Use of Retarded Green's Functions in Exciton Theory*

CRAIG W. DEUTSCHE† AND C. ALDEN MEAD‡
School of Chemistry, University of Minnesota, Minneapolis, Minnesota
(Received 23 October 1964)

It is pointed out that the usual treatment of excitons in interaction with radiation with spatial dispersion does not obviously satisfy causality in all cases, and it is shown explicitly that certain paradoxes and ambiguities can appear even in simple cases. In particular, it is shown for a simple special case that an apparently reasonable choice of boundary condition leads to a reflection coefficient greater than unity; and also that in some cases one must make rather arbitrary and inconsistent requirements in order to get the right number of boundary conditions to determine a solution. The theory is reformulated in a way that is manifestly causal, with both field and polarization being generated by retarded Green's functions. Umklapp processes are also included. This formulation confirms the results of the usual treatment in the cases in which the latter are unambiguous and acceptable, but avoids the difficulties mentioned above. The model used is essentially classical and one dimensional, but nevertheless sufficiently general for the treatment of these problems.

I. INTRODUCTION AND SUMMARY

THE theory of excitons in interaction with radiation including "spatial dispersion" was first discussed by Pekar.[1] A somewhat different approach based on an idea due to Fano,[2] has been used by Hopfield,[3] and by Hopfield and Thomas.[4] In all these theories, umklapp processes are neglected. The result is that, if the exciton energy is a function of the wave number, there are, in general, two or more refractive indices found for each frequency and polarization, even in an isotropic crystal. The wave propagating in the crystal at a given frequency is then a superposition of several waves of different wave numbers. The coefficients are determined by extra boundary conditions based on a detailed consideration of interactions near the edge of the crystal.

One of us[5] has questioned the validity of these theories on the basis of causality. For a special case with two solutions, it was shown that limiting oneself to a single refractive index leads to a violation of the Kronig-Kramers theorem, and hence of causality. Causality can be formally restored by superimposing the two solutions in an appropriate way; but this would require the use of both solutions over the full frequency range. The derivation of the two solutions, however, requires the assumption of long wavelengths (neglect of umklapp processes), and the resulting solutions are such that there is only a rather narrow range for which this assumption is satisfied by both solutions. For this reason, the possibility of restoring causality in this way is referred to in Ref. 5 as a "remote possibility"; but we shall see that it is indeed the correct answer to the problem. In order to discuss it in a self-consistent way, however, it is clear that the theory must be generalized to include umklapp processes, which makes the usual procedure a good deal more difficult. It was also suggested in Ref. 5 that the apparent noncausality might be associated with the periodic boundary conditions customarily imposed on the radiation field, which seem to be incompatible with the use of the retarded Green's function. Hence, a formulation of the theory which is manifestly causal, and does not depend on unphysical boundary conditions, seems desirable.

The model we shall use is essentially classical and one dimensional, but is nevertheless sufficiently general to deal with the problems in which we are interested. In three dimensions, it might be visualized as a "crystal" consisting of infinite plane sheets of polarizable material, perpendicular to the z axis and spaced at regular intervals. The crystal interacts with a radiation field polarized in the x direction and propagating in the z direction. Thus, the vector character of polarization and field plays no role (there being only one nonzero component), and the z dimension is the only spatial dimension that needs to be considered. In addition, quantum-mechanical transitions of the material system are replaced by classical oscillators; for the justification of this, the reader is referred to the work of Fano[2] and Hopfield.[3] Hence, this model is not as unrealistic as it might appear at first glance; and in any case, the problems which we wish to discuss take precisely the same form in our model as in more realistic theories.

The plan of the paper is as follows: In Sec. II, the conventional treatment is discussed. It is shown that, in addition to the doubts about causality, certain paradoxes and ambiguities arise in the usual treatment, even for rather simple cases. In particular, it is shown that apparently reasonable boundary conditions can lead to totally unacceptable results, and that without making rather arbitrary assumptions one does not always arrive at the right number of boundary conditions for the number of refractive indices. In this section,

* This work was supported in part by the Directorate of Chemical Sciences, U. S. Air Force Office of Scientific Research, Grant No. AF AFOSR 13-63.
† National Science Foundation Fellow.
‡ Present address, September 1964 to September 1965: Department of Physics, Birkbeck College, London, England.
[1] S. I. Pekar, Zh. Eksperim. i Teor. Fiz. 34, 1176 (1958) [English transl.: Soviet Phys.—JETP 7, 813 (1958)].
[2] U. Fano, Phys. Rev. 103, 1202 (1956).
[3] J. J. Hopfield, Phys. Rev. 112, 1555 (1958).
[4] J. J. Hopfield and D. G. Thomas, Phys. Rev. 132, 563 (1963).
[5] C. A. Mead, Radiation Res. 20, 101 (1963).

following the usual procedure, the crystal is treated as continuous, i.e., umklapp processes are neglected.

Section III introduces the formulation in terms of causal (retarded) Green's functions. It is shown that the possibility of constructing a properly causal solution automatically entails a restriction on allowable forms for boundary conditions. In particular, those boundary conditions that led to unacceptable results in Sec. II are shown to be unacceptable.

In Sec. IV, the crystal is treated as discrete (all umklapp processes included), with interactions between oscillators of the most general form compatible with the existence of a solution expressible in terms of a finite number of refractive indices. Again retarded Green's functions are used, so the treatment is manifestly causal. An exact solution is obtainable, and it is shown that one always gets exactly the right number of boundary conditions. In the long-wavelength limit, the theory gives the same results as the usual treatment in all cases where the latter gives a definite answer; and in cases where the usual treatment is ambiguous, the present theory leads to a unique result. There is some discussion of the results in Sec. V.

II. PROBLEMS WITH USUAL TREATMENT

The possible difficulty with causality in the usual theories has already been mentioned in Sec. I, and the reader is referred to Ref. 5 for further details. It is appropriate at this point to discuss briefly the consequences of causality in a medium which is not describable by a single refractive index, but which nevertheless responds linearly to incident radiation. Since we shall always be discussing semi-infinite mediums, infinite in the positive z direction, we may confine ourselves to this case. For an incident wave with vector potential given by

$$A_{\text{in}}(z,t)=\int_{-\infty}^{\infty} d\omega A_{\text{in}}(\omega)e^{i\omega(z-t)},$$

we have in general for the transmitted vector potential within the medium

$$A_t(z,t)=\int_{-\infty}^{\infty} d\omega A_{\text{in}}(\omega)B_T(\omega,z)e^{-i\omega t}.$$

Causality requires (at least) that for an incident wave with a sharp front reaching the edge of the medium at time $t=0$,

$$A_t(z,t)=0 \quad \text{for} \quad t<0.$$

This leads in the usual way[6] to the requirement that $B_T(\omega,z)$ be an analytic function of ω in the upper half-plane for all positive z. In the case where there is a single refractive index $\rho(\omega)$, $B_T(\omega,z) =T(\omega)\exp[i\rho(\omega)\omega z]$, and the analyticity of B_T requires the analyticity of ρ. If there are two or more refractive indices, however, one cannot in general require that both be separately analytic. From the analyticity of B_T, one may conclude that if the solution is known for a finite range of frequencies it can be found for all frequencies by analytic continuation; this means, for example, that one must have the same number of refractive indices for all frequencies.

It is also necessary at this point to say a few words about absorption. The model that we are considering does not include the coupling of excitons to phonons, which is the chief damping mechanism leading to absorption in actual crystals. Even without phonons, in a real three-dimensional crystal it might be conceivable that umklapp processes in the x and y directions might bring about an attenuation of the primary beam propagating in the z direction, leading to apparent absorption. Be that as it may, however, it is clear that such processes cannot take place in our one-dimensional model; there simply is not place for the energy to go except into waves propagating in the z direction. Hence it is clear that there cannot be absorption in this case; reflected plus transmitted intensity must equal incident intensity, purely on grounds of energy conservation.

In order to illustrate the problems that can arise, let us consider a particularly simple special case. Suppose we have oscillators (or sheets) with natural angular frequency ν_0 located along the z axis at points

$$z=n\alpha,\quad n=0,1,2,3,\cdots,$$

where α is, of course, the lattice constant. If the dipole moment of the nth oscillator (dipole density per unit area of nth sheet) is μ_n, then the corresponding dipole density is

$$P_n=\mu_n/\alpha.$$

We further suppose that nearest-neighbor oscillators are connected by Hooke's-law springs. This leads to the equation of motion for the nth oscillator:

$$\ddot{P}_n+\nu_0^2P_n-q[(P_{n+1}-P_n)+(P_{n-1}-P_n)] =-\epsilon^2\dot{A}(n\alpha),\quad (1)$$

where q is the spring constant, ϵ^2 is a coupling constant involving the charge and mass of the oscillators, A is the vector potential, and we have used units in which $c=1$. A dot denotes time differentiation.

Equation (1) holds for all oscillators except $n=0$, which has only one nearest neighbor. For this oscillator, Eq. (1) is replaced by

$$\ddot{P}_0+\nu_0^2P_0-q(P_1-P_0)=-\epsilon^2\dot{A}(0).\quad (2)$$

To treat this problem by the usual method, we neglect umklapp processes by assuming that α is small compared with wavelengths of interest. In this case, P can be treated approximately as a continuous function

[6] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, translated by J. B. Sykes and J. S. Bell (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1960), pp. 256–268.

of z, and we can expand P_{n+1} and P_{n-1} in forms similar to

$$P_{n+1}-P_n \cong \alpha P'(\alpha n)+\tfrac{1}{2}\alpha^2 P''(\alpha n),$$

where a prime denotes differentiation with respect to z. Inserting this into Eq. (1), we obtain

$$\ddot{P}+\nu_0^2 P-\beta P''=-\epsilon^2 \dot{A}, \tag{3}$$

where $\beta=q\alpha^2$. At the edge, Eq. (2) becomes

$$\ddot{P}(0)+\nu_0^2 P(0)-\beta\left(\frac{1}{\alpha}P'(0)+\tfrac{1}{2}P''(0)\right)=-\epsilon^2\dot{A}(0). \tag{4}$$

In order for both (3) and (4) to hold, we must have the boundary condition at the edge

$$P'(0)-(\alpha/2)P''(0)=0, \tag{5}$$

or approximately

$$P'(0)=0. \tag{6}$$

Of course, other types of interactions would lead to other boundary conditions. For example, Pekar[1] finds $P(0)=0$ instead of (6).

The Maxwell equation for the vector potential A is (in Heaviside units)

$$\ddot{A}-A''=\dot{P}. \tag{7}$$

We now seek to find a refractive index $\rho(\omega)$ by seeking solutions of (3) and (7) of the form

$$\begin{aligned} A&=A_0\exp\{i\omega[\rho(\omega)z-t]\}; \\ P&=P_0\exp\{i\omega[\rho(\omega)z-t]\}. \end{aligned} \tag{8}$$

Inserting (8) into (3) and (7), we easily obtain

$$P_0=i\omega[\rho(\omega)^2-1]A_0, \tag{9}$$

and

$$\rho^2(\omega)=\frac{(\beta+1)\omega^2-\nu_0^2\pm\{[(1-\beta)\omega^2-\nu_0^2]^2+4\beta\epsilon^2\omega^2\}^{1/2}}{2\beta\omega^2}. \tag{10}$$

According to (10), there are two solutions for ρ^2, and therefore four for ρ itself. The general solution to (3) and (7) for each frequency is, therefore, a superposition of four solutions of the form (8), (9), corresponding to the four values of ρ. We are mainly interested, however, in the waves set up in the crystal by an incident light wave propagating in the positive z direction. For this case, we can also require

$$\mathrm{Re}\rho\geq 0, \tag{11}$$

and

$$\mathrm{Im}\rho\geq 0. \tag{12}$$

Equation (11) corresponds to the requirement that the wave propagate forward, not backward, in the medium; more precisely, it can be shown to be a consequence of the requirement that the time average of the Poynting vector associated with a single wave be in the positive z direction. Equation (12) simply requires that the wave be attenuated, not amplified, as it propagates. Since our crystal is infinite in the positive z direction, this requirement is certainly essential. If β is positive, it is easy to see from (10) that both values of ρ^2 are real for all ω. Hence there will always be exactly two values of ρ that satisfy (10), (11) and (12). Denoting these by ρ_1 and ρ_2, we can fit boundary conditions quite easily for an incident wave

$$A=e^{i\omega(z-t)}$$

by trying a solution of the form

$$A=e^{i\omega(z-t)}+\mathrm{R}e^{-i\omega(z+t)}, \qquad z<0; \tag{13}$$

$$A=T_1e^{i\omega(\rho_1 z-t)}+T_2e^{i\omega(\rho_2 z-t)}, \quad z>0. \tag{14}$$

The unknowns are the reflection coefficient R, and the two transmission coefficients T_1 and T_2. From the usual boundary conditions we have

$$1+R=T_1+T_2; \tag{15}$$

$$1-R=\rho_1T_1+\rho_2T_2. \tag{16}$$

Equations (15) and (16) represent, respectively, the continuity of E and H at the edge. The third condition required to determine the three unknowns is provided by the boundary condition (5) or (6). For example, if (6) is used, we find

$$\rho_1(\rho_1^2-1)T_1+\rho_2(\rho_2^2-1)T_2=0. \tag{17}$$

In this case, therefore, the solution is completely and unambiguously determined by the general requirements which we have made.

One may ask at this point whether the situation is always as happy as in the foregoing treatment; that is, does one always obtain exactly the right number of boundary conditions to fix the solution. In this connection, it should be remembered that more complicated constitutive relations than (3) may lead to more than two refractive indices. In fact, one can even get into difficulties with (3), as we shall now see.

Suppose that we allow β to be negative in (3). This corresponds to a negative effective mass for excitons, or to a tendency for nearest-neighbor dipoles to point in opposite directions. The solution goes through exactly as before, leading again to the result (10). With β negative, however, ρ^2 is no longer real for all frequencies. A glance at (10) reveals that there is a range in which both values of ρ^2 are complex, being complex conjugates of each other. In this range, there is only *one* value of ρ which satisfies both (11) and (12). There are two possible approaches that one might take here:

(a) One might use only the one value of ρ that satisfies (11) and (12). (By analytic continuation, this means that there is only one refractive index for all frequencies.) If one does this, however, it is impossible to satisfy simultaneously all three boundary conditions [continuity of E and H at the boundary, plus Eq. (5)

or (6)]. Moreover, the single refractive index that one gets is complex over a finite range, so that this approach leads to absorption contrary to energy conservation.

(b) One might drop the requirement (11), arguing that it is only the full solution that must propagate forward, not its parts separately. If one does this, one finds two refractive indices satisfying (12), the boundary conditions can all be satisfied, and an easy calculation shows that the reflection coefficient is unity in the region of complex ρ, so there is no true absorption. If Eq. (11) does not represent a valid general requirement, however, how are we justified in requiring it for the case of positive β? For positive β, there are always at least *three* values of ρ satisfying (12), so if we drop the requirement (11) we do not have enough boundary conditions to fix the solution. It seems, therefore, that this procedure is not completely satisfactory either. We are forced to conclude that the usual treatment does not contain a simple prescription which will lead to a unique solution in all cases.

We now present another difficulty. Of the two boundary conditions (5) and (6), it would appear at first glance that (5) is a better approximation to the true (discrete) state of affairs, and should be preferred. We shall show, however, that the use of Eq. (5), far from leading to an improvement, actually gives results which are totally unacceptable.

If Eq. (5) is used, the condition (17) is replaced by

$$\xi_1 T_1+\xi_2 T_2=0, \tag{18}$$

where

$$\xi_j=(\rho_j^2-1)[i\omega\rho_j+\tfrac{1}{2}\alpha\omega^2\rho_j^2]. \tag{19}$$

Solving (15), (16), and (18) for the reflection coefficient R, we find

$$R=\frac{(1-\rho_1)\xi_2-(1-\rho_2)\xi_1}{(1+\rho_1)\xi_2-(1+\rho_2)\xi_1}. \tag{20}$$

We now note from Eq. (10), that one of the solutions (ρ_2, say) is equal to zero when $\omega^2=\nu_0^2+\epsilon^2$. If ω^2 is slightly less than this, we have

$$\rho_2=i\delta, \tag{21}$$

where δ is a small positive number. At the same time, the other solution

$$\rho_1\equiv\rho \tag{22}$$

is real and greater than unity, as may be easily verified. We will now calculate R in this frequency range, to lowest order in δ. We consider ω to be positive. To first order in δ, we have

$$\xi_2=(-\delta^2-1)(-\omega\delta-\tfrac{1}{2}\alpha\omega^2\delta^2)\approx\omega\delta. \tag{23}$$

At the same time, we find for ξ_1:

$$\xi_1=\alpha u+iv, \tag{24}$$

where u and v are both positive. We now substitute (23) and (24) into (20), to find R up to first order in δ:

$$\begin{aligned}R&\approx\frac{(\rho-1)\omega\delta-(1-i\delta)(\alpha u+iv)}{(\rho+1)\omega\delta-(1+i\delta)(\alpha u+iv)}\\&\approx\frac{\alpha u+iv+\delta[\omega(\rho-1)-i(\alpha u+iv)]}{\alpha u+iv-\delta[\omega(\rho+1)-i(\alpha u+iv)]}\\&\approx1+\delta(2\rho\omega/(\alpha u+iv))-2\delta i\\&=1+2\delta\rho\omega\frac{\alpha u-iv}{\alpha^2u^2+v^2}-2i\delta.\end{aligned} \tag{25}$$

The reflected intensity per unit incident intensity is

$$RR^*\approx1+[4\rho\omega\alpha u/(\alpha^2u^2+v^2)]\delta>1. \tag{26}$$

Thus, the boundary condition (5) leads to the obviously unacceptable conclusion that the reflected intensity will be greater than the incident in this frequency range. The fact that RR^*-1 is proportional to α makes it clear that this would not have happened if we had used (6). Hence, an attempt to improve the boundary condition has led to disaster rather than improvement.

The considerations of this section show, we feel, that a more careful and general formulation of the theory is desirable. We construct such a formulation in the following sections.

III. GREEN'S FUNCTION FORMULATION—BOUNDARY CONDITIONS IN CONTINUOUS CASE

In this section, we wish to consider the problem in somewhat greater generality than before, and to express the solution formally in a manifestly causal manner.

Equation (1) or (3), determining the polarization, may be generalized to[7]

$$\ddot{P}+\mathcal{L}P=-\epsilon^2\dot{A}, \tag{27}$$

where $\mathcal{L}$ is some linear operator representing the natural frequencies and interactions of the oscillators. In general, P may be a discrete function defined only for $z=n\alpha$, or it may be a function of a continuous variable z. We will use continuum notation in this section. The vector potential A still obeys (7).

We are interested in solutions of (7) and (27) which represent the causal response of an otherwise unpolarized medium to an incident field. Accordingly, the appropriate solution to (7) is

$$A(z,t)=A_{\text{in}}(z,t)+\int_{-\infty}^{\infty}dt'\int_0^{\infty}dz'G^{\text{ret}}(z-z',t-t')\dot{P}(z',t'), \tag{28}$$

where $G^{\text{ret}}(z,t)$ is the retarded Green's function for the

[7] If desired, damping may be introduced in a phenomenological way by replacing $\ddot{P}$ by $\ddot{P}+g\dot{P}$.

electromagnetic field:

$$G^{\mathrm{ret}}(z,t)=\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{e^{i(\kappa z-\omega t)}d\kappa d\omega}{\kappa^2-(\omega+i\delta)^2}, \qquad (29)$$

in which the limit $\delta\to 0+$ is to be understood. The integral over z' in (28) goes from zero to infinity, of course, because that is the region occupied by the medium. In the solution we want, the polarization is entirely that generated by the field, i.e., there is no "P_{in}." Accordingly, the solution of (27) which we should use is

$$P(z,t)=\int_{-\infty}^{\infty}dt'\int_{-\infty}^{\infty}dz' F^{\mathrm{ret}}(z,z',t-t')[-\epsilon^2 A(z',t')]. \quad (30)$$

Here F^{ret} is a retarded Green's function associated with the operator $\mathcal{L}$. It is written as a function of (z,z') instead of $(z-z')$ because the fact that P is defined only for positive z destroys the translational symmetry. If $\mathcal{L}$ possesses a continuous eigenvalue spectrum with normalized real eigenfunctions $\psi_k(z)$, the F^{ret} has the form

$$F^{\mathrm{ret}}(z,z',t)=\frac{1}{2\pi}\int dk\int_{-\infty}^{\infty}d\omega\frac{\psi_k(z)\psi_k(z')e^{-i\omega t}}{\lambda^2(k)-(\omega+i\delta)^2}. \quad (31)$$

The range of the k integration is unspecified since this depends on the properties of $\mathcal{L}$. $\lambda^2(k)$ is the eigenvalue, defined by

$$\mathcal{L}\psi_k(z)=\lambda^2(k)\psi_k(z).$$

The generalization to the case in which $\mathcal{L}$ has discrete eigenvalues is obvious.

In order for F^{ret} to be a true retarded Green's function, we must have

$$F^{\mathrm{ret}}(z,z',t)=0 \quad \text{for} \quad t<0.$$

This simply means, referring to (30), that there is no polarization anywhere in the medium until there is a field somewhere in the crystal. At negative times, the integration over ω in (31) is carried out by closing the contour in the upper half-plane, so in order to get zero there must be no singularities in the upper half-plane. There are singularities, however, whenever

$$\omega=\pm\lambda(k).$$

We conclude, therefore, that in order for a causal solution to be possible all the $\pm\lambda(k)$ must lie on or below the real axis, which means that the eigenvalues $\lambda^2(k)$ must all be real and positive; in other words, the operator $\mathcal{L}$ must be Hermitian and positive definite.

In the case of functions defined only for positive z, such as we have here, it should be remembered that the hermiticity of an operator depends not only on the form of the operator but also on the boundary conditions. For example, if we use Eq. (3),

$$\mathcal{L}=\nu_0^2-\beta(\partial^2/\partial z^2), \qquad (32)$$

and $\mathcal{L}$ is Hermitian only if the operator $\partial^2/\partial z^2$ is Hermitian. In order for this operator to be Hermitian, we must have

$$\int_0^{\infty}(\psi\phi''-\phi\psi'')dz=0$$

for any allowable functions ψ and ϕ. However, we find

$$\int_0^{\infty}(\psi\phi''-\phi\psi'')dz=\psi'(0)\phi(0)-\phi'(0)\psi(0).$$

This is zero only if the boundary condition obeyed by all allowed functions is

$$\psi(0)+\theta\psi'(0)=0, \qquad (33)$$

where θ is some constant. In particular, the boundary condition (5) is not of this form, and is therefore not allowed. Hence, if one has an operator of the form (32), one may attempt to represent approximately the physical conditions at the edge of the crystal by varying the parameter θ in (33); but any attempt to go beyond this will make a causal solution impossible: there will then always be some polarization which is not "caused" by the field, and this "spontaneous" polarization will lead to unphysical effects such as increased reflected intensity.

If $\mathcal{L}$ is given by (32), its eigenvalues are

$$\lambda^2(k)=\nu_0^2+\beta k^2.$$

This is not positive definite if β is negative, so one might be tempted to use the positive-definiteness requirement to dispose of the dilemma raised in the last section by negative β. This will not do, however, since the same problem arises for

$$\lambda^2(k)=\nu_0^2-2|\beta|(1-\cos k), \qquad (33a)$$

which is positive definite if

$$|\beta|<\nu_0^2/2.$$

The general treatment of this problem requires that the crystals be treated as discrete. We take this up in the next section.

IV. DISCRETE MODEL

A. Exciton Eigenfunctions

In this section we consider the oscillators of our crystal to be located at points regularly spaced along the z axis. The unit of length is chosen so that the distance between neighboring oscillators is unity. It will be convenient to number the oscillators from one to infinity. Thus, the polarization is given by

$$P(z,t)=\sum_{n=1}^{\infty}P_n(t)\delta(z-n),$$

and Eq. (27) becomes

$$\ddot{P}_n(t)+\sum_{n'=1}^{\infty}\mathcal{L}_{nn'}P_{n'}(t)=-\epsilon^2\dot{A}(n,t). \quad (34)$$

Equation (7) takes the form

$$\ddot{A}(z,t)-A''(z,t)=\sum_{n=1}^{\infty}\dot{P}_n(t)\delta(z-n). \quad (35)$$

We have already seen that the matrix $\mathcal{L}$ must be Hermitian and positive definite in order for a causal solution to exist. This requirement alone, however, is not sufficiently restrictive for our purposes. In the type of theory in which we are interested, the optical properties of the medium are represented in terms of a finite number of refractive indices, and the waves propagating in the medium are sinusoidal, perhaps with exponential attenuation. It is clear that an exact solution of (34) and (35) will yield results of this type only if the eigenfunctions of $\mathcal{L}$ are sinusoidal (or exponential) waves. We therefore require that $\mathcal{L}$ be of such a form as to have sinusoidal waves for eigenfunctions. Our treatment will not be the most general possible, therefore, but it will be the most general compatible with the use of refractive indices, and therefore certainly general enough to discuss the problems raised in Sec. II. Of course, in a real crystal there will normally be effects due to the distortion of waves from the sinusoidal form near the edge (a simple model for such effects has been discussed by Hopfield and Thomas[4]), but such effects cannot be treated by means of the usual theories and will not be treated here.

Rather than try to decide what forms of $\mathcal{L}$ will lead to the desired eigenfunctions, it is easier to postulate the eigenfunctions and eigenvalues and then work backwards to find the matrix $\mathcal{L}$. Accordingly, we require that the eigenfunctions of $\mathcal{L}$ be of the form

$$\psi_n(k)=a(k)\sin kn+b(k)\cos kn, \quad (36)$$

where we can require without loss of generality that $a(k)$ and $b(k)$ are real. The range of k is

$$0\leqslant k<\pi.$$

The coefficients $a(k)$ and $b(k)$ must be chosen in such a way that the set is complete and orthonormal. We first consider the orthonormality requirement. Using (36), it is a simple matter to show that

$$\sum_{n=1}^{\infty}\psi_n(k)\psi_n(k') =\tfrac{1}{2}\pi[a(k)a(k')+b(k)b(k')]\delta(k-k')-\tfrac{1}{2}b(k)b(k') +\frac{1}{2}\left[\frac{a(k)b(k')\sin k-b(k)a(k')\sin k'}{\cos k'-\cos k}\right]. \quad (37)$$

In arriving at (37) we have used the completeness property

$$\sum_{n=-\infty}^{\infty}e^{inx}=2\pi\sum_{j=-\infty}^{\infty}\delta(x-2\pi j),$$

and also the fact that the range of k prevents $k-k'$ or $k+k'$ from being equal to 2π. Where necessary, convergence factors e^{-gn}, $g\to 0+$ have been introduced. For orthogonality, then, it is necessary that the non-δ-function terms in (37) vanish for all k, k'. This will happen if and only if

$$a(k)=-b(k)((\cos k+\xi)/\sin k),$$

where ξ is a constant.[8] This can also be expressed by

$$a(k)=f(k)(\cos k+\xi);\quad b(k)=-f(k)\sin k. \quad (38)$$

For normalization we must also have

$$\pi/2[a^2(k)+b^2(k)]=1 =f^2(k)[\xi^2+2\xi\cos k+1]. \quad (39)$$

Hence,

$$f(k)=(2/\pi)^{1/2}[\xi^2+2\xi\cos k+1]^{-1/2}. \quad (40)$$

We find then for the eigenfunction $\psi_n(k)$:

$$\psi_n(k)=(2/\pi)^{1/2}[\xi^2+2\xi\cos k+1]^{-1/2}\times[(\xi+\cos k)\sin kn-\sin k\cos kn] =(2/\pi)^{1/2}[\xi^2+2\xi\cos k+1]^{-1/2}\times[\xi\sin kn+\sin k(n-1)]. \quad (41)$$

The set of functions defined by (41) is an orthonormal set. We must now examine the question of its completeness. The criterion for this is

$$\int_0^{\pi}\psi_n(k)\psi_{n'}(k)dk=\delta_{nn'}.$$

With the functions (41), we find

$$\int_0^{\pi}\psi_n(k)\psi_{n'}(k)dk =-\frac{1}{4\pi}\int_0^{2\pi}[(\xi+e^{-ik})e^{ikn}-(\xi+e^{ik})e^{-ikn}] \times\frac{[(\xi+e^{-ik})e^{ikn'}-(\xi+e^{ik})e^{-ikn'}]}{(\xi+e^{ik})(\xi+e^{-ik})}. \quad (42)$$

In arriving at (42), we have used the evenness and periodicity of the integrand to extend the range of integration from zero to 2π. We have also written the

[8] The role of the constant ξ is similar to that of θ in (33). In the limit of small k, we get from (38)

$$\psi_n(k)=f(k)[(1+\xi)\sin kn-k\cos kn],$$

while (33) would give for sine and cosine functions

$$\psi_n(k)=f(k)[1/\theta\sin kn-k\cos kn].$$

Hence we see that $\theta=(1+\xi)^{-1}$.

sine and cosine in exponential form and done some factoring. With a little rearrangement, we further obtain

$$\int_0^\pi \psi_n(k)\psi_{n'}(k)dk$$
$$=-\frac{1}{4\pi}\int_0^{2\pi}\left\{\left(\frac{\xi+e^{-ik}}{\xi+e^{ik}}\right)e^{ik(n+n')}+\left(\frac{\xi+e^{ik}}{\xi+e^{-ik}}\right)e^{-ik(n+n')}-e^{ik(n-n')}-e^{ik(n'-n)}\right\}dk \quad (43)$$
$$=\delta_{nn'}-\frac{1}{4\pi}\int_0^{2\pi}\left(\frac{\xi+e^{-ik}}{\xi+e^{ik}}\right)e^{ik(n+n')}-\frac{1}{4\pi}\int_0^{2\pi}\left(\frac{\xi+e^{ik}}{\xi+e^{-ik}}\right)e^{-ik(n+n')}. \quad (44)$$

We now make the change of variables $u=e^{ik}$ in the first integral on the right-hand side of (44), and $u=e^{-ik}$ in the second. Equation (44) now becomes

$$\int_0^\pi \psi_n(k)\psi_{n'}(k)dk=\delta_{nn'}-\frac{1}{2\pi i}\oint\frac{(\xi+u^{-1})u^{n+n'}du}{u(\xi+u)} \quad (45)$$

in which the contour goes around the unit circle. Since always $n+n'\geq 2$, the only possible pole is at $u=-\xi$. There are now two possible cases:

(a) $|\xi|\geq 1$. In this case the pole is outside the contour, so the integral in (45) is zero and we obtain

$$\int_0^\pi \psi_n(k)\psi_{n'}(k)dk=\delta_{nn'}. \quad (46)$$

Hence, the set is complete.

(b) $|\xi|<1$. Here we get a contribution from the pole, giving the result

$$\int_0^\pi \psi_n(k)\psi_{n'}(k)dk=\delta_{nn'}-(\xi^{-2}-1)(-\xi)^{n+n'}. \quad (47)$$

The set is not complete in this case, but it can be made so by supplementing it with

$$\phi_n=(\xi^{-2}-1)^{1/2}(-\xi)^n. \quad (48)$$

It is easily verified that ϕ_n is normalized and orthogonal to all the $\psi_n(k)$. Using (47) and (48), one sees immediately that the augmented set is complete;

$$\int_0^\pi \psi_n(k)\psi_{n'}(k)dk+\phi_n\phi_{n'}=\delta_{nn'}. \quad (49)$$

The function ϕ_n is seen to have the properties of a "surface exciton."

The sets of functions defined by (41) with $|\xi|\geq 1$, or by (41) and (48) with $|\xi|<1$ exhaust the possibilities for eigenfunctions acceptable to us. In the next subsection we consider the eigenvalues, and deduce the form of the matrix $\mathfrak{L}$.

B. Eigenvalues and Interactions

The eigenvalues of the matrix $\mathfrak{L}$ are defined by

$$\sum_{n'=1}^{\infty}\mathfrak{L}_{nn'}\psi_{n'}(k)=\lambda^2(k)\psi_n(k), \quad (50)$$

and in the case $|\xi|<1$, also by

$$\sum_{n=1}^{\infty}\mathfrak{L}_{nn'}\phi_{n'}=\Lambda^2\phi_n. \quad (51)$$

The eigenvalues $\lambda^2(k)$ and Λ^2 must, of course, be real and positive.

We first consider the case $|\xi|\geq 1$. Here we have

$$\mathfrak{L}_{nn'}=\int_0^\pi \lambda^2(k)\psi_n(k)\psi_{n'}(k)dk. \quad (52)$$

Apart from the factor $\lambda^2(k)$, the integrand is the same as that discussed in the last section. Thus we find, analogously to (43):

$$\mathfrak{L}_{nn'}=-\frac{1}{4\pi}\int_0^{2\pi}\lambda^2(k)\left\{\left(\frac{\xi+e^{-ik}}{\xi+e^{ik}}\right)e^{ik(n+n')}+\left(\frac{\xi+e^{ik}}{\xi+e^{-ik}}\right)e^{-ik(n+n')}-e^{ik(n-n')}-e^{ik(n'-n)}\right\}dk$$
$$=-\frac{1}{2\pi}\mathrm{Re}\int_0^{2\pi}\lambda^2(k)\left\{\left(\frac{\xi+e^{-ik}}{\xi+e^{ik}}\right)e^{ik(n+n')}-e^{ik(n-n')}\right\}dk. \quad (53)$$

Since $\lambda^2(k)$ has physical meaning only in the range $0\leqslant k<\pi$, we can without loss of generality define it in other regions by requiring that it be an even function of k with period 2π. In this case we can write

$$\lambda^2(k)=\sum_l a_l \cos lk, \quad (54)$$

where the sum goes over $l=0, 1, 2, \cdots$ and may be finite or infinite. Combining (53) and (54) we obtain

$$\mathfrak{L}_{nn'}=-\frac{1}{4\pi}\mathrm{Re}\sum_l a_l\times\int_0^{2\pi}\left\{\left(\frac{\xi+e^{-ik}}{\xi+e^{ik}}\right)[e^{ik(n+n'+l)}+e^{ik(n+n'-l)}]-e^{ik(n-n'+l)}-e^{ik(n-n'-l)}\right\}dk.$$

We now evaluate the last two terms in the integral, and make the substitution $u=e^{ik}$ in the other, with the result

$$\mathcal{L}_{nn'}=\tfrac{1}{2}\sum_l a_l\Big\{\delta_{|n-n'|,l}-\frac{1}{2\pi}\mathrm{Re}\frac{1}{i}\oint\frac{(\xi+u^{-1})(u^{n+n'+l}+u^{n+n'-l})du}{u(\xi+u)}\Big\},$$

where the contour is again the unit circle. Because $|\xi|\geq 1$, the only contribution to the integral is from the pole at $u=0$, whose order depends on n, n', and l. We easily find the result

$$\mathcal{L}_{nn'}=\tfrac{1}{2}\sum_l a_l\{\delta_{|n-n'|,l}-\xi^{-1}\delta_{n+n',l+1}-(1-\xi^{-2})(-\xi)^{n+n'-l}h(l-n-n')\},\quad (55)$$

where

$$h(y)=1,\quad y\geq 0,\qquad h(y)=0,\quad y<0.\quad (56)$$

For each value of l, the first term in (55) is seen to represent an lth nearest-neighbor interaction, while the remaining terms exist only near the edge of the crystal $(n+n'\leqslant l+1)$, and are necessary to assure that the eigenfunctions remain sinusoidal.

The case $|\xi|<1$ is not greatly different. Here we have instead of (53):

$$\mathcal{L}_{nn'}=\int_0^\pi \lambda^2(k)\psi_n(k)\psi_{n'}(k)dk+\Lambda^2\phi_n\phi_{n'}.\quad (57)$$

There are two differences between this and the previous case: first, when we make the change of variables $u=e^{ik}$, we get a pole at $u=-\xi$; second, there is a contribution from the surface exciton state ϕ_n. The final result is

$$\mathcal{L}_{nn'}=\tfrac{1}{2}\sum_l a_l\{\delta_{|n-n'|,l}-\xi^{-1}\delta_{n+n',l+1}-(1-\xi^{-2})(-\xi)^{n+n'-l}h(l-n-n')\}+(\xi^{-2}-1)(\Lambda^2-\bar{\Lambda}^2)(-\xi)^{n+n'},\quad (58)$$

where

$$\bar{\Lambda}^2=\tfrac{1}{2}\sum_l a_l\{(1-\xi)^l+(-\xi)^{-l}\}.\quad (59)$$

It is seen that $\bar{\Lambda}^2$ is just the analytic extension of $\lambda^2(k)$ to $e^{ik}=-\xi$. If $\Lambda^2=\bar{\Lambda}^2$, the interaction has exactly the same form as before.

The case $\xi=0$ requires separate treatment, because here the poles at $u=0$ and $u=-\xi$ coincide, and also $\phi_n=\delta_{n1}$. We find in this case

$$\mathcal{L}_{nn'}=(\Lambda^2-\tfrac{1}{2}a_0)\delta_{n1}\delta_{n'1}+\tfrac{1}{2}\sum_l a_l\{\delta_{|n-n'|,l}-\delta_{n+n',l+2}\}.\quad (60)$$

The interactions given by (55), (58), and (60) seem perhaps somewhat arbitrary, but our treatment has shown that they are the only possible forms for the interaction if one wishes to describe the crystal by means of refractive indices. Thus, whenever one uses refractive indices one is actually assuming that the interaction is one of these forms, or that it can be sufficiently well approximated by one of them. In the next subsection, we consider the interaction of our crystal model with radiation.

C. Interaction with Radiation

As before, we are interested in the case of a previously unpolarized medium responding causally to an incident field. Accordingly, using retarded Green's functions, we express the desired solution of (34) as

$$P_n(t)=\sum_{n'=1}^{\infty}\int_{-\infty}^{\infty}dt'F_{nn'}^{\mathrm{ret}}(t-t')[-\epsilon^2\dot{A}(n',t')],\quad (61)$$

where

$$F_{nn'}^{\mathrm{ret}}(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega\int_0^\pi dk\frac{\psi_n(k)\psi_{n'}(k)e^{-i\omega t}}{\lambda^2(k)-(\omega+i\delta)^2}\quad (62)$$

for $|\xi|\geq 1$, or

$$F_{nn'}^{\mathrm{ret}}(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega\Big[\int_0^\pi\frac{dk\psi_n(k)\psi_{n'}(k)e^{-i\omega t}}{\lambda^2(k)-(\omega+i\delta)^2}+\frac{\phi_n\phi_{n'}e^{-i\omega t}}{\Lambda^2-(\omega+i\delta)^2}\Big],\quad (63)$$

for $|\xi|<1$. The appropriate solution of (35) is

$$A(z,t)=A_{in}(z,t)+\sum_{n'=1}^{\infty}\int_{-\infty}^{\infty}dt'G^{\mathrm{ret}}(z-n',t-t')\dot{P}_{n'}(t'),\quad (64)$$

where G^{ret} is given by (29). We are particularly interested in the response of the medium to an incident wave of a fixed frequency, and we want the solution to be expressible in terms of refractive indices. Accordingly, we seek a solution of (61) and (64) of the form

$$A_{in}(z,t)=Ae^{i\nu(z-t)};\quad A(n,t)=\sum_j A_je^{i\nu(\rho_jn-t)};\quad P_n(t)=\sum_j P_je^{i\nu(\rho_jn-t)}.\quad (65)$$

The sum over j of course corresponds to the different values ρ_j of the refractive index. Each ρ_j is understood to be a function of ν.

Substituting (29) and (65) into (64), and carrying out the integrations over t' and κ, we find

$$\sum_j A_je^{i\rho_j\nu n}=Ae^{i\nu n}+\tfrac{1}{2}\sum_j P_j\sum_{n'=1}^{\infty}e^{i\nu|n-n'|}e^{i\rho_j\nu n'}=\frac{i\sin\nu}{2}\sum_i\frac{P_je^{i\rho_j\nu n}}{\cos\rho_j\nu-\cos\nu}+\Big\{A-\tfrac{1}{2}\sum_i\frac{P_j}{1-e^{i(\rho_j-1)\nu}}\Big\}e^{i\nu n}.\quad (66)$$

Again we have used convergence factors where appropriate to evaluate the sum over n', and have done some rearranging. Equation (66) must be obeyed for all n, so we must equate coefficients of all different functions of n. Equating coefficients of $e^{i\rho_j\nu n}$, we find

$$P_j=(2/i\sin\nu)(\cos\rho_j\nu-\cos\nu)A_j. \tag{67}$$

Since the electric field $E=-\dot{A}$, we have $A_j=(i\nu)^{-1}E_j$, so (67) may also be written as

$$P_j=-(2/\nu\sin\nu)(\cos\rho_j\nu-\cos\nu)E_j.$$

As ν becomes small, this becomes

$$P_j=(\rho_j^2-1)E_j,$$

which is the usual continuum relation between polarization, refractive index, and field. There is no factor of 4π because of the Heaviside units.

Equating coefficients of $e^{i\nu n}$ in (66) we find

$$A=\tfrac{1}{2}\sum_j\frac{P_j}{1-e^{-i(\rho_j-1)\nu}} \tag{68}$$

$$=\frac{1}{i\sin\nu}\sum_j\frac{\cos\rho_j\nu-\cos\nu}{1-e^{-i(\rho_j-1)\nu}}A_j. \tag{69}$$

In the limit of small ν, Eq. (69) takes the form

$$A=\tfrac{1}{2}\sum_j(\rho_j+1)A_j. \tag{70}$$

To get the reflected wave, one inserts (65) into (64) for $z<1$, and also uses (67). The result is

$$A(z,t)=(A\,e^{i\nu z}+\mathrm{R}e^{-i\nu z})e^{-i\nu t},$$

with the reflected amplitude given by

$$R=\frac{1}{i\sin\nu}\sum_j\frac{\cos\rho_j\nu-\cos\nu}{e^{-i(\rho_j+1)\nu}-1}. \tag{71}$$

The limit of (71) for small ν is

$$R=-\tfrac{1}{2}\sum_j(\rho_j-1)A_j. \tag{72}$$

Equations (70) and (72) are easily seen to correspond to the usual ones of the continuity of E and H at the edge. In particular, the sum of (70) and (72) gives the continuity of E, while their differences gives that of H. We see, then, that all we get out of inserting (65) into (64) are the usual relations well known from classical optics, generalized to the discrete case. This is to be expected, since we have not yet made any use of the nature of the interactions between the oscillators. We now proceed to take this into account, considering first the case $|\xi|\geq 1$.

Inserting the solution (65) into (62), (61), we find

$$\sum_j P_je^{i\rho_j\nu n}$$

$$=i\nu\epsilon^2\sum_j A_j\int_0^\pi\frac{\psi_n(k)dk}{\lambda^2(k)-(\nu+i\delta)^2}\sum_{n'=1}^{\infty}\psi_{n'}(k)e^{i\rho_j\nu n'} \tag{73}$$

$$=\frac{2i\nu\epsilon^2}{\pi}\sum_j A_j$$

$$\times\int_0^\pi\frac{[\xi\sin kn+\sin k(n-1)]Q_j(k)dk}{[\lambda^2(k)-(\nu+i\delta)^2](\xi+e^{ik})(\xi+e^{-ik})}, \tag{74}$$

where

$$Q_j(k)=\sum_{n'=1}^{\infty}[\xi\sin kn'+\sin k(n'-1)]e^{i\rho_j\nu n'}$$

$$=\frac{(\xi+e^{i\rho_j\nu})}{2}\frac{\sin k}{(\cos\rho_j\nu-\cos k)}. \tag{75}$$

It is convenient to use the relation

$$\frac{\xi\sin kn+\sin k(n-1)}{(\xi+e^{-ik})(\xi+e^{ik})}=\frac{1}{2i}\left\{\frac{e^{ikn}}{\xi+e^{ik}}-\frac{e^{-ikn}}{\xi+e^{-ik}}\right\}. \tag{76}$$

We now insert (75) and (76) into (74), making the substitution $u=e^{ik}$ in the first term of (76) and $u=e^{-ik}$ in the second, and also use (67). After a little algebra, this gives

$$\sum_j P_je^{i\rho_jn}=\frac{\epsilon^2\nu\sin\nu}{8\pi i}\sum_j\frac{P_j(\xi+e^{i\rho_j\nu})}{\cos\rho_j\nu-\cos\nu}$$

$$\times\oint\frac{u^{n-1}(u-u^{-1})du}{[\cos\rho_j\nu-\frac{1}{2}(u+u^{-1})][\lambda^2(u)-(\nu+i\delta)^2](\xi+u)}, \tag{77}$$

where the contour is the unit circle, as usual. The pole at $u=\exp(i\rho_j\nu)$ in each term in the sum (77) leads to a term on the right-hand side proportional to $\exp(i\rho_j\nu n)$. Equating coefficients of $\exp(i\rho_j\nu n)$, we obtain

$$l=-\frac{\epsilon^2\nu\sin\nu}{2[\lambda^2(\rho_j\nu)-(\nu+i\delta)^2](\cos\rho_j\nu-\cos\nu)} \tag{78}$$

or

$$\cos\rho_j\nu-\cos\nu=-\frac{\epsilon^2\nu\sin\nu}{2[\lambda^2(\rho_j\nu)-(\nu+i\delta)^2]}. \tag{79}$$

In the limit of small ν, (79) becomes

$$\rho_j^2-1=\epsilon^2/(\lambda^2(\rho_j\nu)-\nu^2), \tag{80}$$

which is the usual relation. Equation (79) is, then, the equation that must be obeyed by each refractive index. Note that it is independent of ξ. Actually it just determines $\exp(i\rho_j\nu)$ or $\cos(\rho_j\nu)$, since λ^2 depends only on $\cos k$. The fact that u must lie within the unit circle requires that $\mathrm{Im}(\rho_j\nu)\geq 0$. [Actually, this has already been assumed, since it is needed for the convergence of the sum (75).] Referring to the solution (65), moreover, we see that nothing is changed by adding a multiple of 2π to $\rho_j\nu$. Hence, without loss of generality we can require that the real part of $\rho_j\nu$ lie between zero and 2π. There is now no more ambiguity and the values of the refractive indices are uniquely fixed for each frequency.

In order to treat the problem of boundary conditions, it is necessary to consider the other possible singularities within the unit circle of the integrand of (77). The possibilities are:

(a) Branch points. If λ^2 has a branch point within the unit circle, so will the integrand. In this case, however, one always has some choice in the way branch cuts are drawn, so that the behavior of λ^2 within the unit circle is determined not only by its value on the boundary but also by the choice of branch cuts. This would mean that the values of the refractive indices would be ambiguous, so we must not permit it to happen. Therefore, λ^2 must be chosen in such a way that there are no branch points.

(b) Essential singularities. If the sum (54) is infinite, λ^2 will have an essential singularity at the origin, leading to an infinite number of solutions to (79). We are interested, however, in a medium describable by a *finite* number of refractive indices, so we must not allow this. Accordingly, it is necessary to require the sum (54) to be finite; this in turn guarantees that there will be no essential singularities at other points, and no branch points. There are, therefore, only poles, and only a finite number of these.

The poles of the integrand in (77) other than the ones we have considered occur at the zeros of the denominator $\lambda(u)-(\nu+i\delta)^2$. They will all be simple poles except perhaps for certain values of ν. Each pole, moreover, leads to a relation between the coefficients P_j, i.e., to a boundary condition. For example, if there is a pole at $u=u_0$, its contribution to the right-hand side of (77) will be of the form

$$\sum_j P_j R_j u_0^n,$$

where R_j depends on the value of the residue. This sum must be equal to zero, since there is no term on the left-hand side proportional to u_0^n. We already have one boundary condition from (69). The total number is therefore one plus the number of zeros of the denominator $\lambda^2-(\nu+i\delta)^2$ within the unit circle. Now consider the function

$$S(u)=\tfrac{1}{2}(u+u^{-1})-\cos\nu+\frac{\epsilon^2\nu\sin\nu}{2[\lambda^2(u+u^{-1})-(\nu+i\delta)^2]}, \quad (81)$$

where we have explicitly expressed the fact that λ^2 depends only on $\cos k$, that is only on $u+u^{-1}$. According to (79), the number of zeros Z_S of S within the unit circle is equal to the number of refractive indices. The number of poles P_S is one plus the number of zeros of the denominator of the last term, i.e., equal to the number of boundary conditions. The excess of refractive indices over boundary conditions is then just Z_S-P_S. This should be zero in order for a unique solution to exist.

According to a well-known theorem, however,[9]

$$Z_S-P_S=\frac{1}{2\pi i}\oint\frac{S'(u)du}{S(u)}$$

or

$$Z_S-P_S=\frac{1}{2\pi}\int_0^{2\pi} d(\arg S(k)). \quad (82)$$

As k goes from zero to 2π, however, $\cos k$ goes from one to minus one and then back along the same path. S, which is a function only of $\cos k$, must also trace out some path as k goes from zero to π, and then return along the same path. Hence, the integral in (82) is clearly zero. The small imagninary term in the denominator in (81) guarantees that no singularities are encountered along the path. This proves that in this model one always automatically gets exactly the right number of boundary conditions.

It is also clear from (79) that one gets the same number of refractive indices for all frequencies, since one must solve an algebraic equation of the same order, and exactly half of the solutions are in the unit circle due to the symmetry between u and u^{-1}.

The treatment of the case $|\xi|<1$ is quite similar to the foregoing, so we shall just briefly state the result. If $\Lambda^2=\bar{\Lambda}^2$, the result is exactly the same as the above, since in this case the pole at $u=-\xi$ in (77) and the contribution of ϕ_n exactly cancel. If this is not the case, one must add to the solution (65) another term proportional to $(-\xi)^n$ for each frequency, and one more boundary condition is obtained, keeping the relation between number of boundary conditions and refractive indices.

V. DISCUSSION

The model which we have discussed is free of the difficulties mentioned in Sec. II, and is manifestly causal. We have seen both in Sec. III and in Sec. IV that the requirement of Hermiticity and positive definiteness of the operator $\mathfrak{L}$ puts restrictions on the allowable boundary conditions so as to prohibit unphysical effects such as reflection coefficients greater than unity.

The treatment of the negative value of β, which led to ambiguities in the usual theory, would be handled as follows in our model: First, we would have to use (33a) for the eigenvalue spectrum. Inserting (33a) into (79), we would get instead of (10) two values for $\cos(\rho\nu)$. As in the case of (10) with negative β, there would be a region in which these two values were complex conjugates of one another, but this would cause no difficulty:

[9] E. T. Whittaker and G. N. Watson, *A Course in Modern Analysis* (University Press, Cambridge, England, 1958), p. 119.

Corresponding to each value of $\cos(\rho\nu)$, there is always exactly one value of $(\rho\nu)$ with imaginary part positive and real part between zero and 2π. (Or between $-\pi$ and π, if one prefers.) Hence, procedure (b) for handling the problem of negative β (leading to two refractive indices) is automatically singled out, but there is no difficulty thereby introduced for positive β. The solution of this difficulty, therefore, depends on the correct inclusion of umklapp processes. It was to be expected that the inclusion of umklapp processes would be necessary in a general and self-consistent treatment since, as pointed out in Sec. I, one normally finds in the usual theory solutions which must be included to maintain causality but which do not satisfy the long-wavelength assumption.

In summary, our treatment justifies the usual procedure in cases where the latter leads to an unambiguous answer, and is able to avoid the difficulties that sometimes arise in the usual treatment. Our model is as general as is possible without destroying the validity of the refractive index concept; it is not, however, capable of discussing effects due to the distortion of exciton wave functions from sinusoidal form near the edge of the crystal. One such effect, apparently, is the reflection spike observed by Hopfield and Thomas.[4]

Note added in proof. Some recent papers in the Soviet literature dealing with the problem of acceptable boundary conditions are due to S. I. Pekar,[10] V. L. Ginzburg,[11] and V. I. Sugakov.[12]

[10] S. I. Pekar, Fiz. Tverd. Tela 4, 1301 (1962) [English transl.: Soviet Phys.—Solid State 4, 953 (1962)].
[11] V. L. Ginzburg, Zh. Eksperim. i Teor. Fiz. 34, 1593 (1958) [English transl.: Soviet Phys.—JETP 7, 1096 (1958)].
[12] V. I. Sugakov, Fiz. Tverd. Tela 5, 2207, 2683 (1963) [English transl.: Soviet Phys.—Solid State 5, 1959, 1607 (1963)].

PHYSICAL REVIEW VOLUME 138, NUMBER 1A 5 APRIL 1965

Critical Magnetic Field of a Thin Film of Superconductor*

G. RICKAYZEN†
Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York
(Received 18 August 1964; revised manuscript received 3 December 1964)

A theory of the critical field of a thin superconductive film is given based on the assumption that this is the field in which the normal state becomes unstable to the formation of Cooper pairs. Only the case where l, the mean free path for the electrons is much less than d, the thickness of the film, and where the coherence length is less than d^2/l is discussed. The resulting field for temperatures near T_c agrees with that predicted by the Ginzburg-Landau theory. At lower temperatures, the field is greater than that predicted by an earlier theory of Maki and appears to be in better agreement with the experimental results.

I. INTRODUCTION

THE theory of the critical magnetic field of a thin film of superconductive material has aroused considerable interest in recent years, especially as the experimenters are now reaching temperatures sufficiently low to discriminate between various theories. The first calculation based on the microscopic theory of superconductivity was given by Bardeen.[1] However, his result was later contradicted by Nambu and Tuan[2] and by Maki.[3] The calculations of the latter authors differed in the range of values of the parameters considered. In this paper we offer a calculation of the critical magnetic field for another range of values.

The problem contains four fundamental lengths, the thickness of the film d, the coherence length of pure bulk material $\xi_0(\sim\hbar v_F/kT_c)$, the penetration depth of the film λ, and the mean free path l. Nambu and Tuan and Maki considered the case of $d<\lambda$ and showed that the transition is of the second order. We shall assume that the transition is of the second order and thereby implicitly assume $d<\lambda$.

Nambu and Tuan also considered the case of a pure specimen ($l\gg d$) and specular reflection at the film surface. Maki considered the case more nearly realized in practice of $l\ll d$. He also assumed $(l\xi_0)^{1/2}\gg d$. We consider instead the case of

$$l\ll(l\xi_0)^{1/2}\ll d.$$

Our main interest is in the temperature dependence of the critical magnetic field which is rather different from Maki's.

Near T_c, the critical magnetic field has been calculated also from the Ginzburg-Landau theory.[4] However, it is not obvious that this theory is applicable to films of thickness much less than the coherence length. Gorkov's derivation of the theory, for example, assumes that the Green's function $G(\mathbf{r},\mathbf{r}',t)$ for the propagation of an electron in a magnetic field described by the vector

* Supported by the U. S. Office of Naval Research.
† On leave of absence from The University of Liverpool, Liverpool, England.
[1] J. Bardeen, Rev. Mod. Phys. 34, 667 (1962).
[2] Y. Nambu and S. F. Tuan, Phys. Rev. 133, A1 (1964).
[3] K. Maki, Progr. Theoret. Phys. (Kyoto) 29, 603 (1963).

[4] D. H. Douglass, Phys. Rev. Letters 6, 345 (1961).