Homogenization of Maxwell's equations in periodic composites: Boundary effects and dispersion relations

Vadim A. Markel[*]

Department of Radiology and Graduate Group in Applied Mathematics and Computational Science,
University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

John C. Schotland[†]

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109, USA

We consider the problem of homogenizing the Maxwell equations for periodic composites. The analysis is based on Bloch-Floquet theory. We calculate explicitly the reflection coefficient for a half space and derive and implement a computationally efficient continued-fraction expansion for the effective permittivity. Our results are illustrated by numerical computations for the case of two-dimensional systems. The homogenization theory of this paper is designed to predict various physically measurable quantities rather than to simply approximate certain coefficients in a partial differential equation.

PACS number(s): 41.20.Jb, 78.67.−n, 78.20.Bh

I. INTRODUCTION

Theories of electromagnetic homogenization of composite materials, also known as effective-medium theories (EMTs), have a history that dates to the time of Maxwell. Nevertheless, these theories continue to attract attention and even controversy, as evidenced by recent reviews [1,2] and many references therein. In applied mathematics the theory of homogenization based on multiscale analysis of partial differential equations is also well established [3–6]. However, interest in EMTs has been steadily on the rise for the past ten years with conceptually new approaches continuing to appear [7–10]. This can be explained perhaps by noting that the tasks of relating the existing mathematical theories to physical observables and of determining the range of applicability of a given theory have not been fully addressed, particularly for the case of Maxwell's equations. Indeed, in the past ten years or so homogenization theories have been applied to obtain extreme properties of electromagnetic composites, including the phenomenon of strong artificial magnetism. At the same time significant experimental progress has been made recently in manufacturing deeply subwavelength (in the visible spectral range) periodic metallic nanostructures [11–13]. The question is whether the existing theories are directly applicable or accurate enough to guide the experimental design of periodic nanostructures of desirable properties. Another reason for the renewed interest in homogenization theories is that, in addition to abstract mathematical results, there is a need for efficient, stable computational methods. Thus the question of how to construct physically relevant and computationally effective EMTs and determine their limits of applicability have not been completely settled.

This paper is an attempt to address the above issues for the case of periodic composites; random media are not considered. The framework we develop is based on the Bloch-Floquet

expansion, which is a well-known tool in homogenization theory [5,14–19]. However, in several aspects we go beyond the standard theory. In particular (i) we explicitly account for boundary effects and derive a general expression for the reflection coefficient (many existing homogenization theories consider infinite composites) and (ii) we make use of the integral equation formulation of scattering theory for the Maxwell equations. The resulting formulas for the effective-medium parameters (EMPs) have a different mathematical structure than those derived from partial differential equations; (iii) we develop a computationally efficient algorithm for calculating the EMPs. The algorithm is based on a continued-fraction expansion of the self-energy and is obtained from a result on the resolvent of a linear operator and (iv) a numerical study of stability and convergence is performed for some test cases. Stability is investigated by comparing the results for inclusions of the same volume fraction but different shape and of the same shape but different volume fractions.

It is useful to recognize that all EMTs can be classified as either *standard* or *extended*. A standard EMT is obtained by taking the limit $h \to 0$, where h is the scale of the medium's heterogeneity; in this paper h is the lattice spacing. In standard theories h is viewed as a mathematically and physically independent variable and the resulting EMPs are independent of h as long as the latter is small enough for the theory to be applicable. Another feature of all standard theories is the so-called law of unaltered ratios [20], which states that if a composite medium is made of several constituents with permittivities ϵ_j $(j = 1,2,\dots)$ and if $\epsilon_j \to \lambda\epsilon_j$ $(\lambda > 0)$, then the effective permittivity $\bar{\epsilon}$ also scales as $\bar{\epsilon} \to \lambda\bar{\epsilon}$.

Extended EMTs came to the fore (at least in the physics literature) in Refs. [21,22]. The basic idea of these papers is to note that one can compute the exact electric and magnetic polarizabilities α_e and α_m of a spherical particle through the use of the first Lorenz-Mie coefficients a_1 and b_1 even when the sphere in question is not small compared to the external wavelength. These polarizabilities can be used to construct an extended Maxwell-Garnett approximation. Since a_1 and b_1 are not proportional to the sphere volume, except in the quasistatic

―――――――――
[*]vmarkel@mail.med.upenn.edu
[†]schotland@umich.edu

limit, the resultant EMTs contain the sphere radius explicitly. In Refs. [23,24] a counterintuitive effect of noncommuting limits was demonstrated. Specifically, it was shown that insofar as the effective refractive index of a photonic crystal is computed from the slope of the dispersion curve near the Γ point, different results are generally obtained depending on which of the two limits $h \to 0$ and $\epsilon_1 \to \infty$ is taken first, where ϵ_1 is the permittivity of one of the components of the photonic crystal. A related point is that a complete theory of homogenization requires error estimates. That is, it is essential to determine how the error in the homogenization limit depends upon contrast. Moreover, the reflection and transmission properties of the composite medium have not been considered [21–24].

In this paper we develop a standard EMT. However, when considering reflection and refraction at a planar interface, we derive formulas for the reflection and transmission coefficients, which are valid for finite values of h. Then we show that taking the limit $h \to 0$ results in the standard Fresnel coefficients. In this case the electric and magnetic properties of the medium constituents do not mix, in agreement with Ref. [25]. That is, if we begin with nonmagnetic inclusions, the resultant composite is also nonmagnetic. An extended EMT can be obtained by taking a different limit in which the permittivity of one of the constituents scales as $1/h^2$ [26]. Here we note again the existence of the effect of noncommuting limits [23,24,27], which calls for additional scrutiny of the homogenization results thus obtained. In particular, one would expect that in the limit considered in Ref. [26], Fresnel formulas would also be reproduced, but with a nontrivial magnetic permeability. We have not been able to show that this is the case. In other words, it is not clear whether the EMPs obtained from an extended EMT are independent of the incidence angle or, more generally, of the type of incident wave. This is in accord with Refs. [28–33], which find that the conditions under which metamaterials exhibiting strong magnetic resonances can be assigned purely local (incidence-angle-independent) EMPs are rather restrictive. The same point has been made in the recent review [2].

An additional feature by which EMTs can be classified is the physical model of the medium. In the model of dipole lattices the medium is thought of as being composed of point particles that are completely characterized by their polarizabilities (electric and possibly magnetic) and whose shape and size do not enter into the problem directly [34–36]. Alternatively, one can consider the space as a two-component continuous medium [37–39]. The point-dipole model is appealing because of its simplicity but leads to serious mathematical problems. The so-called dipole sum (also known as the lattice sum or the dipole self-energy), which plays a key role in this model, diverges in the case of three-dimensional (3D) lattices. While it is true that even divergent series can be summed by means of applying various mathematical tricks, the results obtained depend on the particular trick used, a state of affairs that is not very satisfying. Therefore, we will adopt from the start a model of a two-component continuous medium. As the development in this paper progresses, it will become apparent why the point-dipole model is inadequate.

The mathematical development in this paper begins by considering the integral equation obeyed by the polarization

field, which is introduced in Sec. II. In Sec. III we derive a homogenization theory of the standard type for infinite periodic media. Reflection and refraction at a planar boundary are considered in Sec. IV. In Sec. V we discuss the correspondence between the point-dipole model and the continuous-medium model of this paper. A continued fraction expansion of the effective permittivity is derived in Sec. VI and used in the numerical simulations of Sec. VII. The expansion has its origins in a theorem on resolvents of general linear operators (with no special symmetry properties), which is stated in Sec. VI and proved in the Appendixes. A discussion and a summary of results are contained in Secs. VIII and IX.

II. BASIC EQUATIONS

The geometry of the problem we consider is sketched in Fig. 1. The medium consists of two intrinsically nonmagnetic constituents: a host medium of permittivity ϵ_b and periodically arranged inclusions of permittivity ϵ_a. In practice, the host is often a transparent dielectric with $\mathrm{Re}\epsilon_b > 0$ and $0 < \mathrm{Im}\epsilon_b \ll \mathrm{Re}\epsilon_b$ and the inclusions are metallic. However, the theory of this paper places no such restriction on the permittivities and requires only that $\mathrm{Im}\epsilon_b > 0$ and $\mathrm{Im}\epsilon_a > 0$. In the case when the host medium is a vacuum, we will set $\epsilon_b = 1 + i0$. The inclusions are arranged on a cubic lattice of period h. The position vector of the center of each unit cell is denoted by $\mathbf{r}_n$, where n can be viewed as a composite index: $n = (n_x, n_y, n_z)$ and $\mathbf{r}_n = h(\hat{\mathbf{x}}n_x + \hat{\mathbf{y}}n_y + \hat{\mathbf{z}}n_z)$. Whenever a summation over n (or a similar composite index m) appears in the text it is implied that the sum runs over all three integer indices. Inside the nth cell the spatial region Ω_n has the permittivity ϵ_a; the rest of the cell has the background permittivity ϵ_b. All regions Ω_n are identical and differ only by translation. It is assumed that Ω_n can touch but not cross the cell boundaries. No assumption on the connectivity of Ω_n is made. The union of all regions Ω_n is denoted by Ω_{tot} and the volume of each region is denoted by V:

$$\Omega_{\mathrm{tot}} = \bigcup_n \Omega_n, \quad \int_{\Omega_n} d^3r = V. \tag{1}$$

We work in the frequency domain and the common factor $\exp(-i\omega t)$ is suppressed. All frequency-dependent quantities,

$$\mathbf{r}_n = h(n_x\hat{\mathbf{x}} + n_y\hat{\mathbf{y}} + n_z\hat{\mathbf{z}})$$

$$-\infty < n_x, n_y, n_z < \infty$$
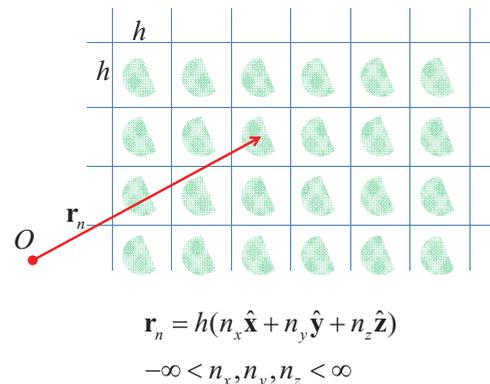
FIG. 1. (Color online) Sketch of the geometry considered: an infinite 3D lattice.

such as the permittivities ϵ_b and ϵ_a, are evaluated at the frequency ω.

The mathematical development in this paper begins with the integral equation

$$\mathbf{P}(\mathbf{r}) = \frac{3\chi}{4\pi}\left[\mathbf{E}_i(\mathbf{r}) + \int_{\Omega_{\text{tot}}} G(\mathbf{r},\mathbf{r}')\mathbf{P}(\mathbf{r}')d^3r'\right], \quad \mathbf{r} \in \Omega_{\text{tot}}. \quad (2)$$

Here $\mathbf{P}(\mathbf{r})$ is the vector of polarization, which is related to the electric field $\mathbf{E}(\mathbf{r})$ by

$$\mathbf{P}(\mathbf{r}) = \frac{\epsilon(\mathbf{r}) - \epsilon_b}{4\pi\epsilon_b}\mathbf{E}(\mathbf{r}), \quad (3)$$

$\mathbf{E}_i(\mathbf{r})$ is the incident electric field, $G(\mathbf{r},\mathbf{r}')$ is the *regular* part of the free-space retarded Green's tensor, and

$$\chi = \frac{\epsilon_a - \epsilon_b}{\epsilon_a + 2\epsilon_b}. \quad (4)$$

Note that $\mathbf{P}(\mathbf{r})$ defined in Eq. (3) is not the true physical polarization, which is given by $[\epsilon(\mathbf{r}) - 1]\mathbf{E}(\mathbf{r})/4\pi$, but rather an auxiliary field; $\mathbf{P}(\mathbf{r})$ vanishes in the host medium while the true polarization does not.

In what follows we will make use of the spatial Fourier transform of the Green's tensor, namely,

$$G(\mathbf{r},\mathbf{r}') = \frac{4\pi}{3}\int \frac{d^3p}{(2\pi)^3}K(\mathbf{p})\exp[i\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')], \quad (5)$$

where

$$K(\mathbf{p}) = \frac{2k_b^2 + p^2 - 3\mathbf{p}\otimes\mathbf{p}}{p^2 - k_b^2} \quad (6)$$

and

$$k_b^2 = \epsilon_b k, \quad k = \frac{\omega}{c}. \quad (7)$$

Here the wave number in the background medium is denoted by k_b and the wave number in vacuum is denoted by k. We note that the integral equation (2) is equivalent to the pair of curl Maxwell equations written in the frequency domain.

III. WAVES IN INFINITE LATTICES

A. Three-dimensional lattices

Consider the propagation of a wave in a three-dimensional infinite lattice. In this case the incident field is absent and Eq. (2) must be satisfied for $\mathbf{E}_i = 0$. We seek the solution to Eq. (2) in the form of a Bloch wave:

$$\mathbf{P}(\mathbf{r}) = \exp(i\mathbf{q}\cdot\mathbf{r}_n)\mathbf{F}(\mathbf{r}-\mathbf{r}_n), \quad \mathbf{r}\in\Omega_n. \quad (8)$$

Here $\mathbf{q}$ is the Bloch wave number and $\mathbf{F}(\mathbf{r})$ is a vector function. Equivalently, if we write $\mathbf{r} = \mathbf{r}_n + \mathbf{R}$, then

$$\mathbf{P}(\mathbf{r}_n + \mathbf{R}) = \exp(i\mathbf{q}\cdot\mathbf{r}_n)\mathbf{F}(\mathbf{R}), \quad \mathbf{R}\in\Omega. \quad (9)$$

In this formula $\Omega \equiv \Omega_0$ is the region centered at the origin of a rectangular reference frame. From the above relation we find the equation obeyed by $\mathbf{F}(\mathbf{R})$:

$$\mathbf{F}(\mathbf{R}) = \frac{3\chi}{4\pi}\int_{\Omega} W(\mathbf{R},\mathbf{R}')\mathbf{F}(\mathbf{R}')d^3R', \quad (10)$$

where

$$W(\mathbf{R},\mathbf{R}') = \sum_m G(\mathbf{r}_n + \mathbf{R},\mathbf{r}_m + \mathbf{R}')\exp[i\mathbf{q}\cdot(\mathbf{r}_m - \mathbf{r}_n)]. \quad (11)$$

It can be seen that W is independent of n. It should also be noted that the summation in Eq. (11) runs over the entire lattice, including the term $m = n$. In theories that consider pointlike particles, the dipole sum is defined as an incomplete lattice sum, which excludes the term $m = n$. This makes application of the Poisson summation formula problematic and unnecessarily complicates the mathematics [36].

Returning to our derivation, we evaluate W as

$$W(\mathbf{R},\mathbf{R}') = \frac{4\pi}{3}\int \frac{d^3p}{(2\pi)^3}K(\mathbf{p})\exp[i\mathbf{p}\cdot(\mathbf{R}-\mathbf{R}')]$$
$$\times \sum_m \exp[i(\mathbf{p}-\mathbf{q})\cdot(\mathbf{r}_n - \mathbf{r}_m)]$$
$$= \frac{4\pi}{3h^3}\sum_{\mathbf{g}} K(\mathbf{q}+\mathbf{g})\exp[i(\mathbf{q}+\mathbf{g})\cdot(\mathbf{R}-\mathbf{R}')], \quad (12)$$

where

$$\mathbf{g} = \frac{2\pi}{h}(\hat{\mathbf{x}}n_x + \hat{\mathbf{y}}n_y + \hat{\mathbf{z}}n_z) \quad (13)$$

are the reciprocal lattice vectors and we have used the Poisson summation formula

$$\sum_m \exp[i(\mathbf{p}-\mathbf{q})\cdot(\mathbf{r}_m - \mathbf{r}_n)] = \left(\frac{2\pi}{h}\right)^3\sum_{\mathbf{g}}\delta(\mathbf{p}-\mathbf{q}-\mathbf{g}). \quad (14)$$

The summation in Eqs. (12) and (14) is over the complete set of reciprocal lattice vectors; equivalently, it can be viewed as a summation over the triplet of indices (n_x, n_y, n_z) that appear in Eq. (13).

The series on the right-hand side of Eq. (12) diverges when $\mathbf{R} = \mathbf{R}'$. This is the well-known divergence of the dipole sum [40], which hinders the analysis of waves in lattices made of pointlike polarizable particles. The model of pointlike dipoles is discussed in more detail in Sec. V. In the equations derived above, the divergence is of no concern because $W(\mathbf{R},\mathbf{R}')$ appears only inside an integral and the singularity in question is integrable.

Upon substitution of Eq. (12) into Eq. (10) we obtain

$$\mathbf{F}(\mathbf{R}) = \frac{\chi}{h^3}\sum_{\mathbf{g}} K(\mathbf{q}+\mathbf{g})\exp[i(\mathbf{q}+\mathbf{g})\cdot\mathbf{R}]$$
$$\times \int_{\Omega} \mathbf{F}(\mathbf{R}')\exp[-i(\mathbf{q}+\mathbf{g})\cdot\mathbf{R}']d^3R'. \quad (15)$$

It follows from Eq. (15) that $\mathbf{F}(\mathbf{R})$ can be expanded as

$$\mathbf{F}(\mathbf{R}) = \sum_{\mathbf{g}}\mathbf{F}_{\mathbf{g}}\exp[i(\mathbf{q}+\mathbf{g})\cdot\mathbf{R}] \quad (16)$$

and that the expansion coefficients satisfy the system of equations

$$\mathbf{F}_{\mathbf{g}} = \rho\chi K(\mathbf{q}+\mathbf{g})\sum_{\mathbf{g}'} M(\mathbf{g}-\mathbf{g}')\mathbf{F}_{\mathbf{g}'}, \quad (17)$$

where $\rho = V/h^3$ is the volume fraction of inclusions and $M(\mathbf{g})$ is defined by the expression

$$M(\mathbf{g}) = \frac{1}{V} \int_{\Omega} \exp(-i\mathbf{g} \cdot \mathbf{R}) d^3 R. \qquad (18)$$

Note that $M(\mathbf{g})$ is defined only by the shape of the inclusions and is invariant with respect to the coordinate rescaling $\mathbf{r} \to \lambda\mathbf{r}$. Some mathematical properties and calculations of $M(\mathbf{g})$ for special geometries are given in Appendix A.

So far we have simply restated the well-known theorem of Floquet. The eigenproblem (17) defines the band structure of a photonic crystal. It is well known that EMTs are not always applicable to photonic crystals. However, there exists a regime in which EMPs can be reasonably introduced; this regime will be explored below. Namely, if $qh, k_b h \ll 1$, we can consider the cases $\mathbf{g} = 0$ and $\mathbf{g} \neq 0$ in Eq. (17) separately. This yields the following equations:

$$\mathbf{F}_0 = \rho \chi K(\mathbf{q}) \left[\mathbf{F}_0 + \sum_{\mathbf{g} \neq \mathbf{0}} M(-\mathbf{g}) \mathbf{F}_{\mathbf{g}} \right], \qquad (19a)$$

$$\mathbf{F}_{\mathbf{g}} = \rho \chi Q(\mathbf{g}) \left[M(\mathbf{g}) \mathbf{F}_0 + \sum_{\mathbf{g}' \neq \mathbf{0}} M(\mathbf{g} - \mathbf{g}') \mathbf{F}_{\mathbf{g}'} \right], \quad \mathbf{g} \neq 0, \qquad (19b)$$

where

$$Q(\mathbf{g}) \equiv \lim_{h \to 0} K(\mathbf{q} + \mathbf{g}) = 1 - 3\hat{\mathbf{g}} \otimes \hat{\mathbf{g}}, \quad \mathbf{g} \neq 0. \qquad (20)$$

Here $\hat{\mathbf{g}} = \mathbf{g}/|\mathbf{g}|$ is a unit vector.

The derivation of Eqs. (19) is one of the key developments of this paper. It can be seen that Eq. (19b) does not contain the variable k or $\mathbf{q}$, but is completely defined by the geometry of inclusions and by the variable χ. Moreover, these equations are invariant with respect to the rescaling $\mathbf{r} \to \lambda\mathbf{r}$. For any given value of $\mathbf{F}_0$, Eq. (19b) can be solved uniquely as $\mathbf{F}_{\mathbf{g}} = A_{\mathbf{g}} \mathbf{F}_0$, where the tensor $A_{\mathbf{g}}$ depends on $\mathbf{g}$, the shape of inclusions, and on χ. Given this result, we can write

$$\sum_{\mathbf{g} \neq \mathbf{0}} M(-\mathbf{g}) \mathbf{F}_{\mathbf{g}} = \sum_{\mathbf{g} \neq \mathbf{0}} M(-\mathbf{g}) A_{\mathbf{g}} \mathbf{F}_0 = \Sigma \mathbf{F}_0, \qquad (21)$$

where the tensor Σ has all the properties of $A_{\mathbf{g}}$ and in addition is independent of $\mathbf{g}$. It will be shown in Sec. VI that Σ plays the role of the self-energy and originates due to the electromagnetic interaction within and between the inclusions. It will also be shown that Σ can be computed as a resolvent of a linear operator, which depends only on the shape of inclusions.

Using the notation introduced in Eq. (21), we can rewrite Eq. (19a) as

$$[1 - \rho \chi K(\mathbf{q})(1 + \Sigma)]\mathbf{F}_0 = 0. \qquad (22)$$

This equation has nontrivial solutions if

$$\det[1 - \rho \chi K(\mathbf{q})(1 + \Sigma)] = 0. \qquad (23)$$

Here the quantity in the square brackets is a 3×3 matrix. For a fixed value of k (that is, at a fixed frequency), the condition (23) is an algebraic equation with respect to the Cartesian components of the Bloch vector $\mathbf{q}$. Roots of this equation, computed at different values of k, determine the dispersion relation $\mathbf{q}(k)$. There can be more than one branch of the dispersion relation corresponding to different polarization states. By polarization of the mode we mean here the direction of the vector $\mathbf{F}_0$.

Effective-medium parameters can be inferred by comparing these results to the polarization states and dispersion relation in a homogeneous medium characterized by tensor permittivity and permeability $\bar{\epsilon}$ and $\bar{\mu}$. However, it is not possible to determine $\bar{\epsilon}$ and $\bar{\mu}$ *simultaneously and uniquely* from consideration of the dispersion relation alone. For example, in an isotropic medium, only the product of these two quantities (the squared refractive index) can be unambiguously obtained. Indeed, the dispersion relation in such a medium is invariant with respect to the transformation $\bar{\epsilon} \to \xi\bar{\epsilon}$ and $\bar{\mu} \to \xi^{-1}\bar{\mu}$, where $\xi \neq 0$ is a complex number. To determine $\bar{\epsilon}$ and $\bar{\mu}$ uniquely one must consider reflection and refraction at the medium boundary. This will be done in Sec. IV. In particular it will be shown that in order to obtain the correct Fresnel reflection coefficients one must set $\bar{\mu} = 1$.

To summarize the results of this section, the electromagnetic modes of a medium can be found if the tensor Σ is known. Computation of the modes involves diagonalization of a 3×3 matrix, while the tensor Σ is uniquely determined by the solution to Eq. (19b). The latter is an infinite set of equations that must be appropriately truncated in numerical computations. Thus we have reduced the homogenization problem to solving a set of algebraic equations in which the shape of the inclusions appears only in the functions $M(\mathbf{g})$.

B. Main homogenization result for three-dimensional composites with well-defined optical axes

The standard description of electromagnetic waves in anisotropic crystals is based on the assumption that the tensors $\bar{\epsilon}$ and $\bar{\mu}$ commute and are simultaneously diagonalizable by a rotation of the reference frame with purely real Euler angles. The axes of the reference frame in which $\bar{\epsilon}$ and $\bar{\mu}$ are diagonal are known as the optical axes. Moreover, standard textbooks often specialize to the case $\bar{\mu} = 1$, which is a very good approximation in crystal optics [41]. In the most general case, however, the tensors $\bar{\epsilon}$ and $\bar{\mu}$ do not commute, which gives rise to two distinct sets of electric and magnetic axes. Furthermore, $\bar{\epsilon}$ and $\bar{\mu}$ are complex valued, symmetric, and hence non-Hermitian matrices. A purely real rotation that diagonalizes any one of these two tensors may not exist. A mathematically tractable dispersion relation for the most general case has been derived only recently [42]; we will use below one particular case of this result.

For the composite medium consisting of nonmagnetic components, which is considered in this paper, the situation is somewhat simpler. It can be seen from Eq. (23) that a unique set of optical axes exists if the tensor Σ is diagonalizable by a real-angle rotation of the reference frame. Thus the issue of commutability of two different tensors does not arise in this case.

In this section we assume that the optical axes of the composite medium (that is, the principal axes of the tensor Σ) exist and moreover coincide with the crystallographic axes of the medium. The latter assumption is not really necessary, but any composite can be cut is such a way that it holds.

In particular, Σ is diagonal in the reference frame defined by the crystallographic axes (which is the laboratory frame in this paper) if the inclusions are symmetric with respect to reflections in each of the xy, xz, and yz planes. The principal values of Σ, denoted by $\Sigma_{\alpha\alpha}$ ($\alpha = x,y,z$), are not necessarily equal in this case. The two familiar examples of reflection-symmetric inclusions that result in all three principal values being different are a general parallelepiped and an ellipsoid with unequal semiaxes. However, if the inclusions also have cubic symmetry (which in addition to reflections includes rotations about each axis by the angle $\pi/4$), then Σ is reduced to a scalar and the effective medium is isotropic.

1. General direction of propagation

Let the tensor Σ be diagonal in the rectangular frame xyz. We then use the expression (6) for $K(\mathbf{q})$, evaluate the determinant in Eq. (23), and obtain the following equation:

$$\frac{\epsilon_b^2 \prod_\alpha [1 + 2\rho\chi(1 + \Sigma_{\alpha\alpha})]}{(q^2 - k_b^2)^2} \mathscr{D}_c(k,\mathbf{q}) = 0, \qquad (24)$$

where

$$\mathscr{D}_c(k,\mathbf{q}) = k^4 - \mathscr{A}_c(\mathbf{q})k^2 + \mathscr{B}_c(\mathbf{q}) \qquad (25)$$

and

$$\mathscr{A}_c(\mathbf{q}) = q_x^2 \left(\frac{1}{\eta_y} + \frac{1}{\eta_z} \right) + q_y^2 \left(\frac{1}{\eta_x} + \frac{1}{\eta_z} \right) + q_z^2 \left(\frac{1}{\eta_x} + \frac{1}{\eta_y} \right), \qquad (26a)$$

$$\mathscr{B}_c(\mathbf{q}) = \frac{q_x^4}{\eta_y\eta_z} + \frac{q_y^4}{\eta_x\eta_z} + \frac{q_z^4}{\eta_x\eta_y} + \frac{q_x^2 q_y^2}{\eta_z}\left(\frac{1}{\eta_x} + \frac{1}{\eta_y} \right) + \frac{q_x^2 q_z^2}{\eta_y}\left(\frac{1}{\eta_x} + \frac{1}{\eta_z} \right) + \frac{q_y^2 q_z^2}{\eta_x}\left(\frac{1}{\eta_y} + \frac{1}{\eta_z} \right). \qquad (26b)$$

The quantities η_α are given by

$$\eta_\alpha = \epsilon_b \frac{1 + 2\rho\chi(1 + \Sigma_{\alpha\alpha})}{1 - \rho\chi(1 + \Sigma_{\alpha\alpha})}, \quad \alpha = x,y,z, \qquad (27)$$

and the subscript in $\mathscr{D}_c$, $\mathscr{A}_c$, and $\mathscr{B}_c$ has been used to emphasize that these expressions are applicable to composite media and have been obtained by evaluating the left-hand side of Eq. (23).

The set of dispersion relations (24)–(26) should be compared to the analogous set of equations in a homogeneous medium characterized by the effective tensors $\bar{\epsilon}$ and $\bar{\mu}$. Generally, the dispersion relation in such media reads

$$\det[(\mathbf{q} \times \bar{\mu}^{-1}\mathbf{q}\times) + k^2\bar{\epsilon}] = 0 \qquad (28a)$$

if $\bar{\mu}^{-1}$ exists or

$$\det[(\mathbf{q} \times \bar{\epsilon}^{-1}\mathbf{q}\times) + k^2\bar{\mu}] = 0 \qquad (28b)$$

if $\bar{\epsilon}^{-1}$ exists. If both $\bar{\mu}$ and $\bar{\epsilon}$ are invertible, the two equations (28a) and (28b) are identical.

For homogenization theory to be applicable, the effective medium must have the same symmetry as the composite. It is therefore evident that the principal axes of Σ should coincide with the optical axes of the effective medium. Denote the principal values of $\bar{\epsilon}$ and $\bar{\mu}$ by $\bar{\epsilon}_{\alpha\alpha}$ and $\bar{\mu}_{\alpha\alpha}$. Let us further assume that $\bar{\mu}$ is invertible. In this case Eq. (28a) takes the

following form:

$$k^2 \bar{\epsilon}_{xx}\bar{\epsilon}_{yy}\bar{\epsilon}_{zz}\mathscr{D}_h(k,\mathbf{q}) = 0, \qquad (29)$$

where

$$\mathscr{D}_h(k,\mathbf{q}) = k^4 - \mathscr{A}_h(\mathbf{q})k^2 + \mathscr{B}_h(\mathbf{q}) \qquad (30)$$

and

$$\mathscr{A}_h(\mathbf{q}) = q_x^2 \left(\frac{1}{\bar{\epsilon}_{yy}\bar{\mu}_{zz}} + \frac{1}{\bar{\epsilon}_{zz}\bar{\mu}_{yy}} \right) + q_y^2 \left(\frac{1}{\bar{\epsilon}_{xx}\bar{\mu}_{zz}} + \frac{1}{\bar{\epsilon}_{zz}\bar{\mu}_{xx}} \right)$$
$$+ q_z^2 \left(\frac{1}{\bar{\epsilon}_{xx}\bar{\mu}_{yy}} + \frac{1}{\bar{\epsilon}_{yy}\bar{\mu}_{xx}} \right), \qquad (31a)$$

$$\mathscr{B}_h(\mathbf{q}) = \frac{q_x^4}{\bar{\epsilon}_{yy}\bar{\epsilon}_{zz}\bar{\mu}_{yy}\bar{\mu}_{zz}} + \frac{q_y^4}{\bar{\epsilon}_{xx}\bar{\epsilon}_{zz}\bar{\mu}_{xx}\bar{\mu}_{zz}} + \frac{q_z^4}{\bar{\epsilon}_{xx}\bar{\epsilon}_{yy}\bar{\mu}_{xx}\bar{\mu}_{yy}}$$
$$+ \frac{q_x^2 q_y^2}{\bar{\epsilon}_{zz}\bar{\mu}_{zz}}\left(\frac{1}{\bar{\epsilon}_{xx}\bar{\mu}_{yy}} + \frac{1}{\bar{\epsilon}_{yy}\bar{\mu}_{xx}} \right)$$
$$+ \frac{q_x^2 q_z^2}{\bar{\epsilon}_{yy}\bar{\mu}_{yy}}\left(\frac{1}{\bar{\epsilon}_{xx}\bar{\mu}_{zz}} + \frac{1}{\bar{\epsilon}_{zz}\bar{\mu}_{xx}} \right)$$
$$+ \frac{q_y^2 q_z^2}{\bar{\epsilon}_{xx}\bar{\mu}_{xx}}\left(\frac{1}{\bar{\epsilon}_{yy}\bar{\mu}_{zz}} + \frac{1}{\bar{\epsilon}_{zz}\bar{\mu}_{yy}} \right). \qquad (31b)$$

Here the subscript in $\mathscr{D}_h$, $\mathscr{A}_h$, and $\mathscr{B}_h$ has been used to emphasize that these expressions are applicable to homogeneous media. In the case $\bar{\mu}_{xx} = \bar{\mu}_{yy} = \bar{\mu}_{zz} = 1$, Eq. (29) reduces to the well-known Fresnel equation.

The prefactors in Eqs. (24) and (29) are almost always nonzero, except in the case of nondissipative plasmas, which can support longitudinal waves. This case will be considered by us separately. Assuming that the prefactors are nonzero, the dispersion relations are $\mathscr{D}_c(k,\mathbf{q}) = 0$ for the composite medium and $\mathscr{D}_h(k,\mathbf{q}) = 0$ for the homogeneous medium. We can introduce EMPs for the composite by observing that these two dispersion relations become identical if we set

$$\bar{\epsilon}_{\alpha\alpha} = \xi\eta_\alpha, \quad \bar{\mu}_{\alpha\alpha} = \frac{1}{\xi}, \qquad (32)$$

where $\xi \neq 0$ is an arbitrary complex number. As already mentioned, the nonuniqueness in the above definition of the EMPs cannot be removed by considering the dispersion relations alone.

Several remarks regarding the dispersion relations obtained above should be made. First, in the general case the functions $\mathscr{D}_c(k,\mathbf{q})$ and $\mathscr{D}_h(k,\mathbf{q})$ cannot be factorized into products of two quadratic forms in the variables k, q_x, q_y, and q_z. However, such a factorization becomes possible for special directions of propagation when one or more of the Cartesian components of $\mathbf{q}$ are zero. Examples will be given below.

Second, the condition (32), which guarantees that $\mathscr{D}_c(k,\mathbf{q}) = \mathscr{D}_h(k,\mathbf{q})$, requires that the effective permeability $\bar{\mu}$ be a scalar. Any deviation of $\bar{\mu}$ from a scalar will result in different laws of dispersion in the composite and in the effective medium with no hope of obtaining the same measurables from these two models. This requirement that $\bar{\mu}$ be a scalar even in a strongly anisotropic composite is difficult to justify on physical grounds unless of course $\bar{\mu} = 1$.

Third, the dispersion relations (23) (for a composite medium) and Eqs. (28a) and (28b) (for a homogeneous medium) appear to have very different mathematical structure. The fact that they reduce to the same equation under the simple condition (32) is quite remarkable. Thus we have shown that if orthogonal optical axes of the composite medium can be defined, its dispersion relation $\mathbf{q}(k)$ and its isofrequency surfaces [defined as the sets containing all $\mathbf{q}$ such that $\mathscr{D}_c(k,\mathbf{q}) = 0$ for each $k = \omega/c$] are equivalent to those obtained in a homogeneous medium with EMPs $\bar{\epsilon}$ and $\bar{\mu}$ given by Eq. (32), where the quantities η_α are defined in Eq. (27).

Since it will be proved below that the correct choice of the parameter ξ in Eq. (32) is $\xi = 1$, we now state the main homogenization result of this paper pertaining to the principal values of the EMPs:

$$\bar{\epsilon}_{\alpha\alpha} = \epsilon_b \frac{1 + 2\rho\chi(1 + \Sigma_{\alpha\alpha})}{1 - \rho\chi(1 + \Sigma_{\alpha\alpha})}, \quad \bar{\mu}_{\alpha\alpha} = 1. \quad (33)$$

It can be seen that the Maxwell-Garnett mixing formula is obtained from Eq. (33) by setting $\Sigma = 0$. Electromagnetic interactions of inclusions in the composite result in a nonzero value of Σ and correspondingly in the deviation of the EMPs from the predications of Maxwell-Garnett theory.

2. Propagation along crystallographic axes

Consider a plane wave propagating along the z axis so that $q_x = q_y = 0$. In this case

$$\begin{aligned}\mathscr{D}_c(k,\mathbf{q}) &= k^4 - k^2 q_z^2 \left(\frac{1}{\eta_x} + \frac{1}{\eta_y}\right) + \frac{q_z^4}{\eta_x\eta_y} \\ &= \left(k^2 - \frac{q_z^2}{\eta_x}\right)\left(k^2 - \frac{q_z^2}{\eta_y}\right).\end{aligned} \quad (34)$$

Thus $\mathscr{D}_c(k,\mathbf{q})$ is factorized into a product of two quadratic forms, giving rise to two branches of the dispersion relation: $q_z^2 = \eta_x k^2$ and $q_z^2 = \eta_y k^2$. Obviously these two branches correspond to x- and y-polarized modes. It can be seen that, in agreement with Eq. (32), the quantities η_α give the effective squared refractive index for the transverse modes of the composite.

In addition to the two transverse modes, a longitudinally polarized mode can also exist under certain conditions. A mode with an arbitrary wave number q_α, which propagates and is polarized along the same axis α, exists if and only if

$$1 + 2\rho\chi(1 + \Sigma_{\alpha\alpha}) = 0. \quad (35)$$

Under this condition, the equality (24) holds, even if $\mathscr{D}_c(k,\mathbf{q}) \neq 0$.

Let us consider briefly the physical conditions for the existence of the longitudinal waves. From the property (A4) (given in Appendix A) it follows that $\lim_{\rho\to 1} \Sigma_{\alpha\alpha} = 0$. Consequently, the longitudinal waves exist in the high-density limit if $1 + 2\chi = 0$, which is possible only if $\epsilon_a = 0$. This is the well-known condition for longitudinal waves in nondissipative plasma. The low-density limit cannot be considered so easily because $\Sigma_{\alpha\alpha}$ does not approach zero when $\rho \to 0$ (see Sec. III C) and can in fact diverge for certain values of χ. However, we can use the reciprocity substitutions $\rho \leftrightarrow 1 - \rho$ and $\epsilon_a \leftrightarrow \epsilon_b$ to see that in the low-density limit the

$$\mathbf{r}_n = h(n_x \hat{\mathbf{x}} + n_y \hat{\mathbf{y}} + n_z \hat{\mathbf{z}})$$
$$-\infty < n_x, n_y < \infty$$
$$1 \le n_z < \infty$$
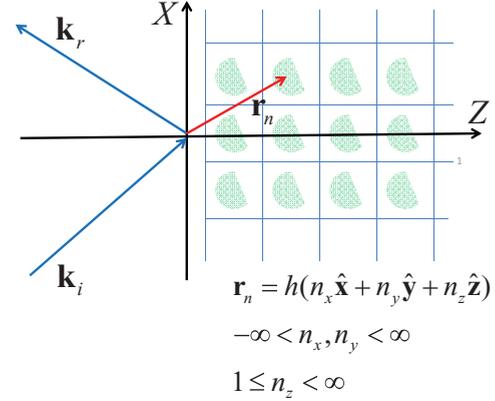
FIG. 2. (Color online) Sketch of the geometry considered: reflection and refraction at a half-space boundary.

condition for the existence of the longitudinal waves is $\epsilon_b = 0$. Quite analogously, longitudinal waves can be obtained by considering the dispersion relation (29) and setting one of the principal values $\bar{\epsilon}_{\alpha\alpha}$ equal to zero.

3. Propagation in a crystallographic plane

We now discuss the case when $\mathbf{q}$ lies in the xz plane. Problems of this type can arise when one considers reflection and refraction at the interface $z = 0$, where the xz plane is the plane of incidence, as shown in Fig. 2. Under the condition $q_y = 0$ we have

$$\begin{aligned}\mathscr{D}_c(k,\mathbf{q}) &= k^4 - k^2\left[q_x^2\left(\frac{1}{\eta_y} + \frac{1}{\eta_z}\right) + q_z^2\left(\frac{1}{\eta_x} + \frac{1}{\eta_y}\right)\right] \\ &\quad + \frac{q_x^4}{\eta_y\eta_z} + \frac{q_z^4}{\eta_x\eta_y} + \frac{q_x^2 q_z^2}{\eta_y}\left(\frac{1}{\eta_x} + \frac{1}{\eta_z}\right) \\ &= \left[k^2 - \left(\frac{q_x^2}{\eta_y} + \frac{q_z^2}{\eta_y}\right)\right]\left[k^2 - \left(\frac{q_x^2}{\eta_z} + \frac{q_z^2}{\eta_x}\right)\right]. \quad (36)\end{aligned}$$

Thus $\mathscr{D}_c(k,\mathbf{q})$ is factorized into a product of two quadratic forms, which correspond to the s- and p-polarized modes.

By equating the first factor in Eq. (36) to zero, we obtain the dispersion relation for the s-polarized wave:

$$\frac{q_z^2}{\eta_y} + \frac{q_x^2}{\eta_y} = k^2. \quad (37)$$

The vector $\mathbf{F}_0$ of the s-polarized wave is aligned with the y axis and is therefore perpendicular to the plane of incidence.

By equating the second factor in Eq. (36) to zero, we obtain the dispersion relation for the p-polarized wave:

$$\frac{q_z^2}{\eta_x} + \frac{q_x^2}{\eta_z} = k^2. \quad (38)$$

We can now find the vector $\mathbf{F}_0$ for the p-polarized wave by considering the nontrivial solutions to Eq. (22). It can be easily seen that $\mathbf{F}_0$ lies in this case in the plane of incidence (its projection onto the y axis is zero) and the x and z components of $\mathbf{F}_0$, F_{0x}, and F_{0z} satisfy the following relation (details of the derivation are given in Appendix B):

$$\frac{F_{0x}}{F_{0z}} = -\frac{1 + 2\rho\chi(1 + \Sigma_{zz})}{1 + 2\rho\chi(1 + \Sigma_{xx})}\frac{q_z}{q_x}. \quad (39)$$

Equation (39) will be used below in Sec. IV to compute the half-space reflection coefficient for the p-polarized incident wave.

C. Low-density and low-contrast limits

Iteration of Eq. (19b) results in the following expansion for the self-energy:

$$\Sigma = \rho\chi \sum_{\mathbf{g}\neq 0} M(-\mathbf{g})Q(\mathbf{g})M(\mathbf{g}) + (\rho\chi)^2$$
$$\times \sum_{\mathbf{g},\mathbf{g}'\neq 0} M(-\mathbf{g})Q(\mathbf{g})M(\mathbf{g}-\mathbf{g}')Q(\mathbf{g}')M(\mathbf{g}') + \cdots . \quad (40)$$

It is important to note that this expansion should be used with caution. Indeed, if χ is of the order of unity or larger, the series in Eq. (40) does not converge, even for arbitrarily small values of the density ρ. This result may seem unexpected, but it is easily understood by observing that the functions $M(\mathbf{g})$ depend on ρ and obey the sum rules (A2).

In Sec. VI a more useful (and always convergent) expansion of Σ will be derived. Here we note that the functions $M(\mathbf{g})$ are independent of χ. Therefore, Eq. (40) is the formal expansion of Σ into the powers of χ. Thus, in the low-contrast limit $\chi \to 0$ we have $\Sigma \to \rho\chi\sigma_1$, where $\sigma_1 = \sum_{\mathbf{g}\neq 0} M(-\mathbf{g})Q(\mathbf{g})M(\mathbf{g})$. In the case of three-dimensional inclusions with cubic symmetry, σ_1 is identically zero. Then the first nonvanishing term in the low-contrast expansion of Σ is given by $(\rho\chi)^2\sigma_2$, where σ_2 grows naturally out of the second term on the right-hand side of Eq. (40).

D. Two-dimensional lattices

Consider a medium in which $\epsilon = \epsilon(x,y)$ is independent of z. As above, we assume that $\epsilon(x,y)$ is periodic on a square lattice with lattice step h. The homogenization theory for this medium can be obtained either by considering a three-dimensional lattice with unequal steps h_x, h_y, and h_z and taking the limit $h_z \to 0$ or by following the derivations of Sec. III A, taking account of the modified geometry. The results obtained are very similar to those in the 3D case, with some obvious modifications. Specifically, we arrive at Eqs. (19a) and (19b) in which, however, we must take $\mathbf{g} = (2\pi/h)(\hat{\mathbf{x}}n_x + \hat{\mathbf{y}}n_y)$. Additionally, in the integrals (18), Ω must be understood as a two-dimensional region (the intersection of an inclusion with the xy plane) and V as the area of Ω and d^3R is replaced by d^2R. The definition of $Q(\mathbf{g})$ (20) remains unchanged, but $Q(\mathbf{g})$ is now a 2×2 tensor.

Consider a wave propagating in the xy plane and polarized along the z axis. In this case $\mathbf{F_g} = \hat{\mathbf{z}}F_\mathbf{g}$, where $F_\mathbf{g}$ is a scalar and Σ can be found analytically in general. Indeed, we have in this case $Q(\mathbf{g})\mathbf{F}_{\mathbf{g}'} = \mathbf{F}_{\mathbf{g}'}$, $Q(\mathbf{g})\mathbf{F}_0 = \mathbf{F}_0$ and Eq. (19b) becomes

$$F_\mathbf{g} = \rho\chi\left[M(\mathbf{g})F_0 + \sum_{\mathbf{g}'\neq 0} M(\mathbf{g}-\mathbf{g}')F_{\mathbf{g}'} \right], \quad \mathbf{g} \neq 0. \quad (41)$$

The solution to this equation is

$$F_\mathbf{g} = \frac{\rho\chi}{1-(1-\rho)\chi}M(\mathbf{g})F_0, \quad (42)$$

where some of the properties in Eq. (A2) have been used (keeping in mind that the term $\mathbf{g} = 0$ must be excluded from the summation). We then have

$$\Sigma_{zz} = \frac{\rho\chi}{1-(1-\rho)\chi} \sum_{\mathbf{g}\neq 0} M(-\mathbf{g})M(\mathbf{g}) = \frac{(1-\rho)\chi}{1-(1-\rho)\chi}. \quad (43)$$

It can be seen from this equation that Σ_{zz} does not approach zero when $\rho \to 0$, as discussed in Sec. III C. Upon substitution of Eq. (43) into Eq. (33) we find that

$$\bar{\epsilon}_{zz} = (1-\rho)\epsilon_b + \rho\epsilon_a = \langle\epsilon\rangle. \quad (44)$$

Thus the effective permittivity for z polarization is given by the arithmetic average of $\epsilon(x,y)$. This is in agreement with Krokhin *et al.* [14,15].

E. Concept of the smooth field

The result (44) for a z-polarized wave could have been anticipated. To understand better why the effective permittivity in this case is given by an arithmetic average, it is instructive to consider the concept of the *smooth field*. The smooth field $\mathbf{S}(\mathbf{r})$ changes slowly on the characteristic scale defined by the heterogeneities in the medium. As a result, one can factorize spatial averages of $\mathbf{S}(\mathbf{r})$ multiplied by any rapidly varying function. For example, we can write $\langle \mathbf{S}\epsilon \rangle = \langle \mathbf{S} \rangle \langle \epsilon \rangle$, etc.

Let us recall some well-known results for 1D periodically layered media [43]. The effective permittivity of such media is $\bar{\epsilon}_\parallel = \langle\epsilon\rangle$ for waves polarized parallel to the layers and $\bar{\epsilon}_\perp = \langle\epsilon^{-1}\rangle^{-1}$ for waves polarized perpendicularly to the layers. These two results can be obtained quite expeditiously by applying the concept of the smooth field. In the case of tangential polarization, the electric field $\mathbf{E}$ is smooth. This follows from the boundary condition that requires that the tangential components of the electric field be continuous at all interfaces. Consequently, we can write

$$\langle \mathbf{D} \rangle = \langle \epsilon \mathbf{E} \rangle = \langle \epsilon \rangle \langle \mathbf{E} \rangle, \quad (45)$$

from which it follows that $\bar{\epsilon}_\parallel = \langle\epsilon\rangle$. For perpendicular polarization, the field $\mathbf{D}$ is smooth. We then write

$$\langle \mathbf{E} \rangle = \langle \epsilon^{-1}\mathbf{D} \rangle = \langle \epsilon^{-1} \rangle \langle \mathbf{D} \rangle \quad (46)$$

and $\bar{\epsilon}_\perp = \langle\epsilon^{-1}\rangle^{-1}$.

Similar considerations can be applied to the 2D problem of Sec. III D. For waves polarized along the z axis, the field $\mathbf{E}$ is smooth, which results in $\bar{\epsilon}_{zz} = \langle\epsilon\rangle$, in agreement with Eq. (44).

One can also consider a more general smooth field of the form $\mathbf{S} = p\mathbf{E} + (1-p)\mathbf{D} = [p+(1-p)\epsilon]\mathbf{E}$, where p is a mixing parameter. Here we consider the 3D case and assume that $\mathbf{S}$ is smooth for any polarization state. Application of the smooth field principle results in the following equalities:

$$\langle \mathbf{E} \rangle = \langle \mathbf{S} \rangle \langle 1/[p+(1-p)\epsilon] \rangle, \quad (47a)$$
$$\langle \mathbf{D} \rangle = \langle \mathbf{S} \rangle \langle \epsilon/[p+(1-p)\epsilon] \rangle, \quad (47b)$$

from which we find the effective permittivity to be

$$\bar{\epsilon}_{\alpha\beta} = \delta_{\alpha\beta} \frac{\langle \epsilon/[\epsilon+p/(1-p)] \rangle}{\langle 1/[\epsilon+p/(1-p)] \rangle}. \quad (48)$$

Equation (48) is in fact the Maxwell-Garnett formula. Although this form is rarely used, the Maxwell-Garnett effective

permittivity can be written as

$$\bar{\epsilon}_{\text{MG}} = \frac{\langle \epsilon/(\epsilon + 2\epsilon_b) \rangle}{\langle 1/(\epsilon + 2\epsilon_b) \rangle}. \tag{49}$$

We see that Eqs. (48) and (49) coincide if $p = 2\epsilon_b/(1 + 2\epsilon_b)$.

Thus the Maxwell-Garnett EMT assumes that the field $\mathbf{S} = [(\epsilon + 2\epsilon_b)/(1 + 2\epsilon_b)]\mathbf{E}$ is smooth. Since the mixing parameter p depends on the permittivity of the host medium, Eq. (49) is not invariant with respect to the substitutions $\epsilon_a \leftrightarrow \epsilon_b$ and $\rho \leftrightarrow 1 - \rho$. The homogenization formula (33) derived in this paper, however, is fully symmetric. Note that Bruggeman's EMT is also symmetric but cannot be easily written in terms of averages. Therefore, it is not clear which form of the smooth field Bruggeman's approximation assumes. In general, the smooth field does not need to be a linear functional of $\mathbf{E}$ and $\mathbf{D}$.

IV. REFLECTION AND REFRACTION AT A HALF-SPACE BOUNDARY

An infinite lattice is a mathematical abstraction. All experimental media are bounded and the physical effects that occur at the boundary are often important. For instance, as mentioned above, it is not possible to determine simultaneously and uniquely the effective permittivity and permeability of a medium from the bulk dispersion relation alone.

The problem of reflection and refraction of a wave at a flat interface is considered in this section. The goals are threefold. First, we will derive the limit in which the correct expression for the Fresnel reflection coefficient is obtained. This will turn out to be the same limit used in Sec. III A. Second, we will show that the correct expression for the reflection coefficients results only if we set $\xi = 1$ in Eq. (32), from which it follows that $\bar{\mu} = 1$. Third, we will provide additional mathematical justification for the results of Sec. III A. Indeed, the derivations of that section contain one dubious step. Namely, the applicability of the Poisson summation formula (14) can be questioned because the variable $\mathbf{q}$ is complex. Strictly speaking, the series on the left-hand side of Eq. (14) diverges for an infinite lattice. The problem can be fixed, in principle, by considering real-valued $\mathbf{q}$'s and then analytically continuing the summation result to the whole complex plane. In this section no such complication will arise since all series in question are convergent.

A. General setup

The geometry considered in this section is sketched in Fig. 2. The medium occupies the right half space and the left half space has the background permittivity ϵ_b. It would be more appropriate to consider the case when the left half space is a vacuum and the right half space is a two-component mixture, so that there are three different components in the problem. This however requires the use of the half-space Green's tensor [44]: a step that is not conceptually difficult, yet mathematically involved. Here we restrict consideration to only two components. This includes the cases when the host medium is a vacuum and also when the incident beam is first refracted from the vacuum into a homogeneous medium of permittivity $\epsilon_b \neq 1$ (at a planar interface that is located at

$z = z_1 \ll -h$ and is not considered explicitly) and then into a heterogeneous medium that is a mixture of a- and b-type components.

Physically, the z coordinate of the effective-medium boundary can be stated only approximately, within an interval of width $\sim h$. It will prove mathematically convenient to place the boundary on the plane $z = 0$ and the centers of the leftmost cells on the plane $z = h$, as shown in Fig. 2. In the EMT developed below, the half space $z > 0$ is assumed to be filled with an effective medium.

A wave cannot propagate in a semi-infinite medium without an external source. Therefore, we must solve the integral equation (2) with a nonzero incident field $\mathbf{E}_i$, which we will take to be a plane wave. We will also find that under appropriate conditions, a uniquely defined reflected plane wave $\mathbf{E}_r$ exists in the region $z < 0$. The incident and the reflected waves are given by

$$\mathbf{E}_i(\mathbf{r}) = \mathbf{A}_i \exp(\mathbf{k}_i \cdot \mathbf{r}), \quad -\infty < z < \infty, \tag{50a}$$

$$\mathbf{E}_r(\mathbf{r}) = \mathbf{A}_r \exp(\mathbf{k}_r \cdot \mathbf{r}), \quad -\infty < z < 0. \tag{50b}$$

Note that the incident wave is defined in the whole space but Eq. (2) is defined only for $\mathbf{r} \in \Omega_{\text{tot}}$. The wave numbers of the incident and the reflected waves can be written as

$$\mathbf{k}_i = \mathbf{k}_\perp + \hat{\mathbf{z}}k_{iz}, \quad \mathbf{k}_r = \mathbf{k}_\perp - \hat{\mathbf{z}}k_{iz}. \tag{51}$$

Henceforth the subscript $\perp$ will be used to denote projections of vectors onto the xy plane. Note that $\mathbf{k}_\perp \cdot \hat{\mathbf{z}} = 0$ and

$$k_i^2 = k_r^2 = k_\perp^2 + k_{iz}^2 = k_b^2 = k^2\epsilon_b = \left(\frac{\omega}{c}\right)^2 \epsilon_b. \tag{52}$$

It is important to note that the vector $\mathbf{k}_\perp$ is purely real. A complex-valued $\mathbf{k}_\perp$ would imply a wave that is evanescent in a direction parallel to the interface. This would necessitate the presence of additional interfaces; such a possibility is not considered here. The vector $\mathbf{k}_\perp$ is real valued even if the host medium is absorbing. Indeed, we should keep in mind that the incident wave enters the host medium from a vacuum and that the tangential component of the wave vector is conserved at any planar interface, even if one of the media is absorbing. However, the z projection of $\mathbf{k}_i$ does not need to be real. In a transparent host ($\epsilon_b > 0$) the incident wave is evanescent and k_{iz} is purely imaginary if $k_\perp > k_b$; in an absorbing host k_{iz} is generally complex.

Note that the reflected wave (50b) does not enter Eq. (2) because it is identically zero in Ω_{tot}. The reflected wave is computed *a posteriori* once the polarization field $\mathbf{P}$ is found. Then the amplitudes $\mathbf{A}_r$ and $\mathbf{A}_i$ can be used to determine the reflection coefficient.

To solve Eq. (2) in the presence of the incident field we decompose $\mathbf{P}$ as

$$\mathbf{P} = \mathbf{P}_B + \mathbf{P}_S, \tag{53}$$

where $\mathbf{P}_B$ is the Bloch wave of the form (8) and $\mathbf{P}_S$ is an additional wave that originates due to the presence of the surface. We seek the condition under which

$$\mathbf{E}_{\text{EO}}(\mathbf{r}) \equiv \int_{\Omega_{\text{tot}}} G(\mathbf{r}, \mathbf{r}')\mathbf{P}_B(\mathbf{r}')d^3r' = \mathbf{E}_B(\mathbf{r}) + \mathbf{E}_{\text{ext}}(\mathbf{r}) + \mathbf{E}_S(\mathbf{r}), \tag{54}$$

where in Ω_{tot}

$$\mathbf{E}_B(\mathbf{r}) = \frac{4\pi}{3\chi} \mathbf{P}_B(\mathbf{r}), \tag{55a}$$

$$\mathbf{E}_{\text{ext}}(\mathbf{r}) = -\mathbf{E}_i(\mathbf{r}), \tag{55b}$$

If Eqs. (53)–(55) hold, then Eq. (2) becomes

$$\mathbf{P}_S(\mathbf{r}) = \frac{3\chi}{4\pi} \left[\mathbf{E}_S(\mathbf{r}) + \int_{\Omega_{\text{tot}}} G(\mathbf{r},\mathbf{r}') \mathbf{P}_S(\mathbf{r}') d^3 r' \right], \quad \mathbf{r} \in \Omega_{\text{tot}}. \tag{56}$$

Note that Eq. (56) contains only quantities that are associated with the surface wave.

Equation (54) is the mathematical formulation of the Ewald-Oseen extinction theorem and we will refer to $\mathbf{E}_{\text{EO}}$ as to the Ewald-Oseen field. We will see that one can determine the reflection coefficient from the conditions (55). We will also see that the surface wave is exponentially localized near the interface and does not contribute to either reflection or transmission coefficients if

$$(\mathbf{k}_\perp + \mathbf{g}_\perp)^2 > k_b^2 \quad \forall \mathbf{g}_\perp \neq 0. \tag{57}$$

Inequality (57) is weaker than what is required for homogenization. It is merely the condition that there is no Bragg diffraction in the medium; if Eq. (57) is violated, the conventional reflection and transmission coefficients cannot be defined. If, however, Eq. (57) holds, we do not need to solve Eq. (56) explicitly; it suffices to know that the surface wave does not contribute to any measurement performed sufficiently far from the interface.

B. Evaluation of the Ewald-Oseen field

To compute the Ewald-Oseen field, we proceed along the lines of Sec. III A to arrive at the following expression:

$$\mathbf{E}_{\text{EO}}(\mathbf{r}) = \frac{4\pi}{3} \int \frac{d^3 p}{(2\pi)^3} K(\mathbf{p}) \int_\Omega d^3 R\, \mathbf{F}(\mathbf{R})$$
$$\times \exp[i\mathbf{p} \cdot (\mathbf{r} - \mathbf{R})] \sum_m \exp[i(\mathbf{q} - \mathbf{p}) \cdot \mathbf{r}_m]. \tag{58}$$

So far, no restrictions on $\mathbf{r}$ have been placed. In particular, $\mathbf{r}$ can be either in the right or left half space. However, when we later substitute the result of integration into Eqs. (55), $\mathbf{r}$ will be restricted to Ω_{tot}.

The sum over m in Eq. (58) can be evaluated as follows. First, we expand the summation as

$$\sum_m \exp[i(\mathbf{q} - \mathbf{p}) \cdot \mathbf{r}_m]$$
$$= \sum_{m_x, m_y = -\infty}^{\infty} \exp[i(q_x - p_x)hm_x + i(q_y - p_y)hm_y]$$
$$\times \sum_{m_z = 1}^{\infty} \exp[i(q_z - p_z)hm_z]. \tag{59}$$

From symmetry considerations we know that $\mathbf{q}_\perp = \mathbf{k}_\perp$. This property is a manifestation of momentum conservation and will be confirmed below by considering the conditions (55). Since, as discussed above, $\mathbf{k}_\perp$ is purely real, q_x and q_y are also real. Therefore, we can compute the sums over m_x and m_y

using the Poisson sum formula. Further, the half-range sum over m_z converges absolutely because the transmitted wave decays into the medium and correspondingly $\text{Im} q_z > 0$. We therefore have

$$\sum_m \exp[i(\mathbf{q} - \mathbf{p}) \cdot \mathbf{r}_m] = \left(\frac{2\pi}{h}\right)^2 f(p_z) \sum_{\mathbf{g}_\perp} \delta(\mathbf{p}_\perp - \mathbf{q}_\perp - \mathbf{g}_\perp), \tag{60}$$

where

$$f(p_z) \equiv \sum_{m_z = 1}^{\infty} \exp[i(q_z - p_z)hm_z] = \frac{1}{\exp[i(p_z - q_z)h] - 1} \tag{61a}$$

$$= \frac{2\pi}{h} \sum_{g_z} \frac{(2\pi i)^{-1}}{p_z - q_z - g_z}. \tag{61b}$$

Here the well-known Laurant expansion of the function $1/[\exp(iz) - 1]$ has been used. The equality (61b) is an important observation. It will allow us to evaluate the Ewald-Oseen field.

We now proceed by substituting Eq. (60) into Eq. (58), which yields

$$\mathbf{E}_{\text{EO}}(\mathbf{r}) = \frac{4\pi}{3h^2} \sum_{\mathbf{g}_\perp} \int_{-\infty}^{\infty} \frac{dp_z}{2\pi} f(p_z) K(\mathbf{q}_\perp + \mathbf{g}_\perp + \hat{\mathbf{z}} p_z)$$
$$\times \int_\Omega d^3 R\, \mathbf{F}(\mathbf{R}) \exp[i(\mathbf{q}_\perp + \mathbf{g}_\perp + \hat{\mathbf{z}} p_z) \cdot (\mathbf{r} - \mathbf{R})]. \tag{62}$$

The integral over p_z can be computed by contour integration since all the poles and residues of the integrand are known. The positions of the poles in the complex p_z plane are shown in Fig. 3. The poles at $p_z = q_z + g_z$ are the singularities of the function $f(p_z)$. Since q_z has a positive imaginary part and all g_z's are real valued, these poles lie in the upper half plane. The remaining poles are the singularities of $K(\mathbf{q}_\perp + \mathbf{g}_\perp + \hat{\mathbf{z}} p_z)$, which is viewed here as a function of p_z. From the definition (6) we find that these singularities are located at $p_z = \pm \mathcal{P}_{\mathbf{g}_\perp}$, where

$$\mathcal{P}_{\mathbf{g}_\perp} = \sqrt{k_b^2 - (\mathbf{q}_\perp + \mathbf{g}_\perp)^2}. \tag{63}$$
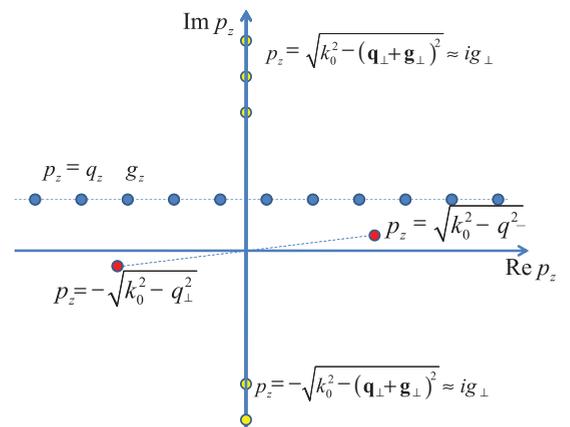
FIG. 3. (Color online) Poles of the integrand of Eq. (62) in the complex p_z plane.

These poles can be considered separately for $\mathbf{g}_\perp = 0$ and $\mathbf{g}_\perp \neq 0$. The two poles corresponding to $\mathbf{g}_\perp = 0$ are $p_z = \pm \mathcal{P}_0 = \pm \sqrt{k_b^2 - q_\perp^2}$. The poles with $\mathbf{g}_\perp \neq 0$ have large (either positive or negative) imaginary parts if $hk_b, hq_\perp \ll 1$, in which case they can be written approximately as $\mathcal{P}_{\mathbf{g}_\perp} \approx ig_\perp$.

Note that in the case of infinite lattices the singularities of $K(\mathbf{p})$ do not contribute to Fourier integrals of the type (12) because the corresponding residues are identically zero [these singularities fall in between the peaks of the δ-function fence given by the right-hand side of Eq. (14)]. We will compute the contributions of the different families of poles to the integral (62) separately. If the vector of position $\mathbf{r}$ is inside one of the inclusions, the integration contour must be closed in the upper half of the complex p_z plane. Correspondingly, only the poles with positive imaginary parts contribute to the integral (62) in this case. The Ewald-Oseen field can also be computed in the left half space. If the point of observation $\mathbf{r}$ is further away from the interface than $h/2$ so that the inequality $\hat{\mathbf{z}} \cdot \mathbf{r} < -h/2$ holds, the integration contour must be closed in the lower half of the complex p_z plane. In what follows it will be shown that the poles at $p_z = q_z + g_z$ yield the Bloch-wave field $\mathbf{E}_B(\mathbf{r})$, the pole at $p_z = \mathcal{P}_0$ yields the extinction field $\mathbf{E}_{\text{ext}}(\mathbf{r})$, the pole at $p_z = -\mathcal{P}_0$ yields the reflected wave, and finally the poles $p_z \approx \pm\mathcal{P}_{\mathbf{g}_\perp}$ with $\mathbf{g}_\perp \neq 0$ yield the fast-decaying surface wave.

1. Bloch wave

We start by computing the Bloch-wave contribution to the Ewald-Oseen field $\mathbf{E}_B(\mathbf{r})$. We place the point of observation in Ω_{tot}, use the expression (61b) for $f(p_z)$, and evaluate the contributions of the poles $p_z = q_z + g_z$ to the integral (62). This results in the following expression:

$$\mathbf{E}_B(\mathbf{r}) = \frac{4\pi}{3h^3} \sum_{\mathbf{g}} \exp[i(\mathbf{q}+\mathbf{g})\cdot\mathbf{r}]K(\mathbf{q}+\mathbf{g})$$
$$\times \int_\Omega \mathbf{F}(\mathbf{R})\exp[-i(\mathbf{q}+\mathbf{g})\cdot\mathbf{R}]d^3R, \quad \mathbf{r} \in \Omega_{\text{tot}}.$$
(64)

Here we have used the equalities $\mathbf{g}_\perp + \hat{\mathbf{z}}g_z = \mathbf{g}$ and $\sum_{\mathbf{g}_\perp}\sum_{g_z} = \sum_{\mathbf{g}}$. Now, if $\mathbf{F}(\mathbf{R})$ is expanded according to Eq. (16) and if the expansion coefficients $\mathbf{F}_{\mathbf{g}}$ satisfy Eq. (17), then the field given by Eq. (64) satisfies $\mathbf{E}_B(\mathbf{r}) = (4\pi/3\chi)\mathbf{P}_B(\mathbf{r})$ for $\mathbf{r} \in \Omega_{\text{tot}}$, where $\mathbf{P}_B$ is of the form (8). Thus Eq. (55a) is satisfied if the Bloch wave of the polarization $\mathbf{P}_B$ is the same as one would find by solving the eigenproblem (17) for an infinite medium. This justifies the use of the Poisson summation formula in Sec. III A.

Equation (17) applies to general photonic crystals that are not necessarily describable by EMPs. As discussed in Sec. III A, homogenization is obtained by taking the limit $h \to 0$. This limit must be computed separately for the equations with $\mathbf{g} = 0$ and $\mathbf{g} \neq 0$, which results in Eq. (19). This system of equations defines an eigenproblem for the Bloch wave vector $\mathbf{q}$, while the polarization vector $\mathbf{F}_0$ is obtained as an eigenvector of Eq. (22). The higher-order expansion coefficients $\mathbf{F}_{\mathbf{g}}$ are uniquely determined by $\mathbf{F}_0$, but $\mathbf{F}_0$ itself is defined by Eq. (19) only up to a multiplicative factor. Next we will show that this factor is fixed by the condition (55b).

2. Extinction wave

We now compute the contribution of the pole located at $p_z = \mathcal{P}_0$. The function $f(p_z)$ is analytic in the vicinity of $\mathcal{P}_0$; therefore, we can use the expression (61a) for $f(p_z)$. Since Eq. (55b) should hold only for $\mathbf{r} \in \Omega_{\text{tot}}$, we close the integration contour in the upper half plane. A straightforward calculation yields

$$\mathbf{E}_{\text{ext}}(\mathbf{r}) = \frac{4\pi i}{h^2} \frac{\exp[i(\mathbf{q}_\perp + \hat{\mathbf{z}}\mathcal{P}_0)\cdot\mathbf{r}]}{\exp[i(\mathcal{P}_0 - q_z)h] - 1}$$
$$\times \frac{k_b^2 - (\mathbf{q}_\perp + \hat{\mathbf{z}}\mathcal{P}_0) \otimes (\mathbf{q}_\perp + \hat{\mathbf{z}}\mathcal{P}_0)}{2\mathcal{P}_0}$$
$$\times \int_\Omega d^3R\, \mathbf{F}(\mathbf{R})\exp[-i(\mathbf{q}_\perp + \hat{\mathbf{z}}\mathcal{P}_0)\cdot\mathbf{R}], \quad \mathbf{r} \in \Omega_{\text{tot}}.$$
(65)

We seek the condition under which $\mathbf{E}_{\text{ext}}(\mathbf{r}) = -\mathbf{E}_i(\mathbf{r})$ for $\mathbf{r} \in \Omega_{\text{tot}}$, where $\mathbf{E}_i(\mathbf{r})$ is given by Eq. (50a). It immediately transpires that the above equality can hold only if $\mathbf{q}_\perp = \mathbf{k}_\perp$. The continuity of the tangential components of all wave vectors, including the incident wave vector $\mathbf{k}_i$, the reflected wave vector $\mathbf{k}_r$, and the Bloch-wave vector of the transmitted wave $\mathbf{q}$, follows from the discrete translational symmetry of the problem. We now find from Eq. (51) that $\mathcal{P}_0 = k_{iz}$ and $\mathbf{q}_\perp + \hat{\mathbf{z}}\mathcal{P}_0 = \mathbf{k}_i$. With the use of these equalities and the notation

$$\tilde{\mathbf{F}}(\mathbf{k}) = \int_\Omega \mathbf{F}(\mathbf{R})\exp(-i\mathbf{k}\cdot\mathbf{R})d^3R,$$
(66)

we can simplify Eq. (65) as

$$\mathbf{E}_{\text{ext}}(\mathbf{r}) = \frac{4\pi i}{h^2} \frac{\exp(i\mathbf{k}_i\cdot\mathbf{r})}{\exp[i(k_{iz}-q_z)h]-1}$$
$$\times \frac{k_b^2 - \mathbf{k}_i \otimes \mathbf{k}_i}{2k_{iz}}\tilde{\mathbf{F}}(\mathbf{k}_i), \quad \mathbf{r} \in \Omega_{\text{tot}}.$$
(67)

The extinction condition then takes the form

$$\mathbf{A}_i = -\frac{2\pi i}{h^2}\frac{k_b^2 - \mathbf{k}_i \otimes \mathbf{k}_i}{\exp[i(k_{iz}-q_z)h]-1}\frac{\tilde{\mathbf{F}}(\mathbf{k}_i)}{k_{iz}}.$$
(68)

So far, no approximations have been made. The homogenization limit is obtained by observing that

$$\lim_{h \to 0} \exp[i(\pm k_{iz}-q_z)h] = 1 + i(\pm k_{iz} - q_z)h,$$
(69a)

$$\lim_{h \to 0} \tilde{\mathbf{F}}(\mathbf{k}_i) = \lim_{h \to 0} \tilde{\mathbf{F}}(\mathbf{k}_r) = V(1+\Sigma)\mathbf{F}_0.$$
(69b)

Once the above limiting expressions are used, the extinction condition becomes of the form

$$\mathbf{A}_i = -2\pi\rho\frac{k_b^2 - \mathbf{k}_i \otimes \mathbf{k}_i}{k_{iz}(k_{iz}-q_z)}(1+\Sigma)\mathbf{F}_0.$$
(70)

This equation couples the amplitude of the incident field $\mathbf{A}_i$ and the amplitude of the Bloch polarization wave $\mathbf{F}_0$. The vector $\mathbf{F}_0$ must simultaneously satisfy the following two conditions: It must (i) be an eigenvector of the tensor in the square brackets in Eq. (22) and (ii) satisfy Eq. (70). These two conditions determine both the direction and the length of $\mathbf{F}_0$.

3. Reflected wave

Consider now the case when the point of observation $\mathbf{r}$ is in the left half space. As discussed above, we will place $\mathbf{r}$ at least $h/2$ away from the interface. This will allow us to close the integration contour in Eq. (62) in the lower half of the complex p_z plane. The reflected wave is obtained by computing the input of the pole $p_z = -\mathcal{P}_0$. We find that the electric field of the reflected wave is of the form (50b), where the amplitude $\mathbf{A}_r$ is given by

$$\mathbf{A}_r = \frac{2\pi i}{h^2} \frac{k_b^2 - \mathbf{k}_r \otimes \mathbf{k}_r}{\exp[-i(k_{iz} + q_z)h] - 1} \frac{\tilde{\mathbf{F}}(\mathbf{k}_r)}{k_{iz}}. \quad (71)$$

This expression contains no approximations. In the homogenization limit we use the limiting expressions (69) and obtain

$$\mathbf{A}_r = -2\pi\rho \frac{k_b^2 - \mathbf{k}_r \otimes \mathbf{k}_r}{k_{iz}(k_{iz} + q_z)}(1 + \Sigma)\mathbf{F}_0. \quad (72)$$

4. Surface wave

Finally, let us evaluate the contribution of the poles $p_z = \mathcal{P}_{\mathbf{g}_\perp}$ with $\mathbf{g}_\perp \neq 0$. For $\mathbf{r} \in \Omega_{\text{tot}}$ we have

$$\mathbf{E}_S(\mathbf{r}) = \frac{2\pi i}{h^2} \sum_{\mathbf{g}_\perp \neq 0} f(\mathcal{P}_{\mathbf{g}_\perp}) \exp(i\mathbf{k}_{\mathbf{g}_\perp} \cdot \mathbf{r})$$
$$\times \frac{k_b^2 - \mathbf{k}_{\mathbf{g}_\perp} \otimes \mathbf{k}_{\mathbf{g}_\perp}}{\mathcal{P}_{\mathbf{g}_\perp}} \tilde{\mathbf{F}}(\mathbf{k}_{\mathbf{g}_\perp}), \quad \mathbf{r} \in \Omega_{\text{tot}}, \quad (73)$$

where

$$\mathbf{k}_{\mathbf{g}_\perp} = \mathbf{q}_\perp + \mathbf{g}_\perp + \hat{\mathbf{z}}\mathcal{P}_{\mathbf{g}_\perp}. \quad (74)$$

If the condition (57) holds, the quantities $\mathcal{P}_{\mathbf{g}_\perp}$ have nonzero imaginary parts even if the host is transparent. Therefore, the surface wave decays exponentially away from the interface. In the homogenization limit the exponential decay is fast. Indeed, in the limit $h \to 0$ we have (for $\mathbf{g}_\perp \neq 0$) $\mathcal{P}_{\mathbf{g}_\perp} \to ig_\perp$, $\mathbf{k}_{\mathbf{g}_\perp} \to \mathbf{g}_\perp + i\hat{\mathbf{z}}g_\perp$, and $f(\mathcal{P}_{\mathbf{g}_\perp}) \to -1/g_\perp h$. With these limits taken into account, the surface wave takes the following form:

$$\mathbf{E}_S(\mathbf{r}) = -\frac{2\pi i}{h^3} \sum_{\mathbf{g}_\perp \neq 0} \frac{k_b^2 - (\mathbf{g}_\perp + i\hat{\mathbf{z}}g_\perp) \otimes (\mathbf{g}_\perp + i\hat{\mathbf{z}}g_\perp)}{g_\perp^2}$$
$$\times \exp[(i\mathbf{g}_\perp - \hat{\mathbf{z}}g_\perp) \cdot \mathbf{r}]\tilde{\mathbf{F}}(\mathbf{g}_\perp + \hat{\mathbf{z}}g_\perp), \quad \mathbf{r} \in \Omega_{\text{tot}}. \quad (75)$$

It can be seen that $\mathbf{E}_S$ decays exponentially on the scale of h. So does the wave of polarization $\mathbf{P}_S$, as both fields are related by the integral equation (56).

Solving Eq. (56) numerically can be a very difficult task. Fortunately, doing so is not necessary if one is concerned with only far-field measurements.

C. Reflection coefficient

We will now utilize the results of Sec. IV B. We will use the assumption of Sec. III B, namely, that the crystallographic and optical axes of the medium coincide so that the tensor Σ is diagonal in the laboratory frame. Apart from other simplifications, media of this type are nonchiral and do not rotate the polarization of the transmitted and reflected waves.

This property holds even beyond the homogenization limit since it is a straightforward consequence of the elementary cell symmetries and it will enable us to consider the s and p polarizations separately.

In this section we will explicitly use the reference frame shown in Fig. 2. That is, we will assume that the plane of incidence is the xz plane and that the projection of the wave vectors $\mathbf{k}_i$, $\mathbf{k}_r$, and $\mathbf{q}$ onto the interface is $\mathbf{k}_\perp = k_x\hat{\mathbf{x}}$.

1. s polarization

In the case of s polarization the incident and reflected waves are polarized perpendicularly to the plane of incidence. Consequently, we have $\mathbf{A}_i, \mathbf{A}_r \propto \hat{\mathbf{y}}$, and the *exact* reflection coefficient is given by

$$r = \frac{\mathbf{A}_r \cdot \hat{\mathbf{y}}}{\mathbf{A}_i \cdot \hat{\mathbf{y}}} = -\frac{\tilde{\mathbf{F}}(\mathbf{k}_r) \cdot \hat{\mathbf{y}}}{\tilde{\mathbf{F}}(\mathbf{k}_i) \cdot \hat{\mathbf{y}}} \frac{\exp[i(k_{iz} - q_z)h] - 1}{\exp[-i(k_{iz} + q_z)h] - 1}. \quad (76)$$

To derive the second equality we have used the expressions (68) and (71) for the amplitudes $\mathbf{A}_i$ and $\mathbf{A}_r$. This is an exact expression that retains its physical meaning as long as Eq. (57) holds. In the homogenization limit, we use the expressions (69) to obtain

$$r = \frac{k_{iz} - q_z}{k_{iz} + q_z}. \quad (77)$$

Here q_z is given by

$$q_z = \sqrt{k^2\eta_y - k_x^2}, \quad (78)$$

which follows from the dispersion relation (37), in which we must take $q_x = k_x$. The square root branch in Eq. (78) is determined by the condition $\text{Im}(q_z) > 0$.

The expressions (77) and (78) should be compared to the corresponding Fresnel coefficient r_F and the dispersion relation for a homogeneous medium characterized by the permittivity and permeability tensors $\bar{\epsilon}$ and $\bar{\mu}$:

$$r_F = \frac{k_{iz} - q_z/\bar{\mu}_{xx}}{k_{iz} + q_z/\bar{\mu}_{xx}}. \quad (79)$$

The wave number q_z in an effective medium satisfies the dispersion relation

$$q_z = \sqrt{k^2\bar{\epsilon}_{yy}\bar{\mu}_{xx} - k_x^2\frac{\bar{\mu}_{xx}}{\bar{\mu}_{zz}}}. \quad (80)$$

As discussed in Sec. III B1, we must impose the condition (32) on the EMPs $\bar{\epsilon}$ and $\bar{\mu}$ in order to obtain the same laws of dispersion in the composite and in the continuous medium models. In particular, this condition guarantees that the quantities q_z given by Eqs. (78) and (80) are equal for all values of k_x. However, if this is so, the only way the two expressions (77) and (79) can yield the same reflection coefficient is if we set $\xi = 1$ in Eq. (32), which corresponds to $\bar{\mu} = 1$.

We note that to reach the above conclusion it is sufficient to consider the reflection coefficient for s polarization only. We will show next that the same conclusion can be reached by considering p polarization only and that the homogenization results obtained in these two cases are consistent.

2. p polarization

In the case of p polarization the reflection coefficient can be conveniently defined by using the ratio of tangential components of the magnetic field for the reflected and incident waves. The magnetic field amplitudes of these waves are given by

$$\mathbf{B}_{i,r} = \frac{1}{k} \mathbf{k}_{i,r} \times \mathbf{A}_{i,r}. \tag{81}$$

As could be anticipated, the amplitudes $\mathbf{B}_{i,r}$ are aligned with the y axis. We can now use the expressions (68) and (71) for the amplitudes $\mathbf{A}_{i,r}$ to find the *exact* reflection coefficient:

$$r = \frac{\mathbf{B}_r \cdot \hat{\mathbf{y}}}{\mathbf{B}_i \cdot \hat{\mathbf{y}}} = -\frac{[\mathbf{k}_r \times \tilde{\mathbf{F}}(\mathbf{k}_r)] \cdot \hat{\mathbf{y}}}{[\mathbf{k}_i \times \tilde{\mathbf{F}}(\mathbf{k}_i)] \cdot \hat{\mathbf{y}}} \frac{\exp[i(k_{iz} - q_z)h] - 1}{\exp[-i(k_{iz} + q_z)h] - 1}. \tag{82}$$

In the homogenization limit, this expression is simplified by using Eq. (69), which leads to

$$r = -\frac{[\mathbf{k}_r \times (1 + \Sigma)\mathbf{F}_0] \cdot \hat{\mathbf{y}}}{[\mathbf{k}_i \times (1 + \Sigma)\mathbf{F}_0] \cdot \hat{\mathbf{y}}} \frac{k_{iz} - q_z}{k_{iz} + q_z}. \tag{83}$$

As shown in Appendix B, Eq. (83) can be further simplified to read

$$r = \frac{k_{iz}/\epsilon_b - q_z/\eta_x}{k_{iz}/\epsilon_b + q_z/\eta_x}. \tag{84}$$

In Eqs. (83) and (84) q_z satisfies the dispersion relation for the p-polarized wave Eq. (38). With the substitution $q_x = k_x$ the latter reads

$$q_z = \sqrt{k^2 \eta_x - k_x^2 \frac{\eta_x}{\eta_z}}. \tag{85}$$

As in the case of s polarization, the branch of the square root is determined by the condition $\mathrm{Im}(q_z) > 0$.

We wish to compare the expressions (84) and (85) to the analogous expressions in a continuous medium with the EMPs $\bar{\epsilon}$ and $\bar{\mu}$. The Fresnel reflection coefficient for a p-polarized incident wave is given by

$$r_F = \frac{k_{iz}/\epsilon_b - q_z/\bar{\epsilon}_{xx}}{k_{iz}/\epsilon_b + q_z/\bar{\epsilon}_{xx}} \tag{86}$$

and the dispersion relation in the effective medium is

$$q_z = \sqrt{k^2 \bar{\epsilon}_x \bar{\mu}_y - k_x^2 \frac{\bar{\epsilon}_x}{\bar{\epsilon}_z}}. \tag{87}$$

As in the case of s polarization, the condition (32) with an arbitrary parameter ξ guarantees that the two expressions (85) and (87) yield the same wave number q_z for all values of k_x. However, the expressions (84) and (86) yield the same reflection coefficient only if we set $\xi = 1$ in Eq. (32).

This completes the proof that the correct choice of the parameter ξ in Eq. (32) is $\xi = 1$ and correspondingly the correct homogenization result is $\bar{\mu} = 1$. A similar proof has been given by us for a one-dimensional layered medium in Ref. [43] for both s and p polarizations.

V. COMPARISON OF POINT-DIPOLE AND CONTINUOUS-MEDIUM MODELS

The model of pointlike polarizable particles arranged on a three-dimensional infinite lattice possesses an intuitive physical appeal. Historically, many authors have used this model and although an exhaustive review is outside of the scope of this paper, Refs. [34–36,40,45,46] can be mentioned. Unfortunately, the model is haunted by divergences. In this section we will discuss the nature and origins of these divergences and some of the commonly used methods for their regularization. We will also attempt, to the degree it is possible, to establish a correspondence between the model of point dipoles and the model of a continuous two-component medium, which is the subject of this paper.

Most previous works on electromagnetic waves in point-dipole lattices assume that the background medium is a vacuum. For compatibility of results and simplicity of notation, we will also make this assumption (in this section only) and set $\epsilon_b = 1 + i0$ and $k_b = k = \omega/c + i0$.

The model of point dipoles considers an array of pointlike particles that have well-defined locations, but no shape or size. Instead of the latter two physical characteristics, the electric dipole polarizability α is used. In some generalizations of the model the magnetic dipole polarizability is also included. The basic idea of this approach is that the electromagnetic response of a particle is completely characterized by its polarizability.

If only the electric polarizability is retained, one arrives, in lieu of the integral equation (2), at the set of algebraic equations

$$\frac{1}{\alpha} \mathbf{d}_n = \mathbf{E}_i(\mathbf{r}_n) + \sum_{m \neq n} G(\mathbf{r}_n, \mathbf{r}_m) \mathbf{d}_m. \tag{88}$$

Here $\mathbf{d}_n$ is the electric dipole moment of the nth particle. Now two important points should be made. First, the summation on the right-hand side of Eq. (88) is restricted only to the indices m that are not equal to n. This reflects the idea that the electric field at the site of the nth dipole is a superposition of the incident wave $\mathbf{E}_i(\mathbf{r}_n)$ and the waves scattered by all other dipoles. Second, energy conservation requires that [47–49] $\mathrm{Im}(1/\alpha) \leqslant -2k^3/3$. If the equality holds, the particles are nonabsorbing. It is convenient to decompose the inverse polarizability as

$$\frac{1}{\alpha} = \frac{1}{\alpha_{\mathrm{LL}}} - i\frac{2k^3}{3}, \tag{89}$$

where α_{LL} is the Lorenz-Lorentz quasistatic polarizability and $-i2k^3/3$ is the first nonvanishing radiative correction to the imaginary part of $1/\alpha$. Radiative corrections to the real part of $1/\alpha$ also exist and are in fact of a lower order in k, but it is the correction to the imaginary part that is physically important and should be retained even in the limit $kh \to 0$. We will see momentarily that the two seemingly unrelated facts mentioned above are mathematically connected.

We now consider an infinite lattice, set the incident field to zero, and seek the solution to Eq. (88) in the form $\mathbf{d}_n = \mathbf{d} \exp(i\mathbf{q} \cdot \mathbf{r}_n)$. This results in the eigenproblem

$$\frac{1}{\alpha} \mathbf{d} = S(\mathbf{q})\mathbf{d}, \tag{90}$$

where

$$S(\mathbf{q}) = \sum_{m \neq n} G(\mathbf{r}_n, \mathbf{r}_m) \exp[-i\mathbf{q} \cdot (\mathbf{r}_n - \mathbf{r}_m)] \qquad (91)$$

is the dipole sum. Using the Fourier representation (5), we rewrite Eq. (91) as

$$S(\mathbf{q}) = \frac{4\pi}{3} \int \frac{d^3 p}{(2\pi)^3} K(\mathbf{p}) \sum_{m \neq n} \exp[i(\mathbf{p} - \mathbf{q}) \cdot (\mathbf{r}_m - \mathbf{r}_n)]. \qquad (92)$$

The first complication encountered in the above is that the summation on the right-hand side of Eq. (92) is incomplete. We can easily fix this problem by adding and subtracting unity to the series, which leads to

$$S(\mathbf{q}) = \frac{4\pi}{3} \left[\frac{1}{h^3} \sum_{\mathbf{g}} K(\mathbf{q} + \mathbf{g}) - \int \frac{d^3 p}{(2\pi)^3} K(\mathbf{p}) \right], \qquad (93)$$

where we have used the Poisson summation formula (14). Still, both terms on the right-hand side of Eq. (93) are divergent. We will deal with the integral first. To this end we utilize the expression for $K(\mathbf{p})$ given in Eq. (6) and notice that the angular integral of the term $p^2 - 3\mathbf{p} \otimes \mathbf{p}$ is zero in three dimensions. Therefore, we have

$$I \equiv \int \frac{d^3 p}{(2\pi)^3} K(\mathbf{p}) = 4\pi \int_0^\infty \frac{p^2 dp}{(2\pi)^3} \frac{2k^2}{p^2 - k^2}. \qquad (94)$$

This is still a divergent integral. We can regularize Eq. (94) by writing

$$I = \lim_{\lambda \to 0} \left\{ 4\pi \int_0^\infty \frac{p^2 dp}{(2\pi)^3} \frac{2k^2}{p^2 - k^2} \exp[-(\lambda p)^2] \right\}. \qquad (95)$$

The above limit indeed exists and is equal to $ik^3/2\pi$, assuming that $\mathrm{Im}\, k > 0$ (which is true if we set $k = \omega/c + i0$). Upon substitution of this result into Eq. (93) we find that

$$\frac{4\pi}{3} \left[-\int \frac{d^3 p}{(2\pi)^3} K(\mathbf{p}) \right] = -i \frac{2k^3}{3}. \qquad (96)$$

We now use the decomposition (89) and notice that the above term is canceled by a similar term on the left-hand side of Eq. (90). Taking into account this cancellation, Eq. (90) becomes

$$\frac{1}{\alpha_{\mathrm{LL}}} \mathbf{d} = \frac{4\pi}{3h^3} \sum_{\mathbf{g}} K(\mathbf{q} + \mathbf{g}) \mathbf{d}. \qquad (97)$$

The mathematical tricks used so far are not very objectionable. The result (96) is a reflection of the fact that

$$\lim_{\lambda \to 0} \left[\frac{3}{4\pi \lambda^3} \int_{|\mathbf{r}' - \mathbf{r}| \leqslant \lambda} G(\mathbf{r}, \mathbf{r}') d^3 r' \right] = -i \frac{2k^3}{3}. \qquad (98)$$

Here we have assumed that the particle is spherically symmetric. The use of a different integration volume in Eq. (98), or of a different regularization function in Eq. (95), would certainly yield a different result. Fortunately, if $kh \ll 1$, only the real part of I is affected by the choice of the regularization function in Eq. (95) while the imaginary part is relatively stable. If $\mathrm{Re}\, I$ is unimportant, e.g., if it is small compared to the sum of real parts of all other contributions in Eq. (93), then Eq. (97)

is a good approximation, regardless of the true shape of the particles.

However, the divergence of the series on the right-hand side of Eq. (97) is truly problematic. One can attempt to regularize this divergence by the same mathematical trick that was used above. However, the result of such a manipulation would indeed depend on the regularization function in a nontrivial way. One can conclude that knowledge of the particle polarizability is in fact insufficient for solving the problem at hand. The shape of the particles is also important and cannot be disregarded.

Another way to look at this is the following. The polarizability α defines the response of a particle to an external electric field that is almost uniform over the particle volume. However, in an infinite three-dimensional lattice the electric field is not uniform over the particle volume, no matter how small the particle is. This is because the lattice Green's function $W(\mathbf{R}, \mathbf{R}')$ given by Eq. (12) experiences an integrable divergence when $\mathbf{R} = \mathbf{R}'$. However, in the point-dipole model, we are attempting to evaluate this function exactly at $\mathbf{R} = \mathbf{R}' = 0$, which is not mathematically reasonable.

It appears that the only feasible approach to regularize the summation in Eq. (97) is to endow the particles with a finite volume, as done, for example, in Ref. [34]. This would naturally lead to a modification of Eq. (97) in which the right-hand side is multiplied by a decaying function $f(\mathbf{g})$, ensuring convergence. Unfortunately, the exact form of $f(\mathbf{g})$ strongly depends on the particle shape and size. If the regularization is carried out in a mathematically consistent way, one would end up with a set of equations that are identical to the equations obtained here for the model of a continuous two-component medium.

Evidently, within the point-dipole model, one wishes to avoid introducing the particle shape and size. Then the only conceivable approach to regularization is simply to truncate the series in Eq. (97), by leaving only the $\mathbf{g} = 0$ term in the summation, which leads to the eigenproblem

$$\frac{1}{\alpha_{\mathrm{LL}}} \mathbf{d} = \frac{4\pi}{3h^3} K(\mathbf{q}) \mathbf{d}. \qquad (99)$$

Regularization of this type is in fact appropriate for small spherical particles. If one also uses the quasistatic polarizability of a sphere of radius a, namely,

$$\alpha_{\mathrm{LL}} = a^3 \frac{\epsilon_a - 1}{\epsilon_a + 2}, \qquad (100)$$

then Eq. (99) becomes equivalent to the Clausius-Mossotti relation and the EMT that follows from it is the standard Maxwell-Garnett approximation.

One may be tempted to forget about the limits of applicability of Eq. (99). In other words, once Eq. (99) is derived, it is technically possible to use it with *any* polarizability α_{LL}. The latter can be obtained independently, i.e., by solving the Laplace equation for a single isolated particle of arbitrary shape. Unfortunately, this approach is mathematically inconsistent. Equation (99) was derived from Eq. (97) by applying a regularization method that is appropriate only for small spheres. Application of Eq. (99) to particles of nonspherical shape is likely to result in errors.

In summary, the model of point dipoles is capable of reproducing the standard Maxwell-Garnett mixing rule for small spheres. Radiative corrections to this result can also be derived [35]. However, in three dimensions the model breaks down and cannot be used when a substantial deviation from the Maxwell-Garnett approximation is expected, i.e., for particles whose volume fraction is not small or whose shape is different from a sphere. In other words, the model does not provide a mathematically consistent way of computing the self-energy Σ that appears in Eqs. (23) or (33) and is therefore usable only in the physical situations when Σ can be neglected. Nevertheless, we note that in systems of lower dimensionality (e.g., in nanoparticle chains), the point-dipole model is useful and can provide significant physical insights.

VI. CONTINUED-FRACTION EXPANSION OF THE SELF-ENERGY AND THE MEAN-FIELD APPROXIMATION

A. Abstract notation

In this section we will find it convenient to rewrite Eqs. (19b) and (21) in Dirac notation. First, we note that in order to recover all components of the tensor Σ one must solve Eq. (19b) for three different right-hand sides: $\mathbf{F}_0 = \hat{\mathbf{x}}$, $\mathbf{F}_0 = \hat{\mathbf{y}}$, and $\mathbf{F}_0 = \hat{\mathbf{z}}$. To this end we introduce a triplet of infinite-dimensional vectors $|a_\beta\rangle$; operators Q, M, and W; and vectors $|b_\beta\rangle$ ($\beta = x,y,z$) according to

$$\langle \alpha \mathbf{g} | a_\beta \rangle = M(\mathbf{g}) \delta_{\alpha\beta}, \tag{101a}$$

$$\langle \alpha \mathbf{g} | Q | \alpha' \mathbf{g}' \rangle = \delta_{\mathbf{g}\mathbf{g}'}(1 - 3\hat{g}_\alpha \hat{g}_{\alpha'}), \tag{101b}$$

$$\langle \alpha \mathbf{g} | M | \alpha' \mathbf{g}' \rangle = \delta_{\alpha\alpha'} M(\mathbf{g} - \mathbf{g}'), \tag{101c}$$

$$W = QM, \tag{101d}$$

$$|b_\beta\rangle = Q|a_\beta\rangle. \tag{101e}$$

Note that Q is diagonal in the index $\mathbf{g}$ and M is diagonal in the index α, but the product of the two, $W = QM$, is not diagonal. We must also keep in mind that the index $\mathbf{g}$ in the above equations is not allowed to take the zero value. We further define the vectors $|F_\beta\rangle$ as the solutions to

$$\left(\frac{1}{\rho\chi} - W \right) |F_\beta\rangle = |b_\beta\rangle. \tag{102}$$

The above is equivalent to the set (19b). The tensor elements of Σ are defined by

$$\Sigma_{\alpha\beta} = \langle a_\alpha | F_\beta \rangle = \langle a_\alpha | \left(\frac{1}{\rho\chi} - W \right)^{-1} | b_\beta \rangle$$

$$= \langle a_\alpha | \left(\frac{1}{\rho\chi} - QM \right)^{-1} Q | a_\beta \rangle. \tag{103}$$

It can be seen that Σ is computed as the resolvent of the operator $W = QM$ and plays the role of the self-energy, which accounts for interactions between the inclusions.

B. Mean-field approximation

The mean-field approximation is often misunderstood. In particular it is unrelated to Maxwell-Garnett theory. Rather it allows one to replace certain operators by appropriately chosen scalar multiples of the identity. The approximation reproduces

the exact zeroth and first moments of the resolvent and serves as the first-order approximation in its continued-fraction expansion. Here the approximation is explained following Berry and Percival [50].

Let us seek the solution to Eq. (102) in the form $|F_\beta\rangle = \lambda|b_\beta\rangle$, where λ is a scalar to be determined. Upon substitution of this ansatz into Eq. (102), we obtain the equation

$$\left(\frac{1}{\rho\chi} - \frac{1}{\lambda} \right) |b_\beta\rangle = W|b_\beta\rangle. \tag{104}$$

Because $|b_\beta\rangle$ is generally not an eigenvector of W, there is no such value of λ for which Eq. (104) would hold. The best we can hope for is that a *projection* of this equation onto a given vector would hold for some λ. Since we are interested not in the whole vector $|F_\beta\rangle$ but in its projection onto $|a_\alpha\rangle$, it seems reasonable to project Eq. (104) onto the latter. This yields

$$\lambda = \frac{\rho\chi}{1 - \rho\chi \frac{\langle a_\alpha|W|b_\beta\rangle}{\langle a_\alpha|b_\beta\rangle}} \tag{105}$$

and the corresponding mean-field approximation for the self-energy is

$$\Sigma_{\alpha\beta} = \frac{\rho\chi\langle a_\alpha|b_\beta\rangle}{1 - \rho\chi \frac{\langle a_\alpha|W|b_\beta\rangle}{\langle a_\alpha|b_\beta\rangle}} = \frac{\rho\chi\langle a_\alpha|Q|a_\beta\rangle}{1 - \rho\chi \frac{\langle a_\alpha|QMQ|a_\beta\rangle}{\langle a_\alpha|Q|a_\beta\rangle}}. \tag{106}$$

As mentioned in Sec. III C, the matrix element

$$\langle a_\alpha|Q|a_\beta\rangle = \sum_{\mathbf{g}\neq 0} [M(-\mathbf{g})Q(\mathbf{g})M(\mathbf{g})]_{\alpha\beta} \tag{107}$$

is identically zero for inclusions with cubic symmetry (in three-dimensional composites) so that Eq. (106) yields in this case zero and is not useful. If $\langle a_\alpha|Q|a_\beta\rangle$ is zero, a nonvanishing mean-field approximation can be obtained by shifting the solution according to $|F_\beta\rangle = \rho\chi|b_\beta\rangle + |F'_\beta\rangle$. The self-energy is then given by $\Sigma_{\alpha\beta} = \langle a_\alpha|F'_\beta\rangle$, where $|F'_\beta\rangle$ satisfies

$$\left(\frac{1}{\rho\chi} - W \right) |F'_\beta\rangle = \rho\chi W|b_\beta\rangle. \tag{108}$$

The mean-field approximation for the shifted equation (108) is

$$\Sigma_{\alpha\beta} = \frac{(\rho\chi)^2\langle a_\alpha|QMQ|a_\beta\rangle}{1 - \rho\chi \frac{\langle a_\alpha|(QM)^2Q|a_\beta\rangle}{\langle a_\alpha|QMQ|a_\beta\rangle}}. \tag{109}$$

C. Continued-fraction expansion of the self-energy

Continued-fraction expansions (CFEs) are very useful in physics [51,52]. The mathematical underpinning of all CFEs is the theory of the correspondence between the formal Laurent series of meromorphic functions and certain continued fractions [53]. There exists a deep mathematical relation between CFEs and the problem of moments, that is, the problem of finding a distribution from the knowledge of its moments.

Continued-fraction expansions can be derived in different ways. Haydock [51] has employed the Lanczos recursion to transform a certain Hamiltonian to tridiagonal form. A diagonal element of the inverse of a tridiagonal matrix can be written as a J-fraction (a continued fraction of Jacobi type). In Ref. [51] this procedure was applied to a Hermitian

operator to compute a diagonal matrix element of the resolvent. In this paper the operator W in Eq. (102) or (103) is not symmetric or Hermitian and we are interested in off-diagonal elements of the resolvent. Therefore, the numerical procedure used by Haydock is not directly applicable. Perhaps it can be generalized to become applicable to the problem at hand; we have not explored this possibility. Instead, we will derive a CFE for the right-hand side of Eq. (103) from the following theorem, which does not require any symmetry properties of the operators involved and yields a CFE for arbitrary off-diagonal elements. The resultant expansion will be an S-fraction (a continued fraction of Stieltjes type). Note that an S-fraction can always be transformed into a J-fraction by the so-called equivalence transformation.

Theorem 1. Let W be a linear operator acting on the Hilbert space $\mathcal{H}$ and $\mathcal{Z}$ be a complex number. Suppose that $|\phi\rangle$, $|\psi\rangle \in \mathcal{H}$. If (i) $\langle\phi|\psi\rangle \neq 0$ and (ii) $(\mathcal{Z} - W)^{-1}$ exists, then

$$\langle\phi|(\mathcal{Z} - W)^{-1}|\psi\rangle = \frac{\mathcal{Z}^{-1}\langle\phi|\psi\rangle}{1 - \frac{\langle\phi|(\mathcal{Z} - WT)^{-1}W|\psi\rangle}{\langle\phi|\psi\rangle}}, \quad (110)$$

where

$$T = 1 - \frac{|\psi\rangle\langle\phi|}{\langle\phi|\psi\rangle}. \quad (111)$$

The proof is given in Appendix C. Note that Eq. (110) has a finite limit when $\mathcal{Z} \to 0$.

The factor $\langle\phi|(\mathcal{Z} - WT)^{-1}W|\psi\rangle$ in the denominator of Eq. (110) can be written as $\langle\phi|(\mathcal{Z} - W_1)^{-1}|\psi_1\rangle$, where $W_1 = WT$ and $|\psi_1\rangle = W|\psi\rangle$. The formula (110) can now be applied to $\langle\phi|(\mathcal{Z} - W_1)^{-1}|\psi_1\rangle$ and so on iteratively. After some manipulation, this yields the following expansion:

$$\langle\phi|(\mathcal{Z} - W)^{-1}|\psi\rangle = \cfrac{\kappa_1}{\mathcal{Z} - \cfrac{\kappa_2}{1 - \cfrac{\kappa_3}{\mathcal{Z} - \cdots}}}, \quad (112)$$

Note the interlacing factors of $\mathcal{Z}$ and 1. The coefficients κ_j $(j = 1, 2, \ldots)$ are obtained from a three-point recursion. Namely, starting from $|\psi_0\rangle = 0$, $|\psi_1\rangle = |\psi\rangle$, and $\kappa_1 = \langle\phi|\psi\rangle$, we compute for $j = 1, 2, \ldots$

$$|\psi_{j+1}\rangle = W(|\psi_j\rangle - \kappa_j|\psi_{j-1}\rangle), \quad \kappa_{j+1} = \frac{\langle\phi|\psi_{j+1}\rangle}{\langle\phi|\psi_j\rangle}. \quad (113)$$

To obtain a CFE of the right-hand side of Eq. (103) we identify $\mathcal{Z} = 1/\rho\chi$, $W = QM$, $|\phi\rangle = |a_\alpha\rangle$, and $|\psi\rangle = |b_\beta\rangle = Q|a_\beta\rangle$.

With the above substitutions taken into account, it transpires that the coefficients κ_j are determined only by the geometry of the composite. Once a set of κ_j have been found for a given geometry, the EMPs can be easily computed for any material parameters of the composite constituents. This is a characteristic feature of a spectral theory and the CFE (112) is in fact a spectral representation of the self-energy Σ.

Finally, we note that in the case of three-dimensional composites with cubic symmetry, the first condition of the Theorem does not hold when the theorem is applied directly to Eq. (103). In this case one can build a CFE starting from the shifted equation (108).
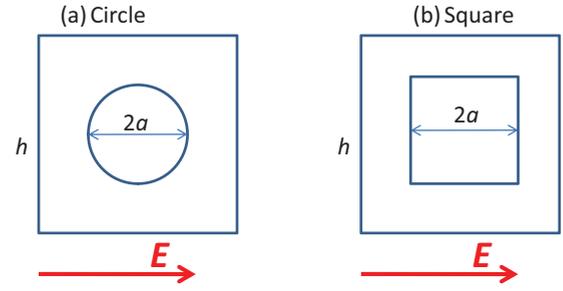
FIG. 4. (Color online) Two types of elementary cells used in numerical simulations.

VII. NUMERICAL SIMULATIONS

A. General setup

Numerical simulations have been performed for a two-dimensional composite. The composite is periodic in the xy plane while the inclusions form infinitely long fibers that are oriented parallel to the z axis and can have different cross sections. The case when the electric field is parallel to the fibers is not considered here since this polarization results in a simple arithmetic average of the type (44). However, when the electric field is polarized in the xy plane, the homogenization problem is nontrivial and can be numerically challenging. We will consider inclusions with circular and square cross sections, as illustrated in Fig. 4. The functions $M(\mathbf{g})$ for these shapes are given in Appendix A.

It is assumed that the host medium is a vacuum and the inclusions are metallic and characterized by a frequency-dependent Drude permittivity of the form

$$\epsilon_a = 1 - \frac{3\omega_F^2}{\omega(\omega + i\gamma)}, \quad \epsilon_b = 1. \quad (114)$$

In Eq. (114) $\omega_F = \omega_p/\sqrt{3}$ is the Frohlich frequency, ω_p is the plasma frequency, and γ is the Drude relaxation constant. We will compute the effective permittivity of the composite $\bar{\epsilon}$ as a function of frequency for $0.1 \leqslant \omega/\omega_F \leqslant 2$ and for the fixed ratio $\gamma/\omega_F = 0.1$. It is assumed that for all frequencies used in the simulations, the basic condition for the validity of a standard EMT, $k_b h, q h \ll 1$, is satisfied.

Numerical simulations will be performed by truncating the infinite set of equations (19b) so that the vectors $\mathbf{g}$ fill the box

$$-2\pi L/h \leqslant g_x, \quad g_y \leqslant 2\pi L/h, \quad (115)$$

where L is an integer. The total number of $\mathbf{g}$ vectors that satisfy the above inequality is $(2L + 1)^2$ and the total number of algebraic equations to be solved is $N = 2[(2L + 1)^2 - 1]$, where we have accounted for the fact that the vector $\mathbf{g} = 0$ is excluded in the set of equations (19b). It can be seen that $N \to 8L^2$ when $L \to \infty$. In the simulations we will use integer powers of 2 for L up to $L = 2^8 = 256$. The latter case corresponds to $N = 526$ and 366 equations.

The truncated set of equations (19b) can be solved by any direct numerical method. The computational complexity of direct methods is $O(N^3)$ and the solution must be obtained anew for every frequency used (we sample the frequency at 200 equidistant points in the interval $0.1 \leqslant \omega/\omega_F \leqslant 2$). This is time consuming but possible for $L \leqslant 64$. For larger values of

L, direct methods become impractical. We will therefore use the CFE of Sec. VI C. The computational complexity of this expansion is $O(j_{max}N^2)$, where j_{max} is the order of truncation of the continued fraction. More specifically, the continued fraction is truncated by assuming that $\kappa_j = 0$ for $j > j_{max}$, so that only the first j_{max} coefficients are used in Eq. (112). For the problem at hand, $j_{max} \approx 50$ will prove sufficient. Other iterative methods such as the conjugate gradient method also have computational complexity $O(j_{max}N^2)$, with j_{max} being the number of iterations. However, the computationally intensive part of the conjugate-gradient solver [when applied to Eq. (19b)] must be repeated for every value of ω, while the coefficients κ_j in Eq. (112) need to be computed only once for a given geometry.

The inclusions shown in Fig. 4 have cubic symmetry. As discussed in Sec. III B, the self-energy Σ is reduced in this case to a scalar. As a result, the effective medium is isotropic in the xy plane. Of course anisotropy can still be revealed if the polarization vector has a component along the
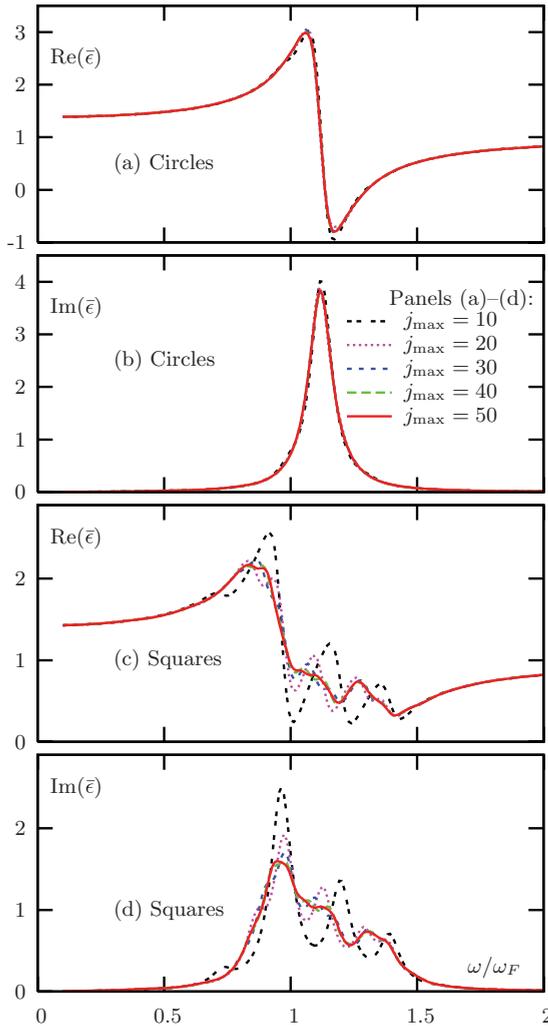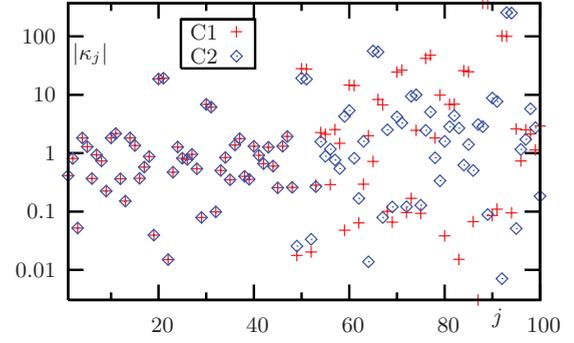
FIG. 6. (Color online) Absolute values of the coefficients κ_j computed for circular inclusions with $\rho = 0.16$ and $L = 64$ on two different computers (C1 and C2). The same FORTRAN code and input data have been used in both cases.

z axis. In the simulations reported below, we have computed Σ by solving Eq. (19b) and using the definition (21). The effective permittivity for transversely polarized waves was then computed by using Eq. (33).

B. Convergence and stability

The convergence of the CFE (112) with the truncation order of the continued fraction j_{max} is illustrated in Fig. 5. Here the real and imaginary parts of the effective permittivity are plotted as functions of frequency. It can be seen that the convergence is very fast for circular inclusions and somewhat slower for square inclusions. In all cases $j_{max} = 50$ is sufficient for convergence.

The three-point recurrence relation (113) is numerically unstable for large values of j. This is illustrated in Fig. 6. Shown in this figure are the coefficients κ_j obtained on two different computers for the geometry described in the figure caption. The same code and input data were used in both cases. The coefficients from the two sets coincide for $j \lesssim 50$ with high precision. However, differences start to appear at $j \sim 50$ and at $j \sim 100$ the coefficients are unreliable. The instability occurs when an iteration step in Eq. (113) asks for a relatively small difference of two large numbers and the numerical precision of the floating-point arithmetic is exceeded.

The instability illustrated in Fig. 6 appears to be troublesome but is in fact of little concern. This is illustrated in Fig. 7, which displays the effective permittivity computed by the CFE (112) for various truncation orders j_{max} and the same quantity computed by solving Eq. (19b) directly. One of the sets of κ_j's displayed in Fig. 6 has been used for computing the data points for Figs. 7(a) and 7(b). Despite the instability, the curves with $j_{max} = 50$ and 100 are indistinguishable and very close to the data points obtained by direct inversion of Eq. (19b). Thus the unreliable coefficients κ_j do not influence the final result. This is one of the nice properties of all CFEs: A numerical instability does not result in numerical imprecision. It is true that increasing the truncation order beyond $j_{max} = 50$ is not useful, but it is not harmful either. This point and some related issues are discussed in more detail in Sec. VIII below.

Having established the convergence properties of the CFE, we next consider convergence with the size of the box L (up to now, all plots have been computed for $L = 64$). In Figs. 8 and

FIG. 5. (Color online) Convergence of the CFE (112) with the truncation order j_{max} for (a) and (b) circular and (c) and (d) quadratic inclusions with the same volume density $\rho = 0.16$. The set of equations (19b) has been truncated using $L = 64$. In (a) and (b) the curves with $j_{max} = 30, 40, 50$ are indistinguishable.
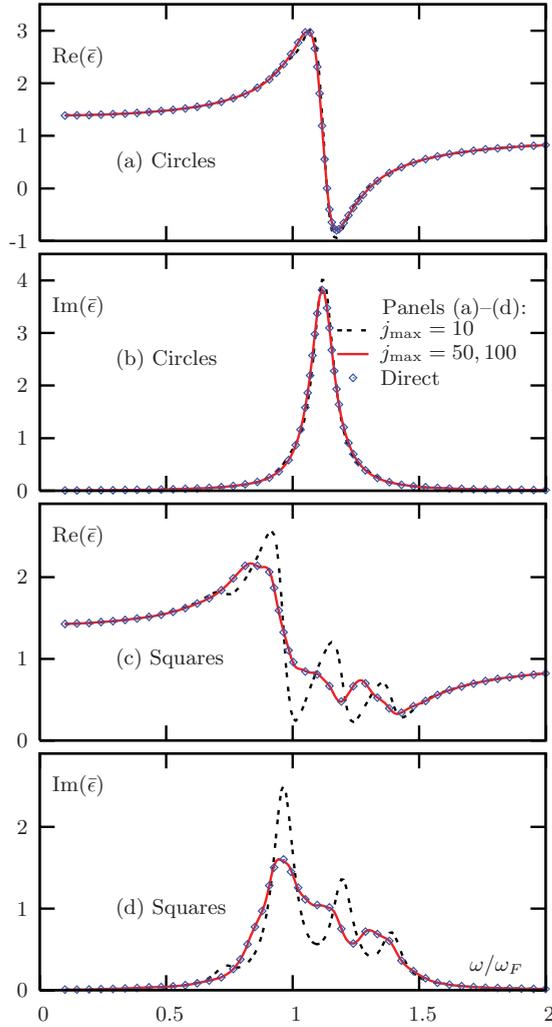
FIG. 7. (Color online) Effective permittivity of (a) and (b) circular and (c) and (d) square inclusions computed using the CFE (112) with the truncation orders $j_{max} = 10$ (dashed line) and $j_{max} = 50,100$ (solid line) and by direct inversion of Eq. (19b) (centered symbols). In all cases $\rho = 0.16$ and $L = 64$. The data points for $j_{max} = 50$ and 100 are visually indistinguishable and therefore represented with the same solid line. The displayed data points for direct inversion (centered symbols) are sampled from the same set of 201 data points; the sampling density in each panel is variable and chosen to accurately represent variations of the corresponding function.

9, $\bar{\epsilon}$ is plotted as functions of frequency for various values of the density ρ and the box size L. Also shown in these figures are the results obtained from the generalized Maxwell-Garnett formula

$$\epsilon_\nu = \epsilon_b \frac{1 + \frac{2\rho}{3} \frac{\epsilon_a - \epsilon_b}{\epsilon_b + \nu(\epsilon_a - \epsilon_b)}}{1 - \frac{\rho}{3} \frac{(\epsilon_a - \epsilon_b)}{\epsilon_b + \nu(\epsilon_a - \epsilon_b)}}, \quad (116)$$

which applies to ellipsoids, with ν being the appropriate depolarization factor. In the case of three-dimensional spheres $\nu = 1/3$ and Eq. (116) coincides with Eq. (33), in which the self-energy Σ is set to zero. In the case of infinite circular cylinders, the depolarization factor, which corresponds to the orthogonal electric polarization, is $\nu = 1/2$.
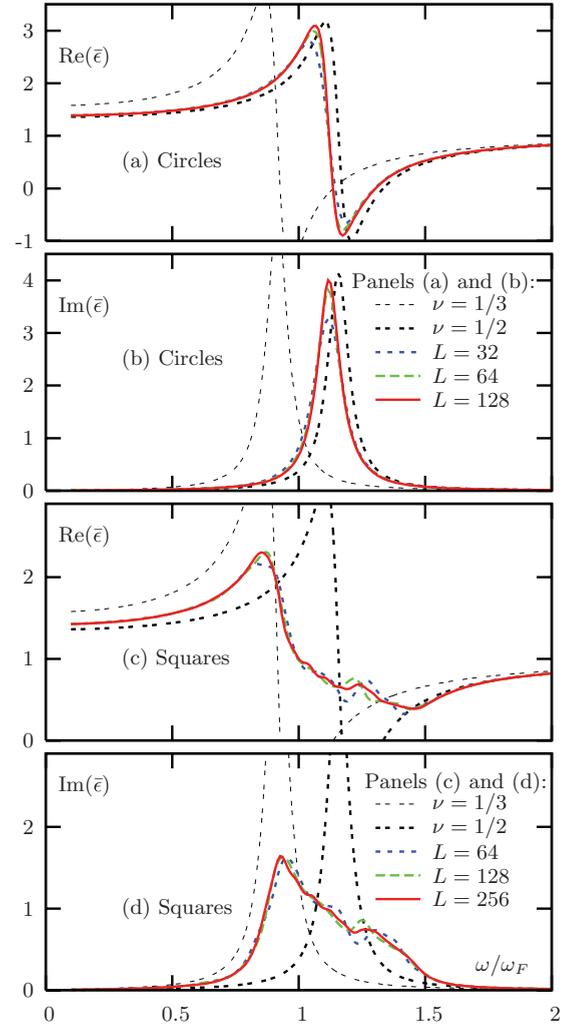
FIG. 8. (Color online) Convergence of the effective permittivity $\bar{\epsilon}$ with the size of the box L for (a) and (b) circular and (c) and (d) square inclusions with $\rho = 0.16$. The curves labeled as $\nu = 1/2$ and $\nu = 1/3$ (thick and thin dashed lines, respectively, with variable spacing) have been obtained from the generalized Maxwell-Garnett mixing formula (116) for the values of ν indicated.

Several conclusions can be drawn from Figs. 8 and 9. First, convergence is obtained for boxes of reasonable size. In all cases shown $L = 256$ yields very accurate results and in some cases $L = 64$ is sufficient. However, it is important to note that we have verified the convergence by doubling the size of the box. Determination of convergence by using linearly sampled values of L (say, $L = 10, 11, 12, \ldots$) can be misleading. This is a typical situation when boundary-value problems are solved numerically. Convergence must be established by at least doubling the size of the mesh used.

Second, it can be seen that convergence is faster for $\rho = 0.32$ than for $\rho = 0.16$. Although the electromagnetic interaction is stronger in the second case, the faster convergence is to be expected. Indeed, the size of the box should be selected so that the sum rules (A2) are satisfied with some reasonable precision,; that is achieved at smaller values of L for larger values of ρ. Even faster convergence is obtain for $\rho = 64$ (data not shown). However, at the percolation threshold
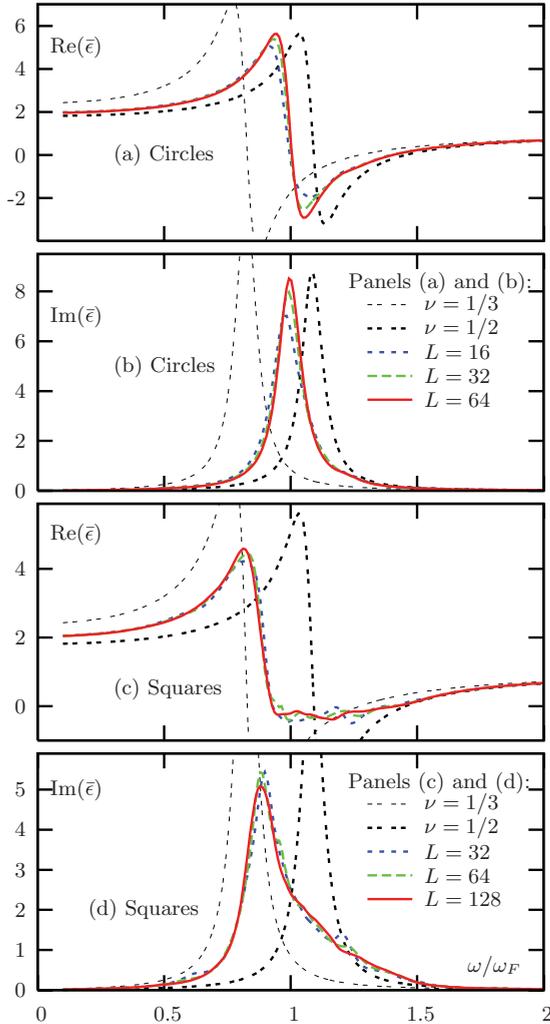
FIG. 9. (Color online) Same as in Fig. 8 but for $\rho = 0.32$ and different values of L as indicated.

($\rho = \pi/4 \approx 0.79$ for circular inclusions), the convergence is relatively slow.

Third, the generalized Maxwell-Garnett formula (116) with $\nu = 1/2$ yields a reasonable result for circular inclusions with $\rho = 0.16$. Even better agreement has been obtained for $\rho = 0.08$ and 0.04 (data not shown). However, as the size of circular inclusions increases, the Maxwell-Garnett approximation becomes less accurate. For square inclusion the approximation is inaccurate even for very small values of ρ. In all cases the electromagnetic interaction tends to shift the absorption peaks from the Maxwell-Garnett prediction towards the lower frequencies. At $\rho = 0.32$ the effect is already quite pronounced.

C. Comparison of inclusions of various size

We finally compare the effective permittivity for circular and square inclusions of different sizes. The results are displayed in Figs. 10 and 11. In the case of circular inclusions there exists a pronounced spectral peak that shifts towards lower frequencies when ρ is increased. However, once the inclusions touch (this happens at $\rho = \pi/4 \approx 0.79$) the single resonance is destroyed and a broad absorption band develops.
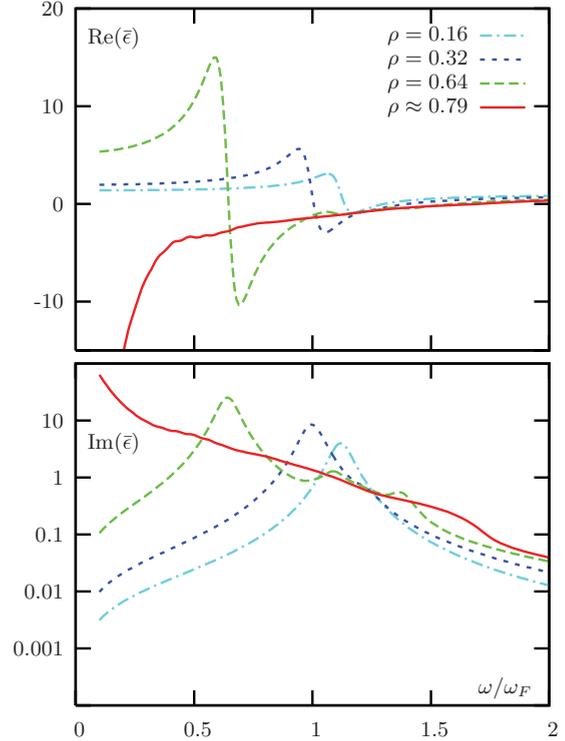
FIG. 10. (Color online) Effective permittivity for circular inclusions of different volume densities. The $\rho \approx 0.79$ case corresponds to the percolation threshold (touching circles).

The lower-frequency behavior of $\bar{\epsilon}$ is in this case metallic since the percolating sample is characterized by a nonzero static conductivity. This result cannot be obtained within the Maxwell-Garnet approximation or the Bruggemann approximation, even at a qualitative level.

The square inclusions do not touch for $\rho < 1$. Correspondingly, the low-frequency behavior of $\bar{\epsilon}$ is not metallic even for large filling fractions, e.g., for $\rho = 0.85$. Interestingly, at relatively small values of ρ, the absorption spectrum forms a band with one main resonance and many minor resonances that are shifted toward the shorter waves. However, as ρ increases, the minor resonances become less pronounced. At $\rho = 0.85$ the spectrum is dominated by a single Lorentzian-type resonance. In the case of circular inclusions, the picture is somewhat different. A single Lorentzian resonance exists at small values of ρ and additional minor resonances develop as ρ increases. These additional resonances are clearly visible in the $\rho = 0.64$ curve shown in the left column of Fig. 10.

VIII. DISCUSSION

A few points that deserve additional discussion are addressed in this section, in no particular order.

A. Conditions of applicability

The EMT derived in this paper describes a composite medium accurately if $qh, kh \ll 1$. There are no additional conditions. In particular, there is no requirement that the permittivity (or conductivity) of any constituent of the composite be bounded. However, if a metallic inclusion has very small
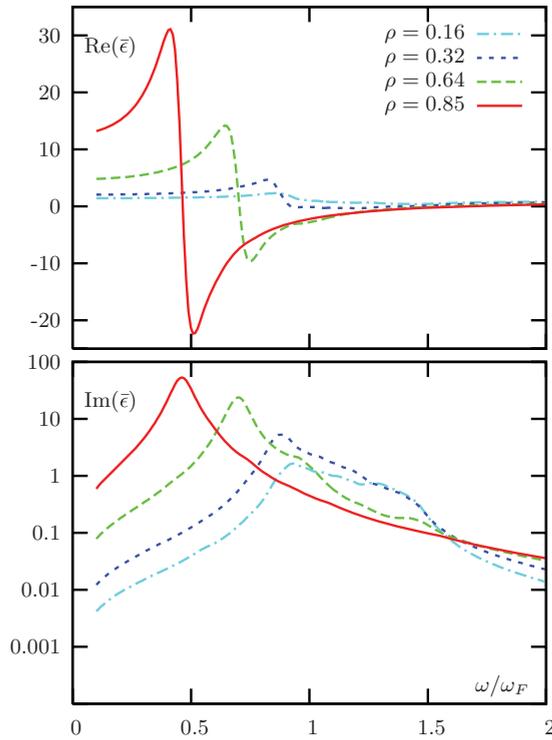
FIG. 11. (Color online) Same as in Fig. 10 but for square inclusions.

losses (very high conductivity), then the effective permittivity computed according to the formulas of this paper can have one or more sharp spectral peaks. These peaks are caused by electromagnetic resonances in the inclusions (which we have not disregarded by any means) and can be seen in Figs. 5–11. In the spectral regions where these resonances take place it is possible that $q \gg k$. This effect is known as the resonance wavelength shortening. Conceivably, the Bloch wave number q can become so large due to this effect that the condition $qh \ll 1$ would break. In this case our theory is inapplicable.

The above consideration can be construed as a justification for the development of extended EMTs, e.g., by taking a limit in which the conductivity of metal inclusions goes to infinity first [23,24,27], or by using other trajectories in the parameter space [26]. However, two important caveats exist. First, in many known applications, EMPs of the order of unity are required, e.g., $\bar{\epsilon} \approx \bar{\mu} \approx -1$ is required for operation of a superlens. In this case, of course, $q \approx k$, there is no resonant wavelength shortening, and our theory applies. The second caveat is that even if metal inclusions have very high conductivity, the imaginary part of the obtained effective permittivity is not small close to a resonance. This can be clearly seen in Figs. 5–11. Therefore, there is not much hope to obtain a resonant effect without simultaneously having strong absorption in the medium. This observation is in agreement with Stockman [54], although we do not pursue here a rigorous mathematical consideration of this point.

Finally, in the case when qh is not actually small compared to unity and our theory does not apply, it appears from considering the *exact* reflection coefficients (76) and (82) that any EMPs that can be introduced in any theory would depend on the angle of incidence. More generally, the EMPs would

depend on the type of illumination. We conclude that the medium is simply not electromagnetically homogeneous in this case.

B. The case of small losses

Another problem associated with high conductivity of metallic inclusions is numerical stability and convergence. The simulations of Sec. VII have been performed for a relatively large loss parameter $\gamma/\omega_F = 0.1$. If this number is substantially reduced, the convergence with the truncation order of the continued fraction $j_{\max}$ is expected to become slower. A general rule of thumb is that the truncation order should not be less than the number of clearly discernible peaks in the function $\text{Im}\bar{\epsilon}(\omega)$ (the absorption spectrum). This is because the CFE truncated at the order $j_{\max}$ captures correctly the first $j_{\max}$ moments of the above function. At sufficiently large values of j, the three-point recursion (113) becomes numerically unstable, as illustrated in Fig. 6. If the required value of $j_{\max}$ is larger than the value of j at which the onset of numerical instability occurs, then the CFE will not yield an accurate numerical result.

The situation outlined above is common for all iterative methods. For example, the convergence of the conjugate-gradient method becomes extremely slow for small ratios of γ/ω_F; at some point, the recurrence relations used in the conjugate-gradient iterations also become numerically unstable. One can hope to improve stability by noting that the nth order tail of the CFE (112), that is, the expression

$$\frac{\kappa_{n+1}}{\mathcal{Z} - \frac{\kappa_{n+2}}{1 - \cdots}},$$

is also an expansion of a certain resolvent and the instability occurs because the parameter ε (defined in the proof of Theorem 1, Appendix C) becomes numerically small. This can be fixed by shifting the operator A as described in Sec. VI B. In this way, a nested set of CFEs can be obtained, where each CFE is numerically stable, as well as the whole expression.

C. Consideration of chirality and polarization conversion

Although the general formalism of this paper allows one to take chiral media into consideration, all derivations that were brought to a logical conclusion have been carried out for the nonchiral case. This has provided a mathematical simplification yet left untouched a wealth of interesting physical phenomena that are associated with chirality and not addressed in the present paper.

Even if the medium is nonchiral, it can exhibit the effect of polarization conversion [55], which has been recently predicted and experimentally observed in deeply subwavelength nanostructures in Ref. [13]. In Sec. IV C we have made an assumption that the plane of incidence coincides with one of the crystallographic planes of the medium. In this case the s- and p-polarized waves are independent and polarization conversion does not occur. However, the homogenization result obtained in this paper is more general and in particular it is applicable to any direction of incidence. If the plane of incidence does not coincide with any crystallographic plane, the geometry of the problem becomes similar to that considered in Ref. [13] and polarization conversion can occur.

In other words, the reflected and transmitted (in the case of a finite slab) waves due to a purely *s*- or *p*-polarized incident wave can have both *s*- and *p*-polarized components and at least theoretically it is possible to design a medium with the conversion coefficient close to unity.

D. Three-dimensional vs two-dimensional simulations

So far we have performed simulations only for 2D media. One can argue that in the 3D case the size of the algebraic problem would become so large as to render the method unusable. Of course three-dimensional electromagnetic problems are always challenging. However, there is reason for optimism. Namely, the formula for the effective permittivity (33) uses the three-dimensional Maxwell-Garnett approximation as the point of departure. In other words, a nonzero value of Σ provides a correction to the three-dimensional Maxwell-Garnett formula. This happens to be true even for two-dimensional media. However, the three-dimensional Maxwell-Garnet formula is inaccurate in the 2D case even for very thin cylinders, as clearly illustrated in Figs. 8 and 9. In the numerical simulations of Sec. VII (for circular inclusions), a lot of effort was spent to compute accurately the self-energy Σ whose effect was essentially to transform the Maxwell-Garnett from a 3D to a 2D form.

In the case of small three-dimensional inclusions one can expect a much faster convergence with L. For example, if the inclusions are small spheres, an accurate result is obtained by starting with $\Sigma = 0$. As the spheres increase in size, the Maxwell-Garnett approximation becomes less accurate and a nonzero value of Σ must be used. However, as we have seen in the numerical simulations, the required values of L are in fact smaller for larger sizes of the inclusions.

Mathematically, the above considerations are related to an interesting fact that was mentioned in Sec. VI. Namely, the matrix element $\langle a_\alpha | Q | a_\beta \rangle$ is identically zero for three-dimensional cells with cubic symmetry. Consequently, the mean-field approximation and the continued-fraction expansion must be derived for the shifted equation (108). As a result, the mean-field formula (109) contains an overall factor of $(\rho\chi)^2$, while in the 2D simulations of Sec. VII this factor was equal to $\rho\chi$.

IX. CONCLUSION

We can draw the following conclusions.

(i) A medium constructed from nonmagnetic components is also nonmagnetic in the limit $h \to 0$. This result is in line with arguments put forth in Ref. [28], the simulations in Ref. [29], and the more formal mathematical theory of Ref. [25].

(ii) The model of pointlike polarizable particles is ill suited for homogenization of three-dimensional periodic composites due to inherent divergences. The point-dipole approximation can still be a useful theoretical tool for studying systems in lower dimensions.

(iii) In agreement with the previous conclusion, we have found numerically that the EMPs are sensitive to the shape of inclusions even if the volume fraction is small. Thus circular and square inclusions in Figs. 7 and 8 have very different spectra of EMPs even though the volume fraction of the

inclusions is $\rho = 0.16$. When the volume fraction becomes larger, the differences between the circular and the square shapes are dramatic. Thus it is shown in Figs. 10 and 11 that the percolation phenomenon occurs for the circular inclusions at the volume fraction $\rho = \pi/4 \approx 0.79$ when the inclusions touch. The composite in this case is conducting. The composite consisting of square inclusions of the volume fill fraction (which do not touch) is still a dielectric.

(iv) We believe that the goal of homogenization theory is to describe a given physical composite. Therefore, rather than studying different limits, which correspond to different trajectories in the parameter space, it is important to delineate regions of the parameter space and to determine to which one of these regions the particular composite belongs. Along similar lines, we note that a satisfactory theory of homogenization requires error estimates. That is, it is critical to understand how the error in the homogenization limit depends upon contrast.

APPENDIX A: MATHEMATICAL PROPERTIES OF $M(\mathbf{g})$ AND SOME SPECIAL CASES

From the definition (18) it follows that

$$M(0) = 1, \quad M(-\mathbf{g}) = M^*(\mathbf{g}). \quad (A1)$$

For the case of inclusions whose center of symmetry coincides with the center of the unit cell we have $M(-\mathbf{g}) = M(\mathbf{g})$ and therefore $M(\mathbf{g})$ is real. If the center of symmetry is displaced by a vector $\mathbf{a}$, the function $\mathbf{M}(\mathbf{g})$ is transformed according to $\mathbf{M}(\mathbf{g}) \to \exp(-i\mathbf{a} \cdot \mathbf{g})\mathbf{M}(\mathbf{g})$.

By applying the Poisson summation formula, we can derive the following sum rules:

$$\sum_{\mathbf{g}} M(\mathbf{g}) = \begin{cases} 1/\rho, & 0 \in \Omega \\ 0, & 0 \notin \Omega, \end{cases} \quad (A2a)$$

$$\sum_{\mathbf{g}} M(-\mathbf{g})M(\mathbf{g}) = \frac{1}{\rho}, \quad (A2b)$$

$$\sum_{\mathbf{g}'} M(\mathbf{g} - \mathbf{g}')M(\mathbf{g}') = \frac{1}{\rho}M(\mathbf{g}). \quad (A2c)$$

These equations hold for inclusions of arbitrary shape. Now define a complimentary function $N(\mathbf{g})$ by

$$N(\mathbf{g}) = \frac{1}{h^3 - V} \int_{C\backslash\Omega} \exp(-i\mathbf{g} \cdot \mathbf{R})d^3R. \quad (A3)$$

Here C denotes the unit cell and $C\backslash\Omega$ is the region complimentary to the inclusion. It can be seen that $N(\mathbf{g})$ has all the properties of $M(\mathbf{g})$ with the substitution $\rho \to 1 - \rho$. Additionally, the functions $M(\mathbf{g})$ and $N(\mathbf{g})$ are related by

$$\rho M(\mathbf{g}) + (1 - \rho)N(\mathbf{g}) = \delta_{\mathbf{g}0}. \quad (A4)$$

From this we obtain the low- and high-density limits

$$\lim_{\rho \to 0} N(\mathbf{g}) = \lim_{\rho \to 1} M(\mathbf{g}) = \delta_{\mathbf{g}0}. \tag{A5}$$

Of course the high-density limit is unreachable for most regular shapes (with the exception of cubes). For example, in the case of spheres, the maximum allowed value of ρ is $\pi/6$.

Some special cases of $M(\mathbf{g})$ are given below. For an inclusion in the shape of either a 3D sphere or a 2D circle of radius $a \leqslant h/2$,

$$M_{3D}(\mathbf{g}) = \frac{3[\sin(ga) - ga\cos(ga)]}{(ga)^3}, \tag{A6a}$$

$$M_{2D}(\mathbf{g}) = \frac{2J_1(ga)}{ga}, \tag{A6b}$$

where $J_1(x)$ is the cylindrical Bessel function of the first kind. For a parallelepiped or rectangle centered at the origin with all faces parallel to the crystallographic planes and sides of length $2a_x$, $2a_y$, and $2a_z$,

$$M_{3D}(\mathbf{g}) = \frac{\sin(g_x a_x)}{g_x a_x} \frac{\sin(g_y a_y)}{g_y a_y} \frac{\sin(g_z a_z)}{g_z a_z}, \tag{A7a}$$

$$M_{2D}(\mathbf{g}) = \frac{\sin(g_x a_x)}{g_x a_x} \frac{\sin(g_y a_y)}{g_y a_y}. \tag{A7b}$$

APPENDIX B: DETAILS OF SOME CALCULATIONS PERTAINING TO THE CASE OF *p* POLARIZATION

To simplify notation we will denote (in this Appendix only)

$$1 + \Sigma \equiv S, \quad \rho\chi \equiv \kappa, \tag{B1}$$

so that

$$\eta_\alpha = \epsilon_b \frac{1 + 2\kappa S_{\alpha\alpha}}{1 - \kappa S_{\alpha\alpha}}. \tag{B2}$$

We start by deriving Eq. (39). To this end we write the wave vector of the *p*-polarized wave as $\mathbf{q} = q_x\hat{\mathbf{x}} + q_z\hat{\mathbf{z}}$ (note that $q_y = 0$) and seek a nontrivial solution to Eq. (22). Multiplying Eq. (22) by the nonzero factor $q^2 - k_b^2$ and using Eq. (6), we obtain the following equation:

$$(q^2 - k_b^2)\mathbf{F}_0 - \kappa(2k_b^2 + q^2)S\mathbf{F}_0 + 3\kappa\mathbf{q}(\mathbf{q} \cdot S\mathbf{F}_0) = 0. \tag{B3}$$

We now account for the fact that the tensors Σ and $S = 1 + \Sigma$ are diagonal in the laboratory frame and write

$$(S\mathbf{F}_0)_\alpha = S_{\alpha\alpha}F_{0\alpha}, \quad \alpha = x,y,z, \tag{B4}$$

and

$$\mathbf{q} \cdot S\mathbf{F}_0 = q_x S_{xx}F_{0x} + q_z S_{zz}F_{0z}. \tag{B5}$$

Using this result and projecting Eq. (B3) onto the y axis, we immediately obtain $F_{0y} = 0$. The two remaining Cartesian components of $\mathbf{F}_0$ satisfy a system of two linear equations, which are obtainable by projecting Eq. (B3) onto the x and z axes. These two equations are not linearly independent provided the dispersion relation (38) holds [otherwise, the only solution to Eq. (B3) is trivial]. It is therefore sufficient to consider one of these equations, say, by projecting Eq. (B3)

onto the x axis. The resultant equation is

$$AF_{0x} + BF_{0z} = 0, \tag{B6}$$

where

$$A = (1 - \kappa S_{xx})q^2 + 3\kappa S_{xx}q_x^2 - (1 + 2\kappa S_{xx})k_b^2, \tag{B7a}$$

$$B = 3\kappa S_{zz}q_x q_z. \tag{B7b}$$

We now simplify the expression (B7a) for the coefficient A. Specifically, we substitute into this expression $q^2 = q_z^2 + q_x^2$ and $k_b^2 = \epsilon_b k^2 = \epsilon_b(q_z^2/\eta_x + q_x^2/\eta_z)$, where we have used the dispersion relation (38). This yields

$$A = (1 - \kappa S_{xx})(q_z^2 + q_x^2) + 3\kappa S_{xx}q_x^2$$
$$- \epsilon_b(1 + 2\kappa S_{xx})\left(\frac{q_z^2}{\eta_x} + \frac{q_x^2}{\eta_z}\right). \tag{B8}$$

We now use Eq. (B2) to write the quantities η_x and η_z in Eq. (B8) in terms of S_{xx} and S_{zz}. It can be seen that the terms proportional to q_z^2 cancel and we obtain

$$A = 3\kappa S_{zz}\frac{1 + 2\kappa S_{xx}}{1 + 2\kappa S_{zz}}q_x^2. \tag{B9}$$

We use this result and the expression (B7b) for B to compute

$$\frac{F_{0x}}{F_{0z}} = -\frac{B}{A} = -\frac{1 + 2\kappa S_{zz}}{1 + 2\kappa S_{xx}}\frac{q_z}{q_x}. \tag{B10}$$

Returning to the original notation in Eq. (B1), we obtain Eq. (39).

Next we show how to derive Eq. (84) from Eq. (83). Equation (83) contains the factor

$$R \equiv \frac{[\mathbf{k}_r \times (1 + \Sigma)\mathbf{F}_0] \cdot \hat{\mathbf{y}}}{[\mathbf{k}_i \times (1 + \Sigma)\mathbf{F}_0] \cdot \hat{\mathbf{y}}} = \frac{[\mathbf{k}_r \times S\mathbf{F}_0] \cdot \hat{\mathbf{y}}}{[\mathbf{k}_i \times S\mathbf{F}_0] \cdot \hat{\mathbf{y}}}, \tag{B11}$$

which we will now evaluate. To compute the vector products we note that $\mathbf{k}_i = \hat{\mathbf{x}}k_x + \hat{\mathbf{z}}k_{iz}$, $\mathbf{k}_r = \hat{\mathbf{x}}k_x - \hat{\mathbf{z}}k_{iz}$, and $S\mathbf{F}_0 = \hat{\mathbf{x}}S_{xx}F_{0x} + \hat{\mathbf{z}}S_{zz}F_{0z}$. From this we find

$$R = \frac{k_x S_{zz}F_{0z} + k_{iz}S_{xx}F_{0x}}{k_x S_{zz}F_{0z} - k_{iz}S_{xx}F_{0x}}. \tag{B12}$$

Next we use the ratio F_{0x}/F_{0z} given by Eq. (B10), account for the conservation of the wave vector projection onto the interface, that is, $q_x = k_x$, and rewrite Eq. (B12) as

$$R = \frac{k_x^2 S_{zz}(1 + 2\kappa S_{xx}) - k_{iz}q_z S_{xx}(1 + 2\kappa S_{zz})}{k_x^2 S_{zz}(1 + 2\kappa S_{xx}) + k_{iz}q_z S_{xx}(1 + 2\kappa S_{zz})}. \tag{B13}$$

To proceed we need to exclude the variable k_x^2 from Eq. (B13). Using the dispersion relations (38) and (52) for the refracted and the incident waves (in the geometry considered, $q_x^2 = k_\perp^2 = k_x^2$), we write

$$\frac{q_z^2}{\eta_x} + \frac{k_x^2}{\eta_x} = k^2 = \frac{1}{\epsilon_b}k_b^2 = \frac{1}{\epsilon_b}(k_x^2 + k_{iz}^2). \tag{B14}$$

Solving Eq. (B14) for k_x^2, we obtain

$$k_x^2 = \frac{k_{iz}^2/\epsilon_b - q_z^2/\eta_x}{1/\eta_z - 1/\epsilon_b} = \frac{1 + 2\kappa S_{zz}}{3\kappa S_{zz}}\left(\frac{1 - \kappa S_{xx}}{1 + 2\kappa S_{xx}}q_z^2 - k_{iz}^2\right), \tag{B15}$$

where we have used Eq. (B2) to obtain the second expression from the first. We now substitute the result given in Eq. (B15)

into Eq. (B13). The factors of $1 + 2\kappa S_{zz}$ in the numerator and the denominator cancel and we obtain

$$R = \frac{(1 - \kappa S_{xx})q_z^2 - (1 + 2\kappa S_{xx})k_{iz}^2 - 3\kappa S_{xx}k_{iz}q_z}{(1 - \kappa S_{xx})q_z^2 - (1 + 2\kappa S_{xx})k_{iz}^2 + 3\kappa S_{xx}k_{iz}q_z}. \quad (B16)$$

At the next step we divide the numerator and the denominator in Eq. (B16) by the factor $1 + 2\kappa S_{xx}$ and, accounting for the identity

$$\frac{3\kappa S_{xx}}{1 + 2\kappa S_{xx}} = \epsilon_b \left(\frac{1}{\epsilon_b} - \frac{1}{\eta_x} \right), \quad (B17)$$

obtain

$$R = \frac{\frac{q_z^2}{\eta_x} - \frac{k_{iz}^2}{\epsilon_b} - \left(\frac{1}{\epsilon_b} - \frac{1}{\eta_x} \right)k_{iz}q_z}{\frac{q_z^2}{\eta_x} - \frac{k_{iz}^2}{\epsilon_b} + \left(\frac{1}{\epsilon_b} - \frac{1}{\eta_x} \right)k_{iz}q_z}. \quad (B18)$$

The expressions in the numerator and denominator can now be factorized and we arrive at the final result

$$R = -\frac{(q_z + k_{iz})\left(\frac{k_{iz}}{\epsilon_b} - \frac{q_z}{\eta_x} \right)}{(q_z - k_{iz})\left(\frac{k_{iz}}{\epsilon_b} + \frac{q_z}{\eta_x} \right)}. \quad (B19)$$

Substitution of this expression into Eq. (83) immediately results in Eq. (84).

APPENDIX C: PROOF OF THEOREM 1

1. An equivalence transformation

To derive the equality (110) we first introduce some notation. Let

$$\varepsilon \equiv \langle \phi | \psi \rangle, \quad (C1a)$$

$$P \equiv |\psi\rangle\langle\phi|, \quad (C1b)$$

$$R(\mathcal{Z}; A) \equiv (\mathcal{Z} - A)^{-1}, \quad (C1c)$$

$$B \equiv R(\mathcal{Z}; W)W, \quad (C1d)$$

$$\sigma \equiv \langle \phi | R(\mathcal{Z}; W) | \psi \rangle. \quad (C1e)$$

Here $R(\mathcal{Z}; A)$ is the resolvent of the linear operator A and $\mathcal{Z}$ is a complex number. In the new notation the operator T defined in Eq. (111) takes the form

$$T = 1 - \frac{1}{\varepsilon}P \quad (C2)$$

and Eq. (110) is rewritten as

$$\sigma = \frac{1}{\mathcal{Z}} \frac{\varepsilon}{1 - \frac{1}{\varepsilon}\langle\phi|R(\mathcal{Z}; WT)W|\psi\rangle}. \quad (C3)$$

Note that, by the first hypothesis of Theorem 1, $\varepsilon \neq 0$.

We now write the following chain of equalities in which the second hypothesis of Theorem 1, namely, that $R(\mathcal{Z}; W)$ exists, has been used:

$$R(\mathcal{Z}; WT) = (\mathcal{Z} - WT)^{-1} = \left(\mathcal{Z} - W + \frac{1}{\varepsilon}WP \right)^{-1}$$

$$= \left(R^{-1}(\mathcal{Z}; W) + \frac{1}{\varepsilon}WP \right)^{-1}$$

$$= \left[R^{-1}(\mathcal{Z}; W)\left(1 + \frac{1}{\varepsilon}R(\mathcal{Z}; W)WP \right) \right]^{-1}$$

$$= \varepsilon[\varepsilon + R(\mathcal{Z}; W)WP]^{-1}R(\mathcal{Z}; W). \quad (C4)$$

Using the last equality in Eq. (C4) and the notation in Eq. (C1d), we rewrite Eq. (C3) identically as

$$\sigma = \frac{1}{\mathcal{Z}} \frac{\varepsilon}{1 - \langle\phi|(\varepsilon + BP)^{-1}B|\psi\rangle}. \quad (C5)$$

2. A useful identity

Below we will frequently use the following identity:

$$\langle\phi|B|\psi\rangle = \mathcal{Z}\sigma - \varepsilon. \quad (C6)$$

The above equation is easily derived by noting that

$$\langle\phi|B|\psi\rangle = \langle\phi|(\mathcal{Z} - W)^{-1}W|\psi\rangle$$
$$= \langle\phi|(\mathcal{Z} - W)^{-1}(W - \mathcal{Z})|\psi\rangle + \mathcal{Z}\langle\phi|(\mathcal{Z} - W)^{-1}|\psi\rangle$$
$$= -\varepsilon + \mathcal{Z}\sigma. \quad (C7a)$$

3. Main derivation

To proceed we need to express the operator $(\varepsilon + BP)^{-1}$, which appears on the right-hand side of Eq. (C5), in a more tractable form. To this end consider the equation

$$(\varepsilon + BP)|x\rangle = |b\rangle, \quad (C8)$$

where $|x\rangle$ is viewed as the unknown and $|b\rangle \neq 0$ is an otherwise arbitrary element of the same Hilbert space. Using the definition of P in Eq. (C1b), we transform Eq. (C8) to

$$\varepsilon|x\rangle + B|\psi\rangle\langle\phi|x\rangle = |b\rangle, \quad (C9)$$

project the result onto $|\phi\rangle$, and find that

$$\langle\phi|x\rangle = \frac{\langle\phi|b\rangle}{\varepsilon + \langle\phi|B|\psi\rangle}. \quad (C10)$$

We now use the previously derived identity (C6) on the right-hand side of Eq. (C10) to obtain

$$\langle\phi|x\rangle = \frac{\langle\phi|b\rangle}{\mathcal{Z}\sigma}. \quad (C11)$$

Upon substitution of Eq. (C11) into Eq. (C9) we find the solution to Eq. (C8) or (C9), namely,

$$|x\rangle = \frac{1}{\varepsilon}\left(1 - \frac{B|\psi\rangle\langle\phi|}{\mathcal{Z}\sigma} \right)|b\rangle. \quad (C12)$$

Since the vector $|b\rangle$ in Eq. (C8) is arbitrary, we conclude that

$$(\varepsilon + BP)^{-1} = \frac{1}{\varepsilon}\left(1 - \frac{B|\psi\rangle\langle\phi|}{\mathcal{Z}\sigma} \right). \quad (C13)$$

This equality can be verified directly by substitution.

4. Putting everything together

We can now put everything together and obtain Eq. (C5). From Eq. (C13) we have

$$\langle\phi|(\varepsilon + BP)^{-1} = \frac{1}{\varepsilon}\left(1 - \frac{\langle\phi|B|\psi\rangle}{\mathcal{Z}\sigma}\right)\langle\phi| = \frac{\langle\phi|}{\mathcal{Z}\sigma}, \quad \text{(C14)}$$

where we have again used Eq. (C6). Now we can write

$$\langle\phi|(\varepsilon + BP)^{-1}B|\psi\rangle = \frac{\langle\phi|B|\psi\rangle}{\mathcal{Z}\sigma} = 1 - \frac{\varepsilon}{\mathcal{Z}\sigma}. \quad \text{(C15)}$$

Upon substitution of this result into the right-hand side of Eq. (C5), we find that the latter is indeed an identity and so are Eqs. (C3) and (110).

[1] C. R. Simovski, Opt. Spectrosc. **107**, 726 (2009).

[2] C. R. Simovski, J. Opt. **13**, 103001 (2011).

[3] A. Bensoussan, J. L. Lions, and G. Papanicolaou, *Asymptotic Analysis for Periodic Structures* (North-Holland, Amsterdam, 1978).

[4] O. A. Oleinik, A. S. Shamaev, and G. A. Yosifian, *Mathematical Problems in Elasiticity and Homogenization* (Elsevier, Amsterdam, 1992).

[5] G. W. Milton, *The Theory of Composites* (Cambridge University Press, Cambridge, 2002).

[6] L. Tartar, *The General Theory of Homogenization* (Springer, Heidelberg, 2009).

[7] M. G. Silveirinha, Phys. Rev. B **75**, 115104 (2007).

[8] I. Tsukerman, J. Opt. Soc. Am. B **28**, 577 (2011).

[9] A. Pors, I. Tsukerman, and S. I. Bozhevolnyi, Phys. Rev. E **84**, 016609 (2011).

[10] M. G. Silveirinha, Phys. Rev. B **83**, 165104 (2011).

[11] L. Feng, Z. Liu, V. Lomakin, and Y. Fainman, Appl. Phys. Lett. **96**, 041112 (2010).

[12] L. Feng, Z. Liu, and Y. Fainman, Appl. Opt. **50**, G1 (2011).

[13] L. Feng, A. Mizrahi, S. Zamek, Z. Liu, V. Lomakin, and Y. Fainman, ACS NANO **5**, 5100 (2011).

[14] A. A. Krokhin, P. Halevi, and J. Arriaga, Phys. Rev. B **65**, 115208 (2002).

[15] A. A. Krokhin and E. Reyes, Phys. Rev. Lett. **93**, 023904 (2004).

[16] K. D. Cherednichenko and S. Guenneau, Waves Random Media **17**, 627 (2007).

[17] S. Guenneau, F. Zolla, and A. Nicolet, Waves Random Media **17**, 653 (2007).

[18] S. Guenneau and F. Zolla, Prog. Electromagn. Res. **27**, 91 (2011).

[19] R. V. Craster, J. Kaplunov, E. Nolde, and S. Guenneau, J. Opt. Soc. Am. A **28**, 1032 (2011).

[20] C. F. Bohren, J. Nanophoton. **3**, 039501 (2009).

[21] G. A. Niklasson, C. G. Granqvist, and O. Hunderi, Appl. Opt. **20**, 26 (1981).

[22] W. T. Doyle, Phys. Rev. B **39**, 9852 (1989).

[23] N. A. Nicorovici, R. C. McPhedran, and L. C. Botten, Phys. Rev. Lett. **75**, 1507 (1995).

[24] N. A. Nicorovici, R. C. McPhedran, and L. C. Botten, Phys. Rev. E **52**, 1135 (1995).

[25] N. Wellander and G. Kristensson, SIAM J. Appl. Math. **64**, 170 (2003).

[26] D. Felbacq and G. Bouchitte, New J. Phys. **7**, 159 (2005).

[27] C. Poulton, S. Guenneau, and A. B. Movchan, Phys. Rev. B **69**, 195112 (2004).

[28] C. F. Bohren, J. Atmos. Sci. **43**, 468 (1986).

[29] C. Menzel, T. Paul, C. Rockstuhl, T. Pertsch, S. Tretyakov, and F. Lederer, Phys. Rev. B **81**, 035320 (2010).

[30] C. Menzel, C. Rockstuhl, R. Iliew, F. Lederer, A. Andryieuski, R. Malureanu, and A. V. Lavrinenko, Phys. Rev. B **81**, 195123 (2010).

[31] C. R. Simovski and S. A. Tretyakov, Photon. Nanostruct. **8**, 254 (2010).

[32] A. Andryieuski, C. Menzel, C. Rockstuhl, R. Malureanu, F. Lederer, and A. Lavrinenko, Phys. Rev. B **82**, 235107 (2010).

[33] T. Paul, C. Menzel, W. Smigaj, C. Rockstuhl, P. Lalanne, and F. Lederer, Phys. Rev. B **84**, 115142 (2011).

[34] J. E. Sipe and J. Van Kranendonk, Phys. Rev. A **9**, 1806 (1974).

[35] B. T. Draine and J. Goodman, Astrophys. J. **405**, 685 (1993).

[36] P. A. Belov and C. R. Simovski, Phys. Rev. E **72**, 026615 (2005).

[37] D. J. Bergman, Phys. Rep. **43**, 377 (1978).

[38] D. J. Bergman, J. Phys. C: Solid State Phys. **12**, 4947 (1979).

[39] D. J. Bergman, Phys. Rev. B **19**, 2359 (1979).

[40] F. J. G. Abajo, Rev. Mod. Phys. **79**, 1267 (2007).

[41] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, Oxford, 1984).

[42] Y. Itin, Phys. Lett. A **374**, 1113 (2010).

[43] V. A. Markel and J. C. Schotland, J. Opt. **12**, 015104 (2010).

[44] A. A. Maradudin and D. L. Mills, Phys. Rev. B **11**, 1392 (1975).

[45] G. D. Mahan and G. Obermair, Phys. Rev. **183**, 834 (1969).

[46] W. Lamb, D. M. Wood, and N. W. Ashcroft, Phys. Rev. B **21**, 2248 (1980).

[47] B. T. Draine, Astrophys. J. **333**, 848 (1988).

[48] V. A. Markel, J. Mod. Opt. **39**, 853 (1992).

[49] V. A. Markel, J. Opt. Soc. Am. B **12**, 1783 (1995).

[50] M. V. Berry and I. C. Percival, Opt. Acta **33**, 577 (1986).

[51] R. Haydock, *Solid State Physics*, Vol. 35 (Academic, New York, 1980), pp. 215–294.

[52] V. A. Markel, V. N. Pustovit, S. V. Karpov, A. V. Obuschenko, V. S. Gerasimov, and I. L. Isaev, Phys. Rev. B **70**, 054202 (2004).

[53] W. B. Jones and W. J. Thron, *Continued Fractions. Analytic Theory and Applications* (Addison-Wesley, Reading, MA, 1980).

[54] M. I. Stockman, Phys. Rev. Lett. **98**, 177404 (2007).

[55] S. J. Elston, G. P. Bryan-Brown, and J. R. Sambles, Phys. Rev. B **44**, 6393 (1991).