcal{R}^\kappa$, $g(x)>0$, $\varphi(x)$ is bounded below with a finite number t of isolated local minima, and $\lim_{|x|\to\infty}\ln|\varphi(x)|/\ln|x|>0$. Although many of these assumptions can be relaxed, it is not necessary to do so for our purposes.

We expect the principal contributions to $I_N(\beta)$ to come from the neighborhoods where $\varphi(x)$ has a local minimum.[13] Assume first that φ has only one local minimum at $x=x^{(1)}$. Expanding $g(x)$, $\varphi(x)$ in a Taylor series results in the expression

$$I_N^{(1)}(\beta)=g(x^{(1)})\exp[-N\beta\varphi(x^{(1)})] \times \int \exp\{-\tfrac{1}{2}(x-x^{(1)})_i\varphi_{ij}(1)(x-x^{(1)})_j\}\,dx + O(N^{-1}), \tag{6.3}$$

$$\varphi_{ij}(1)=\partial^2\varphi(x^{(1)})/\partial x_i\partial x_j. \tag{6.4}$$

The matrix $\varphi_{ij}(1)$ is positive definite and symmetric, since we have assumed that φ has an isolated minimum at $x^{(1)}$. The integral in (6.3) is standard,

$$I_N^{(1)}(\beta)=g(x^{(1)})\exp[-N\beta\varphi(x^{(1)})](2\pi)^{\kappa/2}[\det\varphi_{ij}(1)]^{-1/2}. \tag{6.5}$$

If $\varphi(x)$ has more than one isolated local minimum, then

$$I_N(\beta)=\sum_{b=1}^{t} I_N^{(b)}(\beta)+O(N^{-1}). \tag{6.6}$$

This result can be made rigorous by introducing (δ,ϵ) notation. The proof follows those in Ref. 14 almost *mutatis mutandis*.

The principal contribution to the sum (6.6) will come from the global minimum of φ. If this minimum is t'-fold degenerate, then

$$\begin{aligned} i(\beta)&=\lim_{N\to\infty}-\frac{1}{N\beta}\ln\Big(\sum_{b=1}^{t} I_N^{(b)}(\beta)\Big)\\ &=\varphi(m)-\lim_{N\to\infty}\frac{1}{N\beta}\{\ln t'g(m)+\tfrac{1}{2}\kappa\ln 2\pi-\tfrac{1}{2}\ln\det\varphi_{ij}(m)\}\\ &=\varphi(m). \end{aligned} \tag{6.7}$$

Here m is any one of the t' isolated points at which φ assumes its global minimum value.

The method of maximum contribution described above does not apply directly when φ does not have isolated minima. In this case, $\det\varphi_{ij}=0$. It often happens that a symmetry group G exists which acts transitively[11] on nonisolated minima. In this case, we can decompose the integral appearing in (6.3) into an integral over G and an integral over G orbit representatives $\mathcal{R}^\kappa/G$:

$$\int d^\kappa x=\int_{\mathcal{R}^\kappa/G} d\mu(\mathcal{R}^\kappa/G)\int d\mu(G), \tag{6.8}$$

where $d\mu(\mathcal{R}^\kappa/G)$ and $d\mu(G)$ are the invariant measures on $\mathcal{R}^\kappa/G$ and G, respectively. If G is compact, $\int d\mu(G)$ is finite and does not contribute in the limit (6.7). If φ has a finite number of isolated minima on $\mathcal{R}^\kappa/G$, then Laplace's method can be applied directly, resulting in (6.7).

We now apply Laplace's method to discuss the second step involved in the coupled order parameter method, described in the Introduction. This is done for (5.1); the other Hamiltonians (2.1), (3.1), and (4.1) are special cases of (5.1). The approach used is that of Wang and Hioe.[15]

The free energy for (5.1) is determined by

$$e^{-\beta F}=\mathrm{Tr}\,e^{-\beta H}. \tag{6.9}$$

The trace over the field states is taken conveniently by introducing the Glauber coherent state representation,[16]

$$\mathrm{Tr}\,O=\sum_{n=0}^{\infty}\langle n|O|n\rangle=\int\frac{d^2\alpha}{\pi}\langle\alpha|O|\alpha\rangle \tag{6.10}$$

In this representation the field operators are replaced, to an adequate approximation, by the complex numbers α,α^*. The trace over the 2^N atomic basis states is straightforward. Writing $\alpha=\sqrt{N}(x+iy)$, the integral resulting from (6.9) in the representation (6.10) has the form given in (6.1), with $g(x,y)=N/\pi$, and to $O(N^{-1})$,

$$\varphi(x,y)=(\omega+4\kappa)x^2+\omega y^2+2x\,\mathrm{Re}\,h + 2y\,\mathrm{Im}\,h-(1/\beta)\ln 2\cosh\beta\theta, \tag{6.11}$$

where θ is given by (3.7). This is identical to (5.3). Therefore,

$$\lim_{N\to\infty}F/N=\varphi(m)=f(m)=\lim_{N\to\infty}F_{\mathrm{L}}/N. \tag{6.12}$$

The first equality results from application of the method of maximum contribution, the second from comparison of (6.11) with (5.3), the third by direct computation from the linearized Hamiltonian H_{L} evaluated with order parameters on the global minimum branch of the coupled nonlinear order parameter equations.

As a result, H_{L} is thermodynamically equivalent to H, when H_{L} is obtained from H by the coupled order parameter method using order parameters on the global minimum branch of the coupled nonlinear equations. No modifications are necessary in the case of gauge invariant Hamiltonians and free energies.

7. APPLICATION TO r-LEVEL SYSTEMS

The results of the preceding sections may be applied to more complicated Hamiltonians than those discussed in Secs. 2—5. In this section we extend the coupled order parameter treatment to multimode systems and to multilevel systems.[18]

We consider first an ensemble of N identical 2-level systems interacting with n (finite) modes of the field, and described by the Hamiltonian

$$H=\sum_{i=1}^{n}\omega_i a_i^\dagger a_i+\epsilon\sum_{j=1}^{N}\tfrac{1}{2}\sigma_j^z+(1/\sqrt{N})\sum_{i=1}^{n}\sum_{j=1}^{N}(\lambda_i^* a_i^\dagger\sigma_j^-+\lambda_i a_i\sigma_j^+). \tag{7.1}$$

With more than one mode present, it is no longer possible to choose the relative phases of the ground and excited state wavefunctions so that the coupling constants λ_i are all real.

The Hamiltonian (7.1) is linearized by making the ansatz (2.2), with one disposable c-number parameter for each field mode $[a_i\to\mu_i\sqrt{N}+(a_i-\mu_i\sqrt{N})]$. The resulting coupled equations are

$$\omega_i\mu_i+\lambda_i^*\nu=0,\quad i=1,2,\ldots,n, \tag{7.2a}$$

$$\nu+\frac{1}{2\theta}\Big(\sum_{i=1}^{n}\lambda_i\mu_i\Big)\tanh\beta\theta=0, \tag{7.2b}$$

$$\theta^2=\Big(\frac{\epsilon}{2}\Big)^2+\sum_{i=1}^{n}|\lambda_i\mu_i|^2=\Big(\frac{\epsilon}{2}\Big)^2+|\nu|^2\Big(\sum_{i=1}^{n}\frac{|\lambda_i|^2}{\omega_i}\Big). \tag{7.3}$$

It is convenient to eliminate the parameters μ_i using (7.2a) to obtain a nonlinear equation for ν:

$$\{1-(\Lambda^2/2\theta)\tanh\beta\theta\}\nu=0, \tag{7.4}$$

where

$$\Lambda^2 \equiv \sum_{i=1}^{n}\frac{|\lambda_i|^2}{\omega_i}. \tag{7.5}$$

The free energy in the thermodynamic limit is

$$f(\nu,\nu^*)=\Lambda^2|\nu|^2-(1/\beta)\ln 2\cosh\beta\theta, \tag{7.6}$$

This free energy is to be evaluated on the global minimum branch of (7.2). The results obtained for the multimode case are identical to those obtained in Sec. 2 for the single mode case under the identification $\lambda\to\Lambda$. The Hamiltonians of Secs. 3—5 may be treated similarly, with similar results.

We consider next an ensemble of N identical r-level atoms interacting with $\binom{r}{2}$ modes of the radiation field, with one mode connecting each pair of levels. The Hamiltonian describing this coupled system is

$$H=\sum_{1\le i<j}^{r}\omega_{ji}a_{ji}^{\dagger}a_{ji}+\sum_{l=1}^{N}\sum_{i=1}^{r}\epsilon_i H_i^{(l)} + (1/\sqrt{N})\sum_{l=1}^{N}\sum_{1\le i<j}^{r}(\lambda_{ji}a_{ji}^{\dagger}E_{ij}^{(l)}+\lambda_{ji}^{*}a_{ji}E_{ji}^{(l)}). \tag{7.7}$$

Here the field mode operators $a_{ji}^{\#}$ obey

$$a_{ji}=a_{ij}^{\dagger}, \tag{7.8}$$

$$[a_{j'i'},a_{ji}^{\dagger}]=\delta_{j'j}\delta_{i'i}. \tag{7.9}$$

The operators $H_i^{(l)}$, $E_{ji}^{(l)}$ describing atom l obey independent $u(r)$ commutation relations.[11,17] The operator $E_{ji}^{(l)}$ describes transitions from state i to state j in atom l. The ϵ_i describe the internal levels of each molecule, and ω_{ji} the energy of the photon field mode connecting levels i and j. It is sometimes convenient to assume $\epsilon_1\le\epsilon_2\le\cdots\le\epsilon_r$ and $\omega_{ji}\approx\epsilon_j-\epsilon_i$, although we will not make these assumptions here.

The Hamiltonian (7.7) can be treated by the coupled order parameter method, making the ansatz

$$a_{ji}=\mu_{ji}\sqrt{N}+(a_{ji}-\mu_{ji}\sqrt{N}), \tag{7.10a}$$

$$E_{ij}^{(l)}=\nu_{ij}+(E_{ij}^{(l)}-\nu_{ij}). \tag{7.10b}$$

The nonlinear equations relating μ_{ji} and ν_{ij} are

$$\omega_{ji}\mu_{ji}+\lambda_{ji}\nu_{ij}=0,\quad 1\le i<j\le r, \tag{7.11a}$$

$$(\epsilon_i-\epsilon_j)\nu_{ij}+\lambda_{ji}^{*}\mu_{ji}\langle H_j-H_i\rangle=0,\quad 1\le i<j\le r. \tag{7.11b}$$

The expectation value in (7.11b) is to be taken with respect to the $r\times r$ matrix

$$M(r)=\sum_{i=1}^{r}\epsilon_i H_i+\sum_{1\le i<j}^{r}(\lambda_{ji}\mu_{ji}^{*}E_{ij}+\lambda_{ji}^{*}\mu_{ji}E_{ji}). \tag{7.12}$$

The nonlinear equations (7.11) can be treated either by eliminating the μ_{ij} or the ν_{ij}. Eliminating the latter leads to the $\binom{r}{2}$ equations

$$\left\{\omega_{ji}-\frac{\lambda_{ji}\lambda_{ji}^{*}\langle H_i-H_j\rangle}{\epsilon_j-\epsilon_i}\right\}\mu_{ji}=0. \tag{7.13}$$

The conjugates of these equations need not be considered in the absence of gauge symmetry breaking terms of the form considered in Secs. 3 and 4. The intensive free energy in the thermodynamic limit is

$$f(\mu_{ij},\mu_{ij}^{*})=\sum_{1\le i<j}^{r}\omega_{ji}|\mu_{ji}|^2-(1/\beta)\ln\mathrm{Tr}\exp[-\beta M(r)]. \tag{7.14}$$

This is obtained from the linearized Hamiltonian obtained from (7.7) and evaluated on the global minimum branch of (7.13).

The bifurcation properties of (7.13) are not difficult to discuss in a qualitative way. For sufficiently large temperatures, there is only the trival solution $\mu_{ji}=0$, for all i,j. On this chaotic branch the FESM has inertia $(n_+n_0,n_-)=(r(r-1),0,0)$. As the temperature decreases, two eigenvalues approach zero. At zero, a gauge invariant ordered branch escapes. Just below the bifurcation, the FESM has inertia $(r^2-r-2,0,-2)$ on the thermal branch and $(r^2-r-1,1,0)$ on the ordered branch.

Additional bifurcations may occur from the thermal branch as the temperature decreases. Each time an eigenvalue of the FESM changes sign, it is possible for a solution to escape. Since $M(r)$ (7.12) is a function of $\binom{r}{2}$ complex variables with $(r-1)$ real nonlinear Casimir invariants, there may be as many as $(r-1)^2$ primary bifurcations from the thermal branch onto primary branches for Hamiltonians of the form (7.7). In fact, by direct computation of the fluctuation-transformation matrix,[5] it is possible to verify the existence of no more than four primary bifurcations from the thermal branch for $r=3$. Each of the primary branches may have one secondary bifurcation, but no secondary branch has any bifurcations. In addition, one of the four primary branches has a turn around.

For the r-level system, the primary branches may undergo secondary bifurcations, the secondary branches may have tertiary bifurcations, etc. The process stops on the $(r-1)^{\text{ary}}$ branch. This may happen even if the only nonzero coupling constants are $\lambda_{i,i+1}$ and $\lambda_{i+1,i}$. This occurs because the level of the highest root in the Lie algebra $u(r)$ is $r-1$.

It is generally true that, as a function of decreasing temperature, the first bifurcation to occur from the κ^{ary} branch to a $(\kappa+1)^{\text{ary}}$ branch has a positive semidefinite FESM, $\kappa=0,1,2,\ldots,r-2$. If no other branches have a positive definite or semidefinite FESM, then this is the global minimum solution. It is possible for a turn around to occur on some other branch which changes the FESM from indefinite to positive definite or semidefinite. In this case, there may be an exchange of stability between these branches. Such a stability exchange is associated with a first order phase transition. A model Hamiltonian in which this occurs has been discussed by Thompson.[19]

The linearized Hamiltonian obtained from (7.7) and associated with the globally minimum branch of (7.13) is thermodynamically equivalent to (7.7). If there is only a discrete free energy invariance group, the theorem of Sec. 6 is immediately applicable. If there is a gauge group G, then G is a closed subgroup of $\mathrm{SO}(r^2-r)$ and is therefore compact. Thermodynamic equivalence follows.

Gauge breaking terms of the form considered in Sec. 3 may be added to (7.7). These can be treated following the methods of that section. Their principal effect is to lift degeneracy at bifurcation points.

External sources of the form considered in Sec. 4 may be added to (7.7). Such sources destroy some bifurcations, but not generally all bifurcations. A simple test exists to determine which bifurcations are destroyed and which are not. Let $\mu(T)$ be a nontrivial solution of the homogeneous problem arising by bifurcation at a critical temperature T_c. Let (,) represent the usual Hermitian inner product in $C^{r(r-1)/2}$. Then $|T-T_c|^{1/2}\, d\mu(T_c)/dT$ is finite. If $(h, |T-T_c|^{1/2}\, d\mu(T_c)/dT)=0$, the bifurcation is preserved; if the inner product is nonzero, it is destroyed (cf. Sec. 5).

The Hamiltonian (7.7) may be made more complicated by allowing more than one field mode to couple to each pair of molecular states, as considered in (7.1). The coupled order parameter treatment is then simpler to carry out by eliminating the μ_{ji} instead of the ν_{ij} from equations (7.11). The results are exactly the same as in the single-mode-per-level-pair case (7.7), provided only a finite number of field modes are present and renormalizations of the form (7.5) are carried out for each pair of levels.

8. PHYSICAL INTERPRETATION

The density operator describing a system in thermal equilibrium is $\rho=\exp(-\beta H)/Z$, $Z=\mathrm{Tr}\exp(-\beta H)$. For the computation of intensive parameters, but not for fluctuation quantities, it is sufficient to replace H by H_L for the class of Hamiltonians studied in Secs. 2–5, 7. Here H_L is obtained from H by the coupled order parameter method and evaluated on the global minimum branch. The density operator then factors into a product of density operators $\rho=\rho_F\otimes\rho_A$. The field density operator, ρ_F, describes the equilibrium properties of the field subsystem, ρ_A describes the atomic subsystem properties. For the Hamiltonian (5.1)

$$H_1=\omega a^\dagger a+\kappa(a^\dagger+a)^2+\lambda\sqrt{N}\, a^\dagger(\nu+r^*\nu^*)$$
$$+\lambda\sqrt{N}\, a(\nu^*+r\nu)+h\sqrt{N}\, a^\dagger+h^*\sqrt{N}\, a, \tag{8.1}$$

$$H_2=\sum_{j=1}^{N}\{\epsilon\tfrac{1}{2}\sigma_j^z+\lambda(\mu+r^*\mu^*)\sigma_j^+ +\lambda(\mu^*+r\mu)\sigma_j^-\}. \tag{8.2}$$

The Hamiltonian (8.1) describes a field mode in a statistical superposition of thermal noise and a coherent state produced by classical currents. The classical currents are $h\sqrt{N}$ and the individual atoms, which act classically and drive the field mode through the terms $a^\dagger\lambda\sqrt{N}\,\nu$, etc. To convert (8.1) to a situation describing a field driven by classical sources in the absence of matter, we perform a canonical transformation which eliminates the A^2 term, using the parameters in line 1 of Table I. The coupled nonlinear equations are then solved for μ,ν. The transformed density operator then describes a field mode in vacuum which is a statistical superposition of noise, characterized by noise factor $N=[\exp(\beta\omega')-1]^{-1}$, and a signal, characterized by the coherent state parameter[16] $\alpha=\mu\sqrt{N}$. Here $\mu(\beta)$ is the global minimum solution to the nonlinear equations of the transformed Hamiltonian.

The atomic density operator obtained from (8.2) describes 2-level atoms in a statistical superposition of noise and signal. The thermal part is characterized by weight factors $\exp(\pm\beta E/2)$, where

$$E=2\left[(\epsilon/2)^2+\left|\frac{\epsilon\nu}{\lambda'\langle\sigma^z\rangle}\right|^2\right]^{1/2}$$

is the Stark split level separation caused by a classical external field. The atomic coherent state parameters[20] (θ,φ) describing the signal are related by

$$e\exp(-i\varphi)\sin\theta=\nu/[|\tfrac{1}{2}\lambda'\langle\sigma_z\rangle|^2+|\nu|^2]^{1/2}. \tag{8.3}$$

These results do not generalize to multilevel molecular systems with $r>2$.

9. CONCLUSION

A variation of the coupled order parameter method has been applied to the description of several model Hamiltonians. In this method, the Hamiltonian was linearized by expanding the shift operators appearing in the interaction term about disposable c-number parameters. The free energy of the linearized Hamiltonian was computed, and the parameters were chosen to minimize this free energy. Vanishing of the first derivatives led to a system of coupled nonlinear equations for the disposable parameters. The different branches of the nonlinear problem were labelled by the inertia of the free energy stability matrix. Only branches with a positive definite or semidefinite FESM can provide parameters yielding a linearized Hamiltonian thermodynamically equivalent to the original Hamiltonian.

It is always possible to find solutions of the nonlinear equations for which the FESM is positive definite or semidefinite. In the latter case the free energy is invariant under a gauge group. We proved in Sec. 6 that the global minimum of the nonlinear equations does in fact supply a thermodynamically equivalent Hamiltonian for any model Hamiltonian of Dicke type [i.e. (7.7)] with only a finite number of field modes present.

Applying these results in Sec. 3, we saw that the real or imaginary branch is globally stable if $s<0$ or $s>0$. In Sec. 4 we showed that external sources will destroy the second order phase transition in the simple Dicke model. If internal mechanisms are present which already destroy the gauge symmetry, then the phase transition may persist under certain conditions (Sec. 5). In Sec. 8 we gave a physical interpretation to the density operator obtained from the linearized form of (5.1).

In Sec. 7 we treated generalized Dicke models. The changes brought about by introducing several field modes were easily treated [(7.5)]. The Dicke Hamiltonian describing multimode-multilevel interaction (7.7) was then treated by the coupled order parameter method. The bifurcation properties of the resulting nonlinear equations were treated qualitatively, and it was apparent that there is a much richer spectrum of possible ordered states for 3-level systems than for 2-level systems. The modifications required by allowing more than one mode per pair of molecular levels were discussed. Finally, we presented a criterion for determining whether an external coupling will destroy a bifurcation and therefore a phase transition.

ACKNOWLEDGMENTS

One of us (R.G.) would like to thank Professor N. Bazley (Köln) for very enlightening discussions, and Professor A. Verbeure and Dr. P. N. M. Sission for bringing Ref. 12 to his attention.

*Work supported in part by the US Army Research Office, Durham, North Carolina, Grant DAHC04-72-A-0001.

[1]R.H. Dicke, Phys. Rev. **93**, 99 (1954).

[2]K. Hepp and E.H. Lieb, Ann. Phys. (N.Y.) **76**, 360 (1973).

[3]G.H. Pimbley, Jr., *Eigenfunction Branches of Nonlinear Operators, and Their Bifurcation* (Springer, Berlin, 1969).

[4]R. Gilmore and C.M. Bowden (to appear).

[5]R. Gilmore, to be published.

[6]K. Rzażewski, W. Wódkiewicz, and W. Żakowicz, Phys. Rev. Lett. **35**, 432 (1975).

[7]R. Gilmore, Phys. Lett. A **55**, 459 (1976).

[8]R.W. Gibberd, Aust. J. Phys. **27**, 241 (1974).

[9]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley-Interscience, New York, 1953).

[10]L. Landau and E. Lifshitz, *Physique Statistique* (Editions Mir., Moscow, 1967).

[11]R. Gilmore, *Lie Groups, Lie Alegebras, and Some of Their Applications* (Wiley, New York, 1974).

[12]J.P. Provost, F. Rocca, G. Vallee, and M. Sirugue, to be published.

[13]We reserve the term "method of steepest descent" to integrals of analytic functions of several complex variables. In this connection, cf. H. Jeffries and B.S. Jeffries, *Methods of Mathematical Physics* (University Press, Cambridge, 1966), p. 503.

[14]D.V. Widder, *An Introduction to Transform Theory* (Academic, New York, 1971), pp. 134–40.

[15]Y.K. Wang and F.T. Hioe, Phys. Rev. A **7**, 831 (1973).

[16]R.J. Glauber, Phys. Rev. **130**, 2529 (1963); 131, 2766 (1963).

[17]R. Gilmore, Ann. Phys. (N.Y.) **74**, 391 (1972).

[18]R. Gilmore, C.M. Bowden, and L.M. Narducci, Phys. Rev. A **12**, 1019 (1975).

[19]B.V. Thompson, J. Phys. A **8**, 126 (1975).

[20]F.T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A **6**, 2211 (1972).

Classification of all simple graded Lie algebras whose Lie algebra is reductive. I

M. Scheunert, W. Nahm, and V. Rittenberg

Physikalisches Institut der Universität Bonn, Bonn, Germany
(Received 22 March 1976)

All simple graded Lie algebras whose Lie algebra is reductive are presented, and the classification theorem is proved. Several theorems which may show up to be useful in a different context are also included.

1. INTRODUCTION

Graded Lie algebras made their appearance in the early sixties in mathematics,[1] were rediscovered by physicists in the early seventies,[2] and recently started to have parallel evolutions.[3] In the present paper we will try to phrase the problem in both languages. The readers interested in immediate physical applications will probably be mainly interested in the first two sections.

A graded Lie algebra $\mathfrak{a}$ contains both commutators ([,]) and anticommutators ({ , }),

$$[Q_m, Q_n] = C^p_{mn} Q_p, \tag{1.1}$$

$$[Q_m, V_\alpha] = F^\beta_{m\alpha} V_\beta, \tag{1.2}$$

$$\{V_\alpha, V_\beta\} = A^m_{\alpha\beta} Q_m. \tag{1.3}$$

The even generators (Q_m) form a Lie algebra (even subspace) $\mathfrak{g}$, the odd generators (V_α) generate the odd subspace $\mathfrak{u}$ ($\mathfrak{a} = \mathfrak{g} \oplus \mathfrak{u}$). The odd generators verify the Jacobi identities

$$[Q_m, \{V_\alpha, V_\beta\}] + \{[V_\alpha, Q_m], V_\beta\} + \{[V_\beta, Q_m], V_\alpha\} = 0, \tag{1.4}$$

$$[V_\alpha, \{V_\beta, V_\gamma\}] + [V_\gamma, \{V_\alpha, V_\beta\}] + [V_\beta, \{V_\gamma, V_\alpha\}] = 0, \tag{1.5}$$

and form a representation of the algebra $\mathfrak{g}$.

One can define a *Killing form* through the adjoint representation of the algebra $\mathfrak{a}$ (i.e., with the help of the structure constants) as follows[4]:

$$\begin{aligned} g_{mn} &= g_{nm} = C^p_{mq} C^q_{np} - F^\beta_{m\alpha} F^\alpha_{n\beta}, \\ g_{\alpha\beta} &= -g_{\beta\alpha} = F^\gamma_{m\alpha} A^m_{\beta\gamma} - F^\gamma_{m\beta} A^m_{\alpha\gamma}, \\ g_{m\alpha} &= g_{\alpha m} = 0. \end{aligned} \tag{1.6}$$

It is also possible to define a *trace-form metric* associated with *any* (graded) representation of $\mathfrak{a}$ (not necessarily the adjoint representation).[5]

Let us remind the reader what is the present stage of the problem of classification of the graded Lie algebras.

In Ref. 4 the graded Lie algebras which satisfy the criteria

(a) $\mathfrak{g}$ is simple,

(b) the Killing form of $\mathfrak{a}$ is nondegenerate

have been classified. The classification was possible because of two theorems, the weight-root theorem (which relates the weights of the representation V_α to the roots of the Lie algebra $\mathfrak{g}$) and the C theorem [which relates the $A^m_{\alpha\beta}$ coefficients in Eq. (1.3) to the $F^\beta_{m\alpha}$ coefficients of Eq. (1.2)]. These two theorems will be generalized in Sec. 3.

In Ref. 6 the present authors have given without proof (the proof is contained in the present paper) the classification of the graded Lie algebras which satisfy the criteria

(a) $\mathfrak{a}$ is simple (contains no nontrivial ideals),

(b) the Killing form of $\mathfrak{a}$ is nondegenerate.

These algebras were called *strictly simple*. In Ref. 5 it was shown that a graded Lie algebra which has a nondegenerate Killing form is the direct product of strictly simple algebras.

In an independent investigation, Freund and Kaplansky[7] have classified the algebras for which

(a) $\mathfrak{a}$ is simple,

(b) at least one trace-form metric is nondegenerate.

In the present paper we present all graded Lie algebras for which

(a) $\mathfrak{a}$ is simple,

(b) $\mathfrak{g}$ is reductive ($\mathfrak{g} = \mathfrak{g}_1 \times \mathfrak{g}_2$ where $\mathfrak{g}_1$ is semisimple and $\mathfrak{g}_2$ Abelian).

As will be shown the algebras considered previously[4–7] are all contained in the present class.

After our work was completed, we learnt of a paper by Kac[8] who has classified all simple graded Lie algebras. The algebras we are considering are called by him *classical*.

The classical simple graded Lie algebras have been discovered independently by various authors. Apart from Ref. 8 which contains all of them, we mention that the special linear and orthosymplectic graded Lie algebras have been defined in Refs. 5 and 7, the exceptional graded Lie algebras have been found in Ref. 7, and the remaining classical graded Lie algebras have been constructed in Ref. 5 [among the latter there are the well-known (f, d) algebras of Gell-Mann, Michel, and Radicati[9]].

The line of thinking of this paper was started in Ref. 5 where two of the present authors showed the special role of simple graded Lie algebras whose Lie algebra is reductive. It was shown that for these algebras the odd generators V_α form a completely reducible representation with at most two irreducible components.

This result has far reaching consequences. In fact, assume that $\mathfrak{g}$ is neither simple nor of the particular form $\mathrm{gl}(1)\times\mathfrak{g}_1$ or $\mathrm{sl}(2)\times\mathfrak{g}_1$ with a simple Lie algebra $\mathfrak{g}_1$. Then it is possible to read off the complete structure of $\mathfrak{a}$ directly from the Jacobi identities. The special cases excluded above can then be treated by a careful study of the representation on the odd generators. Furthermore, the Killing form of $\mathfrak{a}$ plays a crucial role in the latter discussion.

Our work is organized as follows. In Sec. 2 we describe the families of simple graded Lie algebras containing a reductive Lie algebra. These families are given in an explicit matrix form so that the calculation of the structure constants in (1.1)—(1.3) is an elementary exercise. The three exceptional algebras are mentioned, but their construction is left for a subsequent paper.[10] The classification theorem is also presented in Sec. 2. Its proof depends on the general results derived in Ref. 5; those which are most important are summarized in a second theorem.

In Sec. 3 we collect some preliminaries and generalize the two theorems of Ref. 4.

The last three sections contain the main proof. We have to distinguish several cases depending on whether the Lie algebra $\mathfrak{g}$ is simple or not and whether the odd subspace $\mathfrak{u}$ is irreducible or not.

Our proof depends essentially on some results on "low dimensional" irreducible representations of simple Lie algebras. These results as well as our notational conventions concerning Lie algebras are collected in four appendices. As a by-product, the classification of the algebras considered in Ref. 4 is reobtained in an elegant way in Appendix C.

The problem of the representations of the graded Lie algebras whose Lie algebra is reductive is not considered here but will be dealt with in another paper.[11]

Our notation concerning graded Lie algebras is that of Ref. 5, in particular we denote the multiplication in a graded Lie algebra [see (1.1)—(1.3)] by a bracket $\langle\ ,\ \rangle$. All vector spaces and algebras are supposed to be finite dimensional over an algebraically closed field K of characteristic zero (for example the field of complex numbers).

2. COLLECTION OF SOME EARLIER RESULTS AND FORMULATION OF THE MAIN THEOREM

In Ref. 5 we constructed several (double) sequences of simple graded Lie algebras. Our starting point was the *general linear graded Lie algebra* $\mathrm{pl}(n,m)$. This algebra is defined as follows: Choose any positive integers $n, m \geq 1$ and let $\mathrm{pl}(n,m)$ be the vector space of all $(n+m)\times(n+m)$ matrices, written in block form

$$X=\begin{pmatrix} A & B\\ C & D\end{pmatrix} \tag{2.1}$$

with

A an arbitrary $n\times n$ matrix, B an arbitrary $n\times m$ matrix,

C an arbitrary $m\times n$ matrix, D an arbitrary $m\times m$ matrix.

The Lie algebra $\mathfrak{g}$ of $\mathrm{pl}(n,m)$ consists of the "diagonal" block matrices $\left(\begin{smallmatrix} A & 0\\ 0 & D\end{smallmatrix}\right)$. The odd subspace $\mathfrak{u}$ of $\mathrm{pl}(n,m)$ consists of the "off-diagonal" block matrices $\left(\begin{smallmatrix} 0 & B\\ C & 0\end{smallmatrix}\right)$. The bracket $\langle X, X'\rangle$ of two elements X, X' of $\mathrm{pl}(n,m)$ is the usual commutator if X or X' is an element of $\mathfrak{g}$; it is the anticommutator if X and X' are elements of $\mathfrak{u}$. Hence if $X'=\left(\begin{smallmatrix} A' & B'\\ C' & D'\end{smallmatrix}\right)$, we obtain

$$\langle X, X'\rangle = \begin{pmatrix} AA'-A'A+BC'+B'C & BD'-B'D+AB'-A'B\\ CA'-C'A+DC'-D'C & DD'-D'D+CB'+C'B\end{pmatrix}. \tag{2.2}$$

Note that the mapping

$$\begin{pmatrix} A & B\\ C & D\end{pmatrix}\to\begin{pmatrix} D & C\\ B & A\end{pmatrix} \tag{2.3}$$

is an isomorphism of $\mathrm{pl}(n,m)$ onto $\mathrm{pl}(m,n)$.

(a) Define the *special linear graded Lie algebra* $\mathrm{spl}(n,m)$ by

$$\mathrm{spl}(n,m)=\left\{\begin{pmatrix} A & B\\ C & D\end{pmatrix}\in \mathrm{pl}(n,m)\,\middle|\,\mathrm{Tr}(A)=\mathrm{Tr}(D)\right\}. \tag{2.4}$$

Its Lie algebra is $\mathrm{sl}(n)\times\mathrm{sl}(m)\times\mathrm{gl}(1)$ [where $\mathrm{gl}(1)$ is the trivial one-dimensional Lie algebra]. If $n\neq m$ then $\mathrm{spl}(n,m)$ is simple. Since $\mathrm{spl}(n,m)$ and $\mathrm{spl}(m,n)$ are isomorphic, it suffices to consider the algebras $\mathrm{spl}(n,m)$ with $n>m\geq 1$.

(b) The algebras $\mathrm{spl}(n,n)$ are not simple. In fact they have a nontrivial center z_n which consists of the scalar multiples of the unit matrix

$$z_n=\left\{\begin{pmatrix} \lambda I_n & 0\\ 0 & \lambda I_n\end{pmatrix}\,\middle|\,\lambda\in K\right\}. \tag{2.5}$$

(Quite generally I_r denotes the r-dimensional unit matrix.) The quotient algebra $\mathrm{spl}(n,n)/z_n$ is simple if $n\geq 2$; its Lie algebra is $\mathrm{sl}(n)\times\mathrm{sl}(n)$.

(c) Suppose that $n=2p$ is an even positive integer and that $m\geq 1$ is an arbitrary positive integer. Define the $2p\times 2p$ matrix

$$G=\begin{pmatrix} 0 & I_p\\ -I_p & 0\end{pmatrix}. \tag{2.6}$$

Then the subalgebra of $\mathrm{pl}(2p,m)$ consisting of all block matrices $\left(\begin{smallmatrix} A & B\\ C & D\end{smallmatrix}\right)$ which satisfy

$$^tAG+GA=0,\quad {}^tD+D=0,\quad C={}^tBG \tag{2.7}$$

is simple. Its Lie algebra is $\mathrm{sp}(2p)\times\mathrm{o}(m)$. Hence this algebra will be called an *orthosymplectic* graded Lie algebra; it will be denoted by $\mathrm{osp}(2p,m)$. Note that the cases $m=1$ and $m=2$ are not excluded; note, furthermore, that $\mathrm{osp}(2,2)$ is isomorphic to $\mathrm{spl}(2,1)$.

(d) Suppose that $n=m\geq 3$. The subalgebra of $\mathrm{pl}(n,n)$ consisting of all block matrices $\left(\begin{smallmatrix} A & B\\ C & D\end{smallmatrix}\right)$ with

$$\begin{aligned} &{}^tA+D=0,\quad {}^tB-B=0,\\ &{}^tC+C=0,\quad \mathrm{Tr}(A)=0\end{aligned} \tag{2.8}$$

is simple. This graded Lie algebra will be denoted by $\mathrm{b}(n)$; its Lie algebra is $\mathrm{sl}(n)$.

(e) Suppose once again that $n=m\geq 3$. Define a subalgebra $\mathrm{d}(n)$ of $\mathrm{pl}(n,n)$ by

$$\mathrm{d}(n)=\left\{\begin{pmatrix}A & B\\ B & A\end{pmatrix}\middle| A\in \mathrm{gl}(n),\ B\in \mathrm{sl}(n)\right\}. \tag{2.9}$$

The center of $\mathrm{d}(n)$ is equal to z_n [see (2.5)] and the quotient algebra $\mathrm{d}(n)/z_n$ is simple. Note that $\mathrm{d}(n)/z_n$ is the (f,d) algebra of Gell-Mann, Michel, and Radicati.[9] Its Lie algebra is equal to $\mathrm{sl}(n)$.

(*f*) In a subsequent paper[10] we shall prove that there exist some additional simple graded Lie algebras whose Lie algebra is reductive. These algebras will be called *exceptional*. There exist:

(i) A one-parameter family of 17-dimensional simple graded Lie algebras whose Lie algebra is equal to $\mathrm{sl}(2)\times\mathrm{sl}(2)\times\mathrm{sl}(2)$ and whose odd subspace carries the tensor product of the three two-dimensional representations of the algebras $\mathrm{sl}(2)$,

(ii) A 31-dimensional simple graded Lie algebra whose Lie algebra is $\mathrm{sl}(2)\times G_2$ and whose odd subspace carries the tensor product of the two-dimensional representation of $\mathrm{sl}(2)$ with the seven-dimensional fundamental representation of G_2,

(iii) A 40-dimensional simple graded Lie algebra whose Lie algebra is $\mathrm{sl}(2)\times\mathrm{o}(7)$ and whose odd subspace carries the tensor product of the two-dimensional representation of $\mathrm{sl}(2)$ with the eight-dimensional spin representation of $\mathrm{o}(7)$.

Now we can formulate the main result of the present work as follows:

Theorem 1: The simple graded Lie algebras described in (a)—(f) constitute a complete list of all simple graded Lie algebras whose Lie algebra is reductive.

Our proof of this theorem will depend crucially on the general results derived in Ref. 5. Those which are most relevant for our purpose are collected in Theorem 2.

Let $\mathfrak{a}=\mathfrak{g}\oplus\mathfrak{u}$ be a graded Lie algebra with Lie algebra $\mathfrak{g}$ and odd subspace $\mathfrak{u}$. Recall that $\mathfrak{u}$ carries a natural representation $\mathrm{ad}_{\mathfrak{u}}$ of $\mathfrak{g}$ which we call the *adjoint representation of* $\mathfrak{g}$ *in* $\mathfrak{u}$.

Theorem 2: Let $\mathfrak{a}=\mathfrak{g}\oplus\mathfrak{u}$ be a simple graded Lie algebra. Then

$$\text{(a)}\ \langle\mathfrak{g},\mathfrak{u}\rangle=\mathfrak{u},\ \langle\mathfrak{u},\mathfrak{u}\rangle=\mathfrak{g} \tag{2.10}$$

and $\mathrm{ad}_{\mathfrak{u}}$ is faithful.

(b) The adjoint representation $\mathrm{ad}_{\mathfrak{u}}$ of $\mathfrak{g}$ in $\mathfrak{u}$ is completely reducible if and only if $\mathfrak{g}$ is reductive.

In the following we suppose that the (equivalent) conditions of (b) are satisfied. Then

(c) The odd subspace $\mathfrak{u}$ of $\mathfrak{a}$ is either $\mathfrak{g}$-irreducible or it decomposes into the direct sum

$$\mathfrak{u}=\mathfrak{u}'\oplus\mathfrak{u}'' \tag{2.11}$$

of two (nontrivial) $\mathfrak{g}$ irreducible subspaces $\mathfrak{u}'$ and $\mathfrak{u}''$ which satisfy

$$\langle\mathfrak{u}',\mathfrak{u}'\rangle=\langle\mathfrak{u}'',\mathfrak{u}''\rangle=\{0\} \tag{2.12}$$

and hence $\langle\mathfrak{u}',\mathfrak{u}''\rangle=\mathfrak{g}$.

(d) Suppose that the center $\mathfrak{g}_0$ of $\mathfrak{g}$ is nontrivial. Then $\mathfrak{g}_0$ is one-dimensional and the representation $\mathrm{ad}_{\mathfrak{u}}$ is reducible. Furthermore, there exists a (unique) element $E\in\mathfrak{g}_0$ such that [with the notations introduced in (c)]

$$\begin{aligned}&\langle E,U'\rangle=U' \quad\text{if } U'\in\mathfrak{u}'\\ &\langle E,U''\rangle=-U'' \text{ if } U''\in\mathfrak{u}''.\end{aligned} \tag{2.13}$$

3. PRELIMINARIES

In this section we shall derive some general results on graded Lie algebras which will be needed in the following. It would interrupt the main line of argumentation if we introduced them just at the places where they were relevant.

The first result is connected with some trivial process by which one can construct a new graded Lie algebra from a given one. It turns out that both algebras are isomorphic. Since we want to classify graded Lie algebras up to isomorphisms, we must be aware of this process in order to avoid a "double counting" of some algebras. The result is described in the following lemma.

Lemma 3.1: Let $\mathfrak{a}=\mathfrak{g}\oplus\mathfrak{u}$ be a graded Lie algebra (whose multiplication is denoted by $\langle\ ,\ \rangle$), let $c\neq 0$ be any element of K and let τ be an automorphism of the Lie algebra $\mathfrak{g}$. Define a new graded Lie algebra $\mathfrak{a}'=\mathfrak{g}\oplus\mathfrak{u}$, whose underlying Lie algebra and odd subspace are again respectively $\mathfrak{g}$ and $\mathfrak{u}$, but whose multiplication $\langle\ ,\ \rangle'$ is defined by

$$\begin{aligned}&\langle G_1,G_2\rangle'=\langle G_1,G_2\rangle,\quad \langle G,U\rangle'=\langle\tau^{-1}(G),U\rangle,\\ &\langle U,G\rangle'=\langle U,\tau^{-1}(G)\rangle,\quad \langle U_1,U_2\rangle'=(1/c^2)\tau(\langle U_1,U_2\rangle)\end{aligned} \tag{3.1}$$

if $G,G_1,G_2\in\mathfrak{g}$ and $U,U_1,U_2\in\mathfrak{u}$. It is easy to see that $\mathfrak{a}'$ is indeed a graded Lie algebra and that the linear mapping

$$\phi:\mathfrak{a}\to\mathfrak{a}', \tag{3.2a}$$

defined by

$$\begin{aligned}&\phi(G)=\tau(G)\quad\text{if } G\in\mathfrak{g},\\ &\phi(U)=cU\quad\text{if } U\in\mathfrak{u},\end{aligned} \tag{3.2b}$$

is an isomorphism of $\mathfrak{a}$ onto $\mathfrak{a}'$.

One should note that the representations of $\mathfrak{g}$ is the odd subspaces of $\mathfrak{a}$ and $\mathfrak{a}'$ respectively are not necessarily equivalent in spite of the fact that $\mathfrak{a}$ and $\mathfrak{a}'$ are isomorphic. The lemma shows in particular that a rescaling of the product mapping $\mathfrak{u}\times\mathfrak{u}\to\mathfrak{g}$ leads to isomorphic graded Lie algebras.

The rest of this section is applicable to any graded Lie algebra $\mathfrak{a}=\mathfrak{g}\oplus\mathfrak{u}$ whose (generalized) Killing form is nondegenerate. In fact what we need to know is that

(a) the Lie algebra $\mathfrak{g}$ is reductive, i.e., $\mathfrak{g}$ is the direct product of its center $\mathfrak{g}_0$ with the derived algebra $\mathfrak{g}'=\langle\mathfrak{g},\mathfrak{g}\rangle$ which is semisimple;

(b) the adjoint representation $\mathrm{ad}_{\mathfrak{u}}$ of $\mathfrak{g}$ in $\mathfrak{u}$ is completely reducible;

(c) $\mathfrak{a}$ has a nondegenerate bilinear form ϕ which is even (with respect to the grading), i.e.,

$$\phi(\mathfrak{g}, \mathfrak{u}) = \phi(\mathfrak{u}, \mathfrak{g}) = \{0\}, \tag{3.3}$$

which is invariant, i.e.,

$$\phi(\langle X, Y\rangle, Z) = \phi(X, \langle Y, Z\rangle) \tag{3.4}$$

for all $X, Y, Z \in \mathfrak{a}$, and which has the symmetry properties

$$\begin{aligned} &\phi(X, Y) = \phi(Y, X) \quad \text{if } X, Y \in \mathfrak{g}, \\ &\phi(X, Y) = -\phi(Y, X) \quad \text{if } X, Y \in \mathfrak{u}. \end{aligned} \tag{3.5}$$

It is obvious that $\mathfrak{g}_0$ and $\mathfrak{g}'$ are orthogonal with respect to ϕ, hence the restrictions of ϕ to $\mathfrak{g}_0$ and $\mathfrak{g}'$ are nondegenerate.

Let us choose a Cartan subalgebra $\mathfrak{h}$ of $\mathfrak{g}$. Then

$$\mathfrak{h} = \mathfrak{g}_0 \times \mathfrak{h}', \tag{3.6}$$

where $\mathfrak{h}'$ is a Cartan subalgebra of the semisimple Lie algebra $\mathfrak{g}'$. We conclude that the restriction of ϕ to $\mathfrak{h}$ is nondegenerate; consequently we can define as usual a nondegenerate symmetric bilinear form $(\,|\,)$ on the dual space $\mathfrak{h}^*$ of $\mathfrak{h}$ (and in particular for the weights of the representations of $\mathfrak{g}$) by following procedure.

Let $\alpha \in \mathfrak{h}^*$ be a linear form on $\mathfrak{h}$. Then there exists exactly one element $H_\alpha \in \mathfrak{h}$ such that

$$\alpha(H) = \phi(H_\alpha, H) \quad \text{for all } H \in \mathfrak{h}. \tag{3.7}$$

If $\alpha, \beta \in \mathfrak{h}^*$ we define

$$(\alpha|\beta) = \phi(H_\alpha, H_\beta) = \alpha(H_\beta) = \beta(H_\alpha). \tag{3.8}$$

Let us apply this formalism to the adjoint representation $\mathrm{ad}_\mathfrak{u}$ of $\mathfrak{g}$ in $\mathfrak{u}$. For any linear form $\alpha \in \mathfrak{h}^*$ we define

$$\mathfrak{g}^\alpha = \{X \in \mathfrak{g} \,|\, \langle H, X\rangle = \alpha(H)X \quad \text{for all } H \in \mathfrak{h}\}, \tag{3.9}$$

$$\mathfrak{u}^\alpha = \{Y \in \mathfrak{u} \,|\, \langle H, Y\rangle = \alpha(H)Y \quad \text{for all } H \in \mathfrak{h}\}. \tag{3.10}$$

In particular we have

$$\mathfrak{g}^0 = \mathfrak{h}. \tag{3.11}$$

The linear forms $\alpha \in \mathfrak{h}^*$ with $\mathfrak{u}^\alpha \neq \{0\}$ (resp. with $\alpha \neq 0$ and $\mathfrak{g}^\alpha \neq \{0\}$) are the weights of $\mathrm{ad}_\mathfrak{u}$ (resp. the roots of $\mathfrak{g}'$) and we know that

$$\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\alpha \neq 0} \mathfrak{g}^\alpha, \tag{3.12}$$

$$\mathfrak{u} = \bigoplus_{\alpha} \mathfrak{u}^\alpha. \tag{3.13}$$

Now the restriction of ϕ to $\mathfrak{u}$ is nondegenerate. Hence $\mathrm{ad}_\mathfrak{u}$ is self-contragredient and $-\alpha$ is a weight of $\mathrm{ad}_\mathfrak{u}$ if and only if α is a weight; furthermore, the restriction of ϕ to $\mathfrak{u}^\alpha \times \mathfrak{u}^{-\alpha}$ is nondegenerate.

Choose $X \in \mathfrak{u}^\alpha$ and $Z \in \mathfrak{u}^{-\alpha}$. Then $\langle X, Z\rangle \in \mathfrak{h}$, and from (3.4) we conclude that

$$\phi(\langle X, Z\rangle, H) = \alpha(H)\phi(X, Z) = \phi(\phi(X, Z)H_\alpha, H) \tag{3.14}$$

for all $H \in \mathfrak{h}$. It follows that

$$\langle X, Z\rangle = \phi(X, Z)H_\alpha \tag{3.15}$$

for all $X \in \mathfrak{u}^\alpha$, $Z \in \mathfrak{u}^{-\alpha}$.

It is now easy to derive the natural generalization of the root-weight theorem of Ref. 4.

Suppose that α is a weight and choose $X, Y \in \mathfrak{u}^\alpha$; $Z \in \mathfrak{u}^{-\alpha}$. Then the (generalized) Jacobi identity

$$\langle\langle X, Y\rangle, Z\rangle + \langle\langle Z, X\rangle, Y\rangle + \langle\langle Y, Z\rangle, X\rangle = 0 \tag{3.16}$$

reads, according to (3.15) and (3.10),

$$\langle\langle X, Y\rangle, Z\rangle = -(\alpha|\alpha)\{\phi(X, Z)Y + \phi(Y, Z)X\}. \tag{3.17}$$

If $(\alpha|\alpha) \neq 0$, the right-hand side of this equation is not identically zero. Now $\langle X, Y\rangle \in \mathfrak{g}^{2\alpha}$ for all $X, Y \in \mathfrak{u}^\alpha$. Since $\mathfrak{g}^{2\alpha}$ is at most one-dimensional, we conclude the following

Lemma 3.2: Let α be a weight of $\mathrm{ad}_\mathfrak{u}$ such that $(\alpha|\alpha) \neq 0$. Then 2α is a root of $\mathfrak{g}$ and

$$\dim \mathfrak{u}^\alpha = \dim \mathfrak{u}^{-\alpha} = 1.$$

A slight modification of the argument above yields the following lemma.

Lemma 3.3: Let α, β be two weights of $\mathrm{ad}_\mathfrak{u}$ such that $\alpha \neq \pm\beta$. If $(\alpha|\beta) \neq 0$, then $\beta + \alpha$ or $\beta - \alpha$ is a root of $\mathfrak{g}$.

Proof: Choose elements $X \in \mathfrak{u}^\alpha$, $Y \in \mathfrak{u}^{-\alpha}$ such that $\phi(X, Y) \neq 0$ and let Z be any nonzero element of $\mathfrak{u}^\beta$. Then the Jacobi identity (3.16) means

$$\langle\langle Z, X\rangle, Y\rangle + \langle\langle Z, Y\rangle, X\rangle = -(\alpha|\beta)\phi(X, Y)Z. \tag{3.18}$$

Since the right-hand side is nonzero, one of the elements $\langle Z, X\rangle \in \mathfrak{g}^{\beta+\alpha}$ and $\langle Z, Y\rangle \in \mathfrak{g}^{\beta-\alpha}$ must be nonzero, and the lemma follows.

Let us now exploit in some greater detail the fact that $\mathfrak{g}$ is reductive. We have already mentioned that $\mathfrak{g} = \mathfrak{g}_0 \times \mathfrak{g}'$, where $\mathfrak{g}_0$ is the center of $\mathfrak{g}$ and $\mathfrak{g}' = \langle \mathfrak{g}, \mathfrak{g}\rangle$ is semisimple. Consequently $\mathfrak{g}$ decomposes into a direct product

$$\mathfrak{g} = \mathfrak{g}_0 \times \mathfrak{g}_1 \times \cdots \times \mathfrak{g}_r \tag{3.19}$$

with simple Lie algebras $\mathfrak{g}_i$, $1 \leq i \leq r$.

It is easy to see that the $\mathfrak{g}_j$, $0 \leq j \leq r$, are mutually orthogonal with respect to ϕ, i.e.,

$$\phi(\mathfrak{g}_j, \mathfrak{g}_k) = \{0\} \quad \text{if } 0 \leq j < k \leq r. \tag{3.20}$$

Hence the restriction ϕ_j of ϕ to $\mathfrak{g}_j$, $0 \leq j \leq r$, is nondegenerate. In particular, we would like to stress the important fact that ϕ_i, $1 \leq i \leq r$, is a nonzero multiple of the Killing form of $\mathfrak{g}_i$.

It is well known that the Cartan subalgebra $\mathfrak{h}'$ of $\mathfrak{g}' = \mathfrak{g}_1 \times \cdots \times \mathfrak{g}_r$ is the direct product of Cartan subalgebras $\mathfrak{h}_i$ of the $\mathfrak{g}_i$, $1 \leq i \leq r$, and therefore

$$\mathfrak{h} = \mathfrak{h}_0 \times \mathfrak{h}_1 \times \cdots \times \mathfrak{h}_r. \tag{3.21}$$

For notational convenience we have defined $\mathfrak{h}_0 = \mathfrak{g}_0$.

The restrictions of ϕ_j to $\mathfrak{h}_j$, $0 \leq j \leq r$, define a nondegenerate symmetric bilinear form $(\,|\,)_j$ on the dual $\mathfrak{h}_j^*$ of $\mathfrak{h}_j$ in the same way as $(\,|\,)$ is defined by the restriction of ϕ to $\mathfrak{h}$.

Now $\mathfrak{h}^*$ is canonically isomorphic to $\mathfrak{h}_0^* \times \cdots \times \mathfrak{h}_r^*$, that is every element $\alpha \in \mathfrak{h}^*$ can be identified with the family $(\alpha_j)_{0 \leq j \leq r}$, where α_j is the restriction of α to $\mathfrak{h}_j$. If $\alpha = (\alpha_j)_{0 \leq j \leq r}$ and $\beta = (\beta_j)_{0 \leq j \leq r}$ are two elements of $\mathfrak{h}^*$, then

$$(\alpha|\beta)=\sum_{j=0}^{r}(\alpha_j|\beta_j)_j. \tag{3.22}$$

In the following we shall omit the index j on $(\ |\ )_j$, since it will be obvious from the context which bilinear form has to be taken.

Finally let us generalize the C theorem of Ref. 4. We remark first that an even bilinear form ϕ on $\mathfrak{g}$ is invariant if and only if it is $\mathfrak{g}$-invariant and if

$$\phi(G,\langle U,V\rangle)=\phi(\langle G,U\rangle,V)=\phi(\langle U,V\rangle,G) \tag{3.23}$$

for all $U,V\in\mathfrak{u}$ and $G\in\mathfrak{g}$.

To exploit this equation we suppose that we are given two $\mathfrak{g}$-invariant $\mathfrak{g}$-irreducible subspaces $\mathfrak{u}'$ and $\mathfrak{u}''$ of $\mathfrak{u}$ such that the restriction of ϕ to $\mathfrak{u}'\times\mathfrak{u}''$ is nondegenerate. (In our applications we shall have either $\mathfrak{u}'=\mathfrak{u}''=\mathfrak{u}$ or $\mathfrak{u}'\oplus\mathfrak{u}''=\mathfrak{u}$.) Then there exist irreducible representations ρ_i' (resp. ρ_i'') of $\mathfrak{g}_i$ in some spaces $\mathfrak{u}_i'$ (resp. $\mathfrak{u}_i''$), $1\le i\le r$, such that

$$\mathfrak{u}'=\mathfrak{u}_1'\otimes\cdots\otimes\mathfrak{u}_r',\quad \mathfrak{u}''=\mathfrak{u}_1''\otimes\cdots\otimes\mathfrak{u}_r'', \tag{3.24}$$

and such that the representation of $\mathfrak{g}'=\mathfrak{g}_1\times\cdots\times\mathfrak{g}_r$ in $\mathfrak{u}'$ (resp. $\mathfrak{u}''$) induced by $\mathrm{ad}_{\mathfrak{u}}$ is the tensor product of the representations ρ_i' (resp. ρ_i''). Since the restriction of ϕ to $\mathfrak{u}'\times\mathfrak{u}''$ is nondegenerate there exist nondegenerate $\mathfrak{g}_i$-invariant bilinear forms ψ_i on $\mathfrak{u}_i'\times\mathfrak{u}_i''$, $1\le i\le r$, and these are determined up to a nonzero factor. Moreover, we have

$$\phi(U_1'\otimes\cdots\otimes U_r',U_1''\otimes\cdots\otimes U_r'')=\sigma\prod_{i=1}^{r}\psi_i(U_i',U_i'') \tag{3.25}$$

for all $U_i'\in\mathfrak{u}_i'$, $U_i''\in\mathfrak{u}_i''$, $1\le i\le r$, with some nonzero constant $\sigma\in K$.

On the other hand, there exist an element $F\in\mathfrak{g}_0$ and, for every i, $1\le i\le r$, a $\mathfrak{g}_i$-invariant bilinear mapping

$$P_i:\mathfrak{u}_i'\times\mathfrak{u}_i''\to\mathfrak{g}_i \tag{3.26}$$

such that

$$\langle U_1'\otimes\cdots\otimes U_r',U_1''\otimes\cdots\otimes U_r''\rangle =\sum_{i=1}^{r}\prod_{k\ne i}\psi_k(U_k',U_k'')\ P_i(U_i',U_i'')+\Big(\prod_{i=1}^{r}\psi_i(U_i',U_i'')\Big)F \tag{3.27}$$

for all $U_i'\in\mathfrak{u}_i'$, $U_i''\in\mathfrak{u}_i''$, $1\le i\le r$.

Then Eq. (3.23) yields for $1\le i\le r$,

$$\phi_i(P_i(U_i',U_i''),G_i)=\sigma\psi_i(\langle G_i,U_i'\rangle,U_i''), \tag{3.28}$$

where $U_i'\in\mathfrak{u}_i'$, $U_i''\in\mathfrak{u}_i''$, $G_i\in\mathfrak{g}_i$, and where $\langle G_i,U_i'\rangle$ denotes the action of G_i on U_i' according to the representation ρ_i'. Conversely it is easy to see that this equation determines uniquely a bilinear $\mathfrak{g}_i$-invariant mapping P_i of $\mathfrak{u}_i'\times\mathfrak{u}_i''$ into $\mathfrak{g}_i$.

The essential fact is now that P_i is fixed up to a factor, once $\mathfrak{g}_i$ and the contragredient representations ρ_i' and ρ_i'' of $\mathfrak{g}_i$ are given, the free factor (which may depend on i) reflects the fact that ϕ_i and ψ_i are (in advance) only known up to a factor. Stated differently: It is evident that the tensor product of the contragredient representations ρ_i' and ρ_i'' contains the adjoint representation of $\mathfrak{g}_i$, i.e., that a nonzero $\mathfrak{g}_i$-invariant bilinear mapping $\mathfrak{u}_i'\times\mathfrak{u}_i''\to\mathfrak{g}_i$ does exist, but since the tensor product might contain the adjoint representation more than once, it is important to know that we must choose P_i according to (3.28).

In the special case where $\mathfrak{u}'=\mathfrak{u}''$ and $\mathfrak{u}_i'=\mathfrak{u}_i''$, $1\le i\le r$, it is well known that every ψ_i is either symmetric or skew-symmetric; then from Eq. (3.28) one can easily deduce that P_i is skew-symmetric (resp. symmetric).

From now on all graded Lie algebras $\mathfrak{a}=\mathfrak{g}\oplus\mathfrak{u}$ which occur are supposed to be simple and to contain a reductive Lie algebra $\mathfrak{g}$. For their classification we recall that $\mathfrak{a}$ satisfies one of the following three conditions (see Theorem 2):

(a) $\mathfrak{g}$ is not simple and $\mathrm{ad}_{\mathfrak{u}}$ is irreducible;

(b) $\mathfrak{g}$ is not simple and $\mathrm{ad}_{\mathfrak{u}}$ decomposes into two irreducible representations;

(c) $\mathfrak{g}$ is simple.

Our procedure will be the following: We assume that we are given a simple graded Lie algebra $\mathfrak{a}=\mathfrak{g}\oplus\mathfrak{u}$ of a certain type and study the product map $\mathfrak{u}\times\mathfrak{u}\to\mathfrak{g}$. This will lead us to identities which fix the possible Lie algebras $\mathfrak{g}$ and the representations $\mathrm{ad}_{\mathfrak{u}}$ of $\mathfrak{g}$ in the odd subspace $\mathfrak{u}$. Apart from three exceptions it will then be evident that the given algebra $\mathfrak{a}$ must be one of the algebras defined in (a)–(e) of Sec. 2. For the exceptional cases our information will be sufficient to construct the algebra $\mathfrak{a}$ explicitly; this will be done in a subsequent paper.[10]

4. GRADED LIE ALGEBRAS FOR WHICH $\mathfrak{g}$ IS NOT SIMPLE AND $\mathrm{ad}_{\mathfrak{u}}$ IS IRREDUCIBLE

As we shall see, this class of simple graded Lie algebras $\mathfrak{a}=\mathfrak{g}\oplus\mathfrak{u}$ is the most difficult one; in particular the exceptional algebras are of this type. To begin with we recall [Theorem 2(d)] that $\mathfrak{g}$ must be semisimple. Since we suppose that $\mathfrak{g}$ is not simple, the Lie algebra $\mathfrak{g}$ decomposes into a direct product of two semisimple subalgebras,

$$\mathfrak{g}=\mathfrak{g}_1\times\mathfrak{g}_2. \tag{4.1}$$

Note that we do not assume that $\mathfrak{g}_1$ or $\mathfrak{g}_2$ are simple. By assumption, the adjoint representation $\mathrm{ad}_{\mathfrak{u}}$ of $\mathfrak{g}$ in $\mathfrak{u}$ is irreducible; furthermore, $\mathrm{ad}_{\mathfrak{u}}$ is faithful (since $\mathfrak{a}$ is simple). Hence there exist irreducible faithful representations ρ_i of $\mathfrak{g}_i$ in some vector spaces $\mathfrak{u}_i$, $i=1,2$, such that

$$\mathfrak{u}=\mathfrak{u}_1\otimes\mathfrak{u}_2 \tag{4.2}$$

and such that $\mathrm{ad}_{\mathfrak{u}}$ is the tensor product of ρ_1 and ρ_2.

As we know the simplicity of $\mathfrak{a}$ implies that

$$\mathfrak{g}=\langle\mathfrak{u},\mathfrak{u}\rangle. \tag{4.3}$$

Consequently there exist for $i=1,2$ a nondegenerate $\mathfrak{g}_i$-invariant bilinear form ψ_i on $\mathfrak{u}_i$ and a nonzero $\mathfrak{g}_i$-invariant bilinear mapping

$$P_i:\mathfrak{u}_i\times\mathfrak{u}_i\to\mathfrak{g}_i \tag{4.4}$$

such that

$$\langle U_1\otimes U_2,V_1\otimes V_2\rangle =\psi_2(U_2,V_2)P_1(U_1,V_1)+\psi_1(U_1,V_1)P_2(U_2,V_2) \tag{4.5}$$

for all $U_1, V_1 \subset \mathfrak{u}_1$, and $U_2, V_2 \in \mathfrak{u}_2$.

We would like to stress that we do not assume the existence of a nondegenerate invariant bilinear form on $\mathfrak{a}$ but that the existence of the forms ψ_i is a consequence of (4.1)–(4.3).

It is well known that ψ_i is determined up to a nonzero factor and that it is either symmetric or skew-symmetric. Therefore, the mappings ψ_2, P_1 must be either both symmetric or both skew-symmetric, and similarly for ψ_1, P_2; this is a consequence of the fact that the product mapping $\mathfrak{u} \times \mathfrak{u} \to \mathfrak{g}$ is symmetric.

We shall now exploit the Jacobi identity with three odd elements,

$$\langle\langle U_1 \otimes U_2, V_1 \otimes V_2\rangle, W_1 \otimes W_2\rangle + \text{cyclic} = 0 \tag{4.6}$$

where $U_i, V_i, W_i \in \mathfrak{u}_i$, $i = 1, 2$.

Let us introduce the abbreviation

$$\widetilde{G}_i = \rho_i(G_i) \quad \text{if} \quad G_i \in \mathfrak{g}_i. \tag{4.7}$$

(A) We shall first assume that

$$\dim \mathfrak{u}_i \geq 3, \quad i = 1, 2. \tag{4.8}$$

Inserting the expression (4.5) into (4.6) we see first that there exist constants ω_i, σ_i, $\tau_i \in K$ such that

$$\begin{aligned}\widetilde{P}_i(U_i, V_i)W_i & \\ &= \omega_i \psi_i(U_i, V_i)W_i + \sigma_i \psi_i(V_i, W_i)U_i + \tau_i \psi_i(W_i, U_i)V_i\end{aligned} \tag{4.9}$$

for all $U_i, V_i, W_i \in \mathfrak{u}_i$, $i = 1, 2$, and deduce then that (4.6) is fulfilled if and only if

$$\omega_1 + \omega_2 = 0, \quad \sigma_1 + \tau_2 = 0, \quad \sigma_2 + \tau_1 = 0. \tag{4.10}$$

Now the bilinear forms ψ_i are $\mathfrak{g}_i$-invariant. In particular we must have

$$\psi_i(\widetilde{P}_i(U_i, V_i)W_i, \overline{W}_i) + \psi_i(W_i, \widetilde{P}_i(U_i, V_i)\overline{W}_i) = 0 \tag{4.11}$$

for all $U_i, V_i, W_i, \overline{W}_i \in \mathfrak{u}_i$, $i = 1, 2$. With (4.9) this condition is fulfilled if and only if

$$\omega_i = 0, \quad \sigma_i + \tau_i = 0, \quad i = 1, 2. \tag{4.12}$$

As a consequence of (4.12) the mapping P_i is symmetric (resp. skew-symmetric) if and only if ψ_i is skew-symmetric (resp. symmetric).

Collecting our results we have shown that

$$\widetilde{P}_i(U_i, V_i)W_i = \sigma\{\psi_i(V_i, W_i)U_i - \psi_i(W_i, U_i)V_i\} \tag{4.13}$$

for all $U_i, V_i, W_i \in \mathfrak{u}_i$, $i = 1, 2$, with some nonzero constant $\sigma \in K$, and, furthermore, that one of the ψ_i is symmetric, the other skew-symmetric.

Without loss of generality we may assume that ψ_1 is skew-symmetric and that ψ_2 is symmetric. It is then easy to see that the linear mappings $\widetilde{P}_i(U_i, V_i)$; $U_i, V_i \in \mathfrak{u}_i$, $i = 1$, resp. $i = 2$, generate a subspace of $\mathrm{gl}(\mathfrak{u}_i)$ (the general linear Lie algebra of $\mathfrak{u}_i$) which is equal to the symplectic Lie algebra $\mathrm{sp}(\psi_1)$ [resp. to the orthogonal Lie algebra $\mathrm{o}(\psi_2)$].

Now we know that ψ_i is $\mathfrak{g}_i$-invariant and that the representation ρ_i of $\mathfrak{g}_i$ in $\mathfrak{u}_i$ is faithful. Therefore, ρ_1 (resp. ρ_2) is an isomorphism of $\mathfrak{g}_1$ (resp. $\mathfrak{g}_2$) onto the symplectic Lie algebra $\mathrm{sp}(\psi_1)$ [resp. onto the orthogonal Lie algebra $\mathrm{o}(\psi_2)$]. Obviously, under this isomorphism the representation ρ_1 (resp. ρ_2) corresponds to the elementary representation of $\mathrm{sp}(\psi_1)$ [resp. of $\mathrm{o}(\psi_2)$].

Furthermore, according to (4.5) and (4.13), the product mapping $\mathfrak{u} \times \mathfrak{u} \to \mathfrak{g}$ is determined up to a factor σ. In view of Lemma 3.1 we have thus shown that $\mathfrak{a}$ must be isomorphic to an orthosymplectic graded Lie algebra $\mathrm{osp}(2p, m)$ with $p \geq 2$, $m \geq 3$.

(B) Let us now consider the case where the condition (4.8) is not fulfilled. Since the representation ρ_i of $\mathfrak{g}_i$ in $\mathfrak{u}_i$ is faithful we conclude that (at least) one of the spaces $\mathfrak{u}_i$ is two-dimensional and that the corresponding Lie algebra $\mathfrak{g}_i$ is isomorphic to $\mathrm{sl}(2)$.

Without loss of generality we may assume that

$$\dim \mathfrak{u}_1 = 2, \quad \mathfrak{g}_1 = \mathrm{sl}(2), \tag{4.14}$$

and that ρ_1 is the elementary representation of $\mathrm{sl}(2)$. It is well known that there exists a nondegenerate skew-symmetric invariant bilinear form ψ_1 on $\mathfrak{u}_1$ and a nonzero invariant bilinear mapping

$$P_1 : \mathfrak{u}_1 \times \mathfrak{u}_1 \to \mathfrak{g}_1. \tag{4.15a}$$

Both ψ_1 and P_1 are unique up to a factor, in particular

$$\widetilde{P}_1(U_1, V_1)W_1 = \sigma_1\{\psi_1(V_1, W_1)U_1 - \psi_1(W_1, U_1)V_1\}, \tag{4.15b}$$

where $U_1, V_1, W_1 \in \mathfrak{u}_1$ and where $\sigma_1 \in K$ is some nonzero constant.

Therefore, the mappings ψ_1 and P_1 in (4.5) are already known. We conclude that ψ_2 (resp. P_2) must be symmetric (resp. skew-symmetric) and hence that ρ_2 must be an orthogonal representation.

It is now easy to see that (4.6) is fulfilled if and only if

$$\begin{aligned}\widetilde{P}_2(U_2, V_2)W_2 - \widetilde{P}_2(V_2, W_2)U_2 - \sigma_1\psi_2(U_2, V_2)W_2 - \sigma_1\psi_2(V_2, W_2)U_2 \\ + 2\sigma_1\psi_2(W_2, U_2)V_2 = 0\end{aligned} \tag{4.16}$$

for all $U_2, V_2, W_2 \in \mathfrak{u}_2$. This condition can be rephrased by demanding that the trilinear mapping

$$\hat{P}_2 : \mathfrak{u}_2 \times \mathfrak{u}_2\ \mathfrak{u}_2 \to \mathfrak{u}_2, \tag{4.17a}$$

defined by

$$\begin{aligned}&\hat{P}_2(U_2, V_2, W_2) \\ &\quad = \widetilde{P}_2(U_2, V_2)W_2 - \sigma_1\{\psi_2(V_2, W_2)U_2 - \psi_2(W_2, U_2)V_2\},\end{aligned} \tag{4.17b}$$

should be totally skew-symmetric.

If $\hat{P}_2 = 0$ then we are back at (4.13) and we can conclude as in (A) that $\mathfrak{a}$ must be isomorphic to an orthosymplectic graded Lie algebra $\mathrm{osp}(2, m)$, $m \geq 3$. But it turns out that $\hat{P}_2$ is not necessarily equal to zero. Nevertheless if $\dim \mathfrak{u}_2 \leq 3$, then evidently $\hat{P}_2 = 0$.

Now $\mathrm{sl}(2) \times \mathrm{sl}(2) \approx \mathrm{o}(4)$ is the only semisimple Lie algebra which has a faithful irreducible orthogonal representation of dimension four. In this case we cannot conclude that $\hat{P}_2 = 0$; in fact it will turn out that this case leads to a one-parameter family of exceptional simple graded Lie algebras.

For the rest of part (B) we may assume that

$$\dim \mathfrak{u}_2 \geq 5 \tag{4.18}$$

and that $\mathfrak{g}_2$ is simple [the case in which $\mathfrak{g}_2$ is not simple leads back to (A)].

To proceed we show first that the Killing form $\phi_{\mathfrak{a}}$ of $\mathfrak{a}$ is nondegenerate. In fact, it is sufficient to prove that $\phi_{\mathfrak{a}}$ is not identically zero, and this follows from the equation

$$\phi_{\mathfrak{a}}(G_1, G_1') = \tfrac{1}{4}(4 - \dim \mathfrak{u}_2)\phi_{\mathfrak{g}_1}(G_1, G_1') \tag{4.19}$$

for all $G_1, G_1' \in \mathfrak{g}_1 = \mathrm{sl}(2)$, where $\phi_{\mathfrak{g}_1}$ is the Killing form of the Lie algebra $\mathfrak{g}_1$.

Hence we are free to apply all the results of Sec. 3. Let us first answer the question of uniqueness.

Lemma 4.1: Suppose that we are given a simple Lie algebra $\mathfrak{g}_2$ and a faithful irreducible orthogonal representation ρ_2 of $\mathfrak{g}_2$ in some vector space $\mathfrak{u}_2$ with $\dim \mathfrak{u}_2 \geq 5$. Then there is up to isomorphism at most one simple graded Lie algebra $\mathfrak{a} = \mathfrak{g} \oplus \mathfrak{u}$ with Lie algebra $\mathfrak{g} = \mathrm{sl}(2) \times \mathfrak{g}_2$ and odd subspace $\mathfrak{u} = \mathfrak{u}_1 \otimes \mathfrak{u}_2$ such that $\mathrm{ad}_{\mathfrak{u}}$ is equal to the tensor product of ρ_1 and ρ_2.

Proof: In fact, from (3.28) we know that P_2 is fixed up to a factor and (4.17b) is totally skew-symmetric for at most one choice of this factor (since the curly bracket is not totally skew-symmetric). Therefore, our assertion follows from Lemma 3.1.

To find out which simple Lie algebras $\mathfrak{g}_2$ and representations ρ_2 of $\mathfrak{g}_2$ are really possible we use the notation and results concerning the roots and weights as described in Sec. 3.

Let μ be one of the two weights of the representation ρ_1. Then the weights of $\mathrm{ad}_{\mathfrak{u}}$ are exactly the linear forms of the type $\alpha = (\pm\mu, \tilde{\alpha})$ where $\tilde{\alpha}$ is a weight of the representation ρ_2 of $\mathfrak{g}_2$. We normalize the even invariant bilinear form ϕ on $\mathfrak{a}$ such that

$$(\mu \mid \mu) = -1. \tag{4.20}$$

If $\tilde{\alpha} \neq 0$, then 2α is certainly not a root of $\mathfrak{g}$ and Lemma 3.2 shows that $(\alpha \mid \alpha) = 0$, i.e.,

$$(\tilde{\alpha} \mid \tilde{\alpha}) = 1. \tag{4.21}$$

Hence we have proved Lemma 4.2.

Lemma 4.2: The restriction of ϕ to $\mathfrak{g}_2$ is a positive multiple of the Killing form of $\mathfrak{g}_2$. All nonzero weights of ρ_2 have the same length.

To proceed we distinguish two cases depending on whether zero is a weight of ρ_2 or not.

A. Zero is a weight of ρ_2

In this case, $(\pm\mu, 0)$ are weights of $\mathrm{ad}_{\mathfrak{u}}$ and Lemma 3.2 shows that these weights are simple. Hence the weight 0 of ρ_2 must be simple, too.

According to Table I of Appendix B the Lie algebra $\mathfrak{g}_2$ must be isomorphic to one of the algebras

$$A_1, \quad B_n, \; n \geq 2, \quad G_2,$$

and ρ_2 must be the elementary representation except in the case $\mathfrak{g}_2 = A_1$ where ρ_2 is the adjoint representation. The possibility $\mathfrak{g}_2 = A_1 = \mathrm{sl}(2)$ has already been treated.

Consider next the case $\mathfrak{g}_2 = B_n = \mathrm{o}(2n+1)$. The Lie algebra of the orthosymplectic graded Lie algebra $\mathrm{osp}(2, 2n+1)$, $n \geq 2$, is equal to $\mathrm{sl}(2) \times \mathrm{o}(2n+1)$ and the corresponding representation ρ_2 is equal to the (orthogonal) elementary representation of $\mathrm{o}(2n+1)$. According to Lemma 4.1 our graded Lie algebra $\mathfrak{a}$ must, therefore, be isomorphic to $\mathrm{osp}(2, 2n+1)$.

Finally, the seven-dimensional elementary representation of G_2 is orthogonal; hence G_2 and this representation are possible candidates for $\mathfrak{g}_2$ and ρ_2. In fact, we shall show that there is indeed an (exceptional) simple graded Lie algebra for which $\mathfrak{g} = \mathrm{sl}(2) \times G_2$ and such that ρ_2 is the seven-dimensional representation of G_2.

B. Zero is not a weight of ρ_2

In this case we need some additional information on the weights of ρ_2. Let $\tilde{\alpha}$, $\tilde{\beta}$ be two weights of ρ_2. Then

$$\alpha = (\mu, \tilde{\alpha}), \quad \beta = (\mu, \tilde{\beta}) \tag{4.22}$$

are two weights of $\mathrm{ad}_{\mathfrak{u}}$ and

$$(\alpha \mid \beta) = (\mu \mid \mu) + (\tilde{\alpha} \mid \tilde{\beta}) = -1 + (\tilde{\alpha} \mid \tilde{\beta}). \tag{4.23}$$

Because of (4.21) this is zero if and only if $\tilde{\alpha} = \tilde{\beta}$. Hence if $\tilde{\alpha} \neq \pm\tilde{\beta}$ then $\alpha + \beta$ is not a root of $\mathfrak{g}$ and using Lemma 3.3 we conclude that $\alpha - \beta = (0, \tilde{\alpha} - \tilde{\beta})$ must be a root of $\mathfrak{g}$, i.e., we have shown the following lemma.

Lemma 4.3: If $\tilde{\alpha}$, $\tilde{\beta}$ are two weights of ρ_2 such that $\tilde{\alpha} \neq \pm\tilde{\beta}$ then $\tilde{\alpha} - \tilde{\beta}$ is a root of $\mathfrak{g}_2$.

Corollary: Suppose that $\tilde{\alpha}$, $\tilde{\beta}$ are two weights of ρ_2.

(I) If $\mathfrak{g}_2$ is one of the algebras A_n, $n \geq 1$; D_m, $m \geq 3$; E_p, $6 \leq p \leq 8$, then $(\tilde{\alpha} \mid \tilde{\beta}) = 0, \pm 1$.

(II) If $\mathfrak{g}_2$ is one of the algebras B_n, C_n, $n \geq 2$; F_4, then $(\tilde{\alpha} \mid \tilde{\beta}) = 0, \pm\frac{1}{3}, \pm 1$.

(III) If $\mathfrak{g}_2 = G_2$, then $(\tilde{\alpha} \mid \tilde{\beta}) = 0, \pm\frac{1}{2}, \pm 1$.

Proof: The representation ρ_2 is orthogonal, hence $-\tilde{\beta}$ is a weight if and only if $\tilde{\beta}$ is a weight.

If $\tilde{\alpha} = \pm\tilde{\beta}$, then $(\tilde{\alpha} \mid \tilde{\beta}) = \pm 1$ because of (4.21).

Suppose now that $\tilde{\alpha} \neq \pm\tilde{\beta}$. Then $\tilde{\alpha} \pm \tilde{\beta}$ are roots of $\mathfrak{g}_2$. In Case (I) all roots of $\mathfrak{g}_2$ have the same length which implies $(\tilde{\alpha} \mid \tilde{\beta}) = 0$. In Case (II) [resp. (III)] the two roots $\tilde{\alpha} \pm \tilde{\beta}$ either have the same length, which implies $(\tilde{\alpha} \mid \tilde{\beta}) = 0$, or the squares of their lengths differ by a factor of 2 (resp. 3), which implies $(\tilde{\alpha} \mid \tilde{\beta}) = \pm\frac{1}{3}$ (resp. $(\tilde{\alpha} \mid \tilde{\beta}) = \pm\frac{1}{2}$).

We are now ready to proceed with our classification. Let us first look for all simple Lie agebras $\mathfrak{g}_2$ and all faithful irreducible orthogonal representations ρ_2 of $\mathfrak{g}_2$ such that all weights of ρ_2 have the same length. From Table II of Appendix B we know that there exist the following possibilities:

$\mathfrak{g}_2$	ρ_2	
A_n	$\rho(\lambda_{(n+1)/2})$	$n = 4m-1$; $m = 1, 2, \cdots$,
B_n	$\rho(\lambda_n)$	$n = 4m-1$ or $n = 4m$; $m = 1, 2, \cdots$,
D_n	$\rho(\lambda_1)$	$n = 3, 4, 5, \cdots$,
	$\rho(\lambda_{n-1})$, $\rho(\lambda_n)$	$n = 4m$; $m = 1, 2, \cdots$.

We consider these cases separately.

Case A_n: One can prove that part (I) of the corollary rules out all algebras A_n except A_3. Now A_3 is isomorphic to D_3 and the representation $\rho(\lambda_2)$ of A_3 corresponds to the representation $\rho(\lambda_1)$ of D_3. Hence we can drop A_3 in favor of D_3.

Case B_n: It is easy to see that the spin representation $\rho(\lambda_n)$ of B_n satisfies the condition of part (II) in the corollary, only if $n=3$. Hence we are left with

$\mathfrak{g}_2 = B_3 = o(7)$ and $\rho_2 = \rho(\lambda_3) =$ spin representation.

We shall show that this possibility indeed corresponds to an (exceptional) simple graded Lie algebra.

Case D_n: The choice $\rho_2 = \rho(\lambda_1)$ leads (because of Lemma 4.1) to the orthosymplectic algebras osp(2, 2n), $n \geq 3$. Part (I) of the corollary rules out the representations $\rho(\lambda_{n-1})$ and $\rho(\lambda_n)$ except in the case $n=4$, but the representations $\rho(\lambda_1)$, $\rho(\lambda_3)$, $\rho(\lambda_4)$ of D_4 are connected by automorphisms of D_4. Because of Lemmas 3.1 and 4.1 the simple graded Lie algebras corresponding to $\rho(\lambda_3)$ and $\rho(\lambda_4)$ are isomorphic to that constructed with $\rho(\lambda_1)$, i.e., to osp(2, 8); consequently they must not be mentioned separately.

5. $\mathfrak{g}$ IS NOT SIMPLE AND $\mathrm{ad}_\mathfrak{u}$ DECOMPOSES INTO TWO IRREDUCIBLE REPRESENTATIONS

In this section we consider simple graded Lie algebras $\mathfrak{a} = \mathfrak{g} \oplus \mathfrak{u}$ for which $\mathfrak{g}$ is reductive but not simple and for which $\mathfrak{u}$ decomposes into the direct sum of two $\mathfrak{g}$-irreducible subspaces $\mathfrak{u}'$ and $\mathfrak{u}''$,

$$\mathfrak{u} = \mathfrak{u}' \oplus \mathfrak{u}''. \tag{5.1}$$

It follows that

$$\langle \mathfrak{u}', \mathfrak{u}' \rangle = \langle \mathfrak{u}'', \mathfrak{u}'' \rangle = \{0\}, \quad \langle \mathfrak{u}', \mathfrak{u}'' \rangle = \mathfrak{g}. \tag{5.2}$$

We know that $\mathfrak{g}$ is the direct product of its center $\mathfrak{g}_0$ and of the semisimple Lie algebra $\mathfrak{g}' = \langle \mathfrak{g}, \mathfrak{g} \rangle$. Furthermore,

$$\dim \mathfrak{g}_0 \leq 1. \tag{5.3}$$

If $\dim \mathfrak{g}_0 = 1$, then the Killing form of $\mathfrak{a}$ is nondegenerate and there exists an element $E \in \mathfrak{g}_0$ such that

$$\begin{aligned} &\langle E, U' \rangle = U' \quad \text{if } U' \in \mathfrak{u}', \\ &\langle E, U'' \rangle = -U'' \quad \text{if } U'' \in \mathfrak{u}''. \end{aligned} \tag{5.4}$$

By using the irreducibility of $\mathfrak{u}'$ and $\mathfrak{u}''$ as well as (5.2) and (5.4) it is easy to see that $\mathfrak{g}$ cannot be Abelian, i.e., that the semisimple factor $\mathfrak{g}'$ of $\mathfrak{g}$ cannot be equal to $\{0\}$.

In the following we shall distinguish two cases depending on whether $\mathfrak{g}'$ is simple or not. In the former case we have $\mathfrak{g}_0 \neq \{0\}$ according to the general assumptions of this section.

A. $\mathfrak{g}'$ is not simple

In this case $\mathfrak{g}$ decomposes into a direct product

$$\mathfrak{g} = \mathfrak{g}_0 \times \mathfrak{g}_1 \times \mathfrak{g}_2, \tag{5.5}$$

where the Lie algebras $\mathfrak{g}_1$ and $\mathfrak{g}_2$ are semisimple (and different from $\{0\}$) and where $\mathfrak{g}_0$ is the center of $\mathfrak{g}$ (which may be equal to $\{0\}$).

The representations ρ' (resp. ρ'') of $\mathfrak{g}_1 \times \mathfrak{g}_2$ in $\mathfrak{u}'$ (resp. $\mathfrak{u}''$) induced by the adjoint representation $\mathrm{ad}_\mathfrak{u}$ of $\mathfrak{g}$ in $\mathfrak{u}$ are irreducible. Hence there exist irreducible representations ρ_i' (resp. ρ_i'') of $\mathfrak{g}_i$ in some vector spaces $\mathfrak{u}_i'$ (resp. $\mathfrak{u}_i''$), $i=1, 2$, such that

$$\mathfrak{u}' = \mathfrak{u}_1' \otimes \mathfrak{u}_2', \quad \mathfrak{u}'' = \mathfrak{u}_1'' \otimes \mathfrak{u}_2'' \tag{5.6}$$

and such that ρ' (resp. ρ'') is the tensor product of the representations ρ_1' and ρ_2' (resp. of ρ_1'' and ρ_2'').

Because of $\langle \mathfrak{u}', \mathfrak{u}'' \rangle = \mathfrak{g}$ there exists for $i=1, 2$ a nondegenerate $\mathfrak{g}_i$-invariant bilinear form ψ_i on $\mathfrak{u}_i' \times \mathfrak{u}_i''$ (which is uniquely determined up to a nonzero factor) and a nonzero $\mathfrak{g}_i$-invariant bilinear mapping

$$P_i : \mathfrak{u}_i' \times \mathfrak{u}_i'' \to \mathfrak{g}_i \tag{5.7}$$

such that

$$\begin{aligned} &\langle U_1' \otimes U_2', U_1'' \otimes U_2'' \rangle \\ &= \psi_2(U_2', U_2'') P_1(U_1', U_1'') + \psi_1(U_1', U_1'') P_2(U_2', U_2'') \\ &\quad + \psi_1(U_1', U_1'') \psi_2(U_2', U_2'') F \end{aligned} \tag{5.8}$$

for all $U_1' \in \mathfrak{u}_1'$, $U_2' \in \mathfrak{u}_2'$, $U_1'' \in \mathfrak{u}_1''$, $U_2'' \in \mathfrak{u}_2''$. Here F is a suitable element of $\mathfrak{g}_0$ which is nonzero if $\mathfrak{g}_0 \neq \{0\}$. It will be useful to define the number $\eta \in K$ by

$$\begin{aligned} &\eta = 0 \quad \text{if } \mathfrak{g}_0 = \{0\}, \\ &F = \eta E \quad \text{if } \mathfrak{g}_0 \neq \{0\}, \end{aligned} \tag{5.9}$$

where $E \in \mathfrak{g}_0$ is the element described in (5.4). Then $\eta = 0$ if and only if $\mathfrak{g}_0 = \{0\}$.

The existence of the bilinear forms ψ_i means that the representations ρ_i' and ρ_i'' are contragredient with respect to each other, $i=1, 2$. Since $\mathrm{ad}_\mathfrak{u}$ is faithful, we conclude that all the four representations ρ_i', ρ_i'' must be faithful, too.

Let us introduce the abbreviations

$$\tilde{G}_i' = \rho_i'(G_i), \quad \tilde{G}_i'' = \rho_i''(G_i) \quad \text{if } G_i \in \mathfrak{g}_i. \tag{5.10}$$

We shall once again exploit the Jacobi identity for three odd elements. Taking two elements from $\mathfrak{u}'$ and one element from $\mathfrak{u}''$ this identity reads

$$\begin{aligned} &\langle\langle U_1' \otimes U_2', U_1'' \otimes U_2'' \rangle, \bar{U}_1' \otimes \bar{U}_2' \rangle \\ &+ \langle\langle \bar{U}_1' \otimes \bar{U}_2', U_1'' \otimes U_2'' \rangle, U_1' \otimes U_2' \rangle = 0 \end{aligned} \tag{5.11}$$

for all

$$U_1', \bar{U}_1' \in \mathfrak{u}_1', \quad U_2', \bar{U}_2' \in \mathfrak{u}_2', \quad U_1'' \in \mathfrak{u}_1'', \quad U_2'' \in \mathfrak{u}_2''.$$

Inserting the expression (5.8) into (5.11), we see first that there exist constants σ_i, $\tau_i \in K$ such that

$$\tilde{P}_i'(U_i', U_i'') \bar{U}_i' = \sigma_i \psi_i(U_i', U_i'') \bar{U}_i' + \tau_i \psi_i(\bar{U}_i', U_i'') U_i' \tag{5.12}$$

for all U_i', $\bar{U}_i' \in \mathfrak{u}_i'$, $U_i'' \in \mathfrak{u}_i''$, $i=1, 2$; then we deduce that (5.11) is fulfilled if and only if

$$\sigma_1 + \sigma_2 + \eta = 0, \quad \tau_1 + \tau_2 = 0. \tag{5.13}$$

Now the bilinear forms ψ_i are $\mathfrak{g}_i$-invariant; in particular we must have

$$\psi_i(\tilde{P}_i'(U_i', U_i'') \bar{U}_i', \bar{U}_i'') + \psi_i(\bar{U}_i', \tilde{P}_i''(U_i', U_i'') \bar{U}_i'') = 0 \tag{5.14}$$

for all U_i', $\bar{U}_i' \in \mathfrak{u}_i'$, U_i'', $\bar{U}_i'' \in \mathfrak{u}_i''$, which implies

$$\tilde{P}_i''(U_i', U_i'') \bar{U}_i'' = -\sigma_i \psi_i(U_i', U_i'') \bar{U}_i'' - \tau_i \psi_i(U_i', \bar{U}_i'') U_i'' \tag{5.15}$$

if $U_i' \subset \mathfrak{u}_i'$, U_i'', $\bar{U}_i'' \in \mathfrak{u}_i''$. The Jacobi identity for one element of $\mathfrak{u}'$ and two elements of $\mathfrak{u}''$ is then automatically satisfied as a consequence of (5.13).

Let us now recall that the images of the elements of a semisimple Lie algebra under any finite-dimensional representation are traceless. Hence we must have

$$\mathrm{Tr}\tilde{P}_i'(U_i', U_i'') = 0 \tag{5.16}$$

for all $U_i' \in \mathfrak{u}_i'$, $U_i'' \in \mathfrak{u}_i''$, $i = 1, 2$, and similarly for $\tilde{P}_i''$. If we define for $i = 1, 2$

$$n_i = \dim \mathfrak{u}_i' = \dim \mathfrak{u}_i'', \tag{5.17}$$

then our trace condition is equivalent to

$$n_i\sigma_i + \tau_i = 0, \quad i = 1, 2. \tag{5.18}$$

Considering the dimensions n_1 and n_2 as fixed, we can rephrase (5.13) and (5.18) by demanding that there exist a nonzero constant $\tau \in K$ such that

$$\begin{aligned} &\sigma_1 = \tau/n_1, \quad \tau_1 = -\tau, \\ &\sigma_2 = -\tau/n_2, \quad \tau_2 = \tau \\ &\eta = \frac{n_1 - n_2}{n_1 n_2}\tau. \end{aligned} \tag{5.19}$$

Note that $\eta = 0$ if and only if $n_1 = n_2$.

It is now easy to show that the linear mappings $\tilde{P}_i'(U_i', U_i'')$, $U_i' \in \mathfrak{u}_i'$, $U_i'' \in \mathfrak{u}_i''$ generate a subspace of $\mathrm{gl}(\mathfrak{u}_i')$ which is equal to the special linear Lie algebra $\mathrm{sl}(\mathfrak{u}_i')$ of $\mathfrak{u}_i'$. Since the representation ρ_i' is faithful, we conclude that ρ_i' is an isomorphism of $\mathfrak{g}_i$ onto $\mathrm{sl}(\mathfrak{u}_i') \approx \mathrm{sl}(n_i)$. Obviously under this isomorphism the representation ρ_i' corresponds to the elementary representation of $\mathrm{sl}(\mathfrak{u}_i')$.

Furthermore, according to (5.8), (5.12), (5.15), and (5.19) the product mapping $\mathfrak{u} \times \mathfrak{u} \to \mathfrak{g}$ is determined up to the factor τ. In view of Lemma 3.1 we have thus shown that $\mathfrak{a}$ must be isomorphic to the special linear graded Lie algebra $\mathrm{spl}(n_1, n_2)$ if $n_1 \neq n_2$ and to $\mathrm{spl}(n_1, n_1)/z_{n_1}$ [with z_n the one-dimensional center of $\mathrm{spl}(n_1, n_1)$] if $n_1 = n_2$. Note that our assumptions imply $n_1, n_2 \geq 2$.

B. $\mathfrak{g}'$ is simple

In this case $\mathfrak{g}$ decomposes into a direct product

$$\mathfrak{g} = \mathfrak{g}_0 \times \mathfrak{g}_1, \tag{5.20}$$

where $\mathfrak{g}_0$ is the one-dimensional center of $\mathfrak{g}$ and where the Lie algebra $\mathfrak{g}_1$ is simple.

The Killing form of $\mathfrak{a}$ is nondegenerate, hence we may use the conventions and results of Sec. 3. We normalize the even invariant bilinear form ϕ on $\mathfrak{a}$ (which is proportional to the Killing form) by the condition

$$\phi(E, E) = -1, \tag{5.21}$$

where $E \in \mathfrak{g}_0$ has been defined in (5.4).

Let ρ' (resp. ρ'') be the representation of $\mathfrak{g}_1$ in $\mathfrak{u}'$ (resp. $\mathfrak{u}''$) induced by $\mathrm{ad}_\mathfrak{u}$. The restriction of ϕ to $\mathfrak{u}' \times \mathfrak{u}''$ is nondegenerate, hence ρ' and ρ'' are contragredient with respect to each other. Furthermore, since $\mathrm{ad}_\mathfrak{u}$ is faithful we conclude that ρ' and ρ'' must be faithful too.

We shall now first settle the question of uniqueness. Our argument will be completely analogous to that used to prove Lemma 4.1. Define the nondegenerate $\mathfrak{g}_1$-invariant bilinear form ψ_1 on $\mathfrak{u}' \times \mathfrak{u}''$ and the nonzero $\mathfrak{g}_1$-invariant bilinear mapping

$$P_1 : \mathfrak{u}' \times \mathfrak{u}'' \to \mathfrak{g}_1 \tag{5.22}$$

by the equation

$$\langle U', U'' \rangle = P_1(U', U'') + \psi_1(U', U'')E \tag{5.23}$$

for all $U' \in \mathfrak{u}'$, $U'' \in \mathfrak{u}''$.

Let us introduce the abbreviations

$$\tilde{G}_1' = \rho'(G_1), \quad \tilde{G}_1'' = \rho''(G_1) \quad \text{if } G_1 \in \mathfrak{g}_1. \tag{5.24}$$

Then the Jacobi identity with two elements from $\mathfrak{u}'$ and one element from $\mathfrak{u}''$ is equivalent to the condition that the expression

$$\langle\langle U', U'' \rangle, \bar{U}' \rangle = \tilde{P}_1'(U', U'')\bar{U}' + \psi_1(U', U'')\bar{U}' \tag{5.25}$$

should be skew-symmetric in U', $\bar{U}' \in \mathfrak{u}'$ for all fixed $U'' \in \mathfrak{u}''$. Since $\psi_1(U', U'')\bar{U}'$ is certainly not skew-symmetric in U', $\bar{U}'$, we conclude from (3.28) and the condition above that the product mapping $\mathfrak{u} \times \mathfrak{u} \to \mathfrak{g}$ is fixed up to a factor once the simple Lie algebra $\mathfrak{g}_1$ and the two (contragredient) representations ρ' and ρ'' are given.

In view of Lemma 3.1 we have thus shown the following lemma.

Lemma 5.1: Suppose that we are given a simple Lie algebra $\mathfrak{g}_1$ and two faithful irreducible representations ρ' and ρ'' of $\mathfrak{g}_1$ in some vector spaces $\mathfrak{u}'$ (resp. $\mathfrak{u}''$) which are contragredient with respect to each other. Then there is up to isomorphism at most one simple graded Lie algebra $\mathfrak{a}$ with Lie algebra $\mathfrak{g} = \mathfrak{g}_0 \times \mathfrak{g}_1$ and odd subspace $\mathfrak{u} = \mathfrak{u}' \oplus \mathfrak{u}''$, such that the restriction of $\mathrm{ad}_\mathfrak{u}$ to $\mathfrak{g}_1$ is equal to the direct sum of ρ' and ρ'' and such that $\mathfrak{g}_0$ acts on $\mathfrak{u}$ as described in (5.4).

Let us now discuss the weights of $\mathrm{ad}_\mathfrak{u}$, the adjoint representation of $\mathfrak{g}$ in $\mathfrak{u}$. Recall that ρ' and ρ'' are contragredient with respect to each other. Hence $\tilde{\alpha}$ is a weight of ρ' if and only if $-\tilde{\alpha}$ is a weight of ρ''.

Define ν to be the linear form on $\mathfrak{g}_0$ such that

$$\nu(E) = 1. \tag{5.26}$$

Then the weights of the representation $\mathrm{ad}_{\mathfrak{u}'}$ (resp. $\mathrm{ad}_{\mathfrak{u}''}$) of $\mathfrak{g}$ induced by $\mathrm{ad}_\mathfrak{u}$ in $\mathfrak{u}'$ (resp. $\mathfrak{u}''$) are exactly the linear forms of the type $\alpha = (\nu, \tilde{\alpha})$ [resp. $-\alpha = (-\nu, -\tilde{\alpha})$], where $\tilde{\alpha}$ is a weight of ρ'.

Evidently 2α is not a root of $\mathfrak{g}$, hence (Lemma 3.2)

$$(\alpha \mid \alpha) = 0. \tag{5.27}$$

But Eq. (5.21) implies

$$(\nu \mid \nu) = -1. \tag{5.28}$$

Hence we conclude

$$(\tilde{\alpha} \mid \tilde{\alpha}) = 1 \tag{5.29}$$

and we have proved Lemma 5.2.

Lemma 5.2: The restriction of ϕ to $\mathfrak{g}_1$ is a positive multiple of the Killing form of $\mathfrak{g}_1$. All weights of ρ' have the same length, in particular zero is not a weight of ρ'.

As a consequence of Lemma 3.3 we derive Lemma 5.3.

Lemma 5.3: If $\tilde{\alpha}$, $\tilde{\beta}$ are two different weights of ρ' then $\tilde{\alpha}-\tilde{\beta}$ is a root of $\mathfrak{g}_1$.

Proof: Define the following weights of $\mathrm{ad}_{\mathfrak{u}'}$ by

$$\alpha=(\nu, \tilde{\alpha}), \quad \beta=(\nu, \tilde{\beta}). \tag{5.30}$$

Then

$$(\alpha|\beta)=(\nu|\nu)+(\tilde{\alpha}|\tilde{\beta})=-1+(\tilde{\alpha}|\tilde{\beta}). \tag{5.31}$$

Because of (5.29) this expression is zero if and only if $\tilde{\alpha}=\tilde{\beta}$, which is not the case. Furthermore, $\alpha+\beta=(2\nu, \tilde{\alpha}+\tilde{\beta})$ is not a root of $\mathfrak{g}$, hence (Lemma 3.3) $\alpha-\beta=(0, \tilde{\alpha}-\tilde{\beta})$ must be a root of $\mathfrak{g}$, as desired.

Corollary: Suppose that $\mathfrak{g}_1$ is one of the algebras B_n, C_n, $n\geq 2$, F_4. If $\tilde{\alpha}$, $\tilde{\beta}$ are two weights of ρ' such that $\tilde{\alpha}\neq\pm\tilde{\beta}$, then $(\tilde{\alpha}|\tilde{\beta})=0$.

Proof: It is well known that the representations of the algebras B_n, C_n, F_4 are self-contragredient. Hence if $\tilde{\alpha}$ is a weight of ρ' then $-\tilde{\alpha}$ is also a weight. According to Lemma 5.3 it is evident that $2\tilde{\alpha}$ must be a long root of $\mathfrak{g}_1$ and $\tilde{\alpha}-\tilde{\beta}$ must be a short root, i.e., we must have

$$(2\tilde{\alpha}|2\tilde{\alpha})=2(\tilde{\alpha}-\tilde{\beta}|\tilde{\alpha}-\tilde{\beta}). \tag{5.32}$$

This implies

$$(\tilde{\alpha}|\tilde{\beta})=0 \tag{5.33}$$

as desired.

Let us now proceed with our classification. We shall distinguish two cases depending on whether the roots of $\mathfrak{g}_1$ have all the same length or not.

(a) $\mathfrak{g}_1$ has roots of different length

According to Lemma 5.2 and to Table II in Appendix B we have the following possibilities for $\mathfrak{g}_1$ and its representation ρ':

$\mathfrak{g}_1$	ρ',	
B_n	$\rho(\lambda_n)$,	$n\geq 2$.
C_n	$\rho(\lambda_1)$,	

The algebras B_n with $n\geq 3$ are ruled out by the corollary. Furthermore, B_2 is isomorphic to C_2 and the representation $\rho(\lambda_2)$ of B_2 corresponds to the representation $\rho(\lambda_1)$ of C_2. Hence we may drop B_2 in favor of C_2 and we are left with

$\mathfrak{g}_1=C_n$, $\rho'=\rho(\lambda_1)=$ elementary representation,

where $n\geq 2$.

It is now evident from Lemma 5.1 that $\mathfrak{a}$ must be isomorphic to the orthosymplectic graded Lie algebra $\mathrm{osp}(2n, 2)$.

(b) All roots of $\mathfrak{g}_1$ have the same length

In this case, too, we could use Appendix B to obtain severe restrictions on $\mathfrak{g}_1$ and ρ', but we prefer to argue more directly.

In fact, according to our assumptions and to Lemma 5.3 there exists a constant $\omega\in K$, $\omega\neq 0$, such that

$$(\tilde{\alpha}-\tilde{\beta}|\tilde{\alpha}-\tilde{\beta})=-2\omega, \tag{5.34}$$

i.e.,

$$(\tilde{\alpha}|\tilde{\beta})=1+\omega \tag{5.35}$$

if $\tilde{\alpha}, \tilde{\beta}$ are two different weights of ρ'. It follows that

$$(\alpha|\beta)=-1+(\tilde{\alpha}|\tilde{\beta})=\omega \tag{5.36}$$

if $\alpha=(\nu, \tilde{\alpha})$ and $\beta=(\nu, \tilde{\beta})$ are two different weights of $\mathrm{ad}_{\mathfrak{u}'}$.

Suppose now that α, β, γ are three weights of $\mathrm{ad}_{\mathfrak{u}'}$ and choose arbitrary elements

$$U'_\alpha\in\mathfrak{u}'\cap\mathfrak{u}^\alpha, \quad U''_{-\beta}\in\mathfrak{u}''\cap\mathfrak{u}^{-\beta}, \quad U'_\gamma\in\mathfrak{u}'\cap\mathfrak{u}^\gamma \tag{5.37}$$

[see (3.10)]. Then

$$\langle\langle U'_\alpha, U''_{-\beta}\rangle, U'_\gamma\rangle\in\mathfrak{u}'\cap\mathfrak{u}^{\alpha-\beta+\gamma}. \tag{5.38}$$

If $\beta\neq\alpha, \gamma$ then

$$(\alpha-\beta+\gamma|\alpha-\beta+\gamma)=2(\alpha|\gamma)-4\omega\neq 0 \tag{5.39}$$

and hence [because of (5.27)]

$$\langle\langle U'_\alpha, U''_{-\beta}\rangle, U'_\gamma\rangle=0. \tag{5.40}$$

On the other hand [see (3.15)],

$$\begin{aligned}\langle\langle U'_\alpha, U''_{-\alpha}\rangle, U'_\gamma\rangle&=(\alpha|\gamma)\phi(U'_\alpha, U''_{-\alpha})U'_\gamma,\\ \langle\langle U'_\alpha, U''_{-\gamma}\rangle, U'_\gamma\rangle&=-(\alpha|\gamma)\phi(U'_\gamma, U''_{-\gamma})U'_\alpha.\end{aligned} \tag{5.41}$$

Using (5.27) and (5.36) the Eqs. (5.40) and (5.41) can be combined to give the general result

$$\langle\langle U'_\alpha, U''_{-\beta}\rangle, U'_\gamma\rangle=\omega\{\phi(U'_\alpha, U''_{-\beta})U'_\gamma-\phi(U'_\gamma, U''_{-\beta})U'_\alpha\} \tag{5.42}$$

for all weights α, β, γ of $\mathrm{ad}_{\mathfrak{u}'}$, i.e., we have

$$\langle\langle U', U''\rangle, \bar{U}'\rangle=\omega\{\phi(U', U'')\bar{U}'-\phi(\bar{U}', U'')U'\} \tag{5.43}$$

for all $U', \bar{U}'\in\mathfrak{u}'$ and $U''\in\mathfrak{u}''$.

It is now easy to see that the linear mappings

$$\bar{U}'\to\omega\{\phi(U', U'')\bar{U}'-\phi(\bar{U}', U'')U'\} \tag{5.44}$$

from $\mathfrak{u}'$ into itself, with $U'\in\mathfrak{u}'$ and $U''\in\mathfrak{u}''$, generate $\mathrm{gl}(\mathfrak{u}')$ as a vector space. Since $\mathrm{ad}_{\mathfrak{u}'}$ is faithful, we conclude that $\mathrm{ad}_{\mathfrak{u}'}$ is an isomorphism of $\mathfrak{g}$ onto the general linear Lie algebra $\mathrm{gl}(\mathfrak{u}')$. Using Lemma 5.1 it follows that $\mathfrak{a}$ must be isomorphic to a special linear graded Lie algebra $\mathrm{spl}(n, 1)$ with $n\geq 2$.

6. $\mathfrak{g}$ IS SIMPLE

We shall now consider simple graded Lie algebras $\mathfrak{a}=\mathfrak{g}\oplus\mathfrak{u}$ whose Lie algebra $\mathfrak{g}$ is simple.

Let us begin with a remark on the invariant bilinear forms on a Lie algebra $\mathfrak{g}$. As is well known, with any representation ρ of $\mathfrak{g}$ there is associated an invariant bilinear form ϕ_ρ on $\mathfrak{g}$ defined by

$$\phi_\rho(G, G') = \mathrm{Tr}(\rho(G)\rho(G')) \tag{6.1}$$

if $G, G' \in \mathfrak{g}$. Taking for ρ the adjoint representation of $\mathfrak{g}$ we obtain the Killing form $\phi_\mathfrak{g}$ of $\mathfrak{g}$.

Suppose now that $\mathfrak{g}$ is simple. Then all invariant bilinear forms on $\mathfrak{g}$ are proportional, hence for every representation ρ of $\mathfrak{g}$ there exists an element $l_\rho \in K$ such that

$$\phi_\rho = l_\rho \phi_\mathfrak{g}. \tag{6.2}$$

The number l_ρ is called the *index of the representation* ρ. It is easy to see that l_ρ is a positive rational number which is nonzero if ρ is faithful. If ρ is the direct sum of the subrepresentations ρ_1 and ρ_2 then, obviously, $l_\rho = l_{\rho_1} + l_{\rho_2}$; hence it is sufficient to calculate l_ρ for the irreducible representations of $\mathfrak{g}$. In fact one can derive a formula[12] which gives the index l_ρ of an irreducible representation ρ in terms of the highest weight of ρ.

Let us apply these results to our graded Lie algebra $\mathfrak{a}$. If $\phi_\mathfrak{a}$ is the (generalized) Killing form of $\mathfrak{a}$, if $\phi_\mathfrak{g}$ is the Killing form of $\mathfrak{g}$, and if $\phi_\mathfrak{u}$ is the invariant bilinear form on $\mathfrak{g}$ associated with the adjoint representation ad $\mathfrak{u}$ of $\mathfrak{g}$ in $\mathfrak{u}$, then

$$\phi_\mathfrak{a}(G, G') = \phi_\mathfrak{g}(G, G') - \phi_\mathfrak{u}(G, G') \tag{6.3}$$

for all $G, G' \in \mathfrak{g}$.

Now assume in addition that $\mathfrak{g}$ is simple. Let $l_\mathfrak{u}$ be the index of $\mathrm{ad}_\mathfrak{u}$. Then (6.2) and (6.3) yield

$$\phi_\mathfrak{a}(G, G') = (1 - l_\mathfrak{u})\phi_\mathfrak{g}(G, G') \tag{6.4}$$

for all $G, G' \in \mathfrak{g}$.

It is known[5] that an invariant bilinear form on a simple graded Lie algebra is either nondegenerate or zero. Hence for the algebras $\mathfrak{a}$ which we consider in this section we have either $l_\mathfrak{u} \neq 1$ and the Killing form $\phi_\mathfrak{a}$ is nondegenerate, or else we have $l_\mathfrak{u} = 1$ and the Killing form $\phi_\mathfrak{a}$ is zero. We shall discuss both cases separately.

A. The Killing form of $\mathfrak{a}$ is nondegenerate. $l_\mathfrak{u} \neq 1$

This class of graded Lie algebras has been treated in Ref. 4. But since, with the results at hand, it is easy to settle this case, we include it for completeness.

In fact, because of (6.4) we conclude from Lemma 3.2 that every nonzero weight of $\mathrm{ad}_\mathfrak{u}$ is half a root of $\mathfrak{g}$ and that all these weights are simple. Using Lemma C.1 of Appendix C it is then easy to see that $\mathfrak{g}$ must be isomorphic to some algebra C_n, $n \geq 1$, and that $\mathrm{ad}_\mathfrak{u}$ must be irreducible and equivalent to the elementary representation $\rho(\lambda_1)$ of C_n. Finally we deduce from (3.28) and Lemma 3.1 that $\mathfrak{a}$ must be isomorphic to the orthosymplectic graded Lie algebra osp$(2n, 1)$.

B. The Killing form of $\mathfrak{a}$ is zero, $l_\mathfrak{u} = 1$

It is appropriate to distinguish two cases depending on whether $\mathfrak{u}$ is irreducible or not.

(a) $\mathfrak{u}$ is irreducible

Since $l_\mathfrak{u} = 1$ we conclude from Appendix D that $\mathrm{ad}_\mathfrak{u}$ is equivalent to the adjoint representation of $\mathfrak{g}$. In order to define the (symmetric) product mapping $\mathfrak{u} \times \mathfrak{u} \to \mathfrak{g}$ we answer the following question: Let $\mathrm{ad}_\mathfrak{g}$ be the adjoint representation of the (simple) Lie algebra $\mathfrak{g}$. Is $\mathrm{ad}_\mathfrak{g}$ contained in the symmetric tensor product of $\mathrm{ad}_\mathfrak{g}$ with itself?

It turns out[13] that this is the case if and only if $\mathfrak{g}$ is one of the algebras A_n, $n \geq 2$, and in this case the symmetric tensor product of $\mathrm{ad}_\mathfrak{g}$ with itself contains $\mathrm{ad}_\mathfrak{g}$ only once. According to Lemma 3.1 it is now evident that $\mathfrak{a}$ must be isomorphic to the (f, d) algebra $d(n+1)/z_{n+1}$ of Gell-Mann, Michel, and Radicati. We note that this result can also be derived without using the results of Ref. 13; instead one may take advantage of the Jacobi identity for three odd elements.

(b) $\mathfrak{u}$ is reducible

In this case we know that $\mathfrak{u}$ decomposes into the direct sum of two $\mathfrak{g}$-irreducible subspaces $\mathfrak{u}'$ and $\mathfrak{u}''$,

$$\mathfrak{u} = \mathfrak{u}' \oplus \mathfrak{u}''. \tag{6.5}$$

Let $\mathrm{ad}_{\mathfrak{u}'}$ (resp. $\mathrm{ad}_{\mathfrak{u}''}$) be the representation of $\mathfrak{g}$ in $\mathfrak{u}'$ (resp. $\mathfrak{u}''$) induced by $\mathrm{ad}_\mathfrak{u}$ and let l' (resp. l'') be its index. According to our assumption, we have

$$l' + l'' = 1. \tag{6.6}$$

Since $\mathfrak{a}$ is simple we know that $\langle \mathfrak{g}, \mathfrak{u} \rangle = \mathfrak{u}$; hence $\mathrm{ad}_{\mathfrak{u}'}$ and $\mathrm{ad}_{\mathfrak{u}''}$ are nontrivial, which implies $l', l'' \neq 0$.

Now we are faced with the following problem: Suppose we are given a simple Lie algebra $\mathfrak{g}$. Find all pairs of faithful irreducible representations of $\mathfrak{g}$ the sum of whose indices is equal to one.

In the following we discuss all simple Lie algebras separately. Using Table III of Appendix D we give all "admissible pairs" of irreducible representations and discuss which of these pairs lead to a simple graded Lie algebra.

Case A_n, $n \geq 1$: This case is the most complicated one.

Admissible pairs of representations:

(1)	$\rho(2\lambda_1)$,	$\rho(\lambda_{n-1})$;	$n \geq 2$,
(2)	$\rho(2\lambda_1)$,	$\rho(\lambda_2)$;	$n \geq 2$,
(1′)	$\rho(2\lambda_n)$,	$\rho(\lambda_2)$;	$n \geq 2$,
(2′)	$\rho(2\lambda_n)$,	$\rho(\lambda_{n-1})$;	$n \geq 2$,
(3)	$\rho(\lambda_3)$,	$\rho(\lambda_3)$;	$n = 5$,
(4)	$\rho(\lambda_1)$,	$\rho(\lambda_3)$;	$n = 7$,
(5)	$\rho(\lambda_1)$,	$\rho(\lambda_5)$;	$n = 7$,
(4′)	$\rho(\lambda_7)$,	$\rho(\lambda_5)$;	$n = 7$,
(5′)	$\rho(\lambda_7)$,	$\rho(\lambda_3)$;	$n = 7$.

The "primed" possibilities are connected with the nonprimed possibilities by an automorphism of A_n. In view of Lemma 3.1 the primed cases may, therefore, be omitted.

(1) The tensor product of $\rho(2\lambda_1)$ with $\rho(\lambda_{n-1})$ contains the adjoint representation of A_n exactly once. In view of Lemma 3.1 it is then clear that the corresponding graded Lie algebra is isomorphic to $b(n+1)$.

(2) We may assume $n \neq 3$ since the case $n=3$ is included in (1). Then the tensor product of $\rho(2\lambda_1)$ with $\rho(\lambda_2)$ does not contain the adjoint representation, hence this case does not lead to a simple graded Lie algebra.

(3) The tensor product of $\rho(\lambda_3)$ with itself contains the adjoint representation exactly once, namely in the symmetric part. The latter property implies that the corresponding product mapping $\mathfrak{u} \times \mathfrak{u} \to \mathfrak{g}$ does not lead to a simple graded Lie algebra.

(4), (5) The tensor product of $\rho(\lambda_1)$ with $\rho(\lambda_3)$ or with $\rho(\lambda_5)$ does not contain the adjoint representation, hence these cases do not lead to a simple graded Lie algebra.

Case C_n, $n \geq 2$: Admissible pair of representations:

$\rho(\lambda_2)$, $\rho(\lambda_2)$; $n=3$.

The tensor product of $\rho(\lambda_2)$ with itself contains the adjoint representation exactly once, namely in the skew-symmetric part. How to define the representation $\rho(\lambda_2)$ in tensor space is well known, and it is then straightforward to construct the candidate for the product mapping $\mathfrak{u} \times \mathfrak{u} \to \mathfrak{g}$. Once this has been done it is easy to see that the Jacobi identity for three odd elements is not satisfied.

Cases B_n, $n \geq 3$; D_m, $m \geq 4$; E_6, E_7, E_8, F_4, G_2: No admissible pairs of representations.

Summarizing the results of this subsection we have shown that only the algebras $b(n)$, $n \geq 3$, belong to case (B), (b).

APPENDIX A

In the appendices we collect our notational conventions concerning *simple* Lie algebras and we discuss some classes of irreducible representations with "low dimensions." Our notation is mainly that of Tits[14]; for our calculations we have made use also of the results collected in the appendices of the treatises by Freudenthal, de Vries,[15] and by Bourbaki.[16]

Let $\mathfrak{g}$ be a simple Lie algebra and let $\mathfrak{h}$ be a Cartan subalgebra of $\mathfrak{g}$. Choose any nondegenerate invariant bilinear form on $\mathfrak{g}$ (all these forms are proportional). By restriction it induces a nondegenerate bilinear form on $\mathfrak{h}$ and, consequently, also a nondegenerate bilinear form on the dual space $\mathfrak{h}^*$ of $\mathfrak{h}$. The bilinear form on $\mathfrak{h}^*$ will be denoted by a bracket $(\;|\;)$.

The roots of $\mathfrak{g}$ as well as the weights of the representations of $\mathfrak{g}$ are elements of $\mathfrak{h}^*$. Let us choose a fundamental system of simple roots $\alpha_1, \ldots, \alpha_n$. Then the fundamental weights $\lambda_1, \ldots, \lambda_n$ are defined by

$$2\frac{(\lambda_i \mid \alpha_j)}{(\alpha_j \mid \alpha_j)} = \delta_{ij}. \tag{A1}$$

Any (finite-dimensional) irreducible representation of $\mathfrak{g}$ is characterized (up to equivalence) by its highest weight. An element $\lambda \in \mathfrak{h}^*$ is the highest weight of a finite-dimensional irreducible representation of $\mathfrak{g}$ if and only if it has the form

$$\lambda = \sum_{i=1}^{n} c_i \lambda_i \tag{A2}$$

with integers $c_i \geq 0$. The corresponding irreducible representation will be denoted by $\rho(\lambda)$.

Unfortunately there seems to be no generally accepted enumeration of the vertices in the Dynkin diagrams (and hence of the simple roots and of the fundamental weights). Therefore, we have to specify our convention as in Fig. 1. We remark that the arrow points towards the short roots.

The representation $\rho(\lambda_1)$ is called *elementary*. In the cases of the Lie algebras A_n, B_n, C_n, D_n this is just the matrix representation by which the algebra is usually defined.

Suppose we are given an irreducible representation ρ of $\mathfrak{g}$. If ρ is equivalent to its contragredient representation then we call ρ *self-contragredient*. This is the case if and only if there exists a nondegenerate invariant bilinear form ψ on the representation space of ρ. It is well known that ψ (if it exists) is uniquely determined up to a nonzero factor; in particular ψ is either symmetric or skew-symmetric. In the former (resp. latter) case the representation ρ is called *orthogonal* (resp. *symplectic*).

APPENDIX B

We discuss some classes of representations with low dimensions. We are well aware of the fact that the results to be derived in the following should be contained somewhere in the mathematical literature.

Let $\mathfrak{g}$ be any *simple* Lie algebra. We want to find all irreducible representations of $\mathfrak{g}$ whose nonzero weights have all the same length.

Let ρ be any nontrivial representation of this type. If $\mu \neq 0$ is a weight of ρ and if α is a root of $\mathfrak{g}$ such that $\mu - \alpha$ is a nonzero weight of ρ, then

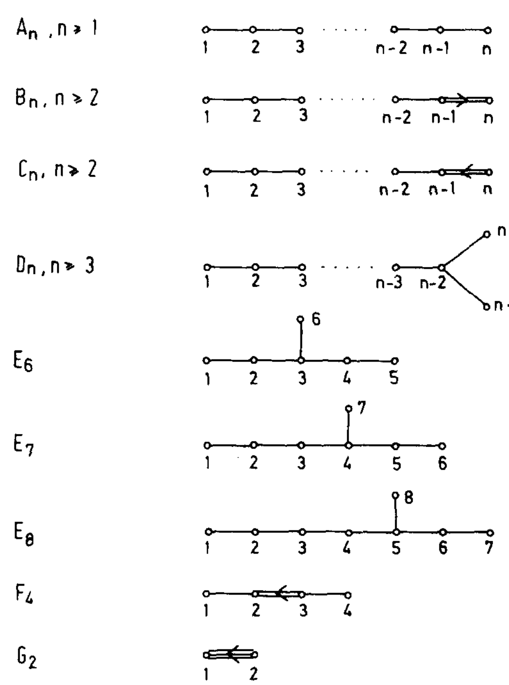

FIG. 1.

TABLE I. Irreducible representations of simple Lie algebras which have zero as a weight and whose nonzero weights have all the same length.

algebra	representation	multiplicity of weight 0
A_n, $n \geq 1$	$\rho(\lambda_1 + \lambda_n)$	n
B_n, $n \geq 2$	$\rho(\lambda_1)$	1
C_n, $n \geq 2$	$\rho(\lambda_2)$	$n-1$
D_n, $n \geq 4$	$\rho(\lambda_2)$	n
E_6	$\rho(\lambda_6)$	6
E_7	$\rho(\lambda_6)$	7
E_8	$\rho(\lambda_1)$	8
F_4	$\rho(\lambda_1)$	2
G_2	$\rho(\lambda_1)$	1

All representations appearing in this table are orthogonal.

$$(\mu - \alpha \mid \mu - \alpha) = (\mu \mid \mu) \tag{B1}$$

and, therefore,

$$\mu - \alpha = \mu - 2\,\frac{(\mu \mid \alpha)}{(\alpha \mid \alpha)}\,\alpha = S_\alpha(\mu), \tag{B2}$$

where S_α is the Weyl reflection defined by the root α.

Now we have the following well-known lemma.

Lemma B.1: (a) The Weyl group operates transitively on the roots of equal length.

(b) The Weyl group permutes the weights of any representation of $\mathfrak{g}$.

(c) If two weights of a representation of $\mathfrak{g}$ are connected by a Weyl transformation, then the corresponding weight-spaces have the same dimension.

Using this lemma as well as (B2) and the fact that ρ is irreducible one can prove the following lemma.

Lemma B.2: The Weyl group operates transitively on the nonzero weights of ρ, in partiuclar all these weights are simple.

We shall now distinguish two cases depending on whether 0 is a weight of ρ or not.

Let us suppose first that 0 is a weight of ρ. Using the irreducibility of ρ as well as Lemmas B.1 and B.2 we deduce that the nonzero weights of ρ are exactly the short roots of $\mathfrak{g}$. By this property the representation ρ is uniquely fixed, and it is then easy to check that there indeed exists an irreducible representation with the desired properties.

Table I gives for every simple Lie algebra the (uniquely determined) nontrivial irreducible representation which has 0 as a weight and whose nonzero weights have all the same length. In the case of the algebras A_n, D_n, E_n this representation is of course the adjoint representation. In Table I all representations are orthogonal.

We next consider the case where 0 is not a weight of ρ. Let λ be the highest weight of ρ,

$$\lambda = \sum_{i=1}^{n} c_i \lambda_i \tag{B3}$$

with integers $c_i \geq 0$.

If α is any positive root of $\mathfrak{g}$ then $\lambda + \alpha$ is not a weight of ρ. Using the well-known formula for the α-ladder through a weight it is easy to see that

$$\begin{aligned} &(\lambda \mid \alpha) = 0 \quad \text{if } \lambda - \alpha \text{ is not a weight,} \\ &2(\lambda \mid \alpha) = (\alpha \mid \alpha) \quad \text{if } \lambda - \alpha \text{ is a weight.} \end{aligned} \tag{B4}$$

The positive root α can be expressed in the form

$$\alpha = \sum_{j=1}^{n} m_j \alpha_j \tag{B5}$$

with suitable integers $m_j \geq 0$; consequently we deduce from (A1), (B3), and (B5) that

$$2(\lambda \mid \alpha) = \sum_{i=1}^{n} c_i m_i (\alpha_i \mid \alpha_i). \tag{B6}$$

The positive roots α, i.e., the allowed n-tuples $(m_1, \ldots, m_n)$, may be taken from Refs. 14–16. For every simple Lie algebra $\mathfrak{g}$ there exists a positive root for which the condition (B4) is most stringent; in fact this is just the root which is the highest weight of the representation given in Table I. The condition which we obtain in this way means that ρ must be equal to one of the representations given in Table II. Conversely one can prove that the weights of the representations given in this table indeed do have the same length.

In the last column of Table II we describe which of the representations are self-contragredient and, if this is the case, whether they are orthogonal or symplectic.[14]

APPENDIX C

As a by-product of Appendix B we prove the following lemma.

Lemma C.1: Let $\mathfrak{g}$ be a simple Lie algebra and let ρ be a nontrivial irreducible representation of $\mathfrak{g}$ whose nonzero weights are equal to half a root of $\mathfrak{g}$. Then $\mathfrak{g}$ is isomorphic to some algebra C_n, $n \geq 1$, and the representation ρ is equivalent to the elementary representation $\rho(\lambda_1)$.

Proof: Since ρ is irreducible and since the double of a

TABLE II. Irreducible representations of simple Lie algebras whose roots have all the same length.

algebra	representation	type of representation if self-contragredient
A_n, $n \geq 1$	$\rho(\lambda_i)$, $1 \leq i \leq n$	$\rho(\lambda_{(n+1)/2})$ is orthog. if $n = 4m-1$ sympl. if $n = 4m+1$
B_n, $n \geq 2$	$\rho(\lambda_n)$	orthog. if $n = 4m-1$, $4m$ sympl. if $n = 4m+1$, $4m+2$
C_n, $n \geq 2$	$\rho(\lambda_1)$	symplectic
D_n, $n \geq 3$	$\rho(\lambda_1)$	orthogonal
	$\rho(\lambda_{n-1})$, $\rho(\lambda_n)$	orthog. if $n = 4m$ sympl. if $n = 4m+2$
E_6	$\rho(\lambda_1)$, $\rho(\lambda_5)$	*
E_7	$\rho(\lambda_1)$	symplectic

TABLE III. Irreducible representations of simple Lie algebras whose index is (strictly) smaller than 1.

algebra	condition on the rank	representation	index
A_n	$n \geq 1$	$\rho(\lambda_1)$, $\rho(\lambda_n)$	$\frac{1}{2(n+1)}$
	$n \geq 2$	$\rho(2\lambda_1)$, $\rho(2\lambda_n)$	$\frac{n+3}{2(n+1)}$
	$n \geq 2$	$\rho(\lambda_2)$, $\rho(\lambda_{n-1})$	$\frac{n-1}{2(n+1)}$
	$3 \leq n \leq 7$	$\rho(\lambda_3)$, $\rho(\lambda_{n-2})$	$\frac{(n-1)(n-2)}{4(n+1)}$
B_n	$n \geq 2$	$\rho(\lambda_1)$	$\frac{1}{2n-1}$
	$2 \leq n \leq 6$	$\rho(\lambda_n)$	$\frac{2^{n-3}}{2n-1}$
C_n	$n \geq 2$	$\rho(\lambda_1)$	$\frac{1}{2(n+1)}$
	$n \geq 3$	$\rho(\lambda_2)$	$\frac{n-1}{n+1}$
	$n = 2, 3$	$\rho(\lambda_n)$	$\frac{1}{2n(n+1)}\binom{2n}{n-1}$
D_n	$n \geq 4$	$\rho(\lambda_1)$	$\frac{1}{2(n-1)}$
	$4 \leq n \leq 7$	$\rho(\lambda_{n-1})$, $\rho(\lambda_n)$	$\frac{2^{n-5}}{n-1}$
E_6		$\rho(\lambda_1)$, $\rho(\lambda_5)$	$\frac{1}{4}$
E_7		$\rho(\lambda_1)$	$\frac{1}{3}$
F_4		$\rho(\lambda_1)$	$\frac{1}{3}$
G_2		$\rho(\lambda_1)$	$\frac{1}{4}$

root of $\mathfrak{g}$ is not a root we see that 0 cannot be a weight of ρ.

Next it is easy to see that at most one of the linear forms $\frac{1}{2}\alpha_j$, $1 \leq j \leq n$, can be a weight of ρ. In fact, suppose that $1 \leq j$, $k \leq n$ and that $\frac{1}{2}\alpha_j$ and $\frac{1}{2}\alpha_k$ are weights of ρ. Since ρ is irreducible there exist integers t_i, $1 \leq i \leq n$, such that

$$\tfrac{1}{2}\alpha_k = \tfrac{1}{2}\alpha_j + \sum_{i=1}^{n} t_i \alpha_i, \tag{C1}$$

which implies $j = k$.

Combining this result with Lemma B.1 it is easy to see that all weights of ρ have the same length. What we have shown implies (in view of Lemma B.1 and of Table II) that there remain the following possibilities for $\mathfrak{g}$ and ρ (up to isomorphism and equivalence)

$\mathfrak{g}$	ρ,
A_1	$\rho(\lambda_1)$,
$B_n, n \geq 2$	$\rho(\lambda_n)$,
$C_n, n \geq 2$	$\rho(\lambda_1)$.

The cases B_n with $n \geq 3$ have to be excluded since $2\lambda_n$ is not a root of B_n if $n \geq 3$. The rest is obvious.

APPENDIX D

We determine all irreducible representations ρ of $\mathfrak{g}$ whose index l_ρ [see (6.2)] satisfies

$$l_\rho \leq 1. \tag{D1}$$

We have already mentioned that there exists a formula[12] which gives l_ρ in terms of the highest weight of ρ. Using this formula as well as Weyl's dimension formula it is straightforward but somewhat cumbersome to determine all irreducible representations ρ of $\mathfrak{g}$ with $l_\rho \leq 1$.

Now the same problem has been solved in the mathematical literature in quite another context[17] and our results agree with those of Ref. 17. The outcome is the following: For any irreducible representation ρ of $\mathfrak{g}$ the index l_ρ and the dimension $\dim\rho$ satisfy

$l_\rho < 1$ if and only if $\dim\rho < \dim\mathfrak{g}$,

$l_\rho > 1$ if and only if $\dim\rho > \dim\mathfrak{g}$,

$l_\rho = 1$ if and only if ρ is equivalent to the adjoint representation of $\mathfrak{g}$.

Table III contains all nontrivial irreducible representations ρ of $\mathfrak{g}$ for which $l_\rho < 1$.

[1]M. Gerstenhaber, Ann. Math. **78**, 267 (1963); **79**, 59 (1964); A. Nijenhuis and R.W. Richardson, Bull. Am. Math. Soc. **70**, 406 (1964).

[2]P. Ramond, Phys. Rev. D **3**, 2415 (1971); A. Neveu and J.H. Schwarz, Nucl. Phys. B **31**, 86 (1971); D.V. Volkov and V.P. Akulov, Phys. Lett. B **46**, 109 (1973); J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974).

[3]For further references see L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975) and the GLA "newsletters" mentioned in Ref. 7.

[4]A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975).

[5]W. Nahm and M. Scheunert, J. Math. Phys. **17**, 868 (1976).

[6]W. Nahm, V. Rittenberg and M. Scheunert, Phys. Lett. **B 61**, 383 (1976).

[7]P.G.O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976). See also the preliminary version of a work on graded Lie algebras, part I, II, by I. Kaplansky.

[8]V.G. Kac, Functional Anal. Appl. **9**, No. 3, 91 (1975). This letter doesn't contain any proof. However, in the meantime an expanded version is available in preprint form; we would like to thank Prof. S. Sternberg for sending a copy to us.

[9]See, for example, L. Michel in *Group Representations in Mathematics and Physics, Batelle Seattle* 1969 *Rencontres* (Springer, Berlin, 1970), p. 136.

[10]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1640 (1976).

[11]M. Scheunert, W. Nahm, and V. Rittenberg, "Graded Lie Algebras: Generatization of Hermitian Representations" (to be published).

[12]See, for example, Ref. **15**, Proposition 43.6.

[13]M. Krämer, Commun. Algebra **3**, 691 (1975), Sec. 5. Ausreduzierung einiger Tensorprodukte. Note that his first formula for D_4 should read

$$S^2\pi_2 = \pi_2^2 + \pi_1^2 + \pi_3^2 + \pi_4^2 + 1.$$

[14]J. Tits, *Tabellen zu den einfachen Liegruppen und ihren Darstellungen*, *Lecture Notes in Mathematics* 40 (Springer, Berlin, 1967).

[15]H. Freudenthal and H. de Vries, *Linear Lie Groups* (Academic, New York, 1969).

[16]N. Bourbaki, *Groupes et algèbres de Lie* (Hermann, Paris, 1968), Chaps. 4–6.

[17]E.M. Andreev, E.B. Vinberg, and A.G. Elashvili, Functional Anal. Appl. **1**, No 4, 3 (1967). The authors are grateful to M. Krämer for drawing their attention to this work.

Classification of all simple graded Lie algebras whose Lie algebra is reductive. II. Construction of the exceptional algebras

M. Scheunert, W. Nahm, and V. Rittenberg

Physikalisches Institut der Universität Bonn, Bonn, Germany
(Received 22 March 1976)

The exceptional simple graded Lie algebras whose existence is suggested by the results of the preceding paper are explicitly constructed. In this way the classification of all simple graded Lie algebras whose Lie algebra is reductive is completed.

In this paper we construct the exceptional simple graded Lie algebras whose existence is suggested by the results of Sec. 4 of the preceding work.[1] These algebras have been discovered by Freund and Kaplansky[2]; their Lie algebras are $\mathrm{sl}(2)\times\mathrm{sl}(2)\times\mathrm{sl}(2)$, $\mathrm{sl}(2)\times G_2$, and $\mathrm{sl}(2)\times\mathrm{o}(7)$. It turns out that the natural framework to study the $\mathrm{sl}(2)\times G_2$ and $\mathrm{sl}(2)\times\mathrm{o}(7)$ cases are the algebras of the octonions, respectively the Clifford algebras; some properties of these algebras are presented for completeness. The reader who is not interested in the "compact" mathematical language but rather in possible physical applications can skip the entire section and use the appendices where the exceptional graded Lie algebras are given in a pedestrian way by their commutation relations.

1. $\mathfrak{g} = \mathrm{sl}(2)\times\mathrm{sl}(2)\times\mathrm{sl}(2)$

To preserve the complete symmetry between the three algebras sl(2) we modify our notation of Sec. 4.[1] For notational convenience we consider the algebras sl(2) as symplectic Lie algebras in vector spaces of dimension two.

Choose for $i=1,2,3$ a two-dimensional vector space $\mathfrak{u}_i$ and a nondegenerate skew-symmetric bilinear form ψ_i on $\mathfrak{u}_i$. Let $\mathfrak{g}_i=\mathrm{sp}(\psi_i)=\mathrm{sl}(\mathfrak{u}_i)$ be the symplectic Lie algebra of all linear mappings of $\mathfrak{u}_i$ into itself which leave ψ_i invariant. We define

$$\mathfrak{g}=\mathfrak{g}_1\times\mathfrak{g}_2\times\mathfrak{g}_3,\quad \mathfrak{u}=\mathfrak{u}_1\otimes\mathfrak{u}_2\otimes\mathfrak{u}_3. \tag{1}$$

Then there exists a natural irreducible representation of $\mathfrak{g}$ in $\mathfrak{u}$ which will be taken as the adjoint representation of $\mathfrak{g}$ in $\mathfrak{u}$.

Now we recall that any $\mathfrak{g}_i$-invariant bilinear mapping

$$P_i:\mathfrak{u}_i\times\mathfrak{u}_i\to\mathfrak{g}_i=\mathrm{sp}(\psi_i) \tag{2a}$$

has the form

$$P_i(U_i,V_i)\,W_i=\sigma_i\{\psi_i(V_i,W_i)\,U_i-\psi_i(W_i,U_i)\,V_i\} \tag{2b}$$

for all $U_i, V_i, W_i\in\mathfrak{u}_i$ and with some element $\sigma_i\in K$. Hence the most general $\mathfrak{g}$-invariant ansatz for the product mapping $\mathfrak{u}\times\mathfrak{u}\to\mathfrak{g}$ is

$$\begin{aligned}\langle U_1\otimes U_2\otimes U_3, V_1\otimes V_2\otimes V_3\rangle &=\psi_2(U_2,V_2)\,\psi_3(U_3,V_3)\,P_1(U_1,V_1)\\ &+\psi_1(U_1,V_1)\,\psi_3(U_3,V_3)\,P_2(U_2,V_2)\\ &+\psi_1(U_1,V_1)\,\psi_2(U_2,V_2)\,P_3(U_3,V_3)\end{aligned} \tag{3}$$

with $U_i, V_i\in\mathfrak{u}_i$, $i=1,2,3$. *A priori* the constants $\sigma_i\in K$ may be chosen arbitrarily except that they must be nonzero in order to ensure that $\langle\mathfrak{u},\mathfrak{u}\rangle=\mathfrak{g}$, which is true for any simple graded Lie algebra. Note that our product mapping $\mathfrak{u}\times\mathfrak{u}\to\mathfrak{g}$ is automatically symmetric.

Using the identity

$$\psi_i(U_i,V_i)\,W_i+\psi_i(V_i,W_i)\,U_i+\psi_i(W_i,U_i)\,V_i=0 \tag{4}$$

for all U_i, V_i, $W_i\in\mathfrak{u}_i$ it is now easy to see that the Jacobi identity for three odd elements [or equivalently that Eq. (4.16) of Ref. 1] is fulfilled if and only if

$$\sigma_1+\sigma_2+\sigma_3=0. \tag{5}$$

Evidently the graded Lie algebra which we have obtained is simple; let us call it $\Gamma(\sigma_1,\sigma_2,\sigma_3)$.[2]

We know from (4.19)[1] that the Killing form of $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ is identically zero. Hence there remains the question whether there exists any nondegenerate even invariant bilinear form ϕ on $\Gamma(\sigma_1,\sigma_2,\sigma_3)$.

If ϕ exists at all it is uniquely determined up to a nonzero factor. Now the restriction of ϕ to $\mathfrak{u}\times\mathfrak{u}$ must be $\mathfrak{g}$-invariant. Hence it is nothing but a normalization of ϕ if we demand that

$$\phi(U_1\otimes U_2\otimes U_3, V_1\otimes V_2\otimes V_3)=\prod_{i=1}^{3}\psi_i(U_i,V_i) \tag{6}$$

for all U_i, $V_i\in\mathfrak{u}_i$, $i=1,2,3$.

On the other hand, we know that the three Lie algebras $\mathfrak{g}_i$ must be orthogonal with respect to ϕ. Therefore, the even bilinear form ϕ is fixed if we know the restriction ϕ_i of ϕ to the Lie algebras $\mathfrak{g}_i$, $i=1,2,3$. It is easy to see that ϕ is invariant if and only if we define

$$\phi_i=-(1/8\sigma_i)\,\phi_{\mathfrak{g}_i}, \tag{7}$$

where $\phi_{\mathfrak{g}_i}$ is the Killing form of the Lie algebra $\mathfrak{g}_i$.

Using this result we can answer the question as to which of the algebras $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ are isomorphic. In fact, suppose we are given two triples $(\sigma_1,\sigma_2,\sigma_3)$ and $(\sigma_1',\sigma_2',\sigma_3')$ of nonzero elements of K such that

$$\sigma_1+\sigma_2+\sigma_3=\sigma_1'+\sigma_2'+\sigma_3'=0. \tag{8}$$

If there exists an isomorphism ω of $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ onto $\Gamma(\sigma_1',\sigma_2',\sigma_3')$, then every nondegenerate even invariant bilinear form ϕ' on $\Gamma(\sigma_1',\sigma_2',\sigma_3')$ corresponds via ω to a nondegenerate even invariant bilinear form ϕ on

$\Gamma(\sigma_1, \sigma_2, \sigma_3)$. In view of (7) this implies that there exists a nonzero number $\tau \in K$ and a permutation π of the set $\{1, 2, 3\}$ such that

$$\sigma_i' = \tau \circ \sigma_{\pi i}; \quad i = 1, 2, 3. \tag{9}$$

Conversely, because of Lemma 3.1 [1] it is evident that the algebras $\Gamma(\sigma_1, \sigma_2, \sigma_3)$ and $\Gamma(\sigma_1', \sigma_2', \sigma_3')$ are isomorphic if (9) is fulfilled.

2. $\mathfrak{g} = \mathrm{sl}(2) \times G_2$

Because of the well-known connection of G_2 and its seven-dimensional fundamental representation with the algebra **O** of octonions (over the field K) it is most appropriate to construct our algebra in this language. Let us, therefore, collect the properties of **O** which are relevant for our purpose.[3,4] The algebra **O** of octonions is an eight-dimensional algebra over K with unit element e which has the following properties:

(a) **O** is not associative but only alternative, i.e., the associator

$$a(x, y, z) = (xy)z - x(yz) \tag{10}$$

is a skew-symmetric trilinear mapping of $\mathbf{O} \times \mathbf{O} \times \mathbf{O}$ into **O**.

If x is any element of **O** we define the mapping L_x (left multiplication by x) and R_x (right multiplication by x) of O into itself by

$$L_x(y) = xy, \quad R_x(y) = yx \tag{11}$$

for all $y \in \mathrm{O}$. Then the alternativity of O has the important consequence that for all $x, y \in \mathrm{O}$ the linear mapping

$$D_{x,y} = [L_x, L_y] + [R_x, R_y] + [L_x, R_y] \tag{12}$$

of **O** into itself is a derivation of the algebra **O**. [Recall that a derivation of an algebra is a linear mapping D of the algebra into itself which satisfies

$$D(uv) = D(u)\, v + uD(v) \tag{13}$$

for all elements u, v of the algebra.] If D is any derivation of **O** then

$$[D, D_{x,y}] = D_{D(x),y} + D_{x,D(y)} \tag{14}$$

for all $x, \; y \in \mathrm{O}$.

(b) **O** is a Cayley algebra, i.e., there exists an involution of **O**, denoted by $x \to \bar{x}$, such that

$$x + \bar{x} \in Ke, \quad x\bar{x} \in Ke \tag{15}$$

for all $x \in \mathbf{O}$. [Recall that an involution of an algebra is a bijective linear mapping τ of the algebra onto itself which satisfies

$$\tau^2(u) = u, \quad \tau(uv) = \tau(v)\, \tau(u) \tag{16}$$

for all elements u, v of the algebra.]

It is customary to define a linear form T on **O** (the trace) and a quadratic form N on O (the norm) by

$$x + \bar{x} = T(x)e, \quad x\bar{x} = N(x)e \tag{17}$$

for all $x \in \mathbf{O}$. Then

$$T(x\bar{y}) = T(y\bar{x}) = N(x + y) - N(x) - N(y) \tag{18}$$

for all $x, y \in \mathbf{O}$ and the bilinear form ψ on **O** defined by

$$\psi(x, y) = \tfrac{1}{2} T(x\bar{y}) \tag{19}$$

for all $x, y \in \mathbf{O}$ is symmetric and nondegenerate.

In the following we denote by O_0 the subspace of traceless octonions, i.e.,

$$\mathbf{O}_0 = \{x \in \mathbf{O} \mid T(x) = 0\}. \tag{20}$$

If D is any derivation of O, then

$$\psi(D(x), y) + \psi(x, D(y)) = 0 \tag{21}$$

and

$$D(\mathrm{O}) \subset \mathbf{O}_0, \quad a(x, y, z) \in \mathrm{O}_0 \tag{22}$$

for all $x, y, z \in \mathbf{O}$.

(c) The Lie algebra of all derivations of **O** is isomorphic (and will be identified) with G_2. It is then obvious that the associator a is a G_2 invariant trilinear mapping of $\mathbf{O} \times \mathbf{O} \times \mathbf{O}$ into **O**; furthermore, Eq. (14) implies that $(x, y) \to D_{x,y}$ is a G_2-invariant bilinear mapping of $\mathbf{O} \times \mathbf{O}$ into G_2 and Eq. (21) means that the bilinear form ψ on **O** is G_2-invariant.

Finally we deduce from (22) that there is a natural seven-dimensional representation of G_2 in $\mathbf{O}_0$; this representation is (equivalent to) the fundamental representation $\rho(\lambda_1)$ of G_2.

It is now easy to prove the existence of the simple graded Lie algebra with $\mathfrak{g} = \mathrm{sl}(2) \times G_2$.

With the notation introduced in Sec. 4 [1] we choose $\mathfrak{u}_2 = \mathrm{O}_0$, $\psi_2 = \psi$ and identify ρ_2 with the natural representation of G_2 in **O** mentioned above.

According to the results of Sec. 4 [1] all we have to do is to look for a totally skew-symmetric G_2 invariant trilinear mapping

$$\hat{P}_2 : \mathbf{O}_0 \times \mathbf{O}_0 \times \mathbf{O}_0 \to \mathrm{O}_0, \tag{23}$$

such that for all $U, V \in \mathrm{O}_0$ the linear mapping $\tilde{P}_2(U, V)$ of O_0 into itself, defined by

$$\tilde{P}_2(U, V)\, W = \hat{P}_2(U, V, W) + \sigma_1\{\psi(V, W)\, U - \psi(W, U)\, V\} \tag{24}$$

for all $W \in \mathbf{O}_0$, is induced by an element of G_2, i.e., by a derivation of **O**.

Now there is a natural candidate for $\hat{P}_2$, namely the restriction of the associator a to $\mathbf{O}_0 \times \mathbf{O}_0 \times \mathbf{O}_0$ [see (10) and (22)]. In fact, up to a factor this is the only possibility since the exterior product of three copies of $\rho(\lambda_1)$ decomposes according to

$$\rho(\lambda_1) \wedge \rho(\lambda_1) \wedge \rho(\lambda_1) = \rho(2\lambda_1) \oplus \rho(\lambda_1) \oplus \rho(0), \tag{25}$$

i.e., it contains $\rho(\lambda_1)$ just once.

On the other hand, we are aware of the derivations $D_{x,y}$ in (12). It is not difficult to bring the definition (12) to a form which is similar to (24) and in particular to show that

$$D_{U,V}(Z) = -a(U, V, Z) - 4\{\psi(V, Z)\, U - \psi(Z, U)\, V\} \tag{26}$$

for all $U, V \in \mathbf{O}_0$ and all $Z \in \mathbf{O}$. Hence our conditions are fulfilled if we define

$$\hat{P}_2(U, V, W) = \tfrac{1}{4}\sigma_1 a(U, V, W) \tag{27}$$

for all $U, V, W \in \mathrm{O}_0$.

The complete definition of the product mapping $\mathfrak{u} \times \mathfrak{u} \to \mathfrak{g}$ is now contained in Eqs. (4.5), (4.15),[1] (24), and (27). In particular we have

$$P_2(U, V) = -\tfrac{1}{4}\sigma_1 D_{U,V} \tag{28}$$

for all $U, V \in \mathbf{O}_0$.[2]

Obviously the graded Lie algebra which emerges is simple and we know from (4.19) [1] that its Killing form is nondegenerate.

3. $\mathfrak{g} = \mathrm{sl}(2) \times \mathrm{o}(7)$

To begin with we recall that (with the notation introduced in Sec. 4 of Ref. 1) the representation ρ_2 of o(7) in $\mathfrak{u}_2$ is the eight-dimensional spin representation. Hence it is appropriate to use Clifford algebra techniques[5] in order to construct our graded Lie algebra.

We first collect some basic results on Clifford algebras. Let Q be a nondegenerate quadratic form on an m-dimensional vector space and let $C(Q)$ be its Clifford algebra. We assume that m is even, $m = 2n$. Then $C(Q)$ is isomorphic to the algebra $L(\boldsymbol{F})$ of all linear mappings of a 2^n-dimensional vector space $\boldsymbol{F}$ into itself. In particular $C(Q)$ is simple and all irreducible representations of $C(Q)$ are equivalent to the representation in $\boldsymbol{F}$.

It follows that there exist $2n$ elements $\Gamma_j \in L(\boldsymbol{F})$, $1 \le j \le 2n$, which generate the algebra $L(\boldsymbol{F})$ and satisfy

$$\Gamma_j\Gamma_k + \Gamma_k\Gamma_j = 2\delta_{jk} \tag{29a}$$

if $1 \le j,\ k \le 2n$. If we introduce the abbreviation

$$\Gamma_{2n+1} = i^n \Gamma_1 \cdots \Gamma_{2n}, \tag{29b}$$

then it is easy to see that Eq. (29a) remains valid for $1 \le j,\ k \le 2n+1$.

Suppose $1 \le j,\ k \le 2n+1$ and let E_{jk} be the $(2n+1) \times (2n+1)$ matrix whose elements are all equal to zero with the exception of the element in the jth row and the kth column, which is equal to one. Then the matrices $E_{jk} - E_{kj}$, $1 \le j < k \le 2n+1$, form a basis of $\mathrm{o}(2n+1)$ and it is easy to see that they obey the same commutation relations as the elements $\frac{1}{2}\Gamma_j\Gamma_k$, $1 \le j < k \le 2n+1$, of $L(\boldsymbol{F})$. Hence we have a natural representation of $\mathrm{o}(2n+1)$ in $\boldsymbol{F}$ which maps $E_{jk} - E_{kj}$ onto $\frac{1}{2}\Gamma_j\Gamma_k$ if $1 \le j, k \le 2n+1$, $j \ne k$. This is the (irreducible) spin representation of $\mathrm{o}(2n+1)$.

It is easy to derive rules for the traces of products of the Γ_j which are similar to those valid for the usual Dirac matrices. For later reference we note that

$$\mathrm{Tr}(\Gamma_j\Gamma_k\Gamma_p\Gamma_q) = -2^n\delta_{jp}\delta_{kq} \tag{30}$$

if $1 \le j < k \le 2n+1$ and $1 \le p < q \le 2n+1$.

Let us come back to Eqs. (29a). These relations are equally satisfied if we replace the elements $\Gamma_j \in L(\boldsymbol{F})$ by $(-1)^n\, {}^t\Gamma_j \in L(\boldsymbol{F}^*)$, $1 \le j \le 2n$. Hence we have a representation of $C(Q)$ in $\boldsymbol{F}^*$ which (according to our previous remarks) must be equivalent to the representation in $\boldsymbol{F}$. This means that there exists a nondegenerate bilinear form ψ on $\boldsymbol{F}$ such that

$$\psi(\Gamma_j(U), V) = (-1)^n\psi(U, \Gamma_j(V)) \tag{31}$$

for $1 \le j \le 2n$ and all $U, V \in \boldsymbol{F}$. (Of course ψ is nothing but a basis-independent version of the charge conjugation matrix.) It is easy to check that this equation remains valid for $j = 2n+1$ and that

$$\psi(\Gamma_j\Gamma_k(U), V) + \psi(U, \Gamma_j\Gamma_k(V)) = 0 \tag{32}$$

if $1 \le j,\ k \le 2n+1$, $j \ne k$, and $U, V \in \boldsymbol{F}$. Equation (32) means that ψ is invariant under $\mathrm{o}(2n+1)$. One can prove that ψ is symmetric if n is congruent to -1 or 0 mod 4 and that ψ is skew-symmetric if n is congruent to 1 or 2 mod 4.

For the construction of our graded Lie algebra we are interested in the case $n = 3$; then $\dim\boldsymbol{F} = 8$ and ψ is symmetric. Using the notation of Sec. 4 of Ref. 1, we choose $\mathfrak{u}_2 = \boldsymbol{F}$, $\psi_2 = \psi$ and identify ρ_2 with the spin representation which has been defined above.

Next we apply Eq. (3.28) [1] to obtain the correct ansatz for $\tilde{P}_2$. Recall that the trace form associated with the spin representation is a nondegenerate invariant bilinear form on o(7). Hence we define in agreement with (3.28) [1] a bilinear mapping

$$\tilde{P}_2 : \boldsymbol{F} \times \boldsymbol{F} \to \rho_2(\mathrm{o}(7)) \tag{33a}$$

by

$$\tfrac{1}{8}\mathrm{Tr}(\tilde{P}_2(U, V)\Gamma_j\Gamma_k) = \tau\psi(\Gamma_j\Gamma_k(U), V) \tag{33b}$$

if $1 \le j < k \le 7$ and $U, V \in \boldsymbol{F}$, with some element $\tau \in K$. In view of (30) this implies

$$\tilde{P}_2(U, V) = -\tau \sum_{j<k} \psi(\Gamma_j\Gamma_k(U), V)\Gamma_j\Gamma_k \tag{34}$$

for all $U, V \in \boldsymbol{F}$.

The graded Lie algebra in question will exist if and only if we can find a nonzero element $\tau \in K$ such that Eq. (4.16) [1] is fulfilled.

We have solved this problem by making a Fierz transformation of (4.16) [1] and by taking advantage of symmetry properties like (31) and (32). Without going into the details (it is not necessary to use a particular representation for the Γ_j) we state that (4.16) [1] is fulfilled if (and only if) we choose

$$\tau = \tfrac{1}{3}\sigma_1. \tag{35}$$

The complete definition of the product mapping $\mathfrak{u} \times \mathfrak{u} \to \mathfrak{g}$ is now contained in Eqs. (4.5), (4.15),[1] (34), and (35). In particular we have

$$P_2(U, V) = \tfrac{1}{3}\sigma_1 \sum_{j,k} \psi(U, \Gamma_j\Gamma_k(V))(E_{jk} - E_{kj}) \tag{36}$$

for all $U, V \in \boldsymbol{F} = \mathfrak{u}_2$.

Evidently the graded Lie algebra which emerges is simple and we know from (4.19) [1] that its Killing form is nondegenerate.

Appendix A: $\mathfrak{g} = \mathrm{sl}(2) \times \mathrm{sl}(2) \times \mathrm{sl}(2)$

Even generators:

$$Q_j^m;\quad 1 \le j \le 3,\quad 1 \le m \le 3 \tag{A1}$$

Odd generators:

$$V_{\alpha\beta\gamma};\quad \alpha, \beta, \gamma = \pm 1 \tag{A2}$$

We use the summation convention [except of course for the upper index of Q which enumerates the three algebras sl(2)].

Commutation relations:

$$[Q_j^m, Q_k^n] = i\delta_{mn}\epsilon_{jkl} Q_l^m, \tag{A3}$$

$$\begin{aligned}
&[Q_j^1, V_{\alpha\beta\gamma}] = \tfrac{1}{2}\tau^j_{\alpha'\alpha} V_{\alpha'\beta\gamma},\\
&[Q_j^2, V_{\alpha\beta\gamma}] = \tfrac{1}{2}\tau^j_{\beta'\beta} V_{\alpha\beta'\gamma},\\
&[Q_j^3, V_{\alpha\beta\gamma}] = \tfrac{1}{2}\tau^j_{\gamma'\gamma} V_{\alpha\beta\gamma'},
\end{aligned} \tag{A4}$$

$$\begin{aligned}
\{V_{\alpha\beta\gamma}, V_{\alpha'\beta'\gamma'}\} = {}& \sigma_1 C_{\beta\beta'} C_{\gamma\gamma'} (C\tau^j)_{\alpha\alpha'} Q_j^1\\
&+ \sigma_2 C_{\alpha\alpha'} C_{\gamma\gamma'} (C\tau^j)_{\beta\beta'} Q_j^2\\
&+ \sigma_3 C_{\alpha\alpha'} C_{\beta\beta'} (C\tau^j)_{\gamma\gamma'} Q_j^3.
\end{aligned} \tag{A5}$$

Here τ^j, $1 \le j \le 3$, are the Pauli matrices and

$$C = i\tau^2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{A6}$$

is the corresponding charge conjugation matrix. Furthermore, σ_1, σ_2, σ_3 are arbitrary nonzero numbers which satisfy

$$\sigma_1 + \sigma_2 + \sigma_3 = 0. \tag{A7}$$

Appendix B: $\mathfrak{g} = \mathrm{sl}(2) \times G_2$

In order to derive concise expressions for the commutation relations of the graded Lie algebra in question we shall first give a description of the Lie algebra G_2 and of its fundamental seven-dimensional representation which might be also useful in other situations.

To begin with, let $e_1, \ldots, e_7$ be a basis of O_0 which is orthonormal with respect to ψ. Then the multiplication in O is given by

$$e_i e_j = -\delta_{ij} e + \xi_{ijk} e_k, \tag{B1}$$

where ξ is a totally skew-symmetric G_2-invariant tensor [see (25)]. Here and in the following, all indices run from 1 to 7; furthermore, we use the summation convention.

For the "usual" basis of O_0 (which will be chosen in the following) the components of ξ are determined by the following prescription: If (i,j,k) is one of the triples

$$\begin{aligned}
&(1,2,3),\quad (1,4,5),\quad (1,7,6),\\
&(2,4,6),\quad (2,5,7),\quad (3,4,7),\quad (3,6,5),
\end{aligned} \tag{B2}$$

then $\xi_{ijk} = 1$. If there is no permutation of $\{1, \ldots, 7\}$ which transforms (i,j,k) into one of the triples (B2), then $\xi_{ijk} = 0$.

Next we define a G_2-invariant tensor η of rank four by

$$a(e_i, e_j, e_k) = 2\eta_{ijkr} e_r. \tag{B3}$$

It turns out that η is totally skew-symmetric. Of course it is possible to express η in terms of ξ; in fact one can prove that

$$\xi_{ijr}\xi_{pqr} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp} + \eta_{ijpq}. \tag{B4}$$

Using this result it is easy to see that $\eta_{ijpq} = 1$ if (i,j,p,q) is one of the quadruples

$$\begin{aligned}
&(1,2,4,7),\quad (1,2,6,5),\quad (1,3,6,4),\quad (1,3,7,5),\\
&(2,3,4,5),\quad (2,3,7,6),\quad (4,5,7,6),
\end{aligned} \tag{B5}$$

and that $\eta_{ijpq} = 0$ if there is no permutation of $\{1, \ldots, 7\}$ which transforms (i,j,p,q) into one of the quadruples (B5).

Finally we introduce the abbreviation

$$D_{ij} = D_{e_i, e_j}. \tag{B6}$$

Note that

$$D_{ij} = -D_{ji}. \tag{B7}$$

Of course the D_{ij} form a set of generators of the vector space G_2. But even the D_{ij} with $i<j$ (for example) are not linearly independent. In fact, since G_2 has dimension 14 there must exist seven independent linear relations among the D_{ij}, $i<j$.

Now one can prove that in any alternative algebra

$$D_{xy,z} + D_{yz,x} + D_{zx,y} = 0 \tag{B8}$$

for all elements x, y, z of the algebra. This equation implies

$$\xi_{ijr} D_{rk} + \xi_{jkr} D_{ri} + \xi_{kir} D_{rj} = 0, \tag{B9}$$

or equivalently

$$\xi_{ijk} D_{ij} = 0 \tag{B10}$$

which is also obvious from general representation theory reasons. These seven equations may be described as follows: Choose any index k, $1 \le k \le 7$. Then there exist just three different pairs of indices (p,p'), (q,q'), (r,r') such that

$$\xi_{pp'k} = \xi_{qq'k} = \xi_{rr'k} = 1. \tag{B11}$$

With these indices we have

$$D_{pp'} + D_{qq'} + D_{rr'} = 0. \tag{B12}$$

$D_{pp'}$, $D_{qq'}$, and $D_{rr'}$ span a Cartan subalgebra of G_2. From (26) we obtain the explicit formula

$$\begin{aligned}
D_{ij}(e_p) &= (4\delta_{ip}\delta_{jq} - 4\delta_{iq}\delta_{jp} - 2\eta_{ijpq}) e_q\\
&= (6\delta_{ip}\delta_{jq} - 6\delta_{iq}\delta_{jp} - 2\xi_{ijk}\xi_{pqk}) e_q,
\end{aligned} \tag{B13}$$

which may be regarded as the definition of the seven-dimensional fundamental representation of G_2.

Applying the second of these equations as well as (B9) we derive from (14) the following commutation relations for the D_{ij}:

$$\begin{aligned}
[D_{ij}, D_{pq}] = {}& 6\delta_{ip} D_{jq} - 6\delta_{jp} D_{iq} + 6\delta_{jq} D_{ip}\\
&- 6\delta_{iq} D_{jp} - 2\xi_{ijk}\xi_{pqr} D_{kr}.
\end{aligned} \tag{B14}$$

We would like to remark that it is a bit cumbersome to derive (B14) directly from (B13).

Using these results, a complete description of the graded Lie algebra in question reads as follows:

Even generators:

$$\begin{aligned}
&Q_j;\quad 1 \le j \le 3,\\
&F_{pq};\quad 1 \le p, q \le 7,
\end{aligned} \tag{B15}$$

where

$$\begin{aligned}
&F_{pq} = -F_{qp},\\
&\xi_{pqr} F_{pq} = 0;
\end{aligned} \tag{B16}$$

Odd generators:

$$V_{\alpha p};\quad \alpha = \pm 1,\ 1 \leqslant p \leqslant 7; \tag{B17}$$

Commutation relations:

$$[Q_j, Q_k] = i\epsilon_{jkl}\, Q_l,$$

$$[F_{pq}, F_{rs}] = 3\delta_{pr} F_{qs} - 3\delta_{qr} F_{ps} + 3\delta_{qs} F_{pr} - 3\delta_{ps} F_{qr} - \xi_{pqu}\xi_{rsv} F_{uv}, \tag{B18}$$

$$[Q_j, F_{pq}] = 0,$$

$$[Q_j, V_{\alpha p}] = \tfrac{1}{2}\tau^j_{\alpha'\alpha} V_{\alpha' p},$$

$$[F_{pq}, V_{\alpha r}] = 2\delta_{pr} V_{\alpha q} - 2\delta_{qr} V_{\alpha p} - \eta_{pqrs} V_{\alpha s}, \tag{B19}$$

$$\{V_{\alpha p}, V_{\beta q}\} = 2\sigma\delta_{pq}(C\tau^j)_{\alpha\beta} Q_j - (\sigma/2)\, C_{\alpha\beta} F_{pq}. \tag{B20}$$

Once again τ^j are the Pauli matrices and $C = i\tau^2$. The G_2-invariant tensors ξ and η have been defined above. Finally σ is an arbitrary nonzero constant.

Appendix C: $\mathfrak{g} = sl(2) \times 0(7)$

Even generators:

$$Q_j;\quad 1 \leqslant j \leqslant 3,$$

$$Q_{pq};\quad 1 \leqslant p, q \leqslant 7, \tag{C1}$$

where

$$G_{pq} = - G_{qp}; \tag{C2}$$

Odd generators:

$$V_{\alpha\mu};\quad \alpha = \pm 1,\ 1 \leqslant \mu \leqslant 8; \tag{C3}$$

Commutation relations:

$$[Q_j, Q_k] = i\epsilon_{jkl} Q_l,$$

$$[G_{pq}, G_{rs}] = - \delta_{pr} G_{qs} + \delta_{qr} G_{ps} - \delta_{qs} G_{pr} + \delta_{ps} G_{qr}, \tag{C4}$$

$$[Q_j, G_{pq}] = 0,$$

$$[Q_j, V_{\alpha\mu}] = \tfrac{1}{2}\tau^j_{\alpha'\alpha} V_{\alpha'\mu},$$

$$[G_{pq}, V_{\alpha\mu}] = \tfrac{1}{2}(\Gamma_p\Gamma_q)_{\mu'\mu} V_{\alpha\mu'},\quad p \neq q, \tag{C5}$$

$$\{V_{\alpha\mu}, V_{\beta\nu}\} = 2\sigma\tilde{C}_{\mu\nu}(C\tau^j)_{\alpha\beta} Q_j + (\sigma/3) C_{\alpha\beta}(\tilde{C}\Gamma_p\Gamma_q)_{\mu\nu} G_{pq}. \tag{C6}$$

Here again τ^j are the Pauli matrices and $C = i\tau^2$.

The Γ_p, $1 \leqslant p \leqslant 7$, are a family of eight×eight matrices which satisfy

$$\Gamma_p\Gamma_q + \Gamma_q\Gamma_p = 2\delta_{pq}. \tag{C7}$$

$\tilde{C}$ is the corresponding charge conjugation matrix with

$${}^t\tilde{C} = \tilde{C},\quad {}^t\Gamma_p\tilde{C} = -\tilde{C}\Gamma_p. \tag{C8}$$

Finally σ is an arbitrary nonzero constant. For a convenient choice of the Γ_p and $\tilde{C}$ matrices see Ref. 6.

[1]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626 (1976).

[2]P. G. O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976). See also I. Kaplansky, "Graded Lie algebras II"; preliminary version, University of Chicago (1975).

[3]R. D. Schafer, *An Introduction to Nonassociative Algebras* (Academic, New York, 1966).

[4]N. Bourbaki, *Algèbre* (Hermann, Paris, 1971), new ed., Chap. III.

[5]N. Bourbaki, *Algèbre* (Hermann, Paris, 1959), Chap. IX.

[6]A. Pais, J. Math. Phys. **3**, 1135 (1962).

Soluble classical spin model with competing interactions*

John Stephenson[†‡]

Department of Physics, Imperial College, London SW7, England
(Received 8 March 1976)

The partition function and spin pair correlation functions have been calculated exactly for a classical linear chain model with alternate next-nearest-neighbor (nnn) interactions, in which the interaction energies between pairs of nearest (nn) and next-nearest (nnn) neighbor spins are arbitrary functions of the angles between the relevant spins. Of special interest is the cosine interaction model described by the Hamiltonian

$$\mathcal{H} = -\Sigma_{i=1}^{N}[J_1(\cos\theta_{2i-1,2i} + \cos\theta_{2i,2i+1}) + J_2\cos\theta_{2i-1,2i+1}].$$

When the nnn interaction is antiferromagnetic ($J_2 < 0$) it competes with the nn interaction J_1, and there can be disorder point(s) at which nnn correlations change from monotonic to oscillatory. The ground state is ferromagnetic when the interaction ratio $r \equiv J_2/|J_1| > -1/2 = r_c$, but is disordered for more negative values. The disorder point locus has been determined. It terminates at zero temperature at $r_D = -1/2^{1/2}$, at which point the ground state energy is a maximum. The result that r_D differs from r_c is thought to be peculiar to one-dimensional models. Over a limited range of values of r there can be two disorder points. The low temperature asymptotic behavior of the partition and correlation functions is analyzed in detail. Also a novel summation formula for spherical Bessel functions is obtained.

1. INTRODUCTION

In this paper we calculate the partition and spin correlation functions for a classical linear chain model with alternate next-nearest-neighbor (nnn) interactions, in which the interaction energies between pairs of nearest (nn) and next-nearest (nnn) neighbor spins are arbitrary functions of the angles between the relevant spins. We shall be especially concerned with the case when the interactions are simple cosine interactions described by the Hamiltonian

$$\begin{aligned}\mathcal{H} &= -\sum_{i=1}^{N}\{J_i(\mathbf{S}_{2i-1}\cdot\mathbf{S}_{2i} + \mathbf{S}_{2i}\cdot\mathbf{S}_{2i+1}) + J_2\mathbf{S}_{2i-1}\cdot\mathbf{S}_{2i+1}\}\\ &= -\sum_{i=1}^{N}\{J_1(\cos\theta_{2i-1,2i} + \cos\theta_{2i,2i+1}) + J_2\cos\theta_{2i-1,2i+1}\}.\end{aligned} \quad (1.1)$$

The inner product between spin variables $\mathbf{S}_i \cdot \mathbf{S}_j$, which would appear in the quantum mechanical Heisenberg model, has been replaced by the corresponding classical cosine interaction. The model is of interest when the nnn interaction J_2 is antiferromagnetic ($J_2 < 0$) and competes with the nn interaction J_1. We take $J_1 > 0$ throughout, without loss of generality.

The present soluble model is simpler than the general quantum mechanical Heisenberg model, with both first and second nearest neighbor interactions, in two important respects. First, as in any classical model, noncommuting finite-dimensional matrix (spin) operators have been replaced by continuous commuting angle (spin) variables, and traces over products of matrices have been replaced by multiple integrations, over the unit sphere in the present case. Second, only *alternate* next-nearest-neighbor interactions have been retained, so the model consists of a chain of triangles, as illustrated in Fig. 1. Consequently, the partition and correlation functions factorize into terms related to individual triangles.

Study of this model is motivated by a paper by Thorpe and Blume[1] in which they solve exactly a classical model containing biquadratic interactions[2] which exhibits a "quadrupolar" disorder point. This model has features which distinguish it qualitatively from the analogous Ising models on linear chains containing nnn interactions.[3] In particular, over a limited range of interaction ratios $r = J_2/J_1$ there are two disorder points, and the disorder point locus terminates at zero temperature at a value of the interaction ratio r_D which differs from the value r_c associated with the breakdown of the ferromagnetic ground state. For values of r less than r_c there is a disordered ground state, with a characteristic angle between adjacent spins.

The cosine interaction model described by (1.1), which is the subject of this paper, is similar to the Thorpe—Blume model as regards the shape of the disorder point locus and the appearance of distinct values of r_D and r_c. However, the ground state energy per lattice site E has a maximum as a function of r which also occurs at r_D. Moreover, the functional dependence of E on r is quite different from both the Thorpe—Blume model, for which E is a monotonic function of r, and from the Ising models, for which $r_D = r_c$ and the graph of E versus r consists of two intersecting straight lines. These matters are discussed in Secs. 10 and 11 of this paper, which contain the results of most physical interest, and can be read independently of the remainder of the paper.

The partition and correlation functions of the classical linear chain and of the decorated classical linear chain with alternate nnn interactions are calculated in Secs. 2—4 and Appendix A, for a general classical interaction Hamiltonian. The special form of the cosine interaction in (1.1) enables us to obtain compact expressions for the partition function and nnn correlation functions in terms of the first two eigenvalues λ_0 and λ_1, of the transfer operator or "matrix." Various methods for analyzing the integral expressions for these eigenvalues are employed successively in: Sec. 6, series expansions; Sec. 7, integration by parts; Sec. 8, Laplace's method; and Sec. 9, representation by special functions. We shall be especially interested in the mathematical techniques required for the extraction of low temperature properties. A by-product of our investigation is a novel summation formula for spherical Bessel functions, obtained in Appendix D.

2. PARTITION AND CORRELATION FUNCTIONS

In this section we review the *direct* calculation of the partition function and generalized spin correlation functions for a one-dimensional assembly of classical spins.[4–6] We shall treat the ring and chain in turn as a cluster of N spins. Label the spins $i=1,2,\ldots,N$, and suppose there is an interaction Hamiltonian $H_{i,i+1}$ between adjacent spins i and $(i+1)$ which depends only on the angle $\theta_{i,i+1}$ between them. Then the partition function for a *ring* in which the Nth spin is linked to the first spin is:

$$Z_N^{(r)}=\int\cdots\int\frac{d\Omega_1\cdots d\Omega_N}{(4\pi)^N}\exp(-\beta H_{12})\cdots\exp(-\beta H_{N1})\quad\text{(ring)}\tag{2.1}$$

where $d\Omega_i=\sin\theta_i\,d\theta_i\,d\phi_i$ denotes the element of solid angle $\Omega_i=(\theta_i,\phi_i)$, θ_i and ϕ_i being polar and azimuthal angles determining the orientation of the ith spin referred to arbitrary but fixed axes, and the integration for each i is over the surface of a unit sphere. Here $\beta=1/k_BT$. This partition function may be evaluated in terms of quantities λ_n, $n=0,1,\ldots,\infty$, defined via the expansion of the Boltzmann factor in a series of Legendre polynomials, and thence, by use of the addition theorem, in terms of (normalized) spherical harmonics, as follows:

$$\exp(-\beta H_{12})=\sum_{n=0}^{\infty}(2n+1)\lambda_nP_n(\cos\theta_{12})\tag{2.2a}$$

$$=4\pi\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\lambda_nY_{nm}(\Omega_1)Y_{nm}^*(\Omega_2).\tag{2.2b}$$

We have written down the expansion for the bond between spins 1 and 2, but we shall suppose, for economy, that the interaction Hamiltonian has the same form for all pairs of neighboring spins. The extension of our results to the case of arbitrary interactions between different pairs of spins is straightforward. The coefficients λ_n are given explicitly by

$$\lambda_n=\tfrac{1}{2}\int_{-1}^{+1}dx\,\exp(-\beta H_{12})P_n(x),\tag{2.3}$$

where $x=\cos\theta_{12}$, and the orthogonality of the Legendre polynomials over the interval $(-1,1)$ has been used. On substituting expansions of the form (2.2b) into each factor in the integrand of the partition function, and employing the orthogonality of the spherical harmonics, we obtain the desired expression for $Z_N^{(r)}$:

$$\begin{aligned}Z_N^{(r)}&=\sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty}\sum_{m_1=-n_1}^{n_1}\cdots\sum_{m_N=-n_N}^{n_N}\lambda_1\cdots\lambda_N\int d\Omega_1\cdots\int d\Omega_N\\&\quad\times Y_{n_1m_1}(\Omega_1)Y_{n_1m_1}^*(\Omega_2)\cdots Y_{n_Nm_N}(\Omega_N)Y_{n_Nm_N}^*(\Omega_1)\\&=\sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty}\lambda_{n_1}\cdots\lambda_{n_N}\delta_{n_1n_2}\cdots\delta_{n_Nn_1}\\&\quad\times\sum_{m_1=-n_1}^{n_1}\cdots\sum_{m_N=-n_N}^{n_N}\delta_{m_1m_2}\cdots\delta_{m_Nm_1}\\&=\sum_{n=0}^{\infty}\lambda_n^N\sum_{m=-n}^{n}1=\sum_{n=0}^{\infty}(2n+1)\lambda_n^N\quad\text{(ring)}.\end{aligned}\tag{2.4}$$

For asymptotically large N, the partition function is determined by λ_0, the largest coefficient, as may easily be seen from (2.3) using the fact that $P_0=1$, and all other Legendre polynomials are less than unity in magnitude in the interval $(-1,1)$. Similarly, one may obtain the partition function for a chain. On setting $H_{N1}=0$, so that there is one less interaction to expand, one observes that only $Y_{00}(\Omega_1)$ and $Y_{00}^*(\Omega_N)$ contribute to the integrals over Ω_1 and Ω_N, and so the sums collapse to a single term $n=m=0$, and

$$Z_N^{(c)}=\lambda_0^{N-1},\quad\text{(chain)}.\tag{2.5}$$

Since we shall only be interested in the limit of asymptotically large N, we shall write down formulas as if for a ring, on the understanding that vanishingly small terms will be omitted in subsequent calculation.

A generalized pair correlation function between classical spins i and $(i+r)$ may be defined in terms of the mean values of $P_n(\cos\theta_{i,i+r})$:

$$\begin{aligned}\langle P_n\rangle=\int\cdots\int\frac{d\Omega_1\cdots d\Omega_N}{Z_N(4\pi)^N}P_n(\cos\theta_{i,i+r})\\\times\exp(-\beta H_{12})\cdots\exp(-\beta H_{N1}).\end{aligned}\tag{2.6}$$

When $n=1$, we retrieve the usual pair correlation function $\langle\cos\theta_{i,i+r}\rangle$. To evaluate the generalized correlation, expand each factor in the integrand using (2.2b), and set

$$P_n(\cos\theta_{i,i+r})=\left(\frac{4\pi}{2n+1}\right)\sum_{m=-n}^{n}Y_{nm}^*(\Omega_i)Y_{nm}(\Omega_{i+r}),\tag{2.7}$$

in which the complex conjugate is placed on the first member of the addition formula. Then retaining only asymptotically important terms, we have, setting $Y_{00}=(4\pi)^{-1/2}$,

$$\begin{aligned}&\langle P_n(\cos\theta_{i,i+r})\rangle\\&=\sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'}\left(\frac{\lambda_{n'}}{\lambda_0}\right)^r\sum_{m=-n}^{n}\frac{4\pi}{2n+1}\int\frac{d\Omega_i}{(4\pi)^{1/2}}Y_{nm}^*(\Omega_i)Y_{n'm'}(\Omega_i)\\&\quad\times\int\frac{d\Omega_{i+r}}{(4\pi)^{1/2}}Y_{nm}(\Omega_{i+r})Y_{n'm'}^*(\Omega_{i+r})\\&=\sum_{m=-n}^{n}\sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'}\left(\frac{\lambda_{n'}}{\lambda_0}\right)^r\frac{\delta_{nn'}\delta_{mm'}}{(2n+1)}=\left(\frac{\lambda_n}{\lambda_0}\right)^r.\end{aligned}\tag{2.8}$$

The result (2.8) for the generalized correlation function is exact for a chain with free ends, and is asymptotically correct for a ring as $N\to\infty$. This may easily be seen by constructing the general expression for the correlation on a ring, and setting $H_{N1}=0$ to obtain a chain, as we did previously in deducing the partition function for the chain from that for the ring, between Eqs. (2.4) and (2.5). The derivation of an exact formula for the correlation function of a finite ring is more involved, and is not considered here.

The first two correlation functions with $n=1$ and 2 for dipolar and quadrupolar order were used by Thorpe and Blume.[1] However, our result is actually a special case ($\nu=3$) of the general result for ν-dimensional classical spins obtained by Liu and Joseph (ν is the spin-space dimensionality).[7,8]

The coefficients λ_n are in fact just the eigenvalues of the integral equation

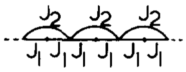

FIG. 1. The linear chain with nearest-neighbour interactions J_1 and alternate next-nearest neighbour interactions J_2.

$$\lambda_n \psi_n(\Omega_1) = \int \frac{d\Omega_2}{4\pi} \exp(-\beta H_{12}) \psi_n(\Omega_2), \tag{2.9}$$

whose eigenfunctions are the spherical harmonics $Y_{nm}(\Omega)$. As is well known, the partition function for a ring of spins can be written in terms of these eigenvalues,[9] as in (2.4) above.

3. PARTITION FUNCTION OF DECORATED CHAIN

The decorated chain is illustrated in Fig. 1. Odd numbered spins interact via a Hamiltonian $H_{2i-1,2i+1}$, whereas even numbered spins are linked with neighboring odd (numbered) spins via Hamiltonians $H_{2i-1,2i}$ and $H_{2i,2i+1}$. We actually have a one-dimensional assembly of spins linked by nearest-neighbor (nn) interactions plus alternate next-nearest-neighbor (nnn) interactions between odd spins. The even spins can be treated as a decoration. That is, we can integrate all the even spin variables and reduce the problem to that of a linear chain with an effective pair interaction Hamiltonian $H^{\mathrm{eff}}_{2i-1,2i+1}$ between odd spins. (This is equivalent to integrating out vertices of degree 2, following Joyce's method[6]). The integral in the partition function involving the $2i$th spin can be developed by expansions of the form (2.2). Set

$$\exp(-\beta H_{2i-1,2i+1}) = \sum_{n=0}^{\infty} (2n+1)\lambda_n^{(0)} P_n(\cos\theta_{2i-1,2i+1}), \tag{3.1a}$$

$$\exp(-\beta H_{2i,2i+1}) = \sum_{n=0}^{\infty} (2n+1)\mu_n P_n(\cos\theta_{2i,2i+1}), \tag{3.1b}$$

etc., and perform the integration over $\Omega_{2i} = (\theta_{2i}, \phi_{2i})$:

$$\begin{aligned}
&\int \frac{d\Omega_{2i}}{4\pi} \exp[-\beta(H_{2i-1,2i} + H_{2i,2i+1})] \\
&= \int \frac{d\Omega_{2i}}{4\pi} (4\pi)^2 \sum_{n=0}^{\infty} \sum_{n'=0}^{\infty} \sum_{m=-n}^{n} \sum_{m'=-n'}^{n'} \mu_n \mu_{n'} \\
&\quad \times Y_{nm}(\Omega_{2i-1}) Y^*_{nm}(\Omega_{2i}) Y_{n'm'}(\Omega_{2i}) Y^*_{n'm'}(\Omega_{2i+1}) \\
&= 4\pi \sum_{n=0}^{\infty} \mu_n^2 \sum_{m=-n}^{n} Y_{nm}(\Omega_{2i-1}) Y^*_{nm}(\Omega_{2i+1}) \\
&= \sum_{n=0}^{\infty} (2n+1)\mu_n^2 P_n(\cos\theta_{2i-1,2i+1}).
\end{aligned} \tag{3.2}$$

Thus once the decoration is integrated out, the effective Hamiltonian will involve only the angle between the spins $2i-1$ and $2i+1$:

$$\begin{aligned}
&\exp(-\beta H^{\mathrm{eff}}_{2i-1,2i+1}) \\
&= \exp(-\beta H_{2i-1,2i+1}) \sum_{n=0}^{\infty} (2n+1)\mu_n^2 P_n(\cos\theta_{2i-1,2i+1}),
\end{aligned} \tag{3.3}$$

with corresponding eigenvalues as in (2.3):

$$\begin{aligned}
\lambda_n &= \tfrac{1}{2}\int_{-1}^{+1} dx \exp(-\beta H_{2i-1,2i+1}) P_n(x) \sum_{n'=0}^{\infty} (2n'+1)\mu_{n'}^2 P_{n'}(x) \\
&= \tfrac{1}{2} \sum_{n'=0}^{\infty} \sum_{n''=0}^{\infty} (2n'+1)(2n''+1)\lambda_{n''}^{(0)} \mu_{n'}^2 \qquad (3.4a) \\
&\quad \times \int_{-1}^{+1} dx P_n(x) P_{n'}(x) P_{n''}(x) \\
&= \tfrac{1}{2} \sum_{n'=0}^{\infty} \sum_{n''=0}^{\infty} (2n'+1)(2n''+1)\lambda_{n''}^{(0)} \mu_{n'}^2 J(n,n',n''), \qquad (3.4b)
\end{aligned}$$

where $J(n,n',n'')$ denotes the final integral over a product of three Legendre polynomials. In order for expressions like (3.2), (3.3), (3.4) to be useful, one needs to perform the indicated summations, if possible, in closed form. We note in passing that if we define expansion coefficients $\nu_{nn'}$ by

$$\exp(-\beta H_{2i-1,2i+1}) P_n(x) = \sum_{n'=0}^{\infty} (2n'+1)\nu_{nn'} P_{n'}(x), \tag{3.5}$$

and substitute in (3.4a), then the eigenvalues λ_n can be expressed as a single sum:

$$\lambda_n = \sum_{n'=0}^{\infty} (2n'+1)\nu_{nn'}\mu_{n'}^2. \tag{3.6}$$

For any selected value of n, one may perform the integration required to evaluate $J(n,n',n'')$ in (3.4b) and express λ_n as a single sum. For example,

$$\lambda_0 = \sum_{n=0}^{\infty} (2n+1)\lambda_n^{(0)}\mu_n^2, \tag{3.7a}$$

$$\lambda_1 = \sum_{n=0}^{\infty} \mu_n^2 [(n+1)\lambda_{n+1}^{(0)} + n\lambda_{n-1}^{(0)}] \tag{3.7b}$$

(3.7a) can of course also be derived directly from (3.6).

4. PAIR CORRELATION FUNCTIONS OF DECORATED CHAIN

The pair correlation function between spins i and $(i+r)$ is defined as the mean value of $\cos\theta_{i,i+r}$ as in (2.6) with $n=1$, and with the Hamiltonian for the decorated chain in the Boltzmann factors. There are three types of correlation to consider, depending on whether the spins involved are linked by nnn bonds (odd numbered spins) or are decorating spins (even). That is, we have odd—odd, even—odd, and even—even spin correlations.

The odd—odd case is simplest, since intervening decorating spins may be integrated out and an effective Hamiltonian introduced, as in (3.2). The desired correlation then reduces directly to the calculation of λ_0 and λ_1, with the effective Hamiltonian, as in (3.7). By analogy with (2.8), setting $n=1$, we have

$$\langle \cos\theta_{2i,2i+r}\rangle = (\lambda_1/\lambda_0)^k, \quad r = 2k. \tag{4.1}$$

The calculation of the pair correlations involving decorating (even) spins is more involved. We may either appeal to a powerful theorem of Joyce [Eq. (5.12) of Ref. 6] or proceed directly following Thorpe and Blume [Eqs. (9) and (10) of Ref. 1]. A detailed derivation is given in Appendix A, where it is shown that for even—odd spins

$$\begin{aligned}
&\langle \cos\theta_{2i,2i+r}\rangle \\
&= \left(\frac{\lambda_1}{\lambda_0}\right)^k \left\{ \sum_{n=0}^{\infty} \frac{\lambda_n^{(0)}}{\lambda_0} \mu_n [(n+1)\mu_{n+1} + n\mu_{n-1}] \right\}, \quad r = 2k+1,
\end{aligned} \tag{4.2}$$

and for even—even spins

$$\langle\cos\theta_{2i,2i+r}\rangle$$
$$=\left(\frac{\lambda_1}{\lambda_0}\right)^k\left\{\sum_{n=0}^{\infty}\frac{\lambda_n^{(0)}}{\lambda_0}\mu_n[(n+1)\mu_{n+1}+n\mu_{n-1}]\right\}^2, \quad r=2k+2. \tag{4.3}$$

The structure of these expressions for pair correlation functions is analogous to the corresponding results for an Ising chain with alternate next-nearest neighbor interactions.[3] One may readily check that the formulas of the last two sections (3 and 4) reduce correctly in special cases when either the nn or the nnn interactions are absent.

5. COSINE INTERACTION MODELS

Now we adopt specific forms for the interaction between adjacent spins. For decorating spins set

$$H_{2i,2i+1}=-J_1\cos\theta_{2i,2i+1} \quad\text{and}\quad K=\beta J_1=J_1/k_BT. \tag{5.1}$$

Then the expansion of the Boltzmann factor becomes

$$\exp(-\beta H_{2i,2i+1})=\exp(K\cos\theta)=\sum_{n=0}^{\infty}(2n+1)i_n(K)P_n(\cos\theta), \tag{5.2}$$

where $i_n(K)$ is a spherical Bessel function of pure imaginary argument (Appendix B). Thus we may identify the coefficients in the expansion (3.1b) as

$$\mu_n=\tfrac{1}{2}\int_{-1}^{+1}dx\,\exp(Kx)P_n(x)=i_n(K). \tag{5.3}$$

We are now in a position to perform the sum needed for the construction of the effective interaction in (3.3). Making a slight generalization to allow for different interactions between a decorating spin and its left- and right-hand neighbors, the required sum is, with $x=\cos\theta$,

$$\sum_{n=0}^{\infty}(2n+1)i_n(K)i_n(K')P_n(x)=\frac{\sinh R(x)}{R(x)}, \tag{5.4}$$

where

$$R(x)=(K^2+K'^2+2KK'x)^{1/2}. \tag{5.5}$$

The eigenvalues are given by the integral formula (3.4a):

$$\lambda_n=\frac{1}{2}\int_{-1}^{+1}dx\,\exp(-\beta H_{2i-1,2i+1})\frac{\sinh R(x)}{R(x)}P_n(x). \tag{5.6}$$

In the further special case when the nnn interaction is also a cosine interaction, with

$$H_{2i-1,2i+1}=-J_2\cos\theta_{2i-1,2i+1} \quad\text{and}\quad L=\beta J_2=J_2/k_BT, \tag{5.7}$$

the integral for the eigenvalues becomes

$$\lambda_n=\frac{1}{2}\int_{-1}^{+1}dx\,\exp(Lx)\frac{\sinh R(x)}{R(x)}P_n(x). \tag{5.8}$$

Then the expansion coefficients $\lambda_n^{(0)}$ in (3.1a) are also spherical Bessel functions,

$$\lambda_n^{(0)}=i_n(L), \tag{5.9}$$

and the expansions (3.7) for the eigenvalues λ_0 and λ_1 become

$$\lambda_0=\sum_{n=0}^{\infty}(2n+1)i_n(K)i_n(K')i_n(L), \tag{5.10}$$

$$\lambda_1=\sum_{n=0}^{\infty}(2n+1)i_n(K)i_n(K')[(n+1)i_{n+1}(L)+ni_{n-1}(L)] \tag{5.11a}$$

$$=\sum_{n=0}^{\infty}(2n+1)i_n(K)i_n(K')i_n'(L). \tag{5.11b}$$

Of course (5.8) is a more useful "closed form" for the sums in these equations. It is interesting to note the symmetry in λ_0 between K, K', and L. In (5.11a) the square bracket contains Bessel functions which combine to give the derivative $i_n'(L)$ so that

$$\lambda_1=\frac{\partial\lambda_0}{\partial L}, \tag{5.12}$$

a result which is obvious from (5.8), and is actually a special case of a general derivative relation obtained by Joyce [Eq. (5.12) of Ref. 6]. One observes in Eqs. (3.7), (4.2), and (4.3) that the square bracket factors contain just those combinations of eigenvalues which in the cosine interaction models reduce to Bessel function derivatives as discussed above. From now on we shall further restrict our analysis to the case when K' and K are equal, so that in Eq. (5.8) for λ_n we have

$$R(x)=K[2(1+x)]^{1/2}. \tag{5.13}$$

Also for cosine interaction models, the extra curly-bracket factors $\{\,\}$ appearing in the even—odd and even—even spin correlation functions in (4.2) and (4.3) can be obtained by differentiation of λ_0. Noting (5.3), (5.9), (5.10) and (5.11), we have

$$\left\{\sum_{n=0}^{\infty}\lambda_n^{(0)}\mu_n\frac{[(n+1)\mu_{n+1}+n\mu_{n-1}]}{\lambda_0}\right\}$$
$$=\lambda_0^{-1}\sum_{n=0}^{\infty}(2n+1)i_n(L)i_n(K)i_n'(K)=\lambda_0^{-1}\frac{1}{2}\frac{\partial\lambda_0}{\partial K}, \tag{5.14}$$

from which it is clear that the extra $\{\,\}$ factors are positive and do not affect the signs of the correlations.

6. SERIES EXPANSIONS

The series expansions for λ_0 and λ_1 are useful at high temperatures, and in certain asymptotic limits also at low temperatures.

At high temperatures, λ_0 and λ_1 may be expanded in powers of K and L by use of the power series representation of Bessel functions. For λ_1 we have, to leading order,

$$\lambda_1\sim\tfrac{1}{3}L+\tfrac{1}{9}K^2+\cdots, \tag{6.1}$$

from which we note that when L is negative ($J_2<0$, antiferromagnetic), λ_1 is negative at sufficiently high temperatures. Now it is λ_1 that determines the nature of the pair correlation decay (Sec. 4). In particular nnn pair correlations are seen to be oscillatory at high temperatures, and we can estimate the disorder point locus by finding the temperature at which λ_1 vanishes. The estimate is valid when $k_BT\gg J_1\gg|J_2|$. From (6.1),

$$\frac{1}{K_1}\equiv\frac{k_BT}{J_1}\sim-\frac{J_1}{3J_2}. \tag{6.2}$$

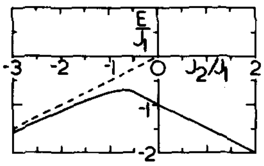

FIG. 2. Graph of the ground state energy per lattice site E/J_1 versus interaction ratio $r=J_2/J_1$. The ferromagnetic phase terminates at $r_c=-\frac{1}{2}$, and the maximum is at $r_D=-1/\sqrt{2}$.

At low temperatures in the asymptotic limit $|J_2| \gg k_BT \gg J_1$, we may use the asymptotic form of the Bessel function,

$$i_n(L) \sim \frac{\exp|L|}{2|L|}\left(1-\frac{n(n+1)}{2|L|}+\cdots\right)\times\begin{cases}1, & L>0,\\ (-)^n, & L<0.\end{cases} \tag{6.3}$$

The leading terms in λ_0 and λ_1 may be extracted from the series

$$\begin{aligned}\lambda_0 &\sim \sum_{n=0}^{\infty}(2n+1)i_n(|L|)[i_n(K)]^2\times\begin{cases}1, & L>0\\ (-)^n, & L<0,\end{cases}\\ &\sim \frac{\exp(|L|)}{2|L|}\sum_{n=0}^{\infty}(2n+1)[i_n(K)]^2\times\begin{cases}1, & L>0\\ (-)^n, & L<0,\end{cases}\\ &= \frac{\exp(|L|)}{2|L|}\times\begin{cases}(\sinh 2K)/2K, & L>0\\ 1, & L<0.\end{cases}\end{aligned} \tag{6.4}$$

For ferromagnetic nnn interactions, $L>0$, the low temperature form

$$\lambda_0 \sim \exp(L+2K)/8KL \tag{6.5}$$

actually agrees with the exact result we shall obtain later, and gives the ground state energy per site correctly (Fig. 2):

$$E=-(J_1+\tfrac{1}{2}J_2). \tag{6.6}$$

For antiferromagnetic nnn interactions, $L<0$, the low temperature form is not adequate to give the ground state. Even with the next correction term in powers of J_1^2/J_2 or K^2/L, so

$$\begin{aligned}\lambda_0 &\sim \frac{\exp(|L|)}{2|L|}\sum_{n=0}^{\infty}(2n+1)[i_n(K)]^2\left\{1-\frac{n(n+1)}{2|L|}+\cdots\right\}\\ &\quad\times(-)^n,\quad L<0\\ &= \frac{\exp(|L|)}{2|L|}\left\{1+\frac{K^2}{3|L|}+\cdots\right\},\end{aligned} \tag{6.7}$$

we still do not get the correct result. This is because we are (invalidly) trying to interchange limiting processes. As far as it goes, (6.7) is in agreement with the full expansion, which is obtained in the next section.

7. ANALYSIS OF INTEGRAL: INTEGRATION BY PARTS

In an attempt to investigate further the asymptotic case $|J_2| \gg k_BT \gg J_1$ when $J_2<0$ we commence by taking the integral for λ_0, and integrate successively by parts. This process generates two infinite series, coming from the upper and lower limits of integration, one of which is summable directly in terms of a confluent hypergeometric function, and contains the expansion (6.7) of the previous section. The integral for λ_0 is, from (5.8) and (5.13),

$$\lambda_0=\frac{1}{2}\int_{-1}^{+1}dx\,\exp(Lx)\frac{\sinh\{K[2(1+x)]^{1/2}\}}{\{K[2(1+x)]^{1/2}\}}. \tag{7.1}$$

We observe that the second factor in the integrand can be expressed in terms of a Bessel function

$$i_0(z)=(\sinh z)/z. \tag{7.2}$$

Then with the abbreviation

$$y=[2(1+x)]^{1/2},\quad\text{so that } y'=1/y, \tag{7.3}$$

the integral becomes

$$\lambda_0=\tfrac{1}{2}\int_{-1}^{+1}dx\,\exp(Lx)i_0(Ky) \tag{7.4}$$

We now integrate by parts successively using the general identity

$$\int_a^b dx\,fg=\sum_{n=0}^{N-1}(-)^n(I^{n+1}f)(D^ng)\Big|_a^b+(-)^N\int_a^b dx\,(I^Nf)(D^Ng), \tag{7.5}$$

where $D\equiv d/dx$ and I denotes indefinite integration. We identify

$$f=\exp(Lx)\quad\text{and } g=i_0(Ky), \tag{7.6}$$

so (Appendix B)

$$I^mf=\exp(Lx)/L^m\quad\text{and}\quad D^mg=K^{2m}i_m(Ky)/(Ky)^m. \tag{7.7}$$

One may easily check that the remaining integral in (7.5) tends to zero as $N\to\infty$ [by placing a uniform bound on the integrand in the interval $(-1,1)$, and using the large N behavior of the Bessel function]. Then, extending the series to infinity, we get

$$\lambda_0=\tfrac{1}{2}\sum_{n=0}^{\infty}\frac{K^{2n}}{(-L)^{n+1}}\left\{e^{-L}\lim_{\alpha\to 0}\frac{i_n(\alpha)}{\alpha^n}-e^L\frac{i_n(2K)}{(2K)^n}\right\} \tag{7.8a}$$

$$=\frac{e^{-L}}{(-2L)}\sum_{n=0}^{\infty}\left(\frac{K^2}{-L}\right)^n\frac{1}{1\cdot 3\cdot 5\cdots(2n+1)}-\frac{e^L}{(-2L)}\times\sum_{n=0}^{\infty}\left(\frac{K}{-2L}\right)^n i_n(2K) \tag{7.8b}$$

$$=\frac{e^{-L}}{(-2L)}M\left(1,\tfrac{3}{2};\frac{-K^2}{2L}\right)-\frac{e^L}{(-2L)}\sum_{n=0}^{\infty}\left(\frac{K}{-2L}\right)^n i_n(2K) \tag{7.8c}$$

$$=\frac{\exp(-L-K^2/2L)}{(-2L)}M\left(\tfrac{1}{2},\tfrac{3}{2};\frac{K^2}{2L}\right)-\frac{e^L}{(-2L)}\times\sum_{n=0}^{\infty}\left(\frac{K}{-2L}\right)^n i_n(2K), \tag{7.8d}$$

where $M(a,b;z)$ is a confluent hypergeometric function (Appendix C), and we have used Kummer's transformation on the final line. The first series in (7.8b) is an expansion in powers of (K^2/L), and is in agreement with (6.7) up to two terms. It is the first term in (7.8) which is asymptotically important when J_2 is negative and $|J_2| \gg k_BT \gg J_1$. The representations of λ_0 in (7.8) are exact, and will be of further interest in Sec. 9. We note here that the asymptotic behavior is described correctly by the first term when $J_2<-\frac{1}{2}J_1$, but that the second term also contributes to the *low* temperature properties when $T\to 0$ with J_1 and J_2 fixed, and $J_2 \geq -\frac{1}{2}J_1$. However, in the above asymptotic limit we can use the asymptotic form of the hypergeometric function, and write

$$\lambda_0 \sim \frac{\exp(-L)\exp(-K^2/2L)}{K(-2L/\pi)^{1/2}}, \tag{7.9}$$

a result which is confirmed in Sec. 8, and is shown to hold rigorously when $L < -\frac{1}{2}K$ in Sec. 9.

8. ANALYSIS OF INTEGRAL: LAPLACE'S METHOD

The low temperature behavior of λ_0 and λ_1 can be extracted from the integrals by application of Laplace's and related methods.[10] After making the variable change

$$x = 2w^2 - 1 \tag{8.1}$$

the integral for λ_n becomes

$$\lambda_n = \frac{\exp(-L)}{K} \int_0^1 dw \exp(2Lw^2) \sinh 2Kw P_n(2w^2-1). \tag{8.2}$$

When $L \gg 0$, but K is fixed, the leading asymptotic term in λ_0 can be extracted by inspection of the behavior of the integrand at the upper limit. Using Laplace's method,[10] we have

$$\lambda_0 \sim \frac{e^L}{2L} \cdot \frac{\sinh 2K}{2K} \tag{8.3}$$

in agreement with the corresponding expression in (6.4).

At *low* temperatures when K *and* $|L|$ are large, the dominant contribution to the integral for λ_n comes from the neighborhood of the point where the factor

$$\exp(2Lw^2)\sinh 2Kw \equiv \exp[g(w)], \quad \text{say}, \tag{8.4}$$

is greatest. Explicitly we have

$$g(w) = 2Lw^2 + \log\sinh 2Kw, \tag{8.5a}$$

$$g'(w) = 4Lw + 2K\coth 2Kw, \tag{8.5b}$$

$$g''(w) = 4L - 4K^2(\operatorname{csch} 2Kw)^2. \tag{8.5c}$$

While $J_2 > -\frac{1}{2}J_1$, or $L > -\frac{1}{2}K$, the maximum value of $g(w)$ occurs at the *upper* limit of integration where $g'(1) > 0$. Then, following Laplace's method,[10] we have

$$\lambda_n = \frac{\exp(-L)}{K}\int_0^1 dw \exp[g(w)] P_n(2w^2-1) \tag{8.6a}$$

$$\sim \frac{\exp(-L)}{K} \cdot \frac{\exp[g(1)]}{g'(1)} P_n(1) \tag{8.6b}$$

$$\sim \frac{\exp(L)\sinh 2K}{2K(K+2L)}, \tag{8.6c}$$

where in the last line we have inserted the low temperature form of $g'(1)$. This confirms our earlier results and shows explicitly the breakdown at $K = -2L$. The ground state energy is given correctly as in (6.6), and as $T \to 0$, the specific heat per site approaches the (physically unacceptable) value k_B.

When J_2 is negative and sufficiently strong so that $J_2 < -\frac{1}{2}J_1 < 0$, or $L < -\frac{1}{2}K$, then $g(w)$ has a maximum within the range of integration at $w = w_0$, say, determined by the vanishing of $g'(w)$ in (8.5b). Following Laplace's method, we now have

$$\lambda_n \sim \frac{\exp(-L)}{K} \frac{\exp[g(w_0)]}{\Delta} P_n(2w_0^2 - 1), \tag{8.7a}$$

with

$$\Delta^2 = -\tfrac{1}{2}g''(w_0)/\pi. \tag{8.7b}$$

As $T \to 0$,

$$w_0 \to -K/2L = -J_1/2J_2 = \cos\theta, \quad \text{say}, \tag{8.8a}$$

and

$$\Delta \sim (-2L/\pi)^{1/2}, \tag{8.8b}$$

and $dw_0/d\beta \to 0$ exponentially fast, from (8.5b). On substituting these low temperature forms in (8.7a) we find the low temperature form of λ_0:

$$\lambda_0 \sim \frac{\exp(-L)\exp(-K^2/2L)}{2K(-2L/\pi)^{1/2}}. \tag{8.9}$$

Now, employing the usual thermodynamic recipes, one finds that the energy per site is (Fig. 2)

$$E = \tfrac{1}{2}J_2 + w_0^2 J_2 + \tfrac{1}{2}d\log(K\Delta)/d\beta \tag{8.10a}$$

$$\to \tfrac{1}{2}(J_2 + J_1^2/2J_2) \quad \text{as } T \to 0, \tag{8.10b}$$

and the specific heat per site is

$$C \to \tfrac{3}{4}k_B, \quad \text{as } T \to 0. \tag{8.11}$$

One of the most interesting results is that for the pair correlation function between nnn spins (4.1), which at $T = 0$ with $r = 2k$ is

$$\langle \cos\theta_{2i,2i+r}\rangle = (\lambda_1/\lambda_0)^k = \begin{cases} 1, & J_2 > -\frac{1}{2}J_1, \\ (\cos 2\theta)^k, & J_2 < -\frac{1}{2}J_1, \end{cases} \tag{8.12}$$

with θ defined as in (8.8a). There is a ferromagnetic ground state when $J_2 > -\frac{1}{2}J_1$, and a disordered ground state when $J_2 < -\frac{1}{2}J_1$. The change over occurs at the critical interaction ratio

$$r_c = J_2/J_1 = -\tfrac{1}{2}. \tag{8.13}$$

In this latter phase, the pair correlation decays exponentially, but changes over from monotonic to oscillatory at $\theta = \pi/4$, where the interaction ratio is

$$r_D = J_2/J_1 = -1/\sqrt{2}, \tag{8.14}$$

at which value the disorder point locus meets the zero temperature axis. The shape of this locus near $T = 0$ can be extracted from λ_1 when an extra higher order factor is included in (8.7a). It may be shown (Sec. 9) that λ_0 is given by (8.9) up to terms which vanish exponentially fast. Hence we can extract λ_1 by differentiation of λ_0 as in (5.12) and gain the required correction factor:

$$\lambda_1 = \frac{\partial\lambda_0}{\partial L} = \frac{\exp(-L)\exp(-K^2/2L)}{2K(-2L/\pi)^{1/2}}\left[-1 + \frac{K^2}{2L^2} - \frac{1}{2L}\right]. \tag{8.15}$$

As $T \to 0$, $\lambda_1/\lambda_0 \to -1 + K^2/2L^2 = \cos 2\theta$, as before, (8.12). λ_1 vanishes when the square bracket vanishes. The low temperature disorder point locus is therefore

$$k_B T/J_1 = 1/K = -2J_2/J_1 + J_1/J_2, \tag{8.16}$$

which we can write with $r = J_2/J_1$ as

$$k_B T/J_1 \equiv f(r) = -2r + (1/r), \tag{8.17}$$

so the initial ($T = 0$) slope is

$$f'(r) = -2 - (1/r^2), \tag{8.18}$$

which is negative, and tends to (-4) as $r \to r_D$. Therefore, the disorder point locus "doubles back" initially as shown in Fig. 3.

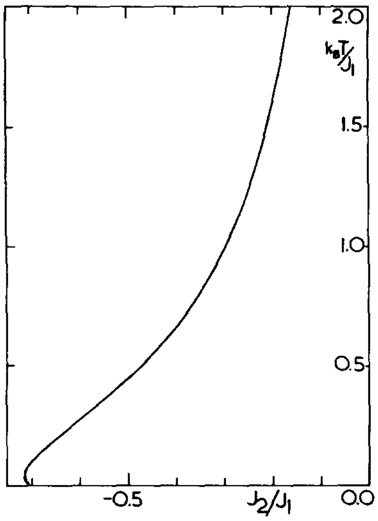

FIG. 3. Graph of the disorder point locus k_BT_D/J_1 versus interaction ratio $r=J_2/J_1$. The locus terminates at $r_D=-1/\sqrt{2}$. For values of r in the range $-0.7149\ldots<r<r_D$ there are two disorder points.

By a similar analysis one can easily show that the extra factors { } in even—odd and even—even spin correlations (4.2) and (4.3) are just

$$\{\ \}=\cos\theta, \quad \text{as } T\to 0. \tag{8.19}$$

Here we have used the derivative relation (5.14) and the asymptotic form (8.9) for λ_0.

9. ANALYSIS OF INTEGRAL: SPECIAL FUNCTIONS

It is possible to express the integral for λ_0 in terms of error functions, or equivalently in terms of confluent hypergeometric functions (Appendix C). This enables us to make a more systematic study, and to gain further appreciation, of how the low temperature properties come from the partition function integral. The exact representation of λ_0 obtained may be compared with that of Sec. 7, (7.8), and an interesting Bessel function identity extracted (Appendix D). The steps are straightforward, and commence by completing the square in the arguments of the exponentials in the integrand. After a short calculation one obtains

$$\lambda_0=\frac{\exp(-L)\exp(-K^2/2L)}{2K}\left\{\left(1+\frac{K}{2L}\right)M\left(\frac{1}{2},\frac{3}{2};2L\left(1+\frac{K}{2L}\right)^2\right)\right.$$
$$-\left(1-\frac{K}{2L}\right)M\left(\frac{1}{2},\frac{3}{2};2L\left(1-\frac{K}{2L}\right)^2\right)$$
$$\left.-\left(\frac{K}{L}\right)M\left(\frac{1}{2},\frac{3}{2};2L\left(\frac{K}{2L}\right)^2\right)\right\}. \tag{9.1}$$

This form is especially suitable for asymptotic analysis when $L<0$. By using Kummer's transformation we obtain a form which is suitable for asymptotic analysis when $L>0$:

$$\lambda_0=\frac{e^L}{2K}\left\{\exp(2K)\left(1+\frac{K}{2L}\right)M\left(1,\frac{3}{2};-2L\left(1+\frac{K}{2L}\right)^2\right)\right.$$
$$-\exp(-2K)\left(1-\frac{K}{2L}\right)M\left(1,\frac{3}{2};-2L\left(1-\frac{K}{2L}\right)^2\right)$$
$$\left.-\exp(-L)\left(\frac{K}{L}\right)M\left(1,\frac{3}{2};-\frac{K^2}{2L}\right)\right\}. \tag{9.2}$$

The asymptotic behavior of λ_0 as $T\to 0$ can be derived from the properties of the confluent hypergeometric function of large argument (Appendix C). The appearance of different asymptotic forms when L is negative according as $L\gtrless-\frac{1}{2}K$ is of some interest.

$L>0$: From (9.2)

$$\lambda_0\sim\frac{e^L}{8KL}\left\{\exp(2K)-\exp(-2K)\,\mathrm{sgn}\left(1-\frac{K}{2L}\right)\right\} \tag{9.3a}$$
$$\sim\frac{e^L}{2L}\,\frac{\sinh 2K}{2K},\quad \text{for large } L \tag{9.3b}$$
$$\sim\frac{\exp(L+2K)}{8KL},\quad \text{for large } K \text{ and } L. \tag{9.3c}$$

This is in agreement with (6.4), (6.5), and (8.6c).

$L<0$: We write out *separately* the three leading order terms from the three hypergeometric functions in (9.1):

$$\lambda_0\sim\frac{\exp(-L)\exp(-K^2/2L)}{4K(-2L/\pi)^{1/2}}\left\{\mathrm{sgn}\left(1+\frac{K}{2L}\right)-1+2\right\}. \tag{9.4}$$

$L<-\frac{1}{2}K$: If $L<-\frac{1}{2}K$ the contributions from the first two hypergeometric functions *cancel*, and the asymptotic behavior is determined by the same hypergeometric function as appeared in the integration by parts development of Sec. 7. Moreover, the correction terms are exponentially small (Appendix D), so that we are justified in differentiating to get $\lambda_1=\partial\lambda_0/\partial L$ as in Sec. 8.

$-\frac{1}{2}K<L<0$: The leading terms in (9.4) *cancel* when $-\frac{1}{2}K<L<0$, and we must retain higher order terms in the asymptotic expansions of the hypergeometric functions. It is now *crucial* to observe that the required extra terms in the asymptotic series are preceded by an *exponential* function (C8). The remaining dominant contribution is actually from the first hypergeometric function. We obtain

$$\lambda_0\sim\frac{\exp(2K+L)}{(-)8KL\,|1+K/2L|}. \tag{9.5}$$

$L=-\frac{1}{2}K$: In this borderline case when the central term in (9.1) drops out, we have contributions from *both* the remaining terms:

$$\lambda_0\sim\pi^{1/2}\exp(3K/2)/4K^{3/2}. \tag{9.6}$$

We note that as $T\to 0$, the specific heat per site $C\to\frac{3}{4}k_B$.

10. GROUND STATE ENERGY

The ground state energy per site can be extracted at once from (9.3), (9.4), (9.5), and (9.6):

$$E=(-)(J_1+\tfrac{1}{2}J_2),\quad J_2\geq-\tfrac{1}{2}J_1, \tag{10.1a}$$
$$E=\tfrac{1}{2}(J_2+J_1^2/2J_2),\quad J_2<-\tfrac{1}{2}J_1. \tag{10.1b}$$

One can, of course, derive this result directly by minimizing the classical energy via the total Hamiltonian of the system (of $2N$ spins):

$$H=(-)\sum_{i=1}^{N}\{J_1(\cos\theta_{2i-1,2i}+\cos\theta_{2i,2i+1})+J_2\cos\theta_{2i-1,2i+1}\}. \tag{10.2}$$

The ferromagnetic solution (10.1a) comes from setting all angles θ equal to zero, and is valid for $J_2>-\frac{1}{2}J_1$.

For more negative values of $J_2 < -\frac{1}{2}J_1$ one finds for a given triple of spins $2i-1$, $2i$, and $2i+1$ that the spins are coplanar, with the odd spins being symmetrically disposed about the central even spin and making a polar angle θ with it, where

$$\cos\theta = -J_1/2J_2. \tag{10.3}$$

Spins involved with a given nnn link (Fig. 1) maintain rigid relative orientations, in a common plane. But there is no correlation between the orientation of the planes associated with different nnn bonds, except that adjacent planes must contain a common odd numbered spin. In this sense the ground state is disordered, with the pair correlation function between nnn spins being given by (8.12). Also, the pair correlation between an even decorating spin and an adjacent odd spin is just $\cos\theta$, from (4.2) and (8.19).

The ground state energy is plotted in Fig. 2 as a function of $r = J_2/J_1$. The change over from an ordered ferromagnetic state to a "disordered" ground state occurs at the critical ratio $r_c = -\frac{1}{2}$, at which point $E = -\frac{3}{4}J_1$, and the left- and right-hand branches of the graph meet with common slope. The maximum value of the ground state energy when $J_2 < -\frac{1}{2}J_1$ is $E = -J_1/\sqrt{2}$, and occurs where the interaction ratio $r = -1/\sqrt{2}$, which is precisely the value r_D at which the disorder point locus terminates at zero temperature.

11. DISORDER POINTS

We gather together in this section previous results for the disorder point locus, along which pair correlations between nnn spins vanish. The desired locus is graphed in Fig. 3. The complete locus was determined numerically by selecting a value of K $(=J_1/k_BT)$ and finding the corresponding value of L $(=J_2/k_BT)$ for which λ_1 in (8.2) vanishes. The analytical form of the locus at high and low temperatures is confirmed by these numerical results. From (6.2) and (8.16), these forms are:

$$\text{at high temperatures,} \quad 1/K \sim -1/3r, \tag{11.1}$$

$$\text{at low temperatures,} \quad 1/K \sim -2r + (1/r) = f(r), \quad \text{say.} \tag{11.2}$$

The locus terminates at

$$r = r_D = -1/\sqrt{2}, \quad \text{where } \cos 2\theta = 0, \tag{11.3}$$

at which point the slope is negative, $f'(r_D) = -4$, so the locus "doubles back" initially, as remarked previously. For values of r in the range

$$-0.7149\cdots < r < r_D \tag{11.4}$$

there are two disorder points, and at the lower limit of this range, $1/K = 0.04(5)$. The over-all shape of the disorder point locus is similar to that for the "quadrupolar" disorder point of Thorpe and Blume.[1] The result that r_D differs from r_c is thought to be peculiar to one-dimensional models, as may be surmised from inspection of Fig. 1 of Ref. 11, which shows typical disorder and critical point loci for some soluble two-dimensional models.

APPENDIX A: CORRELATION FUNCTIONS

It is straightforward to establish the procedure for calculating the general correlation function $\langle P_n(\cos\theta)\rangle$ when even—odd and even—even spins are involved. Expanding $P_n(\cos\theta)$ by the addition theorem, we can express the desired correlations as sums of $(2n+1)$ terms as follows:

$$\begin{aligned}
&\langle P_n(\cos\theta_{2i,2i+r})\rangle \\
&= \left(\frac{4\pi}{2n+1}\right)\sum_{m=-n}^{n}\sum_{n'm'}\left(\frac{\lambda_{n'}}{\lambda_0}\right)^k \Bigg\{\int d\Omega_{2i-1}\int d\Omega_{2i}\int d\Omega_{2i+1} \\
&\quad\times\left[\frac{Y_{00}^*(\Omega_{2i-1})}{\lambda_0(4\pi)^2}\right] Y_{nm}^*(\Omega_{2i})Y_{n'm'}(\Omega_{2i+1})\exp(-\beta H_{2i-1,2i}) \\
&\quad\times\exp(-\beta H_{2i-1,2i+1})\exp(-\beta H_{2i,2i+1})\Bigg\} \\
&\quad\times\left\{\int d\Omega_{2i+r}Y_{n'm'}^*(\Omega_{2i+r})Y_{nm}(\Omega_{2i+r})Y_{00}(\Omega_{2i+r})\right\}
\end{aligned} \tag{A1}$$

for even—odd spins with $r = 2k+1$, and

$$\begin{aligned}
&\langle P_n(\cos\theta_{2i,2i+r})\rangle \\
&= \left(\frac{4\pi}{2n+1}\right)\sum_{m=-n}^{n}\sum_{n'm'}\left(\frac{\lambda_{n'}}{\lambda_0}\right)^k \Bigg\{\int d\Omega_{2i-1}\int d\Omega_{2i}\int d\Omega_{2i+1} \\
&\quad\times\left(\frac{Y_{00}^*(\Omega_{2i-1})}{\lambda_0(4\pi)^2}\right) Y_{nm}^*(\Omega_{2i})Y_{n'm'}(\Omega_{2i+1})\exp(-\beta H_{2i-1,2i}) \\
&\quad\times\exp(-\beta H_{2i-1,2i+1})\exp(-\beta H_{2i,2i+1})\Bigg\} \\
&\quad\times\Bigg\{\int d\Omega_{2i+r-1}\int d\Omega_{2i+r}\int d\Omega_{2i+r+1}Y_{n'm'}(\Omega_{2i+r-1}) \\
&\quad\times Y_{nm}(\Omega_{2i+r})\left(\frac{Y_{00}(\Omega_{2i+r+1})}{\lambda_0(4\pi)^2}\right)\exp(-\beta H_{2i+r-1,2i+r}) \\
&\quad\times\exp(-\beta H_{2i+r-1,2i+r+1})\exp(-\beta H_{2i+r,2i+r+1})\Bigg\}
\end{aligned} \tag{A2}$$

for even—even spins with $r = 2k+2$. In each case k denotes the number of J_2 bonds involved. These bonds can be integrated out to give the factor $(\lambda_{n'}/\lambda_0)^k$, at the expense of introducing the sums over n' and m'. The external factor (4π) cancels the zero order spherical harmonics $Y_{00} = Y_{00}^* = (4\pi)^{-1/2}$. In the case $n=1$, we will show that the sums over n' reduce to three equal terms with $n'=1$ and $m' = \pm 1, 0$, and thence identify the curly bracket factors in (A1) and (A2) with corresponding factors in (4.2) and (4.3). With obvious abbreviations for the curly bracket factors, the above expressions have the structure

$$\langle P_n(\cos\theta_{2i,2i+2k+1})\rangle = (2n+1)^{-1}\sum_{m=-n}^{n}\sum_{n'm'}\left(\frac{\lambda_{n'}}{\lambda_0}\right)^k I_{nm,n'm'}\{\delta_{nn'}\delta_{mm'}\}, \tag{A3}$$

$$\langle P_n(\cos\theta_{2i,2i+2k+2})\rangle = (2n+1)^{-1}\sum_{m=-n}^{n}\sum_{n'm'}\left(\frac{\lambda_{n'}}{\lambda_0}\right)^k I_{nm,n'm'}I_{nm,n'm'}^*. \tag{A4}$$

To evaluate $I_{nm,n'm'}$ expand each of the exponential Boltzmann factors using (3.1) and perform the integrations over Ω_{2i-1}, Ω_{2i}, and Ω_{2i+1} in turn:

$$I_{nm,n'm'} = \left(\frac{4\pi}{\lambda_0}\right)\int d\Omega_{2i-1}\int d\Omega_{2i}\int d\Omega_{2i+1}Y_{nm}^*(\Omega_{2i})Y_{n'm'}(\Omega_{2i+1})$$

$$\times \sum_{n''m''} \mu_{n''} Y_{n''m''}(\Omega_{2i-1}) Y^*_{n''m''}(\Omega_{2i})$$

$$\times \sum_{n^{iv}m^{iv}} \lambda^{(0)}_{n^{iv}} Y^*_{n^{iv}m^{iv}}(\Omega_{2i-1}) Y_{n^{iv}m^{iv}}(\Omega_{2i+1})$$

$$\times \sum_{n'''m'''} \mu_{n'''} Y_{n'''m'''}(\Omega_{2i}) Y^*_{n'''m'''}(\Omega_{2i+1})$$

$$= \frac{4\pi}{\lambda_0} \sum_{n''m''} \sum_{n'''m'''} \mu_{n''}\lambda^{(0)}_{n''}\mu_{n'''} \int d\Omega_{2i} Y^*_{nm}(\Omega_{2i}) Y^*_{n''m''}(\Omega_{2i})$$

$$\times Y_{n'''m'''}(\Omega_{2i}) \int d\Omega_{2i+1} Y_{n'm'}(\Omega_{2i+1}) Y_{n''m''}(\Omega_{2i+1})$$

$$\times Y^*_{n'''m'''}(\Omega_{2i+1}). \tag{A5}$$

In the case $n=1$, the integral over Ω_{2i} is straightforward with the results

$$I_{10,n'm'} = \frac{4\pi}{\lambda_0} \sum_{n''m''} \mu_{n''}\lambda^{(0)}_{n''} \int d\Omega_{2i+1} Y_{n'm'}(\Omega_{2i+1}) Y_{n''m''}(\Omega_{2i+1})$$

$$\times \left(\frac{3}{4\pi}\right)^{1/2} \left[\mu_{n''+1} Y^*_{n''+1,m''} \left\{\frac{(n''-m''+1)(n''+m''+1)}{(2n''+1)(2n''+3)}\right\}^{1/2}\right.$$

$$\left. + \mu_{n''-1} Y^*_{n''-1,m''} \left\{\frac{(n''-m'')(n''+m'')}{(2n''-1)(2n''+1)}\right\}^{1/2}\right], \tag{A6a}$$

$$I_{11,n'm'} = \frac{4\pi}{\lambda_0} \sum_{n''m''} \mu_{n''}\lambda^{(0)}_{n''} \int d\Omega_{2i+1} Y_{n'm'}(\Omega_{2i+1}) Y_{n''m''}(\Omega_{2i+1})$$

$$\times \left(\frac{3}{8\pi}\right)^{1/2} \left[-\mu_{n''+1} Y^*_{n''+1,m''+1} \left\{\frac{(n''+m''+1)(n''+m''+2)}{(2n''+1)(2n''+3)}\right\}^{1/2}\right.$$

$$\left. + \mu_{n''-1} Y^*_{n''-1,m''+1} \left\{\frac{(n''-m'')(n''-m''-1)}{(2n''-1)(2n''+1)}\right\}^{1/2}\right], \tag{A6b}$$

$$I_{1-1,n'm'} = \frac{4\pi}{\lambda_0} \sum_{n''m''} \mu_{n''}\lambda^{(0)}_{n''} \int d\Omega_{2i+1} Y_{n'm'}(\Omega_{2i+1}) Y_{n''m''}(\Omega_{2i+1})$$

$$\times \left(\frac{3}{8\pi}\right)^{1/2} \left[\mu_{n''+1} Y^*_{n''+1,m''-1} \left\{\frac{(n''-m''+1)(n''-m''+2)}{(2n''+1)(2n''+3)}\right\}^{1/2}\right.$$

$$\left. - \mu_{n''-1} Y^*_{n''-1,m''-1} \left\{\frac{(n''+m'')(n''+m''-1)}{(2n''-1)(2n''+1)}\right\}^{1/2}\right]. \tag{A6c}$$

Next, perform the sums over m'' of which there are six (drop the primes and the common arguments Ω_{2i+1} of the spherical harmonics)

$$\sum_m Y_{nm} Y^*_{n+1,m} \left\{\frac{(n-m+1)(n+m+1)}{(2n+1)(2n+3)}\right\}^{1/2} = \frac{(n+1)\cos\theta}{4\pi}, \tag{A7a}$$

$$\sum_m Y_{nm} Y^*_{n-1,m} \left\{\frac{(n-m)(n+m)}{(2n-1)(2n+1)}\right\}^{1/2} = \frac{n\cos\theta}{4\pi}, \tag{A7b}$$

$$\sum_m Y_{nm} Y^*_{n+1,m+1} \left\{\frac{(n+m+1)(n+m+2)}{(2n+1)(2n+3)}\right\}^{1/2}$$

$$= \frac{-(n+1)\exp(-i\phi)\sin\theta}{4\pi}, \tag{A7c}$$

$$\sum_m Y_{nm} Y^*_{n-1,m+1} \left\{\frac{(n-m)(n-m-1)}{(2n-1)(2n+1)}\right\}^{1/2} = \frac{n\exp(-i\phi)\sin\theta}{4\pi}, \tag{A7d}$$

$$\sum_m Y_{nm} Y^*_{n+1,m-1} \left\{\frac{(n-m+1)(n-m+2)}{(2n+1)(2n+3)}\right\}^{1/2}$$

$$= \frac{(n+1)\exp(i\phi)\sin\theta}{4\pi}, \tag{A7e}$$

$$\sum_m Y_{nm} Y^*_{n-1,m-1} \left\{\frac{(n+m)(n+m-1)}{(2n-1)(2n+1)}\right\}^{1/2} = \frac{-n\exp(i\phi)\sin\theta}{4\pi}. \tag{A7f}$$

Express the rhs in terms of spherical harmonics Y_{1m}, and insert in (A6). The remaining integrals over Ω_{2i+1} now involve products of $Y_{n'm'}$ with spherical harmonics of degree 1, viz., Y^*_{1m}. So the $I_{1m,n'm'}$ vanish unless $n'=1$ and $m'=m$. Moreover, the three remaining terms $I_{1m,1m}$ are equal:

$$I_{1m,n'm'} = \delta_{1n'}\delta_{mm'} \left\{\sum_{n''=0}^{\infty} \frac{\mu_{n''}\lambda^{(0)}_{n''}}{\lambda_0} \left[(n''+1)\mu_{n''+1} + n''\mu_{n''-1}\right]\right\}. \tag{A8}$$

Now in (A3) and (A4) the sums over n' collapse to a single term $n'=1$, and the sums over m' and m reduce to multiplication by 3, which just cancels the $(2n+1)$ factors. The final expressions are just (4.2) and (4.3).

APPENDIX B: BESSEL FUNCTIONS

The spherical Bessel function $i_n(z)$ required in the text is defined by

$$i_n(z) = \left(\frac{\frac{1}{2}\pi}{z}\right)^{1/2} I_{n+1/2}(z)$$

$$= \Gamma(\tfrac{3}{2})(\tfrac{1}{2}z)^n \sum_{r=0}^{\infty} \frac{(\frac{1}{2}z)^{2r}}{\Gamma(r+1)\Gamma(n+r+\frac{3}{2})}$$

$$= \frac{z^n}{1\cdot 3\cdot 5\cdots(2n+1)} \left\{1 + \frac{\frac{1}{2}z^2}{1!\,(2n+3)}\right.$$

$$\left. + \frac{(\frac{1}{2}z^2)^2}{2!\,(2n+3)(2n+5)} + \cdots\right\}. \tag{B1}$$

An alternative representation in terms of exponential functions is

$$i_n(z) = \left(\frac{1}{2z}\right)\left\{e^z \sum_{r=0}^{n} \frac{(-)^r (n+r)!}{r!\,(n-r)!\,(2z)^r}\right.$$

$$\left. + (-)^{n+1} e^{-z} \sum_{r=0}^{n} \frac{(n+r)!}{r!\,(n-r)!\,(2z)^r}\right\}, \tag{B2}$$

so that $i_0(z) = (\sinh z)/z$ etc. This form is useful for obtaining the large $|z|$ behavior of $i_n(z)$. The spherical Bessel function satisfies the recurrence and derivative relations

$$\frac{(2n+1)}{z}\, i_n = i_{n-1} - i_{n+1}, \tag{B3a}$$

$$(2n+1)\frac{di_n}{dz} = n i_{n-1} + (n+1) i_{n+1}, \tag{B3b}$$

$$i_{n-1} = \left(\frac{n+1}{z}\right) i_n + \frac{di_n}{dz}, \tag{B3c}$$

$$i_{n+1} = \left(\frac{-n}{z}\right) + \frac{di_n}{dz}. \tag{B3d}$$

In particular $i_0' = i_1$, and in general

$$\left(\frac{1}{z}\frac{d}{dz}\right)^m \left[\frac{i_n(z)}{z^n}\right] = \frac{i_{n+m}}{z^{n+m}}. \tag{B4}$$

In addition to the integral representation (5.3) and the expansions (5.2) and (5.4), the following summation formulae are employed in the text:

$$\sum_{n=0}^{\infty}(2n+1)[i_n(z)]^2=(\sinh 2z)/(2z), \tag{B5a}$$

$$\sum_{n=0}^{\infty}(-)^n(2n+1)[i_n(z)]^2=1, \tag{B5b}$$

$$\sum_{n=0}^{\infty}(-)^n n(n+1)(2n+1)[i_n(z)]^2=-\tfrac{2}{3}z^2. \tag{B5c}$$

APPENDIX C: CONFLUENT HYPERGEOMETRIC AND ERROR FUNCTIONS

The confluent hypergeometric function is defined by

$$M(a,b;z)=\frac{\Gamma(b)}{\Gamma(a)}\sum_{n=0}^{\infty}\frac{\Gamma(a+n)}{\Gamma(b+n)}\frac{z^n}{n!}. \tag{C1}$$

Kummer's transformation is

$$M(a,b;z)=e^zM(b-a,b;-z). \tag{C2}$$

For large $|z|$ we have leading terms in the asymptotic expansion:

$$\mathrm{Re}z>0,\quad M(a,b;z)\sim\frac{\Gamma(b)}{\Gamma(a)}e^z z^{a-b}, \tag{C3a}$$

$$\mathrm{Re}z<0,\quad M(a,b;z)\sim\frac{\Gamma(b)}{\Gamma(b-a)}(-z)^{-a}. \tag{C3b}$$

The error function and the complementary error function are defined by

$$\mathrm{erf}z=\frac{2}{\sqrt{\pi}}\int_0^z\exp(-t^2)\,dt, \tag{C4a}$$

$$\mathrm{erfc}z=\frac{2}{\sqrt{\pi}}\int_z^{\infty}\exp(-t^2)\,dt\equiv 1-\mathrm{erf}z. \tag{C4b}$$

The error function has the following series representations

$$\mathrm{erf}z=\frac{2}{\sqrt{\pi}}\sum_{n=0}^{\infty}\frac{(-)^n z^{2n+1}}{n!\,(2n+1)} \tag{C5a}$$

$$=\frac{2}{\sqrt{\pi}}\exp(-z^2)\sum_{n=0}^{\infty}\frac{2^n z^{2n+1}}{1\cdot 3\cdot 5\cdots(2n+1)} \tag{C5b}$$

$$=\sqrt{2}\sum_{n=0}^{\infty}(-)^n[I_{2n+1/2}(z^2)-I_{2n+3/2}(z^2)]. \tag{C5c}$$

The error function is related to the confluent hypergeometric function by

$$\mathrm{erf}z=\frac{2z}{\sqrt{\pi}}M(\tfrac{1}{2},\tfrac{3}{2};-z^2)=\frac{2z}{\sqrt{\pi}}\exp(-z^2)M(1,\tfrac{3}{2};z^2). \tag{C6}$$

The behavior of erf$z$ for large z ($\arg z<3\pi/4$) is best obtained from the asymptotic expansion for erfc$z$. We have

$$\sqrt{\pi}z\exp(z^2)\,\mathrm{erfc}z\sim 1+\sum_{n=1}^{\infty}(-)^n\frac{1\cdot 3\cdot 5\cdots(2n-1)}{(2z^2)^n} \tag{C7}$$

whence

$$M(\tfrac{1}{2},\tfrac{3}{2};-z^2)=\frac{\sqrt{\pi}}{2z}\,\mathrm{erf}z=\frac{\sqrt{\pi}}{2z}(1-\mathrm{erfc}z)$$
$$\sim\frac{\sqrt{\pi}}{2z}\left\{1-\frac{\exp(-z^2)}{\sqrt{\pi}z}\times\left[1+\sum_{n=1}^{\infty}(-)^n\frac{1\cdot 3\cdot 5\cdots(2n-1)}{(2z^2)^n}\right]\right\}. \tag{C8}$$

This asymptotic expansion is important for Sec. 9 of the text.

APPENDIX D: SUMMATION FORMULA FOR BESSEL FUNCTIONS

As a by-product of the analysis of Secs. 7 and 9 we obtain a formula summing a power series whose coefficients are spherical Bessel functions, a result we believe to be new. By comparing (7.8c) and (9.2) in the text, we observe that they contain a common hypergeometric function on the rhs, which may be cancelled out, leaving the desired Bessel function sum:

$$\sum_{n=0}^{\infty}\left(\frac{-K}{2L}\right)^n i_n(2K)$$
$$=\left(\frac{L}{K}\right)\left[\exp(2K)\left(1+\frac{K}{2L}\right)M\left(1,\frac{3}{2};-2L\left(1+\frac{K}{2L}\right)^2\right)\right.$$
$$\left.-\exp(-2K)\left(1-\frac{K}{2L}\right)M\left(1,\frac{3}{2};-2L\left(1-\frac{K}{2L}\right)^2\right)\right]. \tag{D1}$$

Setting

$$x=-K/2L,\quad y=2K, \tag{D2}$$

we have an expression for the generating function sum

$$\sum_{n=0}^{\infty}x^n i_n(y)=\left[\exp(-y)(1+x)M\left(1,\frac{3}{2};\frac{1}{2}\frac{y}{x}(1+x)^2\right)\right.$$
$$\left.-\exp(y)(1-x)M\left(1,\frac{3}{2};\frac{1}{2}\frac{y}{x}(1-x)^2\right)\right]\Big/(2x). \tag{D3}$$

By Kummer's transformation

$$\sum_{n=0}^{\infty}x^n i_n(y)=\exp[\tfrac{1}{2}y(x+1/x)]\left[(1+x)M\left(\frac{1}{2},\frac{3}{2};-\frac{1}{2}\frac{y}{x}(1+x)^2\right)\right.$$
$$\left.-(1-x)M\left(\frac{1}{2},\frac{3}{2};-\frac{1}{2}\frac{y}{x}(1-x)^2\right)\right]\Big/(2x). \tag{D4}$$

In particular if $x=1$,

$$\sum_{n=0}^{\infty}i_n(y)=\exp(-y)M(1,\tfrac{3}{2};2y)=\exp(y)M(\tfrac{1}{2},\tfrac{3}{2};-2y). \tag{D5}$$

*Work supported in part by the National Research Council of Canada, Grant No. A6595.

†Science Research Council Senior Visiting Fellow.

‡On leave from the Physics Department, University of Alberta, Edmonton, Alberta, Canada.

[1]M.F. Thorpe and M. Blume, Phys. Rev. B 5, 1961 (1972).
[2]L.L. Liu and R.I. Joseph, J. Phys. Chem. Solids 33, 451 (1972).
[3]J. Stephenson, Can. J. Phys. 48, 1724 (1970).
[4]T. Nakamura, J. Phys. Soc. Japan 7, 264 (1952).
[5]M.E. Fisher, Amer. J. Phys. 32, 343 (1964).
[6]G.S. Joyce, Phys. Rev. 155, 478 (1967).
[7]H.E. Stanley, Phys. Rev. 179, 570 (1969).
[8]L.L. Liu and R.I. Joseph, Phys. Rev. Lett. 26, 1378 (1971).
[9]C. Domb, Advan. Phys. 9, 149 (1960).
[10]E.T. Copson, *Asymptotic Expansions*, No. 55, Cambridge Tracts in Mathematics and Mathematical Physics (Cambridge U.P., London, 1965).
[11]J. Stephenson and D.D. Betts, Phys. Rev. B 2, 2702 (1970).

How to calculate the grand partition function in the uncorrelated jet model

J. Gasser*†

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England
and Institute for Theoretical Physics, University of Berne, Berne, Switzerland
(Received 8 March 1976)

We present a new technique for calculating the grand partition function and all quantities of physical interest in the uncorrelated jet model. The method, which is also valid in the large-p_T region, consists of the numerical evaluation of the appropriate integral representation in the complex plane. We analyze in detail the difficulties associated with this approach and show how to overcome them. The numerical results are checked with a new high energy expansion for the grand partition function.

1. INTRODUCTION

In the simplest version of the uncorrelated jet model (UJM),[1–5] the normalized inclusive cross section for the reaction $a+b\rightarrow c+X$ is given by

$$\frac{1}{\sigma_{\rm TOT}}\frac{d\sigma}{d\mu(k)}=f(k^T)\frac{\Omega_g(P-k)}{\Omega_g(P)},$$

where

$$\Omega_g(Q)=\sum_{n=2}^{\infty}(g^n/n!)\Omega_n(Q),$$
$$\Omega_n(Q)=\int\prod_{i=1}^{n}d\mu(p_i)f(p_i^T)\delta^4\left(Q-\sum_{j=1}^{n}p_j\right), \qquad (1.1)$$
$$d\mu(p)=\frac{d^3p}{2(2\pi)^3p^0}.$$

(We consider a theory with only one type of particle with mass m, $k^0\neq\sqrt{s}/2$.) P^μ denotes the total four-momentum of the incoming particles a and b, whereas $p^T=(p_x^2+p_y^2)^{1/2}$ is the transverse component of **p**. Throughout the paper we work in the c. m. system ($\mathbf{P}=(\sqrt{s},0)$), keeping the z axis as the beam or longitudinal direction. The coupling constant g is a free parameter. We will assume in the following that the function $f(p)$ (which cuts off transverse momenta) is normalized according to

$$[\pi/(2\pi)^3]\int_0^\infty dp\,pf(p)=1 \qquad (1.2)$$

whenever $f(p)\neq 1$.

Every physical quantity in the UJM can be calculated once we know the grand partition function $\Omega_g(Q)$ or the partition functions $\Omega_n(Q)$ for all $n\geqslant 2$.[5] Therefore the basic problem is to evaluate $\Omega_g(Q)$ or $\Omega_n(Q)$ in some way. The following three techniques to do this played a central role during the last ten years:

Approach 1: Lurçat and Mazur[6] used the method of steepest descent to approximate $\Omega_n(Q)$ in the case $f=1$. The result is valid for large n with corrections of the order $O(1/\sqrt{n})$. Later on, this procedure has been generalized to $f\neq 1$[2] and to $\Omega_g(Q)$.[7,8] (The analytic expression for Ω_g obtained in this way is valid at high energies.)

Approach 2: Another method to obtain an analytic expression for $\Omega_g(Q)$ at high energies is presented in Refs. 4 and 5. These authors use a version of the Riemann—Lebesque lemma. It has been shown by de Groot[5] that one can evaluate analytically Ω_n for $n\geqslant 2$ in the high energy limit, provided the transverse momentum of the produced particles is fairly small.

Approach 3: One may apply Monte-Carlo techniques to handle the $3n-4$ integrations in $\Omega_n(Q)$.[9–12]

Despite their usefulness, these methods suffer from the following disadvantages:

(a) The errors introduced in the approximation schemes 1 and 2 are cumbersome to estimate.[5–7,11,12] The corrections to the leading asymptotic behavior of both the partition function Ω_n and the grand partition function Ω_g may be large even at ISR energies. See Figs. 2–4 and Ref. 11.

(b) Method 2 is not valid in the region where produced particles have large transverse momentum. See Ref. 5 and Fig. 2. This large p_T region provides an important test for theories of strong interactions at ISR energies.[13,14]

(c) Monte-Carlo calculations are exact in principle. However, this method can be very expensive in terms of computing time. If the latter is scrimped, the results can have large statistical errors.[12]

It is the aim of this article to present a technique, different from Approaches 1–3 above, used previously[15] to evaluate $\Omega_g(Q)$. Our method does not suffer from the drawbacks (a)–(c) just mentioned, and is *valid for all* Q^μ which are not too near the edge of phase space. (See Sec. II.) It consists of a numerical evaluation of the integral representation[4]

$$\Omega_g(Q)=\frac{1}{4\pi^2 i}\int^C dz\,zI_0(zQ_L)\int_0^\infty dx\,xJ_0(xQ_T)\times\{\exp[g\Phi(z,x)]-1-g\Phi(z,x)\},$$

where

$$\Phi(z,x)=\frac{1}{(2\pi)^2}\int_0^\infty dp\,pJ_0(xp)K_0(zm_T)f(p)$$
$$Q_L=(Q_0^2-Q_z^2)^{1/2},\quad Q_T=(Q_x^2+Q_y^2)^{1/2},\quad m_T=(m^2+p^2)^{1/2}. \qquad (1.3)$$

$I_0(z)$ and $K_0(z)$ are modified Bessel functions, whereas $J_0(x)$ is a Bessel function of the first kind.[16] The path C runs from $-i\infty+\epsilon$ to $i\infty+\epsilon$, $\epsilon>0$. In Eq. (1.3) there is essentially a three-dimensional integration, one of the integrations to be carried out in the complex z plane. We also want to stress that the approximation schemes 1 and 2 described above require, in general, one-dimensional numerical integrations. (Besides the numerical solution of a transcendental equation in Approach 1.) As our method delivers an exact result, we believe that the two additional integrations needed here are worth the additional effort.

In our approach, we do not know $\Omega_g(Q)$ analytically, so we cannot directly calculate other quantities from it by, say, differentiation with respect to g. However, there exist integral representations for these quantities, similar to (1.3). They all have the form

$$\frac{1}{4\pi^2 i}\int^C dz\, zI_0(zQ_L)\int_0^\infty dx\, xJ_0(xQ_T)F[\Phi(z,x)]. \qquad (1.4)$$

$F[\Phi(z,x)]$ depends on the specific quantity considered, whereas Φ does not: Φ plays the same role in numerical calculations as does the grand partition function $\Omega_g(Q)$ in analytical calculations. Once we know Φ, we may then calculate everything through the double integration in Eq. (1.4).

The following discussion is restricted to the evaluation of $\Omega_g(Q)$. In this case we have $F[\Phi]=\exp(g\Phi)-1-g\Phi$. However, other functional forms of $F[\Phi]$ could be handled equally well.

The paper is organized as follows. In Sec. 2 we describe the difficulties one is faced with in our approach and show how to get rid of them. We then illustrate with a numerical example our method to evaluate $\Omega_g(Q)$. In Sec. 3 we present a similar (but simpler) second way to evaluate Ω_g in the high energy limit $Q_L\to\infty$ up to terms of order $O[Q_L^{2g-4}]$. The sample case from Sec. 2 is then evaluated with this method and compared with the first (exact) one. In the last section we summarize our results and propose a splitting of phase space according to presumably relevant calculational schemes.

2. EVALUATION OF Ω_g (Q)

The main difficulties connected with the numerical evaluation of the integral representation for Ω_g are the following:

(i) The integrands in (1.3) are highly oscillating, especially in the large p_T region. Formula (1.3) is therefore not well suited for *simple* numerical integration. This is the most baffling feature of our approach.

(ii) One of the integrations has to be carried out in the complex plane from $-i\infty+\epsilon$ to $i\infty+\epsilon$. Which path should we choose?

(iii) We need the values of the Bessel functions $K_0(z)$ and $I_0(z)$ for complex z. Computer programs in general require z to be real. Therefore we have to supply a method for calculating $K_0(z)$ and $I_0(z)$.

We dispose of these difficulties in turn.

A. Integration of rapidly oscillating fucntions

The integration of rapidly oscillating functions is an old problem in numerical mathematics, and there exists a long list of papers which deal with it in special cases like Fourier transformation.[17] Our oscillating functions, however, are only approximately of the trigonometric type, and we could not find in the literature a method directly applicable to our problem.[18] After many trials we finally found it most convenient to proceed in the following way. We first note that the oscillating integrands are of the type

$$h(x)=g(x)\times\{\sin(px) \text{ or } \cos(px)\}, \qquad (2.1)$$

where *$g(x)$ is smooth and slowly oscillating*. Furthermore we recall the Gaussian rule for numerical integration[20]

$$\int_a^b \varphi(x)\,dx\approx\frac{b-a}{2}\sum_{k=1}^{N}A_k^{(N)}\varphi\left[\left(\frac{b-a}{2}\right)x_k^{(N)}+\frac{b+a}{2}\right]. \qquad (2.2)$$

The coefficients $A_k^{(N)}$ denote the weights (independent of φ), and $x_k^{(N)}$ are the roots of the Legendre polynomial of degree N. The Gaussian rule delivers the exact result if $\varphi(x)$ is a polynomial of degree $2N-1$.

We shall use the following *empirical fact*: For $N=64$, the rule (2.2) applies also (with an accuracy of more than $\approx$ nine decimal figures) to $h(x)$ as given in (2.1), provided

$$|b-a|\leq 100/p \qquad (2.3)$$

for large values of p. For small p, there is no danger in using the Gaussian rule.

Stated differently, (2.2) applies to oscillating functions which are of the type in Eq. (2.1) provided that the interval $[a,b]$ does not contain more than ≈ 15 periods of the oscillating function. The fact that the Gaussian rule also applies in this case with a high accuracy is not so surprising: Remember that for $N=64$ (which is the case we are considering) the result of the numerical integration is exact whenever $\varphi(x)$ is a polynomial of degree 127. The assumption on the smoothness of $g(x)$ in Eq. (2.1) then just guarantees that $h(x)$ can be approximated very accurately by such a polynomial in the interval $[a,b]$. Consult the Appendix for illustrating examples.

Our rule has to be taken as a guide and (2.3) must eventually be improved, i.e., $|b-a|$ may have to be smaller than is allowed by (2.3). The need for an improvement in our calculations depends on the function $f(p)$ chosen in (1.3) and on the region in momentum space one is interested in. (See below.)

Now we can get rid of the problem with oscillating integrands as follows. The path C will be fixed in (ii) below. Split the region of integration in the variables (z,x) into surface elements such that in each element (2.3) is true with respect to x at fixed z and with respect to z after integration over x in Eq. (1.3). The evaluation of Φ has to be carried out analogously.

Remark: The function Φ has to be evaluated with an increasingly higher accuracy as Q^μ reaches the edge of phase space. The reason is the following: $\Omega_g(Q)$ goes to zero as Q^μ reaches the edge of phase space. There-

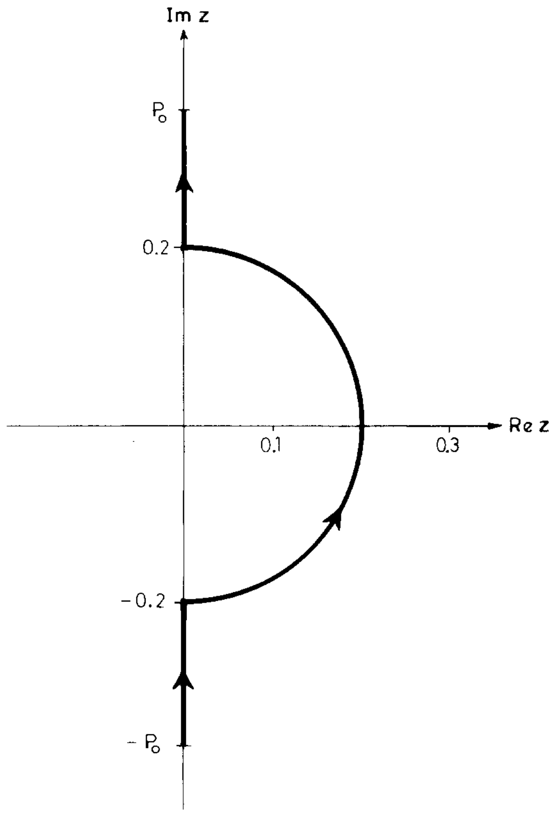

FIG. 1. The path C is the one chosen in the evaluation of the integral (1.3). The value of the cutoff P_0 depends on the accuracy required for the result.

fore the oscillations in the integrand have to enforce the vanishing of the value of Ω_g. The computer has to add and subtract (many times) large numbers in such a way that the result becomes small. Hence in order to obtain a reliable result the quantities $K_0(z)$, $I_0(z)$, and $\Phi(z,x)$ must be very accurately known. Note that small rounding errors in $\Phi(z,x)$ rapidly become uncontrollable since $\Phi(z,x)$ is exponentiated in Eq. (1.3). [In the calculations carried out in Sec. 2, Part D, the accuracy required for $K_0(z)$ and $\Phi(z,x)$ was 8–10 decimal figures.] This fact will most probably impose a limit on the application of our technique for Q^μ near the edge of phase space.

B. Choice of the path C in the complex plane

Since $\Phi(z,x)$ is analytic in z for $\mathrm{Re}z > 0$, Cauchy's theorem tells us that the value of the integral over z in Eq. (1.3) is independent of the choice for the path C, provided the latter runs from $-i\infty$ to $i\infty$ and lies in the half-plane $\mathrm{Re}z > 0$. However, life is not so easy if we are working with the computer: Due to rounding errors, certain paths C will be totally unsuited for numerical integration and will render meaningless results, although, according to Cauchy, we should always obtain the same answer.

For our application[15] we found it most convenient to choose the path shown in Fig. 1. (In the actual calculation, we only need to integrate over $\mathrm{Im}z > 0$. The location of the cutoff point P_0 depends on the accuracy required for the result. See Sec. 2, Part D.) It is certain that there exist other choices of C, equally well suited for our purposes, and it may well be that there exists even a better choice than ours. We have no compelling reason for our decision, but only the following arguments:

(1) The calculation with our choice of C works well and agrees with another method for the evaluation of $\Omega_g(Q)$ at high energies (see Sec. 2, Part D and Sec. 3).

(2) We were interested in our application[15] in the limit where Q_L becomes large. This limit is dominated by the behavior of

$$\exp[g\Phi(z,x)] - 1 - g\Phi(z,x)$$

near $z=0$. We therefore expect to pick up the main contribution to the integral (1.3) at small z. To determine what "small z" means in our case, we proceed as follows.

The integration over z (for fixed x) has the form

$$\int^C dz\, z^{-\rho} I_0(zQ_L) \tag{2.4}$$

with $\rho \approx 1-3$ at high energies and not too large values of x. (We chose $g=2$. This corresponds to a total cross section which behaves like $\sim\mathrm{const}/\ln s$ at high energies.) For a fixed value of ρ, there exists a number $R_{\min}$ such that the integral

$$\int^{C_R} dz\, z^{-\rho} I_0(zQ_L),$$

$$C_R = \{z \,|\, z = R\exp(i\theta),\ -\pi/2 \le \theta \le \pi/2,\ R \text{ fixed}\} \tag{2.5}$$

approximates (2.4) within 1% for *all* $R \ge R_{\min}$. The value of $R_{\min}$ depends on ρ and Q_L, but lies in the range $0.1 \le R_{\min} \le 0.2$ for $1 \le \rho \le 3$ and $30 \le Q_L \le 50$. (These are, in appropriate units, the values of Q_L appearing in Ref. 15.) We therefore expect, for our choice of the value of R, to pick up the main contribution to the integral from the semicircle C_R.

The path C shown in Fig. 1 is suited for the evaluation of the integral (1.3) at high energies and for a definite value of the parameters in the UJM. It would be advantageous to know whether this path is suited for the evaluation of $\Omega_g(Q)$ in other regions of momentum space and (or) for another choice of the parameters. We do not examine this problem in the present article.

C. Evaluation of the Bessel functions $I_0(Z)$ and $K_0(Z)$ for complex argument

According to our choice of the path C and for $Q_L \le 70$, we need to know the values of $I_0(z)$ for $|z| \le 14$, $\mathrm{Re}z > 0$, and for $z = ix$, $|x| \ge 14$. For $|z| \le 14$, a power series expansion for $I_0(z)$ is adequate and gives no problems in what is relevant to the accuracy of the result.[21] The evaluation of $I_0(ix)$, x real, is easy. So we are left with the calculation of $K_0(z)$ for complex argument. The simplest way to proceed is the following.[22] We match the representation valid for small $|z|$, i.e.,

$$K_0(z) = -\left\{\sum_{\nu=0}^{\infty}\left(\frac{z^2}{4}\right)^{\nu}\frac{1}{(\nu!)^2}\right\}\ln\left(\frac{Cz}{2}\right) + \sum_{\nu=1}^{\infty}\left(\frac{z^2}{4}\right)^{\nu} \times \frac{1}{(\nu!)^2}\left\{1+\frac{1}{2}+\cdots+\frac{1}{\nu}\right\}$$

$$C = e^{0.577\cdots}$$

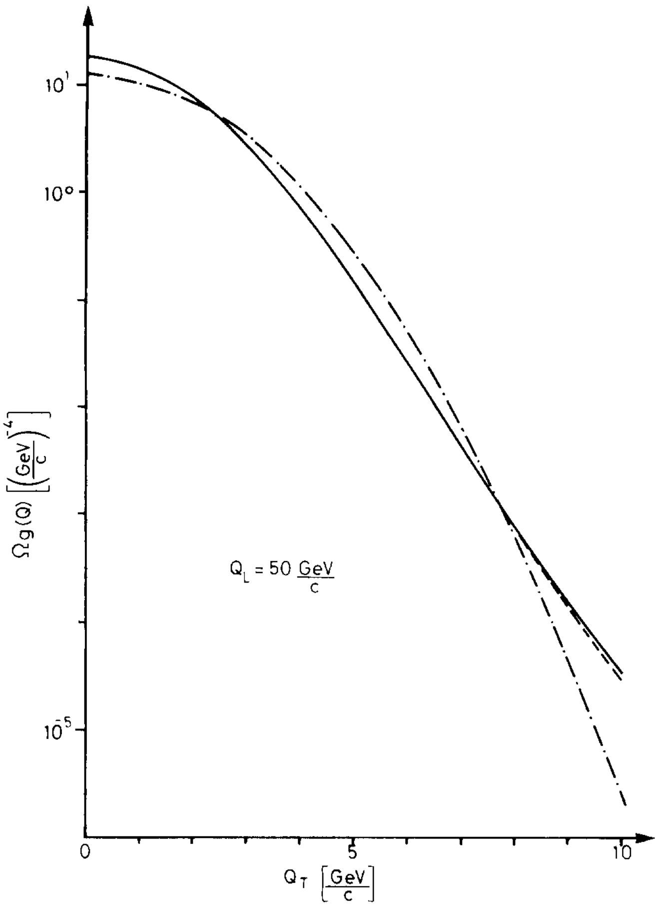

FIG. 2. $\Omega_g(Q)$ is plotted as a function of Q_T at fixed $Q_L = 50(GeV/c)$ for the parameters given in (2.6). ————: Exact result from (1.3). -------: Approximation (3.1). -·-·-: Leading asymptotic behavior (3.2)

with the asymptotic expansion

$$K_0(z) \sim \left(\frac{2\pi}{z}\right)^{1/2} e^{-z}\left\{1 - \frac{1}{8z} + \frac{1^2\cdot 3^2}{2!\,(8z)^2} - \cdots\right\},$$

$$|\arg z| < 3\pi/2, \quad |z| \to \infty$$

at some fixed value $R = |z|$. The discussion of point (i) above shows that the accuracy required for $K_0(z)$ and $I_0(z)$ depends on $f(p)$, and on the region in momentum space in which we want to evaluate $\Omega_g(Q)$. If we work on the computer with extended precision variables (≈ 35 significant figures), we may choose $R = 18$. This results in at least 13 decimal place accuracy for the value of $K_0(z)$. With double precision variables (≈ 16 significant figures), one should use[22] $R \approx 9$. The accuracy then drops to nine decimal figures.[22]

This concludes our discussion of the difficulties (i)—(iii) enumerated above.

D. Details of the calculation. An example

A few comments concerning the details of the calculation are in order at this stage.

As already pointed out in the Introduction, we need to calculate Φ only once for a given choice of the parameters in the UJM. (We may even use the same Φ for

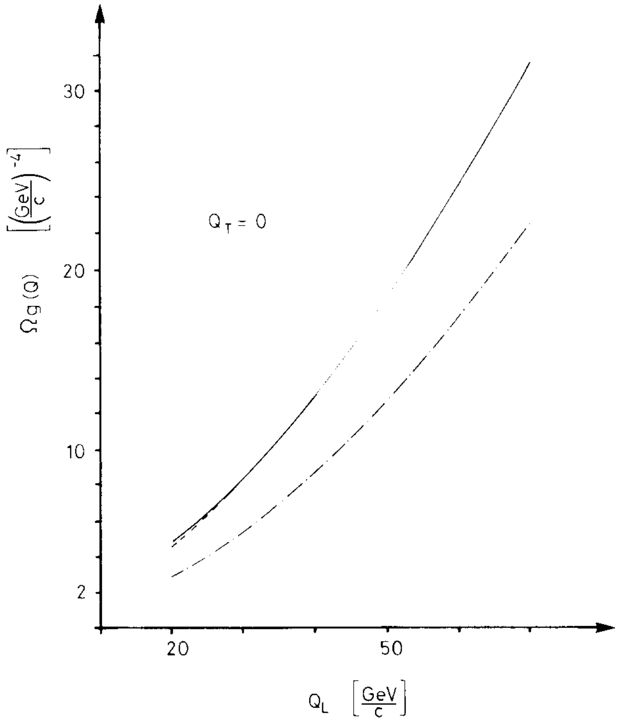

FIG. 3. $\Omega_g(Q)$ is plotted as a function of Q_L at $Q_T = 0$. Parameters and meaning of the lines are the same as in Fig. 2.

diferent values of the coupling constant g.) Φ may then be stored, e.g., on a tape, and recalled for the remaining integrations over x and z. In what concerns the integral over z, we actually need to integrate only in the upper half-plane $\mathrm{Im}z \geq 0$. The contributions from the lower half-plane may then be obtained via the Schwarz reflection principle.

We show in Figs. 2—4 an example of our calculations carried out along these lines. The values chosen for the parameters are

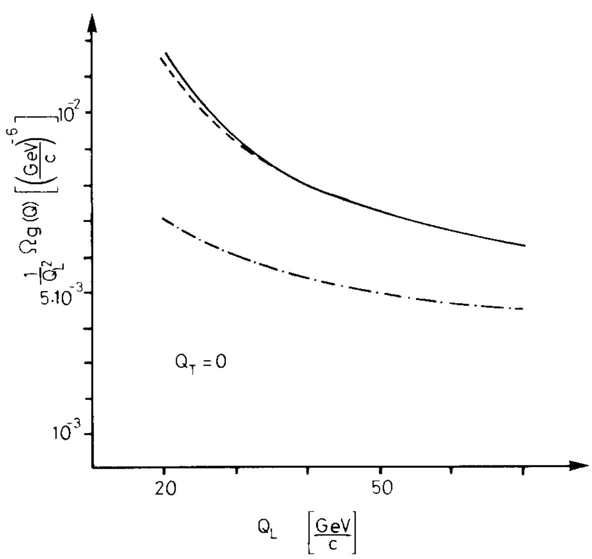

FIG. 4. $\Omega_g(Q)/Q_L^2$ is plotted as a function of Q_L at $Q_T = 0$. Parameters and meaning of the lines are the same as in Fig. 2.

$$f(p)=\frac{6(2\pi)^3}{\pi}\frac{1}{(1+p^2)^4}\left(\frac{GeV}{c}\right)^{-2},$$
$$g=2,\quad m=1\frac{GeV}{c^2}. \tag{2.6}$$

(We have also drawn in these figures the result from two approximate calculations to be discussed in the next section.)

Those curves in Figs. 2–4 which were obtained via the representation (1.3) have been calculated using $P_0=0.4$ in Fig. 1. We remark that the values so obtained agree with the "improved high energy expansion" from Sec. III within $\frac{1}{2}\%$ or better for $Q_T=0$ and $Q_L \geq 50GeV/c$ (Fig. 3). For the values in Fig. 2 we find a 3% discrepancy at $Q_T=5GeV/c$. This discrepancy then increases and becomes 15% at $Q_T=10GeV/c$. It should be noted, however, that the "improved high energy expansion" cannot be true at large Q_T since it does not respect energy–momentum conservation (see Sec. 3). The 15% discrepancy should therefore not be considered as a defect of our exact method. In fact we feel that the above comparison of these two different methods to obtain $\Omega_g(Q)$ reveals that the numbers we obtain from the representation (1.3) are trustworthy at least at the level of a few percent.

3. IMPROVED HIGH-ENERGY LIMIT

In view of the subtleties involved in the numerical evaluation of (1.3), it may be advantageous to have available a simpler method to calculate $\Omega_g(Q)$ approximately. Although the method presented below will be valid only in the high-energy limit and for Q_T not too large, it serves as a welcome test of the procedure described in Sec. 2. At the same time, the technique will go far beyond the leading order approximations mentioned in the Introduction.

The expansion of Ω_g for high energies reads[4,12]

$$\Omega_g(Q)=\frac{1}{\pi Q_L^2}\int_0^\infty dx\, xJ_0(xQ_T)\frac{1}{[\Gamma(gD_{-1}(x))]^2}\left(\frac{Q_L}{2m}\right)^{2gD_{-1}(x)}$$
$$\times\exp[-gD_0(x)]\left\{1+O\left(\frac{\ln Q_L}{Q_L^2}\right)\right\},\quad Q_L\to\infty,\quad Q_T \text{ fixed},$$

where

$$\Phi(z,x)=-2D_{-1}(x)\ln(mz)-D_0(x)+O(z^2\ln z),\quad z\to 0$$

and

$$D_{-1}(x)=\frac{\pi}{(2\pi)^3}\int_0^\infty dp\, pJ_0(xp)f(p),$$
$$D_0(x)=\frac{1}{(2\pi)^2}\int_0^\infty dp\, pJ_0(xp)f(p)\ln\left(\frac{C_{m_T}}{2m}\right). \tag{3.1}$$

$\Phi(z,x)$ is defined in (1.3).

We note that the first part in (3.1) contains only a two-fold integration: Difficulties (ii) and (iii) described at the beginning of the last section have disappeared. The approximation scheme announced above consists in the numerical evaluation of the integrals in Eq. (3.1). An expansion similar to Eq. (3.1) exists for all quantities in the UJM. The integral representation in Eq. (3.1) presents therefore an extremely simple way to obtain a relatively accurate answer and provides a check for the exact results (at high energies).

In Figs. 2–4 we compare the values of Ω_g evaluated by three different methods:

(i) exact calculation as described in Sec. 2,

(ii) approximation according to (3.1),

(iii) leading asymptotic behavior of Ω_g at high energies[4,12]

$$\Omega_g(Q)=\frac{e^{-gD_0(0)}}{8\pi m^2 D_{-1}''(0)}\frac{1}{g[\Gamma(g)]^2}\left(\frac{Q_L}{2m}\right)^{2(g-1)}\frac{1}{\ln(Q_L/2m)}$$
$$\times\exp\left(\frac{-Q_T^2}{4gD_{-1}''(0)\ln(Q_L/2m)}\right)\left\{1+O\left(\frac{1}{\ln Q_L}\right)\right\},$$
$$Q_L\to\infty,\quad Q_T\text{ fixed}$$

where

$$D_{-1}(x)=1-\frac{x^2}{2!}D_{-1}''(0)+O(x^4),\quad x\to 0$$

and

$$D_{-1}''(0)=\frac{\pi}{2(2\pi)^3}\int_0^\infty dp\, p^3 f(p). \tag{3.2}$$

Remarks: (1) The size of the terms neglected in the two high-energy expansions presented above [Eqs. (3.1) and (3.2)] are not the same in both cases. Let $\Delta\Omega_g(Q)$ be the *difference* between $\Omega_g(Q)$ and its approximating expression. Then one finds that $\Delta\Omega_g(Q)$ is of the order $O(Q_L^{2g-4})$ and $O(Q_L^{2g-2}/(\ln Q_L)^2)$ in Eqs. (3.1) and (3.2), respectively.

(2) The difference between the exact result and the leading asymptotic behavior (3.2) diverges as $Q_L\to\infty$, as seen in Fig. 3. This is due to the fact that this difference is of order $O(Q_L^{2g-2}/(\ln Q_L)^2)$, as is mentioned in point (1). For completeness, we show in Fig. 4 the function

$$(1/Q_L^2)\Omega_g(Q)$$

which appears in actual calculations of physical quantities. Clearly the exact result and the leading asymptotic form coincide in this case as $Q_L\to\infty$, their difference being of order $O(1/(\ln Q_L)^2)$ for $g=2$.

(3) The drawback of the expansion (3.1) is the fact that we have lost energy–momentum conservation which was built into the representation (1.3). However, Fig. 2 shows that this affects the value of Ω_g only at fairly large transverse momentum Q_T for our choice of the parameters m, g, and $f(p)$.

(4) It would be easy to include correction terms in the expansion (3.1). We did not check, however, how they affect the result at large Q_T.

4. SUMMARY AND CONCLUSIONS

(i) We propose to use the integral representation (1.3) for the evaluation of the partition function $\Omega_g(Q)$ in the UJM.

(ii) The most severe difficulty of this procedure is due to the rapidly oscillating integrands in (1.3). They are not well suited for naive numerical integration.

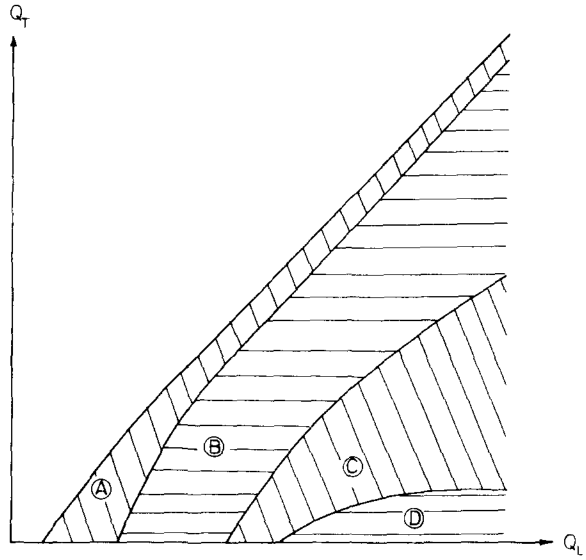

FIG. 5. The region of nonvanishing $\Omega_g(Q)$ is partitioned according to presumably relevant calculational schemes. (A): Monte-Carlo technique. (B): Integral representation (1.3). (C): High-energy approximation (3.1). (D): Leading asymptotic behavior (3.2), statistical methods.[6–8] See also Ref. 5. No scale is given on the coordinate axis.

(iii) This difficulty can be overcome by using the *empirical fact* that the Gaussian rule for numerical integration (2.2) also applies for highly oscillating functions, as amplified in Sec. 2 and in the Appendix.

(iv) Most probably there will be a limitation to the application of our method near the edge of phase space (large Q_T or small Q_L) due to rounding errors.

(v) The simpler representation (3.1) serves as a quick check for the exact calculation at large Q_L and moderate Q_T (see Figs. 2–4).

(vi) We conclude that different techniques are appropriate for the evaluation of $\Omega_g(Q)$ in different regions of phase space. We *propose* a splitting of phase space as shown in Fig. 5. (The Monte–Carlo technique may be useful near the edge of phase space because of the small number of particles produced.) For the scale on the coordinate axis in Fig. 5, see point (vii).

(vii) It would be highly useful to make the statements "large Q_L," "moderate Q_T," "near the edge of phase space," as used above more precisely. Ultimately we would like to supply the scale on the Q_L- and Q_T- axis in Fig. 5. We are not able to do so.

TABLE I. Comparison between exact and numerical integration of the functions $h_i(x)$ defined in the Appendix. We tabulate the number of decimal figures to which the numerical integration agrees with the exact result. The integration interval is $[0, 100/p]$.

Function \ p	3	5	10	30	50
$h_1(x)$	13	13	13	13	13
$h_2(x)$	14	15	13	14	15
$h_3(x)$	16	15	14	15	14
$h_4(x)$	9	12	12	13	13

TABLE II. Comparison between exact and numerical integration of the functions $h_i(x)$ defined in the Appendix. We tabulate the number of decimal figures to which the numerical integration agrees with the exact result. The integration interval is $[10, 10+100/p]$.

Function \ p	3	5	10	30	50
$h_1(x)$	13	13	13	11	13
$h_2(x)$	13	13	12	12	13
$h_3(x)$	13	14	15	11	13
$h_4(x)$	13	14	14	12	13

ACKNOWLEDGMENTS

I wish to thank Professor Polkinghorne for the kind hospitality at DAMTP, where most of this work has been done. I profited from discussions with S. D. Ellis, J. Engels, and J. Fleischer.

APPENDIX

In order to illustrate our empirical rule (2.3), we present in Tables I and II a comparison between exact and numerical integration of the following four functions:

$$h_1(x) = \cos(px),$$

$$h_2(x) = x^3 \sin(px),$$

$$h_3(x) = \exp(-x)\sin(px),$$

$$h_4(x) = 40\left\{\frac{x}{(1+x^2)^3}\sin(px) - p\cos(px)\right\}\exp\frac{1}{4(1+x^2)^2}$$

$$\equiv h_{41}(x)\sin(px) + h_{42}(x)\cos(px).$$

Note that

$$\int h_4(x)\,dx = -40\sin(px)\exp[1/4(1+x^2)^2] + \text{const.}$$

The tables give the number of decimal figures to which the numerical integration

$$\int_a^b h_i(x)\,dx \approx \frac{b-a}{2}\sum_{k=1}^{64} A_k^{(64)} h_i\left[\left(\frac{b-a}{2}\right)x_k^{(64)} + \frac{b-a}{2}\right]$$

agrees with the exact value. The integration interval was chosen to be $[0, 100/p]$ in Table I and $[10, 10+100/p]$ in Table II.

A particularly interesting example is the function $h_4(x)$. Although $h_{41}(x)$ is sharply peaked at $x = 0.42$ (value at the peak $= 12.4$), the result for $p = 5$ is still accurate to 12 decimal figures. Note that in this case the integration interval in Table I is $[0, 20]$ and contains therefore *both* the peak and the tail of $h_{41}(x)$.

*Research supported in part by the European Science Exchange Program and Schweizerischer Nationalfonds.

†Present address: Institute for Theoretical Physics, University of Berne, Sidlerstrasse 5, CH-3012 Berne, Switzerland.

[1] L. Van Hove, Rev. Mod. Phys. **36**, 655 (1964).

[2]A. Krzywicki, Nuovo Cimento **32**, 1067 (1964); J. Math. Phys. **6**, 485 (1965).
[3]E.H. de Groot and Th.W. Ruijgrook, Nucl. Phys. B **27**, 45 (1971).
[4]A. Bassetto, L. Sertorio, and M. Toller, Nuc. Phys. B **34**, 1 (1971).
[5]E.H. de Groot, Nucl. Phys. B **48**, 295 (1972).
[6]F. Lurçat and P. Mazur, Nuovo Cimento **31**, 140 (1964).
[7]D. Sivers and G. Thomas, Phys. Rev. D **7**, 1961 (1972).
[8]S. Humble, G.D. Kaiser, and J.N.J. White, Nucl. Phys. B **56**, 61 (1973).
[9]L. Van Hove, Nucl. Phys. B **9**, 331 (1969).
[10]W. Kittel, L. Van Hove, and W. Wojcik, Comput. Phys. Commun. **1**, 425 (1970).
[11]M.S. Chen and R.F. Peierls, Phys. Rev. D **7**, 2183 (1972).
[12]R. Baier, J. Engels, H. Satz, and K. Schilling, Nuovo Cimento A **28**, 455 (1975).
[13]P.V. Landshoff, rapporteur talk, in Proc. of the 17th *International Conference on High Energy Physics*, edited by J.R. Smith (Science Research Council, London, 1974).
[14]S.D. Ellis, review talk, in Ref. 13.
[15]S.D. Ellis and J. Gasser, "The Role of Momentum Conservation in Large p_T Events," DAMTP preprint (DAMTP 76|10).
[16]M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, (Dover, New York, 1970).
[17]L.N.G. Filon, R. Soc. Edinburgh **49**, 38 (1928); Y.L. Luke, Proc. Cambridge Philos. Soc. **50**, 269 (1954); W.W. Clendenin, Numer. Math. **8**, 422 (1966); A.I. van de Vooren and H.J. van Linde, Math. Comput. **20**, 232 (1966); E.O. Tuck, Math. Comput. **21**, 239 (1967); N. Lyness, Math. Comput. **24**, 101 (1970); Math. Comput. **25**, 59 (1971).
[18]After the completion of this work, we learned that J. Engels and J. Fleischer also use the representation (1.3) for calculational purposes. These authors treat the rapidly oscillating integrands with the method described in Ref. 19.
[19]J.S.R. Chisholm, A. Genz, and G.E. Rowlands, J. Comput. Phys. **10**, 284 (1972).
[20]V.J. Krylov, *Approximate Calculation of Integrals* (Macmillan, New York, 1962), p. 100–11.
[21]H.A. Scarton, J. Comput. Phys. **8**, 295 (1971).
[22]H.H. Lowell, Math. Comput. **14**, 81 (1960).

On the implementability of local gauge transformations in a theory with localized states

J. P. Leveille

Department of Physics, Boston University, Boston, Massachusetts 02215

P. Roman*

Max Planck Institut, D-831 Starnberg, Germany

(Received 3 March 1976; revised manuscript received 5 April 1976)

A conflict between unitary implementability of local gauge transformations (kind one) on the one hand and certain properties of sets of localized states on the other is deduced in the axiomatic framework of relativistic local quantum field theory.

I. INTRODUCTION

In this note we wish to study whether local gauge transformations (of the first kind) are kinematical transformations (in the sense defined by Jauch[1]) in a relativistic quantum field theory with localized states. It has been known for some time that, in a local field theory, it seems to be difficult to obtain representations where both the Poincaré and the gauge group are unitarily implemented. For example, dell'Antonio[2] studied the incompatibility of unitary representations of a gauge group with certain types of (canonical) representations of field operators. In our, rather different, approach we study the conflict between the unitary representability of local gauge transformations on one hand and certain properties of localized states, on the other. In our study we shall use an algebraic field theory framework.[3] We denote by $\mathcal{J}(\Omega)$ the local field algebra associated with the open and bounded region Ω of space–time. Without loss of generality we can assume that $\mathcal{J}(\Omega)$ has been already extended to a von Neumann local field algebra.

2. PREPARATIONS

Following Knight,[4] we start with

Definition 1: A state $|\psi\rangle$ is said to be localized in Ω iff $\langle\psi|F|\psi\rangle=\langle 0|F|0\rangle$ for all $F\in\mathcal{J}(\Omega')$, where $\mathcal{J}(\Omega')$ is the algebra generated by all $\mathcal{J}(\Omega_s)$ with Ω_s spacelike separated from Ω.

We then easily establish the following:

Lemma 1: If U is any unitary operator belonging to $\mathcal{J}(\Omega)$, then the state $|\psi\rangle_\Omega\equiv U|0\rangle$ is localized in Ω.

Proof: ${}_\Omega\langle\psi|F|\psi\rangle_\Omega=\langle 0|U^\dagger FU|0\rangle=\langle 0|U^\dagger UF+U^\dagger[F,U]|0\rangle$. But if $F\in\mathcal{J}(\Omega')$ then, by local commutativity, $[F,U]=0$; hence ${}_\Omega\langle\psi|F|\psi\rangle_\Omega=\langle 0|F|0\rangle$. QED

We use this result to deduce

Lemma 2: Let $\mathcal{L}_\Omega\equiv\{U|0\rangle\,|\,U\in\mathcal{J}(\Omega),\ U \text{ unitary}\}$. Then $\mathcal{L}_\Omega$ is a dense set.

Proof: Since any element of a von Neumann algebra is a linear combination of unitary operators, the linear span of $\mathcal{L}_\Omega$ contains the set $\mathcal{R}\equiv\{F|0\rangle\,|\,F\in\mathcal{J}(\Omega)\}$. The Reeh–Schlieder theorem tells us[5] that $\mathcal{R}$ is dense. But

$$\bar{\mathcal{L}}_\Omega\supset \text{linear span of } \mathcal{L}_\Omega\supset\mathcal{R};$$

hence $\mathcal{L}_\Omega$ is dense. QED

This lemma leads to the important

Theorem 1: Subject to the usual assumptions of local relativistic quantum field theory, the set of states localized in any Ω (as specified by Definition 1) is dense.

Proof: Clearly, every element of $\mathcal{L}_\Omega$ is a localized state and so the set of all localized states contains the dense set $\mathcal{L}_\Omega$.

We conclude this section by formalizing what we mean by a local gauge transformation (of the first kind) in our framework. We generalize the familiar situation of a naive theory {where a local gauge transformation is acting on the state function at x via $\psi(x)\to\exp[i\omega(x)]\psi(x)$} by the following

Definition 2: Let G be a global (i.e., nongauged) group which is unitarily implemented on the Hilbert space $\mathcal{H}$. Then any map ω from the set of all open and bounded space–time regions Ω into the group G, given by

$$\Omega\mapsto\omega(\Omega)\in G, \tag{2.1}$$

will be called a local gauge transformation. [We note here that we do not assume that $\omega(\Omega)$ is the identity outside some bounded region.]

From this definition, and from the fact that we have in mind a generalization of the naive situation, it follows that the action of a local gauge transformation ω on $\mathcal{H}$ will be described by some operator U_ω which has the property that, for any state $|\psi\rangle_\Omega$ localized in Ω, we have

$$U_\omega|\psi\rangle_\Omega=U_{\omega(\Omega)}|\psi\rangle_\Omega, \tag{2.2}$$

where $U_{\omega(\Omega)}$ is the unitary operator that corresponds (in the assumed unitary representation of the global group G on $\mathcal{H}$) to the particular group element $g=\omega(\Omega)$.

3. THE MAIN THEOREM

We now ask: Are local gauge transformations kinematical symmetries, i.e., can U_ω be unitary on $\mathcal{H}$ for all local gauge transformations (2.1)? The essentially negative answer to this question is formalized by

Theorem 2: If the assumptions under which Theorem 1 is valid hold and if U_ω is unitary for all ω, then $\omega(\Omega)$ must not depend on Ω, i.e., the presumed representation U_ω of the local gauge group reduces to a representation of the global group G.

Proof: Let the set of states localized in Ω be

$$\mathcal{S} \equiv \{|\psi_n\rangle_\Omega \,|\, n \in I\},$$

where I is some index set. By Theorem 1, $\mathcal{S}$ is dense.

By Eq. (2.2), we have

$$U_\omega |\psi_n\rangle_\Omega = U_{\omega(\Omega)} |\psi_n\rangle_\Omega \tag{3.1}$$

for all ω and every n. Since $\mathcal{S}$ is dense, we can extend U_ω unitarily and uniquely to all of $\mathcal{H}$. Now, if $|\psi\rangle$ is an arbitrary state in $\mathcal{H}$, then, again because of the denseness of $\mathcal{S}$,

$$|\psi\rangle = \lim_{n\to\infty} \sum_{k=1}^{n} |\psi_k\rangle_\Omega, \tag{3.2}$$

so that

$$\begin{aligned} U_\omega |\psi\rangle &= \lim \sum U_\omega |\psi_k\rangle_\Omega = \lim \sum U_{\omega(\Omega)} |\psi_k\rangle_\Omega \\ &= U_{\omega(\Omega)} \lim \sum |\psi_k\rangle_\Omega = U_{\omega(\Omega)} |\psi\rangle. \end{aligned} \tag{3.3}$$

Choosing now a different region $\hat{\Omega}$ and repeating this calculation, we get

$$U_\omega |\psi\rangle = U_{\omega(\hat{\Omega})} |\psi\rangle. \tag{3.4}$$

Since (3.3) and (3.4) are supposed to hold for all ω, we have

$$U_{\omega(\Omega)} = U_{\omega(\hat{\Omega})} \quad \text{for all } \omega,\ \Omega,\ \hat{\Omega}. \tag{3.5}$$

The only solution is: $U_{\omega(\Omega)}$ is independent of Ω, from which the statement of the theorem follows directly.

4. DISCUSSION

The nonimplementability of the gauge group we deduced above hinges upon the denseness of the sets $\mathcal{S}$ of Ω-localized states. In other words, we found that *unitary implementability of a local gauge group (kind one) and denseness of the set of Ω-localized states are in conflict*. If we wish to insist on implementability, the denseness of $\mathcal{S}$ must be avoided. Admittedly, this is not an easily acceptable step, since intuitively one expects that localized states are "complete" in some sense, and it is hard to see how this could be achieved if $\mathcal{S}$ is not dense. Nevertheless, we may wish to pursue this possibility. Since the essential ingredient of Theorem 1 is the Reeh–Schlieder theorem, we may enquire about relaxing conditions that lead to it.[6] The major (specific) assumptions of the Reeh–Schlieder theorem are[5]

(a) the spectral conditions,

(b) "weak additivity",[7]

(c) positive definite metric of $\mathcal{H}$.

One would not be willing[8] to give up assumption (a). The abandonment of (b) would lead to the nonunitarity of translations. Thus, it appears that the only relaxable assumption is (c). In fact, in a recent paper Strocchi and Wightman[9] demonstrate that in a full gauge theory (with the vector gauge fields present) the use of an indefinite metric is a necessity. Thus, once the gauge fields have been introduced, the lack of a definite metric can invalidate the Reeh–Schlieder theorem[10] and hence may render our Theorem 1 inoperative so that Theorem 2 will not follow. Further, it is well known that the necessary *existence* of gauge fields (though not their equation of motion) follows from demanding gauge *invariance* (via the necessity of introducing covariant derivatives). On the other hand, it is hard to see how such an argument would work outside the framework of a Lagrangian formulation.

It appears that there may be a completely different possibility of achieving unitary implementability of local gauge transformations, without abandoning the denseness of localized states.[11] Indeed, our Definition 2 of local gauge transformations, assuming as a *prerequisite* the existence and unitary implementability of a global group which it "turns into a local group," could be altered/generalized. For example, one may think of replacing the (single) map ω by a family of maps ω_Ω each of which vanishes outside a compact region. Then our conclusions would not necessarily hold.

ACKNOWLEDGMENT

We are obliged to Professor H. Reeh (Göttingen) for having critically commented on this paper while it was in press. Primarily, he emphasized that our definition of local gauge transformations is perhaps too much intuitive and the central theorem could be avoided by adopting a more suitable definition.

*Permanent address: Department of Physics, Boston University, Boston, Mass. 02215.

[1]J.M. Jauch, Helv. Phys. Acta **37**, 284 (1964).

[2]G.F. dell'Antonio, J. Math. Phys. **12**, 148 (1971).

[3]R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964); S. Doplicher, R. Haag, and J.E. Roberts, Comm. Math. Phys. **13**, 1 (1969); **15**, 173 (1969). For a brief review cf. N.N. Bogolubov, A.A. Logunov, and I.T. Todorov, *Introduction to Axiomatic Field Theory* (Benjamin, Reading, Mass., 1975).

[4]J.M. Knight, J. Math. Phys. **2**, 459 (1961). See also R. Haag and J.A. Swieca, Comm. Math. Phys. **1**, 307 (1965), and H. Araki and R. Haag, *ibid.* **4**, 77 (1967).

[5]H. Reeh and S. Schlieder, Nuovo Cimento **22**, 1051 (1961).

[6]Another possibility to avoid denseness of $\mathcal{S}$ would be to change the commonly accepted Definition 1 of localized states to something else (from which Lemmas 1, 2 and hence Theorem 1 do not follow), but of course, there is no hint what such a new definition could be.

[7]Cf. also H.J. Borchers, Comm. Math. Phys. **10**, 269 (1968).

[8]It may be worth mentioning in this context that in a previous paper [P. Roman and J.P. Leveille, J. Math. Phys. **15**, 2053 (1974)] we constructed on purely group theoretical (not field theoretic) grounds a consistent relativistic quantum dynamics where there was no conflict between unitary implementability of a local gauge group and localized statefunctions, but the price we had to pay for this was the emergence of a continuous mass spectrum $-\infty < m^2 < +\infty$.

[9]F. Strocchi and A.S. Wightman, J. Math. Phys. **15**, 2198 (1974).

[10]Strictly speaking, it deserves a detailed study to ascertain how, in the particular structure of the Hilbert space used for gauge theories, the Reeh–Schlieder theorem becomes modified.

[11]We are greatly obliged to Dr. K. Drühl (Max Planck Institut, Starnberg) for calling our attention to this point.

Gauge theories and nonrelativistic cosmological symmetries

P. Roman*

Max-Planck-Institut, D-813 Starnberg, Postfach 1529, Germany

J. Haavisto

Department of Physics, Boston University, Boston, Massachusetts 02215
(Received 22 March 1976)

It is shown that the application of the locality principle in a uniformly curved space leads to the emergence of a dynamical quantum mechanical group which is precisely the Hooke group. The interaction structure is also studied.

1. INTRODUCTION

It is well known that the central extension of the Galilei group provides a complete, concise, and beautiful algebraic description of nonrelativistic quantum dynamics in flat space.[1] It is also common knowledge that the Galilei group is the "speed-space" contraction of the Poincaré group, arising from it by substituting $N_{k0}\to \epsilon N_{k0}$, $P_k \to \epsilon P_k$ and taking the limits[2]

$$\lim_{\epsilon\to 0}\epsilon N_{k0}=\hat{\mathbf{Q}}, \tag{1.1a}$$

$$\lim_{\epsilon\to 0}\epsilon P_k=\hat{\mathbf{P}}. \tag{1.1b}$$

The Poincaré group, in turn, is the "space–time" contraction of the de Sitter group, arising via the substitutions $J_{k4}\to \epsilon J_{k4}$, $J_{04}\to \epsilon J_{04}$ and the subsequent taking of the limits

$$\lim_{\epsilon\to 0}\epsilon J_{k4}=P_k, \tag{1.2a}$$

$$\lim_{\epsilon\to 0}\epsilon J_{04}=P_0. \tag{1.2b}$$

These considerations permit one to attach a geometrical interpretation to Galilean quantum dynamics. The de Sitter world is the simplest cosmological model, possessing highest possible symmetry. Because of (1.1), the flat Minkowski world, with Poincaré symmetry, corresponds to an approximation of the de Sitter world where only small spacelike and small timelike intervals (compared to those of the cosmological model) are considered, but speeds have not been restricted. Because of (1.2), the Euclidean world, with Galilei symmetry, corresponds to a further approximation, where only small spacelike intervals (compared to the timelike intervals) and small speeds (compared to the unit c) are considered.[3]

Less well known is the fact that, as discovered by Bacry and Lévy-Leblond,[4] the cosmological de Sitter group also allows for another important contraction, namely, one of the "space-speed" type: Make the substitutions $J_{k4}\to \epsilon J_{k4}$, $J_{k0}\to \epsilon J_{k0}$ and take the limits

$$\lim_{\epsilon\to 0}\epsilon J_{k4}=\mathbf{P}, \tag{1.3a}$$

$$\lim_{\epsilon\to 0}\epsilon J_{k0}=\mathbf{Q}. \tag{1.3b}$$

Thus, we obtain a world which is an approximation of the de Sitter world in which only small spacelike intervals (compared to those of the de Sitter world) and small speeds (compared to the unit c), *but arbitrarily large timelike intervals* are considered.

The generators $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{J}\equiv J_{kl}$, $H\equiv J_{04}$ form a Lie group[5] which differs from that of the Galilei group only inasmuch that the commutator of H and $\mathbf{P}$ is not zero but is proportional to $\mathbf{Q}$. Since (as in the Galilei group but unlike as in the de Sitter and Poincaré groups) the time translation generator does not occur on the rhs of the space translation and boost commutator, time intervals are not invariant under boosts, from which follows that we have an "absolute time."

This new group and its unitary representations have been further studied in detail by Derome and by Dubois[6–8] who also gave it the now commonly accepted name "Hooke group." The name derives from the fact that the Casimir invariant corresponding to internal energy [cf. Eq. (A5b)] contains a harmonic potential. The presence of this term can be interpreted as the long-range effect of curvature, since we are considering the universe *on a large scale of time*.[9]

In summary, the Hooke group deserves serious interest because it describes low-speed (nonrelativistic) transformations of a universe *at large*, endowed with an absolute time. Thus, it is (in contrast to the Galilei and Poincaré group which are "local") still a "cosmological" group. We may say that it summarizes the dynamics of a nonrelativistic universe at large. We are entitled to call this model a *nonrelativistic cosmological world*.

It will not come as a surprise to note that the Euclidean world (with the Galilei symmetry) is an approximation of this nonrelativistic cosmological world (with its Hooke symmetry): The Galilei group is a "space–time" contraction of the Hooke group,[10] obtained by substituting $\mathbf{P}\to\epsilon\mathbf{P}$, $H\to\epsilon H$ and performing the limits

$$\lim_{\epsilon\to 0}\epsilon\mathbf{P}=\hat{\mathbf{P}}, \tag{1.4a}$$

$$\lim_{\epsilon\to 0}\epsilon H=\hat{H}. \tag{1.4b}$$

Some time ago, being inspired by the earlier work of Jauch[11] which pointed to remarkable connections between Galilean symmetry and gauge symmetry, one of us[12,13] showed that the entire Galilean structure of nonrelativistic quantum dynamics (including the structure of interactions and various superselection rules) in a flat Euclidean world can be simply derived from the basic requirement of local phase (gauge) symmetry and a few additional, rather obvious assumptions. Since, because of the reasons outlined above (see also Appendix A), we believe that the Hooke group can provide considerable insight into quantum dynamics of curved spaces, we find it worthwhile to explore in the present work whether it can be derived from the locality principle, in analogy to the Galilei group, and study what special features arise.

2. THE KINEMATICAL GROUP

Our ultimate aim is to build, from first principles, quantum kinematics and then quantum dynamics in the nonrelativistic approximation for a "cosmological," curved space. For simplicity, and also because it has the highest possible symmetry, we take this three-space to have constant curvature. Accordingly, we adopt

Assumption 1: The space of events is the homogeneous and isotropic three-dimensional space $\mathcal{S}_3$ of constant curvature.

This space can be imbedded in a four-dimensional flat space where it becomes represented as a pseudo three-sphere S with[14] radius r,

$$x_4^2 - \mathbf{x}^2 = r^2. \tag{2.1}$$

In accord with current cosmological beliefs, we decided to assume a positive curvature,[15] $r^2>0$, i.e., to choose for imbedding an $E_{3,1}$ space with metric $g_{44}=-g_{11}=-g_{22}=-g_{33}=1$, $g_{\mu\nu}=0$ for $\mu\neq\nu$. The group of symmetries for $\mathcal{S}_3$ (as defined in Assumption 1) is then equivalent to "rotations" of S, hence isomorphic to SO(3,1) with the Lie algebra

$$[M_{\mu\nu}, M_{\rho\sigma}] = g_{\mu\rho}M_{\nu\sigma} + g_{\nu\sigma}M_{\mu\rho} - g_{\mu\sigma}M_{\nu\rho} - g_{\nu\rho}M_{\mu\sigma} \quad (\mu,\nu=1,2,3,4). \tag{2.2}$$

It is convenient to introduce the notations

$$J_k \equiv \tfrac{1}{2}\epsilon_{klm}M_{lm}, \quad \Pi_k \equiv -r^{-1}M_{k4} \quad (k,l,m,=1,2,3). \tag{2.3}$$

Then we have

$$[\Pi_i,\Pi_j] = ir^{-2}\epsilon_{ijk}J_k, \quad [J_k,\Pi_i] = -i\epsilon_{ikl}\Pi_l, \quad [J_i,J_k] = i\epsilon_{ikl}J_l. \tag{2.4}$$

This algebra can be realized in the Hilbert space of square integrable functions $\psi(\mathbf{x})$ on S as follows[16]:

$$\Pi_k \sim -ir^{-1}(r^2+\mathbf{x}^2)^{-1/2}[x_i(x_k\partial_i - x_i\partial_k) + r^2\partial_k], \tag{2.5a}$$

$$J_k \sim -i\epsilon_{klm}x_l\partial_m. \tag{2.5b}$$

Next, we proceed to formulate the crucial *locality postulate*, i.e., we demand that a local phase transformation be an automorphism of the Hilbert space. More precisely, we introduce[17]

Assumption 2: To every transformation

$$\psi(\mathbf{x}) \to \exp[i\omega(\mathbf{x})]\psi(\mathbf{x}) \tag{2.6}$$

with a differentiable $\omega(\mathbf{x})$ there corresponds in Hilbert space a unitary operator $\mathcal{U}$ such that

$$(\mathcal{U}\psi)(\mathbf{x}) = \exp[i\omega(\mathbf{x})]\psi(\mathbf{x}). \tag{2.7}$$

Using the realization (2.5a) and Eq. (2.1) we now calculate[18]

$$\begin{aligned}(\mathcal{U}\Pi_k\mathcal{U}^{-1}\psi)(\mathbf{x}) &= \exp(i\omega)\{-ir^{-1}x_4^{-1}[x_i(x_k\partial_i - x_i\partial_k) + r^2\partial_k]\}\exp(-i\omega)\psi(\mathbf{x})\\ &= \{\Pi_k - r^{-1}x_4^{-1}[x_ix_k\partial_i\omega - x_ix_i\partial_k\omega + r^2\partial_k\omega]\}\psi(\mathbf{x}),\end{aligned}$$

i.e.,

$$\Pi_k \to \Pi_k - r^{-1}x_4^{-1}(x_ix_k\partial_i\omega - x_ix_i\partial_k\omega + r^2\partial_k\omega). \tag{2.8}$$

Similarly, with (2.5b) we get

$$J_k \to J_k - \epsilon_{klm}x_l\partial_m\omega. \tag{2.9}$$

As in Ref. 12, we wish to insist that the local phase transformation (2.6) be a kinematical transformation in the sense of Jauch,[11] i.e., that, setting $\mathcal{U}=\exp(iF)$ with F self-adjoint, the transformations (2.8) and (2.9) be implementable as

$$\Pi_k \to \exp(iF)\Pi_k\exp(-iF), \tag{2.10}$$

$$J_k \to \exp(iF)J_k\exp(-iF), \tag{2.11}$$

where F is constructed from the algebra of observables. Formally, we postulate

Assumption 3: The algebra of observables is large enough to guarantee that arbitrary local phase transformations with a differentiable $\omega(\mathbf{x})$ are kinematical transformations.

Now, in order to combine (2.8) with (2.10), and (2.9) with (2.11) so as to determine the $[\mathbf{\Pi},F]$ and $[\mathbf{J},F]$ commutators, it is necessary that $[F,[\mathbf{\Pi},F]]=[F,[\mathbf{J},F]]=0$. Furthermore, (2.6) and (2.7) imply $F\psi=\omega\psi$ (with ω a c-number). It is easily seen that these conditions imply that F cannot be expressed as a function of $\mathbf{\Pi}$ and $\mathbf{J}$ alone. Hence, to satisfy Assumption 3, we must enlarge our algebra. To see how, we note that (2.8), (2.10) and (2.9), (2.11), respectively, imply in lowest order

$$i[F,\Pi_k] = -r^{-1}x_4^{-1}[x_ix_k\partial_i\omega - x_ix_i\partial_k\omega + r^2\partial_k\omega], \tag{2.12}$$

$$i[F,J_k] = -\epsilon_{klm}x_l\partial_m\omega. \tag{2.13}$$

Equation (2.12) shows that, if ω is not constant, $[F,\Pi_k]$ must contain at least a c-number term (to produce $r^2\partial_k\omega$) and an operator whose realization in Hilbert space is of the form x_ix_j. Therefore, F must contain a trilinear form of an operator whose realization is x_i. We write

$$\omega(\mathbf{x}) = a_0 + \sum_{n=1}^{\infty}\sum_{i=1}^{3}(a_ix_i)^n \tag{2.14}$$

and in particular, selecting a specific i, we write

$$\omega_i(x) = a_0 + a_ix_i\sum_{n=1}^{\infty}\sum_{k=1}^{3}(a_kx_k)^{n-1}, \tag{2.15}$$

(no summation over i) and denote the generator corresponding to a phase transformation with (2.15) by the symbol F_i. From (2.12) and (2.15) we then obtain

$$\begin{aligned}[F_l,\Pi_k] = ir^{-1}x_4^{-1}\{&a_lx_lx_k\sum_{n,j}n(a_jx_j)^{n-1}\\ &- x_ix_ia_k\sum_{n,j}n(a_jx_j)^{n-1} + r^2a_k\sum_{n,j}n(a_jx_j)^{n-1}\},\end{aligned} \tag{2.16}$$

with no summation over l. To gain insight, let us take the flat space limit $r\to\infty$ and use the notation

$$\lim_{r\to\infty}\Pi_k \equiv \check{\Pi}_k. \tag{2.17}$$

Then (2.8) becomes [in view of (2.1)]

$$\check{\Pi}_k \to \check{\Pi}_k - \partial_k\omega, \tag{2.18}$$

so that $[\check{F}_l, \check{\Pi}_k]$ must be a multiple of the identity operator. Specifically, (2.16) tells us that

$$[\check{F}_l, \check{\Pi}_k] = i\delta_{lk}. \tag{2.19}$$

In a similar manner one finds for the flat space limit

$$[\check{F}_i, \check{J}_k] = i\epsilon_{ikl}\check{F}_l. \tag{2.20}$$

One might think that (2.19) and (2.20) should also hold for the operators prior to taking flat space limits, i.e., that

$$[F_i, \Pi_j] = i\delta_{ij}, \quad [F_i, J_k] = i\epsilon_{ikl}F_l. \tag{2.21}$$

But this is not possible. Comparing powers of x_i in (2.16) and (2.21) we see that

$$a_k = x_4 r^{-2}\delta_{lk}, \tag{2.22}$$

all other coefficients vanishing term by term in x_i. On the other hand, in first order

$$0 = irx_4^{-1}a_k \sum_j 2(a_j x_j),$$

implying $a_k = 0$, which contradicts (2.22). Thus, for a general multilinear series in x_i, lower order terms cannot be determined which would allow (2.21) to hold. In fact, it is not obvious that there exists some unique F with unspecified commutation relations relative to Π_k and J_k which would yield the flat space limit (2.19), (2.20).

However, we do not really face a difficulty here. Since we are concerned with *local* transformations specified by $\omega(\mathbf{x})$, it is entirely consistent to consider *local displacements* only, i.e., displacements that are small compared to r. Therefore, equivalently, the appropriate generators can be taken to be the limits $\check{\Pi}_k$, cf. (2.17). From (2.4) it follows that, naturally, they commute. Their change under an arbitrary local phase transformation is given by (2.18), so that we take over, without change, the statement and proof of Theorem 1 of Ref. 12. Denoting, thus, from now on, $\check{\Pi}_k$ by P_k and introducing $Q_k \equiv MF_k$, we are led to the kinematical group K *which is identical to that of flat space*,

$$K = \mathrm{SU}(2)^J \otimes [T_3^P \otimes (T_3^Q \times T_1^M)]. \tag{2.23}$$

The Lie algebra is

$$[P_i, P_j] = [Q_i, Q_j] = 0, \tag{2.24a}$$

$$[J_k, P_l] = i\epsilon_{klm}P_m, \quad [J_k, Q_l] = i\epsilon_{klm}Q_m, \tag{2.24b}$$

$$[J_k, J_l] = i\epsilon_{klm}J_m; \tag{2.24c}$$

$$[P_k, Q_l] = -iM\delta_{kl}. \tag{2.24d}$$

The realizations are [cf. also (2.5)]

$$P_k \sim i\partial_k, \tag{2.25a}$$

$$Q_k \sim Mx_k, \tag{2.25b}$$

$$J_k \sim -i\epsilon_{klj}x_l\partial_j + \textstyle\sum_k. \tag{2.25c}$$

All remarks (and footnotes) of Ref. 12 on p. 1762 hold also in the present case. In particular, the kinematical group determines mass and spin.

3. THE DYNAMICAL GROUP

In analogy to previous work[12,13] we introduce dynamics by

Definition 1: A development transformation of an isolated system is a kinematical symmetry (in the sense of Jauch[11]) characterized by

$$\mathbf{J} \to \mathbf{J}, \quad \mathbf{P} \to g(\mathbf{Q}, \mathbf{P}, \mathbf{J}), \quad \mathbf{Q} \to f(\mathbf{Q}, \mathbf{P}, \mathbf{J}). \tag{3.1}$$

The motivation here is that the geometry of our space requires that the generator(s) of intrinsic development transformations be invariant under rotations (i.e., under **J**), *but not necessarily invariant under arbitrary large translations* (i.e., under **P**). The difference from the flat world case lies in the behavior of **P**.

Using the same motivation as on p. 1762 of Ref. 12, we next make

Assumption 4: Development transformations form a one-parameter Lie group T_1^H [so that they are represented by $U_\tau = \exp(i\tau H)$].

As in Ref. 12, we demand

Assumption 5: H is contained in the algebra generated by **P**, **Q**, **J**.

Assumptions 4 and 5 together with the invariance requirement implied by Def. 1, determine the form of the development operator,

$$H = H(\mathbf{P}^2, \mathbf{Q}^2, \mathbf{QP}, \mathbf{TP}, \mathbf{TQ}, I). \tag{3.2}$$

Since, as in Ref. 12, the development transformations give rise to an equivalence relation on the algebra of observables generated by K, we again can define a *dynamical group* G by

Assumption 6: The kinematical group K is isomorphic to the quotient group modulo T_1^H of some group G.

Thus, $K \approx G/T_1^H$, i.e., H and the generators of K must form a closed Lie algebra. This restricts the form of H as given by (3.2) to be as follows:

$$H = A\mathbf{P}^2 + B\mathbf{Q}^2 + C(\mathbf{PQ} + \mathbf{QP}) + D, \tag{3.3}$$

where A, B, C, D are as yet arbitrary real constant c-numbers. Using (3.3) we then have the Lie algebra of the dynamical group which consists of (2.24a)–(2.24d) plus the relations

$$[H, P_i] = i2M(BQ_i + CP_i), \tag{3.4a}$$

$$[H, Q_i] = -i2M(CQ_i + AP_i), \tag{3.4b}$$

$$[H, J_i] = 0. \tag{3.4c}$$

In order to fix the constants in (3.3), we must make a further assumption on development transformations.[19] Physical intuition motivates

Assumption 7: The transformation T corresponding to inversion of dynamical development,

$$T: U_\tau \to U_{-\tau}, \tag{3.5}$$

is a kinematical symmetry of the system, i.e., an automorphism of the algebra, realizable in the total state space.

Furthermore, it stands to reason that T be gauge invariant, i.e., that it commutes with arbitrary local phase transformations.[20] Formally, we postulate

Assumption 8: The operator T of development inversion is invariant under local phase transformations,

$$\exp[i\omega(\mathbf{Q})]\,T\exp[-i\omega(\mathbf{Q})]=T. \tag{3.6}$$

From its definition (3.5) and from Assumption 4 we have

$$T\exp(i\tau H)T^{-1}=\exp(-i\tau H) \tag{3.7}$$

which tells us that[21]

$$T(iH)T^{-1}\equiv(iH)'=-iH. \tag{3.8}$$

Transforming (3.4) with T and writing

$$TP_iT^{-1}\equiv P_i',\quad TQ_iT^{-1}\equiv Q_i', \tag{3.9}$$

we get

$$[(iH)',P_i']=-2M(BQ_i'+CP_i'), \tag{3.10a}$$

$$[(iH)',Q_i']=2M(CQ_i'+AP_i'). \tag{3.10b}$$

Taking next in (3.6) the specific case of a linear local phase transformation, $\omega(\mathbf{Q})=M^{-1}c_kQ_k$, we see that $[Q_k,T]=0$, i.e., $Q_k'=Q_k$. Then, with (3.8), Eqs. (3.10) become

$$[H,P_i']=-2iM(BQ_i+CP_i'), \tag{3.11a}$$

$$[H,Q_i]=2iM(CQ_i+AP_i'). \tag{3.11b}$$

Equation (3.11b) is compatible with (3.4b) only if $C=0$ and $P_i'=-P_i$. Then (3.11a) is, without further assumptions, also compatible with (3.4a). Thus, we must take $C=0$ in (3.3), and since H is determined only up to an over-all multiplicative constant, we set, for convenience,

$$A=(2M)^{-1},\quad B=(2M)^{-1}\nu^2,\quad D=\mathcal{C}_1.$$

Here ν is simply a constant determining a scale of units whose significance will become evident later. Thus, the final and *unique* form of H, as determined by our assumptions, is

$$H=(2M)^{-1}\mathbf{P}^2+(2M)^{-1}\nu^2\mathbf{Q}^2+\mathcal{C}_1. \tag{3.12}$$

The Lie brackets (3.4) become

$$[H,P_i]=i\nu^2Q_i, \tag{3.13a}$$

$$[H,Q_i]=-iP_i, \tag{3.13b}$$

$$[H,J_i]=0. \tag{3.13c}$$

We now see that *the algebra of the dynamical group G, given by* (2.24) *and* (3.13), *is precisely the Lie algebra* (A4) *of the centrally extended Hooke group* which has the structure[22]

$$G\equiv\tilde{\mathcal{B}}_4=T_1^H\otimes K=T_1^H\otimes\{(\mathrm{SU}(2)^J\otimes T_3^Q)\otimes(T_3^P\times T_1^M)\}. \tag{3.14}$$

We derived it essentially from the locality principle in a curved space.

If we write $G=\mathcal{B}_4\otimes T_1^M$ and, *for convenience*, decide to represent $\mathcal{B}_4$ on the left coset space $\mathcal{B}_4/\mathrm{SO}(3)^J\otimes T_3^Q$, we easily find with the composition law (A1), and upon identifying the elements of the coset space $(\bar\tau,\bar{\mathbf{a}})$ with the points $(t,\mathbf{x})$ of the space $E_1(t)\times\mathcal{S}_3(\mathbf{x})$, the transformation law of this space as given by Eq. (A3). Thus, the active viewpoint of our abstract dynamical group is to consider it a set of endomorphisms of $E_1(t)\times\mathcal{S}_3(\mathbf{x})$. The importance of this is that it permits us to interpret *"nonrelativistic cosmologic time"* in a purely group theoretic manner. This time variable was not introduced from the outset, but rather arose simply as a convenience, permitting an active characterization of the dynamical group.

We also see from (A3) and the identification of $(\bar\tau,\bar{\mathbf{a}})$ with $(t,\mathbf{x})$ that the constant ν introduced in the course of fixing the constants in H has the dimension of reciprocal nonrelativistic cosmological time, and (3.12) tells us that it is the circular frequency of inertial motion in this world.

We return to the discussion of development reversal T. Because of the identification of the coset space $\mathcal{B}_4/\mathrm{SO}(3)^J\otimes T_3^Q$ with $E_1(t)\times\mathcal{S}_3(\mathbf{x})$, the operation T clearly means *cosmologic time reversal*. In the course of the application of Assumptions 7 and 8 we found that

$$Q_k'=TQ_kT^{-1}=Q_k, \tag{3.15a}$$

$$P_k'=TP_kT^{-1}=-P_k. \tag{3.15b}$$

Further, taking the T transform of (2.24d), we have $[P_k',Q_i']=-[P_i,Q_i]=-T(iM)T^{-1}$. Consistency[23] with (2.24d) demands that T be *antilinear*. Therefore, Eq. (3.8) gives

$$H'=THT^{-1}=H. \tag{3.15c}$$

In summary, T has the same properties as the familiar time reversal operator of the flat space theory.[24]

We now discuss a further consequence of using the above described homogeneous representation space. As in the Galilean case,[12] we are led to define, for each t, a Hilbert space $\mathcal{H}_t$ of square integrable (on S) functions by setting

$$\psi(\mathbf{x};t)=\exp(-itH)\psi(\mathbf{x}), \tag{3.16}$$

and the total Hilbert space is $\mathcal{H}=\oplus\mathcal{H}_t$. On a particular "slice" the realization of the basic observables is easily seen to be

$$P_k\sim-i\cos(\nu t)\partial_k-M\nu x_k\sin(\nu t), \tag{3.17a}$$

$$Q_k\sim Mx_k\cos(\nu t)-i\nu^{-1}\sin(\nu t)\partial_k, \tag{3.17b}$$

$$J_k\sim-i\epsilon_{klj}x_l\partial_j+\Sigma_k, \tag{3.17c}$$

$$H\sim i\partial_t. \tag{3.17d}$$

In particular, H assumed a double role: on each slice, apart from (3.17d), it also has the realization

$$H\sim-\{(2M)^{-1}\partial_k\partial_k+\tfrac{1}{2}(M\nu^2)x_kx_k\}+\mathcal{C}_1, \tag{3.18}$$

which follows easily from inserting (3.17a) and (3.17b) into (3.12). Since the $\mathcal{C}_1$ in (3.12) is precisely the Casimir invariant given in (A5b), it can be taken (up to ray equivalence) to be zero, so that (3.18) and (3.17d) give the familiar[8] Hooke–Schrödinger equation,

$$i\partial_t\psi(\mathbf{x};t)=[(2M)^{-1}\partial_k\partial_k+\tfrac{1}{2}M\nu^2x_kx_k]\psi(\mathbf{x};t). \tag{3.19}$$

In our framework it emerged from the fact that we have selected the "homogeneous Hooke group" $SO(3)^J\otimes T_3^Q$ as the subgroup which defines a homogeneous G space.

4. INTERACTING PARTICLES

The transformation property of the basic observables when a local phase transformation is performed[25] can be found from (3.17) and we get

$$P_k\to P_k-\cos(\nu t)\partial_k\omega, \tag{4.1a}$$

$$Q_k\to Q_k+\nu^{-1}\sin(\nu t)\partial_k\omega, \tag{4.1b}$$

$$J_k\to J_k-\epsilon_{klm}x_l\partial_m\omega, \tag{4.1c}$$

$$H\to H. \tag{4.1d}$$

In particular we see that, except on the slice $t=0$, the position operator is not invariant under local phase transformations. Since, for physical reasons, we do not find it acceptable that localization be dependent on the gauge, we stipulate, as in Ref. 12,

Assumption 9: Localization does not depend on the choice of a phase $\omega(\mathbf{x})$.

In order to satisfy the requirement that $\mathbf{Q}\to\mathbf{Q}$ under arbitrary local phase transformations, we must modify our system, by introducing essentially extraneous degrees of freedom, i.e., by coupling it to some system in a suitable way. Systems for which

(a) Assumption 9 holds, i.e., for which $\mathbf{Q}\to\mathbf{Q}$ under a local phase transformation,

(b) H is invariant under a local phase transformation, i.e., $H\to H$,

(c) H is independent of t,

we shall call *covariantly interacting systems*. Such systems are *uniquely characterized* by the following theorem.

Theorem: The Hamiltonian of a covariantly interacting (spinless) system is given by

$$H=(\tfrac{1}{2}M)(\mathbf{P}-\mathbf{A})^2+(2M)^{-1}\nu^2\mathbf{Q}^2+V, \tag{4.2}$$

where $\mathbf{A}$ depends on $\mathbf{x}$ and t, V depends on $\mathbf{x}$. Further, $\mathbf{A}$ has the realization

$$A_k\sim\mathring{A}_k(\mathbf{x})\cos\nu t, \tag{4.3}$$

and under a local phase transformation

$$A_k\to A_k-\partial_k\omega. \tag{4.4}$$

V does not change under a local phase transformation.

Proof: To satisfy requirement (a) of a covariantly interacting system, we must modify the realization of $\mathbf{Q}$. Since in the absence of interaction we must recover (4.1b), and since on the slice $t=0$ we must recover (2.25b), we are led to set

$$Q_k\sim Mx_k\cos(\nu t)-i\nu^{-1}\sin(\nu t)\partial_k-\nu^{-1}\mathring{A}_k\sin(\nu t). \tag{4.5}$$

Calculating the action of a local phase transformation we find

$$\begin{aligned}(UQ_kU^{-1}\psi)(\mathbf{x};t)&=\exp(i\omega)\,[(Mx_k\cos(\nu t)-i\nu^{-1}\sin(\nu t)\,\partial_k]\\&\quad\times\exp(-i\omega)\psi(\mathbf{x};t)-\nu^{-1}\sin(\nu t)U\mathring{A}_kU^{-1}\psi(\mathbf{x};t)\\&=[Mx_k\cos(\nu t)-i\nu^{-1}\sin(\nu t)\partial_k\\&\quad-\nu^{-1}\sin(\nu t)\partial_k\omega-\nu^{-1}\sin(\nu t)U\mathring{A}_kU^{-1}]\\&\quad\times\psi(\mathbf{x};t).\end{aligned}$$

Thus, $Q_k\to Q_k$ provided $U\mathring{A}_kU^{-1}=\mathring{A}_k-\partial_k\omega$. In other words, requirement (a) is satisfied if (4.4) holds. To show that, if (4.3) holds, then H has the form (4.2) and that then requirements (b) and (c) are also satisfied, we first compute the commutation relations for the interacting system. The realization of the operator H is, by its basic meaning, given by (3.17d). The realization of $\mathbf{P}$ and $\mathbf{Q}$ is given by (3.17a) and (4.5), respectively. The realization of $\mathbf{A}$ is stipulated in (4.3). Using these expressions, we find[26]

$$[Q_i,\,P_j-A_j]=iM\delta_{ij}, \tag{4.6a}$$

$$[H,\,Q_k]=-i(P_k-A_k), \tag{4.6b}$$

$$[H,\,P_k-A_k]=i\nu^2Q_k. \tag{4.6c}$$

Thus we see that

$$\mathcal{P}_k\equiv P_k-A_k \tag{4.7}$$

is the momentum canonically conjugate to Q_k, and it then follows that for any slice t we can write

$$H=(2M)^{-1}\mathcal{P}^2+(2M)^{-1}\nu^2\mathbf{Q}^2+V \tag{4.8}$$

which is precisely (4.2). Now we observe that under a local phase transformation our $Q_k\to Q_k$ (as proved above). Further, since the corresponding transformation of $\mathbf{P}$ is still given by (4.1a), whereas that of $\mathbf{A}$ is defined by (4.4), we have $\mathcal{P}\to\mathcal{P}$. It then follows from (4.8) that $H\to H$, i.e., requirement (b) is fulfilled. Finally, substituting the realizations (3.17a) and (4.5) into the rhs of (4.8), we see that (provided V is independent of t), H is explicitly time independent, so that requirement (c) is satisfied. This concludes the proof.

As in Ref. 12, we may decide to make explicit the superselection rule which was implicitly introduced in Sec. 3 by the choice of the homogeneous space that led to a sequence $\mathcal{H}_t$ of incoherent Hilbert spaces. Without change, we can take over Assumption 8 of Ref. 12 concerning time-dependent gauge transformations ($\omega=\omega(\mathbf{x},t)$). This again leads to the transformation law $H\to H+\partial_t\omega$, and to achieve consistency with the realization (4.2), we must again require that the heretofore arbitrary V transforms according to the law

$$V\to V+\partial_t\omega. \tag{4.9}$$

As for flat space, now it is also possible to simplify the description of the system by performing a particular gauge transformation with

$$\omega(\mathbf{x},t)=-\int_0^t V\,dt,$$

with the result that, while $\mathcal{P}$ and $\mathbf{Q}$ remain unchanged, $H\to H-V$, so that *in this particular gauge*

$$H=(2M)^{-1}\mathcal{P}^2+(2M)^{-1}\nu^2\mathbf{Q}^2, \tag{4.10}$$

and arbitrary gauge transformations are invariance transformations.

Finally, we point out that all considerations of Ref.

12 on p. 1765 concerning superselection rules, carry over unchanged for the curved space theory.

5. CONCLUDING REMARKS

The major result of this study is that, exactly as in a flat space, the locality principle, supplemented by a few rather simple intuitive requirements, has the power to determine a quantum dynamical group in a uniformly curved space background. The major difference from the flat space case is that, since dynamical development need not be invariant under arbitrary large spatial translations, **P** has a more general development transformation. Furthermore, for obtaining a unique development law, it is necessary to *assume* the existence of an additional gauge invariant kinematical symmetry, which later could be interpreted as cosmological time reversal. The dynamical group so obtained is identical with the (abstract) Hooke group, as expected. The additional requirement that localization be gauge independent, leads to the necessity of an interaction and for a unique interaction structure which has the familiar minimal coupling form.

APPENDIX A: THE HOOKE GROUP

We review here that Hooke group $\mathcal{B}_4$ which, in the framework of Bacry and Lévy-Leblond[4] arises from the contraction of SO(3,2).[27]

Denoting the parameters corresponding to the generators H, **P**, **Q**, **J** by τ, **a**, **v**, R, respectively, the *composition law* is

$$(\tau, \mathbf{a}, \mathbf{v}, R)(\bar{\tau}, \bar{\mathbf{a}}, \bar{\mathbf{v}}, \bar{R}) = (\tau+\bar{\tau}, \mathbf{a}\cos\nu\bar{\tau} + \nu^{-1}\mathbf{v}\sin\nu\bar{\tau} + R\bar{\mathbf{a}}, \mathbf{v}\cos\nu\bar{\tau} - \nu\mathbf{a}\sin\nu\bar{\tau} + R\bar{\mathbf{v}}, R\bar{R}), \tag{A1}$$

where ν is an arbitrary dimensionate constant that may be chosen as unit.[28]

The *inverse element* is

$$(\tau, \mathbf{a}, \mathbf{v}, R)^{-1} = (-\tau, \nu^{-1}R^{-1}\mathbf{v}\sin\nu\tau - R^{-1}\mathbf{a}\cos\nu\tau, -\nu R^{-1}\mathbf{a}\sin\nu\tau - R^{-1}\mathbf{v}\cos\nu\tau, R^{-1}). \tag{A2}$$

It is possible to view the group $\mathcal{B}_4$ as a transformation group on a space–time manifold $(\mathbf{x}, t)$, where the *transformation law* is

$$t \to t+\tau, \quad \mathbf{x} \to R\mathbf{x} + \mathbf{a}\cos\nu t + \nu^{-1}\mathbf{v}\sin\nu t. \tag{A3}$$

From this it follows that the *inertial motion* of a particle initially at rest at $\mathbf{x}=0$ is given by $\mathbf{x} = \nu^{-1}\mathbf{v}\sin\nu t$, i.e., it is an oscillatory motion with (circular) period ν. From the viewpoint of a cosmological model, ν^{-1} may be thought of as the "lifetime" of the universe (cf. Ref. 6).

The *Lie algebra of the central extension*[29] $\tilde{\mathcal{B}}_4$ of $\mathcal{B}_4$ is found to be

$$[P_i, P_k] = 0, \quad [Q_i, Q_k] = 0, \tag{A4a}$$

$$[J_k, J_l] = i\epsilon_{klj}J_j, \tag{A4b}$$

$$[J_k, P_l] = i\epsilon_{klj}P_j, \quad [J_k, Q_l] = i\epsilon_{klj}Q_j, \tag{A4c}$$

$$[P_k, Q_l] = -iM\delta_{kl}, \tag{A4d}$$

$$[H, J_k] = 0, \tag{A4e}$$

$$[H, Q_k] = -iP_k, \quad [H, P_k] = i\nu^2 Q_k. \tag{A4f}$$

In (A4d) M is an arbitrary constant associated with the central extension, giving rise to a superselection rule, which, as in the Galilean case, can be interpreted as mass.

The *Casimir invariants* of $\tilde{\mathcal{B}}_4$ are

$$\mathcal{C}_0 = MI, \tag{A5a}$$

$$\mathcal{C}_1 = H - \mathbf{P}^2/2M - (\nu^2/2M)\mathbf{Q}^2, \tag{A5b}$$

$$\mathcal{C}_2 = \mathbf{T}^2, \tag{A5c}$$

where

$$\mathbf{T} \equiv \mathbf{J} - M^{-1}\mathbf{Q}\times\mathbf{P} = \Sigma \tag{A5d}$$

is the spin. Since representations with different $\mathcal{C}_1$ are ray-equivalent, $\mathcal{C}_1$ can be chosen to be zero, so that

$$H = \mathbf{P}^2/2M + (\nu^2/2M)\mathbf{Q}^2, \tag{A5e}$$

i.e., there is an effective potential of harmonic force type, in agreement with the previous statement on inertial motion.

In the Introduction we emphasized the role of $\mathcal{B}_4$ as a nonrelativistic cosmological symmetry group.[30] However, we believe that $\mathcal{B}_4$ may play an important role in particle physics, too. The idea that the internal space–time manifold corresponding to an elementary particle is a curved space (specifically, a de Sitter space) has been put forward many times. Therefore, $\tilde{\mathcal{B}}_4$ describing an approximation of quantum dynamics for such a world, may serve as a tool for understanding elementary particles.[31] These possibilities will be explored at a later time.

APPENDIX B: REALIZATION OF EQ. (2.4) IN $L^2(S)$

Rosen[32] has shown that, in the limit of the vanishing of the extra metric component, the Lie algebra of ISO(p,q) is equivalent to that of SO$(p+1, q)$ or SO$(p, q+1)$. The algebra of these homogeneous groups is[33]

$$[N_{\alpha\beta}, N_{\gamma\delta}] = i(g_{\alpha\gamma}N_{\beta\delta} + g_{\beta\delta}N_{\alpha\gamma} - g_{\alpha\delta}N_{\beta\gamma} - g_{\beta\gamma}N_{\alpha\delta}),$$
$$\alpha, \beta, \gamma, \delta = 0, 1, \ldots, n, \ (p+q=n). \tag{B1}$$

Using indices μ, ν, ρ, σ for $0, 1, \ldots, n-1$, we can rewrite this as

$$[N_{\mu\nu}, N_{\rho\sigma}] = i(g_{\mu\rho}N_{\nu\sigma} + g_{\nu\sigma}N_{\mu\rho} - g_{\mu\sigma}N_{\nu\rho} - g_{\nu\rho}N_{\mu\sigma}), \tag{B2a}$$

$$[N_{\mu\nu}, N_{\rho n}] = i(g_{\mu\rho}N_{\nu n} - g_{\nu\rho}N_{\mu n}), \tag{B2b}$$

$$[N_{\mu n}, N_{\nu n}] = \pm iaN_{\mu\nu}, \tag{B2c}$$

where the sign of $a \equiv g_{nn}$ is $-$ or $+$ depending on whether we take SO$(p+1, q)$ or SO$(p, q+1)$. Clearly, the $N_{\mu\nu}$ obey the SO(p, q) algebra and the $N_{\mu n}$ are vectors for this group, which in the limit $a \to 0$ vanish. The parameter $|a|^{1/2}$ can be considered as the reciprocal of the radius of curvature.[34]

If y_α denote the homogeneous coordinates of an $n+1$ dimensional linear space, then the Lie algebra (B1) can be realized by differential operators acting on the space of square integrable functions $f(y)$ as follows:

$$N_{\alpha\beta} \sim -i(y_\alpha \partial_\beta - y_\beta \partial_\alpha). \tag{B3}$$

Thus,

$$N_{\mu\nu} \sim i(y_\mu \partial_\nu - y_\nu \partial_\mu), \tag{B3a}$$

$$N_{\mu n} \sim -i(y_\mu \partial_n - y_n \partial_\mu). \tag{B3b}$$

To relate the homogeneous coordinates to the inhomogeneous coordinates for n dimensions, we use the equation of the pseudosphere,[35]

$$\sum_{\mu,\nu=0}^{n-1} g_{\mu\nu} y^\mu y^\nu = r^2 - y_n^2, \tag{B4}$$

from which it follows that

$$\partial_n = -y_n^{-1} \sum_{\mu,\nu=0}^{n-1} g_{\mu\nu} y^\mu \partial^\nu. \tag{B5}$$

Introducing the notation

$$\Pi_\mu \equiv r^{-1} N_{\mu n}, \tag{B6}$$

using (B3b), (B5), and expressing y_n by (B4), we obtain

$$\Pi_\mu \sim -ir^{-1}(r^2 - \Sigma g_{\mu\nu} y^\mu y^\nu)^{-1/2} \times [-y_\mu \Sigma g_{\nu\rho} y^\nu \partial^\rho + \Sigma g_{\nu\rho} y^\nu y^\rho \partial_\mu - r^2 \partial_\mu]. \tag{B7}$$

Equation (2.5a) is a special case; Eq. (2.5b) is obvious from (B3a).

*Permanent address: Department of Physics, Boston University, Boston, Mass. 02215.

[1]For an authoritative review of the Galilei group see J.M. Lévy-Leblond, in *Group Theory and Its Applications*, edited by E.M. Loebl (Academic, New York, 1971), Vol. II.

[2]Our notation for generators is as follows: de Sitter group: J_{ab} ($a,b = 0, 1, 2, 3, 4$); Poincaré group: M_{kl} ($k, l = 1, 2, 3$), N_{k0} ($k = 1, 2, 3$), P_μ ($\mu = 0, 1, 2, 3$); Galilei group: $\hat{\mathbf{J}}$, $\hat{\mathbf{Q}}$, $\hat{\mathbf{P}}$, $\hat{\mathbf{H}}$; Hooke group: **J**, **Q**, **P**, **H**.

[3]These statements follow from the fact that the limits in (1.1) and (1.2) imply that the corresponding parameters are small. Thus, for example, (1.1a) says that the parameters associated with $\hat{\mathbf{Q}}$ (i.e., velocities) and the parameters associated with $\hat{\mathbf{P}}$ (i.e., spatial translations distances) are made small by the contraction process. This justifies the use of the terms "speed-space" (or "space—time") contractions, cf. Ref. 4.

[4]H. Bacry and J.-M. Lévy-Leblond, J. Math. Phys. **9**, 1605 (1968).

[5]For convenience, a review of the "Hooke group" is given in Appendix A.

[6]J.R. Derome and J.G. Dubois, Nuovo Cimento B **9**, 351 (1972).

[7]J.G. Dubois, Nuovo Cimento B **15**, 1 (1973).

[8]J.G. Dubois, Can. J. Phys. **51**, 1757 (1973).

[9]No "time contraction" has been performed.

[10]Just as the Poincaré group is a "space—time" contraction of the de Sitter group.

[11]J.M. Jauch, Helv. Phys. Acta **37**, 284 (1964).

[12]P. Roman and J.P. Leveille, J. Math. Phys. **15**, 1760 (1974).

[13]P. Roman, in *Quantum Theory and the Structures of Time and Space*, edited by L. Castell, M. Drieschner, and C.F. von Weizsäcker (Carl Hauser, Munich, 1975), pp. 85—102.

[14]Here and in the following $\mathbf{x}^2$ denotes $x_1^2 + x_1^2 + x_3^2$.

[15]Assuming negative curvature would not essentially affect our results: instead of the Hooke group corresponding to an oscillating and "closed" universe, we would get the Hooke group for an expanding and "open" universe.

[16]A straightforward derivation of Eq. (2.5) is given in Appendix B.

[17]For details of this and subsequent arguments and their motivation, cf. the analogous discussion in Refs. 12, 13.

[18]Summation over repeated indices understood unless otherwise stated.

[19]This is, interestingly, in contrast to the Galilean case where no (essential) constant occurred in the determination of H (cf. Eq. (3.2) of Ref. 12).

[20]This is motivated by the fact that, as it will transpire later, T corresponds physically to time reversal, i.e., is a space-time symmetry, whereas gauge transformations are "internal" symmetries, giving rise to "charges".

[21]We do not assume that T is a *linear* automorphism.

[22]There are other, isomorphic ways to write the structure. We chose the one given in Ref. 6.

[23]Note that M is a real c-number.

[24]It is interesting to note that, because of Assumption 7, the dynamics described by the resulting Hooke group is time reversal invariant. If the global implementability of T is dropped (spontaneous symmetry breaking), the resulting dynamical group will be more general. It seems worthwhile to explore this possibility.

[25]Simultaneously on all slices, i.e., $\psi(\mathbf{x};t) \to \exp[i\omega(\mathbf{x})]\psi(\mathbf{x};t)$.

[26]Commutators involving **J** are as for the noninteracting system and are not relevant at present.

[27]The other (not "oscillatory" but "expanding") type arises from contracting $\mathbf{SO}(4,1)$. The formulas given below remain valid for the latter if one replaces sin and cos by sinh and cosh, and formally changes the sign of ν^2 everywhere.

[28]Letting $\nu \to 0$ corresponds to contracting $\mathcal{B}_4$ to the Galilei group $\mathcal{G}_4$.

[29]For details of the central extension cf. Ref. 7. The Lie algebra of $\mathcal{B}_4$ differs from that of $\widetilde{\mathcal{B}}_4$ only inasmuch as the rhs of (A4d) is zero.

[30]Some interesting consequences of this viewpoint are expressed in Refs. 4, 6, and 7.

[31]If one assumes that SO(4,2) (rather than SO(3,2)) corresponds to the submicroscopic world, as has been suggested, for example, by L. Castell, Nuovo Cimento A **49**, 285 (1967); Nuclear Phys. B **4**, 343 (1967) etc., then, by the contraction procedure one arrives at a relativistic generalization $\widetilde{\mathcal{B}}_5$ of the Hooke group [Castell's group II; see also P.L. Huddleston, M. Lorente, and P. Roman, Found. of Phys. **5**, 75 (1975)]. This may then give an even more interesting handle for exploring particle physics.

[32]J. Rosen, Nuovo Cimento **35**, 1234 (1966).

[33]We use the metric $-g_{00} = \cdots = -g_{qq} = +g_{q+1,q+1} = \cdots = g_{nn} = +1$.

[34]The curvature tensor of a uniformly curved Riemann space is $R_{\mu\nu\rho\sigma} = a(g_{\mu\nu}g_{\rho\sigma} - g_{\mu\sigma}g_{\nu\rho})$.

[35]Note that $g_{n\beta} = \delta_{n\beta}$.

Canonical parameters of the 3j coefficient*

Loren A. Lockwood

Departamento de Física, Universidad de Oriente, Cumaná, Venezuela
(Received 5 April 1976)

The 3j coefficient is expressed as a function of five new parameters which have unique properties. They are completely independent, satisfy simple validity criteria, and display the symmetry properties of the function in a particularly transparent manner. By means of the new parameters, the known 72-element symmetry group is reduced to an eight-element group, and the absolute symmetries are separated in a clear way from those which contain a phase factor.

Wigner's 3j coefficient[1] may be considered to be defined by the equation[2]

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \equiv (3\mathrm{J}) = \delta(m_1+m_2+m_3)(-1)^{j_1-j_2-m_3}$$
$$\times[(j_1+j_2-j_3)!\,(j_2+j_3-j_1)!\,(j_3+j_1-j_2)!/$$
$$\times(j_1+j_2+j_3+1)!\,]^{1/2}$$
$$\times[(j_1+m_1)!\,(j_1-m_1)!\,(j_2+m_2)!\,(j_2-m_2)!$$
$$\times(j_3+m_3)!\,(j_3-m_3)!\,]^{1/2}$$
$$\times\sum_t(-1)^t[(j_1+j_2-j_3-t)!\,(j_1-m_1-t)!$$
$$\times(j_2+m_2-t)!\,(t+j_3-j_2+m_1)!$$
$$\times(t+j_3-j_1-m_2)!\,t!\,]^{-1}, \tag{1}$$

using the j and m quantum numbers of three angular momenta as parameters. The summation index t assumes all values for which none of the factorials becomes undefined.

Regge[3] has shown that the 3j coefficient possesses a 72-element symmetry group. Only twelve elements, those which involve permutations of the angular momenta and space reflection, are simply represented in terms of the j's and m's. The others require replacing certain j's and m's by algebraic expressions involving the original ones.

In obtaining his result, Regge introduced a square symbol, expressing the 3j coefficient as a function of nine parameters, bound by four equations. The symmetry operations are found to involve simultaneous permutations among at least six of Regge's parameters.

In the following, we shall express the 3j coefficient as a function of five new parameters, which have properties not possessed by the quantum numbers of the angular momenta nor by Regge's parameters.

To begin, m_3 is replaced by $-m_1-m_2$ and five intermediate parameters are defined as follows: $k_1=j_2+m_2$, $k_2=j_1-m_1$, $k_3=j_1+j_2-j_3$, $k_4=j_1-j_3+m_2$, $k_5=j_2-j_3-m_1$. When substituted into Eq. (1), the parameters k_1, k_2, and k_3 are found to occur in a symmetrical way.[4] The same is true of k_4, k_5, and the constant zero. To take advantage of this fact, the triplet (k_1,k_2,k_3) is placed in ascending order and the ordered elements are named (p,q,r); i.e., $p\leqslant q\leqslant r$. Similarly, the triplet $(k_4,k_5,0)$ is ordered and named (f,g,h), with $f\leqslant g\leqslant h$.[5] Finally we define $n=p-h$, $a=h-g$, $b=h-f$, $c=q-p$, $d=r-p$, and let the summation index t in Eq. (1) be replaced by $s=t-h$. Then Eq. (1) becomes

$$(3\mathrm{J})=PRT, \tag{2}$$

where

$$P=(-1)^{h+k_4-k_5},$$

$$R=[n!\,(n+a)!\,(n+b)!\,(n+c)!\,(n+d)!\,(n+a+c)!$$
$$\times(n+a+d)!\,(n+b+c)!\,(n+b+d)!/$$
$$(3n+a+b+c+d+1)!\,]^{1/2}, \tag{3}$$

and

$$T=\sum_{s=0}^{n}\frac{(-1)^s}{s!\,(s+a)!\,(s+b)!\,(n-s)!\,(n+c-s)!\,(n+d-s)!}. \tag{4}$$

There are six ways that (f,g,h) may correspond to $(k_4,k_5,0)$, and in all six cases, the phase factor P may be shown to be equal to

$$P=(-1)^{a+b}. \tag{5}$$

We argue that the parameters n, a, b, c, d may be called canonical, on the basis of the following properties.

Firstly, they are independent. The delta function in Eq. (1) implies that there are only five independent parameters there also, but the "independence" of the j's and m's is only partial. The j's must still satisfy the triangular relationship, and the usual restrictions obtain on the absolute value of each m and on the way that integral and half-integral values may combine. Regge's parameters are likewise mutually restricted. The value of any one of the new parameters, however, in no way restricts the values of the others. All five are required simply to be integral and nonnegative. This condition completely satisfies all the restrictions on the j's and m's, as well as those relating to Regge's parameters.

Secondly, the symmetry properties of the 3j coefficient are especially transparent when it is expressed by Eqs. (2)–(5), as will now be shown.

Inasmuch as each of P, R, and T is invariant under the interchange of a and b or of c and d, the following four-element symmetry group is evident:

$$(3J)(n;\ a,b;\ c,d) = (3J)(n;\ b,a;\ c,d)$$
$$= (3J)(n;\ b,a;\ d,c)$$
$$= (3J)(n;\ a,b;\ d,c). \qquad (6)$$

There are 36 ways that the two triplets (p,q,r) and (f,g,h) can simultaneously correspond to (k_1,k_2,k_3) and $(k_4,k_5,0)$, each being, effectively, a mapping from the j's and m's onto n, a, b, c, d. Of the 36, nine may be chosen so that the remaining 27 are simply interchanges of a and b or of c and d, or both. Thus the four symmetries of Eq. (6) represent all 36 of the absolute symmetries of the $3j$ coefficient, a nine-to-one homomorphism.

All the remaining 36 symmetries contain a phase factor, which, in terms of the j's, is equal to $(-1)^{j_1+j_2+j_3}$. In terms of the new parameters, these symmetries correspond to the simultaneous interchange of a and c and of b and d. It can be seen that this operation leaves R unchanged, multiplies T by $(-1)^n$ (by converting s into $n-s$), and multiplies P by $(-1)^{c+d-a-b}$. Thus the symmetries of Eq. (6) may be augmented by the following:

$$(3J)(n;c,d;a,b) = (3J)(n;d,c;a,b)$$
$$= (3J)(n;d,c;b,a)$$
$$= (3J)(n;c,d;b,a)$$
$$= (-1)^{n+a+b+c+d}(3J)(n;a,b;c,d). \qquad (7)$$

This implies that any $3j$ coefficient with $a=c$ and $b=d$ (or $a=d$ and $b=c$) and n odd must vanish. Since $j_1+j_2+j_3=3n+a+b+c+d$, this statement embodies all known symmetry arguments for the vanishing of the $3j$ coefficient.

Equations (6) and (7) together define an eight-element symmetry group which is equivalent to Regge's 72-element group through the previously noted nine-to-one homomorphism. Interestingly, the symmetry operations involve permutations among only four of the new parameters, and there is a clear separation of the absolute and the phase-conditioned symmetries.

Work is in progress to determine whether the existence of these canonical parameters for the $3j$ coefficient may shed additional light on the symmetry properties of the $6j$ and $9j$ coefficients.

*Supported in part by Consejo de Investigación, Universidad de Oriente.

[1]E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959), p. 290.

[2]This equation, originally due to Racah, appears in various forms in a large number of works, e.g., A. de-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963), p. 138. From the physical viewpoint, it is not usually taken to define the $3j$ coefficient, but, mathematically, it may be so used.

[3]T. Regge, Nuovo Cimento **10**, 544 (1958).

[4]In fact, the full 72-element symmetry group may be expressed as the direct product of the six-element permutation group of (k_1,k_2,k_3) with the 12-element group of physical symmetries, i.e., permutations of the angular momenta and space reflection.

[5]All that is required, actually, is that p be the smallest of the first triplet and that h be the largest of the second.

Quantum two-particle scattering in fuzzy phase space*

Eduard Prugovečki

Department of Mathematics, University of Toronto, Toronto, Canada M5S 1A1
(Received 21 January 1976)

The concepts of configuration and momentum representation space for state vectors are generalized to that of fuzzy-phase-space representation spaces $L^2(\Gamma_s)$, $0 < s < \infty$, which are interpolated in between these two standard representations. It is shown that the wavepacket in $L^2(\Gamma_s)$ displays the familiar evanescence property from any region $K_s \times M_s$ in the fuzzy phase space Γ_s if that region is bounded in its configuration part K_s; also, that the probability of detecting the system in $K_s \times M_s$ has a finite asymptotic time limit if K_s is a (fuzzy) cone. For scattering states the existence of free states that are asymptotic in Γ_s is established, and a formula for differential cross section in Γ_s is derived.

1. INTRODUCTION

It has been pointed out recently[1] that in quantum mechanics one can assign probability densities to a simultaneous measurement of position $\mathbf{Q}$ and momentum $\mathbf{P}$ of a particle as long as one recognizes the fact that no such measurement can pinpoint the determined values $\mathbf{q}$ and $\mathbf{p}$ with arbitrary accuracy, since one is limited by Heisenberg's uncertainty relations. Thus, the outcome of such a measurement cannot be described exclusively in terms of the values $(\mathbf{q},\mathbf{p}) \in \mathbb{R}^6$; instead, such a description has to be supplemented by assigning to each $(\mathbf{q},\mathbf{p})$ a *confidence function* $\chi_{\mathbf{q},\mathbf{p}}(\mathbf{x},\mathbf{k})$, to which a straightforward operational meaning can be assigned (cf. Appendix) in terms of the *accuracy calibration* of the instrument used in the measurement of $\mathbf{Q}$ and $\mathbf{P}$: namely when $\chi_{\mathbf{q},\mathbf{p}}$ is normalized to unity, the values

$$\nu_{\mathbf{q},\mathbf{p}}(I_1 \times I_2) = \int_{I_1} d\mathbf{x} \int_{I_2} d\mathbf{k}\, \chi_{\mathbf{q},\mathbf{p}}(\mathbf{x},\mathbf{k}) \tag{1.1}$$

express our confidence that when the reading $(\mathbf{q},\mathbf{p})$ is obtained the *actual* values of $\mathbf{Q}$ and $\mathbf{P}$ are within the intervals I_1 and I_2, respectively. In other words, each such measurement supplies a fuzzy sample point that represents the simultaneous values of $\mathbf{Q}$ and $\mathbf{P}$. Consequently, by generalizing[1] the mathematical framework of probability theory to the case when the sample points are fuzzy, we have managed to relate the description of the statistics of such measurements (carried out on a sample of systems in one and the same quantum mechanical state) to the concept of probability measures on fuzzy events in phase space.

As with the case of conventional probability theory, in building a probability space over fuzzy events, the starting point lies in the specification of the space $\mathcal{S}$ of sample points. For the measurement of $\mathbf{Q}$ and $\mathbf{P}$ of a quantum-mechanical particle we have taken[1] $\mathcal{S}$ to consist of all fuzzy points $(\mathbf{q},\mathbf{p},\chi_{\mathbf{q},\mathbf{p}})$ specified in terms of some $(\mathbf{q},\mathbf{p}) \in \mathbb{R}^6$ and a confidence function $\chi_{\mathbf{q},\mathbf{p}}$ with maximum at $(\mathbf{q},\mathbf{p})$ and having the form

$$\chi_{\mathbf{q},\mathbf{p}}(\mathbf{x},\mathbf{k}) = \int_{\mathbb{R}^9} \chi_{\mathbf{q}'}^{(\mathbf{s})}(\mathbf{x})\chi_{\mathbf{p}'}^{(\mathbf{s})}(\mathbf{k})\, d\mu(\mathbf{q}',\mathbf{p}',\mathbf{s}'); \tag{1.2}$$

here μ is any normalized measure on $\mathbb{R}^9$ and

$$\begin{aligned}\chi_{\mathbf{q}}^{(\mathbf{s})}(\mathbf{x}) &= \pi^{-3/2}\prod_{\alpha=1}^{3} s_\alpha^{-1}\exp[-s_\alpha^{-2}(x_\alpha - q_\alpha)^2],\\ \chi_{\mathbf{p}}^{\prime(\mathbf{s})}(\mathbf{k}) &= \pi^{-3/2}\prod_{\alpha=1}^{3} s_\alpha\exp[-s_\alpha^{2}(k_\alpha - p_\alpha)^2],\end{aligned} \tag{1.3}$$

where $\mathbf{s} = (s_1, s_2, s_3)$ and $0 < s_\alpha < \infty$ for $\alpha = 1, 2, 3$. By assigning to each $(\mathbf{q},\mathbf{p}) \in \mathbb{R}^6$ a unique $\chi_{\mathbf{q},\mathbf{p}}$ we obtain a family $\hat{\Gamma}^6$ of fuzzy points. We refer to any such family as a *fuzzy phase space* provided that (1.2) is continuous as a function on $\mathbb{R}^{12}$.

We note that in this terminology the ordinary phase space Γ^6 can be looked upon as being the set of "fuzzy" points

$$\Gamma^6 = \{(\mathbf{q},\mathbf{p},\chi_{\mathbf{q},\mathbf{p}}) \mid (\mathbf{q},\mathbf{p}) \in \mathbb{R}^6, \; + \chi_{\mathbf{q},\mathbf{p}}(\mathbf{x},\mathbf{k}) = \delta(\mathbf{x}-\mathbf{q})\delta(\mathbf{k}-\mathbf{p})\} \tag{1.4}$$

that have confidence measures $\nu_{\mathbf{q},\mathbf{p}}$, which are Dirac measures centered at $(\mathbf{q},\mathbf{p})$. Classical mechanics allows the possibility of such points being the *optimal* sample points obtainable by measuring $\mathbf{Q}$ and $\mathbf{P}$, i.e., they are the outcome of measurements with perfectly accurate instruments. Naturally, such instruments can be viewed only as idealization of realistic instruments, i.e., they represent an asymptotic limit of a sequence of realistic instruments of ever-increasing precision. Hence, in the classical context the family of fuzzy sample points $(\mathbf{q},\mathbf{p},\chi_{\mathbf{q},\mathbf{p}})$ contains all calibration functions of the form (1.2) with $\chi_{\mathbf{q}}^{(0)}(\mathbf{x}) = \delta(\mathbf{x}-\mathbf{y})$, $\chi_{\mathbf{p}'}^{\prime(0)}(\mathbf{k}) = \delta(\mathbf{k}-\mathbf{p})$ and $d\mu(\mathbf{q},\mathbf{p},\mathbf{s}) = d\mu'(\mathbf{q}',\mathbf{p}')d\mu_\delta(\mathbf{s})$, where $\mu_\delta(\mathbf{s})$ is the Dirac measure and μ' is arbitrary as long as a precise meaning can be given to the resulting formal integral in terms of convolutions of measures.

In complete analogy, the calibration functions (1.2) for the quantum mechanical case are constructed from those in (1.3), since these last ones correspond to optimal sample points obtainable by measuring simultaneously $\mathbf{Q}$ and $\mathbf{P}$ with *optimally* accurate instruments (which cannot be, however, perfectly accurate as long as their accuracy calibration takes into consideration the usual "gedanken experiment" analysis leading to the uncertainty principle).

In order to avoid cumbersome notation, we shall restrict ourselves in this paper to those optimal calibration which are given by confidence functions of the form

$$\chi_{\mathbf{q}}^{(s)}(\mathbf{x}) = (\pi s^2)^{-3/2}\exp[-s^{-2}(\mathbf{x}-\mathbf{q})^2], \tag{1.5a}$$

$$\chi_{\mathbf{p}}^{(s^{-1})}(\mathbf{k}) = (\pi s^{-2})^{-3/2}\exp[-s^{2}(\mathbf{k}-\mathbf{p})^2] \tag{1.5b}$$

They stand out as the Galilean invariant calibrations of optimally accurate instruments. The fuzzy phase space associated with them is denoted by Γ_s.

In Sec. 2 we introduce the representation of the wavepacket in Γ_s, and show that such a representation is in a physical sense a generalization of both configuration and momentum representations. In Sec. 3 we investigate the kinematics of a free wavepacket in this representation, and prove that it has the same evanescence feature as in the configuration representation. This leads to considering in Sec. 4 its asymptotic behavior in fuzzy cones in Γ_s, and yields the strikingly simple formula (4.19).

In Sec. 5 we establish the existence of asymptotic free states in fuzzy phase space for short as well as long-range interactions, and then use (4.19) to derive the expression (5.20) for the differential cross section of scattering into fuzzy solid angles. This formula has exactly the same form as its counterpart for sharp solid angles—a fact for which we give a plausible physical explanation at the end of that section.

2. THE FUZZY PHASE SPACE REPRESENTATION OF A WAVEPACKET

Let us consider a system of two particles without spin. After eliminating the center-of-mass motion, we can describe its internal states in terms of a single particle with reduced mass. For different choices of complete sets of observables we get different representations of these states—the pure states being the elements of the spectral representation space[2] for the chosen complete set. In particular, for the position observables $\mathbf{Q}$ and momentum observables $\mathbf{P}$, the conventional choice of spectral representation space leads to $L^2(\mathbb{R}^3)$. Thus we arrive at the configuration representation $\psi(\mathbf{x})$ and momentum representation $\tilde{\psi}(\mathbf{k})$, respectively, of the same wavepacket ψ [here $\tilde{\psi}(\mathbf{k})$ stands for the Fourier–Plancherel transform $U_F\psi$ of $\psi(\mathbf{x})$, since we adopt units in which $\hbar=1$]. We then arrive at the standard interpretation according to which $|\psi(\mathbf{q})|^2$ and $|\tilde{\psi}(\mathbf{p})|^2$ are the probability densities for having a perfectly accurate measurement of $\mathbf{Q}$ and $\mathbf{P}$, respectively, on the system in the state ψ yield the respective sharp values $\mathbf{q}\in\mathbb{R}^3$ and $\mathbf{p}\in\mathbb{R}^3$.

Let us denote now by Γ_s the fuzzy phase space $\hat{\Gamma}^6$ corresponding to the optimal calibrations (1.5),

$$\Gamma_s=\{(\mathbf{q},\chi_{\mathbf{q}}^{(s)})\times(\mathbf{p},\chi_{\mathbf{p}}^{(s^{-1})})\,|\,(\mathbf{q},\mathbf{p})\in\mathbb{R}^6\}, \tag{2.1}$$

at a fixed finite value of $s>0$. On the basis of previous considerations[1] we propose interpreting

$$\langle\psi\,|F(\mathbf{q},\mathbf{p};s)\psi\rangle_0=|\psi(\mathbf{q},\mathbf{p};s)|^2 \tag{2.2}$$

as the probability density for having a simultaneous measurement of Q and P with an instrument of optimal calibration (1.5) yield the result $(\mathbf{q},\mathbf{p})\in\mathbb{R}^6$, where $\langle\cdot|\cdot\rangle_0$ denotes the L^2-inner product in the configuration representation, and

$$\psi(\mathbf{q},\mathbf{p};s)=(2\pi)^{-3/2}\langle\phi_{\mathbf{q},\mathbf{p}}^{(s)}|\psi\rangle_0, \tag{2.3}$$

$$\phi_{\mathbf{q},\mathbf{p}}^{(s)}(\mathbf{x})=(\pi s^2)^{-3/4}\exp\left[-\frac{(\mathbf{x}-\mathbf{q})^2}{2s^2}+i\mathbf{p}\left(\mathbf{x}-\frac{\mathbf{q}}{2}\right)\right]. \tag{2.4}$$

It is then natural to refer to the space

$$L^2(\Gamma_s)=\{\psi(\mathbf{q},\mathbf{p};s)=(2\pi)^{-3/2}\langle\phi_{\mathbf{q},\mathbf{p}}^{(s)}|\psi\rangle_0\,|\,\psi\in L^2(\mathbb{R}^3)\} \tag{2.5}$$

as the *representation space over the fuzzy phase space* Γ_s.

A simple computation[3] shows that

$$\phi_{\mathbf{q},\mathbf{p}}^{(s)}=\exp[i(\mathbf{pQ}-\mathbf{qP})]\phi_{0,0}^{(s)}=\exp(\boldsymbol{\zeta}_s^*a_s^*-\zeta_s\mathbf{a}_s)\phi_{0,0}^{(s)}, \tag{2.6}$$

where $\zeta_s=2^{-1/2}(s^{-1}\mathbf{q}-is\mathbf{p})$ and

$$\mathbf{a}_s=2^{-1/2}(s^{-1}\mathbf{Q}+is\mathbf{P}),\quad \mathbf{a}_s^*=2^{-1/2}(s^{-1}\mathbf{Q}-is\mathbf{P}). \tag{2.7}$$

Since $[\zeta_s^*\mathbf{a}_s^*,\zeta_s\mathbf{a}_s]=-|\zeta_{s1}|^2-|\zeta_{s2}|^2-|\zeta_{s3}|^2=-\zeta_s^*\zeta_s$, we obtain

$$\psi(\mathbf{q},\mathbf{p},s)=\exp(-\tfrac{1}{2}\zeta_s^*\zeta_s)\langle\exp(\zeta_s^*\mathbf{a}_s^*)\phi_{0,0}^{(s)}|\psi\rangle_0. \tag{2.8}$$

After expanding in a power series in $\zeta_s^*\mathbf{a}_s$ and inserting the result in (2.3), we get

$$\psi(\mathbf{q},\mathbf{p};s)=\exp[-\tfrac{1}{4}(s^{-2}\mathbf{q}^2+s^2\mathbf{p}^2)]f_\psi(\zeta_s), \tag{2.9a}$$

where f_ψ is an entire function on $\mathbb{C}^3$:

$$f_\psi(\zeta_s)=(2\pi)^{-3/2}\sum_{n_1,n_2,n_3=0}^{\infty}\frac{\zeta_{s1}^{n_1}\zeta_{s2}^{n_2}\zeta_{s2}^{n_3}}{n_1!\,n_2!\,n_3!}\langle\phi_{0,0}^{(s)}|a_{s1}^{n_1}a_{s2}^{n_2}a_{s3}^{n_3}\psi\rangle_0. \tag{2.9b}$$

On the other hand, the well-known identity for coherent states[1,3]

$$\pi^{-3}\int_{\mathbb{R}^6}|\phi_{\zeta_s}^{(s)}\rangle\,d\zeta_s\langle\phi_{\zeta_s}^{(s)}|=\mathbb{1} \tag{2.10}$$

implies that the inner product $\langle\cdot|\cdot\rangle_s$ in $L^2(\Gamma_s)$ is

$$\langle\psi_1|\psi_2\rangle_s=\int_{\mathbb{R}^6}\psi_1^*(\mathbf{q},\mathbf{p};s)\psi_2(\mathbf{q},\mathbf{p};s)\,d\mathbf{q}\,d\mathbf{p}. \tag{2.11}$$

Thus $L^2(\Gamma_s)$ is a closed proper subspace of $L^2(\mathbb{R}^6)$; it consists of all $\psi\in L^2(\mathbb{R}^6)$ of form (2.9), where f_ψ belongs to the Fisher space[4] $\mathcal{J}^3$ over $\mathbb{C}^3$. Fisher spaces have been studied in the context of quantum mechanics first by Bargmann.[5] Bargmann's results in Ref. 5 are formulated for the case $s=1$, but can be extended routinely to all $s>0$ and can be easily shown to lead to the conclusion that f_ψ varies over all of $\mathcal{J}^3$ as ψ varies over $L^2(\mathbb{R}^3)$, and that the inverse of the unitary transformation

$$U^{(s,0)}:\psi(\mathbf{x})\to\psi(\mathbf{q},\mathbf{p};s)=\int_{\mathbb{R}^3}\phi_{\mathbf{q},\mathbf{p}}^{(s)*}(\mathbf{x})\psi(\mathbf{x})\,d\mathbf{x} \tag{2.12}$$

of $L^2(\mathbb{R}^3)$ onto $L^2(\Gamma_s)$ is

$$U^{(0,s)}:\psi(\mathbf{q},\mathbf{p};s)\to\psi(\mathbf{x})$$
$$=\int_{\mathbb{R}^6}\phi_{\mathbf{q},\mathbf{p}}^{(s)}(\mathbf{x})\psi(\mathbf{q},\mathbf{p};s)\,d\mathbf{q}\,d\mathbf{p}. \tag{2.13}$$

Thus, the transition $U^{(s',0)}U^{(0,s)}$ from $L^2(\Gamma_s)$ to $L^2(\Gamma_{s'})$ is affected by[6]

$$\begin{aligned}(U^{(s',s)}\psi)(\mathbf{q}',\mathbf{p}';s')&=[2\pi^2(s'^2+s^2)]^{-3/2}(s's)^{3/2}\\&\quad\times\int\exp\{(s's)^3(s'^2+s^2)^{-3/2}[(\mathbf{q}'+\mathbf{q}+i\mathbf{p}'-i\mathbf{p})^2]\}\\&\quad\times\psi(\mathbf{q},\mathbf{p};s)\,d\mathbf{q}\,d\mathbf{p}.\end{aligned}$$

An interesting feature of the fuzzy phase space representation $\psi(\mathbf{q},\mathbf{p};s)$ of the wavepacket is that for $\psi(\mathbf{x})\in C^0\cap L^1\cap L^2$

$$\lim_{s\to+0}(\pi s^{-2})^{3/4}\psi(\mathbf{q},\mathbf{p};s)=\exp(\tfrac{1}{2}i\mathbf{pq})\psi(\mathbf{q}), \tag{2.14a}$$

$$\lim_{s\to+\infty}(\pi s^2)^{3/4}\psi(\mathbf{q},\mathbf{p};s)=\exp(-\tfrac{1}{2}i\mathbf{pq})\tilde{\psi}(\mathbf{p}). \tag{2.14b}$$

This fact is not accidental since a sharp measurement of $\mathbf{Q}$ yielding the value $\mathbf{q}$ could be considered as being a measurement of $\mathbf{Q},\mathbf{P}$ that yields $(\mathbf{q},\mathbf{p})$ with the confidence $\nu_{\mathbf{q}}(\{x\})=1$ for $\mathbf{x}$ being $\mathbf{q}$ and the confidence function $\chi_{\mathbf{p}}(\mathbf{k})\equiv 1$ for $\mathbf{k}$ being $\mathbf{p}$ (i.e., totally undecided in favor of any particular value $\mathbf{p}$ for $\mathbf{P}$). Thus the sample space $\mathbb{R}^3$ corresponding to perfectly accurate measurements of $\mathbf{Q}$ can be replaced by the "fuzzy" phase space

$$\Gamma_0=\{(\mathbf{q},\chi_{\mathbf{q}})\times(\mathbf{p},\chi_{\mathbf{p}})\,|\,\chi_{\mathbf{q}}(\mathbf{x})=\delta(\mathbf{x}-\mathbf{q}),\quad \chi_{\mathbf{p}}(\mathbf{k})\equiv 1,\ \mathbf{q},\mathbf{p}\in\mathbb{R}^3\}. \tag{2.15a}$$

In complete analogy, perfectly sharp measurements of $\mathbf{P}$ yield results in

$$\Gamma_\infty=\{(\mathbf{q},\chi_{\mathbf{q}})\times(\mathbf{p},\chi_{\mathbf{p}})\,|\,\chi_{\mathbf{q}}(\mathbf{x})\equiv 1,\quad \chi_{\mathbf{p}}(\mathbf{k})=\delta(\mathbf{k}-\mathbf{p}),\ \mathbf{q},\mathbf{p}\in\mathbb{R}^3\}. \tag{2.15b}$$

In view of (2.14) and (2.15), we can introduce the suggestive notation

$$L^2(\Gamma_0)=\{\psi(\mathbf{q},\mathbf{p};0)=\exp(\tfrac{1}{2}i\mathbf{pq})\psi(\mathbf{q})\,|\,\psi\in L^2(\mathbb{R}^3)\}, \tag{2.16a}$$

$$L^2(\Gamma_\infty)=\{\psi(\mathbf{q},\mathbf{p};\infty)=\exp(-\tfrac{1}{2}i\mathbf{pq})\tilde{\psi}(\mathbf{p})\,|\,\psi\in L^2(\mathbb{R}^3)\}, \tag{2.16b}$$

and observe that

$$|\psi(\mathbf{q})|^2=|\psi(\mathbf{q},\mathbf{p};0)|^2=\lim_{s\to+0}(\pi s^{-2})^{3/2}|\psi(\mathbf{q},\mathbf{p};s)|^2, \tag{2.17a}$$

$$|\tilde{\psi}(\mathbf{p})|^2=|\psi(\mathbf{q},\mathbf{p};\infty)|^2=\lim_{s\to+\infty}(\pi s^{2})^{3/2}|\psi(\mathbf{q},\mathbf{p};s)|^2. \tag{2.17b}$$

The last two relations, supplemented by the evident observation that for $0<s<\infty$

$$|\psi(\mathbf{q},\mathbf{p};s)|^2=\lim_{s'\to s}|\psi(\mathbf{q},\mathbf{p};s')|^2, \tag{2.18}$$

are a direct reflection of the continuity property (cf. Ref. 1, Sec. 3) of the spectral measure on fuzzy events in the phase space Γ^6. We note that in this context the appearance of the factors $(\pi s^{-2})^{3/2}$ and $(\pi s^2)^{3/2}$ in (2.17a) and (2.17b), respectively, are necessitated by the appearance of the corresponding factors in (1.5b) and (1.5a), respectively. Indeed, those factors have to be compensated for in order to have

$$\lim_{s\to+\infty}(\pi s^2)^{3/2}\chi_{\mathbf{q}}^{(s)}(\mathbf{x})\equiv 1, \tag{2.19a}$$

$$\lim_{s\to+0}(\pi s^{-2})^{3/2}\chi_{\mathbf{p}}^{(s)}(\mathbf{k})\equiv 1. \tag{2.19b}$$

These considerations show that the configuration and momentum representations of a given wavepacket can be looked upon as limiting cases of the fuzzy phase space representations. Nevertheless, there is one key difference between the two limiting cases $s=0$ and $s=\infty$, and the remaining in-between cases $0<s<\infty$: While the complete knowledge of either the configuration probability distribution $|\psi(\mathbf{q})|^2$ or the momentum distribution $|\tilde{\psi}(\mathbf{p})|^2$ does not pinpoint the corresponding wavepacket ψ uniquely, its Γ_s-probability distribution $|\psi(\mathbf{q},\mathbf{p};s)|^2$ does. This fact is an immediate consequence of the global analyticity of $f_\psi(\zeta;s)$ in (2.6). In fact, $|\psi_1(\mathbf{q},\mathbf{p};s)|^2\equiv|\psi_2(\mathbf{q},\mathbf{p};s)|^2$ implies $|f_{\psi_1}(\boldsymbol{\zeta}_s)|\equiv|f_{\psi_2}(\boldsymbol{\zeta}_s)|$, which for entire functions is true if and only if $f_{\psi_2}(\boldsymbol{\zeta}_s)\equiv cf_{\psi_1}(\boldsymbol{\zeta}_s)$ for all $\boldsymbol{\zeta}_s\in\mathbb{C}^3$ and some constant c of absolute value one. (Indeed, the set Z in $\mathbb{C}^3$ on which f_{ψ_1} vanishes has no accumulation points, since that would imply $f_{\psi_1}\equiv 0$. Thus $\mathbb{C}^3\backslash Z$ is connected. Now, by the maximum modulus theorem of complex analysis, $|f_{\psi_2}/f_{\psi_1}|=1$ at all $\boldsymbol{\zeta}_s\in\mathbb{C}^3\backslash Z$ implies $f_{\psi_2}/f_{\psi_1}=c(\mathcal{N})$ in any open neighborhood $\mathcal{N}$ of $\boldsymbol{\zeta}_s$ where f_{ψ_2}/f_{ψ_1} is analytic. The connectedness of $\mathbb{C}^3\backslash Z$ implies that $c(\mathcal{N})$ is the same for all $\boldsymbol{\zeta}_s\in\mathbb{C}^3\backslash Z$, i.e., $f_{\psi_2}\equiv cf_{\psi_1}$, $|c|=1$, at all such $\boldsymbol{\zeta}_s$. Since Z has empty interior, this result is valid globally.)

In view of the present theory of fuzzy phase space this feature is not at all surprising since fuzzy simultaneous measurements of $\mathbf{Q}$ *and* $\mathbf{P}$ obviously supply more information than the corresponding fuzzy measurements of $\mathbf{Q}$ or of $\mathbf{P}$ exclusively. On the other hand the fuzzy localization operators for $\mathbf{Q}$ as well as for $\mathbf{P}$ supply exactly the same amount of information as their sharp counterparts (cf. Ref. 7, Theorem 2).

The above results can be generalized to arbitrary fuzzy phase spaces related to $(\mathbf{Q},\mathbf{P})$ measurements with nonoptimal calibrations of the form (1.2). It is interesting to note that the general Galilean-invariant phase space (Γ^6,μ) is related to Γ_s, $0<s<\infty$, in the same manner fuzzy configuration space[7] $(\mathbb{R}^3,\nu)$ is related to ordinary configuration space $\mathbb{R}^3$: Both require a "smearing" of the optimal confidence functions with the respective Galilean-invariant measures μ and ν. However, only in case of (Γ^6,μ) these optimal calibration functions are the Gaussians (1.5), while for $(\mathbb{R}^3,\nu)$ they are the δ functions $\delta_{\mathbf{q}}$, $\mathbf{q}\in\mathbb{R}^3$.

3. THE EVOLUTION OF THE FREE WAVEPACKET IN $L^2(\Gamma_s)$

Any bounded operator A in $L^2(\Gamma_0)$ can be expressed as an integral operator in $L^2(\Gamma_s)$:

$$(A\psi)(\mathbf{q},\mathbf{p};s)=\int_{\mathbb{R}^6}A(\mathbf{q},\mathbf{p};\mathbf{q}',\mathbf{p}';s)\psi(\mathbf{q},\mathbf{p};s)\,d\mathbf{q}\,d\mathbf{p}. \tag{3.1}$$

In fact, (3.1) is a direct consequence of (2.10) rewritten in the form

$$(2\pi)^{-3}\int_{\mathbb{R}^6}|\phi_{\mathbf{q},\mathbf{p}}^{(s)}\rangle\,d\mathbf{q}\,d\mathbf{p}\langle\phi_{\mathbf{q},\mathbf{p}}^{(s)}|=\mathbb{1}, \tag{3.2}$$

and holds even for an unbounded A provided that $\phi_{\mathbf{q},\mathbf{p}}^{(s)}\in\mathcal{D}_A\cap\mathcal{D}_{A^*}$ for all $\boldsymbol{\zeta}_s\in\mathbb{C}^3$:

$$\begin{aligned}(A\psi)(\mathbf{q},\mathbf{p};s)&=(2\pi)^{-3/2}\langle\phi_{\mathbf{q},\mathbf{p}}^{(s)}|A\psi\rangle_0=(2\pi)^{-3/2}\langle A^*\phi_{\mathbf{q},\mathbf{p}}^{(s)}|\psi\rangle_0\\&=(2\pi)^{-9/2}\int_{\mathbb{R}^6}\langle A^*\phi_{\mathbf{q},\mathbf{p}}^{(s)}|\phi_{\mathbf{q}',\mathbf{p}'}^{(s)}\rangle_0\langle\phi_{\mathbf{q}',\mathbf{p}'}^{(s)}|\psi\rangle_0\,d\mathbf{q}'\,d\mathbf{p}'\\&=(2\pi)^{-3}\int_{\mathbb{R}^6}\langle\phi_{\mathbf{q},\mathbf{p}}^{(s)}|A\phi_{\mathbf{q}',\mathbf{p}'}^{(s)}\rangle\psi(\mathbf{q}',\mathbf{p}',s)\,d\mathbf{q}\,d\mathbf{p}.\end{aligned} \tag{3.3}$$

By comparing with (3.1) we see that

$$A(\mathbf{q},\mathbf{p};\mathbf{q}',\mathbf{p}';s)=(2\pi)^{-3}\langle\phi_{\mathbf{q},\mathbf{p}}^{(s)}|A\phi_{\mathbf{q}',\mathbf{p}'}^{(s)}\rangle_0. \tag{3.4}$$

In order to study the behavior of a free wavepacket $\psi_t=U_t^{(0)}\psi$ in $L^2(\Gamma_s)$, let us compute

$$U_t^{(0)}(\mathbf{q},\mathbf{p};\mathbf{q}',\mathbf{p}';s)=(2\pi)^{-3}\langle\phi_{\mathbf{q},\mathbf{p}}^{(s)}|\exp(-iH_0t)\phi_{\mathbf{q}',\mathbf{p}'}^{(s)}\rangle_0 \tag{3.5}$$

for the free Hamiltonian H_0:

$$(H_0\psi)^\sim(\mathbf{k})=(\mathbf{k}^2/2m)\tilde{\psi}(\mathbf{k}). \tag{3.6}$$

After computing the Fourier transform of $\phi_{\mathbf{q},\mathbf{p}}^{(s)}$,

$$\tilde{\phi}_{\mathbf{q},\mathbf{p}}^{(s)}(\mathbf{k})=(\pi^{-1}s^2)^{3/4}\exp\{-(s^2/2)(\mathbf{k}-\mathbf{p})^2-i(\mathbf{k}-\mathbf{p}/2)\mathbf{q}\}, \tag{3.7}$$

we immediately get

$$U_t^{(0)}(\mathbf{q},\mathbf{p};\mathbf{q}',\mathbf{p}';s)$$
$$=(\pi^{-1}s)^3\exp\{-(s^2/2)(\mathbf{p}^2+\mathbf{p}'^2)-(i/2(\mathbf{pq}-\mathbf{p}'\mathbf{q}')\}$$
$$\times I(\mathbf{q},\mathbf{p};\mathbf{q}',\mathbf{p}';s), \tag{3.8}$$

where[6]

$$I(\mathbf{q},\mathbf{p};\mathbf{q}',\mathbf{p}';s)=(4\pi)^{-3/2}\int_{\mathbb{R}^3}\exp(-\mathbf{k}^2/4\beta_t+\gamma\mathbf{k})\,d\mathbf{k}$$
$$=\beta_t^{3/2}\exp(\beta_t\gamma^2), \tag{3.9}$$

$$4\beta_t=[s^2+(i/2m)t]^{-1},\quad \gamma=s^2(\mathbf{p}+\mathbf{p}')+i(\mathbf{q}-\mathbf{q}'). \tag{3.10}$$

It is instructive to see what happens to (3.8) in the limit $s\to+0$. From (2.14a) we see that we have to multiply $U_t^{(0)}(\mathbf{q},\mathbf{p};\mathbf{q}',\mathbf{p}';s)$ by a corresponding factor before taking this limit. We get

$$\lim_{s\to+0}(\pi s^{-2})^{3/2}\exp[(i/2)(\mathbf{pq}-\mathbf{p}'\mathbf{q}')]$$
$$\times U_t^{(0)}(\mathbf{q},\mathbf{p};\mathbf{q}',\mathbf{p}';s)=(m/2\pi it)^{3/2}$$
$$\times\exp[(im/2t)(\mathbf{q}-\mathbf{q}')^2], \tag{3.11}$$

which is the standard formula for the kernel $U_t^{(0)}(\mathbf{q},\mathbf{q}';0)$ of $U_t^{(0)}$ in the configuration representation. On the other hand, pointwise in $\mathbf{p}$, the limit $s\to+\infty$ does not lead to a function. This is not surprising since from (2.4) and (3.5) we see that, in the sense of distributions,

$$\lim_{s\to\infty}(\pi s^2)^{3/2}\exp[(i/2)(\mathbf{p}'\mathbf{q}'-\mathbf{pq})]U_t^{(0)}(\mathbf{q},\mathbf{p};\mathbf{q}',\mathbf{p}';s)$$
$$=\langle\mathbf{p}\,|\,U_t^{(0)}\,|\,\mathbf{p}'\rangle_0=\exp[-i(\mathbf{p}^2/2m)t]\delta(\mathbf{p}-\mathbf{p}'). \tag{3.12}$$

In configuration space, any wavepacket that is integrable in $\mathbf{q}\in\mathbb{R}^3$, $\psi(\mathbf{q})\in L^2(\mathbb{R}^3)\cap L^1(\mathbb{R}^3)$, displays in time the well-known behavior $|\psi_t(\mathbf{q})|\sim O(|t|^{-3/2})$. The fact $L^2\cap L^1$ is dense in L^2, combined with the unitarity of $U_t^{(0)}$, implies then the evanescence of all free wavepackets from any bounded region B of configuration space.

A similar result holds in Γ_s. In fact, by combining (3.8)–(3.10) we get

$$|\psi_t(\mathbf{q},\mathbf{p};s)|$$
$$\leq(\pi^{-1}s^3|\beta_t|^{3/2}\exp(-s^2p^2/2)\int_{\mathbb{R}^3}d\mathbf{p}'\exp(-s^2\mathbf{p}'^2/2)$$
$$\times\int_{\mathbb{R}^3}d\mathbf{q}\exp\{|\beta_t|(s^2|\mathbf{p}+\mathbf{p}'|+|\mathbf{q}-\mathbf{q}'|)^2\}$$
$$\times|\psi(\mathbf{q}',\mathbf{p}';s)|. \tag{3.13}$$

Now for any $\psi\in L^2(\Gamma_s)$ and all $\mathbf{q},\mathbf{p}\in\mathbb{R}^3$,

$$|\psi(\mathbf{q},\mathbf{p};s)|\leq(2\pi)^{-3/2}\|\psi\|_s \tag{3.14}$$

and therefore (cf. Ref. 5, pp. 193, 197)

$$|\psi^{(\kappa)}(\mathbf{q},\mathbf{p};s)|$$
$$\leq(2\pi)^{-3/2}\|\psi\|_s\exp[-\kappa(s^{-2}\mathbf{q}^2+s^2\mathbf{p}^2)/4],\quad 0\leq\kappa<1, \tag{3.15}$$

$$\psi^{(\kappa)}(\mathbf{q},\mathbf{p};s)$$
$$=\exp\{-\kappa(s^{-2}\mathbf{q}^2+s^2\mathbf{p}^2)/4\}\psi((1-\kappa)^{1/2}\mathbf{q},(1-\kappa)^{1/2}\mathbf{p};s). \tag{3.16}$$

In view of the large-time behavior of β_t,

$$\beta_t=(m/2it)[1+O(|t|^{-1})], \tag{3.17}$$

and the inequality

$$(s^2|\mathbf{p}+\mathbf{p}'|+|\mathbf{q}-\mathbf{q}'|)^2\leq2[s^2(\mathbf{p}^2+\mathbf{p}'^2)+\mathbf{q}^2+\mathbf{q}'^2], \tag{3.18}$$

we conclude that when $|t|$ is sufficiently large (so that, for instance, $16|\beta_t|\leq\min\{\kappa s^{-2},2(\kappa+1)\}$) the integral in (3.13) can be majorized by a constant. Consequently, for such values of t

$$|\psi_t^{(\kappa)}(\mathbf{q},\mathbf{p};s)|\leq\text{const}\,|t|^{-3/2}$$
$$\times\exp[-s^2(1-4|\beta_t|\mathbf{p}^2)+2|\beta_t|\mathbf{q}^2/2]. \tag{3.19}$$

We have (Ref. 5, p. 197, Sec. 1g):

$$\lim_{\kappa\to+0}\|\psi-\psi^{(\kappa)}\|_s=0. \tag{3.20}$$

Consequently, if K and M are Borel sets in $\mathbb{R}^3$ and K is bounded and of Lebesgue measure $|K|$, then

$$\{\int_K d\mathbf{q}\int_M d\mathbf{p}\,|\psi_t(\mathbf{q},\mathbf{p};s)|^2\}^{1/2}$$
$$\leq\|\psi-\psi^{(\kappa)}\|_s+[\int_K d\mathbf{q}\int_M d\mathbf{p}\,|\psi_t^{(\kappa)}(\mathbf{q},\mathbf{p};s)|^2]^{1/2}$$
$$\leq\|\psi-\psi^{(\kappa)}\|+|K|\,C_K[\,|t|^{-3}\int_M\exp(-s^2\mathbf{p}^2/2)\,d\mathbf{p}]^{1/2}. \tag{3.21}$$

This establishes the evanescence of ψ from $B\times K$.

Concerning the Γ_s representation of $U_t^{(0)}$, it is interesting to note that by expanding γ^2 in (3.9) in terms of its constituent vectors and inserting the result in (3.8) we get

$$U_t^{(0)}=U_{-t}^{(2)*}U_t^{(1)}U_t^{(2)},$$
$$(U_t^{(1)}\psi)(\mathbf{q},\mathbf{p};s)=\beta_t^{3/2}\int\exp[2\beta_t(s^2\mathbf{p}+i\mathbf{q})$$
$$\times(s^2\mathbf{p}'-i\mathbf{q}')]\psi(\mathbf{q},\mathbf{p};s)\,d\mathbf{q}\,d\mathbf{p}, \tag{3.22}$$
$$(U_t^{(2)}\psi)(\mathbf{q},\mathbf{p};s)=(\pi^{-1}s)^{3/2}\exp[-s^2\mathbf{p}^2/2+i\mathbf{pq}$$
$$+\beta_t(s^2\mathbf{p}-i\mathbf{q})^2]\psi(\mathbf{q},\mathbf{p};s).$$

This decomposition generalizes a similar one[8] in configuration space (cf. Ref. 2, p. 414) that has proved very useful in scattering theory.[8] It should be noted, however, that $U_t^{(1)}$ and $U_t^{(2)}$ do not leave $L^2(\Gamma_s)$ invariant, and should be regarded as operators on $L^2(\mathbb{R}^6)$.

4. ASYMPTOTIC PROBABILITIES FOR FUZZY CONES IN Γ_s

Consider now a fuzzy Borel set[1]

$$K_s\times M_{s^{-1}}=\{(\mathbf{q},\chi_{\mathbf{q}}^{(s)})\times(\mathbf{p},\chi_{\mathbf{p}}^{(s^{-1})})\,|\,\mathbf{q}\in K,\mathbf{p}\in M\} \tag{4.1}$$

in Γ_s, that corresponds to some Borel sets K, $M\subset\mathbb{R}^3$. We have already seen that the probability

$$P_{\psi_t}(K_s\times M_{s^{-1}})=\int_K d\mathbf{q}\int_M d\mathbf{p}\,|\psi_t(\mathbf{q},\mathbf{p};s)^2 \tag{4.2}$$

of measuring simultaneous fuzzy values $(\mathbf{q},\chi_{\mathbf{q}}^{(s)})$ and $(\mathbf{p},\chi_{\mathbf{p}}^{(s^{-1})})$ that belongs to $K_s\times M_{s^{-1}}$ vanishes in the limit $t\to\pm\infty$ if K is bounded in $\mathbb{R}^3$. Consequently, let K be a cone with apex at the origin, so that for all $\tau>0$

$$\tau K=\{\tau\mathbf{q}\,|\,\mathbf{q}\in K\}=K. \tag{4.3}$$

We shall prove that in this case P_{ψ_t} in (4.2) has for $t\to\pm\infty$ a limiting value which in general is greater than zero.

Using (3.7), we get

$$(2\pi)^{3/2}\,\psi_t(\mathbf{q},\mathbf{p};s)$$
$$=\langle\tilde{\phi}_{\mathbf{q},\mathbf{p}}^{(s)}\,|\,(U_t^{(0)}\psi)^{\sim}\rangle_0=(\pi^{-1}s^2)^{3/4}$$
$$\times\int_{\mathbb{R}^3}\exp[i\mathbf{qk}-i(\mathbf{k}^2/2m)t-s^2(\mathbf{p}-\mathbf{k})^2/2$$
$$+i\mathbf{qp}/2]\tilde{\psi}(\mathbf{k})\,d\mathbf{k}. \tag{4.4}$$

Hence, for any fixed $\mathbf{p} \in \mathbb{R}^3$ we have

$$\psi_t(\mathbf{q},\mathbf{p};s) = (2\pi)^{-3/2} \exp(i\mathbf{q}\mathbf{p}/2) \times \int_{\mathbb{R}^3} \exp(i\mathbf{q}\mathbf{k})(U_t^{(0)}\tilde{g}_{\mathbf{p}})(\mathbf{k})\, d\mathbf{k}, \tag{4.5}$$
$$\tilde{g}_{\mathbf{p}}(\mathbf{k}) = (4\pi^{-1}s^2)^{3/4} \exp\{-s^2(\mathbf{p}-\mathbf{k})^2/2\}\tilde{\psi}(\mathbf{k}).$$

Clearly, $\tilde{g}_{\mathbf{p}}$ is square-integrable in $\mathbf{k} \in \mathbb{R}^3$. Thus we can apply the configuration representation (3.11) for $U_t^{(0)}$ to the inverse Fourier transform $g_{\mathbf{p}}(\mathbf{q})$ of $\tilde{g}_{\mathbf{p}}(\mathbf{k})$:

$$\psi_t(\mathbf{q},\mathbf{p};s) = (m/2\pi i t)^{3/2} \exp(i\mathbf{q}\mathbf{p}/2) \times \int_{\mathbb{R}^3} \exp[(im/2t)(\mathbf{q}-\mathbf{q}')^2] g_{\mathbf{p}}(\mathbf{q}')\, d\mathbf{q}'. \tag{4.6}$$

By inserting the expression for $g_{\mathbf{p}}(\mathbf{q}')$ into (4.6) and then expanding the square $(\mathbf{q}-\mathbf{q}')^2$, we get

$$(U_t^{(0)}\psi)(\mathbf{q},\mathbf{p};s) = (W_t^{(s)}\Phi_t^{(s)}\psi)(\mathbf{q},\mathbf{p};s), \tag{4.7a}$$
$$(\Phi_t^{(s)}\psi)(\mathbf{q},\mathbf{p};s) = \exp\{-i[\mathbf{q}\mathbf{p} - (m/t)\mathbf{q}^2]/2\}\psi(\mathbf{q},\mathbf{p};s), \tag{4.7b}$$
$$(W_t^{(s)}\psi)(\mathbf{q},\mathbf{p};s) = (m/it)^{3/2} \exp\{i[\mathbf{q}\mathbf{p} + (m/t)\mathbf{q}^2]/2\}\hat{\psi}(m\mathbf{q}/t,\mathbf{p};s), \tag{4.7c}$$

where $\hat{\psi}$ denotes the Fourier transform of ψ with respect to $\mathbf{q}$:

$$\hat{\psi}(\mathbf{q}',\mathbf{p};s) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} \exp(-i\mathbf{q}'\mathbf{q})\psi(\mathbf{q},\mathbf{p};s)\, d\mathbf{q}. \tag{4.8}$$

We note that neither $\Phi_t^{(s)}$ nor $W_t^{(s)}$ leave $L^2(\Gamma_s)$ invariant. However, $\Phi_t^{(s)}$ is evidently a unitary operator on $L^2(\mathbb{R}^6)$, and so is $W_t^{(s)}$, as proven by the change of variable $mt^{-1}\mathbf{q} = \mathbf{q}'$ in the following integral:

$$\int_{\mathbb{R}^6} |(W_t^{(s)}\psi)(\mathbf{q},\mathbf{p};s)|^2\, d\mathbf{q}\, d\mathbf{p} = \int_{\mathbb{R}^6} |\hat{\psi}(\mathbf{q}',\mathbf{p};s)|^2\, d\mathbf{q}'\, d\mathbf{p}. \tag{4.9}$$

Thus we have

$$\|\psi_t - W_t^{(s)}\Phi_\infty^{(s)}\psi\| = \|(\Phi_t^{(s)} - \Phi_\infty^{(s)})\psi\|, \tag{4.10}$$

where in accordance with (4.7b)

$$(\Phi_\infty^{(s)}\psi)(\mathbf{q},\mathbf{p};s) = \exp(-i\mathbf{q}\mathbf{p}/2)\psi(\mathbf{q},\mathbf{p};s). \tag{4.11}$$

On the other hand,

$$\|\Phi_\infty^{(s)}\psi - \Phi_t^{(s)}\psi\|_s^2 = \int_{\mathbb{R}^6} |1 - \exp[(im/2t)\mathbf{q}^2]|^2 |\psi(\mathbf{q},\mathbf{p};s)|^2 d\mathbf{q}\, d\mathbf{p} \tag{4.12}$$

converges to zero by Lebesgue's dominated convergence theorem.[2] Thus we can state that

$$\underset{t\to\pm\infty}{\text{s-lim}}(U_t^{(0)} - W_t^{(s)}\Phi_\infty^{(s)}) = 0. \tag{4.13}$$

Since the probabilities P_ψ are expectation values of bounded operators[2] in $L^2(\Gamma_0)$,

$$P_\psi(B_s) = \langle\psi | E^{\mathbf{Q},\mathbf{P}}(B_s)\psi\rangle_0,\ B_s \subset \Gamma_s, \tag{4.14}$$

we conclude from (4.13) that

$$\lim_{t\to\pm\infty}[P_{\psi_t}(K_s \times M_{s^{-1}}) - P_{\psi_t'}(K_s \times M_{s^{-1}})] = 0,$$
$$\psi_t' = W_t^{(s)}\Phi_\infty^{(s)}\psi. \tag{4.15}$$

Making again the transition from the variable $\mathbf{q}$ to $\mathbf{q}' = mt^{-1}\mathbf{q}$ and taking into consideration (4.3), we obtain, for $t \neq 0$,

$$P_{\psi_t'}(K_s \times M_{s^{-1}}) = \int_{\pm K} d\mathbf{q}' \int_M d\mathbf{p}\, |(\Phi_\infty^{(s)}\psi)\hat{\ }(\mathbf{q}',\mathbf{p};s)|^2, \tag{4.16}$$

where $-K = \{-\mathbf{q} \mid \mathbf{q} \in K\}$ corresponds to the case $t < 0$.

We note that (4.16) is actually time-independent. Futhermore, using (4.4) at $t = 0$, we obtain

$$(\Phi_\infty^{(s)}\psi)(\mathbf{q},\mathbf{p};s) = (4\pi s^2)^{3/4} \times \int_{\mathbb{R}^3} \exp[i\mathbf{q}\mathbf{k} - s^2(\mathbf{p}-\mathbf{k})^2/2]\tilde{\psi}(\mathbf{k})\, d\mathbf{k}, \tag{4.17}$$

which inserted in (4.8) leads to the conclusion that

$$(\Phi_\infty^{(s)}\psi)\hat{\ }(\mathbf{q}',\mathbf{p};s) = (\pi^{-1}s^2)^{3/4}\tilde{\psi}(\mathbf{q}')\exp[-s^2(\mathbf{q}'-\mathbf{p})^2/2]. \tag{4.18}$$

Hence the contribution of the integration over M is the same for all $\psi \in L^2(\Gamma_s)$:

$$P_{\psi_{\pm\infty}}(K_s \times M_{s^{-1}}) = \lim_{t\to\pm\infty} P_{\psi_t}(K_s \times M_{s^{-1}}) = (\pi^{-1}s^2)^{3/2} \int_{\pm K} d\mathbf{q}\, |\tilde{\psi}(\mathbf{q})|^2 \int_M d\mathbf{p} \exp[-s^2(\mathbf{p}-\mathbf{q})^2]. \tag{4.19}$$

When $M = \mathbb{R}^3$, the above formula describes the asymptotic probabilities for finding the particle in the fuzzy cone K_s as $t \to \pm\infty$, regardless of the values of its momentum:

$$P_{\psi_\pm}(K_s \times \mathbb{R}^3_{s^{-1}}) = \int_{\pm K} |\tilde{\psi}(\mathbf{q})|^2\, d\mathbf{q}. \tag{4.20}$$

We note the remarkable fact that this probability coincides with that of finding it in the sharp cone K_0.

If we take $K = \mathbb{R}^3$ and afterwards go to the limit $s \to +\infty$, we recover the standard formula

$$P_{\psi_{\pm\infty}}(\mathbb{R}^3_\infty \times M_0) = \int_M |\tilde{\psi}(\mathbf{p})|^2\, d\mathbf{p} \tag{4.21}$$

of detecting the particle in the sharp region M_0 of momentum space.

Consider now the case when $M_{s^{-1}}$ is a fuzzy cone with apex at the origin, and $K \cap M = \emptyset$. We note that in general (4.19) does not vanish although the particle is fuzzy-localized in K_s, and K does not intersect M. This is, however, to be expected since the imperfectly accurate instrument used in determining some $(\mathbf{q}, \chi_{\mathbf{q}}^{(s)}) \in K_s$ can give (fuzzy) readings of momenta whose direction vectors do not lie within $\pm K$. On the other hand, if K and M are very narrow cones around the unit vectors $\mathbf{q}_0$ and $\mathbf{p}_0$, respectively, we see that for the $t = +\infty$ case (4.19) assumes its maximum when $\mathbf{p}_0 = \mathbf{q}_0$ and its minimum when $\mathbf{p}_0 = -\mathbf{q}_0$; naturally, the converse is true for $t = -\infty$.

5. ASYMPTOTIC STATES AND THE DIFFERENTIAL CROSS SECTION IN Γ_s

Consider now the case when the particles interact and the internal energy operator is the Hamiltonian $H = H_0 + V$. If the interacting term V is nonlocal or of short-range and local, then the wave operators $\Omega_\pm$ can be defined by the strong limits[2]

$$\Omega_\pm = \underset{t\to\pm\infty}{\text{s-lim}}\, U_t^* U_t^{(0)}, \quad U_t = \exp(-iHt). \tag{5.1}$$

Consequently, every interacting state represented in Schrodinger picture by $U_t\psi$, will have asymptotic states $\psi_{\rm in}(t) = U_t^{(0)}\psi_-$ and $\psi_{\rm out}(t) = U_t^{(0)}\psi_+$ that are asymptotic in Γ_s, i.e., which satisfy

$$\lim_{t\to\pm\infty}[P_{U_t\psi}(B_s) - P_{U_t^{(0)}\psi_\pm}(B_s)] = 0, \quad \psi_\pm = \Omega_\pm\psi, \tag{5.2}$$

for any fuzzy Borel set $B_s \subset \Gamma_s$. Indeed, according to (4.15),

$$
\begin{aligned}
&|P_{U_t\psi}(B_s) - P_{U_t^{(0)}\psi}(B_s)| \\
&\quad = |\langle U_t\psi | E^{\mathbf{Q},\mathbf{P}}(B_s)U_t\psi\rangle_0 - \langle U_t^{(0)}\psi_\pm | E^{\mathbf{Q},\mathbf{P}}(B_s)U_t^{(0)}\psi_\pm\rangle_0| \\
&\quad \le |\langle U_t\psi | E^{\mathbf{Q},\mathbf{P}}(B_s)(U_t\psi - U_t^{(0)}\psi_\pm)\rangle_0| \\
&\qquad + |\langle U_t\psi - U_t^{(0)}\psi_\pm | E^{\mathbf{Q},\mathbf{P}}(B_s)U_t^{(0)}\psi_\pm\rangle_0|
\end{aligned}
\tag{5.3}
$$

and the expressions on the right-hand side of the above inequality vanish in the limit $t\to\pm\infty$ since $\|U_t\|=\|U_t^{(0)}\| = 1$, $\|E^{\mathbf{Q},\mathbf{P}}(B_s)\| < \infty$, and $\|U_t\psi - U_t^{(0)}\Omega_\pm\psi\| \to 0$ by (4.1).

We shall show now that if we deal with long-range potentials,

$$V(\mathbf{x}) = c|\mathbf{x}|^{-1+\delta} + V_0(\mathbf{x}), \quad \delta \ge 0, \quad V_0(\mathbf{x}) = O(|\mathbf{x}|^{-2+\epsilon}), \tag{5.4}$$

the statement (5.2) stays true at least as long as $\delta < \frac{1}{2}$.

We recall first that for the long-range potentials in (5.4) the strong limit (5.1) does not exist, while the corresponding weak limit is zero.[9] The wave operators are instead defined[8,9] by

$$\Omega_\pm = \operatorname*{s-lim}_{t\to\pm\infty} U_t^* U_t^{(0)} \exp(-iG_t^{(\delta)}), \tag{5.5}$$

where $G_t^{(\delta)}$ is a self-adjoint function of H_0 and therefore also of $\mathbf{P}$; for $t>0$ it can be chosen as follows[8,10]:

$$G_{\pm t}^{(\delta)}(\mathbf{P}) = \begin{cases} \pm mc|\mathbf{P}|^{-1}\ln(2m^{-1}t\mathbf{P}^2), & \delta = 0, \\ \pm m^{1+\delta}c\delta^{-1}|\mathbf{P}|^{-(1-\delta)/2}t^\delta, & 0<\delta<\frac{1}{2}. \end{cases} \tag{5.6}$$

Consequently, by replacing in (5.3) $U_t^{(0)}$ by

$$\tilde{U}_t^{(0)} = U_t^{(0)}\exp(-iG_t^{(\delta)}), \tag{5.7}$$

we immediately obtain

$$\lim_{t\to\pm\infty}[P_{U_t\psi}(B_s) - P_{\tilde{U}_t^{(0)}\psi_t}(B_s)] = 0. \tag{5.8}$$

Thus, in order to establish (4.2) it remains to prove that

$$
\begin{aligned}
P_{\tilde{U}_t^{(0)}\psi}(B_s) - P_{U_t^{(0)}\psi}(B_s) = \int_B \{ &|(\tilde{U}_t^{(0)}\psi)(\mathbf{q},\mathbf{p};s)|^2 \\
&- |(U_t^{(0)}\psi)(\mathbf{q},\mathbf{p};s)|^2\}\, d\mathbf{q}\, d\mathbf{p}
\end{aligned}
\tag{5.9}
$$

vanishes in the limit $t\to\pm\infty$.

By duplicating the argument leading to (4.7) with $U_t^{(0)}$ replaced, however, by $\tilde{U}_t^{(0)}$ we arrive at the conclusion that

$$\tilde{U}_t^{(0)} = \tilde{W}_t^{(s)}\Phi_t^{(s)}, \tag{5.10a}$$

$$
\begin{aligned}
(\tilde{W}_t^{(s)}\psi)(\mathbf{q},\mathbf{p};s) = (m/it)^{3/2}&\exp\{i[\mathbf{qp} + m/t)\mathbf{q}^2]/2\} \\
&\times \hat{\psi}_{\mathrm{ren}}(m\mathbf{q}/t,\mathbf{p};s;t),
\end{aligned}
\tag{5.10b}
$$

$$\hat{\psi}_{\mathrm{ren}}(\mathbf{q}',\mathbf{p};s;t) = \exp[-iG_t^{(\delta)}(\mathbf{q}')]\hat{\psi}(\mathbf{q}',\mathbf{p};s). \tag{5.10c}$$

We easily compute (cf. Ref. 10, p. 105):

$$
\begin{aligned}
\|(\tilde{U}_t^{(0)} - \tilde{W}_t^{(s)}\Phi_\infty^{(s)})\psi\|_s^2 = (2\pi)^{-3}\int_{\mathbb{R}^6} d\mathbf{q}'\, d\mathbf{p} \left|\exp\left(\frac{im\mathbf{q}'^2}{2t}\right) - 1\right|^2 \\
\times |\psi_{\mathrm{ren}}(\mathbf{q}',\mathbf{p};s;t)|^2.
\end{aligned}
\tag{5.11}
$$

From the proof of Lemma 1 in Ref. 10 it follows that

$$
\begin{aligned}
&\int_{\mathbb{R}^3} d\mathbf{q}\,|\exp(im\mathbf{q}^2/2t) - 1|^2|\psi_{\mathrm{ren}}(\mathbf{q},\mathbf{p};s;t)|^2 \\
&\quad \le b(t)\int_{\mathbb{R}^3} d\mathbf{q}\,|\psi(\mathbf{q},\mathbf{p};s)|^2,
\end{aligned}
\tag{5.12}
$$

where $b(t)\to 0$ as $t\to\pm\infty$. By using the estimate (5.12) in (5.11), we immediately arrive at the result

$$\operatorname*{s-lim}_{t\to\pm\infty}[\tilde{U}_t^{(0)} - \tilde{W}_t^{(s)}\Phi_\infty^{(s)}] = 0. \tag{5.13}$$

Since according to (5.10b) and (5.10c)

$$P_{\tilde{W}_t^{(s)}\psi}(B_s) = |m/t|^3\int_B (\Phi_\infty^{(s)}\psi)^\wedge(m\mathbf{q}/t,\mathbf{p};s)^2\, d\mathbf{q}\, d\mathbf{p}, \tag{5.14}$$

we infer from (5.8) that

$$\lim_{t\to\pm\infty}[P_{\tilde{U}_t^{(0)}\psi}(B_s) - \int_{B(tm^{-1})}|(\Phi_\infty^{(s)}\psi)^\wedge(\mathbf{q}',\mathbf{p};s)|^2\, d\mathbf{q}'\, d\mathbf{p}] = 0, \tag{5.15}$$

where $B(\tau) = \{(\tau\mathbf{q},\mathbf{p}) | (\mathbf{q},\mathbf{p}) \in B\}$. On the other hand, we can deduce from (5.7) in exactly the same manner that

$$\lim_{t\to\pm\infty}\{P_{U_t^{(0)}\psi} - \int_{B(tm^{-1})}|(\Phi_\infty^{(s)}\psi)^\wedge(\mathbf{q}',\mathbf{p};s)|^2\, d\mathbf{q}'\, d\mathbf{p}\} = 0. \tag{5.16}$$

In view of (5.8), the relations (5.15) and (5.16) establish our original contention about (5.2) being true also in the long-range case.

With (5.2) established for nonlocal and local short-range as well as long-range potentials (under constraints like $\delta > \frac{1}{2}$, which are of a purely technical nature and can be probably eliminated by a more detailed analysis) we can make the claim that, for all such two-body interactions,

$$
\begin{aligned}
\lim_{t\to\pm\infty} P_{U_t\psi}(K_s\times M_{s^{-1}}) = (\pi^{-1}s^2)^{3/2}&\int_{\pm K} d\mathbf{q}\,|\tilde{\psi}_\pm(\mathbf{q})|^2 \\
&\times\int_M d\mathbf{p}\exp[-s^2(\mathbf{p}-\mathbf{q})^2]
\end{aligned}
\tag{5.17}
$$

when K is a cone in $\mathbb{R}^3$ with apex at the origin; we get this result from (5.2) by replacing in (4.19) ψ with $\psi_\pm$.

Let us therefore introduce[2] the T matrix $T(p;\boldsymbol{\omega}_{\mathrm{out}},\boldsymbol{\omega}_{\mathrm{in}})$ on the energy shell $\mathbf{p}_{\mathrm{in}}^2 = \mathbf{p}_{\mathrm{out}}^2 = \mathbf{p}^2$, and use the polar coordinates $\mathbf{p}_{\mathrm{in}} = (p,\boldsymbol{\omega}_{\mathrm{in}})$ and $\mathbf{p}_{\mathrm{out}} = (p,\boldsymbol{\omega}_{\mathrm{out}})$. Since $\psi_+ = S\psi_-$ and $S = \mathbb{1} - 2\pi iT$, we can write for any nonforward direction $\boldsymbol{\omega}_{\mathrm{out}} \ne \boldsymbol{\omega}_{\mathrm{in}}$ (i.e., any direction that contains no points from the support of $\tilde{\psi}_-$),

$$
\begin{aligned}
&\tilde{\psi}_+(p,\boldsymbol{\omega}_{\mathrm{out}}) \\
&\quad = -2\pi i\int T(\mathbf{p};\boldsymbol{\omega}_{\mathrm{out}},\boldsymbol{\omega}_{\mathrm{in}})\tilde{\psi}_-(p,\boldsymbol{\omega}_{\mathrm{in}})\, d\boldsymbol{\omega}_{\mathrm{in}},
\end{aligned}
\tag{5.18}
$$

where the integration in $\boldsymbol{\omega}_{\mathrm{in}}$ is over the unit sphere in $\mathbb{R}^3$.

If K is a cone with apex at the origin that cuts out on this unit sphere the solid angle Ω, i.e., $K = \{(\mathbf{q},\boldsymbol{\omega}) | \boldsymbol{\omega} \in \Omega\}$, then according to (5.17)

$$
\begin{aligned}
&\lim_{t\to+\infty} P_{U_t\psi}(K_s\times\mathbb{R}^3_{s^{-1}}) \\
&\quad = \int_\Omega d\boldsymbol{\omega}_{\mathrm{out}}\int_0^\infty |\tilde{\psi}_+(q,\boldsymbol{\omega}_{\mathrm{out}})|^2 q^2\, dq.
\end{aligned}
\tag{5.19}
$$

Consider the asymptotic probability density of observing the system in the fuzzy direction specified by the fuzzy ray

$$\hat{r}_{\mathrm{out}}^{(s)} = \{(\mathbf{q},\chi_{\mathbf{q}}^{(s)}) | \mathbf{q} = (q,\boldsymbol{\omega}_{\mathrm{out}}),\ 0 \le q < \infty\} \tag{5.20}$$

in $\mathbb{R}^3_s$. It is obviously obtained when dividing the right-hand side of (5.20) by the area $|\Omega|$ of Ω and going to the limit $\Omega \to \{\boldsymbol{\omega}_{\mathrm{out}}\}$:

$$
\begin{aligned}
\frac{dP}{d\Omega}(\psi_- \to \hat{r}_{\mathrm{out}}^{(s)}) = 4\pi^2\int_0^\infty p^2\, dp \\
\times\left|\int T(p;\boldsymbol{\omega}_{\mathrm{out}},\boldsymbol{\omega}_{\mathrm{in}})\tilde{\psi}_-(p,\boldsymbol{\omega}_{\mathrm{in}})\, d\boldsymbol{\omega}_{\mathrm{in}}\right|^2.
\end{aligned}
\tag{5.21}
$$

Let us now employ the standard derivation (cf. Ref.

2, pp. 407–13 or Ref. 11, pp. 46–51) relating differential cross sections to the T matrix. We then arrive at the expression

$$\frac{d\sigma}{d\Omega}\left(\frac{p^2}{2m},\omega_{\text{out}}\right)=\frac{(2\pi)^4}{p^2}\,|T(p;\omega_{\text{out}},-\omega_{\text{in}})|^2 \qquad (5.22)$$

as being the differential cross section for having a beam, coming in along $-\omega_{\text{in}}$, scatter from a target (placed at the origin and extending in a plane orthogonal to ω_{in}) in the direction $\hat{r}^{(s)}_{\text{out}}$ specified in (5.20)—regardless of the direction of the momentum of the scattered particles.

We see that (5.22) reads the same as when perfectly precise measurement of ω_{out} are performed. Naturally, this fact is a consequence of having obtained for the asymptotic probability (5.19) of scattering within a fuzzy cone K_s the same formula as for scattering in the corresponding sharp cone K_0.

From an intuitive physical point of view this result is not that surprising if we recall the evanescence of the wavepacket from any bounded region: To observe the particle in the direction (5.20) after an infinite time interval $[0,\infty)$ has elapsed, we have to place the detector at infinity along the direction ω_{out}; consequently, the particle is going to "see" an *aperture* whose fuzziness has diminished to zero and had become "sharp" because of the infinite distance at which the apparatus had been placed.

A confirmation of the correctness of this reasoning can be obtained by considering the asymptotic behavior of the probability

$$P_{U_t^{(0)}\psi}(K_s)=\int_K d\mathbf{q}\int_{\mathbb{R}^3}\chi_{\mathbf{q}}^{(s)}(\mathbf{x})\,|(U_t^{(0)}\psi)(\mathbf{x})|^2\,d\mathbf{x} \qquad (5.23)$$

of detecting a particle in the fuzzy cone K_s of the fuzzy configuration space[7] $\mathbb{R}^3_s=\{\mathbf{q},\chi_{\mathbf{q}}^{(s)})\,|\,\mathbf{q}\in\mathbb{R}^3\}$. In view of the fact that[2,8]

$$\lim_{t\to\pm\infty}\int_{\mathbb{R}^3}\left|(U_t^{(0)}\psi)(\mathbf{x})-\left(\frac{m}{it}\right)^{3/2}\exp\left(\frac{im}{2t}\mathbf{x}^2\right)\tilde{\psi}\left(\frac{mx}{t}\right)\right|d\mathbf{x}=0 \qquad (5.24)$$

and that, after using Tonelli's and Fubini's theorems[2] to interchange in (5.23) the orders of integration, we have $\int_K\chi_{\mathbf{q}}^{(s)}(\mathbf{x})\,d\mathbf{q}=1$, we can deduce that

$$\lim_{t\to+\infty}\left(P_{U_t^{(0)}\psi}(K_s)-\left|\frac{m}{t}\right|^3\int_{\mathbb{R}^3}d\mathbf{x}\left|\psi\left(\frac{m\mathbf{x}}{t}\right)\right|^2\int_K d\mathbf{q}\chi_{\mathbf{q}}^{(s)}(\mathbf{x})\right)=0. \qquad (5.25)$$

Hence, by making the transition to the variables $\mathbf{x}'=mt^{-1}\mathbf{x}$ and $\mathbf{q}'=m^{-1}t\mathbf{q}$ we easily compute that

$$\begin{aligned}\lim_{t\to\pm\infty}P_{U_t^{(0)}\psi}(K_s)&=\lim_{t\to\pm\infty}\left[\pi\left(\frac{ms}{t}\right)^2\right]^{-3/2}\\&\quad\times\int_{\pm K}d\mathbf{q}'\int_{\mathbb{R}^3}\exp\left[-\left(\frac{ms}{t}\right)^{-2}(\mathbf{x}'-\mathbf{q}')^2\right]|\tilde{\psi}(\mathbf{x}')|^2\,d\mathbf{x}'\\&=\int_{\pm K}|\tilde{\psi}(\mathbf{q}')|^2\,d\mathbf{q}'.\end{aligned} \qquad (5.26)$$

Thus we see that, indeed, we obtain an expression that is identical to that for the asymptotic probability of scattering within a sharp cone K_0.

6. CONCLUSION

The main conclusion we have arrived at is that two-body scattering theory in phase space based on interpreting (2.2) as the probability density for observing the fuzzy value $(\mathbf{q},\chi_{\mathbf{q}}^{(s)})\times(\mathbf{p},\chi_{\mathbf{p}}^{(s^{-1})})\in\Gamma_s$ leads to exactly the same observational consequences as scattering theory based on perfectly sharp measurements. This is somewhat surprising since, as we have argued in Sec. 2, fuzzy measurements in Γ_s certainly supply more information than both fuzzy or sharp measurements in either configuration space or in momentum space. Thus somehow this additional information gets lost if we wait an infinitely long time.

To understand the mathematical reasons for this phenomenon, we have to look at the relations (4.7): The operators $W_t^{(s)}$ when applied to $\psi(\mathbf{q},\mathbf{p};s)$ destroy the analyticity of its $f_\psi(\boldsymbol{\zeta}_s)$ factor. Yet, it is essentially the contribution of $W_t^{(s)}$ that survives when $t\to\pm\infty$. But as $(W_t^{(s)}\psi)(\mathbf{q},\mathbf{p};s)$ is not in $L^2(\Gamma_s)$, its values cannot be uniquely (up to a multiplicative constant) reconstructed from those of $|(W_t^{(s)}\psi)(\mathbf{q},\mathbf{p};s)|^2$.

On the other hand, $(U_t^{(0)}\psi)(\mathbf{q},\mathbf{p};s)$ belongs to $L^2(\Gamma_s)$ and therefore it can be uniquely specified in terms of [cf. (2.9b)] the function $f_{\psi_t}(\boldsymbol{\zeta}_s)$, which is entire in $\boldsymbol{\zeta}_s\in\mathbb{C}^3$; hence it can be reconstructed from the knowledge of $|(U_t^{(0)}\psi)(\mathbf{q},\mathbf{p},s)|^2$ on Γ_s. Moreover, observe that due to analyticity, $f_{\psi_t}(\boldsymbol{\zeta}_s)$ can be reconstructed if we know it on any characteristic set $\mathfrak{C}$ of values[5] of $\boldsymbol{\zeta}_s\in\mathbb{C}^3$. We note that the images $\mathfrak{C}_s$ in Γ_s of such characteristic sets can be significantly "smaller" than Γ_s itself: countable sets

$$\mathfrak{C}_s=\{(\mathbf{q}_n,\chi_{\mathbf{q}_n}^{(s)})\times(\mathbf{p}_n,\chi_{\mathbf{p}_n}^{(s^{-1})})\}_{n=1}^{\infty}$$

for which $\{\mathbf{q}_n,\mathbf{p}_n\}_{n=1}^{\infty}$ have an accumulation point, or sets like $(\mathbf{q},\chi_{\mathbf{q}}^{(s)})\times\mathbb{R}^3_{s^{-1}}$ specified by a fixed $\mathbf{q}\in\mathbb{R}^3$. Contrast, however, this last case with the asymptotic one when we measure for each outgoing direction ω_{out} the probability density in the momentum at all points in $\mathbb{R}^3_{s^{-1}}$, and yet by (4.17) we still cannot determine ψ_+ beyond the values of $|\tilde{\psi}_+(\mathbf{q})|^2$!

Physically, this paradox can be understood in the light of the remarks made at the end of the last section: as opposed to momentum (which is conserved) the taking of the limit $t\to+\infty$ in the formulas (4.19), (5.17), or (5.19) reflects a physical situation in which the position detector has to be placed *at infinity* in order to detect the particle in the limit $t\to+\infty$. Otherwise, due to evanescence, the particle would be actually detected *before* an infinite period of time had elapsed, and therefore these formulas would not be applicable in a literal sense.

Looked upon in this light, the fact that in actual scattering experiments the scattered particle is detected in some finite time interval after the interaction had taken place is consistent with the above remarks only because some of the information *extractable* from such experiments is completely ignored: Usually the momentum-determination aspect (cf. Appendix) of such measurements is taken into account, while the information on position that is intrinsically gained by such experimental procedures is simply not considered.

The loss of information inherent in the conventional treatment of differential cross-section measurements is reflected in the well-known fact that such treatment, which, as seen from (5.20), yields only the absolute value of the T matrix, cannot be used to pinpoint the T operator itself in all cases. Yet, the experimental procedures themselves, when interpreted as simultaneous fuzzy measurements of $\mathbf{Q}$ and $\mathbf{P}$ (as outlined in the Appendix) can be used, at least in principle, to determine T to an arbitrary degree of accuracy. Indeed, if at the instant t_1 we prepare (in the Schrodinger picture) a state $U_{t_1}\psi$ and, after the collision has taken place, we carry out at time t_2 a Γ_s measurement that yields $|(U_{t_2}\psi)(\mathbf{q},\mathbf{p};s)|^2$, we have complete knowledge of the incoming and outgoing asymptotic states $U_t^{(0)}\psi_\pm$, $\psi=\Omega_\pm\psi_\pm$. As a matter of fact, $|(U_t\psi(\mathbf{q},\mathbf{p};s)|^2$ determines $(U_t\psi)(\mathbf{q},\mathbf{p};s)$ itself. Moreover, by (3.14)

$$\begin{aligned}&|(U_t^{(s)}\psi_\pm)(\mathbf{q},\mathbf{p};s)-(U_t\psi)(\mathbf{q},\mathbf{p};s)|\\&\quad\le\|U_t^{(s)}\psi_\pm-U_t\psi\|_s=\|(U_t^*U_t^{(0)}-\Omega_\pm)\psi_\pm\|_s,\end{aligned}\tag{6.1}$$

and we see that when (5.1) is valid $(U_t^{(0)}\psi_\pm)(\mathbf{q},\mathbf{p};s)$ can be chosen to be arbitrarily close to $(U_t\psi)(\mathbf{q},\mathbf{p};s)$ *uniformly* on all of Γ_s by choosing $-t_1$ and t_2 respectively, sufficiently large; naturally, the same statement, with $U_t^{(0)}\psi_\pm$ replacing $U_t^{(0)}\psi_\pm$, remains true in case of long-range interactions for which (5.5) holds.

The probability of actually determining for any ψ_- the vector $\psi_+=S\psi_-$ by Γ_s measurements confirms that such measurements do indirectly determine S. However, the absolute value $|S(\mathbf{q},\mathbf{p};\mathbf{q}',\mathbf{p}',s)|$ of the Γ_s representation (3.1) of S can be also directly inferred from such measurements since

$$|\psi_+(\mathbf{q},\mathbf{p};s)|^2=|S(\mathbf{q},\mathbf{p};\mathbf{q}_0,\mathbf{p}_0;s)|^2\tag{6.2}$$

if the incoming asymptotic state is so prepared that at $t=0$ it is represented by the coherent state $\phi^{(s)}_{\mathbf{q}_0,\mathbf{p}_0}(\mathbf{x})$. According to the argument leading to (2.9), at any fixed point $\mathbf{q}_0,\mathbf{p}_0\in\mathbb{R}^3$, we have

$$S(\mathbf{q},\mathbf{p};\mathbf{q}_0,\mathbf{p}_0;s)=\exp[-\tfrac14(s^{-2}\mathbf{q}^2+s^2\mathbf{p}^2)]S_{\mathbf{q}_0,\mathbf{p}_0}(\boldsymbol{\zeta}_s),\tag{6.3}$$

where $S_{\mathbf{q}_0,\mathbf{p}_0}$ is analytic in $\boldsymbol{\zeta}_s\in\mathbb{C}^3$. Hence the knowledge of (6.2) on any set of values $(\mathbf{q},\mathbf{p})\in\mathbb{R}^3$ which is open in $\mathbb{R}^6$ (and for which, therefore, the corresponding set of values $\boldsymbol{\zeta}_s\in\mathbb{C}^3$ is open in $\mathbb{C}^3$) determines (6.3) up to an unessential constant c of absolute value one.

Thus, our final conclusion is that the differential cross-section approach to fuzzy phase-space measurements leads to results which are in complete agreement with those based on (sharp or fuzzy) position or momentum measurements, but does not supply any additional data for pinpointing the scattering operator S; yet, the very same experimental procedures, when treated as fuzzy phase-space measurements performed at large but *finite* times before and after scattering, do supply, in principle, all the information necessary for the complete determination of S.

ACKNOWLEDGMENT

I am grateful to Dr. S. T. Ali for supplying me with a preprint of Ref. 7 before publication and for proofreading the manuscript.

APPENDIX: THE OPERATIONAL MEANING OF FUZZY POINTS IN PHASE SPACE

It has been recognized from the earliest days in the inception of modern quantum mechanics that while in the process of measurement on macrosystems, whose behavior is described by classical mechanics, the influence of the instrument on the system can be ignored in the majority of cases (and taken care of by a straightforward reduction of data procedure in the remaining cases) this would be an incorrect approach when a measurement on microsystems is performed. Carried to its ultimate logical conclusion this observation leads to a distinction between *preparatory* measurement and *determinative* measurement.[12,13,14] Roughly speaking, when the values ξ of the physical quantities X are *prepared* at the instant t that means the system at that instant "*has*" the values ξ for X; on the other hand, when we claim the values ξ' are *determined* at t we mean that the system "*would have had*" the values ξ for X *if there had been no disturbance* caused by its interaction with the measuring instrument. In the first approximation[13] this translates into the operational request that when a preparatory measurement of X at t yields ξ, an immediately following determinative measurement of X at $t+\epsilon$ should reproduce ξ (cf. Reproducibility Principle, Ref. 13, pp. 10, 13).

In order to deal fully with quantum measurement, the preceding operational interpretation has to be refined by introducing "fuzzy values" (ξ,χ_ξ) as descriptions of the outcome of measurements. In cases of the simultaneous measurement of $\mathbf{Q}$ and $\mathbf{P}$ such an analysis leads to the following interpretation[14] of the statement that "the fuzzy value $(\mathbf{q},\chi_{\mathbf{q}})\times(\mathbf{p},\chi'_{\mathbf{p}})$ of $(\mathbf{Q},\mathbf{P})$ has been measured": If the measurement is preparatory, then an immediately following determinative measurement of $\mathbf{Q}$ should yield the value $\mathbf{x}$ with the probability density $\chi_{\mathbf{q}}(\mathbf{x})$, while a following determinative measurement of $\mathbf{P}$ should yield (assuming in either case that the particle had not interacted in the meantime with anything else other than the apparatus) the value $\mathbf{k}$ with the probability density $\chi'_{\mathbf{p}}(\mathbf{k})$. Conversely, if the measurement is determinative and immediately prior to it a perfectly accurate preparatory measurement of $\mathbf{Q}$ had been performed, the probability density that the $\mathbf{Q}$-measurement had prepared $\mathbf{x}$ is $\chi_{\mathbf{q}}(\mathbf{x})$; similarly, the probability density that a prior measurement of $\mathbf{P}$ had prepared the value $\mathbf{k}$ is given by $\chi_{\mathbf{p}}(\mathbf{k})$. We note that this interpretation de facto specifies an operational procedure for calibrating an instrument $\mathcal{I}(\mathbf{Q},\mathbf{P})$ used in the simultaneous measurement of $\mathbf{Q}$ and $\mathbf{P}$ provided we already have perfectly precise [or, practically speaking, "very" precise compared to $\mathcal{I}(\mathbf{Q},\mathbf{P})$] instruments $\mathcal{I}(\mathbf{Q})$ and $\mathcal{I}(\mathbf{P})$ for measuring $\mathbf{Q}$ and $\mathbf{P}$ separately.

It is quite easy to present the main features for the blueprint of an instrument $\mathcal{I}(\mathbf{Q},\mathbf{P})$ used in the simultaneous measurement of $\mathbf{Q}$ and $\mathbf{P}$ of a charged particle. Actually, all the main ingredients of such an instrument are already present in many standard apparatuses for measuring $\mathbf{P}$, the only features still lacking in such cases being the accuracy calibration for both $\mathbf{Q}$ and $\mathbf{P}$. Namely, these ingredients are a generator of a homogeneous magnetic field **H**, and the facilities for measuring three *consecutive* fuzzy values $(\mathbf{x}_A, \chi_A)$, $(\mathbf{x}_B, \chi_B)$ and $(\mathbf{x}_C, \chi_C)$ of the position of a particle travelling in the region where this field **H** is contained. Such a set-up can be used then for both preparatory and determinative measurements of $(\mathbf{Q}, \mathbf{P})$, but the obtained data are used in different manners in the two cases: For preparatory measurement we take $\mathbf{q} = \mathbf{x}_C$, $\chi_\mathbf{q} \equiv \chi_C$, and $\mathbf{p} = \mathbf{p}_C$, while for determinative measurement we take $\mathbf{q} = \mathbf{x}_A$, $\chi_\mathbf{q} \equiv \chi_A$, and $\mathbf{p} = \mathbf{p}_A$. Here $|\mathbf{p}_A| = |\mathbf{p}_C| = e|\mathbf{H}|R$, where e is the charge of the particle and R is the radius of the circle passing through $\mathbf{x}_A$, $\mathbf{x}_B$, and $\mathbf{x}_C$, with the vectors $\mathbf{p}_A$ and $\mathbf{p}_C$ being tangential to this circle at the points $\mathbf{x}_A$ and $\mathbf{x}_C$, respectively, and pointing in the direction of motion; furthermore, the position measurements A, B, and C should be treated as preparatory in the first case and as determinative in the second case.

The confidence function $\chi_\mathbf{p}$ can be obtained by an accuracy calibration based, as described earlier, on very accurate measurements of **P** prior (when $\mathbf{p} = \mathbf{p}_A$) or after (when $\mathbf{p} = \mathbf{p}_C$) the measurement yielding **p**. However, $\chi_\mathbf{p}$ can be also inferred from the accuracy calibration of the instruments used in determining $\mathbf{x}_A$, $\mathbf{x}_B$, and $\mathbf{x}_C$ (provided **H** is perfectly homogeneous). We have

$$\chi_\mathbf{p}(\mathbf{k}) = \int_{D(\mathbf{k})} \chi_A(\mathbf{x})\chi_B(\mathbf{y})\chi_C(\mathbf{z})\, d\mathbf{x}\, d\mathbf{y}\, d\mathbf{z},$$

where $D(\mathbf{k})$ is the set of all points $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \mathbb{R}^9$ which lie on any circle of radius $|\mathbf{k}|e^{-1}|\mathbf{H}|^{-1}$ and for which $\mathbf{k} = k\mathbf{r}_x$, where $\mathbf{r}_x$ is the unit tangential vector to the circle at **x**.

In conclusion, we emphasize that in the determinative measurement of $(\mathbf{Q}, \mathbf{P})$, $\mathbf{x}_A$, $\mathbf{x}_B$, and $\mathbf{x}_C$ are the *determined* position vectors, i.e., they describe the points A, B, and C where the particle would have been if the disturbances caused by these three measurements were negligible. Thus, if, for example, A, B, C were obtained[12] in each case by observing through a microscope the direction in which a photon of given momentum had bounced off the particle, the momentum of the photon after the collision has to be measured in order to reduce the obtained data by discounting the change the photon had caused in the momentum of the particle. The natural by-products of such a consideration are the confidence functions χ_A, χ_B, and χ_C: In the case when photons are playing the role of microdetectors $\chi_A, \ldots, \chi_C$ result from the standard type of arguments[12] used in analysing gedanken experiments in the context of the uncertainty relations.

*Supported in part by the National Research Council of Canada.

[1]E. Prugovečki, J. Math. Phys. **17**, 517 (1976).
[2]E. Prugovečki, *Quantum Mechanics in Hilbert Space* (Academic, New York, 1971).
[3]J.R. Klauder and E.C.G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968).
[4]D.J. Neuman and H.S. Shapiro, Bull. Amer. Math. Soc. **72**, 971 (1966).
[5]V. Bargmann, Comm. Pure Appl. Math. **24**, 187 (1961).
[6]Yu. V. Geronimus and M.Y. Tseyllin, *Tables of Integrals, Series and Products* (Academic, New York, 1965), pp. 307, 930.
[7]S.T. Ali and H.D. Doebner, J. Math. Phys. **17**, 1105 (1976).
[8]J.D. Dollard, Comm. Math. Phys. **12**, 193 (1969).
[9]E. Prugovečki and J. Zorbas, J. Math. Phys. **14**, 1398 (1973); Erratum, *ibid.* **15**, 268 (1974).
[10]E.B. Manoukian and E. Prugovečki, Can. J. Phys. **49**, 102 (1971).
[11]J.R. Taylor, *Scattering Theory* (Wiley, New York, 1972).
[12]E. Prugovečki, Can. J. Phys. **45**, 2173 (1967).
[13]E. Prugovečki, Found. Phys. **3**, 3 (1973).
[14]E. Prugovečki, Found. Phys. **5**, 557 (1975).

Quantum mechanical soft springs and reverse correlation inequalities

Richard S. Ellis*

Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01002

Charles M. Newman†

Departments of Mathematics and Physics, Technion-Israel Institute of Technology, Haifa, Israel
(Received 31 March 1976)

Various properties of one-dimensional Schrödinger operators with "soft spring" potentials are derived as a consequence of the fact that the GHS and other correlation inequalities are reversed for certain general Ising modules.

We consider the Hamiltonian, $H_V = -d^2/dx^2 + V(x)$, of a one-dimensional quantum mechanical particle under the influence of a "spring" force, $-dV/dx$. In classical mechanics, a distinction is sometimes drawn between the qualitatively different motions due to "hard springs," where d^2V/dx^2 is increasing in $|x|$, and "soft springs," where d^2V/dx^2 is decreasing in $|x|$ (see Ref. 1, Chap. II). The introduction of statistical mechanical techniques into constructive quantum field theory in recent years (see Ref. 2) has led to some interesting "spin-off" results concerning quantum mechanical hard springs, and the main purpose of this paper is to give the corresponding results for soft springs.

Our soft spring potentials will be real-valued functions in the class

$$\mathcal{V}_s = \{V \mid V(x) = \text{const} + \int_0^x G(y)\,dy \quad \text{with } G(y) = -G(-y)\ \forall\, y,$$
$$G \text{ concave on } [0,\infty), \text{ and } a_V \equiv \lim_{y\to\infty} G(y) > 0\}. \tag{1}$$

We further define

$$\exp(-\mathcal{V}_s) = \{f \mid f = \exp(-V) \text{ for some } V \in \mathcal{V}_s\}. \tag{2}$$

For $-a_V < a < a_V$, $H_V - ax$ [considered as an operator on $L^2(\mathbb{R}, dx)$] has nondegenerate eigenvalues which we list in increasing order as $E_0(a) < E_1(a) < \cdots$ and a normalized ground state Ω^a $[(H_V - ax)\Omega^a = E_0(a)\Omega^a]$ which we choose to be positive.

Theorem 1: Suppose $V \in \mathcal{V}_s$. Then

$$M(a) \equiv (\Omega^a, x\Omega^a) \text{ is convex on } [0, a_V), \tag{3}$$

$$E_1(a) - E_0(a) \text{ is nonincreasing on } [0, a_V), \tag{4}$$

$$E_1(0) - E_0(0) \geq E_2(0) - E_1(0), \tag{5}$$

$$U \in \exp(-\mathcal{V}_s) \Rightarrow \exp(-tH_V)U \in \exp(-\mathcal{V}_s), \quad \text{for } t \geq 0, \tag{6}$$

$$\Omega^0 \in \exp(-\mathcal{V}_s). \tag{7}$$

Remark 2: In the case of hard spring potentials, the analogues of (3) and (4) were first derived in Ref. 3 and the analog of (5) in Ref. 4 for V a quartic polynomial. These three results were then extended to a larger class of V's in Refs. 5 and 6, and finally to all hard spring potentials in Refs. 7 and 8. The analogues of (6) and (7) for hard springs are given in Ref. 8. We do not include a proof of Theorem 1 since (3)–(7) follows from the "reverse" correlation inequalities [in particular (14) and (15)] given below for certain general Ising models in exactly the same way as the hard spring results follow from the usual GHS and Lebowitz inequalities (see Ref. 9, Chap. IX, and Ref. 8 for details).

Remark 3: Property (6) for H_V can be expressed in terms of the diffusion process determined by H_V, exactly as was done for the analogous hard spring result in Ref. 8. Property (5) and its hard spring analog suggest some general relation between convexity properties of V and those of the spectrum of H_V. In particular, we suggest the existence of a natural class of V's for which $E_{j+1}(0) - E_j(0)$ is nonincreasing (resp., nondecreasing) in j.

A general Ising model (with pair interactions) is a collection of "spin" random variables, $\{X_i: i = 1, \ldots, N\}$, with joint probability distribution,

$$Z^{-1}\exp\left(\sum_{i=1}^N h_i x_i + \sum_{i,j=1}^N J_{ij}x_i x_j\right)\prod_{i=1}^N \rho_i(dx_i), \tag{8}$$

where each ρ_i is a measure in $\mathcal{E}$, the set of even Borel measures ρ on $\mathbb{R}$ such that $\int \exp(kx^2)\rho(dx) < \infty$ for some $k > 0$, where Z is chosen so that (8) is a probability measure, and where the J_{ij}'s are real and so small that Z is finite for all real h_i's. We shall always assume that $J_{ij},\ h_i \geq 0$ for all i, j.

In order to discuss our correlation inequalities, we consider four independent copies, $\{X_i^\alpha\}$ $(\alpha = 1, \ldots, 4)$, of the $\{X_i\}$ and define $T_i = (X_i^1 + X_i^2)/\sqrt{2}$, $Q_i = (X_i^1 - X_i^2)/\sqrt{2}$, $W_i^\alpha = \sum_{\beta=1}^4 A_{\alpha\beta}X_i^\beta$, and $Y_i^\alpha = \sum_{\beta=1}^4 B_{\alpha\beta}X_i^\beta$, where A and B are the following 4×4 matrices:

$$A = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$
$$B = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 \end{bmatrix}. \tag{9}$$

Given a finite measure ρ on $\mathbb{R}$ and an invertible 4×4 matrix T, we define $\boldsymbol{\rho}(d\mathbf{x})$ as $\prod_{\alpha=1}^4 \rho(dx^\alpha)$ and $\boldsymbol{\rho}_T(d\mathbf{x})$ as $\boldsymbol{\rho}(d[T^{-1}\mathbf{x}])$, where $\mathbf{x} = (x^1, \ldots, x^4) \in \mathbb{R}^4$. We define $\mathcal{G}_+$ (resp., $\mathcal{G}_-$) as the set of measures ρ in $\mathcal{E}$ such that for

$$\mu = \boldsymbol{\rho}_A \text{ (resp., } \boldsymbol{\rho}_B)$$
$$\int_{\mathbb{R}^4} (x^1)^{l^1} \cdots (x^4)^{l^4} \mu(d\mathbf{x}) \geq 0,$$
$$\forall\ l^1, \ldots, l^4 = 0, 1, 2, \cdots. \tag{10}$$

We also define $\mathcal{G}_s$ (resp., $\mathcal{G}_h$) as the set of finite even measures ρ such that $\boldsymbol{\rho}_A \geq \boldsymbol{\rho}_B$ (resp., $\boldsymbol{\rho}_B \geq \boldsymbol{\rho}_A$) on $\mathbb{R}^4_+ = \{\mathbf{x} : \mathbf{x}^\alpha > 0,\ \forall\ \alpha\}$. (Note that in Ref. 8, $\mathcal{G}_h$ is denoted by $\mathcal{G}$.) We denote a multi-index $(m_1, \ldots, m_N)$ by m, $\prod_{i=1}^N X_i^{m_i}$ by X^m, an expectation $E(H)$ by $\langle H\rangle$, and $\langle X_{i_1} \cdots X_{i_k}\rangle$ by $\langle i_1 \cdots i_k\rangle$.

Theorem 4: A measure ρ in $\mathcal{E}$ belongs to $\mathcal{G}_+$ if it belongs to $\mathcal{G}_s$. If each ρ_i in (8) belongs to $\mathcal{G}_+$, then for any multi-indices $m^1, \ldots, m^4, m, n$, and any $i_1, \ldots, i_4$,

$$\left\langle \prod_{\alpha=1}^4 (W^\alpha)^{m^\alpha} \right\rangle \geq 0, \tag{11}$$

$$\langle Q^m Q^n\rangle - \langle Q^m\rangle\langle Q^n\rangle \geq 0, \tag{12}$$

$$\langle T^m Q^n\rangle - \langle T^m\rangle\langle Q^n\rangle \geq 0, \tag{13}$$

$$\langle i_1 i_2 i_3\rangle - \langle i_1\rangle\langle i_2 i_3\rangle - \langle i_2\rangle\langle i_1 i_3\rangle - \langle i_3\rangle\langle i_1 i_2\rangle + 2\langle i_1\rangle\langle i_2\rangle\langle i_3\rangle \geq 0, \tag{14}$$

$$\langle i_1 i_2 i_3 i_4\rangle - \langle i_1 i_2\rangle\langle i_3 i_4\rangle - \langle i_1 i_3\rangle\langle i_2 i_4\rangle - \langle i_1 i_4\rangle\langle i_2 i_3\rangle \geq 0, \tag{15}$$

when

$$h_j = 0,\ \forall\ j.$$

Remark 5: These results are the "reverse" of the usual correlation inequalities which were originally proved when each $\rho_i(dx) = \delta(x-1) + \delta(x+1)$ in Refs. 10, 11, and 12 and then extended to measures in $\mathcal{G}_-$ in Refs. 5–7. In the case of $\mathcal{G}_-$, the direction of the inequalities (13)–(15) is changed to give the usual GHS and Lebowitz inequalities, W^α in (11) is replaced by Y^α to give the usual Ellis–Monroe inequality, and (12) remains the same.

Proof: The proof is essentially identical to that of the usual inequalities as given in Refs. 6 and 7. We only note that in deriving (12)–(13) from (11), use must be made of the fact that the sign of any two of the bottom three rows of A may be changed without altering the validity of (11).

The next theorem completely characterizes measures in $\mathcal{G}_s$ and is analogous to the characterization of $\mathcal{G}_h$ given in Ref. 8.

Theorem 6: For a finite, even, not identically zero, Borel measure ρ on $\mathbb{R}$, the following three statements are equivalent:

(i) Either $\rho(dx) = C\delta(x)$ for some $C > 0$, or else $\rho(dx) = f(x)\,dx$ for some $f \in \exp(-\mathcal{L}_s)$,

(ii) $\rho \in \mathcal{G}_s$,

(iii) For any $b > 0$

$$\left(\frac{d^3}{dh^3}\right) \ln \int_{-\infty}^{\infty} \exp(hx - bx^2)\rho(dx) \geq 0 \quad \text{for } h \geq 0. \tag{16}$$

Proof: The proof of Theorem 2.4 of Ref. 8 directly yields that (i) $\Rightarrow$ (ii) and (ii) $\Rightarrow$ (iii), and it reduces the proof of (iii) $\Rightarrow$ (i) to showing that if $\rho_b \to \rho$ weakly with $\rho_b(dx) = f_b(x)\,dx$ and $f_b \in \exp(-\mathcal{L}_s)$, then ρ must be as in (i). This latter fact is easily derived by using the proofs of Lemmas 4.6 and 4.10 of Ref. 8.

*On leave from Northwestern University, Evanston, Illinois. Research partially supported by NSF Grant MPS 71-02838 A04.

†NATO postdoctoral fellow on leave from Department of Mathematics, Indiana University, Bloomington, Indiana. Research partially supported by NSF Grant MPS 74-04870 A01.

[1]J. J. Stoker, *Nonlinear Vibrations in Mechanical and Electrical Systems* (Interscience, New York, 1950).
[2]G. Velo and A. S. Wightman, Eds., *Constructive Quantum Field Theory* (Springer-Verlag, Berlin, 1973).
[3]B. Simon and R. B. Griffiths, Commun. Math. Phys. **33**, 145–64 (1973).
[4]J. Glimm, A. Jaffe, and T. Spencer in G. Velo and A. S. Wightman, Eds., *Constructive Quantum Field Theory* (Springer-Verlag, Berlin, 1973).
[5]R. S. Ellis, Bull. Am. Math. Soc. **81**, 925–29 (1975).
[6]G. Sylvester, "Inequalities for Continuous Spin Ising Ferromagnets," J. Stat. Phys. (to appear).
[7]R. S. Ellis, J. L. Monroe, and C. M. Newman, Commun. Math. Phys. **46**, 167–82 (1976).
nets," *Commun. Math. Phys.* (to appear).
[8]R. S. Ellis and C. M. Newman, University of Massachusetts/Technion preprint (1976).
[9]B. Simon, *The $P(\phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U. P., Princeton, N. J., 1974).
[10]R. B. Griffiths, C. A. Hurst, and S. Sherman, J. Math. Phys. **11**, 790–95 (1970).
[11]J. L. Lebowitz, Commun. Math. Phys. **35**, 87–92 (1974).
[12]R. S. Ellis and J. L. Monroe, Commun. Math. Phys. **41**, 33–38 (1975).

Binary mixture with nearest and next nearest neighbor interaction on a one-dimensional lattice

C. C. Yan

Instituto de Física, Universidade Federal de Rio de Janeiro, Rio de Janeiro, Brazil
(Received 27 February 1976; revised manuscript received 27 April 1976)

A mixture of two kinds of molecules on a one-dimensional lattice with free ends is considered. The energy term is assumed to consist of interactions of nearest neighbors and next nearest neighbors and interaction with the uniform external field. The partition function is evaluated by determining the degeneracies.

I. INTRODUCTION

The study of one dimensional systems is perhaps mainly motivated by the mathematical tractability.[1] However, one justification is that such study may cast some light on the more realistic two- or three-dimensional systems. In a lattice model, the lattice sites are enumerable. Hence two- or three-dimensional lattice models can always be mapped onto and represented by one-dimensional lattice models. In the process of mapping, the originally nearest neighbors in the two- or three-dimensional lattice may appear as distant neighbors in the corresponding one-dimensional lattice. This suggests that if farther and farther neighbor interactions are included in the one-dimensional lattice models, the results may not only cast light on but also ultimately become the exact results of the higher-dimensional models. As a first step we shall investigate one-dimensional lattice model with nearest and next nearest neighbor interactions.

In a previous paper[2] we have proposed a method applicable to multicomponent as well as binary systems to determine degeneracies associated with nearest neighbor interactions on a one-dimensional lattice. In the following, we shall generalize the method to include next nearest neighbor interactions while confining ourselves to binary system for simplicity.

II. PRELIMINARY ANALYSIS

The energy associated with the molecules interacting with nearest neighbors and next nearest neighbors on a one-dimensional lattice in a uniform external field can be written as

$$E=\sum_i v_i n_i+\tfrac{1}{2}\sum_{i,j}(1+\delta_{ij})v_{ij}n_{ij}+\tfrac{1}{2}\sum_{i,j}(1+\delta_{ij})u_{ij}m_{ij}, \tag{2.1}$$

where $i,j=1,2,\ldots,k$ indicate different kinds of molecules and runs through these values independently in the summations, n_i is the number of ith kind of molecules, n_{ij} and m_{ij} are respectively the nearest and next nearest neighbor pairs between ith and jth kinds, and v_i, v_{ij}, u_{ij} are the associated interaction energies.

The partition function corresponding to a definite set of $\{n_i\}$ is given by

$$Z(\{n_i\})=\sum_{\{n_{ij}\}\{m_{ij}\}} M(\{n_i\},\{n_{ij}\},\{m_{ij}\})\exp(-\beta E), \tag{2.2}$$

where $M(\{n_i\},\{n_{ij}\},\{m_{ij}\})$ is the multiplicity or degeneracy corresponding to the distributions specified by the sets $\{n_i\}$, $\{n_{ij}\}$, and $\{m_{ij}\}$. If we wish to consider the restriction due to the finite size of the molecules, we shall incorporate the following condition

$$\sum_i l_i n_i = L, \tag{2.3}$$

where L is the total number of lattice sites and l_i is the number of sites each ith kind of molecule would occupy. Summing over the set $\{n_i\}$ in accordance with (2.3), we obtain

$$Z(L)=\sum_{\{n_i\}} Z(\{n_i\}). \tag{2.4}$$

Once the partition functions are known the thermodynamic quantities of interest can be calculated. It is clear that the knowledge of the degeneracy M is essential to the evaluation of the partition functions. Thus the main task is the determination of the degeneracy M. For simplicity, we shall limit ourselves to the case of binary mixture so that i,j take the values of 1 and 2 only. In such case, a state with fixed n_1 and n_2 can be defined by

$$|\phi\rangle=|n_{11},n_{12},n_{22},m_{11},m_{12},m_{22}\rangle \tag{2.5}$$

corresponding to a definite energy E and associated multiplicity M. When $v_{11}=v_{22}=-v_{12}$, $u_{11}=u_{22}=-u_{12}$, the model considered here reduces to the well known Ising model.[3]

As a preliminary step, we shall place all the molecules of the first kind in a row, thus creating (n_1-1) numbers of nearest neighbor pairs and (n_1-2) numbers of next nearest neighbor pairs. This initial state can be written as

$$|\phi_0\rangle=|n_1-1,0,0,n_1-2,0,0\rangle \tag{2.6}$$

with $E=v_1n_1+v_2n_2+v_{11}(n_1-1)+u_{11}(n_1-2)$ and $M=1$. Our problem is now reduced to the proper choice among the ends and intervals to place properly the second kind of molecules in accordance with the specified set of $\{n_{ij}\}$ and $\{m_{ij}\}$. By ends and intervals, we mean the following: Ends are the space left to the leftmost molecule of the first kind or the space right to the rightmost molecule of the first kind. The first left (right) interval is the space between the leftmost (rightmost) molecule of the first kind and the next molecule of the first kind. Interior intervals are the spaces between the nearest neighbor pairs of the first kind of molecules. First intervals and interior intervals are to be called simply intervals when no distinction is needed. To keep track of the changes of states brought

FIG. 1. Representatives of the arrangements corresponding to $A_2BC^3D^2E^2P$ with $n_1=12$, $n_2=10$. The dots represent the first kind of molecules while the bars represent the second kind of molecules.

about by the process to be described in the next section, we now introduce the following:

A_1: one end is exactly singly occupied by the second kind of molecule and has the following operational effect

$$A_1|a,b,c,d,e,f\rangle=|a,b+1,c,d,e+1,f\rangle. \qquad (2.7)$$

A_2: one end is at least doubly occupied and operationally

$$A_2|a,b,c,d,e,f\rangle=|a,b+1,c+1,d,e+2,f\rangle. \qquad (2.8)$$

B: one first interval is at least singly occupied and the end next to it is not occupied. Operationally,

$$B|a,b,c,d,e,f\rangle=|a-1,b+2,c,d,e+1,f\rangle. \qquad (2.9)$$

C: one of the interior intervals or one of the first intervals with occupied end next to it is at least singly occupied. Operationally,

$$C|a,b,c,d,e,f\rangle=|a-1,b+2,c,d-1,e+2,f\rangle \qquad (2.10)$$

D: one interval, be it first interval or interior interval, exclusive of the ends, is at least doubly occupied. Operationally,

$$D|a,b,c,d,e,f\rangle=|a,b,c+1,d-1,e+2,f\rangle. \qquad (2.11)$$

When any given end or interval contains more than two molecules of the second kind, we say there are extras. Then

E: one extra, operationally,

$$E|a,b,c,d,e,f\rangle=|a,b,c+1,d,e,f+1\rangle. \qquad (2.12)$$

P: one pair of occupied nearest neighbor intervals inclusive of the pair formed from the end and the first interval next to it. Operationally,

$$P|a,b,c,d,e,f\rangle=|a,b,c,d+1,e-2,f+1\rangle. \qquad (2.13)$$

In describing a given situation, we apply all the operators that are compatible. Thus, for instance, in a given situation if we find that the left end has one molecule of the second kind and an interior interval has five molecules of the second kind, we would then say the following: One end is exactly singly occupied, one interior interval is said to be at least singly occupied, and also at least doubly occupied, and also has three extras. Hence, we write A_1CDE^3. Now suppose we move all the five molecules belonging to this interior interval to the first left interval, we create one pair of occupied nearest intervals. Since the left end is occupied, we continue to identify the first left interval as C and would now rewrite as A_1CDE^3P. Next, if we move all the five molecules to the first right interval, the pair will be destroyed and the first right interval has to be identified as B as the right end is not occupied. Hence we rewrite as A_1BDE^3. Now if all the five molecules are moved into the right end, we say the following: One end is exactly singly occupied; one end is at least doubly occupied with three extras. Hence we rewrite as $A_1A_2E^3$. Lastly, if all the five molecules are moved into the left end, we would say that one end is at least doubly occupied with four extras. Hence we rewrite as A_2E^4. The meaning of the operators shall become clearer by examining the following example and the accompanying illustrative figure.

Example: When we write $A_2BC^3D^2E^2P$, it means that we may find in the arrangement the following: One end is at least doubly occupied; the first interval near the other end is at least singly occupied; three interior intervals are at least singly occupied; two intervals are at least doubly occupied; plus two extras and one pair of nearest neighbor intervals. The total number of molecules of the second kind is 10, i.e., $n_2=10$.

To be more concrete, we shall take $n_1=12$. Then, two of the possible representations of the above description are shown in Fig. 1.

Operationally, we have

$$A_2BC^3D^2E^2P|11,0,0,10,0,0\rangle=|7,9,5,6,11,3\rangle. \qquad (2.14)$$

It is easy to verify that the two representatives shown in Fig. 1 have the same number of pairs as defined in (2.14)

III. DETERMINATION OF THE DEGENERACY

The determination of the degeneracy will be done by the following procedures and we shall assume, without loss of generality, that $n_1>n_2$.

(*A*) All the n_1 molecules of the first kind are placed in a row so that an initial state is created:

$$|\phi_0\rangle=|n_{11},n_{12},n_{22},m_{11},m_{12},m_{22}\rangle$$
$$=|n_1-1,0,0,n_1-2,0,0\rangle. \qquad (3.1)$$

(*B*) The second kind of molecules are placed into the ends or the first intervals so as to result in the following boundary conditions:

$$(1)\colon \{0,0;0,0\},\quad (2)\colon \{0,1;0,0\},\{0,0;1,0\},$$
$$(3)\colon \{0,1;10\},\quad (4)\colon \{1,0;0,0\},\{0,0;0,1\},$$
$$(5)\colon \{1,0;1,0\},\{0,1;0,1\},\quad (6)\colon \{1,0;0,1\}, \qquad (3.2)$$
$$(7)\colon \{2,0;0,0\},\{0,0;0,2\},\quad (8)\colon \{2,0;1,0\},\{0,1;0,2\},$$
$$(9)\colon \{2,0;0,1\},\{1,0;0,2\},\quad (10)\colon \{2,0;0,2\},$$

where $\{\alpha,\beta;\gamma,\delta\}$ has the following meaning: The left end is occupied by α numbers of molecules of the second kind, the first left interval is occupied by β number of molecules of the second kind, the first right interval is occupied by γ number of molecules of the second kind, and the right end is occupied by δ number of molecules of the second kind.

The above boundary conditions can be represented by the following operators:

$$(1)\colon 1,\quad (2)\colon 2B,\quad (3)\colon B^2,\quad (4)\colon 2A_1,\quad (5)\colon 2A_1B,$$
$$(6)\colon A_1^2,\quad (7)\colon 2A_2,\quad (8)\colon 2A_2B,\quad (9)\colon 2A_2A_1,\quad (10)\colon A_2^2, \qquad (3.3)$$

where the factor 2 is the double degeneracy corresponding to the right left symmetry.

(C) x internal intervals be selected to place x number of molecules of the second kind so that y number of pairs of occupied nearest neighbor intervals is created. This can be represented by the operator

$$C^x P^y. \tag{3.4}$$

To determine the degeneracy resulting from this process, one may regard the ends and intervals as lattice sites and the unoccupied sites as the first kind of molecules while the occupied sites as the second kind of molecules. Then the following can be obtained[2,4] with M_i' denoting the number of ways (multiplicity) to change the state defined by the previous step (B) into the state defined by the present step (C) in accordance with the ith boundary condition.

$$\begin{aligned}
&(1): M_1' = \binom{x-1}{y}\binom{n_1-2-x}{x-y}, \quad (2): M_2' = \binom{x}{y}\binom{n_1-3-x}{x-y},\\
&(3): M_3' = \binom{x+1}{y}\binom{n_1-4-x}{x-y}, \quad (4): M_4' = \binom{x}{y}\binom{n_1-2-x}{x-y},\\
&(5): M_5' = \binom{x+1}{y}\binom{n_1-3-x}{x-y}, \quad (6): M_6' = \binom{x+1}{y}\binom{n_1-2-x}{x-y},\\
&(7): M_7' = \binom{x}{y}\binom{n_1-2-x}{x-y}, \quad (8): M_8' = \binom{x+1}{y}\binom{n_1-3-x}{x-y},\\
&(9): M_9' = \binom{x+1}{y}\binom{n_1-2-x}{x-y}, \quad (10): M_{10}' = \binom{x+1}{y}\binom{n_1-2-x}{x-y}.
\end{aligned} \tag{3.5}$$

By definition, x and y are positive integers. Here and in the following the convention

$$\binom{P}{Q} = 0 \quad \text{whenever } P<0,\ Q<0, \text{ or } P<Q \tag{3.6}$$

is adopted.

(D) z intervals be selected from among the by now singly occupied intervals, excluding the ends, to place z molecules of the second kind so that each of these z intervals will be doubly occupied. The corresponding operator is

$$D^z. \tag{3.7}$$

The degeneracies resulted from this process can be easily found to be

$$\begin{aligned}
&(1): M_1'' = \binom{x}{z}, \quad (2): M_2'' = \binom{x+1}{z}, \quad (3): M_3'' = \binom{x+2}{z},\\
&(4): M_4'' = \binom{x}{z}, \quad (5): M_5'' = \binom{x+1}{z}, \quad (6): M_6'' = \binom{x}{z},\\
&(7): M_7'' = \binom{x}{z}, \quad (8): M_8'' = \binom{x+1}{z}, \quad (9): M_9'' = \binom{x}{z},\\
&(10): M_{10}'' = \binom{x}{z}.
\end{aligned} \tag{3.8}$$

(E) Place whatever the number of molecules of the second kind still unused into the by-now exactly doubly occupied ends and intervals in a manner of Bose–Einstein distribution so that those exactly doubly occupied ends and intervals will become at least doubly occupied.

It is clear that the number of unused molecules of the second kind depends on the boundary conditions. Hence the associated operators and the degeneracies will also depend on the boundary conditions and are given by

$$\begin{aligned}
&(1): E^{(n_2-x-z)};\ M_1''' = \binom{n_2-x-1}{z-1},\\
&(2): E^{(n_2-x-z-1)};\ M_2''' = \binom{n_2-x-2}{z-1},\\
&(3): E^{(n_2-x-z-2)};\ M_3''' = \binom{n_2-x-3}{z-1},\\
&(4): E^{(n_2-x-z-1)};\ M_4''' = \binom{n_2-x-2}{z-1},\\
&(5): E^{(n_2-x-z-2)};\ M_5''' = \binom{n_2-x-3}{z-1},\\
&(6): E^{(n_2-x-y-2)};\ M_6''' = \binom{n_2-x-3}{z-1},\\
&(7): E^{(n_2-x-z-2)};\ M_7''' = \binom{n_2-x-2}{z},\\
&(8): E^{(n_2-x-z-3)};\ M_8''' = \binom{n_2-x-3}{z},\\
&(9): E^{(n_2-x-z-3)};\ M_9''' = \binom{n_2-x-3}{z},\\
&(10): E^{(n_2-x-z-4)};\ M_{10}''' = \binom{n_2-x-3}{z+1}.
\end{aligned} \tag{3.9}$$

When the values of x, y, and z together with one of the ten boundary conditions are specified, a state is defined. The degeneracy corresponding to such a state is obtained by combining (3.5), (3.8), and (3.9) and is given by

$$M_i = \gamma_i (M_i')(M_i'')(M_i'''), \tag{3.10}$$

where i denotes the ith boundary condition and

$$\begin{aligned}
\gamma_i &= 1, \quad i = 1,3,6,10,\\
&= 2, \quad i = 2,4,5,7,8,9.
\end{aligned} \tag{3.11}$$

In order to evaluate the partition function, one needs to know not only the degeneracies but also the states and the associated energies. These can be obtained as follows:

(1):

$$\begin{aligned}
|\phi_1\rangle &= C^x P^y D^z E^{(n_2-x-z)} |\phi_0\rangle = |n_{11}, n_{12}, n_{22}, m_{11}, m_{12}, m_{22}\rangle\\
&= |n_1 - 1 - x, 2x, n_2 - x, n_2 - 2 - x + y - z, 2x - 2y + 2z,\\
&\quad n_2 - x + y - z\rangle.
\end{aligned} \tag{3.12}$$

If n_1, n_2, n_{22}, m_{22} are chosen as variables, x, y, and z can be solved in terms of them. The solutions for this case are

$$x = n_2 - n_{22}, \quad y = m_{22} - n_{22} + z, \quad z = z. \tag{3.13}$$

In terms of n_1, n_2, n_{22}, and m_{22}, one has

$$\begin{aligned}
|\phi_1\rangle = |&n_1 - n_2 - n_{22} - 1, 2(n_2 - n_{22}), n_{22}, n_1 - n_2 + m_{22} - 2,\\
&2(n_2 - m_{22}), m_{22}\rangle,
\end{aligned} \tag{3.14}$$

$$E_1 = E_c - (v_{11} + 2u_{11}), \tag{3.15}$$

where

$$E_c = (v_1 + v_{11} + u_{11})n_1 + (v_2 - v_{11} + 2v_{12} - u_{11} + 2u_{12})n_2 + (v_{11} - 2v_{12} + v_{22})n_{22} + (u_{11} - 2u_{12} + u_{22})m_{22} \quad (3.16)$$

and

$$M_1 = \sum_z \binom{n_2 - n_{22} - 1}{m_{22} - n_{22} + z}\binom{n_1 - n_2 + n_{22} - 2}{n_2 - m_{22} - z}\binom{n_2 - n_{22}}{z}\binom{n_{22} - 1}{z - 1}. \quad (3.17)$$

The summation of z over all positive integers arises owing to the fact that the final state is independent of z.

The rest of the cases can be similarly worked out. The results in terms of n_1, n_2, n_{22}, and m_{22} are

(2):

$$|\phi_2\rangle = |n_1 - n_2 + n_{22} - 1, 2(n_2 - n_{22}), n_{22}, n_1 - n_2 + m_{22} - 1, 2(n_2 - m_{22}) - 1, m_{22}\rangle, \quad (3.18)$$

$$E_2 = E_c - (v_{11} + u_{11} + u_{12}), \quad (3.19)$$

$$M_2 = \sum_z 2\binom{n_2 - n_{22} - 1}{m_{22} - n_{22} + z}\binom{n_1 - n_2 + n_{22} - 2}{n_2 - m_{22} - 1 - z}\binom{n_2 - n_{22}}{z}\binom{n_{22} - 1}{z - 1}. \quad (3.20)$$

(3):

$$|\phi_3\rangle = |n_1 - n_2 + n_{22} - 1, 2(n_2 - n_{22}), n_{22}, n_1 - n_2 + m_{22}, 2(n_2 - m_{22} - 1), m_{22}\rangle, \quad (3.21)$$

$$E_3 = E_c - (v_{11} + 2u_{12}), \quad (3.22)$$

$$M_3 = \sum_z \binom{n_2 - n_{22} - 1}{m_{22} - n_{22} + z}\binom{n_1 - n_2 + n_{22} - 2}{n_2 - m_{22} - 2 - z}\binom{n_2 - n_{22}}{z}\binom{n_{22} - 1}{z - 1}. \quad (3.23)$$

(4):

$$|\phi_4\rangle = |n_1 - n_2 + n_{22}, 2(n_2 - n_{22}) - 1, n_{22}, n_1 - n_2 + m_{22} - 1, 2(n_2 - m_{22}) - 1, m_{22}\rangle, \quad (3.24)$$

$$E_4 = E_c - (v_{12} + u_{11} + u_{12}), \quad (3.25)$$

$$M_4 = \sum_z 2\binom{n_2 - n_{22} - 1}{m_{22} - n_{22} + z}\binom{n_1 - n_2 + n_{22} - 1}{n_2 - m_{22} - 1 - z}\binom{n_2 - n_{22} - 1}{z} \times \binom{n_{22} - 1}{z - 1}. \quad (3.26)$$

(5):

$$|\phi_5\rangle = |n_1 - n_2 + n_{22}, 2(n_2 - n_{22}) - 1, n_{22}, n_1 - n_2 + m_{22}, 2(n_2 - m_{22} - 1), m_{22}\rangle, \quad (3.27)$$

$$E_5 = E_c - (v_{12} + 2u_{12}), \quad (3.28)$$

$$M_5 = \sum_z 2\binom{n_2 - n_{22} - 1}{m_{22} - n_{22} + z}\binom{n_1 - n_2 + n_{22} - 1}{n_2 - m_{22} - 2 - z}\binom{n_2 - n_{22} - 1}{z} \times \binom{n_{22} - 1}{z - 1}. \quad (3.29)$$

(6):

$$|\phi_6\rangle = |n_1 - n_2 + n_{22} + 1, 2(n_2 - n_{22} - 1), n_{22}, n_1 - n_2 + m_{22}, 2(n_2 - m_{22} - 1), m_{22}\rangle, \quad (3.30)$$

$$E_6 = E_c - (2v_{12} + 2u_{12} - v_{11}), \quad (3.31)$$

$$M_6 = \sum_z \binom{n_2 - n_{22} - 1}{m_{22} - n_{22} + z}\binom{n_1 - n_2 + n_{22}}{n_2 - m_{22} - 2 - z}\binom{n_2 - n_{22} - 2}{z} \times \binom{n_{22} - 1}{z - 1}. \quad (3.32)$$

(7):

$$|\phi_7\rangle = |n_1 - n_2 + n_{22}, 2(n_2 - n_{22}) - 1, n_{22}, n_1 - n_2 + m_{22}, 2(n_2 - m_{22} - 1), m_{22}\rangle, \quad (3.33)$$

$$E_7 = E_c - (v_{12} + 2u_{12}), \quad (3.34)$$

$$M_7 = \sum_z 2\binom{n_2 - n_{22} - 1}{m_{22} - n_{22} + 1 + z}\binom{n_1 - n_2 + n_{22} - 1}{n_2 - m_{22} - 1 - z}\binom{n_2 - n_{22} - 1}{z} \times \binom{n_{22} - 1}{z - 1}. \quad (3.35)$$

(8):

$$|\phi_8\rangle = |n_1 - n_2 + n_{22}, 2(n_2 - n_{22}) - 1, n_{22}, n_1 - n_2 + m_{22} + 1, 2(n_2 - m_{22}) - 3, m_{22}\rangle, \quad (3.36)$$

$$E_8 = E_c - (v_{12} + 3u_{12} - u_{11}), \quad (3.37)$$

$$M_8 = \sum_z 2\binom{n_2 - n_{22} - 1}{m_{22} - n_{22} + 1 + z}\binom{n_1 - n_2 - n_{22} - 1}{n_2 - m_{22} - 1 - z}\binom{n_2 - n_{22} - 1}{z} \times \binom{n_{22} - 1}{z - 1}. \quad (3.38)$$

(9):

$$|\phi_9\rangle = |n_1 - n_2 + n_{22} + 1, 2(n_2 - n_{22} - 1), n_{22}, n_1 - n_2 + m_{22} + 1, 2(n_2 - m_{22}) - 3, m_{22}\rangle, \quad (3.39)$$

$$E_9 = E_c - (2v_{12} + 3u_{12} - v_{11} - u_{11}), \quad (3.40)$$

$$M_9 = \sum_z 2\binom{n_2 - n_{22} - 1}{m_{22} - n_{22} + 1 + z}\binom{n_1 - n_2 + n_{22}}{n_2 - m_{22} - 1 - z}\binom{n_2 - n_{22} - 2}{z} \times \binom{n_{22} - 1}{z - 1}. \quad (3.41)$$

(10):

$$|\phi_{10}\rangle = |n_1 - n_2 + n_{22} + 1, 2(n_2 - n_{22} - 1), n_{22}, n_1 - n_2 + m_{22} + 2, 2(n_2 - m_{22} - 2), m_{22}\rangle, \quad (3.42)$$

$$E_{10} = E_c - (2v_{12} + 4u_{12} - v_{11} - 2u_{11}), \quad (3.43)$$

$$M_{10} = \sum_z \binom{n_2 - n_{22} - 1}{m_{22} - n_{22} + 2 + z}\binom{n_1 - n_2 + n_{22}}{n_2 - m_{22} - 4 - z}\binom{n_2 - n_{22} - 2}{z} \times \binom{n_{22} - 1}{z - 1}. \quad (3.44)$$

Finally the partition function is given by

$$Z(\{n_i\}) = \sum_{n_{22}, m_{22}} \left[\sum_{i=1}^{10} M_i \exp(-\beta E_i)\right]. \quad (3.45)$$

[1]Elliott H. Lieb and Daniel C. Mattis, *Mathematical Physics in One Dimension* (Academic, New York, 1966).
[2]C.C. Yan, J. Math. Phys. **17**, 69 (1976).
[3]E. Ising, Z. Physik **31**, 253 (1925).
[4]Richmond B. McQuistan, J. Math. Phys. **13**, 1317 (1972); **15**, 1192 (1974).

On the matrix representation of unbounded operators*

G. Epifanio

Istituto di Fisica dell'Università di Palermo, Palermo, Italy
(Received 19 February 1976)

It is shown that a matrix representation with properties analogous to the ones that hold for the bounded operators in Hilbert space is possible also for important sets of unbounded operators. These sets consist of the $*$-algebras C_D of the linear operators on any noncomplete scalar product space D, which have an adjoint in D. (These algebras have already been studied by the author, in collaboration with others, in previous papers.) Specifically it is proved that for these operators a matrix representation is possible with respect to an arbitrary orthonormal basis within D, in contrast to the situation that has been found by von Neumann for the unbounded closed symmetric operators. The matrix representation of the operators considered here also allows the usual algebraic operations. Besides, the changes of basis induced by automorphisms of D are allowed.

1. INTRODUCTION

It is known that the operators in quantum mechanics are often unbounded[1]: for various reasons they are of course usually closed, hence they cannot be everywhere defined on the Hilbert space H because of the "closed graph theorem.[2]" This situation contrasts with the one that occurs for the bounded operators defined everywhere on a separable Hilbert space, which admit a matrix representation in complete analogy to the operators of finite dimensional spaces.[3]

As early as 1929, von Neumann remarked[4] that if one tries to get a matrix representation of such operators unexpected pathologies occur which make it very difficult to construct a theory on such a basis. [The subject was closely connected with the "new quantum theory" (matrix theory and transformation theory) which was born a few years before and where the unitary transformations of Hermitian unbounded matrices play a fundamental role.]

More precisely, if A is any unbounded operator in H, with dense domain D_A, and if (e_ν) is an orthonormal basis in D_A, then it is always possible to define the matrix $(A_{\mu\nu})$ with $A_{\mu\nu}=(Ae_\nu/e_\mu)$. However in general such a matrix does not represent A. Nevertheless, von Neumann proved[5] that in the case of any closed symmetric operator A it is always possible to find an orthonormal basis (e_ν) in D_A in such a way that the corresponding matrix represents A. However the basis (e_ν) for the representation cannot in general be arbitrarily chosen within D_A. (The matrices which represent closed symmetric operators have the properties

$$A_{\mu\nu}=\bar{A}_{\nu\mu},\quad \sum_\nu |A_{\mu\nu}|^2=\sum_\nu |A_{\nu\mu}|^2<\infty,$$

and are called "Hermitesch quadrierbar" by von Neumann and "C-matrices" by others).

From the foregoing it follows that further difficulties arise when changing the basis (even within D_A). Von Neumann introduces the concept of unitary equivalence for systems "matrix—basis" (two such systems being equivalent whenever they describe the same closed symmetric operator), and he finds in particular that this equivalence is a reciprocal property, but it is not transitive.

Of course von Neumann's theory concerning the pathologies of unbounded closed symmetric operators has not always been taken into account by authors in quantum mechanics and in particular Dirac in his famous treatise[6] states: "Any linear operator is represented by a matrix" and gives properties of such a representation which are certainly true for bounded operators. The arbitrariness of such a statement has also been pointed out in recent papers.[7]

In this work we show that a matrix representation, having properties analogous to those that hold for bounded operators, is also possible for important sets of unbounded operators, defined as follows. (We have already studied them in previous papers[8] in collaboration with other authors.)

Definition 1: Let D be a scalar product space. We say that a linear operator A defined on D has an adjoint A^* in D whenever there exists a linear operator A^* defined on D such that

$$\forall \varphi,\psi\in D,\quad (A\varphi,\psi)=(\varphi,A^*\psi).$$

We call C_D the set of the linear operators that are defined on D and have an adjoint in D.

We also note the following propositions[9]:

Theorem 1: For any scalar product space D, C_D (endowed with the natural operations) is a $*$-algebra of closed operators.

In this paper it is proved, under the assumption of separable D, that the elements of C_D may be represented by matrices with respect to any arbitrary basis in D. Also the changes of basis induced by the authomorphisms of D are allowed. Moreover, the representation provides an isomorphism of the $*$-algebra C_D onto a $*$-algebra of unbounded matrices.

From the mathematical point of view these results support that, as already put into evidence in our previous works, C_D is the most natural algebra of unbounded operators that reduces to[10] $\mathcal{L}(H)$ when D is chosen to be complete.

From the point of view of the applications to quantum physics, these results provide a justification of the use of matrices to represent the algebra of the operators of H that arise in important physical problems. In fact the results may be applied whenever all the operators A_i, together with their adjoints,[11] are defined on a common invariant dense subspace D of H

$$A_i D \subset D, \quad A_i^* D \subset D, \quad \bar{D} = H.$$

That is the A_i's belong to the algebra C_D.

These conditions are well known to be satisfied by the algebra generated by the "smeared" Wightman fields.[12]

They are also known to be satisfied by the algebra generated by the operators describing the canonical coordinates, momenta, and the total energy of a nonrelativistic quantum system of n interacting particles where the potential energy satisfies some regularity conditions,[13] provided D is suitably chosen.

2. MATRIX REPRESENTATION OF OPERATORS OF THE *-ALGEBRA C_D FOR SEPARABLE D

Let us preliminarily introduce the following definition and a theorem which we have already proved in a previous work[14]:

Definition 2: Let D be a scalar product space. We call D_w the space D endowed with the "D-weak" topology determined by the set of seminorms

$$\{\varphi \to |(\varphi, \psi)| \mid \psi \in D\}.$$

Theorem 2: In order that the operator A belong to C_D it is necessary and sufficient that the operator A be continuous in D_w.

We are now ready to prove the possibility of the matrix representation for the operators of C_D with respect to any basis which has been arbitrarily chosen in D.

Definition 3: Let A be a linear operator defined everywhere in the separable scalar product space D, let (e_ν) be an orthonormal basis in D and $M = (A_{\mu\nu})$, an infinite matrix. We say that the matrix M represents the operator A relative to the basis (e_ν) if

$$\forall\, \varphi = \sum_1^\infty {}_\nu\, \xi_\nu e_\nu \in D \quad \text{for } \psi = A\varphi \text{ with } \psi = \sum_1^\infty {}_\nu\, \eta_\nu e_\nu$$

we have

$$\eta_\mu = \sum_1^\infty {}_\nu\, A_{\mu\nu} \xi_\nu.$$

Theorem 3: Every operator $A \in C_D$ (for separable D) admits a matrix representation with respect to any orthonormal basis in D. In this representation to the operator A^* (the adjoint of A) corresponds the matrix[15] $M^* = (A^*_{\mu\nu})$ with $A^*_{\mu\nu} = \bar{A}_{\nu\mu}$.

Proof: Since we consider on D both the norm topology defined by the scalar product and the D-weak topology of Definition 2, we use the symbols "s" (strong) and "w" (weak) to indicate the limits in the first and second topology respectively.

Let (e_ν) be an orthonormal basis in D; we have

$$\forall \varphi \in D, \quad \varphi = \sum_1^\infty {}_\nu\, \xi_\nu e_\nu = \lim_{n\to\infty} {}^s \sum_1^n {}_\nu\, \xi_\nu e_\nu \quad \text{with } \xi_\nu = (\varphi, e_\nu),$$

but since D-weak topology is coarser than the strong one

$$\forall\, \varphi \in D, \quad \varphi = \lim_{n\to\infty} {}^w \sum_1^n {}_\nu\, \xi_\nu e_\nu.$$

Then

$$\forall\, \varphi \in D, \quad (A\varphi)_\mu = (A\varphi, e_\mu) = (A \lim_{n\to\infty} {}^w \sum_1^n {}_\nu\, \xi_\nu e_\nu, e_\mu)$$
$$= \lim_{n\to\infty} (A \sum_1^n {}_\nu\, \xi_\nu e_\nu, e_\mu) = \lim_{n\to\infty} \sum_1^n {}_\nu\, \xi_\nu (Ae_\nu, e_\mu).$$

We have made use here of the linearity of A and of its weak continuity stated in Theorem 2.

Introducing the numbers

$$A_{\mu\nu} = (Ae_\nu, e_\mu)$$

we have

$$\forall \varphi = \sum_1^\infty {}_\nu\, \xi_\nu e_\nu \in D \quad \text{for } \psi = A\varphi$$

$$\text{with } \psi = \sum_1^\infty {}_\nu\, \eta_\nu e_\nu, \quad \eta_\mu = \sum_1^\infty {}_\nu\, A_{\mu\nu} \xi_\nu.$$

So according to Definition 3 it is proved that A admits a matrix representation in D with respect to the basis (e_ν) and its representative matrix is $M(A) = (A_{\mu\nu})$.

Moreover A^* belongs to C_D and, by Theorem 2, it is continuous in D_w, therefore it also admits the matrix representation and one has

$$A^*_{\mu\nu} = (A^* e_\nu, e_\mu) = (e_\nu, Ae_\mu) = \overline{A_{\nu\mu}}.$$

It is obvious that the orthonormal basis (e_ν) may be arbitrarily fixed in D.

We have seen that every operator of C_D generates an infinite matrix $M(A)$. Let us pose the converse problem: What kind of elements $A_{\mu\nu}$ must an infinite matrix have in order to yield an operator of C_D?

Theorem 4: Let D be a separable scalar product space, (e_ν) an orthonormal basis in D and d the linear manifold[16] of l^2 which is canonically isomorphic to D. In order that the matrix $M = (A_{\mu\nu})$ represents an operator $A \in C_D$ with respect to the basis (e_ν) it is necessary and sufficient that

(a) $\forall (\xi_\nu)_{\nu \in N} \in d, \quad (\sum_1^\infty {}_\nu\, A_{\mu\nu} \xi_\nu)_{\mu \in N} \in d,$

(b) $\forall (\xi_\mu)_{\mu \in N} \in d, \quad (\sum_1^\infty {}_\mu\, \overline{A_{\mu\nu}}\, \xi_\mu)_{\nu \in N} \in d,$

(c) $\forall (\xi_\nu)_{\nu \in N} \in d, \quad \forall (\eta_\mu)_{\mu \in N} \in d,$

$$\sum_1^\infty {}_\mu \sum_1^\infty {}_\nu\, \xi_\nu A_{\mu\nu} \bar{\eta}_\mu = \sum_1^\infty {}_\nu \sum_1^\infty {}_\mu\, \xi_\nu A_{\mu\nu} \bar{\eta}_\mu.$$

Proof: The necessity of conditions (a), (b), and (c) is evident. In fact if

$$\varphi = \sum_1^\infty {}_\nu\, \xi_\nu e_\nu$$

then

$$\sum_1^\infty {}_\nu\, A_{\mu\nu} \xi_\nu, \quad \mu = (1, 2, \cdots),$$

are the components of the vector $\psi = A\varphi$. Moreover, since $A^* \in C_D$ and $M(A^*) = M^*(A)$,

$$\sum_1^\infty {}_\mu\, A^*_{\nu\mu} \xi_\mu = \sum_1^\infty {}_\mu\, \overline{A_{\mu\nu}}\, \xi_\mu, \quad \nu = (1, 2, \cdots),$$

are the components of the vector $\chi = A^*\varphi$

Lastly, from the definition of the adjoint (Definition 1) one has

$$\forall\, \varphi = \sum_1^\infty {}_\nu \xi_\nu e_\nu \in D,\ \forall\, \psi = \sum_1^\infty {}_\nu \eta_\nu e_\nu \in D, \quad (A\varphi, \psi) = (\varphi, A^*\psi).$$

Therefore

$$\forall\, (\xi_\nu)_{\nu \in N} \in d, \quad (\eta_\mu)_{\mu \in N} \in d,$$

$$\sum_1^\infty {}_\mu \Big(\sum_1^\infty {}_\nu \xi_\nu A_{\mu\nu}\Big)\bar{\eta}_\mu = \sum_1^\infty {}_\nu \xi_\nu \Big(\sum_1^\infty {}_\mu A_{\mu\nu}\bar{\eta}_\mu\Big).$$

We turn now to the proof of the sufficiency of the above mentioned conditions. If the elements of the matrix $M = (A_{\mu\nu})$ satisfy condition (a), $(A_{\mu\nu})$ represents a linear operator which is everywhere defined in D,

$$A:\ (\xi_\nu)_{\nu \in N} \to \Big(\sum_1^\infty {}_\nu A_{\mu\nu}\xi_\nu\Big)_{\mu \in N}.$$

From condition (b) also $M^* = (A^*_{\mu\nu})$, with $A^*_{\mu\nu} = \overline{A_{\nu\mu}}$, represents a linear operator B which is everywhere defined in D,

$$B:\ (\xi_\mu)_{\mu \in N} \to \Big(\sum_1^\infty {}_\mu A^*_{\nu\mu}\xi_\mu\Big)_{\nu \in N} = \Big(\sum_1^\infty {}_\mu \bar{A}_{\mu\nu}\xi_\mu\Big)_{\nu \in N},$$

and one has

$$\forall\, \varphi,\ \psi \in D \ \text{ with } \varphi = \sum_1^\infty {}_\nu \xi_\nu e_\nu, \quad \psi = \sum_1^\infty {}_\nu \eta_\nu e_\nu,$$

$$(A\varphi, \psi) = \sum_1^\infty {}_\mu \Big(\sum_1^\infty {}_\nu \xi_\nu A_{\mu\nu}\Big)\bar{\eta}_\mu,$$

$$(\varphi, B\psi) = \sum_1^\infty {}_\nu \xi_\nu \Big(\sum_1^\infty {}_\mu \overline{A^*_{\nu\mu}\eta_\mu}\Big) = \sum_1^\infty {}_\nu \xi_\nu \Big(\sum_1^\infty {}_\mu A_{\mu\nu}\bar{\eta}_\mu\Big).$$

Hence from condition (c)

$$\varphi, \psi \in D, \quad (A\varphi, \psi) = (\varphi, B\psi),$$

and, from Definition 1, B is the adjoint of A.

So it is proved that the matrices which satisfy the conditions (a), (b), and (c) represent linear operators everywhere defined on D, with adjoints in D (see Definition 1), hence operators of C_D.

Theorem 5: Let D be a separable scalar product space and (e_ν) an orthonormal basis in D. Let us call d the linear manifold of l^2 which is canonically isomorphic to D and $\mathcal{M}_d$ the set of the matrices that satisfy conditions (a), (b), and (c) of Theorem 4. Then, if we define matrix addition, matrix multiplication, scalar multiplication, and the adjoint matrix as is usual, $\mathcal{M}_d$ is a $*$-algebra and the matrix representation $A \to M(A)$ of the operators of C_D provides an isomorphism of the $*$-algebra C_D onto the $*$-algebra[17] $\mathcal{M}_d$.

Proof: The fact that $A \to M(A)$ is a bijection of C_D onto $\mathcal{M}_d$ follows directly from Theorems 3 and 4 and from Definition 3. In fact, for any fixed orthonormal basis of D, every operator $A \in C_D$, according to Theorem 3, is represented by a matrix $M(A)$ which, according to the necessity of conditions (a), (b), and (c), of Theorem 4, belongs to $\mathcal{M}_d$. So $A \to M(A)$ is a mapping of C_D into $\mathcal{M}_d$. It is implicit in Definition 3 that, whenever an operator is represented by a matrix, the matrix determines uniquely the operator (explicitly, for the operators of C_D, this fact depends on their linearity and continuity in D_w, which is stated by Theorem 2 and used in the proof of Theorem 3): So the mapping $A \to M(A)$ is injective. The surjectivity is stated by the sufficiency of the conditions (a), (b), and (c), of Theorem 4.

There remains to prove

$$M(\alpha A + \beta B) = \alpha M(A) + \beta M(B),$$

$$M(AB) = M(A)M(B), \quad M(A^*) = M^*(A).$$

The proof of the first and the last relation is immediate; concerning the second one we have

$$(M(AB))_{\mu\nu} = (AB_\nu, e_\mu) = (Be_\nu, A^*e_\mu) = \sum_1^\infty {}_\rho (Be_\nu, e_\rho)(e_\rho, A^*e_\mu)$$

$$= \sum_1^\infty {}_\rho (Be_\nu, e_\rho)(Ae_\rho, e_\mu) = \sum_1^\infty {}_\rho (M(B))_{\rho\nu}(M(A))_{\mu\rho}.$$

So the theorem is proved.

In order to develop the analogy between the matrix representation of operators belonging to C_D and operators in finite dimensional spaces, we consider now any change of basis induced by an automorphism U of D. The treatment of more general cases goes beyond the framework of this paper.

Theorem 6: Let (e_ν) and (e'_ν) be two orthonormal bases in the separable scalar product spaces D, $M(A)$, and $M'(A)$, the matrices representing any operator $A \in C_D$ with respect to the bases (e_ν) and (e'_ν) respectively. Let U be the operator associated with the change of basis, that is $Ue_\nu = e'_\nu$; if U is an automorphism of D, then the following relation is valid:

$$M'(A) = M(U^{-1})M(A)M(U),$$

where

$$(M(U))_{\mu\nu} = (Ue_\nu, e_\mu).$$

Proof: Since U is a unitary operator on D, it is an element of C_D, hence we have

$$(M'(A))_{\mu\nu} = (Ae'_\nu, e'_\mu) = (AUe_\nu, Ue_\mu) = (U^{-1}AUe_\nu, e_\mu)$$

$$= (M(U^{-1}AU))_{\mu\nu} = (M(U^{-1})M(A)M(U))_{\mu\nu}.$$

SOME NOTATION

(φ, ψ)	scalar product of the elements φ and ψ
A^*	operator adjoint of the operator A
$A_{\mu\nu}$	matrix element
$(A_{\mu\nu})$	matrix
$(A^*_{\mu\nu})$	matrix adjoint of the matrix $(A_{\mu\nu})$
(e_ν)	orthonormal basis $(e_1,\ e_2, \cdots)$
$(\xi_\nu)_{\nu \in N}$	infinite sequence $(\xi_1, \xi_2, \cdots)$
$\sum_1^n {}_\nu,\ \sum_1^\infty {}_\nu$	summations
$\bar{\xi}$	conjugate of the complex number ξ

ACKNOWLEDGMENTS

The author is grateful to Professor R. Ascoli for his fruitful comments and for a critical reading of the manuscript. She also thanks Dr. G. Riela for stimulating discussions.

*Research partially supported by C.N.R. under contract 75.00422.02, Istituto di Matematica, Università di Palermo.

[1]For definitions and general theorems concerning linear operators in scalar product spaces and their matrix representation see, for instance (a) N.I. Akhiezer and I.M. Glazman, *Theory of Linear Operators in Hilbert Space* (Ungar, New York, 1961), Vol. I.; (b) V.I. Smirnov, *A Course of Higher Mathematics* (Pergamon, New York, Oxford, London, 1964), Vol. V. In the following we call D a complex scalar product space and H a complex complete scalar product space (Hilbert space).

[2]For a proof see, for instance, Ref. 1(b), Sec. 3, n. 186, Theorem 2. It is important to recall that the assumptions of the closed graph theorem require the completeness of the scalar product space.

[3]See Ref. 1(a), Sec. 26 or Ref. 1(b), Sec. 1, n. 134, Sec. 2, n. 163, and Sec. 3, n. 216.

[4](a) J. von Neumann, Math. Ann. **102**, 49 (1929); (b) J. von Neumann, J. f. Math. **161**, 208 (1929).

[5]See, in particular, Ref. 4(a) Anhang III.2. Von Neumann uses the word "Hermitian" in place of the word "symmetric". Various books report the von Neumann theory of matrix representation of closed symmetric operators. See for instance Ref. 1.

[6]P.A.M. Dirac, *The Principles of Quantum Mechanics* (Clarendon, Oxford, 1947), Sec. 17.

[7]See among others J.M. Jauch, "On Bras and Kets" in *Aspects of Quantum Theory*, edited by A. Salam and E.P. Wigner (Cambridge, England, 1972), Sec. 12.

[8](a) R. Ascoli, G. Epifanio, and A. Restivo, Commun. Math. Phys. **18**, 291 (1970); (b) R. Ascoli, G. Epifanio, and A. Restivo, Riv. Math. Univ. Parma 2, 3 (1974).

[9]See Ref. 8(a), Theorem 1, and Ref. 8(b), Theorem 1.

[10]$\mathcal{L}(H)$ is the space of bounded operators in Hilbert space.

[11]The interest to also consider non-Hermitian operators comes from the fact that often these operators, even though they do not correspond to observables, nevertheless are important in some treatments of physical problems.

[12]See, for instance, R.F. Streater and A.S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, Amsterdam, 1964), p. 98.

[13]J.E. Roberts, J. Math. Phys. 7, 1097 (1966).

[14]See Ref. 8(b), Theorem 4.

[15]We recall that the matrix M^*, in the matrix terminology, is called the adjoint matrix of M.

[16]See J. Dieudonné, *Foundation of Modern Analysis* (Academic, New York, London, 1960), Theorem 6.6.2.

[17]If D is complete, and therefore $d=l^2$, the conditions (a), (b), and (c), of Theorem 4 reduce to condition (a) only, because (b) and (c) follow from it. [See Ref. 1(a), p. 51). In such a case $\mathcal{M}_d$ coincides with the algebra $\mathcal{M}$ of all the bounded matrices, and C_D with the algebra $\mathcal{L}(H)$.

Exact dynamics of a model for a three-level atom

Russell Davidson

Department of Economics, University of British Columbia, Vancouver, British Columbia, V6T, 1W5 Canada

John J. Kozak

Department of Chemistry and Radiation Laboratory, University of Notre Dame, Notre Dame, Indiana 46556*
(Received 11 February 1976)

In this paper we investigate the dynamics of a model for a three-level atom in interaction with a radiation field. The exact solution to the spontaneous-emission problem is derived using methods developed earlier by the authors, and expressions are obtained for the probabilities of the atom's being in the first or second excited states at any time t. For the special case that the strengths of the coupling between each of the excited states and the λth mode of the field are proportional, detailed conclusions can be drawn concerning the effects of such factors as system size, coupling function, and level splitting on the temporal evolution of the system. The evolution of excited quantum systems having one versus two modes of decay to the ground state is also compared, and similarities and differences in the temporal behavior are noted. Finally the relevance of the theory presented in this paper to experimental problems in radiation chemistry and physics is indicated in our concluding remarks.

I. INTRODUCTION

The model to be discussed in this paper is one of a three-level quantum system interacting with a one-dimensional radiation field. The three-level system will usually be thought of as an atom with three accessible electronic states between which transitions occur with emission or absorption of radiation. The "atom" could just as well be a molecule, and the radiation field could be a phonon field, a set of closely spaced molecular states, or any of a variety of things, since the model is stripped of all complications—spin, three space dimensions, and so on—in order to make the method of solution as clear as possible and to show its generality.

This work has arisen out of previous work of the authors (Refs. 1–7, hereafter referred to as I–VII, respectively) on a similar model for a two-level quantum system, in particular, the work reported in Paper VII of the series. The techniques evolved there are used for the present model, and are capable of much further extension. Although making the step from a two-level system to a three-level one may not seem very exciting, there is in fact a substantial increase in the structure of the model. Some of the effects which become available for discussion are phosphorescence and fluorescence, competing decay modes, and the like. Such effects may indeed be modelled only crudely by the system described here, but the model is solved exactly, and its extension to more realistic systems is certainly possible.

The atom, or quantum system, then, has three states open to it, $|3\rangle$, $|2\rangle$, and $|1\rangle$. Transitions between any pair of them can occur under the influence of the radiation field, the assumption being that no quantum numbers are involved other than 1, 2, 3 and those of the field. This means that the atom is held fixed in space—it may be thought of as having infinite mass so that it does not recoil when it emits radiation. The field itself is described in terms of the creation and annihilation operators, a_λ^* and a_λ, of the λth mode of the field. Formally the Hamiltonian is

$$H=\hbar\epsilon_3|3\rangle\langle 3|+\hbar\epsilon_2|2\rangle\langle 2|+\sum_\lambda \tfrac{1}{2}\hbar\omega_\lambda a_\lambda^* a_\lambda$$
$$+\sum_\lambda \{h_\lambda^* a_\lambda^* |2\rangle\langle 3|+h_\lambda a_\lambda|3\rangle\langle 2|$$
$$+g_\lambda^* a_\lambda^*|1\rangle\langle 2|+g_\lambda a_\lambda|2\rangle\langle 1|$$
$$+f_\lambda^* a_\lambda^*|1\rangle\langle 3|+f_\lambda a_\lambda|3\rangle\langle 1|\}, \tag{1}$$

where $\hbar\epsilon_3$ and $\hbar\epsilon_2$ are the energies separating states $|3\rangle$ and $|2\rangle$ from $|1\rangle$, $\hbar\omega_\lambda$ is the energy of a photon in the λth mode, and the quantities h_λ, g_λ, f_λ measure the strength of the coupling between the atomic states and the field (they are more or less the transition "matrix elements"). Further, the a_λ operators are defined by

$$\langle n_\lambda|a_\lambda|m_\lambda\rangle=\langle m_\lambda|a_\lambda^*|n_\lambda\rangle$$
$$=[2(n_\lambda+1)]^{1/2}\,\delta^{\mathrm{Kr}}(m_\lambda-n_\lambda-1),$$

where the state $|n_\lambda\rangle$ is that with n_λ photons in the λth mode of the field, and $\delta^{\mathrm{Kr}}(\cdots)$ is the Kronecker delta. The product states

$$|i;\{n_\lambda\}\rangle\equiv|i\rangle\prod_\lambda|n_\lambda\rangle\quad(i=1,2,3;\ n_\lambda=0,1,2,\cdots)$$

define a basis for the Hilbert space of the problem.

We shall be interested throughout this paper in the spontaneous emission of the atom from state $|3\rangle$, that is, in the evolution of the system from the initial state $|3\rangle\Pi_\lambda|0_\lambda\rangle$. At this point, a complication arises that is absent in two-level models. The transitions admitted by the Hamiltonian of Eq. (1) allow the atom to decay in two stages from $|3\rangle$ through $|2\rangle$ to $|1\rangle$, with a photon emitted at each stage. Then one of these photons can be reabsorbed with the atom returning to state $|3\rangle$, still in the presence of the second photon. Such processes do not, of course, conserve energy—and if account is taken of them, the model is no longer in general exactly soluble. For both of these reasons, then, the troublesome processes will not be allowed. This means that the matrix elements of H between those states $|1\rangle|1_{\lambda_1}\rangle|1_{\lambda_2}\rangle\Pi_{\lambda\neq\lambda_1,\lambda_2}|0_\lambda\rangle$ that are accessible from $|3\rangle\Pi_\lambda|0_\lambda\rangle$ and $|3\rangle|1_{\mu_1}\rangle\Pi_{\mu\neq\mu_1}|0_\mu\rangle$ should be zero. This will be rigorously true if h_λ and g_λ are nonzero only on a set of λ for which f_λ is zero. In this case, we are entitled to restrict H to the subspace spanned by the vectors

$|3\rangle \prod_{\lambda} |0_\lambda\rangle$ written now as $|3\rangle$,

$|2\rangle |1_{\lambda_1}\rangle \prod_{\lambda \neq \lambda_1} |0_\lambda\rangle$ written now as $|2;\lambda_1\rangle$,

$|1\rangle |1_{\lambda_1}\rangle |1_{\lambda_2}\rangle \prod_{\lambda \neq \lambda_1, \lambda_2} |0_\lambda\rangle$ written now as $|1;\lambda_1, \lambda_2\rangle$,

with $\lambda_1 < \lambda_2$, for some ordering of the modes,

$|1\rangle |2_{\lambda_1}\rangle \prod_{\lambda \neq \lambda_1} |0_\lambda\rangle$ written now as $|1;2\lambda_1\rangle$,

$|1\rangle |1_{\lambda_1}\rangle \prod_{\lambda \neq \lambda_1} |0_\lambda\rangle$ written now as $|1;\lambda\rangle$

since H has no nonvanishing matrix elements between any of these states and a state not in the subspace spanned by them. What we shall now do is to restrict the Hamiltonian to this subspace, or "sector," anyway, without necessarily imposing the condition given above for the vanishing of f_λ where h_λ and g_λ do not vanish. This procedure in effect changes the Hamiltonian of the problem to the following:

$$\begin{aligned}H = {}& \hbar\epsilon_3 |3\rangle\langle 3| + \sum \hbar(\epsilon_2 + \omega_\lambda) |2;\lambda\rangle\langle 2;\lambda| \\ &+ \sum_{\lambda_1<\lambda_2}\sum \hbar(\omega_{\lambda_1} + \omega_{\lambda_2}) |1;\lambda_1, \lambda_2\rangle\langle 1;\lambda_1, \lambda_2| \\ &+ \sum_\lambda 2\hbar\omega_\lambda |1;2\lambda\rangle\langle 1;2\lambda| + \sum_\lambda \hbar\omega_\lambda |1;\lambda\rangle\langle 1;\lambda| \\ &+ \sqrt{2} \sum_\lambda h_\lambda^* |2;\lambda\rangle\langle 3| + \sum_{\lambda_1>\lambda_2}\sum g_{\lambda_1}^* |1;\lambda_2, \lambda_1\rangle\langle 2;\lambda_2| \\ &+ \sum_{\lambda_1<\lambda_2}\sum \Big\{ g_{\lambda_1}^* |1;\lambda_1, \lambda_2\rangle\langle 2;\lambda_2| + \sum_\lambda g_\lambda^* \sqrt{2}\, |1;2\lambda\rangle\langle 2;\lambda| \\ &+ \sum_\lambda f_\lambda^* |1;\lambda\rangle\langle 3| + \text{Hermitian conjugate} \Big\}. \end{aligned} \tag{2}$$

The factors of $\sqrt{2}$ take account of the matrix elements of the a_λ and a_λ^*. Without the proper condition on f_λ, this is not the same Hamiltonian as that of Eq. (1), but it has the advantage of yielding an exactly soluble system, in which the allowed transitions are as shown in Fig. 1. If there is good physical reason to believe that the scheme diagrammed in Fig. 1 adequately represents the processes of interest in some situation, then the modified Hamiltonian of Eq. (2) should be applicable.

The plan of the paper is now described. The resolvant of the modified Hamiltonian, Eq. (2), is studied in Sec. II, and the exact solution for the spontaneous-emission problem is presented in Sec. III; we obtain, for any time t, the probabilities of the atom's being in the state $|3\rangle$ or state $|2\rangle$. In Sec. IV we explore the consequences of introducing the simplifying assumption that h_λ and g_λ have the same sort of dependence on λ, that they are in fact proportional; a numerical analysis of this special case is reported in Sec. V, and the features which characterize the system's evolution are identified and discussed. The final section is given over to a critique of the model investigated here, and attention is drawn to further problems in radiation theory accessible to exact analysis, given the methods developed in this paper.

II. THE RESOLVENT OF THE MODIFIED HAMILTONIAN

A discussion of the spontaneous emission of the atom from state $|3\rangle$ leads to an initial-value problem. Such problems are usually best handled by use of the resolvent of the Hamiltonian. If our system at time $t=0$ is in the state $|\Psi(0)\rangle$, then at a later time t it is in the state

$$\begin{aligned}|\Psi(t)\rangle &= \exp(-iHt/\hbar)\,|\Psi(0)\rangle \\ &= (1/2\pi i)\int_C dz\, \exp(-izt)(H/\hbar - z)^{-1}\,|\Psi(0)\rangle, \end{aligned} \tag{3}$$

where C is a Bromwich contour taken parallel to the positive direction of the real axis of the complex variable z and above all singularities of the integrand. The operator $(H/\hbar - z)^{-1}$ is what we shall call the resolvent of the Hamiltonian. The spontaneous emission problem will be solved if

$$(H/\hbar - z)^{-1}|3\rangle$$

can be found, that is, if the equation

$$(H/\hbar - z)^{+1}|\Phi\rangle = |3\rangle \tag{4}$$

can be solved for $|\Phi\rangle$.

The way in which Eq. (4) is solved is very similar to that used in Paper VII of the authors' series on two-level atoms (Refs. 1—7), but most of the details will be given here so that this paper may be more or less self-contained. First, it follows straightforwardly from Eq. (2) for the Hamiltonian that

$$(H/\hbar - z)|3\rangle = (\epsilon_3 - z)|3\rangle + (\sqrt{2}/\hbar)\Big\{\sum_\lambda h_\lambda^* |2;\lambda\rangle + \sum_\lambda f_\lambda^* |1;\lambda\rangle\Big\},$$

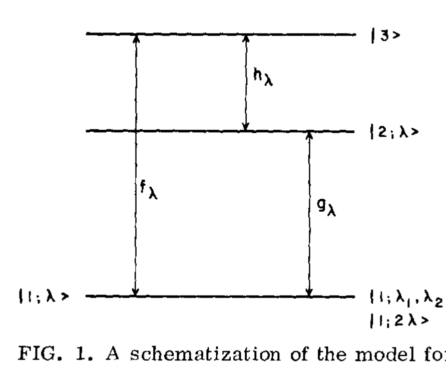

FIG. 1. A schematization of the model for a three-level atom considered in this paper.

$$(H/\hbar - z)|2;\lambda\rangle$$
$$= (\epsilon_2 + \omega_\lambda - z)|2;\lambda\rangle + \sqrt{2}/\hbar\{h_\lambda|3\rangle + \sum_{\mu>\lambda} g_\mu^*|1;\lambda,\mu\rangle$$
$$+ \sum_{\mu<\lambda} g_\mu^*|1;\mu,\lambda\rangle + g_\lambda^*\sqrt{2}|1;2\lambda\rangle\},$$
$$(H/\hbar - z)|1;\lambda_1,\lambda_2\rangle$$
$$= (\omega_{\lambda_1} + \omega_{\lambda_2} - z)|1;\lambda_1,\lambda_2\rangle + (\sqrt{2}/\hbar)\{g_{\lambda_1}|2;\lambda_2\rangle + g_{\lambda_2}|2;\lambda_1\rangle\},$$
$$(H/\hbar - z)|1;2\lambda\rangle$$
$$= (2\omega_\lambda - z)|1;2\lambda\rangle + (\sqrt{2}/\hbar)\{g_\lambda\sqrt{2}|2;\lambda\rangle\},$$
$$(H/\hbar - z)|1;\lambda\rangle = (\omega_\lambda - z)|1;\lambda\rangle + (\sqrt{2}/\hbar)\{f_\lambda|3\rangle\}. \tag{5}$$

The unknown ket $|\Phi\rangle$ can be expanded in terms of the basis states:

$$|\Phi\rangle = \phi_3|3\rangle + \sum_\lambda \phi_{2;\lambda}|2;\lambda\rangle + \sum_{\lambda_1<\lambda_2}\sum \phi_{1;\lambda_1,\lambda_2}|1;\lambda_1,\lambda_2\rangle$$
$$+ \sum_\lambda \phi_{1;2\lambda}|1;2\lambda\rangle + \sum_\lambda \phi_{1;\lambda}|1;\lambda\rangle.$$

This expansion is now substituted in Eq. (4) and use is made of the relations (5). The result is a set of linear equations for the ϕ coefficients:

$$(\epsilon_3 - z)\phi_3 + (\sqrt{2}/\hbar)\left\{\sum_\lambda h_\lambda\phi_{2;\lambda} + \sum_\lambda f_\lambda\phi_{1;\lambda}\right\} = 1,$$
$$(\epsilon_2 + \omega_\lambda - z)\phi_{2;\lambda} + (\sqrt{2}/\hbar)\{h_\lambda^*\phi_3 + \sum_{\mu<\lambda} g_\mu\phi_{1;\mu,\lambda}$$
$$+ \sum_{\mu>\lambda} g_\mu\phi_{1;\lambda,\mu} + g_\lambda\sqrt{2}\,\phi_{1;2\lambda}\} = 0,$$
$$(\omega_{\lambda_1} + \omega_{\lambda_2} - z)\phi_{1;\lambda_1,\lambda_2} + (\sqrt{2}/\hbar)\{g_{\lambda_2}^*\phi_{2;\lambda_1} + g_{\lambda_1}^*\phi_{2;\lambda_2}\} = 0,$$
$$(2\omega_\lambda - z)\phi_{1;2\lambda} + (\sqrt{2}/\hbar)\{g_\lambda^*\sqrt{2}\,\phi_{2;\lambda}\} = 0,$$
$$(\omega_\lambda - z)\phi_{1;\lambda} + (\sqrt{2}/\hbar)\{f_\lambda^*\phi_3\} = 0. \tag{6}$$

From these equations, an equation involving only the coefficients $\phi_{2;\lambda}$ can readily be obtained:

$$\left(\epsilon_2 + \omega_\lambda - z - \sum_\mu \frac{2|g_\mu|^2}{\hbar^2(\omega_\mu + \omega_\lambda - z)}\right)\phi_{2;\lambda} - \sum_\mu \frac{2g_\mu g_\lambda^*\phi_{2;\mu}}{\hbar^2(\omega_\mu + \omega_\lambda - z)}$$
$$= -\frac{h_\lambda^*\sqrt{2}}{\hbar}\left[\epsilon_3 - z - \sum_\mu \frac{2|f_\mu|^2}{\hbar^2(\omega_\mu - z)}\right]^{-1}\left[1 - \frac{\sqrt{2}}{\hbar}\sum_\lambda h_\lambda\phi_{2;\lambda}\right]. \tag{7}$$

This equation is of the same kind as Eq. (7) of Paper VII. It is not surprising that this should be so, since the Hamiltonian that gave rise to that equation was similar to ours of Eq. (2). It modelled in a certain sense one part of the problem being treated here, namely the emission of a two-level system in the presence of a photon. It will become clear that the solution of that problem leads to the solution of the present one.

In addition to the $\phi_{2;\lambda}$, we are interested in ϕ_3. It is readily obtained from the $\phi_{2;\lambda}$ by use of the relation

$$\phi_3 = \left[\epsilon_3 - z - \sum_\lambda \frac{2|f_\lambda|^2}{\hbar^2(\omega_\lambda - z)}\right]^{-1}\left[1 - \frac{\sqrt{2}}{\hbar}\sum_\lambda h_\lambda\phi_{2;\lambda}\right] \tag{8}$$

which follows, as well as Eq. (7), from Eqs. (6). It is noteworthy that

$$-\phi_3 h_\lambda^*\sqrt{2}/\hbar$$

is therefore nothing but the right-hand side of Eq. (7).

Some further definitions will make it possible to simplify the expression of Eq. (7). Let

$$F(z) = \frac{1}{2\pi i}\left(\epsilon_3 - z - \sum_\mu \frac{2|f_\mu|^2}{\hbar^2(\omega_\mu - z)}\right),$$
$$H(z) = \frac{1}{2\pi i}\left(\epsilon_2 - z - \sum_\mu \frac{2|g_\mu|^2}{\hbar^2(\omega_\mu - z)}\right),$$
$$X(z) = \frac{1}{2\pi i}\sum_\mu \frac{g_\mu\phi_{2;\mu}}{\omega_\mu - z}$$
$$A = \frac{1}{2\pi i}\left(1 - \frac{\sqrt{2}}{\hbar}\sum_\lambda h_\lambda\phi_{2;\lambda}\right). \tag{9}$$

The functions F, H, X, depend on z above and are meromorphic functions with poles at the points $z = \omega_\mu$. The residues of X and H at these points are as follows:

$$\mathrm{Res}_{\omega_\lambda} X = -g_\lambda\phi_{2;\lambda}/2\pi i,$$
$$\mathrm{Res}_{\omega_\lambda} H = 2|g_\lambda|^2/2\pi i\hbar^2. \tag{10}$$

Equation (7) is therefore

$$H(z - \omega_\lambda)\,\mathrm{Res}_{\omega_\lambda} X + X(z - \omega_\lambda)\,\mathrm{Res}_{\omega_\lambda} H$$
$$= \frac{g_\lambda h_\lambda^*\sqrt{2}}{(2\pi i)^2\hbar}\cdot\frac{A}{F(z)}. \tag{11}$$

In this equation, the function X is unknown, and z may be regarded as a (nonreal) parameter. The number A depends on the unknown quantities $\phi_{2;\lambda}$ through the sum $\sum_\lambda h_\lambda\phi_{2;\lambda}$, but does *not* depend on the variable ω_λ in Eq. (11). Consequently, A can be regarded as a constant in the solution of Eq. (11), to be determined subsequently. Now, in paper VII, the following equation was solved in detail:

$$H(z - \omega_\lambda)\,\mathrm{Res}_{\omega_\lambda} X + X(z - \omega_\lambda)\,\mathrm{Res}_{\omega_\lambda} H = -g_\lambda(z)/(2\pi i)^2. \tag{12}$$

The solution is rederived for convenience in the Appendix. All that is needed to obtain the solution of Eq. (11) is to substitute for $g_\lambda(z)$ in Eq. (12) the quantity

$$-\frac{g_\lambda h_\lambda^*\sqrt{2}}{\hbar}\cdot\frac{A}{F(z)} \tag{13}$$

This is done in the next section.

III. THE SPONTANEOUS EMISSION SOLUTION

The solution of Eq. (11) follows from Eq. (A12). We have the quantities $\mathrm{Res}_{\omega_\lambda}X$ and $\mathrm{Res}_{\omega_\lambda}H$ from Eq. (10), and, following Eq. (A5), we use Eq. (13) to make the definition

$$\tilde{G}(\xi) = -\frac{A}{F(z)}\frac{\sqrt{2}}{\hbar}\frac{1}{(2\pi i)^2}\sum_\lambda \frac{g_\lambda h_\lambda^*}{\omega_\lambda - \xi}.$$

The solution is then

$$\phi_{2;\lambda} = \frac{2g_\lambda^*}{\hbar^2}\frac{1}{H_1(z)}\sum_\mu \frac{1}{(\xi_\mu - \omega_\lambda)H'(\xi_\mu)H(z - \xi_\mu)}$$
$$\times\sum_\kappa \frac{\tilde{G}(\xi_\mu) - \tilde{G}(\xi_\kappa) + \tilde{G}(z - \xi_\mu) - \tilde{G}(z - \xi_\kappa)}{H'(\xi_\kappa)H(z - \xi_\kappa)}. \tag{14}$$

The ξ_μ are the zeros of the function $H(\xi)$ (as explained in the Appendix) and

$$H_1(z) = \sum_\kappa [1/H'(\xi_\kappa) H(z-\xi_\kappa)]. \tag{15}$$

The function $\tilde{G}$ still involves the unknown number A. It will be convenient to make this explicit:

$$[A/F(z)]\, G(\xi) \equiv \tilde{G}(\xi),$$

where $G(\xi)$ is just

$$-\frac{\sqrt{2}}{\hbar}\frac{1}{(2\pi i)^2}\sum_\lambda \frac{g_\lambda h_\lambda^*}{\omega_\lambda - \xi}. \tag{16}$$

Now A is defined in Eq. (9), and so from Eq. (14):

$$\begin{aligned}
1 - 2\pi i A &= \frac{\sqrt{2}}{\hbar}\sum_\lambda h_\lambda \phi_{2;\lambda} \\
&= \frac{2}{\hbar^2}\frac{1}{H_1(z)}\sum_\mu \frac{\sqrt{2}}{\hbar}\sum_\lambda \frac{g_\lambda^* h_\lambda}{\xi_\mu - \omega_\lambda}\frac{1}{H'(\xi_\mu)H(z-\xi_\mu)}\frac{A}{F(z)} \\
&\quad\times \sum_\kappa \frac{G(\xi_\mu) - G(\xi_\kappa) + G(z-\xi_\mu) - G(z-\xi_\kappa)}{H'(\xi_\kappa)H(z-\xi_\kappa)} \\
&= (2\pi i)^2 \frac{2}{\hbar^2}\frac{A}{F(z)H_1(z)}\sum_\mu \frac{\overline{G}(\xi_\mu)}{H'(\xi_\mu)H(z-\xi_\mu)} \\
&\quad\times \sum_\kappa \frac{G(\xi_\mu) - G(\xi_\kappa) + G(z-\xi_\mu) - G(z-\xi_\kappa)}{H'(\xi_\kappa)H(z-\xi_\kappa)},
\end{aligned}$$

where the bar denotes a complex conjugate. The number A can now be determined:

$$A = \frac{1}{2\pi i}\left\{1 + \frac{2\pi i}{F(z)H_1(z)}\frac{2}{\hbar^2}\sum_\mu \frac{\overline{G}(\xi_\mu)}{H'(\xi_\mu)H(z-\xi_\mu)} \times \sum_\kappa \frac{G(\xi_\mu) - G(\xi_\kappa) + G(z-\xi_\mu) - G(z-\xi_\kappa)}{H'(\xi_\kappa)H(z-\xi_\kappa)}\right\}^{-1}$$

From Eq. (8) we are already able to obtain ϕ_3:

$$\begin{aligned}
\phi_3 &= \frac{A}{F(z)} \\
&= \frac{1}{2\pi i}\left\{F(z) + \frac{2\pi i}{H_1(z)}\frac{2}{\hbar^2}\sum_\mu \frac{\overline{G}(\xi_\mu)}{H'(\xi_\mu)H(z-\xi_\mu)} \times \sum_\kappa \frac{G(\xi_\mu) - G(\xi_\kappa) + G(z-\xi_\mu) - G(z-\xi_\kappa)}{H'(\xi_\kappa)H(z-\xi_\kappa)}\right\}^{-1}.
\end{aligned} \tag{17}$$

The solution for $\phi_{2;\lambda}$ can be expressed in terms of this:

$$\phi_{2;\lambda} = \frac{2g_\lambda^*}{\hbar^2}\frac{\phi_3}{H_1(z)}\sum_\mu \frac{1}{(\xi_\mu - \omega_\lambda)H'(\xi_\mu)H(z-\xi_\mu)} \times \sum_\kappa \frac{G(\xi_\mu) - G(\xi_\kappa) + G(z-\xi_\mu) - G(z-\xi_\kappa)}{H'(\xi_\kappa)H(z-\xi_\kappa)}. \tag{18}$$

With ϕ_3 at our disposal, we may form the probability amplitude for the atom's being in state $|3\rangle$ at time t. From Eq. (3)

$$\langle 3|\Psi(t)\rangle \equiv \phi_3(t) = \frac{1}{2\pi i}\int_C dz\, \exp(-izt)\,\phi_3. \tag{19}$$

This integral can be evaluated if we know the singularities of ϕ_3 as a function of z. From Eq. (17), these are the zeros of the function

$$D(z) \equiv 2\pi i\left\{F(z) + \frac{2\pi i}{H_1(z)}\frac{2}{\hbar^2}\sum_\mu \frac{\overline{G}(\xi_\mu)}{H'(\xi_\mu)H(z-\xi_\mu)} \times \sum_\kappa \frac{G(\xi_\mu) - G(\xi_\kappa) + G(z-\xi_\mu) - G(z-\xi_\kappa)}{H'(\xi_\kappa)H(z-\xi_\kappa)}\right\}. \tag{20}$$

Now this is another meromorphic function, and its poles are at the points $z = \omega_\mu$ [from $F(z)$—see Eq. (9)] and at the zeros of $H_1(z)$. The Mittag—Leffler expansion of $1/H(\xi)$ can be used to furnish the expansion of $H_1(z)$. From

$$\frac{1}{H(\xi)} = \sum_\mu \frac{1}{H'(\xi_\mu)(\xi - \xi_\mu)}$$

and the definition (15) of H_1, it follows that

$$H_1(z) = \sum_\mu\sum_\kappa \frac{1}{H'(\xi_\mu)H'(\xi_\kappa)(z - \xi_\mu - \xi_\kappa)}. \tag{21}$$

Thus H_1 has a series of simple poles at the points $z = \xi_\mu + \xi_\kappa$, and its simple zeros consequently interlace these zeros (all real) by a set ζ_λ (for a more detailed discussion of these matters, see Paper VII). The function D has no other poles, although it might seem that there were others at the points $z = \xi_\rho + \xi_\sigma$. But since H_1 also has poles at these points, it can be seen by a straightforward but rather tedious calculation that D is not singular there. Since, further, the second term on the right-hand side of Eq. (20) tends to zero as $z \to \infty$, D may be written:

$$\begin{aligned}
D(z) &= \epsilon_3 - z - \sum_\mu \frac{2|f_\mu|^2}{\hbar^2(\omega_\mu - z)} \\
&\quad + (2\pi i)^2 \sum_\lambda \frac{1}{H_1'(\zeta_\lambda)(z - \zeta_\lambda)}\frac{2}{\hbar^2}\sum_\mu \frac{\overline{G}(\xi_\mu)}{H'(\xi_\mu)H(\zeta_\lambda - \xi_\mu)} \\
&\quad \times \sum_\kappa \frac{G(\xi_\mu) - G(\xi_\kappa) + G(\zeta_\lambda - \xi_\mu) - G(\zeta_\lambda - \xi_\kappa)}{H'(\xi_\kappa)H(\zeta_\lambda - \xi_\kappa)}.
\end{aligned} \tag{22}$$

This is then the Mittag—Leffler expansion of D, and, by an argument similar to the one used for H_1, the zeros of D, at $z = \phi_\nu$, say, interlace its poles at $z = \omega_\mu$ and $z = \zeta_\lambda$. These zeros are all simple. It is then an immediate consequence of Eq. (19) that

$$\phi_3(t) = \frac{1}{2\pi i}\int_C dz\, \exp(-izt)\frac{1}{D(z)} = \sum_\nu \frac{\exp(-i\phi_\nu t)}{D'(\phi_\nu)}. \tag{23}$$

Equation (18) for $\phi_{2;\lambda}$ can now be examined. We have

$$\langle 2;\lambda|\Psi(t)\rangle \equiv \phi_{2;\lambda}(t) = \frac{1}{2\pi i}\int_C dz\, \exp(-izt)\,\phi_{2;\lambda} \tag{24}$$

and

$$\phi_{2;\lambda} = \frac{2g_\lambda^*}{\hbar^2}\frac{1}{D(z)H_1(z)}\sum_\mu \frac{1}{(\xi_\mu - \omega_\lambda)H'(\xi_\mu)H(z-\xi_\mu)} \times \sum_\kappa \frac{G(\xi_\mu) - G(\xi_\kappa) + G(z-\xi_\mu) - G(z-\xi_\kappa)}{H'(\xi_\kappa)H(z-\xi_\kappa)}.$$

The singularities of this function are at the same points as those of ϕ_3, viz. at $z = \phi_\nu$. As before it can be seen that there is no singularity at the points $z = \xi_\rho + \xi_\sigma$, and, although $H_1(z)$ is zero at the points $z = \zeta_\lambda$, it is clear from Eq. (22) or Eq. (20) that $D(z)H_1(z)$ is not. It then follows at once from Eq. (24) that

$$\begin{aligned}
\phi_{2;\lambda}(t) &= \frac{2g_\lambda^*}{\hbar^2}\sum_\nu \frac{\exp(-i\phi_\nu t)}{D'(\phi_\nu)H_1(\phi_\nu)} \\
&\quad \times \sum_\mu \frac{1}{(\xi_\mu - \omega_\lambda)H'(\xi_\mu)H(\phi_\nu - \xi_\mu)} \\
&\quad \times \sum_\kappa \frac{G(\xi_\mu) - G(\xi_\kappa) + G(\phi_\nu - \xi_\mu) - G(\phi_\nu - \xi_\kappa)}{H'(\xi_\kappa)H(\phi_\nu - \xi_\kappa)}.
\end{aligned}$$

More interesting than $\phi_{2;\lambda}(t)$, perhaps, is the total probability that the atom be in its middle state, $|2\rangle$. This probability is

$$\sum_\lambda |\phi_{2;\lambda}(t)|^2.$$

To evaluate this quantity, one needs the sum

$$\sum_\lambda \frac{2|g_\lambda|^2}{\hbar^2(\xi_\mu-\omega_\lambda)(\xi_{\mu'}-\omega_\lambda)}.$$

It is not difficult to use the definition of $H(z)$ [Eq. (9)] to see that the sum is

$$-1-2\pi i H'(\xi_\mu)\,\delta_{\mu\mu'}.$$

The two terms of this expression will give two contributions to $\sum_\lambda|\phi_{2;\lambda}(t)|^2$. The first (from -1) is zero, since it is

$$\frac{2}{\hbar^2}\left|\sum_\nu \frac{\exp(-i\phi_\nu)}{D'(\phi_\nu)H_1(\phi_\nu)}\sum_\mu \frac{1}{H'(\xi_\mu)H(\phi_\nu-\xi_\mu)}\right.$$
$$\left.\times\sum_\kappa \frac{G(\xi_\mu)-G(\xi_\kappa)+G(\phi_\nu-\xi_\mu)-G(\phi_\nu-\xi_\kappa)}{H'(\xi_\kappa)H(\phi_\nu-\xi_\kappa)}\right|^2,$$

which vanishes because interchanging μ and κ merely changes the sign of the summand. The second contribution, and thus $\sum_\lambda|\phi_{2;\lambda}(t)|^2$, equals

$$(-2\pi i)\frac{2}{\hbar^2}\sum_\mu \frac{1}{\overline{H}'(\xi_\mu)}\left|\sum_\nu \frac{\exp(-i\phi_\nu t)}{D'(\phi_\nu)H_1(\phi_\nu)H(\phi_\nu-\xi_\mu)}\right.$$
$$\left.\times\sum_\kappa \frac{G(\xi_\mu)-G(\xi_\kappa)+G(\phi_\nu-\xi_\mu)-G(\phi_\nu-\xi_\kappa)}{H'(\xi_\kappa)H(\phi_\nu-\xi_\kappa)}\right|^2. \quad (25)$$

This result, along with Eq. (23), completes the solution of the problem. We have obtained, for any time t, the probabilities of the atom's being in state $|3\rangle$ or state $|2\rangle$. In the next section, we shall make a simplifying (and restrictive) assumption which will make our expressions much less complicated. The numerical calculations described in Sec. V will all be based on these simplified expressions.

IV. A PARTICULAR CASE

The simplifying assumption we shall make in the rest of this paper is that h_λ and g_λ have the same sort of dependence on λ, that they are in fact proportional: $h_\lambda = r g_\lambda$, say. In a realistic three-dimensional model, this would mean that the transitions $|3\rangle\to|2\rangle$ and $|2\rangle\to|1\rangle$ (but not necessarily $|3\rangle\to|1\rangle$) were of the same electric or magnetic multipolarity, but not of the same strength. The relative strengths are measured by the constant r, which may be any complex number, although here it will always be taken as real for simplicity. With our assumption, then, the functions G and H become related. From the definitions in Eqs. (9) and (16):

$$G(z) = -\frac{\sqrt{2}}{\hbar}\frac{r}{(2\pi i)^2}\sum_\lambda \frac{|g_\lambda|^2}{\omega_\lambda - z}$$
$$= -\frac{\hbar}{\sqrt{2}}\frac{r}{(2\pi i)^2}[\epsilon_2 - z - 2\pi i H(z)].$$

Then the ubiquitous expression from the preceding section,

$$G(\xi_\mu)-G(\xi_\kappa)+G(z-\xi_\mu)-G(z-\xi_\kappa),$$

simplifies to

$$\frac{\hbar}{\sqrt{2}}\frac{r}{2\pi i}[H(z-\xi_\mu)-H(z-\xi_\kappa)] \quad (26)$$

since

$$H(\xi_\mu)=H(\xi_\kappa)=0.$$

Equation (20) for the function D becomes

$$D(z)=2\pi i\left\{F(z)-\frac{1}{H_1(z)}\frac{r^2}{(2\pi i)^2}\sum_\mu \frac{\epsilon_2-\xi_\mu}{H'(\xi_\mu)H(z-\xi_\mu)}\right.$$
$$\left.\times\sum_\kappa \frac{H(z-\xi_\mu)-H(z-\xi_\kappa)}{H'(\xi_\kappa)H(z-\xi_\kappa)}\right\}. \quad (27)$$

This can be reduced further with the help of the following relations:

$$H_1(z)=\sum_\kappa \frac{1}{H'(\xi_\kappa)H(z-\xi_\kappa)}, \quad (15)$$

$$\sum_\mu \frac{1}{H'(\xi_\mu)} = -2\pi i,$$

$$\sum_\mu \frac{\xi_\mu}{H'(\xi_\mu)} = -2\pi i\epsilon_2, \quad (28)$$

$$\sum_\mu \frac{\xi_\mu}{H'(\xi_\mu)H(z-\xi_\mu)} = 2\pi^2+\tfrac{1}{2}zH_1(z).$$

Equations (28) are all proved similarly. For the first, we notice that

$$\sum_\mu \frac{1}{H'(\xi_\mu)} = \frac{1}{2\pi i}\int_S \frac{d\xi}{H(\xi)} \quad (29)$$

(by the residue theorem) where S is a large circle described in the positive direction. But since, by the definition of H,

$$\lim_{\xi\to\infty}[\xi/H(\xi)] = -2\pi i$$

and since this is the residue at infinity of the integrand in Eq. (29), then

$$\sum_\mu [1/H'(\xi_\mu)] = -2\pi i.$$

For the second of Eqs. (28), the appropriate contour integral is

$$(1/2\pi i)\int_S d\xi\,[G(\xi)/H(\xi)]$$

which is zero, and for the third equation, the integral is

$$(1/2\pi i)\int_S d\xi\,[\xi/H(\xi)H(z-\xi)],$$

which is $4\pi^2$. The details of these derivations are easy and are omitted. Use of Eqs. (28) in Eq. (27) gives

$$D(z)=2\pi i F(z)-r^2[\epsilon_2-\tfrac{1}{2}z-2\pi^2/H_1(z)]. \quad (30)$$

An asymptotic expansion of $H_1(z)$ for large $|z|$ can be derived from Eq. (21) with the help of Eqs. (28). It is

$$H_1(z)\sim -\frac{4\pi^2}{z}-\frac{8\pi^2\epsilon_2}{z^2}+O\left(\frac{1}{z^3}\right),$$

whence one obtains

$$\epsilon_2-\tfrac{1}{2}z-\frac{2\pi^2}{H_1(z)}\sim O\left(\frac{1}{z}\right) \quad \text{as } z\to\infty.$$

Consequently, the Mittag–Leffler expansion is

$$\epsilon_2 - \tfrac{1}{2}z - \frac{2\pi^2}{H_1(z)} = 2\pi^2 \sum_\lambda \frac{1}{H_1'(\zeta_\lambda)(\zeta_\lambda - z)}$$

so that, finally,

$$D(z) = \epsilon_3 - z - \sum_\mu \frac{2\,|f_\mu|^2}{\hbar^2(\omega_\mu - z)} - 2\pi^2 r^2 \sum_\lambda \frac{1}{H_1'(\zeta_\lambda)(\zeta_\lambda - z)}\,. \tag{31}$$

The next expression which can be simplified is that in Eq. (25) for

$$\sum_\lambda |\phi_{2;\lambda}(t)|^2.$$

With Eq. (26), one obtains:

$$\sum_\lambda |\phi_{2;\lambda}(t)|^2 = -2\pi i\, r^2 \sum_\mu \frac{1}{\overline{H}'(\xi_\mu)} \left| \frac{1}{2\pi i} \sum_\nu \frac{\exp(-i\phi_\nu t)}{D'(\phi_\nu)} + \sum_\nu \frac{\exp(-i\phi_\nu t)}{D'(\phi_\nu) H_1(\phi_\nu) H(\phi_\nu - \xi_\mu)} \right|^2$$

Since ξ_μ is real, $\overline{H}'(\xi_\mu) = -H'(\xi_\mu)$. A little more calculation with Eqs. (28) then yields

$$\sum_\lambda |\phi_{2;\lambda}(t)|^2 = r^2 \left\{ -|\phi_3(t)|^2 + 2\pi i \sum_\mu \frac{1}{H'(\xi_\mu)} \left| \sum_\nu \frac{\exp(-i\phi_\nu t)}{D'(\phi_\nu) H_1(\phi_\nu) H(\phi_\nu - \xi_\mu)} \right|^2 \right\}. \tag{32}$$

V. NUMERICAL CALCULATIONS

Before the actual description of the numerical work of computing $|\phi_3(t)|^2$ and $\sum_\lambda |\phi_{2;\lambda}(t)|^2$, the variables employed in the work will be introduced. These variables are dimensionless, and correspond to those used in the computations of Refs. 1–7. They were devised there so that certain limits could be taken easily—weak-coupling, infinite-system, etc.—and may seem rather an excrescence in this paper. But they are of much help in the numerical work, by keeping the number of adjustable parameters in our model to a minimum, and by providing a notation in which the only appearance of non-real numbers is in the exponential time-dependence of $|\phi_3(t)|^2$ and $\sum_\lambda |\phi_{2;\lambda}(t)|^2$. Consistency with the notation of Refs. 1–7 is probably also of some value.

It was remarked in the Introduction that the model was for a one-dimensional system, but this has not yet appeared explicitly except in the assumption that there exists a straightforward ordering in the modes, λ, of the radiation field. Even this was by no means essential to the subsequent discussion. Now the one-dimensionality will be made explicit. The system will be of length L, so that ω_λ, the frequency of mode λ, equal to $c|k_\lambda|$, where c is the speed of light and k_λ is the wave number of the mode, will be

$$\omega_\lambda = c\,|k_\lambda| = 2\pi n c/L, \quad n = 1, 2, 3, \cdots. \tag{33}$$

The various coupling strengths f_λ, g_λ, h_λ, are in general of the form (see Ref. 6):

$$|g_\lambda|^2 = (\alpha\hbar^2\epsilon_2 c/L)\, g(c\,|k_\lambda|/\epsilon_2), \tag{34}$$

where α is a dimensionless coupling constant analogous in one dimension to the fine-structure constant of quantum electrodynamics, and g is some dimensionless function scaled so that $g(1) = 1$ (the argument of g is unity for that mode which is in resonance with the energy gap $\hbar\epsilon_2$). If $|g_\lambda|$ is defined as in Eq. (34), then the other couplings can be defined similarly:

$$|h_\lambda|^2 = \frac{\alpha r^2 \hbar^2 \epsilon_2 c}{L}\; g\!\left(\frac{c\,|k_\lambda|}{\epsilon_2}\right)$$

(since $h_\lambda = r g_\lambda$) and

$$|f_\lambda|^2 = \frac{\alpha s^2 \hbar^2 \epsilon_3 c}{L}\; f\!\left(\frac{c\,|k_\lambda|}{\epsilon_3}\right), \tag{35}$$

where we have used ϵ_3 instead of ϵ_2 so that the requirement $f(1) = 1$ is sensible and where the new parameter s measures the relative strengths of f_λ and g_λ at their respective resonances. Next, the time t and the various frequencies ω_λ, ξ_μ, ζ_ν, ϕ_κ are made dimensionless by the definitions

$$\begin{aligned} &\tau = \alpha\epsilon_3 t, \quad \beta_\lambda = \omega_\lambda/\alpha\epsilon_3, \\ &\gamma_\mu = \xi_\mu/\alpha\epsilon_2, \quad \delta_\nu = \zeta_\nu/\alpha\epsilon_2, \quad \theta_\kappa = \phi_\kappa/\alpha\epsilon_3 \end{aligned} \tag{36}$$

The ratio ϵ_3/ϵ_2 will be denoted by e. The parameter used for the length of the system is related to L by:

$$\sigma^2 = \alpha\epsilon_3 L/c$$

Then Eqs. (34) and (35) become

$$|g_\lambda|^2 = [(\alpha\epsilon_2)^2\hbar^2 e/\sigma^2]\, g(\alpha e \beta_\lambda),$$

$$|f_\lambda|^2 = [s^2(\alpha\epsilon_3)^2\hbar^2/\sigma^2]\, f(\alpha\beta_\lambda).$$

It is convenient to introduce dimensionless functions corresponding to H, H_1, and F:

$$H(\alpha\epsilon_2\xi) = \frac{\alpha\epsilon_2 \hat{H}(\xi)}{2\pi i},$$

$$H_1(\alpha\epsilon_2\xi) = [(2\pi i)^2/\alpha\epsilon_2]\, \hat{H}_1(\xi),$$

and

$$F(\alpha\epsilon_3\xi) = \alpha\epsilon_3 \hat{F}(\xi)/2\pi i$$

so that

$$H'(\alpha\epsilon_2\xi) = (1/2\pi i)\, \hat{H}'(\xi),$$

$$H_1'(\alpha\epsilon_2\xi) = (2\pi i/\alpha\epsilon_2)^2\, \hat{H}_1'(\xi),$$

and

$$F'(\alpha\epsilon_3\xi) = \hat{F}'(\xi)/2\pi i.$$

Then from Eqs. (9) and (15):

$$\hat{H}(\xi) = \frac{1}{\alpha} - \xi - \frac{2e}{\sigma^2} \sum_\mu \frac{g(\alpha e \beta_\mu)}{e\beta_\mu - \xi},$$

$$\hat{H}_1(\xi) = \sum_\kappa \frac{1}{\hat{H}'(\gamma_\kappa) \hat{H}(\xi - \gamma_\kappa)},$$

and

$$\hat{F}(\xi) = \frac{1}{\alpha} - \xi - \frac{2s^2}{\sigma^2} \sum_\mu \frac{f(\alpha\beta_\mu)}{\beta_\mu - \xi}. \tag{37}$$

It can now be seen that the γ_μ and δ_ν are the zeros of $\hat{H}$ and $\hat{H}_1$ respectively, and that the functions $\hat{H}$, $\hat{H}_1$, and F are all real for real values of their arguments.

Next, a function $\hat{D}$ can be defined corresponding to D:

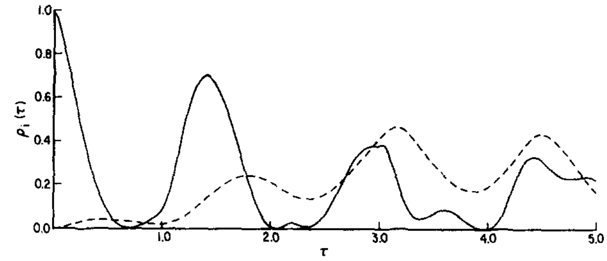

FIG. 2. The temporal behavior of a three-level atom. The parameters which characterize the system are $\alpha = 0.1$, $f(x) = x^{-1/2}$, $\sigma^2 = 1.0$, and $e = 2$; in addition, we set $s = 1.0$ and $r = 0.5$. The solid line describes the evolution of $\rho_3(\tau)$ while the dashed line describes the evolution of $\rho_2(\tau)$. We determine $\rho_3(0) = 0.9927$ and $\rho_2(0) = 0.0010$.

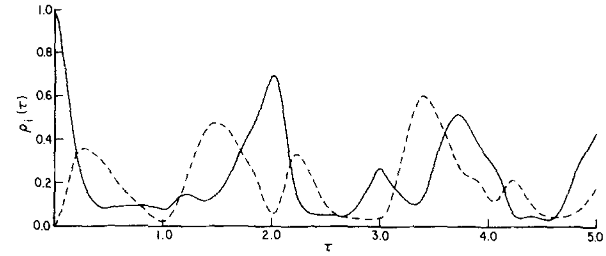

FIG. 4. The temporal behavior of a three-level atom. The conventions and parameter specifications are the same as in Fig. 2, except that here we set $s = 1.0$, $r = 2.0$. In this case, we determine $\rho_3(0) = 0.9927$ and $\rho_2(0) = 0.0173$.

$$D(\alpha\epsilon_3\xi) = \alpha\epsilon_3\hat{D}(\xi)$$

and

$$D'(\alpha\epsilon_3\xi) = \hat{D}'(\xi)$$

whence it follows from Eqs. (30) and (31) that

$$\hat{D}(\xi) = \hat{F}(\xi) - \frac{r^2}{\alpha e} + \tfrac{1}{2}r^2\xi - \frac{r^2}{2e\hat{H}_1(e\xi)}$$

and

$$\hat{D}(\xi) = \frac{1}{\alpha} - \xi - \frac{2s^2}{\sigma^2}\sum_{\mu}\frac{f(\alpha\beta_\mu)}{\beta_\mu - \xi} + \frac{r^2}{2e}\sum_{\lambda}\frac{1}{\hat{H}_1'(\delta_\lambda)(\delta_\lambda - e\xi)}. \tag{38}$$

With all these definitions, then, the equations to be computed, viz. Eqs. (23) and (32), yield

$$\rho_3(\tau) \equiv |\phi_3(t)|^2 = \left|\sum_{\nu}\frac{\exp(-i\theta_\nu\tau)}{\hat{D}'(\theta_\nu)}\right|^2 \tag{39}$$

and

$$\rho_2(\tau) \equiv \sum_{\lambda}|\phi_{2;\lambda}(t)|^2 = -r^2\left\{\rho_3(\tau) + \sum_{\mu}\frac{1}{\hat{H}'(\gamma_\mu)} \times\left|\sum_{\nu}\frac{\exp(-i\theta_\nu\tau)}{\hat{D}'(\theta_\nu)\hat{H}_1(e\theta_\nu)H(e\theta_\nu - \gamma_\mu)}\right|^2\right\}. \tag{40}$$

In the explicit evaluation of these expressions, the β_λ, in accord with Eqs. (33) and (36), are given by

$$\beta_\lambda = |2\pi n/\sigma^2|$$

for some nonzero n, and, for example, from Eqs. (37) and (38) (on differentiation of the latter) we have

$$\hat{H}(\xi) = \frac{1}{\alpha} - \xi + \frac{4e}{\sigma^2}\sum_{n=1}^{\infty}\frac{f(2\pi\alpha n/\sigma^2)}{\xi - 2\pi n/\sigma^2}$$

and

$$\hat{D}'(\theta_\nu) = -1 - \frac{4s^2}{\sigma^2}\sum_{n=1}^{\infty}\frac{f(2\pi\alpha n/\sigma^2)}{(\theta_\nu - 2\pi n/\sigma^2)^2} + \frac{r^2}{2}\sum_{\lambda}\frac{1}{\hat{H}_1'(\delta_\lambda)(\delta_\lambda - e\theta_\nu)^2}.$$

We remark in passing that, with the above specification of β_λ, the eigenfrequencies γ_μ will interlace the poles at $(en)2\pi/\sigma^2$ ($n = 0, 1, 2, \cdots$), the eigenfrequencies δ_ν will interlace the set of terms $(\gamma_i + \gamma_j)$ (all i, j), and the eigenfrequencies θ_κ will interlace the set of terms $\{\delta_i$ (all i), $n2\pi/\sigma^2$ ($n = 0, 1, \cdots$)$\}$. These distributions provide one with checks useful in the numerical search for the eigenfrequencies γ, δ, and θ.

The expressions (39) and (40) for $\rho_3(\tau)$ and $\rho_2(\tau)$ have been computed numerically for a fairly small value of the coupling constant, $\alpha = 0.1$, and for two values of the length parameter, $\sigma^2 = 1.0$ and $\sigma^2 = 10.0$. The coupling functions f and g were chosen so that either $f(x) = g(x) = x^{-1/2}$ or $f(x) = g(x) = x^{-1/4}$. These were the coupling functions most used in Refs. 1–7, and are used here for no better reason. A variety of values for the other parameters, e, s, r, was used, as described later.

As in our earlier work, various checks were performed to assess the reliability of our calculations. Perhaps the most important of these is the one to determine whether the conditions imposed formally at $\tau = 0$ are satisfied numerically. In the present study, the initial conditions are: $\rho_3(0) = 1$ and $\rho_2(0) = 0$. In the captions of Figs 2–10, we indicate the values of $\rho_3(0)$ and $\rho_2(0)$ for the particular cases considered. As a general conclusion, the initial conditions are reproduced satisfactorily for all calculations for which the system size is taken as $\sigma^2 = 1.0$; for all cases considered, $\rho_3(0)$ is effectively 0.99 with $\rho_2(0)$ usually much smaller than 0.01. On the other hand, the calculation of the initial probabilities is somewhat less satisfactory for a system size taken to be $\sigma^2 = 10.0$; in the present study, we determine $\rho_3(0) = 0.9405$ and $\rho_2(0) = 0.0209$. The computational difficulty in satisfying numerically the initial conditions for a system size $\sigma^2 = 10.0$ springs from the fact that an extraordinary large number of eigenfrequencies γ, δ, and θ are required for an accurate determination of $\rho_3(0)$ and $\rho_2(0)$. We remark in passing that this

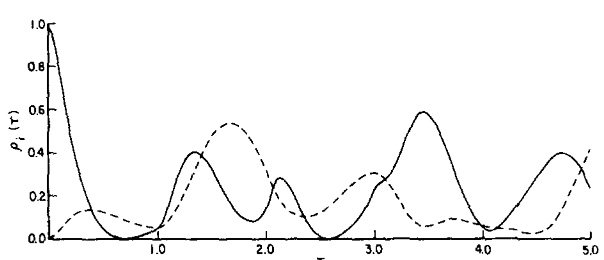

FIG. 3. The temporal behavior of a three-level atom. The conventions and parameter specifications are the same as in Fig. 2, except that here we set $s = 1.0$, $r = 1.0$. In this case, we determine $\rho_3(0) = 0.9928$ and $\rho_2(0) = 0.0042$.

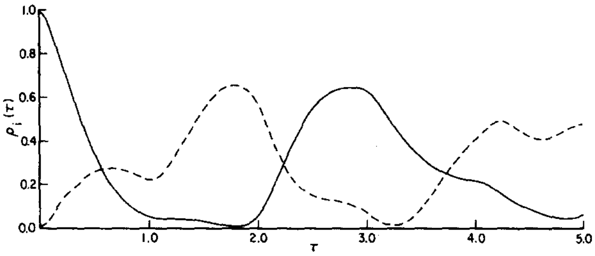

FIG. 5. The temporal behavior of a three-level atom. The conventions and parameter specifications are the same as in Fig. 2, except that here we set $s=0.5$, $r=1.0$. In this case, we determine $\rho_3(0)=0.9980$ and $\rho_2(0)=0.0040$.

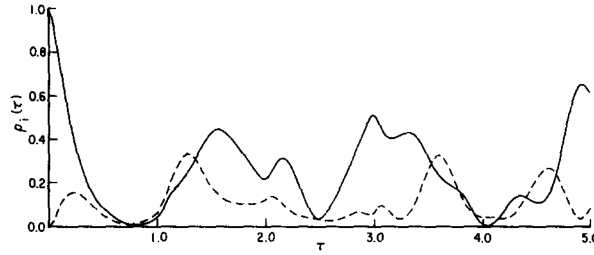

FIG. 7. The temporal behavior of a three-level atom. The parameters which characterize the system are $\alpha=0.1$, $f(x)=x^{-1/2}$, $\sigma=1.0$, and $e=1.1$; here we set $s=1.0$ and $r=1.0$. The solid line describes the evolution of $\rho_3(\tau)$ while the dashed line describes the evolution of $\rho_2(\tau)$. We determine $\rho_3(0)=0.9927$ and $\rho_2(0)=0.0048$.

computational problem is trivial for calculations for which $\sigma^2=1.0$, inasmuch as the number of γ's, δ's and θ's needed to achieve a satisfactory result is quite manageable: Roughly, one needs ~10 γ's, ~10 δ's, and ~20 θ's. When considering $\sigma^2=10.0$, however, one needs 47 γ's, 281 δ's, and 300 θ's just to achieve the value of $\rho_3(0)$ noted above. These difficulties might have been anticipated given our earlier calculations on the Wigner—Weisskopf atom, especially the ones reported in VII; there, as here, the agreement between the value of the initial probability computed numerically and the exact value (for our choice of initial condition), unity, could be improved only by a further, significant investment in computer time. For comparison, we note that the result reported above for $\rho_3(0)$ is slightly better than the value of $\rho(0)$ reported in VII, $\rho(0)=0.9353$, though not as good as the value of $\rho(0)$ computed in IV, 0.9932.

In the first series of figures, Figs. 2—6, $\rho_3(\tau)$ and $\rho_2(\tau)$ are displayed for $\alpha=0.1$, $\sigma^2=1.0$, $f(x)=x^{-1/2}$, and $e=2$. This last value of e means that the three levels of the atom are equally spaced. The choices of r and s are indicated in the caption of each figure, along with the specification of α, σ^2, $f(x)$, and e. If one looks at the three cases where $s=1.0$ (Figs. 2—4), then the effect of changing r, the ratio of the strengths of the couplings h_λ and g_λ (i.e., between states $|3\rangle \rightarrow |2\rangle$ and $|2\rangle \rightarrow |1\rangle$ respectively) can be seen. Increasing r appears to lead to more structure both in ρ_2 and ρ_3, or, more precisely, the time scale of their oscillations decreases. This is no doubt a consequence simply of the increasing strength of one of the decay mechanisms from state $|3\rangle$—a conclusion borne out by the observation that the probability of occupation of state $|2\rangle$ is greater for small τ for greater r. The cases with $r=1.0$ (Figs. 5, 3, 6) manifest the effect of changes in s, the quantity that scales the $|3\rangle \rightarrow |1\rangle$ transition. Again, increasing s decreases the time scale of the oscillations of ρ_3 and ρ_2, and this time, as might be expected, the probability ρ_2 (for τ small) is smaller for greater s.

The second series of figures, Figs. 7—8, holds constant $\alpha=0.1$, $\sigma^2=1.0$, $f(x)=x^{-1/2}$, $s=1$, $r=1$ and examines the consequences of changing e. Since τ is scaled by the frequency ϵ_3, one may imagine the energy gap between $|3\rangle$ and $|1\rangle$ as being fixed, while $|2\rangle$ moves up near $|3\rangle$ for small e (that is, e only slightly greater than unity, since $e>1$ always) and falls to near $|1\rangle$ with large e. Figure 3 can also be included in this series. Since the coupling strengths are fixed here, the time scales are also roughly constant. In fact, changing e produces much less striking effects than changing either r or s. However, it can be seen that the lower the energy of $|2\rangle$ (large e), the more likely it is to be excited *after* the initial decay period, in which, on the contrary, the state $|1\rangle$ is more probable, This effect is rather minor, and in any case in accord with intuition.

Figure 9 keeps the values $\alpha=0.1$, $\sigma^2=1.0$, $e=2$, $s=1$, $r=1$, and changes the coupling function to $f(x)=x^{-1/4}$. This figure is to be compared with Fig. 3, and it is seen at once that it is not very different. In fact, it is difficult to point out any systematic differences in the time evolutions, for either ρ_3 or ρ_2. This is rather fortunate, since as the choice of f is very much ad hoc and has no real physical basis, it is comforting to see that it has only a small effect.

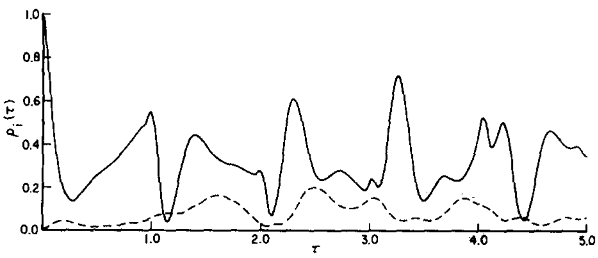

FIG. 6. The temporal behavior of a three-level atom. The conventions and parameter specifications are the same as in Fig. 2, except that here we set $s=2.0$, $r=1.0$. In this case, we determine $\rho_3(0)=0.9895$ and $\rho_2(0)=0.0041$.

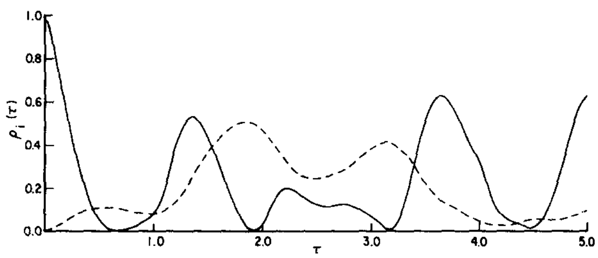

FIG. 8. The temporal behavior of a three-level atom. The conventions and parameter specifications are the same as in the previous figure except that here we set $e=3.0$. We determine $\rho_3(0)=0.9871$ and $\rho_2(0)=0.0073$.

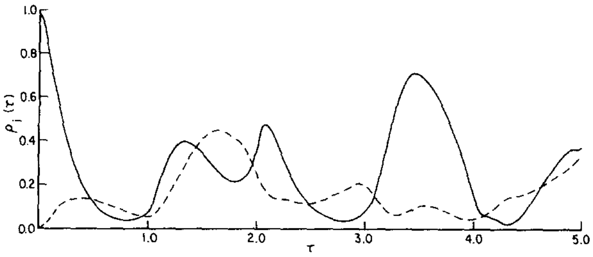

FIG. 9. The temporal behavior of a three-level atom. The parameters which characterize the system are $\alpha=0.1$, $f(x)=x^{-1/4}$, $\sigma^2=1.0$, and $e=2.0$; also, we set $s=1.0$ and $r=1.0$. The solid line describes the evolution of $\rho_3(\tau)$ while the dashed line describes the evolution of $\rho_2(\tau)$. We determine $\rho_3(0)=0.9871$ and $\rho_2(0)=0.0073$.

In Fig. 10, the evolution of ρ_3 and ρ_2 is depicted for the same values of the parameters as in Fig. 3, except that now $\sigma^2=10$. (The τ axis has been much compressed here relative to the $\sigma^2=1.0$ figures.) The result is as expected. The atom decays to state $|1\rangle$ much as before, but now remains there for the longer time required for the emitted photon or photons to bounce back from the edges of the "cavity" in which the system is located. Two times of re-excitation can be seen around $\tau=10$ and $\tau=20$, the second being more diffuse. This effect becomes more pronounced for large τ, viz. $\tau\approx 100$; by that time, the order imparted to the system by the initial condition $\rho_3(0)=1.0$ has been dissipated.

Finally, a comparison can be made between the evolution of a two-level system and our three-level one. The theory presented in Ref. 4 yields the probability ρ as a function of τ of a two-level atom's being excited in circumstances like those of the present model. At first sight it is not clear whether ρ_3 or $\rho_3+\rho_2$ is a better quantity from the present model to use in the comparison with ρ, or given a particular specification of coupling constant, form factor and system length, which choices of r and s ensure that the time scales of the two models are as nearly in accord as possible. However, from an examination of the structure of Eqs. (39) and (40), one anticipates that a choice of r and s which emphasizes the importance of the transition $|3\rangle \rightarrow |1\rangle$ at the expense of the transition $|3\rangle \rightarrow |2\rangle$ should lead to a correspondence in the temporal behavior of the two models.

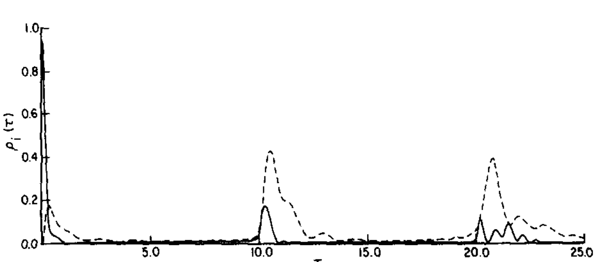

FIG. 10. The temporal behavior of a three-level atom. The parameters which characterize the system are $\alpha=0.1$, $f(x)=x^{-1/2}$, $\sigma^2=10.0$, and $e=2.0$; here we set $s=1.0$ and $r=1.0$. The solid line describes the evolution of $\rho_3(\tau)$ while the dashed line describes the evolution of $\rho_2(\tau)$. We determine $\rho_3(0)=0.9405$ and $\rho_2(0)=0.0209$.

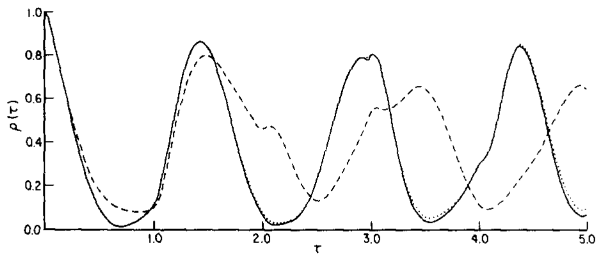

FIG. 11. A comparison of the time evolution of a two-level atom and a three-level one. The parameters common to the two quantum systems are $\alpha=0.1$, $f(x)=x^{-1/2}$, and $\sigma^2=1.0$. The solid line gives $\rho(\tau)$ vs τ for a two-level atom as calculated in Ref. 4. The dotted line gives $\rho_3(\tau)+\rho_2(\tau)$ vs τ for a three-level atom characterized by the parameters $e=2.0$, $s=1.0$, and $r=0.1$; the dashed line describes $\rho_3(\tau)+\rho_2(\tau)$ vs τ for the same three-level system but setting $r=1.0$.

In Fig. 11 we display for $\alpha=0.1$, $f(x)=x^{-1/2}$, and $\sigma^2=1.0$ the temporal evolution of the probabilities ρ (solid line) and $\rho_3+\rho_2$ (dotted line), the latter calculated for the choice $s=1.0$, $r=0.1$; it is seen that the two probabilities are in nearly exact correspondence. On the other hand, if for the same choice of α, $f(x)$, and σ^2 one takes $s=1.0$, $r=1.0$, the correspondence between the two probabilities, ρ (solid line) and $\rho_3+\rho_2$ (dashed line), deteriorates, especially for times greater than $\tau\sim 1.5$. It should also be noted that for this latter choice of r, s the probabilities ρ and ρ_3, or ρ and ρ_2, are in even less good agreement than the probabilities ρ and $\rho_3+\rho_2$, as may be seen by comparing Figs. 3 and 11.

If one performs calculations of ρ and $\rho_3+\rho_2$ for $s=1.0$, $r=1.0$, but for a system characterized by a reduced length $\sigma^2=10.0$ (keeping $\alpha=0.1$ and $f(x)=x^{-1/2}$ fixed), one finds that the similarity in the temporal evolution of the two-level atom and the three-level one persists for times considerably longer than that noted in Fig. 11. In Fig. 12 we have plotted the time development of the probabilities ρ and $\rho_3+\rho_2$ for this set of

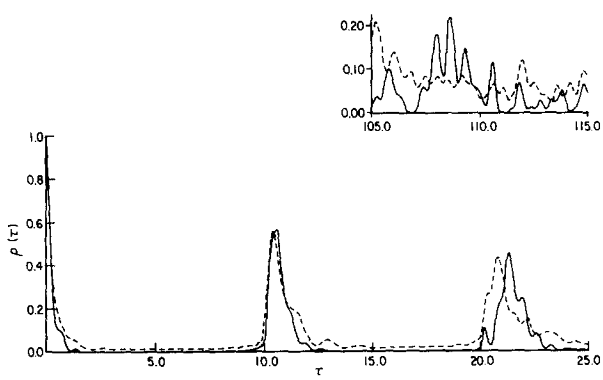

FIG. 12. A comparison of the time evolution of a two-level atom and a three-level one. The parameters common to the two quantum systems are $\alpha=0.1$, $f(x)=x^{-1/2}$, and $\sigma^2=10.0$. The solid line gives $\rho(\tau)$ vs τ for a two-level atom as calculated in Ref. 4. The dashed line gives $\rho_3(\tau)+\rho_2(\tau)$ vs τ for a three-level atom characterized by the parameters $e=2.0$, $s=1.0$, and $r=1.0$. The insert displays the evolution of the two models at a later time τ.

parameters, and it is seen that there is a marked correspondence in the two probabilities over the range of τ considered, $0.0 \leq \tau \leq 25.0$. As shown in the insert, this correspondence deteriorates eventually as the system evolves in time, this due to the dissipation of the initial condition [respectively, $\rho(0)=1.0$ and $\rho_3(0)=1.0$], a property noted in a previous paragraph.

That the evolution of the two-level and three-level system can be brought into nearly exact correspondence for some choices of r, s is not particularly astonishing given the structure of Eqs. (39) and (40). What seems more remarkable is that this correspondence can be achieved for $\sigma^2=1.0$ for a coupling constant characterizing the transition $|3\rangle \to |2\rangle$ only a factor of one-tenth that characterizing the transition $|3\rangle \to |1\rangle$. Increasing the strength of the coupling constant monitoring the transition $|3\rangle \to |2\rangle$ enhances, of course, the importance of the second decay channel open to the three-level system of our model. Indeed, for $\sigma^2=1.0$ when the transitions $|3\rangle \to |1\rangle$ and $|3\rangle \to |2\rangle$ are placed on an equal footing ($s=1.0$, $r=1.0$), the two models exhibit noticeable, quantitative differences for times $\tau > 1.5$. These differences are less pronounced for $\sigma^2=10.0$, at least initially, since the larger system size allows the emitted photons to be "away" from the atom for a longer period of time, thus decreasing the probability of immediate reexcitation of the atom.

VI. DISCUSSION

The chief aim of this paper has been to present the exact solution to a highly stylized problem in the theory of radiation. It is worth insisting on the point that no particular experimental setup has been in mind in the elaboration of the solution, and that consequently the model has no claim, in its present form, to being a paradigm for a specific radiative event. But it seemed important to show that the model was soluble and to give an indication of the sort of behavior it could describe. A considerable number of complications could be incorporated in similar models, which would also be soluble. Spin and three space dimensions have already been mentioned in this connection. It should by now be clear that the methods of this paper are rather general, and so models of quite rich structure should be tractable with their use. Again, since what has been presented here is essentially an exercise in formal quantum mechanics, there is no need to restrict attention to problems involving radiation. Radiationless transitions in aromatic molecules, processes with phonons in solids, and so forth, may well throw up problems similar to that treated here, in addition to the more obvious ones dealing with the phenomena of phosphorescence, fluorescence and such things in atomic and molecular physics.

The treatment of our system as one finite in extent as well as one-dimensional is especially unrealistic from the point of view of radiation theory, where the spectrum is always thought of as continuous rather than discrete. There are two ways in which an infinite system can be considered. One may simply take the results presented here and let σ^2 tend to infinity. This procedure is rather involved, but it is certainly possible: The results are currently being studied. Alternatively, one may formulate a new problem with a Hamiltonian which already has a continuous spectrum. If this is done, the calculations displayed in this paper are replaced by others heavily dependent on the theory of Cauchy integrals (see Ref. 8) rather than on the Mittag–Leffler theorem.

Although it has not been demonstrated in the computations reported here, the influence of the coupling constant, α, on the quantities $\rho_3(\tau)$ and $\rho_2(\tau)$ is not very great for small α. These quantities are not in fact analytic in α at $\alpha=0$ (this matter has been extensively discussed in Refs. 1–7), but even so their limits as $\alpha \to 0$ are well defined, and do not differ markedly from the results given here for $\alpha=0.1$, especially for σ^2 somewhat greater than unity. This claim is borne out in detail for a two-level system in IV. Of course, the variable τ involves α in its definition, so that these remarks do not imply the absurd conclusion that an atom decays from an excited state just as fast if the decay mode is characterized by a weak or a strong coupling to the ground state. It is simply that, once time has been scaled by the coupling constant α, α has little further influence on the evolution of our system.

One obvious increase in the generality of the model, even in its finite-system, one-dimensional form, would be obtained by removing the restriction imposed in Sec. IV that $h_\lambda = r g_\lambda$, and the one imposed in Sec. V that $f(x)=g(x)$. Clearly atomic and molecular transitions are characterized by an electromagnetic multipolarity, as well as spin and parity considerations, and if these are taken into account, the restrictive assumptions cannot be expected to hold. But there is no great difficulty in removing them: The computations merely become more complicated.

Finally, on a more positive note, it is not too far-fetched a claim that the model as it stands in this paper provides an interesting description of a system with two quite different decay modes open to it. The qualitatively sensible results obtained in Sec. V as the parameters r and s are allowed to vary lead one to hope that more complex models of this kind will yield a more detailed description of systems with competing decay channels than has yet been achieved.

APPENDIX

In this appendix a solution will be found for Eq. (12) in the text. The equation is

$$H(z-\omega_\lambda)\,\mathrm{Res}_{\omega_\lambda} X + X(z-\omega_\lambda)\,\mathrm{Res}_{\omega_\lambda} H = -g_\lambda(z)/(2\pi i)^2, \quad \text{(A1)}$$

where z is a parameter. If one considers the function

$$F(\xi) \equiv H(z-\xi)\,X(\xi) + X(z-\xi)\,H(\xi) \quad \text{(A2)}$$

of the complex variable ξ, then one sees that F is meromorphic, and that it has poles where $\xi=\omega_\lambda$ and $\xi = z-\omega_\lambda$. The residues are as follows:

$$\mathrm{Res}_{\omega_\lambda} F = H(z-\omega_\lambda)\,\mathrm{Res}_{\omega_\lambda} X + X(z-\xi)\,\mathrm{Res}_{\omega_\lambda} H,$$

$$\mathrm{Res}_{z-\omega_\lambda} F = -[X(z-\omega_\lambda)\,\mathrm{Res}_{\omega_\lambda} H + H(z-\omega_\lambda)\,\mathrm{Res}_{\omega_\lambda} H]. \quad \text{(A3)}$$

These poles are the only singularities of the function F, which is also finite (or possibly even zero) at infinity. This last remark holds provided $\phi_{2;\lambda}$ is indeed an admissible solution of Eq. (A1) in the sense that $\sum_\lambda |\phi_{2;\lambda}|^2$ is finite. Then F can be expressed in terms of its residues as follows:

$$F(\xi)=\sum_\lambda \frac{\mathrm{Res}_{\omega_\lambda}F}{\xi-\omega_\lambda}+\sum_\lambda \frac{\mathrm{Res}_{z-\omega_\lambda}F}{\xi-(z-\omega_\lambda)}+c(z)$$

by the Mittag–Leffler theorem (see Ref. 9). Here $c(z)$ is just the value of F at infinity. Now by Eqs. (A1) and (A3), this expression for F is

$$F(\xi)=\frac{1}{(2\pi i)^2}\left(\sum_\lambda \frac{g_\lambda(z)}{\omega_\lambda-\xi}+\sum_\lambda \frac{g_\lambda(z)}{\omega_\lambda-(z-\xi)}\right)+c(z)$$
$$=G(\xi)+G(z-\xi)+c(z), \tag{A4}$$

say, where the new meromorphic function G is defined by

$$G(\xi)=\frac{1}{(2\pi i)^2}\sum_\lambda \frac{g_\lambda(z)}{\omega_\lambda-\xi}. \tag{A5}$$

If the result (A4) is put into Eq. (A2) one obtains

$$X(z-\xi)=\frac{1}{H(\xi)}\left[G(\xi)+G(z-\xi)+c(z)-H(z-\xi)X(\xi)\right]. \tag{A6}$$

Now the only poles of $X(z-\xi)$ are where $\xi=z-\omega_\lambda$, and so this must be true also of the right-hand side of Eq. (A6). The function $H(\xi)$ has a set of simple zeros at points $\xi=\xi_\mu$, say, which interlace its poles at $\xi=\omega_\lambda$. (This is an easy consequence of the definition of H, and is shown in detail in Paper IV of the series.) Thus $1/H(\xi)$ has simple poles at $\xi=\xi_\mu$, with residues $1/H'(\xi_\mu)$ (the prime denotes differentiation). But, although the terms on the right-hand side of Eq. (A6) are separately singular at $\xi=\xi_\mu$, this is not so for their sum, and so the separate residues must sum to zero. That is,

$$\frac{G(\xi_\mu)}{H'(\xi_\mu)}+\frac{G(z-\xi_\mu)}{H'(\xi_\mu)}+\frac{c(z)}{H'(\xi_\mu)}-H(z-\xi_\mu)\,\mathrm{Res}_{\xi_\mu}\left(\frac{X}{H}\right)=0.$$

The function X/H is meromorphic with poles only where $\xi=\xi_\mu$ (at $\xi=\omega_\mu$, *both* numerator and denominator vanish, and the ratio is regular) and it vanishes at infinity. The Mittag–Leffler theorem yields

$$X(\xi)=H(\xi)\sum_\mu \frac{1}{\xi-\xi_\mu}\,\frac{G(\xi_\mu)+G(z-\xi_\mu)+c(z)}{H'(\xi_\mu)H(z-\xi_\mu)}. \tag{A7}$$

The quantity $c(z)$ is all that remains to be determined. Since the function X vanishes at infinity, it too has a Mittag–Leffler expansion:

$$X(\xi)=\sum_\lambda \frac{\mathrm{Res}_{\omega_\lambda}X}{\xi-\omega_\lambda}. \tag{A8}$$

From Eq. (A7),

$$\mathrm{Res}_{\omega_\lambda}X=\mathrm{Res}_{\omega_\lambda}H\sum_\mu \frac{G(\xi_\mu)+G(z-\xi_\mu)+c(z)}{(\omega_\lambda-\xi_\mu)H'(\xi_\mu)H(z-\xi_\mu)}, \tag{A9}$$

whence from Eq. (A8)

$$X(\xi)=\sum_\lambda \frac{\mathrm{Res}_{\omega_\lambda}H}{\xi-\omega_\lambda}\sum_\mu \frac{G(\xi_\mu)+G(z-\xi_\mu)+c(z)}{(\omega_\lambda-\xi_\mu)H'(\xi_\mu)H(z-\xi_\mu)}. \tag{A10}$$

Now

$$\sum_\lambda \frac{\mathrm{Res}_{\omega_\lambda}H}{(\xi-\omega_\lambda)(\omega_\lambda-\xi_\mu)}=\frac{1}{\xi-\xi_\mu}\left(\sum_\lambda \frac{\mathrm{Res}_{\omega_\lambda}H}{\xi-\omega_\lambda}-\sum_\lambda \frac{\mathrm{Res}_{\omega_\lambda}H}{\xi_\mu-\omega_\lambda}\right)$$

But the Mittag–Leffler expansion of H is

$$H(\xi)=\frac{1}{2\pi i}(\epsilon_2-\xi)+\sum_\lambda \frac{\mathrm{Res}_{\omega_\lambda}H}{\xi-\omega_\lambda}$$

(the behavior at infinity is important here), and so

$$\sum_\lambda \frac{\mathrm{Res}_{\omega_\lambda}H}{(\xi-\omega_\lambda)(\omega_\lambda-\xi_\mu)}=\frac{H(\xi)}{\xi-\xi_\mu}+\frac{1}{2\pi i}$$

since $H(\xi_\mu)=0$ by definition. When this is put into Eq. (A10) and the result compared with Eq. (A7), an equation for $c(z)$ is obtained:

$$\sum_\mu \frac{G(\xi_\mu)+G(z-\xi_\mu)+c(z)}{H'(\xi_\mu)H(z-\xi_\mu)}=0$$

so that

$$c(z)=-\frac{1}{H_1(z)}\sum_\mu \frac{G(\xi_\mu)+G(z-\xi_\mu)}{H'(\xi_\mu)H(z-\xi_\mu)}, \tag{A11}$$

where the function H_1 is given by

$$H_1(z)=\sum_\mu \frac{1}{H'(\xi_\mu)H(z-\xi_\mu)}.$$

The desired solution to our problem is obtained by substituting Eq. (A11) into Eq. (A9):

$$\mathrm{Res}_{\omega_\lambda}X=\mathrm{Res}_{\omega_\lambda}H\cdot\ -\frac{1}{H_1(z)}\sum_\mu \frac{1}{(\xi_\mu-\omega_\lambda)H'(\xi_\mu)H(z-\xi_\mu)}\times\sum_\kappa \frac{G(\xi_\mu)-G(\xi_\kappa)+G(z-\xi_\mu)-G(z-\xi_\kappa)}{H'(\xi_\kappa)H(z-\xi_\kappa)}. \tag{A12}$$

*The Radiation Laboratory of the University of Notre Dame is operated under contract with the United States Atomic Energy Commission. This is AEC Document No. COO-38-1035.

[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]J.J.-W. Yang, R. Davidson, and J.J. Kozak, J. Math. Phys. **15**, 491 (1974).
[7]R. Davidson and J.J. Kozak, J. Math. Phys. **16**, 1013 (1975).
[8]N.I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953).
[9]P. Dennery and A. Krzywicki, *Mathematics for Physicists* (Harper and Row, New York, 1967), p. 84 et seq.

Fredholm determinants and multiple solitons*

K. M. Case

The Rockefeller University, New York, New York 10021

(Received 29 March 1976)

The discrete inverse scattering problem in one dimension is considered. Exact solutions are obtained using elementary algebraic tools. Expressions found involve determinants of infinite-dimensional matrices. A simple, heuristic, limiting process yields the solution for the continuous problem. When the reflection coefficients do not contribute (the general N soliton case), the determinants reduce to those of given $N\times N$ matrices.

I. INTRODUCTION

In the last few years there has been considerable interest in the Gel'fand—Levitan[1] and Marchenko[2] equations. Part of the reason is the role they play in the remarkable inverse scattering transform method of solving nonlinear partial differential[3] and partial difference[4] equations.

We have two remarks:

(1) It is noted in the literature[5] that in the continuum case (appropriate for partial differential equations) the relevant part of the solutions of these equations are neatly expressed in Fredholm determinants.

(2) Solutions of the Gel'fand—Levitan or Marchenko equations leading to pure N soliton solutions of the related nonlinear evolution equations are simply written in terms of $N\times N$ determinants.[3,4]

Here we consider the solution of the Marchenko equation relevant to the discrete inverse scattering problem. It is shown that the most important quantities are directly expressible in rather simple Fredholm determinants. Besides being useful, the result is of pedagogical interest in that it shows quite clearly how these determinants arise. Finally we show that these (infinite) Fredholm determinants reduce to $N\times N$ determinants in the case of pure N soliton solutions.

To be specific the discrete inverse scattering problem in one dimension is discussed. (This has perhaps the most general interest since the conclusions will pertain to both the Toda lattice and the Korteweg—de Vries equation.) However, it should be emphasized that essentially identical results hold for the discrete inverse scattering problem on the half-line treated either by the Gel'fand—Levitan or Marchenko approach.

II. THE FREDHOLM DETERMINANT SOLUTION

We refer to Refs. 6 and 7 for background and derivations. Briefly stated, the problem is as follows: For the eigenvalue problem

$$a(n+1)\psi(\lambda, n+1) + a(n)\psi(\lambda, n-1) = \lambda\psi(\lambda, n) \tag{1}$$

we are to determine the $a(n)$, or better, the potential $v(n)$ such that

$$a(n) = \tfrac{1}{2}\exp\{-[v(n)+v(n-1)]/2\}, \tag{2a}$$

$$v(n) = \Delta^2 q(n\Delta). \tag{2b}$$

It is assumed that the reflection coefficient $(S_{21}(\lambda))$, the position of the bound states (λ_i) and the bound state normalization constants C_i^2 are given.

The solution procedure that has been described is: Consider the comparison equation

$$a_0(n+1)\psi_0(\lambda,\ n+1) + a_0(n)\psi_0(\lambda, n-1) = \lambda\psi_0(\lambda, n), \tag{3}$$

with known coefficients $a_0(n)$. Define solutions of Eqs. (1) and (3) such that

$$\lim_{n\to\infty}\tfrac{f(\lambda,n)}{f_0(\lambda,n)} \to z^n, \tag{4}$$

where $\lambda = (z+z^{-1})/2$.

Then we have the representation

$$f(\lambda, n) = \sum_{m=n}^{\infty} A(n, m) f_0(\lambda, m), \tag{5}$$

and the solution for $a(n)$ is

$$a(n)/a_0(n) = A(n, n)/A(n-1, n-1). \tag{6}$$

The A is to be obtained so: Let[8]

$$\omega(m, l) = \sum_{i=1}^{N} C_i^2 f_0(\lambda_1, m) f_0(\lambda_i, l) + \int \frac{dz}{2\pi i z}[S_{21}^0(\lambda) - S_{21}(\lambda)] f^0(\lambda, m) f^0(\lambda, l), \tag{7}$$

and

$$\alpha(n, l) = A(n, l)/A(n, n). \tag{8}$$

Then

$$\alpha(n, l) + \omega(n, l) + \sum_{m=n+1}^{\infty}\alpha(n, m)\omega(m, l) = 0, \quad l > n, \tag{9}$$

and

$$\frac{1}{A(n,n)^2} = 1 + \omega(n, n) + \sum_{m=n+1}^{\infty}\alpha(n, m)\omega(n, m). \tag{10}$$

Using Cramer's rule, we can readily write down the solution for $\alpha(n, m)$. Thus

$$\alpha(n, m) = |\quad|m/\det[1+\omega]_{n+1}^{\infty}. \tag{11}$$

Here explicitly

$$\det[1+\omega]_{n+1}^{\infty} = \begin{vmatrix} 1+\omega(n+1,n+1) & \omega(n+1,n+2) & \omega(n+1,n+3) & \cdots \\ \omega(n+2,n+1) & 1+\omega(n+2,n+2) & \omega(n+2,n+3) & \cdots \\ \omega(n+3,n+1) & \omega(n+3,n+2) & 1+\omega(n+3,n+3) & \cdots \\ \vdots & \vdots & \vdots & \end{vmatrix} \tag{12}$$

and $|\ |_m$ is the same as this except that the mth column is replaced by $-\omega(n,n+1)$, $-\omega(n,n+2)$, $-\omega(n,n+3)$, $\cdots$. Thus, as examples,

$$|\ |_{n+1} = \begin{vmatrix} -\omega(n,n+1) & \omega(n+1,n+2) & \omega(n+1,n+3) & \cdots \\ -\omega(n,n+2) & 1+\omega(n+2,n+2) & \omega(n+2,n+3) & \cdots \\ -\omega(n,n+3) & \omega(n+3,n+2) & 1+\omega(n+3,n+3) & \cdots \\ \vdots & \vdots & \vdots & \end{vmatrix}, \tag{13}$$

$$|\ |_{n+2} = \begin{vmatrix} 1+\omega(n+1,n+1) & -\omega(n,n+1) & \omega(n+1,n+3) & \cdots \\ \omega(n+2,n+1) & -\omega(n,n+2) & \omega(n+2,n+3) & \cdots \\ \omega(n+3,n+1) & -\omega(n,n+3) & 1+\omega(n+3,n+3) & \cdots \\ \vdots & \vdots & \vdots & \end{vmatrix}, \tag{14}$$

and

$$|\ |_{n+3} = \begin{vmatrix} 1+\omega(n+1,n+1) & \omega(n+1,n+2) & -\omega(n,n+1) & \cdots \\ \omega(n+3,n+1) & \omega(n+3,n+2) & -\omega(n,n+3) & \cdots \\ \vdots & \vdots & \vdots & \end{vmatrix}. \tag{15}$$

Using Eq. (11) in Eq. (10) we see that

$$\frac{1}{A^2(n,n)} = 1+\omega(n,n) + \sum_{l=n+1}^{\infty} \frac{|\ |_l\,\omega(l,n)}{\det[1+\omega]_{n+1}^{\infty}} = \frac{[1+\omega(n,n)]\det[1+\omega]_{n+1}^{\infty} + \sum_{l=m+1}^{\infty} |\ |_l\,\omega(l,n)}{\det[1+\omega]_{n+1}^{\infty}}. \tag{16}$$

Now consider

$$\det[1+\omega]_{n}^{\infty} = \begin{vmatrix} 1+\omega(n,n) & \omega(n,n+1) & \omega(n,n+2) & \omega(n,n+3) & \cdots \\ \omega(n+1,n) & 1+\omega(n+1,n+1) & \omega(n+1,n+2) & \omega(n+1,n+3) & \cdots \\ \omega(n+2,n) & \omega(n+2,n+1) & 1+\omega(n+2,n+2) & \omega(n+2,n+3) & \cdots \\ \omega(n+3,n) & \omega(n+3,n+1) & \omega(n+3,n+2) & 1+\omega(n+3,n+3) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{vmatrix}. \tag{17}$$

If we expand this in minors of the first row the first term is

$$[1+\omega(n,n)]\det[1+\omega]_{n+1}^{\infty}, \tag{18}$$

the second is

$$\omega(n+1,n)\big|\;\big|_{n+1}, \tag{19}$$

and the third is

$$\omega(n+2,n)\big|\;\big|_{n+2}. \tag{20}$$

In general, the mth term is

$$(n,n+m)\big|\;\big|n+m. \tag{21}$$

Comparing these with Eq. (16) we see that

$$A^2(n,n)=\det[1+\omega]_{n+1}^{\infty}/\det[1+\omega]_n^{\infty}. \tag{22}$$

Finally, from this and Eq. (6) we obtain the following general result for this discrete inverse scattering problem:

$$\frac{a(n)}{a_0(n)}=\left(\frac{\det[1+\omega]_{n+1}^{\infty}\det[1+\omega]_{n-1}^{\infty}}{\{\det[1+\omega]_n^{\infty}\}^2}\right)^{1/2} \tag{23}$$

III. THE CONTINUOUS LIMIT

A heuristic derivation of the continuous form of Eq. (23) is obtained following the approach given in Refs. (6) and (7). In essence we replace all discrete indices $n, m, \cdots$ by $n\Delta$, $m\Delta, \cdots$ and then pass to the limits $n, m, \cdots \to \infty$, $\Delta \to 0$ with $n\Delta = x$, $m\Delta = y, \cdots$ finite. Thus from Eqs. (2) we have

$$q(x)-q_0(x)=\lim_{\Delta\to 0}-\frac{1}{2}\frac{\ln}{\Delta^2}\left[\frac{a(n)}{a_0(n)}^2\right],$$

$$=\lim_{\Delta\to 0}-\frac{1}{2}\frac{1}{\Delta^2}\ln\frac{A^2[(n\Delta,n\Delta)]}{A^2[(m-1)\Delta,(n-1)\Delta]},\quad n\Delta=x. \tag{24}$$

The Marchenko equations (9) become for small Δ,

$$\alpha(n\Delta,l\Delta)+\Delta\omega'(n\Delta,l\Delta)+\Delta\sum_{m=n+1}^{\infty}\alpha(n\Delta,m\Delta)\omega'(m\Delta,l\Delta)$$

$$=0, \tag{25}$$

where $\omega'(n\Delta, l\Delta) \to \omega'(x,y)$ as $\Delta\to 0$, $n\Delta\to x$, $l\Delta\to y$. Explicitly, for the case of no bound states

$$\omega'(x,y)=\frac{1}{2\pi}\int_{-\infty}^{\infty}[S_{12}^{(0)}(k)-S_{12}(k)]f_0(k,y)f_0(k,y)\,dk. \tag{26}$$

Here $S_{12}^{(0)}$, $S_{12}(k)$ are the reflection coefficients for the Schrödinger equation with potentials q_0 and q respectively. $f_0(k,x)$ is the solution with potential q_0 which goes as $\exp(ikx)$ as $x\to\infty$.

From Eq. (25) we see that as $\Delta\to 0$, $\alpha(n\Delta, l\Delta) \to \Delta\alpha'(x,y)$ and thus Eq. (25) becomes the Marchenko integral equation

$$\alpha'(x,y)+\omega'(x,y)+\int_x^{\infty}\alpha'(x,t)\omega'(t,y)\,dt=0. \tag{27}$$

Correspondingly the determinants

$$\det[1+\omega]_n^{\infty}\to\det[1+\omega']_x^{\infty}, \tag{28}$$

where the latter is a Fredholm determinant with kernel

$$\omega'(x,y),\quad \text{defined for } x\le y<\infty. \tag{29}$$

[The limit is particularly clear when the determinants are expressed in terms of traces as described in Eq. (36) given later.]

Now from Eq. (24) we have

$$q(x)-q_0(x)\approx-\frac{1}{2\Delta^2}\ln\frac{A^2[n\Delta,n\Delta]}{A^2[n-1)\Delta,(n-1)\Delta]}$$

$$\approx-\frac{1}{2\Delta^2}\ln\frac{\{A^2[(n-1)\Delta,(n-1)\Delta]+\Delta(\partial/\partial x)A^2(x,x)\}}{A^2[(n-1)\Delta,(n-1)\Delta]}$$

$$\approx-\frac{1}{2\Delta^2}\ln\left[1+\Delta\frac{\partial}{\partial x}\ln A^2(x,x)\right]$$

$$\approx-\frac{1}{2\Delta}\frac{\partial}{\partial x}\ln A^2(x,x), \tag{30}$$

but, from Eq. (22),

$$A^2(n\Delta,n\Delta)=\frac{\det[1+\omega]_{n+1}^{\infty}}{\det[1+\omega]_n^{\infty}}\approx\frac{\det[1+\omega]_n^{\infty}+\Delta(\partial/\partial x)\det[1+\omega']_x^{\infty}}{\det[1+\omega']_x^{\infty}}$$

or

$$A^2(x,x)\approx 1+\Delta\frac{\partial}{\partial x}\ln[1+\omega']_x^{\infty}. \tag{31}$$

Inserting this result in Eq. (30), expanding the logarithm for small Δ and then passing to the limit $\Delta\to 0$ yields the desired result,

$$q(x)-q_0(x)=-\frac{1}{2}\frac{\partial^2}{\partial x^2}\ln\det[1+\omega']_x^{\infty}. \tag{32}$$

IV. THE PURE N-SOLITON CASE

By the term pure N-soliton case we mean the problem when in Eq. (7) the terms involving the reflection coefficients vanish identically, i.e.,

$$\omega(m,l)=\sum_{i=1}^{N}C_i^2 f_0(\lambda_i,m)f_0(\lambda_i,l). \tag{33}$$

It is clear on looking at Eqs. (9) and (10) in this case that the $\alpha(n,m)$ can be expressed in terms of the ratios of determinants of $N\times N$ matrices. What is perhaps not so obvious is that the determinants $\det[1+\omega]_n^{\infty}$ whose ratios determine the $A(n,n)$ are individually expressible in terms of the determinants of $N\times N$ matrices.

The theorem we wish to demonstrate is that if $\omega(m,l)$ is as given by Eq. (33), then

$$\det[1+\omega]_n^{\infty}=\det[1+R_n], \tag{34}$$

where R_n is the $N\times N$ matrix with elements

$$(R_n)_{j1,j2}=C_{j1}C_{j2}\sum_{m=n}^{\infty}f_0(\lambda_{j1},m)f_0(\lambda_{j2},m). \tag{35}$$

A simple, if not necessarily the most elegant, proof is obtained by noting that

$$\det[1+A]=\exp\operatorname{tr}\ln[1+A]=\exp\sum_{l=1}^{\infty}\frac{(-1)^{l+1}}{l}\operatorname{tr}A^l. \tag{36}$$

The proof of Eq. (34) is obtained by showing that

$$\operatorname{tr}\omega^l\big|_n^{\infty}=\operatorname{tr}(R_n)^l,\quad l=1,2,\cdots, \tag{37}$$

but

$$\mathrm{tr}\,\omega^l\Big|_n^\infty = \sum_{\substack{j1,j2\cdots jl\\ 1}}^{N} \sum_{\substack{m_1,m_2\cdots m_l\\ n}}^{\infty} \{\,\}_1$$

with

$$\begin{aligned}\{\,\}_1 = {} & f_0(\lambda_{j1}, m_1) f_0(\lambda_{j1}, m_2)\\ & \cdot f_0(\lambda_{j2}, m_2) f_0(\lambda_{j2}, m_3)\\ & \cdot f_0(\lambda_{j3}, m_3) f_0(\lambda_{j3}, m_4)\\ & \text{----------------}\\ & \cdot f_0(\lambda_{jl-1}, m_{l-1}) f_0(\lambda_{jl-1}, m_l)\\ & \cdot f_0(\lambda_{jl}, m_l) f_0(\lambda_{jl}, m_1),\end{aligned} \tag{38}$$

while

$$\mathrm{tr}(R_n)^l = \sum_{\substack{j1,j2\cdots jl\\ 1}}^{N} \sum_{\substack{m_1',m_2'\cdots m_l'\\ n}}^{\infty} \{\,\}_2,$$

where

$$\begin{aligned}\{\,\}_2 = {} & f_0(\lambda_{j1}, m_1') f_0(\lambda_{j2}, m_1')\\ & \cdot f_0(\lambda_{j2}, m_2') f_0(\lambda_{j3}, m_2')\\ & \cdot f_0(\lambda_{j3}, m_3') f_0(\lambda_{j4}, m_3')\\ & \text{---------------}\\ & \cdot f_0(\lambda_{jl-1}, m_{l-1}') f_0(\lambda_{jl}, m_{l-1}')\\ & \cdot f_0(\lambda_{jl}, m_l') f_0(\lambda_{j1}, m_l').\end{aligned} \tag{39}$$

Relable Eq. (39) so that

$$m_l' = m_1, \quad m_1' = m_2, \quad m_2' = m_3 \cdots, \quad m_{l-1}' = ml.$$

Then shift the entries in $\{\,\}_2$ such that:

(a) The lower right f_0 is put in the top left position.

(b) All others:

(i) If in the left hand column, move to the right.

(ii) If in the right hand column, move one down and one to the left.

We immediately see that the right-hand sides of Eq. (38) and Eq. (39) are equal. This proves Eq. (37) and thus Eq. (34).

V. CONCLUSION

We have considered the discrete inverse scattering problem in one dimension. The exact solution (for the quantities of interest) are obtained using elementary algebra. The expressions are in terms certain infinite dimensional determinants. The solution of the continuous inverse scattering problem is obtained by a simple, if heuristic, limiting procedure. When the reflection coefficients do not contribute (the general N soliton case) the determinants reduce to those of given $N \times N$ matrices.

It is to be emphasized again that essentially identical results for the inverse problem on the half-line can be obtained for either the Gelfand'—Levitan or Marchenko formulation—using exactly the same method.

*Supported in part by the United States Air Force, Grant No. AFOSR 72-2187.

[1]I.M. Gel'fand and B.M. Levittan, Am. Math. Soc. Transl. Ser. 2 1, 253 (1959).
[2]Z.S. Agranovic and V.A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, New York, 1963).
[3]See, for example, A.C. Scott, F.Y.F.Chu, and D.W. McLaughlin, Proc. IEEE **61**, 1443 (1973); M. Ablowitz, D. Kaup, A. Newell, and H. Segur, Stud. Appl. Math. **53**, 255 (1974).
[4]See, for example, H. Flaschka, Prog. Theor. Phys. 51, 703 (1974).
[5]See, for example, F.J. Dyson, Commun. Math. Phys. **47**, 171–83 (1976). This paper strongly motivated our considerations.
[6]K.M. Case and S.C. Chiu, J. Math. Phys. **14**, 1645 (1973).
[7]K.M. Case, J. Math. Phys. **15**, 143 (1974).
[8]For simplicity (only) we are assuming the comparison potential has no bound states.

Derivation of an exact spectral density transport equation for a nonstationary scattering medium

Ioannis M. Besieris

Department of Electrical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061
(Received 3 December 1975)

Within the framework of the quasioptical description and the pure Markovian random process approximation, an exact kinetic equation is derived for the spectral density function in the case of wave propagation in a nondispersive medium characterized by large-scale space–time fluctuations. Also, a quantity, called the *degree of coherence function*, is defined as a quantitative measure of the irreversible effects of randomness.

1. INTRODUCTION

Investigations of electromagnetic wave propagation in nonstationary random media are often based on the equations of classical radiation transport theory, the usual derivation[1,2] of which is based on considerations of energy balance, with no explicit "microscopic" interpretation given to the extinction and scattering coefficients entering into these equations. Moreover, use is frequently made of the random phase approximation which is valid only for incoherent waves (such as stellar radiation). Extensions to this approach introduced by Bugnolo,[3] Stott,[4] and Peacher and Watson[5] are applicable to partially coherent waves and account for multiple scattering effects.

In the past few years, primarily in connection with laser propagation, there has been considerable interest in the investigation of the transformation of the wave spectrum in media characterized by large-scale space—time random fluctuations. Recently reported studies along this direction[6,7] are confined to the quasistatic approximation, with the time dependence of the index of refraction entering parametrically, mostly via a constant or a variable (in the direction of propagation) transverse wind. Furthermore, authors who base their work on radiation transport theory often use uncritically the basic equations of Bugnolo and Peacher and Watson.

It is the intent in this paper to lift several of the aforementioned restrictions and systematically derive an exact spectral density kinetic equation for wave propagation in a nondispersive medium having large-scale space—time random fluctuations within the framework of the quasioptical description and the pure Markovian random process approximation.

2. THE QUASIOPTICAL DESCRIPTION

Ignoring depolarization effects, time-dependent electromagnetic wave propagation in a nondispersive medium with random space—time fluctuations of the refractive index is governed by the stochastic scalar wave equation,

$$\nabla^2 u(\mathbf{r},t)-\frac{1}{c^2}\epsilon_r(\mathbf{r},t)\frac{\partial^2}{\partial t^2}u(\mathbf{r},t)=0. \tag{2.1}$$

Here, c is the velocity of light *in vacuo*, $\epsilon_r(\mathbf{r},t)$ is the relative permittivity which is assumed to be a real random function of space and time, and $u(\mathbf{r},t)$ is a scalar, real, random amplitude function.

For plane- or beam-wave propagation in the z direction, it is convenient to resort to the transformation

$$u(\mathbf{r},t)=\psi(\mathbf{r},t)\exp[ik(z-vt)]+\text{c.c.}, \tag{2.2}$$

where $k=\omega_0/v$, $v=c/\langle\epsilon_r(\mathbf{r},t)\rangle^{1/2}$, and ω_0 is a reference (carrier) frequency. The ensemble average of the random relative permittivity, viz., $\langle\epsilon_r(\mathbf{r},t)\rangle$, is assumed to be constant.

In the quasioptical description, the slowly varying complex random amplitude function $\psi(\mathbf{r},t)$ obeys the nonstationary stochastic parabolic equation[8]

$$\frac{i}{k}\left(\frac{\partial}{\partial z}+\frac{1}{v}\frac{\partial}{\partial t}\right)\psi(\mathbf{x},t;z)=-\frac{1}{2k^2}\nabla_{\mathbf{x}}^2\psi(\mathbf{x},t;z)-\tfrac{1}{2}\epsilon_1(\mathbf{x},t;z)\,\psi(\mathbf{x},t;z),\quad z\geqslant 0, \tag{2.3}$$

where $\mathbf{x}=(x,y)$ and

$$\epsilon_1(\mathbf{x},t;z)=[\epsilon_r(\mathbf{x},t;z)-\langle\epsilon_r(\mathbf{x},t;z)\rangle]/\langle\epsilon_r(\mathbf{x},t;z)\rangle \tag{2.4}$$

is the normalized fluctuating part of the random relative permittivity. Equation (2.3) is rendered closed by specifying the boundary condition $\psi(\mathbf{x},t;0)=\psi_0(\mathbf{x},t)$.

3. THE SPECTRAL DENSITY

A two- (transverse) point, two-time field density function is next introduced as follows in terms of the wavefunction:

$$\rho(\mathbf{x}_2,\mathbf{x}_1,t_2,t_1;z)=\psi^*(\mathbf{x}_2,t_2;z)\,\psi(\mathbf{x}_1,t_1;z). \tag{3.1}$$

It obeys the equation

$$\frac{i}{k}\frac{\partial}{\partial z}\rho(\mathbf{x}_2,\mathbf{x}_1,t_2,t_1;z)=\left[-\frac{i}{k}\frac{1}{v}\left(\frac{\partial}{\partial t_1}+\frac{\partial}{\partial t_2}\right)-\frac{1}{2k^2}\nabla_{\mathbf{x}_1}^2+\frac{1}{2k^2}\nabla_{\mathbf{x}_2}^2-\tfrac{1}{2}\epsilon_1(\mathbf{x}_1,t_1;z)+\tfrac{1}{2}\epsilon_1(\mathbf{x}_2,t_2;z)\right]\rho(\mathbf{x}_2,\mathbf{x}_1,t_2,t_1;z),\quad z\geqslant 0, \tag{3.2a}$$

$$\rho(\mathbf{x}_2,\mathbf{x}_1,t_2,t_1;0)=\rho_0(\mathbf{x}_2,\mathbf{x}_1,t_2,t_1). \tag{3.2b}$$

The "phase-space" analog of the density function is provided by the field spectral density which is defined as follows:

$$f(\mathbf{x},\mathbf{p},t,w;z)=\left(\frac{k}{2\pi}\right)^3\int_{R^2}d\mathbf{y}\int_{R^1}d\tau\,\exp[ik(\mathbf{p}\cdot\mathbf{y}-w\tau)]\times\rho(\mathbf{x}+\tfrac{1}{2}\mathbf{y},\mathbf{x}-\tfrac{1}{2}\mathbf{y},t+\tfrac{1}{2}\tau,t-\tfrac{1}{2}\tau;z). \tag{3.3}$$

This quantity is real, but not necessarily positive everywhere.[9] It will be shown, however, later on in the exposition, that appropriate moments of the spectral density are physical observables.

Using the definition of $f(\mathbf{x},\mathbf{p},t,w;z)$ in conjunction with (3.1) and (2.3), it is found that the spectral density evolves according to the equation

$$\frac{\partial}{\partial z}f(\mathbf{x},\mathbf{p},t,w;z)=Lf(\mathbf{x},\mathbf{p},t,w;z),\quad z\geq 0, \tag{3.4a}$$

$$f(\mathbf{x},\mathbf{p},t,w;0)=f_0(\mathbf{x},p,t,w), \tag{3.4b}$$

$$Lf(\mathbf{x},\mathbf{p},t,w;z)=-\left(\frac{1}{v}\frac{\partial}{\partial t}+\mathbf{p}\cdot\frac{\partial}{\partial\mathbf{x}}\right)f(\mathbf{x},\mathbf{p},t,w;z)+\theta f(\mathbf{x},\mathbf{p},t,w;z). \tag{3.4c}$$

The following representation of the permittivity-dependent term on the right-hand side of (3.4c) will prove useful in the sequel[10]:

$$\theta f(\mathbf{x},\mathbf{p},t,w;z)=\left(\frac{i}{k}\right)^{-1}\left(\frac{2\pi}{k}\right)^{-3}\int_{R^2}d\mathbf{y}\int_{R^1}d\tau\,\exp[ik(\mathbf{p}\cdot\mathbf{y}-w\tau)]$$
$$\times\rho(\mathbf{x}+\tfrac{1}{2}\mathbf{y},\mathbf{x}-\tfrac{1}{2}\mathbf{y},t+\tfrac{1}{2}\tau,t-\tfrac{1}{2}\tau;z)$$
$$\times[\tfrac{1}{2}\epsilon_1(\mathbf{x}+\tfrac{1}{2}\mathbf{y},t+\tfrac{1}{2}\tau;z)-\tfrac{1}{2}\epsilon_1(\mathbf{x}-\tfrac{1}{2}\mathbf{y},t-\tfrac{1}{2}\tau;z)]. \tag{3.5}$$

4. SPECTRAL DENSITY TRANSPORT EQUATION IN THE PURE MARKOVIAN RANDOM PROCESS APPROXIMATION

We consider in this section a statistical analysis of the stochastic equation (3.4). Specifically, we shall derive an exact kinetic equation for the mean spectral density $\langle f(\mathbf{x},\mathbf{p},t,w;z)\rangle$ in the pure Markovian random process approximation.

Averaging both sides of (3.1) yields

$$\left(\frac{\partial}{\partial z}+\frac{1}{v}\frac{\partial}{\partial t}+\mathbf{p}\cdot\frac{\partial}{\partial\mathbf{x}}\right)\langle f(\mathbf{x},\mathbf{p},t;w;z)\rangle=\Theta\langle f(\mathbf{x},\mathbf{p},t,w;z)\rangle, \tag{4.1a}$$

$$\Theta\langle f(\mathbf{x},\mathbf{p},t,w;z)\rangle=\left(\frac{i}{k}\right)^{-1}\left(\frac{2\pi}{k}\right)^{-3}\int_{R^2}d\mathbf{y}\int_{R^1}d\tau$$
$$\times\exp[ik(\mathbf{p}\cdot\mathbf{y}-w\tau)]\langle\rho(\mathbf{x}+\tfrac{1}{2}\mathbf{y},\mathbf{x}-\tfrac{1}{2}\mathbf{y},t+\tfrac{1}{2}\tau,t-\tfrac{1}{2}\tau;z)$$
$$\times[\tfrac{1}{2}\epsilon_1(\mathbf{x}+\tfrac{1}{2}\mathbf{y},t+\tfrac{1}{2}\tau;z)-\tfrac{1}{2}\epsilon_1(\mathbf{x}-\tfrac{1}{2},t-\tfrac{1}{2}\tau;z)]\rangle. \tag{4.1b}$$

We assume that $\epsilon_1(\mathbf{x},t;z)$ is a δ correlated (in z), homogeneous, wide-sense stationary Gaussian process specified completely by the correlation function

$$\langle\epsilon_1(\mathbf{x}_2,t_2;z_2)\epsilon_1(\mathbf{x}_1,t_1;z_1)\rangle=\frac{2\pi}{k}\gamma(\mathbf{x}_2-\mathbf{x}_1,\ t_2-t_1)\delta(z_2-z_1). \tag{4.2}$$

Then, on the basis of the Furutsu–Novikov[11,12] functional formalism, we have

$$\langle\rho(\mathbf{x}_2,\mathbf{x}_1,t_2,t_1;z)[\epsilon_1(\mathbf{x}_2,t_2;z)-\epsilon_1(\mathbf{x}_1,t_1,z]\rangle$$
$$=\int_{R^2}d\mathbf{x}_2'\int_{R^2}d\mathbf{x}_1'\int_{R^1}dt_2'\int_{R^1}dt_1'\int_{R^1}dz'\,\langle[\epsilon_1(\mathbf{x}_2,t_2;z)-\epsilon_1(\mathbf{x}_1,t_1;z)][\epsilon_1(\mathbf{x}_2',t_2';z')-\epsilon_1(\mathbf{x}_1',t_1';z')]\rangle$$
$$\times\langle\delta\rho(\mathbf{x}_2,\mathbf{x}_1,t_2,t_1;z)/\delta[\epsilon_1(\mathbf{x}_2',t_2';z')-\epsilon_1(\mathbf{x}_1',t_1';z')]\rangle$$
$$=\left(\frac{i}{k}\right)^{-1}\left(\frac{2\pi}{k}\right)[\gamma(\mathbf{x}_2-\mathbf{x}_1,\ t_2-t_1)-\gamma(0,0)]\times\langle\rho(\mathbf{x}_2,\mathbf{x}_1,t_2,t_1;z)\rangle. \tag{4.3}$$

[The symbol $\delta(\cdot)$ denotes a functional derivative.] The last equality follows readily from the equation of evolution of the density function [cf. (3.2)] and an extension of the procedure followed by Tatarskii[13] in connection with the time-independent stochastic parabolic equation.

Using the coordinate transformation $\mathbf{x}_{2,1}\to\mathbf{x}+\frac{1}{2}\mathbf{y}$, $t_{2,1}\to t\pm\frac{1}{2}\tau$ in (4.3) and introducing the result into the statistically averaged equation (4.1), we obtain

$$\left(\frac{\partial}{\partial z}+\frac{1}{v}\frac{\partial}{\partial t}+\mathbf{p}\cdot\frac{\partial}{\partial\mathbf{x}}\right)\langle f(\mathbf{x},\mathbf{p},t,w;z)\rangle$$
$$=\left(\frac{\pi k}{2}\right)\left(\frac{k}{2\pi}\right)^3\int_{R^2}d\mathbf{y}\int_{R^1}d\tau\,\exp[ik(\mathbf{p}\cdot\mathbf{y}-w\tau)][\gamma(\mathbf{y},\tau)$$
$$-\gamma(0,0)]\langle\rho(\mathbf{x}+\tfrac{1}{2}\mathbf{y},\mathbf{x}-\tfrac{1}{2}\mathbf{y},\ t+\tfrac{1}{2}\tau,\ t-\tfrac{1}{2}\tau;z)\rangle. \tag{4.4}$$

This equation simplifies considerably upon introducing the spectrum of the space–time correlation function, viz.,

$$\hat{\gamma}(\mathbf{p},w)=\left(\frac{k}{2\pi}\right)^3\int_{R^2}d\mathbf{y}\int_{R^1}d\tau\,\exp[-ik(\mathbf{p}\cdot\mathbf{y}-w\tau)]\gamma(\mathbf{y},\tau), \tag{4.5a}$$

$$\gamma(\mathbf{y},\tau)=\int_{R^2}d\mathbf{p}\int_{R^1}dw\,\exp[ik(\mathbf{p}\cdot\mathbf{y}-w\tau)\hat{\gamma}(\mathbf{p},w). \tag{4.5b}$$

Bearing in mind the definition of the spectral density [cf. (3.3)], (4.4) changes to the simple, convolution-type transport equation

$$\left(\frac{\partial}{\partial z}+\frac{1}{v}\frac{\partial}{\partial t}+\mathbf{p}\cdot\frac{\partial}{\partial\mathbf{x}}+\frac{\pi k}{2}\gamma(0,0)\right)\langle f(\mathbf{x},\mathbf{p},t,w;z)\rangle$$
$$=\frac{\pi k}{2}\int_{R^2}d\mathbf{p}'\int_{R^1}dw'\,\hat{\gamma}(\mathbf{p}-\mathbf{p}',\ w-w')\langle f(\mathbf{x},\mathbf{p}',t,w';z)\rangle. \tag{4.6}$$

It follows from (4.5b) that

$$\gamma(0,0)=\int_{R^2}d\mathbf{p}\int_{R^1}dw\,\hat{\gamma}(\mathbf{p},w). \tag{4.7}$$

The spectrum $\hat{\gamma}(\mathbf{p},w)$, however, is real, nonnegative, and even in both arguments. By virtue of the last property, it is seen that

$$\gamma(0,0)=\int_{R^2}d\mathbf{p}'\int_{R^1}dw'\,\hat{\gamma}(\mathbf{p}-\mathbf{p}',w-w'), \tag{4.8}$$

and Eq. (4.6) can be recast into the form

$$\left(\frac{\partial}{\partial z}+\frac{1}{v}\frac{\partial}{\partial t}+\mathbf{p}\cdot\frac{\partial}{\partial\mathbf{x}}\right)\langle f(\mathbf{x},\mathbf{p},t,w;z)\rangle$$
$$=\int_{R^2}d\mathbf{p}'\int_{R^1}dw'\,W(\mathbf{p},\mathbf{p}',w,w')\times[\langle f(\mathbf{x},\mathbf{p}',t,w';z)-\langle f(\mathbf{x},\mathbf{p},t,w;z)\rangle], \tag{4.9a}$$

$$W(\mathbf{p},\mathbf{p}',w,w')=\frac{\pi k}{2}\hat{\gamma}(\mathbf{p}-\mathbf{p}',w-w'). \tag{4.9b}$$

This expression has the form of a radiation transport equation. [More precisely, if (4.9a) is integrated over w, it becomes a Boltzmann equation for waves (quasi-particles in phase space).] It extends the kinetic equation reported by Klyatskin and Tatarskii[14] in connection with the stationary stochastic parabolic equation, and, in the quasistatic case, it provides a rigorous basis for the work of Fante (cf. Ref. 7).

From what was said earlier about $\hat{\gamma}(\mathbf{p}, w)$, it follows that the *transition probability* (or *scattering indicatrix*) $W(\mathbf{p}, \mathbf{p}', w, w')$ is real, nonnegative, and obeys the (detailed balance) property $W(\mathbf{p}', \mathbf{p}, w', w) = W(\mathbf{p}, \mathbf{p}', w, w')$. The *scattering rate* (also called the *extinction coefficient* or *collision frequency*) is defined in general by

$$\nu(\mathbf{p}, w) = \int_{R^2} d\mathbf{p}' \int_{R^1} dw' W(\mathbf{p}, \mathbf{p}', w, w'). \tag{4.10}$$

In the case under consideration here, the scattering rate is independent of p and w and is given by

$$\nu = (\pi k/2)\,\gamma(0, 0). \tag{4.11}$$

5. PHYSICAL OBSERVABLES

Having established an expression for the mean spectral density by solving the kinetic equation (4.9), the following physically meaningful averaged quantities can be obtained by straightforward integration: (i) the mutual space–time coherence $\langle \rho(\mathbf{x} + \tfrac{1}{2}\mathbf{y}, \mathbf{x} - \tfrac{1}{2}\mathbf{y}, t + \tfrac{1}{2}\tau, t - \tfrac{1}{2}\tau; z)\rangle = \int d\mathbf{p} \int dw \exp[-ik(\mathbf{p}\cdot\mathbf{y} - w\tau)]\langle f(\mathbf{x}, \mathbf{p}, t, w; z)\rangle$; (ii) the mean intensity density $\langle \psi^*(\mathbf{x}, t; z)\psi(\mathbf{x}, t; z)\rangle = \int d\mathbf{p} \int dw \times \langle f(\mathbf{x}, \mathbf{p}, t, w; z)\rangle$; (iii) the intensity density in momentum space $\langle \hat{\rho}(\mathbf{p}, \mathbf{p}, w, w; z)\rangle = \int d\mathbf{p} \int dt \langle f(\mathbf{x}, p, t, w; z)\rangle$, where $\hat{\rho}(\mathbf{p}, \mathbf{p}, w, w; z)$ is the momentum representation of the intensity density; (iv) the mean intensity flux density $\langle \mathbf{J}(\mathbf{x}, t; z)\rangle = \int d\mathbf{p} \int dw\, \mathbf{p} \langle f(\mathbf{x}, p, t, w; z)\rangle$, where $\mathbf{J}(\mathbf{x}, t; z) = (i/2k)[(\nabla_{\mathbf{x}}\psi^*)\psi - \psi^*(\nabla_{\mathbf{x}}\psi)]$ is the intensity flux density. Furthermore, denoting the total mean intensity, viz., $\int d\mathbf{x} \langle \psi^*(\mathbf{x}, t; z)\psi(\mathbf{x}, t; z)\rangle$ by $I(t; z)$, the following two averaged quantities are important in connection with the propagation of spatially bounded beams: (i) the mean "center of gravity" of the beam $\mathbf{x}_c(t; z) = [\int d\mathbf{p} \int dw \int d\mathbf{x}\,\mathbf{x} \times \langle f(\mathbf{x}, \mathbf{p}, t, w; z)\rangle]/I(t; z)$; (ii) spread of a beam $\tfrac{1}{2}\sigma^2(t; z) = [\int d\mathbf{p} \int dw \int d\mathbf{x}(\mathbf{x} - \mathbf{x}_c)^2 \langle f(\mathbf{x}, \mathbf{p}, t, w; z)\rangle]/I(t; z)$.

6. CONSERVATION OF THE MEAN INTENSITY; DEGREE OF COHERENCE

By virtue of the self-adjointness of the operator

$$H_{\mathrm{op}}\left(-\frac{i}{k}\frac{\partial}{\partial \mathbf{x}}, \mathbf{x}, t; z\right) = -\frac{1}{2k^2}\nabla_{\mathbf{x}}^2 - \tfrac{1}{2}\epsilon_1(\mathbf{x}, t; z)$$

appearing on the right-hand side of (2.3), the intensity density function $|\psi(\mathbf{x}, t; z)|^2$ obeys the conservation law[15]

$$\left(\frac{\partial}{\partial z} + \frac{1}{v}\frac{\partial}{\partial t}\right)|\psi(\mathbf{x}, t; z)|^2 + \nabla_{\mathbf{x}}\cdot \mathbf{J}(\mathbf{x}, t; z) = 0, \tag{6.1}$$

where $\mathbf{J}(\mathbf{x}, t; z)$ is the intensity flux density (cf. previous section).

It was pointed out in the previous section that $\langle |\psi(\mathbf{x}, t; z)|^2\rangle = \int d\mathbf{p} \int dw \langle f(\mathbf{x}, \mathbf{p}, t, w; z)\rangle$ and $\langle \mathbf{J}(\mathbf{x}, t; z)\rangle = \int d\mathbf{p} \int dw\, \mathbf{p} \langle f(\mathbf{x}, \mathbf{p}, t, w; z)\rangle$. Bearing in mind these relationships and integrating both sides of (4.9) over p and w results in the following conservation law for the mean intensity:

$$\left(\frac{\partial}{\partial z} + \frac{1}{v}\frac{\partial}{\partial t}\right)\langle |\psi(\mathbf{x}, t; z)|^2\rangle + \nabla_{\mathbf{x}}\cdot\langle \mathbf{J}(\mathbf{x}, t; z)\rangle = 0. \tag{6.2}$$

Integration of this equation over the entire transverse observation plane yields the relation

$$\left(\frac{\partial}{\partial z} + \frac{1}{v}\frac{\partial}{\partial t}\right)\left[\int_{R^2} d\mathbf{x} \langle |\psi(\mathbf{x}, z)|^2\rangle\right] = 0. \tag{6.3}$$

The quantity $D(\mathbf{x}, \mathbf{p}, t, w; z) = \langle f(\mathbf{x}, \mathbf{p}, t, w; z)\rangle^2$ is defined next as the *phase-space degree of coherence density*. Integrating this quantity over p- and w-space we obtain the *configuration-space degree of coherence density* $d(\mathbf{x}, t; z) = \int d\mathbf{p} \int dw\, D(\mathbf{x}, \mathbf{p}, t, w; z)$. Both sides of this last relation are operated on next with $[\partial/\partial z + (1/v)(\partial/\partial t)]$ and use is made of the transport equation (4.9):

$$\begin{aligned}\left(\frac{\partial}{\partial z} + \frac{1}{v}\frac{\partial}{\partial t}\right) & d(\mathbf{x}, t; z) + \nabla_{\mathbf{x}}\cdot \mathbf{K}(\mathbf{x}, t; z) \\ &= 2\int_{R^2} d\mathbf{p} \int_{R^2} d\mathbf{p}' \int_{R^1} dw \int_{R^1} dw'\, W(\mathbf{p}, \mathbf{p}', w, w') \\ &\quad \times [\langle f(\mathbf{x}, \mathbf{p}', t, w'; z)\rangle \langle f(\mathbf{x}, \mathbf{p}, t, w; z)\rangle \\ &\quad - \langle f(\mathbf{x}, \mathbf{p}, t, w; z)\rangle^2],\end{aligned} \tag{6.4}$$

where

$$\mathbf{K}(\mathbf{x}, t; z) = \int_{R^2} d\mathbf{p} \int_{R^1} dw\, \mathbf{p} D(\mathbf{x}, \mathbf{p}, t, w; z) \tag{6.5}$$

is the *configuration-space degree of coherence flux*.

The right-hand side of (6.4) can be rewritten in the more useful form

$$\begin{aligned}&-\int_{R^2} d\mathbf{p} \int_{R^2} d\mathbf{p}' \int_{R^1} dw \int_{R^1} dw'\, W(\mathbf{p}, \mathbf{p}', w, w') \\ &\times [\langle f(\mathbf{x}, \mathbf{p}, t, w; z)\rangle - \langle f(\mathbf{x}, \mathbf{p}', t, w'; z)\rangle]^2 \leq 0\end{aligned} \tag{6.6}$$

on using the following two properties of the transition probability: (i) $W(\mathbf{p}', \mathbf{p}, w', w) = W(\mathbf{p}, \mathbf{p}', w, w')$ (detailed balance); (ii) $W(\mathbf{p}, \mathbf{p}', w, w') \geq 0$ (nonnegativity). Using, then, (6.6) in conjunction with (6.4), it is seen that

$$\left(\frac{\partial}{\partial z} + \frac{1}{v}\frac{\partial}{\partial t}\right) d(\mathbf{x}, t; z) + \nabla_{\mathbf{x}}\cdot \mathbf{K}(\mathbf{x}, t; z) \leq 0. \tag{6.7}$$

Integrating this relation over $\mathbf{x}$ results in the inequality

$$\left(\frac{\partial}{\partial z} + \frac{1}{v}\frac{\partial}{\partial t}\right)\left[\int_{R^2} d\mathbf{x}\, d(\mathbf{x}, t; z)\right] \leq 0 \tag{6.8}$$

which exhibits the monotonic decrease of the total degree of coherence as it is convected along the z direction with the constant velocity v.

It should be noted that inequality (6.8) is analogous to Boltzmann's H theorem is statistical mechanics. In the latter case, the configuration-space degree of coherence density (related to the entropy) would be defined as $d(\mathbf{x}, t, z) = -\int d\mathbf{p} \int dw \langle f\rangle \ln\langle f\rangle$. It has been pointed out, however, that $\langle f\rangle$ can assume negative values; hence, the need for the alternative approach presented in this section.

7. CONCLUDING REMARKS

The transport equation for the spectral density derived in Sec. 4 is an integrodifferential equation of the convolution type which can be integrated formally, i.e., $\langle f(\mathbf{x}, \mathrm{p}, t, w; z)\rangle$ can be expressed in terms of the initial distribution $\langle f(\mathbf{x}, \mathrm{p}, t, w; 0)\rangle$, by a technique analogous to that suggested by Dolin[16] in the case of a stationary scattering medium. This formal solution can then be examined for specific fluctuation spectra (cf. Refs. 17 and 18), in particular, those arising from a constant or a space-dependent (in the z direction) transverse wind (cf. Refs. 6 and 7). It should be noted, however, that the formulation presented in this paper is general enough, and it allows also the investigation of stochastic wave propagation in a space–time-dependent medium to and from moving sources. The latter subject has been recently examined by Strobehn[19] who used a quasistatic approximation and Rytov's method of smooth perturbations.

The discussion in this paper is confined to the mean spectral density, or, equivalently, to the space–time mutual coherence (cf. Sec. 5). This work, however, can be extended in several directions. For example, within the quasioptical assumption and the pure Markovian random process approximation, one can examine longitudinal (in the z direction) correlations, as well as transverse correlations for higher moments. In particular, a kinetic equation for the fourth moment would be important because of its relationship with the physical phenomenon of scintillation.

The main results presented in this paper, as well as the various extensions outlined in the previous paragraph, although interesting by virtue of the fact that they extend the corresponding results for the case of a stationary scattering medium, are, nonetheless, restricted in scope because of the following three underlying assumptions: (i) quasioptical approximation; (ii) nondispersive medium; (iii) pure Markovian random process approximation. Attempts are presently being made towards relaxing these serious restrictions.

[1]V. V. Sobolev, Izv. AN SSSR Ser. Geogr. Geofiz. **8**, 273 (1944).
[2]S. Chandrasekhar, Astrophys. J. **103**, 351 (1946); **104**, 110 (1946).
[3]D. Bugnolo, J. Appl. Phys. **31**, 1176 (1960).
[4]P. E. Stott, J. Phys. A Ser 2, **2**, 75 (1969).
[5]J. Peacher and K. Watson, J. Math. Phys. **11**, 1496 (1970).
[6]A. Ishimaru, Radio Sci. **10**, 45 (1975).
[7]R. Fante, J. Opt. Soc. Am. **64**, 592 (1974); "Frequency Spectrum of Optical Waves Propagating in a Moving Turbulent Atmosphere," TR-74-0135, Air Force Cambridge Research Labs.; Radio Sci. **10**, 77 (1975).
[8]P. Woude and H. Bremmer, Radio Sci. **10**, 23 (1975).

[9]The spectral density defined in (3.3) is analogous to the Wigner distribution used in phase-space approaches to quantum mechanical problems.

[10]I. M. Besieris and F. D. Tappert, J. Math. Phys. **14**, 1829 (1973).
[11]K. Furutsu, J. Res. Natl. Bur. Stand. Sec. D **67**, 303 (1963).
[12]E. A. Novikov, Zh. Eksp. Teor. Fiz **47**, 1919 (1964) [Sov. Phys. JETP **20**, 1290 (1965)].
[13]V. I. Tatarskii, Zh. Eksp. Teor. Fiz. **56**, 2106 (1969) [Sov. Phys. JETP **29**, 1133 (1969)].
[14]V. I. Klyatskin and V. I. Tatarskii, Izv. Vyssh. Uechebn. Zaved. Radiofiz. **13**, 1061 (1970).

[15]Resorting to the transformations $z \rightarrow z$, $t - z/v \rightarrow \tau$, (6.1) assumes the conventional form $(\partial/\partial z)\,|\psi(\mathbf{x}, \tau; z)|^2 + \nabla_x \cdot \mathbf{J}(\mathbf{x}, \tau; z) = 0$, which clearly indicates the convection in the z direction with velocity v.

[16]L. S. Dolin, Izv. Vyssh. Zaved. Radiofiz, **7**, 380 (1967).
[17]S. F. Clifford, J. Opt. Soc. Am. **61**, 1285 (1971).
[18]A. Ishimaru, IEEE Trans. Antennas Propagat. **20**, 10 (1972).
[19]J. W. Strobehn, IEEE Trans. Antennas Propagat. **22**, 303 (1974).

Closed first- and second-order moment equations for stochastic nonlinear problems with applications to model hydrodynamic and Vlasov-plasma turbulence

Ioannis M. Besieris

Department of Electrical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Wojciech B. Stasiak

Department of Engineering Science and Mechanics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

(Received 2 February 1976)

Working along the lines of a procedure outlined by Keller, a technique is developed for deriving closed first- and second-order moment equations for a general class of stochastic nonlinear equations by performing a renormalization at the level of the second moment. The work of Weinstock, as reformulated recently by Balescu and Misguich, is extended in order to obtain two equivalent representations for the second moment using an exact, nonperturbative, statistical approach. These general results, when specialized to the weak-coupling limit, lead to a complete set of closed equations for the first two moments within the framework of an approximation corresponding to Kraichnan's direct-interaction approximation. Additional restrictions result in a self-consistent set of equations for the first two moments in the stochastic quasilinear approximation. Finally, the technique is illustrated by considering its application to two specific physical problems: (1) model hydrodynamic turbulence and (2) Vlasov-plasma turbulence in the presence of an external stochastic electric field.

1. INTRODUCTION

Several significant advances have been made in the area of stochastic nonlinear problems over the past few years. Kraichnan[1] has introduced a technique, known as the *direct-interaction approximation*, wherein the true problems of interest are replaced by stochastic dynamical models that lead, without approximation, to closed equations for covariances and averaged Green's functions. This method has been used extensively in the theories of hydrodynamic turbulence (cf. Ref. 1) and plasma turbulence (cf. Ref. 2). In an effort to understand Kraichnan's direct-interaction approximation, as well as peripheral contributions (cf. Ref. 3) related primarily to the problem of Vlasov-plasma turbulence, Weinstock[4] has presented a generalization based on an exact, nonperturbative statistical approach valid for both strong and weak turbulence. In the weak-coupling limit, Orszag's and Kraichnan's equations for the mean Green's function (cf. Ref. 2), as well as Dupree's turbulence equations (cf. Ref. 3), are recovered. Further restrictions lead to the well-known *quasilinear approximation*. Weinstock's work has been recently reformulated by Balescu and Misguich,[5] and, within the quasilinear approximation, it has been applied to the Vlasov-plasma turbulence problem, with allowance for the presence of an external, stochastic electric field. Furthermore, a modified Weinstock weak-coupling limit, referred to as the *renormalized quasilinear approximation*, has been introduced,[6] and its connection with Kraichnan's direct-interaction approximation has been discussed.

It was pointed out earlier in the introduction that in Kraichnan's direct-interaction approximation, the main results are expressed in terms of closed equations for covariances and averaged Green's functions. On the other hand, Weinstock obtained for a Vlasov plasma a general set of closed equations in terms of smoothed and fluctuating quantities. Although a connection was established in the weak-coupling limit with Orszag's and Kraichnan's equation for the averaged Green's function, no attempt was made to derive closed equations for statistical moments of relevant field quantities. Along the same vein, in Balescu's and Misguich's work on the Vlasov equation with an external stochastic electric field, an equation is established for the first moment in the quasilinear approximation (cf. Ref. 5). This equation, however, is not closed, as it contains a term proportional to the covariance. This difficulty is remedied by solving the equation for the first moment using an iterative procedure. In their most recent work, Misguich and Balescu (cf. Ref. 6) *do* close the equations for the first two moments by resorting to a renormalization at the level of the first moment. Given that $\mu(t)$ is a field quantity of interest, they derive expressions for its mean, $\mathcal{E}\{\mu(t)\}$, and fluctuating, $\delta\mu(t)$, part within the framework of the renormalized quasilinear approximation (a level related to Kraichnan's direct-interaction approximation). From the expression for $\delta\mu(t)$, a relationship is set up for the covariance $\mathcal{E}\{\delta\mu(t)\delta\mu(t')\}$. The relations for $\mathcal{E}\{\mu(t)\}$ and $\mathcal{E}\{\delta\mu(t)\delta\mu(t')\}$, together with an expression for a mean propagator (related to Kraichnan's averaged Green's function), form, then, a self-consistent set.

The procedure followed by Misguich and Balescu in order to close the equations for the first moment and the correlation function, when specialized to linear stochastic problems considered in the first-order smoothing approximation, has led in the past into serious difficulties, as pointed out by Morrison and McKenna.[7] At this stage, it is difficult to assess the degree to which these difficulties are alleviated when working with nonlinear stochastic problems at the level of Misguich and Balescu's renormalized quasilinear approximation. A clarification of this ambiguity is highly desirable; however, it will not be pursued in this paper, especially as a radically different approach to the closure problem will be followed instead.

It is our intent in this paper to present a technique for closing the equations for the first two moments of a field quantity $\mu(t)$ governed by a stochastic nonlinear equation of the form $(\partial/\partial t)\mu(t)=\Omega\mu(t)$ in the special case that the operator Ω depends linearly on $\mu(t)$. This is achieved via the Weinstock—Balescu—Misguich formalism, working, however, at the level of the second moment. We believe this approach is new and eliminates the closure difficulties mentioned earlier in connection with the work of Misguich and Balescu. Our work has been significantly motivated by a procedure outlined by Keller.[8]

In order for the discussion in this paper to be self-contained, the work of Weinstock, as reformulated by Balescu and Misguich, is briefly outlined in Sec. 2. In Sec. 3, the Weinstock—Balescu—Misguich formalism is extended in order to derive two equivalent equations for the second moment using an exact, nonperturbative, statistical approach valid for an arbitrary stochastic nonlinear operator. These general results are specialized in Sec. 4 to the weak-coupling limit, and a complete set of closed equations is obtained for the first two moments of the field $\mu(t)$ on the basis of an approximation corresponding to Kraichnan's direct-interaction approximation. Further simplifications lead to a complete self-consistent set of equations for the first two moments of $\mu(t)$ in the stochastic quasilinear approximation. Finally, the method developed in this paper is applied to two physically important areas: (1) model hydrodynamic turbulence (cf. Sec. 5), and (2) Vlasov-plasma turbulence with an external stochastic electric field (cf. Sec. 6).

2. REVIEW OF THE WEINSTOCK-BALESCU-MISGUICH FORMALISM

Consider the general nonlinear stochastic equation

$$\frac{\partial}{\partial t}\mu(t;\alpha)=\Omega(t;\alpha)\mu(t;\alpha),\quad t\geqslant t_0, \tag{2.1a}$$

$$\mu(t_0;\alpha)=\mu_0(\alpha). \tag{2.1b}$$

Here, $\Omega(t;\alpha)$ is a nonlinear stochastic operator depending on a parameter $\alpha\in\mathcal{A}$, $\mathcal{A}$ being a probability measure space, and $\mu(t;\alpha)$, the random field quantity, is an element of an infinitely dimensional vector space $\mathcal{H}$ and can be either a scalar or a vector quantity. The discussion in this section is general and applies independently of the precise definition of the field $\mu(t;\alpha)$ and the operator $\Omega(t;\alpha)$.[9]

The stochastic operator Ω is split into two parts as follows: $\Omega=\Omega_0+\Omega_1$. The field μ is also decomposed abstractly into two mutually independent terms, viz., $\mu=A\mu+F\mu$ by means of the formal introduction of the two operators A and F. $A\mu$ is called the *average* (or *mean*) component, and $F\mu$ is the *fluctuating* part of μ. The uniqueness of the decomposition as well as the mutual independence of the two components are ensured by prescribing the properties $A+F=I$, $A^2=A$, $F^2=F$, $AF=FA=0$, where I is the identity operator.

The interconnection between the decompositions for the operator Ω and the field μ is contained in the commutation relations $[\Omega_0,A]_-=0$ and $[\Omega_0,F]_-=0$ which constitute a mathematical statement of the fact that the fluctuating part of μ is due only to Ω_1. Therefore, Ω_0 must commute with A, and also with $F=I-A$.

The specific realization of the "projection" operators A and F which will be used in the ensuing work is the following: $A\mu\to\mathcal{E}\{\mu\}$, $F\mu\to\delta\mu$, where $\mathcal{E}\{\mu\}$ and $\delta\mu$ are the ensemble average and fluctuating (incoherent) parts of the random field $\mu(t;\alpha)$, respectively. Within the framework of this specific realization, the aforementioned commutation relations signify that Ω_0 is a deterministic operator and Ω_1 is a generally noncentered stochastic operator.[10]

Operating on (2.1a) with the operator A yields the following equation:

$$\frac{\partial}{\partial t}\mathcal{E}\{\mu(t)\}=\Omega_0(t)\mathcal{E}\{\mu(t)\}+A\Omega_1(t)\delta\mu(t). \tag{2.2}$$

On the other hand, operating on (2.1a) with the operator $F=I-A$ results in the following two equivalent equations for the fluctuating part of the field μ:

$$\frac{\partial}{\partial t}\delta\mu(t)=(I-A)\Omega(t)\delta\mu(t)+\Omega_1(t)\mathcal{E}\{\mu(t)\}, \tag{2.3}$$

$$\frac{\partial}{\partial t}\delta\mu(t)=\Omega_0(t)\delta\mu(t)+(I-A)\Omega_1(t)\delta\mu(t)+\Omega_1(t)\mathcal{E}\{\mu(t)\}. \tag{2.4}$$

[An additional equivalent equation for $\delta\mu(t)$ given by Balescu and Misguich (cf. Ref. 5) is not being presented here since it will not play a significant role in our discussion.]

Equation (2.3) can be solved for $\delta\mu(t)$ in terms of the mean field and the initial value of the fluctuating part of μ,

$$\delta\mu(t)=U_A(t,t_0)\delta\mu(t_0)+\int_{t_0}^{t}dt'\,U_A(t,t')\Omega_1(t')\mathcal{E}\{\mu(t')\}. \tag{2.5}$$

The Weinstock propagator U_A is defined as the solution of the initial value problem

$$\frac{\partial}{\partial t}U_A(t,t_0)=(I-A)\Omega(t)U_A(t,t_0),\quad t\geqslant t_0, \tag{2.6a}$$

$$U_A(t_0,t_0)=I. \tag{2.6b}$$

In the case of infinite space, the solution for the propagator U_A can be written symbolically as

$$U_A(t,t_0)=X\exp\left[\int_{t_0}^{t}dt'\,(I-A)\Omega(t')\right], \tag{2.7}$$

where X denotes a time-ordering operator. [In general, the solution (2.7) must be modified to account for boundary conditions.] Inserting (2.5) into (2.2) results in the equation

$$\frac{\partial}{\partial t}\mathcal{E}\{\mu(t)\}=\Omega_0(t)\mathcal{E}\{\mu(t)\}+A\Omega_1(t)U_A(t,t_0)\delta\mu(t_0)+\int_{t_0}^{t}dt'\,\mathcal{E}\{\Omega_1(t)U_A(t,t')\Omega_1(t')\}\mathcal{E}\{\mu(t')\}. \tag{2.8}$$

In order to integrate (2.4), a propagator, $W(t, t_0)$, is introduced first by means of the equation

$$\frac{\partial}{\partial t} W(t, t_0) = \Omega_0(t) W(t, t_0), \quad t \geq t_0, \tag{2.9a}$$

$$W(t_0, t_0) = I, \tag{2.9b}$$

whose solution, for an unbounded region, can be formally written as follows:

$$W(t, t_0) = X \exp[\int_{t_0}^{t} dt' \, \Omega_0(t')]. \tag{2.10}$$

In terms of this propagator, the integral of (2.4) is given by

$$\delta\mu(t) = W(t, t_0)\delta\mu(t_0) + \int_{t_0}^{t} dt' \, W(t, t') \times [(I - A)\Omega_1(t')\delta\mu(t') + \Omega_1(t')\mathcal{E}\{\mu(t')\}]. \tag{2.11}$$

Iterating the last expression, we obtain the explicit solution

$$\delta\mu(t) = \Lambda_W(t, t_0)\delta\mu(t_0) + \int_{t_0}^{t} dt' \, \Lambda_W(t, t')\Omega_1(t')\mathcal{E}\{\mu(t')\}, \tag{2.12}$$

where

$$\Lambda_W(t, t_0) = \sum_{n=0}^{\infty} [\int_{t_0}^{t} dt' \, W(t, t')(I - A)\Omega_1(t')]^n W(t, t_0). \tag{2.13}$$

Finally, substituting (2.12) into (2.2) we arrive at the following alternative equation for the first moment:

$$\frac{\partial}{\partial t}\mathcal{E}\{\mu(t)\} = \Omega_0(t)\mathcal{E}\{\mu(t)\} + A\Omega_1(t)\Lambda_W(t, t_0)\delta\mu(t_0) + \int_{t_0}^{t} dt' \, \mathcal{E}\{\Omega_1(t)\Lambda_W(t, t')\Omega_1(t')\}\mathcal{E}\{\mu(t')\}. \tag{2.14}$$

The formal expressions (2.8) and (2.14) derived by means of a nonperturbative statistical approach are valid for both weak and strong random fluctuations. It should be pointed out, however, that neither (2.8) nor (2.14) constitutes a closed equation for $\mathcal{E}\{\mu(t)\}$. This would definitely be the case if Ω were a linear operator. Here, however, Ω depends on the field μ by assumption.

In the sequel we shall make extensive use of expressions (2.2), (2.5), (2.8), (2.11), and (2.14). For the sake of simplicity we shall neglect in these relations the parts proportional to $\delta\mu(t_0)$. It must be emphasized, however, that this is a matter of convenience only and it will not detract from the generality of the formalism which will be developed in the following sections.

3. EXTENSION OF THE WEINSTOCK-BALESCU-MISGUICH FORMALISM TO SECOND MOMENTS

The exact, nonperturbative, statistical Weinstock–Balescu–Misguich formalism outlined in the previous section culminated in the derivations of two alternative equations for the first moment [cf. Eqs. (2.8) and (2.14)] which, as pointed out earlier, are not closed by virtue of the nonlinearity of the stochastic operator Ω.

In this section we shall extend the work of Weinstock, Balescu, and Misguich in order to derive two equivalent representations for the second moment [analogous to Eqs. (2.8) and (2.14)] using, again, an exact, nonperturbative, statistical approach.

Let us assume that μ depends on a set of variables $\mathbf{s}$ and on time, viz., $\mu = \mu(\mathbf{s}, t)$. Moreover, let $\Omega = \Omega[\mathbf{s}, \partial/\partial\mathbf{s}, t, \mu(\mathbf{s}, t)]$.[11] [If there is not danger for ambiguity, we shall use in the subsequent discussion the shorter notation $\mu = \mu(t)$ and $\Omega = \Omega(\mathbf{s}, t)$].

Next consider the quantity $R(t, t') \equiv \mu(\mathbf{s}, t)\mu(\mathbf{s}', t')$. Differentiating it with respect to t and using the original Eq. (2.1) for $\mu(\mathbf{s}, t)$ results in the following equation:

$$\frac{\partial}{\partial t} R(t, t') = \Omega(\mathbf{s}, t) R(t, t'). \tag{3.1}$$

Similarly, differentiating $R(t, t')$ with respect to t' and using the equation of evolution for $\mu(\mathbf{s}'t')$, i.e., $(\partial/\partial t')\mu(\mathbf{s}', t') = \Omega(\mathbf{s}', t')\mu(\mathbf{s}', t')$, we obtain the expression

$$\frac{\partial}{\partial t'} R(t, t') = \Omega(\mathbf{s}', t') R(t, t'). \tag{3.2}$$

It should be noted that both Eqs. (3.1) and (3.2) are of the general form (2.1); hence, the Weinstock–Balescu–Misguich formalism introduced in the previous section is applicable. However, several basic departures from their general theory have to be made in order to account for the simultaneous manipulation of Eqs. (3.1) and (3.2).

Operating on (3.1) with the operator A yields the following equation for the coherent part of $R(t, t')$:

$$\frac{\partial}{\partial t}\mathcal{E}\{R(t, t')\} = \Omega_0(\mathbf{s}, t)\mathcal{E}\{R(t, t')\} + A\Omega_1(\mathbf{s}, t)\delta R(t, t'). \tag{3.3}$$

On the other hand, corresponding to (3.1) and (3.2), respectively, and using the Weinstock formulation (2.3), we obtain the following two equations for the fluctuating part of $R(t, t')$:

$$\frac{\partial}{\partial t}\delta R(t, t') = (I - A)\Omega(\mathbf{s}, t)\delta R(t, t') + \Omega_1(\mathbf{s}, t)\mathcal{E}\{R(t, t')\}, \tag{3.4}$$

$$\frac{\partial}{\partial t'}\delta R(t, t') = (I - A)\Omega(\mathbf{s}', t')\delta R(t, t') + \Omega_1(\mathbf{s}', t') \times \mathcal{E}\{R(t, t')\}. \tag{3.5}$$

Proceeding as in (2.5), Eq. (3.4) can be formally integrated as follows:

$$\delta R(t, t') = U_A(\mathbf{s}, t, t_0)\delta R(t_0, t') + \int_{t_0}^{t} d\tau \, U_A(\mathbf{s}, t, \tau)\Omega_1(\mathbf{s}, \tau)\mathcal{E}\{R(\tau, t')\}. \tag{3.6}$$

Analogously, the integral of (3.5) (evaluated at $t = t_0$) is given by

$$\delta R(t_0, t') = U_A(\mathbf{s}', t', t_0)\delta R(t_0, t_0) + \int_{t_0}^{t'} d\tau \, U_A(\mathbf{s}', t', \tau)\Omega_1(\mathbf{s}', \tau)\mathcal{E}\{R(t_0, \tau)\}. \tag{3.7}$$

The first part of the right-hand side of (3.7), i.e., the term proportional to $\delta R(t_0, t_0)$, is neglected for convenience. [It should be stressed, however, that in contradistinction to the initial value term $\delta\mu(t_0)$ in (2.5) and (2.8), which has been shown by Weinstock to be negligible for large t, no justification has ever been made for the neglect of such terms as $\delta R(t_0, t_0)$.] The resulting expression is introduced next in Eq. (3.6),

$$\delta R(t,t') = \int_{t_0}^{t} d\tau\, U_A(\mathbf{s},t,\tau)\Omega_1(\mathbf{s},\tau)\mathcal{E}\{R(\tau,t')\} + \int_{t_0}^{t'} d\tau\, U_A(\mathbf{s},t,t_0)U_A(\mathbf{s}',t',\tau)\Omega_1(\mathbf{s}',\tau)\mathcal{E}\{R(t_0,\tau)\}. \tag{3.8}$$

This expression for $\delta R(t,t')$ is substituted next into (3.3) in order to obtain the final form of the equation for the second moment [analogous to Eq. (2.8) for the first moment],

$$\frac{\partial}{\partial t}\mathcal{E}\{R(t,t')\} = \Omega_0(\mathbf{s},t)\mathcal{E}\{R(t,t')\} + \int_{t_0}^{t} d\tau\,\mathcal{E}\{\Omega_1(\mathbf{s},t)U_A(\mathbf{s},t,\tau)\Omega_1(\mathbf{s},\tau)\} \times\mathcal{E}\{R(\tau,t')\} + \int_{t_0}^{t'} d\tau\,\mathcal{E}\{\Omega_1(\mathbf{s},t)U_A(\mathbf{s},t,t_0) \times U_A(\mathbf{s}',t',\tau)\Omega_1(\mathbf{s}',\tau)\}\mathcal{E}\{R(t_0,\tau)\}. \tag{3.9}$$

In order to derive an alternative equation for the second moment [analogous to Eq. (2.14) for the first moment], we proceed as follows: Corresponding to (3.1) and (3.2), respectively, and using (2.4), we have

$$\frac{\partial}{\partial t}\delta R(t,t') = \Omega_0(\mathbf{s},t)\delta R(t,t') + (I-A)\Omega_1(\mathbf{s},t)\delta R(t,t') + \Omega_1(\mathbf{s},t)\mathcal{E}\{R(t,t')\}, \tag{3.10}$$

$$\frac{\partial}{\partial t'}\delta R(t,t') = \Omega_0(\mathbf{s}',t')\delta R(t,t') + (I-A)\Omega_1(\mathbf{s}',t')\delta R(t,t') + \Omega_1(\mathbf{s}',t')\mathcal{E}\{R(t,t')\}. \tag{3.11}$$

Proceeding as in (2.12), Eq. (3.10) can be integrated formally as follows:

$$\delta R(t,t') = \Lambda_W(\mathbf{s},t,t_0)\delta R(t_0,t') + \int_{t_0}^{t} d\tau\,\Lambda_W(\mathbf{s},t,\tau)\Omega_1(\mathbf{s},\tau)\mathcal{E}\{R(\tau,t')\}. \tag{3.12}$$

Similarly, the integral of (3.11) (evaluated at $t=t_0$) is given by

$$\delta R(t_0,t') = \Lambda_W(\mathbf{s}',t',t_0)\delta R(t_0,t_0) + \int_{t_0}^{t'} d\tau\,\Lambda_W(\mathbf{s}',t',\tau)\Omega_1(\mathbf{s}',\tau)\mathcal{E}\{R(t_0,\tau)\}. \tag{3.13}$$

Neglecting the first term on the right-hand side of (3.13) and using the resulting expression for $\delta R(t_0,t')$ in conjunction with (3.12), we find that

$$\delta R(t,t') = \int_{t_0}^{t} d\tau\,\Lambda_W(\mathbf{s},t,\tau)\Omega_1(\mathbf{s},\tau)\mathcal{E}\{R(\tau,t')\} + \int_{t_0}^{t'} d\tau\,\Lambda_W(\mathbf{s},t,t_0)\Lambda_W(\mathbf{s}',t',\tau)\Omega_1(\mathbf{s}',\tau)\mathcal{E}\{R(t_0,\tau)\}. \tag{3.14}$$

Finally, inserting (3.14) into (3.3) we obtain the desired alternative equation for the second moment,

$$\frac{\partial}{\partial t}\mathcal{E}\{R(t,t')\} = \Omega_0(\mathbf{s},t)\mathcal{E}\{R(t,t')\} + \int_{t_0}^{t} d\tau\,\mathcal{E}\{\Omega_1(\mathbf{s},t)\Lambda_W(\mathbf{s},t,\tau)\Omega_1(\mathbf{s},\tau)\} \times\mathcal{E}\{R(\tau,t')\} + \int_{t_0}^{t'} d\tau\,\mathcal{E}\{\Omega_1(\mathbf{s},t)\Lambda_W(\mathbf{s},t,t_0) \times\Lambda_W(\mathbf{s}',t',\tau)\Omega_1(\mathbf{s}',\tau)\}\mathcal{E}\{R(t_0,\tau)\}. \tag{3.15}$$

We close this section with two important remarks: (1) As in the case of Eqs. (2.8) and (2.14) for the first moment, neither of the equivalent equations (3.9) and (3.15) for the second moment is closed, again because of the nonlinearity of the stochastic operator Ω; (2) the procedure outlined in this section can obviously be extended in order to derive equations for higher moments.

4. CLOSED FIRST- AND SECOND-ORDER MOMENT EQUATIONS

The exact nonperturbative results contained in the previous two sections are valid for an arbitrary nonlinear stochastic operator Ω. In the sequel we shall restrict the discussion to the special case that Ω depends linearly on the field μ. The class of nonlinear stochastic equations (2.1) spanned by Ω under this assumption includes two physically important problems: (1) model hydrodynamic turbulence, and (2) plasma turbulence.

A. Direct-interaction approximation

In the first part of this section we shall present a procedure for obtaining a complete set of closed equations for the first two moments of the field μ in the framework of an approximation corresponding to Kraichnan's direct-interaction approximation.

We introduce the propagator

$$U(t,t_0) = X\exp\left[\int_{t_0}^{t} dt'\,\Omega(t')\right] \tag{4.1}$$

as the solution of the initial value problem

$$\frac{\partial}{\partial t}U(t,t_0) = \Omega(t)U(t,t_0), \quad t \geq t_0, \tag{4.2a}$$

$$U(t_0,t_0) = I \tag{4.2b}$$

in an unbounded domain. This, of course, is the fundamental problem associated with (2.1), viz.,

$$\mu(t) = U(t,t_0)\mu(t_0), \tag{4.3}$$

whence

$$\mathcal{E}\{\mu(t)\} = AU(t,t_0)\mu(t_0), \tag{4.4a}$$

$$\delta\mu(t) = (I-A)U(t,t_0)\mu(t_0), \tag{4.4b}$$

and, consequently, since $\mu(t_0)$ is specified, $\mathcal{E}\{U(t,t_0)\}$ may be chosen as the basic quantity in the place of $\mathcal{E}\{\mu(t)\}$.

Equation (2.8) for the first moment can be expressed in terms of $U(t,t_0)$, instead of $\mu(t)$, by substituting (4.4) as follows:

$$\frac{\partial}{\partial t}AU(t,t_0) = \Omega_0(t)AU(t,t_0) + A\Omega_1(t)U_A(t,t_0)(I-A) + \int_{t_0}^{t} dt'\,\mathcal{E}\{\Omega_1(t)U_A(t,t')\Omega_1(t')\}AU(t',t_0). \tag{4.5}$$

Weinstock has established that in the weak-coupling approximation,

$$U_A(t, t_0)(I-A) \approx (I-A)\mathcal{E}\{U(t, t_0)\}, \tag{4.6}$$

$$U(t, t_0) \approx \mathcal{E}\{U(t, t_0)\}. \tag{4.7}$$

Assuming for simplicity that $\mu(t_0)$ is deterministic, (4.5) simplifies in this case to

$$\frac{\partial}{\partial t}\mathcal{E}\{U(t, t_0)\} = \Omega_0(t)\mathcal{E}\{U(t, t_0)\} + \int_{t_0}^{t} dt' \mathcal{E}\{\Omega_1(t)\mathcal{E}\{U(t, t')\}\Omega_1(t')\}\mathcal{E}\{U(t', t_0)\} \tag{4.8}$$

and (3.9) assumes the form

$$\frac{\partial}{\partial t}\mathcal{E}\{R(t, t')\} = \Omega_0(\mathbf{s}, t)\mathcal{E}\{R(t, t')\} + \int_{t_0}^{t} d\tau \mathcal{E}\{\Omega_1(\mathbf{s}, t)\mathcal{E}\{U(\mathbf{s}, t, \tau)\}\Omega_1(\mathbf{s}, \tau)\}\mathcal{E}\{R(\tau, t')\} + \int_{t_0}^{t'} d\tau \mathcal{E}\{\Omega_1(\mathbf{s}, t)\mathcal{E}\{U(\mathbf{s}', t', \tau)\}\Omega_1(\mathbf{s}', \tau)\}\mathcal{E}\{R(t, \tau)\}. \tag{4.9}$$

In deriving this equation we have made use of the fact that the propagators $U_A(\mathbf{s}, t, t_0)$ and $U_A(\mathbf{s}', t', \tau)$ in (3.9) commute. Furthermore, we have used the relations

$$U_A(\mathbf{s}, t, t_0)(I-A)\mathcal{E}\{R(t_0, \tau)\} \approx (I-A)\mathcal{E}\{U(\mathbf{s}, t, t_0)\}\mathcal{E}\{R(t_0, \tau)\} \approx (I-A)\mathcal{E}\{U(\mathbf{s}, t, t_0)R(t_0, \tau)\} = (I-A)\mathcal{E}\{R(t, \tau)\}, \tag{4.10}$$

the last equality following from the semigroup property of the propagator U.

Equations (4.8) and (4.9), together with (2.2), form a complete set of closed equations for the smoothed quantities $\mathcal{E}\{\mu(t)\}$, $\mathcal{E}\{R(t, t')\}$, and $\mathcal{E}\{U(\mathbf{s}, t, t_0)\}$. It should be noted that since $\Omega_1 \sim \delta\mu$ by assumption, terms proportional to the covariance $\mathcal{E}\{\delta\mu(t)\delta\mu(t')\}$ appear in Eqs. (4.8), (4.9), and (2.2). However, making use of the formula $\mathcal{E}\{\mu(t)\mu(t')\} = (\mathcal{E}\{R(t, t')\}) = \mathcal{E}\{\mu(t)\}\mathcal{E}\{\mu(t')\} + \mathcal{E}\{\delta\mu(t)\delta\mu(t')\}$, the covariance function can be expressed in terms of first and second moments.

The resolution of the closure problem in the weak-coupling limit presented here has been achieved at a level of approximation corresponding to Kraichnan's direct-interaction approximation; hence the title of this subsection. It should be emphasized that in contradistinction to Kraichnan's direct-interaction approximation which is based on a stochastic modeling scheme, our technique has been developed along the lines of a modified Weinstock—Balescu—Misguich formalism. Also, whereas in Kraichnan's work the main results are expressed in terms of closed equations for the mean field, the covariance, and an averaged Green's function, our results are given in terms of closed equations for the first two moments and the mean propagator $\mathcal{E}\{U\}$.[12]

We wish to close this subsection with a few remarks concerning the difference between our technique for closing the equations for the first two moments and that reported recently by Misguich and Balescu (cf. Ref. 6). Their method is directly specialized to the problem of Vlasov-plasma turbulence with an external stochastic electric field. It is presented here in a more general setting so that comparisons with our work can be made more easily.

Starting from Eq. (2.5) for the fluctuating part of μ, viz.,

$$\delta\mu(t) = \int_{t_0}^{t} dt' \, U_A(t, t')\Omega_1(t')\mathcal{E}\{\mu(t')\}, \tag{4.11}$$

where the part proportional to $\delta\mu(t_0)$ is neglected for simplicity, they obtain in the weak-coupling limit

$$\delta\mu(t) = \int_{t_0}^{t} dt' \, \mathcal{E}\{U(t, t')\}\Omega_1(t')\mathcal{E}\{\mu(t')\}, \tag{4.12}$$

and, analogously,

$$\delta\mu(\tau) = \int_{t_0}^{\tau} dt'' \, \mathcal{E}\{U(\tau, t'')\}\Omega_1(t'')\mathcal{E}\{\mu(t'')\}. \tag{4.13}$$

From the last two expressions, a relationship is set up for the covariance,

$$\mathcal{E}\{\delta\mu(t)\delta\mu(\tau)\} = \int_{t_0}^{t} dt' \int_{t_0}^{\tau} dt'' \, \mathcal{E}\{U(t, t')\}\mathcal{E}\{\Omega_1(t')\mathcal{E}\{\mu(t')\} \times \mathcal{E}\{U(\tau, t'')\}\Omega_1(t'')\}\mathcal{E}\{\mu(t'')\}. \tag{4.14}$$

This equation contains only covariances and mean fields; therefore, together with the equation for the mean propagator $\mathcal{E}\{U(t, t')\}$ [cf. Eq. (4.8)] and the equation for the mean field (in what they call the renormalized quasilinear approximation), viz.,

$$\frac{\partial}{\partial t}\mathcal{E}\{\mu(t)\} = \Omega_0(t)\mathcal{E}\{\mu(t)\} + \int_{t_0}^{t} dt' \, \mathcal{E}\{\Omega_1(t)\mathcal{E}\{U(t, t')\}\Omega_1(t')\}\mathcal{E}\{\mu(t')\}, \tag{4.15}$$

constitutes a self-consistent set.

This type of closure, when specialized to linear stochastic problems considered in the first-order smoothing approximation ($U_A \to W$; also cf. the next subsection), has been criticized by Morrison and McKenna (cf. Ref. 7). We feel that our approach to the closure problem, based on a renormalization at the level of the second moment, is fundamentally different from that of Misguich and Balescu and is devoid of the aforementioned difficulties.

B. Quasilinear approximation

In this subsection we shall outline a procedure for closing the equations for the first two moments within the confines of the quasilinear approximation. The latter is essentially a perturbational method at a level lower than the direct-interaction approximation discussed earlier. It is applicable for small correlations and corresponds to retaining only the zero-order term in the series expansion (2.13), viz., $\Lambda_W \to W$. If this approximate expression for Λ_W is substituted into (3.15), one has[13]

$$\frac{\partial}{\partial t}\mathcal{E}\{R(t, t')\} = \Omega_0(\mathbf{s}, t)\mathcal{E}\{R(t, t')\}$$

$$+\int_{t_0}^{t} d\tau\, \mathcal{E}\{\Omega_1(\mathbf{s},t)W(\mathbf{s},t,\tau)\Omega_1(\mathbf{s},\tau)\}\mathcal{E}\{R(\tau,t')\}$$
$$+\int_{t_0}^{t'} d\tau\, \mathcal{E}\{\Omega_1(\mathbf{s},t)W(\mathbf{s}',t',\tau)\Omega_1(\mathbf{s}',\tau)\}\mathcal{E}\{R(t,\tau)\}. \tag{4.16}$$

In deriving this equation we have made use of the commutation of the operators $W(\mathbf{s}',t',\tau)$ and $W(\mathbf{s},t,t_0)$, as well as the semigroup property $W(\mathbf{s},t,t_0)\mathcal{E}\{R(t_0,\tau)\} = \mathcal{E}\{W(\mathbf{s},t,t_0)R(t_0,\tau)\} \approx \mathcal{E}\{R(t,\tau)\}$. (The last equality is valid only in the quasilinear approximation.)

Equations (4.16) and (2.9), together with (2.2), form a complete set of closed equations for the averaged quantities $\mathcal{E}\{\mu(t)\}$, $\mathcal{E}\{R(t,t')\}$, and $W(\mathbf{s},t,t_0)$.

5. HYDRODYNAMIC TURBULENCE

Let the motion of a fluid be described by the Navier–Stokes equations

$$\left(\frac{\partial}{\partial t} - \nu\nabla^2 + \mathbf{v}(\mathbf{x},t)\cdot\frac{\partial}{\partial \mathbf{x}}\right)v_i(\mathbf{x},t) = -\frac{\partial}{\partial x_i}p(\mathbf{x},t) + f_i(\mathbf{x},t), \tag{5.1a}$$

$i=1,2,3$, and the incompressibility condition

$$\frac{\partial}{\partial x_i}v_i(\mathbf{x},t)=0. \tag{5.1b}$$

Here, $\mathbf{v}(\mathbf{x},t)$ is the fluid velocity vector field, ν is the kinematic viscosity, $\mathbf{f}(\mathbf{x},t)$ is an externally supplied vector forcing function, and $p(\mathbf{x},t)$ is the pressure divided by the density.

It is well known that at high Reynolds numbers the character of the fluid motion changes from laminar to turbulent. It is believed that the "chaotic" or turbulent flow is described adequately by the Navier–Stokes equations, the solutions of which (at high Reynolds numbers) are extremely unstable. Predictions concerning the turbulent flow on the basis of the Navier–Stokes equations would require the specification of initial conditions with unrealistic accuracy. Because of this, it is of interest to determine equations for smooth, mean quantities, such as the first and second moments of the velocity field. (The notion "mean" is used here synonymously with the ensemble average over various realizations of the same flow with different initial conditions.)

For simplicity we are going to deal with a "model" hydrodynamic turbulence, assuming that the pressure is uniform throughout the fluid volume. We shall also neglect the external force in (5.1a). The simplified Navier–Stokes equations are then of the general form (2.1), viz.,

$$\frac{\partial}{\partial t}v_i(\mathbf{x},t)=\Omega v_i(\mathbf{x},t) \tag{5.2a}$$

with

$$\Omega=\nu\nabla^2-\mathbf{v}(\mathbf{x},t)\cdot\frac{\partial}{\partial\mathbf{x}}. \tag{5.2b}$$

The mean and fluctuating parts of the operator Ω can be readily written down as follows:

$$\Omega_0=\nu\nabla^2-\mathcal{E}\{\mathbf{v}(\mathbf{x},t)\}\cdot\frac{\partial}{\partial\mathbf{x}}, \tag{5.3a}$$

$$\Omega_1=-\delta\mathbf{v}(\mathbf{x},t)\cdot\frac{\partial}{\partial\mathbf{x}}. \tag{5.3b}$$

A. Direct-interaction approximation

We are now in a position to write down explicitly a complete set of closed equations for the first two moments of the velocity field in the direct-interaction approximation (cf. Sec. 4).

The equation for the mean velocity field [cf. Eq. (2.2)] assumes the form

$$\left(\frac{\partial}{\partial t}-\nu\nabla^2+\mathcal{E}\{\mathbf{v}(\mathbf{x},t)\}\cdot\frac{\partial}{\partial\mathbf{x}}\right)\mathcal{E}\{v_i(\mathbf{x},t)\}$$
$$=-\mathcal{E}\{\delta\mathbf{v}(\mathbf{x},t)\cdot(\partial/\partial\mathbf{x})\delta v_i(\mathbf{x},t)\}. \tag{5.4}$$

Corresponding to Eq. (4.9) for the second moment, we have in this case

$$\left(\frac{\partial}{\partial t}-\nu\nabla^2+\mathcal{E}\{\mathbf{v}(\mathbf{x},t)\}\cdot\frac{\partial}{\partial\mathbf{x}}\right)\mathcal{E}\{R_{ij}(\mathbf{x},\mathbf{x}',t,t')\}$$
$$=\frac{\partial}{\partial x_k}\int_{t_0}^{t}d\tau\,\mathcal{E}\{U(t,\tau)\mathcal{E}\{\delta v_k(\mathbf{x},t)\delta v_l(\mathbf{x},\tau)\}$$
$$\times\frac{\partial}{\partial x_l}\mathcal{E}\{R_{ij}(\mathbf{x},\mathbf{x}',\tau,t')\}+\frac{\partial}{\partial x_k}\int_{t_0}^{t'}d\tau\,\mathcal{E}\{U(t',\tau)\}$$
$$\times\mathcal{E}\{\delta v_k(\mathbf{x},t)\delta v_l(\mathbf{x}',\tau)\}\frac{\partial}{\partial x_l'}\mathcal{E}\{R_{ij}(\mathbf{x},\mathbf{x}',t,\tau)\}, \tag{5.5}$$

where $\mathcal{E}\{R_{ij}(\mathbf{x},\mathbf{x}',t,t')\}=\mathcal{E}\{v_i(\mathbf{x},t)v_j(\mathbf{x}',t')\}$. Finally, the equation for the mean propagator $\mathcal{E}\{U(t,t_0)\}$ [cf. Eq. (4.8)] becomes in this case

$$\left(\frac{\partial}{\partial t}-\nu\nabla^2+\mathcal{E}\{\mathbf{v}(\mathbf{x},t)\}\cdot\frac{\partial}{\partial\mathbf{x}}\right)\mathcal{E}\{U(t,t_0)\}$$
$$=\frac{\partial}{\partial x_i}\int_{t_0}^{t}d\tau\,\mathcal{E}\{U(t,\tau)\}\mathcal{E}\{\delta v_i(\mathbf{x},t)\delta v_j(\mathbf{x},\tau)\}$$
$$\times\frac{\partial}{\partial x_j}\mathcal{E}\{U(\tau,t_0)\}. \tag{5.6}$$

In the derivation of (5.5) and (5.6) we made use of the incompressibility condition (5.1b).[14] The closure of Eqs. (5.4)–(5.6) is more clearly evident on recalling the formula $\mathcal{E}\{\delta v_i(\mathbf{x},t)\delta v_j(\mathbf{x}',t')\}=\mathcal{E}\{R_{ij}(\mathbf{x},\mathbf{x}',t,t')\} - \mathcal{E}\{v_i(\mathbf{x},t)\}\mathcal{E}\{v_j(\mathbf{x}',t)\}$.

B. Quasilinear approximation

In order to write a closed set of equations for the first two moments of the velocity field within the region of applicability of the quasilinear approximation (cf. Sec. 4), Eq. (5.4) for the mean velocity field is retained unaltered; however, in Eq. (5.5) for the second moment, $\mathcal{E}\{U(t,t')\}$ must be replaced by the propagator $W(t,t')$ which, in turn, satisfies in the case of hydrodynamic turbulence the following equation [also cf. Eq. (2.9)]:

$$\left(\frac{\partial}{\partial t}-\nu\nabla^2+\mathcal{E}\{\mathbf{v}(\mathbf{x},t)\}\cdot\frac{\partial}{\partial\mathbf{x}}\right)W(t,t')=0,\quad t\geq t_0, \tag{5.7a}$$

$$W(t_0,t_0)=I. \tag{5.7b}$$

The solution of (5.7) for an unbounded region can be formally written as follows:

$$W(t,t_0)=X\exp\left[\int_{t_0}^{t}dt'\left(\nu\nabla^2-\mathcal{E}\{\mathbf{v}(\mathbf{x},t')\}\cdot\frac{\partial}{\partial\mathbf{x}}\right)\right]. \tag{5.8}$$

The basic closed equations for the first two moments derived on the basis of the quasilinear approximation by means of the procedure outlined above are valid for the general case of nonstationary, inhomogeneous, and anisotropic turbulence. These equations simplify considerably by introducing additional constraints.

As an illustration, we consider here the case of stationary turbulence. In this special case we need only be concerned with the equation of evolution of the correlation tensor of the velocity field, $\Gamma_{ij}(\mathbf{x}, \mathbf{x}', \tau) \equiv \mathcal{E}\{v_i(\mathbf{x}, t) \times v_j(\mathbf{x}', t-\tau)\}$. [The quantity $\mathcal{E}\{v_i(\mathbf{x}, t)\} = \mathcal{E}\{v_i(\mathbf{x}, t_0)\}$ is assumed to be given.] We put, also, $\nu = 0$ in the expression (5.8) for the propagator $W(t, t_0)$. This means that we neglect the effect of viscosity on turbulence, but not on dissipation, an assumption that seems reasonable for well-developed turbulence. Under these restrictions, Eq. (5.8) reduces to

$$W(t, t_0) = \exp\left[\left(-\mathcal{E}\{\mathbf{v}(\mathbf{x}, t_0)\} \cdot \frac{\partial}{\partial \mathbf{x}}\right)(t - t_0)\right] \tag{5.9}$$

(the time-ordering operator is the identity operator in this case), and the correlation tensor evolves in time as follows:

$$\left(\frac{\partial}{\partial t} - \nu\nabla^2 + \mathcal{E}\{\mathbf{v}(\mathbf{x}, t_0)\} \cdot \frac{\partial}{\partial \mathbf{x}}\right)\Gamma_{ij}(\mathbf{x}, \mathbf{x}', \tau)$$
$$= \int_0^\tau dt\, \mathcal{E}\left\{\delta v_k(\mathbf{x}, \tau)\frac{\partial}{\partial x_k}\exp\left[\left(-\mathcal{E}\{\mathbf{v}(\mathbf{x}, t_0)\} \cdot \frac{\partial}{\partial \mathbf{x}}\right)\right.\right.$$
$$\left.\left. \times (\tau - t)\right]\delta v_l(\mathbf{x}, t)\right\}\frac{\partial}{\partial x_l}\Gamma_{ij}(\mathbf{x}, \mathbf{x}', t). \tag{5.10}$$

In order to evaluate explicitly the effect of the operator $\exp[-(\mathcal{E}\{\mathbf{v}\} \cdot (\partial/\partial \mathbf{x}))(\tau - t)]$ in (5.10) we will make additional assumptions, concerning the mean flow. We recall that $\exp[a(\partial/\partial x)]f(x) = f(x+a)$, and $\exp(A+B) = \exp A \exp B$ provided that $[A, B]_- = 0$. We assume that the exponential operator in (5.10) is factorizable, i.e., the commutators are close to zero $[(\partial/\partial x_j)\mathcal{E}\{v_i\} \approx 0, i \neq j]$ or, for example, that we have a parallel mean flow, e.g., $\mathcal{E}\{\mathbf{v}\} = (0, 0, \mathcal{E}\{v_3(\mathbf{x}, t_0)\})$. In this case (5.10) reduces to the simpler equation

$$\left(\frac{\partial}{\partial t} - \nu\nabla^2 + \mathcal{E}\{\mathbf{v}(\mathbf{x}, t_0)\} \cdot \frac{\partial}{\partial \mathbf{x}}\right)\Gamma_{ij}(\mathbf{x}, \mathbf{x}', \tau)$$
$$= + \int_0^\tau dt \frac{\partial}{\partial x_k}\Gamma_{kl}[\mathbf{x}, \mathbf{x} - \mathcal{E}\{\mathbf{v}\}(\tau - t), \tau - t]$$
$$\times \frac{\partial}{\partial x_l}\Gamma_{ij}[\mathbf{x} - \mathcal{E}\{\mathbf{v}\}(\tau - t), \mathbf{x}', t] - \int_0^\tau dt\, \mathcal{E}\{v_k(\mathbf{x}, t_0)\}$$
$$\times \frac{\partial}{\partial x_k}\mathcal{E}\{v_l[\mathbf{x} - \mathcal{E}\{\mathbf{v}\}(\tau - t), t_0]\}$$
$$\times \frac{\partial}{\partial x_l}\Gamma_{ij}[\mathbf{x} - \mathcal{E}\{\mathbf{v}\}(\tau - t), \mathbf{x}', t]. \tag{5.11}$$

This equation for the correlation tensor is rendered closed on specifying the initial condition $\Gamma_{ij}(\mathbf{x}, \mathbf{x}', 0)$.

6. VLASOV-PLASMA TURBULENCE WITH AN EXTERNAL STOCHASTIC ELECTRIC FIELD

A collisionless plasma in the presence of an external stochastic electric field is governed by the one-species, self-consistent Vlasov—Poisson equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{x}} + \frac{e}{m}[\mathbf{E}^s(\mathbf{x}, t) + \mathbf{E}^e(\mathbf{x}, t)] \cdot \frac{\partial}{\partial \mathbf{v}}\right)f(\mathbf{x}, \mathbf{v}, t) = 0, \tag{6.1a}$$

$$\nabla \cdot \mathbf{E}^s(\mathbf{x}, t) = 4\pi e \int_{R^3} d\mathbf{v}\,[f(\mathbf{x}, \mathbf{v}, t) - n_0\delta(\mathbf{v})]. \tag{6.1b}$$

Here, $f(\mathbf{x}, \mathbf{v}, t)$ is the particle distribution function, e and m are the charge and mass of the particle, respectively, n_0 is the density of a uniform background of neutralizing charge, $\mathbf{E}^s(\mathbf{x}, t)$ is the self-electric field, and $\mathbf{E}^e(\mathbf{x}, t)$ denotes the external stochastic electric field.

Equations (6.1) can be brought into the general form (2.1), viz.,

$$\frac{\partial}{\partial t}f(\mathbf{x}, \mathbf{v}, t) = \Omega f(\mathbf{x}, \mathbf{v}, t), \tag{6.2a}$$

with

$$\Omega = -\mathbf{v} \cdot \frac{\partial}{\partial \mathbf{x}} - \frac{e}{m}[\mathbf{E}^s(\mathbf{x}, t) + \mathbf{E}^e(\mathbf{x}, t)] \cdot \frac{\partial}{\partial \mathbf{v}}, \tag{6.2b}$$

on introducing the relationship

$$\mathbf{E}^s(\mathbf{x}, t) = L(\mathbf{x})[f(\mathbf{x}, \mathbf{v}, t) - n_0\delta(\mathbf{v})], \tag{6.3a}$$

where the operator $L(\mathbf{x})$ is defined by

$$L(\mathbf{x})f(\mathbf{x}, \mathbf{v}, t)$$
$$= 4\pi e \frac{\partial}{\partial \mathbf{x}} \int_{R^3} d\mathbf{x}' \int_{R^3} d\mathbf{v}'\, Q(\mathbf{x}, \mathbf{x}')f(\mathbf{x}', \mathbf{v}', t). \tag{6.3b}$$

$Q(\mathbf{x}, \mathbf{x}')$ is the Green's function for the Poisson equation. In the case of an infinite plasma,

$$Q(\mathbf{x}, \mathbf{x}') = -\frac{1}{4\pi}\frac{1}{|\mathbf{x} - \mathbf{x}'|}. \tag{6.3c}$$

The coherent and fluctuating parts of the operator Ω [cf. Eq. (6.2b)] are given as follows:

$$\Omega_0 = -\mathbf{v} \cdot \frac{\partial}{\partial \mathbf{x}} - \frac{e}{m}[\mathcal{E}\{\mathbf{E}^s(\mathbf{x}, t)\} + \mathcal{E}\{\mathbf{E}^e(\mathbf{x}, t)\}] \cdot \frac{\partial}{\partial \mathbf{v}}, \tag{6.4a}$$

$$\Omega_1 = -\frac{e}{m}[\delta\mathbf{E}^s(\mathbf{x}, t) + \delta\mathbf{E}^e(\mathbf{x}, t)] \cdot \frac{\partial}{\partial \mathbf{v}}. \tag{6.4b}$$

A. Direct-interaction approximation

We next present a complete set of closed equations for the first two moments of the particle distribution function in the direct-interaction approximation.

The equation for the mean distribution function is given by

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{x}} + \frac{e}{m}\mathcal{E}\{\mathbf{E}^t(\mathbf{x}, t)\} \cdot \frac{\partial}{\partial \mathbf{v}}\right)\mathcal{E}\{f(\mathbf{x}, \mathbf{v}, t)\}$$
$$= -\frac{e}{m}\mathcal{E}\{\delta\mathbf{E}^t(\mathbf{x}, t) \cdot (\partial/\partial \mathbf{v})\delta f(\mathbf{x}, \mathbf{v}, t)\}, \tag{6.5}$$

where, for simplicity, we have used the shorter notation $\mathbf{E}^t(\mathbf{x}, t) = \mathbf{E}^s(\mathbf{x}, t) + \mathbf{E}^e(\mathbf{x}, t)$. The equation for the second moment assumes in this case the following form:

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{x}} + \frac{e}{m}\mathcal{E}\{\mathbf{E}^t(\mathbf{x}, t)\} \cdot \frac{\partial}{\partial \mathbf{v}}\right)\mathcal{E}\{R(\mathbf{s}, \mathbf{s}', t, t')\}$$
$$= \left(\frac{e}{m}\right)^2 \frac{\partial}{\partial v_i}\int_{t_0}^t d\tau\, \mathcal{E}\{U(\mathbf{s}, t, \tau)\}\mathcal{E}\{\delta E_i^t(\mathbf{x}, t)\delta E_j^t(\mathbf{x}, \tau)\}$$
$$\times \frac{\partial}{\partial v_j}\mathcal{E}\{R(\mathbf{s}, \mathbf{s}', \tau, t')\} + \left(\frac{e}{m}\right)^2 \frac{\partial}{\partial v_i}\int_{t_0}^{t'} d\tau\, \mathcal{E}\{U(\mathbf{s}', t', \tau)\}$$
$$\times \mathcal{E}\{\delta E_i^t(\mathbf{x}, t)\delta E_j^t(\mathbf{x}', \tau)\}\frac{\partial}{\partial v_j'}\mathcal{E}\{R(\mathbf{s}, \mathbf{s}', t, \tau)\}, \tag{6.6}$$

where $\mathbf{s}=(\mathbf{x},\mathbf{v})$ and $\mathcal{E}\{R(\mathbf{s},\mathbf{s}',t,t')\}=\mathcal{E}\{f(\mathbf{s},t)f(\mathbf{s}',t')\}$. Finally, the equation for the mean propagator $\mathcal{E}\{U(\mathbf{s},t,t_0)\}$ becomes in this case

$$\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\frac{\partial}{\partial \mathbf{x}}+\frac{e}{m}\mathcal{E}\{\mathbf{E}^t(\mathbf{x},t)\}\cdot\frac{\partial}{\partial \mathbf{v}}\right)\mathcal{E}\{U(\mathbf{s},t,t_0)\}$$
$$=\left(\frac{e}{m}\right)^2\frac{\partial}{\partial v_i}\int_{t_0}^t d\tau\,\mathcal{E}\{U(\mathbf{s},t,\tau)\}\mathcal{E}\{\delta E_i^t(\mathbf{x},t)\delta E_j^t(\mathbf{x},\tau)\}$$
$$\times\frac{\partial}{\partial v_j}\mathcal{E}\{U(\mathbf{s},\tau,t_0)\}. \qquad (6.7)$$

Equations (6.5)–(6.7) constitute a closed self-consistent set (1) in the absence of an external stochastic electric field, and (2) in the case that $\mathcal{E}\{\delta\mathbf{E}^e(\mathbf{x},t)\delta f(\mathbf{x},\mathbf{v},t)\}=0$. Both of these restrictions can be lifted without too much difficulty. However, we shall not pursue this issue further in this paper.

B. Quasilinear approximation

Within the domain of validity of the quasilinear approximation, Eq. (6.5) for the mean particle distribution function is retained as it stands; however, in Eq. (6.6) for the second moment, $\mathcal{E}\{U(\mathbf{s},t,t')\}$ must be replaced by the propagator $W(\mathbf{s},t,t')$ which, for the problem under consideration here, satisfies the question

$$\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\frac{\partial}{\partial \mathbf{x}}+\frac{e}{m}\mathcal{E}\{\mathbf{E}^t(\mathbf{x},t)\}\cdot\frac{\partial}{\partial \mathbf{v}}\right)W(\mathbf{s},t,t_0)=0,\quad t\geq t_0, \qquad (6.8a)$$

$$W(\mathbf{s},t_0,t_0)=I. \qquad (6.8b)$$

The solution of (6.8) for an unbounded Vlasov plasma can be formally written down as follows:

$$W(\mathbf{s},t,t_0)=X\exp\left[\int_{t_0}^t dt'\left(-\mathbf{v}\cdot\frac{\partial}{\partial \mathbf{x}}-\frac{e}{m}\mathcal{E}\{\mathbf{E}^s(\mathbf{x},t')\}\cdot\frac{\partial}{\partial \mathbf{v}}\right)\right]. \qquad (6.9)$$

The remarks at the end of the previous subsection concerning the closure of the resulting equations for the first two moments apply here as well.

ACKNOWLEDGMENTS

The research reported in this paper was completed while the first author (I. M. B.) participated in the Applied Mathematics Summer Institute, 1975 at Dartmouth College. The Institute was supported by the Office of Naval Research under Contract No. N00014-75-C-0921 with the Applied Institute of Mathematics, Inc. In the course of this research the second author (W. B. S.) was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AF0SR-74-2651A. The authors are indebted to the reviewer for his critical comments and for bringing to their attention the closely related work of J. Weinstock [Bull. APS 20, 1367 (1975)].

[1]R. H. Kraichnan, J. Fluid Mech. **5**, 497 (1959); J. Math. Phys. **2**, 124 (1961).
[2]S. A. Orszag and R. H. Kraichnan, J. Phys. Fluids **10**, 1720 (1967).
[3]A. B. Mikhailovskii, Nucl. Fusion **4**, 321 (1964); B. B. Kadomtsev, *Plasma Turbulence* (Academic, New York, 1965); T. H. Dupree, J. Phys. Fluids **9**, 1773 (1966).
[4]J. Weinstock, J. Phys. Fluids **12**, 1045 (1969); **13**, 2308 (1970); **15**, 454 (1972).
[5]R. Balescu and J. H. Misguich, J. Plasma Phys. **13**, 33 (1975).
[6]J. H. Misguich and R. Balescu, J. Plasma Phys. **13**, 385, 419, 429 (1975).
[7]J. A. Morrison and J. McKenna, J. Math. Phys. **11**, 2361 (1970).
[8]J. B. Keller, "Stochastic Equations and the Theory of Turbulence" in *Geophysical Fluid Dynamics*, course lectures and seminars (Woods Hole Oceanographic Institute, Woods Hole, Mass. 1966), pp. 131-36.
[9]The parameter α will be usually suppressed for convenience.
[10]For the sake of simplicity, we shall assume in the following discussion that Ω_1 has zero mean. This condition is stated mathematically as $A\Omega_1 A=0$.
[11]The explicit appearance of the arguments will permit the free interchange of the order of noncommuting quantities without committing an error.
[12]Strictly speaking, $\mathcal{E}\{U(\mathbf{s},t,t_0)\}=G^*$, where G is Kraichnan's averaged Green's function, and "*" is the operator of convolution with respect to the variables $\mathbf{s}$.
[13]Equation (4.16) can also be obtained directly from (4.9) since Weinstock has shown that $\mathcal{E}\{U\}\rightarrow W$ for weak correlations.
[14]On the right-hand side of the Eqs. (5.5) and (5.6), the propagator $\mathcal{E}\{U(t,\tau)\}$ operates only on those functions that have τ as the time argument.

A new analytic continuation of Appell's hypergeometric series F_2

Krishan Sud and L. E. Wright

Ohio University, Athens, Ohio
(Received 27 April 1976)

The doubly infinite series for Appell's function $F_2(\alpha, a_1, a_2, b_1, b_2; x, y)$ is written in terms of four of Appell's F_3 functions. Analytic continuations are given for the F_3 series, thereby allowing one to obtain a new analytic continuation for the F_2 series. The new doubly infinite series are all absolutely convergent if their variables satisfy $|x| < 1$ and $|y| < 1$, whereas the F_2 series is absolutely convergent only in the domain $|x| + |y| < 1$. The analytic continuations given here are very useful for evaluating the Appell F_2 series when one of the variables is near unity. In particular, our results are useful for calculating radial matrix elements over products of Dirac–Coulomb functions and the electromagnetic interaction Green's function.

I. INTRODUCTION

Radial matrix elements of the electromagnetic interaction between the states of a relativistic electron in the Coulomb field of a point nucleus can be expressed as a Laplace transform of the product of two confluent hypergeometric functions,[1]

$$I = \int_0^\infty dr \exp(-\Delta r) r^{\alpha-1} {}_1F_1(a_1, b_1, k_1 r)\, {}_1F_1(a_2, b_2, k_2 r). \tag{1}$$

This integral can be performed for values of α unequal to zero or a negative integer to obtain,

$$I = \Gamma(\alpha)\Delta^{-\alpha} F_2(\alpha, a_1, a_2, b_1, b_2; k_1/\Delta, k_2/\Delta), \tag{2}$$

where F_2 is Appell's hypergeometric series[2] which is defined as

$$F_2(\alpha, a_1, a_2, b_1, b_2; x, y) = \sum_{m,n} \frac{(\alpha)_{m+n}(a_1)_m(a_2)_n}{(b_1)_m(b_2)_n m!\,n!} x^m y^n. \tag{3}$$

The Appell F_2 series is absolutely convergent if $|x| + |y| < 1$.

Various analytic continuations for the F_2 series have been given in the literature,[3–5] but the convergence condition on the resulting series remains of the form $|x| + |y| < 1$, according to Horn's criteria.[3] Our investigation[6] of Laplace transforms over products of the asymptotic series form of the Whittaker functions led us to consider the possibility of expressing the F_2 series in terms of Appell's hypergeometric series[2] F_3 defined by,

$$F_3(\alpha, \alpha', \beta, \beta', \gamma; x, y) = \sum_{m,n} \frac{(\alpha)_m(\beta)_m(\alpha')_n(\beta')_n}{(\gamma)_{m+n} m!\,n!} x^m y^n, \tag{4}$$

which is absolutely convergent for $|x| < 1$, $|y| < 1$. We will first derive the relation which expresses an F_2 series in terms of four F_3 series which have nonoverlapping convergence domains, and then give two analytic continuations of the F_3 series which combined with the first result give a new analytic continuation of the F_2 series.

These results will be useful in calculating the bremsstrahlung and the virtual photon spectrum associated with high energy electron scattering from the nucleus.

II. THE F_2 SERIES EXPRESSED IN TERMS OF F_3 SERIES

Appell's F_3 series can be analytically continued by use of the Barnes integral representation to four F_2 series by taking the contour on the left side of the complex plane. This result is quite standard and is explicitly given by[2]

$$\begin{aligned} &F_3(\alpha, \alpha', \beta, \beta', \gamma'; 1/x, 1/y) \\ &\quad= f(\alpha, \alpha', \beta, \beta') x^\alpha y^{\alpha'} F_2(\gamma, \alpha, \alpha', \alpha - \beta + 1, \alpha' - \beta' + 1; x, y) \\ &\qquad + f(\beta, \alpha', \alpha, \beta') x^\beta y^{\alpha'} F_2(\beta - \alpha + \gamma, \beta, \alpha', \beta - \alpha + 1, \\ &\qquad \alpha' - \beta' + 1; x, y) + f(\alpha, \beta', \beta, \alpha') x^\alpha y^{\beta'} \\ &\qquad \times F_2(\beta' - \alpha' + \gamma, \alpha, \beta', \alpha - \beta + 1, \beta' - \alpha' + 1; x, y) \\ &\qquad + f(\beta, \beta', \alpha, \alpha') x^\beta y^{\beta'} F_2(\beta + \beta' - \alpha - \alpha' + \gamma, \\ &\qquad \beta, \beta', \beta - \alpha + 1, \beta' - \alpha' + 1; x, y), \end{aligned} \tag{5}$$

where

$$f(\lambda, \mu, \rho, \sigma) = (-1)^\lambda (-1)^\mu \frac{\Gamma(\gamma')\Gamma(\rho - \lambda)\Gamma(\sigma - \mu)}{\Gamma(\rho)\Gamma(\sigma)\Gamma(\gamma' - \lambda - \mu)},$$

$$\gamma = \alpha + \alpha' + 1 - \gamma',$$

and we have written F_3 in terms of $1/x$ and $1/y$ for convenience. We can apply Eq. (5) to the following four F_3 series which have special relations among their variables and parameters:

$$\begin{aligned} &F_3(\alpha, \alpha', \beta, \beta'; \alpha + \alpha' + 1 - \gamma; 1/x, 1/y), \\ &F_3(\alpha, 1 - \beta', \beta, 1 - \alpha', \alpha - \beta' + 2 - \gamma; (1 - y)/x, (y - 1)/y), \\ &F_3(1 - \beta, \alpha', 1 - \alpha, \beta', \alpha' - \beta + 2 - \gamma; (x - 1)/x, (1 - x)/y), \\ &F_3(1 - \beta, 1 - \beta', 1 - \alpha, 1 - \alpha', 3 - \beta - \beta' - \gamma; \\ &\quad (x + y - 1)/x, (x + y - 1)/y). \end{aligned} \tag{6}$$

By making use of the one-term continuation relations for the F_2 series,[3]

$$\begin{aligned} &F_2(\alpha, a_1, a_2, b_1, b_2; x, y) \\ &\quad = (1 - x)^{-\alpha} F_2(\alpha, b_1 - a_1, a_2, b_1, b_2; x/(x - 1), y/(1 - x)) \\ &\quad = (1 - y)^{-\alpha} F_2(\alpha, a_1, b_2 - a_2, b_1, b_2; x/(1 - y), y/(y - 1)) \\ &\quad = (1 - x - y)^{-\alpha} F_2(\alpha, b_1 - a_1, b_2 - a_2, b_1, b_2; \\ &\qquad x/(x + y - 1), y/(x + y - 1), \end{aligned} \tag{7}$$

the four F_2 series obtained from use of Eq. (5) for each of the F_3 series in Eq. (6) can be written in terms of the four F_2 series explicitly appearing in Eq. (5). That is, we have a 4×4 matrix connecting four F_3 series and four F_2 series. This matrix can be inverted to give the following result:

$$\frac{1}{\Gamma(b_1)\Gamma(b_2)\Gamma(\alpha-a_1-a_2)}F_2(1+a_1+a_2-\alpha, a_1, a_2, b_1, b_2; x, y)$$
$$=AF_3\left(a_1, a_2, a_1+1-b_1, a_2+1-b_2, \alpha; \frac{1}{x}, \frac{1}{y}\right)$$
$$+BF_3\left(a_1, b_2-a_2, a_1+1-b_1, 1-a_2, b_2-2a_2+\alpha; \frac{1-y}{x}, \frac{y-1}{y}\right)+CF_3\left(b_1-a_1, a_2, 1-a_1, a_2+1-b_2, b_1-2a_1+\alpha; \frac{x-1}{x}, \frac{1-x}{y}\right)+DF_3\left(b_1-a_1, b_2-a_2, 1-a_1, 1-a_2, b_1+b_2-2a_1-2a_2+\alpha; \frac{x+y-1}{x}, \frac{x+y-1}{y}\right), \tag{8}$$

where

$$A=x^{-a_1}y^{-a_2}\frac{1}{\Gamma(b_1-a_1)\Gamma(b_2-a_2)\Gamma(\alpha)},$$
$$B=x^{-a_1}(-y)^{a_2-b_2}\frac{(1-y)^{\alpha-2a_2+b_2-1}}{\Gamma(b_1-a_1)\Gamma(a_2)\Gamma(b_2-2a_2+\alpha)},$$
$$C=(-x)^{a_1-b_1}y^{-a_2}\frac{(1-x)^{\alpha-2a_1+b_1-1}}{\Gamma(a_1)\Gamma(b_2-a_2)\Gamma(b_1-2a_1+\alpha)},$$
$$D=(-x)^{a_1-b_1}(-y)^{a_2-b_2}\frac{(1-x-y)^{\alpha+b_1+b_2-2a_1-2a_2}}{\Gamma(a_1)\Gamma(a_2)\Gamma(b_1+b_2-2a_1-2a_2+\alpha)}$$

III. ANALYTIC CONTINUATION OF THE F_3 SERIES

In this section we will obtain analytic continuations of the F_3 series which are useful when one of the variables is near unity ($x\approx 1$), and the other variable is either smaller or greater than one.

The Barnes integral representation of the F_3 series is[2]

$$F_3(\alpha, \beta, \alpha', \beta', \gamma; x, y)=\frac{\Gamma(\gamma)}{\Gamma(\alpha')\Gamma(\beta')}\frac{1}{2\pi i}\int_{k-i\infty}^{k+i\infty}dt\frac{\Gamma(\alpha'+t)\Gamma(\beta'+t)}{\Gamma(\gamma+t)}\times\Gamma(-t)(-y)^t\ {}_2F_1(\alpha, \beta, \gamma+t; x), \tag{9}$$

where ${}_2F_1(\alpha, \beta, \gamma+t; x)$ is Gauss's hypergeometric function, and the contour in the t plane parallels the imaginary axis, except that where necessary, it is indented so that the poles of $\Gamma(\alpha'+t)$ and $\Gamma(\beta'+t)$ lie to the left of the contour, and the poles of $\Gamma(-t)$ lie to the right of the contour. The real parameter k is chosen such that, $k=\mathrm{Re}(\alpha+\beta-\gamma)+\epsilon$ where ϵ is a small positive number.

Gauss's hypergeometric function has a number of analytic continuations. We make use of the following one which is valid for $|\arg x|<\pi$ given in Ref. 3, p. 109:

$$F(a, b, c; x)=\frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}\times x^{-a}F(a, a+1-c, a+b+1-c; 1-x^{-1})+\frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}x^{a-c}(1-x)^{c-a-b}\times F(c-a, 1-a, c+1-a-b; 1-x^{-1}). \tag{10}$$

Substituting this continuation into Eq. (9), we obtain two integrals in the t plane. We will write these as $F_3(\alpha, \beta, \alpha', \beta', \gamma; x, y)=I_1+I_2$, where I_1 and I_2 are given explicitly by the following:

$$I_1=\frac{\Gamma(\gamma)}{\Gamma(\alpha')\Gamma(\beta')}\frac{1}{2\pi i}\int_{k-i\infty}^{k+i\infty}(-y)^t\frac{\Gamma(\alpha'+t)\Gamma(\beta'+t)\Gamma(-t)}{\Gamma(\gamma+t-\alpha)\Gamma(\gamma+t-\beta)}\times\Gamma(\gamma+t-\alpha-\beta)x^{-\alpha}\ {}_2F_1(\alpha, \alpha+1-\gamma-t, \alpha+\beta+1-\gamma-t; 1-x^{-1}). \tag{11}$$

$$I_2=\frac{\Gamma(\gamma)}{\Gamma(\alpha')\Gamma(\beta')}\frac{1}{2\pi i}\int_{k-i\infty}^{k+i\infty}dt(-y)^t\frac{\Gamma(\alpha'+t)\Gamma(\beta'+t)\Gamma(-t)}{\Gamma(\alpha)\Gamma(\beta)}\times\Gamma(\alpha+\beta-\gamma-t)x^{\alpha-\gamma-t}(1-x)^{\gamma+t-\alpha-\beta}\times{}_2F_1(\gamma+t-\alpha, 1-\alpha, \gamma+1+t-\alpha-\beta; 1-x^{-1}). \tag{12}$$

Two different analytic continuations of the F_3 function can be obtained by either closing the contour in the t plane on the right in both terms, or by closing the contour on the left for I_1 and on the right for I_2. By using the asymptotic behavior of the gamma function, we find the condition for absolute convergence of I_1 and I_2 when the contour is closed on the right in the t plane to be $\mathrm{Re}(\alpha'+\beta'-\gamma)<1$, $|y|<1$, and $\mathrm{Re}(\alpha'+\beta'+\alpha+\beta-\gamma)<2$, $|y(1-x^{-1})|<1$, respectively. The condition for absolute convergence for I_1 when the contour in the t plane is closed on the left is $\mathrm{Re}(\alpha'+\beta'-\gamma)<1$, $|1/y|<1$.

The integrand of I_1 has ascending sequences of poles at $t=n$ and $t=\alpha+\beta-\gamma+1+n$ lying in the right-hand contour, and decreasing sequences of poles at $t=-\alpha'-n$, $t=-\beta'-n$, and $t=\alpha+\beta-\gamma-n$ lying in the left-hand contour where $n=0, 1, 2, \cdots$. The integrand of I_2 has ascending sequences of poles lying in the right-hand contour at $t=n$ and $t=\alpha+\beta-\gamma+1+n$ for $n=0, 1, 2, \cdots$. Note that the particular separation between the left and right contours depends on the choice of k given, following Eq. (9), but since the original integrand contains no singularities at $\alpha+\beta-\gamma+n$, the final result is independent of the particular choice of k. When closing both contours on the right, the sequences beginning at $\alpha+\beta-\gamma+1$ cancel and we can write $F_3(\alpha, \beta, \alpha', \beta', \gamma; x, y)=Q_1+Q_2$, where Q_1 and Q_2, obtained by explicit integration in the t plane and using the residue theorem, are given by

$$Q_1=x^{-\alpha}\frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)}\times\sum_{m,n}\frac{(\alpha)_m(\beta')_n(\alpha')_n(\alpha+1-\gamma)_{m-n}}{(\gamma-\beta)_n(\alpha+\beta+1-\gamma)_m m!\,n!}(1-x^{-1})^m y^n,$$

$$Q_2=\frac{x^{\alpha-\gamma}(1-x)^{\gamma-\alpha-\beta}\Gamma(\gamma)\Gamma(\alpha+\beta-\gamma)}{\Gamma(\alpha)\Gamma(\beta)}\times\sum_{m,n}\frac{(\alpha')_n(\beta')_n(1-\alpha)_m(\gamma-\alpha)_{m+n}}{(\gamma-\alpha)_n(1+\gamma-\alpha-\beta)_{m+n}m!\,n!}\times(1-x^{-1})^m(y(1-x^{-1}))^n. \tag{13}$$

The convergence condition for these series according to Horn's criteria[3] are $|y|<1$ and $|1-x^{-1}|<1$ for Q_1, and $|y(1-x^{-1})|<1$ and $|1-x^{-1}|<1$ for Q_2.

For $|y|>1$, we need to close the contour in the t plane for I_1 on the left. Doing so we can write $F_3(\alpha, \beta, \alpha', \beta', \gamma; x, y)=C1+C2+C3+C4+C5$ where these series are given explicitly by

$$C_1=\frac{\Gamma(\gamma)\Gamma(\alpha+\beta-\gamma)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-\gamma}(1-x)^{\gamma-\alpha-\beta}$$

$$\times\sum_{m,n}\frac{(\alpha')_n(\beta')_n(1-\alpha)_m(\gamma-\alpha)_{m+n}}{(\gamma-\alpha)_n(1+\gamma-\alpha-\beta)_{m+n}m!\,n!}(1-x^{-1})^m(y(1-x^{-1}))^n,$$

$$C_2=\frac{\Gamma(\gamma)\Gamma(\alpha'+\alpha+\beta-\gamma+1)\Gamma(\beta'+\alpha+\beta-\gamma+1)}{\Gamma(\alpha')\Gamma(\beta')\Gamma(\alpha)\Gamma(\beta)}$$

$$\times\Gamma(\gamma-\alpha-\beta-1)x^{-\beta}y(1-x^{-1})(-y)^{\alpha+\beta-\gamma}$$

$$\times\sum_{m,n}\frac{(\alpha'+\alpha+\beta-\gamma+1)_n(\beta'+\alpha+\beta-\gamma+1)_n}{(\alpha+\beta-\gamma+2)_n(\beta+1)_n(2)_{m+n}m!}$$

$$\times(1-\alpha)_m(\beta+1)_{m+n}(1-x^{-1})^m(y(1-x^{-1}))^n,$$

$$C_3=\frac{\Gamma(\gamma)\Gamma(\beta'-\alpha')\Gamma(\gamma-\alpha-\beta-\alpha')}{\Gamma(\beta')\Gamma(\gamma-\alpha-\alpha')\Gamma(\gamma-\beta-\alpha')}x^{-\alpha}(-y)^{-\alpha'} \tag{14}$$

$$\times\sum_{m,n}\frac{(\alpha')_n(\alpha)_m(1+\alpha'+\beta-\gamma)_n(\alpha+\alpha'+1-\gamma)_{m+n}}{(1+\alpha'-\beta')_n(\alpha+\beta+1+\alpha'-\gamma)_{m+n}m!\,n!}$$

$$\times(1-x^{-1})^m y^{-n},$$

$$C_4=\frac{\Gamma(\gamma)\Gamma(\alpha'-\beta')\Gamma(\gamma-\alpha-\beta-\beta')}{\Gamma(\alpha')\Gamma(\gamma-\alpha-\beta)\Gamma(\gamma-\beta-\beta')}x^{-\alpha}(-y)^{-\beta'}$$

$$\times\sum_{m,n}\frac{(\beta')_n(\alpha)_m(1+\alpha+\beta-\gamma)_n(1+\beta+\beta'-\gamma)_n}{(1+\beta'-\alpha')_n(1+\alpha+\beta'-\gamma)_n(1+\alpha+\beta+\beta'-\gamma)_{m+n}}$$

$$\times(\alpha+1+\beta'-\gamma)_{m+n}\frac{(1-x^{-1})^m}{m!}\frac{y^{-n}}{n!},$$

$$C_5=\frac{\Gamma(\gamma)\Gamma(\alpha+\beta+\beta'-\gamma)\Gamma(\alpha+\alpha'+\beta-\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

$$\times x^{-\alpha}(-y)^{\alpha+\beta-\gamma}$$

$$\times\sum_{m,n}\frac{(\alpha)_m(1-\alpha)_n(\gamma-\alpha-\beta)_n(1-\beta)_{m+n}}{(1+\gamma-\alpha-\beta-\beta')_n(1+\gamma-\alpha-\alpha'-\beta)_n(1)_{m+n}m!}$$

$$\times(1-x^{-1})^m y^{-n}.$$

These series are absolutely convergent if $|y(1-x^{-1})|<1$ and $|1-x^{-1}|<1$ for $C1$ and $C2$ and $|1/y|<1$ and $|1-x^{-1}|<1$ for $C3$, $C4$, $C5$.

IV. CONCLUSION

The results given in Eq. (8) and Eq. (13) or Eq. (14) provide a new analytic continuation for the F_2 function. To demonstrate this more explicitly, consider the case of electron scattering in the presence of a point nucleus of charge Z. The incident and final electron energies (momenta) are E_1 (P_1) and E_2 (P_2), where $P=(E^2-M^2)^{1/2}$ and the energy lost by the electron is $W=E_1-E_2$. The radial integrals describing this process in distorted-wave Born approximation can be expressed in terms of Appell's F_2 function[6] with variables $x=-2P_2/(P_1-P_2-W)$ and $y=2P_1/(P_1-P_2-W)$. For physical values of the kinematic variables, x and y are both very large. Use of Eq. (8) to transform the F_2 function to four F_3 functions results in the following sets of variables: I. $-(P_1-P_2-W)/2P_2$, $(P_1-P_2-W)/2P_1$; II. $(P_1+P_2+W)/2P_2$, $(P_1+P_2+W)/2P_1$; III. $(P_1+P_2-W)/2P_2$, $(P_1+P_2-W)/2P_1$; IV. $-(P_1-P_2+W)/2P_2$, $(P_1-P_2+W)/2P_1$.

Variable set I is very small for all W except at the end point $P_2\approx 0$, and hence the F_3 function with set I variables is absolutely convergent except very near the end point which we do not consider. One of the variables in sets II and III is always very near unity ($y_{\rm II}\lesssim 1$, $x_{\rm III}\gtrsim 1$) for relativistic electrons, the other variable being greater than one in set II; and less than one in set III. An F_3 function with set III is semiconvergent as is, and the use of the analytic continuation given in Eq. (13) will result in absolute convergence. The F_3 with the set II variables has to be continued by means of W, the energy lost by the electron. For $W\lesssim 50\%\ E_1$, both variables are less than one in magnitude and the F_3 series is absolutely convergent. For $W\gtrsim 50\%\ E_1$, $y_{\rm IV}\approx W/P_1<1$, but $|x_{\rm IV}|\approx W/P_2>1$ and the F_3 function needs to be analytically continued. The use of the continuation given in Eq. (14) will result in absolutely convergent series.

We also note that the analytic continuations given in this paper will allow the rapid evaluation of the Appell F_2 function when one of the variables, say x, is near unity. Depending on whether y is less than or greater than unity, the continuations of Eq. (13) or Eq. (14) can be used. The resulting doubly infinite series will be very rapidly convergent for one of the indices, and absolutely convergent for the other. This can be used when both variables are in the neighborhood of the singular point (1, 1), but clearly the summation over one of the indices will be rather slowly converging.

To summarize, we have found a new analytic continuation of the Appell F_2 function in terms of double series which are absolutely convergent if their variables satisfy $|x|<1$ and $|y|<1$. This continuation has a practical use in the analysis of electron scattering from the nucleus, and also permits one to evaluate the Appell function near the singular point $x=1$, $y=1$.

ACKNOWLEDGMENT

We wish to acknowledge useful discussions on this problem with Professor D.S. Onley.

[1]A. Erdélyi, Math. Z. **42**, 125 (1936).

[2]P. Appell and J. Kampé de Fériet, *Fonction Hypergémétriques et Hypérspheriques* (Gauthier-Villars, Paris, 1926).

[3]A. Erdélyi, W. Magnus, F. Oberhettinger, and F.G. Tricemi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1.

[4]W.W. Gargaro and D.S. Onley, Phys. Rev. C **4**, 1032 (1971).

[5]G.E. Hahne, J. Math. Phys. **10**, 524 (1969).

[6]Krishan Sud, L.E. Wright, and D.S. Onley, "Radial Integrals with Finite Energy Loss for Dirac—Coulomb Functions," submitted to J. Math. Phys.

On the solution of nonlinear matrix integral equations in transport theory

R. L. Bowden and P. F. Zweifel*

Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

R. Menikoff†

Theoretical Division, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87544
(Received 21 May 1976)

The coupled nonlinear matrix integral equations for the matrices $X(z)$ and $Y(z)$ which factor the dispersion matrix $\Lambda(z)$ of multigroup transport theory are studied in a Banach space X. By utilizing fixed-point theorems we are able to show that iterative solutions converge uniquely to the "physical solution" in a certain sphere of X. Both isotropic and anisotropic scattering are considered.

I. INTRODUCTION

In a recent paper,[1] the Chandrasekhar H equation has been studied. In particular the following results were shown:

1. An iterative procedure, proved by Bittoni *et al*[2] to converge to a unique solution inside a certain region of the Banach space $L_1(0, 1)$, actually converges to the "physical solution," i.e., the solution which is analytic in the right-half complex plane. (Alternatively, the "physical solution" is the one which obeys the so-called constraining equations.[3,4])

2. The iteration scheme of Bittoni *et al* can be extended to all values of $\|\psi\|$, provided $\psi(\mu) \geq 0$, $\mu \in [0, 1]$, where $\psi(\mu)$ is the "characteristic function ($\|\psi\| = c/2$ in one-speed isotropic neutron transport)." In Ref. 2, only the case $\|\psi\| < 1$ had been studied.

The advantage of these results is that in any "one-group" transport problem, the H functions can be calculated iteratively without the necessity of introducing constraining equations. Furthermore, the knowledge of the region of Banach space in which the solution exists is of considerable help in performing the numerics. In particular, we observe that if the initial estimate is chosen to be zero, the iterative procedure always converges to the "physical solution."

The purpose of this paper is to present a similar iteration scheme for solving the matrix versions of the Chandrasekhar H equations. The solution of these equations provides the Wiener—Hopf *matrix* factorization of the dispersion matrix Λ and is needed to construct the solution of half-space multigroup transport equations.[5,6] [In the one-speed or scalar case the H function is the Wiener—Hopf factorization of the dispersion function $\Lambda(z)$.]

For the multigroup problem it is necessary to consider coupled nonlinear nonsingular matrix equations which have been written in the form[6]

$$X(-z) = C^{-1}\Sigma - z\int_0^1 Y^{-1}(-s)\Sigma^2\Delta(s)\frac{ds}{s+z} \tag{1a}$$

and

$$Y(-z) = \Sigma - z\int_0^1 \Sigma^2\Delta(s)X^{-1}(-s)\frac{ds}{s+z}. \tag{1b}$$

Here Σ is the diagonal cross section matrix with elements $\delta_{ij}\sigma_i$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N = 1$, and C is the group-to-group scattering matrix, while Δ is a diagonal matrix with elements

$$\Delta_{ij}(s) = \delta_{ij}\theta(s - 1/\sigma_i),$$

where θ is the Heavyside function

$$\theta(s - 1/\sigma_i) = 1, \quad s \leq 1/\sigma_i$$
$$= 0, \quad s > 1/\sigma_i.$$

Moreover, X and Y factor the Λ matrix,[6]

$$\Lambda(z) = (\Sigma - 2C)C^{-1}\Sigma - \int_{-1}^{1}\mu[zI - \mu\Sigma^{-1}]^{-1}\,d\mu,$$

in the form

$$\Lambda(z) = Y(-z)X(z), \tag{2}$$

where $Y(z)$ and $X(z)$ are supposed to be analytic and nonsingular for $\mathrm{Re}z < 0$. Because $Y(z)$ and $X(z)$ factor the dispersion matrix $\Lambda(z)$, the requirement that $Y(z)$ and $X(z)$ be analytic and nonsingular for $\mathrm{Re}z < 0$ is equivalent to the constraints[6]

$$\begin{aligned}&\det Y(+\nu_j) = \det X(+\nu_j) = 0,\\ &\mathrm{Re}\nu_j > 0, \quad j = 0, \ldots, d-1,\end{aligned} \tag{3}$$

where $\pm\nu_j$, $j = 0, \ldots, d-1$ are the $2d$ discrete Van Kampen—Case eigenvalues which obey

$$\Omega(\pm\nu_j) \equiv \det\Lambda(\pm\nu_j) = 0.$$

The constraints in Eq. (3) are usually introduced to assure that the solution of Eqs. (1) (or comparable equations) is unique.[7] The solution of Eqs. (1) which obeys the constraints in (3) will be called the "physical solution." In the current analysis, uniqueness is guaranteed by restricting the solution to a certain sphere in Banach space. The resulting solution can then be shown to be the "physical solution."

The factorization of $\Lambda(z)$ (with a somewhat different notation) was originally obtained by Mullikin[8] and, as used in Ref. 6, was restricted to the case $\rho < \frac{1}{2}$, where ρ is the dominant eigenvalue of the nonnegative matrix

$\Sigma^{-1}C$. The results of this paper are restricted to the more restrictive case $\int_0^1 \|\Delta C\|_M(s)\,ds<\frac{1}{2}$ for the case of isotropic scattering presented in Sec. II and a similar restriction for the case of anisotropic scattering presented in Sec. III. Here $\|\ \|_M$ represents the "matrix norm," e.g.,

$$\|A\|_M=\sup_i \sum_j |A_{ij}|. \tag{4}$$

Before presenting our analysis in the next section, we might remark that if C is a symmetric matrix, $C = C^t$, then $\Lambda=\Lambda^t$ and it can be shown quite easily that

$$Y=X^tC.$$

Then the two coupled equations (1) reduce to a simple equation, which after appropriate transformation becomes the "matrix H equation" considered by Siewert and co-workers.[8,9] Thus the equation they studied is a special case of ours.

II. BANACH SPACE ANALYSIS

Equations (1) can be transformed into a more convenient form by defining

$$U_1(z)=C^{-1}\Sigma X^{-1}(-z) \tag{5a}$$

and

$$U_2(z)=Y^{-1}(-z)\Sigma. \tag{5b}$$

For $z\in[0,1]$, Eqs. (1) reduce then to the coupled nonlinear, nonsingular matrix integral equations

$$U_1(z)=I+z\int_0^1 U_1(z)U_2(s)\Sigma\Delta(s)\Sigma^{-1}C\frac{ds}{z+s} \tag{6a}$$

and

$$U_2(z)=I+z\int_0^1 \Sigma\Delta(s)\Sigma^{-1}CU_1(s)U_2(z)\frac{ds}{z+s}. \tag{6b}$$

We consider U_1 and U_2 as elements of a Banach space X_0 with norm[10]

$$\|U_i\|_{X_0}=\int_0^1 \|U_i\|_M(s)\,ds, \tag{7}$$

where $\|\ \|_M$ is the matrix norm already introduced.[11] Now consider the Banach space X, the Cartesian product of X_0 with itself,

$$\mathcal{U}=[U_1,U_2]\in X,\quad U_1,U_2\in X_0, \tag{8a}$$

with norm

$$\|\mathcal{U}\|_X=\int_0^1 \max[\|U_1\|_M,\|U_2\|_M](s)\,ds. \tag{8b}$$

One can readily verify that $\|\ \|_X$ is a norm.

Let us now define $U_1'\in X_0$ and $U_2'\in X_0$ by

$$U_1'(s)=\Sigma\Delta(s)\Sigma^{-1}CU_1(s),\quad s\in[0,1] \tag{9a}$$

and

$$U_2'(s)=U_2(s)\Sigma\Delta(s)\Sigma^{-1}C,\quad s\in[0,1]. \tag{9b}$$

We can then write from Eqs. (6) the single equation for

$$\mathcal{U}'=[U_1',U_2']\in X,\quad \mathcal{U}'=\mathcal{J}+A(\mathcal{U}',\mathcal{U}'), \tag{10a}$$

where

$$\mathcal{J}=[\Sigma\Delta\Sigma^{-1}C,\Sigma\Delta\Sigma^{-1}C]\in X \tag{10b}$$

and A is the bilinear form

$$A(\mathcal{U},\mathcal{V})(z)=\left[z\int_0^1 V_1(z)U_2(s)\frac{ds}{s+z},\ z\int_0^1 U_1(s)V_2(z)\frac{ds}{s+z}\right]. \tag{10c}$$

The following lemma which is proved in Ref. 2 (and restated in Ref. 1) is vital to the subsequent analysis:

Lemma I: Let Y be a Banach space with norm $\|\ \|_Y$ and $B(u,v)$ a bilinear map: $Y\times Y\to Y$ with "norm"

$$\|B\|=\sup\{\|B(u,v)\|_Y:\|u\|_Y=1 \text{ and } \|v\|_Y=1\}.$$

Then for $2\|B+B^*\|_Y\|f\|_Y<1$, the equation

$$u=Tu\equiv f+B(u,u),\quad f\in Y$$

has one and only one solution in the ball

$$S=\{u\in Y:\|u-f\|_Y<\tfrac{1}{2}\}.$$

Furthermore, $TS\subset S$. [Here $B^*(u,v)\equiv B(v,u)$.]

Corollary: For every $u_0\in S$, $\lim_{n\to\infty}T^nu_0$ converges in Y to the unique solution of the equation $u=Tu$.

We now prove

Lemma II: If A is the bilinear form given by Eq. (10c) and $A^*(\mathcal{U},\mathcal{V})\equiv A(\mathcal{V},\mathcal{U})$, then $\|A+A^*\|=1$.

Proof: By direct calculation we find

$$\begin{aligned}
&\|A(\mathcal{U},\mathcal{V})+A^*(\mathcal{U},\mathcal{V})\|_X\\
&=\int_0^1 dx\max\left[\left\|\int_0^1 ds\,V_1(x)U_2(s)\frac{x}{s+x}+\int_0^1 ds\,U_1(x)V_2(s)\frac{x}{s+x}\right\|_M,\ \left\|\int_0^1 ds\,U_1(s)V_2(x)\frac{x}{s+x}+\int_0^1 ds\,V_1(s)U_2(x)\frac{x}{s+x}\right\|_M\right]\\
&\le\int_0^1 dx\max\left[\int_0^1 dx\{\|V_1\|_M(x)\|U_2\|_M(s)+\|U_1\|_M(x)\|V_2\|_M(s)\}\frac{x}{s+x},\ \int_0^1 ds\{\|U_1\|_M(s)\|V_2\|_M(x)+\|V_1\|_M(s)\|U_2\|_M(x)\}\frac{x}{s+x}\right]\\
&\le\int_0^1 dx\int_0^1 ds\Big\{\max[\|V_1\|_M,\|V_2\|_M](x)\times\max[\|U_1\|_M,\|U_2\|_M](s)\frac{x}{s+x}+\max[\|V_1\|_M,\|V_2\|_M](s)\times\max[\|U_1\|_M,\|U_2\|_M](x)\frac{x}{s+x}\Big\}\\
&\le\int_0^1 dx\int_0^1 ds\max[\|V_1\|_M,\|V_2\|_M](x)\times\max[\|U_1\|_M,\|U_2\|_M](s)\\
&\le\|\mathcal{U}\|_X\cdot\|\mathcal{V}\|_X.
\end{aligned}$$

(In going from the third to the fourth relation, the change of variable $x\to s$ has been made.) The above calculations show that $\|A+A^*\|\le 1$. Equality is obtained by setting $\mathcal{U}=\mathcal{V}=[I,I]$. This completes the proof of the lemma.

Noting that

$$\|\mathcal{J}\|_X = \int_0^1 \|\Delta C\|_M(s)\,ds,$$

we combine Lemmas I and II to obtain

Lemma III: If $\int_0^1 \|\Delta C\|_M(s)\,ds < \frac{1}{2}$, Eq. (10) has a unique solution $\hat{\mathcal{U}}$ in the ball

$$S_1 = \{\mathcal{U}' \in X : \|\mathcal{U}' - \mathcal{J}\|_X < \tfrac{1}{2}\}.$$

Furthermore, the iteration procedure defined by

$$\mathcal{U}_n = \mathcal{J} + A(\mathcal{U}_{n-1}, \mathcal{U}_{n-1})$$

converges to $\hat{\mathcal{U}}$ for every $\mathcal{U}_0 \in S_1$.

The convergence can easily be seen to be uniform and pointwise (see Lemma III of Reference 1). We omit the details here.

We now know that we can solve Eq. (10a) iteratively to obtain $\hat{\mathcal{U}}$. To recover $X(z)$ and $Y(z)$ from Eqs. (5) we must first obtain U_1 and U_2 from U_1' and U_2' [Eqs. (9)]. Unfortunately, $\Delta(s)$ is not an invertible matrix. Therefore, we describe below the scheme which can be used. At the same time, this scheme provides the analytic continuation of $\mathcal{U}$ to the rest of the complex plane.

In other words, we wish to show that the solution of Eq. (10) referred to in Lemma III can be used to obtain the matrices $U_1(z)$ and $U_2(z)$ satisfying Eqs. (6). Moreover we shall prove that these matrices are analytic for $\mathrm{Re}z \geq 0$. To this end let us now state

Lemma IV: If $\hat{\mathcal{U}} = [\hat{U}_1, \hat{U}_2]$ is the unique solution to Eq. (9) in the ball S_1 for $\int_0^1 \|\Delta C\|_M(s)\,ds < \frac{1}{2}$, then for $z \in \mathbb{C}$

$$\det\left[I - \int_0^1 \hat{U}_i(s)\frac{z}{z+s}\,ds\right] \neq 0, \quad \mathrm{Re}z \geq 0, \quad i = 1, 2. \qquad (11)$$

Proof: Since $\hat{\mathcal{U}} \in S_1$ and $\|\mathcal{J}\|_X < \frac{1}{2}$, we have

$$\tfrac{1}{2} > \|\hat{\mathcal{U}} - \mathcal{J}\|_X > \big|\,\|\hat{\mathcal{U}}\|_X - \|\mathcal{J}\|_X\big|.$$

Thus we have

$$\|\hat{\mathcal{U}}\| < 1.$$

Hence

$$\|U_i\|_{X_0} < 1, \quad i = 1, 2.$$

Now let $z = \alpha + i\beta$, $\alpha \geq 0$. Suppose for some value of z

$$\det\left(I - \int_0^1 \hat{U}_i(s)\frac{z}{z+s}\,ds\right) = 0, \quad \mathrm{Re}z \geq 0, \quad i = 1, 2.$$

This would imply that there exists a nonzero vector Ψ such that

$$\|\Psi\| \leq \left\| \int_0^1 \hat{U}_i(s)\frac{\alpha + i\beta}{\alpha + s + i\beta}\,ds\right\|_M \cdot \|\Psi\|,$$

where here $\|\Psi\|$ is a vector norm consistent with $\|\ \|_M$. This last relation yields

$$\begin{aligned}
1 &\leq \int_0^1 \|\hat{U}_i\|_M(s)\left|\frac{\alpha + i\beta}{\alpha + s + i\beta}\right| ds \\
&\leq \int_0^1 \|\hat{U}_i\|_M(s)\left(\frac{\alpha^2 + \beta^2}{(\alpha + s)^2 + \beta^2}\right)^{1/2} ds \\
&\leq \int_0^1 \|\hat{U}_i\|_M(s)\,ds \quad (\text{for } \alpha \geq 0) \\
&= \|\hat{U}_i\|_{X_0} < 1
\end{aligned}$$

which is a contradiction. Thus the inequality (11) must hold. This completes the proof of the lemma.

Now define

$$U_1(z) = \left(I - \int_0^1 \hat{U}_2(s)\frac{z}{z+s}\,ds\right)^{-1} \qquad (12a)$$

and

$$U_2(z) = \left(I - \int_0^1 \hat{U}_1(s)\frac{z}{z+s}\,ds\right)^{-1}, \qquad (12b)$$

where $\hat{\mathcal{U}} = [\hat{U}_1, \hat{U}_2]$ is the unique solution of Eq. (10) referred to in Lemma IV. The matrices $U_1(z)$ and $U_2(z)$ are analytic in the complex z plane cut along $[-1, 0]$ with (possible) poles at those values of z for which

$$\det\left[I - \int_0^1 \hat{U}_i(s)\frac{z}{z+s}\,ds\right] = 0, \quad i = 1, 2.$$

In particular we observe from Lemma IV that $U_1(z)$ and $U_2(z)$ are analytic in the complex z plane for $\mathrm{Re}z \geq 0$ ($z \neq 0$). Furthermore, we have

Lemma V: The matrices $U_1(z)$ and $U_2(z)$ defined by Eqs. (12) satisfy Eqs. (6).

Proof: For those values of z such that

$$\det\left(I - \int_0^1 \hat{U}_i(s)\frac{z}{z+s}\,ds\right) \neq 0,$$

$U_1(z)$ and $U_2(z)$ satisfy

$$U_1(z) = I + \int_0^1 U_1(z)\hat{U}_2(s)\frac{z}{z+s}\,ds \qquad (13a)$$

and

$$U_2(z) = I + \int_0^1 \hat{U}_1(s)U_2(z)\frac{z}{z+s}\,ds. \qquad (13b)$$

We then need only to prove that

$$\Sigma\Delta(s)\Sigma^{-1}CU_1(s) = \hat{U}_1(s), \quad s \in [0, 1]$$

and

$$U_2(s)\Sigma\Delta(s)\Sigma^{-1}C = \hat{U}_2(s), \quad s \in [0, 1].$$

However, from Lemma IV, we note that Eqs. (12) are well defined for $z \in [0, 1]$ and

$$\begin{aligned}
\Sigma\Delta(z)\Sigma^{-1}CU_1(z) &= \Sigma\Delta(z)\Sigma^{-1}C\left(I - z\int_0^1 \hat{U}_2(s)\frac{ds}{z+s}\right)^{-1} \\
&= \hat{U}_1(z), \quad z \in [0, 1]
\end{aligned}$$

and

$$\begin{aligned}
U_2(z)\Sigma\Delta(z)\Sigma^{-1}C &= \left(I - z\int_0^1 \hat{U}_1(s)\frac{ds}{z+s}\right)^{-1}\Sigma\Delta(z)\Sigma^{-1}C \\
&= \hat{U}_2(z), \quad z \in [0, 1].
\end{aligned}$$

This completes the proof.

The matrices $U_1(z)$ and $U_2(z)$ are analytic in the left-half complex z plane except for a cut along $[-1, 0]$ and (possible) poles at those values of z for which $\det U_1(z)$ and $\det U_2(z)$ vanish. In this regard, we have

Lemma VI: If $U_1(z)$ and $U_2(z)$ are defined by Eqs. (12), then

$$\det U_1^{-1}(-\nu_j) = \det U_2^{-1}(-\nu_j) = 0,$$

$$\operatorname{Re}\nu_j \geq 0, \quad j = 0, \ldots, d-1, \tag{14}$$

where we recall that $\pm\nu_j$, $j = 0, \ldots, d-1$ are the zeros of $\det\Lambda(z)$.

Proof: From Lemma V, $U_1(z)$ and $U_2(z)$ satisfy Eqs. (6), but by considering

$$[U_2^{-1}(z) - I][U_1^{-1}(-z) - I] = z^2 \int_0^1 ds \int_0^1 dt\, \Sigma\Delta(s)\Sigma^{-1}CU_1(s) \times U_2(t)\Sigma\Delta(t)\Sigma^{-1}C[(z+s)(z-t)]^{-1}$$

one can show that $U_2^{-1}(z)$ form the Wiener–Hopf factorization of $\Lambda(z)$ (cf. Ref. 6),

$$U_2^{-1}(z)U_1^{-1}(-z) = \Sigma^{-1}\Lambda(z)\Sigma^{-1}C. \tag{15}$$

Since $\Lambda(z)$ is even in z, we must have

$$\det U_2^{-1}(\nu_j)\det U_1^{-1}(-\nu_j) = 0, \quad j = 0, \ldots, d-1, \tag{16a}$$

and

$$\det U_2^{-1}(-\nu_j)\det U_1^{-1}(+\nu_j) = 0, \quad j = 0, \ldots, d-1. \tag{16b}$$

The lemma now follows from Lemma IV.

We note that from Eqs. (14) and (16) if ν_j is purely imaginary, then

$$\det U_i^{-1}(+\nu_j) = \det U_i^{-1}(-\nu_j)^* = 0,$$

in contradiction to Lemma IV. We thus have

Corollary to proof of Lemma VI: If $\int_0^1 \|\Delta C\|_M(s)\, ds < \frac{1}{2}$, then there are no purely imaginary zeros of $\det\Lambda(z)$.

We summarize the results of this section with

Theorem I: If $\int_0^1 \|\Delta C\|_M(s)\, ds < \frac{1}{2}$, then the matrices $U_1(z)$ and $U_2(z)$ given by Eqs. (12) satisfy Eqs. (6) with $\hat{U} = [U_1(s), \hat{U}_2(s)]$, $s \in [0, 1]$, being the unique solution to Eq. (10) in the ball S_1. Furthermore, $U_1(z)$ and $U_2(z)$ are analytic in the complex z plane cut along $[-1, 0]$ except for poles at $-\nu_j$, $j = 0, \ldots, d-1$ and factor the dispersion matrix $\Lambda(z)$ according to Eq. (15).

III. ANISOTROPIC SCATTERING

The procedure presented in the preceding section can easily be generalized to the case of anisotropic scattering. The transport equation for a degenerate scattering kernel of the form

$$C(\mu, \mu') = \sum_{i=1}^{M} A_i(\mu)B_i(\mu')$$

has been studied by Larsen and Zweifel.[12] The nonlinear integral equations were written in this reference as

$$X(-z) = I - z\int_0^1 Y^{-1}(-s)\sum_{j=1}^{N} B(s\sigma_j)I_jA(s\sigma_j)\frac{ds}{z+s} \tag{17a}$$

and

$$Y(-z) = I - z\int_0^1 \sum_{j=1}^{N} B(\sigma s_j)I_jA(s\sigma_j)X^{-1}(-s)\frac{ds}{z+s}. \tag{17b}$$

Here, X and Y are $NM\times NM$ matrices (N is the number of groups and M is the order of anisotropy), A is an $N\times NM$ matrix defined by

$$A = (A_1A_2\cdots A_M),$$

and B is the $NM\times N$ matrix defined by

$$B^t = (B_1^tB_2^t\cdots B_M^t).$$

Also, I_j is an $N\times N$ matrix for which the element in the jth row and jth column is unity and all other elements are zero. (We are discussing only the solution of the X and Y equations in this paper; the reader curious as to the reason for the introduction of such a cumbersome structure should consult Ref. 12.) For technical reasons, it is convenient in the anisotropic scattering case to define the Λ matrix slightly differently from that used in isotropic scattering. Specifically the matrix is defined by

$$\Lambda(z) = I - z\int_{-1}^{1} B(s)\Sigma^{-1}(zI - \Sigma^{-1}s)^{-1}A(s)\, ds.$$

Then the X and Y matrices which satsify Eqs. (17) factor $\Lambda(z)$ according to Eq. (2).

The procedure followed in Sec. II can equally well be applied to Eqs. (17). In particular, if we define

$$X^{-1}(-z) = V_1(z) \quad\text{and}\quad Y^{-1}(-z) = V_2(z),$$

Eqs. (17) can be written as

$$V_1(z) = I + z\int_0^1 V_1(z)V_2(s)R(s)\frac{ds}{z+s} \tag{18a}$$

and

$$V_2(z) = I + z\int_0^1 R(s)V_1(s)V_2(z)\frac{ds}{z+s}, \tag{18b}$$

where we have defined

$$R(s) = \sum_{j=0}^{N} B(s\sigma_j)I_jA(s\sigma_j).$$

If we now make the transformation

$$V_1'(s) = R(s)V_1(s), \quad s \in [0, 1]$$

and

$$V_2'(s) = V_2(s)R(s), \quad s \in [0, 1]$$

we can write the single equation for $\mathcal{V}' = [V_1', V_2'] \in X$,

$$\mathcal{V}' = \mathcal{J}' + A(\mathcal{V}', \mathcal{V}'), \tag{19a}$$

where

$$\mathcal{J}' = [R, R] \in X, \tag{19b}$$

and A is the bilinear form given by Eq. (10c). By the analysis of Sec. II we see that if $\|\mathcal{J}'\|_X < \frac{1}{2}$, i.e., if

$$\int_0^1 \|R\|_M(s)\, ds < \tfrac{1}{2},$$

then Eq. (19) has a unique solution $\hat{\mathcal{V}}$ in the ball S_2 given by

$$S_2 = \{\mathcal{V}' \in X;\ \|\mathcal{V}' - \mathcal{J}'\|_X < \tfrac{1}{2}\}.$$

We now define

$$V_1(z) = \left(I - \int_0^1 \hat{V}_2(s)\frac{z}{z+s}\, ds\right)^{-1} \tag{20a}$$

and

$$V_2(z) = \left(I - \int_0^1 \hat{V}_1(s) \frac{z}{z+s}\, ds\right)^{-1} \tag{20b}$$

and state from the results of Sec. II

Theorem II: If $\int_0^1 \|R\|_M(s)\, ds < \frac{1}{2}$, then the matrices $V_1(z)$ and $V_2(z)$ given by Eqs. (20) satisfy Eqs. (18) with $\hat{V} = [V_1(s), V_2(s)]$, $s \in [0, 1]$ being the unique solution of Eq. (19) in the ball S_2. Furthermore, $V_1(z)$ and $V_2(z)$ are analytic in the complex z plane cut along $[-1, 0]$, except for poles at $-\nu_j$, $j = 0, \ldots, d-1$, and factor the dispersion matrix $\Lambda(z)$ according to

$$V_2(z) V_1(-z) \Lambda(z) = I. \tag{21}$$

IV. DISCUSSION

In Sec. II, the transformation from the set $[U_1, U_2]$ to $[U_1', U_2']$ is made. This is a technical convenience, and one could just as well work with Eqs. (6) for $[U_1, U_2]$. However, in Sec. III, where anisotropic scattering is considered, we have not discovered a convenient way to work with the unprimed quantities. The transformation almost seems unavoidable in that case.

We note, further, that in the solution to either Eqs. (6) for U or Eqs. (10) for U' one need only obtain the solution for the ith row of $U_1(s)$ and the ith column of $U_2(s)$ for $0 \le s\sigma_i \le 1$. If the solution is desired for the entire range of s, $0 \le s \le 1$, or for that matter in the remainder of the complex plane, one only needs to carry out the analytic continuation according to Eqs. (12). However, for the solution of the transport equation,[6] one needs only the values of U_1 and U_2 for the restricted range of $[0, 1]$ described above and at the discrete eigenvalues $-\nu_j$, $j = 0, \ldots, d-1$.

Finally, we address ourselves to the question of generalizing our results. If ρ is the dominant eigenvalue of the nonnegative matrix $\Sigma^{-1}C$, the inequality $\rho < \frac{1}{2}$ is the condition that the infinite medium be subcritical.[13] However, we note that

$$\rho \le \left\| \int_0^1 \Sigma\Delta(s)\Sigma^{-1} C\, ds \right\|_M \le \int_0^1 \|\Delta C\|_M(s)\, ds.$$

If we wish to discuss the general case of infinite medium subcriticality for the isotropic scattering case, then the norm condition in Theorem I is too strong, since there may be some systems which obey the infinite medium subcriticality condition but not the norm inequality in Theorem I. A similar argument also applies to Theorem II in the case of anisotropic scattering. Although it might be possible, by appropriately defining norms, to extend the results of Sections II and III to all subcritical parameters, a more fruitful procedure seems to be indicated. That is to try to find a transformation similar to that introduced in Ref. 1 to extend our results to all systems, supercritical, critical, and subcritical. That is the problem that we are currently pursuing.

*Supported in part by the National Science Foundation Grant Number Eng. 75-15882.

†Work performed under the auspices of and supported by the U.S. Energy Research and Development Administration.

[1]R. L. Bowden and P. F. Zweifel, "A Banach Space Analysis of the Chandrasekhar H Equation," submitted to Astrophys. J.

[2]E. Bittoni, G. Casadei, and S. Lorenzutta, Boll. Unione Mathematica **4**, 535 (1969).

[3]I. W. Busbridge, *The Mathematics of Radiative Transfer* (Cambridge U. P., Cambridge, 1960).

[4]T. W. Mullikin, Trans. Am. Math. Soc. **113**, 316 (1964).

[5]R. L. Bowden, S. Sancaktar, and P. F. Zweifel, J. Math. Phys. **17**, 76 (1976).

[6]R. L. Bowden, S. Sancaktar, and P. F. Zweifel, J. Math. Phys. **17**, 82 (1976).

[7]E. E. Burniston, T. W. Mullikin, and C. E. Siewert, J. Math. Phys. **13**, 1461 (1972).

[8]C. E. Siewert, E. E. Burniston and J. T. Kriese, J. Nucl. Energy **26**, 469 (1972).

[9]C. E. Siewert and Y. Ishiguro, J. Nucl. Energy **26**, 251 (1972).

[10]J. A. Ball, Transp. Theory Stat. Phys. **4**, 67 (1975).

[11]Equation (7) says that $\|U_i\|_M$ must be computed at each point $s \in [0,1]$ and then integrated. At first glance, this seems odd, since the elements of U_i are actually equivalence classes of functions on $[0,1]$ modulo sets of Lebesque measure zero. However, because the result is to be integrated over $[0,1]$ it is easily seen that the same result for $\|U_i\|_{X_0}$ is obtained for all elements of the same equivalence class. The same remark applies to Eq. (8b). We are indebted to Profs. J. Ball and W. Green for a discussion of this point.

[12]E. W. Larsen and P. F. Zweifel, "Steady, One-Dimensional Multigroup Neutron Transport with Anisotropic Scattering," submitted to J. Math. Phys.

[13]R. L. Bowden, Transp. Theory Stat. Phys. **4**, 25 (1975).

A continuum theory of deformable ferrimagnetic bodies. I. Field equations

Gérard A. Maugin

Université de Paris VI, Laboratoire de Mécanique Théorique associé au C.N.R.S., Tour 66, 4 Place Jussieu, 75230 Paris, Cedex 05, France
(Received 15 July 1975)

The first part of this work concerns a thorough study of both global and local field equations that govern deformable (not necessarily linear elastic) ferrimagnets and antiferromagnets from a phenomenological viewpoint. The main tool used is a generalized version of d'Alembert's principle, valid for both reversible and irreversible phenomena, simultaneously with the invariance requirement provided by the so-called objectivity and applied *a priori* to generalized internal forces which represent the various interactions. All interactions taking place in such media are thus given a phenomenological description and are introduced via the duality inherent in the method. The development follows a rational and deductive mathematical scheme in which the notion of topological linear space of velocities plays a predominant role, so that particular cases follow by selecting the appropriate member of this space. In the following companion paper the allied thermodynamics and a thorough discussion of the relevant constitutive equations that follow therefrom are given. The formulation so obtained will allow the consideration of slight perturbations superimposed on bias fields.

1. INTRODUCTION

The aim of this paper is to present an attempt (thought successful) at a "rational" phenomenological approach of of the theory of ferrimagnetic and antiferromagnetic continua. Here "rational" must be understood in the sense that the development is made according to a deductive scheme, starting from well-accepted facts of microphysics, using first principles of mechanics, electromagnetism, and energetics, and developing from the general cases to particular ones. That is, linear elastic ferrimagnetic and antiferromagnetic solids are obtained as special cases of the *exact nonlinear* theory of elastic ferrimagnets, which themselves correspond to specific constitutive assumptions made to close the differential system of balance equations.[1] The interest for such a deductive scheme is manifold. First, it allows us to give a precise statement of all the simplifying assumptions made in obtaining the equations at all degrees of approximation. Next, as recently emphasized by Baumhauer and Tiersten,[2] the initial study of the exact nonlinear case allows one to consider from the start the correct rotationally invariant combinations of deformation measures and physical fields, or of their rates, which show up even in the infinitesimal strain theory (see, for instance, the spin—lattice relaxation thus obtained in Part II), and are in fact necessary if one desires to study superimposition of slight (adiabatic or not) perturbations on bias fields.

With regard to the general method used to obtain the field equations, the following remark is in order. As already remarked in a previous paper of ours,[3] one mainly uses either one of the following methods in order to construct phenomenological theories of continuous media in interaction with physical fields such as electromagnetic fields, the magnetic spin field, and the electric polarization field: (i) to consider a nondissipative continuum and derive both field and constitutive equations from a variational principle of the Lagrangian or Hamiltonian type, once the functional dependence of the relevant potential (e.g., internal energy, enthalpy, or free energy) is specified, the peculiar class of materials being selected from the start; (ii) to construct an ingenious model of interactions, thus providing the form of the terms to be added to the classical balance laws, the constitutive equations being deduced and reduced to manageable forms in a second step. In recent papers[3-6] we however proposed a third method, already favored (but not in the same and systematic fashion) by Penfield an Haus,[7] and somewhat more formal than the preceding ones but, by the same token, much more powerful and with a wider range of application, namely, the method of *virtual power* (and *not* work) extended to continuous media and dynamical processes. It may be referred to as the use of d'Alembert's principle. The new point however is that this principle is used simultaneously with the now well-accepted requirement of *objectivity* (or material frame indifference), thus yielding the straightforward satisfaction of the so-called axiom of virtual power of internal forces[8] (after which the virtual power of internal forces vanishes identically in a virtual velocity field that "rigidifies" the material continuum and "freezes in" the interactions). The application to the theory of deformable ferromagnetic media within the framework of quasimagnetostatics[4] has shown how elegant, powerful, and simple this method proves to be. Elastic ferromagnets which exhibit special surface magneto-elastic couplings (via second-order strains and hyperstresses) have been studied in the like manner.[6] It is to be noted that no constitutive assumptions need be made to start with—the medium may be a deformable solid, a fluid or a continuum with an intermediary behavior with hereditary effects—and the theory applies to arbitrary thermodynamical processes in contradistinction to, for instance, Hamiltonian formulations germane to the description of thermodynamically reversible processes. Furthermore, the building of an involved model of electro-magneto-mechanical interactions is here avoided.

By clearly distinguishing between internal, external, prescribed, and inertial forces and requiring the objectivity solely for the internal forces which, from the phenomenological viewpoint, represent *interactions* that take place within the material continuum, be they of purely mechanical nature (e.g., intrinsic stresses) or of other nature (e.g., interations between neighboring magnetic spins), the method allows one to show, without studying peculiar constitutive equations, that all interactions participate in the total (i.e., Cauchy) stress tensor. This essential property applies not only to thermodynamically recoverable phenomena, but also to dissipative phenomena. When applied to elastic ferromagnets this property permits one to show that the spin–spin interactions and the spin relaxation due to spin–lattice interactions intervene not only in the spin precession equation, but also in the balance of linear momentum, along with the usual elastic forces and the viscosity processes in a rotationally invariant manner.[9] These results are extended to ferrimagnets in Part II of the present work.

In the present paper we propose to apply the same method to the more general case of elastic *ferrimagnets*, the case of elastic antiferromagnets being deduced as a special case. So far, no phenomenological theory of a similar degree of generality and rigor, built up in agreement with all principles of modern continuum physics, has been proposed. The only attempts at a theory of *linear* elastic antiferromagnets are those of the Russian school,[10–12] which do not describe the above mentioned couplings. The basic ingredients of the present approach are: (i) the principle of virtual power in the above-recalled generalized form, (ii) the first and second principle of thermodynamics in global form (iii) Maxwell's equations. The model used for the magnetic properties is the *multimagnetic-sublattice model* initiated by Nêel[13], and specialized to the case of a two-magnetic-sublattice model when dealing with antiferromagnets. Although there are no difficulties of principle to construct a fully dynamical theory (compare Refs. 5 and 14 for ferromagnets and dielectrics), we shall consider for the sake of simplicity the framework of quasimagnetostatics in insulators, since we are mostly interested in applications in the magnon–phonon frequency range, far outside the optical range. Thus electrical polarization is discarded. For the essentials of the theory of rigid ferrimagnets and ferrites the reader is referred to several monographs and reviews.[15–20]

The notation, quasimagnetostatic fields and the related energetic identities, and the "kinematics" of magnetic sublattices are recalled in Sec. 2. Local and global balance laws for ferrimagnetic deformable media are deduced from the principle of virtual power alone in Sec. 3. In the Appendix it is shown how an ingenious model of three interacting lattices (one crystal lattice and two magnetic sublattices) can be used to recover the case of magnetically saturated deformable antiferromagnets. In the companion paper numbered II the macroscopic thermodynamics (following Coleman's axiomatics) is given as well as the constitutive theory that follows therefrom for nonlinear elastic antiferromagnetic insulators. The special case of infinitesimal strains is then deduced. Simple dissipative processes are studied with the help of the Onsager–Casimir theory of irreversible processes, from which follow the spin relaxation contributions for large or small damping in the case of deformable antiferromagnets. Remarks pertaining to the case of rigid ferrimagnetic continua are made by way of conclusion.

2. PREREQUISITES

2.1. Motion, deformation field[21]

A. As a rule we use the standard Cartesian tensor notation in rectangular coordinate systems, the summation convention over repeated indices being understood. The direct (dyadic) notation is used when there is no ambiguity. Three-dimensional Euclidean space E^3 is referred to two orthonormal frames $\{\mathbf{g}_k\}$ and $\{\mathbf{G}_K\}$ $(k,K=1,2,3)$, respectively in the present configuration $\mathcal{K}$, with matter density ρ, at Newtonian time t, and in the reference configuration $\mathcal{K}_0$, with matter desity ρ_0, at time t_0. The motion of a continuous (deformable) media is described by the following diffeomorphism of class C^m $(m\geq 2)$ in $\mathcal{D}_t$, an open, bounded, simply connected region of E^3—with smooth boundary $\partial\mathcal{D}_t$ of unit outward normal $\mathbf{n}$—occupied by a material body $\mathcal{B}$ at time t in its configuration $\mathcal{K}$:

$$\mathbf{x}=\mathcal{X}(\mathbf{X},t) \tag{2.1}$$

where $\mathbf{x}$ and $\mathbf{X}$ are Cartesian coordinates in $\mathcal{K}$ and $\mathcal{K}_0$, respectively. A superimposed dot indicating the usual material time derivative, we have

$$\frac{d\mathbf{A}}{dt}\equiv\dot{\mathbf{A}}=\frac{\partial\mathbf{A}}{\partial t}+(\mathbf{U}\cdot\nabla)\mathbf{A} \tag{2.2}$$

for any tensor-valued field $\mathbf{A}(\mathbf{x},t)$, where

$$\mathbf{U}=\left.\frac{\partial\mathcal{X}}{\partial t}\right|_{\mathbf{X}} \tag{2.3}$$

is the classical velocity field. The velocity–gradient tensor, the rate-of-strain tensor, the rate-of-rotation tensor, and the vorticity vector are defined by

$$(\nabla\mathbf{U})_{ij}=U_{i,j}=D_{ij}+\Omega_{ij}, \tag{2.4a}$$

$$D_{ij}\equiv U_{(i,j)}=\tfrac{1}{2}(U_{i,j}+U_{j,i}), \tag{2.4b}$$

$$\Omega_{ij}\equiv U_{[i,j]}=\tfrac{1}{2}(U_{i,j}-U_{j,i}), \tag{2.4c}$$

$$\Omega_i\equiv-\tfrac{1}{2}\epsilon_{ijk}\Omega_{jk}=\tfrac{1}{2}(\nabla\times\mathbf{U})_i, \tag{2.4d}$$

respectively. ϵ_{ijk} is the permutation symbol. The divergence of tensors is here taken on the last index, e.g., $(\mathrm{div}\mathbf{t})_i\equiv t_{ij,j}$.

The Jacobian determinant of the motion (2.1) is given by

$$J=\rho_0/\rho; \tag{2.5}$$

this is one form of the continuity equation (in $\mathcal{K}_0$), the equivalent statement in $\mathcal{K}$ being given by either one of the following well-known forms:

$$\dot{\rho}+\rho D_{kk}=0,\quad \frac{\partial\rho}{\partial t}+\nabla\cdot(\rho\mathbf{U})=0\quad\text{in } \mathcal{D}_t. \tag{2.6}$$

B. The usual differential equation of a *rigid-body motion* (Killing's theorem) is expressed by

$$D_{ij}=0 \tag{2.7}$$

at all points $\mathbf{x}\in\mathcal{D}_t$ and for all times. The integral of Eq. (2.7) reads

$$\hat{x}_j = Q_{ij}(t)x_j + c_i(t) \tag{2.8}$$

where $\mathbf{c}$ is a time-dependent vector and $\mathbf{Q}$ is a *time-dependent* proper orthogonal tensor, i.e.,

$$\mathbf{Q}\mathbf{Q}^T=\mathbf{I},\quad \mathbf{Q}^T=\mathbf{Q}^{-1},\quad \det\mathbf{Q}=+1. \tag{2.9}$$

Coordinate transformations of the type (2.8) play an essential role in the subsequent development, when they are considered as *active* transformations, i.e., when they represent the motion of a *rigid* body $\mathcal{B}$. Then the associated velocity field, described in K, is given by

$$\hat{U}_i(\mathbf{x},t)=\bar{U}_i(t)+\bar{\Omega}_{ij}(t)x_j, \tag{2.10}$$

wherein

$$\bar{U}_{i,j}=0,\quad \bar{\Omega}_{ij}=-\bar{\Omega}_{ji},\quad \bar{\Omega}_{ij,k}=0, \tag{2.11a}$$

$$\bar{\Omega}_{ij}=\dot{Q}_{ki}Q_{kj},\quad \bar{U}_i=\dot{c}_i, \tag{2.11b}$$

at all $\mathbf{x}\in\bar{\mathcal{D}}_t$, where $\bar{\mathcal{D}}_t$ is the closure of $\mathcal{D}_t$. Equipped with, for instance, the norm of uniform convergence, the linear space spanned by all velocity fields (2.10) is a vector-valued *topological linear space* (T.L.S.) of dimension six (three parameters defining the degrees of freedom of translation and three parameters defining the degrees of freedom of rotation, these parameters being uniform throughout the rigid material body, but possibly time dependent) called the *distributor space* of rigid-body motions, $C(\bar{\mathcal{D}}_t)$ for the body $\mathcal{B}$.

Geometrical objects $\mathbf{A}(\mathbf{x},t)$ whose tensor components transform tensorially with respect to the transformations (2.8) (considered as *passive* point transformations) are said either to be *objective* or to satisfy the so-called principle of material frame indifference. In particular, $\mathbf{U}(\mathbf{x},t)$ and $\Omega_{ij}(\mathbf{x},t)$ (for a deformable or rigid body), and the material derivative of an objective tensor field are *not* objective tensor fields, whereas $D_{ij}(\mathbf{x},t)$ is an objective tensor field. Of particular importance for the remainder of this work is the objective time derivative known as the *Jaumann* or *corotational* derivative,[22] noted D_J. For a vector field $\mathbf{A}(\mathbf{x},t)$ and a second-order (in general not symmetric) tensor field $\mathcal{A}(\mathbf{x},t)$ we have

$$(D_J\mathbf{A})_i\equiv\dot{A}_i-\Omega_{ij}A_j=\left[\left(\frac{d}{dt}-\mathbf{\Omega}\times\right)A\right]_i, \tag{2.12a}$$

$$(D_J\mathcal{A})_{ij}\equiv\dot{\mathcal{A}}_{ij}-\Omega_{jk}\mathcal{A}_{ik}-\Omega_{ik}\mathcal{A}_{kj}. \tag{2.12b}$$

For $\mathcal{A}\equiv\nabla\mathbf{A}=\{A_{i,j}\}$, it can be remarked that the following quantity

$$(D_J\nabla\mathbf{A})_{ij}+A_{i,k}D_{kj}=(\dot{A}_i)_{,j}-\Omega_{ik}A_{k,j} \tag{2.13}$$

is also objective and, moreover, is linearly independent of D_{ij}.

Finally, we say that a vector field $\mathbf{A}(\mathbf{x},t)$ is *frozen in* the deformable matter if and only if

$$D_J\mathbf{A}=0. \tag{2.14}$$

Indeed, if this equation is satisfied, then Eq. (2.12a) says that the vector field $\mathbf{A}$ rotates at the same local rate as the deformable matter.

2.2. Quasimagnetostatic fields[23]

Let $\mathbf{B}$, $\mathbf{H}$, Φ, $\mathbf{M}$ and $\mu\equiv\mathbf{M}/\rho$ be the magnetic induction, the magnetic field, the magnetic scalar potential, the magnetization per unit volume, and the magnetization per unit mass in K at time t. Then, in Lorentz–Heaviside units, Maxwell's equations for magnetostatics read:

$$\mathbf{H}=\mathbf{B}-\mathbf{M}=-\nabla\Phi, \tag{2.15}$$

$$\nabla^2\Phi-\nabla\cdot\mathbf{M}=0\quad\text{in }\mathcal{D}_t, \tag{2.16}$$

and

$$[\![\partial\Phi/\partial n]\!]+\mathbf{M}^{\text{in}}\cdot\mathbf{n}=0\quad\text{on }\partial\mathcal{D}_t, \tag{2.17}$$

where the symbolism $[\![\cdots]\!]$ indicates the jump, $\partial/\partial n\equiv\mathbf{n}\cdot\nabla$, and the superscript *in* means the value on the inside face of $\partial\mathcal{D}_t$.

The *ponderomotive force*—the arbitrariness of which must be emphasized—and *couple* acting upon the unit volume element of magnetized matter in $\mathcal{D}_t$ read

$$\mathbf{f}^{\text{em}}=(\nabla\mathbf{B})\cdot\mathbf{M}, \tag{2.18a}$$

$$\mathbf{c}^{\text{em}}=\mathbf{M}\times\mathbf{B}. \tag{2.18b}$$

The electromagnetic stress tensor t^{em}_{ij}, the skewsymmetric stress tensor $C^{\text{em}}_{ij}=-C^{\text{em}}_{ji}$ associated with $\mathbf{c}^{\text{em}}$ and the electromagnetic surface traction $\mathbf{T}^{\text{em}}$ are introduced via the equations

$$\mathbf{f}^{\text{em}}=\operatorname{div}\mathbf{t}^{\text{em}},\quad C^{\text{em}}_{ij}=-t^{\text{em}}_{[ij]}, \tag{2.19}$$

$$t^{\text{em}}_{ij}=H_iB_j-(\tfrac{1}{2}\mathbf{B}^2-\mathbf{M}\cdot\mathbf{B})\delta_{ij}, \tag{2.20}$$

$$C^{\text{em}}_{ij}\equiv\tfrac{1}{2}\epsilon_{ijk}c^{\text{em}}_k,\quad T^{\text{em}}_i=-[\![t^{\text{em}}_{ij}]\!]n_j, \tag{2.21}$$

so that the following global identity is obtained by integrating the first of Eqs. (2.19) over $\mathcal{D}_t$:

$$\int_{\mathcal{D}_t}\mathbf{f}^{\text{em}}\,dv+\int_{\partial\mathcal{D}_t}\mathbf{T}^{\text{em}}\,da=\mathbf{0}. \tag{2.22}$$

No electromagnetic momentum appears in Eqs. (2.19a) and (2.22) because of the magnetostatic hypothesis. Of equal importance for what follows is a global energetic identity for the magnetostatic fields obtained by specializing to the magnetostatic frame the general identity valid for the electrodynamics of moving bodies derived in a previous paper.[14] We have

$$\frac{d}{dt}U^{\text{em}\cdot\text{m}}(\mathcal{D}_t)=-\int_{\mathcal{D}_t}(\mathbf{f}^{\text{em}}\cdot\mathbf{U}+\rho\mathbf{B}\cdot\dot{\mu})\,dv-\int_{\partial\mathcal{D}_t}\mathbf{T}^{\text{em}}\cdot\mathbf{U}\,da, \tag{2.23}$$

or, on account of Eqs. (2.19), (2.21), and (2.4a),

$$\frac{d}{dt}U^{\text{em}\cdot\text{m}}(\mathcal{D}_t)=\int_{\mathcal{D}_t}(t^{\text{em}}_{ij}D_{ij}+C^{\text{em}}_{ij}\Omega_{ji}-\rho\mathbf{B}\cdot\dot{\mu})\,dv, \tag{2.24}$$

where

$$U^{\text{em}\cdot\text{m}}(\mathcal{D}_t)=\int_{\mathcal{D}_t}(\tfrac{1}{2}\mathbf{B}^2-\mathbf{M}\cdot\mathbf{B})\,dv. \tag{2.25}$$

2.3. The magnetic sublattices and their "kinematics"

Consider the following most general continuous description of the magnetization field in a deformable magnetically ordered crystal below its magnetic phase-transition temperature T_{cr}. At each point $\mathbf{x} \in \mathcal{D}_t$ in the configuration $\mathcal{K}$ at time t the magnetization per unit mass is the vectorial resultant of the sum of n magnetization fields $\boldsymbol{\mu}_\alpha$, $\alpha = 1, 2, \cdots, n$, the *magnetic sublattices*, arising at $\mathbf{x}$ from n different ionic species defined by unit mass in $\mathcal{K}$, and having spectroscopic splitting factors g_α (in the usual paramagnetic case) and gyromagnetic ratio $\gamma_\alpha = g_\alpha e/2m_0c$ (e: electronic charge; m_0: rest mass of the electron; c: light velocity in vacuo). In accordance with microscopic considerations, a spin density $\mathbf{s}_\alpha$ per unit mass is associated with each $\boldsymbol{\mu}_\alpha$ via the gyromagnetic relation

$$\mathbf{s}_\alpha(\mathbf{x}, t) = \gamma_\alpha^{-1} \boldsymbol{\mu}_\alpha(\mathbf{x}, t). \tag{2.26}$$

The total spin intrinsic momentum per unit mass is thus given by

$$\mathbf{s} = \sum_\alpha \mathbf{s}_\alpha = \sum_\alpha \gamma_\alpha^{-1} \boldsymbol{\mu}_\alpha, \tag{2.27}$$

whereas

$$\boldsymbol{\mu}(\mathbf{x}, t) = \sum_\alpha \boldsymbol{\mu}_\alpha(\mathbf{x}, t) = \sum_\alpha \boldsymbol{\mu}_\alpha(\mathbf{X}, t). \tag{2.28}$$

Equation (2.1) has been used for writing the last expression.

Of course, if all magnetic moments of the ions arise only from spin, the g_α are all equal to $g_e = 2$, the splitting factor of electrons, and there is no distinction between the different magnetic orders (e.g., ferromagnetism and ferrimagnetism, since, then, $\mathbf{s} = \gamma_e^{-1} \boldsymbol{\mu}$ with $\gamma_e = g_e e/2m_0 c$. In general, however, values of $g = |\sum_\alpha g_\alpha \mathbf{s}_\alpha| / |\sum_\alpha \mathbf{s}_\alpha|$ appreciably different from two are often found, which result is important in discussing the value of resonances.[24] In the case of ferrimagnetic bodies which is our concern, $\boldsymbol{\mu}$, as defined by Eq. (2.28) may be different from zero in absence of external field below the critical temperature. In the case of antiferromagnetic bodies where $T_{cr} = \theta_N$ (θ_N is Nèel temperature) two magnetic sublattices at least, $\boldsymbol{\mu}_\alpha$, $\alpha = A, B$, need be considered for, below θ_N, $\boldsymbol{\mu}$ as given by Eq. (2.28) vanishes in absence of external magnetic field (then $\boldsymbol{\mu}_A$ and $\boldsymbol{\mu}_B$ are antiparallel and of equal magnitude). In all these cases it can be considered at temperatures much below the corresponding critical temperature that each magnetic sublattice has, at each point $\mathbf{x} \in \mathcal{D}_t$, an amplitude independent of time. That is,

$$\boldsymbol{\mu}_\alpha \cdot \boldsymbol{\mu}_\alpha = \mu_\alpha^2(\mathbf{X}), \quad \mu_\alpha^2 \cdot \dot{\boldsymbol{\mu}}_\alpha = 0. \tag{2.29}$$

It follows from the second of these that each $\boldsymbol{\mu}_\alpha$ has necessarily at $\mathbf{x}$ and instantaneously at time t a time-evolution equation of the type

$$\dot{\boldsymbol{\mu}}_\alpha = \boldsymbol{\omega}_\alpha \times \boldsymbol{\mu}_\alpha, \tag{2.30}$$

where $\boldsymbol{\omega}_\alpha(\mathbf{x}, t)$ is the instantaneous and local precessional velocity of $\boldsymbol{\mu}_\alpha$. Of course,

$$\boldsymbol{\omega}_\alpha \cdot \dot{\boldsymbol{\mu}}_\alpha = 0 \tag{2.31}$$

or, on account of Eq. (2.26),

$$\gamma_\alpha^{-1} \dot{\boldsymbol{\mu}}_\alpha \cdot \boldsymbol{\omega}_\alpha = 0. \tag{2.32}$$

This equation means that $\mathbf{s}_\alpha$ is a *d'Alembertian inertia couple* (i.e., a gyroscopic couple) which does not produce any power in a real precessional velocity.

Furthermore, for $\theta \ll T_{cr}$, it is often assumed that the amplitude of each magnetic sublattice is uniform throughout the specimen in the reference configuration $\mathcal{K}_0$; then it is said that each magnetic sublattice is *magnetically saturated*, so that, in supplement to Eqs. (2.29), the following constraints must also be satisfied:

$$\boldsymbol{\mu}_\alpha \cdot \boldsymbol{\mu}_\alpha = \mu_{S\alpha}^2 = \text{const}; \tag{2.33}$$

hence

$$(\partial \mu_{S\alpha}/\partial \mathbf{X}_K) = 0 \quad \text{or} \quad \mu_{\alpha i} \mu_{\alpha i, K} = 0. \tag{2.34}$$

$\mu_{S\alpha}$ is the saturation value of $\boldsymbol{\mu}_\alpha$. Then it is clear that the most important problem to be dealt with is the finding of the expression of $\boldsymbol{\omega}_\alpha(\mathbf{x}, t)$ in function of the different interactions that take place within the magnetic body, and the finding of the angular distribution of the magnetic sublattices at each point and for all times throughout the deformable region $\mathcal{D}_t$, the value of $|\boldsymbol{\mu}_\alpha|$ and $|\boldsymbol{\mu}|$ in function of temperature being not examined in the present context.

If one assumes that each $\boldsymbol{\mu}_\alpha$ is an objective field, and since $\boldsymbol{\mu}_\alpha$ is subjected to rotational motion (cf. Eq. (2.30), it is clear that relevant objective time rates of $\boldsymbol{\mu}_\alpha$ and its spatial gradient $\nabla \boldsymbol{\mu}_\alpha$ are provided by Jaumann derivatives, so that in accordance with Eqs. (2.12a) and (2.13) we define the following objective fields, which prove essential in the sequel ($\alpha = 1, 2, \ldots, n$):

$$\hat{m}_{\alpha i} \equiv (D_J \mu_\alpha)_i = \dot{\mu}_{\alpha i} - \Omega_{ij} \mu_{\alpha j}, \tag{2.35}$$

$$\hat{\mathfrak{M}}_{\alpha ij} \equiv (\dot{\mu}_{\alpha i})_{,j} - \Omega_{ik} \mu_{\alpha k, j}. \tag{2.36}$$

3. FIELD EQUATIONS

3.1. The principle of virtual power for ferrimagnetic continua

A. The generalized velocity field

Generalizing the formulation given for ferromagnetic deformable bodies in a previous paper,[4] we consider a generalized "motion" (set of primitive independent variables) represented by the $(n+1)$-tuple of vectors

$$\{\mathbf{x}, \boldsymbol{\mu}_\alpha; \alpha = 1, 2, \ldots, n\} \tag{3.1}$$

in a ferrimagnetic continuum whose magnetic structure is made of n magnetic sublattices. On account of Eq. (2.1), one can consider

$$\chi(\mathbf{X}, t) = \{\mathcal{X}(\mathbf{X}, t), \boldsymbol{\mu}_\alpha(\mathbf{X}, t); \alpha = 1, 2, \ldots, n\}. \tag{3.2}$$

The generalized velocity field of the present theory of magnetomechanical interactions is assumed to be, at each point $\mathbf{x} \in \bar{\mathcal{D}}_t$, an element of a $3(n+1)$-dimensional T.L.S (the topology being, for instance, that induced by the norm of uniform convergence) such that, for fixed t,

$$v(\mathbf{x}) = \{\mathbf{U}(\mathbf{x}, t), \dot{\boldsymbol{\mu}}_\alpha(\mathbf{x}, t); \alpha = 1, 2, \ldots, n\}. \tag{3.3}$$

Of course, for $\theta \ll T_{cr}$, we can make use of Eq. (2.30) so that, equivalent to (3.3), one may consider the following $3(n+1)$-dimensional T.L.S.:

$$\hat{v}(\mathbf{x}) = \{\mathbf{U}(\mathbf{x},t); \omega_\alpha(\mathbf{x},t); \alpha = 1,2,\ldots,n\}. \tag{3.4}$$

Following the terminology introduced in a previous work work,[4] we consider a so-called *first-order-gradient* theory with respect to $v(\mathbf{x})$, as given by Eq. (3.3), for deformable ferrimagnets. Thus v must be enlarged with the elements obtained by taking the first spatial gradients of its elements to yield a new $12(n+1)$-dimensional T.L.S that we can write, on account of Eq. (2.4a), as

$$\mathcal{V}(\mathbf{x}) = \{U_i, D_{ij}, \Omega_{ij}, \mathring{\mu}_{\alpha i}, (\mathring{\mu}_{\alpha i})_{,j} \mid i,j = 1,2,3; \alpha = 1,2,\ldots,n\}. \tag{3.5}$$

It is clear that the introduction of the first spatial gradients of $\mathbf{U}$ and $\mathring{\boldsymbol{\mu}}_\alpha$ allows us to give a better description of these fields in a neighborhood of a point of $\mathcal{D}_t$ (the underlying idea is that of Taylor expansion in a neighborhood). No higher order gradients are considered for the sake of simplicity and because they are in fact unnecessary to obtain a realistic representation of physical phenomena. Indeed, although the above reasoning concerns velocity (or, more generally, time-rate) fields, because we shall apply these considerations to the expression of various *powers*, the gradient $U_{i,j}$ (or D_{ij} and Ω_{ij}) allows one to describe, in terms of deformation fields, so-called simple materials—according to the term coined by Noll[25]—which include linear elastic or classical Newtonian fluid behaviors as well as nonlinear elastic and non-Newtonian fluids behaviors (and intermediate behaviors such as viscoelasticity), whereas the gradients $\nabla\mathring{\boldsymbol{\mu}}_\alpha$ allows us to describe, in terms of the gradients of the fields $\mathring{\boldsymbol{\mu}}_\alpha$ present in (3.2), the *exchange* and *superexchange* mechanisms arising between neighboring spins within the same sublattice and between spins of different ionic species in a phenomenological manner, as will be shown below. The fact that gradients of the magnetization offer, when introduced as independent variables in the energy density, a satisfactory phenomenological description of exchange (Heisenberg) forces between neighboring spins has been known for some decades, starting with the pioneering work of Landau and Lifshitz[26], and is also emphasized in Brown's monograph.[27] A subspace of $v(\mathbf{x})$ and $\hat{v}(\mathbf{x})$ clearly is the restriction, $\mathcal{C}(\bar{\mathcal{D}}_t)|_{\mathbf{x}}$, of $\mathcal{C}(\bar{\mathcal{D}}_t)$ at a fixed $\mathbf{x} \in \mathcal{D}_t$.

B. The virtual power of internal forces

A virtual power in general is a linear form on a set of virtual velocities or, in other words, is the scalar which results from the scalar product between a force and a virtual velocity. However, *internal* forces that represent phenomenologically interactions (in a broad sense) and for which, ultimately, constitutive equations must be given, are here supposed, in accordance with the now well-accepted principles of modern continuum mechanics, to be objective, i.e., form-invariant under point transformations of the type (2.8). Clearly, the corrresponding generalized velocities which are the cofactors of the generalized internal forces in the resulting virtual power must also be objective in virtue of the trivial invariance of the scalar product. Then the virtual power of generalized internal forces ought to be a linear form (or a linear functional when the whole body is considered and thus requires space integration) on a set of objective generalized velocities. Thus the main problem in every continuum theory approached by using the principle of virtual power as starting point is the building of this adequate set of objective generalized velocities, called $\mathcal{V}_{\text{obj}}(\mathbf{x})$ at each point $\mathbf{x} \in \mathcal{D}_t$, the corresponding generalized forces being formally introduced as cofactors and interpreted by means of dimensional analysis and thanks to the duality inherent in the method.

In the present case consider the primitive generalized velocity field (3.3). It is a simple matter to show that $\mathcal{V}_{\text{obj}}(\mathbf{x})$ is the quotient space

$$\mathcal{V}_{\text{obj}}(\mathbf{x}) = \mathcal{V}(\mathbf{x})/\mathcal{C}(\bar{\mathcal{D}}_t)|_{\mathbf{x}}. \tag{3.6}$$

Thus $\mathcal{V}_{\text{obj}}$ is a $12(n+1)-6 = 6(2n+1)$-dimensional T.L.S. A linearly independent (but obviously not unique) basis of elements which spans this T.L.S. is easily found as follows. $\mathbf{U}$ is not objective and must be rejected (it is not possible to combine it with other fields in the present case in order to construct an objective field). D_{ij} is objective and can be kept as such. Although Ω_{ij} and $\mathring{\boldsymbol{\mu}}_\alpha$ and $\nabla\mathring{\boldsymbol{\mu}}_\alpha$ are not objective, linear objective combinations of these (which do not depend on D_{ij}) can be constructed. A straightforward solution is obtained by considering the objective rates defined by Eqs. (2.35) and (2.36). Thus we can write ($\mathbf{x} \in \mathcal{D}_t$):

$$\mathcal{V}_{\text{obj}}(\mathbf{x}) = \{D_{ij}, \hat{\mathfrak{M}}_{\alpha ij} \mid i,j = 1,2,3; \alpha = 1,2,\ldots,n\}. \tag{3.7}$$

Introducing the set of cofactors (generalized internal forces)

$$\mathcal{F}_{\text{int}}(\mathbf{x}) = \{-\sigma_{ij}, f^L B_{\alpha i}, -\mathcal{B}_{\alpha ij} \mid i,j = 1,2,3; \alpha = 1,2,\ldots,n\}$$

which spans the dual T.L.S of $\mathcal{V}_{\text{obj}}$ [in topological terms, $\mathcal{V}_{\text{obj}}$ and $\mathcal{F}_{\text{int}}$ are placed in *separating duality* via the bilinear form $(\mathbf{A},\mathbf{B}) = \text{tr}(\mathbf{A}\mathbf{B}^T)$, $\mathbf{A} \in \mathcal{F}_{\text{int}}$, $\mathbf{B} \in \mathcal{V}_{\text{obj}}$, T = transpose], the total power developed by the internal forces of the present theory, for the spatial volume $\mathcal{D}_t$ at time t, is given by the following linear continuous functional (an asterisk indicates a virtual field or the value of an expression in such a field):

$$\mathcal{P}^*_{(i)}(\mathcal{D}_t, \mathcal{V}^*_{\text{obj}}) = -\int_{\mathcal{D}_t} \Big(\sigma_{ij} D^*_{ij} - \sum_\alpha \rho^L B_{\alpha i} \hat{m}^*_{\alpha i} + \sum_\alpha \mathcal{B}_{\alpha ij} \hat{\mathfrak{M}}^*_{\alpha ij}\Big)\, dv. \tag{3.8}$$

The signs are chosen, and ρ is introduced, for convenience. The physical interpretation of the elements of $\mathcal{F}_{\text{int}}$ is immediate. The symmetric tensor σ_{ij} represents *intrinsic stresses*, which would be present even in the absence of magnetic effects. The ${}^L\mathbf{B}_\alpha$, $\alpha = 1, 2, \ldots, n$, have the dimension of a magnetic field. If each magnetic sublattice is *frozen in* the material continuum, then $\hat{\mathbf{m}}_\alpha = 0$ according to Eqs. (2.14) and (2.35). Then the ${}^L\mathbf{B}_\alpha$ do not participate in the power consumption. One type of interaction is thus suppressed. It is thus expected that the ${}^L\mathbf{B}_\alpha$ represent in some way an interaction between each magnetic sublattice and the material continuum, i.e., the crystal lattice. This is known as the spin–lattice interaction, which yields the notions of *magnetic anisotropy* (since ${}^L\mathbf{B}_\alpha$ is linked to the relative orientation of the $\boldsymbol{\mu}_\alpha$ with respect to the crystal lattice) and of *magnetocrystalline energy*. Furthermore, in the present case concerned with ferrimagnetism, it will be shown later on that the fields ${}^L\mathbf{B}_\alpha$ also account for the *intermagnetic-sublattice* interactions not due to the spatial disuniformities in the fields μ_α. The fields $\mathcal{B}_{\alpha ij}$, whose dimension is (magnetic field) $\times ML^{-4}$, account for the spatial disuniformities in the magnetization fields $\boldsymbol{\mu}_\alpha$. They represent thus the short-range *intra-* and *inter*magnetic-sublattice interactions. The $\mathcal{B}_{\alpha ij}$ will be referred to as the *spin-interaction* (not symmetric) tensors.

Consider now the following "virtual" situation: the material body $\mathcal{B}$ is "rigidified," thus D_{ij} satisfies the Eq. (2.7), and all magnetic sublattices are frozen in this rigidified body. Then ($\emptyset$ = empty set),

$$R[\mathcal{V}_{\mathrm{obj}}]^* = \{D^*_{ij}=0,\ \hat{\mathbf{m}}^*_\alpha = \mathbf{0}, \hat{\mathfrak{M}}^*_{\alpha ij}=0\} = \emptyset. \tag{3.9}$$

The last expression in $R[\mathcal{V}_{\mathrm{obj}}]^*$ follows from the second and Eq. (2.11a)[3]. Then the corresponding virtual power (3.8) vanishes identically. Equations (3.9) are the differential equations of the generalized rigid body motion of the present theory: all interactions are frozen in. This follows from the algebraic result: $\mathcal{C}(\bar{\mathcal{D}}_t)|_{\mathbf{x}} \cap \mathcal{V}_{\mathrm{obj}}(\mathbf{x}) = \emptyset$, which follows from the definition (3.6). We obviously have: $\mathcal{C}(\mathcal{D}_t) = \text{Kernel}\,[\mathcal{P}_{(i)}]$, i.e.,

$$\mathcal{P}^*_{(i)}(\mathcal{D}_t, V^* \in \mathcal{C}) \equiv 0. \tag{3.10}$$

This statement is none other than the expression of the so-called *axiom of virtual power of internal forces* extended to interactions other than purely mechanical ones.

C. Other virtual powers

There obviously is no restriction of objectivity placed upon the virtual power of external forces and inertia forces (the latter are in fact never objective). External forces are subdivided in two types: those forces which act per unit volume within $\mathcal{D}_t$ and may be considered as the result of *at-a-distance* actions, and those which represent *contact* actions on the boundary $\partial\mathcal{D}_t$ of $\mathcal{D}_t$ at time t. The first type of forces is here prescribed in the sense that their expression is provided by physical theories foreign to continuum mechanics, *per se*, e.g., gravitation and electromagnetism. Forces of the second type have numerically prescribed values (or the dual condition in terms of velocities is prescribed) or they are unknowns to be determined in the process of problem solving. Let $\mathcal{P}_{(d)}$ and $\mathcal{P}_{(c)}$ denote, respectively, the power of *volume* at-a-distance forces and the power of *contact* forces. In general we have ($\mathbb{R}$: real line):

$$\mathcal{P}_{(d)}: \mathcal{V} \mapsto \mathbb{R}, \quad \mathcal{P}_{(c)}: v \mapsto \mathbb{R}. \tag{3.11}$$

However, we may discard in $\mathcal{P}_{(d)}$ all contributions which receive neither theoretical nor experimental support. For instance, if $\mathcal{V}$ is given by Eq. (3.5), we know of no external field which may be the dual of $(\hat{\mu}_i)_{,j}$, so that this term is discarded, and we can write formally on account of (3.11) and (3.5):

$$\mathcal{P}^*_{(d)}(\mathcal{D}_t, \mathcal{V}^*) = \int_{\mathcal{D}_t} (\mathbf{f}\cdot\mathbf{U}^* + \Phi_{ij}D^*_{ij} + C_{ij}\Omega^*_{ij} + \sum_\alpha \rho L_{\alpha i}\hat{\mu}^*_{\alpha i})\, dv. \tag{3.12}$$

The volume force $\mathbf{f}$ may represent the action of gravity if necessary. $\Phi_{ij}=\Phi_{ji}$ and $C_{ij}=-C_{ji}$ represent, respectively, a double symmetric volume force (i.e., a symmetric stress tensor) and a volume couple (or a skew symmetric stress). We think it is logical and convenient to use here the ambiguity in the interpretation of the actions of electromagnetic fields on matter, and to consider them as giving rise solely to volume at-a-distance actions. Thus, following previous works[4,5] and on account of Eq. (2.28) and of the form of Eq. (2.24), we propose the following identification:

$$\Phi_{ij} = -t^{\mathrm{em}}_{(ij)}, \quad C_{ij} = -t^{\mathrm{em}}_{[ij]}, \quad L_{\alpha i} = B_i,\ \forall\ \alpha. \tag{3.13}$$

This means that, in contrast to Eq. (2.23), electromagnetic fields are introduced only in the form of internal stresses and via the power developed by the magnetic sublattices in such fields. Thus the EM fields will not participate in $\mathcal{P}_{(c)}$ below [of course, the alternate formulation using (2.23) can also be considered]. Equation (3.12) reads thus

$$\mathcal{P}^*_{(d)}(\mathcal{D}_t, \mathcal{V}^*) = \int_{\mathcal{D}_t} (\mathbf{f}\cdot\mathbf{U}^* - t^{\mathrm{em}}_{ij}D^*_{ij} - t^{\mathrm{em}}_{[ij]}\Omega^*_{ij} + \rho\mathbf{B}\cdot\hat{\boldsymbol{\mu}}^*)\, dv. \tag{3.14}$$

As explicited by the second of Eqs. (3.11), $\mathcal{P}_{(c)}$ is considered to be a continuous linear functional on the space (3.3) and not on the larger space (3.5). This results from pure mathematical reasons for we assume that the material body $\mathcal{B}$ has, at all times in the course of its motion, a continuous tangent plane (i.e., no edges: for instance, $\mathcal{B}$ may have an ellipsoidal shape), so that terms involving gradients would disppear automatically by using the surface Stokes theorem). Thus,

$$\mathcal{P}^*_{(c)}(\partial\mathcal{D}_t, v^*) = \int_{\partial\mathcal{D}_t} (\mathbf{T}\cdot\mathbf{U}^* + \sum_\alpha \rho\mathcal{T}_\alpha\cdot\dot{\boldsymbol{\mu}}^*_\alpha)\, da. \tag{3.15}$$

$\mathbf{T}$ is a surface traction not due to electromagnetic fields (see an above-made remark). Clearly, the $\mathcal{T}_\alpha$, whose dimension is that of a surface distribution of magnetic dipoles, are the surface "exchange" contact "forces" that correspond to the internal forces $\mathcal{B}_{\alpha ij}$. It is shown hereinafter that they give rise to surface densities of couples (the dual notion being that of pinning and orientation of the magnetic sublattices at the bounding surface).

Finally, two types of inertia arise in the present theory. The first one is the classical inertia $\rho\dot{\mathbf{U}}$ due to the motion of the crystal lattice (viewed macroscopically). The second one is related to the intrinsic spin density associated with each magnetic sublattice. The latter does not work in *real* precessional velocity fields—cf. Eq. (2.32)—but it can be accounted for if virtual precessional velocities $\boldsymbol{\omega}^*_\alpha$ are considered. This clearly is a tremendous advantage of the use of the virtual power principle. Thus,

$$\mathcal{P}_{(a)}: v \mapsto \mathbb{R},$$
$$\mathcal{P}^*_{(a)}(\mathcal{D}_t, v^*) = \int_{\mathcal{D}_t} \rho(\dot{\mathbf{U}}\cdot\mathbf{U}^* + \sum_\alpha \gamma_\alpha^{-1}\dot{\boldsymbol{\mu}}_\alpha\cdot\boldsymbol{\omega}^*_\alpha)\, da, \tag{3.16}$$

where ω^*_α is related to $\dot{\boldsymbol{\mu}}^*_\alpha$ by the equation obtained by inverting Eq. (2.30) for virtual fields:

$$(\mu^2_\alpha\delta_{ij} - \mu_{\alpha i}\mu_{\alpha j})\omega^*_{\alpha j} = (\boldsymbol{\mu}_\alpha \times \dot{\boldsymbol{\mu}}^*_\alpha)_i. \tag{3.17}$$

D. Statement of the principle

In a Galilean frame and for an absolute Newtonian chronology, the virtual power of the inertia forces of a mechanical subsystem S balances the virtual power of all other forces, internal or external, impressed on the system, for any virtual velocity field. Thus

$$\mathcal{P}_{(a)}^*(\mathcal{D}_t, v^*) = \mathcal{P}_{(i)}^*(\mathcal{D}_t, \mathcal{V}_{\text{obj}}^*) + \mathcal{P}_{(d)}^*(\mathcal{D}_t, \mathcal{V}^*) + \mathcal{P}_{(c)}^*(\partial\mathcal{D}_t, v^*), \tag{3.18}$$

where the different expressions are provided by Eqs. (3.16), (3.8), (3.14), and (3.15), respectively. The expression (3.18) is posited to be valid at all times t in the course of the motion and deformation processes of the material body $\mathcal{B}$, for arbitrary *virtual* (or real) velocity fields (3.3) defined at all $\mathbf{x} \in \bar{\mathcal{D}}_t$ and satisfying the constraints (2.30), and for arbitrary small regions within $\mathcal{D}_t$ and on $\partial\mathcal{D}_t$, provided these are sufficiently regular. The remainder of Part I of this work is devoted to exploiting Eq. (3.18) for various virtual fields.

3.2. Local field equations in deformable ferrimagnets

For every couple $(\mathbf{U}^*, \omega_\alpha^*)$ at all points of $\mathcal{D}_t$ and $\partial\mathcal{D}_t$, after using Green–Gauss' theorem when necessary, and on account of Eqs. (2.30), (2.19), and (2.21), we obtain the following results:

Theorem: The local field equations that govern the motion and the interactions in a deformable ferrimagnet, according to the multi-sublattice model and for a theory of the first gradient in quasimagnetostatics, are:

$$* \text{ in } \mathcal{D}_t, \quad t_{ij,j} + f_i + f_i^{\text{em}} = \rho \dot{U}_i, \tag{3.19}$$

$$* \text{ on } \partial\mathcal{D}_t, \quad t_{ij} n_j = T_i + T_i^{\text{em}}, \tag{3.20}$$

and

$$* \text{ in } \mathcal{D}_t, \quad \overset{\circ}{\boldsymbol{\mu}}_\alpha = \boldsymbol{\omega}_\alpha \times \boldsymbol{\mu}_\alpha, \quad \boldsymbol{\omega}_\alpha = -\gamma_\alpha \mathbf{B}_\alpha^{\text{eff}}, \tag{3.21}$$

$$* \text{ on } \partial\mathcal{D}_t, \quad \epsilon_{ipq}(\mathcal{B}_{\alpha pj} n_j - \rho \mathcal{T}_{\alpha p})\mu_{\alpha q} = 0, \tag{3.22}$$

$\alpha = 1, 2, \ldots, n$, where we have defined

$$t_{ij} \equiv \sigma_{ij} + \sum_\alpha (\hat{\sigma}_{\alpha[ij]} + \tilde{\sigma}_{\alpha[ij]}), \tag{3.23}$$

$$\hat{\sigma}_{\alpha ij} \equiv \rho^L B_{\alpha i} \mu_{\alpha j}, \quad \tilde{\sigma}_{\alpha ij} \equiv -\mathcal{B}_{\alpha ik} \mu_{\alpha j,k} \tag{3.24}$$

(no summation over α), and

$$B_{\alpha i}^{\text{eff}} \equiv B_i + {}^L B_{\alpha i} + \rho^{-1} \mathcal{B}_{\alpha ij,j}. \tag{3.25}$$

These equations are supplemented with the continuity equation (2.6) and Maxwell's equations (2.16)–(2.17) which, because of the essentially mechanistic nature of the virtual power principle, cannot be deduced from the latter. $\mathbf{f}^{\text{em}}$ and $\mathbf{T}^{\text{em}}$ as given by Eqs. (2.18a) and (2.21) and (2.20) contain unknown fields. $\mathbf{f}$, $\mathbf{T}$ and $\mathcal{T}_\alpha$ are in general data of a problem. It remains to formulate constitutive equations (cf. Part II) for the elements of $\mathcal{J}_{\text{in}}$ on account of thermodynamical constraints (the so-called thermodynamical admissibility and the dissipation inequality). Furthermore, if the medium is a heat conductor, then the energy equation which provides the heat propagation equation must be adjoined to the above-given equations (see Part II).

The following remarks are in order:

(a) The Eqs. (3.19) are the first Euler–Cauchy equations of the motion, and Eqs. (3.20) are the associated boundary conditions. The second Euler–Cauchy equations, which express the local balance of moment of momentum, are simply obtained by taking the skew symmetric part of t_{ij}—the latter is the *Cauchy* stress. That is, on account of Eq. (3.23),

$$t_{[ij]} = \sum_\alpha (\hat{\sigma}_{\alpha[ij]} + \tilde{\sigma}_{\alpha[ij]}). \tag{3.26}$$

The general expression (3.23), which is valid whatever the mechanical and thermodynamical behaviors of the material are (the only restriction is that of first-order-gradient theory, which, as remarked above, is not a strong limitation), shows that, without studying peculiar constitutive equations, the spin–lattice interactions and the exchange and superexchange forces participate in the the Cauchy stress, along with the usual intrinsic stress (which can be shown to contain the same effects, but in a symmetric combination), in a *nonlinear* theory. In particular, this remark holds true even if the fields ${}^L\mathbf{B}_\alpha$ and $\mathcal{B}_{\alpha ij}$ present dissipative parts which contribute otherwise to the spin relaxation (see Part II).

It must also be remarked that, while Eq. (3.26) describes the skew part of the Cauchy stress, it does not contain (apparently) the intrinsic spins $\gamma_\alpha^{-1}\boldsymbol{\mu}_\alpha$. The transformation of Eq. (3.26) so as to exhibit the presence of these spins is given below in Paragraph 3.4. Also, Eq. (3.26) simplifies in the case of *magnetically saturated* magnetic sublattices in a nonlinear elastic ferrimagnet, for which it can be shown (see Part II, Sec 4) that

$$\tilde{\sigma}_{\alpha[ij]} = 0, \quad \forall\, \alpha. \tag{3.27}$$

The fact that the second Euler–Cauchy equations are somehow contained in the definition of t_{ij}—Eq. (3.23)—results from the application of a rotational invariance (objectivity) in writing $\mathcal{P}_{(i)}^*$.

(b) If one defines a *total* stress tensor τ_{ij} by

$$\tau_{ij} \equiv t_{ij} + t_{ij}^{\text{em}}, \tag{3.28}$$

then Eqs. (3.19), (3.20), and (3.26) transform to

$$\tau_{ij,j} + f_i = \rho \dot{U}_i, \tag{3.29}$$

$$\tau_{ij} n_j = T_i, \tag{3.30}$$

$$\tau_{[ij]} = \sum_\alpha (\hat{\sigma}_{\alpha[ij]} + \tilde{\sigma}_{\alpha[ij]}) - C_{ij}^{\text{em}}, \tag{3.31}$$

on account of Eqs. (2.19) and (2.22). These condensed equations however do not present any advantage in problem solving, although they place in evidence the external contributions $\mathbf{f}$ and $\mathbf{T}$.

(c) Equations (3.21) for $\alpha = 1, 2, \ldots, n$ are the precession equations for the different magnetic sublattices at temperatures much below the critical temperature T_{cr}. They assume the same form as in ferromagnetism for a single magnetic lattice, and thus generalize the usual Larmor precession equation by replacing the simple action of induction $\mathbf{B}$ by a linear combination of $\mathbf{B}$ and of the fields representing the spin–lattice interactions and the intra- and intersublattice interactions.

Here also, ${}^L\mathbf{B}_\alpha$ and $\mathcal{B}_{\alpha ij}$ are general constitutive dependent variables which may present both thermodynamically recoverable and dissipative parts. Equations (3.22) are the associated boundary conditions which, for *zero* $\mathcal{T}_\alpha$—no surface magnetic dipoles—take the obvious form:

$$\mathcal{B}_\alpha \cdot \mathbf{n} + \lambda \boldsymbol{\mu}_\alpha = \mathbf{0} \quad \text{on } \partial\mathcal{D}_t, \tag{3.32}$$

where λ is an unknown such that $\lambda = -\mu_\alpha^{-2}(\boldsymbol{\mu}_\alpha \cdot \mathcal{B}_\alpha \cdot \mathbf{n})$. This allows one to account for the different types of boundary conditions imposed on $\mathcal{B}_\alpha$ or $\boldsymbol{\mu}_\alpha$ depending on whether λ equals zero, infinity or an intermediary value. In fact, Eq. (3.32) is the exact boundary condition for the nonlinear theory ($\mathcal{B}_\alpha$ may be nonlinear), which generalizes the condition derived in a painstaking way in classical treatises.[28]

3.3. Global balance laws in deformable ferrimagnets

A. Balance of momentum

Consider Eq. (3.18) and a virtual rigidifying velocity field that belongs to $\mathcal{C}(\bar{\mathcal{D}}_t)$—Eq. (2.10)—such that

$$U_i^*(\mathbf{x},t) = \bar{U}_i(t), \quad \bar{U}_{i,j} = 0 \tag{3.33}$$

throughout $\mathcal{D}_t$, with

$$\dot{\boldsymbol{\mu}}_\alpha^* = \mathbf{0}, \ \forall\, \alpha. \tag{3.34}$$

Then Eq. (3.10) is satisfied. On account of Eqs. $(3.33)_2$ and (2.19) through (2.20) we obtain

$$\bar{\mathbf{U}} \cdot \left[\int_{\mathcal{D}_t} (\mathbf{f} - \rho\dot{\mathbf{U}})\, dv + \int_{\partial\mathcal{D}_t} \mathbf{T}\, da\right] = 0. \tag{3.35}$$

This is valid for any $\bar{\mathbf{U}}$. Thus, on account of $\dot{\overline{\rho\, dv}} = 0$,

$$\frac{d}{dt}\int_{\mathcal{D}_t} \rho \mathbf{U}\, dv = \int_{\mathcal{D}_t} \mathbf{f}\, dv + \int_{\partial\mathcal{D}_t} \mathbf{T}\, da. \tag{3.36}$$

This is the global balance of linear momentum. No electromagnetic fields appear in this equation, for the electromagnetic momentum is identically zero in quasimagnetostatics and the ponderomotive force is introduced only via the total stress tensor. The latter is introduced in the usual manner if Eq. (3.36) is considered as a first principle. Applying the usual tetrahedron argument, Eq. (3.36) yields Eq. (3.30) on account of Cauchy's principle for stresses: $\mathbf{T} = \mathbf{T}(\mathbf{n},\mathbf{x})$, $\mathbf{x} \in \partial\mathcal{D}_t$. t_{ij}^{em} is hidden in τ_{ij} as shown by Eq. (3.28).

B. Balance of moment of momentum

Consider Eq. (3.18) and a virtual velocity field that belongs to $\mathcal{C}(\bar{\mathcal{D}}_t)$ such that (in *rectangular* coordinates)

$$U_i^*(\mathbf{x},t) = \bar{\Omega}_{ij}(t) x_j, \quad \bar{\Omega}_{ij,k} = 0, \quad \mathbf{x} \in \mathcal{D}_t, \tag{3.37}$$

all the $\boldsymbol{\mu}_\alpha$ being *frozen* in the deformable matter thus rigidified. Then Eq. (3.10) is satisfied and Eq. (3.18) takes the form

$$\begin{aligned}\bar{\Omega}_{ij}\Big[&\int_{\mathcal{D}_t} (f_{[i}x_{j]} - t_{[ij]}^{\text{em}} + B_{[i}M_{j]})\, dv \\ &- \int_{\partial\mathcal{D}_t} (T_{[i}x_{j]} + m_{ij})\, da \\ &+ \int_{\mathcal{D}_t} \rho(\dot{U}_{[i}x_{j]} + \dot{S}_{ij})\, dv\Big] = 0,\end{aligned} \tag{3.38}$$

in which we have defined the following quantities:

$$m_{ij} \equiv \rho \sum_\alpha \mathcal{T}_{\alpha[i}\mu_{\alpha j]}, \tag{3.39}$$

$$S_{ij} \equiv -\tfrac{1}{2}\epsilon_{ijk}s_k = -S_{ji}, \tag{3.40}$$

where $\mathbf{s}$ is the total spin density vector defined by Eq. (2.27). Clearly, m_{ij} is a surface couple density associated with the contact exchange force arising from the magnetic sublattices. It vanishes if Eqs. (3.32) are satisfied.

Remarking that $t_{ij}^{\text{em}} = B_{[i}M_{j]}$ after Eqs. (2.18b) through (2.21), and $\bar{\Omega}_{ij}$ being arbitrary, Eq. (3.38) yields the global balance of moment of momentum for the whole body $\mathcal{B}$ in the form

$$\frac{d}{dt}\int_{\mathcal{D}_t} (U_{[i}x_{j]} + S_{ij})\, dv = \int_{\mathcal{D}_t} f_{[i}x_{j]}\, dv + \int_{\partial\mathcal{D}_t} (T_{[i}x_{j]} + m_{ij})\, da. \tag{3.41}$$

This is the usual form except for the contributions S_{ij} and m_{ij}, which result from ferrimagnetic effects. Again, if Eq. (3.41) is postulated as a first principle, then by applying to it the tetrahedron argument and assuming Cauchy's principle for m_{ij} [$m_{ij} = m_{ij}(\mathbf{n},\mathbf{x})$, $\mathbf{x} \in \partial\mathcal{D}_t$], it is shown that there exists a third order tensor (skew symmetric in its first two indices), M_{ijk}, such such that at $\partial\mathcal{D}_t$,

$$M_{ijk}n_k = m_{ij}. \tag{3.42}$$

M_{ijk} represents the *couple stresses* arising from exchange and superexchange torques. Its relation with the $\mathcal{B}_{\alpha ij}$ is established below in Sec. 3.4. Then the local form corresponding to Eq. (3.41) is easily found to be

$$\rho\dot{S}_{ij} = M_{ijk,k} + \tau_{[ij]} \tag{3.43}$$

on account of Eq. (3.42) and of the local form of Eq. (3.36)—i.e., Eq. (3.29). Assuming that τ_{ij} has the expression (3.28) and $t_{[ij]}^{\text{em}}$ being given by Eq. $(2.19)^2$, Eq. (3.43) reads

$$\rho\mathring{S}_{ij} = M_{ijk,k} + t_{[ij]} - C_{ij}^{\text{em}}. \tag{3.44}$$

This is the canonical form of an equation of balance of moment of momentum.

Remark: the coupled applied, $-C_{ij}^{\text{em}}$, is not the ponderomotive couple, but minus this couple. The reason is that Eq. (3.44) in fact pertains to the total magnetic lattice and not to the crystal lattice (on which it is C_{ij}^{em} that is applied). We thus witness the fact that ponderomotive couples applied to the crystal lattice are entirely transmitted to the spin lattices, the minus sign of Eq. (3.44) resulting from the law of action and reaction. This property is also used in the Appendix.

C. Global balance laws governing the magnetic sublattices

Consider Eq. (3.18) and a special virtual velocity field (3.3) such that

$$U_i^*(\mathbf{x},t) = \bar{U}_i(t), \quad \bar{U}_{i,j} = 0 \ \text{ throughout } \mathcal{D}_t, \tag{3.45}$$

$$\mathring{\boldsymbol{\mu}}_\alpha^* = \bar{\boldsymbol{\omega}}_\alpha^* \times \boldsymbol{\mu}_\alpha, \ \nabla\bar{\boldsymbol{\omega}}_\alpha^* = \mathbf{0} \ \text{ throughout } \mathcal{D}_t, \tag{3.46}$$

where the uniform $\bar{\omega}_\alpha^*$ are otherwise arbitrary. That is, $\bar{\mathbf{U}}$ generates a translational rigid-body motion but the μ_α are *not* frozen in the rigidified matter. Thanks to this the fields (3.45)–(3.46) generate, first and anew, the Eq. (3.36), and next, a global balance law governing each of the magnetic sublattices. On account of Eqs. (3.35) and (3.46), the principle (3.18) yields, for each α,

$$\bar{\omega}_\alpha^*\Big[\int_{D_t}(\rho\gamma_\alpha^{-1}\dot{\boldsymbol{\mu}}_\alpha+\rho\mathbf{B}_\alpha^{\text{eff}}\times\boldsymbol{\mu}_\alpha)\,dv + \int_{\partial D_t}(\mathcal{B}_\alpha\cdot\mathbf{n}-\rho\mathcal{T}_\alpha)\times\boldsymbol{\mu}_\alpha\,da\Big]=0. \tag{3.47}$$

Here $\mathbf{B}_\alpha^{\text{eff}}$ is defined as in Eq. (3.25). Since the $\bar{\omega}_\alpha^*$ are arbitrary and using the fact that $\mathbf{M}_\alpha=\rho\boldsymbol{\mu}_\alpha$ and $\overline{\rho\,dv}^{\boldsymbol{\cdot}}=0$, we obtain the global laws ($\alpha=1,2,\ldots,n$)

$$\frac{d}{dt}\int_{D_t}\gamma_\alpha^{-1}\mathbf{M}_\alpha\,dv = \int_{D_t}(\mathbf{M}_\alpha\times\mathbf{B}_\alpha^{\text{eff}})\,dv+\int_{\partial D_t}\boldsymbol{\mu}_\alpha\times(\mathcal{B}_\alpha\cdot\mathbf{n}-\rho\mathcal{T}_\alpha)\,da. \tag{3.48}$$

This can be further transformed. Indeed, if the boundary conditions (3.22) are satisfied, then the last contribution in Eq. (3.48) vanishes. Furthermore, consider the case of magnetically saturated magnetic sublattices (i.e., at very low temperatures). Then the condition (3.27) is fulfilled and Eq. (3.48) reduces to

$$\frac{d}{dt}\int_{D_t}\gamma_\alpha^{-1}\mathbf{M}_\alpha\,dv=\int_{D_t}\mathbf{M}_\alpha\times(\mathbf{B}+{}^L\mathbf{B}_\alpha)\,dv+\int_{\partial D_t}\mathbf{M}_\alpha\times\mathcal{T}_\alpha\,da \tag{3.49}$$

on account of Eq. (3.25). This equation may be considered as the global balance law that governs, for the whole spatial region D_t at time t, the continuum reppresented by each magnetic sublattice in the case where the latter is saturated. It is an *angular momentum* equation since $\gamma_\alpha^{-1}\mathbf{M}_\alpha$ is an angular momentum per unit deformed volume. As such, this continuum responds only to torques. These torques are of two types according to the form of the right-hand side of the equation. First, there are torques per unit volume due to the action of the Maxwellian magnetic induction $\mathbf{B}$ and to the interactions between the α magnetic sublattice and the crystal lattice and between the α magnetic sublattice and and the neighboring spins of the different ionic species, $\beta\neq\alpha$ (the latter interactions are those which do not result from the disuniformities in the spin repartitions). Second, there are surface torques, via $\mathcal{T}_\alpha$, which represent phenomenologically short-range actions and result from the spatial disuniformities in the magnetic sublattices. Note that if the postulate of global balance laws is considered as the starting point of the theory, then the postulate of Eq. (3.49) clearly requires the consideration of an ad hoc elementary model of interactions (see the Appendix).

3.4. An alternate formulation

We show that Eqs. (3.44) and (3.26) are compatible. On account of Eqs. (3.24), Eq. (3.26) can be written in the form

$$t_{[ij]}=\sum_\alpha(\rho^L B_{\alpha[i}\mu_{\alpha j]}-M_{\alpha ijk,k}+\mathcal{B}_{\alpha[i|k,k|}\mu_{\alpha j]}), \tag{3.50}$$

in which we have defined the effective couple stress tensors due to the magnetic sublattices by

$$M_{\alpha ijk}\equiv\mathcal{B}_{\alpha[i|k|}\mu_{\alpha j]}=-M_{\alpha jik}. \tag{3.51}$$

On account of Eq. (3.40)—written for each sublattice—and of the algebra of ϵ_{ijk}, Eqs. (3.21)[1] can be rewritten as

$$\dot{S}_{\alpha ij}=B^{\text{eff}}_{\alpha[i}\mu_{\alpha j]}. \tag{3.52}$$

That is, with Eq. (3.25),

$$\rho\dot{S}_{\alpha ij}=B_{[i}M_{\alpha j]}+\rho^L B_{\alpha[i}\mu_{\alpha j]}+\mathcal{B}_{\alpha[i|k,k|}\mu_{\alpha j]}. \tag{3.53}$$

Summing over Eqs. (3.51) and (3.52) and combining the results with Eq. (3.50), we are led to the Eqs. (3.44) and (3.42) in which

$$S_{ij}\equiv\sum_\alpha S_{\alpha ij},\quad M_{ijk}\equiv\sum_\alpha M_{\alpha ijk}. \tag{3.54}$$

It is readily shown that Eq. (3.42) is none other than the summation over α of the boundary conditions (3.22) with m_{ij} defined as in (3.39). Of course, we have lost much information in summing over α equations (3.21) and (3.22), so that the latter are still needed to describe the phenomena in a complete fashion. Hence, if a statement of global balance laws is considered as a starting point to approach the present theory, the global balance laws governing the magnetic sublattices must be postulated independently (cf. Eqs. 3.48 and 3.49).

3.5. The principle of virtual power for a real velocity field

Consider now the case for which the virtual fields $\mathbf{U}^*$ and $\boldsymbol{\omega}_\alpha^*$ are none other than the *real* fields, solutions for real problems of the field equations deduced in Sec. 3.2. Then the virtual power of inertia forces (3.16), for real fields (no asterisk), reduces to

$$\mathcal{P}_{(a)}(D_t,v)=\dot{\mathrm{K}}(D_t), \tag{3.55}$$

where $\mathrm{K}(D_t)$ is the total kinetic energy for the moving deformable body at time t, defined as usual by

$$\mathrm{K}(D_t)=\tfrac{1}{2}\int_{D_t}\rho\mathbf{U}^2\,dv. \tag{3.56}$$

The fact that magnetic spins do not produce any power in a real precessional velocity field—cf. Eq. (2.32)—has been accounted for in writing Eq. (3.55). Hence the principle (3.18) reduces to the expression

$$\dot{\mathrm{K}}(D_t)=\mathcal{P}_{(i)}(D_t,\mathcal{V}_{\text{obj}})+\mathcal{P}_{(d)}(D_t,\mathcal{V})+\mathcal{P}_{(c)}(\partial D_t,v), \tag{3.57}$$

where all sets v, $\mathcal{V}$ and $\mathcal{V}_{\text{obj}}$ correspond to *real* velocity fields. When combined with the global statement of the first principle of thermodynamics and with the identity (2.23), Eq. (3.57) yields the so-called *energy theorem* for the whole body (see Part II).

All the above-derived equations can easily be specialized to the case of deformable antiferromagnets by limiting to two the number of simultaneously present magnetic sublattices and assuming that, in absence of an externally applied field below Néel's temperature, the two remaining sublattices $\mu_{(A)}$ and $\mu_{(B)}$ compensate to yield a zero magnetization.

4. CONCLUSION

Clearly, the above-derived equations for deformable ferrimagnets and antiferromagnets are of the prime importance in studying various problems in these media, especially, coupled magnetoelastic waves and the magnetostrictive and piezomagnetic effects. However, to to fulfill that purpose, they need be closed by constructing ad hoc constitutive equations for both thermodynamically recoverable and dissipative phenomena, which will be done in Part II. In the procedure we have to formulate the relevant thermodynamics in which the Eqs. (3.57) and (3.8) will prove to be key points.

The above study which, along with the contents of the Appendix, is quite exhaustive as regards the field equations, has also shown, especially in Sec. 3.4, how close to the recently formulated purely mechanical theories of continua with microstructure (e.g., micropolar media,[29] media with couple stresses[30]) the theory of *elastic magnetically ordered materials* proves to be. In fact, the latter contains all mechanisms present in such theories, e.g., intrinsic spins, couple stresses, surface and volume couples, nonsymmetric Cauchy stress. Nonetheless, an important remark must be made in this regard. Whereas the magnetic sublattices clearly respond to surface couples, the material continuum (i.e., the *material lattice* in the language of the Appendix) does not possess the necessary mechanism to respond to these couples. In this respect Eq. (A5) below is typical. No *mechanical* couple stress tensor appears in this equation (or in Eq. (3.26). As discussed in another paper[6] for the particular case of ferromagnetism, this brings some constraint on the type of boundary conditions regarding magnetic spins which can be accepted. In fact, only that given by Eq. (3.32) is allowed. In order to enlarge the choice with regard to such boundary conditions it would be necessary, as was done in elastic ferromagnets[6] to consider a finer description for the deformation processes, for instance, a second-order-gradient theory so as to make clear the surface magnetoelastic couplings arising from torques. Although such an involved scheme still is manageable in the ferromagnetic case, it may be reasonably conjectured that it would be rather complex in the present case, so that we note this possibility only for memory.

APPENDIX: A MODEL OF THREE INTERACTING CONTINUA FOR DEFORMABLE ANTIFERROMAGNETS

(1) We generalize Tiersten's model[31] for ferromagnets to the case of deformable antiferromagnets. The case of ferrimagnets could be treated along the same lines, but the antiferromagnetic case exemplifies the method in a clearer fashion. This is a model of interactions which, it must be noted, is not necessarily issued directly from microscopic physics, and there is no necessary one-to-one correspondence between the usual microscopic concepts and the phenomenological entities introduced. Basically, it is a model of three *interacting simultaneously present* continua, referred to as *lattices*. One we call the *material* (or crystal) lattice (for short ML), which is the usual material continuum of elasticity theory and thus is the substrate of (nonlinear or linear) elastic deformations and *phonons* (i.e., elastic waves). The other two continua are none other than the *magnetic sublattices* A and B represented in a continuous fashion by the mass magnetization fields $\boldsymbol{\mu}_{(A)}$ and $\boldsymbol{\mu}_{(B)}$, that depend on $\mathbf{x}$ and t in the deformed configuration $\mathcal{K}$ of the deformable body $\mathcal{B}$. The latter two continua support mainly typical antiferromagnetic effects (superexchange forces) and form the substrates of *magnons*, i.e., spin waves. The crystal lattice and the magnetic sublattices interact because magnetization affects the deformations and, reciprocally, the deformations have an influence on the distribution of magnetization in the antiferromagnetic body (the simplest effects being magnetostriction and piezomagnetism). We examine the case of *magnetically saturated* antiferromagnetic elastic insulators within the framework of quasimagnetostatics.

The material lattice (ML) is governed by the global balance laws of mass, momentum, and moment of momentum written in the usual manner:

$$\frac{d}{dt}\int_{D_t}\rho\,dv=0, \tag{A1}$$

$$\frac{d}{dt}\int_{D_t}\rho\mathbf{U}\,dv=\int_{D_t}\tilde{\mathbf{f}}\,dv+\int_{\partial D_t}\tilde{\mathbf{T}}\,da, \tag{A2}$$

$$\frac{d}{dt}\int_{D_t}(\mathbf{x}\times\rho\mathbf{U})\,dv=\int_{D_t}(\mathbf{x}\times\tilde{\mathbf{f}}+\mathbf{c})\,dv+\int_{\partial D_t}(\mathbf{x}\times\tilde{\mathbf{T}})\,da, \tag{A3}$$

where

$$\tilde{\mathbf{f}}=\mathbf{f}+\mathbf{f}^{\mathrm{em}},\quad \tilde{\mathbf{T}}=\mathbf{T}+\mathbf{T}^{\mathrm{em}}, \tag{A4}$$

and $\mathbf{c}$ is the volume couple resulting from the interactions between the magnetic sublattices A and B and ML. The ponderomotive couple is included in the other terms.

From Eqs. (A1), (A2), and (A4) the local field equations (2.6), (3.19), and (3.20) are deduced in the usual manner. ML possesses only orbital angular momentum, that is, no intrinsic spin, so that Eq. (A3) yields the local form

$$\epsilon_{ijk}t_{jk}+c_i=0 \tag{A5}$$

on account of Eqs. (3.19) and (3.20).

(2) Associated with the continua $\boldsymbol{\mu}_{(A)}$ and $\boldsymbol{\mu}_{(B)}$ are the *magnetic spin continua* A and B, with densities $\mathbf{s}_{(A)}=\gamma_A^{-1}\boldsymbol{\mu}_{(A)}$ and $\mathbf{s}_{(B)}=\gamma_B^{-1}\boldsymbol{\mu}_{(B)}$ per unit mass. These continua possess only angular momentum by definition. Since they possess no linear momentum, none of their points can translate with respect to the corresponding points in the material lattice. Therefore, it is clear that that the spin continua expand and contract with the material lattice and must occupy at all times the same volume as the material lattice, so that their volumetric behavior is governed also by Eq. (2.6). Similarly, the conservation of linear momentum simply says that whatever force of magnetic origin is applied to a point of the spin continua A and B, it is transferred directly to the material lattice at that point. However, after their definition, $\mathbf{s}_{(A)}$ and $\mathbf{s}_{(B)}$ respond only to torques. Then it is assumed that each of the magnetic sublattices interacts with the local material lattice by means of a *local magnetic field* (referred to as the magnetic anisot-

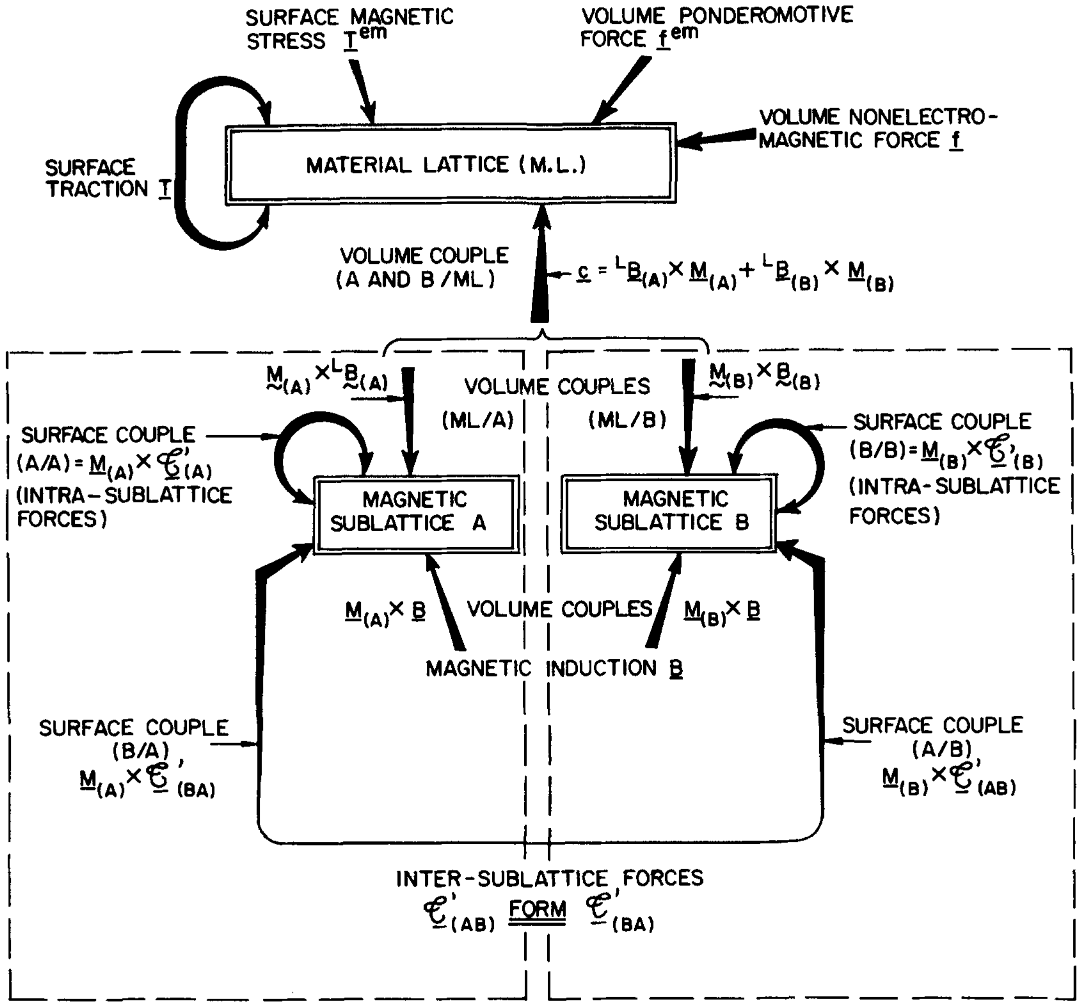

FIG. 1. Interactions in a deformable antiferromagnet.

ropy field in the body of the text), ${}^{L}\mathbf{B}_{(A)}$ or ${}^{L}\mathbf{B}_{(B)}$, which exerts a *couple* per unit volume on its respective magnetization field, $\mathbf{M}_{(A)}$ or $\mathbf{M}_{(B)}$, by means of the "recipe": $\mathbf{M}_{(A)} \times {}^{L}\mathbf{B}_{(A)}$ or $\mathbf{M}_{(B)} \times {}^{L}\mathbf{B}_{(B)}$. These are torques exerted by the local material lattice on the spin continua. Since angular momentum is conserved, equal and opposite torques, ${}^{L}\mathbf{B}_{(A)} \times \mathbf{M}_{(A)}$ and ${}^{L}\mathbf{B}_{(B)} \times \mathbf{M}_{(B)}$, must be exerted by the spin continua on the local material lattice. Then the couple $\mathbf{c}$ appearing in Eq. (A5) is given by

$$\mathbf{c} = {}^{L}\mathbf{B}_{(A)} \times \mathbf{M}_{(A)} + {}^{L}\mathbf{B}_{(B)} \times \mathbf{M}_{(B)}. \tag{A6}$$

It follows from Eqs. (A5) and (A6) that

$$t_{[ij]} = {}^{L}B_{(A)[i}M_{(A)j]} + {}^{L}B_{(B)[i}M_{(B)j]}. \tag{A7}$$

In addition to the couple caused by the material lattice, whose recipe has been given above, each magnetic spin continuum experiences couples due to the ordinary Maxwellian induction, i.e.,

$$\mathbf{c}^{em}_{(A)} = \mathbf{M}_{(A)} \times \mathbf{B}, \quad \mathbf{c}^{em}_{(B)} = \mathbf{M}_{(B)} \times \mathbf{B}. \tag{A8}$$

Furthermore, each individual magnetic spin of each magnetic sublattice experiences from its nearest neighbors within the same sublattice (*intra*sublattice forces) and from neighboring spins that belong to the second magnetic sublattice (*inter*sublattice forces), an action caused by the *exchange* and *superexchange* forces. Given the rapid fall over distance of this type of interactions, we assume that they give rise in a phenomenological manner to *contact*, i.e., surface, actions. In order to account for the forces exerted within each sublattice, we consider a *surface exchange contact force*, *force*, $\mathcal{T}'_{(A)}$ and $\mathcal{T}'_{(B)}$ respectively, which, since the spin continua respond only to couples, produces a couple per unit area equal to $\mathbf{M}_{(A)} \times \mathcal{T}'_{(A)}$ or $\mathbf{M}_{(B)} \times \mathcal{T}'_{(B)}$ depending on the sublattice considered. $\mathcal{T}'_{(A)}$ and $\mathcal{T}'_{(B)}$ have the dimension of a surface distribution of magnetic dipoles. Similarly, in order to account for the superexchange forces produced through intervening ions, i.e., the intersublattice forces, we consider *surface superexchange contact forces*, $\mathcal{T}'_{(BA)}$ and $\mathcal{T}'_{(AB)}$, which produce couples per unit area $\mathbf{M}_{(A)} \times \mathcal{T}'_{(BA)}$ and $\mathbf{M}_{(B)} \times \mathcal{T}'_{(AB)}$ on the A and B sublattices respectively. Since the role of the A and B sublattices can be interchanged, we necessarily have

$$\mathcal{T}'_{(AB)} \overset{\text{form}}{=\!=} \mathcal{T}'_{(BA)}, \tag{A9}$$

where the symbolism $\overset{\text{form}}{=\!=}$ means that both expressions which it relates must be *formally* identical in the interchange of A and B. All interactions thus far introduced are sketched out in Fig. 1.

Analogous to Cauchy's principle for surface tractions, we assume that the surface fields $\mathcal{T}'_{(A)}$, $\mathcal{T}'_{(B)}$, $\mathcal{T}'_{(AB)}$, and $\mathcal{T}'_{(BA)}$ depend on the local normal at the bounding surface and on no geometrical properties of higher order (e.g., the local curvature). Then, on account of the above discussion the equations of balance of angular momentum for the two magnetic spin continua are written in global form as

$$\frac{d}{dt}\int_{\mathcal{D}_t}\rho\gamma_A^{-1}\mu_{(A)}\,dv = \int_{\mathcal{D}_t}\mathbf{M}_{(A)}\times(\mathbf{B}+{}^L B_{(A)})\,dv + \int_{\partial\mathcal{D}_t}\mathbf{M}_{(A)}\times\mathcal{T}_{(A)}\,da \quad \text{(A10)}$$

where

$$\mathcal{T}_{(A)}(\mathbf{n}) = \mathcal{T}'_{(A)}(\mathbf{n}) + \mathcal{T}'_{(AB)}(n). \quad \text{(A11)}$$

Applying the tetrahedron argument to Eq. (A10), we obtain for any $\mathbf{M}_{(A)}$ the linear relationship

$$\mathcal{T}_{(A)i}(\mathbf{n}) = \rho^{-1}\mathcal{B}_{(A)ij}n_j \quad \text{(A12)}$$

on $\partial\mathcal{D}_t$, where $\mathcal{B}_{(A)ij}$ is a linear operator which may obviously be referred to as the *spin-interaction* tensor for the A sublattice. Analogous equations hold true for the B sublattice. According to Eq. (A11), $\mathcal{B}_{(A)ij}$ here represents both the intrasublattice and intersublattice forces acting upon the A sublattice, each contribution being placed in evidence only once constitutive equations are specified (see Part II). It can however be remarked that, similar to Eq. (A9), we must have

$$\mathcal{B}_{(A)ij} \overset{\text{form}}{=\!=} \mathcal{B}_{(B)ij}. \quad \text{(A13)}$$

The local form of Eq. (A10) is easily deduced with the help of Eq. (A12). One obtains

$$\gamma_A^{-1}\dot{\mu}_{(A)i} = \epsilon_{ijk}\mu_{(A)j}(B_k + {}^L B_{(A)k} + \rho^{-1}\mathcal{B}_{(A)km,m}) + \rho^{-1}\epsilon_{ijk}\mu_{(A)j,m}\mathcal{B}_{(A)km}, \quad \text{(A14)}$$

a similar equation describing the spin precession of the B sublattice. The agreement with the equations derived in the main body of the text is obtained if one uses the results (3.27)—derived in Part II—corresponding to *magnetically saturated* sublattices. Then the last term in the right-hand side of Eq. (A14) vanishes and Eqs. (A14), (A10), and (A7) take the same form as Eqs. (3.21) [on account of Eq. (3.25)], (3.49), and (3.26) [on account of Eq. (3.27)].

[1]If there exist fluids or suspensions which present the same magnetic properties as the solids considered in this paper, then they satisfy the same balance equations, including the spin-precession equations and the fact that the stress tensor in general is not symmetric.

[2]J.C. Baumhauer and H.F. Tiersten, J. Acoust. Soc. Am. 54, 1017 (1973).

[3]G.A. Maugin, Lett. Appl. Eng. Sci. 2, 293 (1974).

[4]G.A. Maugin, J. Mec. 13, 73 (1974).

[5]B. Collet and G.A. Maugin, C.R. Acad. Sci. B (Paris) 279, 379 (1974).

[6]B. Collet and G.A. Maugin, C.R. Acad. Sci. A (Paris) 280, 1641 (1975).

[7]P. Penfield and H.A. Haus, *Electrodynamics of Moving Media* (M.I.T. Press, Cambridge, Massachusetts, 1968). These authors do not consider any invariance in applying their formulation of the principle of virtual power.

[9]G.A. Maugin, Physica **81**, 454 (1975). Paris, 1973), Vol. I, p. 26.

[9]G.A. Maugin, Physica A **81**, 454 (1975).

[10]S.V. Peletminskii, Zh. Eksp. Teor. Fiz. 37, 452 (1959) [Sov. Phys. JETP 37, 321 (1960)].

[11]V.G. Bar'yakhtar, M.A. Savchenko, and V.V. Tarasenko, Zh. Eksp. Teor. Fiz. 49, 944 (1965) [Sov. Phys. JETP 22, 657 (1966)].

[12]A.I. Akhiezer, V.G. Bar'yakhtar, and S.V. Peletminskii, *Spin Waves* (English translation) (North-Holland, Amsterdam, 1968).

[13]L. Néel, Ann. Phys. (Paris) 3, 137 (1948).

[14]G.A. Maugin and B. Collet, C.R. Acad. Sci. B(Paris) 279, 439 (1974).

[15]C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1971), third ed., Chap. 16.

[16]S. Chikazumi, *Physics of Magnetism* (Wiley, New York, 1966), Chap. 5.

[17]J. Smit and H.P. Wijn, *Ferrites* (Wiley, New York, 1959).

[18]W.P. Wolf, Rep. Prog. Phys. 24, 218 (1961).

[19]T. Nagamiya, K. Yosida, and R. Kubo, Adv. Phys. 4, 1 (1955).

[20]E.A. Turov, *Physical Properties of Magnetically Ordered Crystals* (English translation) (Academic, New York, 1965).

[21]A.C. Eringen, *Mechanics of Continua* (Wiley, New York, 1967).

[22]See Ref. 8, p. 113.

[23]See Ref. 14, and G.A. Maugin and A.C. Eringen, J. Math. Phys. 13, 143 (1972).

[24]S. Foner, in *Magnetism*, edited by G.T. Rado and H. Suhl (Academic, New York, 1963), Vol. I.

[25]C. Truesdell and W. Noll, in *Handbuch der Physik*, Bd. III/3, edited by S. Flügge (Springer, Berlin, 1965); pp. 60–63.

[26]L.D. Landau and E.M. Lifshitz, Phys. Z. Sowjetunion 8, 153 (1935).

[27]W.F. Brown, *Magnetoelastic Interactions* (Springer, New York, 1966).

[28]A.G. Gurevich, *Magnetic Resonance in Ferrites and Antiferromagnets* (in Russian) (Nauka, Moscow, 1973), p. 140.

[29]A.C. Eringen, in *Fracture: An Advanced Treatise*, edited by H. Liebowitz (Academic, New York, 1968), Vol. II, pp. 621–729.

[30]R.A. Toupin, Arch. Ration. Mech. Anal. 17, 85 (1964).

[31]H.F. Tiersten, J. Math. Phys. 5, 1298 (1964); see also Ref. 23 and the synthesis by the author: "Micromagnetism" in *Continuum Physics*, edited by A.C. Eringen (Academic, New York, to appear early 1976), Vol. III, Part III, pp. 221–312.

A continuum theory of deformable ferrimagnetic bodies. II. Thermodynamics, constitutive theory

Gérard A. Maugin

Université de Paris VI, Laboratoire de Mécanique Théorique associé au C.N.R.S., Tour 66, 4 Place Jussieu, 75230 Paris, Cedex 05, France
(Received 15 July 1975)

In order to close the system of differential field equations developed in Paper I, this article proposes a rational development of the relevant macroscopic thermodynamics and of a constitutive theory. In particular, by following Coleman's thermodynamics, exact nonlinear constitutive equations for thermoelastic antiferromagnetic insulators are formulated. According to the deductive scheme adopted in Paper I, the important case of elastically isotropic antiferromagnets with a magnetic easy axis, and possibly endowed with the property of weak ferromagnetism, is developed in detail by using approximations. In order to supplement the description of thermodynamically recoverable processes and in accordance with the Onsager–Casimir theory of irreversible processes, the constitutive equations governing phenomena such as viscosity, electric and heat conduction, and spin relaxation, the latter either for strong or weak damping, are obtained. Regarding the latter effect, it is shown, thanks to the formalism adopted in Paper I, that both viscosity and spin relaxation participate in the Cauchy equations. The relaxation term of Gilbert is thus generalized to the case of deformable antiferromagnets.

1. INTRODUCTION[1]

General local or global balance laws (independent of the peculiar mechanical and thermodynamical behavior) that govern the motion and the interacting fields in a deformable ferrimagnet or antiferromagnet have been deduced in Part I from a single principle, that of virtual power applied simultaneously with the requirement of objectivity as far as "internal forces"— the fields that represent in a phenomenological manner the different interactions—are concerned. The purpose of this second part is to develop the relevant macroscopic thermodynamics, which allows the construction of constitutive equations for these "internal forces" for both thermodynamically recoverable processes—in accordance with Coleman's thermodynamics[2]—and irreversible processes—along the lines of the Onsager—Casimir theory of irreversible processes.[3] The constitutive equations thus obtained permit one to close the system of differential field equations built in Part I.

Having recalled the main results of Part I in Sec. 2 for the special case of deformable *simple antiferromagnets*, thus with a magnetic structure made of only two magnetic sublattices, we postulate the global form of the first and second principles of thermodynamics according to the scheme of contemporary continuum mechanics. Combined with the expression of the principle of virtual power written for a *real* velocity field, these principles yield the so-called theorem of the energy and the Clausius–Duhem inequality, which proves essential in the subsequent development (Sec. 3). Note that the formalism developed in fact is also valid for a multisublattice structure with more than two sublattices, hence it is valid for the description of deformable *ferrites*. In Sec. 4, following Coleman's thermodynamics, *exact* constitutive equations are obtained for nonlinear thermoelastic antiferromagnetic insulators, especially when, at temperatures much below Néel's temperature, each magnetic sublattice may be considered as saturated. The equations thus obtained, although exact, are in general quite unmanageable because all effects are intricately mixed as a result of the nonlinearity, so that approximations such as those of an expansion of the free energy in its different arguments, and the case of infinitesimal deformations are given much attention in Sec. 5. The different effects such as thermoelasticity, pyromagnetism, magnetocrystalline effects, exchange and superexchange forces, piezomagnetism and magnetostriction are thus placed in evidence. For the purpose of illustration the typical case of an elastically isotropic antiferromagnet with a magnetic easy axis and, possibly, the property of weak ferromagnetism, is given in detail. General dissipative processes such as viscosity, spin relaxation, heat and electricity conductions are looked upon in Sec. 6 in accordance with the classical theory of irreversible processes. Special attention is given to spin-relaxation phenomena, which are seldom examined in detail in ferrimagnetism and/or antiferromagnetism. In particular, the spin-relaxation terms to be considered for *strong damping* in deformable antiferromagnetis are proposed, which generalize our earlier proposal[4,5] concerning the case of deformable ferromagnets. By the same token, Gilbert's expression[6] is generalized to such media, and it is shown in a straightforward manner that the spin relaxation participates in the Cauchy equation of motion, thus exhibiting the fact that this dissipative process may cause the damping of both magnon and phonon branches of the dispersion diagram of coupled magnetoelastic waves in antiferromagnets, especially in the crossover regions. Then an elementary perturbation scheme shows that the generalization to deformable antiferromagnets of the Landau and Lifshitz's relaxation term[7] is valid for *weak damping*. It must be emphasized that the rational formulation of such coupled effects for a wide range of damping results directly from the methodology followed in Part I (especially, the duality between spaces of "forces" and "velocities").

2. RECAPITULATION FOR DEFORMABLE ANTIFERROMAGNETS

2.1. Local balance equations

The equations deduced from a generalized version of the virtual power principle in Part I are here specialized to the case of deformable antiferromagnets whose magnetic structure is built up of two magnetic sublattices numbered $\alpha = A, B$.[8] Let $\mathcal{D}_t$ be the spatial region of E^3 occupied instantaneously at time t, in its present configuration $\mathcal{K}$, by a deformable body $\mathcal{B}$. $\partial\mathcal{D}_t$ is the corresponding bounding surface with unit outward normal n. The relevant equations of the phenomenological description for insulators containing no free charges (in quasimagnetostatics) are the following ones:

(a) Equations governing the crystal lattice:

Continuity

$$\dot{\rho} + \rho U_{k,k} = 0 \quad \text{in } \mathcal{D}_t, \tag{2.1}$$

Cauchy's equations

$$t_{ij,j} + f_i + \mathbf{M}\cdot\mathbf{B}_{,i} = \rho\dot{U}_i \quad \text{in } \mathcal{D}_t, \tag{2.2}$$

$$t_{ij}n_j = T_i - [t_{ij}^{\text{em}}]n_j \quad \text{on } \partial\mathcal{D}_t; \tag{2.3}$$

b. Equations governing the magnetic sublattics ($\alpha = A, B$):

$$\dot{\mu}_{(\alpha)} = -\gamma_\alpha \mathbf{B}_{(\alpha)}^{\text{eff}} \times \mu_{(\alpha)} \quad \text{in } \mathcal{D}_t, \tag{2.4}$$

$$\mathcal{B}_{(\alpha)}\cdot\mathbf{n} + \lambda\mu_{(\alpha)} = 0 \qquad \text{on } \partial\mathcal{D}_t; \tag{2.5}$$

(c) Equations governing the magnetic fields (Lorentz—Heaviside units):

$$\nabla^2\Phi - \nabla\cdot\mathbf{M} = 0 \qquad \text{in } \mathcal{D}_t, \tag{2.6}$$

$$[\partial\Phi/\partial n] + \mathbf{M}_{\text{in}}\cdot\mathbf{n} = 0, \quad \text{on } \partial\mathcal{D}_t; \tag{2.7}$$

(d) Definitions ($\alpha = A, B$):

$$\mathbf{H} = \mathbf{B} - \mathbf{M} = -\nabla\Phi, \tag{2.8}$$

$$\mathbf{M} = \rho\mu = \sum_\alpha \mathbf{M}_{(\alpha)} = \rho\sum_\alpha \mu_{(\alpha)}, \tag{2.9}$$

$$t_{ij} = \sigma_{ij} + \sum_\alpha (\rho\, {}^L B_{(\alpha)[i}\mu_{(\alpha)j]} - \mathcal{B}_{(\alpha)[i|k|}\mu_{(\alpha)j],k}), \tag{2.10}$$

$$B_{(\alpha)i}^{\text{eff}} = B_i + {}^L B_{(\alpha)i} + \rho^{-1}\mathcal{B}_{(\alpha)ij,j}, \tag{2.11}$$

$$t_{ij}^{\text{em}} = H_i B_j - (\tfrac{1}{2}\mathbf{B}^2 - \mathbf{M}\cdot\mathbf{B})\delta_{ij}. \tag{2.12}$$

In these equations the different symbols introduced bear the following significance:

ρ: density of matter in $\mathcal{K}$,

f: volume force (no magnetic effects),

U: matter velocity,

t_{ij}: Cauchy (nonsymmetric) stress tensor,

T_i: surface traction of purely mechanical origin,

t_{ij}^{em}: magnetostatic Maxwell stress tensor,

B: magnetic induction,

H: magnetic field,

Φ: magnetostatic scalar potential,

M: total volume magnetization,

μ: total magnetization per unit mass in $\mathcal{K}$,

$\mu_{(\alpha)}$: magnetization of the α-sublattice per unit mass in $\mathcal{K}$,

γ_α: gyromagnetic ratio of the α-sublattice,

$\mathbf{B}_{(\alpha)}^{\text{eff}}$: effective magnetic induction acting on the α-sublattice;

$\sigma_{ji} = \sigma_{ij}$, ${}^L\mathbf{B}_{(\alpha)}$ and $\mathcal{B}_{(\alpha)}$ are respectively the intrinsic stress tensor, the magnetic anisotropy field of the α-sublattice, and the spin-interaction tensor of the α-sublattice. Constitutive equations must be constructed for these five fields ($\alpha = A, B$).

In writing the boundary conditions (2.5), we have assumed a zero surface exchange contact torque for each magnetic sublattice (See Eq. I. 3.32). λ is a Lagrange multiplier which can be said to measure the *surface magnetic anisotropy*.[9] Depending on whether $\lambda = 0$ or $+\infty$, or an intermediary value, the boundary conditions (2.5) contain all those which have been proposed in the relevant literature.

In a finite deformation theory the classical motion, solution of Cauchy's equations (2.2), is given by the general expression (with a sufficient degree of differentiability)

$$x_k = \mathcal{X}_k(X_K, t) \tag{2.13}$$

where x_k, $k = 1, 2, 3$, and X_K, $K = 1, 2, 3$, denote the position respectively in $\mathcal{K}$ and in the reference configuration $\mathcal{K}_0$ defined at $t = t_0$. t is the Newtonian absolute time. A superimposed dot indicates the material time derivative $\partial/\partial t + \mathbf{U}\cdot\nabla$; $\partial/\partial n \equiv \mathbf{n}\cdot\nabla$ is the normal derivative. The symbolism $[\cdots]$ indicates the jump across $\partial\mathcal{D}_t$.

Of course, in absence of externally applied magnetic field and of other perturbations, the sum given by Eq. (2.9) must vanish at all points in $\mathcal{D}_t$ below Néel's temperature, since we consider the antiferromagnetic case:

$$\mu \equiv \sum_\alpha \mu_{(\alpha)} = 0. \tag{2.14}$$

2.2. The principle of virtual power for a real velocity field

The following expression, that represents the statement of the virtual power principle for *real* velocity fields, has been established in Part I:

$$\dot{\mathbb{K}}(\mathcal{D}_t) = \mathcal{P}_{(i)}(\mathcal{D}_t) + \mathcal{P}_{(d)}(\mathcal{D}_t) + \mathcal{P}_{(c)}(\partial\mathcal{D}_t). \tag{2.15}$$

Here the different contributions represent respectively:

the material derivative of the total kinetic energy:

$$\mathbb{K}(\mathcal{D}_t) = \int_{\mathcal{D}_t} \tfrac{1}{2}\rho\mathbf{U}^2\, dv, \tag{2.16}$$

the power of internal forces:

$$\mathcal{P}_{(i)}(\mathcal{D}_t) = -\int_{\mathcal{D}_t}[\sigma_{ij}D_{ij} - \sum_\alpha(\rho\, {}^L B_{(\alpha)i}\hat{m}_{(\alpha)i} - \mathcal{B}_{(\alpha)ij}\hat{\mathfrak{M}}_{(\alpha)ij})]\, dv, \tag{2.17}$$

the power of volume forces:

$$\mathcal{P}_{(d)}(\mathcal{D}_t) = \int_{\mathcal{D}_t}(\mathbf{f}\cdot\mathbf{U} - t_{ij}^{\text{em}}U_{i,j} + \rho\mathbf{B}\cdot\dot{\mu})\, dv, \tag{2.18}$$

the power of contact forces here written for zero surface exchange contact forces as

$$\mathcal{P}_{(c)}(\partial\mathcal{D}_t)=\int_{\partial\mathcal{D}_t}\mathbf{T}\cdot\mathbf{U}\,da. \tag{2.19}$$

The different *objective*[10] time rates appearing in Eq. (2.17) are defined by

$$D_{ij}\equiv\tfrac{1}{2}(U_{i,j}+U_{j,i})\equiv U_{(i,j)}, \tag{2.20}$$

$$\hat{m}_{(\alpha)i}\equiv\dot{\mu}_{(\alpha)i}-\Omega_{ij}\mu_{(\alpha)j}, \tag{2.21}$$

$$\hat{\mathfrak{M}}_{(\alpha)ij}\equiv(\dot{\mu}_{(\alpha)i})_{,j}-\Omega_{ik}\mu_{(\alpha)k,j}, \tag{2.22}$$

where

$$\Omega_{ij}\equiv\tfrac{1}{2}(U_{i,j}-U_{j,i})\equiv U_{[i,j]}. \tag{2.23}$$

Finally, the following energetic identity for quasimagnetostatic fields proves useful in the sequel (see Eq. I. 2.24):

$$\dot{U}^{\mathrm{em.m}}(\mathcal{D}_t)=\int_{\mathcal{D}_t}(t_{ij}^{\mathrm{em}}U_{i,j}-\rho\mathbf{B}\cdot\dot{\boldsymbol{\mu}})\,dv, \tag{2.24}$$

where $U^{\mathrm{em.m}}$ is the total magnetic energy in the magnetized body $\mathcal{B}$ at time t:

$$U^{\mathrm{em.m}}(\mathcal{D}_t)=\int_{\mathcal{D}_t}(\tfrac{1}{2}\mathbf{B}^2-\mathbf{M}\cdot\mathbf{B})\,dv. \tag{2.25}$$

We are now in a position to construct the macroscopical thermodynamics and constitutive theory that allows us to specify the form of the fields σ_{ij}, ${}^L\mathbf{B}_{(\alpha)}$, and $\mathcal{B}_{(\alpha)}$.

3. THERMODYNAMICAL PRINCIPLES

3.1. Global statements

Let e, h, $\mathbf{q}$, η, and θ be respectively the internal energy per unit mass in K, the volume heat source (e.g., radiation), the heat influx vector through $\partial\mathcal{D}_t$, the entropy per unit mass in K, and the thermodynamical temperature such that $\theta>0$, $\inf\theta=0$. The statement of the first principle of thermodynamics for the whole material moving body expresses the fact that the time rate of change of the total energy, i.e., the sum of the kinetic energy, the internal energy, and the magnetic energy within the body, equals the time rate of change of heat production to which is added the rate of work of *prescribed* forces (i.e., only $\mathbf{f}$ and $\mathbf{T}$ participate in the last quantity if zero surface exchange contact forces are assumed). We thus set at time t

$$\dot{\mathbb{K}}(\mathcal{D}_t)+\dot{\mathbb{E}}(\mathcal{D}_t)+\dot{U}^{\mathrm{em.m}}(\mathcal{D}_t)=\mathring{\mathcal{Q}}(\bar{\mathcal{D}}_t)+\mathcal{P}_{(P)}(\bar{\mathcal{D}}_t), \tag{3.1}$$

where

$$\mathbb{E}(\mathcal{D}_t)=\int_{\mathcal{D}_t}\rho e\,dv, \tag{3.2}$$

$$\mathring{\mathcal{Q}}(\bar{\mathcal{D}}_t)=\int_{\mathcal{D}_t}\rho h\,dv-\int_{\partial\mathcal{D}_t}\mathbf{q}\cdot\mathbf{n}\,da, \tag{3.3}$$

$$\mathcal{P}_{(P)}(\bar{\mathcal{D}}_t)=\int_{\mathcal{D}_t}\mathbf{f}\cdot\mathbf{U}\,dv+\int_{\partial\mathcal{D}_t}\mathbf{T}\cdot\mathbf{U}\,da. \tag{3.4}$$

Note that Eq. (3.1) is a postulate independent of the special form of the virtual power principle (2.15).

The second law of thermodynamics is postulated here in global form as

$$\dot{\mathbb{N}}(\mathcal{D}_t)\geq\mathring{\mathrm{S}}(\bar{\mathcal{D}}_t) \tag{3.5}$$

where

$$\mathbb{N}(\mathcal{D}_t)=\int_{\mathcal{D}_t}\rho\eta\,dv, \tag{3.6}$$

$$\mathring{\mathrm{S}}(\bar{\mathcal{D}}_t)=\int_{\mathcal{D}_t}\rho(h/\theta)\,dv-\int_{\partial\mathcal{D}_t}\theta^{-1}\mathbf{q}\cdot\mathbf{n}\,da. \tag{3.7}$$

Note that the entropy influx vector through $\partial\mathcal{D}_t$ takes here the simple form $\theta^{-1}\mathbf{q}$ because we are in quasimagnetostatics (i.e., the Poynting vector is not involved).[11] The local form of the inequality (3.5) obviously is, for sufficiently smooth fields,

$$\rho\theta\dot{\eta}-\rho h+\nabla\cdot\mathbf{q}-\theta^{-1}\mathbf{q}\cdot\nabla\theta\geq 0 \tag{3.8}$$

at all internal points in $\mathcal{D}_t$.

3.2. The theorem of the energy

Upon combining the expressions (2.15) and (3.1) on account of Eqs. (3.4), (2.12), (2.19), and (2.4), we arrive at the so-called *theorem of the energy*, which expresses the time rate of change of the total internal energy in the form

$$\dot{\mathbb{E}}(\mathcal{D}_t)+\mathcal{P}_{(i)}(\mathcal{D}_t)=\mathring{\mathcal{Q}}(\bar{\mathcal{D}}_t). \tag{3.9}$$

This being posited to be valid for arbitrary regions in $\mathcal{D}_t$, it yields the local equation of the energy on account of Eqs. (3.2), (2.17), (3.3), and of the continuity equation and Gauss' theorem:

$$\rho\dot{e}=\sigma_{ij}D_{ij}-\sum_{\alpha}(\rho\,{}^L\mathbf{B}_{(\alpha)}\cdot\hat{\mathbf{m}}_{(\alpha)}-\mathcal{B}_{(\alpha)ij}\hat{\mathfrak{M}}_{(\alpha)ij})-\nabla\cdot\mathbf{q}+\rho h. \tag{3.10}$$

This yields the heat propagation equation when e is specified.

3.3. The Clausius-Duhem inequality

Substituting from Eq. (3.10) into Eq. (3.8), so that h is eliminated, and defining the free (Helmholtz) energy per unit mass in K by

$$\psi\equiv e-\eta\theta, \tag{3.11}$$

we are led to the local statement of the second principle of thermodynamics in the form known as the *Clausius–Duhem inequality*[12]:

$$-\rho(\dot{\psi}+\eta\dot{\theta})+\sigma_{ij}D_{ij}-\sum_{\alpha}(\rho\,{}^L\mathbf{B}_{(\alpha)}\cdot\hat{\mathbf{m}}_{(\alpha)}-\mathcal{B}_{(\alpha)ij}\hat{\mathfrak{M}}_{(\alpha)ij})-\theta^{-1}\mathbf{q}\cdot\nabla\theta\geq 0. \tag{3.12}$$

This inequality is the ground on which is built the following development since, together with the constitutive dependent variables σ_{ij}, ${}^L\mathbf{B}_{(\alpha)}$, and $\mathcal{B}_{(\alpha)}$, it contains the remaining constitutive dependent variables of the theory, namely, ψ, η, and $\mathbf{q}$. All these variables are *objective* as they should in fact be, as well as their respective cofactors. The importance of applying the systematic method used in Part I to construct the virtual power of internal forces is here made clear, for it is the very expression thus constructed that contributes, in a ready-to-use form, to the statement of the second principle of thermodynamics. If the medium were an electric conductor with finite conductivity, then there would appear a Joule contribution $\mathcal{J}\cdot\mathcal{E}$ in the left-hand side of Eq. (3.12), where $\mathcal{J}$ and $\mathcal{E}$ would be respectively the conduction current and the electromotive intensity in moving media, the latter being such that

$$\mathcal{E} \equiv \mathbf{E} + (1/c)\,\mathbf{U}\times\mathbf{B}$$

where $\mathbf{E}$ is the electric field in a fixed Galilean frame.[13]

4. NONLINEAR THERMOELASTIC ANTIFERROMAGNETIC INSULATORS

4.1. General case

A. Strain measures

Noting $x_{i,K} \equiv \partial\chi_i/\partial X_K$ and $x_{K,i} \equiv \partial X_K/\partial x_i$, the direct and inverse deformation gradients, and remarking that $\mu_{(\alpha)} = \mu_{(\alpha)}(\mathbf{x},t)$ can also be written as

$$\mu_{(\alpha)} = \mu_{(\alpha)}(\mathbf{X},t) \tag{4.1}$$

on account of Eq. (2.13), so that the gradients $\mu_{(\alpha)i,K}$ are well-defined, one can construct the following objective fields which are scalars with respect to coordinate transformations in the configuration K, but are tensor-valued fields in the reference configuration K_0:

$$E_{KL} \equiv \tfrac{1}{2}(x_{i,K}x_{i,L} - \delta_{KL}) = E_{LK}, \tag{4.2}$$

$$m_K^\alpha \equiv \mu_{(\alpha)i}x_{i,K}, \tag{4.3}$$

$$m_{KL}^\alpha \equiv \mu_{(\alpha)i,K}x_{i,L}, \tag{4.4}$$

$$\mathcal{M}_{KL}^\alpha \equiv \mu_{(\alpha)i,K}\mu_{(\alpha)i,L} = \mathcal{M}_{LK}^\alpha, \tag{4.5}$$

$$\mathcal{M}_{KL}^{\alpha\beta} = \mu_{(\alpha)i,K}\mu_{(\beta)i,L} = \mathcal{M}_{KL}^{\beta\alpha} = \mathcal{M}_{LK}^{\alpha\beta}, \quad (\alpha\neq\beta); \tag{4.6}$$

E_{KL} is the usual Lagrangian strain tensor of nonlinear elasticity. It is related to the Cauchy strain tensor $C_{KL} = x_{i,K}x_{i,L}$ via the equation

$$C_{KL} = 2E_{KL} + \delta_{KL}. \tag{4.7}$$

The reciprocal of C_{KL}, such that $\overset{-1}{C}_{MK}C_{KL} = \delta_{ML}$, is easily shown to be given by

$$\overset{-1}{C}_{KL} = X_{K,i}X_{L,i}. \tag{4.8}$$

m_K^α and m_{KL}^α or $\mathcal{M}_{KL}^\alpha$ are measures of the magnetization sublattices and of the disuniformities of these magnetizations, respectively, expressed by convection in the initial configuration K_0. The following scalars can be defined from the vector fields $\mu_{(\alpha)}$, $\alpha = A, B$ (no summation over α):

$$\mu_{S\alpha}^2 = \mu_{(\alpha)}\cdot\mu_{(\alpha)} \tag{4.9}$$

$$\mu^{\alpha\beta} = \mu_{(\alpha)}\cdot\mu_{(\beta)} = \mu^{\beta\alpha}, \quad (\alpha\neq\beta). \tag{4.10}$$

Whereas $\mu_{S\alpha}$ is a measure of the magnitude of each magnetic sublattice, $\mu^{\alpha\beta}$ in fact measures the angle $\varphi^{\alpha\beta}$ between the sublattice directions at a given material point at all times. The last equalities on Eq. (4.6) are established by computing $\mu^{\alpha\beta}_{\cdots,KL}$. Introducing $\bar{m}_K^\alpha$ by

$$\bar{m}_K^\alpha = \mu_{(\alpha)j}X_{K,j} = \overset{-1}{C}_{KL}m_L^\alpha, \tag{4.11}$$

it is readily shown that Eqs. (4.9) and (4.10) can be rewritten as

$$\mu_{S\alpha}^2 = m_K^\alpha\bar{m}_K^\alpha = m_K^\alpha\overset{-1}{C}_{KL}m_L^\alpha, \tag{4.12}$$

$$\mu^{\alpha\beta} = m_K^\alpha\bar{m}_K^\beta = \bar{m}_K^\alpha m_K^\beta = \mu^{\beta\alpha}. \tag{4.13}$$

Also, on account of Eqs. (4.4) and (4.8), Eqs. (4.5) and (4.6) can be written in the form

$$\mathcal{M}_{KL}^\alpha = m_{KP}^\alpha\overset{-1}{C}_{PQ}m_{LQ}^\alpha, \quad \mathcal{M}_{KL}^{\alpha\beta} = m_{KP}^\alpha\overset{-1}{C}_{PQ}m_{LQ}^\beta. \tag{4.14}$$

If each magnetic sublattice is *saturated*, then we have $(\partial\mu_{S\alpha}/\partial X_K) = 0$ for all α. Thus,

$$\mu_{(\alpha)i,K}\mu_{(\alpha)i} = 0. \tag{4.15}$$

These constraints can be written in Lagrangian form with the aid of Eqs. (4.4), (4.8), and (4.11). We have

$$m_{KL}^\alpha\bar{m}_L^\alpha = 0. \tag{4.16}$$

We note that in contrast to the absolute scalar (4.10), one can also form the following pseudoscalar which changes sign by interchange of the roles played by α and β. Then we define $(\alpha\neq\beta)$, $\mathbf{l}$ being a unit vector field defined at the same point as $\mu_{(\alpha)}$ and $\mu_{(\beta)}$:

$$\bar{\mu}^{\alpha\beta}(\mathbf{l}) \equiv (\mu_{(\alpha)}\times\mu_{(\beta)})\cdot\mathbf{l} = -\bar{\mu}^{\beta\alpha}(\mathbf{l}) = -\bar{\mu}^{\alpha\beta}(-\mathbf{l}). \tag{4.17}$$

This quantity is useful in discussing the case of weakly ferromagnetic antiferromagnets.

One can also introduce the gradient of temperature in K_0 via the chain rule of differentiation

$$\theta_{,K} = \theta_{,i}x_{i,K}. \tag{4.18}$$

For subsequent use it is of interest to compute the time rate of the fields defined by Eqs. (4.2)—(4.4). Noting that $\mu_{(\alpha)i,K} = \mu_{(\alpha)i,j}x_{j,K}$ and using Eqs. (I. 2.9) and the definitions (2.20)—(2.22), it is found that

$$\dot{E}_{KL} = D_{ij}x_{i,K}x_{j,L}, \tag{4.19}$$

$$\dot{m}_K^\alpha = [\hat{m}_{(\alpha)i} + D_{ij}\mu_{(\alpha)j}]x_{j,K}, \tag{4.20}$$

$$\dot{m}_{KL}^\alpha = [\hat{\mathfrak{M}}_{(\alpha)ik} + D_{ij}\mu_{(\alpha)j,k}]x_{i,L}x_{k,K}. \tag{4.21}$$

It must be remarked that the objective rates introduced in Part I appear quite naturally in the above calculations.

B. Constitutive equations

Thermoelastic materials are materials which are described by a first-order-gradient theory (see Part I) and have a free energy with the following *a priori* functional dependence:

$$\psi = \psi(x_{i,K}, \mu_{(\alpha)i}, \mu_{(\alpha)i,K}, \theta, \theta_{,K}). \tag{4.22}$$

The same dependence is assumed to hold for the other dependent constitutive variable η, σ_{ij}, ${}^L\mathbf{B}_{(\alpha)}$, $\mathcal{B}_{(\alpha)}$, and $\mathbf{q}$, according to the working hypothesis of *equipresence*.[14] In order that ψ be an objective scalar it is necessary and sufficient, following a classical derivation,[15] that ψ reduce to the following functional form:

$$\psi = \tilde{\psi}(E_{KL}, m_K^\alpha, m_{KL}^\alpha, \theta, \theta_{,K}). \tag{4.23}$$

Assuming $\tilde{\psi}$ to be sufficiently differentiable in its arguments, the time derivative of ψ is computed on account of Eqs. (4.19)—(4.21). One obtains

$$\begin{aligned}\dot{\psi} = &\left[\frac{\partial\tilde{\psi}}{\partial E_{KL}}x_{j,L}x_{i,K} + \sum_\alpha\left(\frac{\partial\tilde{\psi}}{\partial m_K^\alpha}\mu_{(\alpha)j}x_{i,K}\right.\right.\\ &\left.\left.+\frac{\partial\tilde{\psi}}{\partial m_{KL}^\alpha}x_{i,L}\mu_{(\alpha)j,K}\right)\right]D_{ij}\\ &+\sum_\alpha\left[\left(\frac{\partial\tilde{\psi}}{\partial m_K^\alpha}x_{i,K}\right)\hat{m}_{(\alpha)i} + \left(\frac{\partial\tilde{\psi}}{\partial m_{KL}^\alpha}x_{i,L}x_{j,K}\right)\hat{\mathfrak{M}}_{(\alpha)ij}\right]\\ &+\frac{\partial\tilde{\psi}}{\partial\theta}\dot{\theta} + \frac{\partial\tilde{\psi}}{\partial\theta_{,K}}\dot{\overline{\theta_{,K}}}.\end{aligned} \tag{4.24}$$

Substituting from this equation into Eq. (3.12) and the latter being posited to be valid for arbitrary elements of the enlarged T. L. S. of velocities[16]

$$\mathcal{V}_{\mathrm{obj}} \oplus \{\dot{\theta}, \overset{\circ}{\overline{\theta_{,K}}}\}, \tag{4.25}$$

according to Coleman's axiomatics of the thermodynamics of continua,[17] we arrive in the usual fashion at the following result:

Theorem: The constitutive equations of a nonlinear thermoelastic antiferromagnetic insulator are:

$$\begin{aligned}{}^R\sigma_{ij} &= \rho\left\{\frac{\partial\hat{\psi}}{\partial E_{KL}} x_{(i,K} + \sum_\alpha\left(\frac{\partial\hat{\psi}}{\partial m_L^\alpha}\mu_{(\alpha)(i} + \frac{\partial\hat{\psi}}{\partial m^\alpha_{KL}}\mu_{(\alpha)(i,K)}\right)\right\} x_{j),L}\\ &= {}^R\sigma_{ji},\end{aligned} \tag{4.26}$$

$${}^{RL}B_{(\alpha)i} = -\frac{\partial\hat{\psi}}{\partial m_K^\alpha} x_{i,K}, \tag{4.27}$$

$${}^R\mathcal{B}_{(\alpha)ij} = \rho\frac{\partial\hat{\psi}}{\partial m^\alpha_{KL}} x_{i,L}x_{j,K}, \tag{4.28}$$

$$\eta = \frac{\partial\hat{\psi}}{\partial\theta}, \tag{4.29}$$

$$\frac{\partial\hat{\psi}}{\partial\theta_{,K}} = 0, \tag{4.30}$$

$$q_i = x_{i,M}\hat{Q}_M(E_{KL}, m_K^\alpha, m_{KL}^\alpha, \theta, \theta_{,K}), \tag{4.31}$$

the latter satisfying the "continuity" condition (if $\hat{Q}_M$ is assumed to be of class C^1 with respect to its argument $\theta_{,K}$)

$$\hat{Q}_M\Big|_{\theta_{,K}=0} = 0, \tag{4.32}$$

as well as the remaining thermal dissipation inequality

$$\mathbf{q}\cdot\nabla\theta \leqslant 0, \tag{4.33}$$

and ψ being reduced to

$$\psi = \hat{\psi}(E_{KL}, m_K^\alpha, m_{KL}^\alpha, \theta). \tag{4.34}$$

No peculiar material symmetry is here assumed. The left superscript R indicates that the constitutive equations (4.26)–(4.28) are derived from the potential $\hat{\psi}$, which, being a general function (that must however satisfy some positiveness and stability conditions), gives rise to a *nonlinear* behavior, in particular, to the so-called *hyperelastic* behavior as far as mechanical effects are concerned. Thus, apart from heat conduction, all phenomena here described are thermodynamically *recoverable*. Equations (4.27) and (4.28) show that ${}^{RL}\mathbf{B}_{(\alpha)}$ and ${}^R\mathcal{B}_{(\alpha)}$ are primarily determined by the magnetization sublattices and their disuniformities, respectively, so that the interpretations conjectured in Part I for the fields ${}^L\mathbf{B}_{(\alpha)}$ and $\mathcal{B}_{(\alpha)}$ are corroborated by the thermodynamical study. As far as elastic effects and the contributions of magnetic effects to the Cauchy stress tensor are concerned, a simple rearranging of Eq. (2.10) on account of the results (4.26)–(4.28) allows us to put in evidence the different properties. Indeed, call ${}^Rt_{ij}$ the thermodynamically recoverable Cauchy stress thus obtained; on account of the results obtained above it can be rewritten in the following suggestive form:

$${}^Rt_{ij} = {}^Et_{ij} - \sum_\alpha\left(\rho\,{}^{RL}B_{(\alpha)j}\mu_{(\alpha)i} - {}^R\mathcal{B}_{(\alpha)jk}\mu_{(\alpha)i,k}\right), \tag{4.35}$$

where ${}^Et_{ij}$ is the *symmetric* stress tensor defined by

$${}^Et_{ij} \equiv \rho\frac{\partial\hat{\psi}}{\partial E_{KL}} x_{i,K}x_{j,L} = {}^Et_{ji}. \tag{4.36}$$

The latter may be called the *elastic stress tensor*, to which ${}^Rt_{ij}$ would reduce in the absence of magnetic effects; it here includes not only purely elastic effects, but also effects such as magnetostriction, piezomagnetism, and exchange-strictive effects.

The decomposition (4.36), which has been derived only in the case of thermoelastic bodies, shows that in contrast to what could be figured out from the original (but more general) decomposition (2.10), the spin–lattice interactions and the spin–spin interactions (of exchange and superexchange origins) intervene in the Cauchy stress not only via the skewsymmetric combinations of Eq. (2.10), but also via the analogous symmetric combinations, so that the formula (4.36) holds true. Both the decompositions (4.36) and (2.10) can further be simplified if the magnetic sublattices are supposed to be saturated, i.e., of spatially constant magnitude, a reasonable assumption at sufficiently low temperatures.

4.2. Magnetically saturated sublattices

In that case where Eq. (4.15) holds true for each α separately, the constraints $\mu_{S\alpha} = \mathrm{const}$ throughout space, and those represented by Eqs. (4.16)—which exhibit a relationship between the different arguments appearing in $\hat{\psi}$—must be taken into account in computing $\hat{\psi}$ according to Eq. (4.24). One method is to introduce Lagrange multipliers $\mathcal{P}_\alpha$ and $\mathcal{P}_{\alpha K}$, $\alpha = A, B$, $K = 1, 2, 3$, respectively for the constraints $\mu_{S\alpha} = \mathrm{const}$ and Eq. (4.16). That is, we may consider in lieu of $\hat{\psi}$ the following effective free energy density

$$\begin{aligned}\psi^{\mathrm{eff}} &= \hat{\psi}(E_{KL}, m_K^\alpha, m_{KL}^\alpha, \theta)\\ &\quad - \sum_\alpha\{\mathcal{P}_\alpha(m_K^\alpha\bar{m}_K^\alpha - \mu_{S\alpha}^2) + \mathcal{P}_{\alpha K}m_{KL}^\alpha\bar{m}_L^\alpha\}.\end{aligned} \tag{4.37}$$

However, instead of the last constraints involving $\mathcal{P}_{\alpha K}$, noting that Eqs. (4.16) represent six scalar constraints and that m_{KL}^α and $\mathcal{M}_{KL}^\alpha(\alpha = A, B)$ have respectively eighteen and twelve independent components, it is astute to replace the dependence of ψ upon m_{KL}^α by that upon $\mathcal{M}_{KL}^\alpha$ and to discard the Lagrange multipliers $\mathcal{P}_{\alpha K}$. Thus,

$$\psi^{\mathrm{eff}} = \bar{\psi}(E_{KL}, m_K^\alpha, \mathcal{M}_{KL}^\alpha, \theta) - \sum_\alpha \mathcal{P}_\alpha(m_K^\alpha\bar{m}_K^\alpha - \mu_{S\alpha}^2). \tag{4.38}$$

The computations are made much easier with this last effective free energy. Indeed, noting that

$$\frac{\partial\hat{\psi}}{\partial m_{KL}^\alpha} = 2\frac{\partial\bar{\psi}}{\partial\mathcal{M}_{KP}^\alpha}X_{L,k}\mu_{(\alpha)k,P}, \tag{4.39}$$

$$\frac{\partial\hat{\psi}}{\partial E_{KL}} = \frac{\partial\bar{\psi}}{\partial E_{KL}} + \sum_\alpha\left(\frac{\partial\bar{\psi}}{\partial\mathcal{M}_{PQ}^\alpha}\frac{\partial\mathcal{M}_{PQ}^\alpha}{\partial\overset{-1}{C}_{MN}}\right)\frac{\partial\overset{-1}{C}_{MN}}{\partial E_{KL}}, \tag{4.40}$$

of which the second transforms to

$$\frac{\partial\hat{\psi}}{\partial E_{KL}}x_{i,K}x_{j,L} = \frac{\partial\bar{\psi}}{\partial E_{KL}}x_{i,K}x_{j,L} - 2\sum_\alpha\frac{\partial\bar{\psi}}{\partial\mathcal{M}_{MN}^\alpha}\mu_{(\alpha)(i,M}\mu_{(\alpha)j),N}, \tag{4.41}$$

on account of the intermediate results

$$\frac{\partial \mathcal{M}^{\alpha}_{PQ}}{\partial \overset{-1}{C}_{MN}} = m^{\alpha}_{PM} m^{\alpha}_{QN}, \quad \frac{\partial \overset{-1}{C}_{MN}}{\partial E_{KL}} = -2\overset{-1}{C}_{MK}\overset{-1}{C}_{NL}, \tag{4.42}$$

which follow from Eqs. (4.14) and (4.7)–(4.8), the Eqs. (4.35), (4.36), (4.27), and (4.28) are transformed to the following ones:

$$^{R}t_{ij} = {}^{E}t_{ij} - \sum_{\alpha} (\rho\, {}^{RL}B_{(\alpha)j}\, \mu_{(\alpha)i}), \tag{4.43}$$

$$^{E}t_{ij} = \rho \frac{\partial \overline{\psi}}{\partial E_{KL}} x_{i,K} x_{j,L} = {}^{E}t_{ji}, \tag{4.44}$$

$$^{RL}B_{(\alpha)i} = -\frac{\partial \overline{\psi}}{\partial m^{\alpha}_{K}} x_{i,K}, \tag{4.45}$$

$$^{R}\mathcal{B}_{(\alpha)ij} = 2\rho \frac{\partial \overline{\psi}}{\partial \mathcal{M}^{\alpha}_{KL}} \mu_{(\alpha)i,L} x_{j,K}. \tag{4.46}$$

The following comments are in order concerning these results. First, it can be remarked that the $\mathcal{P}_{\alpha}$'s do not contribute to the expression of $^{R}t_{ij}$, as is shown by the calculation; their contribution to $^{RL}\mathbf{B}_{(\alpha)}$ has been discarded since they yield vanishing contributions in the precession equations (2.4). Next, the derivatives $\partial\overline{\psi}/\partial\mathcal{M}^{\alpha}_{KL}$ do not contribute at all to the expression of $^{R}t_{ij}$. This, of course, does not mean that $\mathcal{M}^{\alpha}_{KL}$ cannot appear in this expression since a function of $\mathcal{M}^{\alpha}_{KL}$ may be in a factor of E_{KL} in the free energy density, so that $\mathcal{M}^{\alpha}_{KL}$ will, in general, appear in the expression of $^{E}t_{ij}$, thus yielding an exchange-strictive effect. However, the important point here is that the exchange and superexchange forces do *not* participate in the skewsymmetric part of $^{R}t_{ij}$, i.e., in the effective volume couple acting upon the crystal lattice. Indeed, from the symmetric of $\mathcal{M}^{\alpha}_{KL}$ in K and L we have, with the aid of Eq. (4.46),

$$^{R}\mathcal{B}_{(\alpha)[i|k|}\mu_{(\alpha)j],k} = 2\rho \frac{\partial \overline{\psi}}{\partial \mathcal{M}^{\alpha}_{KL}} \mu_{(\alpha)[i,K}\mu_{(\alpha)j],L} = 0. \tag{4.47}$$

Thus, as shown by Eq. (4.43) and on account of the symmetry of $^{E}t_{ij}$, we have

$$^{R}t_{[ij]} = \sum_{\alpha} \rho\, {}^{RL}B_{(\alpha)[i}\mu_{(\alpha)j]}. \tag{4.48}$$

Equation (4.47) is none other than the constraint *a priori* considered in Part I in order that the global law (I.3.49) for the magnetic sublattices be satisfied and for the model of three interacting continua constructed in the Appendix of Part I to be valid. Then Eq. (4.48) is none other than the local balance law (I.A7) of moment of momentum for the crystal lattice, which in fact expresses the only *direct* interactions which occur between the crystal lattice and the magnetic sublattices. However, as already indicated above, there may be other couplings between the deformations of the crystal lattice and the precession of the magnetic sublattices via magnetostrictive, and exchange-strictive effects. Such effects can be made clear by further specifying the form of the free energy function $\overline{\psi}$ or $\hat{\psi}$ and the material symmetry of the magnetically ordered deformable body. This is examined in the next section.

The constitutive equations (4.43)–(4.46) for *thermoelastic antiferromagnetic insulators with saturated sublattices*, thus at temperatures much below the Nèel's temperature, are of course supplemented with Eq. (4.38) and the heat conduction law

$$q_i = x_{i,M}\overline{Q}_M(E_{KL}, m^{\alpha}_{K}, \mathcal{M}^{\alpha}_{KL}, \theta, \theta_{,K}), \tag{4.49}$$

which replaces Eq. (4.31) and satisfies conditions analogous to Eqs. (4.32) and (4.33).

5. APPROXIMATIONS

5.1. Expansion of the free energy

The exact reduced functional form of the free energy $\hat{\psi}$ or $\overline{\psi}$ can be found for special material symmetries (e.g., full isotropy, hemitropy, orthotropy) with the help of exact representation theorems (see Wang,[18] and Spencer[19]); however, most often, one is satisfied with a reasonable expansion of the free energy in terms of its arguments. We give such an expansion for the case of thermoelastic solids with saturated magnetic sublattices. In these conditions, if E^{0}_{KL} is an initial strain field and $\widetilde{E}_{AB}$ is a perturbation such that

$$E_{AB} = E^{0}_{AB} + \widetilde{E}_{AB}, \quad E^{0}_{AB} = \tfrac{1}{2}(x^{0}_{i,A}x^{0}_{i,B} - \delta_{AB}), \tag{5.1}$$

we obtain

$$\begin{aligned}\rho_0\psi = {} & \overset{1}{G}(m^{\alpha}_{I}, \mathcal{M}^{\alpha}_{IJ}, \theta; E^{0}_{KL}) + \overset{2}{G}_{AB}(m^{\alpha}_{I}, \mathcal{M}^{\alpha}_{IJ}, \theta; E^{0}_{KL})\widetilde{E}_{AB} \\ & + \overset{3}{G}_{ABCD}(m^{\alpha}_{I}, \mathcal{M}^{\alpha}_{IJ}, \theta, E^{0}_{KL})\widetilde{E}_{AB}\widetilde{E}_{CD} \\ & + \text{h.o.t. in } \widetilde{E}_{AB}.\end{aligned} \tag{5.2}$$

Assuming that

$$\begin{aligned} & m^{\alpha}_{I} = m^{\alpha 0}_{I} + \widetilde{m}^{\alpha}_{I}, \quad m^{\alpha 0}_{I} = \mu_{(\alpha)i}x^{0}_{i,I}, \\ & \mathcal{M}^{\alpha}_{IJ} = \mathcal{M}^{\alpha 0}_{IJ} + \widetilde{\mathcal{M}}^{\alpha}_{IJ}, \quad \theta = \theta^{0} + \widetilde{\theta}\end{aligned} \tag{5.3}$$

where θ^{0} is a uniform reference temperature field such that $(\widetilde{\theta}/\theta^{0}) \ll 1$ and $\theta \ll \theta_N$, and that the tensorial coefficients of the expansion (5.2) are sufficiently differentiable in their arguments, we get

$$\begin{aligned}\psi = {} & \psi_0 - \eta_0\widetilde{\theta} - \frac{\gamma}{2\theta^{0}}\widetilde{\theta}^2 + \rho_0^{-1}[N_{KL}\widetilde{E}_{KL} + \sum_{\alpha}(N^{\alpha}_{K}\widetilde{m}^{\alpha}_{K} + M^{\alpha}_{KL}\widetilde{\mathcal{M}}^{\alpha}_{KL})]\widetilde{\theta} \\ & + \sum_{\alpha}\rho_0\chi^{(0\alpha)}_{K}\widetilde{m}^{\alpha}_{K} + \frac{\rho_0}{2}\sum_{\alpha}\chi^{(\alpha)}_{KL}\widetilde{m}^{\alpha}_{L}\widetilde{m}^{\alpha}_{L} + \rho_0\sum_{\alpha\neq\beta}\chi^{(\alpha\beta)}_{KL}\widetilde{m}^{\alpha}_{K}\widetilde{m}^{\beta}_{L} \\ & + \frac{\rho_0}{2}\sum_{\alpha}A^{(\alpha)}_{KL}\widetilde{\mathcal{M}}^{\alpha}_{KL} + \sum_{\alpha}\mathcal{L}^{(\alpha)}_{KPQ}\widetilde{m}^{\alpha}_{K}\widetilde{\mathcal{M}}^{\alpha}_{PQ} + \sum_{\alpha\neq\beta}\mathcal{L}^{(\alpha\beta)}_{KPQ}\widetilde{m}^{\alpha}_{K}\widetilde{\mathcal{M}}^{\beta}_{PQ} \\ & + \rho_0^{-1}(L_{KL}\widetilde{E}_{KL} + \tfrac{1}{2}L_{KLMN}\widetilde{E}_{KL}\widetilde{E}_{MN}) + \sum_{\alpha}\epsilon^{(m\alpha)}_{KLM}\widetilde{m}^{\alpha}_{K}\widetilde{E}_{LM} \\ & + \sum_{\alpha}\rho_0\gamma^{(e\alpha)}_{KLMN}\widetilde{E}_{KL}\widetilde{\mathcal{M}}^{\alpha}_{MN} + \sum_{\alpha\neq\beta}\rho_0\gamma^{(m\alpha\beta)}_{KLMN}\widetilde{E}_{KL}\widetilde{m}^{\alpha}_{M}\widetilde{m}^{\beta}_{N} \\ & + \sum_{\alpha}\rho_0\gamma^{(m\alpha)}_{KLMN}\widetilde{E}_{KL}\widetilde{m}^{\alpha}_{M}\widetilde{m}^{\alpha}_{N} + \cdots,\end{aligned} \tag{5.4}$$

where the different coefficients introduced (η_0, γ, N_{KL}, N^{α}_{K}, etc) have obvious definitions as derivatives of different orders of the G's taken at the zero value of the arguments, and satisfy trivial symmetry conditions that we do not reproduce here. According to the accepted terminology, the different tensorial coefficients, which still depend on the initial state (here assumed nonnatural) represent the following effects: N_{KL}: thermoelasticity; N^{α}_{K}: pyromagnetism; M^{α}_{KL}: pyro-exchange effect; $\chi^{(\alpha)}_{KL}$ and $\chi^{(\alpha\beta)}_{KL}$: magnetocrystalline effects and exchange effects not due to the disuniformities in the magnetization fields; $A_{KL}^{(\alpha)}$: exchange forces; L_{KL}: initial stress; L_{KLMN}: elasticity; $\epsilon_{KLM}^{(m\alpha)}$: piezomagnetism; $\gamma_{KLMN}^{(e\alpha)}$: exchange-strictive effects; $\gamma_{KLMN}^{(m\alpha\beta)}$ and $\gamma_{KLMN}^{(m\alpha)}$: magnetostriction. The terms including $\chi_K^{(0\alpha)}$, $\mathcal{L}_{KPQ}^{(\alpha)}$, and $\mathcal{L}_{KPQ}^{(\alpha\beta)}$ are introduced for the sake of generality, but can be shown, in certain conditions specified below, to be zero if the corresponding terms are to be invariant under the operation of time-reversal $\mathcal{R}$.[20]

Now consider the case where the expansion (5.4) is made about a natural undeformed state that is free of stress and is not magnetized. Then,

$$m_I^{\alpha 0}=0, \quad \mathcal{M}_{IJ}^{\alpha 0}=0, \quad E_{KL}^0=0, \quad L_{KL}=0. \tag{5.5}$$

Let us further assume that $M_{KL}^{\alpha}=0$. Remarking that the remaining tensorial coefficients of the expansion (5.4) now have components which depend only on θ^0 (and possibly on some other parameter of the material such as the matter density), the coefficients $\chi_K^{(0\alpha)}$, $\mathcal{L}_{KPQ}^{(\alpha)}$, and $\mathcal{L}_{KPQ}^{(\alpha\beta)}$ must be zero in virtue of the time-reversal invariance. Let us finally assume that the material is *centro-symmetric*, so that the remaining third-order tensorial coefficients must be zero, for there do not exist representations of such tensors for centrosymmetry. If all these conditions are fulfilled, the expression (5.4) reduces fo the following one:

$$\psi=\psi_{\text{th. el}}+\psi_{\text{m. ex}}+\psi_{\text{m. st}}+\psi_{\text{ex. st}}+\psi_{\text{ex}}, \tag{5.6}$$

where (the tilde is no longer necessary, except for the temperature)

$$\psi_{\text{th. el}}\equiv\psi_0-\eta_0\tilde{\theta}-\frac{\gamma}{2\theta^0}\tilde{\theta}^2-\rho_0^{-1}\tilde{\theta}N_{KL}(\theta^0)E_{KL} + \frac{1}{2\rho_0}L_{KLMN}(\theta^0)E_{KL}E_{MN}, \tag{5.7}$$

$$\psi_{\text{m. ex}}\equiv\frac{\rho_0}{2}\sum_{\alpha}\chi_{KL}^{(\alpha)}(\theta^0)m_K^{\alpha}m_L^{\alpha}+\rho_0\sum_{\alpha\neq\beta}\chi_{KL}^{(\alpha\beta)}(\theta^0)m_K^{\alpha}m_L^{\beta}, \tag{5.8}$$

$$\psi_{\text{m. st}}\equiv\sum_{\alpha}\rho_0\gamma_{KLMN}^{(m\alpha)}(\theta^0)E_{KL}m_M^{\alpha}m_N^{\alpha}+\sum_{\alpha\neq\beta}\rho_0\gamma_{KLMN}^{(m\alpha\beta)}(\theta^0)E_{KL}m_M^{\alpha}m_N^{\beta}, \tag{5.9}$$

$$\psi_{\text{ex. st}}\equiv\sum_{\alpha}\rho_0\gamma_{KLMN}^{(e\alpha)}(\theta^0)E_{KL}\mathcal{M}_{MN}^{\alpha}, \tag{5.10}$$

$$\psi_{\text{ex}}\equiv\frac{\rho_0}{2}\sum_{\alpha}A_{KL}^{\alpha}(\theta^0)\mathcal{M}_{KL}^{\alpha}. \tag{5.11}$$

The expressions (5.7)—(5.11) represent, respectively: (i) the thermoelastic energy; (ii) the magnetocrystalline energy and the exchange energy not due to the disuniformities in the magnetization fields (superexchange forces); (iii) the magnetostrictive energy; (iv) the exchange-strictive energy; (v) the exchange energy which represents the interaction energy between spins of the *same* magnetic sublattice. Concerning the latter, the following important remark must be made. Clearly, by performing the expansion procedure and the approximations represented by Eqs. (5.2)—(5.6), we have partly disconnected the different interactions, in terms of the initial independent variables present in Eq. (4.22), the expansion (5.6) can be written formally as

$$\psi=\psi_1(x_{i,K};\theta)+\psi_2(x_{i,K},\mu_{(\alpha)i};\theta^0) + \psi_3(x_{i,K},\mu_{(\alpha)i,K};\theta^0)+\psi_4(\mu_{(\alpha)i,K};\theta^0), \tag{5.12}$$

where there is no one-to-one correspondence with the energies defined by Eqs. (5.7)—(5.11). Applying the objectivity requirement to each ψ_i, $i=1,2,3,4$ of Eq. (5.12), we find

$$\psi=\bar{\psi}_1(E_{KL};\theta)+\bar{\psi}_2(E_{KL},m_K^{\alpha};\theta^0) + \bar{\psi}_3(E_{KL},\mathcal{M}_{KL}^{\alpha};\theta^0)+\bar{\psi}_4(\mathcal{M}_{KL}^{\alpha},\mathcal{M}_{KL}^{\alpha\beta};\theta^0). \tag{5.13}$$

Here, the representation of ψ_4 that contains the variable defined by Eq. (4.6) follows from the usual Cauchy's theorem.[21] Taking the expansion of the different contributions in Eq. (5.13) and regrouping alike terms, we obtain an expression of the same form as Eq. (5.6), but with an extra term which has been overlooked in the afore-used procedure:

$$\psi_{\text{s. ex}}=\rho_0\sum_{\alpha\neq\beta}A_{KL}^{(\alpha\beta)}\mathcal{M}_{KL}^{\alpha\beta}, \tag{5.14}$$

that represents the superexchange energy resulting from the disuniformities in the magnetization fields of *different* magnetic sublattices. Thus, in fact, the contribution (5.14) must be added to Eq. (5.6), which now reads

$$\psi=\psi_{\text{th. el}}+\psi_{\text{m. ex}}+\psi_{\text{m. st}}+\psi_{\text{ex. st}}+\psi_{\text{ex}}+\psi_{\text{s. ex}}. \tag{5.15}$$

It can also be remarked that Eqs. (5.11) and (5.14) are limited to the first order in $\mathcal{M}_{KL}^{\alpha}$ and $\mathcal{M}_{KL}^{\alpha\beta}$; the reason is that the latter variables are already of second order in the magnetization gradients. Finally, it is not surprising that both expressions (5.10) and (5.11) appear simultaneously, for it can be shown, when only one magnetic sublattice is involved (i.e., in ferromagnetism), that the whole expression

$$\tfrac{1}{2}\rho_0(A_{KL}\mathcal{M}_{KL}+2\gamma_{KLMN}^{(m)}E_{KL}\mathcal{M}_{MN}, \tag{5.16}$$

results as a whole from a semimicroscopic model based on Heisenberg's expression for the spin—spin interaction potential applied to an elastic body subjected to large deformations.[22] However, it must also be emphasized that the last contribution in Eq. (5.16) is of the order of (strain)$\times|\nabla\mu|^2$, so that the contribution (5.10) will in general be discarded in simplified theories.

Since the antiferromagnetic bodies we are interested in are seldom subjected to large deformations, it is of importance to examine the case of *small* deformations. This will offer the opportunity to specify, for a chosen material symmetry, the explicit form of the material tensors, and to compare the resulting form of the internal energy with those postulated in other works.

5.2. Infinitesimal deformations

A. General case

In the case of infinitesimal deformations about a natural undeformed configuration $\mathcal{K}_0$, it is assumed that the displacement u (components u_i) is such that $|\nabla u| < \delta$, where δ is infinitesimally small of the first order. Then, as δ goes to zero, we have

$$x_{i,K}\approx(\delta_{ij}+u_{i,j})\delta_{jK}, \qquad E_{KL}\approx e_{ij}\delta_{iK}\delta_{jL}, \quad e_{ij}\equiv\tfrac{1}{2}(u_{i,j}+u_{j,i}), \tag{5.17}$$

since products of ∇u are $O(\delta^2)$. e_{ij} is the usual linearized Eulerian strain tensor. Consequently, we have

$$
\begin{aligned}
&m_K^\alpha \approx \mu_{(\alpha)i}\delta_{iK},\\
&\mathcal{M}_{KL}^\alpha \approx \mathcal{M}_{kp}^\alpha \delta_{kK}\delta_{pL}, \quad \mathcal{M}_{kp}^\alpha \equiv \mu_{(\alpha)i,k}\mu_{(\alpha)i,p},\\
&\mathcal{M}_{KL}^{\alpha\beta} \approx \mathcal{M}_{kp}^{\alpha\beta}\delta_{kK}\delta_{pL}, \quad \mathcal{M}_{kp}^{\alpha\beta} \equiv \mu_{(\alpha)i,k}\mu_{(\beta)i,p},\\
&\rho \approx \rho_0(1-e_{kk}),
\end{aligned} \tag{5.18}
$$

where we have neglected product terms of the types $\mu\otimes\nabla\mathbf{u}$ and $\nabla\mathbf{u}\otimes\nabla\mu_{(\alpha)}$. Of course, it is no longer distinguished between small and capital Latin indices. The general constitutive equations (4.43)—(4.46), and (4.29), (4.31), and (4.34) take the approximate forms

$$
{}^Rt_{ij} = \rho_0\left(\frac{\partial\hat{\psi}}{\partial e_{ij}} + \sum_\alpha \frac{\partial\hat{\psi}}{\partial\mu_{(\alpha)j}}\mu_{(\alpha)i}\right), \tag{5.19}
$$

$$
{}^{RL}B_{(\alpha)i} = -\frac{\partial\hat{\psi}}{\partial\mu_{(\alpha)i}}, \tag{5.20}
$$

$$
{}^R\mathcal{B}_{(\alpha)ij} = \rho_0\frac{\partial\hat{\psi}}{\partial\mu_{(\alpha)i,j}}, \tag{5.21}
$$

$$
\eta = -\frac{\partial\hat{\psi}}{\partial\tilde{\theta}}, \tag{5.22}
$$

$$
q_p = \hat{q}_p(e_{kl}, \mu_{(\alpha)i}, \mathcal{M}_{ij}^\alpha, \mathcal{M}_{ij}^{\alpha\beta}, \theta, \nabla\theta) \tag{5.23}
$$

$$
\psi = \hat{\psi}(e_{kl}, \mu_{(\alpha)i}, \mathcal{M}_{ij}^\alpha, \mathcal{M}_{ij}^{\alpha\beta}, \theta). \tag{5.24}
$$

On account of Eqs. (5.17) and (5.18) the expressions (5.7)—(5.11) and (5.14) take the following form:

$$
\psi_{\text{th. el}} = \psi_0 - \eta_0\tilde{\theta} - \frac{\gamma}{2\theta^0}\tilde{\theta}^2 - \rho_0^{-1}\tilde{\theta}\,\nu_{ij}(\theta^0)e_{ij} + \frac{1}{2\rho_0}\lambda_{ijkl}(\theta^0)e_{ij}e_{kl}, \tag{5.25}
$$

$$
\psi_{\text{m. ex}} = \frac{\rho_0}{2}\sum_\alpha \chi_{ij}^{(\alpha)}(\theta^0)\mu_{(\alpha)i}\mu_{(\alpha)j} + \rho_0\sum_{\alpha\neq\beta}\chi_{ij}^{(\alpha\beta)}(\theta^0)\mu_{(\alpha)i}\mu_{(\beta)j}, \tag{5.26}
$$

$$
\begin{aligned}
\psi_{\text{m. st}} = &\sum_\alpha \rho_0\gamma_{ijkl}^{(m\alpha)}(\theta^0)e_{ij}\mu_{(\alpha)k}\mu_{(\alpha)l}\\
&+\sum_{\alpha\neq\beta}\rho_0\gamma_{ijkl}^{(m\alpha\beta)}(\theta^0)\,e_{ij}\mu_{(\alpha)k}\mu_{(\beta)l},
\end{aligned} \tag{5.27}
$$

$$
\psi_{\text{ex}} = \frac{\rho_0}{2}\sum_\alpha a_{ij}^{(\alpha)}(\theta^0)\mathcal{M}_{ij}^\alpha, \tag{5.28}
$$

$$
\psi_{\text{s. ex}} = \rho_0\sum_{\alpha\neq\beta} a_{ij}^{(\alpha\beta)}(\theta^0)\mathcal{M}_{ij}^{\alpha\beta}. \tag{5.29}
$$

We have discarded the exchange-strictive effect for an above-given reason. The remaining material tensor coefficients satisfy the following tensor symmetries:

$$
\begin{aligned}
&\nu_{ij} = \nu_{ji}, \quad \lambda_{ijkl} = \lambda_{(ij)(kl)} = \lambda_{klij},\\
&\chi_{ij}^{(\alpha)} = \chi_{ij}^{(\alpha)}, \quad \chi_{ij}^{(\alpha\beta)} = \chi_{ji}^{(\alpha\beta)} = \chi_{ij}^{(\beta\alpha)},\\
&\gamma_{ijkl}^{(m\alpha)} = \gamma_{jikl}^{(m\alpha)} = \gamma_{ijlk}^{(m\alpha)}, \quad \gamma_{ijkl}^{(m\alpha\beta)} = \gamma_{jikl}^{(m\alpha\beta)} = \gamma_{ijkl}^{(m\beta\alpha)},\\
&a_{ij}^{(\alpha)} = a_{ji}^{(\alpha)}, \quad a_{ij}^{(\alpha\beta)} = a_{ij}^{(\beta\alpha)}.
\end{aligned} \tag{5.30}
$$

As regards the heat flux, it is of course assumed that $|\nabla\theta| < \delta_h$, where δ_h is infinitesimally small. On account of the continuity condition (4.32) a Taylor series expansion about $\nabla\theta = 0$ yields the classical Fourier law

$$
q_i = -K_{ij}(\theta^0)\theta_{,j} + o(\delta_h) \tag{5.31}
$$

as δ_h goes to zero. The tensor K_{ij}, which is referred to as the conductivity tensor, is symmetric after the Onsager—Casimir relations, and is semi-positive definite according to the remaining thermal inequality (4.33). In writing the functional dependence of K_{ij}, we have discarded the coupling of heat conduction with the thermodynamically recoverable phenomena. It remains to specify the material symmetry.

B. Elastically isotropic antiferromagnets with a magnetic easy axis

As far as the elastic properties are concerned, the antiferromagnetic solids fall in different crystallographic classes. In this respect many of them are either simply polycrystals, so that they can be considered as *elastically isotropic*, or are of cubic structure. Even if the latter structure is considered, it must be remarked that the typical nondimensional parameter (c_{44}, c_{11}, and c_{12} are usual adiabatic elastic constants for a cubic crystal)[23]

$$
\xi \equiv 1 - 2c_{44}/(c_{11} - c_{12}), \tag{5.32}
$$

usually used to measure the departure of cubic symmetric from isotropy is quite small in typical antiferromagnets, of the order of 0.06. It follows that we shall content ourselves with giving the explicit expressions of the material tensors describing elastic effects only for *isotropy*, the expressions obtained being considered for illustrative purpose. Then the tensor coefficients ν_{ij}, λ_{ijkl}, $\gamma_{ijkl}^{(m\alpha)}$, and $\gamma_{ijkl}^{(m\alpha\beta)}$ take on their isotropic form, which follows from a classical representation theorem due to Racah[24]:

$$
\begin{aligned}
&\nu_{ij} = \nu\delta_{ij},\\
&\lambda_{ijkl} = \lambda_1\delta_{ij}\delta_{kl} + \lambda_2(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}),\\
&\gamma_{ijkl}^{(m\alpha)} = b_1^\alpha\delta_{ij}\delta_{kl} + b_2^\alpha(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}),\\
&\gamma_{ijkl}^{(m\alpha\beta)} = d_1^{(\alpha\beta)}\delta_{ij}\delta_{kl} + d_2^{(\alpha\beta)}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}),
\end{aligned} \tag{5.33}
$$

on account of the symmetries (5.30). ν is the stress-temperature coefficient, λ_1 and λ_2 are the adiabatic Lamé constants, and b_1^α and b_2^α ($\alpha = A, B$), and $d_1^{(\alpha\beta)}$ and $d_2^{(\alpha\beta)}$ ($\alpha\neq\beta$) are the magnetostrictive constants, all in fact dependent on $\theta^0 \ll \theta_N$. Substituting from Eqs. (5.33) in Eqs. (5.25) and (5.27) and noting

$$
M_i^A \equiv \rho_0\mu_{(A)i}, \quad M_i^B \equiv \rho_0\mu_{(B)i}, \tag{5.34}
$$

the sublattice magnetizations per unit volume for the two sublattices $\alpha = A, B$, the resulting expressions of the different energies are positive definite if and only if the following restrictions are imposed on the various constants:

$$
\begin{aligned}
&\gamma > 0, \quad \nu > 0, \quad 3\lambda_1 + 2\lambda_2 > 0, \quad \lambda_2 > 0,\\
&b_1^\alpha = b_2^\alpha = d_2^{(\alpha\beta)} = 0.
\end{aligned} \tag{5.35}
$$

The algebra leading to these results is similar to that given in other works.[25] There is no magnetostrictive effect except through the constant $d_1^{(\alpha\beta)}$.

As far as heat conduction and the magnetization disuniformities are concerned, the material tensors considered are of second order, so that it need not be distinguished between isotropy and cubic symmetry, for the representations are of the same type in both cases. That is, with $\alpha = A, B$,

$$K_{ij}=K(\theta^0)\delta_{ij}, \tag{5.36}$$

$$a_{ij}^{(AA)}=a_{ij}^{(BB)}=\alpha(\theta^0)\delta_{ij}, \quad a_{ij}^{(AB)}=\alpha'(\theta^0)\delta_{ij}. \tag{5.37}$$

The same constant α is used for representing both $a_{ij}^{(AA)}$ and $a_{ij}^{(BB)}$, for the latter represent similar interactions within each magnetic sublattice. The semipositive definiteness condition (4.33), and the positive definiteness of the interaction energy $\psi_{ex}+\psi_{s,ex}$ require that these new coefficients satisfy the following restrictions:

$$K\geq 0, \tag{5.38}$$

$$\alpha>0, \quad (\alpha+\alpha')(\alpha-\alpha')>0. \tag{5.39}$$

Finally, as far as the magnetocrystalline effects and the exchange phenomena not due to the magnetization disuniformities are concerned, a realistic exemplary symmetry is that which corresponds to a *uniaxial* antiferromagnet, the ground state of which in the absence of an external magnetic field is determined by two compensated magnetic sublattices. Let $\mathbf{n}_0$ be the unit vector field pointing in the preferred direction thus distinguished for the magnetic anisotropy properties. The symmetry group under which the material tensors $\chi_{ij}^{(\alpha)}$ and $\chi_{ij}^{(\alpha\beta)}$ must be invariant then is that of rotations $R_{\mathbf{n}_0}^{\varphi}$ by an angle φ, $0<\varphi<2\pi$, about the unit direction $\mathbf{n}_0$. According to a theorem due to Smith and Rivlin,[26] $\chi_{ij}^{(\alpha)}$ is necessarily of the form

$$\chi_{ij}^{(\alpha)}=\beta_2^{\alpha}\delta_{ij}-\beta_1^{\alpha}n_{0i}n_{0j}, \quad \alpha=A,B. \tag{5.40}$$

As to $\chi_{ij}^{(\alpha\beta)}$, if the rotations about the direction $\mathbf{n}_0$ do transform one of the sublattice into the other, then it has a representation analogous to that of $\chi_{ij}^{(\alpha)}$.[27] That is,

$$\chi_{ij}^{(\alpha\beta)}=\beta_2'\delta_{ij}-\beta' n_{0i}n_{0j}. \tag{5.41}$$

If, however, such rotations do not transform one sublattice into the other, then a supplementary joint invariant of $\mathbf{M}^A$ and $\mathbf{M}^B$ must be considered. Equation (5.41) yielded the quadratic invariants $\mathbf{M}^A\cdot\mathbf{M}^B$ and $(\mathbf{M}^A\cdot\mathbf{n}_0)(\mathbf{M}^B\cdot\mathbf{n}_0)$. Now we have in supplement the quadratic invariant $M^{AB}(\mathbf{n}_0)\equiv(\mathbf{M}^A\times\mathbf{M}^B)\cdot\mathbf{n}_0$,[28] which changes sign under interchange of A and B, so that the representation (5.41) must be replaced by the more general one:

$$\chi_{ij}^{(AB)}=\beta_2'\delta_{ij}-\beta' n_{0i}n_{0j}+d\epsilon_{ijk}n_{0k}. \tag{5.42}$$

On account of Eqs. (5.41) and (5.42), the energy (5.26) can be written as

$$\Psi_{m,ex}=\rho_0\psi_{m,ex}=-\tfrac{1}{2}\beta[(\mathbf{M}^A\cdot\mathbf{n}_0)^2+(\mathbf{M}^B\cdot\mathbf{n}_0)^2]+\delta(\mathbf{M}^A\cdot\mathbf{M}^B)$$
$$-\beta'(\mathbf{M}^A\cdot\mathbf{n}_0)(\mathbf{M}^B\cdot\mathbf{n}_0)+d(\mathbf{M}^A\times\mathbf{M}^B)\cdot\mathbf{n}_0, \tag{5.43}$$

where we have set $\beta\equiv\beta_1^A=\beta_2^B$, $\delta\equiv\beta_2'$, and have discarded the terms proportional to $(\mathbf{M}^A)^2$ and $(\mathbf{M}^B)^2$ since they yield pure constants in the case of saturated magnetic sublattices. It remains four material constants: β and β' are the magnetic anisotropy constants. The antiferromagnet is said to be of the "easy axis" type if $\beta-\beta'>0$.[29] The constant δ accounts for the interaction between magnetic sublattices that do not arise from disuniformities in these lattices. The constant d, which is of the same order as the anisotropy constants, is the constant of *weak ferromagnetism*. Indeed, if d differs from zero, then, as has been shown in the original microscopic models of Dzyaloshinskii[30] and Moriya,[31] the presence of the last contribution in Eq. (5.43) may in fact lead to the phenomenon of weak ferromagnetism exhibited, for instance, by CrF_3.[31]

Collecting the expressions (5.25)–(5.29) on account of the representations (5.33), (5.37), and (5.42), and of the notation (5.34), where ρ_0 is assumed to be uniform throughout the body in its initial configuration, we obtain the expression of the free energy for an *elastically isotropic antiferromagnet with weak ferromagnetism and a magnetic easy axis* $\mathbf{n}_0$:

$$\begin{aligned}\Psi\equiv\rho_0\psi=\Psi_0-\rho_0\eta_0\tilde\theta-\frac{\rho_0\gamma}{2\theta^0}\tilde\theta^2+(\tfrac{1}{2}\lambda_1e_{kk}-\nu\tilde\theta+\delta\gamma^s\mathbf{M}^A\cdot\mathbf{M}^B)e_{jj}\\ +\lambda_2e_{ij}e_{ij}-\tfrac{1}{2}\beta[(\mathbf{M}^A\cdot\mathbf{n}_0)^2+(\mathbf{M}^B\cdot\mathbf{n}_0)^2]\\ -\beta'(\mathbf{M}^A\cdot\mathbf{n}_0)(\mathbf{M}^B\cdot\mathbf{n}_0)\\ +\delta\mathbf{M}^A\cdot\mathbf{M}^B+d(\mathbf{M}^A\times\mathbf{M}^B)\cdot\mathbf{n}_0\\ +\tfrac{1}{2}\alpha\left[\left(\frac{\partial\mathbf{M}^A}{\partial x_i}\right)^2+\left(\frac{\partial\mathbf{M}^B}{\partial x_i}\right)^2\right]+\alpha'\frac{\partial\mathbf{M}^A}{\partial x_i}\cdot\frac{\partial\mathbf{M}^B}{\partial x_i},\end{aligned} \tag{5.44}$$

where we have set $\gamma^s\equiv d_1^{(AB)}/\delta$, the remaining magnetostriction constant. Except for the thermal, thermoelastic, and "weak ferromagnetism" terms, this expression coincides with that postulated by Bar'yakhtar *et al.*[32] On account of Eqs. (5.44) and (5.36), the constitutive equations (5.19)–(5.23) and (5.31) read, for $d=0$ (i.e., in absence of weak ferromagnetism):

$$\begin{aligned}{}^Rt_{ij}=(\lambda_1e_{kk}-\nu\tilde\theta+\delta\gamma^s\mathbf{M}^A\cdot\mathbf{M}^B)\delta_{ij}+2\lambda_2e_{ij}\\ +M_i^A\{\delta(1+\gamma^se_{kk})M_j^B-[\beta(\mathbf{M}^A\cdot\mathbf{n}_0)+\beta'(\mathbf{M}^B\cdot\mathbf{n}_0)]n_{0j}\}\\ +M_i^B\{\delta(1+\gamma^se_{kk})M_j^A-[\beta(\mathbf{M}^B\cdot\mathbf{n}_0)+\beta'(\mathbf{M}^A\cdot\mathbf{n}_0)]n_{0j},\end{aligned} \tag{5.45}$$

$${}^{RL}B_{(A)i}=[\beta(\mathbf{M}^A\cdot\mathbf{n}_0)+\beta'(\mathbf{M}^B\cdot\mathbf{n}_0)]n_{0i}-\delta(1+\gamma^se_{kk})M_i^B, \tag{5.46a}$$

$${}^{RL}B_{(B)i}=[\beta(\mathbf{M}^B\cdot\mathbf{n}_0)+\beta'(\mathbf{M}^A\cdot\mathbf{n}_0)]n_{0i}-\delta(1+\gamma^se_{kk})M_i^A, \tag{5.46b}$$

$${}^R\mathcal{B}_{(A)ij}=\rho_0(\alpha M_{i,j}^A+\alpha' M_{i,j}^B), \tag{5.47a}$$

$${}^R\mathcal{B}_{(B)ij}=\rho_0(\alpha M_{i,j}^B+\alpha' M_{i,j}^A), \tag{5.47b}$$

$$\eta=\eta_0+\frac{\gamma\tilde\theta}{\theta^0}+\rho_0^{-1}\nu e_{kk}, \tag{5.48}$$

$$q_i=-K\theta_{,i}. \tag{5.49}$$

Let $M_S=|\mathbf{M}^A|\approx|\mathbf{M}^B|$ be a typical magnitude of a sublattice magnetic moment per unit volume and c_E a typical elastic wave velocity. Then a typical nondimensional parameter useful in studying coupled (via magnetostrictive and ponderomotive effects) magnetoelastic waves will be $\xi=\gamma^s(M_S^2\delta/\rho_0c_E^2)^{1/2}$.[32]

In conclusion of this point it must be noticed that, for the sake of example, we have considered a different material symmetry for each class of effects in order to write the energy (5.44). A more coherent scheme considering only one symmetry (such as cubic symmetry or transverse isotropy) can be formulated without difficulty.[33] As to the fields t_{ij}, ${}^{L}\mathbf{B}_{A}$, ${}^{L}\mathbf{B}_{B}$, $\mathcal{B}_{(A)}$ and $\mathcal{B}_{(B)}$, only the thermodynamical recoverable parts (indiced R) are given by Eqs. (5.45)—(5.47). The next section is devoted to constructing the dissipative parts, which yield viscosity and spin-relaxation phenomena, the latter playing an important role in the damping of coupled magnetoelastic waves, especially in the crossover regions of the dispersion diagram.

6. DISSIPATIVE PROCESSES

6.1. General dissipative processes

Consider the case of an antiferromagnetic deformable heat and electricity conductor whose magnetic structure is made of two magnetic sublattices $\alpha = A, B$. Then the fundamental principle that governs the general thermodynamical processes is Eq. (3.12) in which is added the Joule contribution. That is,[34]

$$-\rho(\dot{\psi}+\eta\dot{\theta})+\sigma_{ij}D_{ij}-\rho({}^{L}\mathbf{B}_{(A)}\cdot\hat{\mathbf{m}}_{(A)}+{}^{L}\mathbf{B}_{(B)}\cdot\hat{\mathbf{m}}_{(B)})$$
$$+(\mathcal{B}_{(A)ij}\hat{\mathfrak{M}}_{(A)ij}+\mathcal{B}_{(B)ij}\hat{\mathfrak{M}}_{(B)ij})+\mathcal{J}\cdot\mathcal{E}-\theta^{-1}\mathbf{q}\cdot\nabla\theta\geq 0. \qquad (6.1)$$

Instead of dealing with this general inequality and considering a nonlinear theory of irreversible processes, we make the following simplifying assumptions: (i) the fields σ_{ij}, ${}^{L}\mathbf{B}_{(\alpha)}$ and $\mathcal{B}_{(\alpha)ij}$ present *additive* thermodynamically reversible and irreversible contributions (the latters indiced D on the left), such that $(\alpha = A, B)$

$$\sigma_{ij}={}^{R}\sigma_{ij}+{}^{D}\sigma_{ij},\quad {}^{L}\mathbf{B}_{(\alpha)}={}^{RL}\mathbf{B}_{(\alpha)}+{}^{DL}\mathbf{B}_{(\alpha)},$$
$$\mathcal{B}_{(\alpha)}={}^{R}\mathcal{B}_{(\alpha)}+{}^{D}\mathcal{B}_{(\alpha)}, \qquad (6.2)$$

where the recoverable contributions and η are derivable from the potential ψ, and have expressions of the type of those derived in previous sections; (ii) the physical significance of each dissipative force is directly related to the interpretation of the interactions represented in a phenomenological manner by the different internal forces (see Part I). Thus, ${}^{D}\sigma_{ij}$ gives rise to viscosity, ${}^{DL}\mathbf{B}_{(A)}$ and ${}^{DL}\mathbf{B}_{(B)}$ represent the transport phenomena associated with, firstly, the interactions between the two magnetic sublattices and the crystal lattice and, secondly, the intermagnetic sublattice interactions that do not result from disuniformities in the magnetic sublattices. Although they theoretically represent the transport phenomena associated with the spin— spin interactions arising from the disuniformities, no microscopic basis can be found, for the time being, for the effects represented by ${}^{D}\mathcal{B}_{(A)}$ and ${}^{D}\mathcal{B}_{(B)}$, so that we shall set these last two fields equal to zero; (iii) We consider a partial uncoupling of the different transport phenomena and use the Onsager—Casimir linear theory of irreversible processes. Then, on account of the fact that,

$$\rho\dot{\psi}=-\rho\eta\dot{\theta}+{}^{R}\sigma_{ij}D_{ij}-\sum_{\alpha}(\rho\,{}^{RL}\mathbf{B}_{(\alpha)}\cdot\hat{\mathbf{m}}_{(\alpha)}-{}^{R}\mathcal{B}_{(\alpha)ij}\hat{\mathfrak{M}}_{(\alpha)ij}), \qquad (6.3)$$

and of Eq. (6.1), the remaining dissipative contributions must satisfy the *dissipation inequality*:

$$\Phi\equiv[{}^{D}\sigma_{ij}(\theta^0,D_{kl})D_{ij}]-\rho_0[{}^{DL}\mathbf{B}_{(A)}(\theta^0,\hat{\mathbf{m}}_{(\alpha)})\cdot\hat{\mathbf{m}}_{(A)}$$
$$+{}^{DL}\mathbf{B}_{(B)}(\theta^0,\hat{\mathbf{m}}_{(\alpha)})\cdot\hat{\mathbf{m}}_{(B)}]+[\mathcal{J}(\theta^0,\mathcal{E},\nabla\theta)\cdot\mathcal{E}$$
$$-\theta^{-1}\mathbf{q}(\theta^0,\mathcal{E},\nabla\theta)\cdot\nabla\theta]\geq 0. \qquad (6.4)$$

In the linear theory of irreversible processes for an *isotropic* medium, we thus have[35]

$${}^{D}\sigma_{ij}=\eta_1(\theta^0)D_{kk}\delta_{ij}+2\eta_2(\theta^0)D_{ij}, \qquad (6.5)$$

$${}^{DL}\mathbf{B}_{(A)}=-\rho_0[\tau_1(\theta^0)\hat{\mathbf{m}}_{(A)}+\tau_{12}(\theta^0)\hat{\mathbf{m}}_{(B)}], \qquad (6.6a)$$

$${}^{DL}\mathbf{B}_{(B)}=-\rho_0[\tau_2(\theta^0)\hat{\mathbf{m}}_{(B)}+\tau_{12}(\theta^0)\hat{\mathbf{m}}_{(A)}], \qquad (6.6b)$$

$$\mathcal{J}=\sigma(\theta^0)\mathcal{E}+K_1(\theta^0)(\theta^0)^{-1}\nabla\theta, \qquad (6.7)$$

$$\mathbf{q}=-K(\theta^0)\nabla\theta-K_1(\theta^0)\mathcal{E}, \qquad (6.8)$$

where η_1 and η_2 are viscosities, τ_1, τ_2, and τ_{12} are relaxation times, σ is the electrical conductivity, K is the heat conductivity, and K_1 is the material constant allowing the representation of the Thomson and Peltier effects. The semi-positive definite character of Φ requires that these material constants satisfy the following inequalities:

$$3\eta_1+2\eta_2\geq 0,\quad \eta_2\geq 0,\quad \tau_1\geq 0,\quad \tau_1\tau_2-\tau_{12}^2\geq 0,$$
$$\sigma\geq 0,\quad \sigma K\theta^0-K_1^2\geq 0. \qquad (6.9)$$

In agreement with the infinitesimal strain theory sketched out in the foregoing section it must be noticed that

$$\rho\approx\rho_0,\quad D_{ij}\approx\dot{u}_{(i,j)},\quad \mathcal{E}\approx\mathbf{E}+\frac{1}{c}\mathring{\mathbf{u}}\times\mathbf{B}. \qquad (6.10)$$

We shall focus our attention on the dissipative phenomena represented by Eqs. (6.6).

6.2. Spin-lattice relaxation

A. Strong damping

On account of the additive character of the decompositions (2.10), (2.11), and (6.2) with respect to the internal forces, we can write

$$t_{ij}={}^{R}t_{ij}+{}^{D}t_{ij}, \qquad (6.11)$$

$$\mathbf{B}^{\mathrm{eff}}_{(\alpha)}={}^{R}\mathbf{B}^{\mathrm{eff}}_{(\alpha)}+{}^{D}\mathbf{B}^{\mathrm{eff}}_{(\alpha)}, \qquad (6.12)$$

where

$${}^{D}t_{ij}={}^{D}\sigma_{ij}+\sum_{\alpha=A,B}(\rho\,{}^{DL}B_{(\alpha)[i}\mu_{(\alpha)j]}), \qquad (6.13)$$

$${}^{R}B^{\mathrm{eff}}_{(\alpha)i}\equiv B_i+{}^{RL}B_{(\alpha)i}+\rho^{-1}\,{}^{R}\mathcal{B}_{(\alpha)ij,j}, \qquad (6.14)$$

$${}^{D}\mathbf{B}^{\mathrm{eff}}_{(\alpha)}\equiv{}^{DL}\mathbf{B}_{(\alpha)}. \qquad (6.15)$$

Equation (6.13) shows that the dissipative fields ${}^{DL}\mathbf{B}_{(\alpha)}$ contribute to the dissipative stresses. As a result of the decomposition (6.12) the spin precession equations (2.4) can be rewritten in the following form $(\alpha = A, B)$:

$$\dot{\mu}_{(\alpha)}=-\gamma^{R}_{\alpha}\mathbf{B}^{\mathrm{eff}}_{(\alpha)}\times\mu_{(\alpha)}+\mathbf{R}_{(\alpha)}, \qquad (6.16)$$

whereas the Cauchy equation (2.2) can be written as $(\mathbf{M}=\mathbf{M}^A+\mathbf{M}^B)$

$$\rho\dot{\mathbf{U}}=\mathrm{div}^{R}\mathbf{t}+{}^{v}\mathbf{f}-\sum_{\alpha=A,B}\left(\nabla\times\frac{\rho\mathbf{R}_{(\alpha)}}{2\gamma_\alpha}\right)+\mathbf{f}+(\nabla\mathbf{B})\cdot\mathbf{M}, \qquad (6.17)$$

where the relaxation terms $\mathbf{R}_{(\alpha)}$ are defined by

$$\mathbf{R}_{(\alpha)} = \gamma_\alpha \mu_{(\alpha)} \times {}^{DL}\mathbf{B}_{(\alpha)}, \quad (\alpha = A, B), \tag{6.18}$$

and the viscous force ${}^v\mathbf{f}$ by

$${}^v\mathbf{f} = \operatorname{div} {}^D\boldsymbol{\sigma}. \tag{6.19}$$

The transformation (6.17) of the Cauchy equation is obtained by noting that, after Eq. (6.18),

$$(\nabla \times \rho \mathbf{R}_{(\alpha)})_i = -2\gamma_\alpha (\rho\, {}^{DL}B_{(\alpha)[i}\mu_{(\alpha)j]})_{,j}. \tag{6.20}$$

In the above-stated equations the constitutive equations of the fields ${}^R t_{ij}$, ${}^{RL}\mathbf{B}_{(\alpha)}$ and ${}^R\mathcal{B}_{(\alpha)}$ are those obtained in Sec. 4 or 5. Note that no hypothesis has been made concerning the magnitude of the constants τ_1, τ_2, and τ_{12}, so that the terms $\mathbf{R}_{(\alpha)}$ correspond to spin relaxation with a possibly *strong* damping. The expressions (6.18) can be made more specific by assuming, first, that $\tau_1 = \tau_2 \equiv \tau$, since the spin—crystal lattice interactions are of the same type for both sublattices. Then, with τ and τ_{12} positive, the fourth of Eqs. (6.9) requires that $\tau_{12} \leq \tau$. Next, in the infinitesimal strain theory, we can define the vorticity vector by

$$\tilde{\Omega}_i = -\tfrac{1}{2}\epsilon_{ijk}\Omega_{jk} = \tfrac{1}{2}(\nabla \times \dot{\mathbf{u}})_i, \tag{6.21}$$

where, from hereon, the superimposed dot indicates the partial time derivative. Then, on account of Eqs. (5.34) and (6.6), Eqs. (6.18) take the forms:

$$\begin{aligned}\mathcal{R}_{(A)} \equiv \rho_0 \mathbf{R}_{(A)} = -\gamma_A \mathbf{M}^A \times [\tau(\dot{\mathbf{M}}^A + \mathbf{M}^A \times \tilde{\mathbf{\Omega}}) \\ + \tau_{AB}(\mathring{\mathbf{M}}^B + \mathbf{M}^B \times \tilde{\mathbf{\Omega}})],\end{aligned} \tag{6.22a}$$

$$\begin{aligned}\mathcal{R}_{(B)} \equiv \rho_0 \mathbf{R}_{(B)} = -\gamma_B \mathbf{M}^B \times [\tau(\dot{\mathbf{M}}^B + \mathbf{M}^B \times \tilde{\mathbf{\Omega}}) \\ + \tau_{AB}(\mathring{\mathbf{M}}^A + \mathbf{M}^A \times \tilde{\mathbf{\Omega}})],\end{aligned} \tag{6.22b}$$

where $\tau_{AB} \equiv \tau_{12}$. By the same token the Cauchy equation (6.17) takes the form

$$\rho_0 \ddot{\mathbf{u}} = \operatorname{div} {}^R\mathbf{t} + {}^v\mathbf{f} - \sum_{\alpha = A, B}\left(\nabla \times \frac{\mathcal{R}_{(\alpha)}}{2\gamma_\alpha}\right) + \mathbf{f} + (\nabla \mathbf{B}) \cdot \mathbf{M}, \tag{6.23}$$

whereas Eqs. (6.16) read $(\alpha = A, B)$

$$\mathring{\mathbf{M}}^\alpha = \gamma_\alpha \mathbf{M}^\alpha \times {}^R\mathbf{B}^{\mathrm{eff}}_{(\alpha)} + \mathcal{R}_{(\alpha)}. \tag{6.24}$$

The relaxation terms defined by Eqs. (6.22) generalize to the case of *deformable* antiferromagnets the spin—lattice relaxation term that we have proposed earlier[4] in deformable ferromagnets to account for a possible *strong* damping of this spin relaxation. They are of the type of the relaxation term proposed by Gilbert[6] in rigid ferromagnets, but with the supplementary effects due to deformations. Whereas the rate of strain partipates in the classical viscosity processes—cf. Eq. (6.5)—the rate of rotation (vorticity) participates in the relaxation of the magnetic sublattices. This shows the interest of using the Jaumann derivative to define an objective time rate of the magnetic sublattices in deformable media. The presence of these relaxation terms in the Cauchy equation (6.23) shows that, especially in the crossover regions of the dispersion diagram of coupled magnetoelastic waves, the damping of elastic waves may partially be caused by the spin—lattice relaxation. On account of the constitutive equations (5.45)—(5.47), and (6.5), the Eqs. (6.23) and (6.24), along with the expressions (6.22), allow a complete study of *damped* magnetoelastic waves in an infinite linear elastic antiferromagnet in the presence of dissipative phenomena resulting from viscosity and spin relaxation (the latter with strong damping).

B. Weak damping

Let us assume that the constants τ and τ_{AB} are infinitesimally small of the first order: $O(\tau) = O(\tau_{AB}) = \epsilon$. Then the contribution $\mathcal{R}_{(\alpha)}$ in the right-hand side of Eqs. (6.24) may be considered as a perturbation. The fields $\mathring{\mathbf{M}}^A$ and $\mathring{\mathbf{M}}^B$ that it contains can be evaluated from the spin precession equations (6.24) in absence of relaxation. That is, up to terms of the order of ϵ^2, we can write

$$\begin{aligned}\mathcal{R}_{(A)} = &-(\tau_A' M_S^2)^{-1}\mathbf{M}^A \times [\mathbf{M}^A \times ({}^R\mathbf{B}^{\mathrm{eff}}_{(A)} + \gamma_A^{-1}\tilde{\mathbf{\Omega}})] \\ &-(\tau_{AB}'' M_S^2)^{-1}\mathbf{M}^A \times [\mathbf{M}^B \times ({}^R B^{\mathrm{eff}}_{(B)} + \gamma_B^{-1}\tilde{\mathbf{\Omega}})], \\ \mathcal{R}_{(B)} = &-(\tau_B' M_S^2)^{-1}\mathbf{M}^B \times [\mathbf{M}^B \times ({}^R\mathbf{B}^{\mathrm{eff}}_{(B)} + \gamma_B^{-1}\tilde{\mathbf{\Omega}})] \\ &-(\tau_{AB}'' M_S^2)^{-1}\mathbf{M}^B \times [\mathbf{M}^A \times ({}^R\mathbf{B}^{\mathrm{eff}}_{(A)} + \gamma_{(A)}^{-1}\tilde{\mathbf{\Omega}})],\end{aligned} \tag{6.25}$$

where we have introduced the new relaxation times

$$\tau_A' \equiv (\gamma_A^2 M_S^2 \tau)^{-1}, \quad \tau_B' \equiv (\gamma_B^2 M_S^2 \tau)^{-1}, \quad \tau_{AB}'' \equiv (\gamma_A \gamma_B M_S^2 \tau_{AB})^{-1}, \tag{6.26}$$

and a typical value M_S of the magnetization per unit volume. Equations (6.25) generalize to the case of *deformable antiferromagnets*, the relaxation term considered by Akhiezer *et al.*[36] in elastic ferromagnets, which term itself generalized the pioneering proposal of Landau and Lifshitz[7,37] for rigid ferromagnets. However, the relaxation terms (6.25) are here obtained at the approximation of weak damping of the magnetic sublattice oscillations. Thus, in a certain sense, Eqs. (6.22) provide relaxation terms valid in a wider range of damping, as is corroborated in rigid ferromagnets by studies based on statistical mechanics[38] and by experiments.[39] In applying the Eqs. (6.25), it may be assumed without too much loss that $\gamma_A = \gamma_B$, so that $\tau_A' = \tau_B'$.

To conclude this section on dissipative processes we note that, by using Eq. (3.11) and the results of Secs. 5 and 6, there is no difficulty in establishing the heat conduction equation which follows from Eq. (3.10) for linear elastic antiferromagnetic heat and electricity conductors in the presence of viscosity and spin relaxation (for both strong and weak dampings).

7. RIGID STATIONARY FERRIMAGNETS

The results of Secs. 5 and 6 are readily specialized to the case of rigid stationary ferrimagnets, for which we need consider only Maxwell's equations and the spin precession equations with $\alpha = 1, 2, \ldots, n$. Then, assuming an isotropic spin—lattice relaxation and defining $\Psi = \rho_0 \psi$ and the total energy

$$W = \Psi(\mathbf{M}_{(\alpha)}, \nabla \mathbf{M}_{(\alpha)}, \theta) - \mathbf{M} \cdot (\mathbf{B} - \mathcal{N} \cdot \mathbf{M}), \tag{7.1}$$

where $\mathcal{N}$ is the demagnetization tensor whose explicit form depends on the shape of the *finite* specimen and $\mathbf{M} = \sum_\alpha \mathbf{M}_{(\alpha)}$, Eqs. (6.16) take the following form for *strong* damping:

$$\dot{\mathbf{M}}_{(\alpha)} = -\gamma_\alpha \mathbf{M}_{(\alpha)} \times \left(\frac{\delta W}{\delta \mathbf{M}_{(\alpha)}} + \sum_\beta \tau_{\alpha\beta} \dot{\mathbf{M}}_\beta\right), \tag{7.2}$$

where $\delta/\delta\mathbf{M}_{(\alpha)}$ indicates the Euler—Lagrange functional derivative, and $\tau_{\alpha\beta}$ are the $n(n+1)/2$ independent relaxation times which obey the dissipation inequality

$$\sum_{\alpha,\beta} \tau_{\alpha\beta}\dot{\mathbf{M}}_{(\alpha)}\cdot\dot{\mathbf{M}}_{(\beta)} \geq 0. \tag{7.3}$$

For weak damping, applying the same perturbation procedure as in Sec. 6.2, Eq. (7.2) is replaced by

$$\dot{\mathbf{M}}_{(\alpha)} = -\gamma_\alpha \tilde{\mathbf{B}}^{eff}_{(\alpha)} \times \mathbf{M}_{(\alpha)}, \tag{7.4}$$

where

$$\tilde{B}^{eff}_{(\alpha)i} = -\sum_{\beta=1}^{n} (\delta_{ij}\delta_{\alpha\beta} + \tau_{\alpha\beta}\gamma_\beta \epsilon_{ijk} M_{(\beta)k}) \frac{\delta W}{\delta M_{(\beta)j}}. \tag{7.5}$$

Equation (7.2) generalizes to ferrimagnets the equation proposed by Gilbert[40] in ferromagnets, whereas Eq. (7.4) generalizes the Landau–Lishitz equations. It must be noted at this point that in most treatments essentially ferrimagnetic multi-sublattice effects concerning relaxation have so far been ignored.[41] For *small* damping, it may further be assumed that the combined effect of the various sublattices can be expressed in terms of a suitable averaged damping term acting on the resultant magnetization **M**. This is what happens here if one sets $\tau_{\alpha\beta} = \tau > 0$ for any α and β. Then Eq. (7.2) yields

$$\dot{\mathbf{M}}_{(\alpha)} = -\gamma_\alpha \mathbf{M}_{(\alpha)} \times \left(\frac{\delta W}{\delta \mathbf{M}_{(\alpha)}} + \tau\dot{\mathbf{M}}\right). \tag{7.6}$$

It is expected that for large damping the description provided by Eq. (7.2) or (7.6) will be more adapted to the physical reality than Eq. (7.4).

8. CONCLUSION

By way of conclusion we specify the range of applicability of the various equations obtained in this work. As already pointed out the equations obtained in Sec. 5 can be applied to the study of coupled magnetoelastic waves in elastic antiferromagnets, possibly endowed with the property of weak ferromagnetism. This study is particularly important in the frequency range where the Magnon–phonon interactions may occur because of the potential use of the conversion of energy thus allowed. In this case the quasimagnetostatic fields can be used without too much loss. In particular, the dissipative phenomena of interest then are only viscosity and spin relaxation with strong or weak damping, which supports the interest for the development of Sec. 6. The electric field then is ignored. The situation is quite different in the frequency range of optical phenomena or if one is interested in the coupling between spin waves and electromagnetic waves. Then the following alterations must be made. The fully dynamical Maxwell's equations must be considered in lieu of Eqs. (2.6)–(2.7), and in the case of an electrical conductor (e.g., in rare earth metals and alloys) one must consider the conduction law (6.7), taking account of the Thomson and Peltier effects if such coupling effects are exhibited by the antiferromagnetic medium. Finally, the electromagnetic momentum must be accounted for in Cauchy's equations of motion in computing the pondermotive force, so that a Lorentz term $(1/c)\mathcal{J}\times\mathbf{B}$ will appear in these equations.[42] The discussion above pertains to the case of a material which is free of stress and magnetization in its initial state. A more involved problem consists in considering perturbations on an initial state defined by an initial stress field and a finite static state of magnetization. In the latter case, the equations governing the perturbing fields superimposed on the bias fields can be deduced from the exact nonlinear equations given in Sec. 2.1 and Sec. 4, according to a scheme similar to that used by other authors[43] in different circumstances. This will be the concern of further works.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Dr. T.L. Gilbert of the Argonne National Laboratory, who made a copy of his unpublished dissertation available to him.

[1]The reader is referred to the Introduction of Part I [G.A. Maugin, J. Math. Phys. **17**, 1727 (1976)] for a general introduction.
[2]B.D. Coleman, Arch. Ration. Mech. Anal. 17, 1 (1964).
[3]Cf. S.R. DeGroot and P. Mazur, *Nonequilibrium Thermodynamics* (North-Holland, Amsterdam, 1962).
[4]G.A. Maugin, J. Phys. A. 5, 1550 (1972).
[5]G.A. Maugin, Physica A (Holland) **81**, 454 (1975).
[6]T.L. Gilbert, Phys. Rev. **100**, 1245 (1955).
[7]L.D. Landau and E.M. Lifshitz, Phys. Z. Sowjetunion 8, 153 (1935).
[8]The notation is that used in Part I. That is, the Cartesian tensor notation (with summation over dummy small and capital Latin indices) is used throughout. The dyadic notation is used where there is no ambiguity. Equations established in Part I are referred to by I followed by their number, e.g., Eq. (I.3.24).
[9]Compare G.T. Rado and J.R. Weertman, J. Phys. Chem. Solids **11**, 315 (1958); R.F. Soohoo, J. Appl. Phys. Suppl. 32, 1485 (1959).
[10]See Part I, Sec. 3.1.
[11]Compare the dynamical equations for ferromagnets in the paper by G.A. Maugin and B. Collet in C.R. Acad. Sci. B (Paris) **279**, 439 (1974).
[12]See, for instance, A.C. Eringen, "Basic Principles," in *Continuum Physics*, edited by A.C. Eringen (Academic, New York, 1975), Vol. II, pp. 1–127.
[13]Cf. Ref. 11.
[14]C. Truesdell, *Rational Thermodynamics* (McGraw-Hill, New York, 1969), p. 38.
[15]C. Truesdell and W. Noll, in *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1965), Vol. III/3, Sec. 43.
[16]See Part I for the definition of V_{obj}.
[17]Cf. Ref. 2.
[18]C.C. Wang, Arch. Ration. Mech. Anal. 36, 166 (1970).
[19]A.J.M. Spencer, "Theory of Invariants" in *Continuum Physics*, edited by A.C. Eringen (Academic, New York, 1971), Vol. I, p. 239–353.
[20]See, for instance, R.R. Birss, Rep. Prog. Phys. 26, 307 (1963).
[21]See, for instance, A.C. Eringen, *Mechanics of Continua* (Wiley, New York, 1967), p. 444.
[22]Cf. W.F. Brown, *Magnetoelastic Interactions* (Springer, New York, 1966), pp. 90–92.
[23]Cf. W. Strauss, "Magnetoelastic Properties of Yttrium-Iron-Garnet" in *Physical Acoustics*, edited by W.P. Mason (Academic, New York, 1968), Vol. 4, Part B, pp. 221–68.
[24]G. Racah, Rend. Accad. Lincei (6) **17**, 386 (1933).
[25]E.g., G.A. Maugin and A.C. Eringen, J. Math. Phys. 13, 1334 (1972).
[26]G.F. Smith and R.S. Rivlin, Q. J. Appl. Math. 15, 308 (1957).
[27]Cf. Ref. 19, p. 294.
[28]Compare Eq. (4.17) and see Ref. 19, p. 302.

[29]A. I. Akhiezer, V. G. Bar'Yakhtar, and S. V. Peletminskii, *Spin Waves* [translation from the Russian] (North-Holland, Amsterdam, 1968), Sec. 7.2.
[30]E. Dzyaloshinskii, Zh. Eksp. Teor. Fiz. 32, 1547 (1957) [Sov. Phys. JETP 5, 1259 (1957)]; J. Phys. Chem. Solids 4, 241 (1958).
[31]T. Moriya, Phys. Rev. 120, 91 (1960).
[32]Compare V. G. Bar'Yakhtar, M. A. Savchenko, and V. V. Tarasenko, Zh. Eksp. Teor. Fiz. 49, 944 (1965) [Sov. Phys. JETP **22**, 657 (1966)].
[33]Compare the ferromagnetic case in Ref. 25.
[34]Compare Eq. (20), valid for moving polarized and deformable ferromagnets, in Ref. 11.
[35]Compare the ferromagnetic case in insulators in Ref. 5.
[36]See Ref. 29, Chap. 4.
[37]For the experimental evidence of the Landau and Lifshitz damping see D. S. Robdell, Phys. Rev. Lett. 13, 471 (1964).
[38]R. Kubo and N. Hashitsume, Suppl. Prog. Theor. Phys. (Japan) 46, 210 (1970).
[39]The superiority of Gilbert's term is shown in V. Kambersky and C. E. Patton, Phys. Rev. B **11**, 2668 (1975); see also J. C. Anderson, *Magnetism and Magnetic Materials* (Chapman, London, 1968), p. 180.
[40]T. L. Gilbert, Ph. D. thesis, I. I. T. Chicago (1956 unpublished).
[41]See, however, R. K. Wangsness, Phys. Rev. 105, 390 (1957).
[42]Cf. B. Collet and G. A. Maugin, C. R. Acad. Sci. B (Paris) 279, 379 (1974).
[43]J. C. Baumhauer and H. F. Tiersten, J. Acoust. Soc. Am. 54, 1017 (1973); Ph. Boulanger, G. Mayne, and R. Van Geen, Int. J. Solids Struc. 9, 1439 (1973).

Inverse wave propagation in an inhomogeneous waveguide

John A. DeSanto

Naval Research Laboratory, Washington, D. C.
*and Admiralty Research Laboratory, Teddington, England**
(Received 27 February 1976)

A solution is given for the problem of inverse propagation of a scalar wave in inhomogeneous rectangular two-dimensional waveguide. The sound speed is assumed to vary in depth and inverse propagation means the calculation of the field at range x_1 in terms of the field at range x_2 where $x_2 > x_1$. The method is analogous to that used by Wolf, Shewell, and Lalor for the inverse diffraction problem in a homogeneous half-space. It is found that the field at x_1 can be expressed in terms of two integrals over the field at x_2. The kernel of the first integral is bounded and expresses physically the result at x_1 of the waves at x_2 reversing their direction of propagation and decay, i.e., they now propagate and decay in the direction of x_1. A reciprocity relation for this term is possible. The kernel of the second integral is singular and expresses the mathematical fact of the residual effect of the evanescent waves at x_1 as they reverse their direction at x_2 and now grow exponentially. Consequences of the neglect of this singular term are discussed.

INTRODUCTION

Sometime ago, Wolf and Shewell[1] and Lalor[2] discussed the solution of the inverse diffraction problem in a homogeneous half-space. Simply, one has a field propagating into a half-space $z > 0$, and assumes the field is known on some plane $z = z_2$. The problem is then to find the field on the plane $z = z_1$, where $z_1 < z_2$. For example, one might wish to calculate the "near" field from the "far" field. The result is expressed as the inverse of one of the Rayleigh diffraction formulas. The kernel of the inversion contains two terms, one of which is singular. Methods for handling the singular term are discussed.

In this paper we briefly present a similar analysis with the problem being the calculation of the inverse field in a two-dimensional rectangular waveguide. Here, in addition, the waveguide is assumed to be inhomogeneous in the sense that the sound speed is a function of depth.

In Sec. 1 we present the basic analysis and express the field at $x_1 < x_2$ as a sum of two terms, each of which is an integral over the field at x_2. The kernel of the first integral is bounded and the term describes that part of the field at x_1 due to waves at x_2 reversing their direction of propagation and decay. The kernel of the second integral is singular and the term describes exponentially growing waves at x_1 due to evanescent waves at x_2 which grow towards x_1. In Sec. 2 the reciprocity relation of the first term is derived, and in Sec. 3 a brief discussion is given of the consequences of neglect of the singular term.

1. GENERAL FORMALISM

In two dimensions the propagation of sound is governed by the Helmholtz equation

$$\phi_{xx} + \phi_{zz} + k^2\eta^2(z)\phi(x, z) = 0 \tag{1}$$

for the velocity potential field ϕ.[3] Here, $\eta(z)$, the index of refraction, is proportional to the inverse of $c(z)$, the sound speed, and $k = 2\pi/\lambda$ is the wavenumber with λ the wavelength. Since c is a function of depth the equation is said to be inhomogeneous. The general problem of sound propagation involves the solution of (1) assuming that ϕ satisfies appropriate boundary conditions. Here we first wish to solve (1) in the region $0 \le z \le D$ and $0 \le x < \infty$ (see Fig. 1), where ϕ satisfies boundary conditions at $z = 0$, D, and $x = 0$, and an outgoing radiation condition as $x \to \infty$. Then we will assume that the field is known on a (far) plane $x = x_2$ and express the field on a (near) plane $x = x_1 < x_2$ in terms of the field on x_2.

The solution of (1) is separable and can be written in terms of an infinite discrete spectral representation

$$\phi(x, z) = \sum_{j=0}^{\infty} A_j \psi_j(z) \exp(ikm_j x), \tag{2}$$

where the eigenfunctions ψ_j satisfy the ordinary differential equation

$$\psi_j'' + k^2[\mu_j - q(z)]\psi_j = 0 \tag{3}$$

with

$$q(z) = 1 - \eta^2(z) \tag{4}$$

and

$$m_j = \begin{cases} (1-\mu_j)^{1/2}, & 0 < \mu_j \le 1, \\ +i(\mu_j - 1)^{1/2}, & \mu_j > 1. \end{cases} \tag{5}$$

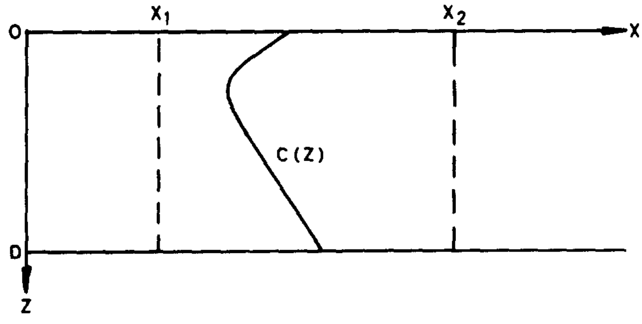

FIG. 1. Inverse propagation in a rectangular two-dimensional waveguide. The sound speed c is a function of depth z. The field is assumed known on the plane $x = x_2$ and the problem is to calculate it on the plane $x = x_1$. Direct propagation proceeds from x_1 to x_2.

The boundary conditions at $z=0$ and D (which we do not specify) yield specific forms for the ψ_j and the discrete eigenvalues μ_j, which we assume, for simplicity, are confined to the positive real axis in the j plane. The choice of branch in (5) is to ensure outgoing or decaying waves as $x\to\infty$. In addition we assume the ψ_j are orthonormal,

$$\int_0^D \psi_j(z)\psi_m(z)\,dz=\delta_{jm}=\begin{cases}1, & j=m,\\ 0, & j\neq m.\end{cases} \tag{6}$$

Multiplying (2) by $\psi_l(z)$, integrating over z from 0 to D and using (6) yields

$$A_j=\exp(-ik\,m_jx)\int_0^D \phi(x,z)\psi_j(z)\,dz. \tag{7}$$

Now let $x=x_1$ and $z=z_1$ in (2), $x=x_2$ and $z=z_2$ in (7), and substitute the resulting (7) into (2) to get

$$\phi(x_1,z_1)=\sum_{j=0}^{\infty}\psi_j(z_1)\exp[ikm_j(x_1-x_2)]\int_0^D\psi_j(z_2)\phi(x_2,z_2)\,dz_2. \tag{8}$$

Next assume $x_1<x_2$ and split the sum in (8) into two parts defined by

$$\sum{}^{-}=\sum_{j=0}^{J},\quad \sum{}^{+}=\sum_{j=J+1}^{\infty}, \tag{9}$$

where $\mu_J<1$ and $\mu_{J+1}>1$. To the result, add and subtract the term

$$\sum{}^{+}\psi_j(z_1)\exp[-k(\mu_j-1)^{1/2}(x_2-x_1)]\int_0^D\psi_j(z_2)\phi(x_2,z_2)\,dz_2, \tag{10}$$

and rewrite the result as the sum of two terms

$$\phi(x_1,z_1)=\phi_1(x_1,z_1)+\phi_2(x_1,z_1), \tag{11}$$

where we define $(m=1,2)$

$$\phi_m(x_1,z_1)=\int K_m(x_1,z_1;x_2,z_2)\phi(x_2,z_2)\,dz_2 \tag{12}$$

with

$$\begin{aligned}K_1(x_1,z_1;\,x_2,z_2)&=\sum{}^{-}\psi_j(z_1)\psi_j(z_2)\exp[ikm_j(x_1-x_2)]\\&\quad+\sum{}^{+}\psi_j(z_1)\psi_j(z_2)\exp[-k(\mu_j-1)^{1/2}(x_2-x_1)]\\&=\sum_{j=0}^{\infty}\psi_j(z_1)\psi_j(z_2)\exp[-ikm_j^*(x_2-x_1)],\end{aligned} \tag{13}$$

where the $*$ is complex conjugation, and

$$\begin{aligned}K_2(x_1,z_1;\,x_2,z_2)&=\sum{}^{+}\psi_j(z_1)\psi_j(z_2)\exp[ikm_j(x_1-x_2)]\\&\quad-\sum{}^{+}\psi_j(z_1)\psi_j(z_2)\exp[-k(\mu_j-1)^{1/2}(x_2-x_1)]\\&\quad-\sum{}^{+}\psi_j(z_1)\psi_j(z_2)\sinh[k(\mu_j-1)^{1/2}(x_2-x_1)].\end{aligned} \tag{14}$$

Thus it is possible to write ϕ at (x_1,z_1) in terms of two integrals over ϕ at (x_2,z_2). The kernel of the first integral, K_1, is bounded and expresses physically the result at x_1 of the waves at x_2 reversing their direction of propagation and decay, i.e., they now propagate and decay in the direction of x_1. The kernel of the second integral, K_2, is singular since the summation in (14) goes to infinity, and the problem becomes ill-posed since a small change in the "initial" condition $\phi(x_2,z_2)$ could produce a large change in $\phi(x_1,z_1)$. This is the mathematical expression of the residual effect of the evanescent waves at x_2 as they reverse their direction and grow exponentially in the direction of x_1. The neglect of this latter term means neglect of large wavenumbers, short wavelength terms, and hence an inability to gather information on an obstacle or process with a characteristic length smaller than a certain amount. There is thus a lower bound on the size of obstacles which can be seen.

2. RECIPROCITY

It is possible to express the ϕ_1 term as the inverse of a diffraction formula analogous to one of the free-space Rayleigh diffraction formulas presented in the references. This is done as follows. The incoming wave Green's function $G^-(x,2;\,x',z')$ satisfies an equation similar to (1) with a delta function source term

$$G^-_{xx}+G^-_{zz}+k^2\eta^2(z)G^-=-\delta(x-x')\delta(z-z') \tag{15}$$

as well as the boundary conditions at $z=0$ and D which are satisfied by the eigenfunctions, and the asymptotic condition of an incoming wave. It can be written as

$$G^-(x,z;\,x',z')=\sum_{j=0}^{\infty}\psi_j(z)\psi_j(z')G_j^-(x,x'), \tag{16}$$

where G_j^- satisfies the differential equation

$$\left(\frac{d^2}{dx^2}+k^2(1-\mu_j)\right)G_j^-(x,x')=-\delta(x-x') \tag{17}$$

and can be written as

$$G_j^-(x,x')=(2ikm_j^*)^{-1}\exp(-ikm_j^*\,|x-x'|), \tag{18}$$

where the complex conjugate of m_j is used in the exponential to ensure that for $j>J$ the function is decaying towards x_1. From (13) it can be easily seen that

$$K_1(x_1,z_1;\,x_2,z_2)=-2\frac{\partial}{\partial x_2}G^-(x_1,z_1;\,x_2,z_2) \tag{19}$$

so that ϕ_1 by (12) can be written as the inverse of a diffraction formula.

3. SUMMARY

To use these results one must be able to neglect the singular term ϕ_2. Neglect of ϕ_2 means neglect of terms of the order of $k(\mu_{J+1}-1)^{1/2}$ and larger, i.e., high frequency terms. The term $k=\omega/c$, where c is some reference sound speed, e.g., the sound speed at the surface. This establishes a characteristic length $L=\lambda/2\pi(\mu_{J+1}-1)^{1/2}$ below which we cannot measure. The higher the frequency of sound the smaller the obstacles we can see, but high frequency sound is rapidly attenuated in many media anyway, so that neglect of ϕ_2 probably yields no worse results than are now available.

*The author's permanent address is NRL. This work was accomplished while the author was at ARL, and he is grateful for their hospitality.

[1]E. Wolf and J.R. Shewell, Phys. Lett. A **25**, 417 (1967); A **26**, 104 (1967); J. Opt. Soc. Am. **58**, 1235 (1968).
[2]E. Lalor, J. Math. Phys. **9**, 2001 (1968). These papers also consider in greater detail mathematical questions similar to those which arise here.
[3]The harmonic time dependence $\exp(-i\omega t)$ is assumed throughout.

Deformations of Poisson brackets, Dirac brackets and applications

M. Flato, A. Lichnerowicz, and D. Sternheimer

Physique Mathématique, Collège de France, F-75231-Paris Cedex 05, France
(Received 21 October 1975)

After a short review of results which we recently obtained on deformations of Lie algebras associated with symplectic manifolds, we discuss physical applications and treat some examples with deformed Poisson brackets. We make explicit a connection between classical and quantum mechanics, and the theory of Dirac brackets for second class constraints, from the viewpoint of deformation theory. Finally we discuss the general Dirac constraints formalism.

INTRODUCTION

In previous papers we calculated, using cohomological methods, the 1-differentiable deformations of the infinite-dimensional Lie algebras of functions endowed with Poisson brackets on symplectic manifolds. This last problem is by no means purely mathematical and possesses potentially a large number of applications to problems in mathematical physics. Here are some examples:

(1) In a given model, to write down the Hamilton equations with the deformed bracket. Integrate then the equations of motion and compare with the nondeformed case.

(2) How much unique is the usual Hamilton mechanics compared with other *a priori* possible "close" (deformed) mechanics?

(3) Can deformation theory of Poisson brackets shed a new light on perturbation theory relative to the usual mechanics?

(4) The Dirac singular Hamiltonian formalism of constrained mechanics has been known since a long time and applied by many authors to construct canonical formalism for the electromagnetic and gravitational cases. Two questions arise:

(a) Can one construct the Dirac formalism in a natural geometric manner?

(b) Can the Dirac bracket be connected to Poisson bracket via deformation theory?

(5) Problems which might be connected with quantum mechanics:

(a) Existence or nonexistence of unitary representations of the exponentiated symmetric polynomial elements of the Poisson algebra.

(b) Is there any possibility of "interpolating" between classical and quantum mechanics?

After a short review of the main mathematical results obtained by us on deformation theory, we try in this article to discuss, present examples, and solve partially the above mentioned problems.

I. DEFORMATIONS OF POISSON BRACKETS

Let W be a symplectic manifold, i.e., a connected paracompact C^∞ real manifold of even dimension $2n$, on which is given a closed 2-form F such that $F^n \neq 0$ everywhere. As is well known, there are symplectic local charts on W with coordinates $(p_\alpha, q_\alpha,\ \alpha=1,\ldots,n)$ for which F takes the usual form $\sum_\alpha dp_\alpha \wedge dq_\alpha$. Let us denote by TW (resp. T^*W) the tangent (resp. cotangent) bundle to W, and by $i(X)$ the interior product by the vector field $X \in T^1(W)$ (cf., e.g., Ref. 1). This enables us to define an important fibre bundle isomorphism $\mu: TW \to T^*W$ by extension from $\mu(X) = -i(X)F$ which associates with the vector field X the 1-form $-i(X)F$. We denote $G = \mu^{-1}(F)$: The Poisson bracket on $N = C^\infty(W, \mathbb{R})$, the space of real-valued infinitely differentiable functions on W, is then defined by $\{u, v\} = i(G)(du \wedge dv)$ for all $u, v \in N$, and endows N with a Lie algebra structure. On a symplectic chart the Poisson bracket takes the usual form $\sum_\alpha(\partial_\alpha u \partial_{\bar\alpha} v - \partial_{\bar\alpha} u \partial_\alpha v)$ with $\partial_\alpha = \partial/\partial p_\alpha$ and $\partial_{\bar\alpha} = \partial/\partial q_\alpha$.

We have been interested in deformations of this Lie algebra, and therefore, according to the general theory of deformations of Lie algebras (cf. Gerstenhaber[2]), in the Chevalley–Eilenberg cohomology[3] of N with values in the adjoint representation, at least in degrees ≤ 3. Little is known on these cohomology spaces in the general case. (Very recently some results have been obtained, when the cochains are given by differential operators and in connection with a specific deformation, by Vey.[4]) However, when cochains are given by order-one differential operators, what we call 1-differentiable cochains, a complete computation has been given by one of us,[5] and this has enabled us[6] to make a complete study of the corresponding deformations. These are the most natural deformations to study since the Poisson bracket itself is 1-differentiable: This property will then be preserved under the deformations. We shall see later another physical motivation for considering only 1-differentiable deformations.

A 1-differentiable p-cochain C with values in N, is a p-linear alternate mapping from N^p to N that can be written $C = A + B$, where A is a p-tensor and B a $(p-1)$-tensor, so that on the domain of a local chart $\{x^k\}$ $(k=1,\ldots,2n)$ of W we have, for $u_1,\ldots,u_p \in N$, denoting $A(k) = A^k$, $\partial_k = \partial(k) = \partial/\partial x^k$, and by $\mathcal{A}$ the alternation over all permutations of the u's, with summation over the k_j:

$$A(u_1,\ldots,u_p) = A(k_1,\ldots,k_p)\,\partial(k_1)u_1\cdots\partial(k_p)u_p, \qquad (1.1)$$

$$B(u_1,\ldots,u_p) = \mathcal{A}(B(k_2,\ldots,k_p)u_1\,\partial(k_2)u_2\cdots\partial(k_p)u_p).$$

When $B=0$, we call the cocycle $C=A$ "pure."

The coboundary ∂C of any C is 1-differentiable if C is so, and is given by

$$\partial C=\mu^{-1}(F\wedge\mu(B)-d\mu(A))+\mu^{-1}(d\mu(B)) \tag{1.2}$$

or, in terms of the Schouten–Nijenhuis brackets[7] which associate with given p-tensor and q-tensor a $(p+q-1)$ tensor in a skew-symmetric or symmetric way (according to the parity of pq), with an accordingly modified version of the Jacobi identity:

$$\partial C=(G\wedge B-[G,A])+[G,B]. \tag{1.2$'$}$$

It therefore makes sense to speak of the 1-differentiable cohomology $H^*(N)$, with values in N.

If we denote by $P^p(W,F)$ (resp. $Q^{p+2}(W;F)$) the kernel (resp. the image) of the map $H^p(W)\to H^{p+2}(W)$ defined on the real cohomology classes of W by the exterior product by F on the p-forms, then (cf. Ref. 5) for the 1-differentiable cohomology:

$$H^p(N)=P^{p-1}(W;F)\oplus H^p(W)/Q^p(W;F), \tag{1.3}$$

where the second summand corresponds to the pure cocycles.

In particular, when F is exact,

$$H^p(N)=H^{p-1}(W)\oplus H^p(W).$$

A formal deformation of the Lie algebra N is a new Lie algebra law

$$[u,v]_\lambda=\sum_{r=0}^{\infty}\lambda^r C_r(u,v), \tag{1.4}$$

where $C_r(u,v)$ are 2-cochains on N, with $C_0(u,v)=\{u,v\}$. The formal Jacobi identity can then be written (denoting by S the sum over circular permutations of u, v, and w):

$$\partial C_t(u,v,w)=S\textstyle\sum^{(t)}C_r(C_s(u,v),w)\equiv E_t(u,v,w) \qquad (1.5)_t$$

for all $t=1,2,\cdots$, where $\sum^{(t)}$ denotes the sum over r,s with $r+s=t$ and $rs\neq 0$.

If $(1.5)_t$ is satisfied for $t=1,\ldots,q-1$, the Jacobi identity is satisfied to order q, i.e.,

$$S[[u,v]_\lambda,w]_\lambda=O(\lambda^q) \tag{1.6}$$

and E_q is a 3-cocycle of N: Its class in the third cohomology space is the obstruction at order q to the construction of a formal deformation of N.

An infinitesimal deformation is given by a 2-cocycle C_1 such that (1.6) is satisfied at order $q=2$.

Now, if we deal with 1-differentiable cochains C_r, the E_q are 1-differentiable,[6] so that $H^3(N)$ is then relevant for the obstructions. When the C_t are pure 1-differentiable 2-cochains A_t, the Jacobi condition $(1.5)_t$ can be written in terms of the Schouten–Nijenhuis brackets as

$$\partial A_t=\tfrac{1}{2}\textstyle\sum^{(t)}[A_r,A_s]. \qquad (1.5')_t$$

We call an infinitesimal 1-differentiable deformation trivial if there exists a 1-differentiable 1-cochain T_1 such that $\partial T_1=C_1$ (such a T_1 is necessarily 1-differentiable if W is noncompact[6]), i.e., $T_\lambda=I+\lambda T_1$ is an infinitesimal automorphism of N:

$$T_\lambda[u,v]_\lambda-\{T_\lambda u,\ T_\lambda v\}=O(\lambda^2). \tag{1.7}$$

Similarly, a 1-differentiable deformation is called trivial if the left-hand side of (1.7) is identically 0 when $T_\lambda=I+\sum_{r=1}^{\infty}\lambda^r T_r$, where the 1-cochains T_r are necessarily (as a consequence of the proposition of Sec. 5, Ref. 6) differential operators when W is noncompact or when they are local.

If G_1 is a pure 2-cocycle, we can define an infinitesimal deformation of the symplectic structure by $G_\lambda=G+\lambda G_1$, giving rise to a new Poisson bracket. We call inessential[6] an infinitesimal 1-differentiable deformation that is of this type up to a trivial deformation, i.e., such that there exist $T_\lambda=I+\lambda T_1$ and $G_\lambda=I+\lambda G_1$ (G_1 a pure cocycle) satisfying

$$T_\lambda[u,v]_\lambda-i(G_\lambda)(dT_\lambda u\wedge dT_\lambda v)=O(\lambda^2)$$

or equivalently that $C_1=G_1+\partial T_1$. Essential is defined as non-inessential.

Thus the space of infinitesimal 1-differentiable deformations, modulo the trivial (resp. the inessential) deformations, is isomorphic to $H^2(N)$ (resp. $P^1(W;F)$). If F is exact and $H^1(W)\neq 0$ but $H^2(W)=0=H^3(W)$, there will exist essential formal 1-differentiable deformations on N [no obstruction will then occur since $H^3(N)=0$].

In some special cases, e.g., the cotangent bundle $W=T^*M$ to a n-dimensional manifold M, which is the most interesting case for physical applications, we shall have a family of rigorous essential deformations of the type[6]

$$[u,v]_\lambda=\{u,v\}+\lambda C_1(u,v)$$

(the Jacobi identity being rigorously satisfied) parametrized by a vector space of cocycles C_1 of dimension $\dim H^1(M)$. We shall use these in the following.

The formal deformation built by Vey[4] (for manifolds with $H^3(W)=0$) is given by cochains C_r that are, in the same sense as above, $(2r+1)$-differentiable 2-cochains, and is not trivial (the class of C_1 in the cohomology of N is not trivial); the order is increasing (and C_1 is of order 3). On a symplectic chart and for polynomials u and v in the local coordinates, we can write $(2r+1)!C_r$ as the $(2r+1)$th-power of the bidifferential operator C_0 (the Poisson bracket); the "Vey bracket" is then, for $\lambda=(\frac{1}{2}i\hbar)^2$, nothing but the Moyal bracket corresponding to commutators of operators in the Weyl quantization procedure. This makes the Moyal–Vey bracket interesting, but in connection with *quantum* mechanics.

The latter example is characteristic of what occurs if we allow differentiable cochains of order greater than one: By construction, if C_r is of order n_r (which is well-defined since C_r is skew-symmetric), the cochains E_t and C_t will be of order (at least) $n_t=\max^{(t)}(n_r+n_s-1)$, where $\max^{(t)}$ means the maximum over r,s with $r+s=t$, $rs\neq 0$. Thus the orders will be increasing if some $n_r>1$ and it is not possible to restrict ourselves to m-differentiable cochains (defined similarly to 1-differentiable ones) for fixed m. The relevant cohomology would then be that studied in part by Vey,[4] which is nontrivial already in the formal case according to Vey.

Apart from the mathematical difficulty of performing a complete study of deformations given by differentiable cochains of unbounded (increasing) order, there are some serious physical reasons making the latter not very suitable for physical applications in classical mechanics. First, as for Poisson brackets, the deformed Hamilton equations

$$\dot{p}_\alpha = [H, p_\alpha]_\lambda, \ \dot{q}_\alpha = [H, q_\alpha]_\lambda \tag{1.8}$$

give locally the equation

$$\dot{f} = [H, p_\alpha]_\lambda \partial_\alpha f_\alpha + [H, q_\alpha]_\lambda \partial_{\bar{\alpha}} f \tag{1.9}$$

for dynamical quantities $f \in N$. The latter could make no sense for general H's and cochains C since the right-hand side in (1.9) involves then (when convergent) an infinite order bidifferential operator (a kind of pseudo-bidifferential operator) for which the local character is lost: $\dot{f}$ at any point could involve the value of H at other points in phase space.

Moreover, even if we restrict ourselves to the first order in λ, whenever the corresponding cochains are m-differentiable with $m > 1$, the connection between the (approximate) deformed Hamilton equations and a variational principle of the Helmholtz type in usual phase space is lost. This also makes the physical interpretation much more uneasy.

Finally one can remark that Poisson brackets behave relatively to products like *derivations* (they have what one can call a derivation character), namely that for all $u, v, w \in N$

$$\{uv, w\} = u\{v, w\} + \{u, w\} v. \tag{1.10}$$

The same formula is true (the products being written in the above order) in the quantum case, when we are dealing with commutators of operators in Hilbert space. Correspondingly, for the Moyal–Vey deformation, a "twisted" (noncommutative) product, defined with the exponential of the Poisson bracket operator, has to be introduced in order that a formula similar to (1.10) holds in that case. For the deformed brackets (1.4) with the ordinary product law and cochains C_r defining the deformation, it is easy to see that the derivation character expressed in (1.10) will hold if and only if the cochains C_r are pure 1-differentiable cochains. The associated infinitesimal deformation is then inessential.

This derivation character has some physical importance. For instance, if the deformed brackets (1.4) have it with respect to the ordinary product law, the deformed local evolution equations (1.9) for $f \in N$ can be written (at least for analytic H's) in the global form

$$\dot{f} = [H, f]_\lambda, \tag{1.11}$$

which is the same as for Poisson bracket. If we want (1.11) to hold for all such H's and $f \in N$, then it is also necessary to have the derivation character. We shall consider more in details in the next section the deformed evolution and Hamilton equations.

II. APPLICATIONS AND EXAMPLES OF DEFORMATIONS

The deformations of the Lie algebra of Poisson brackets, and possibly of other Lie algebras associated with symplectic or contact manifolds, can be physically relevant. We shall begin with a remark concerning the deformed evolution equations (1.9). Deformations given by nonpure cochains (e.g., essential deformations) are in general not equivalent to the global form (1.11). In particular postulating (1.8) we have from (1.9) in this case, for $f = C$, a constant, $[H, C]_\lambda \neq 0 = \dot{C}$ and for $f = H$ we get $\dot{H} \neq 0 = [H, H]_\lambda$: from Hamilton equations (1.8) with brackets deformed by essential deformations we get a new mechanics where energy is not necessarily conserved; the same holds for integrals of motion defined as quantities commuting with H; therefore the corresponding equations will in general be of a new type, not obtainable in the usual classical mechanics. Thus a treatment formally similar to the usual one may describe entirely new situations. This formal similarity would then provide a "deformed canonical formalism" that might be relevant for extending this kind of treatment to field theory and for quantization. It would therefore be of interest to study more in details the underlying physics relative to the deformed brackets. Such an approach might also be useful in cosmology.

Moreover, instead of treating evolution equations $\dot{f} = \{H_\lambda, f\}$ for dynamical quantities f, relative to a perturbed (and possibly not very precisely known) Hamiltonian H_λ, it might be advisable to consider the deformed equations (1.9) relative to a nonperturbed (e.g., free) Hamiltonian H, but with deformed Poisson brackets. In this framework we may express as a deformation the dependence on λ of the dynamical quantities f, i.e., replace f by a formal series f_λ in the above-mentioned equations. We shall thus compare the two treatments when H also is given by a formal series $H_\lambda = H + \sum_{r=1}^{\infty} \lambda^r V_r$.

In order to make these ideas more concrete we shall first compare the treatment with perturbed brackets with that involving a perturbed Hamiltonian, and then present some examples, in classical mechanics, of "motions" relative to essential deformations.

A. Deformations of Hamiltonians

(a) In this approach, we want to replace the treatment of a system described by a perturbed Hamiltonian H_λ, with the usual Poisson bracket formalism, by that of a system described by the free Hamiltonian H in the deformed Poisson brackets formalism. In accordance with our general treatment of deformations, we shall write (with $V_r \in N$):

$$H_\lambda = H + \sum_{r=1}^{\infty} \lambda^r V_r \equiv H + V. \tag{2.1}$$

In addition we shall write the dependence on λ of the dynamical variables as a formal deformation, i.e., as a formal series

$$u_\lambda = u + \sum_{r=1}^{\infty} \lambda^r u_r \tag{2.2}$$

with $u, u_r \in N$, which would in particular express (on a given symplectic chart) a possible formal change in the coordinates (p_α and q_α).

We then require that the Hamilton equations for the dynamical quantities u_λ (with, e.g., $u = p_\alpha$ or q_α on an open chart) with perturbed Hamiltonian H_λ

$$\dot{u}_\lambda = \{H_\lambda, u_\lambda\} \tag{2.3}$$

should have the same dynamical content as the deformed equations which we shall write only in local coordinates:

$$\dot{u}_\lambda = [H, p_\alpha]_\lambda \partial_\alpha u_\lambda + [H, q_\alpha]_\lambda \partial_{\bar\alpha} u_\lambda. \tag{2.4}$$

It is easy to see that (2.4) is equivalent to the global form

$$\dot{u}_\lambda = [H, u_\lambda]_\lambda \tag{2.5}$$

if and only if the coefficients B_t^i of the nonpure part of the cochains C_t satisfy

$$B_t^i \partial_i H = 0. \qquad (2.6)_t$$

In particular if $C_t = \partial T_t$ is a coboundary, with $T_t = a_t^i \partial_i + b_t$, the condition $(2.6)_t$ becomes $\{b_t, H\} = 0$: The nonpure part of the cochain must be given by a symmetry of the free Hamiltonian. Moreover, if $(2.6)_t$ is to be satisfied for all H, we see that C_t must be a pure 1-differentiable 2-cochain (this has been derived earlier from the related "derivation character," relatively to products, of the deformed bracket).

Comparing both values of $\dot{u}_\lambda$ in (2.3) and (2.4) we get

$$\begin{aligned}&\{V_t, u\} + \textstyle\sum^{(t)}\{V_r, u_s\}\\ &= \textstyle\sum^{(t)}(C_r(H, p_\alpha)\partial_\alpha u_s + C_r(H, q_\alpha)\partial_{\bar\alpha} u_s)\\ &\quad + C_t(H, p_\alpha)\partial_\alpha u + C_t(H, q_\alpha)\partial_{\bar\alpha} u. \qquad (2.7)_t\end{aligned}$$

These relations should hold for all u, u_s, and t. Then applying successively $(2.7)_{t'}$ with $t' < t$ to the u_s we see that they are equivalent to

$$C_t(H, q_\alpha)\, dp_\alpha - C_t(H, p_\alpha)\, dq_\alpha = dV_t. \qquad (2.8)_t$$

This implies that the cochains C_t must satisfy, in addition to the deformation conditions $(1.5)_t$, the usual integrability conditions for such systems.

If $(2.6)_t$ is satisfied (e.g., if the cochains C_t are pure), $(2.8)_t$ can also be written

$$\{V_t, u\} = C_t(H, u). \qquad (2.9)_t$$

In particular, for $u = H$, we obtain

$$\{V_t, H\} = 0; \qquad (2.10)_t$$

the allowed perturbations must then be constants of motion for the free Hamiltonian. This still leaves room for interesting perturbations. For instance, since the squared linear momentum $2H = \sum_{\alpha=1}^3 p_\alpha^2$ is a Casimir operator for the Euclidean group $SO(3)\cdot\mathbb{R}^3$, we can take for such an H the angular momentum L^2 or L_α as a possible perturbation V.

If we are looking for infinitesimal deformations, we may restrict ourselves to the first order (in λ). However, if we want these deformations to be rigorous, we have still an additional condition for the cocycle C_1, namely

$$SC_1(C_1(u, v), w) = \partial C_1(u, v, w) = 0. \tag{2.11}$$

In this case, as we shall do in subsection B, it seems preferable to express functionally the dependence of u on λ, in particular in (2.4). When H_λ and $u(\lambda)$ depend analytically on λ (at least for small values of $|\lambda|$), this does not change the compatibility conditions.

(b) Let us now consider more in detail the case [suggested by $(2.6)_t$] where *the cochains C_t are pure.* Denoting $\alpha_t = \mu(C_t)$ and considering the Hamiltonian vector field $Z_H = \mu^{-1}(dH)$, the compatibility conditions $(2.8)_t$ can be written

$$i(Z_H)\alpha_t = -dV_t, \qquad (2.8')_t$$

and therefore the integrability conditions for this equation become, since the Lie derivative $\mathcal{L} = di + id$,

$$\mathcal{L}(Z_H)\alpha_t = i(Z_H)\, d\alpha_t \qquad (2.12)_t$$

and in particular since C_1 is a cocycle,

$$\mathcal{L}(Z_H)\alpha_1 = 0. \qquad (2.12)_1$$

we have thus obtained the direct part of:

Proposition: Any sequence of pure 1-differentiable 2-cochains C_t ($t = 1, 2, \cdots$) satisfying the deformation conditions $(1.5')_t$ and the integrability conditions $(2.12)_t$ for a given H can be associated with a deformation of the "free" Hamiltonian H by "perturbations" V_t commuting with it in such a way that the equations of motion (2.3) and (2.4) will be equivalent; and conversely (locally).

In particular, any pure cocycle C_1 satisfying $(2.12)_1$ defines an infinitesimal deformation of the Poisson brackets in such a way that the infinitesimal deformed equations (2.4) are (to the second order in λ) equivalent to perturbed equations (2.3) with $H_\lambda = H + \lambda V_1$, V_1 commuting with H, and conversely (locally).

To complete the proof, we have to find a deformation of the Poisson bracket equivalent to a perturbation of a given Hamiltonian H by given V_t commuting with H. We shall start with the first order.

Let V_1 be given commuting with H, and let us look locally for C_1 such that $(2.8')_1$ is satisfied. We may take local coordinates such that $Z_H^1 = 1$ and $Z_H^j = 0$ for $j \neq 1$. Then $\partial_1 V_1 = 0$ and if we take for $\alpha = \mu(C_1)$ a *closed* 2-form such that $\alpha_{i1} = \partial_i V_1$ and α_{ij} ($i, j \neq 1$) is independent of x^1, the corresponding C_1 is a solution.

For the general case, given the V_t commuting with H, we shall build $\alpha_t = \mu(C_t)$ by induction. The deformation condition $(1.5')_t$ gives us that $d\alpha_t$ is a known 3-form (expressible in terms of the α_r for $r < t$). Locally, we may choose a 2-form γ_t such that $d\gamma_t = d\alpha_t$. We thus have to look for a closed 2-form β_t such that $i(Z_H)(\beta_t + \gamma_t)$ is the given closed 1-form $-dV_t$. As above with the same local coordinates, dropping the index t for simplicity, we shall select a 2-form β such that $\beta_{i1} = \partial_i V + \gamma_{1i}$ and since β must be closed, such that the β_{ij} ($i, j \neq 1$) satisfy $\partial_1 \beta_{ij} = \partial_i \beta_{1j} - \partial_j \beta_{1i} = \partial_i \gamma_{j1} - \partial_j \gamma_{i1}$, which expresses their x^1 dependence; their dependence on the x^j ($j \neq 1$) is then subject to the only condition that β is closed when x^1 is taken as a parameter. In such a way we may successively construct α_t which satisfy automatically $(2.12)_t$ and define a deformation.

(c) In particular, for a two-dimensional manifold W, taking for C_1 a coboundary (which can always be done locally)

$$C_1(u,v)=\partial T(u,v)=(A+b)\{u,v\}+u\{b,v\}+v\{b,u\}, \tag{2.13}$$

where $T=a_1\partial_1+a_{\bar 1}\partial_{\bar 1}+b$ is a 1-cochain and $A=\partial_1 a_1+\partial_{\bar 1}a_{\bar 1}$, the integrability conditions of the infinitesimal compatibility relations $(2.8)_1$, reduce here to

$$\{H,A\}+\partial_1(p\{H,b\})+\partial_{\bar 1}(q\{H,b\})=0.$$

Taking into account $(2.6)_1$, we see that the coefficients of C_1 must satisfy

$$\{H,A\}=0=\{H,b\}. \tag{2.14}$$

Moreover, since here (2.11) writes $\{A,b\}Su\{v,w\}=0$, the coboundary C_1 will define a rigorous (trivial, in the sense of deformation theory) deformation if in addition to (2.14) we have also $\{A,b\}=0$. Since here we have only one independent constant of motion, we shall thus take A and b functions of H, in which case V will also (as expected) be a function of H: Any (differentiable) function $V(H)$ can be obtained in this way.

Similar conditions can be obtained for a general cocycle (not necessarily a coboundary): In particular $\{H,A\}=0$ is the condition for a pure 2-cocycle $C_1(u,v)=A\{u,v\}$ to define a rigorous (inessential) deformation compatible with a deformation of H by $V(H)$.

(d) In the general case (dimension >2) the conditions take a more complicated form, but the basic principle is the same: Perturbations of Hamiltonians (by integrals of the free motion) can be related to deformations.

For example, let us take $C_1=\partial T$, $T=\sum_{i=1}^{2n}a_i\partial_i+b$, and the free Hamiltonian $H=\frac12\sum_{\alpha=1}^{n}p_\alpha^2$. Let us, moreover, specialize to the case $2n=4$ and $b=0$, so that $(2.10)_1$ is trivially satisfied. Instead of looking for the most general solution of the integrability conditions that ensure the existence of a V_1, we shall try a particular solution, e.g., $V_1=-(p_1a_1+p_2a_2)$ and look for the coefficients of T. Starting with $a_{\bar 1}$ depending explicitly on q_1 and q_2 and some arbitrary functions, and satisfying $(p_1\partial_{\bar 1}+p_2\partial_{\bar 2})a_{\bar 1}\equiv a_1\neq 0$ but $(p_1\partial_{\bar 1}+p_2\partial_{\bar 2})^3a_{\bar 1}=0$ (which is a compatibility condition), we shall be able to find a_2 such that the above-given V_1 is a solution and is given by $(p_1\partial_{\bar 1}-p_2\partial_{\bar 2})a_1=-2p_2\partial_{\bar 1}a_2$; then we can choose any $a_{\bar 2}$ satisfying $p_2\partial_{\bar 2}a_{\bar 2}=p_1\partial_{\bar 2}a_1+\partial_{\bar 1}(p_1a_2)$ and express $\{V_1,H\}=0$.

For instance, with $a_{\bar 1}=k_1(p)q_1^2+k_2(p)q_2^2$ we may have V_1 of the form $V_1=K_1(p)q_1+K_2(p)q_2+p_2K(p)$ (k_1,k_2, and K arbitrary functions of the p, the first two determining K_1 and K_2). In this case $\{H,V_1\}=p_1K_1+p_2K_2$ so that we may choose k_1 and k_2 such that $\{H,V_1\}=0$. One can also try to find solutions such that $V_1=0$ but including terms V_n (for $n>1$).

B. Rigorous essential deformations

As mentioned in Sec. I, for a cotangent bundle $W=T^*M$ with the natural symplectic structure, we have essential regorous deformations if $b_1=\dim H_1(M)\neq 0$. They are built as follows. We denote by $d\omega$ ($\omega=\sum_{\alpha=1}^{n}p^\alpha dq_\alpha$ on a canonical chart) the symplectic form of W, and set $Z=-\mu^{-1}(\omega)$ [locally, $Z=\sum p_\alpha(\partial/\partial p_\alpha)$]. We denote by π the projection $W=T^*M\to M$ and define a vector field $B=\mu^{-1}(\pi^*\beta)$, where β is a closed nonexact 1-form on M (the equivalence classes of such β is a b_1-dimensional space). Then (with summation on $i,j=1,\dots,2n$) we have the following rigorous essential deformations:

$$[u,v]_\lambda=\{u,v\}-\lambda(Z^iB^j-Z^jB^i)\partial_i u\,\partial_j v+\lambda B^i(u\partial_i v-v\partial_i u), \tag{2.15}$$

or in canonical coordinates, if $\beta=\sum_{\alpha=1}^{n}B^\alpha dq_\alpha$ on M:

$$[u,v]_\lambda=(\partial_\alpha u\partial_{\bar\alpha}v-\partial_{\bar\alpha}u\partial_\alpha v)-\lambda(p^\alpha B^{\alpha'}-p^{\alpha'}B^\alpha)\partial_\alpha u\partial_{\alpha'}v+\lambda B^\alpha(u\partial_\alpha v-v\partial_\alpha u). \tag{2.15'}$$

We shall consider in particular the case of cyclical one- or two-dimensional configuration spaces M, i.e., $M=T^1$ (the circle) or $M=T^2$ (the torus). In the first case, for $\beta=dp$ ($\partial_1=\partial/\partial p$, $\partial_{\bar 1}=\partial/\partial q$):

$$[u,v]_\lambda=(\partial_1u\partial_{\bar 1}v-\partial_{\bar 1}u\partial_1v)+\lambda(u\partial_1v-v\partial_1u). \tag{2.16}$$

In the latter case, for $\beta=B_1dq_1+B_2dq_2$, we have

$$[u',v]_\lambda=\{u,v\}+\lambda B_1C_1(u,v)+\lambda B_2C_2(u,v), \tag{2.17}$$

i.e., a combination of two deformations, given by

$$\begin{aligned}C_1(u,v)&=p_2(\partial_1u\partial_2v-\partial_2u\partial_1v)+(u\partial_1v-v\partial_1u),\\ C_2(u,v)&=-p_1(\partial_1u\partial_2v-\partial_2u\partial_1v)+(u\partial_2v-v\partial_2u).\end{aligned} \tag{2.18}$$

Formulas similar to (2.17) and (2.18) can be given for $M=T^n$.

(a) *Free circular motion*: $M=T^1$, $H=\frac12p^2$: In this case, the unperturbed equations give $p=p_0$, $q=p_0t+q_0$. However, with the deformed brackets (2.16), the deformed Hamilton equations

$$\begin{aligned}\dot p&=-\tfrac12\lambda p^2=-\lambda H,\\ \dot q&=p(1-\lambda q)=p+q\dot H/H\end{aligned} \tag{2.19}$$

give $p^{-1}=\frac12\lambda t+p_0^{-1}$, $\lambda q=1-cp^2$ [$c=(1-\lambda q_0)p_0^{-2}$]; hence $p\to 0$ and $q\to\lambda^{-1}$ when $t\to\infty$. Qualitatively, the motion is similar to the asynchrone pendulum.

(b) *Physical pendulum*: $M=T^1$, $H=\frac12I^{-1}p^2+R(1-\cos q)$: Here, the usual Hamilton equations $\dot p=-R\sin q$, $\dot q=pI^{-1}$ give $\ddot q+RI^{-1}\sin q=0$, whence the usual solution

$$t-t_0=\int(2RI^{-1}\cos q+C)^{-1/2}dq \quad (C=\text{const}).$$

With the deformed brackets we have

$$\begin{aligned}\dot p&=-R\sin q+\lambda[R(1-\cos q)-\tfrac12I^{-1}p^2],\\ \dot q&=(1-\lambda q)pI^{-1},\end{aligned}$$

whence $\ddot q+\frac32\lambda\dot q^2(1-\lambda q)^{-1}+RI^{-1}(1-\lambda q)[\sin q-\lambda(1-\cos q)]=0$. We have here (for appropriate values of λ) a kind of "viscosity" term in $\dot q^2$, which cannot be obtained in a natural way with the usual Poisson brackets. This friction term is not so surprising, since with essential deformations energy is not necessarily conserved. The classical integration procedure[8] gives the solution (which coincides with the usual one for $\lambda=0$):

$$t-t_0=\int\{(1-\lambda q)^3[C-2RI^{-1}\int(1-\lambda q)^2\times\sin q(1-\lambda tg\tfrac12q)\,dq]\}^{-1/2}dq.$$

As an approximation, to the first order in q, we obtain $\dot{q}^2 = C(1-3\lambda q) + O(q^2)$ whence $q \sim C(t-t_0) + 2/3\lambda$ for t in the neighborhood of $t_0 - 2/(3C\lambda)$.

(c) *Free Hamiltonian on the torus*: $M = T^2$, $H = \frac{1}{2}(p_1^2 + p_2^2)$: The unperturbed motion is uniform on a straight line on the torus, considered as a rectangle in $\mathbb{R}^2$ with the usual identifications. The deformed brackets (2.17) with $\lambda_\alpha = \lambda B_\alpha$ give ($\alpha = 1, 2$)

$$\dot{p}_\alpha = -\lambda_\alpha H, \quad \dot{q}_\alpha = p_\alpha + q_\alpha \dot{H}/H,$$

which can be integrated, with constants t_0, C_α and $k = \lambda_1 p_2 - \lambda_2 p_1 \neq 0$, if we suppose that $\lambda^2 = \lambda_1^2 + \lambda_2^2 \neq 0$ and define $\lambda_{\alpha+1}$ as λ_2 for $\alpha = 1$ and $-\lambda_1$ for $\alpha = 2$, to

$$\lambda^2 p_\alpha = -k\lambda_{\alpha+1} + k\lambda_\alpha \operatorname{cotg}[\tfrac{1}{2}k(t-t_0)],$$

$$\lambda^2 q_\alpha = \lambda_\alpha + \lambda_{\alpha+1} \operatorname{cotg}[\tfrac{1}{2}k(t-t_0)] - (\tfrac{1}{2}k\,\lambda_{\alpha+1} + \lambda_\alpha C_\alpha)\sin^2[\tfrac{1}{2}k(t-t_0)]$$

[If $k = 0$, we get separation of variables, and for each α the same motion as in case (a) above.] Thus for $t \to t_0$, at least one (both if $\lambda_1\lambda_2 \neq 0$) $p_\alpha \to \infty$, and $q_\alpha \to \infty$, but $q_1/q_2 \to \lambda_1 C_1/\lambda_2 C_2$ and $p_1/p_2 \to \lambda_1/\lambda_2$ if $\lambda_2 \neq 0$: The motion is increasingly accelerated towards a straight line, in the $\mathbb{R}^2$ picture of the torus; afterwards we are in the situation where $k = 0$, and back to the type of motion described in (a), but on a line on the torus.

III. SOME REMARKS ON THE CONNECTION BETWEEN CLASSICAL AND QUANTUM MECHANICS

In this section, we shall introduce a structure which may provide a continuous link between classical and quantum mechanics and discuss the correspondence principle from the point of view of Lie algebra representation theory, namely representations of the dynamical Lie algebra N and their implications for the notion of observable.

(1) Let us write an expression for a bracket

$$[f,g]^* = (1-\lambda)\{f,g\} + \lambda i[f,g], \tag{3.1}$$

the precise meaning of which will be specified in the following.

(a) If we take for f and g differentiable functions on a symplectic manifold, the first bracket in the right-hand side of (3.1), which we shall call the braces, being the Poisson bracket and the second one (the square bracket) being the commutator—which is identically zero—we get, of course, the bracket of classical mechanics with a factor. We get here no deformation of this bracket.

(b) On the other hand if F and G are differential operators with polynomial coefficients in the configuration variables q_α that are formally symmetric, i times their commutator $[F,G]$ has the same property. For the braces in (3.1) we shall take the properly symmetrized differential operator $\{F,G\}^\sim$ obtained from the brutal application of the Poisson bracket operation to F and G considered as functions f and g of the p_α (identified with $-i\partial_{\bar{\alpha}}$) and q_α. This is usually done (see Ref. 9 and references quoted therein) by using some type of ordering (e.g., Weyl). This procedure may also be applied to suitable functions of p_α and q_α. In this case one usually gets that the two operations give the same result, so that for all λ the new bracket (3.1) will be that of quantum mechanics. It should be mentioned that one of the difficulties arising here is due to the fact that there are are[10] formally symmetric differential operators with polynomial coefficients having no self-adjoint extension in $L^2(M)$, M being the configuration space (with coordinates q_α).

(c) Let us now consider the *"mixed" situation*. From what we have mentioned above, one introduces a mapping θ from a subset (e.g., of polynomials) of the Lie algebra N of C^∞ functions in p_α and q_α into a space P of (differential or sometimes pseudodifferential) operators in, e.g., the Hilbert space $L^2(M)$, in such a way that, the squared bracket being the commutator, $i[\theta f, \theta g]$ is equal to $\{\theta f, \theta g\}^\sim$ when f and g have the right properties. The mapping θ is usually called a *quantization*.

More precisely, let P_k be the space of real polynomials of degree $\leqslant k$ ($k = 0, 1, 2, \cdots$) in $(-i\partial_{\bar{\alpha}})$ with real polynomial coefficients in the q_α that are formally symmetric differential operators. $P = \cup P_k$ is a Lie algebra with the Lie law $(a,b) \mapsto i[a,b]$ (for $a, b \in P$) and the P_k define a filtration such that $i[P_h, P_k] \subset P_{h+k-1}$. We shall denote $P^k = P_k/P_{k-1}$, with $P^0 = P_0$ (we consider only the principal part of the differential operator); $\operatorname{gr}(P) = \oplus P^k$ is the graded algebra associated with P. In fact, P can be viewed as a real form of the complex enveloping algebra $\mathcal{U}_{\mathbb{C}}(g_{2n+1})$ of the nilpotent Weyl Lie algebra generated by $2n+1$ elements p_α, q_α, and z, the Lie bracket of p_α and q_α being z and all others vanishing, in the usual Heisenberg representation: The filtration and graduation then become obvious.[11] The space N' of real polynomials in commuting variables p_α and q_α can then be viewed as a real form of the corresponding symmetric algebra.

We then define the mapping $\theta: N' \to \operatorname{gr}(P)$ exactly as the canonical mapping between the symmetric algebra and the graded algebra associated with the filtered enveloping algebra[11]: With every monomial we associate the image in $\operatorname{gr}(P)$ of any symmetrized differential operator obtained by replacing p_α by $(-i\partial_{\bar{\alpha}})$. The reverse operation (replacement of $-i\partial_{\bar{\alpha}}$ by p_α) is the *symbol* mapping σ, which can be defined on P but is best defined in $\operatorname{gr}(P)$.

We are thus led to define the new bracket of the type (3.1) in the direct sum $N' \oplus \operatorname{gr}(P)$ as follows, for $f, g \in N'$ and $F, G \in \operatorname{gr}(P)$:

$$[f+F,\ g+G]^* = (1-\lambda)\{f+\sigma F,\ g+\sigma G\} + \lambda i[\theta f + F,\ \theta g + G]. \tag{3.2}$$

From the definition it follows that this bracket satisfies the Jacobi identity. For $\lambda = 0$ we have the bracket of classical mechanics by restriction to N', and for $\lambda = 1$ we have the bracket of quantum mechanics by "restriction" to suitable elements of P: we have here defined a kind of *interpolation between classical and quantum mechanics*.

Remark: We might here take for P also operators on functions of p_α and q_α in several ways. For instance we may represent p_α by $-i\beta\partial_{\bar\alpha} + (2\beta')^{-1}p_\alpha$ (or by $-i\beta_{\bar\alpha} + \beta' p_\alpha$) and q_α by $-i\beta'\partial_\alpha + (2\beta)^{-1} q_\alpha$ (or respectively by $\beta^{-1} q_\alpha$), so that the commutator is again i. This would keep both variables in the functional space and permit us to consider N' also as acting in this space by multiplication. We have here what can be called a *prequantization*. Replacing $1-\lambda$ by 1 in (3.2) we would thus get a kind of deformation of the Poisson bracket algebra N', considered as an algebra of multiplication operators, to a more general operator algebra having some features of the quantum mechanical algebra.

A similar prequantization procedure has been given by R. Rączka in connection with quantum field theory and gives, for quantum mechanics in the case $2n=2$,

$$f \mapsto F = f - \tfrac{1}{2}(q\,\partial_q f + p\partial_p f) - i(\partial_q f \partial_p - \partial_p f \partial_q).$$

(2) *Some remarks on the representations of N in Hilbert space*: These remarks are based on a recent result which we shall quote here.

Lemma (*Arnal*[10]): Let G be a real Lie group with noncompact Lie algebra and enveloping algebra $\mathcal{U}$. Then there exists a Hermitian element u in $\mathcal{U}$ such that, for any unitary representation U of G with faithful differential dU on the Lie algebra and for any domain D of differentiable vectors dense in the Hilbert space of the representation and invariant under $U(G)$, the restriction of the operator $dU(u)$ to D is symmetric with no self-adjoint extension.

In particular, for the Lie algebra $\mathbf{g}_3$ of the Heisenberg group, one can choose[10] in $\mathcal{U}_{\mathbf{C}}(\mathbf{g}_3)$, $u = iz((p,q)_+, q^2)_+$ where $(p,q)_+ = pq + qp$ is the anticommutator; $dU(u)$ has deficiency indices $(0,1)$ in any faithful unitary irreducible representation of the Heisenberg group, and therefore[10] also no self-adjoint extension in any faithful representation of $\mathbf{g}_3$ that is infinitesimally unitary in the sense of Harish-Chandra, i.e., integrable to a unitary representation (of the Heisenberg group). The same holds for the Weyl algebras $\mathbf{g}_{2n+1}$. But all representations of $\mathcal{U}_{\mathbf{C}}(\mathbf{g}_{2n+1})$ that are scalar on the center z (e.g., the irreducible ones) are representations of the Poisson subalgebra N' of N consisting of real polynomials in the p_α and q_α endowed with the induced structure of both Lie and associative (non-Abelian) algebra: We consider here the flat case $W = \mathbb{R}^{2n}$. We shall call these representations infinitesimally unitary if the symmetric elements are represented by essentially self-adjoint operators. The following result is therefore true:

Theorem: There are no faithful infinitesimally unitary representations of the Poisson algebra N' in the flat (or algebraic variety) case.

One may mention here that a faithful (not infinitesimally unitary) representation of the Lie algebra N by differential operators defined on vector fields, globally on W, has been found by Kerner[12] but only for very particular sympletic manifolds with curvature $-F \otimes Id$.

Moreover, the above example of $dU(u)$ shows that there are symmetric polynomials in p and q that are *not observables* in the strict quantum-mechanical sense. Therefore, either one has to change the usual meaning of observables and include also, e.g., maximal symmetric operators among observables, in which case the one-parameter unitary group structure has to be replaced by a semi-group structure—this would be an extension to the one-parameter case of the motion of local (nonintegrable) representation of Lie algebras introduced by two of us (M. F. and D. S.)[13]; or one has to give criteria excluding some symmetric elements of N from the family of quantum mechanical observables. In both cases, and in addition to other reasons that may suggest it also, a *reassessment of the motion of observable seems to be needed.*

IV. DIRAC BRACKETS (FOR SECOND CLASS CONSTRAINTS) AND DEFORMATIONS

Let N be the Lie algebra (for Poisson brackets) of the differentiable functions over a symplectic manifold W, and let $k_i \in N$ $(i=1,\ldots,S)$ be a set of second-class constraints in the sense of Dirac (cf., e.g., Ref. 14), so that the matrix $(\{k_i, k_j\})$ is regular and has an inverse (C_{ij}). Then the Dirac bracket relative to this situation is defined, for any $u, v \in N$, by

$$[u,v] = \{u,v\} + C(u,v) \tag{4.1}$$

with a 2-cochain $C(u,v)$ given by (with summation over i and j)

$$C(u,v) = -\{u,k_i\}C_{ij}\{k_j,v\}. \tag{4.2}$$

We first notice that, for the bracket $[u,v]'_\lambda = \{u,v\} + \lambda\, C(u,v)$, we have

$$S[[u,v]'_\lambda, w]'_\lambda = (\lambda-1)\,S\{u,k_i\}\{v,k_j\}\{C_{ij},w\},$$

where S means summation over circular permutations of u, v, and $w \in N$, which shows that in general the Jacobi identity will be satisfied only for $\lambda=0$ (Poisson bracket) and $\lambda=1$ (Dirac bracket). We shall, however, relate the Dirac bracket to deformations.

(1) *Some special cases*: We limit ourselves here to two constraints k_1, k_2 on $\mathbb{R}^{2n}$:

$$C(u,v) = (\{u,k_1\}\{k_2,v\} - \{u,k_2\}\{k_1,v\})\{k_1,k_2\}^{-1}.$$

(a) If we take $k_1 = q_1$, $k_2 = p_1$, then $C(u,v) = \{u,v\}_1$, the Poisson bracket relative to the subspace $\mathbb{R}^2$ with coordinates (p_1, q_1) and $[u,v]$ is the Poisson bracket relative to the complementary subspace $\mathbb{R}^{2n-2}$ with coordinates $(p_\alpha, q_\alpha,\ \alpha \neq 1)$. In this case, the above defined bracket $[u,v]'_\lambda$ is a deformation (for all values of λ). The restriction of the Poisson bracket to a symplectic submanifold defined by two conjugate constraints $(\{k_1,k_2\}=1)$ is thus an instant of a rigorous and first-order deformation.

(b) Let us now choose $k_1 = q_1 - \mu f(p)$, $k_2 = p_1 - \mu g(q)$. Then $\{k_1,k_2\} = \mu^2 \partial_\alpha f \partial_{\bar\alpha} g - 1$, whence a formal series expansion

$$C(u,v) = \{u,v\}_1 + \mu\partial T_1(u,v) + \sum_{r=2}^{\infty} \mu^r C_r(u,v), \tag{4.3}$$

where we can choose for instance $T_1 = g\,\partial_1 + f\,\partial_{\bar 1}$ and where the cochains C_r can be computed by multiplying the polynomial of order 2 in μ expressing $\{k_1,k_2\}\,C(u,v)$ and the power series of $\{k_1,k_2\}^{-1}$. The first two terms in the right-hand side of (4.3) can also be written $\partial T(u,v)$, with $T = a\partial_1 + b\partial_{\bar{1}}$, $a = k_1 - \frac{1}{2}p_1$ and $b = k_1 - \frac{1}{2}q_1$ for instance. We then get here, for every μ, an instant of a two-parameter deformation (n-parameter deformations can be defined along the same lines as we did for one-parameter ones).

(2) *General formal case: Lie algebra brackets and deformations*: Let us consider the vector space $N^F = R[[W]]$ of formal series in the p_α and q_α, coordinates of $W = \mathbb{R}^{2n}$ endowed with the Poisson bracket relative to the canonical symplectic form $F = \sum_\alpha dp_\alpha \wedge dq_\alpha$, and let (4.1) be a new Lie algebra law on the same vector space, which we shall suppose also 1-differentiable, i.e.,

C is 1-differentiable pure 2-cochain on N^F; in particular, $C(u,v) = -C(v,u)$. (4.4a)

$$\partial C(u,v,w) = SC(C(u,v),w) \equiv E(u,v,w), \qquad (4.4b)$$

where ∂C is the coboundary of C in $H^*(N^F)$, which expresses the Jacobi identity for the new law.

On the other hand, a formal 1-differentiable deformation of the Lie algebra N^F of formal series is given by a new law (1.4) with cochains C_r, that we shall suppose here pure 1-differentiable satisfying the relations $(1.5)_t$ for all t. The following problem then arises:

Problem: Can we consider formally all laws (4.1), and in particular the Dirac bracket law, as instants of a formal deformation (1.4), for a specialization, say $\lambda = 1$, of the parameter?

From relations $(1.5)_t$ we see that there is a high indetermination entering at each level t, since every C_t can be modified by a 2-coboundary T_t without altering the relation $(1.5)_t$ for the given t. Moreover, if all relations $(1.5)_t$ are satisfied, then the formal sum $C = \sum_{t=1}^{\infty} C_t$ satisfies (4.4). This formal sum is defined on formal series $u, v \in N^F$, at least as long as their coefficients are not given specific numerical values (we shall not enter here into the convergence problem for the coefficients, since we limit ourselves to the formal level). We can thus define a map $\sigma: \{C_t\} \mapsto \sum C_t = C$ from the space $\mathcal{D}$ of sequences $\{C_t\}$ of formal 2-cochains satisfying (1.5) to the space Δ of formal 2-cochains C satisfying (4.4). The above mentioned problem will then receive a *positive answer* if we prove:

Proposition: The map $\sigma: \{C_t\} \mapsto C = \sum C_t$ is onto.

Indeed, let us write, with summation over $i,j = 1, \ldots, 2n$ and over multi-indices $(k) = (k_1, \ldots, k_{2n})$, $k_i \geq 0$ integer, and with $x^{(k)} = \exp(k_1 \log x_1) \cdots \exp(k_{2n} \log x_{2n})$ and $|k| = \sum_i k_i$:

$$u = u_{(k)} x^{(k)}, \quad C_t(u,v) = A_t^{ij} \partial_i u \partial_j v$$

and similarly

$$C(u,v) = A^{ij} \partial_i u \partial_j v, \quad T_t u = a_t^i \partial_i$$

where $A^{ij} = (A^{ij})_{(k)} x^{(k)}$ and similarly for A_t^{ij} and a_t^i $(t \geq 1)$.

We must therefore have

$$(A^{ij})_{(k)} = \textstyle\sum_t (A_t^{ij})_{(k)}.$$

Moreover, the condition (4.4) can be expressed as a series of relations between linear combinations of the $(A^{mn})_{(k)}$ and of the $(A^{pq})_{(h)}(A^{rs})_{(l)}$ with $|k| = |h| + |l|$. The condition $(1.5)_t$ for a given t is expressed by the same relations but between $(A_t^{mn})_{(k)}$ and $(A_y^{pq})_{(h)}(A_z^{rs})_{(l)}$ with summation over y and z satisfying $t = y + z$. If cochains C_t satisfying the latter conditions are found, the former will be automatically satisfied for $C = \sum C_t$.

But Poincaré lemma (triviality of cohomology for closed differential forms) is true in the formal case, and thus H^3 for the 1-differentiable cohomology of N^F is trivial (this follows from the proof in Ref. 5). Therefore, if $(1.5)_y$ is satisfied for $y < t$, the cocycle[2] E_t can be written as ∂C_t for some (nonunique) C_t. These cochains can then be found successively when we start with an arbitrary cocycle C_1.

Now the space of the $(A^{ij})_{(k)}$, considered as coordinates in a vector space, for all (k) with $|k| \leq k_0$ a fixed finite number, is finite-dimensional: They define a point on an algebraic variety in a finite-dimensional space.

But the $(A_t^{ij})_{(k)}$ are defined only up to an infinite number of arbitrary coefficients $(a_r^m)_{(h)}$ with $r \leq t$ and where, for fixed (k), only (h)'s satisfying $|h| \leq |k| + t$ will appear. We can therefore find $(A_t^{ij})_{(k)}$ satisfying relations (4.5) for all (k) with $|k| \leq k_0$ fixed, in an infinite number of ways for any given $(A^{ij})_{(k)}$. We can similarly continue this procedure for another set of (k)'s without altering the already constructed $(A_t^{ij})_{(k)}$, and so on, whence the surjectivity of σ. We have thus proved that *all laws* (4.1) *with cochains C satisfying* (4.4) *are, in the formal case, instants of deformation.* In particular:

Proposition: The Dirac bracket law can be considered as an instant of a formal deformation of the Poisson bracket law on formal series.

Remark: If we suppose that the constraints form a Lie algebra, we can consider $\{k_i, k_j\}$ as a new constraint k_{ij}. We set $k_i = \lambda k_i'$. For $\lambda = 0$, the constraint k_i, expressed with k_i', disappears. For $\lambda \neq 0$, we redefine $C_{ij}' = \lambda C_{ij}$: Then $C(u,v) = \lambda C'(u,v)$, where C' has same form as C but with primed quantities. For $\lambda \neq 0$, we still have Dirac bracket; if $\lambda \to 0$, we get the Poisson bracket, which makes thus N appear as a *contraction* of the Dirac bracket algebra (when the constraints vanish).

V. FURTHER REMARKS ON DIRAC BRACKETS AND THEIR RELATION TO THE NEW NAMBU DYNAMICS

A. Dirac approach

Let (W, F) be a symplectic manifold of dimension $2n$, N its dynamical algebra (C^∞ functions, with Poisson bracket). Let C_0 be a subset of N, called the set of constraints, which we shall here suppose defining a submanifold M of W of codimension k, the common null set of all functions in C_0 (if necessary, we modify W so that this is the case). Without loss of generality, we may then suppose that C_0 is a vector subspace of N. In the terminology of Bergmann and Dirac (cf., e.g., Ref.

14), a basis of C_0 will be the set of both primary and secondary constraints, and is supposed finite-dimensional. Dirac[14] then calls weakly null quantities all functions in $N \otimes C_0$, i.e., linear combinations of constraints (null on M) with arbitrary coefficients, and first-class quantities all functions having weakly null Poisson brackets with all the constraints. He thus introduces the "normalizer," the space of first-class functions:

$$B_0 = \{f \in N; \{f, \varphi\} \in N \otimes C_0 \text{ for all } \varphi \in C_0\}.$$

The first-class constraints form then, of course, the h-dimensional space $A_0 = B_0 \cap C_0$ and all others are called second-class. Dirac's procedure (Ref. 14, p. 38) then amounts to choosing a basis $\{k_j\}$ of a subspace supplementary to A_0 in C_0, which is necessarily of even dimension $k-h$. These constraints enable him to define his *new bracket* by (4.1) and (4.2), which is nothing but the Poisson bracket on some symplectic submanifold $(\tilde{W}, \tilde{F})$ of codimension $(k\text{-}h)$ of (W, F), a "second-class" submanifold, and *therefore satisfies trivially Jacobi identity* (no computation is needed). From the construction it is obvious that the intermediate manifold $\tilde{W}$ is not uniquely defined once C_0 is given, except, of course, if all constraints are second-class ($M = \tilde{W}$).

Moreover, as mentioned by Dirac (in a somewhat unprecise manner), the physical states are "overdescribed" by M since with first-class constraints are associated canonical transformations which do not affect the physical state. The latter has in fact $2n-k-h$ degrees of freedom.

B. Geometric description

While for practical purposes the above-mentioned description is often more appropriate, it may be of interest to give it a more intrinsic formulation. This was given partly in Ref. 15, and with a somewhat different interpretation recently by one of us (A.L., Ref. 16). We shall give here the main results of the latter. Instead of C_0 one considers the spaces C_U of all functions which are constants on $M \cap U$, U being any chart domain on W intersecting M, and instead of B_0 one introduces the space B_U of all functions f such that $\{f, \varphi\}$ is zero on $M \cap U$ for all $\varphi \in C_U$, which is a Lie subalgebra of N with $A_U = B_U \cap C_U$ as an ideal.

One then supposes that the restriction to M of the 2-form F has fixed rank $2n-k-h$. The integrable distribution of h-planes (in all $x \in M$)

$$N_x = \{v \in T_x(M); i(v) F|_M(x) = 0\}$$

defines a foliation on M, and thus a quotient space $\hat{M}$ which we suppose here to be endowed by the projection $p: M \to \hat{M}$ with a $(2n-k-h)$-dimensional manifold structure such that p is a submersion. F then defines in a natural manner a symplectic form $\hat{F}$ on $\hat{M}$, and $(\hat{M}, \hat{F})$ is the manifold of physical states in the sense of Dirac.

We say that M is first-class if $h=k$, and second-class if $h=0$. Then it can be proved[15,16] that, under the above-mentioned hypotheses, there are second-class submanifolds $(\tilde{W}, \tilde{F})$ of (W, F) of codimension $k-h$, such that M is a first-class submanifold of $(\tilde{W}, \tilde{F})$. This is the analog of the Dirac procedure described above. The Dirac bracket is then the Poisson bracket on $\tilde{W}$, defined with the $\tilde{G}$ associated with the manifold $(\tilde{W}, \tilde{F})$.

C. Relation to Nambu dynamics

Recently, Nambu[17] has proposed a new structure, which might be connected with a new mechanics. Bayen and one of us (M. F., cf. Ref. 18) have shown that it contains the same dynamical informations as a singular Hamiltonian mechanics. For instance, in the most interesting (and most extensively studied) case of a three-dimensional space, it has been shown[18] that this space can be linearly imbedded into a six-dimensional phase space W, with three constraints. There are two second-class constraints, which appear in the Dirac bracket, and one first-class constraint. In view of what has been said above, equivalent but more involved (nonlinear) imbeddings can be exhibited for which there will be only one first-class constraint, in symplectic manifolds $\tilde{W}$ of dimension four. This is exactly what has been done by N. Mukunda and E. C. G. Sudarshan.[19] In both cases, the arbitrariness due to the first-class constraint appears through an arbitrary time-rescaling (the function v of Ref. 18), since classical mechanics is in fact done in the product of phase space by the time axis, or more generally in a "canonical manifold" in the sense of Refs. 20 and 21.

1S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963).
2M. Gerstenhaber, Ann. Math. 79, 59 (1964).
3C. Chevalley and S. Eilenberg, Trans. Amer. Math. Soc. 63, 85 (1948).
4J. Vey, "Déformation du crochet de Poisson sur une variété symplectique," Comment. Helv. Math. (to be published).
5A. Lichnerowicz, J. Math. Pures Appl. 53, 459 (1974) and C.R. Acad. Sci. Paris A 277, 215 (1973).
6M. Flato, A. Lichnerowicz, and D. Sternheimer, C.R. Acad. Sci. Paris A 279, 877 (1975); Compositio Mathematica 31, 47 (1975).
7A. Nijenhuis, Indag. Math. 17, 390 (1955).
8E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
9G.S. Agarwal and E. Wolf, Phys. Rev. D 2, 2161 (1970).
10D. Arnal, "Symmetric non self-adjoint operators in an enveloping algebra," J. Funct. Anal. (to be published).
11N. Bourbaki, *Algèbres de Lie* (Hermann, Paris, 1960).
12R. Kerner, C.R. Acad. Sci. Paris A 269, 175 (1969).
13M. Flato and D. Sternheimer, Phys. Rev. Letters 16, 1185 (1966); Commun. Math. Phys. 12, 296 (1969).
14P.A.M. Dirac, *Lectures on Quantum Mechanics*, Belfer Graduate School of Sciences Monograph Series No. 2 (Yeshiva University, New York, 1964).
15J. Sniatycki, Ann. Inst. H. Poincaré 20, 365 (1974).
16A. Lichnerowicz, C.R. Acad. Sci. Paris A 280, 523 (1974).
17Y. Nambu, Phys. Rev. D 7, 2405 (1973).
18F. Bayen and M. Flato, Phys. Rev. D 11, 3049 (1975).
19N. Mukunda and E.C.G. Sudarshan, Phys. Rev. D 13, 2846 (1976).
20A. Lichnerowicz, C.R. Acad. Sci. Paris A 280, 37 (1975); A 281 (1975); Variétés symplectiques, variétés canoniques et systèmes dynamiques (E. T. Davies memorial volume, to be published); Les variétés de Poisson et leurs algèbres de Lie associées (to be published).
21M. Flato, A. Lichnerowicz, and D. Sternheimer, "Algèbres de Lie attachées à une variété canonique," J. Math. Pures Appl. 54, 445 (1975).

Off-shell Jost function and T matrix for the Woods–Saxon potential*

M. N. Sinha Roy, B. Talukdar, and D. Chattarji

Department of Physics, Visva-Bharati University, Santiniketan, West Bengal, India
(Received 20 February 1975; final revised manuscript received 7 November 1975)

The s-wave van Leeuwan and Reiner equation for the Woods–Saxon potential is solved. Analytical expressions for the off-shell Jost solution and Jost function are derived. The results are used to obtain the T matrix.

I. INTRODUCTION

Recent interest in the two-particle T matrix has been stimulated mainly by the discovery of Faddeev[1] equations. Particularly, the off-shell two-body T matrix elements happen to be the main input for the three-body equations of Faddeev. The off-shell two-body transition amplitude also plays a significant role in the theories of nucleon—nucleon bremsstrahlung,[2] nuclear matter,[3] and finite nuclei.[4]

The purpose of the present paper is to obtain the s-wave part of the off-shell two-particle T matrix for the Woods—Saxon potential which is of importance in nuclear interactions. We do this by first obtaining the off-shell Jost solution and Jost function for this problem. Our derivation will be based on the differential equation approach of van Leeuwen and Reiner[5] to off-shell scattering as recently used by Fuda and Whiting.[6] In this approach the T matrix is obtained from an inhomogeneous form of the Schrödinger equation in which the inhomogeneous term represents a departure from elastic scattering. In other words, the equation is characterized by two momenta $\mathbf{k}$ and $\mathbf{q}$, where $\mathbf{k}$ is an on-shell momentum related to the energy by $E=k^2$ and $\mathbf{q}$ is an off-shell momentum. When $\mathbf{q}=\mathbf{k}$, the equation reduces to the conventional Schrödinger equation. The solution of the van Leeuwen—Reiner equation has sometimes been called the off-shell wavefunction.[7]

According to Fuda and Whiting[6] the off-shell Jost solution $f(k,q,r)$ for $l=0$ satisfies the inhomogeneous equation

$$\left(k^2+\frac{d^2}{dr^2}-V(r)\right)f(k,q,r)=(k^2-q^2)\exp(iqr). \tag{1}$$

Equation (1) has been written in units in which $\hbar^2/2m$ is unity. In close analogy with the theory of the ordinary Jost function,[8] the object $f(k,q,r)$ satisfies the asymptotic boundary condition $f(k,q,r)\sim\exp(iqr)$. Its behavior near the origin determines the off-shell Jost function. The off-shell Jost solution $f(k,q,r)$ and Jost function $f(k,q)$ give the appropriate on-shell quantities in the limit $q\rightarrow\pm k$. It is of interest to note that the regular solution $\Phi(k,q,r)$, which satisfies the inhomogeneous Schrödinger-like equation with $(k^2-q^2)\sin qr$ as the inhomogeneous term, can be expressed in terms of the functions $f(k,\pm q,r)$ and $f(k,r)$ $[=f(k,k,r)]$. The regular solution is given by[6]

$$\Phi(k,q,r)=-\tfrac{1}{2}\pi qT(k,q,s)f(k,r)+(1/2i)[f(k,q,r)-f(k,-q,r)], \tag{2}$$

where $T(k,q,s)$ is the half off-shell T matrix. It satisfies the relation

$$T(k,q,s)=[f(k,q)-f(k,-q)]/\pi iqf(k) \tag{3}$$

with

$$s=k^2+i\epsilon,\quad \epsilon\ll 1.$$

In these equations the functions $f(k,r)$ and $f(k)$ stand for the ordinary Jost solution and Jost function respectively:

$$f(k,r)=f(k,k,r),\quad\text{and}\quad f(k)=f(k,k). \tag{4}$$

In terms of $\Phi(k,q,r)$ the off-shell T matrix is

$$T(p,q,s)=\frac{2}{\pi pq}\int_0^\infty dr\,\sin pr\,V(r)\Phi(k,q,r). \tag{5}$$

Looking at Eqs. (2), (3), and (4), we see that we must first try to obtain suitable analytic expressions for the off-shell Jost solution and Jost function in order that we may use Eq. (5) to derive the off-shell T matrix. In Sec. II we derive expressions for the off-shell Jost solution and Jost function for the Woods—Saxon potential. In Sec. III we use these results to obtain the T matrix.

II. JOST SOLUTION AND JOST FUNCTION

The Woods—Saxon potential is given by

$$V(r)=-V_0\{1+\exp[(r-R)/a]\}^{-1}, \tag{6}$$

where R and a are nuclear radial and diffuseness parameters and V_0 the strength of the potential. Inserting Eq. (6) in Eq. (1), we get

$$\left(k^2+\frac{d^2}{dr^2}+\frac{V_0}{1+\exp[(r-R)/a]}\right)f(k,q,r)=(k^2-q^2)\exp(iqr). \tag{7}$$

By using the transformation

$$z=1/\{1+\exp[(r-R)/a]\} \tag{8}$$

Eq. (7) gives

$$\left(z(1-z)\frac{d^2}{dz^2}+(1-2z)\frac{d}{dz}+\frac{k^2a^2+V_0a^2z}{z(1-z)}\right)f(k,q,z)=a^2(k^2-q^2)\exp(iqR)(1-z)^{iqa-1}z^{-iqa-1}. \tag{9}$$

To reduce the left-hand side of Eq. (9) to a known form, we now consider the corresponding homogeneous equation. We proceed as follows.

(i) We note that for large r (i.e., for small z) it has a solution of the form z^{-ika} regardless of the values of the ratio R/a.

(ii) For most nuclei described by this potential, R/a ranges from about 6.0 to 9.0. Thus for $r \to 0$ (i.e., $z \to 1$) the solution of the homogeneous equation has the form $(1-z)^{ia(k^2+V_0)^{1/2}}$

From (i) and (ii) we see that the exact solution of Eq. (9) can be put in the form

$$f(k,q,z) = z^{-ika}(1-z)^{ia(k^2+V_0)^{1/2}} W(z), \tag{10}$$

where $W(z)$ is normalized asymptotically such that

$$f(k,q,r) \underset{r\to\infty}{\sim} \exp(iqr).$$

Substitution of Eq. (10) into Eq. (9) yields

$$z(1-z)W'' + \{C - (A+B+1)z\}W' - ABW$$
$$= a^2(k^2-q^2)\exp(iqR) z^{\sigma+1}(1-z)^{\tau-1}, \tag{11}$$

where

$$A = -ika + ia(k^2+V_0)^{1/2}, \quad B = -ika + ia(k^2+V_0)^{1/2} + 1, \tag{12}$$

$$C = 1 - 2ika, \quad \sigma = i(k-q)a, \quad \tau = ia[q - (k^2+V_0)^{1/2}].$$

The prime on $W(z)$ denotes differentiation with respect to z.

Equation (11) represents an ordinary nonhomogeneous second order differential equation. It can be integrated to give two complementary functions and one particular integral. To decide which of these three functions corresponds to the off-shell Jost solution, we note that the Jost solution derived from one of the complementary functions for Eq. (11) represents the ordinary Jost solution. The other, however, does not satisfy the appropriate boundary conditions. The particular integral gives the off-shell Jost solution. (This point has been discussed in some detail by Fuda and Whiting.) It is given by[9]

$$W(z) = a^2(k^2-q^2)\exp(iqR) \times \sum_{n=0}^{\infty} \frac{\Gamma(n+1-\tau) z^{\sigma+n}}{\Gamma(1-\tau) n!\,(\sigma+n)(\sigma+n+C-1)} \times {}_3F_2\left(\begin{matrix}1, n+\sigma+A, n+\sigma+B\\ n+\sigma+1, n+\sigma+C\end{matrix}\middle| z\right). \tag{13}$$

In Eq. (13) ${}_3F_2$ is a special case of the generalized hypergeometric function ${}_pF_q(\begin{smallmatrix}\alpha_p\\ \beta_q\end{smallmatrix}|z)$ defined by Luke.[10] Using Eq. (13) in Eq. (11) the off-shell Jost solution is obtained in the form

$$f(k,q,r) = a^2(k^2-q^2)\exp(iqR) \times \sum_{n=0}^{\infty} \frac{\Gamma(n+1-\tau)}{\Gamma(1-\tau)n!}\cdot\frac{z^{\sigma+n-ika}}{\sigma+n}\,\frac{(1-z)^{ia(k^2+V_0)^{1/2}}}{(n+\sigma+C-1)} \times {}_3F_2\left(\begin{matrix}1, n+\sigma+A, n+\sigma+B\\ n+\sigma+1, n+\sigma+C\end{matrix}\middle| z\right). \tag{14}$$

The series in Eq. (14) is uniformly convergent for all values of the independent variable r because the maximum value of z is less than unity.

To see that $f(k,q,r)$ given by Eq. (14) satisfies the asymptotic boundary condition $f(k,q,r) \underset{r\to\infty}{\sim} \exp(iqr)$, we rewrite this equation as follows:

$$f(k,q,r) = \exp(iqR) z^{-iqa}(1-z)^{ia(k^2+V_0)^{1/2}} \times {}_3F_2\left(\begin{matrix}1, \sigma+A, \sigma+B\\ \sigma+1, \sigma+C\end{matrix}\middle| z\right) + a^2(k^2-q^2)\exp(iqR) \times \sum_{n=1}^{\infty} \frac{z^{\sigma+n-ika}(1-z)^{ia(k^2+V_0)^{1/2}}}{(\sigma+n)(\sigma+n+C-1)} \times {}_3F_2\left(\begin{matrix}1, \sigma+n+A, \sigma+n+B\\ \sigma+n+1, \sigma+n+C\end{matrix}\middle| z\right) \tag{14$'$}$$

As $r\to\infty$, $z \to \exp[-(r-R)/a] \ll 1$, so that ${}_3F_2(\cdots|z) \approx 1$. Thus the first term in Eq. (14′) becomes equal to $\exp(iqr)$ while the others go to zero because of the factor z^n.

By using Eq. (14) the off-shell Jost function is found to be

$$f(k,q) = a^2(k^2-q^2)\exp(iqR) \times \sum_{n=0}^{\infty} \frac{\Gamma(n+1-\tau)\exp[-iR(k^2+V_0)^{1/2}]}{\Gamma(1-\tau)n!\,(\sigma+n)(\sigma+n+C-1)} \times {}_3F_2\left(\begin{matrix}1, n+\sigma+A, n+\sigma+B\\ n+\sigma+1, n+\sigma+C\end{matrix}\middle| 1-\eta\right) \tag{15}$$

with $\eta = \exp(-R/a)$. In writing Eq. (15) we have used $R \gg a$.

Equations (14) and (15) yield in the on-shell limit

$$f(k,r) \equiv \lim_{q\to k} f(k,q,r) = \exp(ikR) z^{-ika}(1-z)^{ia(k^2+V_0)^{1/2}}\, {}_2F_1\left(\begin{matrix}A, B\\ C\end{matrix}\middle| z\right) \tag{16}$$

and

$$f(k) \equiv \lim_{q\to k} f(k,q) = \exp(ikR)[1-\exp(-R/a)]^{-ika}\exp[-iR(k^2+V_0)^{1/2}] \times {}_2F_1\left(\begin{matrix}A, B\\ C\end{matrix}\middle| 1-\eta\right). \tag{17}$$

Asymptotically the function $f(k,r)$ in Eq. (16) $\sim \exp(ikr)$, which is the correct behavior prescribed for the on-shell Jost solution. Further, the analytic continuation of $f(k,r)$ in the upper half of the complex, k plane gives the bound state wavefunction for the Woods–Saxon potential given by Flügge.[11]

III. *T* MATRIX

In Sec. II we have obtained analytic expressions for the off-shell and on-shell Jost solutions. These results can be utilized to write the off-shell wavefunction regular at origin. Combining Eqs. (14), (16), and (1), we obtain

$$\Phi(k,q,r) = -\tfrac{1}{2}\pi q T(k,q,s)\left(\frac{1}{1+\exp[(r-R)/a]}\right)^{i(k'-k)a} \times {}_2F_1\left(\begin{matrix}A, B\\ C\end{matrix}\middle| \frac{1}{1+\exp[(r-R)/a]}\right) + \frac{a^2(k^2-q^2)}{2i}$$

$$\times\sum_{n=0}^{\infty}\left[\frac{\exp(iqR)\Gamma(n+1-\tau)\exp[ik'(r-R)]}{\Gamma(1-\tau)n!(\sigma+n)(\sigma+n+C-1)}\right.$$
$$\times\left(\frac{1}{1+\exp[(r-R)/a]}\right)^{n-iqa+ik'a}$$
$$\times{}_3F_2\left(\begin{matrix}1,n+\sigma+A,n+\sigma+B\\ n+\sigma+1,n+\sigma+C\end{matrix}\,\middle|\,\frac{1}{1+\exp[(r-R)/a]}\right)$$
$$-\frac{\exp(-iqR)\Gamma(n+1-\tau')\exp[ik'(r-R)]}{\Gamma(1-\tau')n!(\sigma'+n)(n+\sigma'+C-1)}$$
$$\times\left(\frac{1}{1+\exp[(r-R)/a]}\right)^{n+i(q+k')a}$$
$$\left.\times{}_3F_2\left(\begin{matrix}1,n+\sigma'+A,n+\sigma'+B\\ n+\sigma'+1,n+\sigma'+C\end{matrix}\,\middle|\,\frac{1}{1+\exp[(r-R)/a]}\right)\right], \tag{18}$$

where

$$k'=(k^2+V_0)^{1/2},\quad \sigma'=i(k+q)a,\quad \tau'=-ia(q+k'). \tag{19}$$

Since $z<1$ for all values of r, the series representation of the generalized hypergeometric function ${}_pF_q(^{\alpha_p}_{\beta_q}|z)$ in ascending powers of z can now be used to rewrite $\Phi(k,q,r)$ in the form

$$\Phi(k,q,r)=\sum_{m=0}^{\infty}\left\{Q_1G_m(A,B,C,R)\exp(ik'r)\left(\frac{1}{1+\exp[(r-R)/a]}\right)^{m+i(k'-k)a}\right.$$
$$+Q_2\sum_{n=0}^{\infty}\left[H_{mn}(A,B,C,q,\sigma,\tau,R)\exp(ik'r)\left(\frac{1}{1+\exp[(r-R)/a]}\right)^{m+n+ia(k'-q)}\right.$$
$$-H_{mn}(A,B,C,-q,\sigma',\tau',R)\exp(ik'r)$$
$$\left.\left.\times\left(\frac{1}{1+\exp[(r-R)/a]}\right)^{m+n+ia(k'+q)}\right]\right\}, \tag{20}$$

where

$$G_m(A,B,C,R)=\Gamma\begin{bmatrix}A+m,B+m,C\\ A,B,C+m,m+1\end{bmatrix}\exp[i(k-k')R], \tag{21a}$$

$$Q_1=-\tfrac{1}{2}\pi qT(k,q,s),\quad Q_2=a^2(k^2-q^2)/2i, \tag{21b}$$

and

$$H_{mn}(A,B,C,\pm q,X,Y,R)$$
$$=\Gamma\begin{bmatrix}m+n+X+A,m+n+X+B,n+X+1,n+X+C\\ n+X+A,n+X+B,m+n+X+1,m+n+X+C\end{bmatrix}$$
$$\times\exp[iR(\pm q-k')\frac{\Gamma(n+1-Y)}{\Gamma(1-Y)n!(X+n)(X+n+C-1)}. \tag{21c}$$

In writing out Eqs. (20b) and (20c) we have used the notation

$$\frac{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)\cdots\Gamma(a_A)}{\Gamma(b_1)\Gamma(b_2)\Gamma(b_3)\cdots\Gamma(b_B)}=\Gamma\begin{bmatrix}a_1,a_2,\ldots,a_A\\ b_1,b_2,\ldots,b_B\end{bmatrix}. \tag{22}$$

Making use of Eqs. (5), (6), and (20) we can write the s-wave part of the off-shell T matrix element in the form

$$T(p,q,s)=\frac{2V_0}{\pi pq}\sum_{m=0}^{\infty}\left\{Q_1G_m(A,B,C,R)\int_0^{\infty}\sin pr\exp(ik'r)\left(\frac{1}{1+\exp[(r-R)/a]}\right)^{m+1+ia(k'-k)}dr\right.$$
$$+Q_2\sum_{n=0}^{\infty}\left[H_{mn}(A,B,C,q,\sigma,\tau,R)\int_0^{\infty}\sin pr\exp(ik'r)\left(\frac{1}{1+\exp[(r-R)/a]}\right)^{1+m+n+ia(k'-q)}dr\right.$$
$$-H_{mn}(A,B,C,-q,\sigma',\tau',R)\int_0^{\infty}\sin pr\exp(ik'r)$$
$$\left.\left.\times\left(\frac{1}{1+\exp[(r-R)/a]}\right)^{1+m+n+ia(k'+q)}dr\right]\right\}. \tag{23}$$

Each integral in Eq. (23) can now be written as[12]

$$\int_0^{\infty}dr\cdots=\int_0^R dr\cdots+\int_R^{\infty}dr\cdots.$$

Now by using the relation[13]

$$(1+y)^{-\alpha}=\sum_{s=0}^{\infty}(-)^s\frac{(\alpha)_sy^s}{s!}, \tag{24}$$

with

$$(\alpha)_s=\Gamma(\alpha+s)/\Gamma(\alpha), \tag{25}$$

the integrals may be obtained in terms of elementary transcendental functions. We thus obtain the off-shell T matrix

$$T(p,q,s)=\frac{2V_0}{\pi pq}\sum_{m=0}^{\infty}\sum_{s=0}^{\infty}\left(Q_1G_m(A,B,C,R)[I^{(1)}_{m0s}(k)+I^{(2)}_{m0s}(k)]\right.$$
$$+Q_2\sum_{n=0}^{\infty}\{H_{mn}(A,B,C,q,\sigma,\tau,R)[I^{(1)}_{mns}(q)+I^{(2)}_{mns}(q)]$$
$$\left.-H_{mn}(A,B,C,-q,\sigma',\tau',R)[I^{(1)}_{mns}(-q)+I^{(2)}_{mns}(-q)]\}\right). \tag{26}$$

To evaluate $T(p,q,s)$, we need the half-off-shell T matrix which is determined by Eqs. (3), (15), and (17).

In Eq. (26),

$$I^{(1)}_{\mu\nu s}(\xi)=M^{(1)}_{\mu\nu s}(\xi)\{[(s+ik'a)\sin pR-ap\cos pR]\exp(ik'R)+pa\exp[-(R/a)s]\} \tag{27a}$$

$$I^{(2)}_{\mu\nu s}(\xi)=M^{(2)}_{\mu\nu s}(\xi)[(s+\mu+\nu+1-ia\xi)\sin pR+\cos pR]\times\exp(ik'R) \tag{27b}$$

with

$$M^{(1)}_{\mu\nu s}(\xi)=\frac{(-1)^s a(\mu+\nu+1-i\xi a+iak')_s}{s![a^2p^2+(s+ik'a)^2]}, \tag{28a}$$

$$M^{(2)}_{\mu\nu s}(\xi)=\frac{(-1)^s a(\mu+\nu+1-i\xi a+ik'a)_s}{s![a^2p^2+(s+\mu+\nu+1-i\xi a)^2]}. \tag{28b}$$

The triple series in Eq. (26) is uniformly convergent. This can be seen as follows. The n sum arises from the solution of the nonhomogeneous differential Eq. (11). We have already noted that this sum is uniformly convergent if z is less than unity (see Ref. 9, p. 211). The m sum results from the series expansion of the hypergeometric function ${}_{p+1}F_p(\because \mid z)$. In obtaining Eq. (26) by using Eq. (14) via Eq. (20) we have integrated these convergent series expansion within the circle of convergence. The n and m series can therefore be easily summed up on a computer for a fixed s.

As for the s sum we note that this arises from the binomial expansion of $[1/(1+\exp[(r-R)/a])]^\alpha$. It is involved in Eqs. (27a) to (28b). The convergence of this sum can be shown by using the relation $\lim_{n\to\infty}(a)_n = 1/\Gamma(a)$ (see Ref. 13, p. 3). For example, the integrals $I_{\mu\nu s}^{(1,2)} \to 0$ like $1/s^2$ for large s and fixed μ and ν.

*Work supported in part by the Department of Atomic Energy, Government of India.

[1]L. D. Faddeev, Zh. Eksp. Teor. Fiz. 39, 1495 (1960) [Sov. Phys. JETP 12, 1014 (1961)].
[2]A. H. Cromer and M. I. Sobel, Phys. Rev. 152, 1351 (1966); 162, 1174(E) (1966).
[3]M. Baranger, in *Nuclear Structure and Nuclear Reactions*, edited by M. Jean and R. A. Ricci (Academic, New York, 1969).
[4]R. L. Becker, A. D. Mackellar, and B. M. Morris, Phys. Rev. 174, 1264 (1968).
[5]J. M. J. van Leeuwen and A. S. Reiner, Physica 27, 99 (1961).
[6]M. G. Fuda and J. S. Whiting, Phys. Rev. C 8, 1255 (1973).
[7]O. P. Bahethi and M. G. Fuda, J. Math. Phys. 12, 2076 (1971).
[8]R. Jost, Helv. Phys. Acta 20, 256 (1947).
[9]A. W. Babister, *Trancendental Function Satisfying Nonhomogeneous Linear Differential Equations* (MacMillan, New York, 1967).
[10]Y. L. Luke, *The Special Functions and Their Approximations* (Academic, New York, London, 1969), Vol. 1.
[11]S. Flügge, *Practical Quantum Mechanics* (Springer-Verlag, Berlin, Heidelberg, New York, 1971).
[12]L. R. B. Elton, *Nuclear Sizes* (Oxford U. P., Oxford, 1961); J. R. Rook, Nucl. Phys. 39, 479 (1962).
[13]L. J. Slater, *Generalised Hypergeometric Function* (Cambridge U. P., Cambridge, 1966).

Critical length of a transport process in rod geometry

C. V. Pao

Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27607
(Received 14 July 1975)

In this paper, we study a two-point boundary-value problem which governs the transport of n different type of particles in a rod of finite length. Through the construction of an upper solution we establish a simple relation between the rod length and the physical parameters of the transport medium under which the maximal and the minimal sequences obtained in an earlier paper converge, respectively, to a maximal and a minimal solution of the problem. This relation leads to a lower bound for the critical length of the rod when fission occurs in the system. The convergence of the constructed sequences gives a mathematical justification for the existence of a physically meaningful solution to the system. It is also shown, under a slightly stronger condition on the rod length, that the maximal solution coincides with the minimal solution and the physical system is subcritical. In addition, an explicit recursion formula for the calculation of approximate solutions is given.

1. INTRODUCTION

In the transport process of n-different types of particles in a rod of length L the forward-moving particle's density $u_1, \ldots, u_n$ and the backward-moving particle's density $v_1, \ldots, v_n$ are governed by the following coupled equations (cf. Refs. 1, 2, 3)

$$\begin{aligned} &\frac{du}{dx} + A_0(x)u = A_1(x)u + A_2(x)v + p(x), \\ &-\frac{dv}{dx} + B_0(x)v = B_1(x)u + B_2(x)v + q(x), \end{aligned} \qquad (0 \le x \le L), \tag{1.1}$$

where $u \equiv (u_1, \ldots, u_n)$, $v \equiv (v_1, \ldots, v_n)$ together with p, q are n-vectors and A_l, B_l $(l = 0, 1, 2)$ are $n \times n$ matrices. When the ends of the rod are subjected to incident fluxes, we have the boundary conditions

$$u(0) = u_0, \quad v(L) = v_L, \tag{1.2}$$

where u_0, v_L are the incident fluxes at $x = 0$ and $x = L$, respectively. Eq. (1.1) is obtained from the particle's balance relation in which $A_0 u$, $B_0 v$ represent, respectively, the loss of forward and backward moving particles due to absorption while $A_1 u + A_2 v$, $B_1 u + B_2 v$ are the gain due to scattering and fission. The vectors p, q denote any other possible external sources in the system. A fundamental question concerning the above system is under what conditions on the length of the rod and the physical parameters of the transport medium the boundary-value problem (1.1), (1.2) has or has no nonnegative solutions. From a physical point of view, if the effect of absorption dominates the effect of scattering, the transport system should have a nonnegative solution and the existence of such a solution is independent of the length L (see Theorem 3.2 or the Corollary to Theorem 3.1). However, if scattering (and fission) dominates absorption this system does not necessarily have a nonnegative solution unless there is some restriction on L (see the example in Sec. 2). The determination of the critical length L_c so that the system (1.1), (1.2) has a physical meaningful solution is a very interesting (and difficult) problem in the study of transport phenomena. The purpose of this paper is to investigate this problem by studying the mathematical structure of the system (1.1), (1.2). Specifically, we establish some explicit conditions between L and A_l, B_l $(l = 0, 1, 2)$ to insure the existence of at least one nonnegative solution to (1.1), (1.2). The usefulness of this condition is that it leads to a lower bound for the critical length L_c in terms of the physical parameters of the medium when scattering dominates absorption. As is to be expected, this lower bound is independent of the sources p, q and the boundary data u_0, v_L. We also give a mathematical justification on the fact that the system (1.1), (1.2) has exactly one nonnegative solution for any length L when absorption dominates scattering.

The problem (1.1), (1.2) has been investigated by many authors using the approach of invariant imbedding (cf. Refs. 4–7). The corresponding time-dependent transport problem has been discussed by Bellman[8] and more recently by the author.[9] By studying a corresponding Riccati initial-value problem for (1.1), (1.2), Boland and Nelson[10] obtained an upper bound for the critical length of the rod. A more general transport problem in slab and spherical geometry was investigated by Case and Zweifel[11] and by Nelson.[12] In this paper, we use a different approach which is based on the results of an earlier work by the author[3] using the method of successive approximation. This method involves the construction of two monotone sequences which converge monotonically to a maximal and a minimal solution, respectively, provided that the system has an upper solution (see Definition 2.1). The present treatment amplifies the previous results by constructing an upper solution from which we can obtain some simple relations between A_l, B_l and L so that the physical system is either subcritical or critical. The subcriticality means that the system (1.1), (1.2) has a unique nonnegative solution for any nonnegative sources and boundary data, and criticality means more than one nonnegative solution.

In Sec. 2 we describe our process of successive approximation and state a result from Ref. 3 concerning the convergence of the approximations to a maximal or a minimal solution, depending on the initial iteration. Section 3 is devoted to the construction of an upper solution from which we establish some relations between A_l, B_l and L to insure the existence of at least one nonnegative solution. This relation leads to a lower bound for the critical length L_c when fission occurs in the system. Finally, we show in Sec. 4 that under suitable conditions on A_l, B_l and L the maximal solution coincides with the minimal solution and the system has exactly one nonnegative solution.

2. UPPER SOLUTION

In this section we describe our process of approximations and state a result from Ref. 3 for the existence problem. It turns out that the convergence or divergence of the sequence of approximations depends on the existence or nonexistence of an upper solution. Throughout the paper, we assume by physical reasons that all the elements in A_l, B_l $(l=0,1,2)$, p, q, u_0, v_L are nonnegative piecewise continuous functions on $[0,L]$ and A_0, B_0 are diagonal matrices whose elements are positive on $[0,L]$. The elements of A_0, B_0 are denoted by $a_i^{(0)}$, $b_i^{(0)}$ while those of A_l, B_l are denoted by $a_{ij}^{(l)}$, $b_{ij}^{(l)}$ $(l=1,2,\ i,j=1,\dots,n)$. Similar notations will be used for the vectors u, v, p, q.

As we have indicated in the introduction, if the probability of particle's gain due to scattering is more than the loss by collision, that is,

$$a_i^{(0)} < \sum_{j=1}^n (a_{ij}^{(1)} + a_{ij}^{(2)}), \qquad b_i^{(0)} < \sum_{j=1}^n (b_{ij}^{(1)} + b_{ij}^{(2)}), \qquad i=1,\dots,n, \tag{2.1}$$

then the system (1.1), (1.2) may not have a solution unless there are some restrictions on the length L. To demonstrate this, we consider the following simple example:

$$\frac{du}{dx} + u = v, \quad -\frac{dv}{dx} + v = (1+\alpha^2)u, \tag{2.2}$$

$$u(0)=0, \quad v(L)=\eta, \tag{2.3}$$

where u, v are scalar functions and α, η are nonnegative constants. It is easily seen from (2.2) and $u(0)=0$ that $u = c\sin\alpha x$ and $v = c(\alpha\cos\alpha x + \sin\alpha x)$, where c is an arbitrary constant. Now, if $\alpha\cos\alpha L + \sin\alpha L \neq 0$, then (2.2), (2.3) has a unique solution. However, if $\alpha\cos\alpha L + \sin\alpha L = 0$, that is,

$$L = (1/\alpha)[m\pi - \tan^{-1}(1/\alpha)] \quad (0<\alpha<\infty), \quad m=1,2,\cdots, \tag{2.4}$$

the problem (2.2), (2.3) has no solution unless $\eta = 0$. In the latter case it has infinitely many solutions. The value of L given by (2.4) is therefore the critical length of the system for each m. The above example demonstrates that if (2.1) holds, one cannot expect the system (1.1), (1.2) to have a nonnegative solution without suitable restriction on L.

In order to insure the existence of a solution to (1.1), (1.2) we employ the method of successive approximations used in Ref. 3 for the construction of solutions. For convenience, we define the diagonal matrices D_α, D_β by

$$D_\alpha(r,x) = \mathrm{diag}(\alpha_1(r,x),\dots,\alpha_n(r,x)), \qquad D_\beta(x,s) = \mathrm{diag}(\beta_1(x,s),\dots,\beta_n(x,s)), \tag{2.5}$$

where for each $i=1,\dots,n$,

$$\alpha_i(r,x) = \exp[-\textstyle\int_r^x a_i^{(0)}(\eta)\,d\eta], \qquad \beta_i(x,s) = \exp[-\textstyle\int_x^s b_i^{(0)}(\eta)\,d\eta]. \tag{2.6}$$

By a suitable choice of the initial iteration $(u^{(0)}, v^{(0)})$ we construct a sequence $\{u^{(k)}, v^{(k)}\}$ from the recursion formula (cf. Ref. 3)

$$\begin{aligned} u^{(k)}(x) &= D_\alpha(0,x)u_0 + \int_0^x D_\alpha(\xi,x)[A_1(\xi)u^{(k-1)}(\xi) + A_2(\xi)v^{(k-1)}(\xi) + p(\xi)]\,d\xi, \\ &\qquad k=1,2,\cdots. \\ v^{(k)}(x) &= D_\beta(x,L)v_L + \int_x^L D_\beta(x,\xi)[B_1(\xi)u^{(k-1)}(\xi) + B_2(\xi)v^{(k-1)}(\xi) + q(\xi)]\,d\xi, \end{aligned} \tag{2.7}$$

To insure the convergence of the above sequence to a solution of (1.1), (1.2), we need to find an upper solution which is defined as follows:

Definition 2.1: A pair of nonnegative functions $(\mathbf{u},\mathbf{v})$ is called an upper solution of (1.1), (1.2) if it is differentiable at every point where A_l, B_l, p, q are continuous and satisfies the conditions

$$\begin{aligned} &\frac{d\mathbf{u}}{dx} + A_0\mathbf{u} \geq A_1\mathbf{u} + A_2\mathbf{v} + p, \qquad \mathbf{u}(0) \geq u_0, \\ &-\frac{d\mathbf{v}}{dx} + B_0\mathbf{v} \geq B_1\mathbf{u} + B_2\mathbf{v} + q, \quad \mathbf{v}(L) \geq v_L. \end{aligned} \tag{2.8}$$

In the above definition the inequality $u \geq v$ for vectors u, v means that $u_i \geq v_i$ for every $i=1,\dots,n$. An immediate consequence of this definition is that every nonnegative solution of (1.1), (1.2) is also an upper solution.

Let $(\mathbf{u},\mathbf{v})$ be a given upper solution and let $u^{(0)}=\mathbf{u}$, $v^{(0)}=\mathbf{v}$. The sequence from (2.7) with the initial iteration $u^{(0)}=\mathbf{u}$, $v^{(0)}=\mathbf{v}$ is called a maximal sequence and is denoted by $\{\overline{u}^{(k)}, \overline{v}^{(k)}\}$. On the other hand, the sequence from (2.7) with $u^{(0)}=v^{(0)}=0$ is called the minimal sequence and is denoted by $\{\underline{u}^{(k)}, \underline{v}^{(k)}\}$. The following result from Ref. 3 insures the convergence of these sequences.

Theorem 2.1: If there exists an upper solution $(\mathbf{u},\mathbf{v})$ then the maximal sequence $\{\overline{u}^{(k)}, \overline{v}^{(k)}\}$ converges uniformly to a solution $(\overline{u},\overline{v})$ of (1.1), (1.2) while the minimal sequence $\{\underline{u}^{(k)}, \underline{v}^{(k)}\}$ converges uniformly to a solution $(\underline{u},\underline{v})$. Furthermore,

$$\begin{aligned} &0 \leq \underline{u}^{(1)} \leq \underline{u}^{(2)} \leq \cdots \leq \underline{u} \leq \overline{u} \leq \cdots \leq \overline{u}^{(2)} \leq \overline{u}^{(1)} \leq \mathbf{u}, \\ &0 \leq \underline{v}^{(1)} \leq \underline{v}^{(2)} \leq \cdots \leq \underline{v} \leq \overline{v} \leq \cdots \leq \overline{v}^{(2)} \leq \overline{v}^{(1)} \leq \mathbf{v}. \end{aligned} \tag{2.9}$$

The solutions $(\overline{u},\overline{v})$ and $(\underline{u},\underline{v})$ in Theorem 2.1 are called, respectively, maximal and minimal solutions of (1.1), (1.2) in the sense that any other nonnegative solution (u,v) of (1.1), (1.2) with $u \leq \mathbf{u}$, $v \leq \mathbf{v}$ satisfies the relation $u \leq \overline{u}$, $v \leq \overline{v}$ (resp. $u \geq \underline{u}$, $v \geq \underline{v}$). Notice that a maximal sequence depends on the corresponding upper solution but the minimal sequence is independent of upper solutions. Nevertheless, the convergence of both sequences depend on the existence of an upper solution. Since every nonnegative solution is also an upper solution, the minimal sequence must converge to a nonnegative solution unless the problem (1.1), (1.2) has no nonnegative solution. This observation leads to the following:

Theorem 2.2: The problem (1.1), (1.2) has a nonnegative solution if and only if it has an upper solution.

The usefulness of Theorems 2.1 and 2.2 is that upper solutions are required to satisfy only the inequality (2.8) which gives considerable flexibility in the choice of such functions. In the succeeding section we will construct an upper solution under some restrictions on the physical parameters A_l, B_l and the length L (but not on the data p, q, u_0, v_L).

3. A LOWER BOUND FOR THE CRITICAL LENGTH

In order to establish an explicit relation between the matrices A_l, B_l and the length L so that the problem (1.1), (1.2) has a nonnegative solution, we choose a particular pair of functions $(\mathbf{u}, \mathbf{v})$ as a possible upper solution. Specifically, we define

$$\begin{aligned}\mathbf{u}(x) &= D_\alpha(0,x)u_0 + \left[\int_0^x D_\alpha(\xi,x)\,d\xi\right]E_\gamma,\\ \mathbf{v}(x) &= D_\beta(x,L)v_L + \left[\int_x^L D_\beta(x,\xi)\,d\xi\right]E_\gamma,\end{aligned} \tag{3.1}$$

where D_α, D_β are the diagonal matrices given by (2.5) and $E_\gamma = (\gamma, \ldots, \gamma)$ is the vector in E^n with its components all equal to a constant $\gamma > 0$. The value of γ will be chosen so that the pair $(\mathbf{u}, \mathbf{v})$ is an upper solution [see (3.11)]. Since $\mathbf{u}(0) = u_0$, $\mathbf{v}(L) = v_L$ and, by direct differentiation,

$$\frac{d\mathbf{u}}{dx} + A_0\mathbf{u} = E_\gamma, \quad -\frac{d\mathbf{v}}{dx} + B_0\mathbf{v} = E_\gamma, \tag{3.2}$$

this pair will be an upper solution if

$$\begin{aligned}&A_1(x)\{D_\alpha(0,x)u_0 + [\textstyle\int_0^x D_\alpha(\xi,x)\,d\xi]E_\gamma\}\\ &\quad + A_2(x)\{D_\beta(x,L)v_L + [\textstyle\int_x^L D_\beta(x,\xi)\,d\xi]E_\gamma\} + p(x) \le E_\gamma,\end{aligned} \tag{3.3}$$

$$\begin{aligned}&B_1(x)\{D_\alpha(0,x)u_0 + [\textstyle\int_0^x D_\alpha(\xi,x)\,d\xi]E_\gamma\}\\ &\quad + B_2(x)\{D_\beta(x,L)v_L + [\textstyle\int_x^L D_\beta(x,\xi)\,d\xi]E_\gamma\} + q(x) \le E_\gamma.\end{aligned}$$

By letting

$$\begin{aligned}\phi(x) &\equiv A_1(x)D_\alpha(0,x)u_0 + A_2(x)D_\beta(x,L)v_L + p(x),\\ \psi(x) &\equiv B_1(x)D_\alpha(0,x)u_0 + B_2(x)D_\beta(x,L)v_L + q(x),\end{aligned} \tag{3.4}$$

the condition (3.3) is equivalent to

$$\begin{aligned}&\gamma^{-1}\phi(x) + A_1(x)[\textstyle\int_0^x D_\alpha(\xi,x)\,d\xi]E_1\\ &\quad + A_2(x)[\textstyle\int_x^L D_\beta(x,\xi)\,d\xi]E_1 \le E_1,\\ &\gamma^{-1}\psi(x) + B_1(x)[\textstyle\int_0^x D_\alpha(\xi,x)\,d\xi]E_1\\ &\quad + B_2(x)[\textstyle\int_x^L D_\beta(x,\xi)\,d\xi]E_1 \le E_1,\end{aligned} \tag{3.5}$$

where $E^1 \in E^n$ is the vector with its components all equal to one.

Consider first the special case where A_l, B_l are symmetric. Then A_l and B_l commute with any diagonal matrix and thus the inequalities in (3.5) may be written, in terms of their respective components, as

$$\begin{aligned}&\gamma^{-1}\phi_i(x) + \left(\sum_{j=1}^n a_{ij}^{(1)}(x)\right)\int_0^x \alpha_i(\xi,x)\,d\xi\\ &\quad + \left(\sum_{j=1}^n a_{ij}^{(2)}(x)\right)\int_x^L \beta_i(x,\xi)\,d\xi \le 1,\\ &\gamma^{-1}\psi_i(x) + \left(\sum_{j=1}^n b_{ij}^{(1)}(x)\right)\int_0^x \alpha_i(\xi,x)\,d\xi\\ &\quad + \left(\sum_{j=1}^n b_{ij}^{(2)}(x)\right)\int_x^L \beta_i(x,\xi)\,d\xi \le 1,\end{aligned} \tag{3.6}$$

for $i = 1, \ldots, n$, where ϕ_i, ψ_i are the components of ϕ, ψ, respectively. Let M be an upper bound of ϕ_i, ψ_i for all $i = 1, \ldots, n$ and define

$$\begin{aligned}&\underline{a}_i^{(0)} = \inf_{0\le x\le L}[a_i^{(0)}(x)], \quad \underline{b}_i^{(0)} = \inf_{0\le x\le L}[b_i^{(0)}(x)],\\ &\qquad\qquad (i = 1, \ldots, n,\ l = 1, 2).\\ &\overline{a}_i^{(l)} = \sup_{0\le x\le L}\left(\sum_{j=1}^n a_{ij}^{(l)}(x)\right), \quad \overline{b}_i^{(l)} = \sup_{0\le x\le L}\left(\sum_{j=1}^n b_{ij}^{(l)}(x)\right),\end{aligned} \tag{3.7}$$

Then, since

$$\begin{aligned}\int_0^x \alpha_i(\xi,x)\,d\xi &\le \int_0^x \exp[-\underline{a}_i^{(0)}(x-\xi)]\,d\xi\\ &= [1 - \exp(-\underline{a}_i^{(0)}x)]/\underline{a}_i^{(0)},\\ \int_x^L \beta_i(\xi,x)\,d\xi &\le \int_x^L \exp[-\underline{b}_i^{(0)}(\xi - x)]\,d\xi\\ &= \{1 - \exp[-\underline{b}_i^{(0)}(L-x)]\}/\underline{b}_i^{(0)},\end{aligned} \tag{3.8}$$

the condition (3.6) is certainly satisfied if

$$\begin{aligned}&\gamma^{-1}M + (\overline{a}_i^{(1)}/\underline{a}_i^{(0)})[1 - \exp(-\underline{a}_i^{(0)}x)]\\ &\quad + (\overline{a}_i^{(2)}/\underline{b}_i^{(0)})\{1 - \exp[-\underline{b}_i^{(0)}(L-x)]\} \le 1,\\ &\gamma^{-1}M + (\overline{b}_i^{(1)}/\underline{a}_i^{(0)})[1 - \exp(-\underline{a}_i^{(0)}x)]\\ &\quad + (\overline{b}_i^{(2)}/\underline{b}_i^{(0)})\{1 - \exp[-\underline{b}_i^{(0)}(L-x)]\} \le 1.\end{aligned} \tag{3.9}$$

Now if for each $i = 1, \ldots, n$ the functions

$$\begin{aligned}\rho_i^{(1)}(x) &\equiv (\overline{a}_i^{(1)}/\underline{a}_i^{(0)})[1 - \exp(-\underline{a}_i^{(0)}x)]\\ &\quad + (\overline{a}_i^{(2)}/\underline{b}_i^{(0)})\{1 - \exp[-\underline{b}_i^{(0)}(L-x)]\},\\ \rho_i^{(2)}(x) &\equiv (\overline{b}_i^{(1)}/\underline{a}_i^{(0)})[1 - \exp(-\underline{a}_i^{(0)}x)]\\ &\quad + (\overline{b}_i^{(2)}/\underline{b}_i^{(0)}\{1 - \exp[-\underline{b}_i^{(0)}(L-x)]\end{aligned} \tag{3.10}$$

are strictly less than one, then by letting $\overline{\rho}_i^{(l)}$ be the maximum value of $\rho_i^{(l)}(x)$ on $[0, L]$ and choosing

$$\gamma \ge M(1 - \overline{\rho}_i^{(l)})^{-1} \quad \text{for all } i = 1, \ldots, n,\ l = 1, 2. \tag{3.11}$$

the condition (3.9) [and thus (3.6)] is clearly satisfied. With this choice of γ, the pair $(\mathbf{u}, \mathbf{v})$ given by (3.1) becomes an upper solution. By an application of Theorem 2.1 we have the following conclusion:

Theorem 3.1: Assume that A_l, B_l $(l = 1, 2)$ are symmetric and

$$\overline{\rho}_i^{(l)} \equiv \max\{\rho_i^{(l)}(x); 0 \le x \le L\} < 1 \quad (l = 1, 2,\ i = 1, \ldots, n), \tag{3.12}$$

where $\rho_i^{(l)}$ are given by (3.10). Then the problem (1.1), (1.2) has at least one nonnegative solution. Specifically, the maximal sequence $\{\overline{u}^{(k)}, \overline{v}^{(k)}\}$ [with respect to the upper solution given by (3.1)] converges to a maximal solution $(\overline{u}, \overline{v})$ while the minimal sequence $\{\underline{u}^{(k)}, \underline{v}^{(k)}\}$ converges to the minimal solution $(\underline{u}, \underline{v})$.

An immediate consequence of Theorem 3.1 is the following:

Corollary: If A_l, B_l are symmetric and if

$$(\bar{a}_i^{(1)}/\underline{a}_i^{(0)}) + (\bar{a}_i^{(2)}/\underline{b}_i^{(0)}) \le 1,$$
$$(\bar{b}_i^{(1)}/\underline{a}_i^{(0)}) + (\bar{b}_i^{(2)}/\underline{b}_i^{(0)}) \le 1, \tag{3.13}$$

then for any length $L < \infty$ the problem (1,1), (1.2) has at least one nonnegative solution.

Proof: It is obvious that under the condition (3.13) the requirement (3.12) is fulfilled for any $L < \infty$. The result follows immediately from Theorem 3.1.

We next consider nonsymmetric matrices A_l, B_l. Since the inequalities in (3.5) are equivalent to

$$\gamma^{-1}\phi_i(x) + \sum_{j=1}^{n}\left[a_{ij}^{(1)}(x)\int_0^x \alpha_j(\xi,x)\,d\xi + a_{ij}^{(2)}(x)\int_x^L \beta_j(x,\xi)\,d\xi\right] \le 1 \quad (i=1,\ldots,n), \tag{3.14}$$

$$\gamma^{-1}\psi_i(x) + \sum_{j=1}^{n}\left[b_{ij}^{(1)}(x)\int_0^x \alpha_j(\xi,x)\,d\xi + b_{ij}^{(2)}(x)\int_x^L \beta_j(x,\xi)\,d\xi\right] \le 1,$$

if we define

$$\sigma_i^{(1)} = \sum_{j=1}^{n}\left((a_{ij}^{(1)}(x)/\underline{a}_j^{(0)})[1-\exp(-\underline{a}_j^{(0)}x)] + [a_{ij}^{(2)}(x)/\underline{b}_j^{(0)}]\{1-\exp[-\underline{b}_j^{(0)}(L-x)]\}\right), \quad (i=1,\ldots,n), \tag{3.15}$$

$$\sigma_i^{(2)}(x) = \sum_{j=1}^{n}\left((b_{ij}^{(1)}(x)/\underline{a}_j^{(0)})[1-\exp(-\underline{a}_j^{(0)}x)] + [b_{ij}^{(2)}(x)/\underline{b}_j^{(0)}]\{1-\exp[-\underline{b}_j^{(0)}(L-x)]\}\right),$$

then by (3.8) the relations in (3.14) hold if

$$\gamma^{-1}M + \bar{\sigma}_i^{(l)} \le 1 \quad \text{for every } l=1,2,\ i=1,\ldots,n, \tag{3.16}$$

where $\bar{\sigma}^{(l)}$ is the least upper bound of $\sigma_i^{(l)}(x)$ on $[0,L]$. Now, if $\bar{\sigma}_i^l < 1$, then (3.16) holds for any γ satisfying the inequality

$$\gamma \ge M(1-\bar{\sigma}_i^{(l)})^{-1} \quad \text{for } l=1,2,\ i=1,\ldots,n. \tag{3.17}$$

With this choice of γ the pair $(\mathbf{u},\mathbf{v})$ is again an upper solution. This leads to the following conclusion for the general case where A_l, B_l are not necessarily symmetric.

Theorem 3.2: Assume that

$$\bar{\sigma}_i^{(l)} \equiv \sup\{\bar{\sigma}_i^{(l)}(x); 0 \le x \le L\} < 1 \quad (l=1,2,\ i=1,\ldots,n), \tag{3.18}$$

where $\sigma_i^{(l)}(x)$ is given by (3.15). Then the problem (1.1), (1.2) has at least one nonnegative solution. In fact, all the conclusions in Theorem 3.1 hold.

The physical meaning of the results in the above theorems is that if the particles gain due to scattering is no more than the loss due to absorption, then for any $L < \infty$ the transport system cannot be supercritical as is to be expected. On the other hand, if scattering dominates absorption, then these results can be used to obtain a lower bound for the critical length of the rod. To see this, we assume, for convenience, that A_l, B_l are symmetric. Since, by (3.10), each function $\rho_i^{(l)}(x)$ is in the form of

$$\rho(x) = c_1[1-\exp(-c_2x)] + c_3\{1-\exp[-c_4(L-x)]\}, \quad (c_j > 0,\ j=1,\ldots,4)$$

an elementary calculation shows that $\rho''(x) < 0$ for all $x \in [0,L]$ and $\rho(x)$ has the maximum value at

$$x_m = (c_2+c_4)^{-1}[c_4L + \ln(c_1c_2/c_3c_4)].$$

Hence the maximum of $\rho(x)$ is $\bar{\rho} = \rho(x_m)$ when $0 \le x_m \le L$ and $\bar{\rho} = \max\{\rho(0),\rho(L)\}$ when $x_m < 0$ or $x_m > L$. In any case we can obtain a more explicit relation for $\bar{\rho}_i^{(l)}$ in terms of the elements of A_l, B_l. This relation can then be used to determine the length L from (3.12) and thus gives a lower bound for the critical length L_c. As an illustration, we consider the special case

$$\underline{a}_i^{(0)} = \underline{b}_i^{(0)}, \quad \bar{a}_i^{(1)} = \bar{a}_i^{(2)}, \quad \bar{b}_i^{(1)} = \bar{b}_i^{(2)}. \tag{3.19}$$

Then for each $l=1,2$, $i=1,\ldots,n$, the maximum of $\rho_i^{(l)}(x)$ occurs at $x_m = L/2$ and thus

$$\bar{\rho}_i^{(1)} = \rho_i^{(1)}(L/2) = 2(\bar{a}_i^{(1)}/\underline{a}_i^{(0)})[1-\exp(-\underline{a}_i^{(0)}L/2)],$$
$$\bar{\rho}_i^{(2)} = \rho_i^{(2)}(L/2) = 2(\bar{b}_i^{(1)}/\underline{b}_i^{(0)})[1-\exp(-\underline{b}_i^{(0)}L/2)]. \tag{3.20}$$

The above relation implies that (3.12) is satisfied if

$$\bar{a}_i^{(1)} + \bar{a}_i^{(2)} < \underline{a}_i^{(0)}[1-\exp(-\underline{a}_i^{(0)}L/2)]^{-1},$$
$$\bar{b}_i^{(1)} + \bar{b}_i^{(2)} < \underline{b}_i^{(0)}[1-\exp(-\underline{b}_i^{(0)}L/2)]^{-1}. \tag{3.21}$$

This observation leads to the following.

Theorem 3.3: Assume that A_l, B_l are symmetric and (3.19) holds. Then the problem (1.1), (1.2) has at least one nonnegative solution for any $L < \infty$ when

$$\bar{a}_i^{(1)} + \bar{a}_i^{(2)} \le \underline{a}_i^{(0)}, \quad \bar{b}_i^{(1)} + \bar{b}_i^{(2)} \le \underline{b}_i^{(0)} \quad (i=1,\ldots,n). \tag{3.22}$$

On the other hand, if

$$\bar{a}_i^{(1)} + \bar{a}_i^{(2)} > \underline{a}_i^{(0)}, \quad \bar{b}_i^{(1)} + \bar{b}_i^{(2)} > \underline{b}_i^{(0)} \quad (i=1,\ldots,n), \tag{3.23}$$

then it has at least one nonnegative solution provided that

$$L < \min\left\{\frac{2}{\underline{a}_i^{(0)}}\ln\left(\frac{\bar{a}_i^{(1)}+\bar{a}_i^{(2)}}{\bar{a}_i^{(1)}+\bar{a}_i^{(2)}-\underline{a}_i^{(0)}}\right), \frac{2}{\underline{b}_i^{(0)}}\ln\left(\frac{\bar{b}_i^{(1)}+\bar{b}_i^{(2)}}{\bar{b}_i^{(1)}+\bar{b}_i^{(2)}-\underline{b}_i^{(0)}}\right)\right\}. \tag{3.24}$$

Proof: The first part of the theorem follows from the corollary to Theorem 3.1. Since a simple manipulation shows that the relation (3.24) is equivalent to (3.21), the conclusions of the second part follows immediately from Theorem 3.1.

The condition (3.24) gives a more explicit lower bound for the critical length L_c in terms of the physical parameters of the transport medium when A_l, B_l are symmetric and (3.19) holds. In case these conditions are not satisfied, we can still find a lower bound for L_c by evaluating $\bar{\sigma}_i^l$ and then apply Theorem 3.2. It should be pointed out that the lower bound for L_c given above is based on the particular choice of the upper solution $(\mathbf{u}, \mathbf{v})$. A suitable choice of other upper solutions may improve these estimates.

4. UNIQUENESS OF THE SOLUTION

In the preceding section we have constructed an upper solution which insures the convergence of the maximal and the minimal sequence to a maximal and a minimal solution, respectively. It has been shown in Ref. 3 that if the elements of A_l, B_l satisfy the conditions

$$\begin{aligned}&\sum_{i=1}^{n} [a_{ij}^{(1)}(x) + b_{ij}^{(1)}(x)] < a_j^{(0)}(x),\\ &\sum_{i=1}^{n} [a_{ij}^{(2)}(x) + b_{ij}^{(2)}(x)] < b_j^{(0)}(x),\end{aligned} \qquad j = 1, \ldots, n, \tag{4.1}$$

then the maximal solution coincides with the minimal solution and the coincidence is independent of the length L. However, when fission occurs, this uniqueness property may not hold unless there is a restriction on L. This is demonstrated by the example given in (2.2), (2.3). The purpose of this section is to establish some uniqueness results by imposing some conditions on L when scattering (including fission) dominates absorption. Here we do not assume the symmetric property of A_l, B_l. Our first result is the following

Theorem 4.1: Assume that

$$\begin{aligned}&\bar{a}_i^{(1)} + \bar{a}_i^{(2)} < \underline{a}_i^{(0)}[1 - \exp(-\underline{a}_i^{(0)}L)]^{-1},\\ &\bar{b}_i^{(1)} + \bar{b}_i^{(2)} < \underline{b}_i^{(0)}[1 - \exp(-\underline{b}_i^{(0)}L)]^{-1},\end{aligned} \qquad (i = 1, \ldots, n). \tag{4.2}$$

Then any maximal solution $(\bar{u}, \bar{v})$ coincides with the minimal solution $(\underline{u}, \underline{v})$. Furthermore, the problem (1.1), (1.2) has at most one nonnegative solution.

Proof: Let $u = \bar{u} - \underline{u}$, $v = \bar{v} - \underline{v}$. Then $u \geq 0$, $v \geq 0$ and (u, v) satisfies the system (1.1), (1.2) with $p = q = u_0 = v_L = 0$. From the integral representation of the system, the pair (u, v) satisfies the integral equations [see (2.7)]

$$\begin{aligned}u(x) &= \int_0^x D_\alpha(\xi, x)[A_1(\xi)u(\xi) + A_2(\xi)v(\xi)]\,d\xi,\\ &\qquad\qquad (x \in [0, L]).\\ v(x) &= \int_x^L D_\beta(x, \xi)[B_1(\xi)u(\xi) + B_2(\xi)v(\xi)]\,d\xi,\end{aligned} \tag{4.3}$$

In terms of the components of u, v, Eq. (4.3) is equivalent to

$$\begin{aligned}u_i(x) &= \int_0^x \alpha_i(\xi, x)\left(\sum_{j=1}^{n} [a_{ij}^{(1)}(\xi)u_j(\xi) + a_{ij}^{(2)}(\xi)v_j(\xi)\right)d\xi,\\ &\qquad\qquad (i = 1, \ldots, n).\\ v_i(x) &= \int_x^L \beta_i(x, \xi)\left(\sum_{j=1}^{n} [b_{ij}^{(1)}(\xi)u_j(\xi) + b_{ij}^{(2)}(\xi)v_j(\xi)]\right)d\xi,\end{aligned} \tag{4.4}$$

Let the indices i_0, i_1 and the points $x_0, x_1 \in [0, L]$ be such that

$$\begin{aligned}u_{i_0}(x_0) &= \max_{i=1,\ldots,n}\left\{\max_{0 \leq x \leq L}[u_i(x)]\right\},\\ v_{i_1}(x_1) &= \max_{i=1,\ldots,0}\left\{\max_{0 \leq x \leq L}[v_i(x)]\right\}.\end{aligned} \tag{4.5}$$

Then by letting $x = x_0$, $i = i_0$ in the first equation of (4.4) and $x = x_1$, $i = i_1$ in the second equation we obtain

$$\begin{aligned}u_{i_0}(x_0) &\leq \int_0^{x_0} \exp[-\underline{a}_{i_0}^{(0)}(x_0 - \xi)]\\ &\quad\times\left[\left(\sum_{j=1}^{n} a_{i_0j}^{(1)}(\xi)\right)u_{i_0}(x_0) + \left(\sum_{j=1}^{n} a_{i_0j}^{(2)}(\xi)\right)v_{i_1}(x_1)\right]d\xi\\ &\leq (\underline{a}_{i_0}^{(0)})^{-1}[1 - \exp(-\underline{a}_{i_0}^{(0)}x_0)]\\ &\quad\times[\bar{a}_{i_0}^{(1)}u_{i_0}(x_0) + \bar{a}_{i_0}^{(2)}v_{i_1}(x_1)],\end{aligned} \tag{4.6}$$

$$\begin{aligned}v_{i_1}(x_1) &\leq \int_{x_1}^{L} \exp[-\underline{b}_{i_1}^{(0)}(\xi - x_1)]\\ &\quad\times\left[\left(\sum_{j=1}^{n} b_{i_1j}^{(1)}(\xi)\right)u_{i_0}(x_0) + \left(\sum_{j=1}^{n} b_{i_1j}^{(2)}(\xi)\right)v_{i_1}(x_1)\right]d\xi\\ &\leq (\underline{b}_{i_1}^{(0)})^{-1}\{1 - \exp[-\underline{b}_{i_1}^{(0)}(L - x_1)]\}\\ &\quad\times[\bar{b}_{i_1}^{(1)}u_{i_0}(x_0) + \bar{b}_{i_1}^{(2)}v_{i_1}(x_1)].\end{aligned} \tag{4.7}$$

In obtaining the inequalities in (4.6), (4.7) we have used the nonnegative properties of the components of u, v, A_l, B_l and the fact that $\alpha_i(\xi, x) \leq \exp[-\underline{a}_i^{(0)}(x - \xi)]$ and $\beta_i(x, \xi) \leq \exp[-\underline{b}_i^{(0)}(\xi - x)]$. Now if $u_{i_0}(x_0) \geq v_{i_1}(x_1)$, then, by (4.6),

$$\begin{aligned}u_{i_0}(x_0) &\leq (\underline{a}_{i_0}^{(0)})^{-1}[1 - \exp(-\underline{a}_{i_0}^{(0)}x_0)](\bar{a}_{i_0}^{(1)} + \bar{a}_{i_0}^{(2)})u_{i_0}(x_0)\\ &\leq (\underline{a}_{i_0}^{(0)})^{-1}[1 - \exp(-\underline{a}_{i_0}^{(0)}L)](\bar{a}_{i_0}^{(1)} + \bar{a}_{i_0}^{(2)})u_{i0}(x_0).\end{aligned} \tag{4.8}$$

However, the condition (4.2) shows that (4.8) cannot hold unless $u_{i_0}(x_0) = 0$. Hence we must have $u_i(x) = v_i(x) = 0$ for every $i = 1, \ldots, n$ since u_i, v_i are all nonnegative on $[0, L]$. In case $v_{i_1}(x_1) \geq u_{i_0}(x_0)$, then (4.7) implies that

$$v_{i_1}(x_1) \leq (\underline{b}_{i_1}^{(0)})^{-1}[1 - \exp(-\underline{b}_{i_1}^{(0)}L)](\bar{b}_{i_1}^{(1)} + \bar{b}_{i_1}^{(2)})v_{i_1}(x_1). \tag{4.9}$$

This is again impossible unless $v_{i_1}(x_1) = 0$. By the nonnegative property of u_i, v_i we have again $u_i(x) = v_i(x) = 0$ for each i. This proves that $\bar{u} = \underline{u}$, $\bar{v} = \underline{v}$. To show the uniqueness problem we let (u^*, v^*) be any nonnegative solution of (1.1), (1.2). Since every nonnegative solution is also an upper solution, the sequence $\{\bar{u}^{(k)}, \bar{v}^{(k)}\}$ with $\bar{u}^{(k)} = u^*$, $\bar{v}^{(k)} = v^*$ for every $k = 1, 2, \cdots$ is clearly a maximal sequence and converges to (u^*, v^*). It follows from the conclusion of the first part of the theorem that $u^* = \underline{u}$, $v^* = \underline{v}$. Hence the problem (1.1), (1.2) cannot have more than one nonnegative solution. This completes the proof of the theorem.

It is to be noted that in proving the uniqueness problem in the above theorem we have used the nonnegative property of $\bar{u} - \underline{u} \geq 0$, $\bar{v} - \underline{v} \geq 0$ which may not hold for arbitrary pair of nonnegative solutions of (1.1), (1.2). This is another interesting application of the monotone approach for boundary-value problems.

The results of Theorem 3.2 and 4.1 lead immediately to the following conclusion:

Theorem 4.2: If (3.18), (4.2) hold, then the problem (1.1), (1.2) has exactly one nonnegative solution and thus the system is subcritical. Moreover, this solution can be determined from the recursion formula (2.7) with $u^{(0)} = v^{(0)} = 0$.

Corollary: If A_l, B_l are symmetric and if either

$$\underline{a}_i^{(0)} = \underline{b}_i^{(0)}, \quad \bar{a}_i^{(1)} + \bar{a}_i^{(2)} \leq \underline{a}_i^{(0)}, \quad \bar{b}_i^{(1)} + \bar{b}_i^{(2)} \leq \underline{b}_i^{(0)}, \qquad i = 1, \ldots, n \tag{4.10}$$

or the conditions (3.19), (4.2) hold, then the problem (1.1), (1.2) has a unique nonnegative solution.

Proof: Since the condition (4.10) implies both (3.13) and (4.2) for any $L < \infty$, the conclusion for the first case follows from Theorem 4.1 and the Corollary to Theorem 3.1. In fact, the existence of a unique solution is independent of L. In case (3.19), (4.2) hold, then by (4.2) the condition (3.21) is satisfied. Since (3.21) and (3.19) imply (3.12), the conclusion for the second case follows from Theorems 3.1 and 4.1.

[1]G.M. Wing, *An Introduction to Transport Theory* (Wiley, New York, 1962).
[2]R. Bellman, *Matrix Analysis* (McGraw-Hill, New York, 1970), 2nd ed.
[3]C.V. Pao, J. Math. Phys. **16**, 1134 (1975).
[4]R.C. Allen, Jr., and G.M. Wing, J. Math. Anal. Appl. **29**, 141 (1970).
[5]R. Bellman, S. Ueno, and R. Vasudevan, J. Math. Anal. Appl. **44**, 472 (1973).
[6]G.H. Meyer, *Initial Value Methods for Boundary Value Problems* (Academic, New York, 1973).
[7]J.O. Mingle, J. Math. Anal. Appl. **38** (1972).
[8]R. Bellman, J. Math. Anal. Appl. **43**, 232 (1973).
[9]C.V. Pao, J. Math. Anal. Appl. **49**, 545 (1975).
[10]W.R. Boland and P. Nelson, Jr., Appl. Math. Computation **1**, 67 (1975).
[11]K.M. Case and P.F. Zweifel, J. Math. Phys. **4**, 1376 (1963).
[12]P. Nelson, Jr., J. Math. Anal. Appl. **35**, 90 (1971).