PHYSICAL REVIEW B VOLUME 15, NUMBER 2 15 JANUARY 1977

Exactly soluble model for crystal with spatial dispersion

C. Alden Mead
Chemistry Department, University of Minnesota, Minneapolis, Minnesota 55455
(Received 15 March 1976; revised manuscript received 24 May 1976)

We consider a discrete, semi-infinite, one-dimensional crystal exhibiting "spatial dispersion," with the interaction between crystal sites falling off exponentially with distance; a model which has also been treated by Sipe and Van Kranendonk. The problem of the interaction of such a crystal with the electromagnetic field is exactly soluble. Results for physical properties are compared with those obtained using the "dielectric approximation," in which the polariton Green's function for full translational invariance is used to approximate the true one; and also with those obtained from the "near-neighbor approximation," in which the interaction is cut off after N nearest neighbors. In both cases, it is shown that the approximate results do not agree with the exact ones in all respects, even in the limit $N \to \infty$ for the near-neighbor approximation. Arguments are given to support the conclusion that these pessimistic results probably are general, and not merely artifacts of the particular model considered. The possibility of using this model, or generalizations of it, in practical calculations is briefly discussed.

I. INTRODUCTION AND SUMMARY

The phenomenon of "spatial dispersion," caused by nonlocal interactions between the electric polarization at different locations in a crystal, was first described theoretically by Pekar[1] in 1958. When these nonlocal interactions are taken into account, the electric susceptibility becomes formally a function of wave number as well as frequency, with the result that the wave number which actually propagates in the crystal becomes a many-valued function of frequency, even for isotropic media.

In the years following Pekar's discovery, a number of investigations of the phenomenon were carried out, of which the experimental and theoretical work of Hopfield and Thomas[2] perhaps deserves special mention. The work through about 1965 is well summarized in the book by Agranovich and Ginzburg.[3]

In the early work, the question of the so-called additional boundary conditions (ABC's) was a source of difficulty and ambiguity. Owing to the multiple valuedness of the wave number as function of frequency, monochromatic light incident on the surface of a spatially dispersive medium normally gives rise to several transmitted waves with different wave numbers, each with a transmission coefficient which must be calculated. The known boundary conditions of electromagnetic theory suffice only if there is only one transmitted wave, so additional conditions are needed in the spatially dispersive case. Without making somewhat arbitrary assumptions about the behavior of polarization waves near the surface, there seemed to be no unambiguous way of determining these; still it was felt that, in a complete theory, they should be consequences of the basic equations, just as the known conditions are consequences of the Maxwell equations. It was also pointed out[4] that the methods usually used could lead to conflict with the analyticity properties imposed by causality.

These questions were clarified somewhat in a 1965 paper by Deutsche and the present author.[5] We considered a one-dimensional model of a semi-infinite crystal in which oscillators were located at the points $x = 1, 2, 3, \ldots$. We also required that the "bare" polariton functions (polarization waves in the absence of coupling to the electromagnetic field) be sinusoidal in nature, i.e., that, for every κ in a Brillouin zone, there exist a solution of the uncoupled equations of the form

$$P_n(t) = [a(\kappa)\sin\kappa n + b(\kappa)\cos\kappa n]e^{-i\omega t}, \tag{1}$$

with real ω, where $P_n(t)$ is the polarization of the nth oscillator at time t. If the interaction between oscillators reaches beyond the nearest neighbors, the enforcement of (1) usually (but not always) requires a distortion of the interaction near the surface.[5] For this model, it was shown that one can set up the problem so as to automatically satisfy causality, and that one obtains just the right number of ABC's without additional assumptions.

However, it was subsequently pointed out by Mahan and Obermair[6] that the requirement (1) is unduly restrictive. Not only is it usually untrue, but, more to the point, it is not—as we had supposed—an essential approximation if the solution obtained is to be expressible in terms of a finite number of refractive indices. It could be replaced by a superposition of a finite number of sinusoidal waves without impairing this property. Mahan and Obermair introduced another model, also basically one dimensional, in which the interaction between

oscillators depends only on the distance between them (and hence is not distorted near the surface), and vanishes beyond the N nearest neighbors. We call this the near-neighbor approximation (NA). Mahan and Obermair showed how to solve this model, for arbitrary but finite N, without additional approximations. There are $N+1$ transmitted waves at each frequency, and one automatically finds the correct number of ABC's. A number of model calculations have been carried out by Philpott using the NA (restricted for simplicity to $N=0$, or 1), both for semi-infinite crystals,[7] and crystals of finite thickness.[8] In this work, it is also shown how to treat arbitrary angles of incidence while still basically using the one-dimensional model.

At about the same time, a completely different approach was developed independently by a number of authors. The method consists in calculating the frequency- and wave-number-dependent susceptibility for an infinite medium, Fourier transforming it to a space-time- or space-frequency-dependent susceptibility, and then assuming that this same susceptibility function determines the polarization also for the semi-infinite or finite medium. Mathematically (for a continuous medium, the only case for which this method has been used), this assumption takes the form

$$\vec{P}(\vec{r},t)=\int_{-\infty}^{t}dt'\int_{v}d^3r'\chi\left(\left|\vec{r}-\vec{r}'\right|,t-t'\right)\vec{E}(\vec{r}',t'), \tag{2}$$

where $\vec{P}(\vec{r},t)$ and $\vec{E}(\vec{r},t)$ are polarization and electric field, respectively, $\chi(|r|,t)$ is the susceptibility mentioned above, and the volume integration goes over that portion of space occupied by the medium. The susceptibility function is that calculated for the infinite medium, which has complete translational invariance, and therefore depends only on the absolute value of the distance. Special cases of (2) were first studied by Sein and Birman,[9-11] and the theory was later developed in greater generality by Agarwal, Pattanayak, and Wolf, first in a series of short, preliminary publications,[12-15] and later in a longer article, in which the theory was given definitive form and some model calculations were made.[16] Maradudin and Mills[17] developed independently an essentially equivalent theory. Further development and applications are to be found in the articles by Agarwal,[18-20] Agarwal, Pattanayak, and Wolf,[21] and Foley and Devaney.[22] These last authors have named (2) the dielectric approximation (DA). With the approximation (2), the problem can be solved without further assumptions, and the ABC's are automatically determined.[16] Attempts to test calculations using the DA experimentally have been inconclusive.[23]

The theories of Refs. 5, 6, and 16 possess one trait in common: all claim a greater degree of rigor and generality than they actually possess. The restrictive nature of Eq. (1) has already been discussed. The NA, however, is equally restrictive, depending as it does on the assumption that the interaction extends only to N nearest neighbors, and leading to the curious result that the predicted number of polariton modes at each frequency is a function of the cutoff. It is to be emphasized that the cutoff after N nearest neighbors is not merely a computationally convenient approximation in the approach of Mahan and Obermair, but is essential to their method of obtaining boundary conditions. As for the DA, it should be clear that the assumption (2) is just as arbitrary as the original boundary condition of Pekar,[1] as Eq. (1), or as the NA. The work so far published using the DA also suffers from the restriction to a continuous medium, although this restriction is not required by the DA and could be removed. This is essentially a long-wavelength assumption, and it has been shown[4,5] that this is never satisfied by all the transmitted waves over the frequency range of interest.

In the present paper, we consider again the one-dimensional, semi-infinite crystal, and assume that the interaction between different oscillators falls off exponentially with distance, and retains its form near the surface. This model has been treated by Sipe and Van Kranendonk,[24] who showed that it is exactly soluble without further assumptions, and also without much difficulty. There are two reasons why we feel that the discussion of such a model is worthwhile: first, the one-dimensional model does in fact describe the case of a semi-infinite slab of material with light incident normally, and it is known[6,25-28] that for such a slab the interactions do fall off approximately exponentially. Hence, it is at least possible that such a model may be useful in practical calculations. The second reason is that a soluble model is always useful as a "guinea pig" for testing various approximate theories.

In the present article, we confine ourselves to the "guinea pig" aspects of the soluble model, leaving the prospect of practical calculations to a later publication. In Sec. II, the model is set up and solved for refractive indices, reflection, and transmission coefficients for all frequencies. It turns out that there are *two* transmitted wave numbers for each frequency. The solution is given both for continuous and discrete cases. In Sec. III, the DA is considered, first from a general point of view and then by application to our model, comparing the DA results to the exact

ones. It is first shown quite generally, by putting the equations in Hamiltonian form, that the DA violates conservation of energy. A simple physical explanation of this is given. Then, the reflection and transmission coefficients calculated with the DA are compared with the exact ones, and shown not to be good approximations. In Sec. IV, we consider the NA, particularly for large N. It is shown that the NA for the reflection coefficient converges to the correct value for large N, but that there are other well-defined physical properties for which it converges to the wrong value, or does not converge at all. The reasons for this are discussed, and it is concluded that it is probably not an artifact of the particular model used, but is to be expected for any interaction that falls off no faster than exponentially. In Sec. V, there is some further discussion of the advantages and disadvantages of the NA and DA, and of the prospect of using this model, or generalizations of it, in practical calculations.

II. SOLUBLE MODEL

The model which we consider has already been treated by Sipe and Van Kranendonk.[24] Nevertheless, we present the solution in this section, partly to make the present article self-contained and partly to emphasize those aspects of it which are important for our purposes.

A. Formulation; bare polaritons

As in Ref. 5 and 6, we consider a one-dimensional problem, in which "oscillators" are arranged with uniform spacing along the x axis, and interact with each other and with a "field" which is defined everywhere on the x axis. We choose our unit of length to be the spacing between oscillators, so that they are located at the points $x=1,2,\ldots$; our unit of time is chosen so that the speed of light is unity, and a factor of 4π is absorbed into the definition of the polarization. If p_n is the polarization of the nth oscillator, located at $x=n$, then the polarization density $\mathcal{P}(x)$ is

$$\mathcal{P}(x)=\sum_{n=1}^{\infty} p_n\delta(x-n).$$

The "vector potential" $a(x,t)$ obeys the Maxwell equation

$$\ddot{a}(x,t)-a''(x,t)=\dot{\mathcal{P}}(x,t)=\sum_{n=1}^{\infty}\dot{p}_n(t)\delta(x-n). \tag{3}$$

The interaction between different oscillators is assumed in this model to fall off exponentially with distance, so the equation governing the oscillators is

$$\ddot{p}_n(t)=-\nu^2 p_n(t)+g\sum_{\substack{n'=1\\ \neq n}}^{\infty} p_{n'}(t)\,e^{-\gamma|n-n'|}-\epsilon^2\dot{a}(n,t). \tag{4}$$

Since the same equations would be obeyed by an array of parallel, infinite, planar slabs of oscillating material interacting with a field propagating in a direction perpendicular to them, this is also a model for a three-dimensional semi-infinite crystal interacting with light at normal incidence. The coupling constant g in Eq. (4) may be either positive or negative, while the other parameters ν^2, γ, and ϵ^2 are necessarily positive.

We seek solutions of (3) and (4) for fixed frequency ω:

$$a(x,t)=a(x)e^{i\omega t};$$

$$p_n(t)=p_n e^{-i\omega t}.$$

With these substitutions, the causal solution of (3) is[5]

$$a(x)=a_{in}(x)+\frac{1}{2}\sum_{n=1}^{\infty} p_n e^{i\omega|x-n|}, \tag{5}$$

while (4) becomes

$$(\nu^2-\omega^2)p_n=g\sum_{\substack{n'=1\\ \neq n}}^{\infty} p_{n'}e^{-\gamma|n-n'|}+i\omega\epsilon^2 a(n). \tag{6}$$

In this section, we concern ourselves with the "bare polariton" solutions, i.e., with solutions of (6) in the limit $\epsilon^2=0$. In particular, we seek solutions of the form (1), which we write

$$p_n=\alpha e^{i\kappa n}+\beta e^{-i\kappa n}. \tag{7}$$

When (7) is substituted into (6), with $\epsilon^2=0$, the sums are easily carried out. For example, the term proportional to α on the right-hand side is

$$g\alpha\sum_{\substack{n'=1\\ \neq n}}^{\infty} e^{i\kappa n'}e^{-\gamma|n-n'|}=g\alpha\left(e^{-\gamma n}\sum_{n'=1}^{n-1}e^{(i\kappa+\gamma)n'}+e^{\gamma n}\sum_{n'=n+1}^{\infty}e^{(i\kappa-\gamma)n'}\right)$$
$$=g\alpha\left[e^{-\gamma n}\left(\frac{e^{i\kappa+\gamma}-e^{(i\kappa+\gamma)n}}{1-e^{i\kappa+\gamma}}\right)+e^{\gamma n}\left(\frac{e^{-(i\kappa-\gamma)(n+1)}}{1-e^{i\kappa-\gamma}}\right)\right]=g\alpha\left(\frac{e^{-\gamma n}}{e^{-i\kappa-\gamma}-1}+e^{i\kappa n}\frac{\cos\kappa-e^{-\gamma}}{\cosh\gamma-\cos\kappa}\right).$$

Summing the β contribution on the same way, we find

$$(\nu^2-\omega^2)(\alpha e^{i\kappa n}+\beta e^{-i\kappa n})=ge^{-\gamma n}\left(\frac{\alpha}{e^{-i\kappa-\gamma}-1}+\frac{\beta}{e^{i\kappa-\gamma}-1}\right)+g(\alpha e^{i\kappa n}+\beta e^{-i\kappa n})\left(\frac{\cos\kappa-e^{-\gamma}}{\cosh\gamma-\cos\kappa}\right). \tag{8}$$

Equation (8) is satisfied for all n if and only if

$$\alpha/(e^{-i\kappa-\gamma}-1)+\beta/(e^{i\kappa-\gamma}-1)=0 \tag{9}$$

and

$$\nu^2-\omega^2=(\cos\kappa-e^{-\gamma})/(\cosh\gamma-\cos\kappa). \tag{10}$$

Equation (10) determines ω as a function of κ, while (9) is a boundary condition. We see, therefore, that there is a solution of the form (1) or (7) for each κ, which means that the formalism of Ref. 5 can be directly applied to this model. This shows, incidentally, that, Ref. 6 to the contrary notwithstanding, the validity of Eq. (1) is by no means restricted to nearest-neighbor interactions. It can be made valid for arbitrary interactions by appropriate distortion of the interaction near the surface, and it is, as we have just seen, also satisfied by the exponential interaction. It is true that it is not valid in general, and that it is compatible with the NA only for $N=0$ or 1. The NA, however, does not exhaust all possibilities.

If desired, one can now use the methods of Ref. 5 directly to solve Eqs. (5) and (6). It is somewhat simpler, however, and more convenient for comparison with other theories, to solve it directly. This we do in Sec. II B.

B. Solution

We seek a solution of (5) and (6) with

$$a_{in}(x)=e^{i\omega x},\quad a(n)=\sum_j a_j e^{i\kappa_j n},\qquad p_n=\sum_j p_j e^{i\kappa_j n}, \tag{11}$$

where the sums go over an unspecified (for the moment) number of transmitted modes. When the solution has been found, one can calculate $a(x)$ for negative and noninteger x by direct use of (5). It is a simple matter to insert (11) into (5) and (6) to arrive at equations for the κ_j, a_j, and p_j. The sums are carried out just as before, except that the sum on the right-hand side of (5) does not omit the term $n'=n$. Since the sums go to infinity in the positive, though not in the negative, direction, it is assumed that $|e^{i\kappa_j}|\leq 1$. Substitution of (11) into (5) gives the result

$$\sum_j a_j e^{i\kappa_j n}=e^{i\omega n}+\tfrac{1}{2}e^{i\omega n}\sum_j \frac{p_j}{e^{-i(\kappa_j-\omega)}-1}+\sum_j p_j e^{i\kappa_j n}\frac{i}{2}\frac{\sin\omega}{\cos\kappa_j-\cos\omega}. \tag{12}$$

Equation (12) can be satisfied for all n only if

$$\sum_j \frac{p_j}{e^{-i(\kappa_j-\omega)}-1}=-2 \tag{13}$$

and

$$a_j=\frac{i}{2}\frac{\sin\omega}{\cos\kappa_j-\cos\omega}p_j \tag{14}$$

Insertion of (11) and (14) into (6) gives

$$(\nu^2-\omega^2)\sum_j p_j e^{i\kappa_j n}=ge^{-\gamma n}\sum_j \frac{p_j}{e^{-i\kappa_j-\gamma}-1}+g\sum_j p_j e^{i\kappa_j n}\left(\frac{\cos\kappa_j-e^{-\gamma}}{\cosh\gamma-\cos\kappa_j}\right)-\epsilon^2\sum_j p_j e^{i\kappa_j n}\left(\frac{\omega\sin\omega}{2(\cos\kappa_j-\cos\omega)}\right), \tag{15}$$

which is satisfied for all n only if

$$\sum_j \frac{p_j}{e^{i\kappa_j-\gamma}-1}=0, \tag{16}$$

and

$$\nu^2-\omega^2=g\frac{\cos\kappa_j-e^{-\gamma}}{\cosh\gamma-\cos\kappa_j}+\frac{\epsilon^2\omega\sin\omega}{2(\cos\omega-\cos\kappa_j)}. \tag{17}$$

Equation (17) determines the κ_j as functions of ω, while (13) and (16) are essentially boundary conditions. For each value of ω, Eq. (17) has two solutions for $\cos\kappa_j$, and for each of these we may choose $e^{i\kappa_j}$ to be less than or equal to 1 in absolute value. This will determine $e^{i\kappa_j}$ completely unless $\cos\kappa_j$ is real and lies between ± 1. In this case, it can either be determined by adding a small damping term to (4) or (6) and choosing the solution which then becomes less than 1 in absolute value,[6] or by the methods of Ref. 5. Thus, there are always two transmitted waves, the sums over j in (13) and (16) involve only two terms, and these two equations are therefore sufficient to determine the two transmission coefficients p_1 and p_2.

For $x<0$, Eqs. (5) and (11) give the result

$$a(x)=e^{i\omega x}+Re^{-i\omega x}, \tag{18}$$

with the reflection coefficient R given by

$$R=\frac{1}{2}\sum_j \frac{p_j}{e^{-i(\omega+\kappa_j)}-1}. \tag{19}$$

Equations (13), (16), and (19) can be solved easily to give

$$R = -e^{2i\omega}\frac{(e^{\gamma}-e^{i\omega})(e^{-i\kappa_1}-e^{-i\omega})(e^{-i\kappa_2}-e^{-i\omega})}{(e^{\gamma}-e^{-i\omega})(e^{-i\kappa_1}-e^{i\omega})(e^{-i\kappa_2}-e^{i\omega})}. \tag{20}$$

We see, therefore, that a complete solution for this model is obtained by first solving the quadratic equation (17) for the transmitted wave numbers, then (13) and (16) simultaneously for the transmission coefficients. The electric field within the medium is found with the aid of (14), and the reflection coefficient is given by (20). We note in passing that Eq. (20) for the reflection coefficient satisfies the requirements of causality.[4,5] Causality requires that R be an analytic function of ω in the upper half-plane, in other words, where $e^{i\omega}$ lies inside the unit circle. One verifies immediately that this is satisfied by (20). The apparent pole at $\omega = i\gamma$ is cancelled because, as $\omega \to i\gamma$, one of the solutions for $\cos\kappa$ approaches $\cosh\gamma$ (cf. Figs. 1, 2).

C. Continuum limit

For purposes of comparison with DA calculations, it is desirable to consider the continuum limit of our model, even though this is neither realistic[4,5] nor much simpler. Accordingly, we replace our discrete crystal with a continuous one occupying the region $0<x<\infty$. Both field and polarization are now continuous functions of x. In an obvious notation, Eqs. (5) and (6) are replaced by

$$a(x) = a_{in}(x) + \frac{1}{2}\int_0^{\infty} p(x')\,e^{i\omega|x-x'|}\,dx'; \tag{21}$$

$$(\nu^2-\omega^2)p(x) = g\int_0^{\infty} p(x')e^{-\gamma|x-x'|}\,dx' + i\omega\epsilon^2 a(x). \tag{22}$$

We seek a solution with $a_{in} = e^{i\omega x}$, and with, for $x>0$,

$$a(x) = \sum_j a_j e^{i\kappa_j x}, \qquad p(x) = \sum_j p_j e^{i\kappa_j x}. \tag{23}$$

The analogs of (13), (14), (16), (17), (19) are, respectively,

$$\sum_j \frac{p_j}{i(\kappa_j-\omega)} = 2, \tag{24}$$

$$a_j = [\omega/i(\kappa_j^2-\omega^2)]p_j, \tag{25}$$

$$\sum_j \frac{p_j}{i\kappa_j+\gamma} = 0, \tag{26}$$

$$\nu^2-\omega^2 = \frac{2g\gamma}{\gamma^2+\kappa^2} + \frac{\epsilon^2\omega^2}{\kappa^2-\omega^2}, \tag{27}$$

$$R = -\frac{1}{2}\sum_j \frac{p_j}{i(\kappa_j+\omega)}. \tag{28}$$

Equation (27), like (17), has two solutions. Solving (24), (26), (28) for R, one finds

$$R = \frac{(i\omega-\gamma)(\kappa_1-\omega)(\kappa_2-\omega)}{(i\omega+\gamma)(\kappa_1+\omega)(\kappa_2+\omega)}. \tag{29}$$

III. DIELECTRIC APPROXIMATION

A. General remarks; Hamiltonian formulation

In this section, following the procedure of Ref. 9–22, we restrict ourselves to the continuum limit. In this subsection, however, we do not require that the interaction between polarization at different points fall off exponentially, but allow it to be an arbitrary function of distance. The exact field equations are taken to be

$$\ddot{a}(x,t) - a''(x,t) = \dot{p}(x,t) \tag{30}$$

and

$$\ddot{p}(x,t) + \nu^2 p(x,t) = \int_0^{\infty} u(|x-x'|)\,p(x',t)\,dx' - \epsilon^2\dot{a}(x,t), \tag{31}$$

in which it is understood that the polarization field $p(x)$ is only defined for $x>0$.

The DA consists in replacing the true polariton Green's function by that calculated for the infinite crystal, i.e., with complete translational invariance; in other words, the interaction of the polarization field with itself (though not with the electromagnetic field) is assumed to have translational invariance, i.e., the integral on the right-hand side of (31) is formally extended to $-\infty$. It is obvious without a detailed calculation that this approach leads to nonconservation of energy, at least for bare polaritons; for, if such a polariton propagates toward the surface from within the medium, it experiences no change whatever in the interactions influencing it as it nears and reaches the surface. Accordingly, it will simply "pass through" the surface as though it were not there, and formally "propagate" as a fictitious polarization outside the surface, leaving nothing inside.

To see that the same thing happens in the full theory, it is useful to cast the theory into Lagrangian-Hamiltonian form. An appropriate Lagrangian for Eqs. (30) and (31) is

$$L=\frac{1}{2}\int_{-\infty}^{\infty}[\dot{a}^2(x)-(a')^2(x)]\,dx + \frac{1}{2\epsilon^2}\int_0^{\infty}[\dot{p}^2(x)-\nu^2p^2(x)]\,dx + \frac{1}{2\epsilon^2}\int_0^{\infty}\int_0^{\infty}u(|x-x'|)p(x)p(x')\,dx\,dx' + \int_0^{\infty}a(x)\dot{p}(x)\,dx,$$

with corresponding Hamiltonian

$$H=\frac{1}{2}\int_{-\infty}^{\infty}[\dot{a}^2(x)+(a')^2(x)]\,dx+\frac{1}{2\epsilon^2}\int_0^{\infty}[\dot{p}^2(x)+\nu^2p^2(x)] - \frac{1}{2\epsilon^2}\int_0^{\infty}\int_0^{\infty}u(|x-x'|)p(x)p(x')\,dx\,dx'. \quad (32)$$

In DA, one still has (30), but instead of (31), p is given by

$$p(x,t)=-\epsilon^2\int_{-\infty}^{\infty}dt'\int_0^{\infty}dx'\,G(x-x',t-t')\dot{a}(x',t')$$

or

$$p(x,t)=-\epsilon^2\int_{-\infty}^{\infty}dt'\int_{-\infty}^{\infty}dx' \times G(x-x',t-t')\dot{a}(x',t')h(x'), \quad (33)$$

where $h(x)$ is the step function, equal to zero when its argument is negative, and equal to unity when it is positive; and $G(x-x',t-t')$ is the Green's function for the crystal infinite in both directions, defined by

$$\frac{\partial^2}{\partial t^2}+\nu^2\ G(x-x',t-t') - \int_{-\infty}^{\infty}u(|x-x''|)\ G(x''-x',t-t')\,dx'' = \delta(x-x')\delta(t-t'). \quad (34)$$

The right-hand side of (33), of course, is equally well defined if x is negative, and we can thus use (33) to formally define a function $p(x,t)$ which is then well defined for all x. This formal $p(x,t)$, of course, is fictitious for $x<0$, and is equal to the actual polarization only for $x>0$. With this formally redefined $p(x,t)$, and with the aid of (33) and (34), we see that Eqs. (30) and (31) are to be replaced in DA by

$$\ddot{a}(x,t)-a''(x,t)=\dot{p}(x,t)h(x), \quad (30')$$

$$\ddot{p}(x,t)+\nu^2p(x,t)=\int_{-\infty}^{\infty}u(|x-x'|)\,p(x',t)-\epsilon^2\dot{a}(x,t)h(x). \quad (31')$$

Both fields are formally defined for the whole range $-\infty<x<\infty$, but $p(x)$ for negative x is fictitious, being simply a mathematical device to reproduce the DA. A Lagrangian for Eqs. (30′) and (31′) is

$$L_{\mathrm{DA}}=\frac{1}{2}\int_{-\infty}^{\infty}[\dot{a}^2(x)-a')^2(x)]\,dx + \frac{1}{2\epsilon^2}\int_{-\infty}^{\infty}[\dot{p}^2(x)-\nu^2p^2(x)]\,dx + \frac{1}{2\epsilon^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}u(|x-x'|)p(x)p(x')\,dx\,dx' + \int_0^{\infty}a(x)\dot{p}(x)\,dx,$$

with Hamiltonian

$$H_{\mathrm{DA}}=\frac{1}{2}\int_{-\infty}^{\infty}[\dot{a}^2(x)+(a')^2(x)]\,dx + \frac{1}{2\epsilon^2}\int_{-\infty}^{\infty}[\dot{p}^2(x)+\nu^2p^2(x)]\,dx - \frac{1}{2\epsilon^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}u(|x-x'|)p(x)p(x')\,dx\,dx'. \quad (35)$$

In the exact theory, the Hamiltonian (32) represents a conserved energy. In the DA, the conserved energy is given by (35), and includes a contribution from the fictitious polarization in the region of negative x. It is also evident from Eqs. (30′) and (31′) that this fictitious polarization is in general neither zero nor particularly small compared with the "real" polarization for $x>0$. The DA, therefore, always entails a loss of energy to this fictitious field, a quite unphysical effect. In the author's view, this should be sufficient to cast grave doubt on the usefulness of this approximation. It is conceivable, however, that the effect of this might be negligible in special cases.

In Sec. III B, we return to our exponential model in the continuum limit, calculate reflection and transmission coefficients using the DA, and compare these with the exact results of Sec. II C.

B. Comparison with exact solution

We now return to our exponential model in the continuum limit, and consider the solution when the DA is made. We thus seek solutions of (30′) and (31′) with fixed frequency. The resulting Eq. (21) is unchanged, while (22) is replaced by

$$(\nu^2-\omega^2)p(x)=g\int_{-\infty}^{\infty}e^{-\gamma|x-x'|}p(x')\,dx' + i\omega\epsilon^2a(x)h(x). \quad (36)$$

We again seek a solution of the form (23) for x

>0, and again obtain (24), (25), and (28) when this is inserted into (21). This time, however, there is also a polarization, albeit fictitious, for $x<0$. In this region, we assume a polarization of the form

$$p(x)=\sum_{\alpha} p_{\alpha} e^{-iq_{\alpha}x}. \tag{37}$$

Equations (23), (36), and (37) combine to give

$$(\nu^2-\omega^2)\sum_j p_j e^{i\kappa_j x}=ge^{-\gamma x}\left(-\sum_j \frac{p_j}{\gamma+i\kappa_j}+\sum_{\alpha}\frac{p_{\alpha}}{\gamma-iq\alpha}\right) + 2\gamma g\sum_j p_j \frac{e^{i\kappa_j x}}{\gamma^2+\kappa_j^2} + \sum_j p_j e^{i\kappa_j x}\frac{\epsilon^2\omega^2}{\kappa_j^2-\omega^2} \tag{38}$$

for $x>0$, and

$$(\nu^2-\omega^2)\sum_{\alpha} p_{\alpha} e^{-iq_{\alpha}x}=ge^{\gamma x}\left(\sum_j \frac{p_j}{\gamma-i\kappa_j}-\sum_{\alpha}\frac{q_{\alpha}}{\gamma+iq_{\alpha}}\right) + 2\gamma g\sum_{\alpha} p_{\alpha}\frac{e^{-iq_{\alpha}x}}{\gamma^2+q_{\alpha}^2} \tag{39}$$

for $x<0$. Equations (38) and (39) require that (27) be satisfied, and also

$$\sum_j \frac{p_j}{\gamma+i\kappa_j}-\sum_{\alpha}\frac{p_{\alpha}}{\gamma-iq_{\alpha}}=0, \tag{40}$$

$$\sum_j \frac{p_j}{\gamma-i\kappa_j}-\sum_{\alpha}\frac{p_{\alpha}}{\gamma+iq_{\alpha}}=0, \tag{41}$$

$$\nu^2-\omega^2=2\gamma g/(\gamma^2+q_{\alpha}^2). \tag{42}$$

Equation (42) has only one solution, so there is only one wave number propagating in the fictitious region, which we name simply q, its coefficient p_f (f for fictitious). One again obtains (28) for the reflection coefficient. Solving (24), (40), (41), and (28), we obtain for R

$$R=\frac{[(\kappa_1+\kappa_2)(\gamma^2-q\omega)+(\gamma^2-\kappa_1\kappa_2)(q+\omega)](\kappa_1-\omega)(\kappa_2-\omega)}{[(\kappa_1+\kappa_2)(\gamma^2+q\omega)+(\gamma^2-\kappa_1\kappa_2)(q-\omega)](\kappa_1+\omega)(\kappa_2+\omega)}. \tag{43}$$

Equation (43) bears little resemblance to the exact result (29), and in general will lead to quite different results, especially, as may easily be verified, in the case of greatest interest: κ_1, κ_2, q all real and of the same order of magnitude. This is because the contribution of the fictitious polarization is generally of the same order of magnitude as the real polarization in the medium. This last may be seen by evaluation p_f/p_1. The result is

$$p_f/p_1=(\kappa_2-\kappa_1)(\gamma^2+q^2)/(\kappa_2+q)(\gamma^2+\kappa_1^2),$$

with p_f/p_2 obtainable, of course, just by interchanging 1 and 2. It is easily seen that p_f/p_1 and p_f/p_2 are never both small over a wide frequency range. For example, for frequencies well off resonance, one has $\kappa_1 \sim q \sim i\gamma$, $\kappa_2 \sim \omega$, and $|p_f/p_1| \sim 1$. The same is true near resonance, when κ_1, κ_2, q are all real and comparable in magnitude. The fictitious polarization is therefore not negligible compared with the real contributions, and its effect on calculated reflection coefficients will normally be at least as great as that of spatial dispersion itself. We conclude that the DA, in addition to violating energy conservation, is in fact not a good approximation, at least for the model considered here.

IV. NEAR-NEIGHBOR APPROXIMATION

A. Solution with NA

The prescription of Mahan and Obermair[6] is first to solve for the κ_j for a medium extending from plus to minus infinity, then to impose boundary conditions in such a way that there is no fictitious polarization outside the medium. Applying this to our model, we replace Eqs. (5) and (6) by

$$a(x)=\frac{1}{2}\sum_{n=-\infty}^{\infty} p_n e^{i\omega|x-n|}, \tag{44}$$

and

$$(\nu^2-\omega^2)p_n=g\sum_{\substack{n'=n-N\\ \neq n}}^{n+N} p_{n'}e^{-\gamma|n-n'|}+i\omega\epsilon^2 a(n). \tag{45}$$

Inserting

$$a(n)=ae^{i\kappa n},\quad p_n=pe^{i\kappa n}, \tag{46}$$

one finds

$$\nu^2-\omega^2=g\left(\frac{\cos\kappa-e^{-\gamma}+[e^{-\gamma(N+1)}\cos\kappa N-e^{-\gamma N}\cos\kappa(N+1)]}{\cosh\gamma-\cos\kappa}\right)+\frac{\epsilon^2\omega\sin\omega}{2(\cos\omega-\cos\kappa)}. \tag{47}$$

Equation (47) determines the allowed values of $\cos\kappa$ in NA. It differs from the exact result (17) by the expression in square brackets on the right-hand side. For each value of ω, it has $(N+1)$ roots. For large N, the term in square brackets oscillates rapidly for real κ ($\cos\kappa$ real and between -1 and 1), and rapidly becomes large for $|\cos\kappa| > \cosh\gamma$. There is no singularity at $\cos\kappa = \cosh\gamma$, since the term proportional to g in (47) is, despite its appearance, really just a polynomial of Nth order in $\cos\kappa$. For sufficiently large N, it is easily verified that the quantity in curly brackets is very nearly equal to N for $\cos\kappa = \cosh\gamma$.

In Fig. 1 (for positive g), and Fig. 2 (for negative g), the behavior of the right-hand side of (47) for large N is compared qualitatively with that of the exact expression (17). Real roots occur where the curves cross the horizontal line $(\nu^2 - \omega^2)$. If the maximum of (17) in Fig. 2 occurs for $\cos\kappa < 1$, as shown in the figure, there will be a narrow frequency range about the tangent at this maximum for which the oscillations in (47) cause the number of roots in this vicinity to vary erratically with N. The width of this frequency range decreases with N, but in this vicinity the NA certainly does not converge *uniformly* to the right result. For positive g (Fig. 1), one sees from the graphs that there are always two real roots for the exact result (17). For the NA, there are two real roots for odd N, while for even N there are either three or one, depending on whether $\nu^2 - \omega^2$ lies above or below the minimum of the dotted line in Fig. 1. For $\nu^2 - \omega^2 > -g$, the two real roots (or two of the three, as the case may be) approach for large N the two exact roots, while for $\nu^2 - \omega^2 < -g$, this is true only for one of the real roots. In general, one

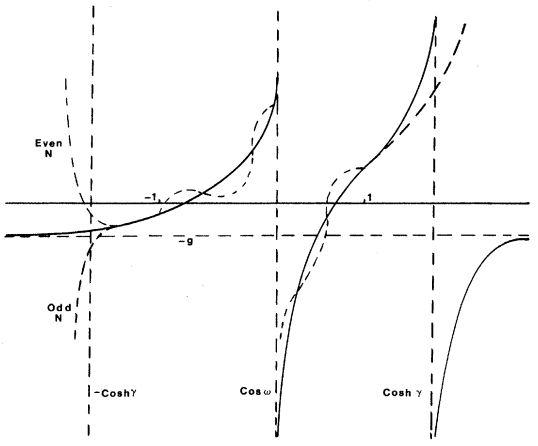

FIG. 1. Qualitative plot of the right-hand sides of Eqs. (17) (solid curve) and (47) with N large (dashed curve) as functions of $\cos\theta$ for $g > 0$.

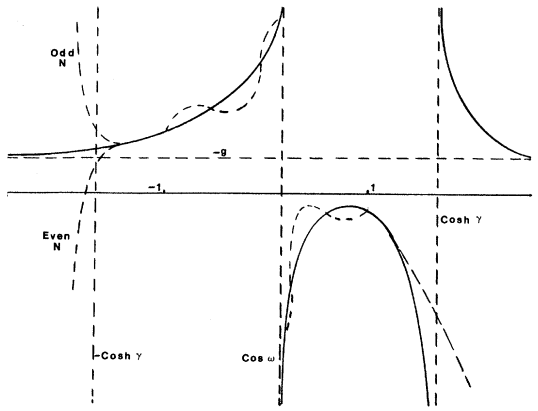

FIG. 2. Qualitative plot of the right-hand sides of Eqs. (17) (solid curve) and (47) with N large (dashed curve) as functions of $\cos\theta$ for $g > 0$.

sees from the figures that the NA for large N sometimes furnishes a good approximation for both of the actual roots, sometimes for only one of them, and sometimes (for negative g) for neither one.

It appears, therefore, that one is justified at this point in harboring doubts about the usefulness of the NA. The great majority of the roots that it leads to are spurious in the sense that they seem to have nothing to do with roots of the exact equation. In some cases, moreover, the $N+1$ roots include a good approximation for only one (not both) of the exact roots, or for neither one. However, a little reflection shows that one should not jump to this conclusion too hastily. One is actually interested in physical quantities such as the reflection coefficient, rather than the refractive indices themselves, and it is still possible that the NA may converge to the correct answer for these. In particular, we will concentrate on a generalization of the reflection coefficient, to be described in the next Sec. IV B.

B. Generalization of reflection coefficient

We make the definitions

$$q_n = -\tfrac{1}{2} e^{-i\omega} p_n, \tag{48}$$

$$\mathcal{R}(s) = s^{-1} \sum_{n=1}^{\infty} q_n s^n, \tag{49}$$

where s is a complex number. If the solution has the form (11), then

$$\mathcal{R}(s) = \sum_j \frac{q_j}{x_j - s}, \tag{50}$$

where

$$q_j = -\tfrac{1}{2} e^{-i\omega} p_j, \tag{51}$$

$$x_j = e^{-i\kappa_j}. \tag{52}$$

If the allowed wave numbers are ordered in such a way that

$$|x_1| \le |x_2| \le |x_3| \le \cdots$$

then (49) converges for all s with $|s| < |x_1|$. If desired, the function $\mathfrak{R}(s)$ may be defined elsewhere by analytic continuation. If, for example, the solution has the form (11), Eq. (50) provides an analytic continuation for other values of s. We will adopt the convention that $\mathfrak{R}(s)$ is defined only by the series (49), and is undefined where that series fails to converge. We define $\tilde{\mathfrak{R}}(s)$ as the analytic continuation of $\mathfrak{R}(s)$. It is clear that $\mathfrak{R}(s)$, where it is defined at all, is a physical observable; it is perhaps debatable whether $\tilde{\mathfrak{R}}(s)$ is observable where $\mathfrak{R}(s)$ is undefined. For purposes of theoretical calculations, it is more convenient to deal with $\tilde{\mathfrak{R}}(s)$, and then to define $\mathfrak{R}(s)$ as being equal to $\tilde{\mathfrak{R}}(s)$ within the circle about the origin passing through the nearest singularity, and undefined elsewhere. Comparison of (49), and (50) with (5) and (19) shows that the reflection coefficient R is given by

$$R = -e^{2i\omega}\mathfrak{R}(e^{i\omega}).$$

Under certain conditions, one can define an observable related to $\mathfrak{R}(s)$ in regions where $\mathfrak{R}(s)$ is not defined by making subtractions. Thus, if (11) holds and $|x_1| < |x_2|$ (not equal), then one can define an asymptotic form of q_n as follows:

$$q_n^{(a)} = \lim_{\nu\to\infty} q_\nu \left(\frac{q_{\nu+1}}{q_\nu}\right)^{n-\nu} = q_1 e^{i\kappa_1 n}. \tag{53}$$

This can then be subtracted off from the observed q_n, and the result used to define a "subtracted $\mathfrak{R}(s)$":

$$q_n^{(1)} = q_n - q_n^{(a)}, \tag{54}$$

$$\mathfrak{R}^{(1)}(s) = s^{-1}\sum_{n=1}^{\infty} q_n^{(1)} s^n = \sum_{j\ge 2} \frac{q_j}{x_j - s}. \tag{55}$$

$\mathfrak{R}^{(1)}(s)$ is defined for $|s| < |x_2|$. Its analytic continuation $\tilde{\mathfrak{R}}^{(1)}(s)$ has one fewer pole than $\tilde{\mathfrak{R}}(s)$. Similarly, if also $|x_2| < |x_3|$, one can make a second subtraction to obtain a function which is analytic for $|s| < |x_3|$:

$$\mathfrak{R}^{(2)}(s) = \sum_{j\ge 3} \frac{q_j}{x_j - s}. \tag{56}$$

For our exact solution one finds, using Eqs. (13), (16), (48), (50), (55), and (56),

$$\tilde{\mathfrak{R}}(s) = \frac{(x_1 - Y)(x_2 - Y)(Z - s)}{(x_1 - s)(x_2 - s)(Z - Y)}, \tag{57}$$

$$\tilde{\mathfrak{R}}^{(1)}(s) = -\frac{(x_1 - Y)(x_2 - Y)(x_2 - Z)}{(x_1 - x_2)(x_2 - s)(Z - Y)}, \tag{58}$$

$$\tilde{\mathfrak{R}}^{(2)}(s) = 0, \tag{59}$$

where $Y = e^{-i\omega}, Z = e^{\gamma}$. We see that $\mathfrak{R}(s)$ is defined for $|s| < |x_1|$, $\mathfrak{R}^{(1)}$ is defined for $|s| < |x_2|$, and $\mathfrak{R}^{(2)}(s)$ is defined everywhere.

For the NA, the methods of Mahan and Obermair[6] give the results

$$\tilde{\mathfrak{R}}_N(s) = \prod_{j=1}^{N+1} \frac{(x_j - Y)}{(x_j - s)}, \tag{60}$$

$$\tilde{\mathfrak{R}}_N^{(1)}(s) = \prod_{j=1}^{N+1} \frac{(x_j - Y)}{(x_j - s)} - \frac{(x_1 - Y)}{(x_1 - s)} \prod_{j=2}^{N+1} \frac{(x_j - Y)}{(x_j - x_1)}, \tag{61}$$

$$\begin{aligned}\tilde{\mathfrak{R}}_N^{(2)}(s) = {} & \prod_{j=1}^{N+1} \frac{(x_j - Y)}{(x_j - s)} \\ & - \frac{(x_1 - Y)(x_2 - Y)}{(x_1 - s)(x_2 - x_1)} \prod_{j=3}^{N+1} \frac{(x_j - Y)}{(x_j - x_1)} \\ & - \frac{(x_1 - Y)(x_2 - Y)}{(x_2 - s)(x_1 - x_2)} \prod_{j=3}^{N+1} \frac{(x_j - Y)}{(x_j - x_2)}.\end{aligned} \tag{62}$$

Derivations of Eqs. (60)–(62) are given in the Appendix.

Comparison of (57) with (60) shows that the NA, in cases where it approximates one or more of the exact roots, gives correctly at least some of the factors contained in $\tilde{\mathfrak{R}}(s)$. It remains to be determined whether the product of the other factors in (60) is equal to the remaining factor(s) in (57). The answer to this question requires an analysis of the other roots of (47), a matter which we take up in Sec. III C.

C. Analysis of roots

We first make a few general remarks about the meaning of the NA for our model. As far as its effect on the calculated transmitted wave numbers is concerned, the NA consists in replacing the term proportional to g on the right-hand side of (17) with the corresponding term in (47), which is just the power series expansion of the correct expression in $e^{i\kappa}$ and $e^{-i\kappa}$, taken through the Nth power. The derivation of (17) is valid for all values of $\cos\kappa$, since it involves an infinite series only in positive powers of $e^{i\kappa}$, and this can be chosen ≤ 1 in absolute value for all values of $\cos\kappa$. The power series, however, does not represent the function at all points, but converges to the function only within a circle of convergence whose radius is the distance from the origin to the nearest singularity, in our case $\cos\kappa = \cosh\gamma, e^{-i\kappa} = e^{\gamma}$. There are always $N+1$ roots of (47), of which at most two are approximations to roots of the exact expression. We call these "physical roots," the

other roots will be called "unphysical roots." For large N, we do not expect to find the unphysical roots in the region well within the circle of convergence, since in this region the power series represents the function well, and we would expect it to have at least approximately the same roots. In the region well outside the circle of convergence, the power series diverges, and the Nth order polynomial will be very large in absolute value for large N, larger in particular than the other terms in Eq. (47). Therefore, we do not expect to find roots here either. The remaining possibility is that they lie near the circle of convergence, so this is where we shall seek them. It should be noted that the tendency for the unphysical roots to converge to the circle of convergence will not be uniform in ω. A glance at Fig. 1, or at Eq. (47), reveals that, for fixed N, there are always values of $\nu^2-\omega^2$ for which there are roots well outside the circle. For fixed ω, however, one can always find N large enough that the extra roots will lie near the circle of convergence.

If (47) is multiplied through by $(\cosh\gamma-\cos\kappa)$, the resulting equation will have the same roots, plus a spurious one at $\cos\kappa=\cosh\gamma$. Doing this, and rearranging somewhat, we obtain

$$G(\cos\kappa)\equiv g[e^{-\gamma(N+1)}\cos N\kappa - e^{-\gamma N}\cos(N+1)\kappa] = F(\cos\kappa), \tag{63}$$

where

$$F(\cos\kappa)=(\cosh\gamma-\cos\kappa)\times\left(\nu^2-\omega^2-\frac{\epsilon^2\omega\sin\omega}{2(\cos\omega-\cos\kappa)}\right)-g(\cos\kappa-e^{-\gamma}) \tag{64}$$

is a relatively smooth, well-behaved function of $\cos\kappa$. The relevant circle of convergence is that of (47) considered as a function of $e^{-i\kappa}$. Near this circle we have

$$e^{-i\kappa}=e^{\gamma+\delta}e^{i\theta}, \quad \text{where } |\delta| \text{ is small.}$$

Inserting this into (63), we find for the function G

$$G=\tfrac{1}{2}g[e^{-\gamma(N+1)}(e^{(\gamma+\delta)N}e^{iN\theta}+e^{-(\gamma+\delta)N}e^{-iN\theta}) - e^{-\gamma N}(e^{(\gamma+\delta)(N+1)}e^{i(N+1)\theta}+e^{-(\gamma+\delta)(N+1)}e^{-i(N+1)\theta})], \tag{65}$$

which for sufficiently large N becomes, apart from terms falling off exponentially with N,

$$G\cong\tfrac{1}{2}g(e^{-\gamma}e^{N\delta}e^{iN\theta}-e^{\gamma}e^{(N+1)\delta}e^{i(N+1)\theta}). \tag{66}$$

Comparison of (64) with (17) shows that F, considered as a function of $z=e^{-i\kappa}$, has zeroes at the exact roots $z=x_1, x_1^{-1}, x_2$, and x_2^{-1}. Inspection shows that it has simple poles at Y, Y^{-1}, and O. Thus, F can be written

$$F(z)=K\frac{(z-x_1)(z-x_1^{-1})(z-x_2)(z-x_2^{-1})}{(z-Y)(z-Y^{-1})(z)}, \tag{67}$$

where K is a constant.

In solving Eqs. (63)–(67), and using the results to evaluate $\tilde{\mathcal{R}}_N(s)$, we must distinguish between three cases, according to how many of the exact solutions lie within the circle of convergence.

Case (a). Both exact roots within circle of convergence, hence both well approximated by NA for large N; NA has two physical roots, $N-1$ unphysical roots. For large N, $\tilde{\mathcal{R}}_N(s)$ can be written

$$\tilde{\mathcal{R}}_N(s)=\tilde{\mathcal{R}}_{Np}(s)\tilde{\mathcal{R}}_{Nu}(s), \tag{68}$$

$$\tilde{\mathcal{R}}_{Np}(s)=\frac{(x_1-Y)(x_2-Y)}{(x_1-s)(x_2-s)}, \tag{69a}$$

$$\tilde{\mathcal{R}}_{Nu}(s)=\prod_{j=3}^{N+1}\frac{x_j-Y}{x_j-s}. \tag{70a}$$

Case (b). One exact root within circle of convergence and well-approximated by NA for large N; other exact root outside circle; NA has one physical root, N unphysical ones. $\tilde{\mathcal{R}}_N(s)$ can be expressed by (68), together with

$$\tilde{\mathcal{R}}_{Np}(s)=\frac{x_1-Y}{x_1-s}, \tag{69b}$$

$$\tilde{\mathcal{R}}_{Nu}(s)=\prod_{j=2}^{N+1}\frac{x_j-Y}{x_j-s}. \tag{70b}$$

Case (c). Both exact roots lie outside circle of convergence, neither one well approximated by NA; all $N+1$ roots of NA are unphysical. $\tilde{\mathcal{R}}_N(s)$ has the form

$$\tilde{\mathcal{R}}_N(s)=\tilde{\mathcal{R}}_{Nu}(s)=\prod_{j=1}^{N+1}\frac{x_j-Y}{x_j-s}. \tag{70c}$$

For all three cases, our task is to evaluate $\tilde{\mathcal{R}}_{Nu}(s)$, the contribution of the unphysical roots. The number of such roots is different for each case, but for sufficiently large N all lie close to the circle of convergence, so we can safely use (66). We rewrite the function G in (66) as

$$G(z)=e^{L(\delta+i\theta)}H(z), \tag{71}$$

where L and H are defined differently for each case, as follows:

case (a): $L=N$, $H(z)=\frac{1}{2}g(Z^{-1}-z)$, (72a)

case (b): $L=N+1$, $H(z)=\frac{1}{2}g(z^{-1}-Z)$, (72b)

case (c): $L=N+2$, $H(z)=\frac{1}{2}g(Z/z)(z^{-1}-Z)$. (72c)

In each case, we are seeking $L-1$ roots, apart from the spurious one at $z=Z$. In each case, the

roots are the solutions of

$$e^{L(\delta+i\theta)} = (z/Z)^L = F(z)/H(z) = Q(z). \tag{73}$$

Taking logs of both sides of (73), remembering that the log is only determined up to an additive multiple of $2\pi i$, and dividing by L, we find

$$\delta + i\theta = \ln(\tfrac{1}{2}z) = 2\pi i l/L + \ln Q(z)/L, \tag{74}$$

where $l = 1, 2, \ldots, L-1$ is not necessarily small compared to L. For $l = 0$ we obtain the spurious root mentioned above, and higher values of l repeat the same roots again. Thus, to within $O(N^{-1})$, the desired roots are

$$z_l = z_{l0} = Ze^{2\pi i l/L}.$$

In this approximation, we obtain for $\tilde{\mathcal{R}}_{Nu}(s)$

$$\tilde{\mathcal{R}}^{(0)}_{Nu}(s) = \prod_{l=1}^{L-1} \frac{Y - Ze^{2\pi i l/L}}{s - Ze^{2\pi i l/L}} = \frac{(Y^L - Z^L)(s - Z)}{(Y - Z)(s^L - Z^L)}. \tag{75}$$

In arriving at (75), we have used the fact that

$$\prod_{l=0}^{L-1} (Y - Ze^{2\pi i l/L}) = Y^L - Z^L,$$

so that

$$\prod_{l=1}^{L-1} (Y - Ze^{2\pi i l/L}) = \frac{Y^L - Z^L}{Y - Z}.$$

As $N, L \to \infty$, we see that

$$\tilde{\mathcal{R}}^{(0)}_{Nu}(s) \to \frac{s - Z}{Y - Z}, \quad |s| < Z; \tag{76}$$

$$\tilde{\mathcal{R}}^{(0)}_{Nu}(s) \to 0, \quad |s| > Z. \tag{76'}$$

In the case (76'), $|s| > Z$, the NA predicts zero for $\tilde{\mathcal{R}}(s)$, since neither $\tilde{\mathcal{R}}_{Nu}(s)$ nor the correction involving $\ln Q$ is infinite. In this region, $\mathcal{R}(s)$ is undefined according to the NA, as are all the subtracted functions, since in the NA for large N, there are arbitrarily many poles of $\tilde{\mathcal{R}}(s)$ near the circle of convergence. The exact result (57) is not zero in this region. Moreover, $\mathcal{R}(s)$ is defined in part of this region for case (c), $\mathcal{R}^{(1)}(s)$ for case (b), and $\mathcal{R}^{(2)}(s)$ for case (a). It is easy to verify that the subtracted functions for NA also disagree with the exact results. There are, therefore, well-defined observable quantities for which the NA converges to the wrong result as N becomes large.

It remains to be decided whether the NA gives the right result for $\tilde{\mathcal{R}}(s)$ in the other region, $|s| < Z$. The answer to this requires the evaluation of the correction term in (74) proportional to $\ln Q$. We rewrite (74) as

$$\ln(z_l/z_{l0}) = (1/L)\ln Q(z_l). \tag{77}$$

Using (77), (75), and (70a)–(70c), defining $z_0 = Ze^{i\theta}$, $\theta = 2\pi l/L$, and replacing a sum by an integral, we obtain

$$\begin{aligned}\ln\left(\frac{\tilde{\mathcal{R}}_{Nu}(s)}{\tilde{\mathcal{R}}^{(0)}_{Nu}(s)}\right) &= \sum_{l=1}^{L-1} \ln\left(\frac{(z_l - Y)(z_{l0} - s)}{(z_{l0} - Y)(z_l - s)}\right) \\ &\approx \frac{L}{2\pi}\int_0^{2\pi} \ln\left(\frac{(z - Y)(z_0 - s)}{(z_0 - Y)(z - s)}\right) d\theta \\ &= \frac{L}{2\pi i}\oint \ln\left(\frac{(z - Y)(z_0 - s)}{(z_0 - Y)(z - s)}\right)\frac{dz_0}{z_0},\end{aligned} \tag{78}$$

where the contour is the circle with radius Z.

From (64), (71), and (72a)–(72c), and (73), one sees that $z = z_0$ at $z = Z$, i.e., that $\ln Q(Z) = 0$. For convenience, we draw all cuts from the singularities of the various logarithms through this point. Thus, $\ln z, \ln z_0, \ln(z - Y)$, etc., all have discontinuities of $2\pi i$ at $z = z_0 = Z$.

Now consider the typical term in (78)

$$\oint \ln(z - Y)\frac{dz_0}{z_0} = \oint \ln(z - Y)\, d\ln z_0.$$

It can be transformed by means of partial integration into

$$[\ln(z - Y)\ln z_0]_z^{ze^{2\pi i}} - \oint \ln z_0\, d\ln(z - Y).$$

The integrated term is just $2\pi i \ln[Z(Z - Y)] - 4\pi^2$, since $z = z_0$ at both endpoints, and because of the discontinuity in the logs. The other term can be transformed by means of (77) and another partial integration to give

$$\begin{aligned}-\oint \ln z_0\, d\ln(z - Y) &= -\oint \ln z\, d\ln(z - Y) + \frac{1}{L}\oint \ln Q(z)\frac{dz}{z - Y} \\ &= -2\pi i \ln[Z(Z - Y)] + 4\pi^2 + \oint \ln(z - Y)\, d\ln z + \frac{1}{L}\oint \ln Q(z)\frac{dz}{z - Y}.\end{aligned}$$

The integrated parts now cancel, and the first integral is cancelled by the term which is the same function integrated over a slightly different contour, and with no singularity in the region between the contours. Thus, only the term in $\ln Q$ remains. Doing the same thing for the other terms in (78), one obtains

$$\ln\left(\frac{\tilde{\mathcal{R}}_{Nu}(s)}{\tilde{\mathcal{R}}^{(0)}_{Nu}(s)}\right) = \frac{1}{2\pi i}\oint \ln Q(z)\left(\frac{1}{z - Y} - \frac{1}{z - s}\right)dz. \tag{79}$$

The contour is now slightly distorted from the circle $|z| = Z$, and passes through the exact solutions. Its evaluation is different for the three cases.

Case (*a*). For this case, one sees from (67) and (72a) that Q has the form

$$Q(z) = K_a \frac{(z - x_1)(z - x_1^{-1})(z - x_2)(z - x_2^{-1})}{z(z - Y)(z - Y^{-1})(z - Z^{-1})} . \quad (80)$$

Q thus has four zeroes and four poles, all within the contour of integration. The integral (79) breaks down into a sum of integrals of the form

$$\oint \frac{dz}{z - a} \ln\left(\frac{z - b}{z - c}\right), \quad (81)$$

which is readily seen to be zero if a, b, c are all inside the contour, as in the present case. The correction is thus zero in this case. From (68), (69a), and (76), we see that the exact result (57) is reproduced.

Case (*b*). This time we have

$$Q = K_b \frac{(z - x_1)(z - x_1^{-1})(z - x_2)(z - x_2^{-1})}{(z - Y)(z - Y^{-1})(z - Z^{-1})} .$$

The zero at $z = x_2$ is now outside the contour, the other zeroes and poles inside. The integral (79) is a sum of terms of the form (81), which contribute nothing, plus a term

$$\frac{1}{2\pi i} \oint \ln(z - x_2)\left(\frac{1}{z - Y} - \frac{1}{z - s}\right) = \ln\left(\frac{x_2 - Y}{x_2 - s}\right). \quad (82)$$

Equation (82), together with (68), (69b), and (76), again reproduces the exact result (57).

Case (*c*). This is similar to case (b). We have this time

$$Q = K_c \frac{z(z - x_1)(z - x_1^{-1})(z - x_2)(z - x_2^{-1})}{(z - Y)(z - Y^{-1})(z - Z^{-1})} .$$

There are now two zeroes outside the contour, plus terms of the form (81), leading to a correction of

$$\ln\left(\frac{(x_1 - Y)(x_2 - Y)}{(x_1 - s)(x_2 - s)}\right),$$

which again leads to the exact result (57).

We conclude, then, that the NA for large N converges to the correct result for $\mathfrak{R}(s)$ or $\tilde{\mathfrak{R}}(s)$ for $|s| < Z$, but not for $|s| > Z$. In particular, it gives the right result for the reflection coefficient. Where it converges, the convergence is fairly rapid. The only approximations made in the above derivation were the omission of terms falling off exponentially with N in arriving at (66), and the replacement of the sum by the integral in (78). The error thus falls off exponentially with N. The convergence is not uniform in ω, however. When one of the exact solutions is sufficiently near to the circle $|z| = Z$, the replacement of the sum by the integral in (78) is not a good approximation for those terms lying near this exact root. A look at Figs. 1 and 2 shows that one of the exact solutions always passes through $-Z$ for some ω. At this point, and for a narrow frequency range around it, there will be an appreciable error in the NA even for arbitrarily large N. The frequency range in which the error is appreciable will, of course, become narrower as N becomes large.

D. Conclusions

According to the foregoing analysis, the NA converges to the correct result for some physically observable quantities (including those most frequently of practical interest), but not for all. Its usefulness, therefore, depends on the specific purpose for which it is to be used. One can use it to calculate reflection coefficients, and, for small N, also as a model for gaining insight into qualitative properties of the solution. There are always observable properties, however, for which it gives the wrong answer, and one must ascertain in each case whether the quantity being calculated is one of those for which the NA is applicable. The criterion for this would seem to be whether the quantity in question involves use of the polynomial approximation outside the radius of convergence of the power series.

As N becomes large, the analytic structure of the NA solution [as expressed, e.g., through the singularities of $\tilde{\mathfrak{R}}(s)$] becomes increasingly different from that of the exact solution. The NA, therefore, would appear to be a highly inappropriate model for the purpose of gaining insight into the mathematical structure of the solution.

Apart from quantitative details, the derivation just given depended not so much on the particular model considered in this paper, but on the fact that the power-series expansion possesses a finite radius of convergence. Whenever this is the case, most of the NA roots for large N will lie near the circle of convergence, i.e., in a region where the actual function is not well approximated by the Nth-order polynomial, and one would expect to obtain erroneous results for some quantities by taking these roots seriously, as the NA procedure requires one to do. The radius of convergence is finite, and the NA therefore suspect, whenever the falloff of the interaction is no faster than ex-

ponential. When the interaction falls off faster than exponentially (Gaussian, factorial), one would expect the NA to converge to the correct answer for all quantities of interest.

One final question is whether, in view of the finiteness of actual crystals, it might be better to use the limit of the NA for large N than the exact result for the semi-infinite crystal. The answer to this question turns out to be in the negative. If the crystal is of finite thickness, one still obtains Eq. (17) unaltered for the allowed wave numbers within the crystal. The difference is that now there are no infinite sums, so one cannot require $|e^{ik}| \leq 1$, which means that each solution of (17) leads to two solutions for e^{ik} (one propagating forward, the other backward). One also finds twice as many boundary conditions, so the problem is still soluble, and the solution still differs from the NA.

V. DISCUSSION

For the soluble model considered in this paper, it has been shown that neither the dielectric nor the near-neighbor approximation leads to correct results for all physically observable properties. In both cases, arguments have been given for the belief that this is not an artifact of our model, but a general conclusion.

Of the two approximations considered, the NA appears definitely to be the better. Despite its failure to converge to the exact result, for all properties, it remains useful as a model which violates no fundamental principles, possesses no blatantly unphysical properties, and actually gives correct results for many properties, including those usually of most interest.

The DA, on the other hand, possesses the highly unphysical property that energy is always being leaked into fictitious polarization waves outside the material medium. Because of this, and because other models are available which are at least equally easy to use and which lack such unphysical properties, the author feels that the DA in the future should be used, if at all, only with the greatest caution.

Another matter worth pursuing is the use of the model of this paper in practical calculations. Actual planewise interactions in crystals are not exactly exponential in their distance dependence, but they can be represented as an infinite, rapidly converging sum of exponentials.[25-28] The present model can easily be extended to treat an interaction represented by a *finite* sum of exponentials; the amount of labor involved for N exponentials is essentially the same as that for N nearest neighbors. Before making elaborate and expensive calculations, however, one should first ascertain whether this approach encounters convergence problems similar to those of the NA. The author hopes to make this the subject of a future publication. If the prognosis is favorable, it would be worthwhile to perform some calculations, using the methods of Philpott[7] to permit the treatment of actual crystals with other than normal incidence.

ACKNOWLEDGMENT

The author is most grateful to Professor G. D. Mahan for pointing out an error in the original version of this paper.

APPENDIX: DERIVATION OF EQS' (60)–(62)

In the Mahan-Obermair theory, one has[6]

$$\sum_j \frac{q_j}{x_j - Y} = 1, \tag{A1}$$

$$\sum_j q_j x_j^l = 0, \quad l = 0, 1, 2, \ldots, (N-1). \tag{A2}$$

Thus, one immediately finds

$$\tilde{\mathcal{R}}(s) = \sum_j \frac{q_j}{x_j - s} = \frac{D(s)}{D(Y)}, \tag{A3}$$

with

$$D(\xi) = \begin{vmatrix} (x_1 - \xi)^{-1} & (x_2 - \xi)^{-1} & \cdots \\ 1 & 1 & \cdots \\ x_1 & x_2 & \cdots \\ x_1^2 & x_2^2 & \cdots \\ \cdots & \cdots & \cdots \end{vmatrix} . \tag{A4}$$

Expanding the determinant (A4) in cofactors of elements in the first row, one finds that each cofactor is a determinant of the type

$$\pm \prod_{\substack{j<h \\ j,K\neq l}} (x_j - x_K).$$

Since this depends only on the differences of the x_j, it will not affect the value of the determinant if each x_j outside the first row is replaced by $x_j - \xi$. With this replacement, one can factor out $\prod(x_j - \xi)^{-1}$ and then *add* ξ to the various x_j to obtain

$$D(\xi)=\begin{vmatrix} (x_1-\xi)^{-1} & (x_2-\xi)^{-1} & \cdots \\ 1 & 1 & \cdots \\ (x_1-\xi) & (x_2-\xi) & \cdots \\ (x_1-\xi)^2 & (x_2-\xi)^2 & \cdots \\ \cdots & \cdots & \cdots \end{vmatrix} = \prod_{j=1}^{N+1}(x_j-\xi)^{-1} \begin{vmatrix} 1 & 1 & \cdots \\ (x_1-\xi) & (x_2-\xi) & \cdots \\ (x_1-\xi)^2 & (x_1-\xi)^2 & \cdots \\ \cdots & \cdots & \cdots \end{vmatrix}$$

$$= \prod_{j=1}^{N+1}(x_j-\xi)^{-1} \begin{vmatrix} 1 & 1 & \cdots \\ x_1 & x_2 & \cdots \\ x_1^2 & x_2^2 & \cdots \\ \cdots & \cdots & \cdots \end{vmatrix} = (-1)^{N(N+1)/2} \prod_{j=1}^{N+1}(x_j-\xi)^{-1} \prod_{j<h}(x_j-x_h). \tag{A5}$$

From (A3) and (A5), one immediately obtains (60). Equations (61) and (62) are obtained by subtracting off the contributions of the poles at $z=x_1, x_2$.

[1]S. I. Pekar, Zh. Eksp. Teor. Fiz. 34, 1176 (1958) [Sov. Phys.–JETP 7, 813 (1958)].
[2]J. J. Hopfield and D. G. Thomas, Phys. Rev. 132, 563 (1963).
[3]V. M. Agranovich and V. L. Ginzburg, *Spatial Dispersion in Crystal Optics and the Theory of Excitons* (Interscience, New York, 1966).
[4]C. A. Mead, Radiat. Res. 20, 101 (1963).
[5]C. W. Deutsche and C. A. Mead, Phys. Rev. 138, A63 (1965).
[6]G. D. Mahan and G. Obermair, Phys. Rev. 183, 834 (1969).
[7]M. R. Philpott, J. Chem. Phys. 60, 1410 (1974).
[8]M. R. Philpott, J. Chem. Phys. 60, 2520 (1974).
[9]J. J. Sein, Phys. Lett. A 32, 141 (1970).
[10]J. J. Sein, J. Opt. Soc. Am. 62, 1037 (1972).
[11]J. L. Birman and J. J. Sein, Phys. Rev. B 6, 2482 (1972).
[12]G. S. Agarwal, D. N. Pattanayak, and E. Wolf, Phys. Rev. Lett. 27, 1022 (1971).
[13]G. S. Agarwal, D. N. Pattanayak, and E. Wolf, Opt. Commun. 4, 255 (1971).
[14]G. S. Agarwal, D. N. Pattanayak, and E. Wolf, Opt. Comm. 4, 260 (1971).
[15]G. S. Agarwal, D. N. Pattanayak, and E. Wolf, Phys. Lett. A 40, 279 (1972).
[16]G. S. Agarwal, D. N. Pattanayak, and E. Wolf, Phys. Rev. B 10, 1447 (1974).
[17]A. A. Maradudin and D. L. Mills, Phys. Rev. B 7, 2787 (1973).
[18]G. S. Agarwal, Phys. Rev. A 11, 230 (1975).
[19]G. S. Agarwal, Phys. Rev. A 11, 243 (1975).
[20]G. S. Agarwal, Phys. Rev. A 11, 253 (1975).
[21]G. S. Agarwal, D. N. Pattanayak, and E. Wolf, Phys. Rev. B 11, 1342 (1975).
[22]J. T. Foley and A. J. Devaney, Phys. Rev. B 12, 3104 (1975).
[23]M. Grosmann, J. Biellmann, and S. Nikitina, in *Springer Tracts in Modern Physics*, edited by G. Höher (Springer, Berlin, 1975), Vol. 73, pp. 242 ff.
[24]J. E. Sipe and J. Van Kranendonk, Can. J. Phys. 53, 2095 (1975).
[25]B. R. Nijboer and F. W. de Wette, Physica (Utr.) 24, 422 (1958).
[26]F. W. de Wette and G. E. Schacher, Phys. Rev. 137, A78 (1965).
[27]A. Lucas, J. Chem. Phys. 48, 3156 (1968).
[28]M. R. Philpott, J. Chem. Phys. 58, 588 (1973).