A Brief Survey of Computational Photonics

Steven G. Johnson, J. N. Winn, R. D. Meade, and J. D. Joannopoulos

October 23, 2006

Introduction to Book Excerpt

The following text is excerpted from the upcoming second edition of the book *Photonic Crystals: Molding the Flow of Light* (Princeton University Press), scheduled for publication in 2007. In particular, although the book is mainly focused on the general principles of photonic crystals and related devices, it includes a survey of computational methods in the form of an appendix, which we reproduce below.

Computational Photonics

In chapters **??** and **??**, we presented the equations of classical electromagnetism, and derived some of the general properties of the solutions based on symmetry and linear algebra. In the chapters that followed, we relied on a wide array of numerical solutions of the equations for particular crystals, waveguides, cavities, and other structures, to illustrate the concepts. Where did these solutions come from? Band diagrams, transmission spectra, field patterns, and other results do not just spring forth from the equations. Considerable effort has gone into the numerical solution of the Maxwell equations. Although our focus in this book has been on general principles rather than numerical techniques, this appendix will serve to draw back the curtain and introduce the reader to the world of computational photonics.

Even as recently as twenty years ago, it would have been unusual to present solutions to the Maxwell equations without first describing the computational method in great detail. Since then, photonics research has undergone the same profound change that has swept through all areas of science and engineering in the last half century, catalyzed by the availability of ever more powerful computers. The solution of a system of partial differential equations in a mere three or four dimensions is now an unremarkable feat. Almost none of the computations in this book required more than a few hours on a personal computer. Most required only a few minutes.

The situation in photonics is especially favorable for computation because the Maxwell equations are practically exact, the relevant material properties

are well known, and the length scales are not too small. Therefore, an exciting aspect of this field is that quantitative theoretical predictions can be made *ab initio* (from first principles), without any questionable assumptions or simplifications. The results of such computations have consistently agreed with experiments. This makes it possible and preferable to optimize the design of photonic crystals on a computer, and then manufacture them. The computer becomes the pre-laboratory.

Many standard numerical techniques for the solution of partial differential equations have been applied to electromagnetism, and each has its own particular strengths and weaknesses. High-quality "black-box" software is widely available, including free, open-source programs (some of which are described at the end of this appendix). Indeed, computational photonics has matured so much that many practitioners have stopped worrying about the finer details of the numerics, and are familiar only with the general principles and capabilities of the different tools. Here, we will summarize some of the most important methods, and work through a specific example of a band-structure calculation in more detail.

Generalities

Broadly speaking, there are three categories of problems in computational photonics:

- **Frequency-domain eigensolvers**: find the band structure $\omega(\mathbf{k})$ and the associated fields, by expressing the problem as a finite matrix eigenproblem $Ax = \omega^2 Bx$ and applying linear-algebra techniques to find a few of the eigenvectors x and eigenvalues ω^2.[1]

- **Frequency-domain solvers**: given a current distribution $\mathbf{J}(\mathbf{x})e^{-i\omega t}$ at a fixed frequency ω, find the resulting fields by expressing the problem as a finite matrix equation $Ax = b$ and applying linear-algebra techniques to solve for x.

- **Time-domain simulations**: simulate the fields $\mathbf{E}(\mathbf{x}, t)$ and $\mathbf{H}(\mathbf{x}, t)$ propagating in time, usually starting with some time-dependent current source $\mathbf{J}(\mathbf{x}, t)$.

Although most of this book has focused on the band structures and eigenfields, the other two problems are also important. For example, one often desires the transmission or reflection spectrum from a *finite* structure. This cannot be easily obtained from the band diagram, but at the same time, knowledge of the gaps and eigenmodes is crucial for the meaningful interpretation of such spectra (as we saw in chapter **??**). There is, of course, some overlap between the problems;

[1] Instead of looking for eigenvalues ω^2 at a fixed $\mathbf{k}$, it is possible to formulate the eigenproblem at a fixed ω for the wave vector k at along a single periodic (or uniform) direction as a generalized Hermitian eigenproblem with eigenvalue k (Johnson et al., 2001; Johnson et al., 2002).

for example, time-domain simulations can be used to compute band structures, as we will see later in this appendix. When implemented properly, all of the problems listed above require a computational effort that scales roughly linearly with the size of the system (as opposed to quadratically or worse). This makes computational photonics tractable with relatively modest resources.

Another way to categorize numerical methods for partial differential equations is by the method that is used to reduce the infinite number of unknowns (e.g., the fields at every point in space) to a finite number (N) of **discretized** unknowns. Four important classes of discretization schemes are:

- **Finite differences**: represent unknown functions $f(x)$ by their values $f_n \approx f(n\Delta x)$ at discrete points on a grid, and derivatives by differences on the grid. The most straightforward case is a uniform Cartesian grid, e.g. $df/dx \approx (f_{n+1} - f_{n-1})/2\Delta x$.

- **Finite elements**: divide space into a set of finite geometric elements (e.g., irregular triangles or tetrahedra), and represent unknown functions by simple approximations defined on each element (typically, low-degree polynomials). In a sense, this method is a generalization of finite differences.

- **Spectral methods**: represent unknown functions as a series expansion in a complete basis set of smooth functions, truncating the series to have a finite number of terms. Archetypically, a Fourier series is used; this is also called a **planewave method** in two or three dimensions (where the terms in the Fourier series are plane waves). More generally, when the boundary conditions are not periodic, it can be advantageous to employ other basis functions such as Chebyshev polynomials. One can also use **spectral elements**, which are similar to finite elements but use a more complicated spectral basis for each element.

- **Boundary-element methods**: instead of discretizing all space, discretize only the *boundaries* between homogeneous regions. The homogeneous regions are treated analytically. The discretization can use a finite-element or a spectral basis. A **multipole method** (Yasumoto, 2005) is essentially a boundary-element method with a specialized spectral basis for cylindrical or spherical boundaries (not to be confused with the **fast multipole method**, an algorithm to quickly evaluate matrix-vector products for boundary-element methods). A related idea can be found in **transfer-matrix** or **coupled-wave** methods, which propagate light in a given direction by breaking space into a sequence of uniform regions, and deriving a scattering matrix that relates the values of the fields at each transition.

Of these, the simplest methods to implement and analyze are those that operate on a uniform grid: finite difference methods and the spectral method with a planewave basis. Finite and spectral elements offer the ability to use different spatial resolutions in different regions via an unstructured grid. This can be a

huge advantage for problems with complex geometries and a mixture of very different length scales, such as metallo-dielectric systems (in which micrometer wavelengths can have nanometer skin depths). On the other hand, this flexibility comes at a price in complexity. Simpler methods may well be more efficient in dielectric systems where the index contrast (and thus the length-scale contrast) is not too great. Some authors tout spectral methods for their extraordinary accuracy: in principle, the error decreases *exponentially* with the number of spectral basis functions. However, that only occurs if all of the discontinuities are accounted for analytically in the basis (or elements), and this is rare for dielectric structures because discontinuities occur at every interface. Boundary-element methods are in a class of their own, and have powerful advantages when one has small surfaces in a large volume. For example, to compute the scattered fields from an object, the infinite amount of empty space surrounding the object is treated analytically, and need not be discretized or truncated in any way. Open problems remain, however, in formulating boundary-element methods in some cases with surfaces that extend to infinity (for example, to treat a waveguide bend), and their advantages over finite-element methods in photonic crystals with many surfaces remain debatable.

Once we choose a set of basis functions to represent the discretized unknowns, we must transform the partial differential (or integral) equation into a set of algebraic equations. Aside from finite differences, the most common way to form these algebraic equations is known by several names: **weighed residual** methods, **Petrov-Galerkin** methods, or the **method of moments**, with **Galerkin** and **collocation** methods as special cases. Although we will not be concerned with the details here, the general idea is simple and worth reviewing (see, e.g., Boyd, 2001). Suppose we are solving a linear equation $\hat{L}f(\mathbf{x}) = g(\mathbf{x})$, with a differential (or integral) operator $\hat{L}$, for an unknown function $f(\mathbf{x})$ that we have expanded in a basis of N functions $b_n(\mathbf{x})$ (e.g. finite elements or a spectral basis). That is, we write $f(\mathbf{x}) = \sum_n c_n b_n(\mathbf{x})$ for some unknown expansion coefficients c_n. To get a set of linear algebraic equations for the c_n, we simply take the inner-product of both sides of the equation with some N **weight functions** $w_m(\mathbf{x})$. This leads to: $\sum_n (w_m, \hat{L}b_n)c_n = (w_m, g)$, which is in the form of an $N \times N$ matrix equation $Ax = b$ that can be solved for the c_n. The choice $w_m = b_m$ gives a *Galerkin method*, which corresponds to solving $\hat{L}f = g$ up to an error $\hat{L}f - g$ (the **residual**) that is required to be orthogonal to the basis functions b_n. The choice $w_m(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_m)$, for a set of N points $\mathbf{x}_m$, gives a *collocation method*, which corresponds to requiring that our equation $\hat{L}f = g$ be satisfied exactly at the N *collocation points* $\mathbf{x}_m$.

We should also alert the reader to a class of computational methods that should be *avoided* except in special circumstances: any methods that are restricted to low index contrasts, which are not appropriate for general photonic-crystal structures. Such methods include scalar and semi-vectorial approximations, as well as techniques such as the beam-propagation method (BPM) that are restricted to structures which are slowly varying in at least one direction.

Frequency-domain Eigensolvers

A frequency-domain eigensolver solves the Maxwell eigenproblem for the frequencies of a periodic system (or non-periodic, as described below), as given by equation (**??**) of chapter **??**:

$$\left[(i\mathbf{k}+\nabla)\times\frac{1}{\varepsilon(\mathbf{r})}(i\mathbf{k}+\nabla)\times\right]\mathbf{u}_\mathbf{k}(\mathbf{r})=\hat{\Theta}_\mathbf{k}\mathbf{u}_\mathbf{k}(\mathbf{r})=\frac{\omega(\mathbf{k})^2}{c^2}\mathbf{u}_\mathbf{k}(\mathbf{r}), \qquad (1)$$

where $\mathbf{u}_\mathbf{k}(\mathbf{r})$ is the periodic Bloch envelope of the magnetic field $\mathbf{H}_\mathbf{k}=e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{u}_\mathbf{k}(\mathbf{r})$. Since $\mathbf{u}_\mathbf{k}(\mathbf{r})$ is periodic, the computation only needs to consider the finite unit cell of the structure. In addition to the eigenequation, $\mathbf{u}_\mathbf{k}(\mathbf{r})$ must satisfy the transversality constraint:

$$(i\mathbf{k}+\nabla)\cdot\mathbf{u}_\mathbf{k}=0 \qquad (2)$$

The solution to equation (1) as a function of $\mathbf{k}$ yields the band structure of the system, a result whose utility we have exploited for most of this book.

On a computer, this eigenequation must be discretized into N degrees of freedom using one of the methods mentioned above, such as the planewave expansion (as will be described in more detail later). In general, such a discretization yields a finite **generalized eigenproblem** $Ax=\omega^2Bx$, where A and B are $N\times N$ matrices and x is the eigenvector. Since the original eigenproblem is Hermitian, the discretization can be chosen so that A and B are Hermitian and B is positive-definite,[2] which are important properties for the numerical methods below. One difficulty is the transversality constraint, which we must somehow impose in addition to the eigenequation. Technically, solutions that violate the transversality constraint form "spurious modes" existing at $\omega=0$.[3] The simplest way to impose transversality is to choose a basis that is automatically transverse, such as the planewave basis described below.

Given such a finite eigenproblem, there are two ways to proceed. One is to use a standard linear-algebra package such as LAPACK (Anderson et al., 1999) to find the eigenvectors x and the eigenvalues ω^2. Unfortunately, this requires computation memory proportional to N^2 and time proportional to N^3. Since N may be in the millions for large three-dimensional systems, this method is problematic. Instead, we can exploit the fact that only a *few* of the eigenvalues are really needed. For example, to compute the band diagrams in this book, we only needed the smallest few eigenvalues $\omega_n(\mathbf{k})$ at each $\mathbf{k}$.

This realization leads to **iterative methods**, which compute a small number p of the eigenvalues and eigenvectors, such as the p smallest eigenvalues. There are many such methods (Bai et al., 2000), but they share a few critical features. First, they work by taking a starting guess for x (e.g., random numbers) and applying some process to iteratively improve the guess, converging

[2] For example, these properties are always preserved by a *Galerkin* discretization, which uses matrix elements $A_{mn}=(\mathbf{b}_m,\hat{\Theta}_\mathbf{k}\mathbf{b}_n)$ and $B_{mn}=(\mathbf{b}_m,\mathbf{b}_n)$, given basis functions $\mathbf{b}_n(\mathbf{x})$.

[3] This fact can be derived by taking the divergence of both sides of the eigenequation. Since the divergence of the curl is zero, one is left with an expression $\omega^2(i\mathbf{k}+\nabla)\cdot\mathbf{u}_\mathbf{k}=0$. That is, the eigenequation itself implies that transversality is satisfied if $\omega\neq0$.

quickly to the true eigenvector. In this way, any desired accuracy can be obtained in a small number of steps. Second, they only require you to supply a fast way to compute the matrix-vector products Ax and Bx. In finite-element methods, these matrices are **sparse** (mostly zeros) and Ax or Bx can be computed in $O(N)$ operations, while for spectral methods other fast algorithms are available as described below. Because of this, A and B need never be stored explicitly, and only $O(Np)$ storage is required (for the eigenvectors). Third, given an $O(N)$ matrix-vector product, the computation time grows as $O(Np^2)$ multiplied by the number of iterations; for $p \ll N$, this process is usually much quicker than the $O(N^3)$ explicit solution.

A simple example of such an iterative method can be constructed from the variational theorem that we proved in the section "**??**" of chapter **??**. We proved the variational theorem for the Maxwell eigenproblem, but the same is true of *any* Hermitian eigenproblem, and in particular for our finite eigenproblem $Ax = \omega^2 Bx$. That is, the smallest eigenvalue ω_0^2 satisfies:

$$\omega_0^2 = \min_x \frac{x^\dagger Ax}{x^\dagger Bx}, \tag{3}$$

where $x^\dagger$ denotes the conjugate-transpose (adjoint) of the column vector x. This is known as *Rayleigh-quotient minimization*, with x_0 at the minimum being the eigenvector. We can perform this minimization using any one of the myriad numerical techniques for optimizing a function of several variables, such as a preconditioned nonlinear conjugate-gradient method (Bai et al., 2000). With the planewave method described below, convergence in 10–30 steps is typical. Then, to find the next eigenvalue ω_1, we simply minimize the same Rayleigh quotient but constrain x by the orthogonality relation (which follows for any Hermitian eigenproblem as in chapter **??**): $x^\dagger Bx_0 = 0$. This process continues for ω_2, ω_3, and so on.

This suffices for periodic structures, but what about non-periodic structures, such as line or point defects in photonic crystals? The simplest case to handle is that of localized modes, such as the waveguide and cavity modes trapped around line and point defects, respectively. In this case, we can use a **supercell approximation**: periodic boundary conditions, but with a large computational cell surrounding the localized mode so that the boundaries are irrelevant. That is, we imagine a cavity or waveguide structure that is periodically repeated at large intervals in space. Because the modes of interest are exponentially localized to the defect (or defects), the solution converges exponentially fast to the desired isolated-defect solution as the size of the computational cell increases. (In a photonic crystal with a large gap, the boundaries typically become irrelevant after only a few extra periods of bulk crystal are included around the defect.) The **k** vector along a supercell direction determines the phase relation between these artificially repeated structures, and this phase relation also becomes irrelevant exponentially fast as the cell size increases.

For non-exponentially localized modes in non-periodic structures, such as the "leaky" cavity modes of chapters **??** and **??**, matters are more complicated.

Typically, one imposes some absorbing boundary conditions or regions (such as the perfectly matched layers described below). The problem becomes non-Hermitian (allowing complex ω to account for radiation losses), leading to more complicated numerical methods.

Frequency-domain Solvers

Although band structures and eigenstates are useful, they are not the only quantities of interest in photonic devices. For example, we often want to know the transmission and reflection through a finite structure from a given source at a given frequency. Additionally, the response of a structure to currents placed at various points can reveal a host of interesting phenomena, from the enhancement (or suppression) of spontaneous emission to scattering losses from surface roughness.[4]

Here, the classic problem is to find the fields $\mathbf{E}(\mathbf{r})e^{-i\omega t}$ (and also $\mathbf{H} = -\frac{i}{\omega\mu_0}\nabla \times \mathbf{E}$) that are generated in response to some constant-frequency current source $\mathbf{J}(\mathbf{r})e^{-i\omega t}$ in linear media. By solving the Maxwell equations for $\mathbf{E}$ in terms of $\mathbf{J}$, we obtain the following *linear equation*:

$$\left[(\nabla \times \nabla \times) - \frac{\omega^2}{c^2}\varepsilon(\mathbf{r})\right]\mathbf{E}(\mathbf{r}) = i\omega\mu_0\mathbf{J}(\mathbf{r}). \tag{4}$$

When this equation is discretized into N unknowns, one obtains an $N \times N$ *matrix equation* of the form $Ax = b$ for the unknown "fields" (column vector) x in terms of the known "currents" b. Although solving such a set of equations directly would require $O(N^2)$ storage and $O(N^3)$ time, iterative methods are available, as for the eigenproblem above, that require $O(N)$ storage and roughly $O(N)$ time (see, e.g., Barrett et al., 1994), given a fast way to compute the matrix-vector product Ax.

Transmission and scattering calculations, however, typically require "open" boundaries. This means that the scattered fields must radiate to infinity instead of reflecting when they hit the edge of the (finite) computational region. Except for boundary-element methods, in which open boundaries are automatic, this problem is typically handled by adding a **perfectly matched layer** (**PML**) around the edges of the computational region. A PML is an artificial absorbing material designed so that there are (theoretically) no reflections from the edge of the material.[5]

In our analysis of the Maxwell eigenproblem, the Hermitian property of the eigenoperator played a central role. The analogous role for the frequency-domain problem (4) is occupied by a closely related property known as **Lorentz reciprocity**.[6] In particular, if $\hat{\Xi}$ is the linear operator on the left-hand side of

[4] See, for example, Fan et al. (1997) and Johnson et al. (2005).

[5] PML was initially designed for time-domain methods (see, e.g., Taflove and Hagness, 2000; Chew et al., 2001), but in frequency domain the same idea applies (and, in fact, is even simpler because only a single frequency ω need be handled).

[6] See, for example, Landau et al. (1984, §69) and Potton (2004).

equation (4), then reciprocity tells us that $(\mathbf{E}_1, \hat{\Xi}\mathbf{E}_2) = (\hat{\Xi}\mathbf{E}_1, \mathbf{E}_2)$ for the *unconjugated* inner product $(\mathbf{F}, \mathbf{G}) = \int \mathbf{F} \cdot \mathbf{G}$. Therefore, $(\mathbf{E}_1, \mathbf{J}_2) = (\mathbf{J}_1, \mathbf{E}_2)$. This theorem holds even for *complex*-valued ε (e.g. for PML absorbing boundaries), unlike the Hermitian property.[7]

Time-domain Simulations

Arguably the most general numerical methods for electromagnetism are those that simulate the full time-dependent Maxwell equations, propagating the fields in both space and time. Such *time-domain* methods can easily support strongly nonlinear or active (time-varying) media. Frequency-domain methods have more difficulty with those cases because frequency is no longer conserved. Time-domain methods can also be used solve the frequency-domain problems above, with some advantages and disadvantages as described below.

By far the most common technique for time-domain simulations is the **finite-difference time-domain** method, or **FDTD**. As the name implies, FDTD divides space and time into a grid (usually uniform) of discrete points and approximates the derivatives ($\nabla\times$ and $\partial/\partial t$) of the Maxwell equations by finite differences. The propagation in time, in particular, uses a "leap-frog" scheme where the $\mathbf{E}$ fields at time t are computed from the $\mathbf{E}$ fields at time $t - \Delta t$ along with the $\mathbf{H}$ fields at time $t - \Delta t/2$, and vice versa for $\mathbf{H}$ at $t + \Delta t/2$. In this way, the $\mathbf{E}$ and $\mathbf{H}$ field patterns are marched through time, offset by half of a time step Δt. The details of such methods, which employ a special staggered "Yee" grid in which the different components of each vector are associated with different locations on the grid cell, are well described in textbooks such as Taflove and Hagness (2000). Because FDTD software is widely available, it is more important to know how it is used and how it compares to frequency-domain methods.

The FDTD method is commonly employed to compute transmission and reflection spectra, much like the frequency-domain solvers of the previous section. Unlike a frequency-domain solver, however, time-domain methods can compute the response of a linear system at *many* frequencies with a *single* computation. The trick is to take the Fourier transform of the response to a short pulse. For example, suppose you want to know the transmitted flux $\text{Re}\int \mathbf{E}^* \times \mathbf{H}/2$ through a filter structure like those of chapter **??**, as a function of frequency. You use an FDTD code to send a short pulse (which has a broad bandwidth) into the structure, and observe the resulting fields $\mathbf{E}(t)$ and $\mathbf{H}(t)$ at the output plane. These are Fourier-transformed to yield $\mathbf{E}(\omega)$ and $\mathbf{H}(\omega)$, from which the flux is obtained at each ω. As for frequency-domain methods above, PML absorbing layers are used to simulate open boundaries.

[7] Technically, we require that ε and μ be symmetric 3×3 matrices, which is almost always true except in magneto-optic materials (see footnote **??** on page **??**). More generally, one can also formulate a reciprocity relation for integrals over finite volumes by including an appropriate surface term. The unconjugated inner product, here, not only allows us to use complex ε, but also turns out to be essential in order to make this surface term vanish for an integral over all space.

Why, then, would anyone ever use a frequency-domain solver? There are several reasons. First, because of the uncertainty principle of the Fourier transform, a time-domain method requires a long time to resolve a sharp spectral feature. Second, if you are interested in the steady-state response to a time-harmonic current source $\mathbf{J}(\mathbf{x})e^{-i\omega t}$, then with a time-domain method you must smoothly "turn on" the current and wait a long time for transient effects to die away. A frequency-domain method may be more efficient. Third, frequency-domain methods allow one to exploit finite-element or boundary-element methods that more efficiently discretize the problem, especially in cases for which highly non-uniform resolution is beneficial. (Moreover, if high spatial resolution is required, then FDTD methods require high temporal resolution as well, in order to maintain numerical stability. Thus, the time for a 3d FDTD simulation scales as resolution to the *fourth* power instead of to the third power as you might expect.) Although finite-element methods may be used for time-domain simulations too, they typically require *implicit time-stepping* in order to remain stable, which means that an $N \times N$ matrix equation must be solved at every time step. Boundary-element methods are even more complicated to implement in the time domain, because different points on surfaces are related by "retarded" Green's functions that are non-local in time.

Similarly, FDTD and other time-domain methods can be used to extract frequency eigenvalues. A time-domain eigensolver works by looking at the response of a structure to a short pulse. The eigenfrequencies are then identified as the peaks in the spectrum of the response. This method can even be used to identify resonant or leaky modes, because the width of the peak is related to the loss rate. In practice, one does not simply look for peaks in the Fourier transform; there are sophisticated signal-processing techniques that are even more accurate than the Fourier uncertainty principle would seem to imply (Mandelshtam and Taylor, 1997). Band structures are computed by imposing Bloch-periodic boundary conditions. Advantages of this technique are that one learns many eigenfrequencies at once, one can easily hunt for the eigenfrequencies in a specific portion of the spectrum (e.g., inside the gap, for a defect-mode calculation), and one can determine loss rates just as easily as frequencies. One disadvantage is that resolving degenerate or near-degenerate modes may take a long time, especially if the field patterns are desired in addition to the frequencies. (Obtaining the eigenfield pattern corresponding to a given eigenfrequency requires a separate simulation with a narrow-bandwidth source.) And unfortunately, the signal processing techniques involved in peak identification offer few guarantees. It is possible to miss eigenfrequencies, or locate spurious ones. Frequency-domain eigensolvers are more straightforward. They are essentially bulletproof, and are often faster at computing a few eigenfrequencies (especially at high resolutions).

A Planewave Eigensolver

In this book, the single most common type of computation we have performed is that of the band structure and eigenmodes. Therefore, it makes sense to explain our computational method for this frequency-domain eigenproblem in more detail. For a more thorough discussion of this scheme, which we have employed in our research for many years and have successfully compared to experimental results, see Meade et al. (1993) and Johnson and Joannopoulos (2001).

We employ a spectral method with a planewave basis. To see how that works, we begin in one dimension where it corresponds to the familiar **Fourier series**. In particular, we are solving (1) for a periodic function $u_k(x) = u_k(x + a)$ with period a. It is a remarkable fact, first postulated by Joseph Fourier at the beginning of the 19th century (amid some controversy), that any reasonable periodic function can be represented by an infinite sum of sines and cosines.[8] Or, in terms of complex exponentials,

$$u_k(x) = \sum_{n=-\infty}^{\infty} c_n(k) e^{i\frac{2\pi n}{a}x}, \tag{5}$$

for complex Fourier-series coefficients $c_n(k) = \frac{1}{a}\int_0^a e^{-i\frac{2\pi n}{a}x} u_k(x)$. Note that each term in the sum is a periodic function with period a. To use this representation on a computer, we need to truncate the sum to have a finite number (N) of terms. This is feasible because the coefficients c_n are decaying with $|n|$.[9] Thus, we use the N lowest-$|n|$ terms (say, $-N/2$ to $N/2-1$). We have transformed the problem from finding $u_k(x)$ to solving a set of linear equations for the N unknowns c_n. We will show how to write down these equations after first returning to the full vectorial problem.

We can generalize the Fourier series to several dimensions by recognizing that the $2\pi n/a$ in the complex exponential is none other than a reciprocal lattice vector of the one-dimensional lattice with period a. By analogy, the multi-dimensional Fourier series is

$$\mathbf{u_k}(\mathbf{r}) = \sum_{\mathbf{G}} \mathbf{c_G}(\mathbf{k}) e^{i\mathbf{G}\cdot\mathbf{r}}, \tag{6}$$

where the sum is over all of the reciprocal lattice vectors $\mathbf{G}$ (see appendix **??**), and $\mathbf{c_G} = \frac{1}{V}\int e^{-i\mathbf{G}\cdot\mathbf{r}}\mathbf{u_k}(\mathbf{r})$ where V is the unit-cell volume. By construction, each term in the sum is periodic in $\mathbf{r}$ with respect to the lattice vectors $\mathbf{R}$, since $\mathbf{G}\cdot\mathbf{R}$ is a multiple of 2π by definition. Note that, since $\mathbf{u_k}$ is a vector field, our Fourier-series coefficients $\mathbf{c_G}$ are now vectors as well. If we apply the transversality constraint (2) to equation (6), we obtain a simple constraint on the coefficients:

$$(\mathbf{k} + \mathbf{G}) \cdot \mathbf{c_G} = 0. \tag{7}$$

[8] A more rigorous statement, historical background, and pathological exceptions to this rule can be found in, e.g., Körner (1988).

[9] The rate of convergence depends on the smoothness of $u_k(x)$; if $u_k(x)$ is ℓ-times differentiable, then $|c_n|$ decreases faster than $1/|n|^\ell$ (Katznelson, 1968).

We see that transversality is automatically obeyed if we build the field $\mathbf{H} = \mathbf{u_k} e^{i\mathbf{k}\cdot\mathbf{r}}$ out of plane waves that are themselves transverse. Therefore, for each $\mathbf{G}$ we choose two perpendicular unit vectors $\hat{\mathbf{e}}_\mathbf{G}^{(1)}$ and $\hat{\mathbf{e}}_\mathbf{G}^{(2)}$ orthogonal to $\mathbf{k}+\mathbf{G}$, and write $\mathbf{c_G} = c_\mathbf{G}^{(1)}\hat{\mathbf{e}}_\mathbf{G}^{(1)} + c_\mathbf{G}^{(2)}\hat{\mathbf{e}}_\mathbf{G}^{(2)}$. We have thus reduced the problem to two unknowns $c_\mathbf{G}^{(1)}$ and $c_\mathbf{G}^{(2)}$ per $\mathbf{G}$, and we need not worry about transversality any more.

Given the transverse Fourier-series representation (6), we now derive a set of equations to determine the coefficients $\mathbf{c_G}$ by substituting (6) into the eigenequation (1). By Fourier transforming both sides of equation (1) (i.e., integrating with $\int e^{-i\mathbf{G}'\cdot\mathbf{r}}$), we obtain equations:

$$\sum_\mathbf{G} \left[-\varepsilon^{-1}_{\mathbf{G}'-\mathbf{G}} \cdot (\mathbf{k}+\mathbf{G}') \times (\mathbf{k}+\mathbf{G}) \times \right] \mathbf{c_G} = \frac{\omega^2}{c^2} \mathbf{c}_{\mathbf{G}'} \tag{8}$$

in terms of the Fourier transform (series coefficients) $\varepsilon^{-1}_\mathbf{G}$ of $\varepsilon^{-1}(\mathbf{r})$.

Equation (**??**) is an *infinite* set of linear equations for the infinite set of unknowns represented by $\mathbf{c_G}$. There are two related methods for truncating this infinite set of equations. First, we can simply take equation (**??**) for a finite set of plane waves $\mathbf{G}$ (e.g., a sphere around the origin), and throw out terms corresponding to larger $|\mathbf{G}|$ values on the assumption that they are small. This would involve computing the *exact* Fourier transform $\varepsilon^{-1}_\mathbf{G}$ of the inverse dielectric function, which might require many expensive numerical integrations. However, since we are going to throw out large $\mathbf{G}$ components anyway, we might as well go one step further and approximate $\varepsilon^{-1}_\mathbf{G}$ using the **discrete Fourier transform** (DFT). The DFT essentially replaces the Fourier transform by a discrete sum.[10]

Once we have truncated to a finite set of $\mathbf{G}$ values, equation (**??**) is a *finite* matrix eigenequation of the form $Ax = \omega^2 x$, where x is the column-vector of our unknown $c_\mathbf{G}^{(\ell)}$'s and A is the matrix of the coefficients on the left-hand side.[11] Viewed in this way, we have a problem: because the coefficients $\varepsilon^{-1}_{\mathbf{G}'-\mathbf{G}}$ are generally non-zero for all $\mathbf{G}'$ and $\mathbf{G}$, our matrix A is *dense* (mostly non-zero), and multiplying Ax takes $O(N^2)$ time. This is a death-knell for iterative methods, which require the multiplication Ax to be very rapid.

The saving grace for the planewave method is the existence of **fast Fourier transform** (FFT) algorithms, which can compute the multidimensional DFT over N points in $O(N \log N)$ time.[12] This means we can multiply $c_\mathbf{G}$ by the operator on the left-hand side of equation (**??**) via a three-step process. First, we take the cross-product $(\mathbf{k}+\mathbf{G}) \times \mathbf{c_G}$, which takes $O(N)$ time. Then, we

[10] Technically, the difference here between computing the Fourier transform exactly and via the DFT is the difference between a *Galerkin* method and a *collocation* method, where the latter means that we are enforcing the eigenequation at a set of discrete points (Boyd, 2001).

[11] The reason why we don't get a generalized eigenproblem $Ax = \omega^2 Bx$, or rather why B here is the identity, is that the planewave basis functions are orthogonal to one another.

[12] See, for example, Brigham (1988).

compute the (inverse) FFT to transform into position ($\mathbf{r}$) space, where we can multiply by $\varepsilon^{-1}(\mathbf{r})$ in $O(N)$ time. Finally, we FFT back to $\mathbf{G}$ coordinates to perform the final cross-product $(\mathbf{k}+\mathbf{G}')\times$. In all, this process takes $O(N \log N)$ time and requires $O(N)$ storage, which is fast enough for iterative methods to be efficient.[13]

We should also mention another important technical advantage of the planewave representation for iterative eigensolvers, having to do with efficient **preconditioners**. A preconditioner, in an iterative method, is essentially an approximate solution to the equation that is used to accelerate each step of the iteration. A good preconditioner can speed up the solution by *orders of magnitude*, from thousands of iterations to tens, but the development of such a preconditioner is a difficult and problem-dependent task. For the planewave method, however, efficient preconditioning is simple: one can precondition by considering only the diagonal entries of A, which are just $|\mathbf{k}+\mathbf{G}|^2$, since these entries dominate the problem for large $|\mathbf{G}|$.[14]

The accuracy of the planewave method is determined by the rate of convergence of the $\mathbf{c}_\mathbf{G}$ Fourier coefficients, since our errors are determined by the size of the large-$|\mathbf{G}|$ coefficients that we discarded. Unfortunately, for discontinuous dielectric structures, the corresponding Fourier transform converges rather slowly (the Fourier coefficients of ε^{-1} decrease proportional to $1/|\mathbf{G}|$), which leads to problems noted by Sözüer et al. (1992). (Related problems arise in finite-difference methods, where they are known as "staircasing" of dielectric interfaces.) Fortunately, these difficulties can be greatly reduced by a properly designed interpolation scheme, which smooths out the sharp dielectric interfaces without itself adding new errors to the frequency (Meade et al., 1993; Johnson and Joannopoulos, 2001). Similar benefits from smoothing accrue in other methods such as FDTD, and the underlying principle guiding the choice of interpolation scheme derives from the perturbation theory of chapter **??** (Farjadpour et al., 2006).

Further Reading and Free Software

A review of iterative planewave eigensolver methods for photonic crystals can be found in Johnson and Joannopoulos (2001), and a good reference on the finite-difference time-domain (FDTD) method is Taflove and Hagness (2000). An overview of finite-element and boundary-element methods in electromagnetism can be found in Chew et al. (2001), and several other methods are described in Yasumoto (2005). For a general (not specifically electromagnetic) introduction to: spectral methods, see e.g. Boyd (2001); boundary-element methods, see e.g. Bonnet (1999); finite-difference methods, see e.g. Strikwerda

[13] Put another way, we are taking advantage of the fact that equation (**??**) is in the form of a discrete *convolution*, which can be evaluated in $O(N \log N)$ operations by a pair of FFTs using the convolution theorem (DFTs turn convolutions into pointwise products).

[14] This is motivated by "kinetic-energy" preconditioners from quantum mechanics (Payne et al., 1992; Johnson and Joannopoulos, 2001). Other strategies to find preconditioners are described in Barrett et al. (1994).

(1989). Broad surveys of iterative methods for linear equations and eigenproblems can be found in Barrett et al. (1994) and Bai et al. (2000), respectively.

Numerous commercial software products are available for electromagnetic problems and can easily be found in the usual catalogs. In the course of our own research, we have become strong proponents of *free software* (a.k.a. *open-source* code), which has many advantages. Besides having low cost, it is also portable, customizable, and vendor-independent. In particular, we have developed and released a free program called *MPB* (ab-initio.mit.edu/mpb) for computing band structures and eigenmodes by a planewave method, and a program called *Meep* (ab-initio.mit.edu/meep) that implements the FDTD method. These two programs performed all the calculations in this book. Another free program that we have found useful is *CAMFR* (camfr.sourceforge.net) by Bienstman (2001), which is a transfer-matrix frequency-domain solver variant that is especially efficient for structures that can be subdivided into a sequence of uniform cross-sections along a given direction.

References

Anderson, E., Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. 1999. *LAPACK Users' Guide*. 3rd ed. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Bai, Zhaojun, James Demmel, Jack Dongarra, Axel Ruhe, and Henk Van Der Vorst, eds. 2000. *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*. Philadelphia: SIAM.

Barrett, R., M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst. 1994. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. 2nd ed. Philadelphia, PA: SIAM.

Bienstman, Peter. 2001. "Rigorous and efficient modelling of wavelength scale photonic components." PhD thesis, Ghent University. Ghent, Belgium.

Bonnet, Marc. 1999. *Boundary Integral Equation Methods for Solids and Fluids*. Chichester, England: Wiley.

Boyd, J. P. 2001. *Chebyshev and Fourier Spectral Methods*. 2nd ed. New York: Dover.

Brigham, E. O. 1988. *The Fast Fourier Transform and Its Applications*. Englewood Cliffs, NJ: Prentice-Hall.

Chew, Weng Cho, Jian-Ming Jin, Eric Michielssen, and Jiming Song, eds. 2001. *Fast and Efficient Algorithms in Computational Electromagnetics*. Norwood, MA: Artech.

Fan, Shanhui, Pierre R. Villeneuve, J. D. Joannopoulos, and E. F. Schubert. 1997. "High extraction efficiency of spontaneous emission from slabs of photonic crystals." *Phys. Rev. Lett.* 78:3294–3297.

Farjadpour, A., David Roundy, Alejandro Rodriguez, M. Ibanescu, Peter Bermel, J. D. Joannopoulos, Steven G. Johnson, and G. W. Burr. 2006. "Improving accuracy by subpixel smoothing in the finite-difference time domain." *Opt. Lett.* 31:2972–2974.

Johnson, Steven G., and J. D. Joannopoulos. 2001. "Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis." *Opt. Express* 8(3):173–190.

Johnson, Steven G., Mihai Ibanescu, M. Skorobogatiy, Ori Weisberg, Torkel D. Engeness, Marin Soljačić, Steven A. Jacobs, J. D. Joannopoulos, and Yoel Fink. 2001. "Low-loss asymptotically single-mode propagation in large-core OmniGuide fibers." *Opt. Express* 9(13):748–779.

Johnson, Steven G., Peter Bienstman, M. Skorobogatiy, Mihai Ibanescu, Elefterios Lidorikis, and J. D. Joannopoulos. 2002. "Adiabatic theorem and continuous coupled-mode theory for efficient taper transitions in photonic crystals." *Phys. Rev. E* 66:066608.

Johnson, Steven G., M. L. Povinelli, M. Soljačić, A. Karalis, S. Jacobs, and J. D. Joannopoulos. 2005. "Roughness losses and volume-current methods in photonic-crystal waveguides." *Appl. Phys. B* 81:283–293.

Katznelson, Yitzhak. 1968. *An Introduction to Harmonic Analysis*. New York: Dover.

Körner, T. W. 1988. *Fourier Analysis*. Cambridge Univ. Press.

Landau, L., E. M. Lifshitz, and L. P. Pitaevskii. 1984. *Electrodynamics of Continuous Media*. 2nd ed. Oxford: Butterworth-Heinemann.

Mandelshtam, V. A., and H. S. Taylor. 1997. "Harmonic inversion of time signals and its applications." *J. Chem. Phys.* 107(17):6756–6769.

Meade, Robert D., A. M. Rappe, K. D. Brommer, J. D. Joannopoulos, and O. L. Alerhand. 1993. "Accurate theoretical analysis of photonic band-gap materials." *Phys. Rev. B* 48:8434–8437. Erratum: S. G. Johnson, *ibid.* **55**, 15942 (1997).

Payne, M. C., M. P. Tater, D. C. Allan, T. A. Arias, and J. D. Joannopoulos. 1992. "Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients." *Rev. Mod. Phys.* 64:1045–1097.

Potton, R. J. 2004. "Reciprocity in optics." *Reports Prog. Phys.* 67:717–754.

Sözüer, H. S., J. W. Haus, and R. Inguva. 1992. "Photonic bands: Convergence problems with the plane-wave method." *Phys. Rev. B* 45:13962–13972.

Strikwerda, John C. 1989. *Finite Difference Schemes and Partial Differential Equations*. Pacific Grove, CA: Wadsworth and Brooks/Cole.

Taflove, Allen, and Susan C. Hagness. 2000. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Norwood, MA: Artech.

Yasumoto, Kiyotoshi, ed. 2005. *Electromagnetic Theory and Applications for Photonic Crystals*. CRC Press.