Applicability of effective medium description to photonic crystals in higher bands: Theory and numerical analysis

Vadim A. Markel and Igor Tsukerman

Can photonic crystals be homogenized in higher bands?

Vadim A. Markel[*]
Aix-Marseille Université, CNRS, Centrale Marseille,
Institut Fresnel UMR 7249, 13013 Marseille, France

Igor Tsukerman
Department of Electrical and Computer Engineering,
The University of Akron, OH 44325-3904, USA[†]
(Dated: May 9, 2016)

We consider conditions under which photonic crystals (PCs) can be homogenized in the higher photonic bands and, in particular, near the Γ-point. By homogenization we mean introducing some effective local parameters ϵ_{eff} and μ_{eff} that describe reflection, refraction and propagation of electromagnetic waves in the PC adequately. The parameters ϵ_{eff} and μ_{eff} can be associated with a hypothetical homogeneous effective medium. In particular, if the PC is homogenizable, the dispersion relations and isofrequency lines in the effective medium and in the PC should coincide to some level of approximation. We can view this requirement as a necessary condition of homogenizability. In the vicinity of a Γ-point, real isofrequency lines of two-dimensional PCs can be close to mathematical circles, just like in the case of isotropic homogeneous materials. Thus, one may be tempted to conclude that introduction of an effective medium is possible and, at least, the necessary condition of homogenizability holds in this case. We, however, show that this conclusion is incorrect: complex dispersion points must be included into consideration even in the case of strictly non-absorbing materials. By analyzing the complex dispersion relations and the corresponding isofrequency lines, we have found that two-dimensional PCs with C_4 and C_6 symmetries are not homogenizable in the higher photonic bands. We also draw a distinction between spurious Γ-point frequencies that are due to Brillouin-zone folding of Bloch bands and "true" Γ-point frequencies that are due to multiple scattering. Understanding of the physically different phenomena that lead to the appearance of spurious and "true" Γ-point frequencies is important for the theory of homogenization.

I. INTRODUCTION

The theory of homogenization of periodic electromagnetic media continues to attract significant attention[1–5] due to the proposed important applications such as sub-wavelength optical imaging[6]. However, some fundamental questions remain open in this field of research. Perhaps the most important of these questions is the following: to what extent can the "exotic" effective parameters obtained via one of the several recently proposed homogenization theories be used without restriction, like the constitutive parameters of homogeneous natural media are conventionally used? Indeed, while it is generally understood that homogenization is an approximate procedure, the accuracy and the applicability range of a given theory is very difficult to ascertain quantitatively.

In this paper, we investigate whether a photonic crystal (PC) can be homogenized at frequencies above the first bandgap and, particularly, near the Γ-point. Answering this question is important for the following reason. It is well known that PCs can be characterized by negative dispersion in the higher photonic bands even if the material from which the PC is made has no absorption. On the other hand, in purely dielectric homogeneous media, negative dispersion is obtained only within the absorption bands. A homogeneous material can be simultaneously transparent and characterized by negative dispersion only if it has a nontrivial magnetic permeability[7,8]. Therefore, to obtain this result by homogenizing an intrinsically non-magnetic PC, one has to consider sufficiently large frequencies where the dispersion is negative[9]. Typically, these frequencies are above the first bandgap of the crystal. We note that some intuitive arguments exist suggesting that a PC can be homogenized sufficiently close to the the Γ-point frequencies (that is, the frequencies for which the Bloch wave number vanishes). However, a careful consideration reveals that the first of these frequencies, $\omega_1 = 0$, is fundamentally different from the higher ones $\omega_n > 0$ ($n = 2, 3, \ldots$). While homogenization can be arbitrarily accurate near ω_1, the same is not true near any ω_n with $n > 1$.

The above conclusion is consistent with some of the previous numerical investigations of the homogenization problem[10–12]. In fact, the work reported here is conceptually close to Ref. 10, and one of the main ideas on which the present paper is based has been stated in that reference. Namely, it was noticed that the right-hand side of the dispersion equation $\omega = f(\mathbf{q})$ can be expanded in powers of the Cartesian components of $\mathbf{q}$ if ω is close to one of the Γ-point frequencies ω_n. Here $\mathbf{q}$ is the Bloch wave vector. Further, for s-polarized waves in two-dimensional PCs with a center of symmetry, this expansion is of the form $f(\mathbf{q}) = \omega_n + \beta_x q_x^2 + \beta_y q_y^2 \ldots$, or in the cases of C_4 and C_6 symmetries that are considered in this paper, $f(\mathbf{q}) = \omega_n + \beta q^2 \ldots$ Here we have written explicitly the first two non-vanishing terms of the expansion and ω is assumed to be in the pass-band. By truncating the latter expansion at the second order, we obtain the isotropic isofrequency line $q^2 = (\omega - \omega_n)/\beta$ so that the wave number q does not depend on the direction of prop-

agation. Another observation made in Ref. 10 was that, for incident waves with a sufficiently large projection of the wave vector onto the interface (which includes but is not limited to evanescent waves), higher-order terms must be retained in the expansion of $f(\mathbf{q})$ no matter how close the frequency is to a Γ-point frequency, and that the resultant law of dispersion is no longer isotropic.

Here we develop this basic idea of Ref. 10 theoretically and illustrate it with numerical examples of a different kind. While this reference is focused on the transmission and imaging properties of a PC slab, here we consider in detail the dispersion relations and isofrequency surfaces. We work in a 2D geometry and in s-polarization (one-component electric field) so that Maxwell's equations are reduced to a scalar wave equation. Additionally, the medium will have either C_4 or C_6 symmetry (square or triangular lattice). Under these conditions, only scalar effective parameters can be introduced[9]. The corresponding effective medium is isotropic and so are its isofrequency surfaces. However, we will show that, in the actual PC, the isotropy of isofrequency surfaces is lost in the higher photonic bands once we include waves with sufficiently large projections of the wave vector onto a given axis. Therefore the PC can not be described by local effective parameters.

Thus, we will view two-dimensional PCs with C_4 or C_6 symmetry whose isofrequency surfaces are not isotropic as not homogenizable[13]. We note that the lack of isotropy can be confused with nonlocality (spatial dispersion). Indeed, nonlocality of material parameters and anisotropy of *local* parameters can result in somewhat similar phenomena. But they are not the same effect and should be distinguished. The distinction is especially evident when boundary value problems are considered. For example, scattering from an anisotropic sphere is an analytically-solvable problem of mathematical physics[14,15]. However, solving the same problem for a nonlocal sphere is much more complicated and will, generally, yield a different result.

To complicate things further, it is frequently stated that there exists a complete physical equivalence between two alternative descriptions of the electromagnetic properties of continuous media[16–19]. In one description the medium is assigned a nonlocal permittivity $\epsilon(\omega, \mathbf{q})$ and a trivial permeability $\mu = 1$. This is the so-called Landau-Lifshitz approach[19]. In the other description, the medium is assigned two local parameters $\epsilon(\omega)$ and $\mu(\omega)$. We have previously argued that the two descriptions are not physically equivalent in general[20], but it is true that, in the so-called weak nonlocality regime[21], such an equivalence exists for the refractive index of the medium (but not for the impedance).

For the purpose of this paper, it is unimportant whether the equivalence mentioned above exists or not. If it does exist, then a homogenizable heterogeneous medium should be characterized by some local parameters $\epsilon_{\text{eff}}(\omega)$ and $\mu_{\text{eff}}(\omega)$. Whether these local parameters correspond to some effective nonlocal parameter $\epsilon_{\text{eff}}(\omega, \mathbf{q})$ and $\mu_{\text{eff}} = 1$ in the alternative description is irrelevant. On the other hand, if such local parameters do not exist, then the medium can not be reasonably homogenized and the introduction of the nonlocal permittivity does not solve the problem. Indeed, the knowledge of $\epsilon_{\text{eff}}(\omega, \mathbf{q})$ for all values of its arguments is not sufficient to solve any boundary value problems in a finite sample[20], unlike the knowledge of the local parameters $\epsilon_{\text{eff}}(\omega)$ and $\mu_{\text{eff}}(\omega)$. Besides, the typical applications discussed in the literature such as subwavelength imaging require local $\epsilon_{\text{eff}}(\omega)$ and $\mu_{\text{eff}}(\omega)$. Therefore, we say that, in order for a PC to be homogenizable, its dispersion relation must be (at least, approximately) the same as in a hypothetical homogeneous effective medium with some local parameters $\epsilon_{\text{eff}}(\omega)$ and $\mu_{\text{eff}}(\omega)$. We emphasize that the above condition is necessary but, in general, not sufficient because it does not include the impedance. But we will show that even this necessary condition of homogenizability does not hold in PCs above the first bandgap.

We illustrate the theoretical arguments of this paper with numerical examples using a rather simple but physically relevant model. As was mentioned above, we consider two-dimensional PCs (hollow inclusions in a high-index host) with a one-component electric field polarized perpendicularly to the plane of periodicity. Such PCs have been previously considered in the literature as homogenizable in the higher photonic bands[22–24]. Similarly to these works, we neglect in the majority of cases frequency dispersion and absorption in the host material. This is done not for computational convenience (our codes can handle the more general case with equal efficiency) but rather to analyze the exact cases that were previously considered in the literature.

However, to illustrate the effects of absorption, we have also performed simulations for a square-lattice PC with a dispersive and absorptive host. In this case, the higher Γ-point frequencies can be defined only approximately and the case for homogenizability is even harder to make. Since no qualitatively new phenomena emerge in an absorptive PC, at least, as far as its homogenezability is concerned, we have restricted simulations of triangular-lattice PCs to the case of non-absorbing host.

We finally note that our results can be understood in a more general framework of the *uncertainty principle* of homogenization[25]. According to this principle, the larger the deviation of the effective magnetic permeability from unity (according to a given theory), the less accurate this theory is in predicting physical observables such as the transmission and reflection coefficients of a composite slab.

The remainder of this paper is organized as follows. We start with some general theoretical considerations relevant to the problem at hand in Sec. II where we explain why the circularity (or sphericity) of a real isofrequency line is not a sufficient condition for homogenizability. Our computational formulas are written down in Sec. III. Extensive numerical examples for 2D square- and triangular-lattice PCs are adduced in Sec. IV. Sec. V

contains a discussion of the results obtained. The two appendices contain mathematical details pertinent to the law of dispersion in 2D periodic structures. These developments are important for clarification of several subtle points such as classification of the Γ-point frequencies. Appendix B also gives some rather involved mathematical formulas that can be used in a numerical implementation of the perturbation theory in which the Bloch wave number is viewed as the expansion parameter.

II. GENERAL CONSIDERATIONS

For Bloch waves with a nonzero fundamental harmonic (as defined in the Appendices) propagating in three-dimensional, intrinsically non-magnetic PCs, the dispersion equation can be written in the following general form[20,26]

$$\det\left[(\mathbf{q}\times\mathbf{q}\times) + k^2\Sigma(\omega,\mathbf{q})\right] = 0 \ . \tag{1}$$

Here $k = \omega/c$ is the free-space wave number and $\Sigma(\omega,\mathbf{q})$ is a 3D tensor, which is completely determined by the PC geometry and composition and by the two arguments ω and $\mathbf{q}$. The latter can be considered as mathematically-independent variables and take arbitrary complex values. We will say that a complex vector $\mathbf{q}$ is the Bloch wave vector of a PC at some frequency ω if the pair $(\omega,\mathbf{q})$ satisfies (1).

Below we work in the frequency domain and restrict the frequency to be real and positive. The set of all $\mathbf{q}$'s that satisfy (1) for a given $\omega > 0$ forms an isofrequency surface in 3D or a line in 2D. However, the words "surface" and "line" should not be understood literally because $\mathbf{q}$ is in general complex. For example, in the 2D geometry that we consider below, the set of complex $\mathbf{q}$'s that satisfy (1) for some $\omega > 0$ is a four-dimensional manifold. The isofrequency lines that are commonly displayed are the intersections of this manifold and various two-dimensional sub-spaces.

The function $\Sigma(\omega,\mathbf{q})$ arises in various physical contexts[27–29] and is sometimes interpreted as the nonlocal permittivity tensor of the medium due to its explicit dependence on $\mathbf{q}$. We have shown[20] that the knowledge of $\Sigma(\omega,\mathbf{q})$ for all values of its arguments is insufficient for solving boundary value problems in finite samples. However, in this paper, we restrict attention to dispersion relations, and to this end the knowledge of $\Sigma(\omega,\mathbf{q})$ is sufficient.

In homogeneous nonlocal media, (1) is obtained in a straightforward manner by the spatial Fourier transform of the nonlocal susceptibility function. In heterogeneous periodic media, the derivation is more involved. Sometimes (1) is derived for such media by employing an external excitation in the form of an "impressed" current that overlaps with the medium and has the mathematical form of a plane wave[27–29]. However, this approach is not necessary and introduction of such currents into the model is not physically justified[30]. Previously, we have derived (1) and defined $\Sigma(\omega,\mathbf{q})$ for general three-dimensional PCs without appealing to the concept of "impressed" currents[20,26]. Below, in Appendix A, we present a basis-free derivation of (1) for a 2-dimensional PC. In Appendix B, we repeat the derivation using the basis of plane waves, which yields expressions that are directly amenable to numerical computation.

We can use (1) to formally define the Γ-point frequencies. As one could expect, the definition is confounded by the vector nature of electromagnetic fields. In particular, the Γ-point frequencies can be polarization-dependent. However, the simulations discussed below have been performed for the special case of transverse Bloch waves, which satisfy the condition $\mathbf{q}\cdot\mathbf{E}_0 = 0$, where $\mathbf{E}_0$ is the amplitude of the fundamental harmonic of the Bloch wave for a given linear polarization state[26]. In this case, (1) simplifies to

$$q^2 = k^2\Sigma(\omega,\mathbf{q}) \ , \tag{2}$$

where $\mathbf{q}$ is now a two-dimensional vector orthogonal to $\mathbf{E}_0$ and $\Sigma(\omega,\mathbf{q})$ is a scalar (a principal value of the tensor Σ that corresponds to the direction of $\mathbf{E}_0$). Note that $\mathbf{E}_0$ can be an eigenvector of $\Sigma(\omega,\mathbf{q})$ for all ω and $\mathbf{q}$, typically, as a consequence of the PC symmetry. Also, we use the same notation Σ for the tensor and for its principal values, but this should not cause confusion since only the latter interpretation will be used below.

In what follows, we will consider only the scalar equation (2) and assume that the two-dimensional Bloch vector $\mathbf{q} = (q_x, q_y)$ lies in the XY plane of a rectangular frame while $\mathbf{E}_0 = E_0\hat{\mathbf{z}}$ is collinear with the Z-axis[31]. By focusing on the special case of transverse waves, we do not disregard any important effects but rather focus on the essential features of the theory.

We now proceed with the analysis of Eq. (2). We will say that a frequency ω_n is the n-th Γ-point frequency of a PC if the following conditions hold:

(i) $\omega_n^2\Sigma(\omega_n,0) = 0$; $\qquad\qquad$ (3a)

(ii) $\Sigma(\omega,\mathbf{q})$ is an analytic function of $\qquad$ (3b)
$\qquad$ ω and $\mathbf{q}$ in the vicinity of $\omega = \omega_n, \mathbf{q} = 0$

The first trivial solution to (3a) is $\omega_1 = 0$, unless $\Sigma(\omega,0)$ diverges as $1/\omega^2$ or faster at $\omega = 0$. This possibility can be safely ignored and the first (fundamental) Γ-point frequency exists in all PCs, even if they are made of conducting constituents. Also, $\Sigma(0,\mathbf{q})$ is usually analytic near $\mathbf{q} = 0$ so the condition (3b) is also satisfied for ω_1. We do not have a proof of this statement but will see that all counterexamples to (3b) occur at $\omega_n > 0$.

The higher Γ-point frequencies can be determined from the equation

$$\Sigma(\omega_n, 0) = 0 \ , \quad \omega_n > 0 \ . \tag{4}$$

The condition (3b) should also hold. It therefore can be seen that the nature of the first and the higher Γ-point frequencies is different. At the first frequency,

$\Sigma(\omega_1, 0) = \Sigma(0, 0)$ does not turn to zero; in fact, it can be readily seen that $\Sigma(0, 0) = n_{\mathrm{eff}}^2$, where n_{eff} is the effective refractive index of the medium in the homogenization limit. But for the higher frequencies $\omega_n \neq 0$, we have $\Sigma(\omega_n, 0) = 0$. This equality can not hold in any truly homogeneous medium (see Appendix A for more details).

At this point, a few remarks are in order. First, purely real Γ-point frequencies of order higher than 1 do not generally exist in PCs with non-negligible absorption. This is so because (4) is in this case a complex equation: both real and imaginary parts of (4) must be satisfied simultaneously. This is unlikely to happen at a purely real ω. However, if, as is frequently done in the PC literature, we assume that the dielectric permittivity of the PC constituents is purely real, then (4) can be expected to have real roots.

Second, derivation of equations (1) or (2) does not require that $\mathrm{Re}(\mathbf{q})$ be in the first Brillouin zone (FBZ) of the lattice. However, if $\mathbf{q}$ is a solution to one of these equations, then $\mathbf{q} + \mathbf{g}$ is also a solution, where $\mathbf{g}$ is any reciprocal lattice vector. Therefore, it is sufficient to consider only the solutions with $\mathrm{Re}(\mathbf{q}) \in \mathrm{FBZ}$.

Third, and related to the above, homogeneous media do not possess higher-order Γ-point frequencies in the sense of definition (3). It is true that the "artificially folded" dispersion curve of a homogeneous medium, such as the one shown in figure 2 of Ref. 32, crosses the vertical axis at some frequencies ω_n, and one can conclude that at $\omega = \omega_n$ we also have $\mathbf{q} = 0$. However, condition (3b) is not satisfied at these frequencies. This point is illustrated in Fig. 1 where we plot the purely real isofrequency lines for a homogeneous medium that is artificially discretized on two-dimensional square and triangular lattices. It can be seen that the isofrequency lines are very far from circular and become quasi-chaotic for large band indexes n. This behavior distinguishes these pseudo-Γ-point frequencies from the true ones, which can exist in PCs due to strong multiple scattering. For a more detailed analysis, see Appendix A.

Returning to the question of homogenizability, we can now see why one might be tempted think that the medium to be homogenizable in the vicinity of a true Γ-point frequency. Let ω be in the transmission band of a PC but close to a Γ-point frequency ω_n ($n > 1$), say, slightly below ω_n. Let us fix the frequency and expand $\Sigma(\omega, \mathbf{q})$ in powers of $\mathbf{q}$. This is possible because of (3b). If, as is often the case, the PC has a center of symmetry, the expansion is to lowest order in $\mathbf{q}$ of the form

$$\Sigma(\omega, \mathbf{q}) = \alpha + \beta_x \frac{q_x^2}{k^2} + \beta_y \frac{q_y^2}{k^2} + \ldots, \quad \alpha \equiv \Sigma(\omega, 0) \ll 1. \quad (5)$$

In the above equation, ω is a fixed parameter and therefore the explicit dependence of the expansion coefficients α, β on ω is suppressed. Also, it is easy to see that in PCs obeying C_4 or C_6 symmetry, $\beta_x = \beta_y = \beta$. If we now treat the first two terms in the expansion (5) as an approximation, write $\Sigma(\omega, \mathbf{q}) \approx \alpha + \beta q^2/k^2$ and substi-

tute the result into the dispersion equation (2), we will obtain an approximate solution of the form

$$q^2 = k^2 \frac{\alpha}{1 - \beta}. \quad (6)$$

We remark briefly that it is rather typical to interpret α as the effective permittivity of the medium, ϵ_{eff}, and $1/(1 - \beta)$ as the effective permeability, μ_{eff}. This approach is discussed in detail in Ref. 20. Since we consider only dispersion relations in infinite media in this paper (a connection to the problem of transmission and reflection of waves is made by observing that the tangential component of the Bloch wave vector is conserved at planar interfaces), the breakdown of the squared refractive index n_{eff}^2 into the product of ϵ_{eff} and μ_{eff} is irrelevant. What is important for our purposes is that (6) describes a perfectly circular isofrequency line. Of course, circular isofrequency lines are also characteristic of homogeneous materials. The conclusion is then drawn that sufficiently close to a Γ-point frequency, a PC is indistinguishable from a homogeneous medium.

However, the above line of argument has the following deficiency. It is not really true that all Cartesian components of $\mathbf{q}$ must be small near a higher-order Γ-point frequency. A more precise statement is that the scalar $q^2 = \mathbf{q} \cdot \mathbf{q}$ is small. The Cartesian components of $\mathbf{q}$ can still be arbitrarily large for complex $\mathbf{q}$. An obvious example is the vector $\mathbf{q} = (p, ip)$ where p is a real number.

In Appendix B, we develop a perturbation theory for $\Sigma(\omega, \mathbf{q})$. The expansion is obtained for two-dimensional PCs with either C_4 or C_6 symmetry in terms of the Cartesian components of $\mathbf{q}$ for ω in the vicinity of one of the Γ-point frequencies ω_n as defined in (3). The result is a special case of the more general expansion (5) (which does not assume any special symmetry) and is of the form

$$\Sigma(\omega, \mathbf{q}) = \alpha + \beta_2 q^2 + \beta_4 q^4 + \beta_6 q^6 + \beta_8 q^8 + \gamma_4 q_x^2 q_y^2$$
$$+ \left[\delta_{6x} q_x^2 \left(q_x^4 - 6 q_x^2 q_y^2 + 9 q_y^4 \right) \right.$$
$$+ \left. \delta_{6y} q_y^2 \left(q_y^4 - 6 q_x^2 q_y^2 + 9 q_x^4 \right) \right] + \ldots \quad (7)$$

Here $q^4 = (q_x^2 + q_y^2)^2$, etc. It can be seen that the expansion terms with the coefficients β_{2n} are all circularly-symmetric. The first term that breaks the circular symmetry is

$$q_x^2 q_y^2. \quad (8)$$

We can say that, starting from fourth order in $\mathbf{q}$, the function $\Sigma(\omega, \mathbf{q})$ starts to bear the traces of the underlying lattice, which is not circularly-symmetric.

However, the coefficient γ_4 is zero in the case of C_6 symmetry of the structure. This is so because expression (8) is not invariant with respect to rotation by $\pi/3$. The simplest anisotropic terms that are invariants of C_6 arise in sixth order in $\mathbf{q}$ and are of the form

$$q_x^2 \left(q_x^4 - 6 q_x^2 q_y^2 + 9 q_y^4 \right), \quad (9a)$$
$$q_y^2 \left(q_y^4 - 6 q_x^2 q_y^2 + 9 q_x^4 \right). \quad (9b)$$

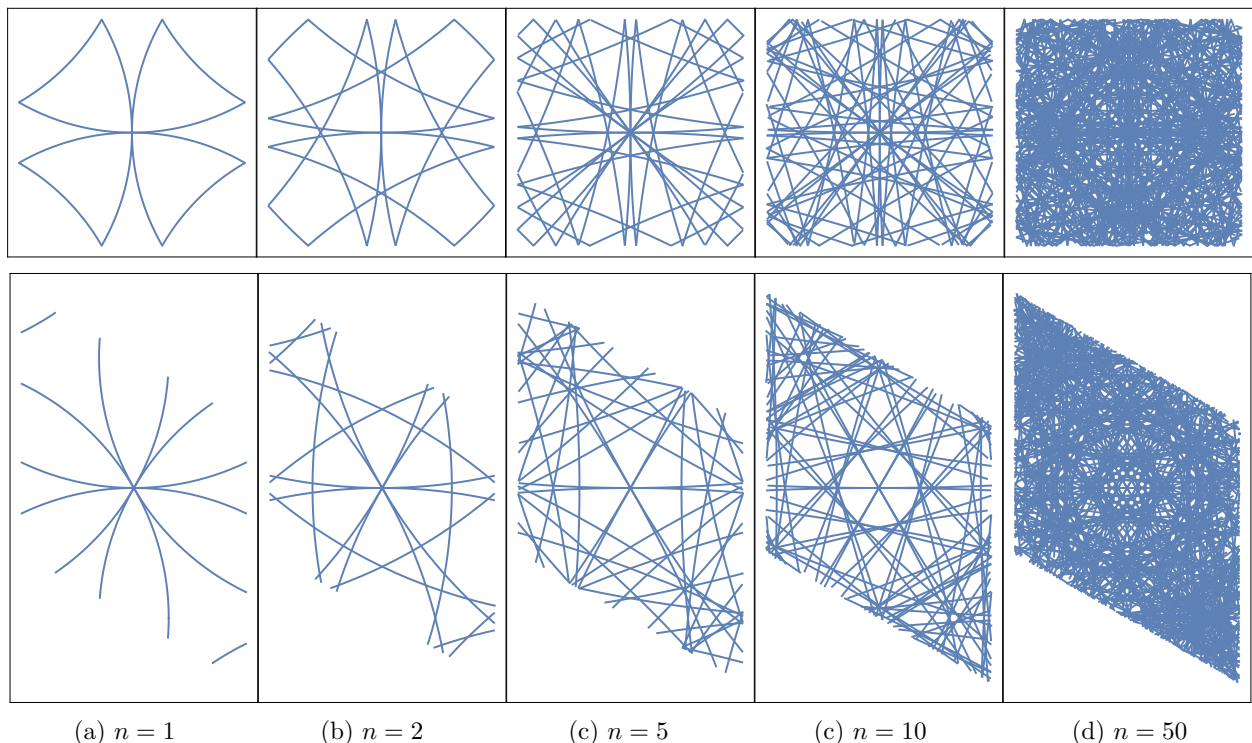(a) $n = 1$ (b) $n = 2$ (c) $n = 5$ (c) $n = 10$ (d) $n = 50$

FIG. 1. (color online) Isofrequency lines for a two-dimensional homogeneous medium artificially discretized on a square (top) and triangular (bottom) lattices. The band index is labeled by n. The plots depict the isofrequency lines contained in the FBZ of the lattice and can be periodically replicated in all directions. A mathematical definition of the folding operation is given in Eq. (22).

The coefficients δ_{6x} and δ_{6y} in front of these terms in (7) are rather complicated and we have not computed them explicitly in Appendix B. We note, however, that the X and Y-directions are not equivalent in a triangular lattice. Therefore, $\delta_{6x} \neq \delta_{6y}$. If these two coefficients were equal, the term in the square brackets in (7) would reduce to q^6. This is exactly what happens in the case of a square lattice where the X and Y-directions are equivalent and the corresponding coefficients are absorbed in β_6.

The terms (8) and (9a) are graphically illustrated in Fig. 2.

We thus see why the triangular lattice is more amenable to homogenization: the anisotropic terms start to appear in this case only in sixth order in $\mathbf{q}$. However, as soon as these terms yield a noticeable contribution to $\Sigma(\omega, \mathbf{q})$, isotropy is lost very fast. This observation was made in Ref. 10 and illustrated by considering the transmission coefficient of a PC slab. Below we will illustrate this observation by plotting the isofrequency lines for an extended range of q_x, which includes not only propagating incident waves (in a vacuum or in a PC or in both), but also evanescent incident waves.

In what follows, we will consider the following problem. Fix the frequency and assume that the Bloch wave vector has a known and purely real projection q_x onto an axis that is tangential to the PC/air interface (although

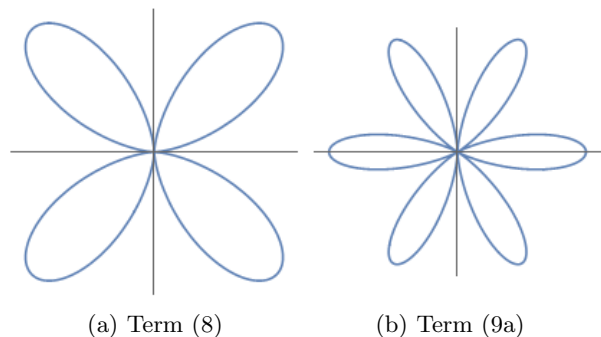(a) Term (8)   (b) Term (9a)

FIG. 2. Graphical illustration of the terms (8) and (9a). The parametric plots show the dependence of the magnitude of each term on the direction of the (purely real) vector $\mathbf{q}$. Note that the lines shown in plots (a) and (b) are similar (but not completely identical) to the isofrequency lines shown in Fig. 1 for $n = 1$. In the latter case, the lines are reflected by the boundaries of the FBZ. However, in the vicinity of the origin, the lines are identical [up to a $\pi/6$ rotation for the term (9a)]. This is a general manifestation of the applicable rotational symmetry group [C_4 in (a) and C_6 in (b)].

the interface is not considered explicitly). Then compute the corresponding values of q_y that satisfy the dispersion equation. Here q_y can be real, imaginary or complex, even if the PC is made of lossless components. In a loss-

less PC, the set of purely real solutions (q_x, q_y) would form a traditional isofrequency line. However, the dispersion equation has a solution (in fact, infinitely many solutions) for any q_x. Some of these solutions are complex. When q_y has a nonzero imaginary part, the corresponding Bloch wave is evanescent. Evanescent Bloch waves can be excited in the PC by an incident plane wave that is either propagating or evanescent in a vacuum. It is important to note, however, that the projection of the incident wave vector onto any flat interface is equal to q_x. Thus, to make sure that a homogenization theory is applicable to a sufficiently large class of incident waves, which must necessarily include evanescent waves.

III. COMPUTATIONAL FORMULAS AND NUMERICAL METHODS

In the numerical examples we consider a two-dimensional PC whose exact permittivity $\tilde{\epsilon}(\mathbf{r})$ satisfies the periodicity relation

$$\tilde{\epsilon}(\mathbf{r} + n_1\mathbf{a}_1 + n_2\mathbf{a}_2) = \tilde{\epsilon}(\mathbf{r}) , \tag{10}$$

where $\mathbf{a}_1$ and $\mathbf{a}_2$ are two primitive lattice vectors and n_1, n_2 are integers. Here and below the overhead tilde will be used to denote functions obeying the lattice periodicity (10). We will work in an orthogonal reference frame whose Z-axis is perpendicular to both $\mathbf{a}_1$ and $\mathbf{a}_2$. Then these two vectors lie in the XY plane. We will further consider electromagnetic waves with one-component electric field $\hat{\mathbf{E}} = (0, 0, E)$ and two-component magnetic field $\mathbf{H} = (H_x, H_y, 0)$. In this case, Maxwell's equations are reduced to the scalar wave equation for the electric field

$$\left[\nabla^2 + k^2\tilde{\epsilon}(\mathbf{r})\right] E(\mathbf{r}) = 0 , \tag{11}$$

where $\mathbf{r} = (x, y)$.

The formulas of this section are applicable to arbitrary primitive vectors $\mathbf{a}_1$ and $\mathbf{a}_2$. However, the simulations will be performed for the square lattice ($a_1 = a_2 = a$ and $\mathbf{a}_1 \cdot \mathbf{a}_2 = 0$) and equilateral triangular lattice ($a_1 = a_2 = a$ and $\mathbf{a}_1 \cdot \mathbf{a}_2 = a^2/2$).

Let $\mathbf{b}_1$ and $\mathbf{b}_2$ be the primitive vectors of the reciprocal lattice such that

$$\mathbf{a}_j \cdot \mathbf{b}_k = 2\pi\delta_{jk} . \tag{12}$$

A generic reciprocal lattice vector $\mathbf{g}$ can be written as

$$\mathbf{g} = \mathbf{b}_1 n_1 + \mathbf{b}_2 n_2 , \tag{13}$$

where n_1, n_2 are integers. We can view $\mathbf{g}$ as a discrete composite index, which maps one-to-one to the pair (n_1, n_2). In particular, the exact permittivity of the medium is expandable as

$$\tilde{\epsilon}(\mathbf{r}) = \sum_{\mathbf{g}} \epsilon_{\mathbf{g}} e^{i\mathbf{g}\cdot\mathbf{r}} , \quad \epsilon_{\mathbf{g}} = \frac{1}{S[\mathbb{C}]} \int_{\mathbb{C}} \tilde{\epsilon}(\mathbf{r}) e^{-i\mathbf{g}\cdot\mathbf{r}} d^2 r , \tag{14}$$

where $\mathbb{C}$ is an elementary cell of the medium and $S[\mathbb{C}]$ is its area, i.e., a^2 for a square lattice or $|(\mathbf{a}_1 \times \mathbf{a}_2) \cdot \hat{\mathbf{z}}|$ in the more general case.

Our simulations were performed for a two-component PC and we therefore adduce the computational formulas that are specific to this case. However, we believe that the conclusions of this paper are not limited to two-component media. For a two-component medium we can write

$$\tilde{\epsilon}(\mathbf{r}) = \epsilon_h + (\epsilon_i - \epsilon_h)\tilde{\Theta}(\mathbf{r}) , \tag{15}$$

where ϵ_h and ϵ_i are the permittivities of the host and the inclusions and $\tilde{\Theta}(\mathbf{r})$ is the lattice-periodic shape function. Let the region of the inclusion be $\Omega \in \mathbb{C}$. Then $\tilde{\Theta}(\mathbf{r}) = 1$ if $\mathbf{r} \in \Omega$ and $\tilde{\Theta}(\mathbf{r}) = 0$ otherwise. Upon Fourier transformation, we find that

$$\epsilon_{\mathbf{g}} = \epsilon_h \delta_{\mathbf{g}0} + \rho\chi M(\mathbf{g}) , \tag{16}$$

where

$$M(\mathbf{g}) = \frac{1}{S[\Omega]} \int_{\Omega} e^{-i\mathbf{g}\cdot\mathbf{r}} d^2 r , \tag{17a}$$

$$\rho = S[\Omega]/S[\mathbb{C}] , \quad \chi = \epsilon_i - \epsilon_h . \tag{17b}$$

In the above equations, ρ is the area fraction of the inclusions and χ is the contrast. The function $M(\mathbf{g})$ contains information about the inclusion geometry but is independent of ϵ_h and ϵ_i.

Bloch-periodic functions such as the electric field $E(\mathbf{r})$ and displacement $D(\mathbf{r})$ can be expanded as

$$E(\mathbf{r}) = \sum_{\mathbf{g}} E_{\mathbf{g}} e^{i(\mathbf{q}+\mathbf{g})\cdot\mathbf{r}} , \quad D(\mathbf{r}) = \sum_{\mathbf{g}} D_{\mathbf{g}} e^{i(\mathbf{q}+\mathbf{g})\cdot\mathbf{r}} , \tag{18}$$

where $\mathbf{q}$ is the Bloch wave vector, which must be determined by substituting (18) into (11). The equation $D(\mathbf{r}) = \tilde{\epsilon}(\mathbf{r})E(\mathbf{r})$ takes the form $D_{\mathbf{g}} = \sum_{\mathbf{p}} \epsilon_{\mathbf{g}-\mathbf{p}}E_{\mathbf{p}}$ while (11) is reduced in this case to

$$(\mathbf{g} + \mathbf{q})^2 E_{\mathbf{g}} = k^2 D_{\mathbf{g}} . \tag{19}$$

Combining the wave equation and constitutive relation together we obtain the eigenproblem

$$(\mathbf{g} + \mathbf{q})^2 E_{\mathbf{g}} = k^2 \sum_{\mathbf{p}} \epsilon_{\mathbf{g}-\mathbf{p}}E_{\mathbf{p}} , \tag{20}$$

which determines the allowable values of the vector $\mathbf{q}$ for each frequency ω or for the corresponding wavenumber $k = \omega/c$. We now use the expression for $\epsilon_{\mathbf{g}}$ (16), which is specific to two-component PCs. Upon substitution of (16) into (20), we obtain the infinite set of equations

$$\left[(\mathbf{g} + \mathbf{q})^2 - k^2\epsilon_h\right] E_{\mathbf{g}} = \rho\chi k^2 \sum_{\mathbf{p}} M(\mathbf{g} - \mathbf{p})E_{\mathbf{p}} . \tag{21}$$

For each real value of q_x (and $k > 0$), the above equation defines an infinite set of q_y.

In the simulations, we have solved (21) by linearization and truncation of the grid of $\mathbf{g}'s$. Linearization can be achieved by the standard method of defining the new variable $\mathbf{H_g} = (\mathbf{g} + \mathbf{q})E_\mathbf{g}$. For any finite truncation of the grid of $\mathbf{g}$'s, the operation of linearization increases the number of equations by the factor of 3. The resultant linear eigenproblem is neither Hermitian nor symmetric and, therefore, its solutions are in general complex.

We have solved the above eigenproblem by using Intel's MKL library of subroutines for Fortran. Convergence of the results with the grid size was checked by (i) consecutively doubling the size of the grid and (ii) by comparison with a high-order finite difference method (FLAME)[33]. We have obtained both an excellent convergence with respect to the grid size and an excellent agreement with the finite-difference method, which allows us to conclude that the numerical results shown below are accurate.

Below, we compare the solutions obtained numerically for the actual PC to similar solutions in a homogeneous effective medium, which is artificially discretized on the same lattice as that of the PC. The refractive index n_eff of the medium is determined from the approximate radius of the purely real quasi-circular lobe of the isofrequency line of the PC (or from a more general fit in the case of losses). This way, the effective medium mimics the dispersion law of the PC as accurately as possible for the directions of propagation that correspond to the points on the quasi-circular lobe. To perform the comparison for more general Bloch wave vectors $\mathbf{q}$, we need to introduce the operation of "folding" into the FBZ of the lattice. This operation can be formally defined as

$$[\mathbf{q}]_\mathrm{FBZ} = \mathbf{q} - (n_1\mathbf{b}_1 + n_2\mathbf{b}_2) \ , \tag{22a}$$

$$n_i = \mathrm{Nint}\left(\frac{\mathbf{q} \cdot \mathbf{a}_i}{2\pi}\right) \ . \tag{22b}$$

In the above equations, $\mathrm{Nint}(z)$ is the nearest integer to the complex number z. The vector $\mathbf{q}$ in the right-hand side of (22a) is not restricted to the FBZ and, in the effective medium, it satisfies the dispersion equation $q_x^2 + q_y^2 = n_\mathrm{eff}^2 k^2$ where n_eff is the effective index of refraction. We can view q_x as a mathematically-independent real-valued variable, compute q_y as $q_y = \pm\sqrt{n_\mathrm{eff}^2 k^2 - q_x^2}$, substitute the resultant pair (q_x, q_y) in the right-hand side of (22a), and this will yield the dispersion equation of the artificially discretized homogeneous effective medium. Examples of such folding will be shown in the figures below, and Fig. 2 was obtained by plotting all real-valued points $[\mathbf{q}]_\mathrm{FBZ}$ in the plane (q_x, q_y) at the particular frequencies for which these lines cross the origin.

IV. NUMERICAL EXAMPLES

A. Square lattice with non-dispersive and non-absorbing host

Consider a two-dimensional square lattice of infinite hollow cylinders embedded in a high-index host. The
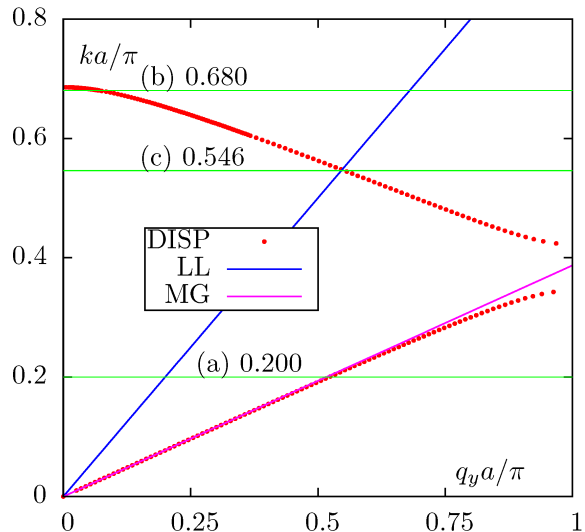

FIG. 3. (color online) Dispersion diagram for the square lattice of hollow cylinders. Frequency is scanned from zero to slightly above the second Γ-point frequency. The variable $ka/\pi = \omega a/\pi c = 2a/\lambda$ is the dimensionless frequency. The line labeled LL is the light line and the line labeled MG gives the Maxwell-Garnett approximation to the dispersion curve. Thin horizontal lines mark the frequencies at which further simulations have been performed.

primitive lattice vectors are in this case $\mathbf{a}_1 = a(1,0)$, $\mathbf{a}_2 = a(0,1)$ and for the reciprocal lattice, $\mathbf{b}_1 = \frac{2\pi}{a}(0,1)$ and $\mathbf{b}_2 = \frac{2\pi}{a}(1,0)$. The radius of the cylinders is taken to be $R = 0.33a$. The host permittivity is $\epsilon_h = 9.61$ and the inclusions (cylinders) are assumed to be a vacuum with $\epsilon_i = 1$. In this example, we disregard absorption either in the host or in the inclusions. This is an "inversion" of the model used in Ref. 34 where a 2D square lattice of aluminum oxide rods in air was considered.

We start with the purely real dispersion diagram. The latter is obtained by setting $q_x = 0$, computing q_y for a range of electromagnetic frequencies, and by keeping only real solutions q_y to the dispersion equation. That is, we will disregard for the moment all complex and imaginary wave numbers q_y (these solutions will be considered later). This approach is conventional for purely real permittivities of the constituent material. Note that we compute the dispersion diagram for a given propagation direction (along the Y-axis). Dispersion curves for different propagation directions are, generally, different. However, if a higher Γ-point frequency ω_n exists according to the definition (4), then ω_n is independent of the propagation direction.

The dispersion diagram for the square lattice of hollow cylinders is shown in Fig. 3. The horizontal lines in this figure mark the dimensionless frequencies at which further simulations have been performed.

The frequency $ka/\pi = 0.200$ [Case (a)] is in the first photonic band. The squared effective refractive index at

this frequency is $n_{\text{eff}}^2 = 6.81$. Note that the Maxwell-Garnett homogenization result for Case (a) is slightly different: $n_{\text{eff}}^2 = 6.66$.

The frequency $ka/\pi = 0.680$ [Case (b)] is in the second transmission band, very close to and slightly below the second Γ-point frequency. This is the main case we are interested in.

Finally, the frequency $ka/\pi = 0.546$ [Case (c)] is at the intersection of the second dispersion branch and the light line. One can expect $n_{\text{eff}}^2 \approx 1$ and $n_{\text{eff}} \approx -1$ at this frequency, since the dispersion in the second photonic band is negative. However, we will see that the medium is very far from being homogenizable in cases (b) and (c) and can not be characterized by local effective parameters.

The isofrequency lines for the three frequencies noted above are shown in Figs. 4-6. As was explained in Sec. II, we fix the frequency and view q_x as a mathematically-independent and purely real variable, which is scanned in some interval. For each q_x considered, we find all values of q_y whose imaginary parts are restricted to some sufficiently large range.

In Fig. 4, we plot the real and imaginary parts of q_y as functions of q_x. For each q_x, there exist infinitely many solutions to the dispersion equation with $\text{Re}(q_y) = 0$ and $\text{Im}(q_y) \neq 0$ and we can not display all such points in the plots. However, the number of solutions with $\text{Re}(q_y) \neq 0$ is finite, and all such data points are shown in the figure.

Note that, for each solution (q_x, q_y) in which both q_x and q_y are real, there is also a solution (q_y, q_x). However, this symmetry is broken if q_y has a nonzero imaginary part. For this reason, the upper plots are not completely symmetric with respect to the line $q_y = q_x$. However, the lobes in the lower-left and upper-right corners consist of purely real solutions and these lobes are symmetric. The line that connects these two lobes consists of complex solutions q_y with both real and imaginary parts different from zero. Therefore, these connecting lines are not symmetric. We note that these complex solutions can not be obtained in a homogeneous medium and are therefore a manifestation of heterogeneity of the PC.

Referring to the data of Fig. 4, we can conclude that, at the frequency $ka/\pi = 0.200$, the dispersion relation in the PC is almost indistinguishable from the dispersion relation in a homogeneous effective medium with $n_{\text{eff}}^2 = 6.81$. The correspondence holds well into the evanescent waves. Indeed, the range of q_x shown in the figure for this frequency is $0 \leq q_x \leq 5k$, and we can expect that the PC is homogenizable for $-5k \leq q_x \leq 5k$. In fact, the PC is homogenizable in an even wider interval of q_x. Indeed, the correspondence still holds in the "reflected" segments of the lines in the bottom plot. In the effective medium, these reflected segments correspond to $|q_x| > 5k$. The only slight discrepancy between the PC and the homogeneous medium can be observed near the edge of the FBZ ($q_x \approx 5k$). Overall, at this frequency, the dispersion relations in the PC mimic those of a homogeneous medium with a high precision and in a wide range of q_x.

Now let us move to Case (b), $ka/\pi = 0.680$. This

frequency is slightly below the second Γ-point frequency. As one could expect from the theoretical arguments of Sec. II, the upper plot has a nearly circular, purely real lobe in the lower-left corner of the frame. If only this lobe is considered, one might erroneously conclude that, at $ka/\pi = 0.680$, the law of dispersion is almost the same as in an effective homogeneous medium. But a quick purview of the scale of the upper horizontal axis reveals that this isotropic behavior holds only in a very narrow range $-0.15 \lesssim q_x/k \lesssim 0.15$. In problems of transmission through a flat interface, this corresponds to incident angles of less than about $6°$ with respect to the normal. Outside of the above range of q_x, the isotropy is lost as can be seen in the bottom plot. We emphasize that the isotropy is lost in this case for $|q_x| < k$. These values of q_x correspond to propagating waves in a vacuum. It is true that these waves are evanescent in the PC but, upon transmission through a finite slab, incoming propagating waves are always transformed into outgoing propagating waves. Therefore, the transmission properties of the PC at this frequency are very different from those of any homogeneous medium.

Note that, at sufficiently large values of q_x, the waves in PC switch from being evanescent to being propagating again (the upper-right corner lobe). This behavior is not characteristic of any homogeneous medium. Moreover, it can be seen that this PC is characterized by birefringence in some range of q_x even though no anisotropy is involved - again, an effect not observed in homogeneous media. By birefringence we mean the effect when, for a given value of q_x, there exist Bloch waves with two real but different values of q_y, which corresponds to two different directions of propagation. This phenomenon can not occur in any local homogeneous medium.

In the Case (c), $ka/\pi = 0.546$, there is obviously no hope to approximate the law of dispersion of the PC by the law of dispersion of any homogeneous effective medium. The real-valued lobe of the isofrequency line is severely distorted and birefringence is quite prominent and occurs in a wide range of q_x. As expected, the distortion of the real-valued lobe is consistent with C_4 symmetry of the problem. Note, however, that this type of distortion is suppressed in the triangular lattice, as will be shown below.

Next, we show the purely real quasi-circular lobes of the isofrequency lines in more detail in Fig. 5. It can be seen that the lobe is almost indistinguishable from a mathematical circle at the frequency $ka/\pi = 0.200$. A distortion consistent with C_4 symmetry is clearly visible at the frequency $ka/\pi = 0.680$. At $ka/\pi = 0.546$ (far from the Γ-point), this distortion is quite severe.

Perhaps the most clear illustration of the departure from the homogeneous behavior in the cases (b) and (c) is shown in Fig. 6, where we plot $\text{Re}(q_y^2)$ as a function of q_x^2. In a homogeneous medium, this function is simply a straight line. This line is folded into the FBZ of the lattice as is shown in Fig. 6(a) if the medium is artificially discretized (we note that for the particular parameters
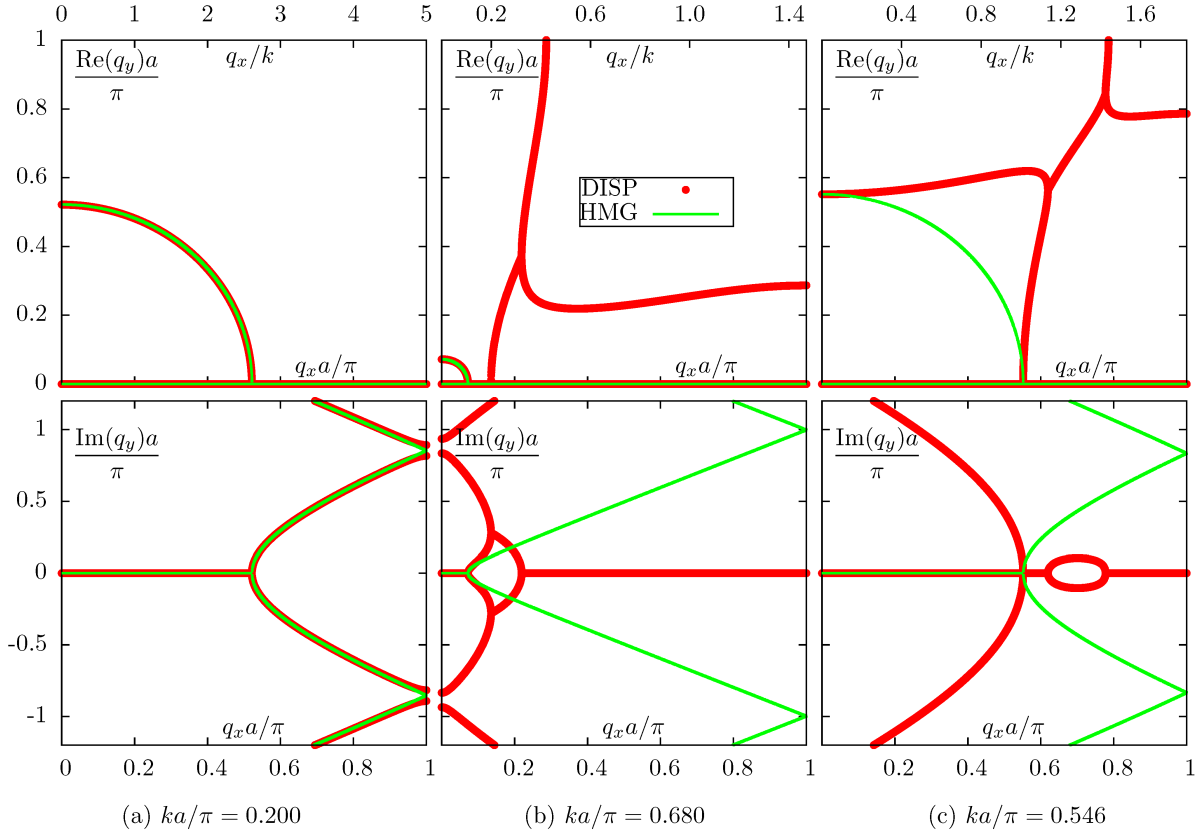
FIG. 4. (color online) Real (upper row of plots) and imaginary (bottom row) parts of q_y as functions of q_x (real by definition) for the square lattice of hollow cylinders at various values of the dimensionless frequency ka/π. Dielectric permittivities of the host and inclusions are $\epsilon_h = 9.61$, $\epsilon_i = 1$; the ratio of the cylinder radius to the lattice period is $R/a = 0.33$. Only one quadrant of the FBZ is shown in the upper row of plots; this quadrant can be replicated to cover the whole FBZ. The scale of the lower horizontal axes is π/a and the scale of the upper horizontal axes is $1/k$, as labeled. All data points with $\text{Re}(q_y) = 0$ (shown in the upper row) have nonzero imaginary parts except for the data point $q_y = 0$ at the intersection of the quasi-circular lobe and the horizontal axis. The corresponding values of q_y are purely imaginary. The imaginary parts of some of these data points are outside of the plotting range in the lower row of plots. Data points are labeled 'DISP' for the actual PC and 'HMG' for a homogeneous medium that is artificially discretized on the same lattice. The HMG isofrequency lines were computed according to (22) with $n_{\text{eff}}^2 = 6.81$ (a), $n_{\text{eff}}^2 = 0.011$ (b) and $n_{\text{eff}}^2 = 0.91$ (c). In the case (c), the choice of n_{eff}^2 guarantees the correct wave number for propagation along X- or Y-axes but not for the intermediate directions.
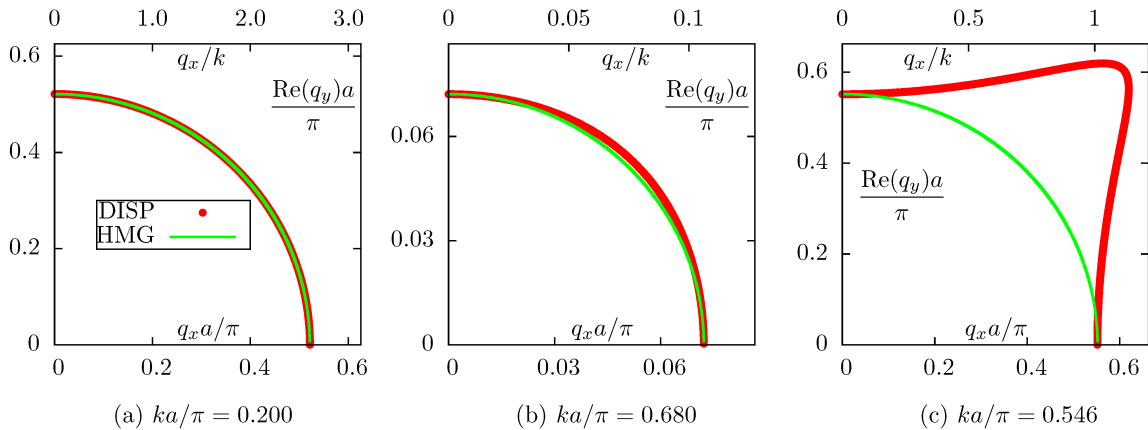

FIG. 5. (color online) Purely real parts of the isofrequency lines shown in Fig. 4. The lines marked 'HMG' are ideal circles.
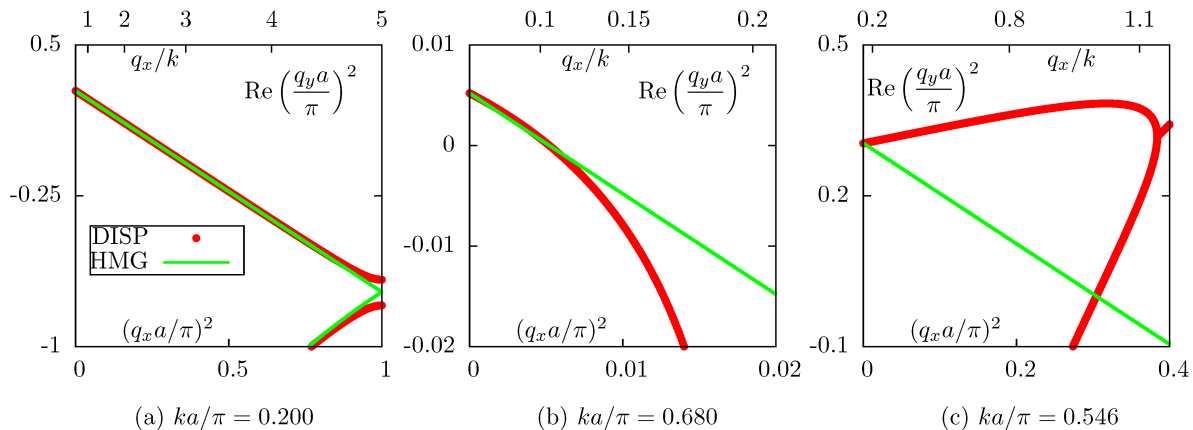
FIG. 6. (color online) Isofrequency lines for the squares of the Cartesian components of Bloch wave vector **q**. More precisely, $\mathrm{Re}(q_y^2)$ is plotted as a function of q_x^2 for the same set of fixed frequencies that were considered in Figs. 4 and 5. In an effective homogeneous medium artificially discretized on a square lattice, this dependence is piece-wise linear as is shown in panel (a).

and the lattice type considered at the moment, the folding results only in linear segments; in a more general case the isofrequency line can acquire curvature due to the folding and an example will be shown below). This behavior is reproduced with very good accuracy at the frequency $ka/\pi = 0.200$ [Case (a)]. However, in Cases (b) and (c) the departure from the linear behavior is obvious and dramatic. We will observe a similar behavior in the triangular lattice as well.

We finally note that the complex branches of the isofrequency lines that connect the purely imaginary and purely real segments of the data point sets is a peculiar feature of the PC, which can not be reproduced in any homogeneous medium with a real refractive index. In the latter case, q_y is either purely real or purely imaginary and q_y^2 is always real.

B. Square lattice with dispersive and absorbing host material

To illustrate the effects of dispersion and absorption in the host material, we now assume that ϵ_h is a function of the frequency ω and is given by the following standard expression:

$$\epsilon_h(\omega) = 1 + [\epsilon_h(0) - 1]\frac{\omega_0^2}{\omega_0^2 - \omega^2 - i\gamma\omega} \ . \qquad (23)$$

Here ω_0 is the resonance frequency and γ is the relaxation constant. In the simulations, we have taken $\epsilon_h(0) = 9.3628$, $\omega_0 a/\pi c = k_0 h/\pi = 4$ and $\gamma/\omega_0 = 1/4\pi \approx 0.07958$. Then, at the frequency corresponding to $ka/\pi = 0.680$ (at which we compute the isofrequency line), we have $\epsilon_h \approx 9.61 + 0.12i$. The real part of this permittivity is the same as that of the non-absorbing host considered in the previous subsection.

The dispersion diagram of a PC with exactly the same geometry as above but with the host permittivity de-

scribed by (23) is shown in Fig. 7. In this figure we compare the dispersion diagrams for the dispersive and non-dispersive hosts and it can be seen that they are barely distinguishable. The working frequency corresponding to $ka/\pi = 0.680$ is still just below the Γ-point frequency in the second photonic band. However, the Γ-point frequency is not precisely defined for the dispersive medium because **q** is now complex at all frequencies. At the apparent Γ-point frequency that is visible in the figure, only $\mathrm{Re}(\mathbf{q})$ turns to zero while $\mathrm{Im}(\mathbf{q})$ does not. Correspondingly, the condition (4) does not really hold at this frequency.

The isofrequency lines computed at $ka/\pi = 0.680$ for the absorbing PC are shown in Fig. 8. The lines are compared to those of a non-absorbing PC (thin blue line) and of a homogeneous effective medium whose refractive index n_{eff} was obtained by fitting the dispersion points to the analytical formula $q_y^2 = n_{\mathrm{eff}}^2 k^2 - q_x^2$ in the interval $0 \leq q_x a/\pi \leq 0.075$. The effective refractive index obtained from the above procedure is $n_{\mathrm{eff}}^2 = 0.0117 - 0.00987i$.

It can be seen from Fig. 8 that account of absorption does not have a significant effect on the homogenizability of the PC. The analysis, however, becomes more complicated because all values of q_y are now complex. An interesting result that is not directly relevant to homogenization is the following. In the case of non-absorbing host, there exists a region of q_x for which q_y has nonzero real and imaginary parts [see Fig. 4(b)]. The corresponding segment of the isofrequency line connects the purely imaginary solutions to the purely real lobe in the upper-right corner of the plot. In the case of an absorbing host, these complex solutions split into two distinct branches with slightly different imaginary and real parts; the upper-right lobe is also deformed and split into two disjoint segments. As a result, absorption in the host material enhances the effect of birefringence. for almost all values of q_x, the PC will transmit two distinct Bloch waves whose rate of spatial decay is not very different.
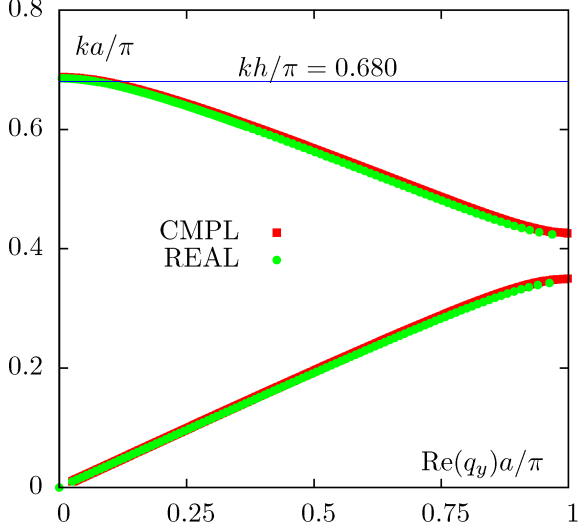
FIG. 7. (color online) Comparison of dispersion diagrams for the host material with a constant real permittivity $\epsilon_h = 9.61$ (REAL) and with dispersive and complex permittivity given by (23) (CMPL). Geometrical parameters are the same as in Fig. 3. The horizontal line marks the frequency at which the isofrequency line has been computed for both cases. In the case of dispersive media, q_y is complex and the quantity shown on the horizontal axis is $\mathrm{Re}(q_y)$.

Of course, there will also be an infinite number of Bloch waves with very fast spatial decay.

The above observation that non-zero absorption results in splitting of the isofrequency line into distinct and disconnected branches may be interesting and deserving an additional investigation, but it has no direct bearing on the main subject of this paper, which is homogenization. At least, the appearance of birefringence makes homogenization only harder.

C. Triangular lattice with non-dispersive and non-absorbing host

We now turn to the case of the triangular lattice. The primitive vectors for the real-space and reciprocal lattices are

$$\mathbf{a}_1 = a(1, 0) \ , \ \mathbf{a}_2 = a(1/2, \sqrt{3}/2) \ ,$$

$$\mathbf{b}_1 = \frac{2\pi}{a}(1, -1/\sqrt{3}) \ , \quad \mathbf{b}_2 = \frac{2\pi}{a}(0, 2/\sqrt{3}) \ .$$

At the center of each elementary cell, which is now rhombic, we place a hollow cylinder of the radius $R = 0.42a$. These inclusions do not cross the boundaries of the elementary cell. It can be seen that the inclusions form a triangular lattice. Since the inclusion obeys the same symmetry as the lattice, the whole structure is C_6-symmetric. The permittivity of the host matrix is taken to be $\epsilon_h = 12.25$. This model was used in Ref. 35.
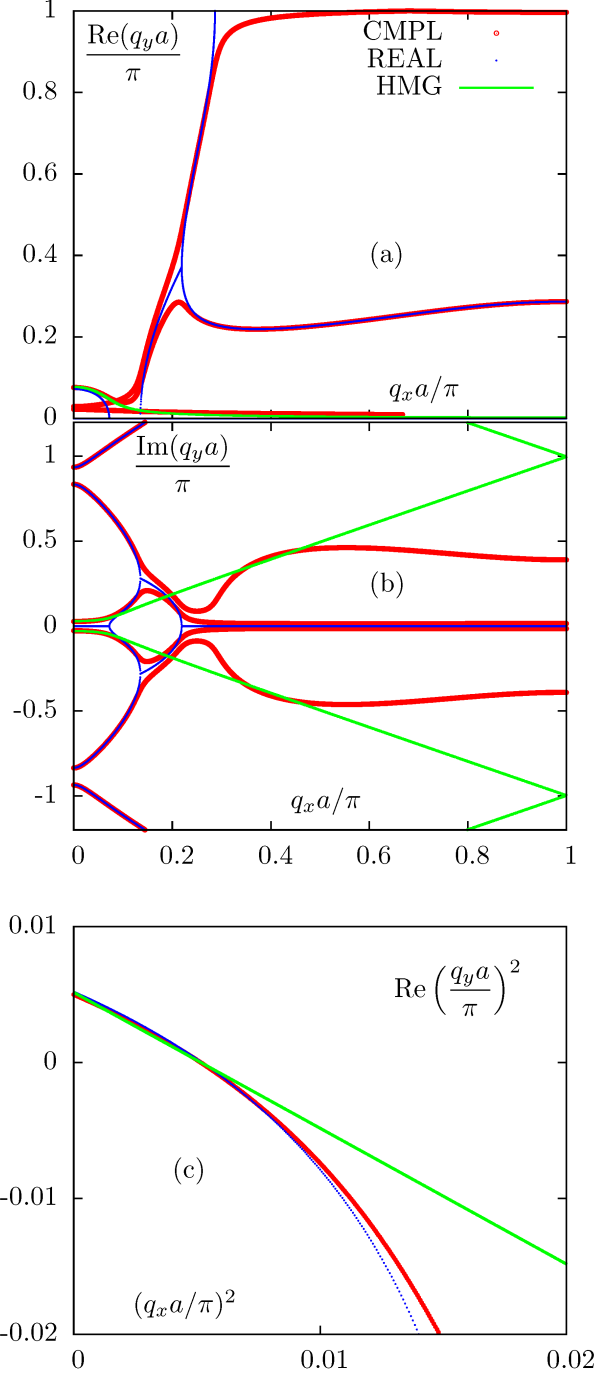

FIG. 8. (color online) Isofrequency lines for the complex permittivity of the host $\epsilon_h = 9.61 + 0.12i$ and $ka/\pi = 0.680$. Isofrequency lines of the absorbing PC are compared to the respective results for a PC with the purely real $\epsilon_h = 9.61$ (thin blue line) and with the homogeneous effective medium with the refractive index $n_{\mathrm{eff}}^2 = 0.0117 - 0.00987i$ (intermediate green line). The top two panels are analogous to the column (b) in Fig. 4 (real and imaginary parts of q_y as functions of q_x). The bottom panel is analogous to the column (b) in Fig. 6 (real part of q_y^2 as a function of q_x^2).
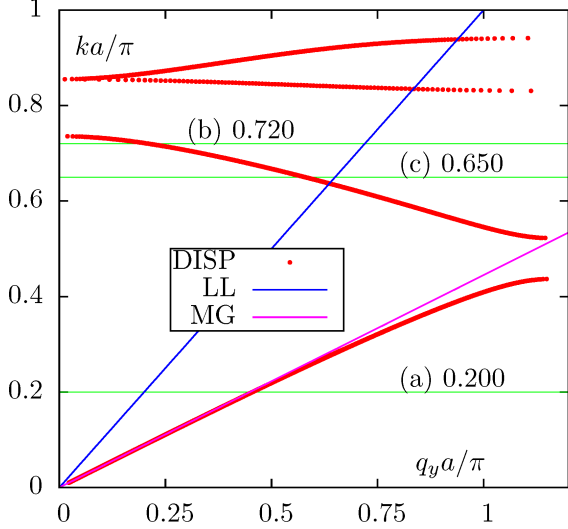
FIG. 9. (color online) Same as in Fig. 3 but for the triangular lattice of hollow cylinders of the radius $R = 0.42a$ in the host medium of the permittivity $\epsilon = 12.25$.

The FBZ of the lattice described above is a rhombus and any of the two triangles created by drawing a rhombus diagonal are equivalent. One can construct a hexagon of two complete rhombuses touching at one corner and two such triangles. All six triangles forming this hexagon are equivalent and, moreover, the dispersion points can be replicated periodically on the hexagonal lattice. It is a common practice to plot the isofrequency lines of such lattices inside a hexagonal region of the reciprocal lattice. However, we do not use this approach in the paper and limit attention to the actual FBZ of the lattice, which is shown in all figures below.

The dispersion diagram for the direction of propagation along the Y-axis ($q_x = 0$) is shown in Fig. 9. As previously, three special frequencies are marked by the horizontal lines in Fig. 9. These frequencies are $ka/\pi = 0.200$ [Case (a)], $ka/\pi = 0.720$ [Case (b)] and $ka/\pi = 0.650$ [Case (c)]. Just as was the case for the square lattice, the frequency (a) is in the first photonic band, the frequency (b) is in the second band slightly below the second Γ-point frequency and the frequency (c) is close to the intersection of the light line and the second branch of the dispersion curve.

The complex isofrequency lines for the three frequencies mentioned above are shown in Fig. 10. This figure is analogous to Fig. 4 and the same quantities are plotted using the same scales of the axes. However, the FBZ of a triangular lattice is more complex geometrically. Therefore, in the top row of plots of Fig. 10, we have shown the complete FBZ of the lattice (the black rhombus) while in Fig. 4, only one quarter of the FBZ was shown.

We now analyze Fig. 10 in more detail. First, focus on the real parts of q_y (the upper row of plots). The central

quasi-circular lobe and the horizontal line $\mathrm{Re}(q_y) = 0$ are analogous to the similar features of the isofrequency lines for a square lattice. The quasi-circular lobes consist of purely real solutions while the line $q_y = 0$ corresponds to purely imaginary solutions. The upper and lower horizontal lines in the upper row of plots correspond to complex solutions that are specific to the triangular lattice. Unlike in the case of a square lattice, these complex solutions do not connect purely imaginary and purely real segments of the isofrequency line (within the FBZ). Another distinction is that the complex solutions in a triangular lattice will appear due to folding of the dispersion equation of an artificially discretized homogeneous medium. For a square lattice, the appearance of complex solutions is a result of interaction that can not be obtained by artificial folding.

We can understand the appearance of the complex solutions shown in Fig. 10 by replicating the central rhombus of the top row of plots in all directions and noting that the horizontal line $q_y = 0$ in the central rhombus will connect to the upper or lower horizontal lines in the replicated rhombuses. We can also understand these solutions qualitatively by considering a homogeneous medium artificially discretized on a triangular lattice. In the infinite (non-discretized) medium, the law of dispersion allows for purely imaginary solutions of the form $q_y = \pm i\sqrt{q_x^2 - n^2k^2}$ for $|q_x| > nk$, $n > 0$ being the index of refraction. Now let us fold these solutions to the FBZ of a triangular lattice according to (22). We can ignore the imaginary part of $\mathbf{q}$ for the purpose of computing the integers n_1, n_2 that are used in this equation. The result of the folding is

$$[q_x]_{\mathrm{FBZ}} = q_x - \frac{2\pi}{a}n_1 \ , \tag{24a}$$

$$[q_y]_{\mathrm{FBZ}} = \pm i\sqrt{q_x^2 - n^2k^2} - \frac{2\pi}{a\sqrt{3}}\left(2n_2 - n_1\right) \ , \tag{24b}$$

$$n_1 = \mathrm{Nint}\left(\frac{q_x a}{2\pi}\right) \ , \quad n_2 = \mathrm{Nint}\left(\frac{q_x a}{4\pi}\right) \ . \tag{24c}$$

These equations are valid for $|q_x| > nk$. We note that the integer index $m = 2n_2 - n_1$ can take only three values: $0, \pm 1$. Correspondingly, for the real and imaginary parts of $[q_y]_{\mathrm{FBZ}}$, we have the following results:

$$[\mathrm{Re}(q_y)]_{\mathrm{FBZ}} = m\frac{2\pi}{a\sqrt{3}} \ , \quad m = 0, \pm 1 \ , \tag{25a}$$

$$[\mathrm{Im}(q_y)]_{\mathrm{FBZ}} = \pm\sqrt{q_x^2 - n^2k^2} \ . \tag{25b}$$

We will obtain a dispersion equation containing only the quantities $[q_x]_{\mathrm{FBZ}}$ and $[q_y]_{\mathrm{FBZ}}$ if we substitute q_x in (25b) from (24a), i.e., use the relation

$$q_x^2 = \left\{[q_x]_{\mathrm{FBZ}} + \frac{2\pi}{a}n_1\right\}^2 \tag{26}$$

to obtain the following closed-form equation:

$$[\mathrm{Im}(q_y)]_{\mathrm{FBZ}} = \pm\sqrt{\left\{[q_x]_{\mathrm{FBZ}} + \frac{2\pi}{a}n_1\right\}^2 - n^2k^2} \tag{27}$$
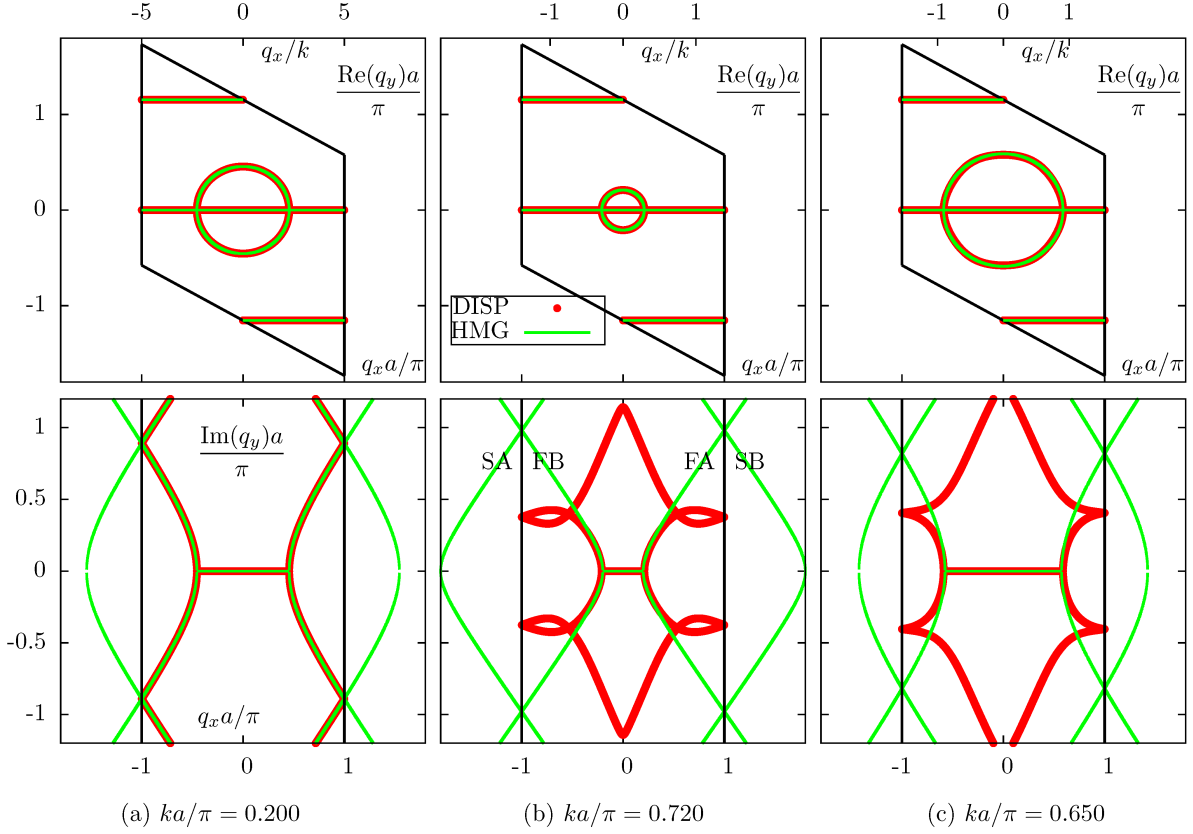
FIG. 10. (color online) Same as in Fig. 4 but for the triangular lattice of hollow cylinders. The 'HMG' isofrequency lines were computed according to (22) with $n_{\text{eff}}^2 = 5.20$ (a), $n_{\text{eff}}^2 = 0.085$ (b) and $n_{\text{eff}}^2 = 0.82$ (c). In the top row of plots, the FBZ of the lattice is shown by the black rhombus. In the lower plot (b), FA and FB mark the first pair of hyperbolas defined by (27), and SA, SB mark the second pair of hyperbolas.

$$\text{for} \quad \left| [q_x]_{\text{FBZ}} + \frac{2\pi}{a} n_1 \right| \geq nk \ .$$

This equation describes a family of hyperbolas parameterized by n_1. The first pair of these hyperbolas (corresponding to $n_1 = 0$) are labeled as FA and FB in Fig. 10. The second pair or hyperbolas (including two branches with $n_1 = \pm 1$) are labeled as SA and SB. Infinitely many similar curves can be generated by translations along the q_x-axis, which correspond to arbitrary integer values of n_1. Note that the two integers n_1 and n_2 that label different "reflected" branches of the solution to the dispersion equation are subject to the selection rule $m = n_1 - 2n_2 = 0, \pm 1$. Therefore, for each n_1, the set of allowable values of n_2 is restricted.

Thus, the FBZ folding in a triangular lattice transforms purely imaginary solutions q_y into complex solutions $[q_y]_{\text{FBZ}}$, which explains the appearance of the upper and lower horizontal lines in the top row of plots in Fig. 10.

Of course, the analytical folding described above is valid only for a homogeneous medium. However, at the frequency $ka/\pi = 0.200$, the actual dispersion relation in the PC mimics the dispersion relation in a homogeneous effective medium with $n_{\text{eff}}^2 = 5.20$ very closely. This conclusion can be drawn from the data shown in the lower plot for the Case (a) in Fig. 10 and is further confirmed and illustrated in Fig. 11. In the latter figure, we plot $[\text{Im}(q_y)]_{\text{FBZ}}$ as a function of $[q_x]_{\text{FBZ}}$ (the symbol $[\ldots]_{\text{FBZ}}$ that signifies Bloch-periodicity of all data points is omitted in this and other figures for brevity). The comparison is made between the respective quantities obtained for the actual PC and a homogeneous effective medium artificially discretized on the same lattice. The individual hyperbolas described by (27) are shown with green lines and the actual dispersion solutions in the PC are shown by the red points. Both solutions are only valid within the FBZ (between the two vertical lines) and beyond these two lines they must be periodically replicated. The green lines outside of the FBZ are shown in the figure only to guide the eye (to help visually identify individual hyperbolas).

It can be seen that, at the dimensionless frequency $ka/\pi = 0.200$, the law of dispersion in the PC is almost indistinguishable from the law of dispersion in a homogeneous medium, at least up to $|\text{Im}(q_y)a/\pi| \leq 5$, which corresponds, approximately, to $|q_x| \lesssim 25k$, that is,
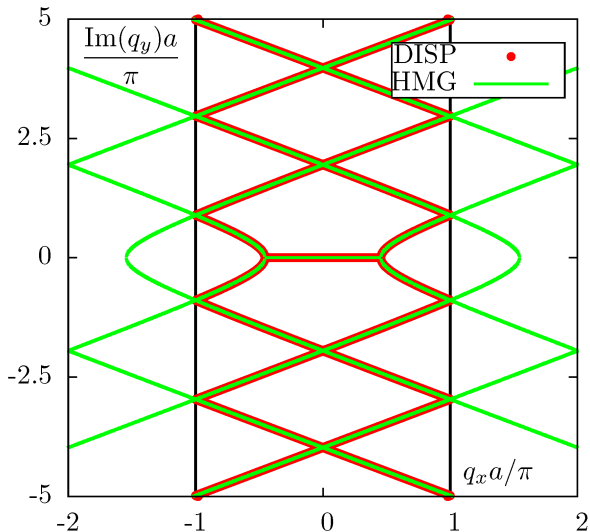
FIG. 11. (color online) An expanded view of $\mathrm{Im}(q_y)$ as a function of q_x for $ka/\pi = 0.200$ [Case (a)]. Here several 'reflected' segments of the isofrequency line are shown. These segments are obtained by folding the corresponding isofrequency line for an infinite homogeneous medium into the FBZ of the triangular lattice according to (27). The analytical lines outside the FBZ (shown by the two vertical lines) are not the actual solutions and are shown only to guide the eye.

very far into the evanescent spectrum. Therefore, we can claim that, at this particular frequency, the PC can be homogenized for many practical purposes.

The situation is quite different at the other frequencies considered. In Case (b) ($ka/\pi = 0.720$, just below the second Γ-point frequency), the dependence of $\mathrm{Re}(q_y)$ on q_x still looks very "homogeneous". However, as soon as we look at $\mathrm{Im}(q_y)$, it becomes obvious that the law of dispersion departs from that of a homogeneous medium quite dramatically as soon as q_x approaches the region of evanescent waves ($|q_x| \gtrsim k$). Definitely, homogenization is not possible for incident evanescent waves, and it is inaccurate for propagating waves with large angles of incidence, e.g., for $\theta_{\mathrm{inc}} \gtrsim 70°$. For the frequency (c) [$ka/\pi = 0.650$], the quasi-circular lobe is much larger and appears to not be distorted significantly. However, the lower plot again clearly indicates a lack of correspondence between the law of dispersion of a homogeneous medium and the law of dispersion in the PC.

We therefore conclude that the PC is not homogenizable at the frequencies (b) and (c). Assigning the PC some effective parameters at these frequencies can potentially be a valid approximation only for a limited range of incident angles that, at the very least, do not include evanescent waves.

Next, as was done above for the square lattice, we show in Fig. 12 the quasi-circular lobes of the top row of Fig. 10 in more detail. In order to make the small deviations from circularity more visible, we plot in this figure only

one quarter of each quasi-circular lobe. The lobes appear to be indistinguishable from mathematical circles in Cases (a) and (b) but some small distortions consistent with the C_6 symmetry are visible in Case (c). The high quality of the quasi-circular lobes can be explained by the fact that the terms of the form (8) do not enter the expansion of $\Sigma(\omega, \mathbf{q})$ because they are not invariants of C_6 symmetry. Therefore, the distortions first appear in six-order terms (9). However, as soon as the terms (9) become non-negligible, the isotropy of $\Sigma(\omega, \mathbf{q})$ is lost quickly. Therefore, the high quality of the quasi-circular lobes shown in Fig. 12 is not a sufficient condition for homogenizability.

But the most convincing demonstration of non-homogenizability of the PC at the frequencies (b) and (c) can be obtained by considering the squares of the Cartesian components of $\mathbf{q}$. In Fig. 13, we plot $\mathrm{Re}(q_y^2)$ vs q_x^2. In a homogeneous effective medium artificially discretized on a triangular lattice, this plot consists of one linear and several curved segments. The linear segment corresponds to the eigenvalues that are either purely real or purely imaginary. The curved segments correspond to the peculiar complex eigenvalues that are specific to the triangular lattice and described mathematically by the equations (24) through (27). One such curved segment is shown in every panel of Fig. 13; other curved segments lie outside of the plot frames. We now observe that at the frequency (a) there is a complete correspondence between the artificially discretized homogeneous medium and the PC. This, of course, could have been expected from the data shown in Figs. 10, 11,12. At the frequencies (b) and (c), the correspondence is broken in the linear segment. In addition the curved segments in the homogeneous medium and in the PC are completely different. The last point is significant and deserves an additional discussion.

Let us assume that a slab of a *homogeneous* material contained between two planes $y = 0$ and $y = L$ is illuminated by a plane wave with the incidence angle such that $q_x < \pi/a$, where a is the period of artificial discretization in the x-direction. We can describe the medium as homogeneous (the traditional approach) or as a PC (by using artificial discretization). Both approaches are mathematically equivalent and will yield the same results for all observables. Assume that we have decided to describe the medium as a PC. In this case, for a given purely real q_x, there will be infinitely many eigenvalues q_y. However, the incident radiation will excite only one mode, namely, the mode with q_y that corresponds to the dispersion relation of the homogeneous material. The modes with other values of q_y can be excited if we take q_x to be outside of the FBZ of the lattice. In any case, for a given q_x, only one mode is excited in the material.

In the case of an actual PC whose law of dispersion closely mimics that of a homogeneous medium, as was the case at the frequency (a), we can expect that the same selection rules will work: at any given q_x, only one mode will be excited in the PC. Then the transmission

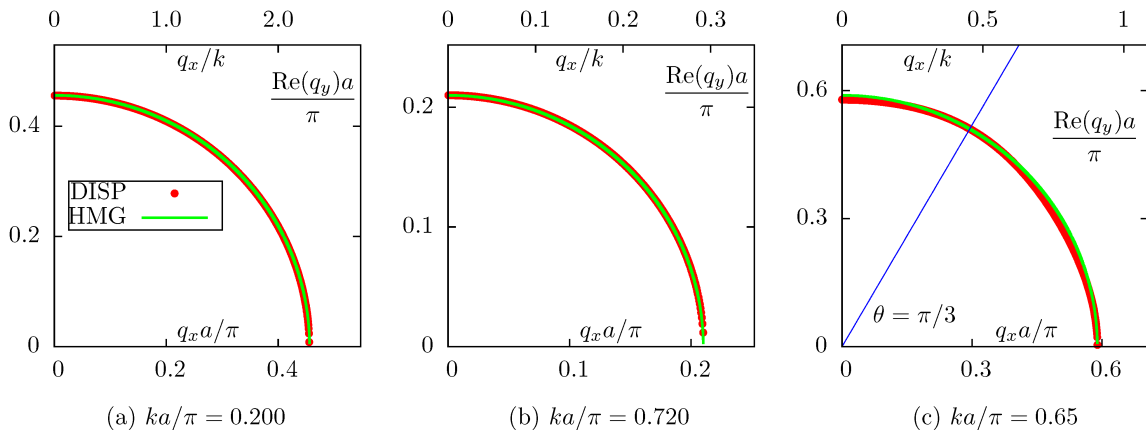(a) $ka/\pi = 0.200$ (b) $ka/\pi = 0.720$ (c) $ka/\pi = 0.65$

FIG. 12. (color online) Same as in Fig. 5 but for the triangular lattice. Distortions of the quasi-circular lobes are consistent with C_6 symmetry. The straight line in panel (c) makes the $\pi/3$ angle with the horizontal axis.

and reflection coefficients can be expected to be the same in the PC and the homogeneous medium. This is indeed the case in the homogenization limit[26,36] $a \to 0$.

If we consider the PC with a constant nonzero a at sufficiently high excitation frequencies (such as the frequencies (b) and (c) in the above examples), there is no reason to believe that the selection rules will work the same way as in a homogeneous medium. In other words, at a given q_x, modes with several different values of q_y can be excited in the PC. Granted, these additional modes have q_y with nonzero imaginary parts and are therefore exponentially decaying inside the medium. However, it is a mistake to neglect these modes completely. Indeed, even the fields associated with these modes are exponentially decaying with y, they are not negligibly small at the interface $y = 0$. Moreover, these modes do not generally average to zero over the surface of the elementary cell. In these respect, they are different from the surface waves discussed by us previously[26,37]. Therefore, excitation of these additional modes will have an adverse effect on homogenizability.

In Fig. 13, two different effects are illustrated. The first effect is the deviation of the law of dispersion in the PC from that in a homogeneous effective medium for the "fundamental mode" (for which q_y^2 is always purely real). This effect is manifest at sufficiently large incident angles and, in particular, for incident evanescent waves. The second, more subtle effect is manifest even at small incident angles. Namely, it can be seen that, at $q_x \sim 0$, the PC has additional modes (allowable values of q_y) that are dramatically different from the respective values in the effective medium. Appearance of these additional solutions can be expected to influence the medium impedance in an angle-dependent manner and have an additional degrading effect on the PC homogenizability.

V. DISCUSSION

The main message of this paper is that, in many applications of practical interest, it is insufficient to consider only the purely real segments of the isofrequency lines to decide whether a given photonic crystal (PC) is homogenizable - that is, electromagnetically similar to a homogeneous medium. In the case of two-dimensional PCs that we have considered, these purely real segments can be circular with reasonable precision in the higher photonic bands. In this case, all propagation directions appear to be equivalent and the physical effects of the medium discreteness (the presence of the lattice) appear to be minimized or absent. The medium can also be characterized by negative dispersion, which means that the real part of the Bloch wave number tends to decrease with frequency while the imaginary part is negligibly small. Nevertheless, the PC is not homogenizable in this case and can not be characterized by angle-independent, purely local effective parameters ϵ_{eff} and μ_{eff}. This conclusion can be drawn by considering the angular dependence of the effective parameters and by including evanescent waves into consideration.

A difficulty one faces when restricting consideration to the dispersion relation is that infinite media have no interfaces and therefore it is easy to overlook the important features of the solutions. This is exactly what happens when one restricts consideration to purely real isofrequency lines. In this paper, we have generalized this approach by considering two orthogonal directions in space, X and Y, and assuming that the projection of the Bloch wave vector $\mathbf{q}$ on one of these axes (X in our case) is a mathematically-independent variable, which is preserved by the process of reflection and refraction at any planar interface $y = \text{const}$.

The results obtained above are consistent with an earlier prediction made by one of the authors regarding the impossibility of negative refraction[38]. The essential as-
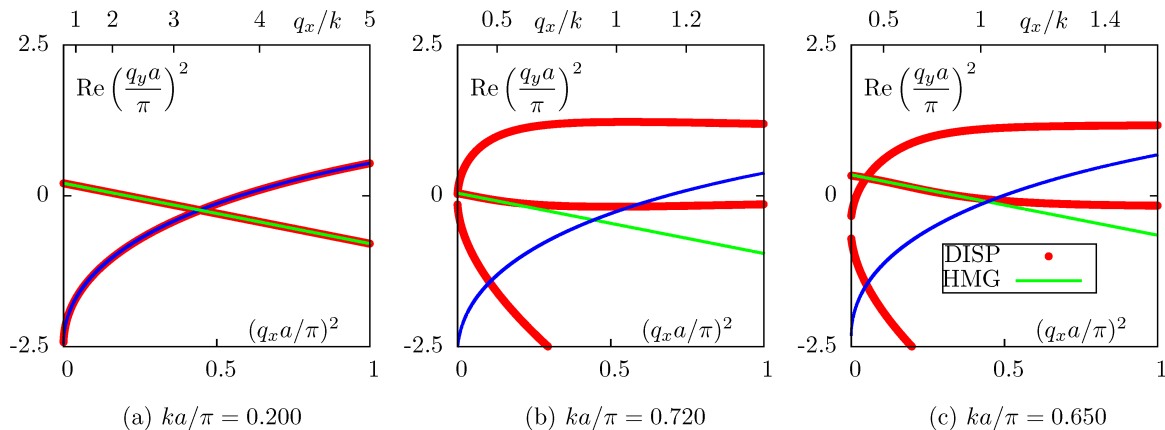
FIG. 13. (color online) Same as in Fig. 6 but for the triangular lattice that is considered in this subsection.

sumption of the above reference was that the medium in question is electromagnetically homogeneous. However, this requirement was not clearly defined. On the other hand, it is a common knowledge that some discrete systems such as photonic crystals or chains of interacting particles or resonators can be characterized by *negative dispersion*. This is true, in particular, for PCs in higher photonic bands. This may seem to contradict the conclusions of Ref. 38. In the present work, we show that there is no contradiction since PCs are not homogenizable in the higher bands. We also define more clearly what we mean by the requirement that the medium is "electromagnetically homogeneous". Specifically, we have formulated a *necessary* condition of homogenizability of a heterogeneous medium. This condition is based on the correspondence of the dispersion relations in a hypothetical homogeneous effective medium and the actual PC. A *sufficient* condition must include the impedance[36], which can not be defined unambiguously without explicit consideration of the medium boundary[20]. Thus, we do not discuss a sufficient condition of homogenizability in this paper. Instead, we show that even the necessary condition is violated in the higher photonic bands.

Note that the lack of homogenizability at sufficiently high frequencies was demonstrated earlier for the special case of 1D periodic media[39]. The latter case is, however, complicated by the fact that a 1D layered medium is always anisotropic. Certain types of indefinite anisotropic media are capable of refracting a narrow incident beam on the same side of the normal. This phenomenon can easily be confused with negative refraction. In this paper, we consider a system without anisotropy (either magnetic or electric).

A more general result closely related to the main conclusion of this paper can be stated in the form of an *uncertainty principle* of homogenization[25], namely: the more an effective magnetic permeability deviates from unity, the less accurate the corresponding homogenization result is. In the former reference as well as in this paper, we define the effective medium parameters ϵ_{eff} and

μ_{eff} by requiring that a heterogeneous and an "effective" homogeneous samples (of the same overall shape) can not be distinguished by performing *external* measurements of reflected, transmitted and scattered waves in a sufficiently broad range of illumination conditions. We have previously formulated this requirement quantitatively for the case of a plane-parallel slab[36]. We believe, however, that, if the medium is truly homogenizable, the effective parameters can also be used to compute various physical quantities that are *internal* to the medium such as dissipated heat per unit volume per unit time. These quantities are usually quadratic in the fields. In particular, we believe that the effective permittivity ϵ_{eff}, when it can be introduced, satisfies the usual physical requirements of passivity and causality. According to the homogenization uncertainty principle[25], the effective permeability μ_{eff} can not be much different from unity, and the physical requirements applicable to this quantity may be more nuanced[40]. We note that these questions are of current interest in the theory of homogenization but, unfortunately, outside of the scope of this paper wherein we do not compute or consider spatial distribution and fluctuations of the electromagnetic fields in a PC, or the medium boundaries.

In Appendix A, we explain the mathematical distinction between spurious Γ-point frequencies that are due to Brillouin-zone folding of Bloch bands and "genuine" Γ-point frequencies that are due to multiple scattering. Understanding this distinction is important for the theory of homogenization.

We finally note that this work is based upon some important theoretical observations made by Li, Holt and Efros[10]. We have developed these observations further with a specific focus on dispersion relations in 2D PCs with square and triangular lattices.

---

[*] On leave from the Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; vmarkel@fresnel.fr,vmarkel@mail.med.upenn.edu

[†] igor@uakron.edu

[1] U. C. Hasar, G. Buldu, M. Bute, J. J. Barroso, T. Karacali, and M. Ertugrul, AIP Advances **4**, 107116 (2014).

[2] T. D. Karamanos, S. D. Assimonis, A. I. Dimitriadis, and N. V. Kantartzis, Photonics and Nanostructures **12**, 291 (2014).

[3] N. C. J. Clausen, S. Arslanagic, and O. Breinbjerg, Photonics and Nanostructures **12**, 419 (2014).

[4] A. Ciattoni and C. Rizza, Phys. Rev. B **91**, 184207 (2015).

[5] V. Sozio, A. Vallecchi, M. Albani, and F. Capolino, Phys. Rev. B **91**, 205127 (2015).

[6] J. B. Pendry, Phys. Rev. Lett. **85**, 3966 (2000).

[7] A. L. Pokrovsky and A. L. Efros, Solid State Comm. **124**, 283 (2002).

[8] S. G. Rautian, Phys. Usp. **51**, 981 (2008).

[9] There can be no electric anisotropy because the electric field in the considered geometry is always aligned with the same axis, say, Z. One can attempt to introduce magnetic anisotropy as a phenomenological adjustable parameter, but then one would obtain either a non-symmetric magnetic tensor and hence non-reciprocity, which is not physically present in the system, or otherwise a tensor whose principal axes do not coincide with the crystallographic axes (e.g., for the triangular lattice), or else a tensor that is trivially proportional to the identity (e.g., for the square lattice). Therefore, anisotropic effective magnetic tensor is also not a possibility.

[10] C. Y. Li, J. M. Holt, and A. L. Efros, J. Opt. Soc. Am. B **23**, 963 (2006).

[11] C. Menzel, C. Rockstuhl, T. Paul, F. Lederer, and T. Pertsch, Phys. Rev. B **77**, 195328 (2008).

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

[13] In a more general geometry, isotropy is not required for homogenizability. Thus, for example, a three-dimensional lattice of general parallelepipeds can be homogenizable and characterized by the tensor ϵ_{eff} whose all three principal values are different[26].

[14] B. Stout, M. Neviere, and E. Popov, J. Opt. Soc. Am. A **23**, 1111 (2006).

[15] B. Stout, M. Neviere, and E. Popov, J. Opt. Soc. Am. A **24**, 1120 (2007).

[16] V. Agranovich and V. Ginzburg, *Spatial Dispersion in Crystal Optics and the Theory of Excitons* (Wiley-Interscience, New York, 1966).

[17] L. D. Landau and L. P. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon Press, Oxford, 1984).

[18] V. M. Agranovich and Y. N. Gartstein, Phys. Usp. **49**, 1029 (2006).

[19] V. M. Agranovich and Y. N. Gartstein, Metamaterials **3**, 1 (2009).

[20] V. A. Markel and I. Tsukerman, Phys. Rev. B **88**, 125131 (2013).

[21] Weak nonlocality is usually understood as the regime in

which Taylor expansion of $\Sigma(\omega, \mathbf{q})$ to second order in $\mathbf{q}$ is an accurate approximation. For more general nonlinear dependence of $\Sigma(\omega, \mathbf{q})$ on $\mathbf{q}$, the dispersion equations (1) or (2) will have multiple solutions that are not obtainable by the operation of folding to first Brillouin zone of the lattice.

[22] C. Luo, S. G. Johnson, J. D. Joannopoulos, and J. B. Pendry, Phys. Rev. B **65**, 201104 (2002).

[23] C. Luo, S. G. Johnson, and J. D. Joannopoulos, Appl. Phys. Lett. **81**, 2352 (2002).

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

[25] I. Tsukerman and V. A. Markel, Phys. Rev. B **93**, 024418 (2016).

[26] V. A. Markel and J. C. Schotland, Phys. Rev. E **85**, 066603 (2012).

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

[28] C. Fietz and G. Shvets, Physica B **405**, 2930 (2010).

[29] A. Alu, Phys. Rev. B **84**, 075153 (2011).

[30] V. A. Markel, J. Phys.: Condens. Matter **22**, 485401 (2010).

[31] These assumptions can hold in both 2D and 3D PCs. However, in the 2D geometry, even stronger conditions are satisfied: (i) not only $\mathbf{E}_0$ but the total electric field $\mathbf{E}(x, y) = \hat{\mathbf{z}} E(x, y)$ is collinear with the Z-axis and (ii) the electric field $E(x, y)$ is independent of z. The latter two properties do not generally hold in three-dimensional PCs.

[32] J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals: Molding the Flow of Light* (Princeton University Press, Princeton, N.J., 2008).

[33] I. Tsukerman and F. Čajko, IEEE Trans. Magnetics **44**, 1382 (2008).

[34] R. Gajic, R. Meisels, F. Kuchar, and K. Hingerl, Opt. Expr. **13**, 8596 (2005).

[35] T.-H. Pei and Y.-T. Huang, J. Opt. Soc. Am. B **29**, 2334 (2012).

[36] I. Tsukerman and V. A. Markel, Proc. R. Soc. A **470**, 0245 (2014).

[37] X. Y. Z. Xiong, L. J. Jiang, V. A. Markel, and I. Tsukerman, Opt. Expr. **21**, 10412 (2013).

[38] V. A. Markel, Opt. Expr. **16**, 19152 (2008).

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

[40] V. A. Markel, Phys. Rev. E **78**, 026608 (2008).

[41] I. Tsukerman, J. Opt. Soc. Am. B **25**, 927 (2008).

Appendix A: Dispersion equation in a basis-independent form

Let the lattice-periodic permittivity of the medium $\tilde{\epsilon}(\mathbf{r})$ be a bounded, lattice periodic function, possibly a constant. A Bloch mode solution to the scalar wave equation (11) can be written in the form

$$E(\mathbf{r}) = \tilde{E}(\mathbf{r}) \exp(i\mathbf{q} \cdot \mathbf{r}) \ ,$$

where $\mathbf{q}$ is the Bloch wave vector and $\tilde{E}(\mathbf{r}) \neq 0$ is a nonzero, twice-differentiable, lattice-periodic function, which satisfies the equation

$$\mathcal{L}(\omega, \mathbf{q})\tilde{E}(\mathbf{r}) = 0 \qquad (A1)$$

with the differential operator $\mathcal{L}(\omega, \mathbf{q})$ given by

$$\mathcal{L}(\omega, \mathbf{q}) = (\nabla + i\mathbf{q})^2 + k^2\tilde{\epsilon}(\mathbf{r}) \ .$$

Recall that $k = \omega/c$; hence the dependence of $\mathcal{L}$ on ω. The other reason for this dependence is frequency dispersion (dependence of $\tilde{\epsilon}$ on ω), which is not indicated here explicitly but is taken into account. The requirement that (A1) has a nontrivial solution $\tilde{E}(\mathbf{r}) \neq 0$ defines the dispersion equation $f(\omega, \mathbf{q}) = 0$. We will assume without proof the following statement: if for some pair $(\omega, \mathbf{q})$ (A1) has a nontrivial solution, then this solution is unique up to multiplication by a constant. This is not generally true in 3D where different polarization states can correspond to the same pair $(\omega, \mathbf{q})$.

We will refer to the pairs $(\omega, \mathbf{q})$ for which (A1) has a nontrivial solution as to the solutions to the dispersion equation. Here the variables ω and $\mathbf{q}$ are not restricted and can take general complex values. However, when studying stationary processes, one is typically interested only in solutions with $\omega > 0$. We also note that, if $(\omega, \mathbf{q})$ is a solution, then $(\omega, \mathbf{q} + \mathbf{g})$ is also a solution, where $\mathbf{g}$ is any reciprocal lattice vector. This mathematical property of Bloch waves gives rise to "folding" of the solutions to the dispersion equation.

It is a nontrivial mathematical question how to distinguish between the solutions that occur due to folding from those that occur due to multiple scattering (interaction). In particular, it is quite plausible that solutions $(\omega, \mathbf{q})$ and $(\omega, \mathbf{q}')$ of different physical origin can co-exist at the same frequency ω (or at two very close frequencies). In this Appendix, we present an approach to mathematical classification of these two physically-different solutions.

As was done in Ref. 41, we decompose $\tilde{E}(\mathbf{r})$ and $\tilde{\epsilon}(\mathbf{r})$ into the constant and zero-mean components according to

$$\tilde{E}(\mathbf{r}) = E_0 + F(\mathbf{r}) \ , \quad \langle F(\mathbf{r}) \rangle_{\mathbb{C}} = 0 \ ,$$
$$\tilde{\epsilon}(\mathbf{r}) = \epsilon_0 + \eta(\mathbf{r}) \ , \quad \langle \eta(\mathbf{r}) \rangle_{\mathbb{C}} = 0 \ .$$

where $\langle \ldots \rangle_{\mathbb{C}}$ indicates averaging over the elementary cell $\mathbb{C}$. We note that E_0 is the amplitude of the fundamental harmonic of the Bloch wave. Upon substitution of the above decomposition into (A1), we obtain

$$\left[k^2(\epsilon_0 + \eta(\mathbf{r})) - q^2 \right] E_0 + \mathcal{L}(\omega, \mathbf{q})F(\mathbf{r}) = 0 \ . \qquad (A2)$$

We now introduce the averaging operator $\mathcal{O}$ and the operator $\mathcal{P} = 1 - \mathcal{O}$ as projections onto the complementary subspaces of constant and zero-mean functions so that $\mathcal{O}\tilde{E}(\mathbf{r}) = E_0$, $\mathcal{P}\tilde{E}(\mathbf{r}) = F(\mathbf{r})$ and $\mathcal{O}\tilde{\epsilon}(\mathbf{r}) = \epsilon_0$, $\mathcal{P}\tilde{\epsilon}(\mathbf{r}) = \eta(\mathbf{r})$. By acting on (A2) with $\mathcal{O}$ and $\mathcal{P}$ from the left, we obtain the following two equations

$$\left(k^2\epsilon_0 - q^2 \right) E_0 + k^2 \langle \eta(\mathbf{r})F(\mathbf{r}) \rangle_{\mathbb{C}} = 0 \ , \qquad (A3a)$$

$$\eta(\mathbf{r})E_0 + \mathcal{W}(\omega, \mathbf{q})F(\mathbf{r}) = 0 \ . \qquad (A3b)$$

In the first equation above, we have used the equality $\mathcal{O}\mathcal{L}F(\mathbf{r}) = k^2\mathcal{O}\eta(\mathbf{r})F(\mathbf{r}) = k^2 \langle \eta(\mathbf{r})F(\mathbf{r}) \rangle_{\mathbb{C}}$, which can be proved by using cell-periodicity of $F(\mathbf{r})$ and integral theorems. In the second equation, we have used $(\mathcal{P}\mathcal{L})F(\mathbf{r}) = (\mathcal{P}\mathcal{L}\mathcal{P})F(\mathbf{r})$ and defined the operator

$$\mathcal{W}(\omega, \mathbf{q}) = k^{-2}\mathcal{P}\mathcal{L}(\omega, \mathbf{q})\mathcal{P} \ .$$

We can now consider the following two different kinds of solutions to (A3):

1. Solutions with $E_0 \neq 0$. Bloch waves $E(\mathbf{r})$ with nonzero fundamental harmonic $E_0 \neq 0$ can satisfy the wave equation (11) for a given pair $(\omega, \mathbf{q})$ only if the equation $\mathcal{W}(\omega, \mathbf{q})\phi = \eta$ has a nonzero solution. According to the assumption of uniqueness of $\tilde{E}$ stated above, this solution is unique if it exists. We will say that ϕ is given in this case by the inverse of $\mathcal{W}$, viz, $\phi = \mathcal{W}^{-1}\eta$. Then it follows from (A3b) that $F = -E_0\mathcal{W}^{-1}\eta$. Substituting this expression into (A3a), we arrive at the dispersion relation (2) in which $\Sigma(\omega, \mathbf{q})$ is given by the following basis-independent expression:

$$\Sigma(\omega, \mathbf{q}) = \epsilon_0 - \langle \eta(\mathbf{r})\mathcal{W}^{-1}(\omega, \mathbf{q})\eta(\mathbf{r}) \rangle_{\mathbb{C}} \ . \qquad (A4)$$

2. Solutions with $E_0 = 0$. There can exist another class of solutions, that is, Bloch waves with a zero fundamental harmonic. For such solutions to exist, the equation $\mathcal{W}(\omega, \mathbf{q})\phi = 0$ must have a nontrivial solution such that $\langle \eta(\mathbf{r})\phi(\mathbf{r}) \rangle_{\mathbb{C}} = 0$. In this case, $\mathcal{W}(\omega, \mathbf{q})$ is singular. Solutions of this kind exist in the case of a homogeneous medium that is artificially discretized on an arbitrary lattice (see below). However, solutions with $E_0 = 0$ (or, in practice, with E_0 in some sense very small) can also exist in heterogeneous periodic media with $\eta \neq 0$.

We will say that solutions to the dispersion equation of the first kind generate "true" Γ-point frequencies that result from multiple scattering in inhomogeneous photonic crystals. Solutions of the second kind are due to purely geometrical folding and generate spurious Γ-point frequencies that do not satisfy the condition (3b).

This classification can be illustrated by considering the special case of a homogeneous medium with $\epsilon(\mathbf{r}) = \epsilon_0$ that is artificially discretized on an arbitrary lattice of a finite pitch. In this case, it is easy to see that $\mathcal{W}(\omega, \mathbf{q})$ is singular if $(\mathbf{q} + \mathbf{g})^2 = k^2\epsilon_0$ where $\mathbf{g} \neq 0$ is any *nonzero* reciprocal lattice vector, and invertible otherwise. Therefore, the dispersion equation for the solutions of the first kind is of the form

$$q^2 = k^2\epsilon_0 \ , \quad \Sigma(\omega, \mathbf{q}) = \epsilon_0 \ \ (\text{first kind}) \ .$$

As could be expected, the corresponding lattice-periodic function $\tilde{E}(\mathbf{r}) = E_0$ is just a constant. Thus, the fundamental harmonic is dominating in this solution.

Solutions of the second kind are of the form

$$(\mathbf{q}+\mathbf{g})^2 = k^2\epsilon_0 \ \text{ for } \mathbf{g}\neq 0 \ \text{ (second kind) .} \quad (A5)$$

For any pair $(\omega,\mathbf{q})$ satisfying the above condition, $\mathcal{W}(\omega,\mathbf{q})$ is singular and we have $\langle\tilde{E}\rangle_{\mathbb{C}} = 0$. This lattice-periodic function is dominated by higher-order harmonics.

We can investigate the spurious Γ-point frequencies generated by the solution of the second kind as follows. If $\mathrm{Im}(\epsilon_0) = 0$, equation (A5) is satisfied by a pair $(\omega_{\mathbf{p}},0)$ where $\omega_{\mathbf{p}}^2 = (cp)^2/\epsilon_0$ and $\mathbf{p}\neq 0$ is any nonzero reciprocal lattice vector. The isofrequency line at $\omega = \omega_{\mathbf{p}}$ is then obtained from the equation

$$(\mathbf{q}+\mathbf{g})^2 = p^2 > 0 \ \text{ for any } \mathbf{g}\neq 0 \ . \quad (A6)$$

Consider for simplicity a square lattice and let $\epsilon_0 > 0$ and $\mathbf{p} = (2\pi/a)(0,1)$. The corresponding quantity $\omega_{\mathbf{p}} = 2\pi c/a\sqrt{\epsilon_0}$ is the first spurious Γ-point frequency out of the infinite sequence. In principle, equation (A6) defines an infinite number of curves but, at the first Γ-point frequency, most of these curves coincide. To obtain the complete solution, we can start by taking $\mathbf{g} = \mathbf{p}$. It can be seen that (A6) defines in this case a circular arc in the (q_x, q_y) plane, and that this arc crosses the origin. On a square lattice, there are four reciprocal vectors of the same length $2\pi/a$ and pointing in the directions $(0,\pm 1)$ and $(\pm 1,0)$. As a result, the isofrequency line contains four circular arcs intersecting at the origin (six arcs in the case of a triangular lattice) as is shown in Fig. 1(a). The arcs are truncated and reflected at the edge of the FBZ. These additional "reflected" segments of the isofrequency line can be obtained by considering additional vectors $\mathbf{p}$ in (A6).

The above result is quite trivial and we could have obtained it without using the mathematical formalism of this Appendix. However, the derivation is useful to show that the formalism developed herein is consistent with the limit of zero contrast. More importantly, it provides us with a clue how to treat the case when E_0 is small but nonzero. Let $\gamma = \|F/E_0\|_2$ where $\|\cdot\|_2$ is the L_2 norm. We believe that it is physically meaningful to consider the phase and group velocities of a Bloch wave only if γ is sufficiently small so that the fundamental harmonic of the Bloch wave is in some sense dominant. This condition holds if the smallest singular value of $\mathcal{W}$ is sufficiently far from zero. If this is not so, then γ can be very large or even diverge. In this case, introducing the characteristics of a plane wave such as the phase and group velocities is devoid of physical meaning, even if this can be done formally by considering the dependence $q(\omega)$.

We therefore conjecture that there are two fundamentally different regimes of propagation, $\gamma \ll 1$ and $\gamma \gg 1$, and in the second regime the group velocity computed as, say, $\nabla_{\mathbf{q}}\omega$ does not correspond to any physically measurable quantity and should not be invested with any particular interpretation. It is not clear, though, how to treat the borderline case $\gamma \sim 1$; perhaps, it can only be investigated numerically.

It should be noted that the above discussion is applicable to the case of s-polarization when the scalar electric field is smooth. For the p-polarization, the electric field can jump at the discontinuities of $\tilde{\epsilon}(\mathbf{r})$, and a more careful analysis is required. We also note that the condition on γ stated above is closely related to the concept of the smooth field introduced in Ref. 26.

Appendix B: Dispersion equation and perturbation theory in the basis of plane waves

In this Appendix, we derive equation (2) from (21), define the function $\Sigma(\omega,\mathbf{q})$ algebraically using the basis of plane waves and then obtain the expansion of $\Sigma(\omega,\mathbf{q})$ in powers of the Cartesian components of $\mathbf{q}$ for ω in the vicinity of one of the Γ-point frequencies ω_n [as defined by (3)]. We note that a basis-independent definition of $\Sigma(\omega,\mathbf{q})$ is given in Appendix A by equation (A4). In order to derive (A4), we have assumed that the fundamental Bloch harmonic E_0 is non-zero. We will use this assumption in this Appendix as well. We are therefore restricting ourselves to the solutions of the first kind according to the terminology of Appendix A.

As was done by us previously[20,26], we consider the equation (21) for the cases $\mathbf{g} = 0$ and $\mathbf{g}\neq 0$ separately. Since we assume that $E_0\neq 0$, we can scale all coefficients $E_{\mathbf{g}}$ by E_0. Let $e_{\mathbf{g}} = E_{\mathbf{g}}/E_0$, so that $e_0 = 1$. First, we take $\mathbf{g} = 0$ in (21). This results in

$$q^2 = k^2\left[\epsilon_h + \rho\chi\sum_{\mathbf{p}}M(-\mathbf{p})e_{\mathbf{p}}\right]$$

$$= k^2\left[\epsilon_h + \rho\chi M(0)e_0 + \rho\chi\sum_{\mathbf{p}\neq 0}M(-\mathbf{p})e_{\mathbf{p}}\right] \ . \ (B1)$$

Since $M(0)e_0 = 1$ and $\chi = \epsilon_i - \epsilon_h$, we see that (B1) is equivalent to (2) in which

$$\Sigma(\omega,\mathbf{q}) = \epsilon_0 + \rho\chi\sum_{\mathbf{g}\neq 0}M(-\mathbf{g})e_{\mathbf{g}} \quad (B2)$$

and $\epsilon_0 = \rho\epsilon_i + (1-\rho)\epsilon_h$ is the cell-averaged permittivity of the PC, the same quantity as the zeroth component in the Fourier expansion (14).

We now must show that the expression in the right-hand side of (B2) is a well-defined quantity and a function of the two arguments ω and $\mathbf{q}$. To this end we note that this expression contains the relative amplitudes $e_{\mathbf{g}} = E_{\mathbf{g}}/E_0$ with $\mathbf{g}\neq 0$. Therefore, to compute $\Sigma(\omega,\mathbf{q})$ algebraically, we must consider (21) for $\mathbf{g}\neq 0$. This yields the following set of equations

$$[(\mathbf{q}+\mathbf{g})^2 - k^2\epsilon_h]e_{\mathbf{g}} = \rho\chi k^2$$

$$\times\left[M(\mathbf{g}) + \sum_{\mathbf{p}\neq 0}M(\mathbf{g}-\mathbf{p})e_{\mathbf{p}}\right] \ , \ \ \mathbf{g}\neq 0 \ . \ (B3)$$

It is important to note that (B3) is a closed (albeit an infinite) set of equations with a nonzero free term $\rho\chi k^2 M(\mathbf{g})$. Consequently, (B3) is not an eigenproblem but rather a linear set of equations that can be, in principle, solved by matrix inversion. This proves that (B2) is a mathematically-consistent definition of the function $\Sigma(\omega, \mathbf{q})$.

To complete the algebraic definition of $\Sigma(\omega, \mathbf{q})$, we can proceed as follows. Define the matrices and vectors:

$$D_{\mathbf{gp}} = \left(g^2 - k^2 \epsilon_h\right) \delta_{\mathbf{gp}} ,$$
$$Q_{\mathbf{gp}} = \left(q^2 + 2\mathbf{q} \cdot \mathbf{g}\right) \delta_{\mathbf{gp}} ,$$
$$M_{\mathbf{gp}} = M(\mathbf{g} - \mathbf{p}) ,$$
$$b_{\mathbf{g}} = M(\mathbf{g}) .$$

Here all indexes are restricted to $\mathbf{g} \neq 0$ and $\mathbf{p} \neq 0$. Also, only the matrix Q depends on $\mathbf{q}$. We have separated the diagonal term Q from the $\mathbf{q}$-independent diagonal term D because we are interested in building a perturbation theory in Cartesian components of $\mathbf{q}$. Also note the normalization rule $\langle b|b\rangle = 1/\rho - 1$. We can now re-write (B3) as follows:

$$(Q + D - \rho\chi k^2 M)|e\rangle = \rho\chi k^2 |b\rangle . \qquad (B4)$$

From this we find that

$$\Sigma(\omega, \mathbf{q}) = \epsilon_0 + (\rho\chi k)^2 \langle b|(D + Q - \rho\chi k^2 M)^{-1}|b\rangle .$$

This expression gives a closed-form algebraic definition of $\Sigma(\omega, \mathbf{q})$.

We now proceed with building the perturbation theory. We assume that (i) a higher-order Γ-point frequency ω_n exists according to the definition (3) and (ii) ω is in the vicinity and just below ω_n (in the pass-band). Under this condition, $\Sigma(\omega, 0)$ is small but nonzero and positive. Otherwise, the frequency ω would be is inside a bandgap. We can define the T-matrix of the problem as follows:

$$T = \left(D - \rho\chi k^2 M\right)^{-1} . \qquad (B5)$$

Note that T is the matrix that we need in order to compute $\Sigma(\omega, 0)$. Indeed, for $\mathbf{q} = 0$, we also have $Q = 0$ and

$$\Sigma(\omega, 0) = \epsilon_0 + (\rho\chi k)^2 \langle b|T|b\rangle$$
$$= \epsilon_0 + (\rho\chi k)^2 \sum_{\mathbf{g}_1, \mathbf{g}_2 \neq 0} M(-\mathbf{g}_1) T_{\mathbf{g}_1 \mathbf{g}_2} M(\mathbf{g}_2) .$$

Let us assume that $T = T(\omega)$ has been computed at the working frequency ω by inverting matrix in the right-hand side of (B5). Computationally, this requires truncation of the basis and one matrix inversion operation. Now, the equation we intend to iterate is of the form

$$|e\rangle = \rho\chi k^2 T|b\rangle - TQ|e\rangle ,$$

which follows directly from (B4) and (B5). The formal power series solution is of the form

$$|e\rangle = \rho\chi k^2 \sum_{n=0}^{\infty} (-TQ)^n T|b\rangle ,$$

and for Σ, we have

$$\Sigma = \epsilon_0 + (\rho\chi k)^2 \sum_{n=0}^{\infty} \langle b|(-TQ)^n T|b\rangle$$
$$= \epsilon_0 + (\rho\chi k)^2 \sum_{n=0}^{\infty} \sigma_n ,$$

where

$$\sigma_n = (-1)^n \langle b|(TQ)^n T|b\rangle .$$

Note that the order in Q is not the same as the order in $\mathbf{q}$. Also note that terms of the form

$$(-1)^j q^{2j} \langle b|T^{j+1}|b\rangle \qquad (B6)$$

are generated in all orders of the expansion. These terms are, of course, perfectly isotropic. However, starting from fourth order, more complicated terms appear in the expansion.

In the formulas of this Appendix, we will use the following notations:

(i) The symbol $\circ$ denotes direct (Hadamard) product of two matrices, e.g., $(A \circ B)_{ij} = A_{ij} B_{ij}$.

(ii) The matrices G, G_{xx}, etc., are defined in terms of the reciprocal lattice vectors as follows:

$$(G_{\alpha\beta})_{\mathbf{gp}} = g_\alpha p_\beta , \quad \alpha, \beta = x, y ;$$
$$G_{\mathbf{gp}} = \sum_\alpha G_{\alpha\alpha} = \mathbf{g} \cdot \mathbf{p} .$$

(iii) In all expressions shown below we imply that, for example, $q^6 = (q_x^2 + q_y^2)^3$, etc.

a. Zeroth order. We start with zero order ($n = 0$), which is rather trivial:

$$\sigma_0 = \langle b|T|b\rangle = \sum_{\mathbf{g}_1, \mathbf{g}_2} M(-\mathbf{g}_1) T_{\mathbf{g}_1 \mathbf{g}_2} M(\mathbf{g}_2) .$$

Thus, σ_0 is a $\mathbf{q}$-independent constant. Obviously, $\Sigma(\omega, 0) = \epsilon_0 + (\rho\chi k)^2 \sigma_0(k)$, where we have indicated the dependence of σ_0 on k explicitly.

b. First order. Here we have

$$\sigma_1 = - \sum_{\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}} M(-\mathbf{g}_1) T_{\mathbf{g}_1 \mathbf{g}} \left(q^2 + 2\mathbf{q} \cdot \mathbf{g}\right)$$
$$\times T_{\mathbf{g}\mathbf{g}_2} M(\mathbf{g}_2) .$$

The contribution of the second term in the brackets is zero due to the symmetry $[\sigma_n(-\mathbf{q}) = \sigma_n(\mathbf{q})]$. Therefore,

$$\sigma_1 = -q^2 \sum_{\mathbf{g}_1, \mathbf{g}_2} M(-\mathbf{g}_1) T_{\mathbf{g}_1 \mathbf{g}_2}^2 M(-\mathbf{g}_2)$$
$$= -q^2 \langle b|T^2|b\rangle .$$

c. Second order. Here we have

$$\sigma_2 = \sum_{\{\mathbf{g}\}} M(-\mathbf{g}_1) T_{\mathbf{g}_1\mathbf{g}_2} T_{\mathbf{g}_2\mathbf{g}_3} T_{\mathbf{g}_3\mathbf{g}_4} M(\mathbf{g}_4)$$
$$\times \left[q^2 + 2(\mathbf{q} \cdot \mathbf{g}_2) \right] \left[q^2 + 2(\mathbf{q} \cdot \mathbf{g}_3) \right] ,$$

where $\sum_{\{\mathbf{g}\}}$ denotes summation over all relevant indexes. We now expand the product of the square brackets and notice that the terms linear in $\mathbf{q}$ sum to zero for the reason of inversion symmetry. The two terms that produce nonzero result upon summation are

$$q^4 \quad \text{and} \quad 4(\mathbf{q} \cdot \mathbf{g}_2)(\mathbf{q} \cdot \mathbf{g}_3) .$$

The first of these generates the result of the form (B6) and the second term requires some additional consideration. We can write

$$(\mathbf{q} \cdot \mathbf{g}_2)(\mathbf{q} \cdot \mathbf{g}_3) = (q_x g_{2x} + q_y g_{2y})(q_x g_{3x} + q_y g_{3y})$$
$$= q_x^2 g_{2x} g_{3x} + q_y^2 g_{2y} g_{3y} + q_x q_y (g_{2x} g_{3y} + g_{2y} g_{3x}) .$$

The term proportional to $q_x q_y$ sums to zero due the inversion symmetry. Moreover, from the symmetry properties of both triangular and square lattices, we find that summation of the coefficients in front of q_x^2 and q_y^2 must yield the same result. Indeed, if this were not so, we would have obtained a term of the form $\beta_x q_x^2 + \beta_y q_y^2$ with $\beta_x \neq \beta_y$, describing an ellipse of unequal semiaxes, which is inconsistent with both C_4 and C_6 symmetries. Therefore, we can replace the above expression (inside the summation) by

$$\frac{1}{2}(q_x^2 + q_y^2)(g_{2x} g_{3x} + g_{2y} g_{3y}) = \frac{1}{2} q^2 (\mathbf{g}_2 \cdot \mathbf{g}_3) .$$

Collecting everything together, we find that

$$\sigma_2 = q^4 \langle b|T^3|b \rangle + 2q^2 \langle b|T(T \circ G)T|b \rangle .$$

We thus see that all terms generated in the second order are still circularly-symmetric.

d. Third order. Here we have

$$\sigma_3 = -\sum_{\{\mathbf{g}\}} M(-\mathbf{g}_1) T_{\mathbf{g}_1\mathbf{g}_2} T_{\mathbf{g}_2\mathbf{g}_3} T_{\mathbf{g}_3\mathbf{g}_4} T_{\mathbf{g}_4\mathbf{g}_5} M(\mathbf{g}_5)$$
$$\times \left[q^2 + 2(\mathbf{q} \cdot \mathbf{g}_2) \right] \left[q^2 + 2(\mathbf{q} \cdot \mathbf{g}_3) \right] \left[q^2 + 2(\mathbf{q} \cdot \mathbf{g}_4) \right] .$$

After expanding the brackets and keeping only the terms that do not sum to zero, we find that $\sigma_3 = \sigma_3^{(a)} + \sigma_3^{(b)}$ where

$$\sigma_3^{(a)} = -q^6 \langle b|T^4|b \rangle \qquad (B7)$$

is of the form (B6) and

$$\sigma_3^{(b)} = -4q^2 \sum_{\{\mathbf{g}\}} M(-\mathbf{g}_1) T_{\mathbf{g}_1\mathbf{g}_2} T_{\mathbf{g}_2\mathbf{g}_3} T_{\mathbf{g}_3\mathbf{g}_4} T_{\mathbf{g}_4\mathbf{g}_5} M(\mathbf{g}_5)$$
$$\times \left[(\mathbf{q} \cdot \mathbf{g}_2)(\mathbf{q} \cdot \mathbf{g}_3) + (\mathbf{q} \cdot \mathbf{g}_2)(\mathbf{q} \cdot \mathbf{g}_4) + (\mathbf{q} \cdot \mathbf{g}_3)(\mathbf{q} \cdot \mathbf{g}_4) \right] .$$

We can use the same transformation as was used in the second order to transform the terms of the form $(\mathbf{q} \cdot \mathbf{g}_2)(\mathbf{q} \cdot \mathbf{g}_3)$ to the form $(1/2)q^2(\mathbf{g}_2 \cdot \mathbf{g}_3)$. We thus obtain

$$\sigma_3^{(b)} = 2(qh)^4 \Big[\langle b|T^2(T \circ G)T|b \rangle$$
$$+ \langle b|T(T^2 \circ G)T|b \rangle + \langle b|T(T \circ G)T^2|b \rangle \Big] . \quad (B8)$$

The first and last terms in this expression are in fact equal but are written separately for symmetry of expression. Still, all expressions arising to third order in Q are circularly-symmetric.

e. Fourth order. In the fourth order, we can write

$$\sigma_4 = \sigma_4^{(a)} + \sigma_4^{(b)} + \sigma_4^{(c)} ,$$

where the expressions for $\sigma_4^{(a)}$ and $\sigma_4^{(b)}$ are obtained in the manner very similar to what was done above. Omitting the intermediate steps, we write the final result for these two terms, viz,

$$\sigma_4^{(a)} = q^8 \langle b|T^5|b \rangle \qquad (B9)$$

$$\sigma_4^{(b)} = 2q^6$$
$$\Big[\langle b|T(T \circ G)T^3|b \rangle + \langle b|T^3(T \circ G)T|b \rangle$$
$$+ \langle b|T(T^2 \circ G)T^2|b \rangle + \langle b|T^2(T^2 \circ G)T|b \rangle$$
$$+ \langle b|T^3(T \circ G)T|b \rangle + \langle b|T^2(T \circ G)T^2|b \rangle \Big] . \quad (B10)$$

Note that the terms appearing on each line of the above expression are pair-wise equal. The expressions (B9) and (B10) could, in fact, be anticipated and are directly analogous to the expressions (B7) and (B8). A simple diagrammatic technique can be devised to generate similar expressions that appear in the higher orders of the perturbation theory.

However, $\sigma_4^{(c)}$ is a term of a different kind and it is the first term we encounter that does not obey the circular symmetry and is consistent with C_4 (but not C_6) symmetry. The term is

$$\sigma_4^{(c)} = 16 \sum_{\{\mathbf{g}\}} M(-\mathbf{g}_1) T_{\mathbf{g}_1\mathbf{g}_2} \; \cdots \; T_{\mathbf{g}_5\mathbf{g}_6} M(\mathbf{g}_6)$$
$$\times (\mathbf{q} \cdot \mathbf{g}_2)(\mathbf{q} \cdot \mathbf{g}_3)(\mathbf{q} \cdot \mathbf{g}_4)(\mathbf{q} \cdot \mathbf{g}_5) . \quad (B11)$$

We can not use the same trick as was used above to transform the factor

$$P = (\mathbf{q} \cdot \mathbf{g}_2)(\mathbf{q} \cdot \mathbf{g}_3)(\mathbf{q} \cdot \mathbf{g}_4)(\mathbf{q} \cdot \mathbf{g}_5)$$

to the form βq^4 where β is a scalar expressible in terms of the dot products $\mathbf{g}_i \cdot \mathbf{g}_j$. Indeed, let us write the factor P in terms of Cartesian components of all vectors involved:

$$P = (q_x g_{2x} + q_y g_{2y})(q_x g_{3x} + q_y g_{3y})$$
$$\times (q_x g_{4x} + q_y g_{4y})(q_x g_{5x} + q_y g_{5y})$$
$$= P' + q_x^4 g_{2x} g_{3x} g_{4x} g_{5x} + q_y^4 g_{2y} g_{3y} g_{4y} g_{5y}$$

$$+q_x^2 q_y^2 \big[g_{2x} g_{3x} g_{4y} g_{5y} + g_{2y} g_{3y} g_{4x} g_{5x}$$
$$+ g_{2x} g_{3y} g_{4x} g_{5y} + g_{2y} g_{3x} g_{4y} g_{5x}$$
$$+ g_{2x} g_{3y} g_{4y} g_{5x} + g_{2y} g_{3x} g_{4x} g_{5y} \big] \; .$$

Here P' is the term that sums to zero by symmetry. We can introduce the notations $\Delta = P - P'$ and Π_x, Π_y, Π_{xy} (definition of the last three quantities will be clear from the next equation) and write

$$\Delta = \Pi_x q_x^4 + \Pi_y q_y^4 + \Pi_{xy} q_x^2 q_y^2$$

Now, if it happens so that

$$\sum_{\{\mathbf{g}\}} F[\{\mathbf{g}\}] \Pi_x = \sum_{\{\mathbf{g}\}} F[\{\mathbf{g}\}] \Pi_y = \frac{1}{2} \sum_{\{\mathbf{g}\}} F[\{\mathbf{g}\}] \Pi_{xy} \; ,$$

where $F[\{\mathbf{g}\}]$ is the coefficient appearing on the first line of (B11), then we would obtain the result $\sigma_4^{(c)} = \beta q^4$. This is what we can expect to happen in triangular lattices with C_6 symmetry. However, there is no general or obvious reason why the second equality in the above expression should hold in the case of C_4 symmetry, and

there are sufficient grounds to believe that it does not. As a result, the function $\Delta(\mathbf{q})$ and the summation result $\sum_{\{\mathbf{g}\}} F[\{\mathbf{g}\}] \Delta(\mathbf{q})$ are not circularly-symmetric.

We can write the result for $\sigma_4^{(c)}$ in a form similar to that used in lower orders if we account for the identity $q_x^4 + q_y^4 = q^4 - 2q_x^2 q_y^2$. Then

$$\sigma_4^{(c)} = 8 \left[q^4 - 2(q_x^2 q_y^2) \right]$$
$$\times \left[\langle b | T(T \circ G_x) T(T \circ G_x) T | b \rangle + (G_x \to G_y) \right]$$
$$+ q_x^2 q_y^2$$
$$\times \Big[\langle b | T(T \circ G_x) T(T \circ G_y) T | b \rangle + (G_x \leftrightarrow G_y)$$
$$+ \langle b | T(T \circ G_{xy}) T(T \circ G_{xy}) T | b \rangle + (G_{xy} \to G_{yx})$$
$$+ \langle b | T(T \circ G_{xy}) T(T \circ G_{yx}) T | b \rangle + (G_{xy} \leftrightarrow G_{yx}) \Big] \; .$$

The terms of the form (9) appear only in the sixth order of the perturbation theory. The corresponding coefficients are very complicated and we do not compute them here.