Theory of light refraction at the surface of a photonic crystal

Wei Jiang and Ray T. Chen
Department of Electrical and Computer Engineering and Microelectronics Research Center, the University of Texas, Austin, Texas 78758, USA

Xuejun Lu
Department of Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell, Massachusetts 01854, USA

(Received 12 February 2004; revised manuscript received 2 February 2005; published 28 June 2005)

A rigorous theory is developed for light refraction by a photonic crystal (PC) with arbitrary lattice-type and surface orientation. First, the refraction of a planar wave incident upon a photonic crystal surface is analyzed. We rigorously prove the equal partition of propagating PC modes by a surface under a general condition. The concept of surface-orientation-dependent mode degeneracy has been proposed and its relationship to quasiperiodic surfaces unfolded. With modes partitioned and the degeneracy properly recognized, a subset of the solved PC modes is identified to uniquely represent all PC modes that can be excited by an incident wave. The refraction problem can thus be rigorously solved in the plane-wave formulation. Essentially, we need to solve the field in only a single cell on the surface to solve the refraction problem. We further discuss the case where a Bloch wave illuminates the surface from inside a photonic crystal, which enables us to compute the transmissions along a complete light path through a series of interfaces. In addition, the transmission of a Gaussian beam is discussed, and the insertion loss formulas are presented. Other realistic beam profiles are discussed for designing photonic crystal devices. With all these issues solved, a complete theoretical framework of the photonic crystal refraction and transmission has thus been established. The theory has been applied to design a wavelength-division multiplexing demultiplexer that exhibits lower than 3 dB loss over a 25-nm spectrum. In examining the refraction by a quasiperiodic surface, a slight change of surface orientation is predicted to split one beam into an infinite number of beams.

DOI: 10.1103/PhysRevB.71.245115 PACS number(s): 42.70.Qs, 42.25.Gy, 42.79.Sz

I. INTRODUCTION

Anomalous refraction of light at an interface between a photonic crystal[1,2] and a homogeneous medium was experimentally studied recently.[3,4] The refraction angle was found to be sensitive to the incident angle and wavelength in certain cases. These phenomena were called the "superprism" effect accordingly and were proposed to be utilized for wavelength-division multiplexing (WDM) and other applications.[3–11]

The physics behind the anomalous refraction is related to the coupling of the incident light with the propagating modes of a photonic crystal (PC). Although the light beam direction inside the PC can be easily determined from the dispersion surface, other characteristics of photonic crystal refraction deserve further study. Particularly important is the amplitude of each excited mode, which is commonly calculated using finite-difference time-domain (FDTD) techniques.[7–9] However, FDTD simulations are often time consuming and, in many cases, prohibitive for studying anomalous refraction. To study such an effect characterized by high wavelength and angular sensitivity, fine spatial and temporal grids and a large simulation region are inevitable and are frequently beyond the capacity of commonly available computing facilities. In addition, FDTD simulations give the insertion loss values case by case; they cannot accurately reveal some general features or trends when certain parameters vary continuously. Particularly, the effect of varying surface orientation is rarely addressed,[12] and refraction at a surface of a PC with a *general* lattice type is rarely seen in the literature. A general PC refraction theory that can handle any type of lattice and any surface orientation is of fundamental importance from a basic research point of view. Such a theory also provides great flexibility in exploring novel device geometries proposed for valuable applications.[7]

A variety of numerical and theoretical methods[13–22] other than FDTD have been employed to study the transmitted optical-field amplitude through photonic crystals including the transfer matrix method,[14–16] the scattering theory of lattices of dielectric cylinders or spheres[17–20] or multiple-scattering method,[10] and the internal-field expansion method.[21,22]

Using these methods, plane-parallel slabs have been extensively studied for their transmission spectra,[14,17–19,21,22] which exhibit interference features due to multiple reflections between two surfaces.[21,22] Stefanou *et al.* discussed complex band structures and evanescent modes in a slab.[18,19] For frequencies inside a photonic band, Sakoda also demonstrated the possible existence of uncoupled modes owing to mirror symmetry.[21,22] Planar incident waves have been assumed in these calculations. In realistic situations, due to the finite width of the incident beam, the interference may not be in effect or may not occur in a way that resembles the planar-incident-wave case. For instance, the secondary beams generated in the slab by successive internal reflections may not overlap in space and, therefore, do not interfere with each other. In addition, there are many practically valuable cases where the entry and exit surfaces are not parallel to each

other.[7] Study of the single-surface transmission problem is necessary to understand these diverse situations, which could frequently arise in design of valuable devices.[7] Note that the single-surface transmission problem for a photonic crystal is nothing but refraction. Moreover, a PC slab transmission theory can be obtained from a refraction theory (as done in conventional thin-film interference theory), but not reversely. As Li and Ho pointed out, for a planar incident wave (of infinite beam width), the mathematical solution for the internal field deep in an extremely thick slab does not converge to the true solution of the field in a semi-infinite photonic crystal because multiple reflections always exist in the slab regardless of the separation between its surfaces.[16]

The refraction problem has been discussed for cubic lattices, with the surface orientations limited to simple cases such as (100) or (111).[15,20] The latest work by Li and Ho showed convincing results that the transfer matrix method calculated the transmission through (10) surfaces of two-dimensional (2D) rectangular lattices with success.[16] In their work, wave propagation in a semi-infinite "photonic crystal" was also studied for other important purposes such as analyzing endface coupling into a photonic crystal waveguide and waveguide-cavity resonant coupling. Based on a supercell approach, the waveguide endface coupling problem was formulated as transmission through the surface of a "photonic crystal" that has a rectangular lattice (i.e., the supercell is rectangular). We also note that Botten *et al.* earlier calculated reflection coefficients for semi-infinite photonic crystals of these special surface orientations.[23] Moreover, in prior studies, the surface is often parallel to a mirror symmetry plane of the lattice.

We shall emphasize that behind these apparent lattice, symmetry, and orientation limitations of these prior theories are more fundamental limitations related to the understanding of the photonic crystal refraction phenomena. Some fundamental issues of the photonic crystal refraction do not emerge if we consider only these simple lattice types and surface orientations. First, in order to satisfy the boundary conditions, the number of eigenmodes propagating in the forward direction (determined with respect to the incident beam) must be the same as the number of modes propagating backwards. In many prior works, this was usually tacitly assumed, but not proved. Such an assumption is usually valid for the (10), (01), or (100) type of surfaces of a rectangular or cubic lattice or other cases where the surface is parallel to a mirror symmetry plane of the crystal.[24] But it is not necessarily true for an arbitrary lattice that has an arbitrary surface orientation. In the study of photonic crystals, this problem has been recognized,[23] though not solved for cases where the mirror symmetry is absent. Second, as we shall see, when the surface orientation is not as simple as {10} or {100}, the degeneracy of the photonic crystal eigenmodes will often be modified. To understand the degeneracy, we need to consider the eigenmodes not only in the first Brillouin zone (BZ), but in the whole reciprocal space. Adding to the complexity are the cases where the surface cannot be described by regular Miller indices. In such cases, the surface is quasiperiodic and the degeneracy may completely disappear. This aspect of the refraction problem has not been discussed in the literature. As we shall see, only when *analytically studying the refraction problem for an arbitrary lattice and an arbitrary surface orientation* do some fundamental physics issues of PC refraction begin to unfold.

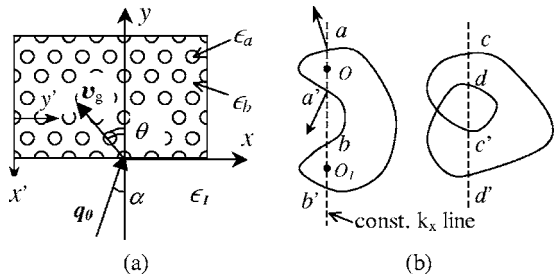

FIG. 1. Schematic drawings: (a) A planar wave illuminates a photonic crystal surface. The dielectric constant of the incident medium is ϵ_1. For the photonic crystal, the background medium has a dielectric constant ϵ_b, whereas the materials inside the columns have a dielectric constant ϵ_a. Each surface has its own local coordinate system for refractions through a series of surfaces. (b) Conceptual dispersion surfaces in 2D. The vertical lines are constant-k_x lines. Note that the contours in the right graph may not be realistic.

This paper will introduce a general, rigorous PC refraction theory that can deal with any lattice type and surface orientation.[25] Two key issues—the partition of the forward and backward modes and the surface-dependent mode degeneracy—will be discussed in detail in Sec. II. The natural emergence of a quasiperiodic surface[26,27] will be discussed as well as its relation to the surface-dependent mode degeneracy. Refraction of realistic beams with finite beam widths will be discussed in Sec. III. A schematic design of a WDM demultiplexer will be presented with low-loss and high-wavelength resolution in Sec. IV, where the physical interpretation and the numerical aspect of the current theory will be discussed. This paper is primarily focused on 2D PC's, but most conclusions can be generalized to 3D easily. Though the theory itself is rigorous for a single-surface (or semi-infinite) problem, to apply this theory to realistic devices often requires us to make certain assumptions to simplify the problem, as in any physics modeling. These simplifications or approximations and the accompanied limitations will be discussed in Sec. IV B.

Part of the theory utilizes some knowledge of topology. We have made a thorough effort to present it in a way accessible to most physics researchers in the main text, leaving more complicated discussions to one of the Appendixes.

II. REFRACTION THEORY FOR A PLANAR INCIDENT WAVE

Consider a TM wave (electric field E normal to the plane) incident upon a 2D PC as illustrated in Fig. 1(a). The incident light is a planar wave $e^{i\mathbf{q}_0\cdot\mathbf{x}}$ with frequency ω. By Fourier expansion, the field equation in PC is (let the speed of light $c=1$)

$$-[(k_x+G_x)^2+(k_y+G_y)^2]E(\mathbf{G})+\omega^2\sum_{\mathbf{G}'}\epsilon(\mathbf{G}-\mathbf{G}')E(\mathbf{G}')=0, \tag{1}$$

where $k_x \equiv q_{0x}$ and $E(\mathbf{G})$ and $\epsilon(\mathbf{G})$ are the Fourier coefficients of $E(\mathbf{x})$ and PC dielectric function $\epsilon(\mathbf{x})$, respectively; $\mathbf{G}$ is a reciprocal lattice vector, explicitly

$$\mathbf{G}_{lm} = l\mathbf{b}_1 + m\mathbf{b}_2,$$

where $\mathbf{b}_1$ and $\mathbf{b}_2$ are the basis vectors of the reciprocal lattice. Here we assume that the constituents of the photonic crystal are lossless as desired in many important applications. For convenience, we sometimes refer to the $y<0$ region as region I.

This set of linear equations for $E(\mathbf{G})$ can be converted into a matrix equation

$$[W][E] = (k_y^2[A] + k_y[B] + [C])[E] = 0,$$

where $[A]$, $[B]$, and $[C]$ are matrices whose elements are functions of ω and k_x. The eigenvalues k_y of Eq. (1) must satisfy $\det[W]=0$. The elements of the column vector $[E]$ are

$$E(\mathbf{G}_{lm}), \quad -L \leqslant l \leqslant L, \quad -M \leqslant m \leqslant M.$$

Let

$$N = (2L+1)(2M+1).$$

Clearly, $[W]$ is an $N\times N$ matrix and $[E]$ is an $N\times 1$ matrix. There is a general numerical scheme for such an eigenvalue problem.[28,29] However, for this particular problem, $[B]$ and $[C]$ depend on k_x and ω in a specific way; note that $[A]$ is the identity matrix.

One readily finds that the determinant can be written as a real-coefficient polynomial

$$\det[W] = F(k_x, k_y, \omega) = \sum_{lmn} c_{lmn} k_x^l k_y^m \omega^n,$$

where the degree of each variable is $2N$ or less. Actually, each element of $[W]$ is a polynomial of k_x, k_y, and ω. By the definition of matrix determinant, $\det[W]$ must be a polynomial as well. The fact that all c_{lmn}'s must be real follows from the fact that $\epsilon(\mathbf{G})$ is Hermitian. Note that k_x^2 and k_y^2 appear on every diagonal element of $[W]$ and ω^2 appears in every element; one finds that in general their highest powers in F are all $2N$. The natural continuity of polynomial F is important for the ensuing proof of equal partition of modes.

In principle, the eigenvalue problem amounts to solving $\det[W]=0$ for $2N$ roots $k_y(s)$, $s=1,2,\ldots,2N$, for given ω and k_x. Then each $k_y(s)$ is substituted back into $[W]$ to find the corresponding eigenvector $[E_s]$, which gives the Fourier transform of the eigenmodes $E_s(\mathbf{x})$. We note that each solved eigenmode $E_s(\mathbf{x})$ satisfies the real-space wave equation everywhere inside the PC in Fig. 1(a) despite the presence of a surface.

A. Separation of up and down modes

Only a subset of the solved eigenmodes from Eq. (1) is allowed in the region $y>0$. Complex $k_y(s)$ roots always appear in conjugate pairs (because F has real coefficients), which was known through other approaches, in Ref. 23, for example. They are allowed only if $\mathrm{Im}\, k_y>0$; otherwise, divergence results. As we will see later, the real $k_y(s)$ must be partitioned into an equal number of up-propagating modes ("up mode") and down-propagating modes ("down mode"); otherwise, the boundary equations cannot be satisfied. A similar partition problem arose in the transfer matrix method and was solved only if a lattice has mirror symmetry M_y.[23,30] The problem is not solved for an arbitrary type of lattice and an arbitrary surface orientation. Also we note that in many transfer-matrix-based approaches, the eigenvalues are not k_y's, but $e^{ik_y d}$'s, where d is a certain distance. Because of this, the pairing of complex eigenmodes and the partition problem may be formulated differently in form.

To study this problem, first we derive the expressions of the Poynting vector $\mathbf{S}$ and $\nabla_{\mathbf{k}}\omega$, which will be given in Appendix A. It is interesting to notice that both the Poynting vector and the group velocity are proportional to a weighted sum of wave vectors of the Fourier components

$$\mathbf{S}, \quad \mathbf{v}_g \propto \sum_{\mathbf{G}} (\mathbf{k}+\mathbf{G})|E(\mathbf{G})|^2.$$

However, for the TE wave, such a simple form is absent. The Poynting vector is

$$\mathbf{S} = (4\omega\epsilon_0)^{-1} \sum_{\mathbf{G},\mathbf{G}'} H_s^*(\mathbf{G}) \left(\frac{1}{\epsilon}\right)_{\mathbf{G}\mathbf{G}'} [2\mathbf{k}_s + \mathbf{G} + \mathbf{G}'] H_s(\mathbf{G}'),$$

where $(1/\epsilon)_{\mathbf{G}\mathbf{G}'}$ is a Fourier component of $1/\epsilon(\mathbf{x})$ and $H(\mathbf{G})$ corresponds to the out-of-plane component of the magnetic field.

Define the sets of up modes and down modes as

$$\mathcal{M}^+ = \{\text{PC mode} | S_y \geqslant 0 \text{ or } \mathrm{Im}\, k_y > 0\},$$

$$\mathcal{M}^- = \{\text{PC mode} | S_y \leqslant 0 \text{ or } \mathrm{Im}\, k_y < 0\}.$$

Note that the sign of S_y is meaningful for real k_y modes only, whereas the sign of $\mathrm{Im}\, k_y$ is meaningful for complex-k_y modes only. Now we prove that half of the real-k_y modes have $S_y \geqslant 0$. Due to the relation between $\mathbf{S}$ and $\nabla_{\mathbf{k}}\omega = v_g\mathbf{n}$ ($\mathbf{n}$ is a dispersion surface normal), it amounts to proving that a constant-k_x line always intersects the dispersion surface of a given ω at an equal number of positive n_y and negative n_y points.

Consider the equation $F(k_x,k_y,\omega)=0$ for a given ω. As $L,M\to\infty$, it describes an infinite set of contours, i.e., the dispersion surface repeated in all BZ's, on the (k_x,k_y) plane. Here we temporarily assume that the dispersion surface consists of *closed* curves, which appear to be typical for 2D photonic crystals. According to topology[31] (and intuitively), a straight line must intersect a closed contour an even number of times except for the tangent case.

For an arbitrary contour shown in the left graph of Fig. 1(b), assume the dispersion contour expands as the frequency increases. Establish a *local* polar coordinate system (ρ,ϕ) for the shorter contour segment between a and a' such that the components of $\mathbf{k}$ are

$$k_x = \rho\cos\phi + k_{Ox},$$

$$k_y = \rho\sin\phi + k_{Oy},$$

where k_{Ox} and k_{Oy} are the coordinates of point O. The outward normal is given by

$$\mathbf{n} = \frac{dk_y}{d\phi}\mathbf{e}_x - \frac{dk_x}{d\phi}\mathbf{e}_y. \tag{2}$$

Note that $\mathbf{n}$ need not be a unit vector here. One easily computes

$$n_y = -\frac{d\rho}{d\phi}\cos\phi + \rho\sin\phi.$$

In a regular polar coordinate system $\phi_a = \pi/2$ and $\phi_{a'} = 3\pi/2$; therefore,

$$n_{ya} = \rho_a > 0, \quad n_{ya'} = -\rho_{a'} < 0.$$

If the dispersion contour shrinks as the frequency increases, one may use the inward normal. The signs of n_{ya} and $n_{ya'}$ will both be reversed; hence, they are still opposite. It is not difficult to prove that n_{ya} and $n_{ya'}$ always have opposite signs regardless of the choice of azimuthal direction $\phi=0$.

For contours that have complicated connectivity (although perhaps unrealistic) as in the right graph of Fig. 1(b), the above argument always applies if one associates the neighboring crossings into pairs—for example, (cc') and (dd'). Note that an equally valid pairing is (cd') and $(c'd)$. An accurate account of the pairing scheme can be found in Appendix A. Therefore, $n_y>0$ and $n_y<0$ always appear in pairs no matter how twisted a contour is. This implies a topological nature. There are some undesirable features of the above proof. One issue is that a key step requires $\phi_a=\pi/2$ and $\phi_{a'}=3\pi/2$, which seems to depend on the choice of direction $\phi=0$. Furthermore, the preceding proof seems not applicable to 3D cases. In 3D cases, a section of the dispersion surface could consist of closed contours, but the surface normal vector is likely out of the sectioning plane. Also, the preceding proof is not applicable to open contours. A complete, yet more complicated, proof that addresses all these issues is presented in Appendix B.

The case where a constant-k_x line is tangent to a contour can be understood "dynamically." Consider that the constant-k_x line in the left graph of Fig. 1(b) sweeps to the right from its current position. When this line is tangent to the contour, two intersections with the contour will merge into one. Mathematically, this means that two k_y roots of $F(k_x,k_y,\omega)=0$ become one doubly degenerate root. These two roots were in $\mathcal{M}^+$ and $\mathcal{M}^-$, respectively, before the constant-k_x line sweeps to the tangent position; they should remain in their own sets at the tangent position. In other words, the doubly degenerate roots should be treated as two roots, one in $\mathcal{M}^+$ and one in $\mathcal{M}^-$. A rigorous treatment of this issue shows that more complicated situations must be considered. Around the tangent point $\mathbf{k}_t=(k_{xt},k_{yt})$, the equation $F(k_x,k_y,\omega)=0$ describing the contour for a given frequency ω can be rewritten as $k_x=\kappa(k_y)$, where $\kappa(k_y)$ is a single-valued function. The local expansion gives

$$k_x = k_{xt} + \frac{d\kappa}{dk_y}(k_y - k_{yt}) + \frac{1}{2}\frac{d^2\kappa}{dk_y^2}(k_y - k_{yt})^2 + \cdots.$$

For a given k_x, this equation can be solved for k_y. As the constant-k_x line is tangent to the contour at $\mathbf{k}_t$, we have $d\kappa/dk_y=0$. A doubly degenerate root results if $d^2\kappa/dk_y^2 \neq 0$. Generally, it can be shown that if the order of the first nonvanishing derivative is even at $\mathbf{k}_t$, the degree of degeneracy will be even. On the other hand, if the order is odd (an inflection point), then it can be shown that a single real root persists when the constant-k_x line sweeps across the tangent point and the sign of n_y is the same on both sides. Such inflection points exist in realistic dispersion contours which are nonconvex. An example is the starlike contours in triangular lattices.[3,7]

If it happens to be multiple contours tangent to the constant k_x line at one point in reciprocal space, one can apply the preceding argument to each contour separately and the equal partition still holds.

We shall emphasize that there could be a number of other interesting ways of pairing the eigenvalues,[23] for example, based on their values. But fundamentally, only pairing the eigenvalues based on their derivatives as rigorously proved here leads to the proper separation of up and down modes, because the propagation direction of each mode is determined by the group velocity, not the wave vector $\mathbf{k}_s$.

B. Surface-orientation-dependent degeneracy

Now we find the coupling amplitude t_s for each "up mode." The field in the PC is

$$E(\mathbf{x}) = \sum_{s\in\mathcal{M}^+}\sum_{\mathbf{G}} t_s E_s(\mathbf{G}) e^{i[k_x+G_x)x + i(k_y(s)+G_y]y}.$$

On the surface, such a PC field excites reflected waves $e^{i\mathbf{p}(\mathbf{G})\mathbf{x}}$, where

$$p_x(\mathbf{G}) = q_{0x} + G_x,$$

$$p_y(\mathbf{G}) = -\sqrt{\epsilon_I\omega^2 - q_x(\mathbf{G})}$$

Throughout this section, square roots are taken on the Riemann surface with argument $0\leqslant\phi<2\pi$. Note that if $G_x=G_x'$, then $\mathbf{p}(\mathbf{G})=\mathbf{p}(\mathbf{G}')$ even if $G_y\neq G_y'$. Denoting the number of reflected waves—i.e., the number of *distinct* $\mathbf{p}(\mathbf{G})$ by $N_\mathbf{p}$—one immediately sees that, in general, $N_\mathbf{p}\leqslant N$. The field in the region $y<0$ (region I) is then

$$E_I(\mathbf{x}) = e^{i\mathbf{q}_0\cdot\mathbf{x}} + \sum_{\mathbf{G}} r_{\mathbf{p}(\mathbf{G})} e^{i\mathbf{p}(\mathbf{G})\mathbf{x}}. \tag{3}$$

By matching $E(x,0)$ and $H_x\propto\partial E(x,0)/\partial y$ for each $e^{ip_x(\mathbf{G})x}$ wave, the boundary conditions are given as

$$\delta_{\mathbf{G},0} + r_{\mathbf{p}(\mathbf{G})} = \sum_{s\in\mathcal{M}^+} t_s E_s(\mathbf{G}), \tag{4a}$$

$$p_y(\mathbf{G})[-\delta_{\mathbf{G},0} + r_{\mathbf{p}(\mathbf{G})}] = \sum_{s\in\mathcal{M}^+} t_s[k_y(s)+G_y]E_s(\mathbf{G}). \tag{4b}$$

Evidently, the number of $r_{\mathbf{p}(\mathbf{G})}$ is $N_\mathbf{p}$ and the number of equations in Eqs. (4) is $2N_\mathbf{p}$.

Consider a simple example, the (01) surface of a rectangular lattice—i.e., $\mathbf{a}_1=a\mathbf{e}_x$ and $\mathbf{a}_2=a'\mathbf{e}_y$. The distinct $p_x(\mathbf{G}_{lm})=p_{0x}+lb_1$ depends on l only; hence, $N_\mathbf{p}=2L+1<N$. Apparently, the number of up modes equals N. However, due

to the periodicity of the BZs, all $k_y^{(m)}=k_y+mb_2$, $m=-M$, $-M+1,\ldots,M$ are just one "degenerate" solution. Therefore, the number of distinct up modes (and t_s) is

$$N_+=\frac{N}{2M+1}=2L+1.$$

The number of unknowns $\{r_{\mathbf{p}(\mathbf{G})},t_s\}$ in Eqs. (4) matches the number of equations. In general, because k_x is given (or fixed) in a refraction problem, one easily sees that the degeneracy exists only if the constant-k_x line passes multiple points of the periodic set

$$\mathcal{U}(k_s)=\{\mathbf{k}_s+l\mathbf{b}_1+m\mathbf{b}_2\}.$$

We call such a degeneracy "surface-orientation-dependent degeneracy." The presence of degeneracy and the degree of it need more analysis if $\mathbf{b}_1$, $\mathbf{b}_2$ are not along the x, y axes. We shall show that the degeneracies of the solved eigenmodes are always correlated with the reduction of surface harmonic waves. In general, two cases must be considered.

First, consider the case where the surface of an *arbitrary* lattice has Miller indices (h_1h_2). It turns out the degeneracy also leads to the coincidence $\mathbf{p}(\mathbf{G})=\mathbf{p}(\mathbf{G}')$ for certain $\mathbf{G}$, $\mathbf{G}'$. Without loss of generality, assume $0<h_1<|h_2|$, and h_1 and h_2 are coprime; the basis vectors $\mathbf{a}_1$, $\mathbf{a}_2$ may have *arbitrary* lengths and angles. Define new basis vectors as

$$\mathbf{A}_1=h_2\mathbf{a}_1-h_1\mathbf{a}_2,\quad \mathbf{A}_2=\mathbf{a}_1,$$

$$\mathbf{B}_1=-(1/h_1)\mathbf{b}_2,\quad \mathbf{B}_2=\mathbf{b}_1+(h_2/h_1)\mathbf{b}_2. \tag{5}$$

One can verify that $\mathbf{A}_1$ is the surface basis vector (along x) and $\mathbf{B}_2$ is along the y axis.

Now start over, expand the E field in the cell spanned by $\mathbf{A}_1$ and $\mathbf{A}_2$, and obtain the new Eq. (1) with $\mathbf{G}_{lm}=l\mathbf{B}_1+m\mathbf{B}_2$; then, solve for $k_y(s)$, E_s. For the new $\mathbf{G}_{lm}$, one finds $\mathbf{p}(\mathbf{G}_{lm})\equiv\mathbf{p}(\mathbf{G}_{l0})$. Let $\mathbf{p}_l=\mathbf{p}(\mathbf{G}_{l0})$, and the distinct reflected waves are $e^{i\mathbf{p}_l\mathbf{x}}$; hence, $N_\mathbf{p}=2L+1$. Now Eqs. (4) become

$$(\delta_{l,0}+r_l)=\sum_{s\in\mathcal{M}^+}\sum_m t_sE_s(\mathbf{G}_{lm}), \tag{6a}$$

$$p_{ly}(-\delta_{l,0}+r_l)=\sum_{s\in\mathcal{M}^+}\sum_m t_s[k_y(s)+lB_{1y}+mB_2]E_s. \tag{6b}$$

One readily shows that $k_y(s)$ (real or complex) and $k_y^{(m)}=k_y(s)+mB_2$ lead to identical eigenfunctions in real space: $E_s(\mathbf{x})\equiv E^{(m)}(\mathbf{x})$. Therefore, all of the nondegenerate k_y are in the new one-dimensional Brillouin zone,

$$-B_2/2\leqslant \mathrm{Re}\,k_y<B_2/2,$$

and the number of independent $[E_s]$ (or t_s) is $N_+=N/(2M+1)=2L+1$. Hence, the numbers of equations and unknowns always match for any (h_1h_2) surface by choosing $\mathbf{B}_1$ and $\mathbf{B}_2$ given in Eqs. (5).

Second, if no lattice vector is parallel to the surface, then no reciprocal lattice vector can be orthogonal to the surface, which means the constant-k_x line cannot pass more than one point of the periodic set $\mathcal{U}(\mathbf{k}_s)$. Furthermore, the x components of reciprocal lattice vectors must be different from each other, which means that no two $\mathbf{p}(\mathbf{G})$ (or G_x) coincide. One easily recognizes that the crystal surface is quasiperiodic,[26] or one may say that h_1/h_2 is an irrational number. In such a case, one needs to solve $2N$ equations in Eqs. (4) for $2N$ unknowns.

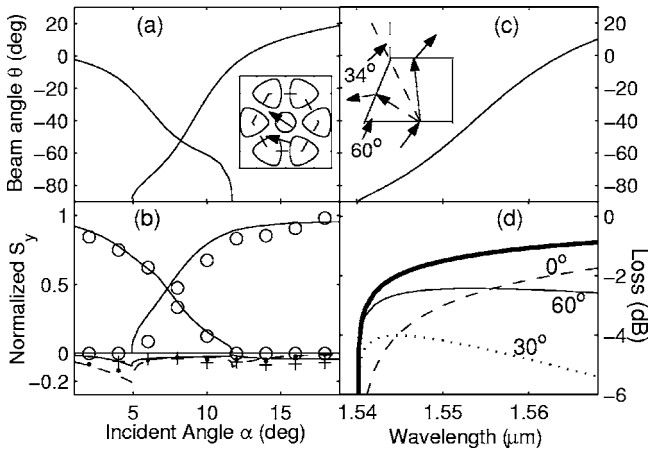

FIG. 2. (a) Refraction angles of two PC modes for a structure $\epsilon_b=\epsilon_I=2.25$, $r/a=0.3$, and $a/\lambda=0.677$. The inset shows the dispersion surface with the BZ drawn by dashed lines. (b) Same structure as in (a), S_y for PC modes ($S_y>0$) and reflected waves ($S_y<0$). FDTD: markers. Theory: lines. (c) Refraction angle of a PC mode for a structure $\epsilon_b=\epsilon_I=12$, $r/a=0.3$, and $a=0.29$ μm, $\alpha=40°$. The inset sketches a demultiplexer design. (d) Insertion losses for the structure in (c). Thick curve: insertion loss after entering PC. Other curves: losses after exiting PC through tilted surfaces.

From a higher point of view, the presence of the surface breaks the discrete translational symmetry. This causes the 2D periodicities in the real and reciprocal spaces to be sectioned along the x and k_y directions, respectively, which may lead to periodic or quasiperiodic 1D sections.[26,27] Furthermore, to calculate the photonic bands of 2D and 3D photonic crystals, one has an infinite number of choices of primitive translation vectors (or basis vectors of unit cell) giving equivalent results. However, in the refraction problem, the presence of the surface and the lowered translational symmetry limit the choices of primitive translation vectors to a subset of the choices for an infinite crystal, as shown in Eq. (5).

As we have seen, the degeneracy of eigenmodes and coincidence of reflected wave vectors $\mathbf{p}(\mathbf{G})=\mathbf{p}(\mathbf{G}')$ are always correlated. This will be further discussed in Sec. IV.

For a quasiperiodic surface, the down modes will not have any surface-dependent degeneracy, either. One may easily extend the above discussions to prove this.

C. Numerical examples

In Figs. 2(a) and 2(b), we present the result of our theory for an incident-angle-sensitive case, compared with the FDTD simulation results. Hexagonal lattices with air holes ($\epsilon_a=1$) are used in this work, with $\epsilon(\mathbf{G})$ calculated analytically.[21] In most cases, about $N=120$ planar waves are sufficient to yield an S_y accuracy of 0.3%. The FDTD S_y data are obtained through Eq. (A3) based on the simulated $E(\mathbf{x},t)$. The two sets of data agree well. The slight difference comes

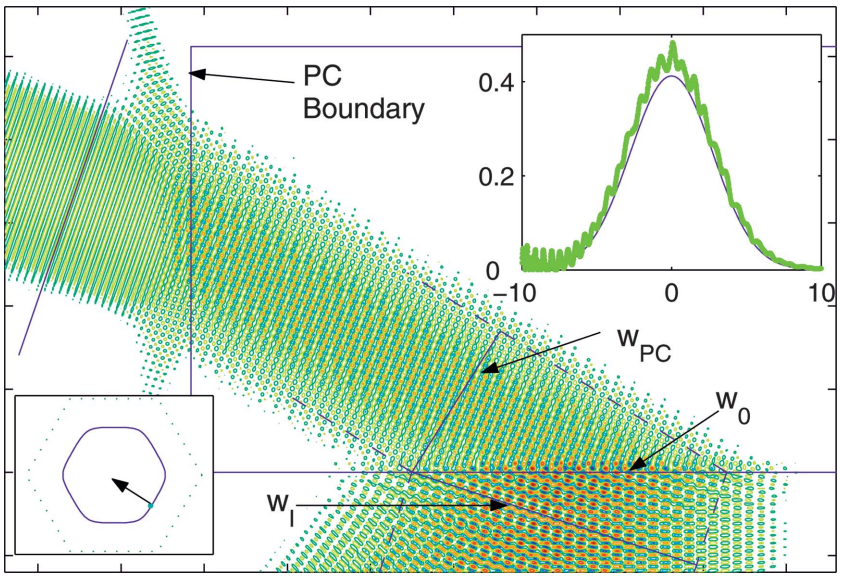

FIG. 3. (Color) Transmission of a Gaussian beam through a PC ($\epsilon_b=\epsilon_I=12, r/a=0.3$). The beam widths w_I, w_{PC}, and w_0 are defined in Eq. (16). The observation plane for the exiting beam is indicated by a line. The top inset shows the exiting beam profile (FDTD, dense dots; theory, line). The bottom inset shows the dispersion surface.

from the FDTD discretization error. A wavelength-sensitive case is presented in Figs. 2(c) and 2(d). Only one PC mode is excited here. The FDTD simulations are almost impossible for this case because all phenomena are due to a 2% change of the wavelength λ, whereas a typical simulation grid spacing is around 2%–3% of λ (or the simulation region will be too small). Our theory does not assume a special shape of scatterers and need not approximate a PC cell by many thin layers. For each incident beam, one essentially solves the field in only one cell per surface. Hence, this theory gives a general, efficient, and accurate calculation method.

D. Light refraction for a Bloch mode exiting a PC

This theory can also calculate the case where a propagating PC mode E_0 with a Bloch vector $\mathbf{k}_0$ is incident upon a boundary from inside a PC (when a beam exits the PC). One readily shows that the boundary conditions must be

$$t_l = \sum_{s\in\mathcal{M}^+}\sum_m r_s E_s(\mathbf{G}_{lm}) + \sum_m E_0(\mathbf{G}_{lm}),$$

$$p_{ly}t_l = \sum_{s\in\mathcal{M}^+}\sum_m r_s[k_y(s) + lB_{1y} + mB_2]E_s(\mathbf{G}_{lm}) + \sum_m (k_{0y} + lB_{1y} + mB_2)E_0(\mathbf{G}_{lm}).$$

Furthermore, this theory allows us to explore device geometries other than slabs,[7,10] provided we define local y axes normal to each surface along a light path [as y' in Fig. 1(a)] and solve the refraction problems surface by surface. In this way, one essentially solves the field in only a few surface cells to obtain the field in the whole space. Figure 3 demonstrates the power of this theory by an extreme case: a vertical exit surface.

III. REFRACTION OF A REALISTIC BEAM: GAUSSIAN OR OTHER PROFILES

To fully understand Fig. 3, one needs to analyze the refraction of a Gaussian beam. The standard technique first decomposes the Gaussian beam into planar-wave components, finds the refracted wave for each planar wave, and then adds them together to find the total refracted beam.[32] Here we use the coordinate systems naturally defined by the directions of incident and refracted beams. Each beam will be described by its own coordinate system, which has one axis along the beam and the other axis perpendicular to the beam. Consider an incident beam with a distribution in q space,

$$a(\mathbf{q}) = a(\mathbf{q}_0 + \delta\mathbf{q}) = \frac{\sqrt{2\pi}}{\Delta q_\perp}\exp\left(-\frac{\delta q_\perp^2}{2\Delta q_\perp^2}\right)2\pi\delta(\delta q_\parallel), \quad (7)$$

where $\delta\mathbf{q}$ is decomposed into components $\delta q_\parallel$ and $\delta q_\perp$, parallel to $\mathbf{q}_0$ and perpendicular to $\mathbf{q}_0$, respectively. Note that $\Delta q_\perp$ is a constant indicating lateral spread of $\mathbf{q}$. The incident field is

$$E_{in}(\mathbf{x}) = (2\pi)^{-2}\int a(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{x}}d^2q. \quad (8)$$

Using the identity

$$(\sqrt{2\pi}a)^{-1}\int_{-\infty}^{\infty} e^{-x^2/2a^2+ibx}dx = e^{-a^2b^2/2}, \quad (9)$$

the incident field is evaluated:

$$E_{in}(\mathbf{x}) = e^{i\mathbf{q}_0\cdot\mathbf{x}}e^{-\Delta q_\perp^2 x_\perp^2/2}. \quad (10)$$

The field in PC is $E_{PC}(\mathbf{x})=\Sigma_s\tilde{E}_s(\mathbf{x})$, where

$$\tilde{E}_s(\mathbf{x}) = (2\pi)^{-2}\int a(\mathbf{q})t_s E_s(\mathbf{x})d^2q. \quad (11)$$

Note that t_s and $E_s(\mathbf{x})$ both depend on the incident wave vector $\mathbf{q}$. If the beam is sufficiently wide, one can use Eq. (9) to evaluate the integral and find

$$\tilde{E}_s(\mathbf{x}) \approx t_s^{(0)} E_s^{(0)}(\mathbf{x}) e^{-(1/2)\Delta k_\perp^2 x_\perp^2}, \tag{12}$$

where the term $t_s^{(0)}E_s^{(0)}$ is evaluated at $\mathbf{q}_0$, $\Delta k_\perp \equiv (\partial k_\perp/\partial q_\perp)_\omega \Delta q_\perp$ gives the lateral spread of $\mathbf{k}_s$, and $x_\perp$ is the component of $\mathbf{x}$ along $k_\perp$. For simplicity, the mode index s for $k_\perp$ is omitted. Each beam inside a photonic crystal retains a Gaussian form as generally expected.[6] An expression equivalent to Eq. (12) was obtained earlier.[32]

The derivative $(\partial k_\perp/\partial q_\perp)_\omega$ can be handily calculated from

$$\left(\frac{\partial k_\perp}{\partial q_\perp}\right)_\omega = \left(\frac{\partial k_\perp}{\partial k_x}\right)_\omega \left(\frac{\partial k_x}{\partial q_x}\right)_\omega \left(\frac{\partial q_x}{\partial q_\perp}\right)_\omega. \tag{13}$$

Note that $k_x \equiv q_x$. It is straightforward to find

$$\left(\frac{\partial k_\perp}{\partial q_\perp}\right)_\omega = \frac{\cos\alpha}{\cos\theta}. \tag{14}$$

Define the beamwidths

$$w_I = \frac{\sqrt{8}}{\Delta q_\perp}, \quad w_{PC} = \frac{\sqrt{8}}{\Delta k_\perp}. \tag{15}$$

Then it follows that

$$\frac{w_I}{\cos\alpha} = \frac{w_{PC}}{\cos\theta} = w_0. \tag{16}$$

This means that the cross sections of the two beams along the boundary have the same width as shown in Fig. 3. Note that the reflection smears the true w_I in Fig. 3.

The peak of $\tilde{E}_s(\mathbf{x})$ remains at the value of the planar-incident-wave case if the peak of the incident Gaussian beam is unity. With the cell-averaged Poynting vector, the power for a Gaussian beam in the PC can be calculated:

$$P = \int S dx_\perp \approx (\sqrt{\pi/8})\bar{S} w_{PC} = \sqrt{\pi/8} w_0 (\bar{S}\cos\theta), \tag{17}$$

where $\bar{S}$ is the peak value of the Poynting vector. One thus finds that the conservation of $S_y = S\cos\theta$ relates to the conservation of total power P as

$$\sum_l R_l + \sum_s T_s = \sum_l \frac{|S_{ly}|}{S_{0y}} + \sum_s \frac{S_{sy}}{S_{0y}} = 1, \tag{18}$$

where R_l and T_s are the normalized powers of a reflected beam and a PC mode $\tilde{E}_s$, respectively; $\mathbf{S}_l$ and $\mathbf{S}_s$ are the corresponding Poynting vectors, and $\mathbf{S}_0$ is the incident Poynting vector. The summations should be limited to the propagating modes. The insertion loss is given by $10\log_{10}(T_s)$ for a PC mode. A beam coming from inside a PC can be treated similarly. Figure 3 shows that the beam peak and width given by the planar-wave theory is in good agreement with the FDTD simulations. Note that a Gaussian function is fully specified by the peak and width. Actually, our theory best fits the FDTD data shown if $r/a = 0.32$ is used in our theory. Therefore, it is obvious that the difference is owing to the fact that the FDTD simulation cannot resolve a hole-radius difference of $0.02a$ with a grid spacing of about the same magnitude.

For a general beam profile, our theory also provides a convenient way of calculating the refracted and reflected beams. Given a photonic crystal with certain surface orientation, one can easily compute and record the quantities

$$r_l, \quad t_s, \quad E_s(\mathbf{G}_{lm}), \tag{19}$$

for, say, 500 incident planar waves whose incident angles are equally spaced.[33] Using these quantities, one can finally integrate to obtain the refracted beam according to Eq. (11) for an arbitrary incident beam. This option offered by our theory is extremely advantageous for real-world design and optimization of photonic crystal devices. For example, to design a photonic-crystal-based demultiplexer, one often needs to vary the incident beamwidth to find the optimized device design in terms of device size, diffraction and beam shape distortion, optical loss, and other issues. And the incident beam may not be Gaussian. With the technique presented above, whenever the ending width of the tapered waveguide or the incident angle changes, one only needs to compute the integral in Eq. (11) with a different set of $a(\mathbf{q})$. In contrast, the FDTD methods would have to perform a time-consuming simulation each time the incident angle or beamwidth changes.

IV. DISCUSSIONS AND APPLICATIONS

A. Physical interpretations

The importance of our analysis of the PC surface refraction goes far beyond merely giving amplitudes $\{t_s, r_l\}$. First, incident light recognizes a lattice by its "face." For a square lattice, if the incident surface has Miller indices other than (10) or (01), the light may nevertheless "see" an oblique lattice, according to Eqs. (5). This "misinterpretation" completely shuffles the intrinsic mode degeneracy of the original lattice.

Second, for an *ideal* periodic surface, the coincidence of the wave vectors of different reflection orders, $\mathbf{p}(\mathbf{G}_{lm}) \equiv \mathbf{p}(\mathbf{G}_{l0})$, means the reflected waves carry only the information of the surface periodicity. The surface BZ (Ref. 34) remains a useful concept. Although the new surface basis-vector $\mathbf{B}_1$ is generally not parallel to the surface, its x component still plays some role because

$$p_x(\mathbf{G}_{l,m}) - p_x(\mathbf{G}_{l-1,m}) = B_{1x} = \frac{2\pi}{A_1}.$$

Note that due to lowered translational symmetry, a semi-infinite 2D photonic crystal has two one-dimensional BZs (associated with $\mathbf{B}_1$ and $\mathbf{B}_2$, respectively), instead of one 2D BZ. It is interesting to notice that the degeneracy of the PC modes is governed by the surface-normal one-dimensional BZ associated with $\mathbf{B}_2$, whereas the wave vector difference of the reflected waves is dictated by the surface BZ through $\mathbf{B}_1$. From another point of view, because the reflected waves carry only the information of the surface periodicity A_1 for a periodic surface, one may say other Bragg planes inside the PC are hidden. Simply from the reflected waves, one could not tell whether the crystal is a 1D grating or a 2D photonic crystal. This confirms that the light may not recognize the

original lattice structure of a semi-infinite lattice, but has a "biased" view, based on what it sees on the surface. Actually, it has been known in all previous studies of (10), (100), or (111) types of surfaces that the reflected modes carry only the information of the surface periodicity in their wave vectors.[16,18–20] However, such a phenomenon has not been correlated with the mode degeneracy of the photonic crystal as we reveal here. Furthermore, it is interesting to see from our analysis that if a surface is quasiperiodic, the reflected waves carry the information of all Bragg planes, not just the surface BZ. This is obvious because no two $\mathbf{p}(\mathbf{G})$ coincide, and for a quasiperiodic surface, there is actually no surface BZ owing to the lack of surface periodicity.

Third, it can be shown that a slight change of surface orientation may split one PC beam into many beams. Consider the rectangular lattice example we discussed in Sec. II. For the (10) surface, suppose there is only one propagating mode among $2L+1$ distinct up modes. One can easily show that when L increases, the new modes introduced will all be evanescent modes with complex $k_y(s)$; the change of M does not affect the number of distinct up modes. On the other hand, it is a drastically different case for a quasiperiodic surface that could be merely 0.0001° from the (10) direction. Owing to the lack of degeneracy, the number of distinct up modes will increase when M increases. Particularly, the number of distinct propagating modes will increase as M, which means more beams will be present in the crystal. How to observe such a sensitive phenomenon is an interesting question. Note that a quasiperiodic section of acceptable quality cannot be achieved in an atomic crystal because an atom cannot be divided or "cut" into fractions. Whereas artificial structures such as photonic crystals can form an ideal "flat" surface, an atomic crystal surface intended to be a quasiperiodic section will in general appear ragged or have lattice voids.

B. WDM demultiplexer

WDM demultiplexers are among the most important applications of photonic crystals.[3] Following the example in Figs. 2(c) and 2(d), we calculate the insertion losses for a demultiplexer assuming the exit surface has an angle of 0°, 30°, or 60° with the entering surface. Generally, we find the *total* transmission will be enhanced if the exit surface is perpendicular to the beam in PC. The 30° case seems to be an exception because there is another exiting beam (not plotted) of comparable strength. The loss spectra in Fig. 2(d) suggest that for $\theta<-34°$, one should use a 60° exit surface. Note that if the 0° exit surface is used for PC beams with θ up to $-84°$, the device lateral size will increase 7 times. The corresponding demultiplexer design is sketched in Fig. 2(c), inset. This design demonstrates lower than 3 dB loss over a 25-nm wavelength span (covering >60 WDM channels if 0.4 nm resolution is possible[6]).

To apply our rigorous refraction theory to a realistic photonic crystal device, such as one shown in Fig. 2(c), always requires certain simplifications—as in any physics modeling. First, we usually need to assume that the device is sufficiently large (>10 lattice constants, for example) such that the evanescent modes originating from one facet (or surface) would be negligible on the other facets. Second, the concept of beam must be valid. These assumptions were commonly employed in many prior studies[5–7,9–11] and were met in many real superprism devices.[3,4] These assumptions can be waived under certain circumstances, possibly accompanied by some adaptation of our theory. For example, the depth requirement can be waived for a plane-parallel slab, a candidate for superlenses. In fact, one can extend the current theory to rigorously calculate the transmission of a planar incident wave through an (arbitrarily thin) PC slab. This can be done by solving the boundary conditions on the front and back surfaces of the slab altogether, the results of which, we found, agree perfectly with existing rigorous theories for slabs. Such a rigorous theory for PC slabs falls out of the scope of this paper and will be presented elsewhere. Besides such thin PC slabs, we are not aware that any other superprism device of wide interest requires us to abandon these two assumptions. Therefore we have not been motivated to study other cases where these assumptions are not appropriate. Nonetheless, such devices could emerge, and our theory may not be applicable to them. Finally, note that when the size of a photonic crystal is, for example, below five lattice constants in any dimension, the nonpropagating modes localized near the surfaces could dominate over the propagating modes in the whole crystal. Under such circumstances, the Bloch modes of a bulk photonic crystal gradually become less helpful in studying such a photonic crystal fragment. Similar phenomena are commonly known as the size effect or quantum confinement effect in other fields. It could be more advantageous to treat such a small structure as a collection of single scatterers rather than a fragment of a periodic lattice, and scattering theories[17,19,20,23] could be the preferred tool. Such small structures are at the brim of photonic crystal research, although they could be interesting in the general area of nanophotonics.

Also, the effect of the secondary beams generated by multiple internal reflections inside the photonic crystal is negligible in this example. Generally, in designing useful devices, one should suppress such secondary beams through choosing proper device structures and optimizing design parameters.[8,16] Detailed discussions of design optimization issues are beyond the scope of this paper.

C. Numerical advantages

The numerical methods for studying beam refraction and propagating in photonic crystals can be roughly divided into three categories.

The first category is what we called the whole-space method, which calculates the field in every cell of the photonic crystals. The computational workload of these methods grows with the volume (or area for 2D PC) of a photonic crystal. This category includes the FDTD method and the multiple-scattering method.[10] These methods are most versatile and can deal with photonic crystals of arbitrary shapes. Unfortunately, in reality, these methods are usually very costly in terms of computational time and data storage such that they are often found incapable of simulating a practical

device of reasonably large size with acceptable accuracy. This situation has been illustrated in the design of a demultiplexer.

The second category includes a variety of supercell methods where the number of cells in a supercell grows with the linear dimension of the crystal under study. For example, the internal-field method[21] computes the field in a supercell extending perpendicularly to the surface of a PC slab. These methods are numerically more efficient than those belonging to the first category, although less versatile in dealing with various geometry; for example, the internal-field method is restricted to slabs.

The last category is what we call the single-cell method—for example, the transfer matrix method developed by Li and Ho.[16] In these methods, refraction and propagation are treated separately. Each time a beam hits a surface (from inside or outside a PC), the refraction algorithm is invoked to compute the field, essentially in one cell on the surface. Inside the body of the PC, the beam amplitudes are given by the Bloch theorem from their values on the surfaces, and they can be calculated handily whenever needed. As these methods solve the field in a few surface cells only, they are capable of accurately computing for a large device with small amounts of computational time and data storage. Despite the efficiency, the prior single-cell methods[15,16,20] were restricted to certain special lattice types and surface orientations, owing to the lack of physical insight of the PC refraction problem (as we discussed in the Introduction). The single-cell method presented here clarifies the physics, overcomes the lattice and surface-orientation limitations, and further improves the efficiency and accuracy of the prior single-cell methods.

A detailed comparison of the numerical implementation of our theory and the layer transfer matrix method (TMM) is beyond the scope of this work. We just point out that in the latter method, a large number L_{TMM} of layers are usually needed for adequate accuracy. Meanwhile, the number of matrix inversion operations needed by a stable TMM algorithm[16,30] grows with L_{TMM}. Besides, the eigenvalues solved in our method are $k_y(s)$ themselves, whereas eigenvalues solved by TMM are actually $e^{ik_y(s)d}$, where d is a certain length. Note that the eigenvalues in the transfer matrix method stretch over a much larger scale.

We should point out that the discussions in Sec. II and Sec. IV A regarding the surface-orientation-dependent degeneracy are rigorous only as $L, M \to \infty$. Whenever the problem is formulated into a numerical form, a Fourier series with finite L and M will be used in Eq. (1). In practice, the calculated $k_y(s)$ for those modes far outside the first BZ would generally have substantial errors. Using larger L and M will improve the accuracy of these modes but new modes are introduced even farther from the first BZ. The eigenvalue $k_y(s)$ of these new modes usually have even larger errors. These numerical errors directly lead to a severe departure from exact degeneracy for the numerically calculated eigenmodes. In general, one may find N_m apparently distinct up modes. The number N_m generally can be any value between $2L+1$ and N, depending on the details of the numerical algorithm. Generally, when L and M increase, the ratio of $N_m/(2L+1)$ will approach, but never reach, unity. Without analytically recognizing the degeneracy, the amplitudes of N_m crystal modes, along with the amplitudes of $2L+1$ reflected modes, can never be solved from $2(2L+1)$ boundary equations no matter how large L and M are.

D. Partial spatial interference in the slab

The theory developed here can also be used to study the problem that a beam of finite width transmits through a relatively thick PC slab. Consider a simple situation where each internal reflection generates only a single reflected beam inside the PC slab. After a round trip of reflections, the beam will be shifted laterally with respect to the original beam in the PC slab. If the reflection angles are relatively large and/or the beamwidths are relatively narrow, the secondary beams generated due to multiple reflections may have little or no spatial overlap with respect to the original beam. Outside the photonic crystal, an observer may see a series of parallel beams exiting each surface rather than a single beam that contains the interference effect. The slab transmission theories for the planar incident waves cannot predict the strength of each exiting beam in such a case. The single-surface refraction theory developed here must be used. In cases where multiple beams are generated upon each reflection, the beams may overlap each other in a complicated manner inside the PC slab. If the beamwidths are finite, again the conventional slab theory would usually not be applicable; the refraction theory presented here shall be used.

V. SUMMARY

A rigorous theory has been developed for light refraction by a photonic crystal with arbitrary lattice type and surface orientation. First, the refraction of a planar incident wave at a photonic crystal surface is rigorously analyzed. To ensure the generality of the theory, we have rigorously proved the equal partition of forward and backward propagating modes. The concept of surface-orientation-dependent mode degeneracy has been proposed and its relationship to quasiperiodic surfaces unfolded. A slight change of surface orientation is predicted to split one beam into an infinite number of beams in certain cases. We have further discussed the case where a Bloch wave illuminates an interface from inside a PC, which enables us to compute the transmissions along a complete light path through a series of interface. In addition, the transmission of a Gaussian beam is analytically discussed and the insertion loss formulas are presented. Other realistic beam profiles are discussed for designing photonic crystal devices. With all these issues solved, a complete theoretical framework of the photonic crystal refraction and transmission has thus been established. The theory has been applied to design a high-channel-count, low-loss WDM demultiplexer.

This theory offers a new picture of the refraction processes and opens the door to a broad range of problems that previously required prohibitive computation resources based on prior methods. In addition, there are many new theoretical problems unveiled by this theory.

The method presented here provides an efficient alternative to study the waveguide endface coupling problem with

the supercell approach.[16] It can be used to study an interface between two PC's of incommensurate lattice constants as well. In addition, this theory can be extended to study the transmission through a PC slab as well by considering the boundary conditions at two surfaces simultaneously. Negative refraction in photonic crystals has been subject to intensive study recently. The prospect of making superlenses with resolutions beyond the Rayleigh limit[35] has stirred significant interest in the research community. Plane-parallel photonic crystal slabs are one of the candidates for superlens applications.[8] The rigorous theory developed here sheds new light on the problem. With this theory, some ideas can now be evaluated on a more rigorous basis.

In a similar fashion, one may develop a rigorous mode theory for 2D photonic crystal waveguides as these waveguides can be regarded as an air slab sandwiched between two semi-infinite photonic crystals. One readily sees that this theory for PC slabs can be used to study grating diffraction as well.

We hope this work will stimulate further study and possibly be extended to areas such as surface physics and x-ray and electron diffraction, where surface-related problems abound in various forms of periodic lattices.

ACKNOWLEDGMENTS

We thank Dr. P. Binding and Dr. L. Li for helpful communications. We are grateful to Dr. Herbert L. Berk, Dr. Joe C. Campbell, Dr. Michael Baker, and Dr. Ananth Dodabalapur for offering valuable comments on this work.

APPENDIX A: GROUP VELOCITY FOR 2D TM WAVES

For the operator matrix $[W]$ and any of its eigenvector $[E_s]$, we define

$$\mathcal{H}(E_s) = [E_s]^\dagger [W][E_s].$$

Note that $\mathcal{H}(E_s) \equiv 0$. Each element of $[W]$ and $[E_s]$ is a function of k_x, k_y, and ω; therefore, $\mathcal{H}$ can also be considered a function of these variables. For real eigenvalues, $[W][E_s] = 0$ yields $[E_s]^\dagger [W] = 0$; thus, the derivative of $\mathcal{H}$ can be simplified,

$$\frac{\partial}{\partial \omega}\mathcal{H}(E_s) = [E_s]^\dagger \frac{\partial [W]}{\partial \omega}[E_s],$$

where a technique similar to the Hellmann-Feynman theorem is applied to avoid the differentiation of $[E_s]$. Similarly,

$$\nabla_{\mathbf{k}}\mathcal{H}(E_s) = [E_s]^\dagger (\nabla_{\mathbf{k}}[W])[E_s].$$

Now consider $\mathcal{H}(k_x, k_y, \omega) \equiv 0$, where $\omega = \omega(k_x, k_y)$ describes the dispersion contours in reciprocal space. The total derivative of $\mathcal{H}$ with respect to k_y gives

$$\frac{\partial \mathcal{H}}{\partial k_y} + \frac{\partial \mathcal{H}}{\partial \omega}\frac{\partial \omega}{\partial k_y} = 0, \tag{A1}$$

from which the y component of group velocity can be calculated from the derivatives of $\mathcal{H}$. A similar procedure yields the x component. Thus one finds

$$\nabla_{\mathbf{k}}\omega = \frac{\epsilon_0}{2\omega \mathcal{E}_{em}(s)} \sum_{\mathbf{G}} (\mathbf{k}_s + \mathbf{G})|E_s(\mathbf{G})|^2, \tag{A2}$$

where $\mathbf{k}_s = k_x \mathbf{e}_x + k_y(s)\mathbf{e}_y$ and

$$\mathcal{E}_{em}(s) = \frac{\epsilon_0}{2\omega^2} \sum_{\mathbf{G}} (\mathbf{k}_s + \mathbf{G})^2 |E_s(\mathbf{G})|^2$$

is non-negative for real $\mathbf{k}$ modes.

For real $\mathbf{k}_s$ modes, the Poynting vector averaged per cell per period of time is readily computed:

$$\langle \mathbf{S} \rangle_{cell,T} = (1/2\mu_0\omega) \sum_{\mathbf{G}} (\mathbf{k}_s + \mathbf{G})|E_s(\mathbf{G})|^2, \tag{A3}$$

where μ_0 is the vacuum permeability. Hereafter, we omit the average brackets of $\mathbf{S}$. Our calculation of $\nabla_{\mathbf{k}}\omega$ differs from the calculation by the Hellmann-Feynman theorem[36] as the operator we have used depends on the frequency, which allows us to differentiate the operator with respect to ω. The expression we obtain, Eq. (A2), appears to give a more obvious correlation between the group and Poynting vector. Actually, one readily finds

$$\mathbf{S} = \mathcal{E}_{em} \nabla_{\mathbf{k}}\omega. \tag{A4}$$

Note that Botten *et al.* derived Eq. (A4) with Green's functions.[23] Even earlier, Yeh derived Eq. (A4) for 1D periodic structures as well.[37] The same relation, Eq. (A4), holds for TE waves, although

$$\mathbf{S} = (4\omega\epsilon_0)^{-1} \sum H_s^*(\mathbf{G}) \left(\frac{1}{\epsilon}\right)_{\mathbf{G}\mathbf{G}'} [2\mathbf{k}_s + \mathbf{G} + \mathbf{G}'] H_s(\mathbf{G}').$$

APPENDIX B: GENERAL PROOF OF THE EQUAL PARTITION OF UP AND DOWN MODES

In Sec. II, we have presented an intuitive proof of the equal partition of up and down modes for 2D photonic crystals. In this section, we shall progressively develop a more rigorous and complete proof that can be applied to 3D photonic crystals and open contours as well.

Consider an arbitrary 2D dispersion contour shown in the left graph of Fig. 1(b). With the Jordan curve theorem,[31] one readily shows that any constant-k_x line intersects a closed contour even number of times. Instead of considering the normal vector, we now consider the tangent vector along the contour. Any closed contour can be given a direction, along which a 1D creature can traverse the contour and return to its starting point. As the creature travels, it counts the points C_i, $i = 1, \ldots, 2N_c$, where it crosses the constant-k_x line. It will pair these points as

$$(C_{2i-1}, C_{2i}), \quad i = 1, \ldots, N_c. \tag{B1}$$

Consider an arbitrary pair of intersecting points, which correspond to two consecutive times that the creature crosses the constant-k_x line. The constant-k_x line divides the reciprocal space into two disconnected part. Whenever the creature crosses the line, it switches from one part to the other. Therefore, if the creature crosses the line from right to the left at

C_{2i-1}, then it must return to the right from the left at C_{2i}, and vice versa. Accordingly, the x component of the tangent vector must have opposite signs at C_{2i-1} and C_{2i}. For example, the tangent vector at C_{2i-1} has an angle between $\pi/2$ and $3\pi/2$ with respect to the $+x$ axis, whereas it has an angle between $-\pi/2$ and $\pi/2$ at C_{2i}. Assume that the normal vector has an angle $-\pi/2$ with respect to the tangent vector (if the angle is $+\pi/2$, the final conclusion will be same). Thus, the normal vector at C_{2i-1} has a polar angle between 0 and π, whereas the normal vector has a polar angle between $-\pi$ and 0 at C_{2i}. This means that the y component of the normal vector has opposite signs at the pair of points. For a 3D photonic crystal, its dispersion surface is 2D surface in 3D reciprocal space. The constant-k_x line will become the constant-(k_x,k_z) line, assuming the coordinate system is chosen such that the y axis remains normal to the crystal surface. It will be difficult to work directly with the surface normal vector in 3D because it may not be in one plane when the creature travels along a section of the dispersion surface. Fortunately, the preceding proof presented in this part primarily relies on the tangent vector. Consider the plane passing the constant-(k_x,k_z) line and incident wave vector. One readily shows that the intersection of this plane and the dispersion surface consists of closed contours provided that the dispersion surface consists of components topologically equivalent to a sphere. Note that the tangent vector of each contour is also a tangent vector of the dispersion surface. Also note that the constant-(k_x,k_z) line is parallel the y axis so that the projection of the normal vector on this plane maintains its y component n_y. Then one can apply the preceding proof with each contour obtained by intersection and show that the y component of the surface normal has opposite sign for each pair of crossings.

As witnessed in some periodic systems that can be regarded as 1D photonic crystals embedded in a 2D or 3D space, the equifrequency surface may not be closed.[38–40] Generally, such a situation is fairly complicated to analyze directly. The preceding discussions on the closed surfaces provide certain insight for investigating this situation although the adiabatic approach shall be followed if one intends to utilize the preceding discussions.

For a given photonic crystal, we parametrize its dielectric function $\epsilon(\mathbf{x})$ in the following manner:

$$\epsilon(\eta,\mathbf{x}) = \bar{\epsilon} + \eta\epsilon_p(\mathbf{x}), \tag{B2}$$

where $\bar{\epsilon}$ is certain spatial average of $\epsilon(\mathbf{x})$ and $\epsilon_p(\mathbf{x}) = \epsilon(\mathbf{x}) - \bar{\epsilon}$ represents the periodic perturbation in the original crystal. Obviously,

$$\epsilon(1,\mathbf{x}) = \epsilon(\mathbf{x}). \tag{B3}$$

Consider the dispersion surface for a photonic crystal represented by a dielectric function $\epsilon(0,\mathbf{x}) = \bar{\epsilon}$. Obviously, this is a homogeneous medium. The dispersion surface should be a circle (2D) or a sphere (3D). Actually, if one solves the eigenvalue problem, Eq. (1), for the TM polarization in 2D, one finds that the dispersion surface consists of periodic replicas of circle centered at reciprocal lattice vertices. A similar replication occurs for a 3D photonic crystal. Applying the preceding discussions to each circle or sphere individually, one readily shows the equal partition in this case because each circle or sphere is closed. Note that these circles or spheres may intersect each other in reciprocal space, often at the BZ boundaries. The dispersion surface around each intersection point will undergo a topological mutation upon an infinitesimal periodic perturbation. Such a mutation of topology may lead to the presence of the open contours or surfaces. One approach is to investigate the general characteristics of the topological mutation and prove that the correlated mutations generate open contours pairwise and maintain the equal partition. By expanding the polynomial $F(k_x,k_y,\omega)$ in the neighborhood of a mutation point, one can show that such a mutation usually leaves a pair of open surfaces (which may be far separated in space), whose corresponding normal vectors have the relevant component in opposite signs. This method works fairly well for those 2D problems where the situation is similar to gratings. It becomes considerably more complicated in other cases, particularly for 3D problems. Here we present another proof.

Assume an infinitesimal perturbation $\eta\epsilon_p(\mathbf{x})$ is imposed on the averaged uniform medium. Because $\nabla_k\omega = v_g\mathbf{n}$, we can use $\partial\omega/\partial k_y$ in place of n_y for equivalent results in the following discussions. Consider a constant-k_x line—or a constant-(k_x,k_z) line in 3D—that does not pass any intersection point of the unperturbed circles (or spheres in 3D). If the perturbation is small enough and the constant k_x line is sufficiently far from all of the intersection points, the dispersion surface should not be mutated where this constant-k_x line crosses the dispersion surface. And the equal partition will be maintained for this value of k_x based on the preceding discussions on the averaged homogeneous medium. Now let the constant-k_x line sweep. Owing to the continuity of the dispersion surface, the values of $\partial\omega/\partial k_y$ at the points where the constant-k_x line crosses the dispersion surface cannot abruptly switch its sign as k_x varies continuously, and the equal partition is maintained. The only possibility that $\partial\omega/\partial k_y$ may change its sign is when $\partial\omega/\partial k_y = 0$. This corresponds to the cases where the constant-k_x line is tangent to the dispersion contour. As discussed in Sec. II, when the constant-k_x line sweeps across such a tangent point, either a pair of real-valued up and down modes will disappear (or appear) or a single real-valued eigenmode will maintain its sign of $\partial\omega/\partial k_y$. In either case, the equal partition shall not be affected. Thus, for any constant-k_x line, the equal partition of the up and down modes is maintained for a photonic crystal with an infinitesimal periodic perturbation regardless of the presence of open contours. As the strength of the periodic perturbation continuously increases (so that η varies from 0 to 1), if there is any further topological mutation, one can apply the above argument and show that the partition of up and down modes remains equal. In conclusion, the equal partition holds for any value of η, including $\eta=1$, which corresponds to the photonic crystal structure we intend to solve.

We note that without this adiabatic approach, a direct analytic proof of the equal partition is difficult. For example, the fact that the up and down modes appear in pairs immediately

reminds us of Vieta's formulas, yet it is surprisingly hard to construct certain convenient forms of the derivative pairs of the roots $k_y(s)$. One factor that one must keep in mind is that these forms must be sufficiently simple so that they can be easily evaluated from the coefficients for the polynomial $F(k_x, k_y, \omega)$, or ultimately from the elements of $[W]$. This often complicates the situation severely, especially when the problem is extended to 3D photonic crystals.

[1] E. Yablonovitch, Phys. Rev. Lett. **58**, 2059 (1987).
[2] S. John, Phys. Rev. Lett. **58**, 2486 (1987).
[3] H. Kosaka, T. Kawashima, A. Tomita, M. Notomi, T. Tamamura, T. Sato, and S. Kawakami, Phys. Rev. B **58**, R10096 (1998).
[4] L. Wu, M. Mazilu, T. Karle, and T. F. Krauss, IEEE J. Quantum Electron. **38**, 915 (2002).
[5] M. Notomi, Phys. Rev. B **62**, 10696 (2000).
[6] T. Baba and T. Matsumoto, Appl. Phys. Lett. **81**, 2325 (2002).
[7] T. Baba and M. Nakamura, IEEE J. Quantum Electron. **38**, 909 (2002).
[8] C. Luo, S. G. Johnson, J. D. Joannopoulos, and J. B. Pendry, Phys. Rev. B **65**, 201104(R) (2002).
[9] K. B. Chung and S. W. Hong, Appl. Phys. Lett. **81**, 1549 (2002).
[10] J. Bravo-Abad, T. Ochiai, and J. Sanchez-Dehesa, Phys. Rev. B **67**, 115116 (2003) and references therein.
[11] T. Prasad, V. Colvin, and D. Mittleman, Phys. Rev. B **67**, 165103 (2003).
[12] Notes added in the revision: internal reflection on photonic crystal surfaces that are not aligned with special crystal directions was recently reported by another research group. See X. Yu and S. Fan, Phys. Rev. E **70**, 055601(R) (2004).
[13] The long-wavelength limit for refraction is also discussed. See P. Halevi, A. A. Krokhin, and J. Arriaga, Phys. Rev. Lett. **82**, 719 (1999).
[14] J. B. Pendry and A. MacKinnon, Phys. Rev. Lett. **69**, 2772 (1992).
[15] T. Minami, H. Ajiki, and K. Cho, Physica E (Amsterdam) **13**, 432 (2002).
[16] Z. Y. Li and K. M. Ho, Phys. Rev. B **68**, 155101 (2003).
[17] K. Ohtaka, T. Ueta, and K. Amemiya, Phys. Rev. B **57**, 2550 (1998); K. Ohtaka, J. Phys. C **13**, 667 (1980).
[18] N. Stefanou, V. Karathanos, and A. Modinos, J. Phys.: Condens. Matter **4**, 7389 (1992).
[19] N. Stefanou, V. Yannopapas, and A. Modinos, Comput. Phys. Commun. **132**, 189 (2000).
[20] T. Ochiai and J. Sanchez-Dehesa, Phys. Rev. B **64**, 245113 (2001).
[21] K. Sakoda, Phys. Rev. B **52**, 7982 (1995).
[22] K. Sakoda, Phys. Rev. B **52**, 8992 (1995).
[23] L. C. Botten, N. A. Nicorovici, R. C. McPhedran, C. Martijn de Sterke, and A. A. Asatryan, Phys. Rev. E **64**, 046603 (2001).
[24] Here we are discussing refraction (or the single- surface transmission), not the slab transmission problem. In the slab transmission problem, for a planar incident wave, there is no need to distinguish the forward- and backward-propagating modes.
[25] This work was first presented by W. Jiang *et al.*, in *Photonics West, 2004*; a summary of the work is included in Proc. SPIE **5360**, 190 (2004).
[26] A. L. Mackay, Physica A **114**, 609 (1982).
[27] Y. S. Chan, C. T. Chan, and Z. Y. Liu, Phys. Rev. Lett. **80**, 956 (1998) and references therein.
[28] J. E. Dennis, J. F. Traub, and R. P. Weber, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **13**, 831 (1976), and references therein.
[29] P. Lancaster, *Lambda-matrices and Vibrating Systems* (Pergamon, Oxford, 1966).
[30] L. Li, J. Opt. Soc. Am. A **13**, 1024 (1996). The introduction of this reference contains a concise review of the history of the layered grating approach.
[31] L. C. Kinsey, *Topology of Surfaces* (Springer-Verlag, New York, 1993).
[32] D. Maystre, J. Mod. Opt. **50**, 1431 (2003).
[33] Note that these quantities for each incident angle only occupy a negligible amount of computer memory or data storage because N is on the order of hundreds. Furthermore, in practical cases, only a few dominant modes, mostly propagating modes, need to be recorded.
[34] J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals* (Princeton University, Princeton, 1995); R. D. Meade, K. D. Brommer, A. M. Rappe, and J. D. Joannopoulos, Phys. Rev. B **44**, 10961 (1991).
[35] J. B. Pendry, Phys. Rev. Lett. **85**, 3966 (2000).
[36] K. Sakoda, *Optical Properties of Photonic Crystals* (Springer, Berlin, 2001).
[37] P. Yeh, J. Opt. Soc. Am. **69**, 742 (1979).
[38] P. St. J. Russell, Electron. Lett. **20**, 72 (1984).
[39] R. Zengerle, J. Mod. Opt. **34**, 1589 (1987).
[40] P. Yeh, *Optical Waves in Layered Media* (Wiley, New York, 1988).