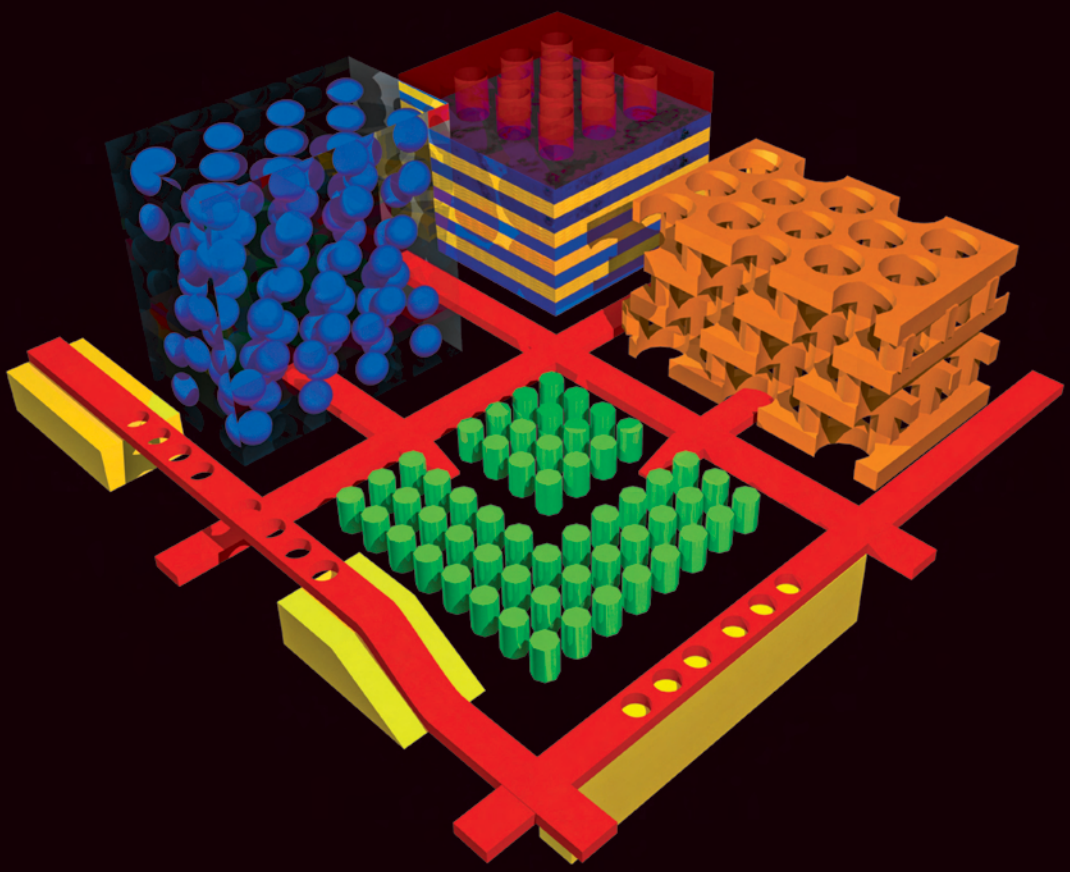
Photonic Crystals

Molding the Flow of Light

SECOND EDITION



John D. Joannopoulos
Steven G. Johnson
Joshua N. Winn
Robert D. Meade

Photonic Crystals

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Photonic Crystals

Molding the Flow of Light

Second Edition

John D. Joannopoulos

Steven G. Johnson

Joshua N. Winn

Robert D. Meade

PRINCETON UNIVERSITY PRESS • PRINCETON AND OXFORD

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Copyright © 2008 by Princeton University Press

Published by Princeton University Press, 41 William Street, Princeton,
New Jersey 08540
In the United Kingdom: Princeton University Press, 3 Market Place,
Woodstock, Oxfordshire OX20 1SY

All Rights Reserved

Library of Congress Cataloging-in-Publication Data
Joannopoulos, J. D. (John D.), 1947-
Photonic crystals: molding the flow of light/John D. Joannopoulos . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN: 978-0-691-12456-8 (acid-free paper)
1. Photons. 2. Crystal optics. I. Joannopoulos, J. D. (John D.), 1947- II. Title.
QC793.5.P427 J63 2008
548′.9–dc22 2007061025

British Library Cataloging-in-Publication Data is available

This book has been composed in Palatino

Printed on acid-free paper. ∞

press.princeton.edu

Printed in Singapore

10 9 8 7 6 5 4 3 2 1

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

To Kyriaki and G. G. G.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

To see a World in a Grain of Sand,
And a Heaven in a Wild Flower,
Hold Infinity in the palm of your hand
And Eternity in an hour.

—*William Blake, Auguries of Innocence (1803)*

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

CONTENTS

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

PREFACE TO THE SECOND EDITION

We were delighted with the positive response to the first edition of this book. There is, naturally, always some sense of trepidation when one writes the first text book at the birth of a new field. One dearly hopes the field will continue to grow and blossom, but then again, will the subject matter of the book quickly become obsolete? To attempt to alleviate the latter, we made a conscious effort in the first edition to focus on the fundamental concepts and building blocks of this new field and leave out any speculative areas. Given the continuing interest in the first edition, even after a decade of exponential growth of the field, it appears that we may have succeeded in this regard. Of course, with great growth come many new phenomena and a deeper understanding of old phenomena. We felt, therefore, that the time was now ripe for an updated and expanded second edition.

As before, we strove in this edition to include new concepts, phenomena and descriptions that are well understood—material that would stand the test of advancements over time.

Many of the original chapters are expanded with new sections, in addition to innumerable revisions to the old sections. For example, chapter 2 now contains a section introducing the useful technique of perturbation analysis and a section on understanding the subtle differences between discrete and continuous frequency ranges. Chapter 3 includes a section describing the basics of index guiding and a section on how to understand the Bloch-wave propagation velocity. Chapter 4 includes a section on how to best quantify the band gap of a photonic crystal and a section describing the novel phenomenon of omnidirectional reflectivity in multilayer film systems. Chapter 5 now contains an expanded section on point defects and a section on linear defects and waveguides. Chapter 6 was revised considerably to focus on many new aspects of 3D photonic crystal structures, including the photonic structure of several well known geometries. Chapters 7 through 9 are all new, describing hybrid photonic-crystal structures consisting, respectively, of 1D-periodic dielectric waveguides, 2D-periodic photonic-crystal slabs, and photonic-crystal fibers. The final chapter, chapter 10 (chapter 7 in the first edition), is again focused on designing photonic crystals for applications, but now contains many more examples. This chapter has also been expanded to include an introduction and practical guide to temporal coupled-mode theory. This is a very simple, convenient, yet powerful analytical technique for understanding and predicting the behavior of many types of photonic devices.

Two of the original appendices have also been considerably expanded. Appendix C now includes plots of gap size and optimal parameters vs. index contrast for both 2D and 3D photonic crystals. Appendix D now provides a completely new description of computational photonics, surveying computations in both the frequency and time domains.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

The second edition also includes two other major changes. The first is a change to SI units. Admittedly, this affects only some of the equations in chapters 2 and 3; the "master equation" remains unaltered. The second change is to a new color table for plotting the electric and magnetic fields. We hope the reader will agree that the new color table is a significant improvement over the old color table, providing a much cleaner and clearer description of the localization and sign-dependence of the fields.

In preparing the second edition, we should like to express our sincere gratitude to Margaret O'Meara, the administrative assistant of the Condensed Matter Theory Group at MIT, for all the time and effort she unselfishly provided. We should also like to give a big *Thank You*! to our editor Ingrid Gnerlich for her patience and understanding when deadlines were not met and for her remarkable good will with all aspects of the process.

We are also very grateful to many colleagues: Eli Yablonovitch, David Norris, Marko Lončar, Shawn Lin, Leslie Kolodziejski, Karl Koch, and Kiyoshi Asakawa, for providing us with illustrations of their original work, and Yoel Fink, Shanhui Fan, Peter Bienstman, Mihai Ibanescu, Michelle Povinelli, Marin Soljacic, Maksim Skorobogatiy, Lionel Kimerling, Lefteris Lidorikis, K. C. Huang, Jerry Chen, Hermann Haus, Henry Smith, Evan Reed, Erich Ippen, Edwin Thomas, David Roundy, David Chan, Chiyan Luo, Attila Mekis, Aristos Karalis, Ardavan Farjadpour, and Alejandro Rodriguez, for numerous collaborations.

Cambridge, Massachusetts, 2006

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

PREFACE TO THE FIRST EDITION

It is always difficult to write a book about a topic that is still a subject of active research. Part of the challenge lies in translating research papers directly into a text. Without the benefit of decades of classroom instruction, there is no existing body of pedagogical arguments and exercises to draw from.

Even more challenging is the task of deciding which material to include. Who knows which approaches will withstand the test of time? It is impossible to know, so in this text we have tried to include only those subjects of the field which we consider most likely to be timeless. That is, we present the fundamentals and the proven results, hoping that afterwards the reader will be prepared to read and understand the current literature. Certainly there is much to add to this material as the research continues, but we have tried to take care that nothing need be subtracted. Of course this has come at the expense of leaving out new and exciting results which are a bit more speculative.

If we have succeeded in these tasks, it is only because of the assistance of dozens of colleagues and friends. In particular, we have benefited from collaborations with Oscar Alerhand, G. Arjavalingam, Karl Brommer, Shanhui Fan, Ilya Kurland, Andrew Rappe, Bill Robertson, and Eli Yablonovitch. We also thank Paul Gourley and Pierre Villeneuve for their contributions to this book. In addition, we gratefully thank Tomas Arias and Kyeongjae Cho for helpful insights and productive conversations. Finally, we would like to acknowledge the partial support of the Office of Naval Research and the Army Research Office while this manuscript was being prepared.

Cambridge, Massachusetts, 1995

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Photonic Crystals

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

1

Introduction

Controlling the Properties of Materials

Many of the true breakthroughs in our technology have resulted from a deeper understanding of the properties of materials. The rise of our ancestors from the Stone Age through the Iron Age is largely a story of humanity's increasing recognition of the utility of natural materials. Prehistoric people fashioned tools based on their knowledge of the durability of stone and the hardness of iron. In each case, humankind learned to extract a material from the Earth whose fixed properties proved useful.

Eventually, early engineers learned to do more than just take what the Earth provides in raw form. By tinkering with existing materials, they produced substances with even more desirable properties, from the luster of early bronze alloys to the reliability of modern steel and concrete. Today we boast a collection of wholly artificial materials with a tremendous range of *mechanical* properties, thanks to advances in metallurgy, ceramics, and plastics.

In this century, our control over materials has spread to include their *electrical* properties. Advances in semiconductor physics have allowed us to tailor the conducting properties of certain materials, thereby initiating the transistor revolution in electronics. It is hard to overstate the impact that the advances in these fields have had on our society. With new alloys and ceramics, scientists have invented high-temperature superconductors and other exotic materials that may form the basis of future technologies.

In the last few decades, a new frontier has opened up. The goal in this case is to control the *optical* properties of materials. An enormous range of technological developments would become possible if we could engineer materials that respond to light waves over a desired range of frequencies by perfectly reflecting them, or allowing them to propagate only in certain directions, or confining them within a specified volume. Already, fiber-optic cables, which simply guide light, have revolutionized the telecommunications industry. Laser engineering, high-speed computing, and spectroscopy are just a few of the fields next in line to reap the

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

benefits from the advances in optical materials. It is with these goals in mind that this book is written.

Photonic Crystals

What sort of material can afford us complete control over light propagation? To answer this question, we rely on an analogy with our successful electronic materials. A crystal is a periodic arrangement of atoms or molecules. The pattern with which the atoms or molecules are repeated in space is the crystal **lattice**. The crystal presents a periodic potential to an electron propagating through it, and both the constituents of the crystal and the geometry of the lattice dictate the conduction properties of the crystal.

The theory of quantum mechanics in a periodic potential explains what was once a great mystery of physics: In a conducting crystal, why do electrons propagate like a diffuse gas of free particles? How do they avoid scattering from the constituents of the crystal lattice? The answer is that electrons propagate as waves, and waves that meet certain criteria can travel through a periodic potential without scattering (although they will be scattered by defects and impurities).

Importantly, however, the lattice can also prohibit the propagation of certain waves. There may be **gaps** in the energy band structure of the crystal, meaning that electrons are forbidden to propagate with certain energies in certain directions. If the lattice potential is strong enough, the gap can extend to cover all possible propagation directions, resulting in a **complete band gap**. For example, a semiconductor has a complete band gap between the valence and conduction energy bands.

The optical analogue is the *photonic* crystal, in which the atoms or molecules are replaced by macroscopic media with differing dielectric constants, and the periodic potential is replaced by a periodic dielectric function (or, equivalently, a periodic index of refraction). If the dielectric constants of the materials in the crystal are sufficiently different, and if the absorption of light by the materials is minimal, then the refractions and reflections of light from all of the various interfaces can produce many of the same phenomena for *photons* (light modes) that the atomic potential produces for *electrons*. One solution to the problem of optical control and manipulation is thus a **photonic crystal**, a low-loss periodic dielectric medium. In particular, we can design and construct photonic crystals with **photonic band gaps**, preventing light from propagating in certain directions with specified frequencies (i.e., a certain range of wavelengths, or "colors," of light). We will also see that a photonic crystal can allow propagation in anomalous and useful ways.

To develop this concept further, consider how metallic waveguides and cavities relate to photonic crystals. Metallic waveguides and cavities are widely used to control microwave propagation. The walls of a metallic cavity prohibit the propagation of electromagnetic waves with frequencies below a certain threshold frequency, and a metallic waveguide allows propagation only along its axis. It would be extremely useful to have these same capabilities for electromagnetic

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

waves with frequencies outside the microwave regime, such as visible light. However, visible light energy is quickly dissipated within metallic components, which makes this method of optical control impossible to generalize. Photonic crystals allow the useful properties of cavities and waveguides to be generalized and scaled to encompass a wider range of frequencies. We may construct a photonic crystal of a given geometry with millimeter dimensions for microwave control, or with micron dimensions for infrared control.

Another widely used optical device is a multilayer dielectric mirror, such as a *quarter-wave stack*, consisting of alternating layers of material with different dielectric constants. Light of the proper wavelength, when incident on such a layered material, is completely reflected. The reason is that the light wave is partially reflected at each layer interface and, if the spacing is periodic, the multiple reflections of the incident wave interfere destructively to eliminate the forward-propagating wave. This well-known phenomenon, first explained by Lord Rayleigh in 1887, is the basis of many devices, including dielectric mirrors, dielectric Fabry–Perot filters, and distributed feedback lasers. All contain low-loss dielectrics that are periodic in one dimension, and by our definition they are *one-dimensional* photonic crystals. Even these simplest of photonic crystals can have surprising properties. We will see that layered media can be designed to reflect light that is incident from any angle, with any polarization—an omnidirectional reflector—despite the common intuition that reflection can be arranged only for near-normal incidence.

If, for some frequency range, a photonic crystal prohibits the propagation of electromagnetic waves of *any* polarization traveling in *any* direction from *any* source, we say that the crystal has a **complete photonic band gap**. A crystal with a complete band gap will obviously be an omnidirectional reflector, but the converse is not necessarily true. As we shall see, the layered dielectric medium mentioned above, which cannot have a complete gap (because material interfaces occur only along one axis), can still be designed to exhibit omnidirectional reflection—but only for light sources far from the crystal. Usually, in order to create a complete photonic band gap, one must arrange for the dielectric lattice to be periodic along three axes, constituting a *three-dimensional* photonic crystal. However, there are exceptions. A small amount of disorder in an otherwise periodic medium will not destroy a band gap (Fan et al., 1995*b*; Rodriguez et al., 2005), and even a highly disordered medium can prevent propagation in a useful way through the mechanism of **Anderson localization** (John, 1984). Another interesting nonperiodic class of materials that can have complete photonic band gaps are *quasi-crystalline* structures (Chan et al., 1998).

An Overview of the Text

Our goal in writing this textbook was to provide a comprehensive description of the propagation of light in photonic crystals. We discuss the properties of photonic crystals of gradually increasing complexity, beginning with the simplest

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

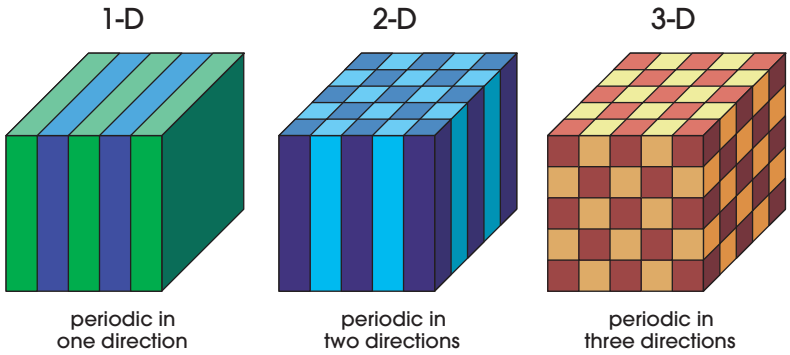


Figure 1: Simple examples of one-, two-, and three-dimensional photonic crystals. The different colors represent materials with different dielectric constants. The defining feature of a photonic crystal is the periodicity of dielectric material along one or more axes.

case of one-dimensional crystals, and proceeding to the more intricate and useful properties of two- and three-dimensional systems (see figure 1). After equipping ourselves with the appropriate theoretical tools, we attempt to convey a useful intuition about *which* structures yield *what* properties, and *why*?

This textbook is designed for a broad audience. The only prerequisites are a familiarity with the macroscopic Maxwell equations and the notion of harmonic modes (which are often referred to by other names, such as eigenmodes, normal modes, and Fourier modes). From these building blocks, we develop all of the needed mathematical and physical tools. We hope that interested undergraduates will find the text approachable, and that professional researchers will find our heuristics and results to be useful in designing photonic crystals for their own applications.

Readers who are familiar with quantum mechanics and solid-state physics are at some advantage, because our formalism owes a great deal to the techniques and nomenclature of those fields. Appendix A explores this analogy in detail. Photonic crystals are a marriage of solid-state physics and electromagnetism. Crystal structures are citizens of solid-state physics, but in photonic crystals the electrons are replaced by electromagnetic waves. Accordingly, we present the basic concepts of both subjects before launching into an analysis of photonic crystals. In chapter 2, we discuss the macroscopic Maxwell equations as they apply to dielectric media. These equations are cast as a single Hermitian differential equation, a form in which many useful properties become easy to demonstrate: the orthogonality of modes, the electromagnetic variational theorem, and the scaling laws of dielectric systems.

Chapter 3 presents some basic concepts of solid-state physics and symmetry theory as they apply to photonic crystals. It is common to apply symmetry arguments to understand the propagation of electrons in a periodic crystal potential. Similar arguments also apply to the case of light propagating in a photonic crystal. We examine the consequences of translational, rotational,

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

mirror-reflection, inversion, and time-reversal symmetries in photonic crystals, while introducing some terminology from solid-state physics.

To develop the basic notions underlying photonic crystals, we begin by reviewing the properties of one-dimensional photonic crystals. In chapter 4, we will see that one-dimensional systems can exhibit three important phenomena: photonic band gaps, localized modes, and surface states. Because the index contrast is only along one direction, the band gaps and the bound states are limited to that direction. Nevertheless, this simple and traditional system illustrates most of the physical features of the more complex two- and three-dimensional photonic crystals, and can even exhibit omnidirectional reflection.

In chapter 5, we discuss the properties of two-dimensional photonic crystals, which are periodic in two directions and homogeneous in the third. These systems can have a photonic band gap in the plane of periodicity. By analyzing field patterns of some electromagnetic modes in different crystals, we gain insight into the nature of band gaps in complex periodic media. We will see that defects in such two-dimensional crystals can localize modes in the plane, and that the faces of the crystal can support surface states.

Chapter 6 addresses three-dimensional photonic crystals, which are periodic along three axes. It is a remarkable fact that such a system can have a *complete* photonic band gap, so that no propagating states are allowed in any direction in the crystal. The discovery of particular dielectric structures that possess a complete photonic band gap was one of the most important achievements in this field. These crystals are sufficiently complex to allow localization of light at point defects and propagation along linear defects.

Chapters 7 and 8 consider *hybrid* structures that combine band gaps in one or two directions with *index-guiding* (a generalization of total internal reflection) in the other directions. Such structures *approximate* the three-dimensional control over light that is afforded by a complete three-dimensional band gap, but at the same time are much easier to fabricate. Chapter 9 describes a different kind of incomplete-gap structure, **photonic-crystal fibers**, which use band gaps or index-guiding from one- or two-dimensional periodicity to guide light along an optical fiber.

Finally, in chapter 10, we use the tools and ideas that were introduced in previous chapters to design some simple optical components. Specifically, we see how resonant cavities and waveguides can be combined to form filters, bends, splitters, nonlinear "transistors," and other devices. In doing so, we develop a powerful analytical framework known as *temporal coupled-mode theory*, which allows us to easily predict the behavior of such combinations. We also examine the reflection and refraction phenomena that occur when light strikes an interface of a photonic crystal. These examples not only illustrate the device applications of photonic crystals, but also provide a brief review of the material contained elsewhere in the text.

We should also mention the appendices, which provide a brief overview of the reciprocal-lattice concept from solid-state physics, survey the gaps that arise in various two- and three-dimensional photonic crystals, and outline the numerical methods that are available for computer simulations of photonic structures.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

2

Electromagnetism in Mixed Dielectric Media

In order to study the propagation of light in a photonic crystal, we begin with the Maxwell equations. After specializing to the case of a mixed dielectric medium, we cast the Maxwell equations as a linear Hermitian eigenvalue problem. This brings the electromagnetic problem into a close analogy with the Schrödinger equation, and allows us to take advantage of some well-established results from quantum mechanics, such as the orthogonality of modes, the variational theorem, and perturbation theory. One way in which the electromagnetic case differs from the quantum-mechanical case is that photonic crystals do not generally have a fundamental scale, in either the spatial coordinate or in the potential strength (the dielectric constant). This makes photonic crystals scalable in a way that traditional crystals are not, as we will see later in this chapter.

The Macroscopic Maxwell Equations

All of macroscopic electromagnetism, including the propagation of light in a photonic crystal, is governed by the four macroscopic **Maxwell equations**. In SI units,[1] they are

$$\begin{aligned} \nabla \cdot \mathbf{B} = 0 \quad \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \\ \nabla \cdot \mathbf{D} = \rho \quad \nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J} \end{aligned} \tag{1}$$

[1] The first edition used cgs units, in which constants such as ε_0 and μ_0 are replaced by factors of 4π and c here and there, but the choice of units is mostly irrelevant in the end. They do not effect the form of our "master" equation (7). Moreover, we will express all quantities of interest—frequencies, geometries, gap sizes, and so on—as dimensionless ratios; see also the section The Size of the Band Gap of chapter 4.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

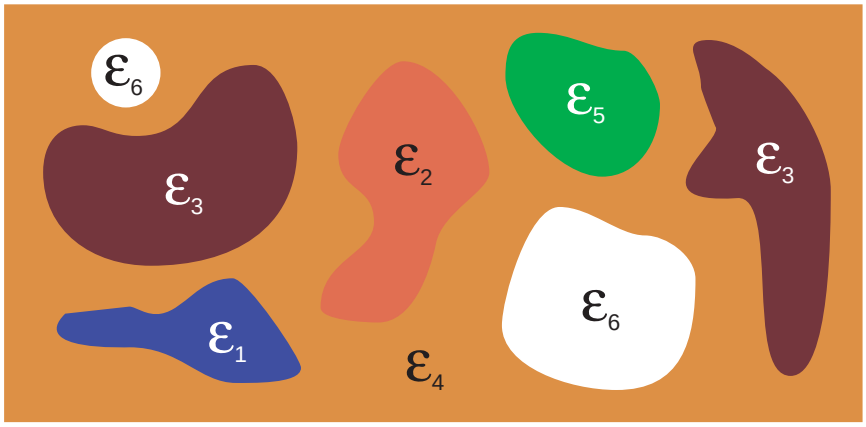


Figure 1: A composite of macroscopic regions of homogeneous dielectric media. There are no charges or currents. In general, $\varepsilon(\mathbf{r})$ in equation (1) can have any prescribed spatial dependence, but our attention will focus on materials with patches of homogeneous dielectric, such as the one illustrated here.

where (respectively) **E** and **H** are the macroscopic electric and magnetic fields, **D** and **B** are the displacement and magnetic induction fields, and ρ and **J** are the free charge and current densities. An excellent derivation of these equations from their microscopic counterparts is given in Jackson (1998).

We will restrict ourselves to propagation within a **mixed dielectric medium**, a composite of regions of homogeneous dielectric material as a function of the (cartesian) position vector **r**, in which the structure does not vary with time, and there are no free charges or currents. This composite need not be periodic, as illustrated in figure 1. With this type of medium in mind, in which light propagates but there are no sources of light, we can set $\rho = 0$ and $\mathbf{J} = 0$.

Next we relate **D** to **E** and **B** to **H** with the constitutive relations appropriate for our problem. Quite generally, the components D_i of the displacement field **D** are related to the components E_i of the electric field **E** via a power series, as in Bloembergen (1965):

$$D_i/\varepsilon_0 = \sum_j \varepsilon_{ij} E_j + \sum_{j,k} \chi_{ijk} E_j E_k + O(E^3), \tag{2}$$

where $\varepsilon_0 \approx 8.854 \times 10^{-12}$ Farad/m is the vacuum permittivity. However, for many dielectric materials, it is reasonable to use the following approximations. First, we assume the field strengths are small enough so that we are in the linear regime, so that χ_{ijk} (and all higher-order terms) can be neglected. Second, we assume the material is macroscopic and isotropic,[2] so that $\mathbf{E}(\mathbf{r},\omega)$ and $\mathbf{D}(\mathbf{r},\omega)$ are related by ε_0 multipled by a scalar **dielectric function** $\varepsilon(\mathbf{r},\omega)$, also called

[2] It is straightforward to generalize this formalism to anisotropic media in which **D** and **E** are related by a Hermitian dielectric tensor $\varepsilon_0 \varepsilon_{ij}$.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

the **relative permittivity**.[3] Third, we ignore any explicit frequency dependence (**material dispersion**) of the dielectric constant. Instead, we simply choose the value of the dielectric constant appropriate to the frequency range of the physical system we are considering. Fourth, we focus primarily on **transparent materials**, which means we can treat $\varepsilon(\mathbf{r})$ as purely real[4] and positive.[5]

Assuming these four approximations to be valid, we have $\mathbf{D}(\mathbf{r}) = \varepsilon_0\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r})$. A similar equation relates $\mathbf{B}(\mathbf{r}) = \mu_0\mu(\mathbf{r})\mathbf{H}(\mathbf{r})$ (where $\mu_0 = 4\pi \times 10^{-7}$ Henry/m is the vacuum permeability), but for most dielectric materials of interest the relative magnetic permeability $\mu(\mathbf{r})$ is very close to unity and we may set $\mathbf{B} = \mu_0\mathbf{H}$ for simplicity.[6] In that case, ε is the square of the **refractive index** n that may be familiar from Snell's law and other formulas of classical optics. (In general, $n = \sqrt{\varepsilon\mu}$.)

With all of these assumptions in place, the Maxwell equations (1) become

$$\begin{aligned} \nabla \cdot \mathbf{H}(\mathbf{r},t) = 0 \qquad & \nabla \times \mathbf{E}(\mathbf{r},t) + \mu_0 \frac{\partial \mathbf{H}(\mathbf{r},t)}{\partial t} = 0 \\ \nabla \cdot [\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r},t)] = 0 \quad & \nabla \times \mathbf{H}(\mathbf{r},t) - \varepsilon_0\varepsilon(\mathbf{r}) \frac{\partial \mathbf{E}(\mathbf{r},t)}{\partial t} = 0. \end{aligned} \tag{3}$$

The reader might reasonably wonder whether we are missing out on interesting physical phenomena by restricting ourselves to linear and lossless materials. We certainly are, and we will return to this question in the section The Effect of Small Perturbations and in chapter 10. Nevertheless, it is a remarkable fact that many interesting and useful properties arise from the elementary case of linear, lossless materials. In addition, the theory of these materials is much simpler to understand and is practically exact, making it an excellent foundation on which to build the theory of more complex media. For these reasons, we will be concerned with linear and lossless materials for most of this text.

In general, both $\mathbf{E}$ and $\mathbf{H}$ are complicated functions of both time and space. Because the Maxwell equations are linear, however, we can separate the time dependence from the spatial dependence by expanding the fields into a set of **harmonic modes**. In this and the following sections we will examine the restrictions that the Maxwell equations impose on a field pattern that varies

[3] Some authors use ε_r (or K, or k, or κ) for the *relative* permittivity and ε for the *permittivity* $\varepsilon_0\varepsilon_r$. Here, we adopt the common convention of dropping the r subscript, since we work only with the dimensionless ε_r.

[4] Complex dielectric constants are used to account for absorption, as in Jackson (1998). Later, in the section The Effect of Small Perturbations, we will show how to include *small* absorption losses.

[5] A negative dielectric constant is indeed a useful description of some materials, such as metals. The limit $\varepsilon \to -\infty$ corresponds to a **perfect metal** into which light cannot penetrate. Combinations of metals and transparent dielectrics can also be used to create photonic crystals (for some early work in this area, see e.g., McGurn and Maradudin, 1993; Kuzmiak et al., 1994; Sigalas et al., 1995; Brown and McMahon, 1995; Fan et al., 1995*c*; Sievenpiper et al., 1996), a topic we return to in the subsection The scalar limit and LP modes of chapter 9.

[6] It is straightforward to include $\mu \neq 1$; see footnote 17 on page 17.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

sinusoidally (harmonically) with time. This is no great limitation, since we know by Fourier analysis that we can build *any* solution with an appropriate combination of these harmonic modes. Often we will refer to them simply as *modes* or *states* of the system.

For mathematical convenience, we employ the standard trick of using a complex-valued field and remembering to take the real part to obtain the physical fields. This allows us to write a harmonic mode as a spatial pattern (or "mode profile") times a complex exponential:

$$\begin{aligned} \mathbf{H}(\mathbf{r},t) &= \mathbf{H}(\mathbf{r})e^{-i\omega t} \\ \mathbf{E}(\mathbf{r},t) &= \mathbf{E}(\mathbf{r})e^{-i\omega t}. \end{aligned} \tag{4}$$

To find the equations governing the mode profiles for a given frequency, we insert the above equations into (3). The two divergence equations give the conditions

$$\nabla \cdot \mathbf{H}(\mathbf{r}) = 0, \quad \nabla \cdot [\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r})] = 0, \tag{5}$$

which have a simple physical interpretation: there are no point sources or sinks of displacement and magnetic fields in the medium. Equivalently, the field configurations are built up of electromagnetic waves that are **transverse**. That is, if we have a **plane wave** $\mathbf{H}(\mathbf{r}) = \mathbf{a}\exp(i\mathbf{k}\cdot\mathbf{r})$, for some wave vector $\mathbf{k}$, equation (5) requires that $\mathbf{a}\cdot\mathbf{k}=0$. We can now focus our attention only on the other two of the Maxwell equations as long as we are always careful to enforce this transversality requirement.

The two curl equations relate $\mathbf{E}(\mathbf{r})$ to $\mathbf{H}(\mathbf{r})$:

$$\begin{aligned} \nabla \times \mathbf{E}(\mathbf{r}) - i\omega\mu_0\mathbf{H}(\mathbf{r}) &= 0 \\ \nabla \times \mathbf{H}(\mathbf{r}) + i\omega\varepsilon_0\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r}) &= 0. \end{aligned} \tag{6}$$

We can decouple these equations in the following way. Divide the bottom equation of (6) by $\varepsilon(\mathbf{r})$, and then take the curl. Then use the first equation to eliminate $\mathbf{E}(\mathbf{r})$. Morever, the constants ε_0 and μ_0 can be combined to yield the vacuum speed of light, $c = 1/\sqrt{\varepsilon_0\mu_0}$. The result is an equation entirely in $\mathbf{H}(\mathbf{r})$:

$$\boxed{\nabla \times \left(\frac{1}{\varepsilon(\mathbf{r})}\nabla \times \mathbf{H}(\mathbf{r})\right) = \left(\frac{\omega}{c}\right)^2 \mathbf{H}(\mathbf{r}).} \tag{7}$$

This is the **master equation**. Together with the divergence equation (5), it tells us everything we need to know about $\mathbf{H}(\mathbf{r})$. Our strategy will be as follows: for a given structure $\varepsilon(\mathbf{r})$, solve the master equation to find the modes $\mathbf{H}(\mathbf{r})$ and the corresponding frequencies, subject to the transversality requirement. Then use

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

the second equation of (6) to recover $\mathbf{E}(\mathbf{r})$:

$$\mathbf{E}(\mathbf{r}) = \frac{i}{\omega\varepsilon_0\varepsilon(\mathbf{r})}\nabla \times \mathbf{H}(\mathbf{r}). \tag{8}$$

Using this procedure guarantees that $\mathbf{E}$ satisfies the transversality requirement $\nabla \cdot \varepsilon\mathbf{E} = 0$, because the divergence of a curl is always zero. Thus, we need only impose one transversality constraint, rather than two. The reason why we chose to formulate the problem in terms of $\mathbf{H}(\mathbf{r})$ and not $\mathbf{E}(\mathbf{r})$ is merely one of mathematical convenience, as will be discussed in the section Magnetic vs. Electric Fields. For now, we note that we can also find $\mathbf{H}$ from $\mathbf{E}$ via the first equation of (6):

$$\mathbf{H}(\mathbf{r}) = -\frac{i}{\omega\mu_0}\nabla \times \mathbf{E}(\mathbf{r}). \tag{9}$$

Electromagnetism as an Eigenvalue Problem

As discussed in the previous section, the heart of the Maxwell equations for a harmonic mode in a mixed dielectric medium is a differential equation for $\mathbf{H}(\mathbf{r})$, given by equation (7). The content of the equation is this: perform a series of operations on a function $\mathbf{H}(\mathbf{r})$, and if $\mathbf{H}(\mathbf{r})$ is really an allowable electromagnetic mode, the result will be a constant times the original function $\mathbf{H}(\mathbf{r})$. This situation arises often in mathematical physics, and is called an **eigenvalue problem**. If the result of an operation on a function is just the function itself, multiplied by some constant, then the function is called an **eigenfunction** or **eigenvector**[7] of that operator, and the multiplicative constant is called the **eigenvalue**.

In this case, we identify the left side of the master equation as an operator $\hat{\Theta}$ acting on $\mathbf{H}(\mathbf{r})$ to make it look more like a traditional eigenvalue problem:

$$\hat{\Theta}\mathbf{H}(\mathbf{r}) = \left(\frac{\omega}{c}\right)^2 \mathbf{H}(\mathbf{r}). \tag{10}$$

We have identified $\hat{\Theta}$ as the differential operator that takes the curl, then divides by $\varepsilon(\mathbf{r})$, and then takes the curl again:

$$\hat{\Theta}\mathbf{H}(\mathbf{r}) \triangleq \nabla \times \left(\frac{1}{\varepsilon(\mathbf{r})}\nabla \times \mathbf{H}(\mathbf{r})\right). \tag{11}$$

The eigenvectors $\mathbf{H}(\mathbf{r})$ are the spatial patterns of the harmonic modes, and the eigenvalues $(\omega/c)^2$ are proportional to the squared frequencies of those modes. An important thing to notice is that the operator $\hat{\Theta}$ is a **linear operator**. That is,

[7] Instead of *eigenvector*, physicists tend to stick *eigen* in front of any natural name for the solution. Hence, we also use terms like *eigenfield*, *eigenmode*, *eigenstate*, and so on.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

any linear combination of solutions is itself a solution; if $\mathbf{H}_1(\mathbf{r})$ and $\mathbf{H}_2(\mathbf{r})$ are both solutions of (10) with the same frequency ω, then so is $\alpha\mathbf{H}_1(\mathbf{r}) + \beta\mathbf{H}_2(\mathbf{r})$, where α and β are constants. For example, given a certain mode profile, we can construct another legitimate mode profile with the same frequency by simply doubling the field strength everywhere ($\alpha = 2$, $\beta = 0$). For this reason we consider two field patterns that differ only by an overall multiplier to be the same mode.

Our operator notation is reminiscent of quantum mechanics, in which we obtain an eigenvalue equation by operating on the wave function with the Hamiltonian. A reader familiar with quantum mechanics might recall some key properties of the eigenfunctions of the Hamiltonian: they have real eigenvalues, they are orthogonal, they can be obtained by a variational principle, and they may be catalogued by their symmetry properties (see, for example Shankar, 1982).

All of these same useful properties hold for our formulation of electromagnetism. In both cases, the properties rely on the fact that the main operator is a special type of linear operator known as a **Hermitian operator**. In the coming sections we will develop these properties one by one. We conclude this section by showing what it means for an operator to be Hermitian. First, in analogy with the inner product of two wave functions, we define the inner product of two vector fields $\mathbf{F}(\mathbf{r})$ and $\mathbf{G}(\mathbf{r})$ as

$$(\mathbf{F}, \mathbf{G}) \triangleq \int d^3\mathbf{r}\, \mathbf{F}^*(\mathbf{r}) \cdot \mathbf{G}(\mathbf{r}), \tag{12}$$

where "$*$" denotes complex conjugation. Note that a simple consequence of this definition is that $(\mathbf{F}, \mathbf{G}) = (\mathbf{G}, \mathbf{F})^*$ for any $\mathbf{F}$ and $\mathbf{G}$. Also note that $(\mathbf{F}, \mathbf{F})$ is always real and nonnegative, even if $\mathbf{F}$ itself is complex. In fact, if $\mathbf{F}(\mathbf{r})$ is a harmonic mode of our electromagnetic system, we can always set $(\mathbf{F}, \mathbf{F}) = 1$ by using our freedom to scale any mode by an overall multiplier.[8] Given $\mathbf{F}'(\mathbf{r})$ with $(\mathbf{F}', \mathbf{F}') \neq 1$, create

$$\mathbf{F}(\mathbf{r}) = \frac{\mathbf{F}'(\mathbf{r})}{\sqrt{(\mathbf{F}', \mathbf{F}')}}. \tag{13}$$

From our previous discussion, $\mathbf{F}(\mathbf{r})$ is really the same mode as $\mathbf{F}'(\mathbf{r})$, since it differs only by an overall multiplier, but now (as the reader can easily verify) we have $(\mathbf{F}, \mathbf{F}) = 1$. We say that $\mathbf{F}(\mathbf{r})$ has been **normalized**. Normalized modes are very useful in formal arguments. If, however, one is interested in the physical energy of the field and not just its spatial profile, the overall multiplier is important.[9]

Next, we say that an operator $\hat{\Xi}$ is **Hermitian** if $(\mathbf{F}, \hat{\Xi}\mathbf{G}) = (\hat{\Xi}\mathbf{F}, \mathbf{G})$ for any vector fields $\mathbf{F}(\mathbf{r})$ and $\mathbf{G}(\mathbf{r})$. That is, it does not matter which function is operated upon before taking the inner product. Clearly, not all operators are Hermitian. To show

[8] The trivial solution $\mathbf{F} = 0$ is not considered to be a proper eigenfunction.
[9] This distinction is discussed again after equation (24).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

that $\hat{\Theta}$ is Hermitian,[10] we perform an integration by parts[11] twice:

$$\begin{aligned}(\mathbf{F},\hat{\Theta}\mathbf{G}) &= \int d^3\mathbf{r}\,\mathbf{F}^*\cdot\nabla\times\left(\frac{1}{\varepsilon}\nabla\times\mathbf{G}\right)\\ &= \int d^3\mathbf{r}\,(\nabla\times\mathbf{F})^*\cdot\frac{1}{\varepsilon}\nabla\times\mathbf{G}\\ &= \int d^3\mathbf{r}\left[\nabla\times\left(\frac{1}{\varepsilon}\nabla\times\mathbf{F}\right)\right]^*\cdot\mathbf{G} = (\hat{\Theta}\mathbf{F},\mathbf{G}).\end{aligned} \tag{14}$$

In performing the integrations by parts, we neglected the surface terms that involve the values of the fields at the boundaries of integration. This is because in all cases of interest, one of two things will be true: either the fields decay to zero at large distances, or the fields are periodic in the region of integration. In either case, the surface terms vanish.

General Properties of the Harmonic Modes

Having established that $\hat{\Theta}$ is Hermitian, we can now show that the eigenvalues of $\hat{\Theta}$ must be real numbers. Suppose $\mathbf{H}(\mathbf{r})$ is an eigenvector of $\hat{\Theta}$ with eigenvalue $(\omega/c)^2$. Take the inner product of the master equation (7) with $\mathbf{H}(\mathbf{r})$:

$$\begin{aligned}&\hat{\Theta}\mathbf{H}(\mathbf{r}) = (\omega^2/c^2)\mathbf{H}(\mathbf{r})\\ \Longrightarrow\ &(\mathbf{H},\hat{\Theta}\mathbf{H}) = (\omega^2/c^2)(\mathbf{H},\mathbf{H})\\ \Longrightarrow\ &(\mathbf{H},\hat{\Theta}\mathbf{H})^* = (\omega^2/c^2)^*(\mathbf{H},\mathbf{H}).\end{aligned} \tag{15}$$

Because $\hat{\Theta}$ is Hermitian, we know that $(\mathbf{H},\hat{\Theta}\mathbf{H}) = (\hat{\Theta}\mathbf{H},\mathbf{H})$. Additionally, from the definition of the inner product we know that $(\mathbf{H},\hat{\Xi}\mathbf{H}) = (\hat{\Xi}\mathbf{H},\mathbf{H})^*$ for *any* operator $\hat{\Xi}$. Using these two pieces of information, we continue:

$$\begin{aligned}(\mathbf{H},\hat{\Theta}\mathbf{H})^* = (\omega^2/c^2)^*(\mathbf{H},\mathbf{H}) = (\hat{\Theta}\mathbf{H},\mathbf{H}) &= (\omega^2/c^2)(\mathbf{H},\mathbf{H})\\ \Longrightarrow\ (\omega^2/c^2)^* &= (\omega^2/c^2).\end{aligned} \tag{16}$$

It follows that $\omega^2 = (\omega^2)^*$, or that ω^2 is real. By a different argument, we can also show that ω^2 is always nonnegative for $\varepsilon > 0$. Set $\mathbf{F}=\mathbf{G}=\mathbf{H}$ in the middle equation

[10] The property that $\hat{\Theta}$ is Hermitian is closely related to the Lorentz reciprocity theorem, as described in the section Frequency-Domain Responses of appendix D.

[11] In particular, we use the vector identity that $\nabla\cdot(\mathbf{F}\times\mathbf{G}) = (\nabla\times\mathbf{F})\cdot\mathbf{G} - \mathbf{F}\cdot(\nabla\times\mathbf{G})$. Integrating both sides and applying the divergence theorem, we find that $\int\mathbf{F}\cdot(\nabla\times\mathbf{G}) = \int(\nabla\times\mathbf{F})\cdot\mathbf{G}$ plus a surface term, from the integral of $\nabla\cdot(\mathbf{F}\times\mathbf{G})$, that vanishes as described above.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

of (14), to obtain

$$(\mathbf{H},\mathbf{H})\left(\frac{\omega}{c}\right)^2 = (\mathbf{H},\hat{\Theta}\mathbf{H}) = \int d^3\mathbf{r}\,\frac{1}{\varepsilon}|\nabla\times\mathbf{H}|^2. \tag{17}$$

Since $\varepsilon(\mathbf{r}) > 0$ everywhere, the integrand on the right-hand side is everywhere nonnegative. The operator $\hat{\Theta}$ is said to be **positive semi-definite**. Therefore all of the eigenvalues ω^2 are nonnegative, and ω is real.

In addition, the Hermiticity of $\hat{\Theta}$ forces any two harmonic modes $\mathbf{H}_1(\mathbf{r})$ and $\mathbf{H}_2(\mathbf{r})$ with different frequencies ω_1 and ω_2 to have an inner product of zero. Consider two normalized modes, $\mathbf{H}_1(\mathbf{r})$ and $\mathbf{H}_2(\mathbf{r})$, with frequencies ω_1 and ω_2:

$$\begin{aligned}\omega_1^2(\mathbf{H}_2,\mathbf{H}_1) = c^2(\mathbf{H}_2,\hat{\Theta}\mathbf{H}_1) = c^2(\hat{\Theta}\mathbf{H}_2,\mathbf{H}_1) = \omega_2^2(\mathbf{H}_2,\mathbf{H}_1)\\ \Longrightarrow\ (\omega_1^2-\omega_2^2)(\mathbf{H}_2,\mathbf{H}_1) = 0.\end{aligned} \tag{18}$$

If $\omega_1 \neq \omega_2$, then we must have $(\mathbf{H}_1,\mathbf{H}_2)=0$ and we say $\mathbf{H}_1$ and $\mathbf{H}_2$ are **orthogonal** modes. If two harmonic modes have equal frequencies $\omega_1=\omega_2$, then we say they are **degenerate** and they are not necessarily orthogonal. For two modes to be degenerate requires what seems, on the face of it, to be an astonishing coincidence: two different field patterns happen to have precisely the same frequency. Usually there is a symmetry that is responsible for the "coincidence". For example, if the dielectric configuration is invariant under a 120° rotation, modes that differ only by a 120° rotation are expected to have the same frequency. Such modes are degenerate and are not necessarily orthogonal.

However, since $\hat{\Theta}$ is linear, any linear combination of these degenerate modes is itself a mode with that same frequency. As in quantum mechanics, we can always choose to work with linear combinations that *are* orthogonal (see, e.g., Merzbacher, 1961). This allows us to say quite generally that different modes are orthogonal, or can be arranged to be orthogonal.

The concept of orthogonality is most easily grasped by considering one-dimensional functions. What follows is a brief explanation (not mathematically rigorous, but perhaps useful to the intuition) that may help in understanding the significance of orthogonality. For two real one-dimensional functions $f(x)$ and $g(x)$ to be orthogonal means that

$$(f,g) = \int f(x)g(x)dx = 0. \tag{19}$$

In a sense, the product fg must be negative at least as much as it is positive over the interval of interest, so that the net integral vanishes. For example, the familiar set of functions $f_n(x) = \sin(n\pi x/L)$ are all orthogonal in the interval from $x = 0$ to $x = L$. Note that each of these functions has a different number of nodes (locations where $f_n(x) = 0$, not including the end points). In particular, f_n has $n-1$ nodes. The product of any two different f_n is positive as often as it is negative, and the inner product vanishes.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

The extension to a higher number of dimensions is a bit unclear, because the integration is more complicated. But the notion that orthogonal modes of different frequency have different numbers of spatial nodes holds rather generally. In fact, a given harmonic mode will generally contain more nodes than lower-frequency modes. This is analogous to the statement that each vibrational mode of a string with fixed ends contains one more node than the one below it. This will be important for our discussion in chapter 5.

Electromagnetic Energy and the Variational Principle

Although the harmonic modes in a dielectric medium can be quite complicated, there is a simple way to understand some of their qualitative features. Roughly, a mode tends to concentrate its electric-field energy in regions of high dielectric constant, while remaining orthogonal to the modes below it in frequency. This useful but somewhat vague notion can be expressed precisely through the electromagnetic **variational theorem**, which is analogous to the variational principle of quantum mechanics. In particular, the smallest eigenvalue ω_0^2/c^2, and thus the lowest-frequency mode, corresponds to the field pattern that minimizes the functional:

$$U_f(\mathbf{H}) \triangleq \frac{(\mathbf{H}, \hat{\Theta}\mathbf{H})}{(\mathbf{H}, \mathbf{H})}. \tag{20}$$

That is, ω_0^2/c^2 is the minimum of $U_f(\mathbf{H})$ over all conceivable field patterns $\mathbf{H}$ (subject to the transversality constraint $\nabla \cdot \mathbf{H} = 0$). The functional U_f is sometimes called the **Rayleigh quotient**, and appears in a similar variational theorem for any Hermitian operator. We will refer to U_f as the electromagnetic "energy" functional, in order to stress the analogy with analogous variational theorems in quantum and classical mechanics that minimize a physical energy.

To verify the claim that U_f is minimized for the lowest-frequency mode, we consider how small variations in $\mathbf{H}(\mathbf{r})$ affect the energy functional. Suppose that we perturb the field $\mathbf{H}(\mathbf{r})$ by adding a small-amplitude function $\delta\mathbf{H}(\mathbf{r})$. What is the resulting small change δU_f in the energy functional? It should be zero if the energy functional is really at a minimum, just as the ordinary derivative of a function vanishes at an extremum. To find out, we evaluate the energy functional at $\mathbf{H} + \delta\mathbf{H}$ and at $\mathbf{H}$, and then compute the difference δU_f:

$$\begin{aligned} U_f(\mathbf{H} + \delta\mathbf{H}) &= \frac{(\mathbf{H} + \delta\mathbf{H}, \hat{\Theta}\mathbf{H} + \hat{\Theta}\delta\mathbf{H})}{(\mathbf{H} + \delta\mathbf{H}, \mathbf{H} + \delta\mathbf{H})} \\ U_f(\mathbf{H}) &= \frac{(\mathbf{H}, \hat{\Theta}\mathbf{H})}{(\mathbf{H}, \mathbf{H})} \\ \delta U_f(\mathbf{H}) &\triangleq U_f(\mathbf{H} + \delta\mathbf{H}) - U_f(\mathbf{H}) \end{aligned} \tag{21}$$

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Ignoring terms higher than first order in $\delta\mathbf{H}$, we can write δU_f in the form $\delta U_f \approx [(\delta\mathbf{H}, \mathbf{G}) + (\mathbf{G}, \delta\mathbf{H})]/2$, where $\mathbf{G}$ is given by

$$\mathbf{G}(\mathbf{H}) = \frac{2}{(\mathbf{H},\mathbf{H})}\left(\hat{\Theta}\mathbf{H} - \left[\frac{(\mathbf{H}, \hat{\Theta}\mathbf{H})}{(\mathbf{H},\mathbf{H})}\right]\mathbf{H}\right). \tag{22}$$

This $\mathbf{G}$ can be interpreted as the *gradient* (rate of change) of the functional U_f with respect to $\mathbf{H}$.[12] At an extremum, δU_f must vanish for *all* possible shifts $\delta\mathbf{H}$, which implies that $\mathbf{G} = 0$. This implies that the parenthesized quantity in (22) is zero, which is equivalent to saying that $\mathbf{H}$ is an eigenvector of $\hat{\Theta}$. Therefore, U_f is at an extremum if and only if $\mathbf{H}$ is a harmonic mode. More careful considerations show that the lowest-ω electromagnetic eigenmode $\mathbf{H}_0$ *minimizes* U_f. The next-lowest-ω eigenmode will minimize U_f within the subspace of functions that are orthogonal to $\mathbf{H}_0$, and so on.

In addition to providing a useful characterization of the modes of $\hat{\Theta}$, the variational theorem is also the source of the heuristic rules about modes that were alluded to earlier in this section. This is most easily seen after rewriting the energy functional in terms of $\mathbf{E}$. Beginning with an eigenmode $\mathbf{H}$ that minimizes U_f, we rewrite the numerator of (20) using (11), (8), and (9), and we rewrite the denominator using (17) and (8). The result is:

$$\begin{aligned} U_f(\mathbf{H}) &= \frac{(\nabla\times\mathbf{E}, \nabla\times\mathbf{E})}{(\mathbf{E}, \varepsilon(\mathbf{r})\mathbf{E})} \\ &= \frac{\int d^3\mathbf{r}\,|\nabla\times\mathbf{E}(\mathbf{r})|^2}{\int d^3\mathbf{r}\,\varepsilon(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2}. \end{aligned} \tag{23}$$

From this expression, we can see that the way to minimize U_f is to concentrate the electric field $\mathbf{E}$ in regions of high dielectric constant ε (thereby maximizing the denominator) and to minimize the amount of spatial oscillations (thereby minimizing the numerator) while remaining orthogonal to lower-frequency modes.[13] Although we derived (23) by starting with an eigenmode $\mathbf{H}$ and rewriting the (minimized) energy functional in terms of $\mathbf{E}$, it can be shown (using the $\mathbf{E}$ eigenproblem of the next section) that (23) is also a valid variational theorem: the lowest-frequency eigenmode is given by the $\mathbf{E}$ field that minimizes (23), subject to $\nabla\cdot\varepsilon\mathbf{E} = 0$.

The energy functional must be distinguished from the *physical* **energy** stored in the electromagnetic field. The *time-averaged* physical energy can be separated

[12] The analogous and perhaps more familiar expression for functions $f(\mathbf{x})$ of a real vector $\mathbf{x}$ is $\delta f \approx \delta\mathbf{x}\cdot\nabla f = [\delta\mathbf{x}\cdot\nabla f + \nabla f\cdot\delta\mathbf{x}]/2$, in terms of the gradient ∇f. This is the first-order change in f when $\mathbf{x}$ is perturbed by a small amount $\delta\mathbf{x}$.

[13] The analogous heuristic rule in quantum mechanics is to concentrate the wave function in regions of low potential energy, while minimizing the kinetic energy and remaining orthogonal to lower-energy eigenstates.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

into a contribution from the electric field, and a contribution from the magnetic field:

$$\begin{aligned} U_{\mathbf{E}} &\triangleq \frac{\varepsilon_0}{4}\int d^3\mathbf{r}\,\varepsilon(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2 \\ U_{\mathbf{H}} &\triangleq \frac{\mu_0}{4}\int d^3\mathbf{r}\,|\mathbf{H}(\mathbf{r})|^2. \end{aligned} \tag{24}$$

In a harmonic mode, the physical energy is periodically exchanged between the electric and magnetic fields, and one can show that $U_{\mathbf{E}} = U_{\mathbf{H}}$.[14] The physical energy and the energy functional are related, but there is an important difference. The energy functional has fields in both the numerator and the denominator, and is therefore independent of the field strength. The physical energy is proportional to the square of the field strength. In other words, multiplying the fields by a constant affects the physical energy but does not affect the energy functional. If we are interested in the physical energy, we must pay attention to the amplitudes of our modes, but if we are interested only in mode profiles, we might as well normalize our modes.

Finally, we should also mention the expression for the rate of energy *transport*, which is given by the **Poynting vector**, $\mathbf{S}$:

$$\mathbf{S} \triangleq \frac{1}{2}\,\mathrm{Re}\left[\mathbf{E}^* \times \mathbf{H}\right], \tag{25}$$

where Re denotes the real part. This is the time-average flux of electromagnetic energy in the direction of $\mathbf{S}$, per unit time and per unit area, for a time-harmonic field. We also sometimes refer to the component of $\mathbf{S}$ in a given direction as the light **intensity**. The ratio of the energy flux to the energy density defines the *velocity* of energy transport, a subject we return to in the section Bloch-Wave Propagation Velocity of chapter 3.[15]

Magnetic vs. Electric Fields

We digress here to answer a question that commonly arises at this stage: why work with the magnetic field instead of the electric field? In the previous sections, we reformulated the Maxwell equations as an eigenvalue equation for the harmonic modes of the magnetic field $\mathbf{H}(\mathbf{r})$. The idea was that for a given frequency, we could solve for $\mathbf{H}(\mathbf{r})$ and then determine the $\mathbf{E}(\mathbf{r})$ via equation (8). But we could

[14] This can be shown from equations (8) and (9) combined with the fact that $\nabla\times$ is a Hermitian operator (see footnote 11 on page 12). Thus, $(\mu_0\mathbf{H},\mathbf{H}) = (\mu_0\mathbf{H}, -\frac{i}{\omega\mu_0}\nabla\times\mathbf{E}) = (+\varepsilon_0\varepsilon\frac{i}{\omega\varepsilon_0\varepsilon}\nabla\times\mathbf{H},\mathbf{E}) = (\varepsilon_0\varepsilon\mathbf{E},\mathbf{E})$.

[15] These equations for energy density and flux are derived in, for example, Jackson (1998) from the principle of conservation of energy. Note that the energy equations change in the presence of nonnegligible material dispersion.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

have equally well tried the alternate approach: solve for the *electric* field in (6) and then determine the magnetic field with (9). Why didn't we choose this route?

By pursuing this alternate approach, one finds the condition on the electric field to be

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) = \left(\frac{\omega}{c}\right)^2 \varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r}). \tag{26}$$

Because there are operators on both sides of this equation, it is referred to as a *generalized* eigenproblem. It is a simple matter to convert this into an ordinary eigenproblem by dividing (26) by ε, but then the operator is no longer Hermitian. If we stick to the generalized eigenproblem, however, then simple theorems analogous to those of the previous section can be developed because the two operators of the generalized eigenproblem, $\nabla \times \nabla \times$ and $\varepsilon(\mathbf{r})$, are easily shown to be both Hermitian and positive semi-definite.[16] In particular, it can be shown that ω is real, and that two solutions $\mathbf{E}_1$ and $\mathbf{E}_2$ with different frequencies satisfy an orthogonality relation: $(\mathbf{E}_1, \varepsilon\mathbf{E}_2) = 0$.

For some analytical calculations, such as the derivation of the variational equation (23) or the perturbation theory discussed in the next section, the **E** eigenproblem is the most convenient starting point. However, it has one feature that turns out to be undesirable for numerical computation: the transversality constraint $\nabla \cdot \varepsilon\mathbf{E} = 0$ depends on ε.

We can restore a simpler transversality constraint by using **D** instead of **E**, since $\nabla \cdot \mathbf{D} = 0$. Substituting $\mathbf{D}/\varepsilon_0\varepsilon$ for **E** in (26) and dividing both sides by ε (to keep the operator Hermitian) yields

$$\frac{1}{\varepsilon(\mathbf{r})}\nabla \times \nabla \times \frac{1}{\varepsilon(\mathbf{r})}\mathbf{D}(\mathbf{r}) = \left(\frac{\omega}{c}\right)^2 \frac{1}{\varepsilon(\mathbf{r})}\mathbf{D}(\mathbf{r}). \tag{27}$$

This is a perfectly valid formulation of the problem, but it seems unnecessarily complicated because of the three factors of $1/\varepsilon$ (as opposed to the single factor in the **H** or **E** formulations). For these reasons of mathematical convenience, we tend to prefer the **H** form for numerical calculations.[17]

The Effect of Small Perturbations

A perfectly linear and lossless material is a very useful idealization, and many real materials are excellent approximations of this idealization. But of course no material is perfectly linear and transparent. We can enlarge the scope of our formalism

[16] The $\varepsilon(\mathbf{r})$ operator on the right-hand side is actually **positive definite**: $(\mathbf{E}, \varepsilon\mathbf{E})$ is strictly positive for any nonzero **E**. This is necessary for the generalized eigenproblem to be well behaved.

[17] If a relative permeability $\mu \neq 1$ is included, the **E** and **H** eigenproblems take on similar forms. In that case, $\nabla \times \frac{1}{\varepsilon}\nabla \times \mathbf{H} = \left(\frac{\omega}{c}\right)^2 \mu\mathbf{H}$ with $\nabla \cdot \mu\mathbf{H} = 0$, compared to $\nabla \times \frac{1}{\mu}\nabla \times \mathbf{E} = \left(\frac{\omega}{c}\right)^2 \varepsilon\mathbf{E}$ with $\nabla \cdot \varepsilon\mathbf{E} = 0$. See, e.g., Sigalas et al. (1997) and Drikis et al. (2004).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

considerably by allowing for small nonlinearities and material absorption, using the well-developed **perturbation theory** for linear Hermitian eigenproblems. More generally, we may be interested in many types of small deviations from an initial problem. The idea is to begin with the harmonic modes of the idealized problem, and use analytical tools to approximately evaluate the effect of small changes in the dielectric function on the modes and their frequencies. For many realistic problems, the error in this approximation is negligible.

The derivation of perturbation theory for a Hermitian eigenproblem is straightforward and is covered in many texts on quantum mechanics, such as Sakurai (1994). Suppose a Hermitian operator $\hat{O}$ is altered by a small amount $\Delta\hat{O}$. The resulting eigenvalues and eigenvectors of the perturbed operator can be written as series expansions, in terms that depend on increasing powers of the perturbation strength $\Delta\hat{O}$. The resulting equation can be solved order-by-order using only the eigenmodes of the unperturbed operator.

Since we are interested in changes $\Delta\varepsilon(\mathbf{r})$, the combination of $\varepsilon(\mathbf{r})$ with curls in equation (7) is inconvenient, and it turns out to be easier to work with equation (26). By applying the perturbation procedure to equation (26), we obtain a simple formula for the frequency shift $\Delta\omega$ that results from a small perturbation $\Delta\varepsilon$ of the dielectric function:

$$\Delta\omega = -\frac{\omega}{2}\frac{\int d^3\mathbf{r}\,\Delta\varepsilon(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2}{\int d^3\mathbf{r}\,\varepsilon(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2} + O(\Delta\varepsilon^2). \tag{28}$$

In this equation, ω and $\mathbf{E}$ are the frequency and the mode profile for the perfectly linear and lossless (unperturbed) dielectric function ε. The error in this approximation is proportional to the square of $\Delta\varepsilon$ and can be neglected in many practical cases, for which $|\Delta\varepsilon|/\varepsilon$ is 1% or smaller.

Although we refer the reader to other texts for a rigorous derivation of equation (28), we point out an intuitive interpretation. Consider the case of a material with a refractive index $n = \sqrt{\varepsilon}$, in which the index is perturbed in some regions by an amount Δn. The volume integral in the numerator of equation (28) has nonzero contributions only from the perturbed regions. Writing $\Delta\varepsilon \approx \varepsilon \cdot 2\Delta n/n$, and supposing that $\Delta n/n$ is the same in all the perturbed regions (and can therefore be brought outside the integral), we obtain

$$\frac{\Delta\omega}{\omega} \approx -\frac{\Delta n}{n} \cdot (\text{fraction of } \int \varepsilon|\mathbf{E}|^2 \text{ in the perturbed regions }). \tag{29}$$

We see that the fractional change in frequency is equal to the fractional change in index multiplied by the fraction of the electric-field energy inside the perturbed regions. The minus sign appears because an increase in the refractive index lowers the mode frequencies, as can be understood from the variational equation (23).

A small absorption loss can be represented by a small imaginary part of the dielectric function. This does not present any obstacle to the perturbation theory, which requires only that the *unperturbed* problem be Hermitian; the perturbation

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

can be non-Hermitian. Thus, a small imaginary $\Delta\varepsilon = i\delta$ leads to a small imaginary $\Delta\omega = -i\gamma/2$, where $\gamma = \omega \int |\mathbf{E}|^2\delta / \int \varepsilon|\mathbf{E}|^2$. This corresponds to a field that is exponentially decaying in time as $e^{-\gamma t/2}$. The factor γ is the decay rate for the energy of the mode. It is also possible to consider a **gain** medium, in which an external energy source is pumping atomic or molecular excited states, by reversing the sign of the imaginary $\Delta\varepsilon$. The corresponding modes experience exponential growth, although in any real system the growth must eventually halt at some finite value.

If a material is only weakly nonlinear, then there will be a small shift $\Delta\varepsilon$ in the dielectric function that is proportional to either the amplitude of the field or its intensity (depending on the material). Perturbation theory is nearly exact for many problems with optical nonlinearities because the maximum changes in the refractive index are typically much less than 1%. Despite this small perturbation strength, the consequences can be profound and fascinating if the perturbations are allowed to accumulate for a long time. A full appraisal of the riches of nonlinear systems is generally beyond the scope of this book, although we will examine restricted examples in chapters 9 and 10.

The formula (28) is applicable to a wide range of possible perturbations. Some of the most interesting of these are time-variable external perturbations, such as are imposed by an external electromagnetic field or the variation of dielectric constant with temperature. However, we warn that there are cases in which the formula is not applicable. For example, a small displacement of the *boundary* between two materials certainly counts as a small perturbation of the system, but if the materials have highly dissimilar dielectric constants ε_1 and ε_2, then the moving discontinuity in the dielectric function renders equation (28) invalid. In this case, if a block of the ε_1-material is moved towards the ε_2-material by a distance Δh (perpendicular to the boundary), the correct expression for the frequency shift involves a surface integral over the interface (Johnson et al., 2002*a*):

$$\Delta\omega = -\frac{\omega}{2}\frac{\iint d^2\mathbf{r}\left[(\varepsilon_1-\varepsilon_2)|\mathbf{E}_\parallel(\mathbf{r})|^2 - \left(\frac{1}{\varepsilon_1}-\frac{1}{\varepsilon_2}\right)|\varepsilon\mathbf{E}_\perp(\mathbf{r})|^2\right]\Delta h}{\int d^3\mathbf{r}\,\varepsilon(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2} + O(\Delta h^2). \qquad (30)$$

In this expression, $\mathbf{E}_\parallel$ is the component of $\mathbf{E}$ that is parallel to the surface, and $\varepsilon\mathbf{E}_\perp$ is the component of $\varepsilon\mathbf{E}$ that is perpendicular to the surface. (Both of these components are guaranteed to be continuous across a dielectric interface.) This expression assumes that Δh is small compared to the transverse extent of the shifted portion of the material. If instead the surface-parallel extent of the shifted material is comparable to Δh or smaller (so that the perturbation is more like a "bump" than a shifted interface), then a more complicated correction is needed (Johnson et al., 2005).

The preceding example is one of several new developments in perturbation theory that can be found in the literature. New twists on the classic perturbative approaches have been required to deal correctly with the high material contrasts and strong periodicities that characterize photonic crystals.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Scaling Properties of the Maxwell Equations

One interesting feature of electromagnetism in dielectric media is that there is no fundamental length scale other than the assumption that the system is macroscopic. In atomic physics, the spatial scale of the potential function is generally set by the fundamental length scale of the Bohr radius. Consequently, configurations of material that differ only in their overall spatial scale nevertheless have very different physical properties. For photonic crystals, there is no fundamental constant with the dimensions of length—the master equation is *scale invariant*. This leads to simple relationships between electromagnetic problems that differ only by a contraction or expansion of all distances.

Suppose, for example, we have an electromagnetic eigenmode $\mathbf{H}(\mathbf{r})$ of frequency ω in a dielectric configuration $\varepsilon(\mathbf{r})$. We recall the master equation (7):

$$\nabla \times \left(\frac{1}{\varepsilon(\mathbf{r})} \nabla \times \mathbf{H}(\mathbf{r}) \right) = \left(\frac{\omega}{c} \right)^2 \mathbf{H}(\mathbf{r}). \tag{31}$$

Now suppose we are curious about the harmonic modes in a configuration of dielectric $\varepsilon'(\mathbf{r})$ that is just a compressed or expanded version of $\varepsilon(\mathbf{r})$: $\varepsilon'(\mathbf{r}) = \varepsilon(\mathbf{r}/s)$ for some scale parameter s. We make a change of variables in (31), using $\mathbf{r}' = s\mathbf{r}$ and $\nabla' = \nabla/s$:

$$s\nabla' \times \left(\frac{1}{\varepsilon(\mathbf{r}'/s)} s\nabla' \times \mathbf{H}(\mathbf{r}'/s) \right) = \left(\frac{\omega}{c} \right)^2 \mathbf{H}(\mathbf{r}'/s). \tag{32}$$

But $\varepsilon(\mathbf{r}'/s)$ is none other than $\varepsilon'(\mathbf{r}')$. Dividing out the s's shows that

$$\nabla' \times \left(\frac{1}{\varepsilon'(\mathbf{r}')} \nabla' \times \mathbf{H}(\mathbf{r}'/s) \right) = \left(\frac{\omega}{cs} \right)^2 \mathbf{H}(\mathbf{r}'/s). \tag{33}$$

This is just the master equation again, this time with mode profile $\mathbf{H}'(\mathbf{r}') = \mathbf{H}(\mathbf{r}'/s)$ and frequency $\omega' = \omega/s$. What this means is that the new mode profile and its corresponding frequency can be obtained by simply rescaling the old mode profile and its frequency. The solution of the problem at one length scale determines the solutions at all other length scales.

This simple fact is of considerable practical importance. For example, the microfabrication of complex micron-scale photonic crystals can be quite difficult. But models can be easily made and tested in the microwave regime, at the much larger length scale of centimeters, if materials can be found that have nearly the same dielectric constant. The considerations in this section guarantee that the model will have the same electromagnetic properties.

Just as there is no fundamental *length* scale, there is also no fundamental value of the dielectric constant. Suppose we know the harmonic modes of a system with dielectric configuration $\varepsilon(\mathbf{r})$, and we are curious about the modes of a system

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

with a dielectric configuration that differs by a constant factor everywhere, so that $\varepsilon'(\mathbf{r}) = \varepsilon(\mathbf{r})/s^2$. Substituting $s^2\varepsilon'(\mathbf{r})$ for $\varepsilon(\mathbf{r})$ in (31) yields

$$\nabla \times \left(\frac{1}{\varepsilon'(\mathbf{r})} \nabla \times \mathbf{H}(\mathbf{r}) \right) = \left(\frac{s\omega}{c} \right)^2 \mathbf{H}(\mathbf{r}). \tag{34}$$

The harmonic modes $\mathbf{H}(\mathbf{r})$ of the new system are unchanged,[18] but the frequencies are all scaled by a factor s: $\omega \to \omega' = s\omega$. If we multiply the dielectric constant everywhere by a factor of $1/4$, the mode patterns are unchanged but their frequencies double.

Combining the above two relations, we see that if we scale ε by s^2 and *also* rescale the coordinates by s, the frequency ω is unchanged. This simple scaling invariance is a special case of more general coordinate transformations. Amazingly, it turns out that *any* coordinate transformation can be replaced simply by a change of ε and μ while keeping ω fixed (Ward and Pendry, 1996). This can be a powerful conceptual tool, because it allows one to warp and distort a structure in complicated ways while retaining a similar form for the Maxwell equations. In general, however, this change in ε and μ is not merely a multiplication by a constant, as it is here.

Discrete vs. Continuous Frequency Ranges

The **spectrum** of a photonic crystal is the totality of all of the eigenvalues ω. What does this spectrum look like? Is it a continuous range of values, like a rainbow, or do the frequencies form a discrete sequence $\omega_0, \omega_1, \ldots$, like the vibration frequencies of a piano string? The next chapter will feature some specific examples of spectra, but in this section we discuss this question in general terms.

The answer depends on the *spatial domain* of the mode profiles $\mathbf{H}(\mathbf{r})$ (or $\mathbf{E}$). If the fields are spatially bounded, either because they are localized around a particular point or because they are periodic in all three dimensions (and therefore represent a bounded profile repeated indefinitely), then the frequencies ω form a discrete set. Otherwise they can form a single continuous range, a set of continuous ranges, or a combination of continuous ranges and discrete sets (for a combination of localized and extended modes).

This property is quite general for many Hermitian eigenproblems. We will argue below that it follows from the orthogonality of the modes. A host of seemingly unrelated physical phenomena can be attributed to this abstract mathematical result: from the discrete energy levels in the spectrum of hydrogen gas (in which the electron wave functions are localized around the nucleus) to the distinct overtones of an organ pipe (in which the vibrating modes dwell within a finite length). Other cases that are familiar to physics students are the

[18] Note, however, that the relationship between $\mathbf{E}$ and $\mathbf{H}$ *has* changed by a factor of s, from equation (8).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

quantum-mechanical problem of a particle in a box (as in Liboff, 1992) and the electromagnetic problem of microwaves in a metallic cavity (as in Jackson, 1998). We will see in the chapters to come that this result, applied to photonic crystals, leads to the concepts of discrete frequency bands and of localized modes near crystal defects.

An intuitive explanation for the relation between the bounded spatial domain of the eigenmodes and the discreteness of the frequency spectrum is as follows.[19] Suppose that we have a continuous range of eigenvalues, so that we can vary the frequency ω continuously and get some eigenmode $\mathbf{H}_\omega(\mathbf{r})$ for each ω. We now argue that this **continuum** cannot be the spectrum of spatially bounded modes. It is reasonable to suppose that, as we change ω continuously, the field $\mathbf{H}_\omega$ can be made to change continuously as well, so that for an arbitrarily small change $\delta\omega$ there is a correspondingly small change $\delta\mathbf{H}$. Any drastic difference in the fields would correspond to a very different value of the electromagnetic energy functional and hence of the frequency. (An exception is made for systems with spatial symmetries that produce degeneracies, as discussed in the next chapter, but a similar argument implies that a bounded system has at most a finite number of degenerate modes with a given eigenvalue.) On the other hand, two spatially bounded modes $\mathbf{H}$ and $\mathbf{H} + \delta\mathbf{H}$ that are arbitrarily similar cannot also be orthogonal: their inner product is $(\mathbf{H}, \mathbf{H}) + (\mathbf{H}, \delta\mathbf{H})$, where the first term is positive and the second term is arbitrarily small for integration over a finite domain, i.e. a system with spatially bounded modes. Thus, the continuous spectrum is incompatible with the required orthogonality of the modes, unless the modes are of unbounded spatial extent.

We will see in the next chapter that many interesting electromagnetic systems exhibit *both* discrete localized modes and a continuum of extended states. This is not too different from the case of a hydrogen atom, which has both bound electron states with discrete energy levels and also a continuum of freely propagating states for electrons with a kinetic energy greater than the ionization energy.

Electrodynamics and Quantum Mechanics Compared

As a compact summary of the topics in this chapter, and for the benefit of those readers familiar with quantum mechanics, we now present some similarities between our formulation of electrodynamics in dielectric media and the quantum mechanics of noninteracting electrons (see table 1). This analogy is developed further in appendix A.

In both cases, we decompose the fields into harmonic modes that oscillate with a phase factor $e^{-i\omega t}$. In quantum mechanics, the wave function is a complex scalar field. In electrodynamics, the magnetic field is a real vector field and the complex exponential is just a mathematical convenience.

[19] For a more formal discussion, see e.g. Courant and Hilbert (1953, chap. 6).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Table 1

	Quantum Mechanics	*Electrodynamics*
Field	$\Psi(\mathbf{r},t) = \Psi(\mathbf{r})e^{-iEt/\hbar}$	$\mathbf{H}(\mathbf{r},t) = \mathbf{H}(\mathbf{r})e^{-i\omega t}$
Eigenvalue problem	$\hat{H}\Psi = E\Psi$	$\hat{\Theta}\mathbf{H} = \left(\frac{\omega}{c}\right)^2 \mathbf{H}$
Hermitian operator	$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})$	$\hat{\Theta} = \nabla \times \frac{1}{\varepsilon(\mathbf{r})}\nabla \times$

Comparison of quantum mechanics and electrodynamics.

In both cases, the modes of the system are determined by a Hermitian eigenvalue equation. In quantum mechanics, the frequency ω is related to the eigenvalue via $E = \hbar\omega$, which is meaningful only up to an overall additive constant V_0.[20] In electrodynamics, the eigenvalue is proportional to the square of the frequency, and there is no arbitrary additive constant.

One difference we did not discuss, but is apparent from Table 1, is that in quantum mechanics, the Hamiltonian is **separable** if $V(\mathbf{r})$ is separable. For example, if $V(\mathbf{r})$ is the sum of one-dimensional functions $V_x(x) + V_y(y) + V_z(z)$, then we can write Ψ as a product $\Psi(\mathbf{r}) = X(x)Y(y)Z(z)$ and the problem separates into three more manageable problems, one for each direction. In electrodynamics, such a factorization is not generally possible: the differential operator, $\hat{\Theta}$, couples the different coordinates even if $\varepsilon(\mathbf{r})$ is separable. This makes analytical solutions rare, and generally confined to very simple systems.[21] To demonstrate most of the interesting phenomena associated with photonic crystals, we will usually make use of numerical solutions.

In quantum mechanics, the lowest eigenstates typically have the amplitude of the wave function concentrated in regions of low potential, while in electrodynamics the lowest modes have their electric-field energy concentrated in regions of high dielectric constant. Both of these statements are made quantitative by a variational theorem.

Finally, in quantum mechanics, there is usually a fundamental length scale that prevents us from relating solutions to potentials that differ by a scale factor. Electrodynamics is free from such a length scale, and the solutions we obtain are easily scaled up or down in length scale and frequency.

[20] Here $\hbar \triangleq h/2\pi$ is given by Planck's constant h, a fundamental constant with an approximate value $h \approx 6.626 \times 10^{-34}$ J sec.

[21] It is possible to achieve a similar separation of the Maxwell equations in two dimensions, or in systems with cylindrical symmetry, but even in these cases the separation is usually achieved only for a particular polarization (Chen, 1981; Kawakami, 2002; Watts et al., 2002). In these special cases, the Maxwell equations can be written in a Schrödinger-like form. [The separable cases of the Schrödinger equation were enumerated by Eisenhart (1948).] On the other hand, if ε does not depend on a particular coordinate, then that particular dimension of the problem is always separable, as we will see in the section Continuous Translational Symmetry of chapter 3.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Further Reading

A particularly lucid undergraduate text on electromagnetism is Griffiths (1989). A more advanced and complete treatment of the macroscopic Maxwell equations, including a derivation from their microscopic counterparts, is contained in Jackson (1998). To explore the analogy between our formalism and the Schrödinger equation of quantum mechanics, consult the first few chapters of any introductory quantum mechanics text. In particular, Shankar (1982), Liboff (1992), and Sakurai (1994) develop the properties of the eigenstates of a Hermitian operator with proofs very similar to our own. The first two are undergraduate texts; the third is at the graduate level. A more formal mathematical approach to the subject of Hermitian operators leads to the field of *functional analysis*, as introduced in, e.g., Gohberg et al. (2000).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

3

Symmetries and Solid-State Electromagnetism

IF A DIELECTRIC STRUCTURE has a certain symmetry, then the symmetry offers a convenient way to categorize the electromagnetic modes of that system. In this chapter, we will investigate what various symmetries of a system can tell us about its electromagnetic modes. Translational symmetries (both discrete and continuous) are important because photonic *crystals* are *periodic* dielectrics, and because they provide a natural setting for the discussion of band gaps. Some of the terminology of solid-state physics is appropriate, and will be introduced. We will also investigate rotational, mirror, inversion, and time-reversal symmetries.

Using Symmetries to Classify Electromagnetic Modes

In both classical mechanics and quantum mechanics, we learn the lesson that the *symmetries* of a system allow one to make general statements about that system's behavior. Because of the mathematical analogy we pursued in the last chapter, it is not too surprising that careful attention to symmetry also helps to understand the properties of electromagnetic systems. We will begin with a concrete example of a symmetry and the conclusion we may draw from it, and will then pass on to a more formal discussion of symmetries in electromagnetism.

Suppose we want to find the modes that are allowed in the two-dimensional metal cavity shown in figure 1. Its shape is somewhat arbitrary, which would make it difficult to write down the exact boundary condition and solve the problem analytically. But the cavity has an important symmetry: if you invert the cavity about its center, you end up with exactly the same cavity shape. So if, somehow, we find that the particular pattern $\mathbf{H}(\mathbf{r})$ is a mode with frequency ω, then the pattern $\mathbf{H}(-\mathbf{r})$ must also be a mode with frequency ω. The cavity cannot distinguish between these two modes, since it cannot tell $\mathbf{r}$ from $-\mathbf{r}$.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

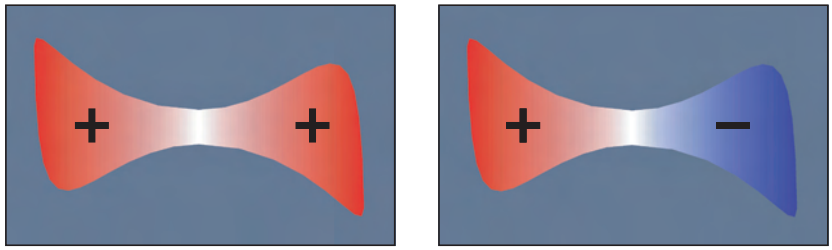


Figure 1: A two-dimensional metallic cavity with inversion symmetry. Red and blue suggest positive and negative fields. On the left, an **even** mode occupies the cavity, for which $\mathbf{H}(\mathbf{r}) = \mathbf{H}(-\mathbf{r})$. On the right, an **odd** mode occupies the cavity, for which $\mathbf{H}(\mathbf{r}) = -\mathbf{H}(-\mathbf{r})$.

Recall from chapter 2 that different modes with the same frequency are said to be **degenerate**. Unless $\mathbf{H}(\mathbf{r})$ is a member of a degenerate family of modes, then if $\mathbf{H}(-\mathbf{r})$ has the same frequency it must be the *same mode*. It must be nothing more than a multiple of $\mathbf{H}(\mathbf{r})$: $\mathbf{H}(-\mathbf{r}) = \alpha\mathbf{H}(\mathbf{r})$. But what is α? If we invert the system *twice*, picking up another factor of α, then we return to the original function $\mathbf{H}(\mathbf{r})$. Therefore $\alpha^2\mathbf{H}(\mathbf{r}) = \mathbf{H}(\mathbf{r})$, and we see that $\alpha = 1$ or -1. A given nondegenerate mode must be one of two types: either it is invariant under inversion, $\mathbf{H}(-\mathbf{r}) = \mathbf{H}(\mathbf{r})$, and we call it **even**; or, it becomes its own opposite, $-\mathbf{H}(-\mathbf{r}) = \mathbf{H}(\mathbf{r})$, and we call it **odd**.[1] These possibilities are depicted in figure 1. We have classified the modes of the system based on how they respond to one of its symmetry operations.

With this example in mind, we can capture the essential idea in more abstract language. Suppose I is an operator (a 3×3 matrix) that inverts vectors (3×1 matrices), so that $I\mathbf{a} = -\mathbf{a}$. To invert a vector *field*, we use an operator $\hat{O}_I$ that inverts both the vector $\mathbf{f}$ and its argument $\mathbf{r}$: $\hat{O}_I\mathbf{f}(\mathbf{r}) = I\mathbf{f}(I\mathbf{r})$.[2] What is the mathematical expression of the statement that our system has inversion symmetry? Since inversion is a symmetry of our system, it does not matter whether we operate with $\hat{\Theta}$ or we first invert the coordinates, then operate with $\hat{\Theta}$, and then change them back:

$$\hat{\Theta} = \hat{O}_I^{-1}\hat{\Theta}\hat{O}_I. \tag{1}$$

[1] This dichotomy is not automatically true of degenerate modes. But we can always form new modes that *are* even or odd, by taking appropriate linear combinations of the degenerate modes.

[2] This is a special case of the operator defined later in equation (14). There is a minor complication here because $\mathbf{H}$ is a pseudovector and $\mathbf{E}$ is a vector, as proved in Jackson (1998). This means that $\mathbf{H}$ transforms with a plus sign ($I\mathbf{H} = +\mathbf{H}$), while $\mathbf{E}$ transforms with a minus sign ($I\mathbf{E} = -\mathbf{E}$). That is, $\hat{O}_I\mathbf{H}(\mathbf{r}) = +\mathbf{H}(-\mathbf{r})$, and $\hat{O}_I\mathbf{E}(\mathbf{r}) = -\mathbf{E}(-\mathbf{r})$. An even mode is defined as one that is invariant under the inversion $\hat{O}_I$, which means that an even mode has $\mathbf{H}(\mathbf{r}) = \mathbf{H}(-\mathbf{r})$ and $\mathbf{E}(\mathbf{r}) = -\mathbf{E}(-\mathbf{r})$. Similarly, an odd mode is defined as one that acquires a minus sign under the inversion $\hat{O}_I$, so that $\mathbf{H}(\mathbf{r}) = -\mathbf{H}(-\mathbf{r})$ and $\mathbf{E}(\mathbf{r}) = \mathbf{E}(-\mathbf{r})$.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

This equation can be rearranged as $\hat{O}_I\hat{\Theta} - \hat{\Theta}\hat{O}_I = 0$. Following this cue, we define the **commutator** $[\hat{A}, \hat{B}]$ of two operators $\hat{A}$ and $\hat{B}$ just like the commutator in quantum mechanics:

$$[\hat{A}, \hat{B}] \triangleq \hat{A}\hat{B} - \hat{B}\hat{A}. \tag{2}$$

Note that the commutator is itself an operator. We have shown that our system is symmetric under inversion only if the inversion operator commutes with $\hat{\Theta}$; that is, we must have $[\hat{O}_I, \hat{\Theta}] = 0$. If we now operate with this commutator on any mode of the system $\mathbf{H}(\mathbf{r})$, we obtain

$$\begin{aligned} [\hat{O}_I, \hat{\Theta}]\mathbf{H} &= \hat{O}_I(\hat{\Theta}\mathbf{H}) - \hat{\Theta}(\hat{O}_I\mathbf{H}) = 0 \\ \Longrightarrow\ \hat{\Theta}(\hat{O}_I\mathbf{H}) &= \hat{O}_I(\hat{\Theta}\mathbf{H}) = \frac{\omega^2}{c^2}(\hat{O}_I\mathbf{H}). \end{aligned} \tag{3}$$

This equation tells us that if $\mathbf{H}$ is a harmonic mode with frequency ω, then $\hat{O}_I\mathbf{H}$ is *also* a mode with frequency ω. If there is no degeneracy, then there can only be one mode per frequency, so $\mathbf{H}$ and $\hat{O}_I\mathbf{H}$ can be different only by a multiplicative factor: $\hat{O}_I\mathbf{H} = \alpha\mathbf{H}$. But this is just the eigenvalue equation for $\hat{O}_I$, and we already know that the eigenvalues α must be either 1 or -1. Thus, we can classify the eigenvectors $\mathbf{H}(\mathbf{r})$ according to whether they are even ($\mathbf{H} \to +\mathbf{H}$) or odd ($\mathbf{H} \to -\mathbf{H}$) under the inversion symmetry operation $\hat{O}_I$.

What if there *is* degeneracy in the system? Then two modes may have the same frequency, but might not be related by a simple multiplier. Although we will not demonstrate it, we can always form linear combinations of such degenerate modes to make modes that are themselves even or odd.

Generally speaking, whenever two operators commute, one can construct simultaneous eigenfunctions of both operators. One reason why this is convenient is that eigenfunctions and eigenvalues of simple symmetry operators like $\hat{O}_I$ are easily determined, whereas those for $\hat{\Theta}$ are not. But if $\hat{\Theta}$ commutes with a symmetry operator $\hat{S}$, we can construct and catalogue the eigenfunctions of $\hat{\Theta}$ using their $\hat{S}$ properties. In the case of inversion symmetry, we can classify the $\hat{\Theta}$ eigenfunctions as either odd or even. We will find this approach useful in later sections when we introduce translational, rotational, and mirror symmetries.

Continuous Translational Symmetry

Another symmetry that a system might have is continuous translation symmetry. Such a system is unchanged if we translate everything through the same distance in a certain direction. Given this information, we can determine the functional form of the system's modes.

A system with translational symmetry is unchanged by a translation through a displacement $\mathbf{d}$. For each $\mathbf{d}$, we can define a translation operator $\hat{T}_\mathbf{d}$ which, when

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

operating on a function $\mathbf{f}(\mathbf{r})$, shifts the argument by $\mathbf{d}$. Suppose our system is translationally invariant; then we have $\hat{T}_\mathbf{d}\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r}-\mathbf{d}) = \varepsilon(\mathbf{r})$, or equivalently, $[\hat{T}_\mathbf{d}, \hat{\Theta}] = 0$. The modes of $\hat{\Theta}$ can now be classified according to how they behave under $\hat{T}_\mathbf{d}$.

A system with **continuous** translation symmetry in the z direction is invariant under all of the $\hat{T}_\mathbf{d}$'s for that direction. What sort of function is an eigenfunction of all the $\hat{T}_\mathbf{d}$'s? We can prove that a mode with the functional form e^{ikz} is an eigenfunction of any translation operator in the z direction:

$$\hat{T}_{d\hat{\mathbf{z}}}e^{ikz} = e^{ik(z-d)} = (e^{-ikd})e^{ikz}. \tag{4}$$

The corresponding eigenvalue is e^{-ikd}. With a little more work, one can show the converse, too: any eigenfunction of $\hat{T}_\mathbf{d}$ for all $\mathbf{d} = d\hat{\mathbf{z}}$ must be proportional to e^{ikz} for some k.[3] The modes of our system can be chosen to be eigenfunctions of all the $\hat{T}_\mathbf{d}$'s, so we therefore know they should have a z dependence of the functional form e^{ikz} (the z dependence is *separable*). We can classify them by the particular values for k, the **wave vector**. (k must be a real number in an infinite system where we require the modes to have bounded amplitudes at infinity.)

A system that has continuous translational symmetry in *all three* directions is a **homogeneous medium**: $\varepsilon(\mathbf{r})$ is a constant ε (= 1 for free space). Following a line of argument similar to the one above, we can deduce that the modes must have the form

$$\mathbf{H}_\mathbf{k}(\mathbf{r}) = \mathbf{H}_0 e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{5}$$

where $\mathbf{H}_0$ is any constant vector. These are **plane waves**, polarized in the direction of $\mathbf{H}_0$. Imposing the transversality requirement—equation (5) of chapter 2—gives the further restriction $\mathbf{k}\cdot\mathbf{H}_0 = 0$. The reader can also verify that these plane waves are in fact solutions of the master equation with eigenvalues $(\omega/c)^2 = |\mathbf{k}|^2/\varepsilon$, yielding the **dispersion relation** $\omega = c|\mathbf{k}|/\sqrt{\varepsilon}$. We classify a plane wave by its wave vector $\mathbf{k}$, which specifies how the mode is transformed by a continuous translation operation.

Another simple system with continuous translational symmetry is an infinite plane of glass, as shown in figure 2. In this case, the dielectric function varies in the z direction, but not in the x or y directions: $\varepsilon(\mathbf{r}) = \varepsilon(z)$. The system is invariant under all of the translation operators of the xy plane. We can classify the modes according to their in-plane wave vectors, $\mathbf{k} = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}}$. The x and y dependence must once again be a complex exponential (a plane wave):

$$\mathbf{H}_\mathbf{k}(\mathbf{r}) = e^{i\mathbf{k}\cdot\boldsymbol{\rho}}\mathbf{h}(z). \tag{6}$$

[3] If $f(x) \neq 0$ is such an eigenfunction, then $f(x-d) = \lambda(d)f(x)$ for all d and some eigenvalues $\lambda(d)$. Scale $f(x)$ so that $f(0) = 1$ and thus $f(x) = f(0-[-x]) = \lambda(-x)$. Therefore, $f(x+y) = f(x)f(y)$, and the only anywhere-continuous functions with this property are $f(x) = e^{cx}$ for some constant c (see, e.g., Rudin, 1964, ch. 8 exercise 6).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

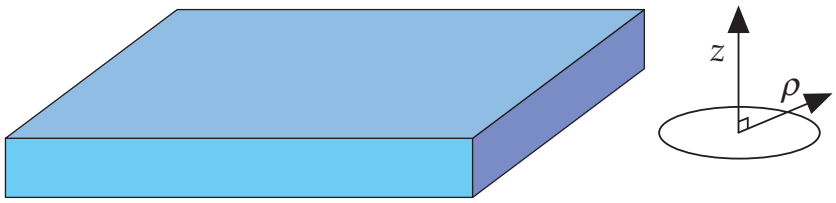


Figure 2: A plane of glass. If the glass extends much farther in the x and y directions than in the z direction, we may consider this system to be one-dimensional: the dielectric function $\varepsilon(\mathbf{r})$ varies in the z direction, but has no dependence on the in-plane coordinate $\boldsymbol{\rho}$.

This is the first of many occasions in which we will use the symbol $\boldsymbol{\rho}$ to denote a vector that is confined to the xy plane. The function $\mathbf{h}(z)$ (which depends on $\mathbf{k}$) cannot be determined by this line of reasoning, because the system does not have translational symmetry in that direction. (The transversality condition does imply one restriction on $\mathbf{h}$: substitution of (6) into $\nabla \cdot \mathbf{H}_\mathbf{k} = 0$ gives $\mathbf{k} \cdot \mathbf{h} = i\partial h_z/\partial z$.)

The reason why the modes are described by equation (6) can also be understood with an intuitive argument. Consider three non-collinear neighboring points at $\mathbf{r}, \mathbf{r} + d\mathbf{x}$, and $\mathbf{r} + d\mathbf{y}$, all of which have the same z value. Due to symmetry, these three points should be treated equally, and should have the same magnetic field amplitude. The only conceivable difference could be the variation in the phase between the points. But once we choose the phase differences between these three points, we set the phase relationships between *all* the points. We have effectively specified k_x and k_y at one point, but they must be universal to the plane. Otherwise we could distinguish different locations in the plane by their phase relationships. Along the z direction, however, this restriction does not hold. Each plane is at a different distance from the bottom of the glass structure and can conceivably have a different amplitude and phase.

We have seen that we can classify the modes by their values of $\mathbf{k}$. Although we cannot yet say anything about $\mathbf{h}(z)$, we can nevertheless line up the modes (whatever they may be) in order of increasing frequency for a given value of $\mathbf{k}$. Let n stand for a particular mode's place in line of increasing frequency, so that we can identify any mode by its unique name $(\mathbf{k}, n)$. If there is degeneracy, then we might have to include an additional index to name the degenerate modes that have the same n and $\mathbf{k}$.

We call n the **band number**. If the spectrum is discrete for a given $\mathbf{k}$, we can use integers for n, but sometimes the band number is actually a continuous variable. As the n value grows, so too does the frequency of the mode. If we make a plot of wave vector versus mode frequency for the plane of glass, the different bands correspond to different lines that rise uniformly in frequency. This **band structure** (also called a **band diagram** or **dispersion relation**) is shown in figure 3 and is considered in more detail below. We computed it by solving the master equation (7) of chapter 2 numerically.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

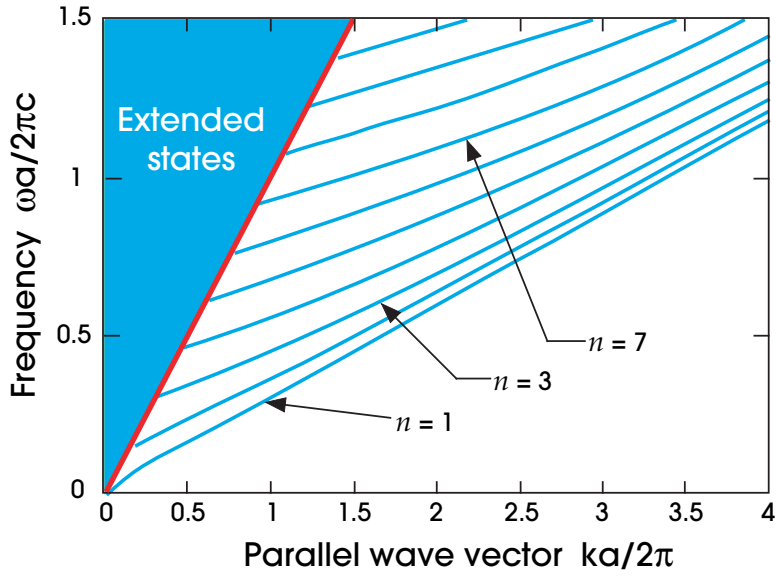


Figure 3: Harmonic mode frequencies for a plane of glass of thickness a and $\varepsilon = 11.4$. Blue lines correspond to modes that are localized in the glass. The shaded blue region is a continuum of states that extend into both the glass and the air around it. The red line is the **light line** $\omega = ck$. This plot shows modes of only one polarization, for which $\mathbf{H}$ is perpendicular to both the z and k directions.

Another way to state the significance of continuous translational symmetry is that the components of the wave vector $\mathbf{k}$ along the symmetry directions are **conserved quantities**. If a field pattern starts out with a particular eigenvalue e^{-ikd} of $\hat{T}_{\mathbf{d}}$ (which commutes with $\hat{\Theta}$), then it will have that eigenvalue at all future times. Such conservation laws have far-reaching consequences, as we will see in subsequent sections.[4]

Index guiding

Returning to the infinite plane of glass, we now discuss one of the most well-known phenomena in classical optics, **total internal reflection**. The familiar description of this phenomenon is that light rays within the glass that strike the interface with the air (or any lower-index medium) at too shallow an angle are totally reflected, and remain confined to the glass (forming a planar waveguide).

[4] For experts, we can state more generally that the irreducible representation of the symmetry group is conserved in a linear system. This is easily proved from the fact that the projection operator for the group representation (see, e.g. Inui et al., 1996) commutes with the time-evolution operator.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

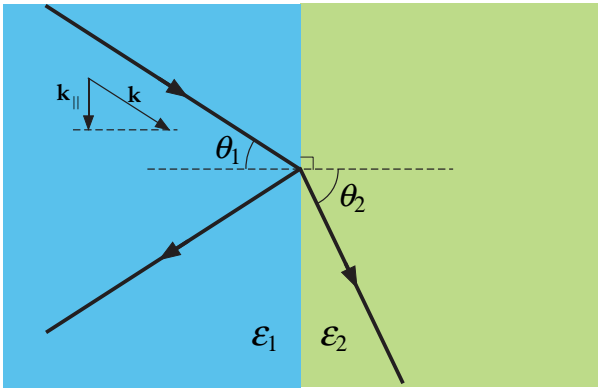


Figure 4: For a flat interface between two dielectrics ε_1 and ε_2, light can be described by a ray with an incident angle θ_1 and a refracted angle θ_2 given by Snell's law. When $\varepsilon_2 < \varepsilon_1$, we can have no solution θ_2 for certain θ_1, and the light undergoes total internal reflection. A generalization of this result follows from translational symmetry, which tells us that $k_\parallel$ is conserved.

In this section, we will describe this familiar phenomenon with the symmetry language developed earlier in this chapter. We will see that, in a sense, the confinement of light is a consequence of translational symmetry. Moreover, symmetry considerations lead to a more general concept of **index guiding** than the ray-optics picture might suggest, a result that forms the foundation of chapters 7 and 8.

The refraction of a light ray at an interface between two dielectrics ε_1 and ε_2, illustrated in figure 4, is usually described in terms of **Snell's law**: $n_1 \sin\theta_1 = n_2 \sin\theta_2$, where n_i is the refractive index $\sqrt{\varepsilon_i}$ and θ_i is the angle the ray makes with the normal to the interface. If $\theta_1 > \sin^{-1}(n_2/n_1)$, then the law would demand $\sin\theta_2 > 1$, for which there is no real solution; the interpretation is that the ray is totally reflected. The *critical angle* $\theta_c = \sin^{-1}(n_2/n_1)$ exists only for $n_2 < n_1$, so total internal reflection occurs only within the higher-index medium. Snell's law, however, is simply the combination of two conservation laws that follow from symmetry: conservation of frequency ω (from the linearity and time-invariance of the Maxwell equations), and conservation of the component $k_\parallel$ of $\mathbf{k}$ that is parallel to the interface (from the continuous translational symmetry along the interface, as we noted above). In particular, $k_\parallel = |\mathbf{k}| \sin\theta$, and $|\mathbf{k}| = n\omega/c$ from the dispersion relation. We obtain Snell's law by setting $k_\parallel$ equal on both sides of the interface. The advantage of this way of thinking about the problem is that we are now in a position to generalize beyond the ray-optics regime (which is valid only on length scales much larger than the wavelength of light).

Let us now be a little more concrete. Consider a plane of glass of width a centered about the origin. We now wish to understand the band structure of the electromagnetic modes, by which we mean the frequency ω versus the wave

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

vector $k_\parallel$ (both of which are conserved quantities). This plot, shown in figure 3, is described below.

First, consider the modes that are not confined to the glass, and extend into the air and out to infinity. Far away from the glass, these modes must closely resemble free-space plane waves. These are superpositions of plane waves with $\omega = c|\mathbf{k}| = c\sqrt{k_\parallel^2 + k_\perp^2}$ for some perpendicular real wave vector component $k_\perp$. For a given value of $k_\parallel$, there will be modes with every possible frequency greater than $ck_\parallel$, because $k_\perp$ can take any value. Thus the spectrum of states is continuous for all frequencies above the **light line** $\omega = ck_\parallel$, which is marked with a red line in figure 3. The region of the band structure with $\omega > ck_\parallel$ is called the **light cone**. The modes in the light cone are solutions of Snell's law (less than the critical angle).

In addition to the light cone, the glass plate introduces new electromagnetic solutions that lie *below* the light line. Because ε is larger in the glass than in air, these modes have *lower* frequencies relative to the values the corresponding modes would have in free space (as demanded by the variational theorem, equation (23) of chapter 2). These new solutions must be localized in the vicinity of the glass. Below the light line, the only solutions in air are those with *imaginary* $k_\perp = \pm i\sqrt{k_\parallel^2 - \omega^2/c^2}$, corresponding to fields that *decay exponentially* (are **evanescent**) away from the glass. We call these the **index-guided** modes, and from the section Discrete vs. Continuous Frequency Ranges of chapter 2 we expect that for a given $k_\parallel$ they form a set of discrete frequencies, because they are localized in one direction. Thus, we obtain the discrete bands $\omega_n(k_\parallel)$ below the light line in figure 3. In the limit of larger and larger $|k_\parallel|$, one obtains more and more guided bands, and eventually one approaches the ray-optics limit of totally internally reflected rays with a continuum of angles $\theta > \theta_c$.

Discrete Translational Symmetry

Photonic crystals, like traditional crystals of atoms or molecules, do not have continuous translational symmetry. Instead, they have *discrete* translational symmetry. That is, they are not invariant under translations of *any* distance, but rather, only distances that are a multiple of some fixed step length. The simplest example of such a system is a structure that is repetitive in one direction, like the configuration in figure 5.

For this system we still have continuous translational symmetry in the x direction, but now we have discrete translational symmetry in the y direction. The basic step length is the **lattice constant** a, and the basic step vector is called the **primitive lattice vector**, which in this case is $\mathbf{a} = a\hat{\mathbf{y}}$. Because of this discrete symmetry, $\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r} \pm \mathbf{a})$. By repeating this translation, we see that $\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r} + \mathbf{R})$ for any $\mathbf{R}$ that is an integral multiple of $\mathbf{a}$; that is, $\mathbf{R} = \ell\mathbf{a}$, where ℓ is an integer.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

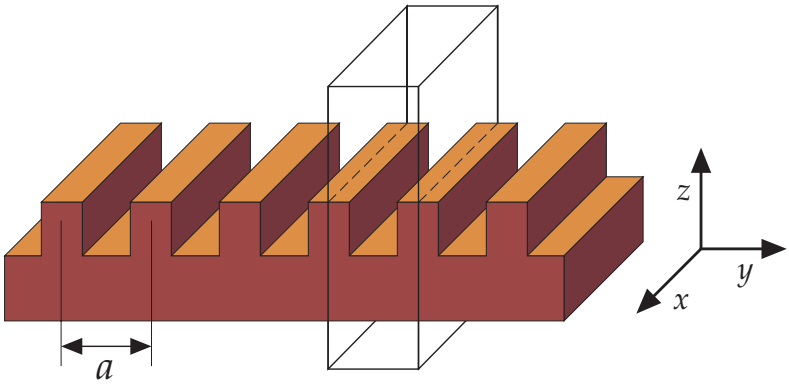


Figure 5: A dielectric configuration with discrete translational symmetry. If we imagine that the system continues forever in the y direction, then shifting the system by an integral multiple of a in the y direction leaves it unchanged. The repeated unit of this periodic system is framed with a box. This particular configuration is employed in distributed-feedback lasers, as in Yariv (1997).

The dielectric unit that we consider to be repeated over and over, highlighted in figure 5 with a box, is known as the **unit cell**. In this example, the unit cell is an xz slab of dielectric material with width a in the y direction.

Because of the translational symmetries, $\hat{\Theta}$ must commute with *all* of the translation operators in the x direction, as well as the translation operators for lattice vectors $\mathbf{R} = \ell a\hat{\mathbf{y}}$ in the y direction. With this knowledge, we can identify the modes of $\hat{\Theta}$ as simultaneous eigenfunctions of both translation operators. As before, these eigenfunctions are plane waves:

$$\begin{aligned} \hat{T}_{d\hat{\mathbf{x}}} e^{ik_x x} &= e^{ik_x(x-d)} = (e^{-ik_x d})e^{ik_x x} \\ \hat{T}_{\mathbf{R}} e^{ik_y y} &= e^{ik_y(y-\ell a)} = (e^{-ik_y \ell a})e^{ik_y y}. \end{aligned} \tag{7}$$

We can begin to classify the modes by specifying k_x and k_y. However, not all values of k_y yield different eigenvalues. Consider two modes, one with wave vector k_y and the other with wave vector $k_y + 2\pi/a$. A quick insertion into (7) shows that they have the same $\hat{T}_{\mathbf{R}}$ eigenvalues. In fact, all of the modes with wave vectors of the form $k_y + m(2\pi/a)$, where m is an integer, form a degenerate set; they all have the same $\hat{T}_{\mathbf{R}}$ eigenvalue of $e^{-i(k_y \ell a)}$. Augmenting k_y by an integral multiple of $b = 2\pi/a$ leaves the state unchanged. We call $\mathbf{b} = b\hat{\mathbf{y}}$ the primitive **reciprocal lattice vector**.

Since any linear combination of these degenerate eigenfunctions is itself an eigenfunction with the same eigenvalue, we can take linear combinations of our

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

original modes to put them in the form

$$\begin{aligned}\mathbf{H}_{k_x,k_y}(\mathbf{r}) &= e^{ik_x x}\sum_m \mathbf{c}_{k_y,m}(z)e^{i(k_y+mb)y}\\ &= e^{ik_x x}\cdot e^{ik_y y}\cdot\sum_m \mathbf{c}_{k_y,m}(z)e^{imby}\\ &= e^{ik_x x}\cdot e^{ik_y y}\cdot \mathbf{u}_{k_y}(y,z),\end{aligned} \tag{8}$$

where the $\mathbf{c}$'s are expansion coefficients to be determined by explicit solution, and $\mathbf{u}(y,z)$ is (by construction) a periodic function in y: by inspection of equation (8), we can verify that $\mathbf{u}(y+\ell a,z)=\mathbf{u}(y,z)$.

The discrete periodicity in the y direction leads to a y dependence for $\mathbf{H}$ that is simply the product of a plane wave with a y-periodic function. We can think of it as a plane wave, as it would be in free space, but modulated by a periodic function because of the periodic lattice:

$$\mathbf{H}(\ldots,y,\ldots)\propto e^{ik_y y}\cdot\mathbf{u}_{k_y}(y,\ldots). \tag{9}$$

This result is commonly known as **Bloch's theorem**. In solid-state physics, the form of (9) is known as a **Bloch state** (as in Kittel, 1996), and in mechanics as a **Floquet mode** (as in Mathews and Walker, 1964). We will use the former name.[5]

One key fact about Bloch states is that the Bloch state with wave vector k_y and the Bloch state with wave vector k_y+mb are identical. The k_y's that differ by integral multiples of $b=2\pi/a$ are not different from a physical point of view. Thus, the mode frequencies must also be periodic in k_y: $\omega(k_y)=\omega(k_y+mb)$. In fact, we need only consider k_y to exist in the range $-\pi/a<k_y\leqslant\pi/a$. This region of important, nonredundant values of k_y is called the **Brillouin zone**.[6] Readers unfamiliar with the notion of a reciprocal lattice or a Brillouin zone might find appendix B a useful introduction to that material.

We digress briefly to make the analogous statements that apply when the dielectric is periodic in three dimensions; here we skip the details and summarize the results. In this case the dielectric is invariant under translations through a multitude of lattice vectors $\mathbf{R}$ in three dimensions. Any one of these lattice vectors can be written as a particular combination of three primitive lattice vectors $(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)$ that are said to "span" the space of lattice vectors. In other words, every $\mathbf{R}=\ell\mathbf{a}_1+m\mathbf{a}_2+n\mathbf{a}_3$ for some integers ℓ, m, and n. As explained in appendix B, the vectors $(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)$ give rise to three primitive *reciprocal* lattice vectors $(\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3)$ defined so that $\mathbf{a}_i\cdot\mathbf{b}_j=2\pi\delta_{ij}$. These reciprocal vectors span a **reciprocal lattice** of their own which is inhabited by wave vectors.

[5] In fact, the essentials of this theorem were discovered independently at least four different times (in four languages), by Hill (1877), Floquet (1883), Lyapunov (1892), and Bloch (1928). With this in mind, it is not surprising that the nomenclature is often confusing.

[6] Strictly speaking, this is known as the *first* Brillouin zone.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

The modes of a three-dimensional periodic system are Bloch states that can be labelled by a **Bloch wave vector** $\mathbf{k} = k_1\mathbf{b}_1 + k_2\mathbf{b}_2 + k_3\mathbf{b}_3$ where $\mathbf{k}$ lies in the Brillouin zone. For example, for a crystal in which the unit cell is a rectangular box, the Brillouin zone is given by $|k_i| \leqslant 1/2$. Each value of the wave vector $\mathbf{k}$ inside the Brillouin zone identifies an eigenstate of $\hat{\Theta}$ with frequency $\omega(\mathbf{k})$ and an eigenvector $\mathbf{H}_\mathbf{k}$ of the form

$$\mathbf{H}_\mathbf{k}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{u}_\mathbf{k}(\mathbf{r}), \tag{10}$$

where $\mathbf{u}_\mathbf{k}(\mathbf{r})$ is a periodic function on the lattice: $\mathbf{u}_\mathbf{k}(\mathbf{r}) = \mathbf{u}_\mathbf{k}(\mathbf{r}+\mathbf{R})$ for all lattice vectors $\mathbf{R}$.

Just as continuous translational symmetry leads to the conservation of the wave vector, a corollary of Bloch's theorem is that $\mathbf{k}$ is a conserved quantity in a periodic system, modulo the addition of reciprocal lattice vectors. Addition of a reciprocal lattice vector does not change an eigenstate or its propagation direction; it is essentially a mere change of label, as discussed further in the section Bloch-Wave Propagation Velocity. This is quite different from the free-space case, in which all wave vectors represent physically distinct states. In the section The Physical Origin of Photonic Band Gaps of chapter 4, we return to this question by considering a plane wave in free space and imagining that a periodicity is slowly turned on by gradually increasing the strength of a periodic dielectric perturbation.

Photonic Band Structures

From very general symmetry principles, we have just suggested that the electromagnetic modes of a photonic crystal with discrete periodicity in three dimensions can be written as Bloch states, as in equation (10). All of the information about such a mode is given by the wave vector $\mathbf{k}$ and the periodic function $\mathbf{u}_\mathbf{k}(\mathbf{r})$. To solve for $\mathbf{u}_\mathbf{k}(\mathbf{r})$, we insert the Bloch state into the master equation (7) of chapter 2:

$$\begin{aligned}
\hat{\Theta}\mathbf{H}_\mathbf{k} &= (\omega(\mathbf{k})/c)^2\mathbf{H}_\mathbf{k} \\
\nabla \times \frac{1}{\varepsilon(\mathbf{r})}\nabla \times e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{u}_\mathbf{k}(\mathbf{r}) &= (\omega(\mathbf{k})/c)^2 e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{u}_\mathbf{k}(\mathbf{r}) \\
(i\mathbf{k}+\nabla) \times \frac{1}{\varepsilon(\mathbf{r})}(i\mathbf{k}+\nabla) \times \mathbf{u}_\mathbf{k}(\mathbf{r}) &= (\omega(\mathbf{k})/c)^2\mathbf{u}_\mathbf{k}(\mathbf{r}) \\
\hat{\Theta}_\mathbf{k}\mathbf{u}_\mathbf{k}(\mathbf{r}) &= (\omega(\mathbf{k})/c)^2\mathbf{u}_\mathbf{k}(\mathbf{r}).
\end{aligned} \tag{11}$$

Here we have defined $\hat{\Theta}_\mathbf{k}$ as a new Hermitian operator that appears in this substitution and depends on $\mathbf{k}$:

$$\hat{\Theta}_\mathbf{k} \triangleq (i\mathbf{k}+\nabla) \times \frac{1}{\varepsilon(\mathbf{r})}(i\mathbf{k}+\nabla) \times . \tag{12}$$

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

The function $\mathbf{u}$, and therefore the mode profiles, are determined by the eigenvalue problem in the fourth equation of (11), subject to transversality $(i\mathbf{k}+\nabla)\cdot\mathbf{u}_\mathbf{k}=0$ and the periodicity condition

$$\mathbf{u}_\mathbf{k}(\mathbf{r})=\mathbf{u}_\mathbf{k}(\mathbf{r}+\mathbf{R}). \tag{13}$$

Because of this periodic boundary condition, we can regard the eigenvalue problem as being restricted to a single unit cell of the photonic crystal. As was discussed in the section Discrete vs. Continuous Frequency Ranges of chapter 2, restricting a Hermitian eigenvalue problem to a finite volume leads to a discrete spectrum of eigenvalues. We can expect to find, for each value of $\mathbf{k}$, an infinite set of modes with discretely spaced frequencies, which we can label by a band index n.

Since $\mathbf{k}$ enters as a continuous parameter in $\hat{\Theta}$, we expect the frequency of each band, for a given n, to vary continuously as $\mathbf{k}$ varies. In this way we arrive at the description of the modes of a photonic crystal: They are a family of continuous functions, $\omega_n(\mathbf{k})$, indexed in order of increasing frequency by the band number. The information contained in these functions is called the **band structure** of the photonic crystal. Studying the band structure of a crystal supplies us with most of the information we need to predict its optical properties, as we will see.

For a given photonic crystal $\varepsilon(\mathbf{r})$, how can we calculate the band structure functions $\omega_n(\mathbf{k})$? Powerful computational techniques are available for the task, but we will not discuss them extensively. The focus of this text is on concepts and results, not on the numerical studies of the equations. A brief outline of the technique that was used to generate the band structures in this text is in appendix D. In essence, the technique relies on the fact that the last equation of (11) is a standard eigenvalue equation that is readily solvable by an iterative minimization technique for each value of $\mathbf{k}$.

Rotational Symmetry and the Irreducible Brillouin Zone

Photonic crystals might have symmetries other than discrete translations. A given crystal might also be left invariant after a rotation, a mirror reflection, or an inversion is performed. To begin, we examine the conclusions we can draw about the modes of a system with *rotational* symmetry.

Suppose the operator (3×3 matrix) $\mathcal{R}(\hat{\mathbf{n}},\alpha)$ rotates vectors by an angle α about the $\hat{\mathbf{n}}$ axis. Abbreviate $\mathcal{R}(\hat{\mathbf{n}},\alpha)$ by $\mathcal{R}$. To rotate a vector field $\mathbf{f}(\mathbf{r})$, we take the vector $\mathbf{f}$ and rotate it with $\mathcal{R}$ to give $\mathbf{f}'=\mathcal{R}\mathbf{f}$. We also rotate the argument $\mathbf{r}$ of the vector field: $\mathbf{r}'=\mathcal{R}^{-1}\mathbf{r}$. Therefore $\mathbf{f}'(\mathbf{r}')=\mathcal{R}\mathbf{f}(\mathbf{r}')=\mathcal{R}\mathbf{f}(\mathcal{R}^{-1}\mathbf{r})$. Accordingly, we define the vector field rotator $\hat{O}_\mathcal{R}$ as

$$\hat{O}_\mathcal{R}\cdot\mathbf{f}(\mathbf{r})=\mathcal{R}\mathbf{f}(\mathcal{R}^{-1}\mathbf{r}). \tag{14}$$

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

If rotation by $\mathcal{R}$ leaves the system invariant, then we conclude (as before) that $[\hat{\Theta}, \hat{O}_{\mathcal{R}}] = 0$. Therefore, we may carry out the following manipulation:

$$\hat{\Theta}(\hat{O}_{\mathcal{R}}\mathbf{H}_{\mathbf{k}n}) = \hat{O}_{\mathcal{R}}(\hat{\Theta}\mathbf{H}_{\mathbf{k}n}) = \left(\frac{\omega_n(\mathbf{k})}{c}\right)^2 (\hat{O}_{\mathcal{R}}\mathbf{H}_{\mathbf{k}n}). \tag{15}$$

We see that $\hat{O}_{\mathcal{R}}\mathbf{H}_{\mathbf{k}n}$ also satisfies the master equation, with the same eigenvalue as $\mathbf{H}_{\mathbf{k}n}$. This means that the rotated mode is itself an allowed mode, with the same frequency. We can further prove that the state $\hat{O}_{\mathcal{R}}\mathbf{H}_{\mathbf{k}n}$ is none other than the Bloch state with wave vector $\mathcal{R}\mathbf{k}$. To do this, we must show that $\hat{O}_{\mathcal{R}}\mathbf{H}_{\mathbf{k}n}$ is an eigenfunction of the translation operator $\hat{T}_{\mathbf{R}}$ with eigenvalue $e^{-i\mathcal{R}\mathbf{k}\cdot\mathbf{R}}$ where $\mathbf{R}$ is a lattice vector. We can do just that, using the fact that $\hat{\Theta}$ and $\hat{O}_{\mathcal{R}}$ commute and thus $\mathcal{R}^{-1}\mathbf{R}$ must also be a lattice vector:

$$\begin{aligned}
\hat{T}_{\mathbf{R}}(\hat{O}_{\mathcal{R}}\mathbf{H}_{\mathbf{k}n}) &= \hat{O}_{\mathcal{R}}(\hat{T}_{\mathcal{R}^{-1}\mathbf{R}}\mathbf{H}_{\mathbf{k}n}) \\
&= \hat{O}_{\mathcal{R}}(e^{-i(\mathbf{k}\cdot\mathcal{R}^{-1}\mathbf{R})}\mathbf{H}_{\mathbf{k}n}) \\
&= e^{-i(\mathbf{k}\cdot\mathcal{R}^{-1}\mathbf{R})}(\hat{O}_{\mathcal{R}}\mathbf{H}_{\mathbf{k}n}) \\
&= e^{-i(\mathcal{R}\mathbf{k}\cdot\mathbf{R})}\hat{O}_{\mathcal{R}}\mathbf{H}_{\mathbf{k}n}.
\end{aligned} \tag{16}$$

Since $\hat{O}_{\mathcal{R}}\mathbf{H}_{\mathbf{k}n}$ is the Bloch state with wave vector $\mathcal{R}\mathbf{k}$ and has the same eigenvalue as $\mathbf{H}_{\mathbf{k}n}$, it follows that

$$\omega_n(\mathcal{R}\mathbf{k}) = \omega_n(\mathbf{k}). \tag{17}$$

We conclude that when there is rotational symmetry in the lattice, the frequency bands $\omega_n(\mathbf{k})$ have additional redundancies within the Brillouin zone. In a similar manner, we can show that whenever a photonic crystal has a rotation, mirror-reflection, or inversion symmetry, the $\omega_n(\mathbf{k})$ functions have that symmetry as well. This particular collection of symmetry operations (rotations, reflections, and inversions) is called the **point group** of the crystal.

Since the functions $\omega_n(\mathbf{k})$ possess the full symmetry of the point group, we need not consider them at every $\mathbf{k}$ point in the Brillouin zone. The smallest region within the Brillouin zone for which the $\omega_n(\mathbf{k})$ are not related by symmetry is called the **irreducible Brillouin zone**. For example, a photonic crystal with the symmetry of a simple square lattice has a square Brillouin zone centered at $\mathbf{k} = 0$, as depicted in figure 6. (See appendix B for a fuller discussion of the reciprocal lattice and the Brillouin zone.) The *irreducible* zone is a triangular wedge with $1/8$ the area of the full Brillouin zone; the rest of the Brillouin zone consists of redundant copies of the irreducible zone.

Mirror Symmetry and the Separation of Modes

Mirror reflection symmetry in a photonic crystal deserves special attention. Under certain conditions it allows us to separate the eigenvalue equation for $\hat{\Theta}_{\mathbf{k}}$ into two

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

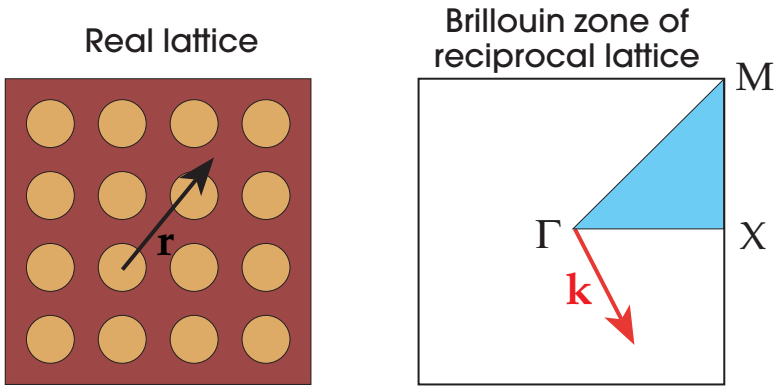


Figure 6: *Left:* A photonic crystal made using a square lattice. An arbitrary vector $\mathbf{r}$ is shown. *Right:* The Brillouin zone of the square lattice, centered at the origin (Γ). An arbitrary wave vector $\mathbf{k}$ is shown. The **irreducible zone** is the light blue triangular wedge. The special points at the center, corner, and face are conventionally known as Γ, M, and X.

separate equations, one for each field polarization. As we will see, in one case $\mathbf{H_k}$ is perpendicular to the mirror plane and $\mathbf{E_k}$ is parallel; while in the other case, $\mathbf{H_k}$ is in the plane and $\mathbf{E_k}$ is perpendicular. This simplification is convenient, because it provides immediate information about the mode symmetries and also facilitates the numerical calculation of their frequencies.

To show how this separation of modes comes about, let us turn again to the dielectric system illustrated in figure 5, the notched dielectric. This system is invariant under mirror reflections in the yz and xz planes. We focus on reflections M_x in the yz plane (M_x changes $\hat{\mathbf{x}}$ to $-\hat{\mathbf{x}}$ and leaves $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ alone).[7] In analogy with our rotation operator, we define a mirror reflection operator $\hat{O}_{M_x}$, which reflects a vector field by using M_x on both its input and its output:

$$\hat{O}_{M_x}\mathbf{f}(\mathbf{r}) = M_x\mathbf{f}(M_x\mathbf{r}). \tag{18}$$

Two applications of the mirror reflection operator restore any system to its original state, so the possible eigenvalues of $\hat{O}_{M_x}$ are $+1$ and -1. Because the dielectric is symmetric under a mirror reflection in the yz plane, $\hat{O}_{M_x}$ commutes with $\hat{\Theta}$: $[\hat{\Theta}, \hat{O}_{M_x}] = 0$. As before, if we operate on $\mathbf{H_k}$ with this commutator we can show that $\hat{O}_{M_x}\mathbf{H_k}$ is just the Bloch state with the reflected wave vector $M_x\mathbf{k}$:

$$\hat{O}_{M_x}\mathbf{H_k} = e^{i\phi}\mathbf{H}_{M_x\mathbf{k}}. \tag{19}$$

Here, ϕ is an arbitrary phase. This relation does not restrict the reflection properties of $\mathbf{H_k}$ very much, unless $\mathbf{k}$ happens to be pointed in such a way that $M_x\mathbf{k} = \mathbf{k}$.

[7] Note that *any* slice perpendicular to the x axis is a valid mirror plane for our system. Thus for any $\mathbf{r}$ in the crystal we can always find a plane such that $M_x\mathbf{r} = \mathbf{r}$. This is not true for M_y.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

When this is true, (19) becomes an eigenvalue problem and, using (18), we see that $\mathbf{H_k}$ must obey

$$\hat{O}_{M_x}\mathbf{H_k}(\mathbf{r}) = \pm\mathbf{H_k}(\mathbf{r}) = M_x\mathbf{H_k}(M_x\mathbf{r}). \tag{20}$$

Although we will not show it explicitly, the electric field $\mathbf{E_k}$ obeys a similar equation, so that both the electric and magnetic fields must be either *even* or *odd* under the $\hat{O}_{M_x}$ operation. But $M_x\mathbf{r} = \mathbf{r}$ for any $\mathbf{r}$ in our dielectric (taken in the two-dimensional yz plane). Therefore, since $\mathbf{E}$ transforms like a vector and $\mathbf{H}$ transforms like a pseudovector (see footnote 2 on page 26), the only nonzero field components of an $\hat{O}_{M_x}$-*even* mode must be H_x, E_y, and E_z. The *odd* modes are described by the components E_x, H_y, and H_z.

In general, given a reflection M such that $[\hat{\Theta}, \hat{O}_M] = 0$, this separation of modes is only possible at $M\mathbf{r} = \mathbf{r}$ for $M\mathbf{k} = \mathbf{k}$. Note from (11) that $\hat{\Theta}_\mathbf{k}$ and $\hat{O}_M$ will not commute unless $M\mathbf{k} = \mathbf{k}$. It appears that the separation of polarizations holds only under fairly restricted conditions and is not that useful for three-dimensional photonic-crystal analyses. (However, a generalization is discussed in the section Symmetry and Polarization of chapter 7.)

On the other hand, these conditions can always be met for two-dimensional photonic crystals. Two-dimensional crystals are periodic in a certain plane, but are uniform along an axis perpendicular to that plane. Calling that axis the z axis, we know that the operation $\hat{\mathbf{z}} \rightarrow -\hat{\mathbf{z}}$ is a symmetry of the crystal for any choice of origin. It also follows that $M_z\mathbf{k}_\parallel = \mathbf{k}_\parallel$ for all wave vectors $\mathbf{k}_\parallel$ in the two-dimensional Brillouin zone. Thus the modes of *every* two-dimensional photonic crystal can be classified into two distinct polarizations: either (E_x, E_y, H_z) or (H_x, H_y, E_z). The former, in which the *electric* field is confined to the xy plane, are called transverse-electric (**TE**) modes. The latter, in which the *magnetic* field is confined to the xy plane, are called transverse-magnetic (**TM**) modes.[8]

Time-Reversal Invariance

We will discuss one more symmetry in detail, and it is of global significance: the *time-reversal* symmetry. If we take the complex conjugate of the master equation for $\hat{\Theta}$ [equation (7) of chapter 2], and use the fact that the eigenvalues are real for lossless materials, we obtain

$$\begin{aligned}(\hat{\Theta}\mathbf{H}_{\mathbf{k}n})^* &= \frac{\omega_n^2(\mathbf{k})}{c^2}\mathbf{H}^*_{\mathbf{k}n} \\ \hat{\Theta}\mathbf{H}^*_{\mathbf{k}n} &= \frac{\omega_n^2(\mathbf{k})}{c^2}\mathbf{H}^*_{\mathbf{k}n}.\end{aligned} \tag{21}$$

[8] Classical waveguide theory sometimes uses different conventions for the meaning of "TE" and "TM." Our notation is common in the photonic-crystal literature, but it is wise to be aware that other authors use different nomenclatures.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

By this manipulation, we see that $\mathbf{H}^*_{\mathbf{k}n}$ satisfies the same equation as $\mathbf{H}_{\mathbf{k}n}$, with the very same eigenvalue. But from (10) we see that $\mathbf{H}^*_{\mathbf{k}n}$ is just a Bloch state at $-\mathbf{k}$. It follows that

$$\omega_n(\mathbf{k}) = \omega_n(-\mathbf{k}). \tag{22}$$

The above relation holds for almost all photonic crystals.[9] The frequency bands have inversion symmetry even if the crystal does not. Taking the complex conjugate of $\mathbf{H}_{\mathbf{k}n}$ is equivalent to reversing the sign of time t in the Maxwell equations, as can be verified from equation (5) of chapter 2. For this reason, we say that (22) is a consequence of the *time-reversal* symmetry of the Maxwell equations.

Bloch-Wave Propagation Velocity

At this point, some remarks are warranted on the physical interpretation of the Bloch state, to ward off common points of confusion. The Bloch state $\mathbf{H}_{\mathbf{k}}(\mathbf{r})e^{-i\omega t}$ is a plane wave $e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$ multiplied by a periodic "envelope" function $\mathbf{u}_{\mathbf{k}}(\mathbf{r})$. It propagates through the crystal without scattering, because $\mathbf{k}$ is conserved (apart from addition of reciprocal lattice vectors, which is merely a relabeling). Or, equivalently, all of the scattering events are *coherent* and result in the periodic shape of $\mathbf{u}_{\mathbf{k}}$.

In the familiar case of a homogeneous, isotropic medium, $\mathbf{k}$ is the direction in which the wave propagates, but this is *not* necessarily true in a periodic medium. Rather, the direction and the speed with which electromagnetic energy passes through the crystal are given by the **group velocity** $\mathbf{v}$, which is a function of both the band index n and the wave vector $\mathbf{k}$:

$$\mathbf{v}_n(\mathbf{k}) \triangleq \nabla_{\mathbf{k}}\omega_n \triangleq \frac{\partial\omega_n}{\partial k_x}\hat{\mathbf{x}} + \frac{\partial\omega_n}{\partial k_y}\hat{\mathbf{y}} + \frac{\partial\omega_n}{\partial k_z}\hat{\mathbf{z}}, \tag{23}$$

where $\nabla_{\mathbf{k}}$ is the gradient with respect to $\mathbf{k}$. The group velocity is the energy-transport velocity whenever the medium is lossless, the material dispersion is small, and the wave vector is real. This fact is often derived, in a homogeneous medium, by considering the propagation of a broad pulse of energy through the medium (as in Jackson, 1998).[10] In our case, it is perhaps easier to employ the

[9] As an exception, **magneto-optic** materials can break time-reversal symmetry. Such materials are described by a dielectric tensor ε that is a 3×3 complex Hermitian matrix (Landau et al., 1984). In this case, $\varepsilon^* \neq \varepsilon$ even for a lossless medium.

[10] Several classic papers on wave velocity are collected in Brillouin (1960). Early discussions of group velocity in periodic structures can be found in Brillouin (1946) and Yeh (1979). Beware that the issue of the energy-transport velocity is greatly complicated if one includes nonnegligible losses, a frequency-dependent ε (material dispersion), or complex values of $\mathbf{k}$ (evanescent modes, such as those of the section Evanescent Modes in Photonic Band Gaps of chapter 4).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

formalism we have already developed in previous chapters. We differentiate the eigenequation $\hat{\Theta}_\mathbf{k}\mathbf{u}_\mathbf{k} = (\omega/c)^2\mathbf{u}_\mathbf{k}$ with respect to $\mathbf{k}$, and then take the inner product with $\mathbf{u}_\mathbf{k}$ on both sides:[11]

$$\left(\mathbf{u}_\mathbf{k}, \nabla_\mathbf{k}\left[\hat{\Theta}_\mathbf{k}\mathbf{u}_\mathbf{k}\right]\right) = \left(\mathbf{u}_\mathbf{k}, \nabla_\mathbf{k}\left[\frac{\omega^2}{c^2}\mathbf{u}_\mathbf{k}\right]\right), \tag{24}$$

giving

$$\left(\mathbf{u}_\mathbf{k}, [\nabla_\mathbf{k}\hat{\Theta}_\mathbf{k}]\mathbf{u}_\mathbf{k} + \hat{\Theta}_\mathbf{k}\nabla_\mathbf{k}\mathbf{u}_\mathbf{k}\right) = \left(\mathbf{u}_\mathbf{k}, 2\frac{\omega}{c^2}\mathbf{v}\mathbf{u}_\mathbf{k} + \frac{\omega^2}{c^2}\nabla_\mathbf{k}\mathbf{u}_\mathbf{k}\right). \tag{25}$$

The $\nabla_\mathbf{k}\mathbf{u}_\mathbf{k}$ terms on both sides cancel one another, because the Hermitian operator $\hat{\Theta}_\mathbf{k}$ that appears in the second term on the left-hand side can operate leftwards and produce $(\omega/c)^2$, matching the second term on the right-hand side. The remaining terms can be solved for $\mathbf{v} = \nabla_\mathbf{k}\omega$, yielding

$$\mathbf{v} = \frac{c^2}{2\omega}\frac{\left(\mathbf{u}_\mathbf{k}, [\nabla_\mathbf{k}\hat{\Theta}_\mathbf{k}]\mathbf{u}_\mathbf{k}\right)}{(\mathbf{u}_\mathbf{k}, \mathbf{u}_\mathbf{k})}. \tag{26}$$

The right-hand side can now be rearranged to a more suggestive form by writing $\mathbf{u}_\mathbf{k} = e^{-i\mathbf{k}\cdot\mathbf{r}}\mathbf{H}_\mathbf{k}$ from equation (10). The denominator then becomes $4U_\mathbf{H}/\mu_0 = 2(U_\mathbf{E} + U_\mathbf{H})/\mu_0$, given the definitions [equation (24) of chapter 2] of the time-averaged electric and magnetic energy. The numerator (with the $c^2/2\omega$ factor) is in fact $2/\mu_0$ times the average electromagnetic energy flux: that is, it is $(2/\mu_0)\int \mathbf{S} = (2/\mu_0)\,\mathrm{Re}\int d^3\mathbf{r}\,\mathbf{E}^* \times \mathbf{H}/2$ (the integral of the Poynting vector $\mathbf{S}$ from equation (25) of chapter 2). This can be seen by differentiating $\hat{\Theta}_\mathbf{k}$ from (12) and applying equation (8) of chapter 2. The final result is that $\mathbf{v}$ is the ratio of the energy flux to the energy density (averaged in time and over the unit cell):

$$\nabla_\mathbf{k}\omega = \mathbf{v} = \frac{\frac{1}{2}\,\mathrm{Re}\int d^3\mathbf{r}\,\mathbf{E}^* \times \mathbf{H}}{\frac{1}{4}\int d^3\mathbf{r}\,(\mu_0|\mathbf{H}|^2 + \varepsilon_0\varepsilon|\mathbf{E}|^2)} = \frac{\int d^3\mathbf{r}\,\mathbf{S}}{U_\mathbf{E} + U_\mathbf{H}}, \tag{27}$$

which by definition is the velocity of energy propagation. For a real $\mathbf{k}$ and a real dielectric function $\varepsilon \geqslant 1$ that is independent of frequency, this speed $|\mathbf{v}|$ is always $\leqslant c$.[12] (In more general cases, the definition of velocity itself is more subtle.)

[11] As in solid-state physics, the process of differentiating the eigenequation with $\mathbf{k}$ is the entry point into "$\mathbf{k}\cdot\mathbf{p}$ theory" (Sipe, 2000), and is also equivalent to the Hellman–Feynman theorem of quantum mechanics. The result is analogous to the first-order perturbation theory described in the section The Effect of Small Perturbations of chapter 2, but with $\Delta\mathbf{k}$ replacing $\Delta\varepsilon$ as the perturbation.

[12] $|\mathbf{v}| \leqslant c$ for $\varepsilon \geqslant 1$ can be proved by applying a sequence of elementary inequalities to the numerator of $|\mathbf{v}| = |\text{flux}|/\text{energy}$. In particular, $\left|\mathrm{Re}\int \mathbf{E}^* \times \mathbf{H}\right| \leqslant \int |\mathbf{E}^* \times \mathbf{H}| \leqslant \int |\sqrt{\varepsilon}\mathbf{E}| \cdot |\mathbf{H}| \leqslant \sqrt{\int \varepsilon|\mathbf{E}|^2}\sqrt{\int |\mathbf{H}|^2} = 4cU_\mathbf{H}$ (where the final $\leqslant$ was the Cauchy–Schwarz inequality), which cancels the denominator $2(U_\mathbf{E} + U_\mathbf{H}) = 4U_\mathbf{H}$, leaving c.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Table 1

	Quantum Mechanics	*Electrodynamics*
Discrete translational symmetry	$V(\mathbf{r}) = V(\mathbf{r}+\mathbf{R})$	$\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r}+\mathbf{R})$
Commutation relationships	$[\hat{H}, \hat{T}_{\mathbf{R}}] = 0$	$[\hat{\Theta}, \hat{T}_{\mathbf{R}}] = 0$
Bloch's theorem	$\Psi_{\mathbf{k}n}(\mathbf{r}) = u_{\mathbf{k}n}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$	$\mathbf{H}_{\mathbf{k}n}(\mathbf{r}) = \mathbf{u}_{\mathbf{k}n}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$

Quantum mechanics vs. electrodynamics in periodic systems.

Another concept with which the reader may be familiar is the **phase velocity**, usually given by $\omega\mathbf{k}/|\mathbf{k}|^2$. For a photonic crystal, however, the phase velocity is difficult to define because $\mathbf{k}$ is not unique: it is equivalent to $\mathbf{k}+\mathbf{G}$ for any reciprocal lattice vector $\mathbf{G}$. Equivalently, because of the periodic envelope function $\mathbf{u}_{\mathbf{k}}$ modulating the plane wave, it is hard to identify unique phase fronts whose velocity is to be measured.[13] In contrast, adding a reciprocal lattice vector to $\mathbf{k}$ has no effect on the group velocity.

Electrodynamics vs. Quantum Mechanics Again

As in the previous chapter, we summarize by way of analogy with quantum mechanics. Table 1 compares the system containing an electron propagating in a periodic potential with the system of electromagnetic modes in a photonic crystal. appendix A develops this analogy further.

In both cases, the systems have translational symmetry: in quantum mechanics the potential $V(\mathbf{r})$ is periodic, and in the electromagnetic case the dielectric function $\varepsilon(\mathbf{r})$ is periodic. This periodicity implies that the discrete translation operator commutes with the major differential operator of the problem, whether with the Hamiltonian or with $\hat{\Theta}$.

We can index the eigenstates ($\Psi_{\mathbf{k}n}$ or $\mathbf{H}_{\mathbf{k}n}$) using the translation operator eigenvalues. These can be labelled in terms of the wave vectors and bands in the Brillouin zone. All of the eigenstates can be cast in Bloch form: a periodic function modulated by a plane wave. The field can propagate through the crystal in a coherent manner, as a Bloch wave. The understanding of Bloch waves for electrons explained one of the great mysteries of nineteenth-century physics: Why do electrons behave like free particles in many examples of conducting crystals? In the same way, a photonic crystal provides a *synthetic medium* in which light can propagate, but in ways quite different from propagation in a homogeneous medium.

[13] For purposes of *qualitative* description, one sometimes defines a "phase velocity" in a photonic crystal by arbitrarily restricting $\mathbf{k}$ to the first Brillouin zone, or alternatively by picking the $\mathbf{k}+\mathbf{G}$ corresponding to the largest Fourier component. Care must be taken in interpreting such a quantity, however.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Further Reading

The study of symmetry in the most general context falls under the mathematical subject of **group theory** or, more specifically, group representation theory. Perhaps most useful references on this subject are texts that apply the formalism of group theory to specific physical disciplines. For example, Tinkham (2003) and Inui et al. (1996) cover both molecular and solid-state applications, and include useful tables of the possible crystallographic symmetry groups.

Readers completely unfamiliar with concepts like the reciprocal lattice, the Brillouin zone, or Bloch's theorem might find it useful to consult the first few chapters of Kittel (1996). There, the concepts are introduced where they find common use, in conventional solid-state physics. Additionally, appendix B of this text contains a brief introduction to the reciprocal lattice and the Brillouin zone. The earliest works on waves in periodic media are reviewed in, for example, Brillouin (1946).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

4

The Multilayer Film: A One-Dimensional Photonic Crystal

We begin our study of photonic crystals by considering the simplest possible case, a one-dimensional system, and applying the principles of electromagnetism and symmetry that we developed in the previous chapters. Even in this simple system, we can discern some of the most important features of photonic crystals in general, such as photonic band gaps and modes that are localized around defects. The optical properties of a one-dimensional layered system may be familiar, but by expressing the results in the language of band structures and band gaps, we can discover new phenomena such as omnidirectional reflectivity, as well as prepare for the more complicated two- and three-dimensional systems that lie ahead.

The Multilayer Film

The simplest possible photonic crystal, shown in figure 1, consists of alternating layers of material with different dielectric constants: a **multilayer film**. This arrangement is not a new idea. Lord Rayleigh (1887) published one of the first analyses of the optical properties of multilayer films. As we will see, this type of photonic crystal can act as a *mirror* (a **Bragg mirror**) for light with a frequency within a specified range, and it can localize light modes if there are any defects in its structure. These concepts are commonly used in dielectric mirrors and optical filters (as in, e.g., Hecht and Zajac, 1997).

The traditional way to analyze this system, pioneered by Lord Rayleigh (1917), is to imagine that a plane wave propagates through the material and to consider the sum of the multiple reflections and refractions that occur at each interface. In this chapter, we will use a different approach—the analysis of

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

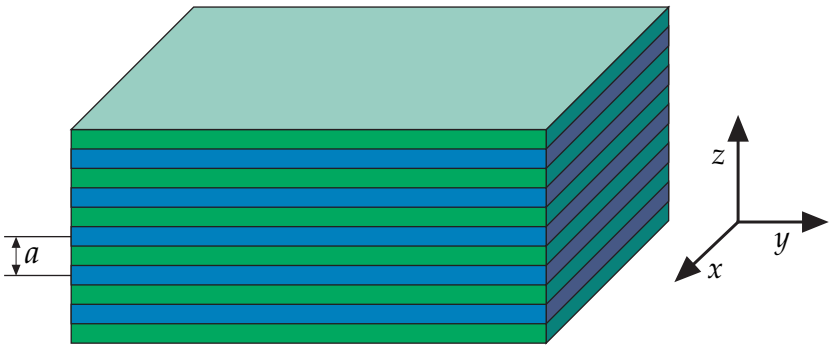


Figure 1: The multilayer film, a one-dimensional photonic crystal. The term "one-dimensional" is used because the dielectric function $\varepsilon(z)$ varies along one direction (z) only. The system consists of alternating layers of materials (blue and green) with different dielectric constants, with a spatial period a. We imagine that each layer is uniform and extends to infinity along the x and y directions, and we imagine that the periodicity in the z direction also extends to infinity.

band structures—that is easily generalized to the more complex two- and three-dimensional photonic crystals.[1]

We begin in the spirit of the previous chapter. By applying symmetry arguments, we can describe the electromagnetic modes sustainable by the crystal. The material is periodic in the z direction, and homogeneous in the xy plane. As we saw in the previous chapter, this allows us to classify the modes using $\mathbf{k}_\parallel$, k_z, and n: the wave vector in the plane, the wave vector in the z direction, and the band number. The wave vectors specify how the mode transforms under translation operators, and the band number increases with frequency. We can write the modes in the Bloch form:

$$\mathbf{H}_{n,k_z,\mathbf{k}_\parallel}(\mathbf{r}) = e^{i\mathbf{k}_\parallel\cdot\rho}e^{ik_z z}\mathbf{u}_{n,k_z,\mathbf{k}_\parallel}(z). \tag{1}$$

The function $\mathbf{u}(z)$ is periodic, with the property $\mathbf{u}(z) = \mathbf{u}(z+R)$ whenever R is an integral multiple of the spatial period a. Because the crystal has continuous translational symmetry in the xy plane, the wave vector $\mathbf{k}_\parallel$ can assume any value. However, the wave vector k_z can be restricted to a finite interval, the one-dimensional Brillouin zone, because the crystal has discrete translational symmetry in the z direction. Using the prescriptions of the previous chapter, if the primitive lattice vector is $a\hat{\mathbf{z}}$ then the primitive reciprocal lattice vector is $(2\pi/a)\hat{\mathbf{z}}$ and the Brillouin zone is $-\pi/a < k_z \leq \pi/a$.

[1] Interestingly, in Lord Rayleigh's first attack on the problem in 1887, he used a cumbersome predecessor of Bloch's theorem that had been worked out by Hill (1877). In modern terminology, Rayleigh was able to show that any one-dimensional photonic crystal has a band gap. When he returned to the problem in 1917, however, he switched to the sum-of-reflections technique.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

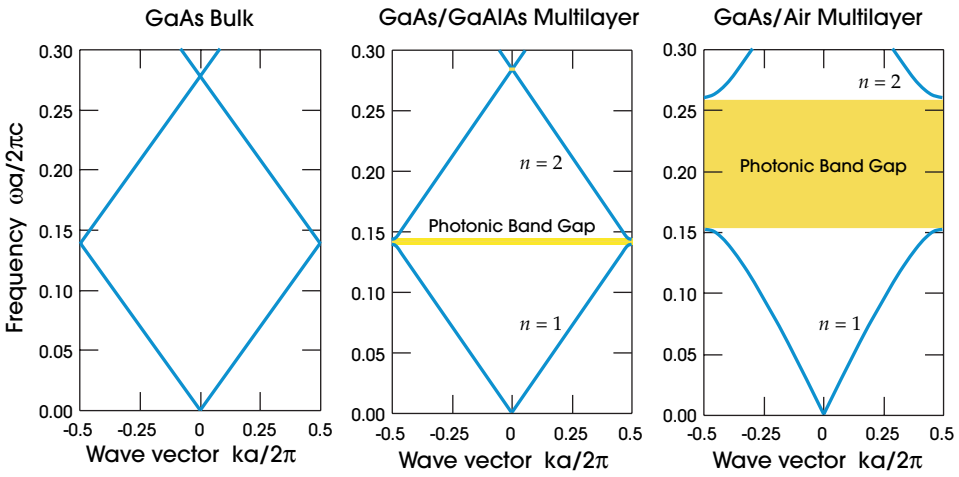


Figure 2: The photonic band structures for on-axis propagation, as computed for three different multilayer films. In all three cases, each layer has a width $0.5a$. *Left:* every layer has the same dielectric constant $\varepsilon = 13$. *Center:* layers alternate between ε of 13 and 12. *Right:* layers alternate between ε of 13 and 1.

The Physical Origin of Photonic Band Gaps

For now, consider waves that propagate entirely in the z direction, crossing the sheets of dielectric at normal incidence. In this case, $\mathbf{k}_{\parallel} = 0$ and only the wave vector component k_z is important. Without possibility of confusion, we can abbreviate k_z by k.

In figure 2, we plot $\omega_n(k)$ for three different multilayer films. The left-hand plot is for a system in which all of the layers have the same dielectric constant; the medium is actually uniform in all three directions. The center plot is for a structure with alternating dielectric constants of 13 and 12, and the right-hand plot is for a structure with a much higher dielectric contrast of 13 to 1.[2]

The left-hand plot is for a homogeneous dielectric medium for which we have arbitrarily assigned a periodicity of a. But we already know that in a homogeneous medium, the speed of light is reduced by the index of refraction. The modes lie along the *light line* (as in the subsection Index guiding of chapter 3), given by

$$\omega(k) = \frac{ck}{\sqrt{\varepsilon}}. \tag{2}$$

[2] We use these particular values because the static dielectric constant of gallium arsenide (GaAs) is about 13, and for gallium aluminum arsenide (GaAlAs) it is about 12, as reported in Sze (1981). These and similar materials are commonly used in devices. Air has a dielectric constant very nearly equal to 1.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Because we have insisted that k repeat itself outside the Brillouin zone, the light line folds back into the zone when it reaches an edge. One can regard this as simply an eccentric way of relabeling of the solutions, in which $k + 2\pi/a$ is replaced by k.[3] The center plot, which is for a *nearly*-homogeneous medium, looks like the homogeneous case with one important difference: there is a gap in frequency between the upper and lower branches of the lines. There is no allowed mode in the crystal that has a frequency within this gap, regardless of k. We call such a gap a **photonic band gap**. The right-hand plot shows that the gap widens considerably as the dielectric contrast is increased.

We will devote a considerable amount of attention to photonic band gaps, and with good reason. Many of the promising applications of two- and three-dimensional photonic crystals to date hinge on the location and width of photonic band gaps. For example, a crystal with a band gap might make a very good, narrow-band filter, by rejecting all (and only) frequencies in the gap. A resonant cavity, carved out of a photonic crystal, would have perfectly reflecting walls for frequencies in the gap.

Why does the photonic band gap appear? We can understand the gap's physical origin by considering the electric field mode profiles for the states immediately *above* and *below* the gap. The gap between bands $n = 1$ and $n = 2$ occurs at the edge of the Brillouin zone, at $k = \pi/a$. For now, we focus on the band structure in the center panel of figure 2, corresponding to the configuration that is a small perturbation of the homogeneous system. For $k = \pi/a$, the modes have a wavelength of $2a$, twice the crystal's spatial period (or *lattice constant*). There are two ways to center a mode of this type. We can position the nodes in each low-ε layer, as in figure 3(a), or in each high-ε layer, as in figure 3(b). Any other position would violate the symmetry of the unit cell about its center.

In our study of the electromagnetic variational theorem, in the section Electromagnetic Energy and the Variational Principle of chapter 2, we found that the low-frequency modes concentrate their energy in the high-ε regions, and the high-frequency modes have a larger fraction of their energy (although not necessarily a majority) in the low-ε regions. With this in mind, it is understandable why there is a frequency difference between the two cases. The mode just *under* the gap has more of its energy concentrated in the $\varepsilon = 13$ regions as shown in figure 3(c), giving it a lower frequency than the next band, most of whose energy is in the $\varepsilon = 12$ regions as shown in figure 3(d).

The bands above and below the gap can be distinguished by where the energy of their modes is concentrated: in the high-ε regions, or in the low-ε regions. Often, especially in the two- and three-dimensional crystals of the later chapters, the low-ε regions are air regions. For this reason, it is convenient to refer to the band *above* a photonic band gap as the **air band**, and the band *below* a gap as the **dielectric band**. The situation is analogous to the electronic band structure

[3] The reader may be familiar with the relabeling of $k + 2\pi/a$ as k from the phenomenon of **quasi–phase-matching**, in which states at the same frequency can couple to one another if their k values differ by multiples of $2\pi/a$, when a weak periodicity a is introduced into a medium.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

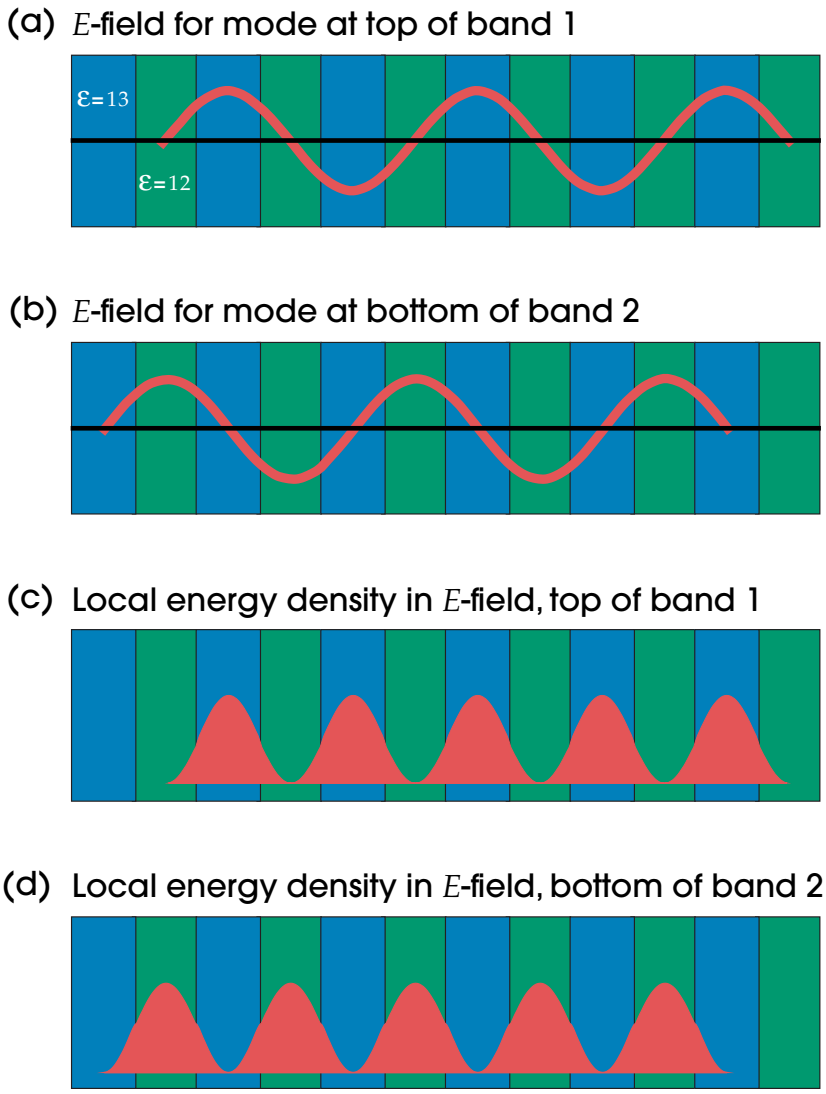


Figure 3: The modes associated with the lowest band gap of the band structure plotted in the center panel of figure 2, at $k = \pi/a$. (a) Electric field of band 1; (b) electric field of band 2; (c) electric-field energy density $\varepsilon\,|\mathbf{E}|^2/8\pi$ of band 1; (d) electric-field energy density of band 2. In the depiction of the multilayer film, blue indicates the region of *higher* dielectric constant ($\varepsilon = 13$).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

of semiconductors, in which the *conduction band* and the *valence band* bracket the fundamental gap.

This heuristic, based on the variational theorem, can be extended to describe the configuration with a large dielectric contrast. In this case, we find that the field energy for *both* bands is primarily concentrated in the high-ε layers, but in different ways—the first band being more concentrated than the second. These fields are shown in figure 4, corresponding to the right panel of figure 2. The gap arises from this difference in field energy location. Consequently, we will still refer to the upper band as the air band and the lower as the dielectric band.

We conclude this section with the observation that in one dimension, a gap usually occurs between *every* set of bands, at either the Brillouin zone's edge or its center.[4] This is illustrated for the band structure of a multilayer film in figure 5. Finally, we emphasize that band gaps always appear in a one-dimensional photonic crystal for *any* dielectric contrast. The smaller the contrast, the smaller the gaps, but the gaps open up as soon as $\varepsilon_1/\varepsilon_2 \neq 1$. This statement is quantified in the following section.

The Size of the Band Gap

The extent of a photonic band gap can be characterized by its frequency width $\Delta\omega$, but this is not a really useful measure. Remember from the section Scaling Properties of the Maxwell Equations of chapter 2 that all of our results are scalable. If the crystal were expanded by a factor s, the corresponding band gap would have a width $\Delta\omega/s$. A more useful characterization, which is independent of the scale of the crystal, is the **gap–midgap ratio**. Letting ω_m be the frequency at the middle of the gap, we define the gap–midgap ratio as $\Delta\omega/\omega_m$, generally expressed as a percentage (e.g., a "10% gap" refers to a gap–midgap ratio of 0.1). If the system is scaled up or down, all of the frequencies scale accordingly, but the gap–midgap ratio remains the same. Thus, when we refer to the "size" of a gap, we are generally referring to the gap–midgap ratio. For the same reason, in the band diagrams in figure 2, as well as all of the other band diagrams in this book, the frequency and wave vector are plotted in *dimensionless* units $\omega a/2\pi c$ and $ka/2\pi$. The dimensionless frequency is equivalent to a/λ, where λ is the vacuum wavelength (given by $\lambda = 2\pi c/\omega$).

We are emphasizing general principles of periodic systems that will apply equally well to the more complicated two- and three-dimensional structures of the later chapters. It is worthwhile, however, to point out a few exceedingly useful *analytical* results that are only possible for the special case of one-dimensional problems.

[4] There is a special exception for the *quarter-wave stack* described in the next section. In that case, while there is always a gap at the edge of the Brillouin zone, there is no gap at the center, because every successive pair of bands is degenerate at $k = 0$.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

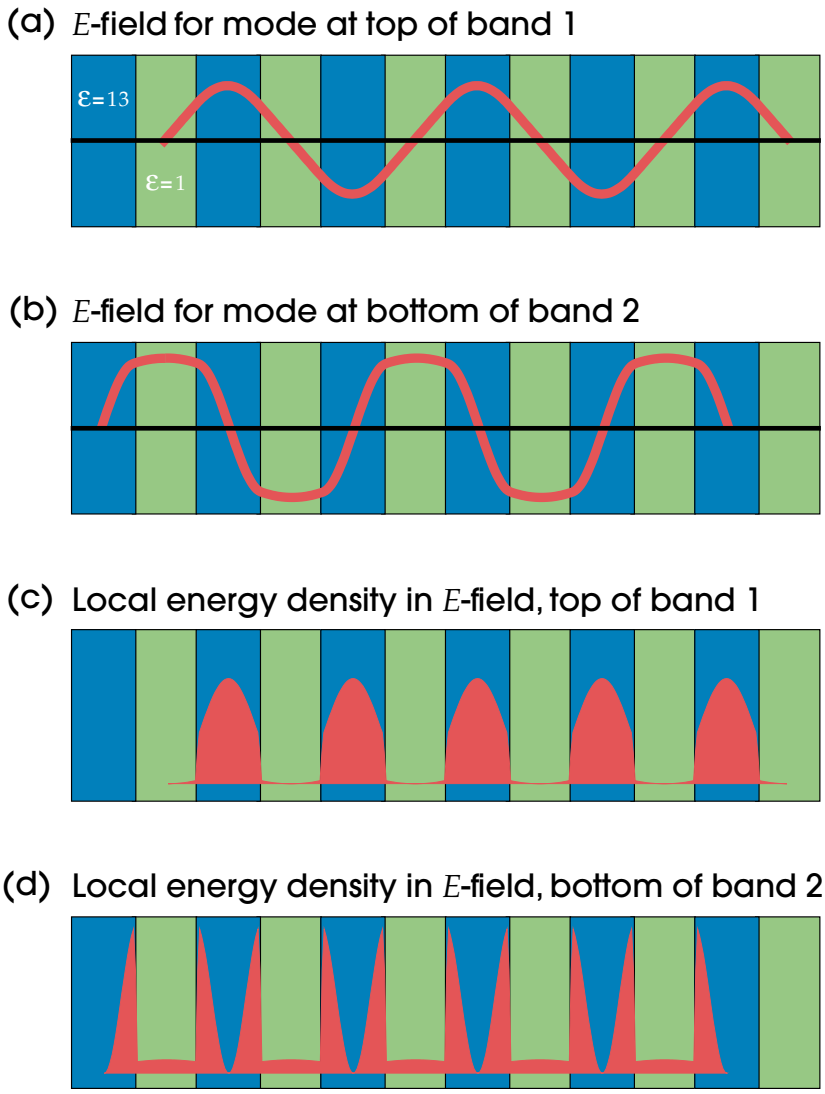


Figure 4: The modes associated with the lowest band gap that is shown in the band structure of the right-hand panel of figure 2, at $k = \pi/a$. The situation is similar to that of figure 3, but the dielectric contrast is larger. The blue and green regions correspond to ε of 13 and 1, respectively.

In a multilayer film with *weak* periodicity, we can derive a simple formula for the size of the band gap from the perturbation theory of the section The Effect of Small Perturbations of chapter 2. Suppose that the two materials in a multilayer film have dielectric constants ε and $\varepsilon + \Delta\varepsilon$, and thicknesses $a - d$ and d. If *either* the dielectric contrast is weak ($\Delta\varepsilon/\varepsilon \ll 1$) *or* the thickness d/a is small, then the

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

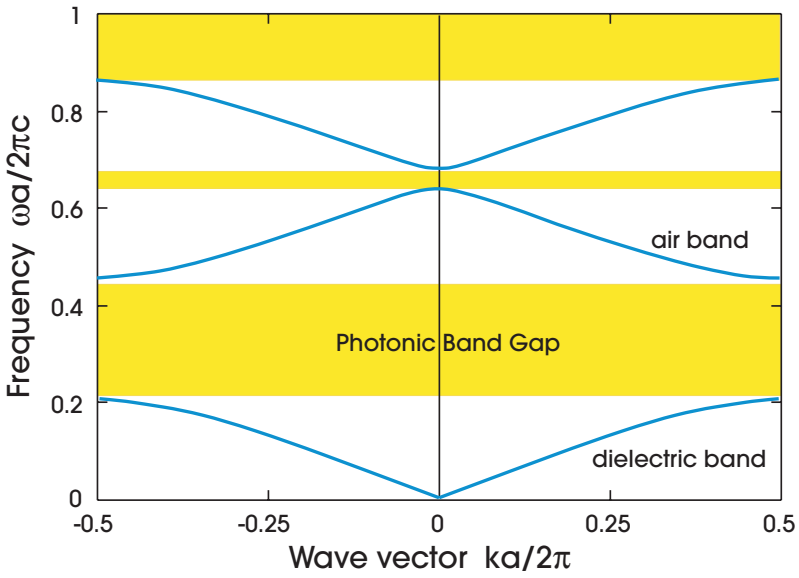


Figure 5: The photonic band structure of a multilayer film with lattice constant a and alternating layers of different widths. The width of the $\varepsilon = 13$ layer is $0.2a$, and the width of the $\varepsilon = 1$ layer is $0.8a$.

gap–midgap ratio between the first two bands is approximately

$$\frac{\Delta\omega}{\omega_{\mathrm{m}}} \approx \frac{\Delta\varepsilon}{\varepsilon} \cdot \frac{\sin(\pi d/a)}{\pi}. \tag{3}$$

This quantifies our previous statement that even an arbitrarily weak periodicity gives rise to a band gap in a one-dimensional crystal. For one of the structures considered in the previous section, with $\Delta\varepsilon/\varepsilon = 1/12$ and $d = 0.5a$ (see the center panel of figure 2), the perturbative formula (3) predicts a 2.65% gap, which is in good agreement with the results of a more accurate numerical calculation (2.55%).

Equation (3) would predict that the gap–midgap ratio is maximized for $d = 0.5a$, but this is valid only for small $\Delta\varepsilon/\varepsilon$. More generally, one can obtain a number of analytical results for arbitrary $\Delta\varepsilon/\varepsilon$, which we summarize here and are derived in, for example, Yeh (1988). For two materials with refractive indices ($\sqrt{\varepsilon}$) n_1 and n_2 and thicknesses d_1 and $d_2 = a - d_1$, respectively, the normal-incidence gap is maximized when $d_1 n_1 = d_2 n_2$, or, equivalently, $d_1 = a n_2/(n_1 + n_2)$. In this specific case, it can be shown that the midgap frequency ω_{m} is

$$\omega_{\mathrm{m}} = \frac{n_1 + n_2}{4 n_1 n_2} \cdot \frac{2\pi c}{a}. \tag{4}$$

The corresponding vacuum wavelength $\lambda_{\mathrm{m}} = 2\pi c/\omega_{\mathrm{m}}$ satisfies the relations $\lambda_{\mathrm{m}}/n_1 = 4d_1$ and $\lambda_{\mathrm{m}}/n_2 = 4d_2$, which means that the individual layers are exactly

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

a quarter-wavelength in thickness. For this reason, this type of multilayer film is called a **quarter-wave stack**. The reason why the gap is maximized for a quarter-wave stack is related to the property that the reflected waves from each layer are all exactly in phase at the midgap frequency. For the gap between the first two bands of a quarter-wave stack, the gap–midgap ratio is

$$\frac{\Delta\omega}{\omega_m} = \frac{4}{\pi}\sin^{-1}\left(\frac{|n_1 - n_2|}{n_1 + n_2}\right). \tag{5}$$

Returning to figure 2, the case that is shown in the right-hand panel is a multilayer film with a dielectric contrast of 13:1 and $d_1 = d_2 = 0.5a$, which is *not* a quarter-wave stack. Numerically, we find that this structure produces a 51.9% gap. If instead we had chosen $d_1 \approx 0.217$, the structure would be a quarter-wave stack with a 76.6% gap, as computed from equation (5). Figure 5 shows the results for $d_1 = 0.2a$, which is nearly a quarter-wave stack, and has a computed band gap of 76.3%. Note also the small gap at $k = 0$, which would go to zero for a quarter-wave stack.

Evanescent Modes in Photonic Band Gaps

The key observation of the section The Physical Origin of Photonic Band Gaps was that the periodicity of the crystal induced a gap in its band structure. No electromagnetic modes are allowed to have frequencies in the gap. But if this is indeed the case, what happens when we send a light wave (with a frequency in the photonic band gap) onto the face of the crystal from outside?

No purely real wave vector exists for any mode at that frequency. Instead, the wave vector is complex. The wave amplitude decays exponentially into the crystal. When we say that there are no states in the photonic band gap, we mean that there are no *extended* states like the mode given by equation (1). Instead, the modes are **evanescent**, decaying exponentially:

$$\mathbf{H}(\mathbf{r}) = e^{ikz}\mathbf{u}(z)e^{-\kappa z}. \tag{6}$$

They are just like the Bloch modes we constructed in equation (1), but with a complex wave vector $k + i\kappa$. The imaginary component of the wave vector causes the decay on a length scale of $1/\kappa$.

We would like to understand how these evanescent modes originate, and what determines κ. This can be accomplished by examining the bands in the immediate vicinity of the gap. Return to the right-hand plot of figure 2. Suppose we try to approximate the second band near the gap by expanding $\omega_2(k)$ in powers of k about the zone edge $k = \pi/a$. Because of time-reversal symmetry, the expansion cannot contain odd powers of k, so to lowest order:

$$\Delta\omega = \omega_2(k) - \omega_2\left(\frac{\pi}{a}\right) \approx \alpha\left(k - \frac{\pi}{a}\right)^2 = \alpha(\Delta k)^2, \tag{7}$$

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

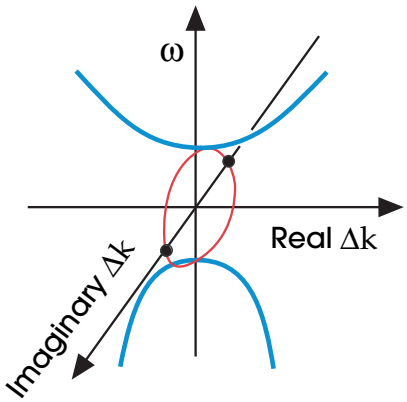


Figure 6: Schematic illustration of the complex band structure of the multilayer film. The upper and lower blue lines correspond to the bottom of band 2 and the top of band 1, respectively. The evanescent states occur on the red line, which extends along the imaginary-k axis running out of the page. The maximum decay occurs roughly at the center of the gap.

where α is a constant depending on the curvature of the band (i.e., the second derivative).

Now we can see where the complex wave vector originates. For frequencies slightly higher than the top of the gap, $\Delta\omega > 0$. In this case, Δk is purely real, and we are within band 2. However, for $\Delta\omega < 0$, when we are within the gap, Δk is purely imaginary.[5] The states decay exponentially since $\Delta k = i\kappa$. As we traverse the gap, the decay constant κ grows as the frequency reaches the gap's center, then disappears again at the lower gap edge. This behavior is depicted in figure 6. By the same token, larger gaps usually result in a larger κ at midgap, and thus less penetration of light into the crystal; for a multilayer film, minimal penetration is therefore achieved in the quarter-wave stack described by the preceding section.

Although evanescent modes are genuine solutions of the eigenvalue problem, they diverge as z goes to $\pm\infty$ (depending on the sign of κ). Consequently, there is no physical way to excite them within an idealized crystal of infinite extent. However, a defect or an edge in an otherwise perfect crystal can terminate this exponential growth and thereby sustain an evanescent mode. If one or more evanescent modes is compatible with the structure and symmetry of a given crystal defect, we can then excite a *localized* mode within the photonic band gap. And, as a general rule of thumb, we can localize states near the middle of the gap much more tightly than states near the gap's edge.[6]

[5] Technically, we are exploiting concepts from complex analysis. The eigenvalues ω are generally analytic functions of any smooth parameter of the operator $\hat{\Theta}_{\mathbf{k}}$, so we can use the analytic continuation of $\omega(\mathbf{k})$ into the complex $\mathbf{k}$ domain via its Taylor expansion.

[6] There are subtle exceptions to this rule. For example, with certain band structures in two and three dimensions, saddle points in the bands can lead to strong localization away from midgap (Ibanescu et al., 2006).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

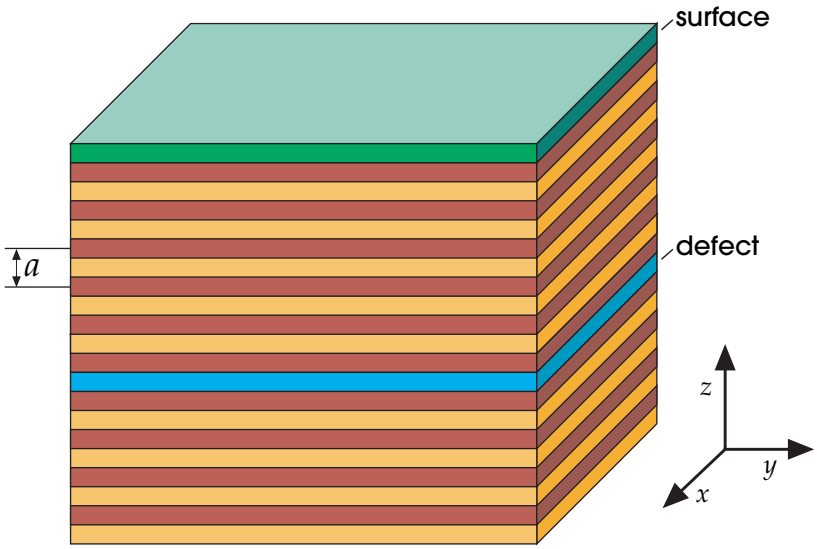


Figure 7: Schematic illustration of possible sites of localized states for a one-dimensional photonic crystal. The states are planar and would be localized differently near the differently colored regions, which break the symmetry in the z direction. We will call a mode at the edge of the crystal (green) a **surface state**, and a mode within the bulk of the crystal (blue) a **defect state**.

Of course, one-dimensional photonic crystals can localize states only in one dimension. The state is confined to a given *plane*, as shown in figure 7. In the section Localized Modes at Defects, we will discuss the nature of such states when they lie deep within the bulk of a photonic crystal. In certain circumstances, however, an evanescent mode can exist at the face of the crystal. We will also discuss these **surface states** in the section Surface States.

Off-Axis Propagation

So far, we have considered the modes of a one-dimensional photonic crystal which happen to have $\mathbf{k}_{\parallel} = 0$; that is, modes that propagate only in the z direction. In this section we will discuss off-axis modes. Figure 8 shows the band structure for modes with $\mathbf{k} = k_y\hat{\mathbf{y}}$ for the one-dimensional photonic crystal described in the caption of figure 5.

The most important difference between on-axis and off-axis propagation is that there are *no band gaps* for off-axis propagation when all possible k_y are included. This is always the case for a multilayer film, because the off-axis direction contains no periodic dielectric regions to coherently scatter the light and split open a gap. (Despite this, we will see in the section Omnidirectional Multilayer Mirrors that it is

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

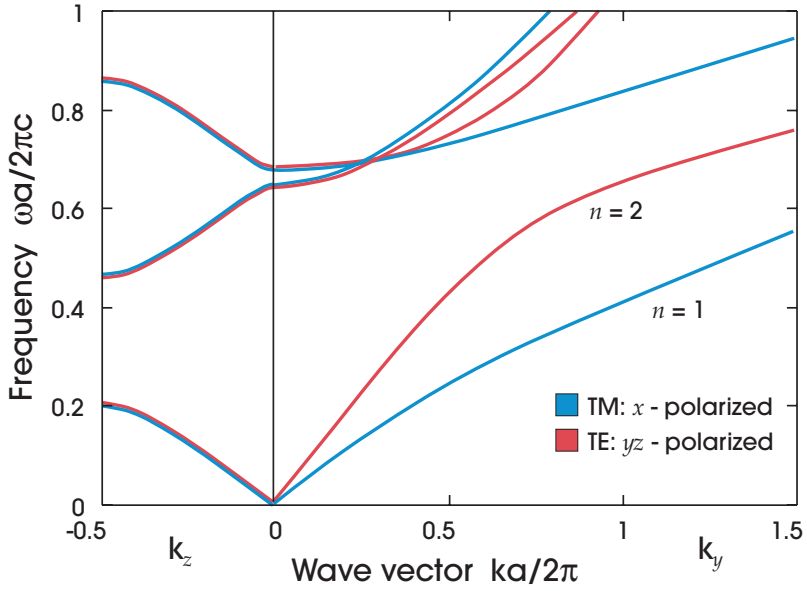


Figure 8: The band structure of a multilayer film. The on-axis bands $(0, 0, k_z)$ are shown on the left side, and an off-axis band structure $(0, k_y, 0)$ is displayed on the right. On-axis, the bands overlap—they are degenerate. Along k_y, the bands split into two distinct polarizations. Blue indicates TM modes polarized so that the electric field points in the x direction, and red indicates TE modes polarized in the yz plane. The layered structure is the same as the one described in the caption of figure 5.

still possible to design a multilayer film that reflects *external* plane waves that are incident from any angle.)

Another difference between the on-axis and off-axis cases involves the *degeneracy* of the bands. For the case of on-axis propagation, the electric field is oriented in the xy plane. We might choose the two basic polarizations as the x and y directions. Since those two modes differ only by a rotational symmetry which the crystal possesses, they must be degenerate. (How could the crystal distinguish between the two polarizations?)

However, for a mode propagating with an arbitrary direction of $\mathbf{k}$, this symmetry is broken. The degeneracy is lifted. There are other symmetries; for example, notice that the system is invariant under reflection through the yz plane. For the special case of propagation down the dielectric sheets, in the y direction, we know from the symmetry discussion of chapter 3 that the possible polarizations are in the x direction (TM) or in the yz plane (TE). But there is no rotational symmetry relationship between these two bands, so they will generally have different frequencies. All of these phenomena are displayed in figure 8.

Although $\omega(\mathbf{k})$ for the two different polarizations have different slopes, both are approximately linear at long wavelengths ($\omega \to 0$). This long-wavelength

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

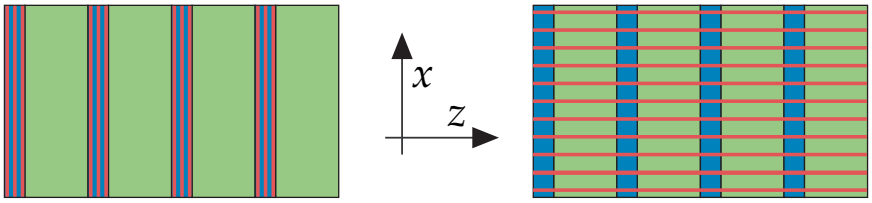


Figure 9: A sketch of the displacement field lines for a long-wavelength mode traveling in the y direction (out of the page). In the left figure, the fields are oriented along x. In the right figure, the fields are oriented primarily along z. The blue regions correspond to high ε.

behavior is characteristic of *all* photonic crystals, regardless of geometry or dimensionality:

$$\omega_\nu(\mathbf{k}) = c_\nu(\hat{\mathbf{k}})k. \tag{8}$$

Here ν is a label that refers to one of the two possible polarizations or, equivalently, one of the first two bands. In general, c_ν will depend on both the direction of $\mathbf{k}$ and on ν.

Why is the dispersion relation always linear at long wavelengths? At long wavelengths, the electromagnetic wave doesn't probe the fine structure of the crystal lattice. Instead, the light effectively sees a homogeneous dielectric medium, with an effective dielectric constant that is a weighted average over all of the "microscopic" variations in ε.

In many cases, the averaged dielectric constant will be a function of the polarization (the direction of $\mathbf{E}$). In those cases, the effective medium is *anisotropic* and the dielectric function is a 3-by-3 matrix, a tensor. The "principal axes" are the basic symmetry axes for which the dielectric tensor is diagonalized. Procedurally, we can imagine measuring the effective dielectric constant along each of three directions by applying a static field in a capacitance measurement. A general analytical expression for the effective dielectric constants of a photonic crystal is not available, but the constants can be calculated numerically.[7]

Returning to the multilayer film, we would like to understand why modes polarized in the x direction (band 1 in figure 8) have a lower frequency than modes polarized in the yz plane (band 2). Once again we use our heuristic: the lower modes concentrate their electrical energy in the high-ε regions. In this case, we focus on the long-wavelength limit of each mode. The fields for both bands are shown schematically in figure 9. For the x-polarized wave, the displacement fields lie in the high-ε regions. But at long wavelengths, the polarization of

[7] One useful analytical constraint for the effective dielectric constant of a general photonic crystal is provided by the Weiner bounds, as in Aspnes (1982). Specifically, for a two-material composite, each effective dielectric constant ε_α is bounded by

$$(f_1\varepsilon_1^{-1} + f_2\varepsilon_2^{-1})^{-1} \le \varepsilon_\alpha \le f_1\varepsilon_1 + f_2\varepsilon_2$$

where f_1 and f_2 are the volume fractions of the materials with dielectric constants ε_1 and ε_2.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

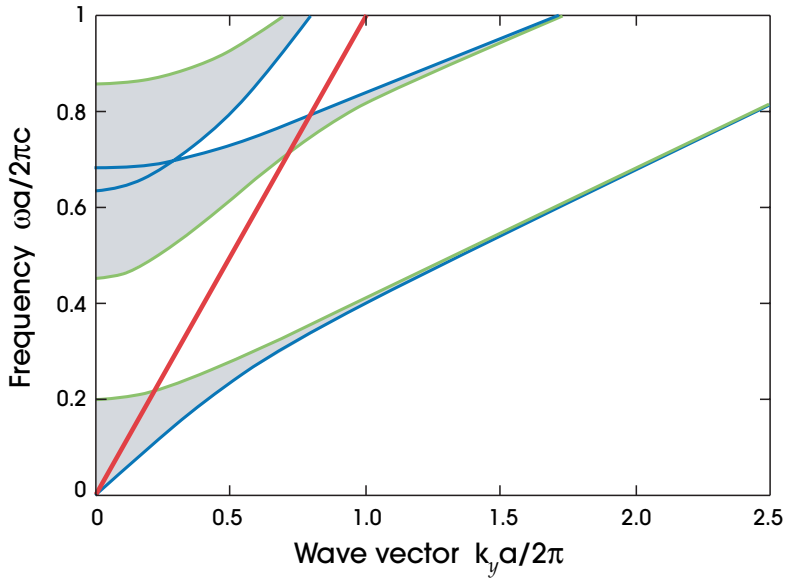


Figure 10: Two superimposed x-polarized band structures of a multilayer film, showing how the bandwidths vary with k_y. The blue lines refer to bands along $(0, k_y, 0)$, while the green lines beside them refer to the same bands along $(0, k_y, \pi/a)$. The regions in between are shaded gray to indicate where the continuum of bands for intermediate k_z would lie. Only modes with electric field oriented along the x direction are shown. The straight red line is the **light line** $\omega = ck_y$. The layered material is the same as the one described in the caption of figure 5.

band 2 is almost entirely along the z direction, crossing both the low-ε and the high-ε regions. Continuity forces the field to penetrate the low-ε region, leading to a higher frequency.

We can also understand the short-wavelength, large-k limit of the band structure. In figure 8, notice that the width of each band is determined by the difference between frequencies at the zone center $(k_z = 0)$ and the zone edge $(k_z = \pi/a)$. As k_y is increased, the bandwidths decrease to zero. This is illustrated in figure 10, which shows the superposition of two band structures. The blue lines represent states along $\mathbf{k} = (0, k_y, 0)$ and the green lines represent states along $\mathbf{k} = (0, k_y, k_z = \pi/a)$. As we saw for the case of a plane of glass, once the frequency is below the light line $\omega = ck_y$, the modes are index guided and decay exponentially into the air region. As k_y is increased, the overlap between modes within neighboring layers of high-ε material goes exponentially to zero. The coupling between neighboring planes becomes small, and each plane essentially guides its own mode independent of its neighbors.[8] In this case, the mode

[8] In solid-state physics, the analogous system is the tight-binding model in the limit of small hopping. See, for example, Harrison (1980).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

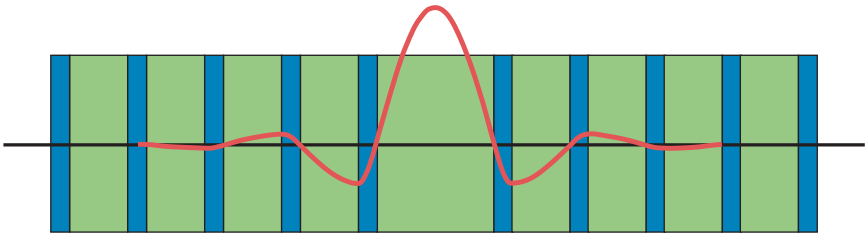


Figure 11: A defect in a multilayer film, formed by doubling the thickness of a single *low-ε* layer in the structure of figure 5. Note that this can be considered to be an interface between two perfect multilayer films. The red curve is the electric-field strength of the defect state associated with this structure (for on-axis propagation).

frequency becomes independent of the on-axis wave vector, and every mode in the band becomes the frequency of a guided mode that is trapped by the high-ε layers.

Localized Modes at Defects

Now that we understand the band structure of a perfectly periodic system, we can examine systems in which the translational symmetry has been broken by a defect. Suppose that the defect consists of a single layer of the one-dimensional photonic crystal that has a different width than the rest. Such a system is shown in figure 11. We no longer have a perfectly periodic lattice. However, we expect intuitively that if we move many wavelengths away from the defect, the modes should look similar to the corresponding modes of a perfect crystal.

For now, we restrict our attention to on-axis propagation and consider a mode with frequency ω in the photonic band gap. There are no extended modes with frequency ω inside the periodic lattice. Introducing the defect does not change that fact. The destruction of periodicity prevents us from describing the modes of the system with wave vector k, but we can still employ our knowledge of the band structure to determine whether a certain frequency will support extended states inside the rest of the crystal. In this way, we can divide up the frequency range into regions in which the states are extended and regions in which they are evanescent, as in figure 12.

Defects may permit *localized* modes to exist, with frequencies inside photonic band gaps. If a mode has a frequency in the gap, then it must exponentially decay once it enters the crystal. The multilayer films on both sides of the defect behave like frequency-specific mirrors. If two such films are oriented parallel to one another, any z-propagating light trapped between them will just bounce back and forth between these two mirrors. And because the mirrors localize light within a finite region, the modes are *quantized* into discrete frequencies, as described in the section Discrete vs. Continuous Frequency Ranges of chapter 2.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

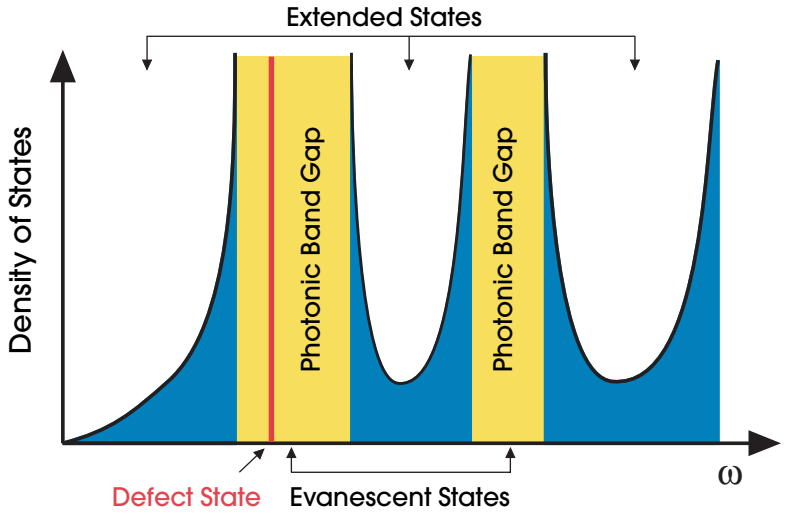


Figure 12: The division of frequency space into extended and evanescent states. In this sketch, the **density of states** (the number of allowed modes per unit frequency) is zero in the band gaps of the crystal (yellow). Modes are allowed to exist in these regions only if they are evanescent, and only if the translational symmetry is broken by a defect. Such a mode is shown in red.

Consider the family of localized states generated by continuously increasing the thickness of the defect layer (shifting the crystal to either side of the defect). The bound mode(s) associated with each member of this family will have a different frequency. As the thickness is increased, the frequency will decrease, because the mode has more space to oscillate; this decreases the numerator of the variational theorem and thus the frequency (all other things equal). In fact, as the thickness increases, a *sequence* of discrete modes are pulled down into the gap from the upper bands. The first such mode is shown in figure 11, pulled down by doubling the thickness of a low-ε layer. On the other hand, if we were to keep the thickness fixed and either increase or decrease the dielectric constant ε of a single layer, we would decrease or increase the frequency, respectively, due to the change in the denominator of the variational theorem. So, in general, a defect may either pull modes down into the gap from the upper bands, or push modes up into the gap from the lower band. Moreover, the degree of localization of the defect mode will be greatest when the frequency is near the center of the gap, as shown in figure 6. States with frequencies in the center of the gap will be most strongly attached to the defect.

The **density of states** of a system is the number of allowed states per unit interval of ω. If a single state is introduced into the photonic band gap, then the density of states of the system in figure 12 is zero in the photonic band gap,

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

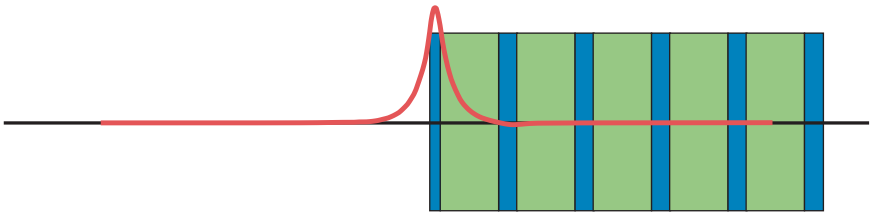


Figure 13: The electric field strength associated with a localized mode at the surface of a multilayer film. In particular, the mode at $k_y = 2\pi/a$ from figure 14 is shown. (This mode actually oscillates in sign with each period of the crystal, but with an amplitude too small to see clearly here.)

except for a single peak associated with the defect.[9] This property is exploited in the bandpass filter known as the **dielectric Fabry–Perot filter**, a case that will be discussed in chapter 10. It is particularly useful at visible-light frequencies because of the relatively low losses of dielectric materials.

Similarly, if we include off-axis wave vectors, we obtain states that are localized in the z direction, but that propagate (are *guided*) along the interface ($k_z = i\kappa, \mathbf{k}_\parallel \neq 0$). These guided modes, forming a planar waveguide, can differ markedly from the conventional solution of total internal reflection, discussed in the subsection Index guiding of chapter 3; for example, as in figure 11, they can be guided in a *lower-*ε region.[10] This idea can be generalized to include the case of an interface between two different multilayer films with different spatial periods. Localized states can exist as long as the band gaps of the two photonic crystals overlap.

Surface States

We have seen under what conditions we can localize electromagnetic modes at defects in a multilayer film. In a similar fashion, we can also localize modes at its *surface*, called **surface states**. In the previous section, the mode was bound because its frequency was within the photonic band gap of the films on both sides. But at the surface, there is a band gap on only one side of the interface. The exterior medium (air, in our typical example) does not have a band gap.

In this case, light is bound to the surface if its frequency is below the light line. Such a wave is index-guided, a phenomenon that is a generalization of total internal reflection (see the subsection Index guiding of chapter 3). An example is depicted in figure 13. At a surface, we must consider whether the modes are extended or localized in both the air region and the layered material, and we

[9] This peak is described by a Dirac delta function in the density of states.

[10] This was pointed out by Yeh and Yariv (1976). We return to an analogous three-dimensional structure, the Bragg fiber, in chapter 9.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

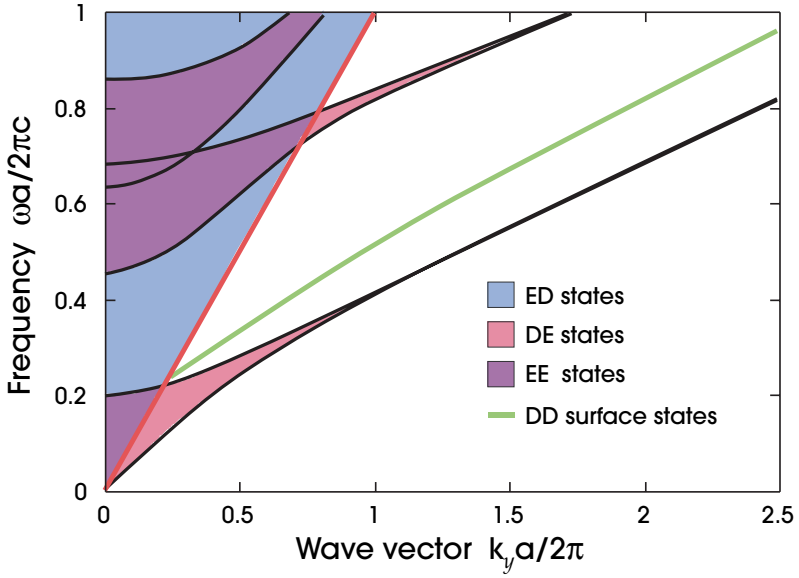


Figure 14: The x-polarized band structure at the surface of a multilayer film. The shaded regions describe states that are extended in the air region (blue, ED), in the layered material (pink, DE), or in both (purple, EE). The green (DD) line represents a band of surface states confined at the interface. The layered material is the same as the one described in figure 5. The surface is terminated by a layer of high dielectric with a width of $0.1a$ (half of its usual thickness), as shown in figure 13.

must consider all possibilities for $\mathbf{k}_{\|}$. The appropriate band structure is shown in figure 14 (restricting ourselves to $\hat{\mathbf{x}}$ polarizations as in figure 10). We divide the band structure into four regions, depending on whether they are localized or extended in the air and crystal regions. For example, the label "DE" means that modes in that region *Decay* in the air region, and are *Extended* in the crystal region.

The EE modes are spatially extended on both sides of the surface, the DE modes decay into the air region and are extended into the crystal, and the ED modes are extended in the air region, but decay inside the crystal. Only when a mode is evanescent on both sides of the surface—it is both in a band gap and below the light line—can we have a surface wave. This is possible in any of the white regions of figure 14, and one such localized mode is labelled DD. In fact, *every* periodic material has surface modes for *some* choice of termination, a phenomenon we shall discuss again in the next two chapters.

Omnidirectional Multilayer Mirrors

As described in the section Off-Axis Propagation, a multilayer film does not have a complete band gap, once one allows for a component of the wave vector that is

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

parallel to the layers. Another way to state this is that for every choice of ω, there exist extended modes in the film for *some* wave vectors $(\mathbf{k}_\parallel, k_z)$. Given this fact, it may seem paradoxical that a properly designed multilayer structure can still reflect light waves that are incident from *any* angle, with *any* polarization, if it has a frequency that is within a specified range.

Such a device, an **omnidirectional mirror**, relies on two physical properties. First, $\mathbf{k}_\parallel$ is conserved at any interface parallel to the layers, if the light source is far enough away that it does not interrupt the translational symmetry of the structure in the $\parallel$ direction. Second, light that is incident from air must have $\omega > c|\mathbf{k}_\parallel|$, corresponding to the freely propagating modes above the light line, just as we saw for index guiding and surface states (see the subsection Index guiding of chapter 3 and the section Surface States). Modes that are below the light line are evanescent modes that cannot reach the mirror from a faraway source. Because of these two properties, the modes that the crystal harbors below the light line are irrelevant for the purpose of reflection.

To investigate omnidirectional reflection, we plot ω vs. k_y, just as we did in figure 14. The yz plane is the plane of incidence, with the y direction parallel to the layers, and the z direction perpendicular to the layers. Now, however, we must consider *both* of the possible polarization states: TM, in which the electric field is perpendicular to the plane of incidence (as in figure 14); and TE, in which the electric field is within the plane of incidence. In this context, it is common to refer to TM modes as *s*-polarized and TE modes as *p*-polarized. Our example structure, here, is a quarter-wave stack (see the section The Size of the Band Gap) consisting of layers with ε of 13 and 2 , rather than 13 and 1 as before. The resulting band diagram is shown in figure 15. Indeed, there is a range of frequencies (yellow) within which all of the modes of the multilayer film are below the light line. Within this range, any incident plane waves cannot couple to the extended states in the layers. Instead, their fields must decay exponentially within the quarter-wave stack. The transmission through such a mirror will drop exponentially with the number of layers: the light will be perfectly reflected, insofar as material absorption can be neglected.

Omnidirectional reflection is not a *general* property of one-dimensional photonic crystals. There are two necessary conditions. First, the dielectric contrast between the two mirror materials must be sufficiently large so that the point labelled U (upper) is above the point L (lower) in figure 15. If the band gap is too narrow, we will hit the top U of the $k_y = 0$ gap before the bottom of the gap has exited the light cone at L. Second, the *smaller* of the two dielectric constants (ε_1) must be larger than the dielectric constant of the ambient medium (ε_a) by a critical amount. This critical contrast with the ambient medium is reached when the p (TE) bands are pulled *down* in frequency far enough that point B in figure 15 (where the first and second p bands intersect) does not fall above the light line. This second criterion is why we chose $\varepsilon_1 = 2$ rather than 1 in our example.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

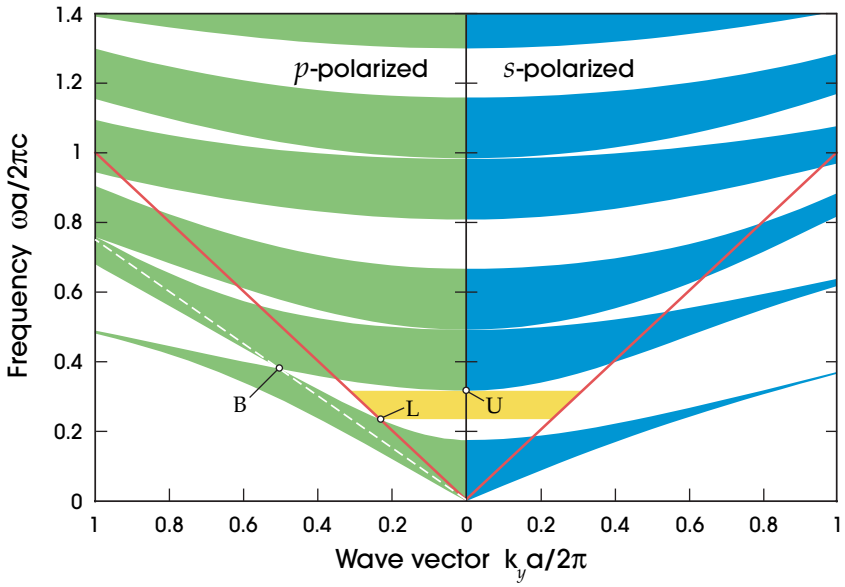


Figure 15: Extended modes (shaded regions) for off-axis propagation vectors $(0, k_y, k_z)$ in a quarter-wave stack with ε of 13 and 2. The right side (blue) indicates modes with $\mathbf{E}$ fields polarized in the x direction (TM or s-polarized), similar to figure 10. The left side (green) indicates modes with fields polarized in the yz incidence plane (TE or p-polarized). The straight red line is the **light line** $\omega = ck_y$, above which extended modes exist in air. In yellow is shaded the first frequency range of **omnidirectional reflection** (with lower and upper edges at L and U, respectively). The dashed white line corresponds to Brewster's angle, which gives rise to the crossing at B.

The B point falls on a line that corresponds to **Brewster's angle**, at which p-polarized light has no reflection at the $\varepsilon_1/\varepsilon_2$ interface.[11] The lack of reflection is what permits the bands to intersect. Combining these two criteria, figure 16 shows the size of the "omnidirectional gap" as a function of the ratios $\sqrt{\varepsilon_2/\varepsilon_1}$ and $\sqrt{\varepsilon_1/\varepsilon_a}$, for the case of quarter-wave stacks. Strictly speaking, a quarter-wave stack does *not* maximize the size of the omnidirectional gap, but in practice it very nearly does so. In figure 16, had we used the optimal layer spacings instead of the quarter-wave spacings, the contours would be displaced by less than about 2% along either axis.

A suitably designed multilayer film can therefore function as an omnidirectional mirror, but there are some things it cannot do. Its reflective property depends on the translational symmetry of the interface, and consequently it cannot

[11] The existence of Brewster's angle (see, e.g., Jackson, 1998) is related to the observation that light reflected obliquely from water, glass, or even asphalt is preferentially s-polarized. This, in turn, is why polarizing sunglasses are effective at reducing glare from the road.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

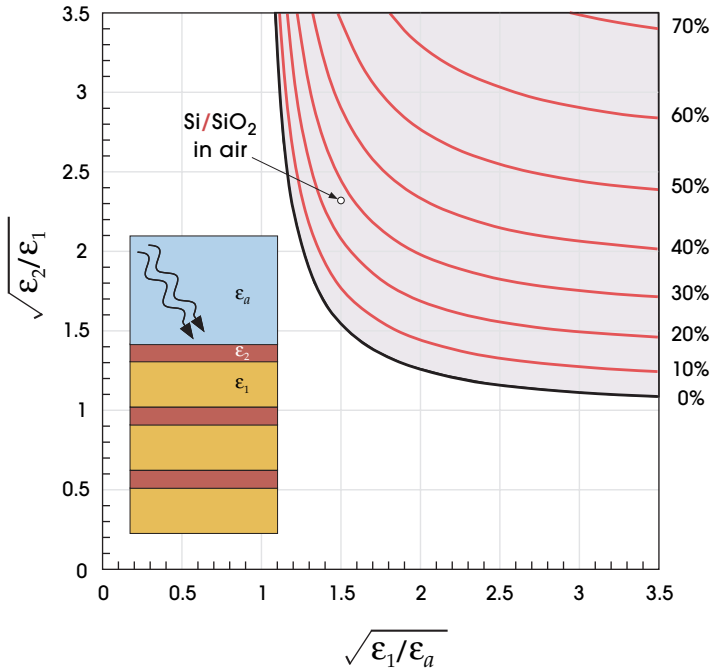


Figure 16: The size of the omnidirectional gap as a function of the dielectric constants of the layers, for a quarter-wave stack. The gap–midgap ratios are labelled on the right side of each curve. The system is illustrated in the inset, in which light is incident from an ambient medium with dielectric constant ε_a. The two materials of the film have dielectric constants ε_1 and ε_2, with $\varepsilon_1 < \varepsilon_2$. It does not matter which material forms the edge of the mirror. The pink shaded area is the region in which there is a nonzero omnidirectional gap. Some common materials, such as the silicon/silica/air combination indicated by the arrow, fall within this region.

confine a mode in three dimensions. In addition, if the interface is not flat, or if there is an object (or a light source) close to the surface, then $\mathbf{k}_\parallel$ is not conserved. In that case, light will generally couple to extended modes propagating in the mirror and will be transmitted. There is an interesting exception, however: if the mirror is curved around a hollow sphere or cylinder, then the *continuous rotational* symmetry can substitute for translational symmetry, and light can be localized within the core. As with the planar mirror, the leakage rate from the core to the exterior decreases exponentially with the number of layers. The cylindrical case (discussed further in chapter 9) was called a **Bragg fiber** by Yeh et al. (1978), and the spherical case has been dubbed a **Bragg onion** by Xu et al. (2003). Here, one did not require omnidirectional mirrors to obtain localized modes, because a mode's rotational symmetry imposes restrictions on the angles that it can escape into at large radii.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Further Reading

Many of the theorems we have developed and the properties we have observed for photonic crystals have analogues in quantum mechanics and solid-state physics. For readers familiar with those fields, appendix A provides a comprehensive listing of these analogies.

The conventional treatment of the multilayer film, including the calculation of absorption and reflection coefficients, can be found in Hecht and Zajac (1997). The use of the multilayer films in optoelectronic devices is widespread in current literature. For example, Fowles (1975) outlines their use in Fabry–Perot filters, and Yeh (1988, p. 337) explains how they are incorporated into distributed feedback lasers. A historical review of the study of electromagnetic waves in one-dimensionally periodic structures can be found in Elachi (1976). Despite the long history of multilayer films, the possibility of achieving omnidirectional reflection was not apparent until 1998, when the idea was proposed (Winn et al., 1998) and realized experimentally (Fink et al., 1998).

The details of the computational scheme we used to compute band structures and eigenmodes can be found in Johnson and Joannopoulos (2001), and are also summarized in appendix D along with various alternative methods.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

5

Two-Dimensional Photonic Crystals

Now that we have discussed some interesting properties of one-dimensional photonic crystals, in this chapter we will see how the situation changes when the crystal is periodic in two directions and homogeneous in the third. Photonic band gaps appear in the plane of periodicity. For light propagating in this plane, the harmonic modes can be divided into two independent polarizations, each with its own band structure. As before, we can introduce defects in order to localize light modes, but in this case we can localize a mode in two dimensions, rather than just one dimension.

Two-Dimensional Bloch States

A two-dimensional photonic crystal is periodic along two of its axes and homogeneous along the third axis. A typical specimen, consisting of a square lattice of dielectric columns, is shown in figure 1. We imagine the columns to be infinitely tall; the case of a *finite* extent in the third direction is treated in chapter 8. For certain values of the column spacing, this crystal can have a photonic band gap in the xy plane. Inside this gap, no extended states are permitted, and incident light is reflected. Unlike the multilayer film, this two-dimensional photonic crystal can prevent light from propagating in *any* direction within the plane.

As always, we can use the symmetries of the crystal to characterize its electromagnetic modes. Because the system is homogeneous in the z direction, we know that the modes must be oscillatory in that direction, with no restrictions on the wave vector k_z. In addition, the system has discrete translational symmetry in the xy plane. Specifically, $\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r} + \mathbf{R})$, as long as $\mathbf{R}$ is any linear combination of the primitive lattice vectors $a\hat{\mathbf{x}}$ and $a\hat{\mathbf{y}}$. By applying Bloch's theorem, we can focus our attention on the values of $\mathbf{k}_\parallel$ that are in the Brillouin zone. As before, we use the label n (band number) to label the modes in order of increasing frequency.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

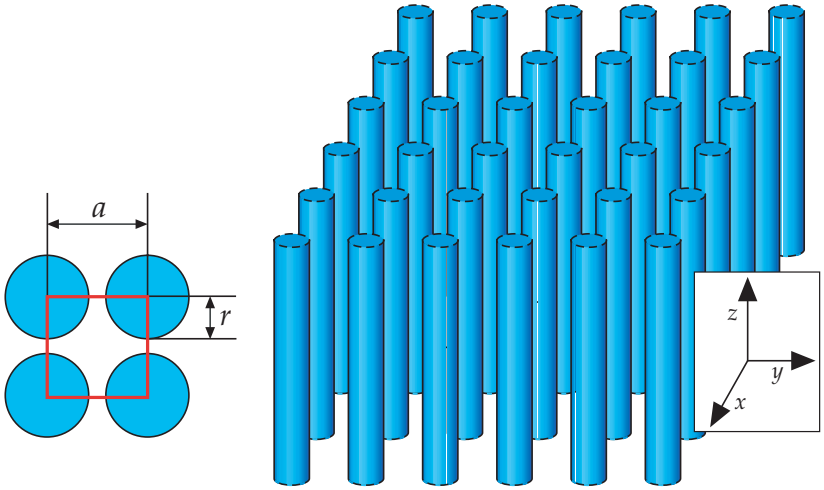


Figure 1: A two-dimensional photonic crystal. This material is a square lattice of dielectric columns, with radius r and dielectric constant ε. The material is homogeneous along the z direction (we imagine the cylinders are very tall), and periodic along x and y with lattice constant a. The left inset shows the square lattice from above, with the unit cell framed in red.

Indexing the modes of the crystal by k_z, $\mathbf{k}_\parallel$, and n, they take the now-familiar form of Bloch states

$$\mathbf{H}_{(n,k_z,\mathbf{k}_\parallel)}(\mathbf{r}) = e^{i\mathbf{k}_\parallel\cdot\boldsymbol{\rho}} e^{ik_z z} \mathbf{u}_{(n,k_z,\mathbf{k}_\parallel)}(\boldsymbol{\rho}). \tag{1}$$

In this equation, $\boldsymbol{\rho}$ is the projection of $\mathbf{r}$ in the xy plane and $\mathbf{u}(\boldsymbol{\rho})$ is a periodic function, $\mathbf{u}(\boldsymbol{\rho}) = \mathbf{u}(\boldsymbol{\rho} + \mathbf{R})$, for all lattice vectors $\mathbf{R}$. The modes of this system look similar to those of the multilayer film that we saw in equation (1) of chapter 4. The key difference is that in the present case, $\mathbf{k}_\parallel$ is restricted to the Brillouin zone and k_z is unrestricted. In the multilayer film, the roles of these two wave vectors were reversed. Also, $\mathbf{u}$ is now periodic in the plane, and not in the z direction as before.

Any modes with $k_z = 0$ (i.e. that propagate strictly parallel to the xy plane) are invariant under reflections through the xy plane. As discussed in chapter 3, this mirror symmetry allows us to classify the modes by separating them into two distinct polarizations. Transverse-electric (TE) modes have $\mathbf{H}$ normal to the plane, $\mathbf{H} = H(\boldsymbol{\rho})\hat{\mathbf{z}}$, and $\mathbf{E}$ in the plane, $\mathbf{E}(\boldsymbol{\rho}) \cdot \hat{\mathbf{z}} = 0$. Transverse-magnetic (TM) modes have just the reverse: $\mathbf{E} = E(\boldsymbol{\rho})\hat{\mathbf{z}}$ and $\mathbf{H}(\boldsymbol{\rho}) \cdot \hat{\mathbf{z}} = 0$.

The band structures for TE and TM modes can be completely different. It is possible, for example, that there are photonic band gaps for one polarization but not for the other polarization. In the coming sections, we will investigate the TE and TM band structures for two different two-dimensional photonic crystals, mainly restricting ourselves to in-plane ($k_z = 0$) propagation. The results will provide some useful insights into the appearance of band gaps.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

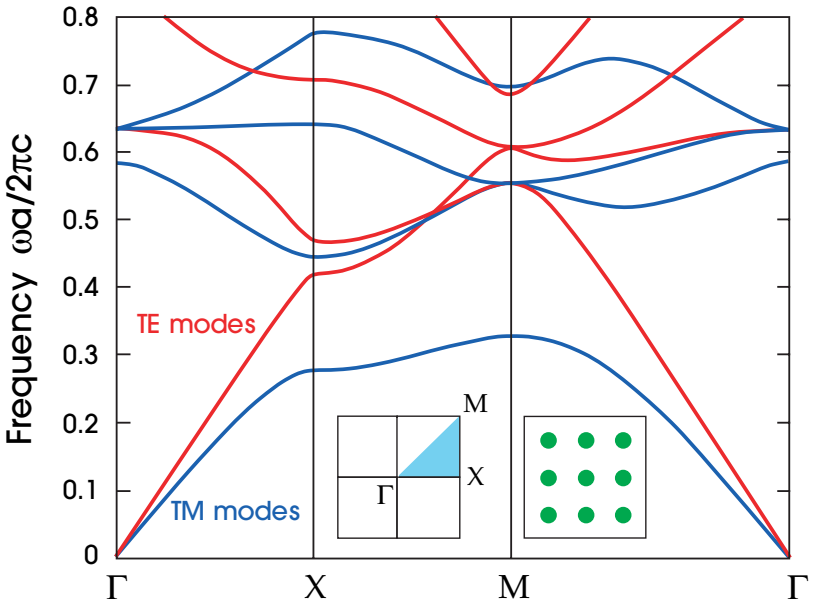


Figure 2: The photonic band structure for a square array of dielectric columns with $r = 0.2a$. The blue bands represent TM modes and the red bands represent TE modes. The left inset shows the Brillouin zone, with the irreducible zone shaded light blue. The right inset shows a cross-sectional view of the dielectric function. The columns ($\varepsilon = 8.9$, as for alumina) are embedded in air ($\varepsilon = 1$).

A Square Lattice of Dielectric Columns

Consider light that propagates in the xy plane of a square array of dielectric columns, such as the structure depicted in figure 1, with lattice constant a. The band structure for a crystal consisting of alumina ($\varepsilon = 8.9$) rods in air, with radius $r/a = 0.2$, is plotted in figure 2. Both the TE and the TM band structures are shown. (As described in the section The Size of the Band Gap of chapter 4, the frequency is expressed as a dimensionless ratio $\omega a/2\pi c$.) The horizontal axis shows the value of the in-plane wave vector $\mathbf{k}_\parallel$. As we move from left to right, $\mathbf{k}_\parallel$ moves along the triangular edge of the irreducible Brillouin zone, from Γ to X to M, as shown in the inset to figure 2.

Since this is the first photonic crystal we have encountered that exhibits a complicated band structure, we will discuss it in detail. Specifically, we will describe the nature of the modes when $\mathbf{k}_\parallel$ is right at the special symmetry points of the Brillouin zone, and we will investigate the appearance of the band gaps. The reason why we have plotted $\mathbf{k}_\parallel$ only along the edge of the Brillouin zone is that the minima and the maxima of a given band (which determine the band gap) almost always occur at the zone edges, and often at a corner. While this is not guaranteed,

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

it is true in most cases of interest, such as all of the structures discussed in this chapter.

The square lattice array has a square Brillouin zone, which is illustrated in the inset panel of figure 2. The irreducible Brillouin zone is the triangular wedge in the upper-right corner; the rest of the Brillouin zone can be related to this wedge by rotational symmetry. The three special points Γ, X, and M correspond (respectively) to $\mathbf{k}_{\|} = 0$, $\mathbf{k}_{\|} = \pi/a\hat{\mathbf{x}}$, and $\mathbf{k}_{\|} = \pi/a\hat{\mathbf{x}} + \pi/a\hat{\mathbf{y}}$. What do the field profiles of the electromagnetic modes at these points look like?

The field patterns of the TM modes of the first band (*dielectric band*) and second band (*air band*) are shown in figure 3. For modes at the Γ point, the field pattern is exactly the same in each unit cell. For modes at the X point (the zone edge), the fields alternate in sign in each unit cell along the direction of the wave vector k_x, forming wave fronts parallel to the y direction. For modes at the M point, the signs of the fields alternate in neighboring cells, forming a checkerboard pattern. Although the X and M patterns may look like wave fronts of a propagating wave, in fact the modes at these particular $\mathbf{k}$ points do not propagate at all—they are **standing waves** with zero group velocity. The field patterns of the TE modes at the X point for the first and second bands are shown in figure 4.

For the TM modes, this photonic crystal has a complete band gap between the first and second bands, with a 31.4% gap–midgap ratio (defined as in the section The Size of the Band Gap of chapter 4). In contrast, for the TE modes there is no complete band gap. We should be able to explain such a significant fact, and we can, by examining the field patterns in figures 3 and 4. The field associated with the lowest TM mode (the dielectric band) is strongly concentrated in the dielectric regions. This is in sharp contrast to the field pattern of the air band. There, a nodal plane cuts through the dielectric columns, expelling some of the displacement field amplitude from the high-ε region.

As we found in chapter 2, a mode concentrates most of its electric-field energy in the high-ε regions in order to lower its frequency, but upper bands must be orthogonal to lower bands. This statement of the variational theorem explains the large splitting between these two bands. The first band has most of its energy in the dielectric regions, and has a low frequency; the second must have a nodal plane in order to be orthogonal to the first, and thus has most of its energy in the air region with a correspondingly higher frequency. We can quantify this statement. An appropriate measure of the degree of concentration of the electric fields in the high-ε regions is the **concentration factor**, which we define as

$$\text{concentration factor} \triangleq \frac{\int_{\varepsilon=8.9} d^3\mathbf{r}\,\varepsilon(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2}{\int d^3\mathbf{r}\,\varepsilon(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2}. \tag{2}$$

Comparing this expression with equation (24) of chapter 2, we see that the concentration factor measures the fraction of electric-field energy located inside the high-ε regions. Table 1 shows the concentration factors for the fields we are considering. The dielectric-band TM mode has a concentration factor of 83%, while the air-band TM mode has a concentration factor of only 32%. This difference in the

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

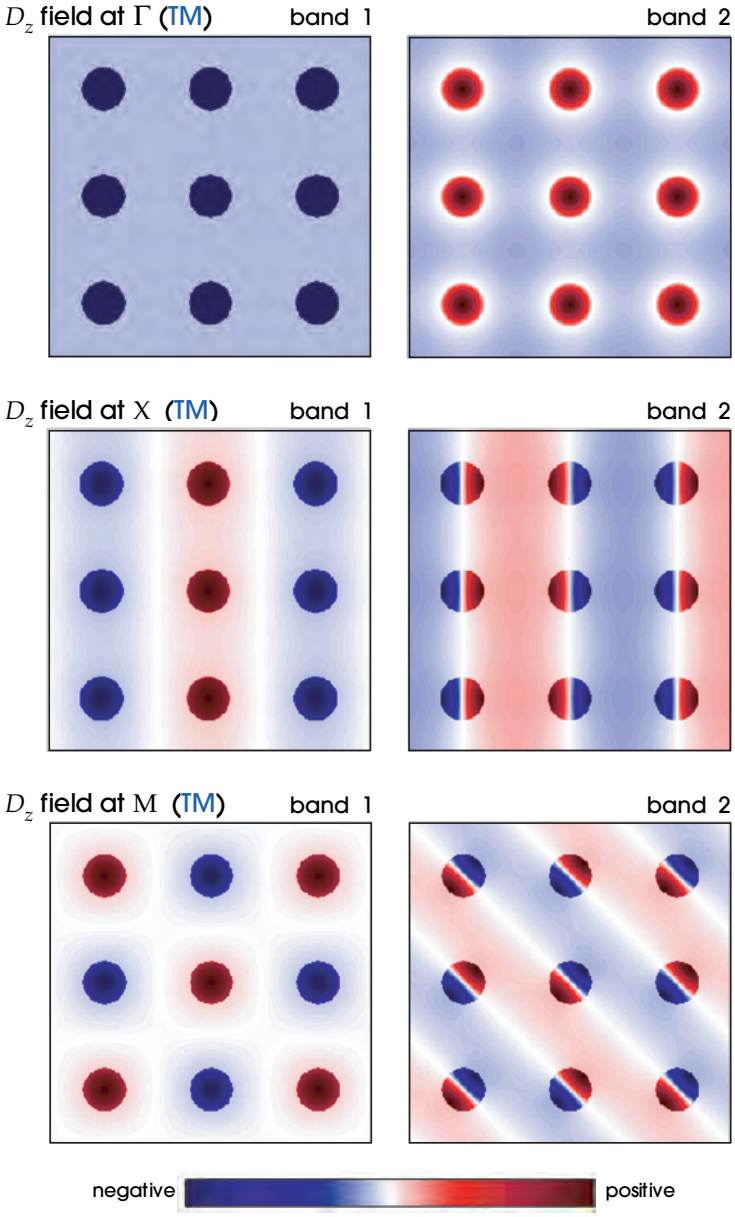


Figure 3: *Displacement* fields of TM states inside a square array of dielectric ($\varepsilon = 8.9$) columns in air. The color indicates the amplitude of the displacement field, which points in the z direction. Modes are shown at the Γ point (top), the X point (middle), and the M point (bottom). In each set, the dielectric band is on the left and the air band is on the right. The fields of the air band at the M point are from one of a pair of degenerate states.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

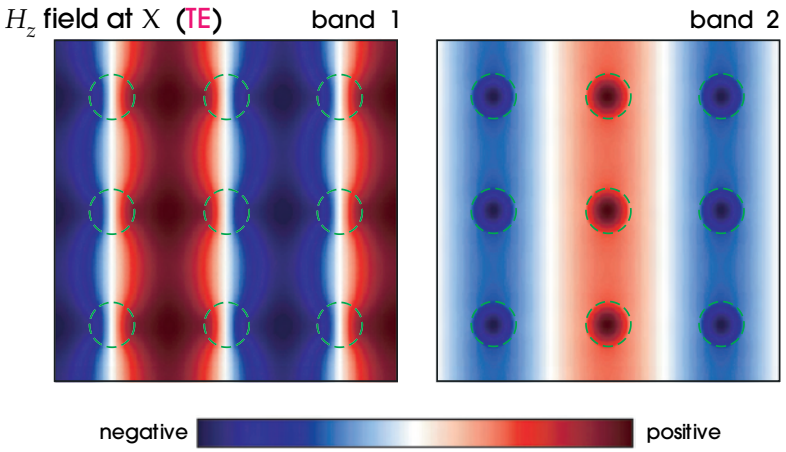


Figure 4: *Magnetic* fields of X-point TE states inside a square array of dielectric ($\varepsilon = 8.9$) columns in air. The column positions are indicated by dashed green outlines, and the color indicates the amplitude of the magnetic field. The dielectric band is on the left and the air band is on the right. Since $\mathbf{D}$ is largest along nodal planes of $\mathbf{H}$, the white regions are where the displacement energy is concentrated. TE modes have $\mathbf{D}$ lying in the xy plane.

Table 1

	TM	TE
Dielectric band	83%	23%
Air band	32%	9%

Concentration factors for the lowest two bands of the square lattice of dielectric rods at the X point of the Brillouin zone.

energy distribution of consecutive modes is responsible for the large TM photonic band gap.

The concentration factors for the TE modes do not contrast as strongly. This is reflected in the field configurations for the lowest two bands, shown in figure 4. We have actually plotted the magnetic field $\mathbf{H}$, since it is a scalar for TE modes and easy to visualize. From equation (8) of chapter 2, we know that the displacement field $\mathbf{D}$ tends to be largest along the nodal planes of the magnetic field $\mathbf{H}$. The displacement field of both modes has a significant amplitude in the air regions, raising the mode frequencies. But in this case there is no choice: there is no continuous pathway between the rods that can contain the field lines of $\mathbf{D}$. The field lines must be continuous, and they are consequently forced to penetrate the air regions. This is the origin of the low concentration factors, and the explanation for the absence of a band gap for TE modes.

The *vector* nature of the electromagnetic field, and in particular its *discontinuous boundary conditions* at material interfaces, is central to this phenomenon. When we move across a dielectric boundary from a high dielectric ε_1 to some lower $\varepsilon_2 < \varepsilon_1$,

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

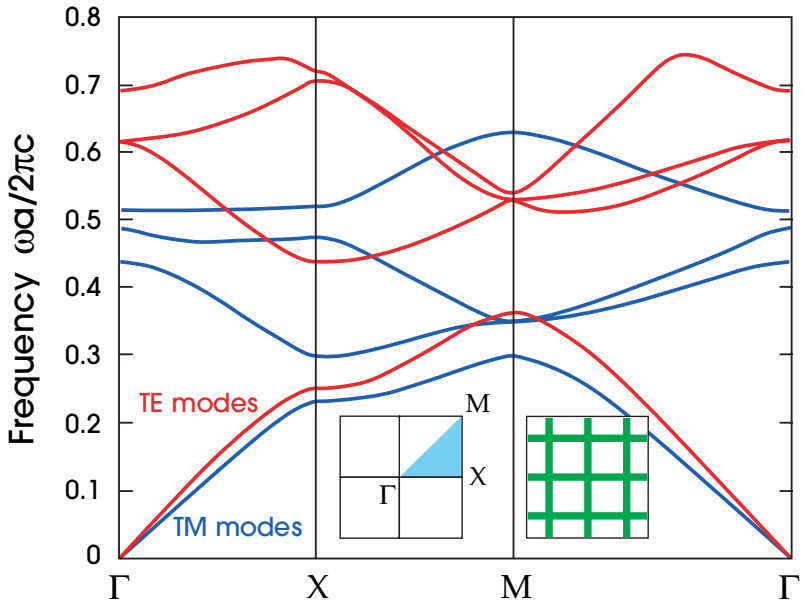


Figure 5: The photonic band structure for the lowest-frequency modes of a square array of dielectric ($\varepsilon = 8.9$) veins (thickness $0.165a$) in air. The blue lines are TM bands and the red lines are TE bands. The left inset shows the high-symmetry points at the corners of the irreducible Brillouin zone (shaded light blue). The right inset shows a cross-sectional view of the dielectric function.

the energy density $\varepsilon|\mathbf{E}|^2$ will *decrease* discontinuously by $\varepsilon_2/\varepsilon_1$ if $\mathbf{E}$ is *parallel* to the interface (since $\mathbf{E}_\parallel$ is continuous) and will *increase* discontinuously by $\varepsilon_1/\varepsilon_2$ if $\mathbf{E}$ is *perpendicular* to the interface (since $\varepsilon\mathbf{E}_\perp$ is continuous). In the TM case, $\mathbf{E}$ is parallel to all dielectric interfaces and so a large concentration factor is possible. In the TE case, however, the electric field lines *must* cross a boundary at some point, forcing electric-field energy out of the rods and preventing a large concentration factor. As a result, consecutive TE modes cannot exhibit markedly different concentration factors, and band gaps do not appear.

A Square Lattice of Dielectric Veins

Another two-dimensional photonic crystal that we will investigate is a square grid of dielectric veins (thickness $0.165a$, $\varepsilon = 8.9$), shown as an inset in figure 5. In a sense, this structure is complementary to the square lattice of dielectric columns, because it is a *connected* structure. The high-ε regions form a continuous path in the xy plane, instead of discrete spots. The complementary nature is reflected in the band structure of figure 5. Here, there is an 18.9% gap in the TE band structure, but there is no gap for the TM modes. This is the opposite of the situation for the square lattice of dielectric columns.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

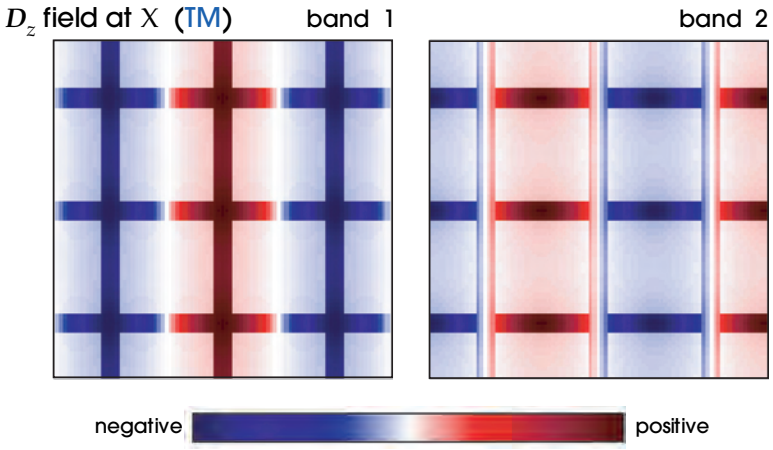


Figure 6: *Displacement* fields of X-point TM modes for a square array of dielectric ($\varepsilon = 8.9$) veins in air. The color indicates the amplitude of the displacement field, which is oriented in the z direction (out of the page). The dielectric band is on the left, and the air band is on the right.

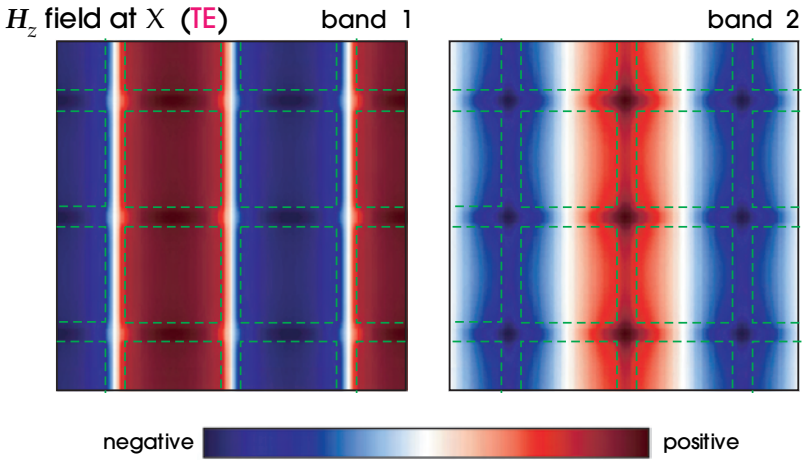


Figure 7: *Magnetic* fields of X-point TE modes for a square array of dielectric ($\varepsilon = 8.9$) veins in air. The green dashed lines indicate the veins, and the color indicates the amplitude of the magnetic field (which is oriented in the z direction). The dielectric band is on the left, and the air band is on the right.

Again, we turn to the field patterns of the modes in the two lowest bands to understand the appearance of the band gap. The fields are displayed for the TM and TE modes in figures 6 and 7, respectively.

Looking at the TM-field patterns in the first two bands, we see that both modes are mainly concentrated within the high-ε regions. The fields of the dielectric band

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Table 2

	TM	TE
Dielectric band	89%	83%
Air band	77%	14%

Concentration factors for the lowest two bands of the square lattice of veins at the X point.

are confined to the dielectric crosses and vertical veins, whereas the fields of the air band are concentrated in the horizontal dielectric veins connecting the square lattice sites. The consecutive modes both manage to concentrate in high-ε regions, thanks to the arrangement of the dielectric veins, so there is no large jump in frequency. This claim is verified by calculating the concentration factors for the field configurations. The results are in table 2.

On the other hand, the TE band structure has a photonic band gap between the first two bands. What is the difference between *this* lattice and the square lattice of dielectric columns, in which no TE gap appeared? In this case, the continuous field lines of the transverse electric field lines can extend to neighboring lattice sites without ever leaving the high-ε regions. The veins provide high-ε roads for the fields to travel on, and for $n = 1$ the fields stay almost entirely on them. As before, since the **D** field will be largest along the nodal (white) regions of the **H** field, we can infer from figure 7 that the **D** field of the lowest band is strongly localized in the vertical dielectric veins.

The **D** field of the next TE band $(n = 2)$ is forced to have a node passing through the vertical high-ε region, to make it orthogonal to the previous band. Some of its energy is thereby forced into the low-ε regions. (The maxima in the displacement energy coincide with the white regions of **H** in figure 7.) This causes a sizable jump in frequency between the two bands. This hypothesis is supported quantitatively by concentration factor calculations in table 2. We find a large concentration factor for the dielectric band and a small one for the air band. This jump in concentration factor between consecutive bands causes the formation of a band gap. In this case, it is the *connectivity* of the lattice that is crucial to the production of TE band gaps.

A Complete Band Gap for All Polarizations

In the previous two sections, we used the field patterns as our guide to understand which aspects of two-dimensional photonic crystals lead to TM and TE band gaps. By combining our observations, we can design a photonic crystal that has band gaps for *both* polarizations. By adjusting the dimensions of the lattice, we can even arrange for the band gaps to overlap, resulting in a **complete band gap** for all polarizations.

Earlier, we found that the isolated high-ε spots of the square lattice of dielectric columns forced consecutive TM modes to have different concentration factors, due

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

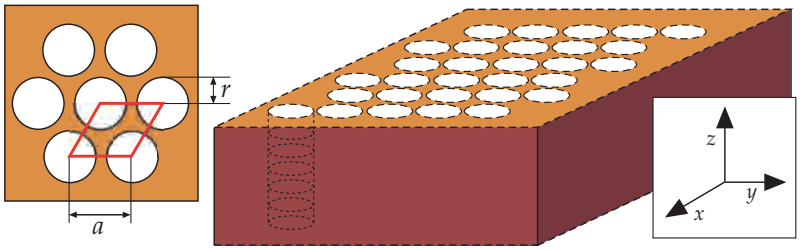


Figure 8: A two-dimensional photonic crystal of air columns in a dielectric substrate (which we imagine to extend indefinitely in the z direction). The columns have radius r and dielectric constant $\varepsilon = 1$. The left inset shows a view of the triangular lattice from above, with the unit cell framed in red. It has lattice constant a.

to the appearance of a node in the higher-frequency mode. This, in turn, led to the large TM photonic band gap. The lattice of dielectric veins presented a more spread-out distribution of high-ε material, leading to greater similarity between the concentration factors of successive modes.

On the other hand, the connectivity of the veins was the key to achieving gaps in the TE band structure. In the square lattice of dielectric rods, the TE modes were forced to penetrate the low-ε regions, since the field lines had to cross dielectric boundaries. As a result, the concentration factors for consecutive modes were both low and not very far apart. This problem disappeared for the lattice of dielectric veins, since the fields could follow the high-ε paths from site to site, and the additional node in the higher mode corresponded to a large frequency jump.

To summarize our rule of thumb, *TM band gaps are favored in a lattice of isolated high-ε regions, and TE band gaps are favored in a connected lattice.*

It seems impossible to arrange a photonic crystal with both isolated spots and connected regions of dielectric material. The answer is a sort of compromise: we can imagine crystals with high-ε regions that are both practically isolated *and* linked by narrow veins. An example of such a system is the **triangular** lattice of air columns, shown in figure 8.

The idea is to put a triangular lattice of low-ε columns inside a medium with high ε. If the radius of the columns is large enough, the spots between columns look like localized regions of high-ε material, which are connected (through a narrow squeeze between columns) to adjacent spots. This is shown in figure 9. The band structure for this lattice, shown in figure 10, has photonic band gaps for both the TE and TM polarizations. In fact, for the particular radius $r/a = 0.48$ and dielectric constant $\varepsilon = 13$, these gaps overlap, and we obtain an 18.6% complete photonic band gap.

Out-of-Plane Propagation

Until now we focused exclusively on modes that propagate in the plane of periodicity, with $k_z = 0$. However, for some applications we must understand

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

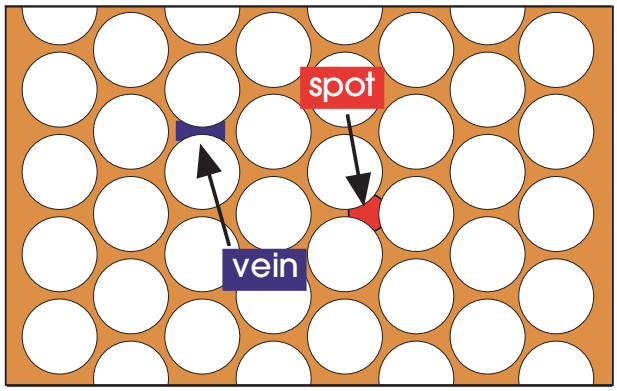


Figure 9: The spots and veins of a triangular lattice. Between the columns are narrow veins, connecting the spots surrounded by three columns.

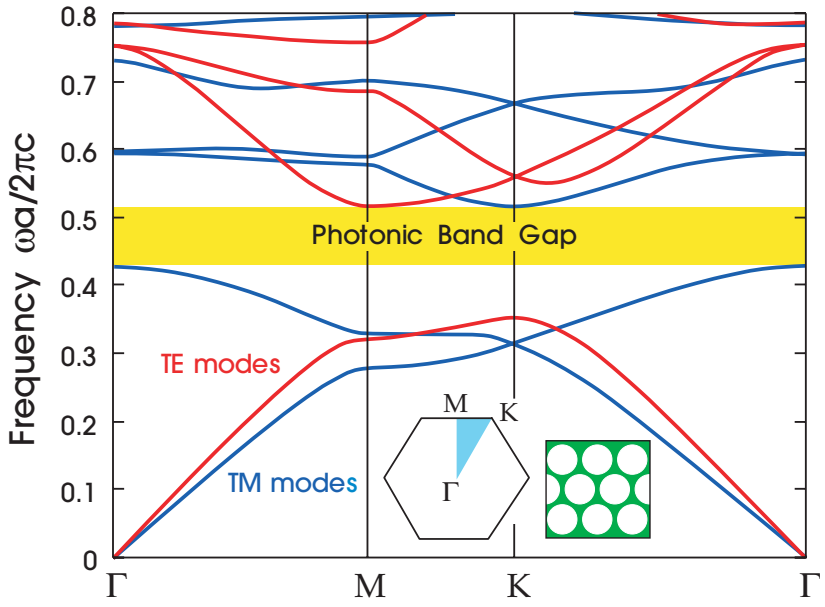


Figure 10: The photonic band structure for the modes of a triangular array of air columns drilled in a dielectric substrate ($\varepsilon = 13$). The blue lines represent TM bands and the red lines represent TE bands. The inset shows the high-symmetry points at the corners of the irreducible Brillouin zone (shaded light blue). Note the complete photonic band gap.

the propagation of light in an arbitrary direction. We will investigate the out-of-plane band structure by considering the $k_z > 0$ modes of the triangular lattice of air columns, the lattice which we discussed in the previous section. The out-of-plane band structure for this photonic crystal is shown in figure 11. Many

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

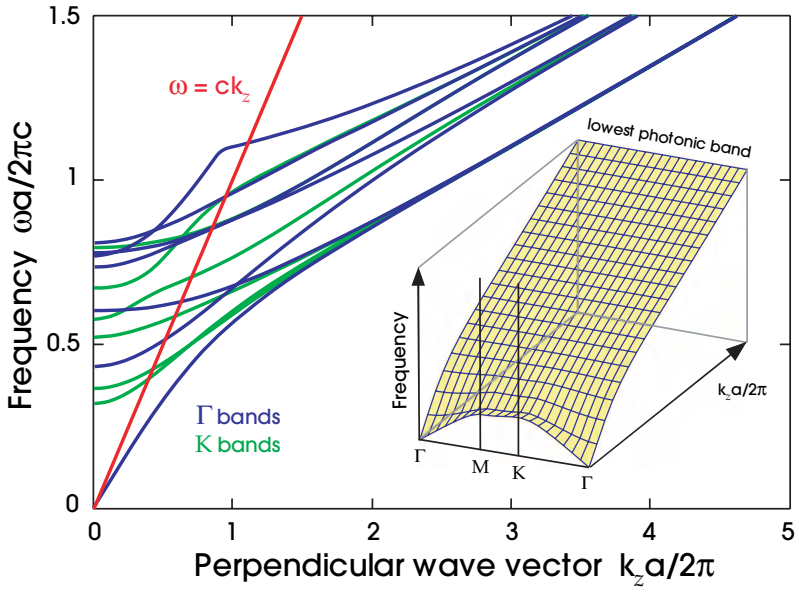


Figure 11: The out-of-plane band structure of the triangular lattice of air columns for the first few bands. The bands that start at Γ, $\omega(\Gamma, k_z)$, are plotted with blue lines, whereas the bands that start at K, $\omega(\mathrm{K}, k_z)$, are plotted with green lines. The light line $\omega = ck_z$ (red) separates the modes that are oscillatory ($\omega \geq ck_z$) in the air regions from those that are evanescent ($\omega < ck_z$) in the air regions. The inset shows the frequency dependence of the lowest band as k_z varies. Note that as k_z increases, the lowest band flattens.

of the qualitative features of the out-of-plane band structure for this crystal are common to all two-dimensional crystals. In fact, these features are just the natural extensions of the corresponding notions in multilayer films, which we developed in the previous chapter.

The first thing to notice about the out-of-plane band structure is that there are no band gaps for propagation in the z direction, just as we found for one-dimensional crystals in the section Off-Axis Propagation of chapter 4. This is a consequence of the homogeneity of the crystal in that direction. Intuitively, this is because no scattering occurs along that direction; band gaps require multiple scatterings from regions of different dielectric constant. Another important point is that there is no longer a useful distinction between the TE and TM polarizations, because the mirror symmetry is broken for $k_z \neq 0$.

Also note that the bands become flat with increasing k_z. The inset to figure 11 shows the frequency dependence of the lowest band as k_z is varied. When $k_z = 0$, this lowest band spans a broad range of frequencies, as we saw in figure 10. As k_z increases, the lowest band flattens and the bandwidth (the range of allowed frequencies for a given k_z) decreases to zero. Since the bandwidth is typically determined by the difference between the frequencies at Γ and K, figure 11 also

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

shows both $\omega(\Gamma, k_z)$ and $\omega(\mathrm{K}, k_z)$ for the first few bands. As k_z increases, the bandwidth of each of the bands vanishes. Why is this the case?

There is a simple explanation. For large k_z, the light is trapped by index guiding inside the dielectric regions, just as it would be in an optical waveguide. The light modes that are trapped in neighboring high-ε regions have very little overlap, so the modes decouple and the bandwidth shrinks to zero. This is especially true for modes that have $\omega \ll ck_z$.[1] In this regime, the fields decay exponentially outside the high-ε regions, and the overlap between modes in neighboring high-ε regions vanishes. This behavior is displayed by the bands in figure 11, which have a large dispersion for $\omega > ck_z$, but a small dispersion for $\omega \ll ck_z$.

In later chapters we will expand upon these notions. In chapter 8, we will consider the case of cylinders of finite height—structures that are periodic in two dimensions, but are not strictly homogeneous in the third direction. In that case, k_z is not conserved, and we cannot ignore out-of-plane propagation vectors. In chapter 9, we will consider the new kinds of band gaps that can arise for large k_z, distinct from the TE/TM gaps in this chapter, and which have important applications in *fibers* with 2D-crystal cross sections.

Localization of Light by Point Defects

Previously, we found two-dimensional photonic crystals with band gaps for in-plane propagation. Within the band gap, no modes are allowed; the density of states (the number of possible modes per unit frequency) is zero. By perturbing a single lattice site, we can create a single localized mode or a set of closely spaced modes that have frequencies within the gap. For the case of the multilayer film, we found that we could localize light near a particular plane by perturbing the dielectric constant of that plane.

In two dimensions, we have many options. As depicted in figure 12, we can remove a single column from the crystal, or replace it with another whose size, shape, or dielectric constant is different than the original. Perturbing just one site ruins the translational symmetry of the lattice. Strictly speaking, we can no longer classify the modes by an in-plane wave vector. However, the mirror-reflection symmetry is still intact for $k_z = 0$. Therefore we can still restrict our attention to in-plane propagation, and the TE and TM modes still decouple. That is, we can discuss the band structures for the two polarizations independently, as before. Perturbing a single lattice site causes a defect along a *line* in the z direction. But because we are considering propagation only in the plane of periodicity, and the perturbation is localized to a particular point in that plane, we refer to this perturbation as a **point defect**.

Removing one column may introduce a peak into the crystal's density of states within the photonic band gap. If this happens, then the defect-induced state must

[1] For $\omega > ck_z$, there are also modes that are mostly trapped in the *low*-ε regions, since Fresnel reflection goes to 100% in the glancing-angle ray-optics limit.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

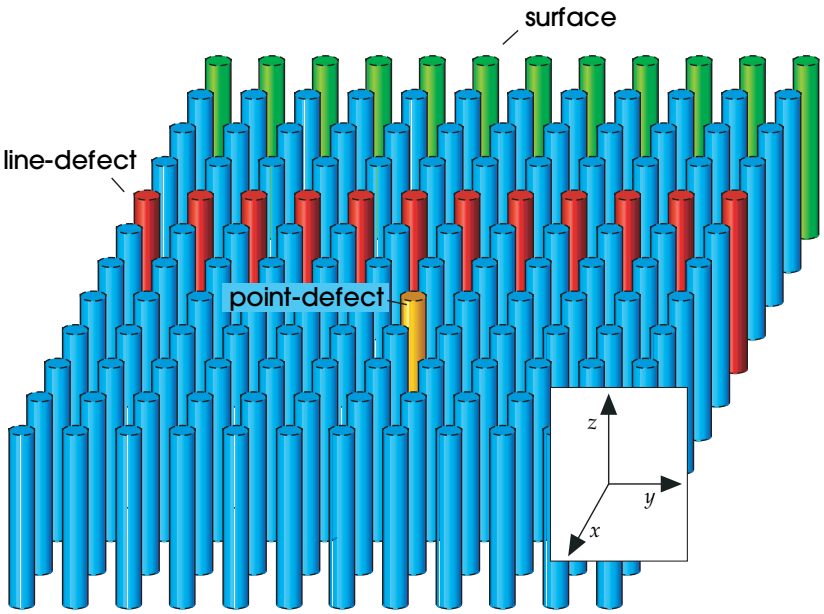


Figure 12: Schematic illustration of possible sites of point, line, and surface defects. Perturbing one column in the bulk of the crystal (yellow) might allow a defect state to be localized in both x and y. Perturbing one row in the bulk of the crystal (red) or truncating the crystal at a surface (green) might allow a state to be localized in one direction (x). The rods are assumed to extend indefinitely in the z direction.

be evanescent. The defect mode cannot penetrate into the rest of the crystal, since it has a frequency in the band gap. The analysis of chapter 4 is easily generalized to the case of two dimensions, allowing us to conclude that any defect modes decay *exponentially* away from the defect. They are localized in the xy plane, but extend in the z direction.

We reiterate the simple explanation for the localizing power of defects: the photonic crystal, because of its band gap, reflects light of certain frequencies. By removing a rod from the lattice, we create a **cavity** that is effectively surrounded by reflecting walls. If the cavity has the proper size to support a mode in the band gap, then light cannot escape, and we can pin the mode to the defect.[2]

[2] A reader familiar with semiconductor physics can understand this result by analogy with impurities in semiconductors. In that case, atomic impurities create localized electronic states in the band gap of a semiconductor (see, e.g., Pantelides, 1978). Attractive potentials create a state at the conduction band edge, and repulsive potentials create a state at the valence band edge. In the photonic case, we can put the defect mode within the band gap with a suitable choice of $\varepsilon_{\text{defect}}$. In the electronic case, we use the effective-mass approximation to predict the frequencies and wave functions of defect modes. Within each unit cell the wave functions are oscillatory, but the oscillatory functions are modulated by an evanescent envelope. An analogous treatment is applicable to the photonic case (Istrate and Sargent, 2006).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

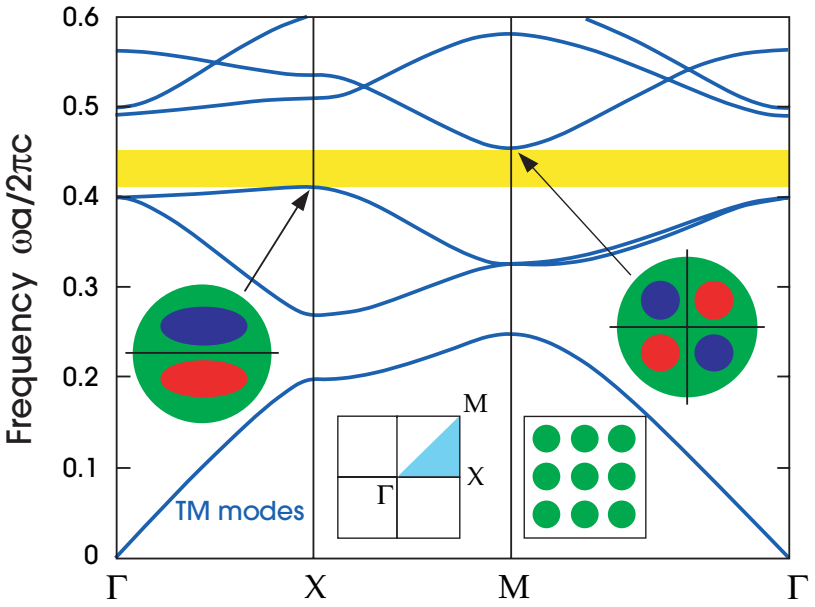


Figure 13: The TM modes of a square array of dielectric ($\varepsilon = 8.9$) columns in air, with $r = 0.38a$. The center inset shows the irreducible Brillouin zone (shaded light blue). The other two insets suggest the field patterns of the modes, within each column (red/blue for positive/negative fields). the left inset shows the π-like pattern for band 3; the right shows the δ-like pattern for the bottom of band 4.

We illustrate this discussion of localized modes in two-dimensional photonic crystals with a system that has been studied both experimentally[3] and theoretically[4]: a square lattice of alumina columns in air. Unlike the infinitely long columns of figure 1, the real columns were sandwiched between metal plates, which introduced a cutoff TE frequency that was larger than the frequencies being investigated. The plates also insured $k_z = 0$ propagation. In this way, the experimentalists created a system that displays only TM modes with $k_z = 0$.

The in-plane band structure of this system is reproduced in figure 13. The system has a 10.1% photonic band gap between the third and fourth bands. By examining the field patterns, as we did in the first two sections of this chapter, we would find that the first band is composed primarily of states that have no nodal planes passing through the high-ε columns. In analogy with the nomenclature of molecular orbitals, we describe such a nodeless field pattern as a "σ-like" band. The second and third bands are composed of "π-like" bands, with one nodal plane passing through each column. The bottom of the fourth band has a "δ-like" pattern with two nodal planes per column (see insets to figure 13). Remember

[3] See McCall et al. (1991).
[4] See Meade et al. (1993).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

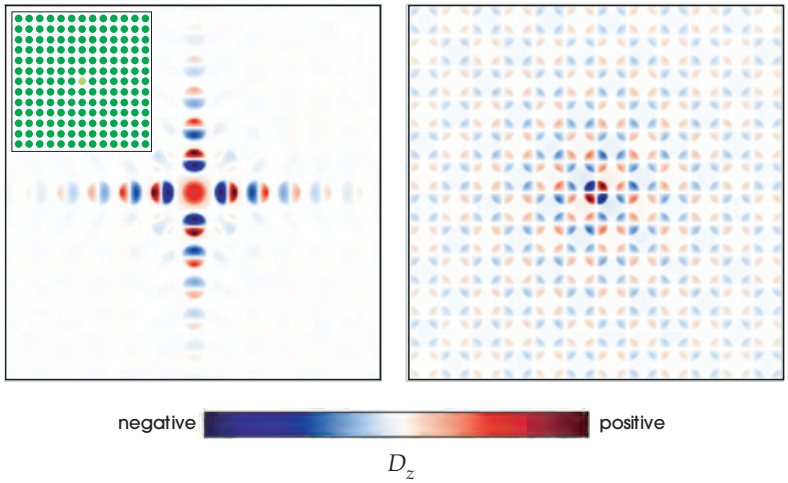


Figure 14: Displacement fields (D_z) of states localized about a defect in a square lattice of alumina rods ($\varepsilon = 8.9$) in air. The color indicates the magnitude of the field. The defect on the left was created by *reducing* the dielectric constant of a single rod. This mode has a **monopole** pattern with a single lobe in the defect and rotational symmetry. The defect on the right was created by *increasing* the dielectric constant of a single rod. This mode has a **quadrupole** pattern with two nodal planes in the defect, and transforms like the function $f(\boldsymbol{\rho}) = xy$ under rotations.

that additional nodal planes in the high-ε regions correspond to larger amplitudes in the low-ε regions, which increases the frequency.

A defect in this array introduces a localized mode, as shown in figure 14. Experimentally, this defect was created by replacing one of the columns with a column of a different radius. Computationally, the defect was introduced by varying the dielectric constant of a single column. In terms of the index of refraction $n = \sqrt{\varepsilon}$, the defect varied from $\Delta n = n_{\text{alumina}} - n_{\text{defect}} = 0$ to $\Delta n = 2$ (one column completely gone). The results of the computation are shown in figure 15.

The largest photonic band gap in this structure is between the third (π-like) band of states with one nodal line and the fourth (δ-like) band of states with two nodal lines in the high-ε regions. When the index of refraction is less than 3, a state leaves the π band and enters the photonic band gap. As Δn is increased between 0 and 0.8, this doubly degenerate mode sweeps across the gap. At $\Delta n = 1.4$, a nondegenerate state enters the gap, sweeps across, and penetrates the δ band at $\Delta n = 1.8$. This mode is displayed at $\Delta n = 1.58$, before entering the δ band, in the left panel of figure 14.

Note that this state has one nodal line passing through each dielectric column other than the center, showing that it retains the essential character of the π band; however, because it has no nodal lines in the *central* (defect) column, we describe this state as a **monopole** mode. Similarly, defects with $\Delta n < 0$ pull states out of the δ band. The resulting localized states retain their δ-like character, with *two* nodal

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

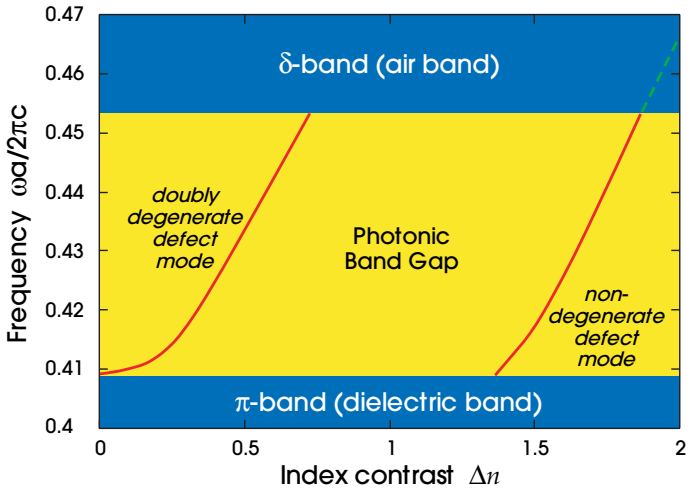


Figure 15: The evolution of localized modes associated with a defect column in an otherwise perfect square lattice, as the defect's index of refraction decreases. An index change $\Delta n = 0$ corresponds to the perfect crystal; $\Delta n = 2$ corresponds to the complete removal of one column. The horizontal lines indicate the band-gap edges. In the gap, the frequencies (red lines) are associated with localized states, but after the line punctures the continuum it becomes a resonance with broadened frequency (green line). The state at $\Delta n = 1.58$ has the field pattern shown in the left panel of figure 14.

planes per column, as shown on the right panel of figure 14. Because there are also two nodal planes in the defect, we call this a **quadrupole** mode.

The defect mode frequency increases as $\varepsilon_{\text{defect}}$ decreases, as we can see in figure 15. A simple way to understand this is to examine the effects that a small variation in $\varepsilon(\mathbf{r})$ has on the frequency of a mode via the perturbation theory of the section The Effect of Small Perturbations of chapter 2. From equation (28) of chapter 2, we see that for a negative value of $\Delta\varepsilon$ (i.e., removing a bit of dielectric), the corresponding frequency shift $\Delta\omega$ is positive, and a state can pop out of the top of the π band (dielectric band). Increasing $\Delta\varepsilon$ pushes the state deeper into the gap. Conversely, for $\Delta\varepsilon$ positive, the frequency shift is negative and a state at the bottom of the δ band (air band) can fall into the gap.[5]

Although the defect destroys the translational symmetry of the crystal, many types of defects still allow the crystal to retain some point symmetries. For

[5] An analogy between this case and the case of impurities in a semiconductor might benefit the reader familiar with that subject. Since the wavelength of light is shorter in a dielectric than in air, these regions are analogous to regions of deep potential in the semiconductor. Decreasing the dielectric constant at one photonic crystal site is analogous to adding a repulsive potential to one atomic site, which pushes a state out of the π band. Decreasing the dielectric constant further increases the repulsive character of the defect, and pushes the localized mode to higher frequencies. Conversely, increasing the dielectric region is analogous to adding an *attractive* potential that pulls a state out of the δ band.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

instance, in the inset to figure 14, we see that after removing one column from the lattice, we can still rotate the crystal by 90° about the z axis and leave it unchanged. If a defect does retain a point symmetry, then we can use that symmetry to classify the defect modes, just as we did in chapter 3.

For example, since the defect in the left panel of figure 14 is unchanged under 90° rotations, we can immediately predict the symmetry properties of the doubly degenerate modes that cross the gap for $0 < \Delta n < 0.8$. They must be a pair of **dipole** modes (with one nodal plane in the defect) that transform into each other under a 90° rotation, since that is the only doubly degenerate way to reproduce the symmetry of the surroundings.[6]

What happens to the defect mode after it reaches the δ (air) band? When its frequency is above the bottom of the δ band, the defect mode is no longer trapped in the band gap; it can leak into the continuum of states that makes up the δ band. The defect no longer creates a truly *localized* mode. However, since we expect a smooth transition between localized and continuum behavior, the mode still concentrates much of its field energy near the defect. But in this case, the defect is not surrounded by reflecting walls, so the energy will leak away into the continuum of states at some rate γ. We call this a **leaky mode** or **resonance**.[7] The peak that the defect creates in the crystal's density of states widens in proportion to γ; as the mode penetrates farther into the continuum, farther and farther from the true bound state, the resonance broadens away and melts into the continuum.

In a crystal with a *finite* number of rods, then even the defect modes *inside* the gap are resonances, rather than truly localized modes. Each resonance has a nonzero leakage rate γ into the surrounding medium. Because these defect modes decay exponentially into the crystal, however, the value of γ also decreases exponentially with the number of surrounding crystal periods. Thus the loss rate can be made as small as desired. We return to resonance and losses in the section Quality Factors of Lossy Cavities of chapter 7, and discuss their impact on device applications in chapter 10.

Point defects in a larger gap

The same principles apply equally well to other types of defects in other crystal structures. We conclude this section with the example of a defect formed by

[6] In the language of group representation theory: the defect has the symmetry of the C_{4v} point group, and the only 2×2 (doubly degenerate) irreducible representation of that group is a pair of dipole-like functions. A monopole state, such as the one in the left panel of figure 14, corresponds to the trivial representation (1). See Inui et al. (1996, Table B.7).

[7] This name is chosen to suggest the correspondence with the phenomenon of resonance scattering in optics and quantum mechanics. In a scattering experiment, tuning the incident energy near the energy of a true bound state causes a peak in the cross section due to the resonance of the incident beam with the potential. Here, there is a peak in the density of states because our defect mode is very close to a true localized mode. The rate of decay of the resonance is proportional to the peak width in both cases, as analyzed in more detail by chapter 10.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

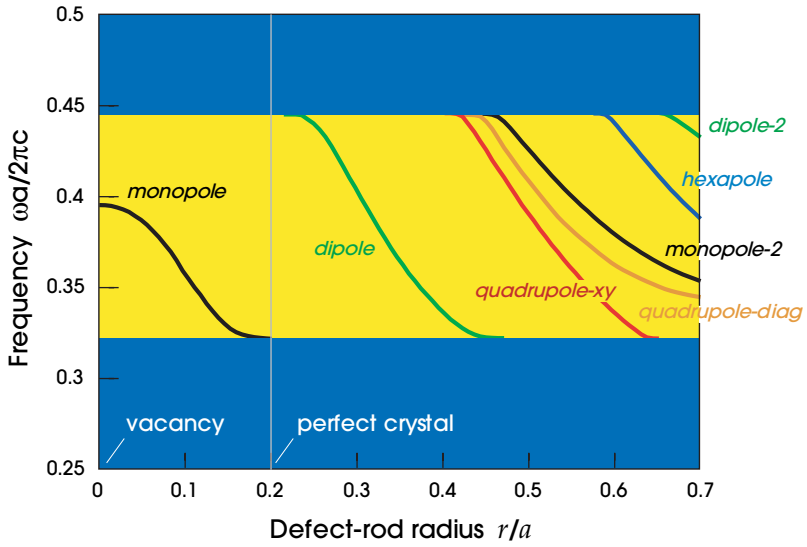


Figure 16: The evolution of localized modes associated with a defect column in an otherwise perfect square lattice with the parameters of figure 2, as the defect's *radius* is changed. A radius of $r = 0.2a$ corresponds to the perfect crystal. The TM band-gap frequency range, about 0.32–0.44 $2\pi c/a$, is shaded yellow. If the radius is *decreased* then a single *monopole* state is pushed *up* into the gap. If the radius is *increased* then a sequence of higher-order modes (with more nodal planes) are pulled *down* into the gap. The *dipole*, *hexapole*, and *dipole-2* states are doubly degenerate, while the other states are nondegenerate. The field patterns for these states and explanations for their names are in figure 17.

changing the *radius* of a single rod in the photonic crystal of figure 2 (a square array of alumina rods with $r = 0.2a$ instead of $0.38a$), using the large *fundamental* gap between the first and second bands (which are σ-like and π-like, respectively). Since the band gap here is three times as large as in the preceding example, the modes are localized more strongly. As a practical matter, one typically works with as large a gap as possible to maximize localization, and the defect modes here are closely related to cases we will examine in chapters 6 and 10.

The frequencies of the various defect states are plotted as a function of the defect radius in figure 16, from a radius of 0 (a rod completely removed) to $0.7a$ (more than tripled in diameter), with the corresponding field patterns in figure 17. For a missing rod, one obtains a monopole state that is pushed up from the dielectric (σ-like) band, whereas increasing the radius initially pulls down a pair of doubly degenerate dipole modes from the air (π-like) band.

For larger radii, we pull down a sequence of higher-order modes with an increasing number of nodal lines: two different quadrupole states, a higher-order monopole with an extra node in the radial direction, and two higher-order dipole

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

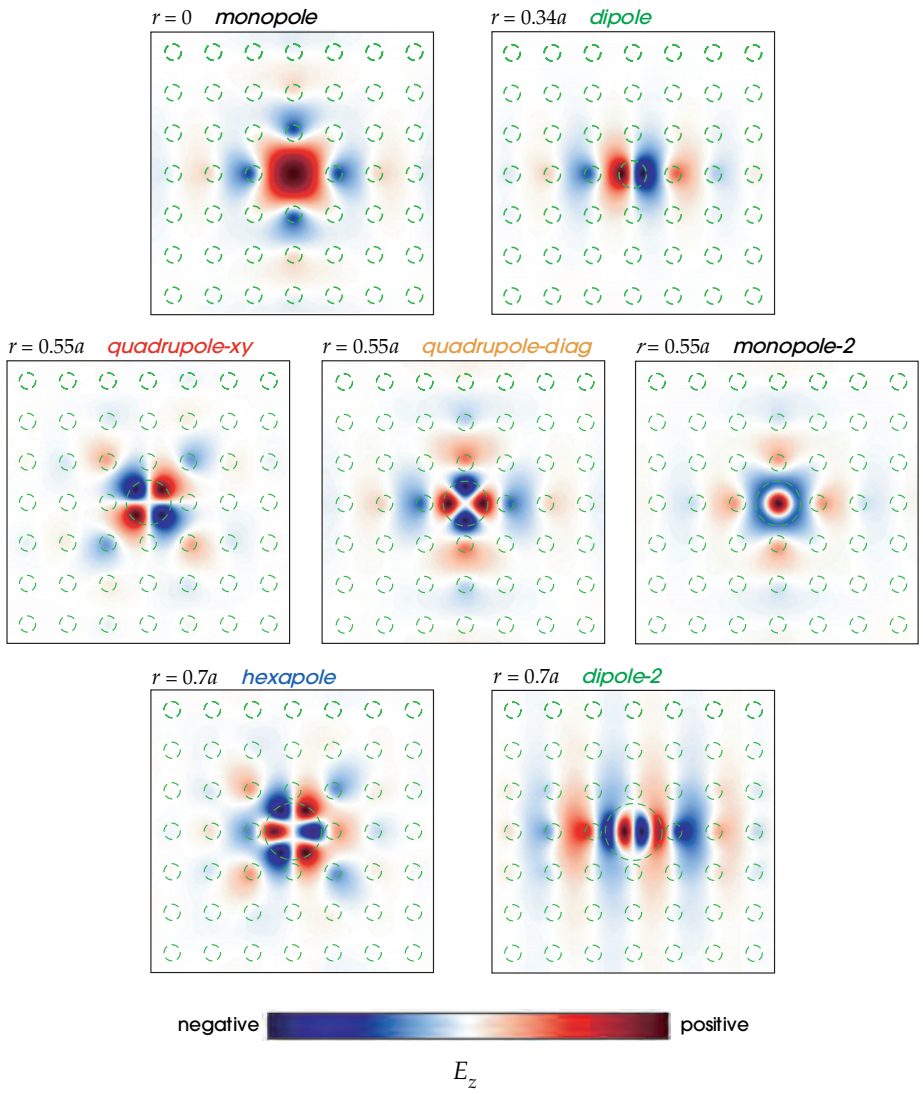


Figure 17: Electric-field (E_z) patterns of states localized about a defect in a square lattice of rods in air, where the radius r of a single rod has been changed. The rods are shown as dashed green outlines. *Top left:* a **monopole** state formed by completely removing a rod ($r = 0$), which has a single lobe in the defect and high symmetry. *Top right:* **dipole** state for an increased radius $r = 0.34a$, which is doubly degenerate with the other state (not shown) formed by a 90° rotation. *Middle row:* three nondegenerate states formed by an increased radius $r = 0.55a$. There are two **quadrupole** states, with the left (*quadrupole-xy*) having nodal planes along the x and y axes and transforming like $f(\boldsymbol{\rho}) = xy$ under rotations, while the center (*quadrupole-diag*) state has diagonal nodal planes and transforms like $x^2 - y^2$. The *monopole-2* state at right is a second-order monopole state, with an extra nodal plane in the radial direction. *Bottom row:* two states formed by an increased radius $r = 0.7a$, *each* of which is doubly degenerate with another state (not shown) formed by a 90° rotation. The **hexapole** state has six lobes in the defect, while *dipole-2* is a second-order dipole pattern.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

states (one of which we label as a **hexapole** to reflect the six lobes in its field pattern). All of the degenerate pairs, regardless of the complexity of their field patterns, have the same dipole symmetry: the degenerate partner differs by a 90° rotation, they transform the same way under the $x = 0$ and $y = 0$ mirror planes as well as under 180° rotation, and the degenerate partner is obtained by a 90° rotation (or a diagonal mirror plane).[8]

Linear Defects and Waveguides

We can use point defects in photonic crystals to trap light, as we have just seen. By using **linear defects**, we can also *guide* light from one location to another. The basic idea is to carve a waveguide out of an otherwise perfect photonic crystal by modifying a linear *sequence* of unit cells, as shown schematically in figure 12.[9] Light that propagates in the waveguide with a frequency within the band gap of the crystal is confined to the defect, and can be directed along the defect. An example is illustrated in figure 18, in which a column of rods has been removed along the y direction from the square-lattice crystal of figure 2.

A system with a linear defect still has *one* direction within the plane for which discrete translational symmetry is preserved. In this case, it is the y direction. Thus, the corresponding wave vector k_y is still a conserved quantity. We have also maintained continuous translational symmetry in z. It follows that k_z is conserved and we can restrict ourselves to in-plane ($k_z = 0$) propagation. We consider only the TM polarization. The band diagram of ω versus k_y is given in figure 19, wherein the introduction of the defect results in a discrete **guided band** lying inside the TM band gap. By virtue of the gap, the mode represented by this band is evanescent within the crystal, and is localized within the defect, as shown in figure 18. As usual, from the section Discrete vs. Continuous Frequency Ranges of chapter 2, the localization explains why the mode exists as a discrete band rather than a continuous range of frequencies for a given wave vector. Outside of the gap, in the blue region of figure 19, the modes extend within the crystal and correspondingly cover a continuous range of frequencies.

We should emphasize a key difference between linear defects and the point defects that were discussed in the previous section. For a point defect, a mode can be localized whenever its frequency is in the photonic band gap. For a linear defect, we consider the mode's behavior not only as a function of frequency, but also as a function of its wave vector k_y. A guided mode need only have a combination (k_y, ω_0) that is disallowed in the crystal; it is not necessary that ω_0 alone is disallowed. To test for the existence of a guided mode, we pick a specific

[8] In general, there are only five possible representations of the C_{4v} symmetry group into which the states can fall, corresponding to the monopole, dipole, and two quadrupole states shown in figure 17, as well as an *octopole* possibility that does not occur for this range of defect radii. See Inui et al. (1996, Table B.7).

[9] This is just one simple example; to create a linear defect, one need not modify *every* unit cell identically, or even at all, as long as the defect has discrete translational symmetry.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

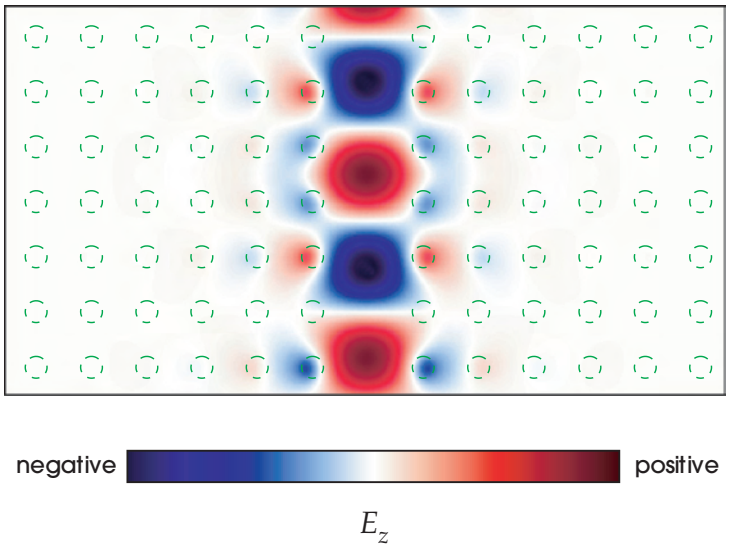


Figure 18: Electric-field (E_z) pattern associated with a linear defect formed by removing a column of rods from an otherwise-perfect square lattice of rods in air. The resulting field, shown here for a wave vector $k_y = 0.3\,(2\pi/a)$ along the defect, is a **waveguide mode** propagating along the defect. The rods are shown as dashed green outlines.

(k_y, ω_0) and ask if there is any k_y which will put that mode on a band. That is, by a suitable choice of k_x can we arrange for some band n such that $\omega_0 = \omega_n(k_x, k_y)$? If we can, then there is at least one extended state in the crystal with that combination (k_y, ω_0). If we tried to establish a guided mode with those parameters, it would leak into the crystal. This process of selecting a value of k_y (which is conserved) and examining all possible k_x (which is no longer conserved) is called **projecting the band structure** of the infinite crystal. The result is the blue region of figure 19, which is the *projection* of the states in the bulk crystal onto their (k_y, ω) coordinates.

The conservation of k_y is also why we are justified in calling this structure a **waveguide**. It is truly a channel along which light *propagates*, as opposed to being just a region where light is localized. A mode with a positive group velocity $d\omega/dk_y$ (as defined in the section Bloch-Wave Propagation Velocity of chapter 3) is a forward-propagating mode; here, these modes lie at $k_y > 0$, but this is not always the case. By flipping the sign of k_y, one obtains a mode with $d\omega/dk_y < 0$, a backward-propagating mode. These two modes do not couple with one another, because they have different values of the conserved quantity k_y. Each mode will propagate in the same direction forever, or at least as far as the translational symmetry is maintained.

The waveguide depicted in figure 18 is a demonstration of a remarkable property of photonic crystals: the ability to guide light primarily in the air. Conventional dielectric waveguides guide light through the mechanism of index

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

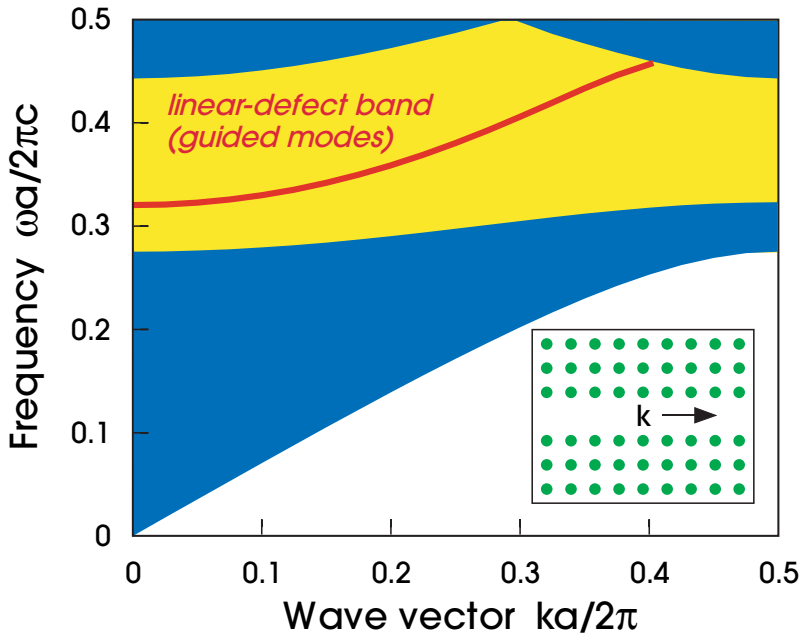


Figure 19: The projected band structure of the line defect (inset) formed by removing a row (or column) of rods from an otherwise perfect square lattice from figure 2, plotted versus the wave vector component k along the defect. The extended modes in the crystal become continuum regions (blue), whereas inside the band gap (yellow) a defect band (red) is introduced corresponding to a localized state as in figure 18.

guiding (total internal reflection), as described in the subsection Index guiding of chapter 3, but that mechanism can confine light only within a *higher-ε* region.[10] In contrast, the guidance mechanism in this case is the band gap of the surrounding crystal, which is independent of the properties of the material filling the core. This capability is important for a number of applications in which it is desirable to reduce the interactions, such as absorption or nonlinearity, between the light and the dielectric materials. We will return to this subject in chapter 9.

One can imagine a variety of ways in which a linear defect could be introduced into a crystal, and a corresponding variety of guided modes. The only requirement is that we maintain discrete translational symmetry in one direction. One possibility would be to remove every rod, or every nth rod (as discussed in more detail by the section Some Other Possibilities of chapter 10), along a single row. Another possibility is to remove every rod along multiple rows of rods. Our earlier choice to remove only a single row resulted in a **single-mode** waveguide, which has the

[10] However, a very *weakly* confined (e.g. low ε contrast) index-guided mode can have the majority of its energy in the exponential tails outside the higher-ε region.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

property that there is at most one guided mode at a given frequency. Removing multiple rows results in a **multi-mode** waveguide. These are generally undesirable for applications involving information transmission, because if signals can propagate with more than one velocity, then they will become scrambled (**modal dispersion**). To estimate the requisite width of a single-mode waveguide, think of the waveguide as an empty space between two perfectly reflecting walls. The appropriate boundary condition for such a propagating waveguide mode is that the field vanishes at the walls. For maximum confinement, we should operate near the midgap frequency $\omega_m = 0.38\,(2\pi c/a) = a/\lambda_m$. For the lowest-frequency mode, one-half the wavelength should span the width of the waveguide. Taken together, we conclude that the width should be roughly $\lambda_m/2 = 1.3a$, which is achieved (at least approximately) by removing only one row of dielectric rods.

Surface States

Much of our discussion has concerned the interior of photonic crystals of infinite extent. But real crystals are necessarily bounded. What happens at the surface of a two-dimensional photonic crystal? In this section, we explore the **surface modes** that photonic crystals are capable of sustaining, much like the one-dimensional case in the section Surface States of chapter 4. For a surface mode, light is localized at a surface plane, as shown schematically in figure 12. The field amplitudes decay exponentially away from the surface. They decay within the crystal because of the band gap, and they decay within the air region because they are index-guided (as in the subsection Index guiding of chapter 3).

We can characterize a given surface by its **inclination** and its **termination**. Surface inclination specifies the angles between the surface normal and the crystal axes. Surface termination specifies exactly where the surface cuts across the unit cell; for example, we can end a two-dimensional lattice of circles by stopping after some whole number of circles, or by cutting each circle in half at the boundary, or by cutting off some arbitrary fraction.

We will focus on the surface states of a square lattice of dielectric columns. Many of the arguments and results that we present, however, are quite general. Specifically, we will return to the square array of alumina rods from figure 2, with $\varepsilon = 8.9$. Consider the TM band structure, which has a photonic band gap between the first and second bands. For our surface inclination, we choose planes of constant x.[11] We will look at two different terminations: either we will draw the boundary just outside a line of whole columns, or we will cut the outermost columns in half. These two terminations are depicted in the insets to figures 20 and 21, respectively.

[11] This is known as the (10) surface of the square lattice, in the language of Miller indices from appendix B.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

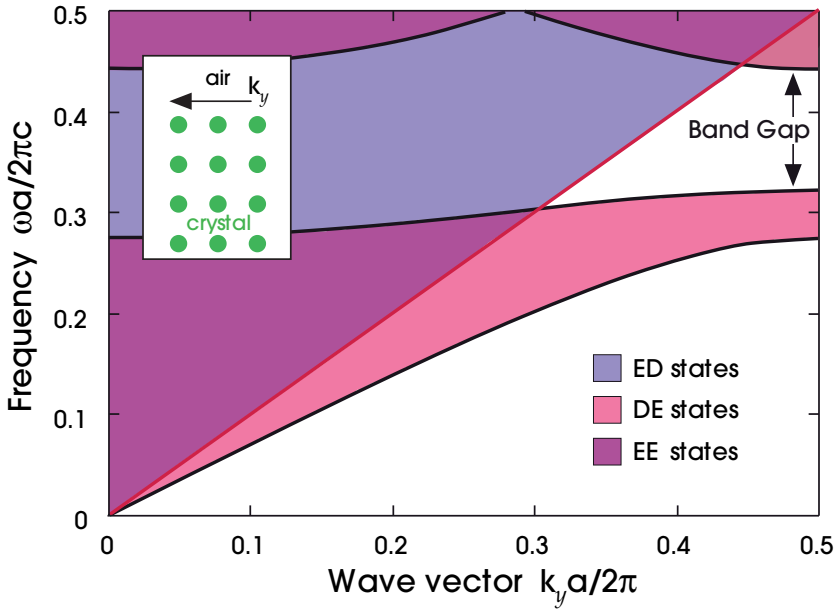


Figure 20: The projected band structure of the constant-x surface of the square lattice of alumina rods in air. The shading denotes regions in which light is transmitted (purple EE states), internally reflected (red DE states), and externally reflected (blue ED states). The crystal is terminated as shown in the inset; this termination has no surface states.

This structure has continuous translational symmetry in the z direction, so we can describe the modes by a k_z wave vector, but we restrict ourselves to $k_z = 0$ where we have a TM band gap. Just as with the linear defects of the previous section, we have broken translational symmetry in one direction (x) but have maintained it in another direction (y) in the plane. Thus, k_y is conserved, but k_x is not conserved. Therefore, it is useful to compute the *projected band structure* of the infinite crystal, in which (k_y, ω) is plotted for each $\omega_n(k_x, k_y)$.

Figure 20 shows the projected band structure of the constant-x surface of the square lattice of dielectric rods. In order to understand it, we first consider the projected band structures of the outside air and the photonic crystal separately. As before, we label each section of the plot with two letters. The first letter indicates whether the states are Extended or Decaying in the air region. The second letter gives the same information for the crystal region. The union of regions EE and ED shown in figure 20 is the projection of the free light modes onto the surface Brillouin zone. For a given k_y, there are light modes at all frequencies $\omega \geq c|k_y|$; this is the *light cone* that was defined in the subsection Index guiding of chapter 3. Along the **light line** $\omega = ck_y$, the light travels parallel to the surface, and increasing ω at fixed k_y corresponds to increasing k_x. Similarly, the union of regions EE and

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

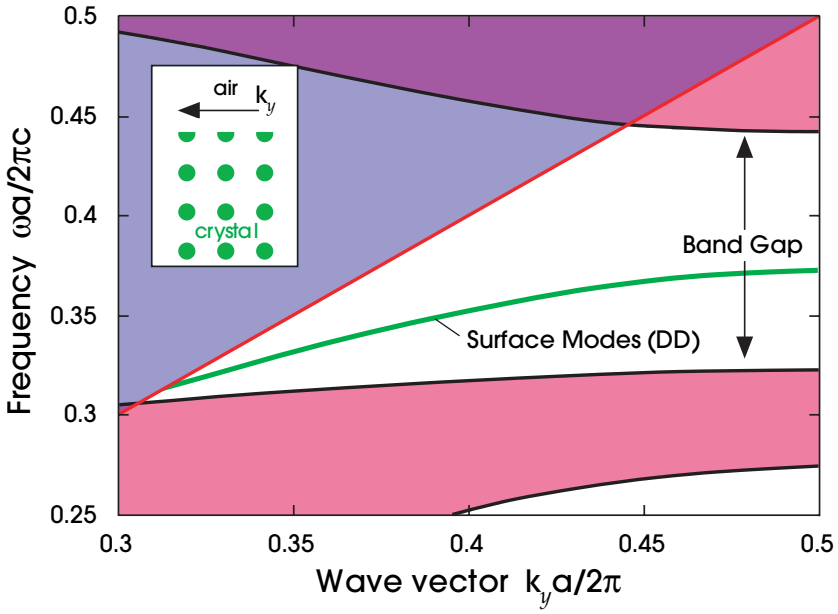


Figure 21: The projected band structure of the constant-x surface of the square lattice of alumina rods in air, like figure 20 above except with a different termination shown in the inset. The line in the gap corresponds to a surface band in which light is exponentially localized to the surface (green, DD).

DE represents the projected band structure of the photonic crystal. Note that the photonic crystal contains a TM band gap at $0.32 < (\omega a/2\pi c) < 0.44$.

Now we can understand the three types of surface states of the projected surface Brillouin zone: light that is transmitted (EE), light that is internally reflected (DE), and light that is externally reflected (ED). In the region of (ω, k_y) marked EE, the modes are extended in both the air and in the crystal, making it is possible to transmit light with those parameters through the crystal. In the DE region, there are modes in the crystal, but they are beneath the light line of the air modes. Thus the light can extend into the crystal, but decays exponentially into the surrounding air. This is nothing but the familiar phenomenon of total internal reflection. In the ED region, the situation is reversed. There, the modes can extend into the air, but they decay exponentially into the crystal.

Finally, there might exist bona fide surface modes, which decay exponentially away from the surface in both directions (DD). Such a mode appears in figure 21, which shows the band structure of the constant-x surface terminated by cutting the columns in half. The modes in the DD region are below the light line of the air modes, and are also within the band gap of the crystal. The fields therefore decay exponentially in both directions, and the mode is pinned to the surface plane, as

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

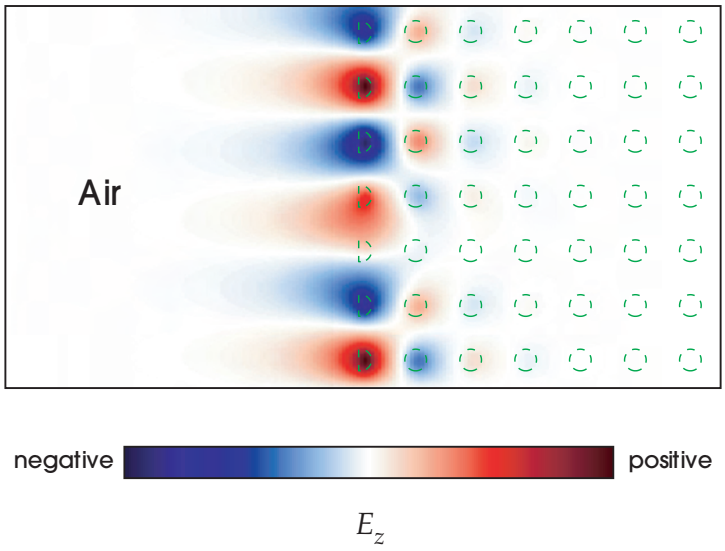


Figure 22: Electric-field (E_z) pattern associated with a surface-localized state formed by truncating a square lattice of alumina rods in air, cutting each rod in half at the boundary as in figure 21. The mode shown corresponds to a surface-parallel wave vector $k_y = 0.4\,(2\pi/a)$. The dielectric rods are shown as dashed green outlines.

shown in figure 22. By exciting such modes, we can imprison light at the surface of the crystal.[12]

Further Reading

Appendix A catalogues many of the analogies between the field of photonic crystals and the disciplines of quantum mechanics and solid-state physics. Appendix B provides a more detailed discussion of the Brillouin zones of the crystal geometries we have studied in this chapter. Appendix C provides the band gap locations for a variety of two-dimensional photonic crystals.

Some early experimental and theoretical approaches to two-dimensional band-gap systems can be found in McCall et al. (1991), Smith et al. (1993), Plihal and Maradudin (1991), and Villeneuve and Piché (1992). Meade et al. (1991*a*; 1991*b*) contains a systematic treatment of the square lattices of dielectric rods and veins.

[12] Note that surface modes cannot be excited by light propagating from either the air or the crystal, precisely because it does not overlap these modes. To excite a surface state, you either need to break translational symmetry (e.g., by ending the surface at a corner), or bring a light source so close to the surface that it can couple to the evanescent field of the surface state.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Winn et al. (1994) contains a more systematic treatment of square and triangular lattices of columns. Surface states on the interface between two different materials were reported in Meade et al. (1991*b*). The experimental investigation of the square lattice of dielectric columns for bulk states can be found in Robertson et al. (1992) and for surface states in Robertson et al. (1993).

The earliest theoretical studies of two-dimensional crystals relied on perturbative techniques that were limited to weak periodicity (e.g., Brillouin, 1946), in which a band gap cannot appear. There is an analytically tractable two-dimensional crystal with strong periodicity, however: a *separable* structure[13] for the TM polarization first solved by Chen (1981), who did not comment on the possibility of band gaps or localized states. This separable system was rediscovered by Kawakami (2002), who also [along with Čtyroký (2001) and Watts et al. (2002)] analyzed its localized defect modes. (In this atypical case, localization does not require a band gap.) A Kronig-Penney–like model of two-dimensionally periodic Dirac delta-function dielectrics is also analytically solvable for both TE and TM polarizations (Shepherd and Roberts, 1995; Axmann et al., 1999). Another possibility, especially in the long-wavelength limit, is an *effective medium* approximation that replaces the photonic crystal by some average homogeneous material (Aspnes, 1982; Smith et al., 2005). Special methods are also available for defects formed by *slow* spatial variations in photonic crystals (Johnson et al., 2002*b*; Istrate and Sargent, 2006). The structures in this chapter and in most of the literature, however, require numerical computations using methods such as those summarized in appendix D.

[13] The references cited here for the separable structure use the opposite convention for naming polarization states: our TE is their TM, and vice versa.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

6

Three-Dimensional Photonic Crystals

THE OPTICAL ANALOGUE of an ordinary crystal is a three-dimensional photonic crystal: a dielectric structure that is periodic along three different axes. Three-dimensional photonic crystals can have the novel properties we discussed in the previous two chapters, including band gaps, defect modes, and surface states. In this chapter, we will present several examples of three-dimensional crystals with complete band gaps: a diamond lattice of air holes, a drilled dielectric known as *Yablonovite*, a *woodpile* stack of orthogonal dielectric columns, an *inverse opal*, and a stack of alternating two-dimensional crystals of rods and holes. Just as in two dimensions, we can localize light at a defect or at a surface, but with a three-dimensional crystal we have the additional capability to localize light in *all three* dimensions.

Three-Dimensional Lattices

Although there are an infinite number of possible geometries for a three-dimensional photonic crystal, we are especially interested in those geometries that promote the existence of photonic band gaps. The results of the previous chapter suggest that we try structures that contain connected networks of dielectric channels. More precisely, the two-dimensional crystal with a complete band gap had dielectric "veins" for the TE polarization (in the plane) and dielectric spots for the TM polarization (out of the plane). However, the spots were actually long channels running in the z direction, parallel to the TM field. So, more generally, we want a network of dielectric channels running along all the directions in which the electric field can point. In three dimensions, then, we might try creating our crystals with arrays of tubes and spheres, analogous to the veins and spots in the last chapter or to the bonds and atoms of crystalline atomic lattices. It turns out

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

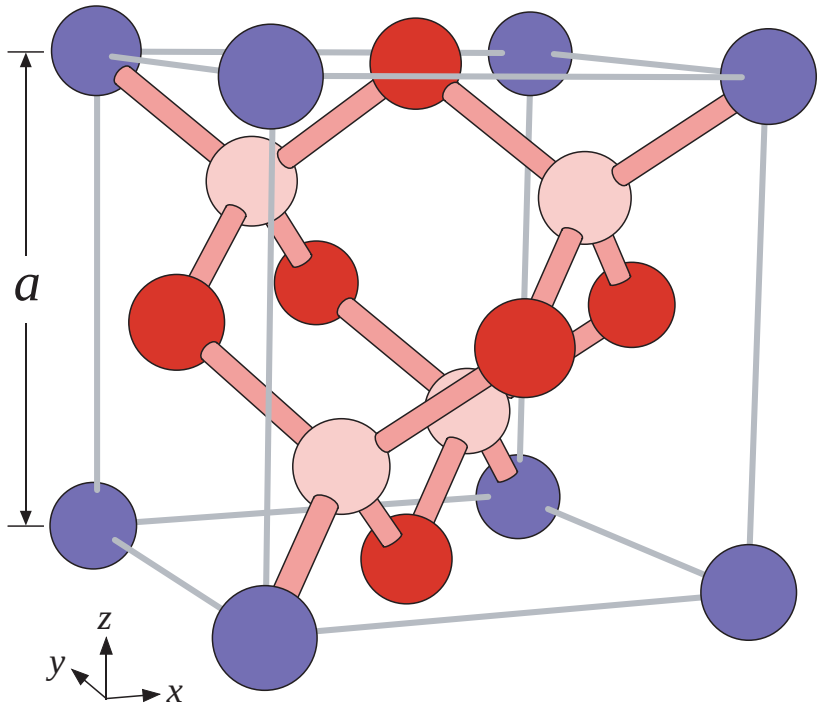


Figure 1: Ball-and-stick ("atomic") representation of several three-dimensional lattices in a cubic supercell, with a lattice constant a. The blue balls alone form a simple cubic lattice. Adding the dark red balls produces a face-centered cubic (fcc) lattice. Adding the pink balls as well produces a diamond lattice, with stick "bonds" (four bonds per ball).

that the choice of the lattice and how it is connected is critical in determining how easily we obtain a band gap. We will investigate several possibilities.

Figure 1 is a schematic representation of several three-dimensional lattices of spheres in a cubic cell. The simplest lattice is formed by the blue spheres at the corners of the cube, with primitive lattice vectors $a\hat{\mathbf{x}}$, $a\hat{\mathbf{y}}$, and $a\hat{\mathbf{z}}$. This is the **simple cubic** lattice. If we add the dark red spheres at the centers of the faces, we obtain a **face-centered cubic** (or **fcc**) lattice. The fcc lattice vectors are $(\hat{\mathbf{x}}+\hat{\mathbf{y}})a/2$, $(\hat{\mathbf{y}}+\hat{\mathbf{z}})a/2$, and $(\hat{\mathbf{x}}+\hat{\mathbf{z}})a/2$. For the fcc lattice, the smallest repeating unit (the **primitive cell**) is not the cubic cell shown in figure 1. Rather, the primitive cell is a rhombohedron (with volume $a^3/4$) whose edges are the three lattice vectors. The cubic cell contains four copies of this primitive cell, and is an example of a **supercell**.

Finally, if we add the pink spheres, which represent another fcc lattice that is shifted by $(a/4, a/4, a/4)$ relative to the blue spheres, then we obtain a **diamond lattice**. The periodicity in this case is the same as for the fcc lattice, but there are two "atoms" per rhombohedral primitive cell. In an atomic diamond

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

lattice, the spheres would be carbon atoms, with each atom bonding to its four nearest neighbors, as illustrated in the figure. Returning to the dielectric case, we imagine these bonds to be dielectric veins. They provide the diamond lattice with the requisite channels along which the electric field lines can run. In fact, we shall see that all known photonic crystals with large band gaps (15% or larger with a dielectric contrast of 13 to 1) are closely related to the diamond structure.

Consider a photonic crystal composed of only two different substances. We will examine two basic topologies, differing in whether the "atoms" and "bonds" are composed of the higher-ϵ material, or the lower-ϵ material. As we showed in chapter 2, it is the *ratio* of the two dielectric constants that matters most, not the individual values. Scaling the entire dielectric function by some constant factor, $\varepsilon(\mathbf{r}) \rightarrow \varepsilon(\mathbf{r})/s^2$, results in a trivial rescaling of the band structure, $\omega \rightarrow s\omega$. We define the **dielectric contrast** as the ratio of the dielectric constants of the high-ε and low-ε materials: $\varepsilon_{\text{high}}/\varepsilon_{\text{low}}$.

As we have seen in earlier chapters, band gaps tend to appear in structures with a high dielectric contrast. The more significant the scattering of light, the more likely a gap will open up. One might wonder whether *any* dielectric lattice has a photonic band gap for a sufficiently high dielectric contrast. This is in fact the case for most two-dimensional crystals, at least for one polarization (see appendix C).

For three-dimensional crystals, complete photonic band gaps are rarer. The gap must smother the entire three-dimensional Brillouin zone, not just any one plane or line. For example, in figure 2 we show the band structure for a face-centered cubic (fcc) lattice of close-packed (tangent) high-dielectric spheres ($\varepsilon = 13$) in air. Although the dielectric contrast is very large, there is no complete photonic band gap.

Nevertheless, a number of three-dimensional crystals have been discovered that do yield sizable complete photonic band gaps. Several of these structures are described in the next section. In most of the theoretical studies of these structures undertaken to date, the results are as follows. For a given crystal lattice, there is no photonic band gap until the dielectric contrast is increased to some threshold value. Above this nonzero threshold, the gap opens up and its width usually increases monotonically with the dielectric contrast, assuming optimal parameters are chosen. These optimal structural parameters (e.g., the radius of the tubes or spheres), which maximize the width of the band gap, vary with the dielectric contrast. Examples of such results can be found in appendix C.

Crystals with Complete Band Gaps

The possibility of three-dimensional photonic band gaps in periodic structures was first suggested by Yablonovitch in 1987, exactly one century after Lord Rayleigh (1887) described one-dimensional band gaps. It took three more years, however, before a specific dielectric structure was correctly predicted to have a complete band gap in three dimensions, by Ho et al. (1990). Subsequently, a large number of systems with band gaps have been proposed based on theoretical

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

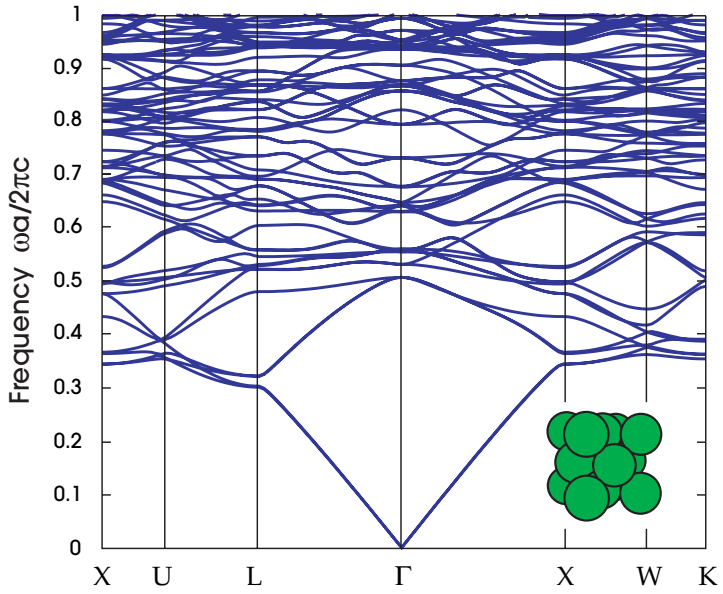


Figure 2: The photonic band structure for the lowest-frequency electromagnetic modes of a face-centered cubic (fcc) lattice of close-packed dielectric spheres ($\varepsilon = 13$) in air (inset). Note the *absence* of a complete photonic band gap. The wave vector varies across the irreducible Brillouin zone between the labelled high-symmetry points; see appendix B for a discussion of the Brillouin zone of an fcc lattice.

calculations, and in many cases these structures have been fabricated and characterized at wavelengths ranging from the microwave regime to the infrared regime.

In fact, there are too many designs for us to describe them all here, and we must concentrate on the essentials. Many of the successful designs share the same basic topology, and differ mainly in the envisioned fabrication method. We will discuss a few of the most historically important structures, and conclude with a detailed examination of a crystal that is closely related to the two-dimensional systems of the previous chapter.

Spheres in a diamond lattice

Ho et al. (1990) found the first structure with a complete three-dimensional photonic band gap by considering a diamond lattice of spheres, similar to the one shown in figure 1, except that the radius of each sphere is large enough to cause the spheres to overlap. This removes the need for any connecting "bonds." It was found that a complete photonic band gap exists whether one embeds dielectric spheres in air or air spheres in a dielectric medium, as long as the sphere radius is chosen appropriately.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

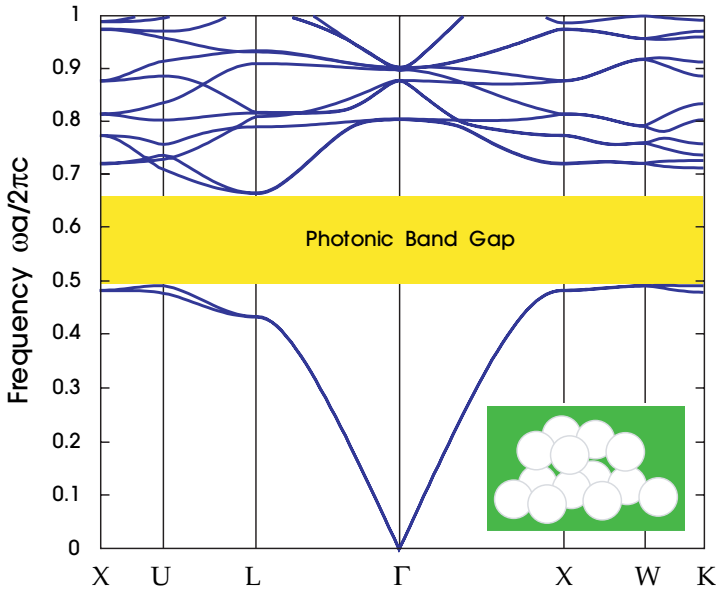


Figure 3: The photonic band structure for the lowest bands of a diamond lattice of air spheres in a high dielectric ($\varepsilon = 13$) material (inset). A complete photonic band gap is shown in yellow. The wave vector varies across the irreducible Brillouin zone between the labelled high-symmetry points; see appendix B for a discussion of the Brillouin zone of an fcc lattice.

The band structure for a lattice of air spheres within a dielectric medium is shown in figure 3. To maximize the size of the band gap, the sphere radius r is chosen to be $0.325a$, where a is the lattice constant of the cubic supercell (see figure 1). Between the second and third bands, there is a complete gap, with a gap–midgap ratio of 29.6%. A further survey of the gaps of this structure can be found in appendix C.

Most of this structure (81% by volume) is air. As noted above, the diameter of the spheres ($0.65a$) is larger than the distance between them ($a\sqrt{3}/4$), causing the air spheres to overlap. Both the air and the dielectric regions are connected; there are no isolated spots of either material. We can think of this crystal as two interpenetrating diamond lattices, one of which is composed of connected air spheres, and the other of which is composed of connected dielectric "remnants." These remnants, which follow the pattern of the "bonds" in figure 1, are the channels along which the electric field lines can run, for the modes in the two lowest bands. However, they are narrow enough that higher bands are forced out, and the corresponding frequency difference produces the photonic band gap.

Some further remarks are in order about why the diamond lattice is so amenable to complete photonic band gaps. In the section The Physical Origin of Photonic Band Gaps of chapter 4, when examining a one-dimensional crystal,

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

we found that a photonic band gap opens up at the Brillouin-zone edge ($\omega = c|\mathbf{k}| = c\pi/a$) even for arbitrarily small values of the dielectric contrast. Given this fact, one might expect that in three dimensions we could achieve the same property—a complete photonic band gap for a small dielectric contrast—if the edge of the Brillouin zone had the same magnitude $|\mathbf{k}|$ in all directions, corresponding to a *spherical* Brillouin zone. However, there is no three-dimensional crystal with a spherical Brillouin zone. It is generally a polygonal solid, just as it was a square or a hexagon in the two-dimensional examples considered in the previous chapter. Thus, the band gaps in different directions generally occur at different frequencies. Only if the dielectric contrast is large can we arrange for all of these directional gaps to be wide enough to create a mutual overlap. What distinguishes the fcc lattice (and the diamond lattice, which shares the same lattice vectors and Brillouin zone) is that the Brillouin zone is *nearly* spherical.[1] In some sense it is the most spherical of all possible three-dimensional lattices. Equivalently, the spatial period of the fcc lattice is most nearly independent of the spatial direction. This seems to be the key property that makes the fcc and diamond structures the most favorable cases for creating three-dimensional band gaps.

Yablonovite

The first case of a laboratory realization of a three-dimensional photonic crystal with a complete photonic band gap was not the simple diamond structure of the previous section. Rather, it was a related structure that proved to be simpler to manufacture. It consists of a dielectric medium that has been drilled along the three lattice vectors of the fcc lattice, as shown in figure 4. This has been named **Yablonovite**, after its discoverer Eli Yablonovitch et al. (1991*a*). Yablonovite was first fabricated on centimeter scales for measurements of microwave propagation.

Drilling holes with a radius of $0.234a$ results in a structure with a complete photonic band gap of 19%, as shown in figure 5. Like the diamond lattice of air spheres, we can think of Yablonovite as two interpenetrating "diamond-like" lattices, one of which is a connected region of dielectric, and the other being a connected air region.[2]

[1] More information on the fcc Brillouin zone can be found in appendix B.

[2] A note for cognoscenti: Yablonovite is not a *true* diamond structure, which would require drilling holes down the six axes [110], [101], [011], $[\bar{1}10]$, $[\bar{1}01]$ and $[\bar{0}11]$. (See the section Miller Indices of appendix B for the meaning of this notation.) By drilling along only three axes, the [110] direction is singled out. For this reason, Yablonovite does not have the full diamond symmetry, but only a D_{3d} symmetry (a threefold rotation axis [111] as well as mirror planes and inversion symmetry). The lowering of symmetry is responsible for broken degeneracies at some of the special points of the Brillouin zone. Nevertheless, Yablonovite is "diamond-like." It has the same topology as a diamond lattice. It consists of a set of dielectric veins that connect the sites of a diamond lattice. However, because only three rods are drilled, the veins along the [111] direction have a larger diameter than those along the $[\bar{1}\bar{1}1]$, $[1\bar{1}\bar{1}]$ and $[\bar{1}1\bar{1}]$ directions.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

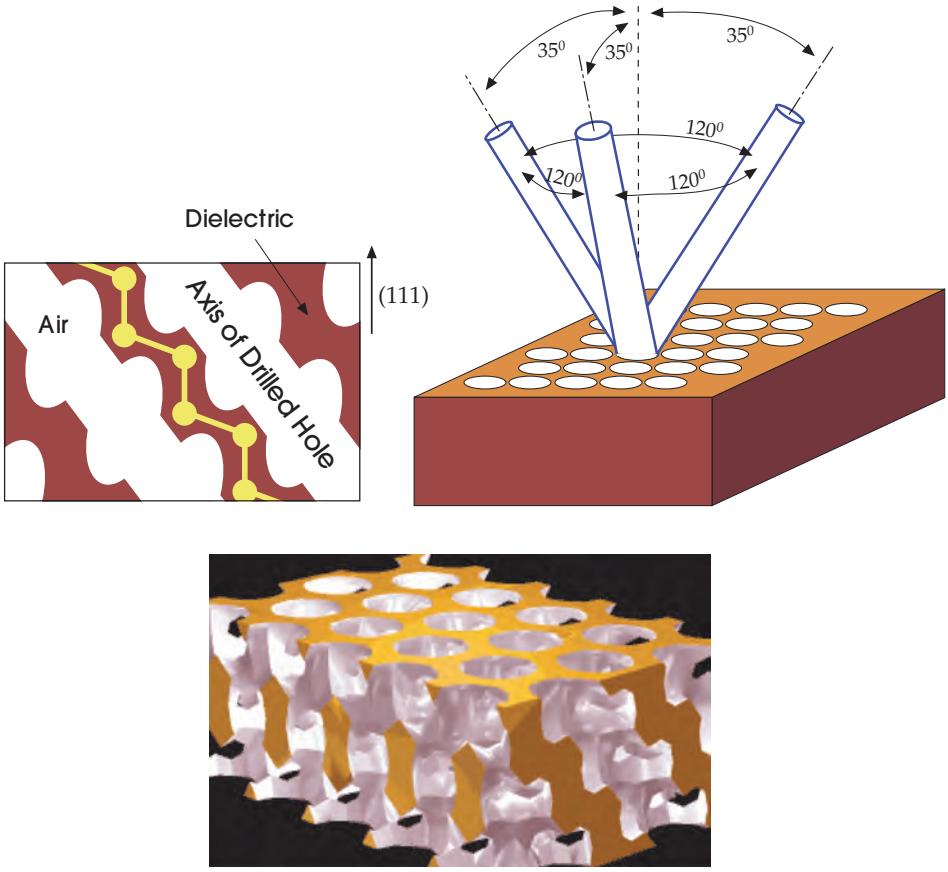


Figure 4: The method for constructing Yablonovite: a slab of dielectric is covered by a mask consisting of a triangular array of holes. Each hole is drilled three times (right), at an angle of 35.26° away from the normal and spread out 120° on the azimuth. This results in a three-dimensional structure whose $(1\bar{1}0)$ cross section is shown on the left. The dielectric connects the sites of a diamond lattice, shown schematically in yellow. The dielectric veins oriented vertically $[111]$ have greater width than those oriented diagonally $[11\bar{1}]$. *Bottom:* computer rendering of the structure (image courtesy E. Yablonovitch).

The woodpile crystal

The first three-dimensional photonic crystal with a complete gap to be fabricated on *micron* scales, for light at *infrared* wavelengths, was the crystal shown in figure 6. This structure was proposed independently by Ho et al. (1994) and Sözüer and Dowling (1994), and was dubbed a **woodpile** structure by the latter

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

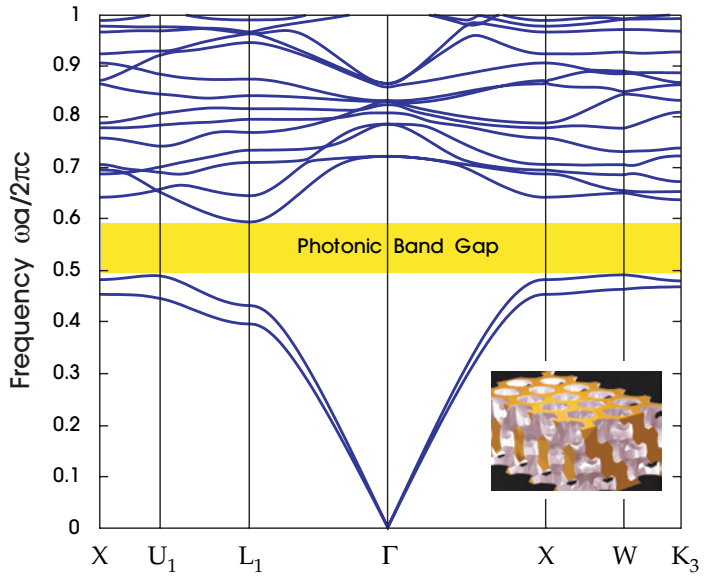


Figure 5: The photonic band structure for the lowest bands of Yablonovite (inset, from figure 4). Wave vectors are shown for a portion of the irreducible Brillouin zone that includes the edges of the complete gap (yellow). A detailed discussion of this band structure can be found in Yablonovitch et al. (1991*a*).

authors.[3] The woodpile crystal is formed by a stack of dielectric "logs" (generally rectangular) with alternating orthogonal orientations. The main advantage of the woodpile, as compared to crystals that had been proposed earlier, is that the woodpile can be fabricated as a sequence of layers deposited and patterned by lithographic techniques developed for the semiconductor electronics industry. Using just such a process, the woodpile structure was fabricated out of silicon ($\varepsilon \approx 12$) logs by Lin et al. (1998*b*) and a band gap was measured around a wavelength of 12 μm. Subsequently, Lin and Fleming (1999) were able to reduce the size by nearly a factor of eight, resulting in a band gap around a wavelength of 1.6 μm.

The simplest woodpile-like stack would be an $ABAB\cdots$ sequence, where A denotes one log orientation and B denotes the orthogonal orientation, but this sequence does not produce a significant gap. Instead, it turns out that a four-layer $ABCDABCD\cdots$ sequence is better, in which C and D are layers with the same orientation as A and B, but are *offset* by half of the horizontal spacing, as illustrated

[3] There was initially some debate over this terminology because the 1994 Sözüer and Dowling paper primarily described a different two-layer structure that had only a 2–3% complete gap. Ho et al. called the four-layer crystal here the "layer-by-layer" structure.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

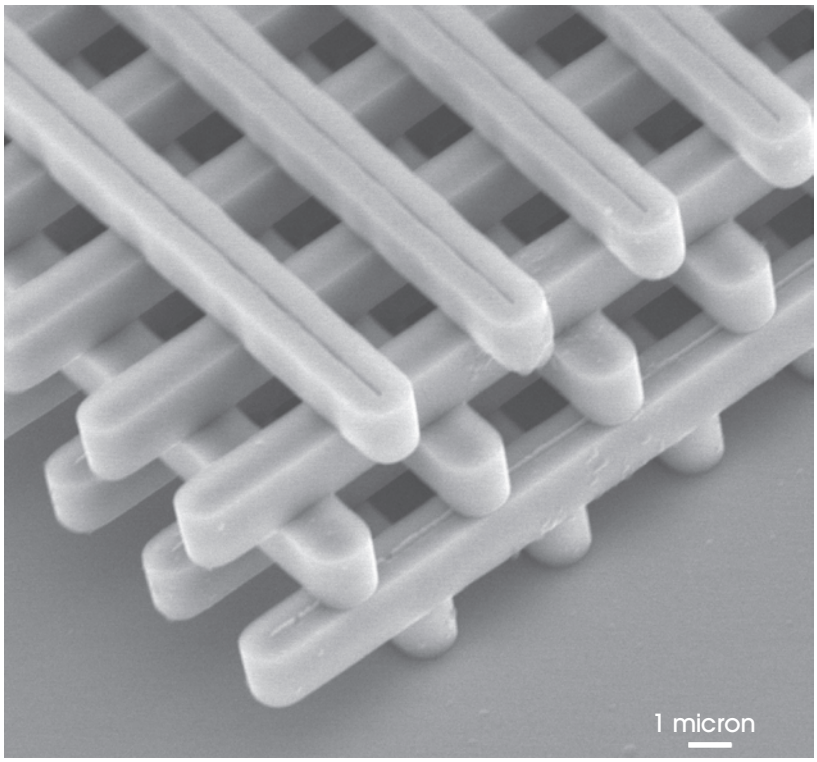


Figure 6: Electron-microscope image of a "woodpile" photonic crystal. The crystal is made of silicon and has a complete band gap centered at a wavelength of approximately 12 microns (Lin et al., 1998*b*). The dielectric "logs" form an fcc (or fct) lattice stacked in the [001] direction. (Image courtesy S.-Y. Lin.)

in figure 6. The band diagram of this structure, for a dielectric contrast of 13:1, is shown in figure 7. It exhibits a 19.5% complete photonic band gap between the second and third bands. This structure has the periodicity of an fcc lattice, in which each fcc "atom" from figure 1 is replaced by a pair of orthogonal logs in the $\hat{\mathbf{x}} + \hat{\mathbf{y}}$ [110] and $\hat{\mathbf{x}} - \hat{\mathbf{y}}$ [$1\bar{1}0$] directions.[4] In fact, the woodpile crystal, like Yablonovite, can also be understood as a distorted form of the diamond lattice (with lower symmetry). If we imagine taking the diamond of figure 1 and *flattening* the bonds so that they lie parallel to the xy plane, then we obtain a woodpile stacking.

[4] The periodicity is strictly fcc only when the log thickness (in z) is $a/4$ and the in-plane period of a sequence of parallel logs is exactly $a/\sqrt{2}$. More generally, it forms a face-centered tetragonal (fct) lattice, which can be thought of as an fcc lattice that has been stretched or compressed in the z [001] direction. In figure 7 we used an fcc lattice with a log width of $0.2a$, which is close to the optimum value for a dielectric contrast of 13:1.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

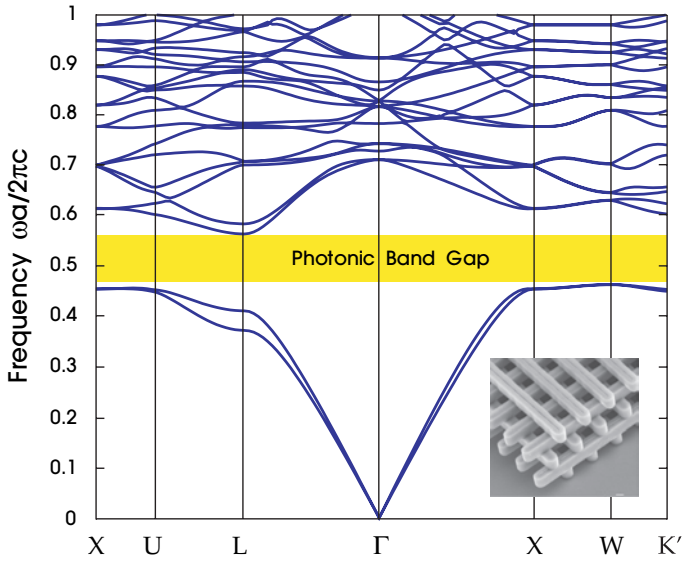


Figure 7: The photonic band structure for the lowest bands of the woodpile structure (inset, from figure 6) with $\varepsilon = 13$ logs in air. The irreducible Brillouin zone is larger than that of the fcc lattice described in appendix B, because of reduced symmetry—only a portion is shown, including the edges of the complete photonic band gap (yellow).

Inverse opals

In figure 2, we showed that an fcc lattice of dielectric spheres does *not* have a complete band gap. Nevertheless, it still has interesting optical properties that are responsible for the brilliant appearance of natural opals. Moreover, it turns out that the *inverse* structure, fcc air holes in high dielectric, *can* have a complete photonic band gap.

Using electron microscopy, Sanders (1964) found that precious opal mineraloids are formed of close-packed[5] arrangements of submicron-diameter silica spheres in a silica–water matrix, with a relatively low dielectric contrast. Just as for the case of an fcc lattice of close-packed dielectric spheres (see figure 2), small gaps appear only at particular points in the band diagram. The **k** vectors of these partial gaps[6] correspond to particular *directions* at which a particular wavelength, and therefore a particular color, is reflected. The narrowness and directionality of these gaps are the source of the bright, iridescent colors that make opal gems

[5] They can be arranged in an fcc lattice, but more often they are random hexagonal close-packed (rhcp) crystals. Defects (e.g. stacking faults) can give rise to visible striations in the color. See, e.g., Norris et al. (2004) for a review of natural and synthetic opal formation, and Graetsch (1994) for opaline minerals.

[6] These partial gaps also give rise to dips in the density of states, known as **pseudogaps**.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

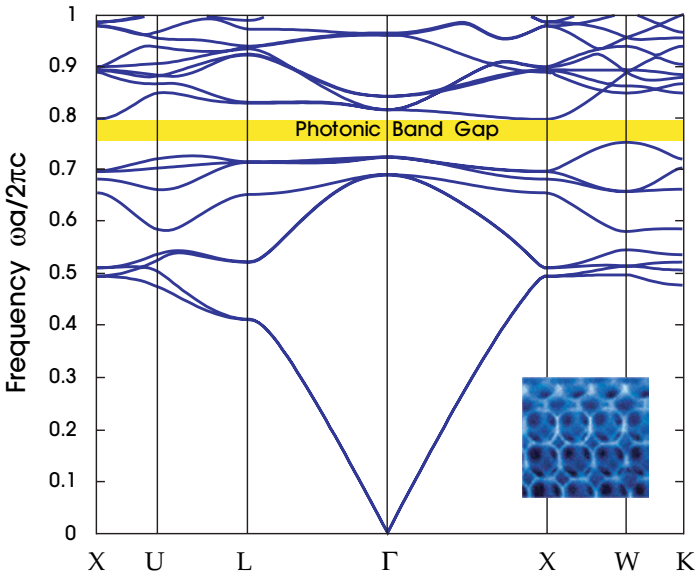


Figure 8: The photonic band structure for the lowest bands of an "inverse opal" structure: a face-centered cubic (fcc) lattice of close-packed air spheres in dielectric ($\varepsilon = 13$). (Inset shows fabricated structure from figure 9.) There is a complete photonic band gap (yellow) between the eighth and ninth bands. The wave vector varies across the irreducible Brillouin zone between the labelled high-symmetry points; see appendix B for a discussion of the Brillouin zone of an fcc lattice.

so attractive. This sort of **structural color** (as opposed to colors that results from absorption by chemical pigments) is responsible for many of the iridescent colors found in nature, from butterfly wings (e.g., Gralak et al., 2001; Biró et al., 2003) to peacock feathers (Zi et al., 2003) to certain beetles (Parker et al., 2003) and jellyfish (McPhedran et al., 2003).

A **synthetic opal** that is similar to natural opals can be readily fabricated, because microscopic spheres can be induced to *self-assemble* into an fcc lattice as one evaporates a solution in which they are suspended (a **colloid**). One then *inverts* the structure by infiltrating the space between the spheres with a high-dielectric material and then dissolving away the spheres, leaving behind an **inverse opal** of air holes. This structure can have a complete band gap, as first predicted by Sözüer et al. (1992). The band structure is shown in figure 8. This structure is an fcc lattice of air spheres that are just touching one another ($r = a/\sqrt{8}$), embedded in a dielectric medium with $\varepsilon = 13$. The gap in this case is *not* between the second and third bands, as it was for the diamond-based structures that were discussed previously. Instead, the gap is between the eighth and ninth bands and has only a 6% gap–midgap ratio.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

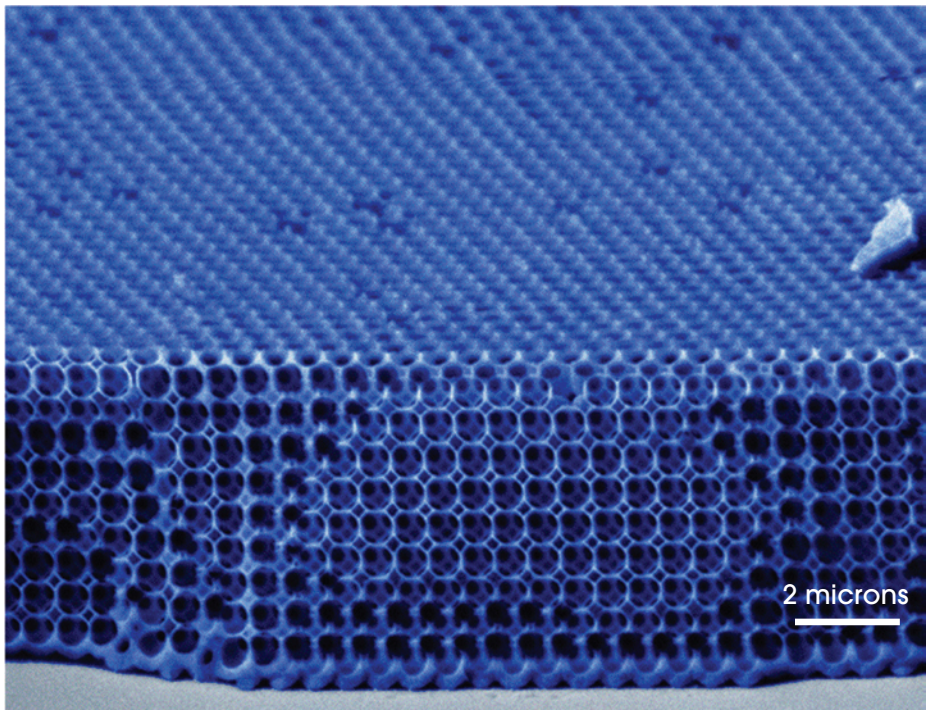


Figure 9: Electron-microscope image (artificial coloring) of inverse-opal structure demonstrated to have a complete band gap around a wavelength of 1.3 microns in Vlasov et al. (2001). Unlike figure 8, this is actually an fcc lattice of hollow dielectric (silicon) *shells*, which increases the size of the gap. (Image courtesy D. J. Norris.)

An inverse opal structure was first fabricated and demonstrated to have a complete band gap by Vlasov et al. (2001). An electron micrograph is shown in figure 9. The fabricated structure actually has a somewhat more complex geometry than that of figure 8: it is an fcc lattice of overlapping hollow spherical *shells*, with pores where they overlap. These shells are a byproduct of the fabrication method. As suggested by Busch and John (1998), the pores act to increase the band gap. For a dielectric contrast of 13:1, one would obtain a 13% gap for a shell inner radius of $0.36a$ (overlapping) and thickness of $0.07a$.

A stack of two-dimensional crystals

The final three-dimensional crystal that we will discuss is also formed of a stacked sequence of layers, like the woodpile. It has a somewhat larger gap than the woodpile, and it repeats every three layers ($ABCABC\cdots$) rather than every four. Our main reason for discussing this structure, however, is that it is particularly

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

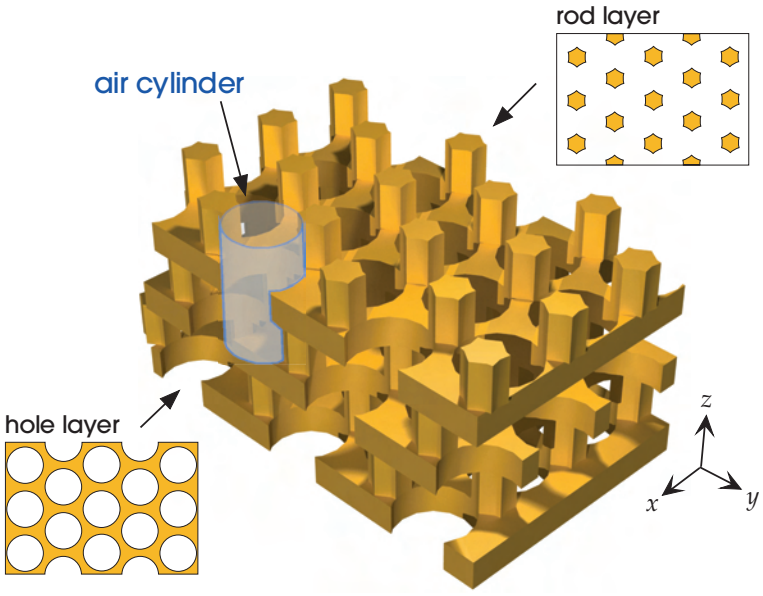


Figure 10: Three-dimensional photonic crystal formed by a stack of layers with two-dimensional cross sections: triangular lattices of dielectric rods in air (upper-right inset) and holes in dielectric (lower-left inset). These layers are paired in bilayers with an $ABCABC\cdots$ stacking, and are formed from a face-centered cubic (fcc) lattice of overlapping air cylinders in dielectric (one is highlighted) oriented along the [111] direction.

simple to visualize and understand: it consists of a stack of finite-thickness "two-dimensional" rod and hole crystals analogous to those of chapter 5.[7]

The structure is shown in figure 10 along with its horizontal cross sections, which fall into two categories: **rod layers**, which are a triangular lattice of high-dielectric rods in air; and **hole layers**, which are a triangular lattice of cylindrical air holes in high dielectric. For a dielectric contrast of 12:1 (similar to common semiconductor materials at infrared wavelengths[8]), the structure has a 21% complete photonic band gap, as shown in the band diagram of figure 11.[9] Again, this structure turns out to be a distorted diamond lattice that is closely related to Yablonovite, and hence the gap is between the second and third bands. To see that it is diamond-like, think of the spot between three holes in a hole layer as

[7] The band gap of this structure was described by Johnson and Joannopoulos (2000).

[8] At $\lambda \approx 1.5\,\mu$m, the dielectric constants are about 11.4 for gallium arsenide and 12.1 for silicon (Si) (Palik, 1998).

[9] For this 12:1 dielectric contrast, the gap can be increased to 25% for an fcc lattice of hexagonal cylinders, and to 27% if one fabricates circular hole and rod layers separately and optimizes their radii independently.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

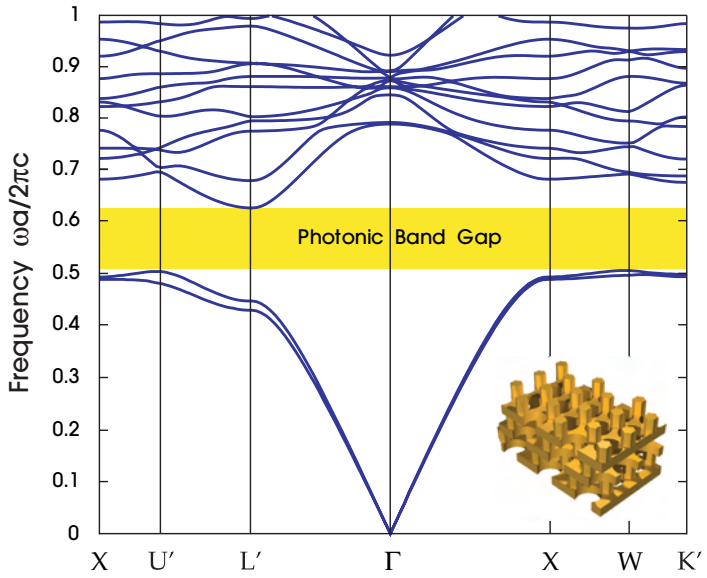


Figure 11: The photonic band structure for the lowest bands of the layered structure from figure 10 (inset). The irreducible Brillouin zone is larger than that of the fcc lattice described in appendix B, because of reduced symmetry—only a portion is shown, including the edges of the complete photonic band gap (yellow). A more detailed discussion of this band structure can be found in Johnson and Joannopoulos (2000).

an "atom," with four "bonds" formed by the three in-plane veins to neighboring spots and the one rod either under or over the spot.

The rods have a strange-looking shape because they are formed from the *remnant* left over from six overlapping cylindrical air holes. The entire structure is an fcc lattice of overlapping air cylinders (with height $h = 0.93a$ and radius $r = 0.293a$) in a dielectric medium with $\varepsilon = 12$.[10] The $ABCABC\cdots$ stacking of these cylinders is shown in figure 12, in which the stacking direction is the body-diagonal [111] direction of the fcc lattice shown in figure 1. In this way, a rod and hole bilayer is formed from each layer of cylindrical air holes. Precisely such a fabrication process was employed to create this structure for infrared wavelengths by Qi et al. (2004), as shown in the bottom panel of figure 12.

From the previous chapter, we know that each individual rod layer resembles a two-dimensional photonic crystal with a TM band gap. Likewise, each individual hole layer resembles a two-dimensional photonic crystal with a TE band gap. We will explore this idea in the subsequent sections. The resulting electromagnetic

[10] More generally, one can employ a *trigonal* lattice formed by stretching or compressing the structure along the [111] direction.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

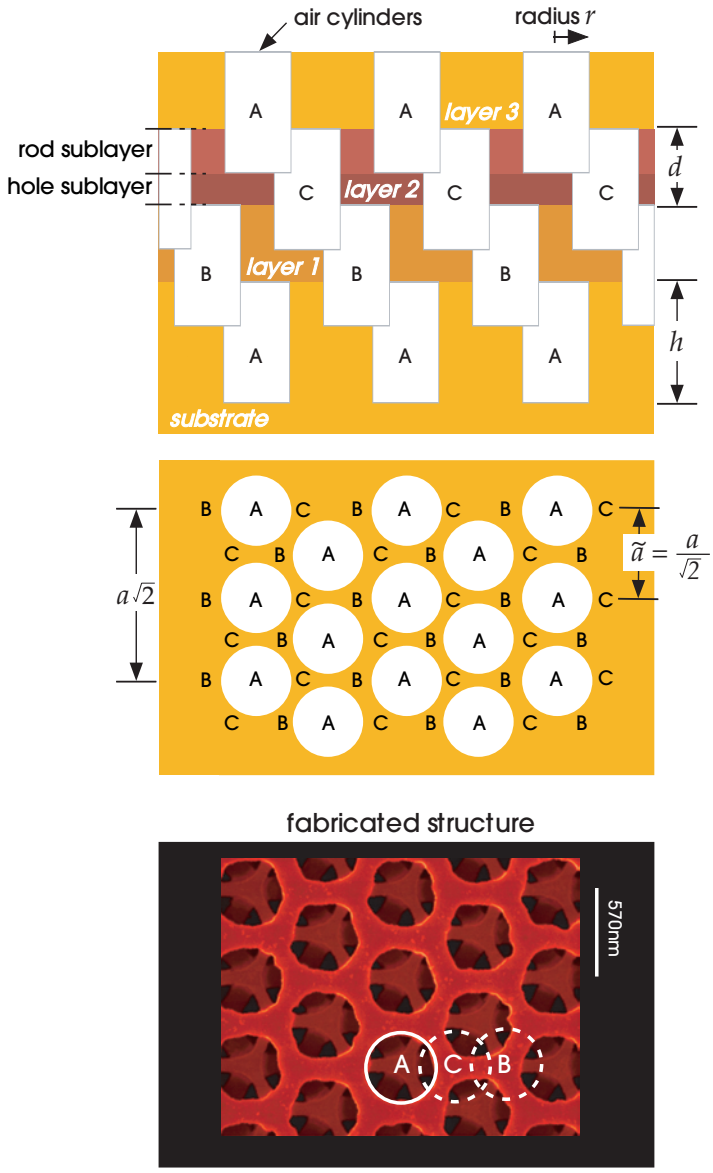


Figure 12: *Top left:* vertical cross section of layered structure from figure 10, showing air cylinders in an fcc lattice forming an ABC stacking in the $[111]$ direction. The structure is fabricated by depositing a sequence of dielectric layers with thickness $d = a/\sqrt{3}$, where a is the fcc lattice constant, into each of which is drilled a lattice of holes offset as shown. Rod/hole sublayer cross sections are formed by regions indicated, where adjacent holes do/don't overlap, respectively. *Middle:* horizontal cross section of a hole sublayer A, showing triangular lattice of holes and offsets of holes in subsequent layers. The horizontal lattice constant is $\tilde{a} = a/\sqrt{2}$. *Bottom:* top view of silicon structure fabricated by Qi et al. (2004) for a gap around 1.3 microns, with three layers of holes visible (artificial coloring).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

modes are directly related to their two-dimensional analogues, which makes them easier to visualize and understand than most other cases of three-dimensional crystals with complete band gaps.

Localization at a Point Defect

Now that we have reviewed several structures that have photonic band gaps, we can discuss some of the novel features that result. We have already seen that defects in photonic crystals can localize light modes. In one dimension, this meant we could confine light to a single plane. In two dimensions, we could localize light to a single line, which can also be considered a single point in the xy plane. In three dimensions, we can perturb a single lattice site, and thereby localize light to a single point in the crystal. It is trapped in all three dimensions. As in the previous chapter, the point defect pulls a state from the continuum above or below the gap *into* the gap itself, resulting in a localized mode.

All of the three-dimensional photonic crystals reviewed in the previous sections support localized modes at point defects. For pedagogical purposes, we will focus on the crystal from the subsection A stack of two-dimensional crystals of chapter 6. To create a point defect, we will modify a *single layer* of either rods or holes in exactly the same way as we created defects in two-dimensional crystals. It turns out that the resulting localized mode has a cross section that closely resembles either the TM or TE defect state from the two-dimensional rod or hole lattice, respectively. This is *not* true for more general point defects or for other three-dimensional crystals; typically, a point-defect mode will have a complex three-dimensional field pattern. Rather, it is a peculiarity of this particular structure that we exploit here in order to make the defect modes easier to study.[11]

Two simple ways to perturb a single lattice site are to add extra dielectric material where it does not belong, or to remove some of the dielectric material that should be there. We refer to the first case as a **dielectric defect** and the second case as an **air defect**. Examples of both are illustrated in figure 13, in which we have removed a single rod in a rod layer to create an air defect (left), and we have increased its radius to create a dielectric defect (right). These are the same types of defects that were considered in the section Localization of Light by Point Defects of chapter 5, and our discussion will simply develop and extend those earlier results.

By inserting a point defect, we ruin the discrete translational symmetry of the lattice. Strictly speaking, we can no longer classify the modes of the system with a wave vector $\mathbf{k}$. Instead, we examine the crystal's density of states. The defect creates a single peak of newly allowed states into the density of states, at a frequency that may be engineered to lie in the photonic band gap. The width of this peak tends to zero as the crystal size tends to infinity. Since no extended states are allowed in the crystal with frequencies within the band gap, the new peak

[11] The 2D-like defect modes of this crystal were first described by Povinelli et al. (2001).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

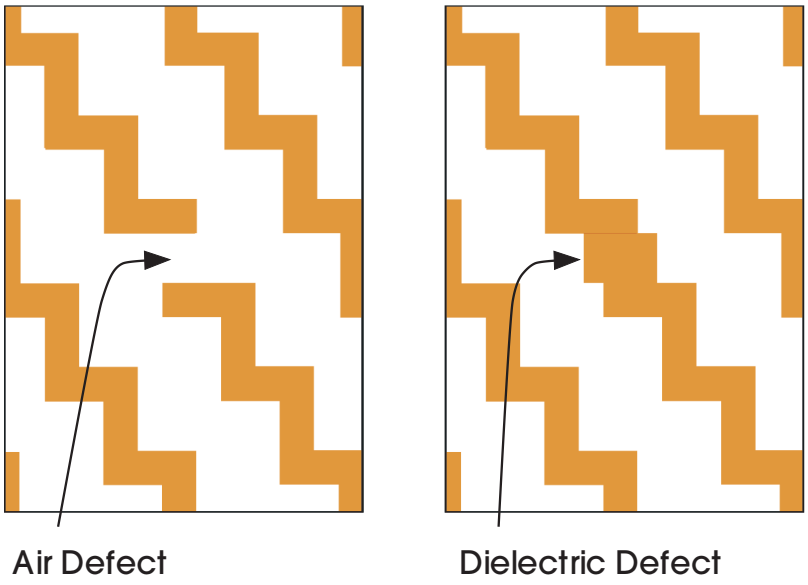


Figure 13: Vertical cross section of the layered structure from figure 10, corresponding to the top panel of figure 12, showing examples of how a point defect may be created by modifying a *single rod*: a rod can be removed (left) to form an **air defect**, or the radius of a rod can be increased (right) to form a **dielectric defect**.

must consist of localized states. Modes with frequencies in the band gap decay exponentially away from the defect. But in this case they decay exponentially in all *three* dimensions: they are trapped near a single point.

Why does a defect localize electromagnetic modes? According to the intuitive picture developed in earlier chapters, the defect is like a cavity with perfectly reflecting walls. If light with a frequency within the band gap somehow winds up near the defect, it cannot leave, because the crystal does not allow extended states at that frequency. Therefore if the defect allows a mode to be excited with a frequency within the band gap, that mode is forever trapped.

The defects create localized modes within the photonic band gap. The air defect of figure 13 introduces a single, nondegenerate state into the photonic band gap, which crosses from the dielectric to the air band as the defect frequency is raised. The field patterns of this state are shown in the top panels of figure 14. The cross section corresponding to the mid-plane of the rods has a *monopole* pattern localized to the defect, with a **TM-like** electric field polarized mostly perpendicular to the cross section, just like the corresponding two-dimensional defect in the section Localization of Light by Point Defects of chapter 5. In fact, the two-dimensional defect with *exactly* the same cross section is shown at the top right of figure 13, and underlines the similarity of the three-dimensional field pattern. Unlike the

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

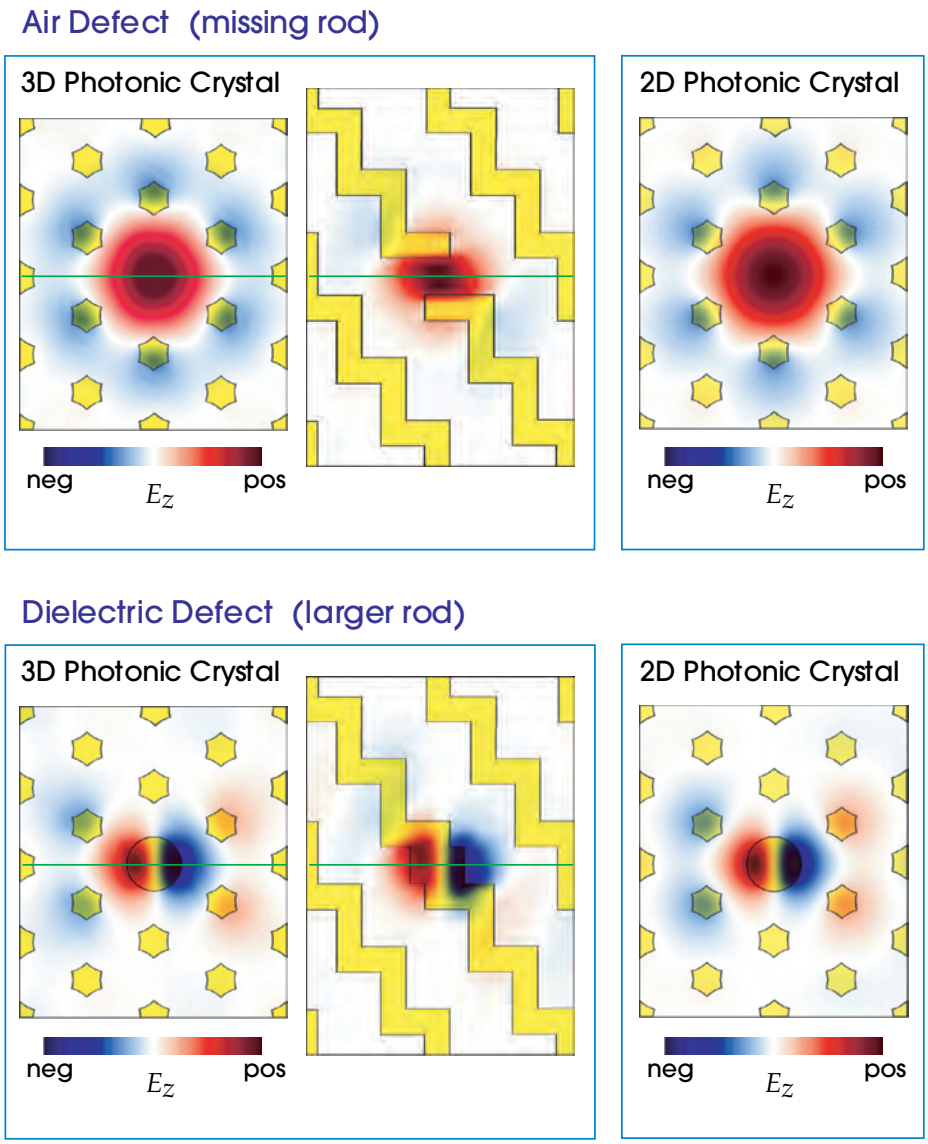


Figure 14: Horizontal and vertical cross sections (intersecting at green lines) of E_z field patterns of point defects in the layered structure (left), formed by modifying a single rod as in figure 13, compared to the field for the corresponding TM state in a *two*-dimensional crystal (right) with the same *rod* cross section. Dielectric material is shown in yellow. *Top:* nondegenerate **monopole** state trapped by completely removing a rod. *Bottom:* dielectric defect formed by replacing a rod with an enlarged dielectric cylinder of radius $0.35\tilde{a}$ (where $\tilde{a}$ is the in-plane lattice constant as in figure 12). This is a doubly degenerate **dipole** state, where the degenerate partner is *roughly* its 90° rotation.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

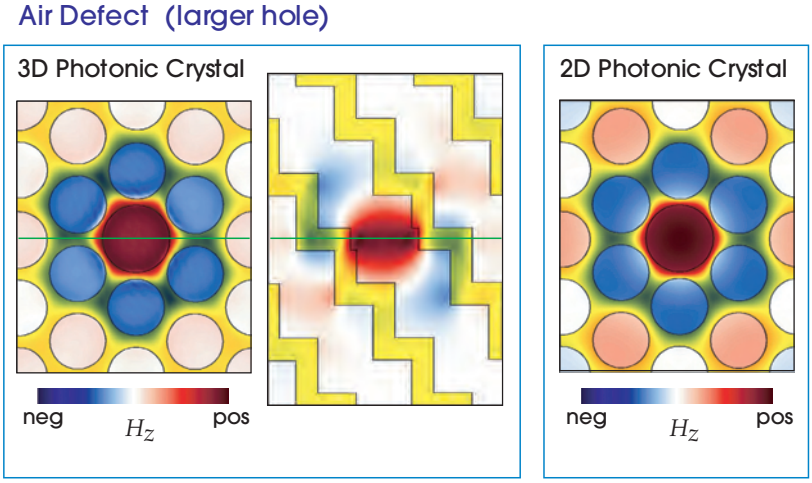


Figure 15: Horizontal and vertical cross sections (intersecting at green lines) of H_z field patterns of point defects in the layered structure from figure 10 (left), formed by modifying increasing the radius of a single hole in a hole layer, compared to the field for the corresponding TE state in a *two*-dimensional crystal (right) with the same *hole* cross section. Dielectric material is shown in yellow.

two-dimensional state, however, the vertical cross section in the top middle of figure 14 demonstrates that the field is localized vertically as well (although it is no longer polarized outside of the mid-plane). Similarly, the dielectric defect of figure 13 introduces a doubly degenerate TM-like *dipole* state into the gap, shown in the bottom of figure 14 along with its two-dimensional counterpart.[12]

In contrast, creating an air defect by increasing the radius of a single *hole* layer, as shown in figure 15, results in a state resembling a localized state in a two-dimensional triangular lattice of air holes (see the section A Complete Band Gap for All Polarizations of chapter 5). In two dimensions, the hole structure most easily creates a TE gap. Likewise, if we look at the cross section of this three-dimensional defect in the mid-plane of the hole layer, we find that it has a **TE-like** magnetic field mostly perpendicular to the plane. Again, its cross section closely resembles the TE mode we obtain in the corresponding two-dimensional

[12] The dipole degeneracy in this case is less intuitive than in the section Localization of Light by Point Defects of chapter 5, because the structure has 120° rotational symmetry, corresponding to the C_{3v} symmetry group, but *not* 90° rotational symmetry. (A close mechanical analogue is the well-known fact that a triangle spinning about its center has two orthogonal principal axes with equal moments of inertia.) Thus, the degenerate partner of the dipole mode in figure 14 is *not* simply its 90° rotation. Nevertheless, the C_{3v} group *does* admit a 2×2 irreducible representation corresponding to doubly degenerate dipole-like states that are *roughly* 90° rotations. More precisely, the orthogonal state is the difference of +120° and −120° rotations of the other state. See Inui et al. (1996, Table B.15).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

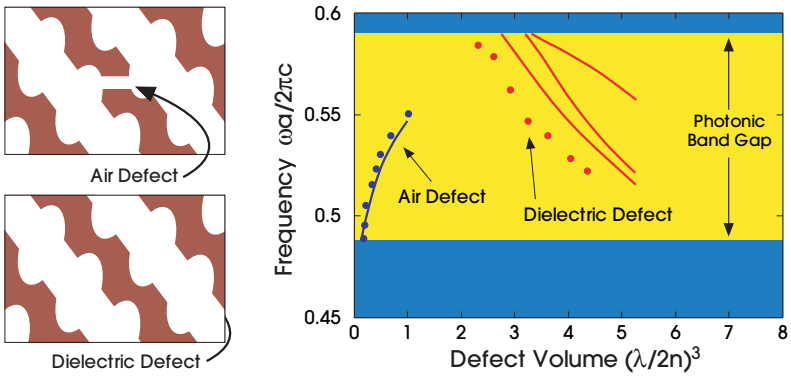


Figure 16: Plotted are the frequencies of the localized modes of Yablonovite as the defect size varies. The dots indicate measured values (Yablonovitch et al., 1991*b*), the lines indicate computed values (Meade et al., 1993), and the yellow region is the photonic band gap. The modes on the blue line (leftmost) result from an air defect (top left), while the modes on the red lines (rightmost) result from a dielectric defect (bottom left). The strength of the defect is expressed in terms of the volume of dielectric added or removed, in units of $(\lambda/2n)^3$, where λ is the midgap vacuum wavelength and $n = \sqrt{\varepsilon}$ is the refractive index.

crystal (shown in the right panel of figure 15), but is exponentially localized in the vertical dimension as well.

Although a point defect destroys the translational symmetry of the lattice, it can be chosen to retain point symmetries about its center. For instance, in the defects of figures 14 and 15, we see that our defects have been chosen to preserve the threefold rotation symmetry of the crystal.[13] We could still classify the modes, including the new localized modes, by how they transform under a 120° rotation in the plane of the picture. It is also possible to engineer defects that preserve other symmetries, such as mirror reflections and inversions.

Experimental defect modes in Yablonovite

Localized modes similar to the ones discussed in the previous section have been realized experimentally. The earliest example of this was for defects in Yablonovite, the structure described in the section Yablonovite. The left panel of figure 16 depicts the Yablonovite defects (both air and dielectric) that were studied. By systematically measuring microwave transmission levels, the experimenters mapped the frequencies of the defects, which are shown along with the theoretical values in the right-hand panel of figure 16.

[13] Although the horizontal cross sections may seem to have sixfold rotation symmetry, this is reduced to threefold by the layers above and below, as can be seen from the middle panel of figure 12.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Figure 16 shows that the defect frequency is an increasing function of the volume of the air defect (removing a slice of a dielectric vein), and a decreasing function of the volume of the dielectric defect. We found a similar result in the two-dimensional case, and the reason is the same (see the section Localization of Light by Point Defects of chapter 5). The dielectric defect (adding a dielectric sphere), on the other hand, introduces states that cross in the opposite way, from the air band to the dielectric band.

There is one significant difference in the three-dimensional case, however: the defect must be larger than some critical size to localize light. As the defect is increased from zero size, it must pass through some nonzero threshold before its localizing power begins. In contrast, for one-dimensional and two-dimensional crystals, it typically appears that even arbitrarily small defects can localize modes. This is the electromagnetic analogue of a famous theorem of quantum mechanics. The theorem states that an arbitrarily weak attractive potential can bind a state in one and two dimensions, but not in three dimensions (see, e.g., Simon, 1976; Yang and de Llano, 1989).

Localization at a Linear Defect

Another class of defects are **linear** defects, which modify a sequence of unit cells in one direction to form a *waveguide*. The basic principles of linear defects in three-dimensional photonic crystals are the same as those described in the section Linear Defects and Waveguides of chapter 5 for two-dimensional crystals, except that we can now confine light in *both* directions transverse to the waveguide.

Linear-defect waveguides are possible to create in any structure with a complete band gap, but again we focus on the crystal of the subsection A stack of two-dimensional crystals of chapter 6 because of its resemblance to the simpler two-dimensional systems. In particular, just as we removed a row of rods to create a two-dimensional waveguide in the section Linear Defects and Waveguides of chapter 5, here we will remove a row of rods from a single rod layer of the three-dimensional crystal. As shown in figure 17, this results in a waveguide state (left panel) much like that of corresponding two-dimensional structure (right panel): it is mostly TM-polarized in the mid-plane of the rods and is confined primarily in the *air* of the defect row. Here, however, the mode is also confined in the vertical direction, exponentially decaying away from the waveguide.

Just as in two dimensions, the critical feature of a waveguide is that it retains *one* direction of translational symmetry and thus *one* conserved wave vector component k_x along the waveguide. This means that, unlike the point defect, we can understand the waveguide mode by calculating a band diagram $\omega(k_x)$, as shown in figure 18. As in the section Linear Defects and Waveguides of chapter 5, figure 18 includes continuous regions which are the *projection* of all the extended states propagating in the bulk region. The discrete guided band is localized

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

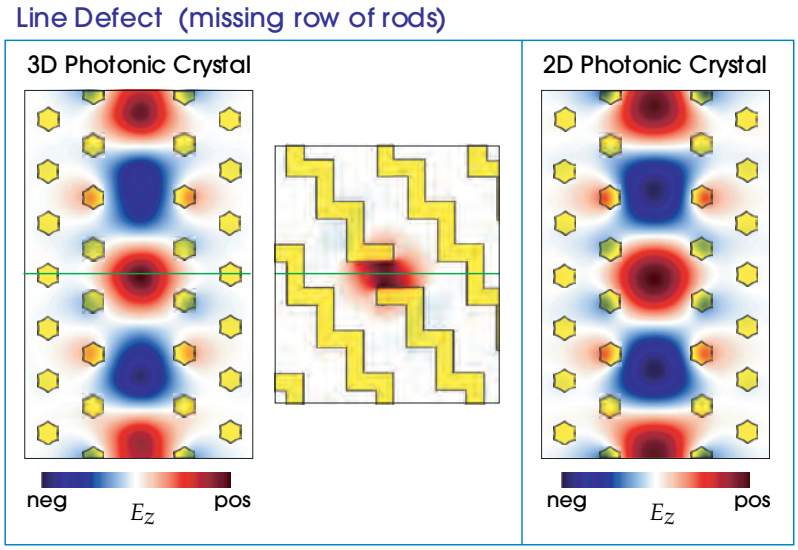


Figure 17: Horizontal and vertical cross sections (intersecting at green lines) of E_z field patterns of a linear defect in the layered structure (left), formed by removing a row of rods from a single rod layer, compared to the field for the corresponding TM state in a *two*-dimensional crystal (right) with the same *rod* cross section. These fields correspond to a wave vector $k_x\tilde{a}/2\pi = 0.3$ along the waveguide, where $\tilde{a}$ is the in-plane lattice constant from figure 12. Dielectric material is shown in yellow.

because it lies in the band gap, where it cannot couple to any of these extended modes.

Such states transport electromagnetic energy along the line defect with a *group velocity* $d\omega/dk_x$ (as in the section Bloch-Wave Propagation Velocity of chapter 3). In this way, linear defects are analogous to metallic waveguides: light is trapped in a tube with perfectly reflecting walls. Like in a hollow metallic tube, the waveguide mode depicted in figure 18 exhibits an interesting feature: there is a point at the edge of the band where the slope goes to zero. Thus, it becomes a **slow light** waveguide in which the velocity of light can be arbitrarily slowed, in principle.

It is tempting to think of the linear defect as a sequence of point defects. The example of figure 17 is only crudely described by such a picture, however, because the subsequent defects are adjacent. This causes the modes in each defect to strongly influence one another, and the field pattern deviates significantly from that of isolated point defects. On the other hand, a sequence of point defects separated by one or more unit cells of unperturbed crystal will couple only through their exponential *tails*, and their fields indeed closely match those of the isolated defects. Because there is still periodicity, it is still a waveguide, but the

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

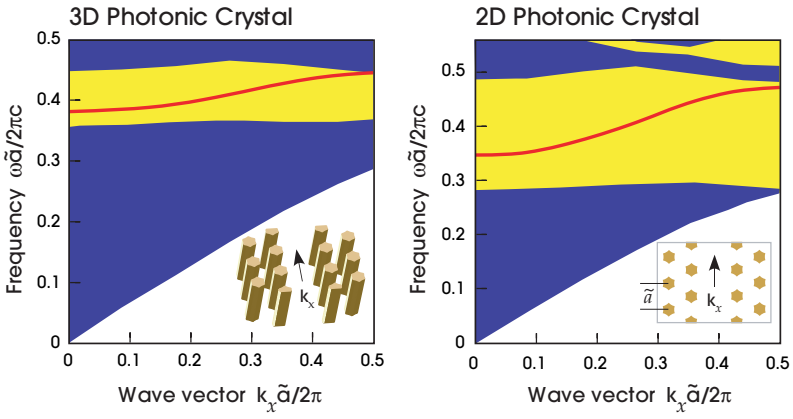


Figure 18: *Left:* Projected band diagram of the line defect from figure 17, formed by a missing row of rods from a single layer of the three-dimensional crystal from figure 10. The red line is the guided band in the complete photonic band gap (yellow). Inset shows a single rod layer with the defect (other layers of the crystal not shown). *Right:* Projected band diagram of TM states for the corresponding *two*-dimensional crystal with an identical rod cross section (inset).

group velocity decreases exponentially with the defect separation. This is known as a **coupled-cavity waveguide**.[14]

Localization at the Surface

Finally, as in the preceding chapters, we will discuss the effect of terminating a photonic crystal by examining its surface band structure. Again we will search for electromagnetic surface modes, which are localized to a surface plane. They are localized in one direction by the band gap of the bulk crystal, and in the other direction by index guiding. In fact, although we focus on surfaces, these types of localized states may also occur at planar defects within the crystal.

Suppose we terminate a three-dimensional photonic crystal in the z direction. By doing this, we destroy the translational symmetry in that direction, and we can no longer classify the states of the crystal with a definite k_z. However, the crystal still has translational symmetry parallel to the surface, and the electromagnetic modes *do* have a definite $\mathbf{k}_\parallel$. We must project the full three-dimensional band structure onto the surface Brillouin zone, with the procedure we described in the last section of chapter 5.

[14] Such waveguides were proposed by Yariv et al. (1999); see also the section Some Other Possibilities of chapter 10.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

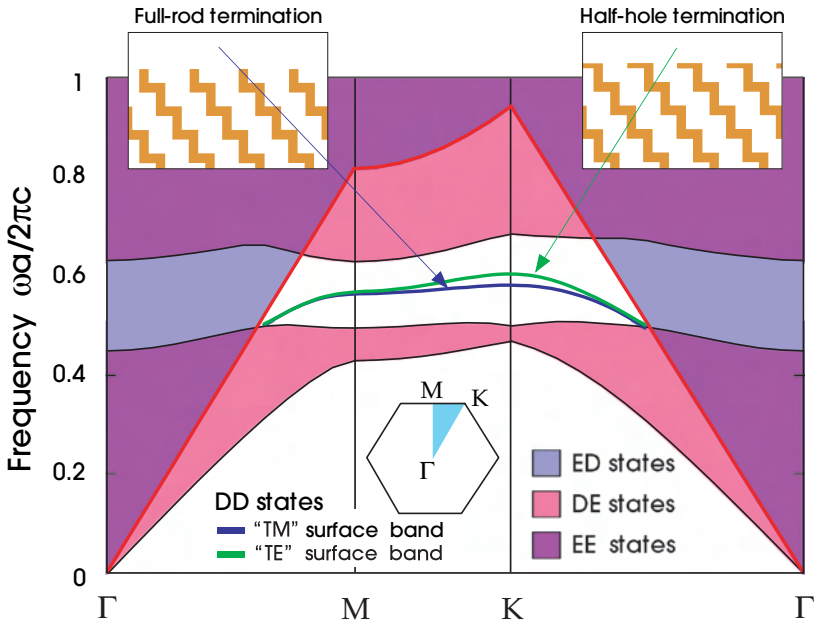


Figure 19: The band structure of the (111) surface of the layered structure from figure 10, along special directions in the surface irreducible Brillouin zone (center inset). The shading denotes regions in which light is transmitted (purple, EE), internally reflected (pink, DE), and externally reflected (blue, ED). The lines in the gap correspond to DD surface bands in which light is localized at the surface, for two terminations. The green line (upper) is the "TE-like" band that results when the surface is terminated with half of a hole layer (top right inset). The blue line (lower) is the "TM-like" band that results when the surface is terminated with a full rod layer (top left inset).

We illustrate our discussion of the general features of surface states by truncating the photonic crystal of the subsection A stack of two-dimensional crystals of chapter 6 in one of its rod or hole layers. This is a (111) surface, in the language of Miller indices as described in appendix B. Note that this system is invariant under 120° rotations about an axis through one of the rods or holes. Before we consider the band structure of the interface, let us first consider the projected band structures of the air and photonic crystals separately. As before, we use an "E" to stand for regions of extended states, and a "D" for regions of decaying states. States are labelled by either E or D on both the air and the crystal sides.

For example, the union of regions EE and ED shown in figure 19 is the projection of the extended air states ($E\_$) onto the surface Brillouin zone of the crystal. For a given $\mathbf{k}_\parallel$, there are extended light modes for all frequencies $\omega \geq c|\mathbf{k}_\parallel|$, which is the *light cone* first described in the subsection Index guiding of chapter 3. Along the *light line* $\omega = c|\mathbf{k}_\parallel|$, the light travels parallel to the surface,

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

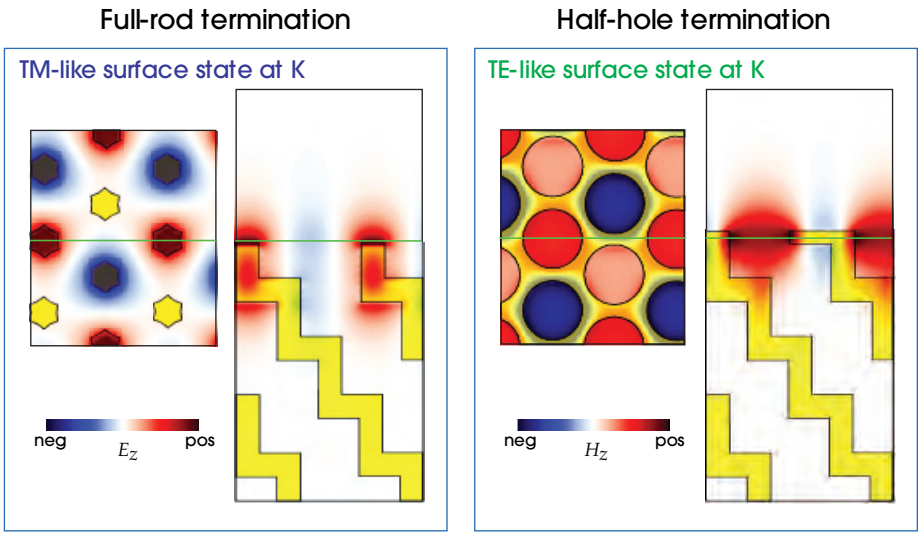


Figure 20: Horizontal and vertical cross sections (intersecting at green lines) of field patterns for the two crystal terminations from figure 19, at the K point. Dielectric material is shown in yellow. *Left:* E_z for "TM-like" band when the surface is terminated with a full rod layer. *Right:* H_z for "TE-like" band when the surface is terminated with half of a hole layer.

and increasing ω corresponds to increasing angles out of the plane. Likewise, the union of regions EE and DE represents the projected band structure of the photonic crystal (_E). Note that this photonic crystal has a complete gap at $0.509 < \omega a/(2\pi c) < 0.630$, in which no extended states are allowed for any $\mathbf{k}_\parallel$.

Thus, we divide the modes into four different categories: transmitted (EE), internally reflected (DE), externally reflected (ED), and surface modes (DD). In the region of $(\mathbf{k}_\parallel, \omega)$ marked EE, the modes are extended in both the air and in the dielectric, so it is possible for incident light to traverse the crystal. In the DE region, modes extend in the crystal, but they are beneath the band edge for air states. In the ED region, the situation is reversed; there are extended states in the air but they lie in the gap of the crystal. Finally, in the region marked DD, the states are below the band edge of the light in the air, as well as in the gap of the crystal. The light decays exponentially in both directions away from the surface plane.

In figure 19, we show the bands for two possible surface terminations: one in which *half* of a hole layer is left on top of the crystal, and one in which a *full* rod layer is left on top. The Brillouin zone of $\mathbf{k}_\parallel$ is the same as that of the triangular lattice of the section A Complete Band Gap for All Polarizations of chapter 5, with the K point corresponding to the nearest-neighbor direction. The fields for these two terminations at the K point is shown are shown in figure 20, and exhibit the expected localization at the surface, in either the surface rod or hole layer. Just as for the point and line defects of the previous section, the mode that is

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

localized in the rod layer is TM-like in its mid-plane, and the mode that is localized in the hole layer is TE-like in its mid-plane. For this reason, we plot only the vertical (z) component of **E** for the rod-layer mode, and the vertical component of **H** for the hole-layer mode. As usual, most of the energy of the field is concentrated in the high-ε regions.

Until now, we have considered only the particular choices of surface terminations shown in the inset to figure 19. However, the surface can be terminated in a variety of ways, as shown in the insets to figure 21. We must specify not only the inclination of the surface plane, but also where in the unit cell we will cut the crystal. In figure 21, we expand the treatment of figure 19 to include multiple possible "cuts" of the rod or hole layers.

In the top panel of figure 21, we show the Γ–K band diagram for different cuts of a *rod* layer: when 1/4, 1/2, 3/4, or a whole rod (as in figure 19) is left on top of the crystal. In all of these cases, there is a "TM-like" surface state localized in the rods, whose frequency *decreases* as we add *more* dielectric on top of the crystal—the usual inverse relationship that we first learned from the variational principle in the section Electromagnetic Energy and the Variational Principle of chapter 2.

We can also look at how different terminations influence the "TE-like" state localized in the top-most *hole* layer. The bottom panel of figure 21 shows the TE-like mode's Γ–K band diagram when 1/4, 1/2, or 3/4 of a hole layer is left on top of the crystal. Again, the frequency decreases as we add more dielectric.

There are two especially interesting terminations that we can observe from the bottom panel of figure 21. First, the case where an *entire* hole layer is left on top of the crystal is omitted, because this termination has *no* surface states. Second, when 1/4 of a hole layer is left on top, there are *two* surface states, one TM-like in the rod layer and one TE-like in the hole layer.

By raising the surface termination starting from an entire hole layer (no surface states), gradually adding a rod layer, then gradually adding another hole layer, we arrive back at the original case of no surface states. In the meantime, we have swept first a TM-like and then a TE-like state through the gap. Overall, we have added one bulk unit cell per surface unit cell, and have increased the total number of states in the dielectric band by two. This is true for each point in the Brillouin zone.

This suggests a general claim: *for a crystal with a band gap, and a surface of a given inclination, we can always find some termination that allows localized surface modes.* We sketch the argument briefly. Since the crystal as a whole has a gap, the surface Brillouin zone must also have a gap. As before, we increase the termination continuously until we arrive back at the original termination geometry, having added b crystal unit cells introduced per surface unit cell. There must then be $2b$ new states transferred from the air band to dielectric band. As the frequencies of these states decrease from the bottom of the air band to the top of the dielectric band, they sweep through the gap. These are localized surface states. A similar argument for the existence of surface states also applies to the cases of multilayer films and two-dimensional photonic crystals in the previous chapters.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

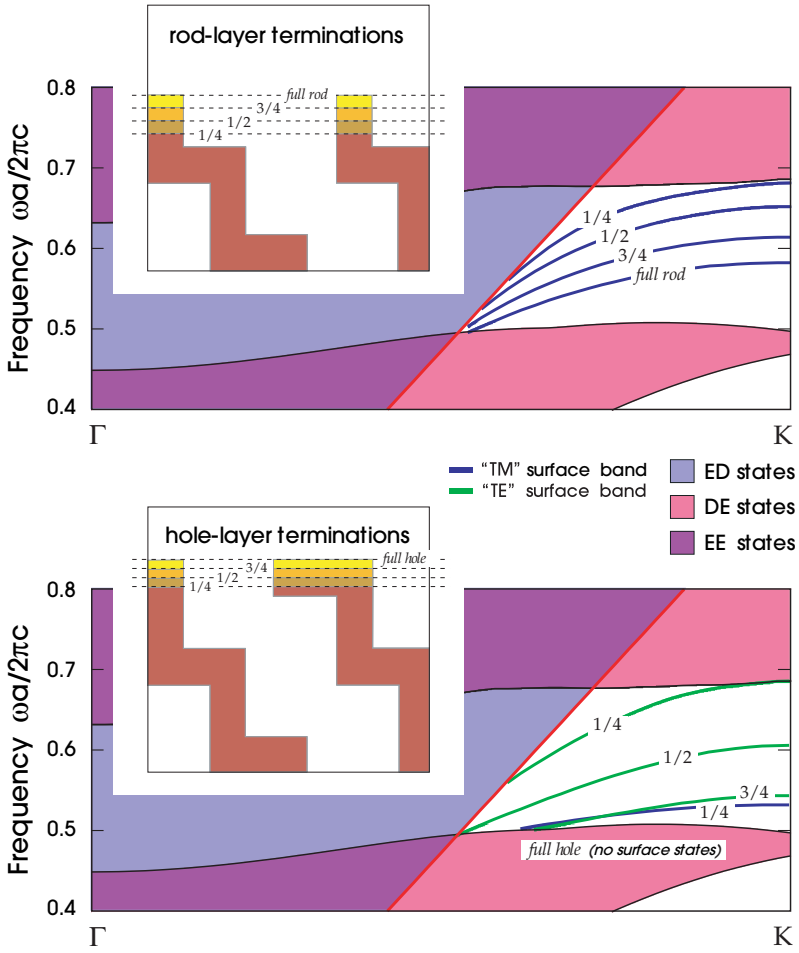


Figure 21: The Γ–K band structures of the (111) surface of the layered structure from figure 10, for various terminations ending at a rod layer (top) or a hole layer (bottom). The figure zooms in on the gap region of figure 19, using the same classification scheme *EE* etc. The insets indicate the corresponding surface terminations. The rod-layer terminations (top) all support a "TM-like" surface state localized in the topmost rods (blue lines). The hole-layer terminations (bottom) localize a "TE-like" state in the topmost hole layer (green lines); leaving 1/4 of a hole layer also supports a "TM-like" surface band in the adjacent rods (bottom-most blue line). As the surface termination increases in the vertical direction, dielectric material is added and the corresponding bands are pulled *down* in frequency.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Further Reading

Appendix B describes the reciprocal lattice and the Brillouin zone in more detail, including the reciprocal lattice of the face-centered cubic lattice. Occasionally in this chapter we used Miller indices to refer to crystal planes and axes. Interested readers can find a brief description of this notation in appendix B. For a more complete treatment of these topics, consult the first few chapters of a text on solid-state physics (for example, Kittel, 1996).

The first suggestion that a material with a complete photonic band gap in three dimensions might be fabricated appeared in Yablonovitch (1987), along with a related suggestion shortly afterwards by John (1987). The history of the discovery of the first realistic materials with a complete photonic band gap can be followed in Yablonovitch and Gmitter (1989), Satpathy et al. (1990), Leung and Liu (1990), Zhang and Satpathy (1990), Ho et al. (1990), Chan et al. (1991), and Yablonovitch et al. (1991*a*). Subsequently, many other three-dimensional structures with complete gaps were proposed. Most of these were based on distorted diamond lattices, as reviewed in Maldovan and Thomas (2004). Non-diamondlike structures with complete gaps include inverse opals, body-centered cubic (bcc) structures (Chutinan and Noda, 1998; Maldovan et al., 2002; Luo et al., 2002*a*), and even a small gap for a simple-cubic lattice (Sözüer and Haus, 1993). Proposed fabrication techniques have ranged from the drilled holes, layer-by-layer lithography, and colloidal self-assembly mentioned earlier in this chapter, to holographic lithography (Sharp et al., 2002), glancing-angle "spiral" deposition (Toader and John, 2001), block copolymers (Fink et al., 1999*a*), and nanorobotic stacking of planar layers (Aoki et al., 2002) or microspheres (Garcia-Santamaria et al., 2002), among other variations.

Yablonovitch (1987) also suggested that a crystal with a complete three-dimensional photonic band gap would totally inhibit the spontaneous emission of atoms inside the crystal, for the frequency range of the gap. Experimental observations of partial inhibition at visible wavelengths have subsequently been reported in three-dimensional photonic crystals by several groups (Martorell and Lawandy, 1990; Petrov et al., 1998; Lodahl et al., 2004).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

7

Periodic Dielectric Waveguides

As we saw in chapter 6, a three-dimensional photonic crystal can confine light in all three dimensions. Suitably engineered materials can pin light to the vicinity of a single point (an optical cavity), direct it along a particular line (a waveguide), or trap it at a two-dimensional surface. Nevertheless, the actual fabrication of a structure that is periodic in all three dimensions remains difficult. In this chapter, we will examine how far we can get by relying on simpler structures: **periodic dielectric waveguides**, which have only a *one*-dimensionally periodic pattern (or *grating*) along the direction of propagation, but have a finite thickness and a finite width.

Overview

Many periodic-waveguide structures are possible, such as those shown in figure 1. It will turn out that, regardless of the geometry, all such structures exhibit common phenomena: they have a form of photonic band gap along their periodic direction, and can confine light in the other directions by the principle of *index guiding*.

The next two chapters will examine different forms of such hybrid systems, all of which use a combination of periodicity and other mechanisms to confine light in three dimensions. Chapter 8 will look at periodic *planar* waveguides known as **photonic-crystal slabs**, which use *two*-dimensional periodicity combined with vertical index-guiding. Chapter 9 examines **photonic-crystal fibers**, which are waveguides in which the periodicity is *transverse* to the direction of propagation.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

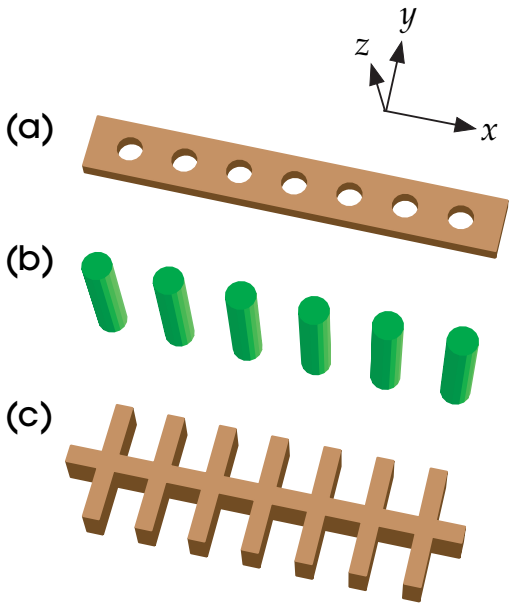


Figure 1: Examples of **periodic dielectric waveguides**, which combine one-dimensional periodicity (in x) and index-guiding in two transverse directions.

A Two-Dimensional Model

Although the real motivation for this chapter is confinement of light in three dimensions, we will begin by considering a simpler, *two-dimensional* model that captures the essential physics. We will combine index guiding in one direction with a photonic band gap in the other direction.

Consider a strip of material extending in the x direction, as shown in figure 2(a). The strip confines light in the y direction by index guiding, as described in the subsection Index guiding of chapter 3, but it is *uniform* in the z direction. As in chapter 5, we can restrict ourselves to light propagating in the xy plane ($k_z = 0$), and we will furthermore restrict ourselves to the TM polarization (E_z only). Next, we will introduce periodic interruptions in the strip, along the x direction, to form a set of dielectric squares, as shown in figure 2(b). For specificity, suppose that the material has a dielectric constant $\varepsilon = 12$, the spatial period of the squares is a, and each square is $0.4a \times 0.4a$ in size.[1]

Both structures have translational symmetry in the x direction: the uniform strip has continuous translational symmetry, and the line of squares has discrete

[1] Early analyses of a similar structure can be found in Meade et al. (1994), Fan et al. (1995*a*), and Atkin et al. (1996).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

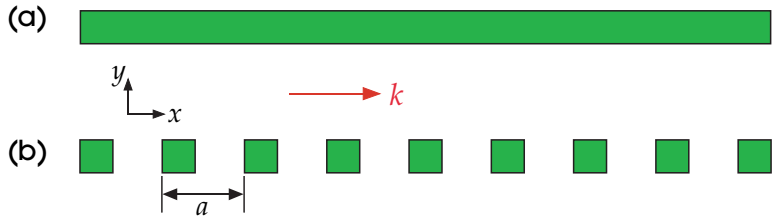


Figure 2: (a) Two-dimensional dielectric waveguide ($\varepsilon = 12$) of width $0.4a$. (b) Periodic waveguide: a period-a sequence of $0.4a \times 0.4a$ dielectric squares. In both (a) and (b) there is a conserved wave vector k along the direction x of translational symmetry, resulting in guided modes.

translational symmetry. Neither structure has translational symmetry in the y direction. Thus, k_x is conserved, but k_y is not, and we will simplify the notation by referring to k_x simply as k.[2] As discussed in the section Linear Defects and Waveguides of chapter 5, it is useful to compute the **projected band structure** $\omega_n(k)$, by which we mean that the mode frequencies are plotted as a function of k only (even though they are really a function of the full vector **k** for modes far from the waveguide).

First, consider the uniform strip, for which the projected band diagram is shown in the left panel of figure 3. Only the range $0 < k < 2\pi/a$ is shown, even though k is unrestricted because of the continuous translational symmetry. As in the section Index guiding we refer to the region $\omega \geq ck$ as the **light cone**, in which there exist extended states propagating in the air. Beneath the light cone, the higher index of the waveguide pulls down discrete **guided modes** that must be localized. As discussed previously, this is essentially total internal reflection.

The waveguide is symmetric under reflections through the plane $y = 0$ that bisects it. Consequently, all of the guided modes can be classified as *even* or *odd* with respect to mirror reflections in this plane. (It might seem that there are additional mirror planes perpendicular to the waveguide axis, but these symmetries are broken for $k \neq 0$.) In figure 3, we see one even band and one odd band. The even band is the **fundamental mode**, for which the mode profile has the fewest nodes and the lowest frequency.

Next, consider the discontiguous strip of figure 2(b). It might seem impossible to rely on total internal reflection to guide light in the x direction, because light rays cannot stay inside the dielectric squares, let alone maintain an angle of incidence smaller than the critical angle. Likewise, according to standard waveguide lore,

[2] In the waveguide literature, the wave vector k is often referred to as the **wave number** or the **propagation constant** and is commonly denoted by β. Furthermore, some authors use k to signify ω/c, also called the wave number; this should not be confused with our definition of k as the wave vector.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

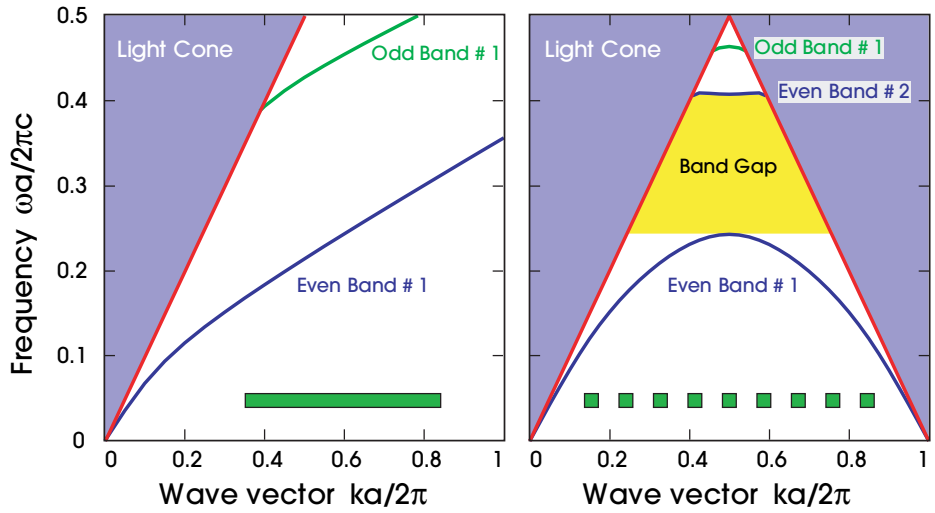


Figure 3: Band diagrams of waveguides from figure 2, for TM-polarized in-plane ($k_z = 0$) light only. *Left:* uniform waveguide of figure 2(a). *Right:* periodic waveguide of figure 2(b), including twice the irreducible Brillouin zone. Blue shaded region is light cone (extended states propagating in air). Discrete guided bands are labelled *even* or *odd* according to the $y = 0$ mirror symmetry plane.

junctions between two different waveguides are best avoided because they result in radiative scattering and losses. How much worse must this structure be, with an infinite sequence of junctions! However, these reasons for pessimism are mistaken, because they fail to take into account Bloch's theorem: *a periodic structure need not scatter waves*. In particular, the periodicity of the system guarantees that the wave vector k is still conserved, and therefore there is still a light cone beneath which can exist localized bands. These bands are seen in the projected band diagram of figure 3. They are truly guided modes: they will propagate along the waveguide indefinitely.

Unlike the uniform strip, however, the Brillouin zone for the discontiguous strip is finite, and there is an interesting consequence. The range $\pi/a < k < 2\pi/a$ is equivalent to the range $-\pi/a < k < 0$, which in turn is the reverse of the $0 < k < \pi/a$ irreducible Brillouin zone. There must be a light cone within each of these zones. The tip of the original light cone, which occurred at $k = 0$ for the uniform strip, is now repeated periodically at $k = 2\pi/a, 4\pi/a, \cdots$. Similarly, the lowest band, starting at zero frequency from $k = 0$, must *flatten* at $k = \pi/a$ and then bend downwards to return to zero frequency at $k = 2\pi/a$. This causes a band gap to open up between the first two guided modes, just as in the one-dimensional crystals we considered in chapter 4. In this case, however, the gap is considered **incomplete**, because only the *guided* modes are excluded from the gap, whereas *radiating* modes (the modes within the light cone) exist for any ω.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

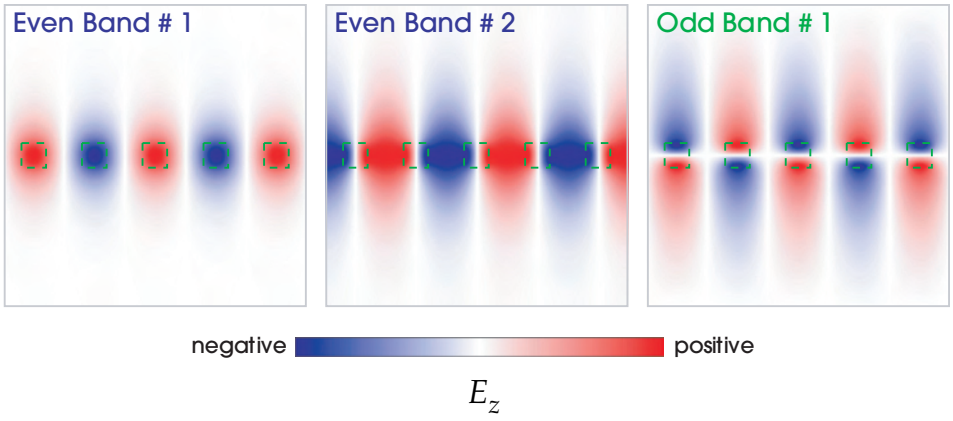


Figure 4: E_z field patterns of the periodic waveguide from figure 2(b) at $k = \pi/a$, the Brillouin-zone edge, labelled as in figure 3. Left and middle panels correspond to the edges of the gap in the even guided modes, while the right panel has odd symmetry with respect to the $y = 0$ mirror plane. The dielectric squares are shown as dashed green lines.

(Recall that the concepts of light cones and index-guided modes in periodic structures also appeared in the study of *surface states* in chapters 5 and 6.)

The field profiles for the three guided modes at $k = \pi/a$ are shown in figure 4. The two even-symmetry modes are reminiscent of the modes considered in the section The Physical Origin of Photonic Band Gaps of chapter 4: the lower band is peaked in the dielectric, and the next-higher band has a node in the dielectric. The odd band has a nodal line along the x axis, within the dielectric, which naturally raises its frequency. It is less tightly confined to the waveguide than the even modes, because it is closer to the light cone. One might wonder why there is no second odd band, with two nodal lines in each block. The answer is that the frequency of such a state is high enough to push it into the light cone; it is not guided. Periodic dielectric waveguides have only a finite number of guided bands, whereas a uniform dielectric waveguide usually has an *infinite* number of guided bands.[3]

Indeed, the periodic replication of the light cone enforces an *upper frequency cutoff* of $\omega = c\pi/a$ for guided modes.[4] This means that, for short wavelengths where ray optics is valid, our intuitive understanding is restored: total internal reflection *cannot* guide light along a periodic structure *in the ray-optics limit*.

[3] An exception is described in chapter 9.

[4] This upper frequency cutoff can be violated in some special cases, such as for a special choice of materials, polarization, and geometry yielding a separable structure (Čtyroký, 2001; Kawakami, 2002; Watts et al., 2002). Also, lossless standing-wave modes can exist above the light line precisely at the Γ point ($\mathbf{k} = 0$) in structures with sufficient symmetry (Paddon and Young, 2000; Ochiai and Sakoda, 2001; Fan and Joannopoulos, 2002). Finally, there can exist *leaky* modes above the light line that propagate with a small rate of radiative losses.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Periodic Dielectric Waveguides in Three Dimensions

Now that we understand the basic principles of periodic dielectric waveguides, it is straightforward to apply them to three-dimensional structures. In particular, we will examine the waveguide of figure 1(a), a dielectric strip in which we have punched a sequence of cylindrical air holes. For concreteness, the holes have a spacing a and radius $0.25a$, and the strip has a dielectric constant of 12, width a, and thickness $0.4a$. For now, we imagine the waveguide to be suspended in air.[5] (Later in this chapter, we will consider the effects of placing the waveguide on a substrate.)

The projected band diagram of this structure is shown in figure 5. Only the range of wave vectors within the irreducible Brillouin zone is shown. As before, there is a conserved wave vector k along the direction of periodicity, a light cone $\omega \geq ck$ of the extended states in air, and discrete guided bands below the light line. The guided bands are more numerous and complex than in the previous example, because in this case we have included *all* the modes, whereas previously we considered only the TM polarization.

The three-dimensional waveguide has *two* mirror-symmetry planes: $z = 0$ (perpendicular to the hole axes) and $y = 0$ (parallel to the hole axes). This allows us to classify all the modes as either even or odd with respect to *both* types of reflection, z and y. We refer to the z-even modes as "E" modes, because these modes are TE-like, as explained below. Likewise, the z-odd modes are labelled "M" because they are TM-like. We attach a subscript to the label, "e" or "o," to specify whether the mode is even or odd under y reflection, and we attach another subscript, n, to identify the band number. For example, the second band of modes that are z-even and y-odd is labelled $\mathrm{E}_{(\mathrm{o},2)}$.

If we consider each symmetry type individually (e.g. only the $\mathrm{E}_{(\mathrm{o},n)}$ bands), then there are band gaps, as before. The largest gap is between $\mathrm{E}_{(\mathrm{o},1)}$ and $\mathrm{E}_{(\mathrm{o},2)}$, with a 21% gap–midgap ratio. As we will see in the next section, the modes in these bands are strongly localized. The dominant component of the magnetic field is H_z, which is pictured in figure 6.

Symmetry and Polarization

In the two-dimensional structures of chapter 5, there is a fundamental distinction between the TM and TE polarizations. It is common for a structure to have a gap for modes of one polarization, but not the other. In three dimensions, the electromagnetic modes cannot generally be divided into two distinct polarizations. However, for the special case of *thin* structures with a mirror symmetry, the fields can be said to be *mostly* polarized, a concept we will now explore.

[5] A similar design was proposed by Villeneuve et al. (1995) and later fabricated by Ripin et al. (1999).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

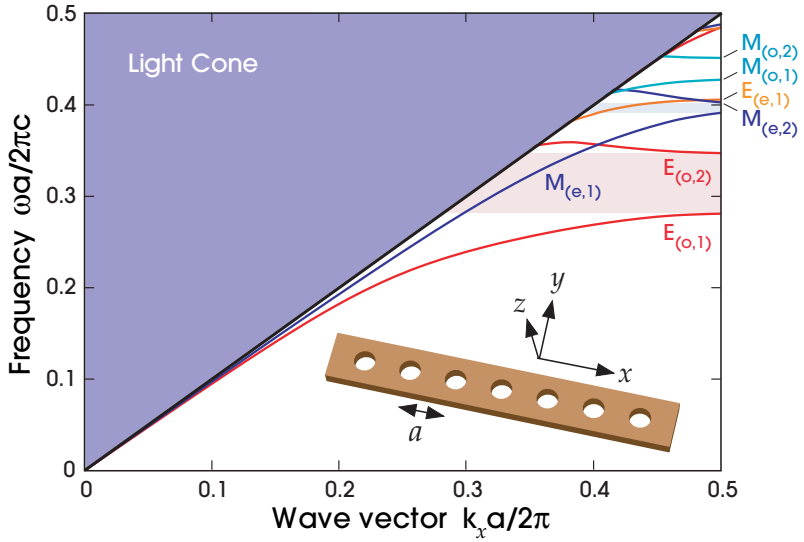


Figure 5: Band diagram for the waveguide of figure 1(a) (inset): a three-dimensional dielectric strip, suspended in air, with a period-a sequence of cylindrical air holes. Only the irreducible Brillouin zone is shown. The discrete guided modes are labelled according to their symmetry as described in the text, with the fundamental E and M band gaps shaded light red and blue, respectively (the light cone is shaded darker blue, bounded by the light line in black).

Returning to the example of the previous section, consider the electric field profiles of the modes that are either even or odd with respect to z reflections. These field profiles are depicted schematically in figure 7. Within the symmetry plane ($z = 0$), the even and odd modes must be purely TE and TM polarized, respectively. If we look *away* from the mirror plane, then by continuity the fields should be *mostly* **TE-like** and **TM-like**, as long as the waveguide thickness is smaller than the wavelength.

The additional mirror-symmetry plane, $y = 0$, provides an extra plane along which the mode is purely polarized, except that now the symmetries are reversed. As can be seen from figure 7, a TE-like mode will be *odd* with respect to y, while a TM-like mode will be *even*. The states that are most TE-like will be those which are even with respect to z and odd with respect to y, such as the $\mathrm{E}_{(\mathrm{o},n)}$ states above, and vice versa for TM-like [$\mathrm{M}_{(\mathrm{e},n)}$].

For example, consider the $\mathrm{E}_{(\mathrm{o},1)}$ and $\mathrm{E}_{(\mathrm{o},2)}$ states from figure 5. Since they are both TE-like, it is reasonable to plot only the H_z field component (which is the *only* **H** field component in the $z = 0$ plane), as we did in figure 6. This plot is instructive because it exhibits a subtle fact about symmetry and polarization: the **H** and **E** fields seem to have *opposite* symmetry. The H_z plot appears to be *even* with respect to $y=0$, and *odd* with respect to $z=0$. Actually, there is no difference in the

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

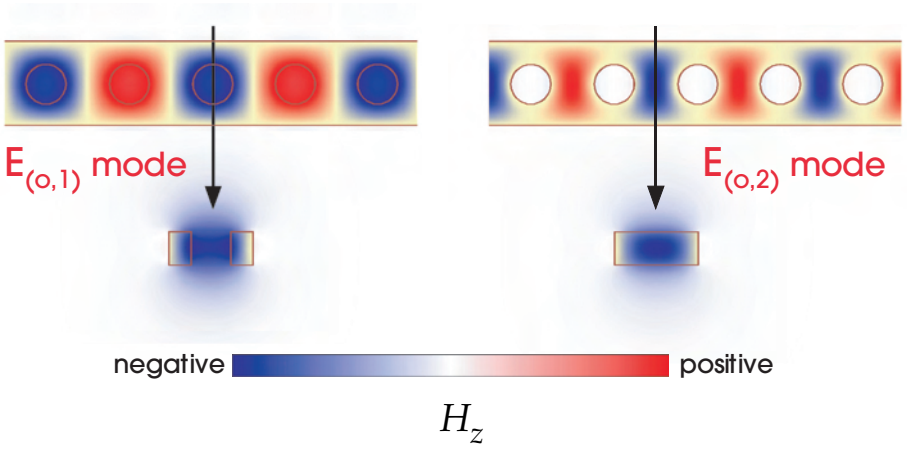


Figure 6: Cross sections of H_z field for lowest-order TE-like modes from figure 5 at the Brillouin-zone edge, $k = \pi/a$. *Left:* first band (lower edge of gap). *Right:* second band (upper edge of gap). The upper plot shows the $z = 0$ horizontal cross sections, in which H_z is the only magnetic-field component. The lower plots show x cross sections at the positions indicated by the arrows (through the H_z peaks). The dielectric material is shown as translucent yellow shading.

symmetry. The subtlety here is that **H** is a **pseudovector**, which is multiplied by an additional factor of -1 under mirror reflections.[6] We usually avoid this subtlety by referring only to the symmetry of **E**.

It is often desirable for a waveguide to have two different mirror symmetries, because this allows a TE/TM-like band gap to remain even if one of the symmetries is broken. For example, if we place the waveguide on a low-index substrate material instead of suspending it in air, then the $z = 0$ mirror symmetry is broken. However, the y symmetry is preserved, and as a result the large gap between the y-odd modes of figure 5 will remain.

It is also important that the substrate have a much smaller dielectric constant than the waveguide, so that the guided modes and the band gap remain well below the light cone. A common pairing at infrared wavelengths is silicon (Si) on silica (SiO_2), which has a dielectric contrast of roughly 12:2.[7] The substrate acts to lower the frequencies of the light cone (and the guided modes, since they penetrate slightly into the substrate). The substrate has additional effects on point defects, as described in the next section.

In short, for thin structures with mirror symmetry, such as those shown in figure 1, the modes can be classified as "TE-like" or "TM-like." In two

[6] The -1 factor arises for any **improper rotation** (a 3×3 rotation matrix with a determinant of -1). See footnote 2 on page 26, and also Jackson (1998).

[7] At a wavelength $\lambda = 1.5\,\mu$m, the dielectric constant of silicon is about 12.1 (Palik, 1998) and of silica is about 2.1 (Malitson, 1965).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

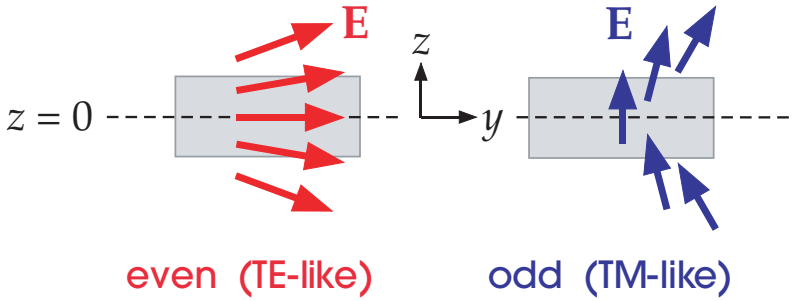


Figure 7: Schematic depiction of electric field lines (**E**) for a thin dielectric structure (gray shading) with a $z = 0$ mirror symmetry plane. Modes that are even with respect to this mirror plane (left) are **TE-like**: **E** is mostly parallel to the mirror plane (and is exactly parallel at $z = 0$). Modes that are odd (right) are **TM-like**: **E** is mostly perpendicular to the mirror plane (exactly perpendicular at $z = 0$).

dimensions, the favorable geometry for TM band gaps was dielectric spots in air, and for TE band gaps it was air holes in dielectric. Likewise, the structures from figure 1(a, c) typically favor TE-like gaps, while the one from figure 1(b) favors TM-like gaps.

Point Defects in Periodic Dielectric Waveguides

One respect in which periodic dielectric waveguides differ from the systems considered in previous chapters is that defect states are not perfectly localized. Consider once again the periodic waveguide of figure 1(a). While there are many ways to create a point defect (by, say, enlarging or reducing the radius of a hole), we will focus here on one particular choice: we will change the spacing between one pair of holes. If we increase the spacing from a to $1.4a$, the extra dielectric material pulls down a mode from the upper $[\mathrm{E}_{(o,2)}]$ band into the gap. The resulting defect state is TE-like, as shown in the left panel of figure 8. The frequency of this state is $\omega_0 \approx 0.308\,(2\pi c/a)$.

The crucial difference between this defect state and the defect states considered in previous chapters is that the band gap is incomplete: there are light-cone modes for any choice of frequency. Consequently, the defect state is a **leaky mode**, or **resonance**, and the point defect forms a **resonant cavity**. Because the translational symmetry is broken, k is no longer conserved and the defect mode couples to light-cone modes with the same ω. The result is *intrinsic radiative loss*. The right panel of figure 8 shows the defect mode profile, using a color scheme that highlights the small-amplitude, outward-radiating wave.

The intrinsic radiative loss of point defects is the main disadvantage of incomplete-gap systems, relative to the complete photonic band gaps of chapter 6.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

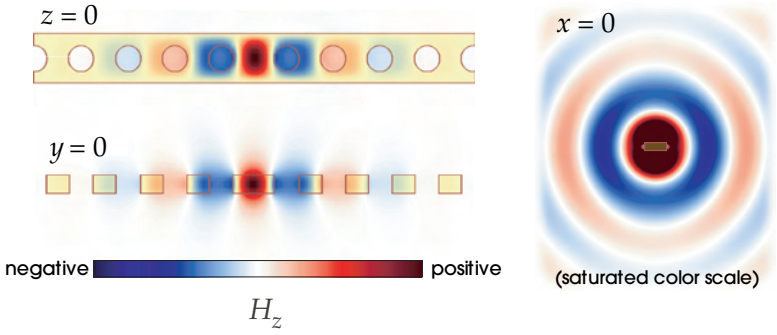


Figure 8: H_z field patterns of a localized resonant mode in a cavity formed by a defect in the periodic waveguide (suspended in air) of figure 1(a): the spacing between one pair of holes is increased from a to $1.4a$. The strong localization, exponentially decaying in the waveguide, is seen in the cross sections at left; the dielectric structure is shaded translucent yellow. The field decays only inversely with distance in the lateral directions, though, due to slow radiative leakage shown by the figure at right, which uses a saturated color scale to exaggerate small field values.

It presents several challenges. First, we must quantify the loss, as described in the next section. Second, we must determine how large a loss we can tolerate for a given application; this is discussed in more detail in chapter 10. Finally, we must reduce the loss, if necessary. We return to this subject at the end of chapter 8, where we outline two possible loss-reduction mechanisms for such systems.

Quality Factors of Lossy Cavities

The mode that exists in a resonant cavity decays slowly. It acts like a mode with a complex frequency $\omega_c = \omega_0 - i\gamma/2$, where the imaginary part of the frequency corresponds to an exponential decay.[8]

If the field decays as $e^{-\gamma t/2}$, then the energy within the cavity decays as $e^{-\gamma t}$. We could characterize the loss rate by γ, but because of the scale-invariance of the Maxwell equations, we prefer the dimensionless quantity $Q \triangleq \omega_0/\gamma$. This quantity, known as the **quality factor**, arises in any discussion of resonance, and can be interpreted in many ways. First, $1/Q$ is a dimensionless decay rate, and is

[8] Strictly speaking, this cannot be an eigenmode of the Maxwell equations. We showed in chapter 2 that for a real and lossless dielectric, all of the eigenvalues ω are real. In reality, the cavity mode is a superposition of extended modes, each with a real value of ω, that can be accurately approximated in the near field by an exponentially decaying function for slow decay, ($\gamma/2 \ll \omega_0$). The mathematical theory of such *leaky modes* is rather subtle; see, e.g., Snyder and Love (1983).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

equivalent to

$$\frac{1}{Q} = \frac{P}{\omega_0 U} \tag{1}$$

where P is the outgoing power and U is the electromagnetic energy localized in the cavity. Second, Q is a dimensionless *lifetime*, the number of optical periods that elapse before the energy decays by $e^{-2\pi}$. Third, $1/Q$ is the *fractional bandwidth* of the resonance. The Fourier transform of the time-varying field in the cavity has a squared amplitude that is a **Lorentzian peak**, proportional to $1/[(\omega - \omega_0)^2 + (\omega_0/2Q)^2]$, and $1/Q$ is the peak's fractional width at half-maximum. Finally, Q plays a critical role in temporal coupled-mode theory, which we will discuss in chapter 10.

If a resonance has more than one decay mechanism, then it is useful to characterize each mechanism with its own Q. Consider, for example, the point-defect cavity in the "air bridge" waveguide of the previous section. The net Q of the resonant mode, as a function of the number N of holes on either side of the cavity, is plotted in the top (blue) curve of figure 9. In a photonic crystal with a complete gap, Q would increase exponentially with N. In this case, Q does rise with N, but it *saturates* as N becomes large. The reason is that there are *two* decay mechanisms for the mode in this cavity: it can decay into the uniform dielectric strip that lies on either side of the holes, or it can radiate into the surrounding air. The net dimensionless decay rate $1/Q$ can be written as the sum of two decay rates: $1/Q = 1/Q_w + 1/Q_r$, where $1/Q_w$ and $1/Q_r$ are the waveguide and radiative decay rates, respectively. (The individual decay rates are defined by subdividing equation (1) into two loss powers, $P = P_w + P_r$.) To a first approximation, when the Q-values are large, they can be treated as independent: Q_w increases exponentially with N (because of the band gap), while Q_r is roughly independent of N. This explains the saturation observed in figure 9. As $N \to \infty$, we find that $Q \to Q_r$ (with $Q_r \approx 1200$ in this example).

In many device applications, the radiated power would be considered loss, and to minimize this loss we typically want $Q_r \gg Q_w$. We will quantify this notion in chapter 10.

Figure 9 also illustrates the effect of a substrate. The black and red lines show Q versus N for two different possible substrates. In both cases, the waveguide has $\varepsilon = 12$, and the underlying substrate has $\varepsilon = 2.25$. The first case (red) is a **monorail** substrate, which has the same cross-section as the waveguide, including the holes.[9] The second case (black) is a solid substrate with no holes.

In both cases, the band gap is preserved because the $y = 0$ mirror plane remains, as foreseen in the section Symmetry and Polarization. However, the substrate tends to decrease the radiative lifetime Q_r. The simplest way to see this is to consider the Fourier decomposition of the localized mode. As we first learned in the section Evanescent Modes in Photonic Band Gaps of chapter 4, the localized field in the

[9] This type of monorail design was fabricated by Foresi et al. (1997).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

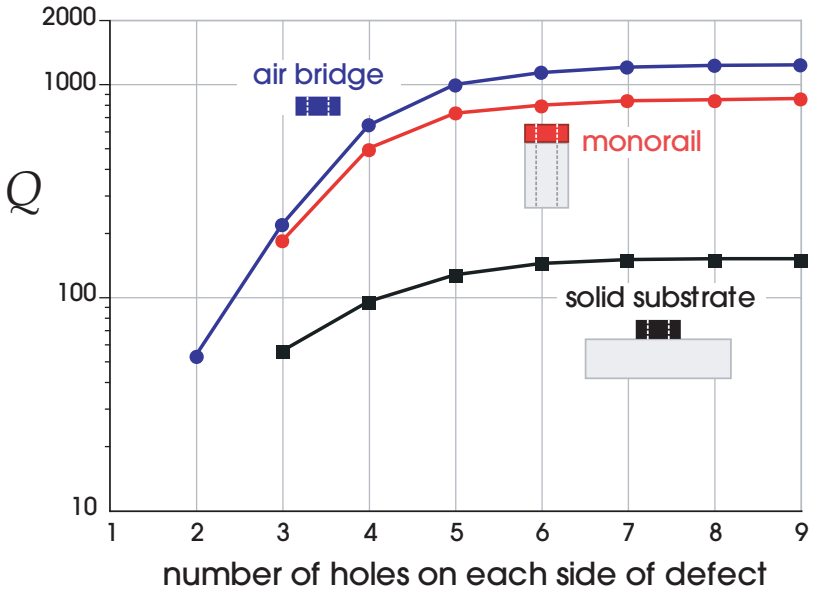


Figure 9: Total Q of the point defect state from figure 8, as a function of the number N of holes on either side of the defect. (After the holes, the structure is simply a uniform waveguide.) Three structure variations are shown (see insets): the *air bridge* cavity suspended in air (top), a *monorail* structure with a $\varepsilon = 2.25$ substrate of the same cross section (middle), and a *solid substrate* of $\varepsilon = 2.25$ (bottom). The Q saturates for large N to the intrinsic radiation loss, or Q_r.

crystal takes the approximate form of a decaying exponential multiplied by an $e^{i\pi x/a}$ oscillation. The Fourier transform of such a pattern yields a Lorentzian-like peak centered at the Brillouin-zone edge, $k = \pi/a$. The radiation losses come from the *tails* of this Fourier peak that lie within the light cone. As the substrate index is increased, the light cone is pulled down, causing a greater portion of these tails to fall within it and radiate.[10]

The monorail substrate has a weaker effect than the solid substrate, because the monorail substrate is mainly air and thus has less effect on the light cone.[11] The monorail substrate reduces Q_r by about 30%, while the solid substrate reduces Q_r by almost a factor of ten.

[10] As an exception, if the material both above and below the waveguide has low contrast, there is a limit where the radiative losses decrease (Benisty et al., 2000); see also the subsection Delocalization of chapter 8.

[11] For the monorail or any other inhomogeneous substrate, the light line is no longer straight. It is instead the lowest-frequency mode of the two-dimensional structure with the same cross section. We return to an analogous situation in chapter 9.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Further Reading

The concepts of the light cone, guided modes, and band gaps (or "stop bands") of periodic dielectric waveguides have been known for several decades, as reviewed by Elachi (1976). The oldest work relied on perturbative methods, restricted to weak periodicity, but early examples of rigorous nonperturbative calculations for variable-reactance surfaces (Oliner and Hessel, 1959) and dielectric waveguides (e.g., Peng et al., 1975) can also be found. A more recent review can be found in Villeneuve et al. (1998), and two early experimental demonstrations of infrared-scale cavities formed by defects in periodic dielectric waveguides are Foresi et al. (1997) and Ripin et al. (1999).

An important practical context in which periodic dielectric waveguides have been employed is the **fiber Bragg grating**. This is a standard glass fiber, which guides light by index guiding, that has been modified to include a weak periodic variation of the refractive index along the fiber axis. It was first demonstrated by Hill et al. (1978) that such a variation could be created by a photorefractive effect: illuminating the fiber with an interference pattern from an intense light source changes the index of the material. The resulting index variations are tiny, at most 1% (Lemaire et al., 1993) and usually much less, so the gaps are very small (and can be analyzed by perturbative methods). This is compensated by the long length of the fiber, however, and fiber gratings have been applied to problems from filtering to sensors to dispersion compensation (Ramaswami and Sivarajan, 1998).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

8

Photonic-Crystal Slabs

In the previous chapter, we saw how simple structures with only one-dimensional periodicity can be used to confine light in three dimensions by a combination of band gaps and index guiding. Now, we will carry that idea one step further, by investigating structures with *two*-dimensional periodicity but a finite thickness. Such hybrid structures are known as **photonic-crystal slabs** or **planar photonic crystals**. They are *not* "two-dimensional" photonic crystals, despite the resemblance: the finite thickness in the vertical (z) direction introduces qualitatively new behavior, just as the periodic dielectric waveguides of the previous chapter differed from photonic crystals in one dimension. As in the three-dimensionally periodic crystals of chapter 6, defects in photonic-crystal slabs can be used to form waveguides and cavities. With such building blocks, many interesting devices have been experimentally realized using standard lithographic techniques based on two-dimensional patterns. This ease of fabrication comes at a price, however: careful design is required to minimize losses at cavities and similar breaks in the periodicity.

Rod and Hole Slabs

Two examples of photonic-crystal slabs are shown in figure 1. Just as in chapter 5, we will study two basic topologies: a square lattice of dielectric rods in air [figure 1(a)]; and a triangular lattice of air holes in dielectric [figure 1(b)]. We will refer to these structures as the **rod slab** and the **hole slab**, respectively. In our rod-slab example, the rods have a radius $r = 0.2a$ and the slab has a thickness $2a$, whereas in the hole-slab example, the holes have a radius $r = 0.3a$ and the slab has a thickness $0.6a$. (We will discuss the optimization of these

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

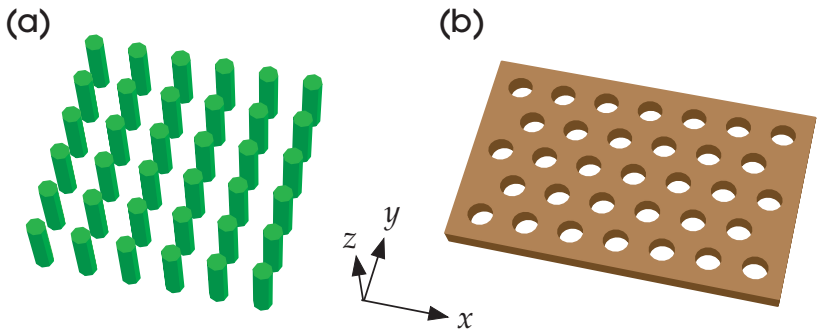


Figure 1: Examples of **photonic-crystal slabs**, which combine two-dimensional periodicity (in the xy directions) and index-guiding in the vertical (z) direction. (a) The **rod slab**, a square lattice of dielectric rods in air. (b) The **hole slab**, a triangular lattice of air holes in a dielectric slab.

parameters later.) We initially consider **suspended membrane** structures that are surrounded entirely by air,[1] and later discuss the effect of a substrate.

With discrete translational symmetry in two directions, the *in-plane* wave vector $\mathbf{k}_\parallel = (k_x, k_y)$ is conserved, but the *vertical* wave vector k_z is not conserved. As before, it is useful to plot the projected band structure, which in this case is a plot of ω versus $\mathbf{k}_\parallel$ in the irreducible Brillouin zone of the two-dimensional lattice. The projected band diagrams for the rod slab and for the hole slab are shown in figure 2. Many features of these diagrams should be familiar from chapter 7. The extended modes propagating in air form a light cone for $\omega \geq c|\mathbf{k}_\parallel|$. Below the light cone, the higher dielectric constant of the slab has pulled down discrete guided bands. Eigenstates in these bands decay exponentially in the vertical direction (away from the slab).

As in the section Symmetry and Polarization of chapter 7, the system is invariant under reflections through the $z = 0$ plane, which allows us to classify the modes as *TE-like* (even) and *TM-like* (odd). (In this case, however, there are generally no y or x mirror symmetries; these are broken by the $\mathbf{k}_\parallel$ vector.) In figure 2, the TE-like modes are plotted with red lines, and the TM-like modes are plotted with blue lines. The rod slab favors a TM-like gap, and the hole slab favors a TE-like gap, as one might expect from our knowledge of the corresponding two-dimensional photonic crystals. However, as in the previous chapter, the gaps in photonic-crystal slabs are **incomplete**: they refer only to the guided modes, and not the light cone.

[1] For the hole slab, such suspended structures have been fabricated by many groups; see, e.g., Kanskar et al. (1997) and Scherer et al. (1998) for early examples.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

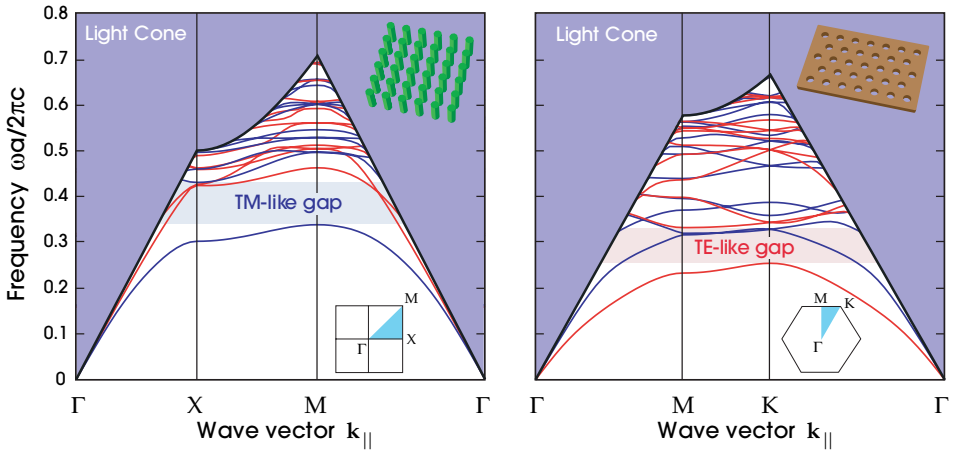


Figure 2: Band diagrams for photonic crystal slabs of figure 1, suspended in air (inset): the **rod slab** (left) and the **hole slab** (right). The blue shaded area is the *light cone*, all of the extended modes propagating in air. Below it are the guided bands localized to the slab: blue/red bands indicate TM/TE-like modes, respectively (odd/even with respect to the $z=0$ mirror plane). The rod/hole slabs have gaps in the TM/TE-like modes, which are shaded light blue/red respectively.

Polarization and Slab Thickness

Why did we choose the rod slab to be nearly four times thicker than the hole slab? This question invites a discussion of the optimal slab thickness, which turns out to be strongly related to polarization. Figure 3 shows the gap sizes of the rod and hole slabs as a function of slab thickness, demonstrating that there is indeed an optimal thickness. Indeed, if the slab is too thick the gap can disappear entirely.

This can be understood intuitively by considering the extreme cases of a very thin or very thick slab. If the slab is too thin, then the bands are weakly guided, if they are guided at all. States that lie just below the light line decay very slowly into the air regions. The frequency difference between the lowest band and the light line becomes too small for much of a gap, and the modes become so delocalized that the periodicity is unimportant. On the other hand, for a very thick slab, the gap in the fundamental (node-free) guided mode approaches that of the infinite two-dimensional system. However, higher-order modes (with more vertical nodes) are pulled down and populate the gap. In the framework of the variational theorem (chapter 2), the cost of extra vertical nodes becomes very small.

Using these considerations, we expect the ideal thickness to be approximately half a wavelength in thickness. This is thick enough for the fundamental mode to be well confined, yet thin enough to prevent higher-order modes from fitting

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

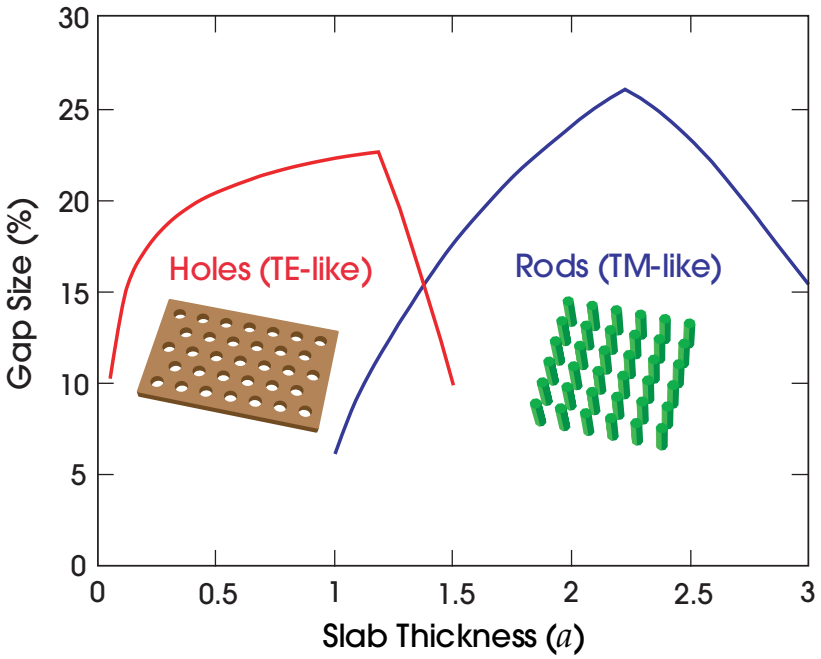


Figure 3: The size of the gap in the guided modes (gap/midgap ratio) of the rod and hole slabs from figure 2 (inset), as a function of the slab thickness in units of the periodicity (a). The hole slab has a gap in the TE-like modes (red, left), and the rod slab has a gap in the TM-like modes (blue, right). As the slab thickness increases, the gap initially increases, approaching the corresponding two-dimensional TE/TM gap; at some point, however, a higher-order mode is pulled into the gap, which abruptly begins decreasing.

within the slab. But which wavelength are we referring to: the wavelength in air, or in the dielectric? The answer lies somewhere in between. The "effective" dielectric constant of the material is mode-dependent, depending on spatial average of the ε profile weighted in some sense by the field profile. The exact prescription for the weighting is the domain of *effective-medium theories*, which we will not discuss in detail, except to quote the result that the weighting depends strongly on polarization.[2] The effective vertical wavelength of the *TE*-like modes is determined mostly by the *high* dielectric material, while for the *TM*-like modes it is the *low* dielectric regions that matter most.

This explains why, in figure 3, the gap in the hole slab opens up at a much smaller thickness than for the rod slab—the former's gap is in the TE-like modes, while the latter's is in the TM-like modes. As these gaps open, they

[2] For a more detailed analysis of the effective-medium description for photonic-crystal slabs, see Johnson et al. (1999).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

initially asymptote towards the gaps of the corresponding two-dimensional structures: 28% for the holes and 39% for the rods. However, at a certain critical thickness ($\sim 1.2a$ for the holes and $\sim 2.25a$ for the rods), a higher-order mode is pulled into the gap and begins to sharply decrease it.

You might wonder why, in figure 2, we didn't choose the slab thicknesses ($0.6a$ and $2a$ for holes and rods, respectively) that created the maximum band gap. The main reason for this, as we shall see in the subsequent sections, is that in slab structures the limiting factor is not usually the *size* of the band gap, so it is not necessary to pick the size that is strictly optimal. Rather, our designs are constrained by things like the light cone, higher-order modes in defects, substrates, and so on. Therefore, we chose a smaller thickness that is easier to fabricate and still has a large gap, corresponding to common experimental parameters.[3] Nevertheless, a large difference in slab thickness is required between the rods and holes, which has helped to make the latter more experimentally attractive.

Linear Defects in Slabs

As in chapters 5 and 6, by introducing a linear defect into the periodic structure, we can create a waveguide mode that propagates along the defect. Now, however, the localization of the waveguide mode relies on both the band gap within the plane of periodicity and also on index guiding in the vertical direction, and this will restrict the kinds of modes that we can guide.

Reduced-radius rods

In our discussion of two- and three-dimensional crystals, we formed a waveguide by removing a row of rods (as in the section Linear Defects and Waveguides of chapter 5). In this case, we will remove the row *gradually*, to show exactly how the defect mode forms. Specifically, we consider shrinking the radius of all of the rods in a particular row—a fabricated example of a similar waveguide is shown in figure 4. In figure 5, we plot the projected band diagram for different choices of the shrunken radius, restricting our attention to TM-like modes. The extended modes of the crystal are in the dark blue region, and the band gap appears in the light blue region. Inside the band gap, there is a single guided band for each choice of radius, corresponding to the TM-like waveguide mode shown in figure 6. This mode is localized to the line defect. It cannot couple to extended modes in the crystal, because it is within the band gap. Nor can it couple to extended modes in air, because it is below the light line. Although the band gap is incomplete, the waveguide mode still exists, because k_x is conserved. Just like the modes of the discontiguous structure in

[3] It is for a similar reason that we chose a relatively small hole radius $r = 0.3a$, even though larger radii yield larger gaps.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

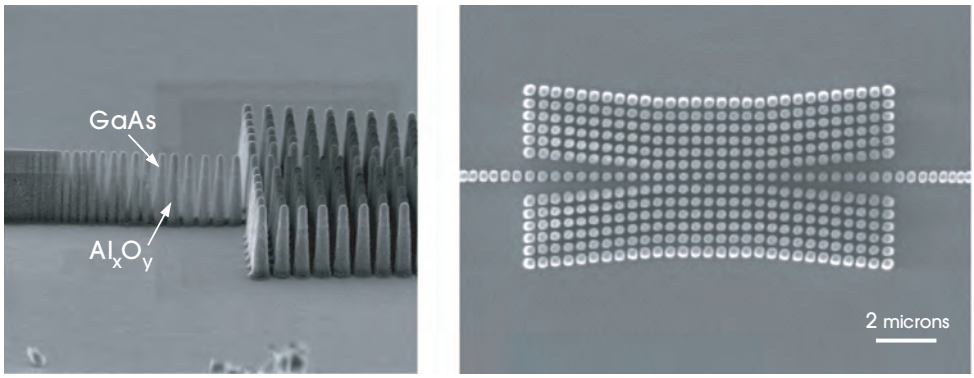


Figure 4: Two views of a reduced-radius waveguide fabricated in a rod slab by Assefa et al. (2004), designed to operate at near-infrared wavelengths. (GaAs rods on low-index aluminum-oxide pedestals.) The structure was designed to couple the reduced-radius waveguide via an adiabatic taper to a dielectric strip waveguide at the ends. (Images courtesy L. Kolodziejski.)

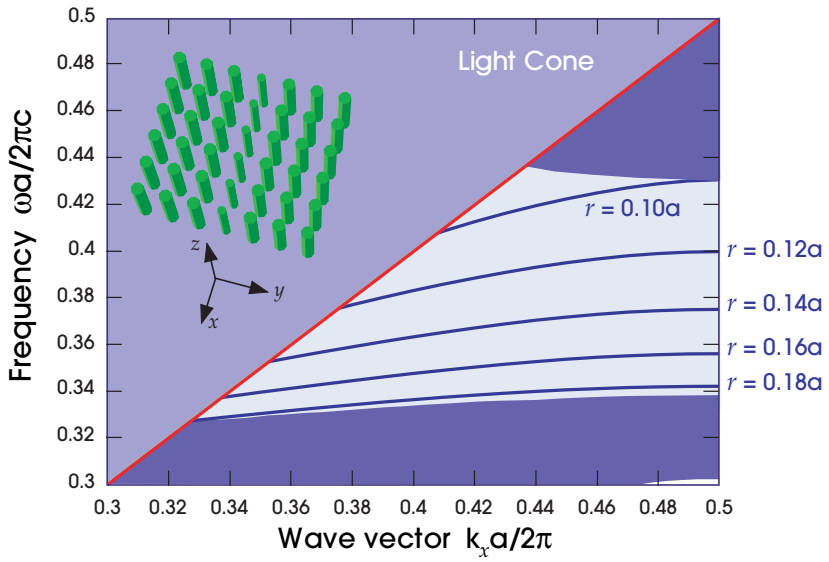


Figure 5: Projected band diagram of TM-like (z-odd) states in a linear waveguide in the rod slab of figure 2(left), formed by reducing the radius of a row of rods (inset), as a function of the wave vector k_x along the defect. The dark blue shaded regions below the light line are extended modes in the crystal. The rods in the bulk crystal have radius $r = 0.2a$; the localized waveguide bands for five different defect radii are labelled in the gap.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

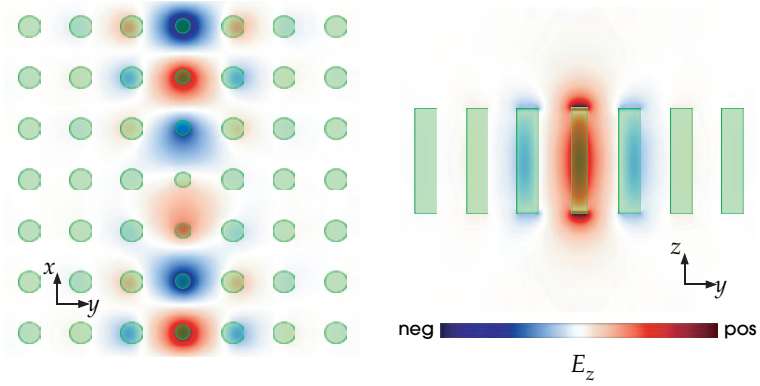


Figure 6: E_z field cross sections in reduced-radius line-defect waveguide from figure 5, for a defect rod radius of $r = 0.14a$ at a wave vector $k_x a/2\pi = 0.42$. The dielectric material is shaded translucent green. *Left:* horizontal ($z = 0$) cross section (in which E_z is the only nonzero electric-field component). *Right:* vertical ($x = 0$) cross section bisecting the rods perpendicular to the waveguide. The field decays exponentially away from the waveguide.

the section A Two-Dimensional Model of chapter 7, this waveguide mode will propagate indefinitely in a perfect, periodic system.

A primary difference between the case of a slab and the case of a truly two-dimensional photonic crystal is that the band diagram of the slab has a light cone. The light cone places an interesting constraint on waveguide modes: we cannot completely remove a row of rods and still support a waveguide mode.[4] As shown in figure 5, by the time we have reduced the rod radius to $0.10a$, the mode hugs the light cone at the very top of the gap. If we reduce the radius further, the mode will cease to be guided. We cannot guide light primarily in the air between the rods, because such a mode would not be confined vertically by index guiding.

To draw out the comparison further, we consider three different structures that share the same periodicity in two dimensions, but differ in the third (vertical) dimension. All three structures have exactly the same $z = 0$ cross section: a triangular lattice of rods in which one row of rods has been removed. The first structure is a two-dimensional photonic crystal with a complete TM band gap. The rods extend to infinity in the vertical direction. The second structure is a three-dimensional photonic crystal with a complete band gap, as in the section Localization at a Linear Defect of chapter 6. The third structure is a photonic-crystal slab. In figure 7, we show the projected band diagrams for these three structures. Here, we can see that the air-guided mode of the first two crystals lies almost entirely above the light line of the slab.

[4] If we terminate the crystal differently, for example by cutting the columns in half adjacent to the missing row, then it is possible to localize surface states, as in the section Surface States of chapter 5. Such modes are still not localized primarily in the air, however.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

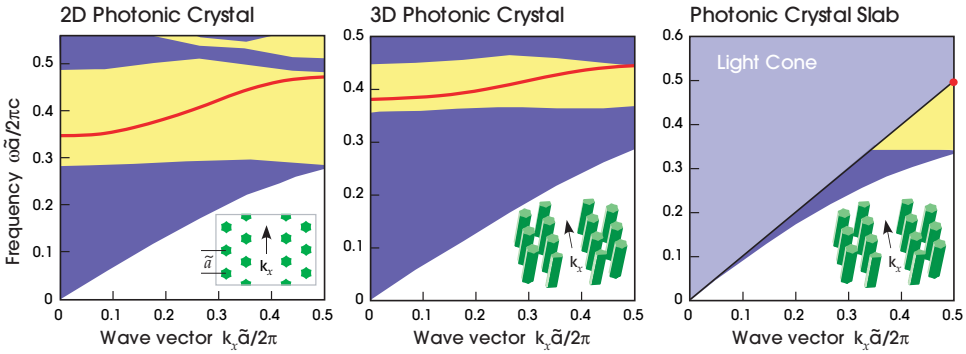


Figure 7: Projected band diagrams for three waveguides, all of whose cross sections are an identical triangular lattice (period $\tilde{a}$) of dielectric ($\varepsilon = 12$) rods in air with a missing row of rods. The guided band is shown as a red line, and extended modes of the crystal are shown in shaded dark blue, with the band gap(s) shaded yellow. *Left:* TM mode of two-dimensional crystal (uniform in z, $k_z = 0$) with a TM gap. *Middle:* TM-like mode of the three-dimensional crystal from chapter 6 (see the section Localization at a Linear Defect), which has a *complete* band gap. *Right:* a slab of rods (thickness $2a$) suspended in air, which has a gap in the TM-like modes; because of its light cone, it has at most a very weakly guided state right at the gap edge (red dot).

Removed holes

In the previous section, we created a defect by reducing the average dielectric constant along a particular line. This pushed a guided band up from the lower edge of the gap. We can also increase the average dielectric constant, thereby pulling bands down from the upper edge of the gap. This is illustrated in figure 8, which is the projected band diagram for a hole slab in which we have filled in a row of holes.[5] A fabricated example of such a waveguide is in a suspended membrane is shown in figure 9.

As in the previous section, this waveguide has a series of guided modes that are confined horizontally by the band gap and vertically by index guiding. However, in this case, there is a second category of guided modes, which arises because the waveguide has a higher average dielectric constant than the surrounding air. These new guided bands, which lie below the extended modes of the crystal, are *index-guided* in all directions. Since all of the guided modes are TE-like and are fundamental (no nodes) in the z direction, we can visualize them by plotting H_z in the $z = 0$ plane. Figure 10 shows the field patterns for the five guided modes that are identified with letters in figure 8.

[5] This commonly used design is called a "W1" defect. More generally, a "Wn" defect involves the removal of n rows of holes, after Olivier et al. (2001). An early fabrication of such a waveguide in a suspended membrane is described by Lončar et al. (2000).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

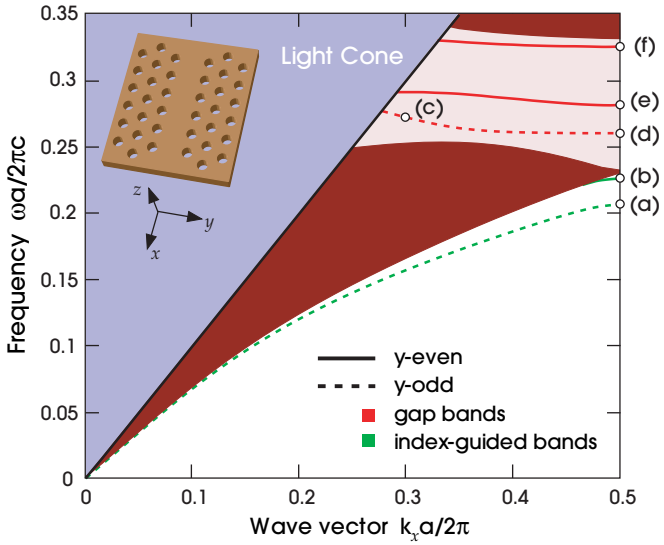


Figure 8: Projected band diagram of TE-like (z-even) states for a "W1" defect in the hole slab of figure 2(right), formed by a missing row of nearest-neighbor holes along the x direction. Dark-red shaded regions indicate extended TE-like modes of the crystal. Guided modes are introduced both in the gap (red bands in pink shaded region) and below all of the extended modes of the crystal (green bands below red shaded region). The latter are index-guided. The guided modes are classified as y-even (solid lines) or y-odd (dashed lines) according to the $y = 0$ mirror symmetry plane bisecting the waveguide. The fields corresponding to the labelled points (a–f) are shown in figure 10.

Because the system is invariant under reflections in the $y = 0$ plane, we can classify all of these modes as either odd or even under reflection in this plane. Recall that $\mathbf{H}$ is a pseudovector, which means that an H_z field pattern that looks y-even is actually y-odd (see the section Symmetry and Polarization of chapter 7). The fundamental mode is y-odd. Most of our attention will focus on the y-odd modes, because they are the ones most easily excited by a planewave input beam.

The field profiles of the second y-odd band (which lies in the band gap) show a surprising evolution as the wave vector varies from the light cone to the Brillouin-zone edge. Figure 10(b) shows the pattern for $k_x = 0.3\pi/a$, and figure 10(c) shows the pattern for $k_x = 0.5\pi/a$. The latter mode has an additional pair of nodes in the y direction. This drastic change is the result of an **anti-crossing**: two bands that are expected to *intersect* will instead couple to one another (unless some symmetry prevents it) and the bands repel.[6]

[6] The anti-crossing origin of the waveguide mode in this structure was first noted by Notomi et al. (2001), and in two dimensions by Olivier et al. (2001). A related system in which a different symmetry allows the bands to intersect was studied by Kuchinsky et al. (2000).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

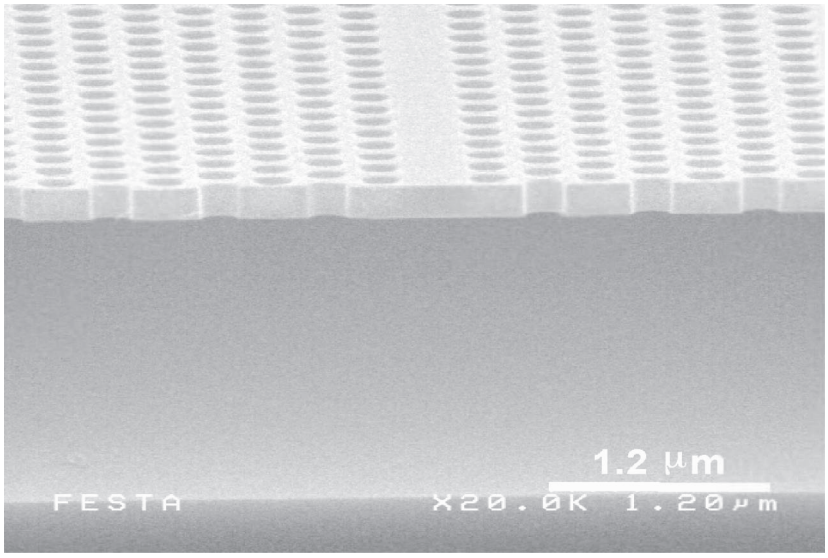


Figure 9: SEM image of a waveguide formed by a missing row of holes in a suspended-membrane hole slab (Sugimoto et al., 2004). (Image courtesy K. Asakawa.)

How can we understand this phenomenon? It is depicted schematically in figure 11. First, we imagine the y-odd bands for the case of a nonperiodic waveguide (shown at left). Our index-guided mode (green) is folded back at the "artificially" imposed edge of the Brillouin zone. The higher-order y mode (red) lies at a higher frequency and is also folded back. Next, we turn on the periodicity in the slab (shown at right). The band gaps appear, and the bands repel one another. This happens not only at the edge of the Brillouin zone, but also where the red and green bands intersect. Thus, as k_x increases along the second band, the field pattern transforms continuously from the red mode [figure 10(b)] to the green mode [figure 10(c)]. This phenomenon can be exploited to produce unusual dispersion effects, such as ultraflat quartic band edges and zero-dispersion inflection points.[7]

Substrates, dispersion, and loss

In the preceding sections, we have concentrated on photonic-crystal slabs floating in air. These are special structures not only because the index contrast between the slab and the surrounding medium is maximized, but also because they have $z = 0$ reflection symmetry. Placing the slab on a substrate will break the reflection symmetry, causing the TE-like and TM-like modes to couple, and destroys any

[7] See, for example, Petrov and Eich (2004).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

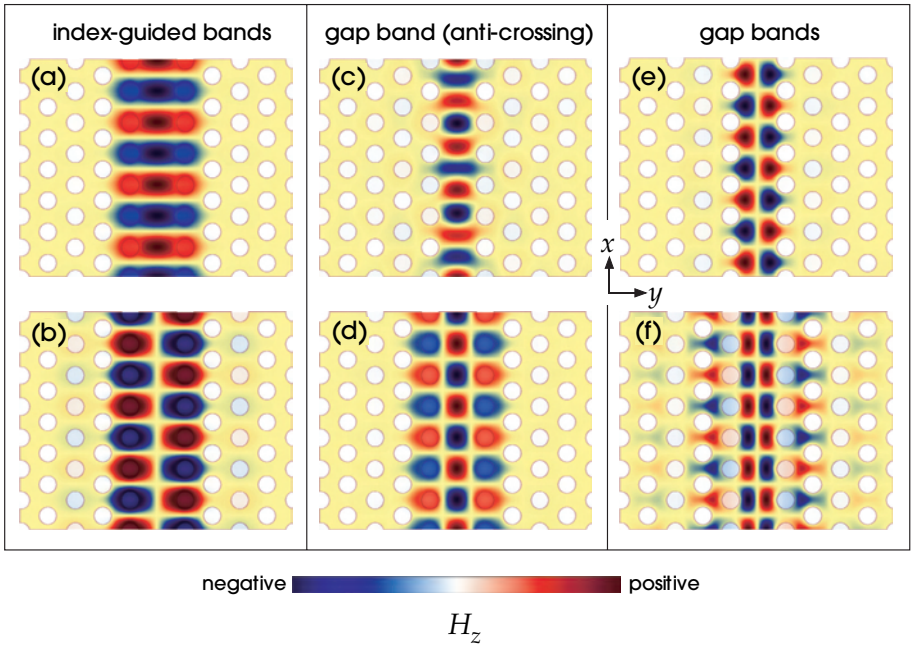


Figure 10: H_z field cross section for the missing-hole waveguide, with (a–f) corresponding to the labelled points of figure 8. Dielectric material is shown as translucent yellow. *Left:* y-odd (top) and y-even (bottom) index-guided modes at $k_x = \pi/a$, which have lower frequencies than any extended mode of the crystal or air at that k_x. (Because $\mathbf{H}$ is a pseudovector, the even modes look odd and vice versa.) *Middle:* two points in the same y-odd gap-guided band, at $k_x = 0.32\pi/a$ (top) and $k_x = \pi/a$ (bottom); the drastic change in field pattern corresponds to an anti-crossing as described in the text. *Right:* two higher-order y-even gap-guided bands.

TM/TE-only band gap (unlike the periodic waveguides of chapter 7, where there was additional symmetry). Thus, when a photonic-crystal slab with a linear waveguide is placed on a substrate, any waveguide modes that were formerly confined by the gap become *leaky*. A formerly guided TE-like mode will couple to the extended TM-like modes, and vice versa, causing the mode to radiate away within the plane of periodicity.[8]

Of course, this is a problem only if the leakage rate is unacceptable for a particular application, and numerous experiments have shown that effective waveguides can be made with oxide substrates (with $\varepsilon \approx 2$).[9] Moreover, we can

[8] Some subtleties due to finite-size effects are discussed in Vlasov et al. (2004).

[9] Early fabrications of waveguides in hole slabs with oxide substrates include Baba et al. (1999), Phillips et al. (1999), Lin et al. (2000), Lončar et al. (2000), Tokushima et al. (2000), and Chow et al. (2001). For the rod slab see, e.g., Assefa et al. (2004) and Tokushima et al. (2004).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

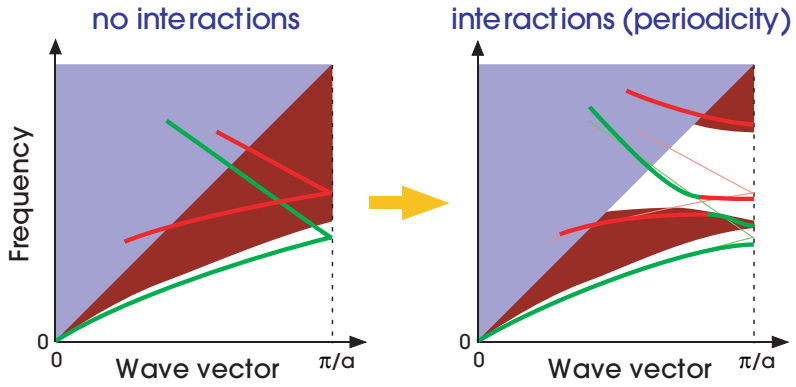


Figure 11: Schematic of anti-crossing(s) that occur when periodicity is added (right) to a dielectric waveguide in some "averaged" slab without periodicity (left). The fundamental index-guided mode (lower line, green) is folded back at the Brillouin zone edge, and therefore intersects with a higher-order mode (upper line, red), which is leaky (above the light line) at this wavelength. Periodicity (right) opens up a gap and also causes the bands to couple at their intersection point, causing them to split and form a hybrid band (red/green line in gap) that transitions from one field pattern to the other.

reduce the polarization mixing by etching the periodic pattern into the substrate as well as the slab, much like the monorail substrate in the section Quality Factors of Lossy Cavities of chapter 7. Or, one can restore the z symmetry to some degree, by depositing a "superstrate" on top of the slab that is similar to the substrate.

Even for a symmetric structure, the guided modes in the gap under the light cone lie relatively near the edge of the Brillouin zone. Because the bands flatten as they approach the edge, the group velocity v_g of these modes approaches zero (see the section Bloch-Wave Propagation Velocity of chapter 3). Likewise, the **group-velocity dispersion** $\sim dv_g/d\omega$, which tells us how fast pulses will spread,[10] diverges at the edge. Thus, as we see from figures 5 and 8, our waveguide modes may have a low group velocity and a strong dispersion over most of their bandwidth.

Sometimes, this is desirable. For example, as described by Soljačić et al. (2002*b*), a low group velocity (*slow light*) can be used to enhance the effect of optical nonlinearities. In other applications, it would be better to have a broad bandwidth of low dispersion and a more typical group velocity. This can be arranged by altering the waveguide design. For example, one can surround a dielectric strip waveguide with the photonic-crystal slab (taking care to avoid terminations with

[10] See, e.g., Ramaswami and Sivarajan (1998). More precisely, it is conventional to define a **dispersion parameter** $D = -(2\pi c/\lambda^2)\, d^2k/d\omega^2 = (2\pi c/v_g^2\lambda^2)\, dv_g/d\omega$, which gives the pulse spreading Δt per unit bandwidth $\Delta\lambda$ per unit distance travelled.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

surface states).[11] A waveguide mode in this structure is not too different from that of an isolated strip, but it remains useful as a channel to reach point defects and other devices embedded in the photonic crystal.[12]

Finally, although the waveguide modes in a perfect and symmetric slab are lossless, in reality there will always be some degree of loss. In addition to the substrate losses described above, losses can be caused by material absorption, or by radiative scattering from the inevitable irregularities that arise during fabrication (*fabrication disorder*). Disorder will violate the translational symmetry of the slab, and allow waveguide modes to couple to states with different values of k. Disorder can also cause reflections into the reverse-direction waveguide mode at $-k$. Although a detailed study of disorder is beyond the scope of this book, there is a simple scaling relation that applies for modes near the $v_g = 0$ band edge.[13] All other things equal, the loss (per unit distance) due to disorder scattering into the reflected mode grows as $1/v_g^2$, while the loss rates into *other* modes (e.g., radiation or absorption) grow as $1/v_g$.

Sometimes, losses are inevitable features of the design. If we want the waveguide to bend, or to undergo a transition into a different waveguide, we must break the translational symmetry of the slab. Transition losses can be minimized by making the transition as gradual as possible (*adiabatic*).[14] Bend losses can be minimized using techniques that we will describe in chapter 10. The most challenging losses to cope with are those that arise when we want the light to stay in one spot for a long time, either because v_g is small or because, as in the next section, we want a high-Q cavity.

Point Defects in Slabs

A point defect in a photonic-crystal slab traps a localized mode, which resembles the corresponding point-defect mode in an infinite two-dimensional crystal. However, just as we found in the section Point Defects in Periodic Dielectric Waveguides of chapter 7, the presence of the light cone means that localized modes in the slab are leaky resonances, with intrinsic vertical radiation losses.

For example, suppose that we want to create a "monopole" state in the rod slab. In a two- or three-dimensional photonic crystal, the easiest way to create a monopole state is to remove a single rod, but that will not work in this case. The vertical confinement would be too weak. Instead of completely removing a rod, however, we could reduce its radius or dielectric constant by a small amount. Figure 12 shows the field pattern of the defect mode that results from reducing the

[11] See Johnson et al. (2000) and Lau and Fan (2002).

[12] A strip waveguide surrounded by a photonic-crystal slab may also have reduced in-plane radiation losses when translational symmetry is broken, e.g. in the presence of fabrication disorder (Johnson et al., 2005).

[13] See Hughes et al. (2005) and Johnson et al. (2005).

[14] For a proof of the adiabatic theorem in periodic systems and the design conditions it imposes, see Johnson et al. (2002*b*).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

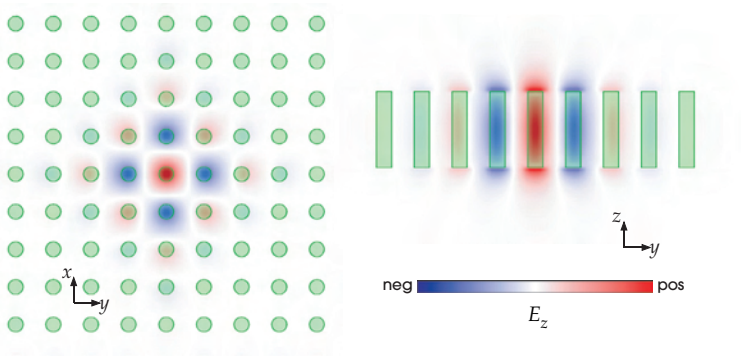


Figure 12: E_z cross sections for resonant "monopole" mode of a point defect in the rod slab (dielectric material shown as translucent green), formed by reducing the dielectric constant of the center rod and its four nearest neighbors from $\varepsilon = 12$ to $\varepsilon = 9$. The mode decays exponentially in the plane of the slab, but slowly radiates away vertically.

Table 1

	Symmetric	Asymmetric
Suspended membrane	13000	—
$\varepsilon = 2.25$ pillars	7200	8100
Solid $\varepsilon = 2.25$ substrate	380	370

Intrinsic radiative lifetimes Q_r for the point-defect structure of figure 12 resting on various substrates: air, $\varepsilon = 2.25$ pillars with the same cross section as the rod slab, and solid $\varepsilon = 2.25$, similar in spirit to the configurations in figure 9 of chapter 7. Both the symmetric case (substrate both above and below the slab) and the asymmetric case (substrate only below the slab) are included.

ε of a single rod *and* its four nearest neighbors from $\varepsilon = 12$ to $\varepsilon = 9$. (We chose to modify the neighboring rods for reasons to be given in the next section.) It is a monopole-pattern TM-like mode with a radiative lifetime $Q_r \approx 13000$.

As we saw in the section Quality Factors of Lossy Cavities of chapter 7, the lifetime of a resonant mode in a slab is strongly affected by the presence of a substrate. Table 1 compares the radiative lifetime Q_r of the monopole state, for various different choices of the substrate. Once again, substrate losses are reduced when the substrate has the same cross section as the slab. Another choice is whether or not to place a layer of the substrate material on top of the slab, thereby restoring the z symmetry. The advantage of this idea is that it prevents polarization mixing and the attendant in-plane radiative losses. The disadvantage is that, by increasing the mean dielectric constant of the structure, the local density of radiative states is enhanced. As can be seen in Table 1, these two effects very nearly counteract one another, and there is no compelling reason to symmetrize the system in this particular case.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Mechanisms for High Q with Incomplete Gaps

When designing a resonant cavity for applications, the central concern is usually minimizing radiation losses, or, equivalently, maximizing the radiative quality factor Q_r. In chapter 10 we will see that Q_r needs to exceed 10^5 for certain devices, such as narrow bandwidth filters. This is a challenge in a system with an incomplete band gap, and every situation presents tradeoffs and constraints. No clearly optimal solution is currently known, although several high-performance designs have been achieved and are cited below. Rather than a futile attempt to present the "best" design, we will focus on more fundamental questions: Why are high-Q_r cavities possible at all without a complete gap? Is there any upper bound to the attainable Q_r? What physical mechanisms can we exploit to increase Q_r, and what tradeoffs do they impose?

Delocalization

A pervasive principle in the design of electromagnetic resonators is that one can usually *trade off localization for loss*. For example, consider a conventional dielectric cavity, the **ring resonator**, which is a dielectric strip waveguide that has been curved around to form a circular loop. Because the waveguide is not straight, there will be radiation losses, but as the radius becomes larger (and the curvature smaller) the losses shrink. Marcatilli (1969) showed that the Q_r of a ring resonator actually increases *exponentially* with the ring radius. The price paid is that the resonator mode spreads out over an increasing volume, which is problematic for reasons given below. In addition, the system supports a larger number of modes, and the frequency separation between modes (the **free spectral range**) decreases.

The situation with point defects in photonic-crystal slabs is similar. The process analogous to increasing the radius of the ring resonator would be to make the point-defect mode larger and more multi-mode. An alternative is to make the defect *weaker* by reducing the dielectric contrast relative to the rest of the slab. For example, if we reduce the dielectric constant of a single rod by a small amount $\Delta\varepsilon$, then the resulting monopole mode will rise by only a small amount $\Delta\omega$ into the gap from the dielectric band. As we first learned in the section Evanescent Modes in Photonic Band Gaps of chapter 4, this means that the mode will be delocalized and will slowly decay into the slab. We might therefore guess that Q_r will grow as $\Delta\omega$ decreases, and this is precisely the result that we see in figure 13. To delocalize the mode even more for the same $\Delta\omega$, we also consider changing both the central rod and its four nearest neighbors by $\Delta\varepsilon$, as in figure 12. This five-rod defect has a Q_r almost two orders of magnitude greater than that of the single-rod defect for the same $\Delta\omega$.[15]

[15] More generally, we might consider a structure with smooth "adiabatic" transitions from the defect to the bulk crystal (e.g., Srinivasan et al., 2004; Istrate and Sargent, 2006), for which an appropriate design should regain the exponential growth of Q_r with size that we saw from the ring resonator.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

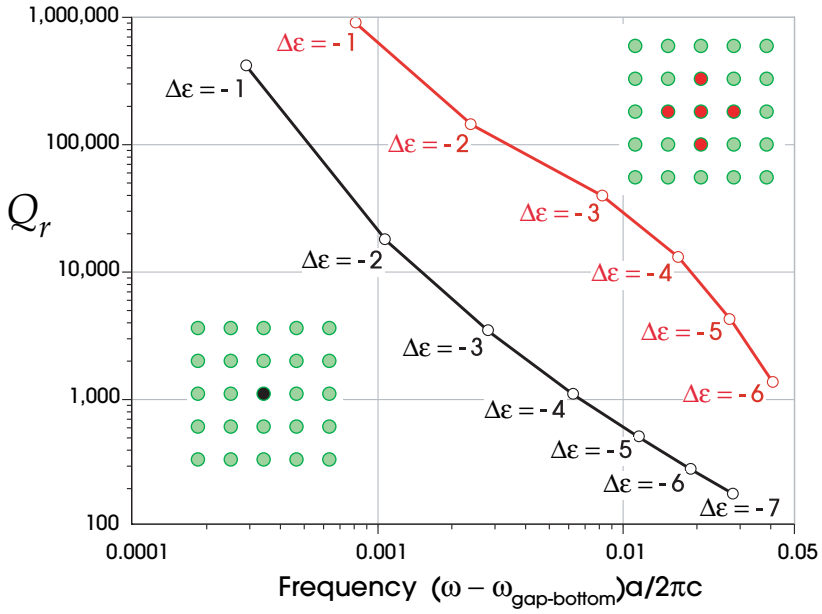


Figure 13: Radiative lifetime Q_r for resonant-cavity modes in the rod slab, versus the $\Delta\omega$ above the gap bottom, on a log-log scale. As the gap bottom is approached, the mode becomes more delocalized in the plane and Q_r increases. The lower curve (black) is for a cavity formed by decreasing the ε of a single rod (lower inset) by the $\Delta\varepsilon$ shown. The upper curve (red) is for a cavity formed by decreasing the ε of a five rods (upper inset) by the $\Delta\varepsilon$ shown, as in figure 12.

What is the reason for the increased Q_r in our delocalized monopole state? One way to understand this phenomenon is as follows. As the mode becomes delocalized in the xy plane, it bears a greater resemblance to the extended modes of the corresponding photonic crystal. Those modes are *guided*, and therefore the only radiative loss comes from the small portion of the field that scatters off the $\Delta\varepsilon$ defect. Equivalently, if one looks at the Fourier components of the field pattern, the more delocalized it is in space, the more localized its Fourier transform can be—this allows most of the Fourier components to lie inside the light cone and not radiate.

Why is it a disadvantage to increase the modal volume (V)?[16] One reason is that it limits the degree of miniaturization of components. Another reason is that a

[16] The precise definition of the modal volume depends on the specific physical process being optimized. For nonlinearity, see the section the section Nonlinear Filters and Bistability of chapter. 10. For spontaneous emission, one uses $V \triangleq \int d^3\mathbf{r}\varepsilon|\mathbf{E}|^2/\max(\varepsilon|\mathbf{E}|^2)$ or similar (Purcell, 1946; Coccioli et al., 1998; Robinson et al., 2005). Typically, V is given in scale-invariant units of $(\lambda/2n)^3$, i.e. cubic half-wavelengths in the material $n = \sqrt{\max\varepsilon}$.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

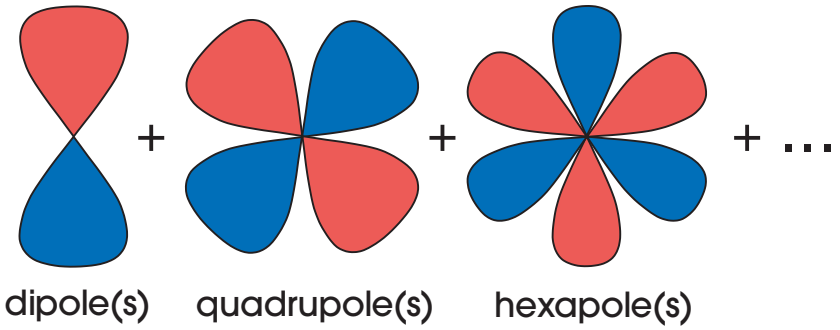


Figure 14: Schematic depiction of multipole expansion: the field radiated by a localized source is expanded as a sum of dipole, quadrupole, et cetera contributions, corresponding to an expansion in spherical harmonics. There are generally several terms of each type (e.g. dipoles oriented in different directions).

large modal volume often translates into a large power requirement. For example, in chapter 10 we will discuss how the power for certain nonlinear optical devices increases as V/Q^2. As another example, Purcell (1946) showed that the rate of spontaneous emission (which is an important factor in the performance of lasers and light-emitting diodes) is enhanced proportional to Q/V in a resonant cavity with the same frequency as the emitter. Still, since Q_r can increase exponentially faster than V from our ring-resonator example, it might seem that the optimum would be an infinitely large structure. This is not the case in practice, however, because Q is ultimately limited by other factors such as bandwidth considerations, material absorption, or fabrication disorder, and so one would usually prefer to minimize V for a given Q_r.

From a practical standpoint, the crucial point is that a modest amount of delocalization can often buy us a large gain in Q_r. The increase in V is often so small, in fact, that this physical mechanism is sometimes more easily noticed as a tradeoff in free spectral range—either because the cavity becomes multi-mode, as with the ring resonator, or because we are approaching a band edge as with the monopole cavity.

Cancellation

Often, it is possible to increase Q_r dramatically without increasing the modal volume or decreasing the free spectral range, simply by optimizing over a small number of structural parameters. Usually the physical mechanism that is responsible is *cancellation*: the dominant component of the radiation is eliminated by a forced balance of scattered fields with opposite signs.

We need to find a compact decomposition of the radiation into loss "components" in order to investigate this concept. Because we are considering the

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

radiation pattern from a localized source, it is natural to represent the radiated field by a **multipole expansion**.[17] Depicted schematically in figure 14, this is a decomposition of the field into a sum of dipole, quadrupole, hexapole, etc., radiation patterns. It turns out that the radiated power is simply a sum of the contributions from each multipole term (there is no inter-term interference). If we can arrange for the dominant term in the scattered field (usually the dipole term) to vanish, we will greatly enhance Q_r, and the far field will take on the characteristics of a higher-order radiation pattern.

We illustrate the mechanism of cancellation with a simple two-dimensional example. Consider a single line of dielectric cylinders, much like the structure of the section A Two-Dimensional Model of chapter 7 except with circles instead of squares. This structure has a TM band gap, and we can create a localized resonant mode by increasing the radius of a single rod, as depicted in upper-right panel of figure 15. In the lower-right panel, we plot $Q_r(\omega)$ as a function of the resonance frequency, which is itself a function of the radius of the defect rod. It is striking that there is a *peak* in Q_r, at a frequency in the middle of the gap rather than at the gap edge (as one might expect from our delocalization discussion). The peak arises from a cancellation of the lowest-order multipole moment, as we can confirm by examining the far-field pattern, which is shown in the left panel of figure 15. For the resonant mode that maximizes Q_r, the far-field radiation pattern has an extra nodal plane indicative of a higher-order multipole. The field in the cavity is hardly changed, but the far-field pattern has a different symmetry and is much reduced in amplitude. What has occurred is that the dipole-like field oscillation from the central rod has been cancelled at the peak by contributions associated with the adjacent rods, where the field has the opposite sign.

Cancellation also occurs in three-dimensional systems. For example, figure 16 shows a point-defect mode that is created in a hole slab by increasing the dielectric constant of a single hole. We parameterize the structure by p, the distance by which a single row of holes (including the defect) is *stretched*, as shown in the upper right.[18] The top left panel of figure 16 shows the frequency and the Q_r of one of the cavity modes as we continuously tune the parameter p. Over a range of p for which the frequency changes by only $\pm 5\%$, a factor-of-ten enhancement in Q_r is possible. Again, this is due to a dipole cancellation, as can be seen by inspecting the far-field intensity $\hat{\mathbf{z}} \cdot (\mathbf{E} \times \mathbf{H})$ in a plane well above the slab (as shown in the bottom panels). At the maximum Q_r, the radiation field has transitioned to

[17] Technically, one expands the radiated fields in terms of vector spherical harmonics; this should not be confused with the multipole expansion of charge distributions from electrostatics. See Jackson (1998).

[18] This design and its discovery are described in Vučković et al. (2002). Note that, for the range of p considered here, the line defect is *not* strong enough to create a waveguide mode that impinges on the cavity frequency. This cavity has a dipole-like pattern that supports *two* modes, which are doubly degenerate for $p = 0$; we show only the higher-Q_r mode, which has its $\mathbf{E}$ field pointing mostly parallel to the line defect.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

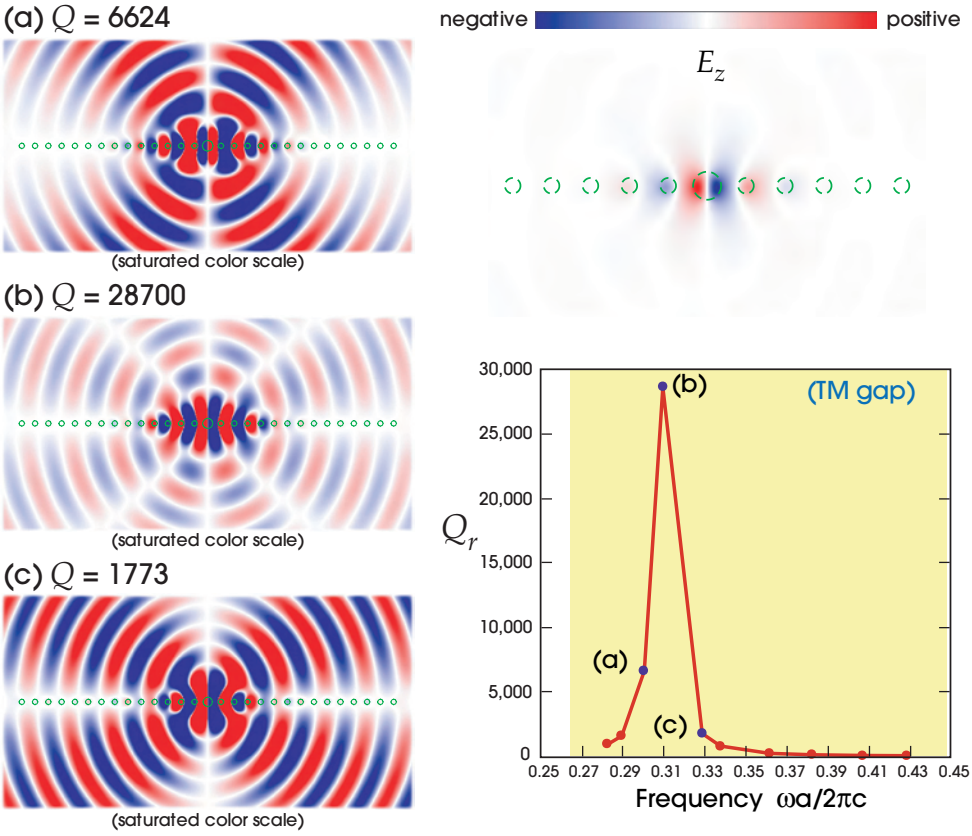


Figure 15: Increase in Q_r due to a cancellation of the lowest-order multipole moment in the two-dimensional cavity shown at upper right: an increased-radius dielectric rod in a period-a sequence of rods (green dashed outlines), trapping the mode whose E_z is shown. As the defect radius is varied, the frequency changes, and the resulting Q_r versus frequency is shown at lower right (with the TM gap in yellow). There is a sharp peak (b) of Q_r in the interior of the gap, corresponding to a field pattern (left middle) that has an extra pair of nodal planes compared to points (a) and (c) away from the peak (left top and bottom). The color scale for the left plots is saturated to exaggerate the small radiated field.

a higher-order quadrupole pattern.[19] As before, the mode pattern within the slab is nearly unchanged, even though the far field has been drastically altered.

A related structure, in which the defect was created by decreasing the radius of the central hole rather than increasing its dielectric constant, was fabricated to operate at $\lambda = 1.5\,\mu$m, and is shown in figure 17 (Lončar et al., 2002).

[19] We are grateful to A. Rodriguez at MIT for these calculations.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

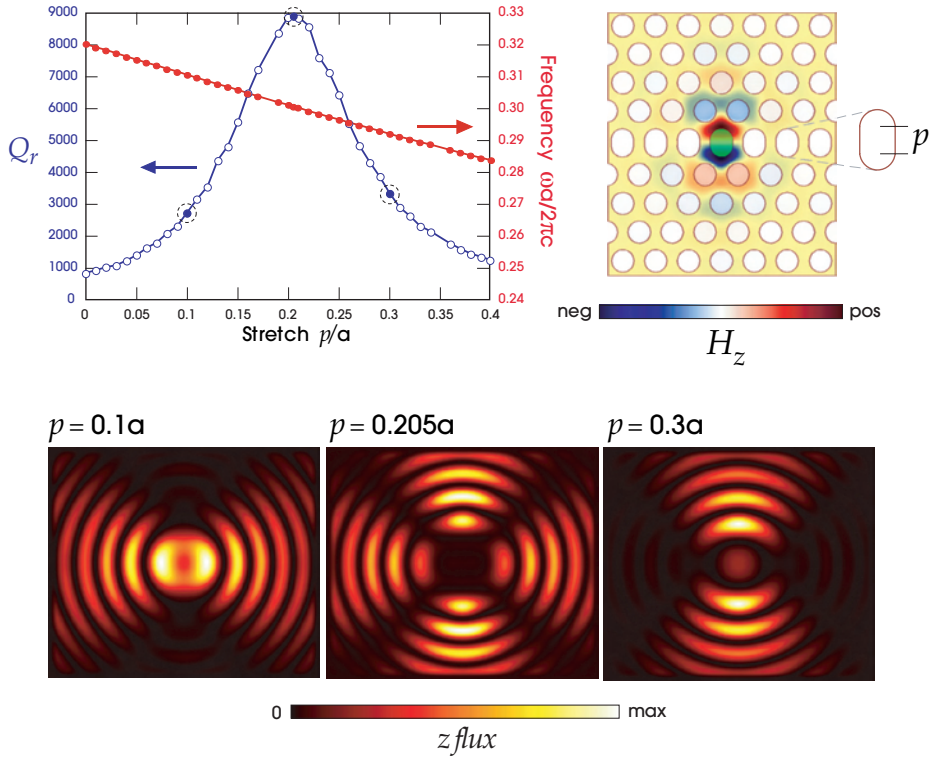


Figure 16: 3D multipole-cancellation effect in a hole-slab with an increased-ε point defect (green) and a tuning parameter p describing the elongation of a single row of holes (top right). *Top left:* Q_r (blue open circles) and frequency (red filled circles) vs. p. *Top right:* field pattern H_z in slab mid-plane for $p = 0.205a$ (peak Q_r), with dielectric material shaded yellow. *Bottom:* vertical (z) component of flux in a plane $3.5a$ above the slab, for mode at $p = 0.1a$ (left), $p = 0.205a$ (center), and $p = 0.3a$ (right), corresponding to the circled points on the graph.

The same cancellation mechanism can be viewed in a variety of ways, for example as cancelling a dominant $\mathbf{k}_\parallel$ Fourier component inside the light cone.[20] For cases with a patterned substrate, the radiated field itself may contain discrete guided modes that can be cancelled individually (Karalis et al., 2004). The defining characteristics, regardless, are a sharp peak in Q_r far from any band edge and the introduction of additional nodal planes in the radiation pattern.

[20] See, for example, Vučković et al. (2002) and Srinivasan and Painter (2002).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

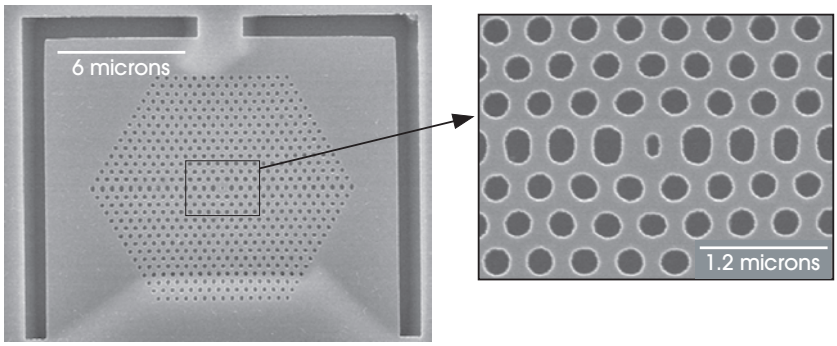


Figure 17: Electron-microscope image of experimental cavity structure in a hole-slab membrane (Lončar et al., 2002). Similar to the cavity of figure 16, but with reduced radius instead of increased ε for the central defect hole. (Image courtesy M. Lončar.)

Further Reading

The concept of index-guided modes below the light cone of a slab with two-dimensional periodicity was known as early as Ulrich and Tacke (1973), who describe an experimental measurement of the band diagram of a guided mode for a two-dimensionally periodic metallic surface. Theoretical considerations of waveguides with two-dimensional periodicity (as in, e.g., Zengerle, 1987) seem to have been largely confined to systems with weak dielectric contrast until the late 1990s, however. As experimental work began to probe higher-contrast regimes (e.g., Krauss et al., 1996), a rigorous three-dimensional theory became necessary.

The use of band gaps in the index-guided modes of two-dimensionally periodic photonic-crystal slabs with strong dielectric contrast was proposed at least as early as Meade et al. (1994). Calculations of the guided bands in such periodic slabs followed in Fan et al. (1997), Villeneuve et al. (1998), and Coccioli et al. (1998), the latter two also considering point-defect cavities. A comprehensive study of the slab guided modes and their band gaps can be found in Johnson et al. (1999). The theory of lossless linear-defect waveguides in slabs was described in Johnson et al. (2000) and Kuchinsky et al. (2000).

Early descriptions of the delocalization mechanism for high-Q cavities in two-dimensionally periodic slabs include Villeneuve et al. (1998) and Johnson et al. (2001*a*), or Benisty et al. (2000) for *vertical* delocalization. The cancellation mechanism was described in Johnson et al. (2001*c*). Some early examples of wavelength-scale cavities fabricated in high-contrast 2D-periodic slab structures include Painter et al. (1999), Lin et al. (2001), and Lončar et al. (2002); a notable recent high-Q cavity was reported in Song et al. (2005).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

9

Photonic-Crystal Fibers

THE MOST IMPORTANT CONDUIT for modern telecommunications is the **optical fiber**: a long filament of glass (or sometimes plastic) that guides light, often for a distance of many kilometers. Optical fibers are also used in a range of other applications, ranging from astrophysics to medicine. A traditional optical fiber consists of a central **core** that is surrounded by a **cladding** of slightly lower dielectric constant, which confines the light by index guiding (as described in the subsection Index guiding of chapter 3). In this chapter, we will find that new regimes are opened for fiber operation by incorporating periodic structures in the cladding: **photonic-crystal fibers**.[1]

Mechanisms of Confinement

Photonic-crystal fibers, also called **microstructured optical fibers**, can be divided into a few broad classes, according to whether they use index guiding or band gaps for optical confinement, and whether the periodicity of the structure is one-dimensional or two-dimensional.

Photonic-bandgap fibers confine light using a band gap rather than index guiding. Band-gap confinement is attractive because it allows light to be guided within a hollow core (much as in the section Linear Defects and Waveguides of chapter 5). This minimizes the effects of losses, undesired nonlinearities, and any other unwanted properties of the bulk materials that are available. Band-gap fibers with a *one*-dimensional periodicity—a cladding with a series of concentric

[1] It is sometimes suggested that the term "photonic-crystal fiber" be restricted to photonic-bandgap fibers. The term "photonic crystal" was coined in analogy with atomic crystals, however, and refers to the periodicity of the structure, rather than to the presence or exploitation of a band gap.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

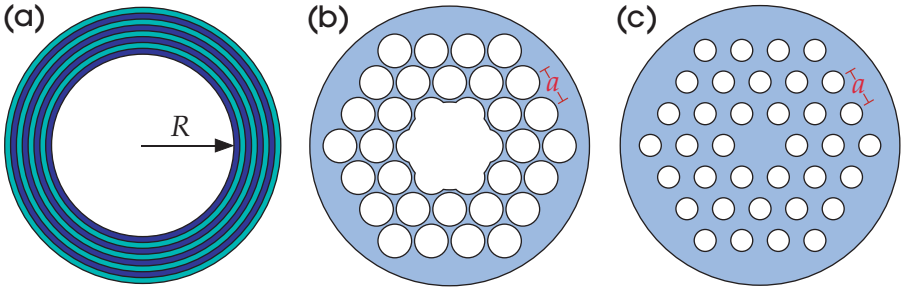


Figure 1: Three examples of photonic-crystal fibers. (a) Bragg fiber, with a one-dimensionally periodic cladding of concentric layers. (b) Two-dimensionally periodic structure (a triangular lattice of air holes, or "holey fiber"), confining light in a hollow core by a band gap. (c) Holey fiber that confines light in a solid core by index guiding.

layers as in figure 1(a)—were first analyzed precisely[2] by Yeh et al. (1978), who called them **Bragg fibers**. Band-gap fibers with *two*-dimensionally periodic claddings, as in figure 1(b), were described by Knight et al. (1998). The most commonly used design is a **holey fiber**, such as the one shown here, in which the cross section is a periodic array of *air holes* running the whole length of the fiber.

Another possibility is an index-guiding **photonic-crystal fiber**, in which the periodic structure is not employed for its band gap, but rather to form an effective low-index cladding around the core. One way to accomplish this is with a solid-core holey-fiber structure, as in figure 1(c). In this way, one can obtain a much higher dielectric contrast than is generally possible with solid fiber materials, leading to a greater strength of optical confinement. This is often useful as a means of enhancing nonlinear effects, or of creating unusual dispersion phenomena.

A final possibility is a periodic modulation of the structure along the fiber's direction of propagation. This is known as a **fiber Bragg grating** (not to be confused with a *Bragg fiber*, above). Such a fiber is simply a kind of periodic dielectric waveguide, as we already discussed at the end of chapter 7. Thus, in this chapter we will focus exclusively on the cases in which the periodicity is in the *transverse* directions.

Photonic-crystal fibers have an enormous practical advantage over the periodic structures that we discussed in previous chapters: fibers can be created through a **drawing** process. In the first step of this process, a scale model of the fiber (or **preform**) is created, typically centimeters in size. Next, the preform is heated and pulled (*drawn*), stretching it like bubble gum into a thin strand whose cross section is a scaled-down version of the preform's.[3] In this way, hundreds of meters or

[2] An earlier paper by Melekhin and Manenkov (1968) proposed and analyzed a similar cylindrical structure in the large-core (ray-optics) approximation.

[3] For fibers with air holes, it is possible for the holes to partially (or totally) collapse, depending on the draw conditions.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

even kilometers of fiber can be drawn from a single preform, with near-perfect uniformity.

In this chapter, we proceed in an order that is somewhat the reverse of the previous chapters. We begin with index-guiding fibers, since those are the easiest to understand. They also demonstrate the important concept of the high-frequency scalar limit, a rigorous asymptotic form of the eigenmodes that we use to explain many phenomena in this chapter. Then, for photonic-bandgap fibers, we start with the case of *two*-dimensional periodicity and follow with one-dimensional periodicity. The reason for this reversal is not only that the holey bandgap fibers are closely related to their index-guiding counterparts, but also that Bragg fibers introduce some new concepts: continuous rotational symmetry and a new form of "tm" and "te" polarizations. For each fiber, we give examples of potential applications, but these examples are by no means exhaustive; our primary intention is to illustrate the fundamental principles by which these fibers are both connected to and distinguished from their predecessors.

Index-Guiding Photonic-Crystal Fibers

The easiest photonic-crystal fibers to understand are those that employ index guiding. They guide light by virtue of the smaller average refractive index of the cladding relative to the core. A typical example is the *holey fiber* of figure 1(c), in which the cladding has a cross section that is a triangular lattice of air holes within an otherwise uniform dielectric medium. In our example, we use a spatial period a, with holes of radius $0.3a$ and a background dielectric constant of $\varepsilon = 2.1$ (approximately that of silica glass at $\lambda = 1.55\,\mu$m). The "core" is really just the location of a missing hole in the center.[4] One might hope that it would be sufficient to consider only some "average" index contrast between core and cladding, but in fact a full understanding of this case requires an analysis of the band diagram.

Because a fiber has translational symmetry along the fiber axis (which we take to be the z axis), k_z is conserved, and we can write the field in the usual Bloch form: $\mathbf{H}(x,y,t) = \mathbf{H}(x,y)e^{ik_z z - i\omega t}$. We then plot ω versus k_z to obtain the band diagram (or **dispersion relation**), which is shown as an inset diagram in figure 2. As in the subsection Index guiding of chapter 3, and also as in chapters 7 and 8, the projected band diagram consists of two parts: a continuum of frequencies (the light cone) representing all of the possible extended states within the cladding, and a discrete set of guided bands with frequencies lying below the light cone. If the cladding material were uniform with a dielectric constant ε (independent of ω), then the light line (the lower boundary of the light cone) would be a straight line, $\omega = ck_z/\sqrt{\varepsilon}$. For a nonuniform cladding such as our lattice of holes, the light line is not straight. Instead, it is given by the **fundamental space-filling mode** of the cladding, which is the lowest-ω extended mode of the cladding at each k_z.

[4] An early fabrication of such a solid-core holey fiber was reported by Knight et al. (1996).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

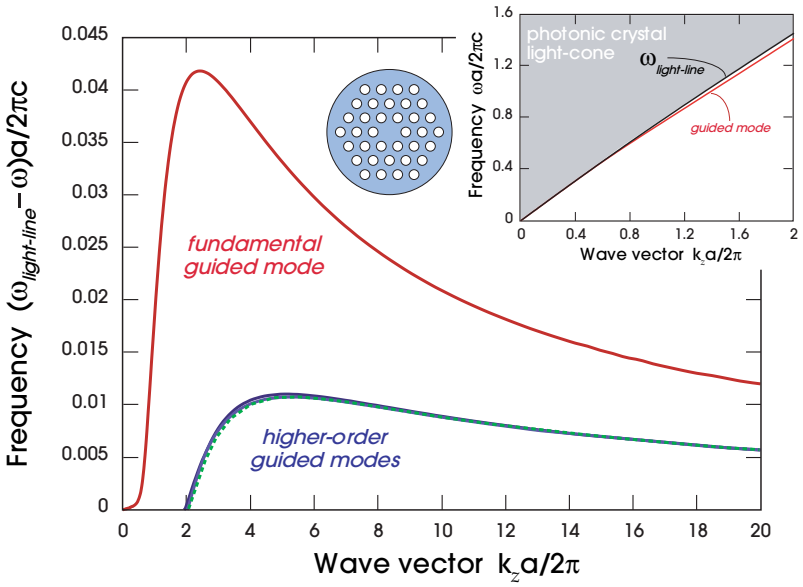


Figure 2: Band diagram of solid-core holey fiber as a function of axial wave vector k_z. The usual ω plot is inset, but for clarity we also plot the $\Delta\omega$ between the guided bands and the light line. The higher-order guided modes are three bands that are nearly on top of one another.

In this structure, however, the guided mode is so close to the light cone that it is more convenient to plot the difference $\Delta\omega = \omega_{\text{lc}} - \omega$ between the light line ω_{lc} and the guided band ω, rather than plotting ω itself. This is done in figure 2. We have defined $\Delta\omega$ such that it is positive for an index-guided mode.

We can calculate the light line in the same way we did previously when considering the continua of extended modes for linear defects or surface states in chapters 5 and 6. For each k_z, we find all of the extended modes of the infinite periodic cladding (i.e., without the core) for all possible transverse wave vectors (k_x, k_y). Then we plot the resulting frequencies as a function of k_z; the lowest frequency for each k_z defines the light line. These extended modes are analyzed by considering the periodic cladding by itself, exactly as in the section Out-of-Plane Propagation of chapter 5: one need only consider (k_x, k_y) in the irreducible Brillouin zone of the triangular lattice. The extended modes take the Bloch form of a plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$ multiplied by a periodic envelope function $\mathbf{H}_{\mathbf{k}}(x, y)$.

The increased ε of the core introduces one or more guided modes, by pulling down modes beneath the light line. Because they are below the light line, these modes must decay exponentially into the cladding. The farther below the light line they are pulled, the faster the transverse decay. For the case of figure 1(c), a doubly degenerate band is localized in the core, whose field patterns are shown

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

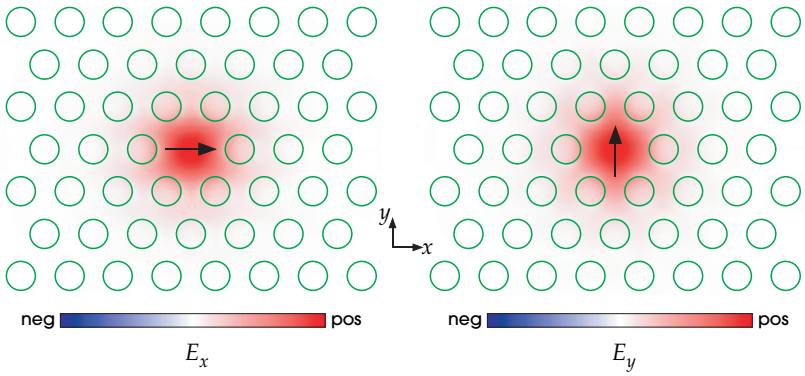


Figure 3: Electric-field pattern for the doubly degenerate fundamental mode of figure 2. Their polarizations are nearly orthogonal everywhere: the mode pictured at left is mostly E_x, and the mode pictured at right is mostly E_y. The green circles show the locations of the air holes.

in figure 3. We call this the **fundamental mode**. In general, the fundamental mode is defined as the mode with the largest k_z for any given ω or, equivalently, the smallest ω for any given k_z. This is the analogue of the two degenerate, orthogonal, linearly polarized "LP_{01}" modes[5] that propagate within standard "single-mode" silica fibers. Here, due to the large index contrast and sixfold symmetry, the two orthogonal modes are neither purely linearly polarized nor are they exactly related by a 90° rotation.[6]

For larger values of k_z, three additional guided bands are localized. These appear below the light line at $k_z a/2\pi \approx 2$ in figure 2. One of these bands is doubly degenerate, and is essentially a higher-order version of the fundamental mode, with an extra nodal plane perpendicular to the direction of polarization. The other two bands are nondegenerate. None of these higher-order modes can be excited by a planewave source incident in the z direction, because the source and the modes would have different symmetries.[7] It might seem surprising that

[5] In standard silica fiber, the index contrast is so low (less than 1%) that one can describe the fields by a *scalar* approximation with a single polarization direction; see the subsection The scalar limit and LP modes of chapter.

[6] This waveguide is described by the C_{6v} symmetry group, which has a doubly degenerate representation corresponding conceptually to orthogonal polarizations. Although the two degenerate states do *not* differ merely by a 90° rotation, they *do* have opposite even/odd symmetry under orthogonal mirror planes, and thus can be selectively excited from a planewave light source. (More precisely, one state is the average of 60° and 120° rotations of the other state.) See Inui et al. (1996, Table B.12).

[7] They transform according to different representations of the C_{6v} symmetry group. With respect to the labelling in Inui et al. (1996, Table B.12), both the fundamental mode and planewave light transform as Γ_6, the higher-order doubly degenerate band is Γ_5 (which transforms like xy and $x^2 - y^2$), and the other two bands are Γ_1 (fully symmetric) and Γ_2 (invariant under rotations, but odd under mirror planes).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

these three higher-order bands lie almost exactly on top of one another; this apparent coincidence is explained in the subsection The scalar limit and LP modes below.

Endlessly single-mode fibers

In an ordinary index-guided waveguide, as one goes to higher and higher ω (smaller wavelength λ), more and more guided modes are pulled below the light line. We first saw this phenomenon in figure 3 of chapter 3. Eventually, one approaches the ray-optics limit, in which the guided modes are described by a continuum of angles greater than the critical angle for total internal reflection. However, as first pointed out by Birks et al. (1997), this need not be true of photonic-crystal fibers: they can remain *endlessly single-mode*, regardless of wavelength (limited only by the material properties).

In figure 2, we saw that our holey fiber guides up to four bands, so it is not endlessly *single*-mode. However, it still displays the essential features of this phenomenon, because as we go to higher and higher k_z (or larger ω) we do *not* get more and more bands—the number of bands never exceeds four. We could reduce the number of modes to one (or, to be precise, one pair of doubly degenerate states) by halving the hole radius to $0.15a$, thereby weakening the strength of confinement.

Why are the higher-order modes absent? The reason is that the effective index contrast between the core and the cladding in the holey fiber *decreases* at smaller wavelengths, rather than remaining fixed as it would for a homogeneous cladding. Thus, the strength of confinement is weaker at smaller wavelengths, and higher-order guided modes remain above the (lowered) light line. To be more concrete, we define the **effective index** of a mode as ck_z/ω, the factor by which the phase velocity ω/k_z is decreased from c; the effective index equals the ordinary refractive index for a plane wave in a homogeneous medium. At a given ω, an index-guided mode obviously has a larger effective index than that of the light line. To show the decrease in effective-index contrast with wavelength, figure 4 shows the effective index versus vacuum wavelength $\lambda/a = 2\pi c/\omega a$, based on the band diagram given in figure 2.

There is an intuitive explanation for why the effective-index contrast decreases with wavelength. As we first showed from the variational theorem in the section Electromagnetic Energy and the Variational Principle of chapter 2, the fundamental (light-line) mode of the cladding "wants" to be concentrated as much as possible in the high-dielectric material. For long wavelengths $\lambda \gg a$, the light cannot lie entirely within the high dielectric, because the field cannot vary faster than the wavelength. As we go to shorter and shorter λ, however, more of the light can "fit" in the dielectric between the holes. In the limit of small $\lambda \ll a$, the ray-optics limit applies, and the light is guided by total internal reflection and remains almost entirely within the dielectric material, with an effective index that approaches the index of the dielectric material. Since the core is made of the same

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

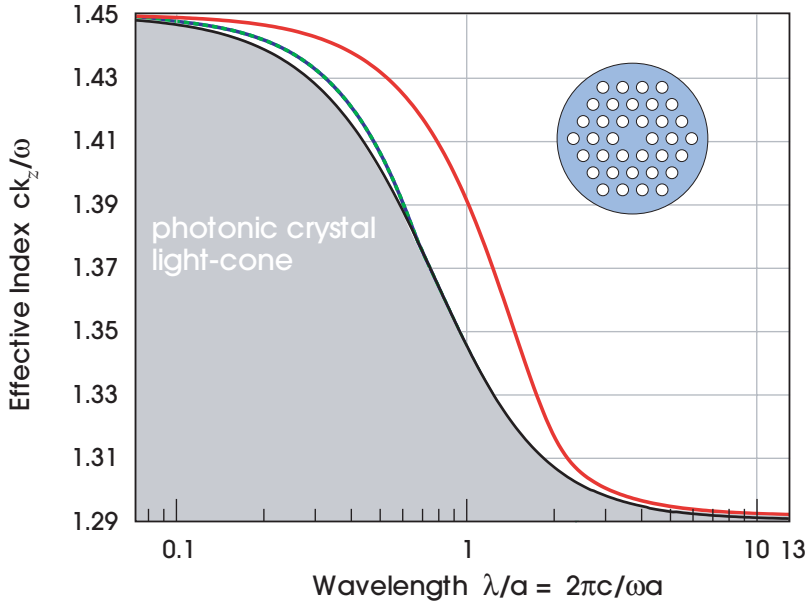


Figure 4: Bands of the solid-core holey fiber, from figure 2, plotted as "effective index" ck_z/ω versus wavelength, along with the light cone (gray region). In the limit of small λ, the effective indices of both the modes and the light line approach the index 1.45 of the bulk silica.

dielectric material, the effective index of the guided mode must also approach the same value as the wavelength decreases. Precisely these limits are seen in figure 4, where both the light line and the guided mode approach the index $\sqrt{\varepsilon} = 1.45$ of the bulk silica.

However, this explanation is not complete. For example, could the effective-index contrast decrease so fast that the modes become less and less confined to the core for small λ? Or perhaps the effective-index contrast does not decrease fast enough to asymptotically exclude higher-order modes? Neither of these is the case: as is derived more rigorously in the next subsection, in the limit $k_z \to \infty$ we obtain a finite number of modes with fixed field patterns.

Of course, in a real material we must eventually take into account the fact that ε is a function of frequency, and indeed the material may cease to be transparent at some ω. On the other hand, we can equivalently keep ω fixed and rescale the *structure*, in which case the above analysis is exact. The endlessly single-mode property, in this case, means that we can make the waveguide arbitrarily large and still guide only a single (doubly degenerate) waveguide mode. This could be useful for reducing the effects of material nonlinearities, although one is eventually limited by the fact that bending losses tend to increase with mode size.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

The scalar limit and LP modes

The key to a quantitative understanding of the large-k_z limit is to realize that this regime is asymptotically described by a *scalar* wave equation that is independent of k_z. Consequently, for large k_z, the modes approach k_z-independent "linearly polarized" (**LP**) field patterns. Indeed, we shall see that this scalar limit is useful for understanding other fiber phenomena as well, such as the existence of photonic band gaps.

Traditionally, the scalar approximation in electromagnetism is formulated only for structures with a *small* dielectric contrast. The dielectric function of such a medium can be described as the sum of a constant ε_c and a small perturbation $\delta\varepsilon(x,y) \ll \varepsilon_c$. In that case, if we neglect terms of order $|\nabla\delta\varepsilon|$, then the Maxwell equations for the electric field $\mathbf{E}$ can be written[8]

$$\left[\nabla^2 + \frac{\omega^2}{c^2}\varepsilon(x,y)\right]\mathbf{E} = 0. \tag{1}$$

In this approximation, the different components of $\mathbf{E}$ are decoupled from one another, although they are not completely independent because of the transversality constraint $\nabla \cdot \varepsilon\mathbf{E} = 0$. (This constraint allows E_z to be determined from E_x and E_y, for example.) If we combine these results with Bloch's theorem for the waveguide modes, it follows that we can write the *transverse* (xy) components of $\mathbf{E}$ in terms of a single *scalar* function $\psi(x,y)$ (an LP mode):

$$\mathbf{E}_t = [p_x\hat{\mathbf{x}} + p_y\hat{\mathbf{y}}]\psi(x,y)e^{ik_z z}, \tag{2}$$

where p_x and p_y are constants that specify the amplitude and the direction of polarization, and the subscript t stands for *transverse*. The function ψ satisfies the eigenequation

$$\left[-\nabla_t^2 - \frac{\omega^2}{c^2}\delta\varepsilon(x,y)\right]\psi = k_t^2\psi, \tag{3}$$

reminiscent of the Schrödinger equation of quantum mechanics. In this equation, ∇_t represents the transverse (x and y) components of ∇, and k_t is a *transverse wave number* defined as

$$k_t \triangleq \sqrt{\frac{\omega^2}{c^2}\varepsilon_c - k_z^2}. \tag{4}$$

In contrast to this traditional approximation, a photonic-crystal fiber generally has a large index contrast. For this reason, it may be surprising that a photonic-crystal fiber can also be accurately described by a scalar approximation. It works

[8] See, for example, Jackson (1998), chapter 8.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

as follows. Suppose that in addition to the small variations $\delta\varepsilon$, we also have some very low-index regions (e.g., the air holes) with a dielectric constant $\varepsilon = \varepsilon_c - \Delta\varepsilon$. In this case, $\Delta\varepsilon$ is large and positive. The key fact is that for large k_z, the fields within the very low-index regions become very small, because of index guiding. We may therefore use the scalar approximation (3) in the regions where $\Delta\varepsilon = 0$, and simply set $\psi = 0$ where $\Delta\varepsilon \neq 0$.[9] Seen this way, the effect of the air holes is to impose a boundary condition on ψ.

To be more explicit, recall from the subsection Index guiding of chapter 3 that the fields fall off exponentially into a low-index region with a spatial decay constant of $\kappa = \sqrt{k_z^2 - \frac{\omega^2}{c^2}\varepsilon}$. In terms of k_t, $\kappa \approx k_z\sqrt{\Delta\varepsilon/\varepsilon_c}[1 - O(k_t^2/k_z^2)]$. The fields in this region can therefore be neglected when $\kappa \sim k_z \gg k_t$ (i.e. when the field decays much faster than the transverse ψ oscillations). However, the condition $k_t \ll k_z$ is equivalent to the condition that the effective index ck_z/ω approaches $\sqrt{\varepsilon_c}$, which we already saw must be true at high frequencies (large k_z).

Why is this important? The scalar limit for large k_z has several interesting consequences.

First, if $\delta\varepsilon = 0$, as it is for our holey fiber, then ψ satisfies an eigenequation $-\nabla_t^2\psi = k_t^2\psi$ (with $\psi = 0$ in the holes) in which neither k_z nor ω appear explicitly. Thus, the values of neither k_z nor ω will affect the solution ψ or the eigenvalue k_t^2. We can conclude that, for large k_z, the modes approach fixed field patterns obeying a dispersion relation $\omega^2 = c^2(k_t^2 + k_z^2)/\varepsilon_c$.

Second, each mode ψ in the scalar limit, a so-called **LP mode** (Gloge, 1971), corresponds to *several* vectorial solutions of the Maxwell equations for the *same* $|\psi|^2$ intensity pattern and the same eigenvalue k_t. There are two possibilities. If ψ is a nondegenerate mode, then we get two vectorial modes $\psi\hat{\mathbf{x}}$ and $\psi\hat{\mathbf{y}}$, corresponding to a doubly degenerate "linearly polarized" state. If ψ itself is a doubly degenerate state with two solutions $\psi^{(1)}$ and $\psi^{(2)}$, then we get *four* vectorial modes $\psi^{(\ell)}\hat{\mathbf{x}}$ and $\psi^{(\ell)}\hat{\mathbf{y}}$ for $\ell = 1, 2$. For finite k_z, the scalar approximation is not exact and such degeneracies break (leaving at most doubly degenerate pairs). The states divide into linear combinations corresponding to different vectorial eigenmodes. It is precisely such LP modes that we saw in figure 2: a pair of doubly degenerate modes corresponding to a nondegenerate (monopole-like) ψ, and four *nearly* degenerate modes (including one doubly degenerate pair) corresponding to a doubly degenerate (dipole-like) ψ.[10]

[9] See also Birks et al. (1997) for further discussion of this approximation and some of its consequences.

[10] The way that these LP modes break down into vectorial modes can be predicted from group theory. The combination $\psi\hat{\mathbf{x}}$ transforms as a *product representation* of the symmetry-group representations corresponding to ψ and $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}\}$, and such a product can be decomposed only into certain irreducible representations corresponding to the vectorial modes (see, e.g., Inui et al., 1996). Thus, for example, only some symmetries of ψ will lead to modes that can couple to a planewave input light from the z direction. In structures with C_{6v} symmetry, as here, it also follows that the LP modes will always correspond to either a single doubly degenerate mode *or* to two nondegenerate modes and a doubly degenerate mode.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

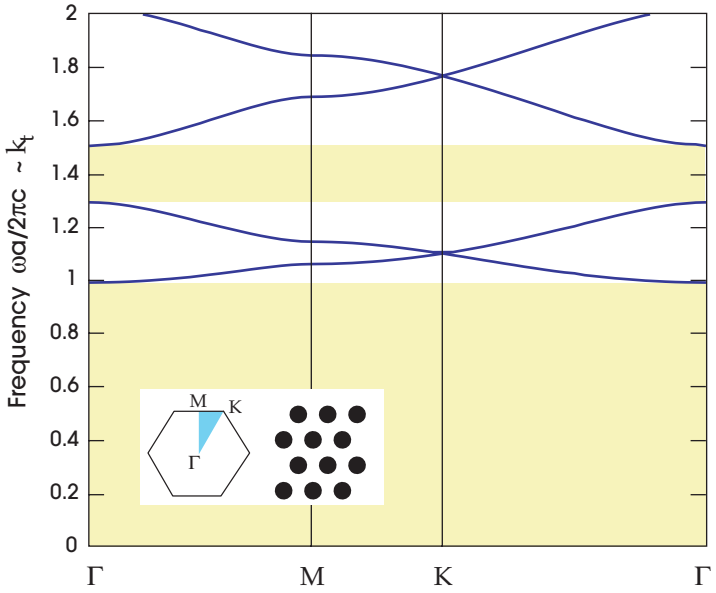


Figure 5: Band diagram for TM-polarized modes of a two-dimensional triangular lattice of perfect-metal cylinders. Two band gaps are shown (shaded yellow), where the lowest band has a low-frequency cutoff characteristic of metallic structures. These bands are equivalent to the modes of the holey fiber in the scalar (large k_z) limit.

Third, we can now predict whether the $k_z \to \infty$ limit will yield a finite or an infinite number of guided modes. Again, we suppose $\delta\varepsilon = 0$, for simplicity. If the low-index ($\Delta\varepsilon$) regions completely surround the core, then in the scalar limit the field behaves like the familiar quantum problem of a "particle in a box" with infinite potential barriers. The "box" supports arbitrarily many modes, limited only by the $k_t \ll k_z$ approximation. On the other hand, for a connected structure, when the eigenvalues k_t are large the scalar field ψ can "leak out" between the holes, and the modes are not guided. Mathematically, this situation is identical to a two-dimensional (2d) photonic crystal of *perfect-metal* rods ($\varepsilon \to -\infty$), for the case of the TM polarization: k_t^2 corresponds to the 2D frequency eigenvalue ω^2/c^2, and ψ corresponds to E_z. The band diagram of this analogous 2D metallic structure is shown in figure 5. It exhibits a well-known property of **metallo–dielectric photonic crystals**: there is a band gap starting at $k_t = 0$ and extending to the minimum of the first band.[11] This finite gap, in turn, corresponds

[11] This gap is somewhat analogous to the **plasma frequency** ω_p for the free-electron dispersion model $\varepsilon(\omega) = 1 - \omega_p^2/\omega^2$ (Jackson, 1998), where for $\omega < \omega_p$ we have $\varepsilon < 0$ and the material is opaque. An early description of this "plasmonic" gap in two-dimensional metallo–dielectric photonic crystals can be found in Kuzmiak et al. (1994).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

to a finite number of discrete-k_t localized modes supported by a defect. Another important feature that appears in figure 5 is that (in this scalar/metallic limit) there are also "ordinary" photonic band gaps that appear between higher bands of the structure. We will return to this point in the section Band-Gap Guidance in Holey Fibers.

Enhancement of nonlinear effects

One of the primary applications of index-guiding photonic-crystal fibers has been to enhance the strength of nonlinear optical effects. Nonlinear phenomena in fibers are typically due to the **Kerr effect**, in which the index varies in proportion to the intensity of the field. For a plane wave in an infinite isotropic medium, the refractive index $n = \sqrt{\varepsilon}$ is altered by $\Delta n = n_2 I$ on average, where I is the time-averaged intensity and n_2 is called the **Kerr coefficient**. When this effect is significant, the frequency ω is *not* conserved, and one can perform a number of useful tricks, such as creating modes that do not disperse (**solitons**), converting a signal from one frequency to another, or generating a whole range of frequencies from a single input frequency (**supercontinuum generation**).[12] However, most materials have rather weak nonlinearities (small n_2),[13] and it is necessary to use very high field intensities or to propagate over very long distances before the effects of Δn become significant. This is why we have been able to neglect nonlinearities for most of this book. Photonic-crystal fibers provide an opportunity to enhance nonlinear phenomena by concentrating the field into a smaller spatial volume, thereby increasing the intensity for a given total input power. Moreover, the large index contrast in the structure permits drastic modifications to the band diagram. For example, one can arrange for modes at different frequencies to have equal group velocities, thereby greatly increasing any interactions between the modes.

In general, the description of light propagation in nonlinear waveguides is an intricate subject involving the solution of nonlinear partial differential equations.[14] For our purposes, though, the most important result is that the strength of nonlinear effects for a particular waveguide mode can be characterized by a single number γ, which is the change in its wave vector (Δk_z) per unit power (P) propagating in the mode. Given this number $\gamma \triangleq \Delta k_z / P$, the inverse $1/\Delta k_z = 1/\gamma P$ is the length scale after which nonlinear effects become significant. Over much shorter distances, they can be ignored. The number γ appears in the nonlinear "Schrödinger" equation that governs propagation in the nonlinear regime and, as we will see, γ can be calculated from a single equation involving the eigenmodes of the *linear* waveguide structure.

[12] Demonstrations of supercontinuum generation in photonic-crystal fibers were reported in Ranka et al. (2000) and Wadsworth et al. (2002).

[13] For silica glass, n_2 is around 3×10^{-8} μm^2/W.

[14] For a detailed discussion of nonlinear waveguide propagation, see Agrawal (2001).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

In the limit of low dielectric contrast, there is a simple and frequently used expression for Δk_z:

$$\Delta k_z = \gamma P = \frac{\omega}{c}\frac{n_2 P}{A_{\text{eff}}}, \tag{5}$$

where A_{eff} is the **effective area** of the mode, a measure of its lateral spread that is described in more detail below. Equation (5) says that $\Delta k_z = \omega \Delta n / c$, where $\Delta n = n_2 \bar{I}$ is determined by the average intensity $\bar{I} = P/A_{\text{eff}}$ in the material. In a photonic-crystal fiber, however, it is not sufficient to know the area alone; one must also know where the mode is concentrated. Furthermore, there are vectorial effects in the nonlinear susceptibility that become significant for large dielectric contrasts (away from the large-k_z limit), but that are neglected in the common scalar version of equation (5).

It is useful, instead, to adopt equation (5) as the *definition* of the effective area A_{eff} (restricting ourselves to a system with a single nonlinear n_2 material, for simplicity). For a given structure, we can then solve for A_{eff} of equation (5) using the perturbation theory of the section The Effect of Small Perturbations of chapter 2. It is safe to use perturbation theory because the nonlinear shift Δn of realistic materials is almost always well below 1%. (In fact, first-order perturbation theory is implicitly assumed when one supposes Δk_z to be proportional to P.) From equation (28) in chapter 2,[15] we find $\Delta\omega$ from a given $\Delta\varepsilon \sim |\mathbf{E}(\mathbf{r},t)|^2$, and then obtain $\Delta k_z = \Delta\omega / v_g$ where $v_g = d\omega/dk_z$ is the group velocity. After some algebra, recalling that group velocity is equal to the ratio of power to energy density, one obtains[16]

$$A_{\text{eff}} = \frac{\mu_0}{\varepsilon_0}\frac{\left[\text{Re}\int d^2\mathbf{r}\ (\mathbf{E}^* \times \mathbf{H})\cdot\hat{\mathbf{z}}\right]^2}{\int_{n_2} d^2\mathbf{r}\ \frac{\varepsilon}{3}\left(|\mathbf{E}\cdot\mathbf{E}|^2 + 2|\mathbf{E}|^4\right)}, \tag{6}$$

where $\mathbf{E}$ and $\mathbf{H}$ are the (complex) eigenfields of the *linear* $\varepsilon(\mathbf{r})$ ($n_2 = 0$), and the $\int_{n_2}$ is an integral over the portion of the cross section in which the nonlinear n_2 appears.

The effective area of the fundamental mode of the holey index-guided fiber is shown in figure 6 as a function of λ/a. For comparison, we also plot the A_{eff} of the fundamental mode of a single cylinder with diameter a and $\varepsilon = 2.1$(silica),

[15] Applying perturbation theory here does require some care, because in a nonlinear system we must work with the physical, real-valued field $\mathbf{E}_r(\mathbf{r},t) = (\mathbf{E}(\mathbf{r})e^{-i\omega t} + \mathbf{E}(\mathbf{r})^* e^{+i\omega t})/2$.

[16] This effective area was derived by Tzolov et al. (1995). Near a zero group-velocity point it is more appropriate to work with $\Delta\omega$ rather than Δk_z, and the corresponding expression was derived by Lidorikis et al. (2004); note that the latter authors use a definition of n_2 that is 4/3 of the one here, however. (Note also that $|\mathbf{E}\cdot\mathbf{E}|^2$ is *not* the same thing as $|\mathbf{E}|^4 = (\mathbf{E}^*\cdot\mathbf{E})^2$.) In the scalar limit, where $\mathbf{E}$ is a single transverse polarization with real amplitude ψ, (6) reduces to the standard expression $A_{\text{eff}} = (\int |\psi|^2)^2 / \int |\psi|^4$.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

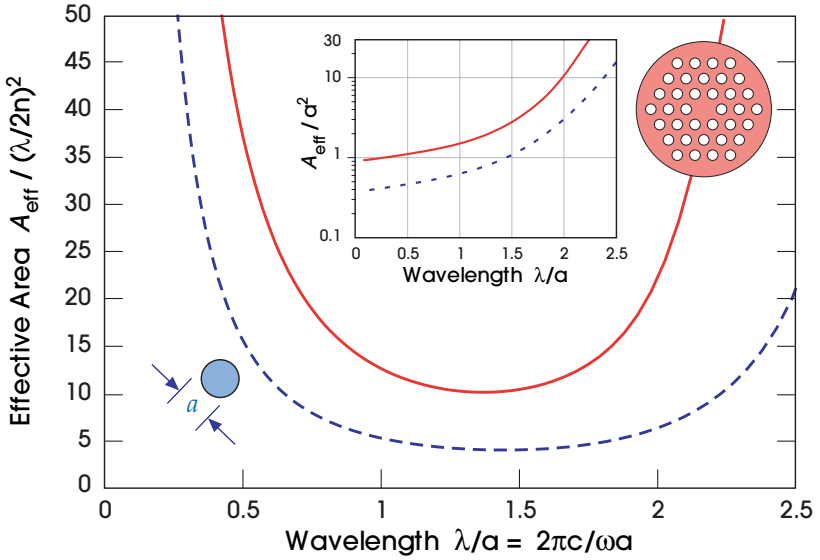


Figure 6: Effective area, in units of the wavelength squared, as a function of wavelength, for the solid-core holey fiber with period a (solid red line) and a simple dielectric cylinder of diameter a (dashed blue line). Inset is the same plot, but with the effective area in units of a^2.

surrounded by air.[17] (We neglect the nonlinearity of air itself.) The main plot shows A_{eff} in units of $(\lambda/2n)^2$, where $n=\sqrt{\varepsilon}$ is the refractive index of the silica. This scale-invariant ratio tells us the size of the mode relative to the "natural diameter" of half a wavelength in the dielectric; it also lets us compare effective areas for the case where we keep λ fixed and vary a. In both the holey fiber and the dielectric cylinder, there is a value of λ/a for which A_{eff} is minimized: for the holey fiber, $A_{\text{eff}} \approx 10.1\,(\lambda/2n)^2$ at $\lambda/a \approx 1.37$ ($k_z a/2\pi \approx 1.0$), and for the dielectric cylinder, $A_{\text{eff}} \approx 4.04\,(\lambda/2n)^2$ at $\lambda/a \approx 1.46$. The reason that an optimum exists is straightforward. If λ is too large compared to a, the mode sees the core as a small perturbation and is therefore only weakly guided, and has a large area. If λ is too small, then many wavelengths fit into the solid core. For both waveguides, the inset of figure 6 shows A_{eff} in units of a^2 (i.e., for when we keep a fixed and vary λ). We see that in the $\lambda \ll a$ limit, A_{eff} in these units approaches a constant that is somewhat smaller than the core area. This is consistent with the asymptotic prediction of the scalar limit, which told us that we should obtain a fixed field pattern (and hence fixed A_{eff}) for small λ.

In contrast, a standard single-mode doped-silica fiber has an effective area of 50–80 μm^2 at $\lambda = 1.55\,\mu\text{m}$, or 175–280 $(\lambda/2n)^2$, and certain highly nonlinear

[17] Such subwavelength silica cylinders have been fabricated by drawing down standard silica fiber (Birks et al., 2000; Tong et al., 2003).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

doped-silica fibers[18] have an A_{eff} of around $10\,\mu\text{m}^2 = 35\,(\lambda/2n)^2$. These larger effective areas translate directly into larger power requirements for nonlinear devices. Their nonlinear applications are further limited by the difficulty of drastic modifications to group-velocity dispersion in low-contrast waveguides.

Conversely, the hollow-core fibers of the following sections can have *weaker* nonlinearities, sometimes by orders of magnitude, than a corresponding mode propagating in the solid constituent materials. This may ultimately lead to highly *linear* fibers that can carry signals over great distances without the distortions that arise from nonlinear effects (such as crosstalk between different frequency channels).

Band-Gap Guidance in Holey Fibers

Index guiding can be relied upon to confine light only within regions of higher effective index. In contrast, a photonic band gap can localize light in a waveguide with a *lower* index, such as the hollow core in figure 1(b). Of course, a fiber cannot have a complete band gap, because of its continuous translational symmetry in the z direction, but a complete band gap is not necessary. Because of the translational symmetry, the wave vector k_z is conserved, and it is therefore still useful to have a band gap over some finite range of k_z. But how might such a gap arise in silica holey fibers such as those of the previous sections? And how can we use it to confine light in air?

Origin of the band gap in holey fibers

We begin by considering the periodic cladding by itself, without any core. At any given k_z value, the solutions are the usual Bloch modes, comprising a band structure in a two-dimensional Brillouin zone. We would like to find a range of k_z for which the band structure has a gap between two bands.

Since we have a two-dimensionally periodic structure, our first impulse might be to return to the results of chapter 5. Are the previously discussed two-dimensional gaps of any use here? The answer, unfortunately, is *no*, because the gaps discussed in chapter 5 correspond to $k_z = 0$. In order to be useful in a waveguide, the gaps must extend over a range of nonzero k_z. If the crystal has a *complete* (overlapping TE and TM) gap at $k_z = 0$, then indeed there will be a range of values of $k_z \neq 0$ over which the gap will persist (much like the gap in the section Bragg Fibers, described below). But the silica/air dielectric contrast of 2.1:1 is not sufficient to obtain such a complete two-dimensional gap (at least, not for these simple periodic geometries).[19] The silica/air structure can have a TE gap,

[18] See e.g. Okuno et al. (1999).

[19] With different materials, on the other hand, a complete gap is possible at $k_z = 0$ and thus for a range of nearby k_z; for example, some chalcogenide glasses have indices of 2.7 or higher, which is theoretically sufficient for this purpose.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

but not an overlapping TM gap, and for $k_z \neq 0$ both the TE/TM distinction and the gap disappear.

What other recourse do we have to find a gap in a holey fiber? Since $k_z = 0$ was unhelpful, let us consider $k_z \to \infty$ instead. In this limit, as described in the subsection The scalar limit and LP modes of chapter, the system is again equivalent to a two-dimensional system—one in which the holes are replaced by perfect-metal rods and only an analogue of the TM polarization is present. Such a structure can indeed have a gap between two bands. For example, in figure 5 we saw that a metallic rod radius of $r = 0.3a$ led to a gap between the second and third bands. Moreover, this band gap will appear not only for silica/air structures, but for *any* index contrast with the same geometry, as long as we go to a large enough k_z. In this case, the first two "LP" bands in the scalar limit correspond to *four* vectorial modes, so we expect to see a gap open between the fourth and fifth bands for sufficiently large k_z.

As we will explain below, when guiding in an air core it is important that the gap open up when k_z is not too large, in order for the gap to extend above the light line of air ($\omega = ck_z$). Therefore, we increase the strength of the gap by enlarging the holes to $r = 0.47a$. The resulting projected band diagram is shown in figure 7, where we plot all the modes of this *periodic cladding* (no defect/core) as a function of k_z. This is the *light cone* of the crystal, but unlike the light cone of a uniform medium it has openings *above* its lowermost boundary: the photonic band gaps.[20]

Just as we predicted from the scalar limit, the lowest gap is indeed between the fourth and fifth bands. This can be seen in the exact vectorial band diagram for the $r = 0.47a$ holey fiber at a particular $k_z a/2\pi = 1.7$, shown in figure 8. There is, of course, a gap below the first band, corresponding to the index-guided region below the light cone, and the next gap is after the fourth band. Even the shape of the first four bands is reminiscent of the scalar band diagram from figure 5, where each of the scalar bands has been split into two vectorial bands. Since this gap comes from the scalar limit, it remains open (and indeed, increases monotonically) as k_z is increased. For larger k_z, higher-order gaps from the scalar limit also open.

(Interestingly, there is at least one small gap visible in figure 7 that does *not* correspond directly to a gap in the scalar limit and therefore is harder to anticipate. This gap opens around $k_z a/2\pi = 1.85$ and then closes at a *finite* wave vector around $k_z a/2\pi = 2.5$. In complicated high-contrast structures, gaps that have no simple analytical explanation are not unusual.)

We close this section by pointing out two other interesting gap properties that can be understood from the scalar limit. First, whereas we saw in chapter 5 and in appendix C that two-dimensional band gaps open only for some minimum index contrast (around 1.4:1 for the triangular lattice of circular holes), this is not true for fibers. The band structure approaches a scalar limit of "metallic" rods, with the *same* gaps, for *any* index contrast, no matter how small, although for a small index contrast the gaps may open only for a large k_z far below the air light

[20] Such out-of-plane band gaps were first identified by Birks et al. (1995).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

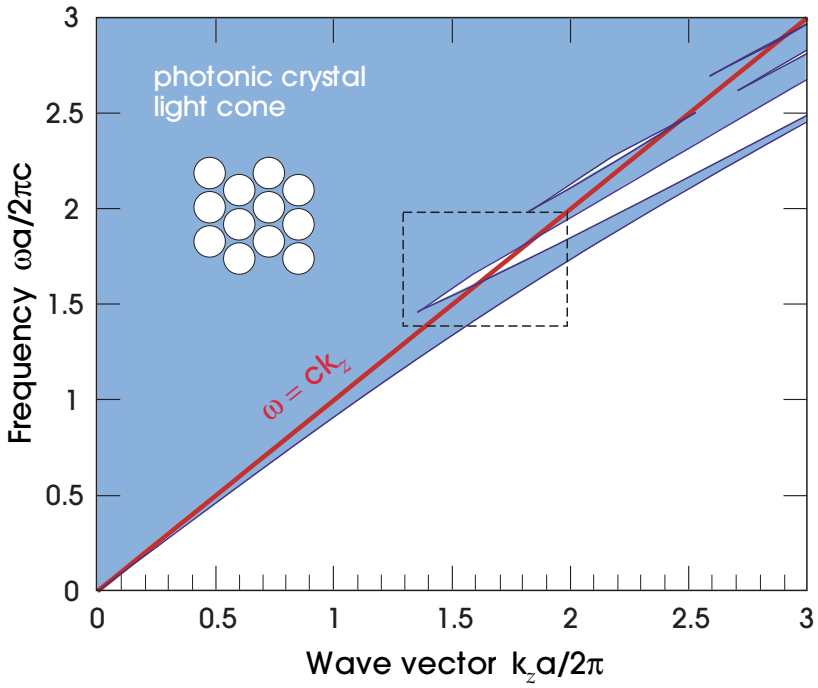


Figure 7: Projected band diagram, as a function of out-of-plane wave vector k_z, for a triangular lattice of air holes (inset: period a, radius $0.47a$) in $\varepsilon = 2.1$. This forms the light cone of the holey fiber from figure 1(b), with gaps appearing as open regions. The light line of air, $\omega = ck_z$, is shown in red. (Dashed box indicates region plotted in figures 10 and 12 for the defect modes.)

line.[21] Second, consider what happens for the inverse case of *higher*-index rods surrounded by a lower-index material. For this case, in the scalar limit one obtains light modes that are 100% confined inside the rods, yielding bands that are independent of the in-plane Bloch wavevector (k_x, k_y). That is, just as we observed in the section Out-of-Plane Propagation of chapter 5, the bandwidths of the lowest photonic bands become very narrow, approaching a discrete set of bands corresponding to the scalar modes in cylindrical metallic cavities. In between these bands are gaps, but the gaps are largely insensitive to the positions of the rods, since in the scalar limit the rods form noninteracting cavities whose frequencies are determined by the rod geometry alone. The localization of modes via this sort of phenomenon has been dubbed "anti-resonant reflecting optical waveguiding," or the *ARROW* model (Litchinitser et al., 2003).

[21] Band-gap guidance has been observed experimentally for index contrasts as low as 1% (Argyros et al., 2005).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

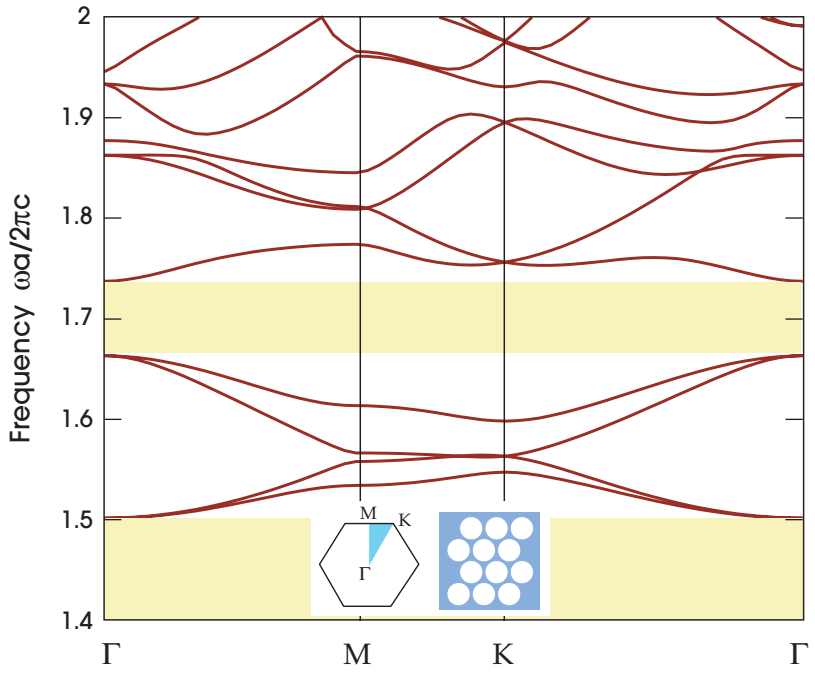


Figure 8: Band diagram versus in-plane wave vector in the irreducible Brillouin zone (inset) for the triangular lattice of air holes from figure 7, at an out-of-plane wave vector $k_z a/2\pi = 1.7$. Gaps are shaded yellow: the lower gap corresponds to the index-guiding region, and the upper gap corresponds to one of the band gaps inside the light cone where guiding in an air core is possible.

Guided modes in a hollow core

By now we are familiar with the proposition that, given a band gap, introducing a defect in the crystal can produce localized states. This phenomenon is exploited to guide light in a hollow-core photonic-crystal fiber.[22] Figure 9 shows the cross section of an experimental holey silica fiber with a hollow core covering the area of seven holes of the periodic structure. Theoretically, we will form a similar cross section by enlarging a single hole to a radius of $1.202a$, and we will focus on the modes within the first gap of figure 7.

The resulting band diagram is shown in figure 10, and exhibits a bewildering variety of guided modes. We can categorize these modes in two ways: by symmetry, and by whether they are surface states or air-core modes. Lines in different colors correspond to different symmetries. The thick red lines are doubly degenerate modes and are the only states that can couple to a source of plane

[22] This was first demonstrated experimentally by Cregan et al. (1999), with notable later improvements reported by Smith et al. (2003) and Mangan et al. (2004).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

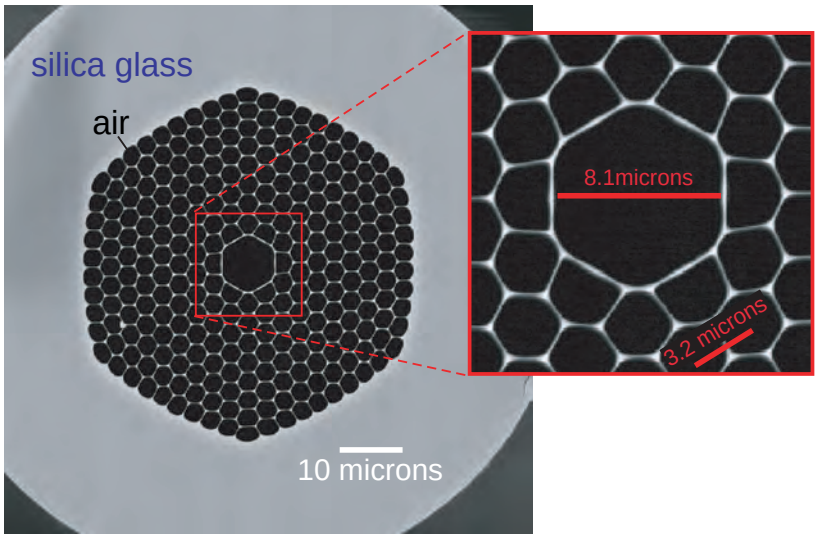


Figure 9: Electron-microscope image of hollow-core holey-fiber cross section (black regions are air holes, and gray regions are silica glass). Central air defect, replacing 7 holes, supports gap-guided modes around a wavelength of 1060 nm. (Image courtesy Karl Koch and Corning, Inc.)

waves incident in the z direction. We will focus on those modes. (The green lines are doubly degenerate of a different symmetry, and the blue lines are nondegenerate modes.[23]) There are three such doubly degenerate bands, and we plot their intensity patterns in figure 11, corresponding to the dots marked in figure 10.

The intensity patterns reveal a striking difference between the two bands that lie *above* the light line of air ($\omega = ck_z$) and the one band that lies *below* the light line of air. The former are concentrated in the air core, while the latter is concentrated around the *surface* of the air core. This is an example of a *surface state*, much like those we have seen in previous chapters: it is evanescent in the crystal because it is in the band gap, and is evanescent in the air core because it is below the air light line.

In fact, we see four surface states of various symmetries below the air light line in figure 10. Why so many? To understand this, let us compare this case with the surface states of a two-dimensional crystal from chapter 5. In two dimensions, we considered only $k_z = 0$, and we found a continuous band of surface states that propagated along a *flat* interface. Here, we have a curved, *finite* interface. Instead of a continuous set of surface states, we have a *discrete set* of surface states at

[23] Technically, by different "symmetries" we mean different irreducible representations of the C_{6v} symmetry group. There are actually four types of nondegenerate representation, but we don't distinguish them here.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

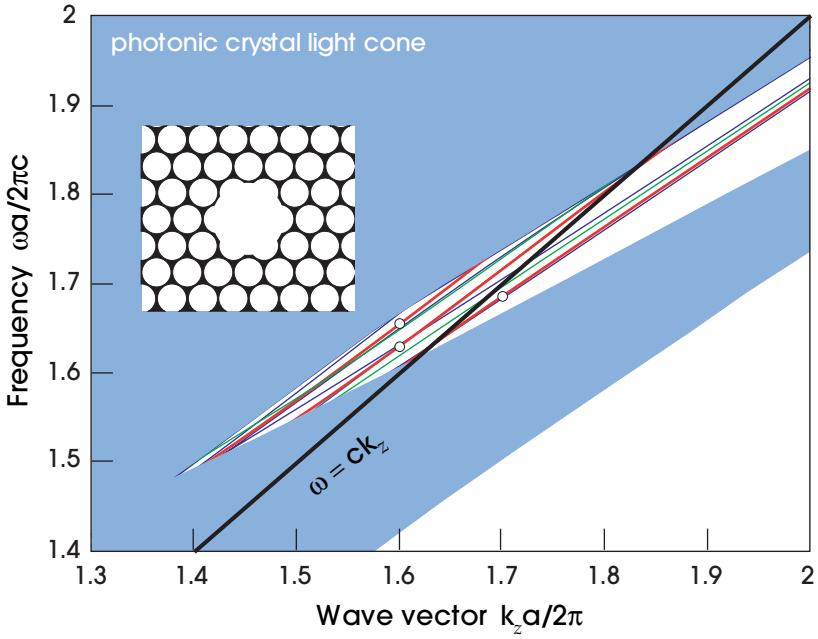


Figure 10: Band diagram showing guided modes of hollow-core holey-fiber structure (inset, similar to experimental structure of figure 9), corresponding to dashed region of figure 7. (Air core is formed by a radius-1.202a air hole.) Three thick red lines indicate doubly degenerate bands that have the correct symmetry to couple to planewave input light. Thin green lines indicate doubly degenerate bands with a different symmetry, and thin blue lines indicate nondegenerate bands. Bands below the light line (thick black) are surface states confined to the edge of the core. Three dots indicate the modes plotted in figure 11.

each k_z (which form continuous bands as we vary k_z). This happens in much the same way that a finite piano string supports only a discrete set of harmonics, and is yet another instance of the general rule we introduced in the section Discrete vs. Continuous Frequency Ranges of chapter. If we were to make the core larger, and the interface longer, then we would get more surface states which are more closely spaced. However, we must also take into account the crucial role of the crystal termination.

As we have seen in previous chapters, the existence of surface states depends on how we terminate the crystal. For example, does the edge of the air core occur at the edges of the holes, or does it cut them in half? It should be possible to improve the performance of the fiber by adjusting the termination to *eliminate* the surface states.[24] Surface modes degrade a fiber's performance primarily because they may have greater losses than the other guided modes. For example, scattering

[24] The importance of crystal termination for air-core fiber performance was analyzed by Saitoh et al. (2004), West et al. (2004), and Kim et al. (2004).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

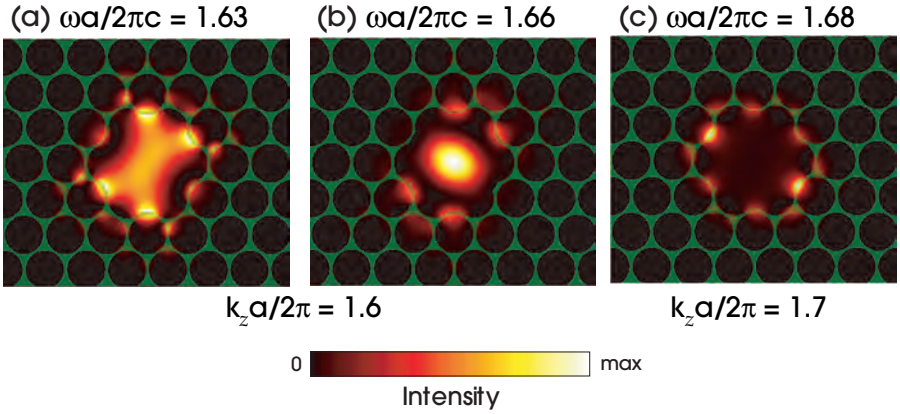


Figure 11: Intensity patterns ($\hat{\mathbf{z}} \cdot \mathrm{Re}[\mathbf{E}^* \times \mathbf{H}]$) of three doubly degenerate modes of a hollow-core holey fiber (ε shaded green), corresponding to the dots on the thick red lines in figure 10. (a) and (b) lie above the air light line at $k_z a/2\pi = 1.6$, while (c) is a surface state lying below the air light line at $k_z a/2\pi = 1.7$.

due to surface roughness is much worse for a mode concentrated at the surface than for a mode concentrated in the core. One could attempt to operate exclusively in the air-core mode of figure 11(a), but this proves difficult in practice. Any small imperfection or asymmetry will tend to couple energy from one mode to another, especially at the points where modes of different symmetry cross in the band diagram.

The effect of an alternate core termination is shown in figure 12, for which we have used a defect hole of an enlarged radius 1.4*a*. Perhaps counterintuitively, enlarging the defect has *reduced* the number of defect modes. This is because the new termination has eliminated the surface states. In fact, there is only one doubly degenerate mode in this fiber core that is of the correct symmetry to couple to planewave input light (thick red line). The other three modes (two nondegenerate and one doubly degenerate) lie almost on top of one another, much like the higher-order modes of figure 2. The intensity pattern of this mode is shown in the inset, and is strongly localized in the air core. Indeed it is even more strongly localized than in the previous structure, a consequence of the lack of surface states to interact with.

Bragg Fibers

Rather than using two-dimensional periodicity to make a band-gap fiber, we can instead use a one-dimensional periodicity, and simply wrap a multilayer film around the core. Such a structure, depicted schematically in figure 1(a), is known as a *Bragg fiber* and was proposed long ago by Yeh et al. (1978), and even earlier by Melekhin and Manenkov (1968). Fink et al. (1999*b*) suggested the use of

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

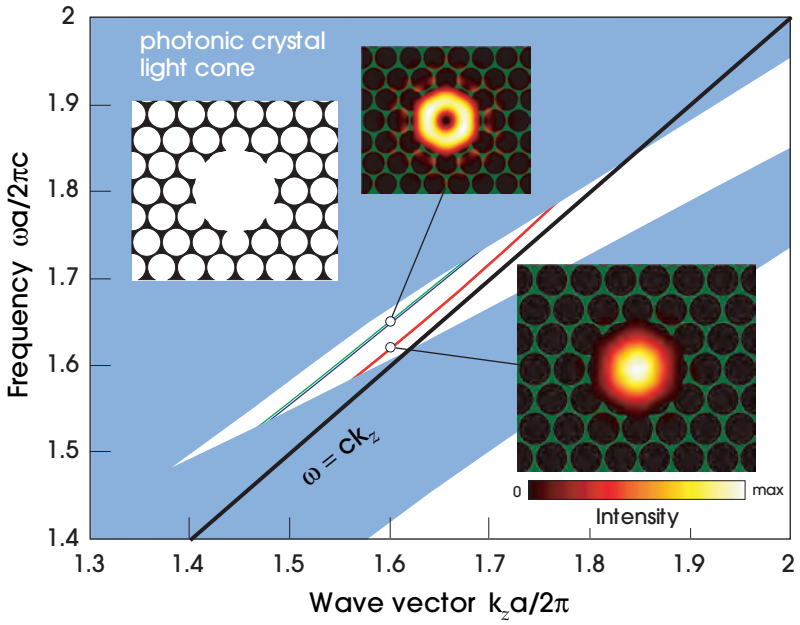


Figure 12: Band diagram and mode intensity patterns (insets) of a hollow-core holey fiber with a larger air core (radius $1.4a$, left inset) corresponding to a crystal termination that eliminates surface states. Thick red line (lower) indicates doubly degenerate mode with correct symmetry to couple to input plane waves. Thin blue and green lines indicate other symmetries (blue is nondegenerate and green is doubly degenerate); there are actually two blue lines that are difficult to distinguish here. Intensity patterns are shown for $k_z a/2\pi = 1.6$, indicated by the dots. (The four higher-order modes all have similar intensity patterns; a nondegenerate mode is shown.)

omnidirectional mirrors for Bragg fibers. As discussed below, the omndirectional regime is correlated with that of the strongest optical confinement. Hollow-core omnidirectional-mirror Bragg fibers have been created in the laboratory, as shown in figure 13, with layers made of a low-index polymer and a high-index chalcogenide glass that have compatible thermal properties for drawing fibers. Such fibers can confine light within a hollow core, much like the holey fibers of the previous section. In fact, they have already been used to guide high-power lasers for endoscopic surgery at wavelengths for which solid materials are too lossy (Torres et al., 2005).

Analysis of cylindrical fibers

The analysis of Bragg fibers is greatly simplified because of their rotational symmetry. As we first learned in the section Continuous Translational Symmetry of chapter 3, the continuous translational symmetry in z means that the z dependence

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

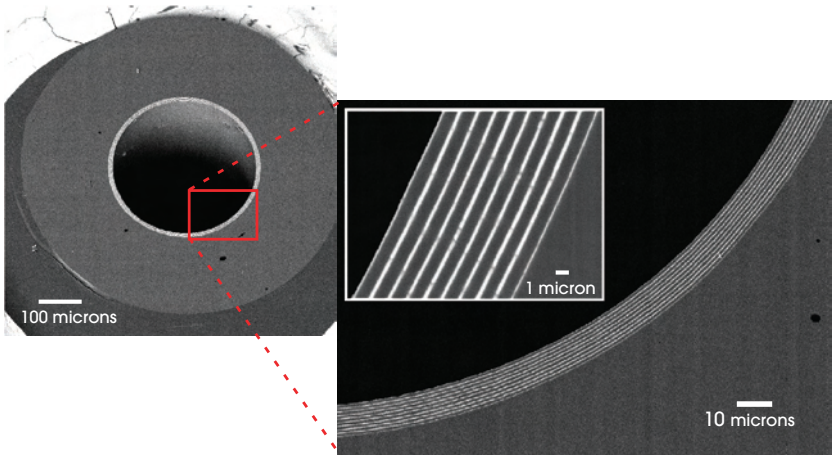


Figure 13: Electron-microscope image of a hollow-core omnidirectional-mirror Bragg-fiber cross section, with insets showing enlarged views of multilayer structures. Thin white layers are a chalcogenide glass and gray regions are a polymer. This fiber was designed to operate at a wavelength of 10.6 μm (Temelkuran et al., 2002).

of the fields can be chosen as $e^{ik_z z}$ for some wave vector k_z. By exactly the same reasoning, the continuous rotational symmetry in φ (the azimuthal angle) means that the φ dependence of the fields can be chosen as $e^{im\varphi}$ for some number m, which is known as the **angular mode number**. It must be an integer because $\varphi = 0$ and $\varphi = 2\pi$ are equivalent, and consequently $e^{im\cdot 2\pi} = 1$.[25] Thus, the field of an eigenstate can be written in the separable form:

$$\mathbf{H}_{k_z,m} = e^{ik_z z + im\varphi}\mathbf{h}_{k_z,m}(r), \tag{7}$$

which has been reduced to a *one*-dimensional problem for the radial (r) dependence $\mathbf{h}_{k_z,m}(r)$.[26]

Equation (7) corresponds to a "circularly" polarized mode, since by including the time dependence $e^{i(m\varphi-\omega t)}$ we see that the field pattern at a fixed z is rotating with an angular phase velocity ω/m for $m \neq 0$. Many authors instead use a φ dependence of $\sin(m\varphi)$ or $\cos(m\varphi)$, forming "linearly" polarized states that are not rotating (although the polarization is not generally uniform over the

[25] m is also called the "angular momentum" in analogy with quantum mechanics, where the angular momentum L_z of such a wave function is $\hbar m$.

[26] The equation for $\mathbf{h}_{k_z,m}(r)$ in a region of uniform refractive index turns out to be analytically solvable in terms of Bessel functions. In a multilayer fiber, the Bessel solutions in each layer are then connected by means of "transfer matrices" that match the boundary conditions at each interface (Yeh et al., 1978; Johnson et al., 2001*b*).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

cross section). That is, the mirror symmetry $\varphi \to -\varphi$ implies that the $\pm m$ eigenmodes are degenerate, and one can therefore form $e^{+im\varphi} \pm e^{-im\varphi}$ combinations to obtain cosines and sines.

Band gaps of Bragg fibers

In order to understand confinement in Bragg fibers, we must first solve for the band gaps. This might seem at first like a difficult task, because concentric rings are not a periodic structure in the sense required by Bloch's theorem. Their curvature decreases with r. However, all that matters for light to be confined is that it cannot escape to $r \to \infty$. In this limit, the curvature of the rings approaches zero, and the structure approaches that of a planar multilayer film. The band diagram of the flat multilayer structure (as discussed in the section Off-Axis Propagation of chapter 4) gives the exact solutions in the $r \to \infty$ limit. If the wave vector k_z and frequency ω of a mode lie in one of the "one-dimensional" gaps, then the mode cannot escape to large radii. It is localized in the fiber core.

Moreover, we can exploit the analytical results for multilayer films given in the section The Size of the Band Gap of chapter 4. A good rule of thumb for the multilayer film was that optimal confinement was obtained for a *quarter-wave stack*. This criterion must be modified here because waveguide modes will not generally be normally incident on the layers. In particular, when we are guiding light in a hollow core, the low-order modes approach glancing incidence as the core size is increased. Equivalently, they approach the air light line $\omega = ck_z$. In this limit, the quarter-wave condition for the thicknesses d_1 and d_2 of the materials with indices n_1 and n_2 becomes $d_1\tilde{n}_1 = d_2\tilde{n}_2$, where $\tilde{n} \triangleq \sqrt{n^2 - (ck_z)^2/\omega^2} \approx \sqrt{n^2 - 1}$.[27] The corresponding quarter-wave frequency is $\omega a/2\pi c = (\tilde{n}_1 + \tilde{n}_2)/4\tilde{n}_1\tilde{n}_2$. Notice that we must have $n_{1,2} > 1$ in order to have $\tilde{n} > 0$ and satisfy the glancing-angle quarter-wave condition with finite-thickness layers. Quarter-wave thicknesses are *not* required for guidance—as we saw in chapter 4, gaps arise for *any* periodic layers—but they help us to optimize confinement for a given index contrast.

Figure 14 shows the projected band diagram, ω versus k_z, for a glancing-angle quarter-wave stack of indices $n_1 = 1.6$ and $n_2 = 2.7$, similar to the polymer and chalcogenide indices of figure 13. We see numerous band gaps that arise, shown as open spaces in the band diagram. This structure is actually an *omnidirectional* reflector as described in the section Omnidirectional Multilayer Mirrors of chapter 4, with the omnidirectional gaps highlighted in yellow. The omnidirectional property is *not* required for guidance in such fibers—any gap at a mode's (k_z, ω) will do—but its appearance is not a coincidence. For strong confinement, we will generally want large index contrast, and we also want $n_{1,2} > 1$ in order to satisfy the glancing-angle quarter-wave condition. These two criteria are precisely

[27] This formula arises from the fact that the radial wave vector k_r in n is $k_r = \sqrt{n^2(\omega/c)^2 - k_z^2}$, and the quarter-wave condition is $k_r d = \pi/2$.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

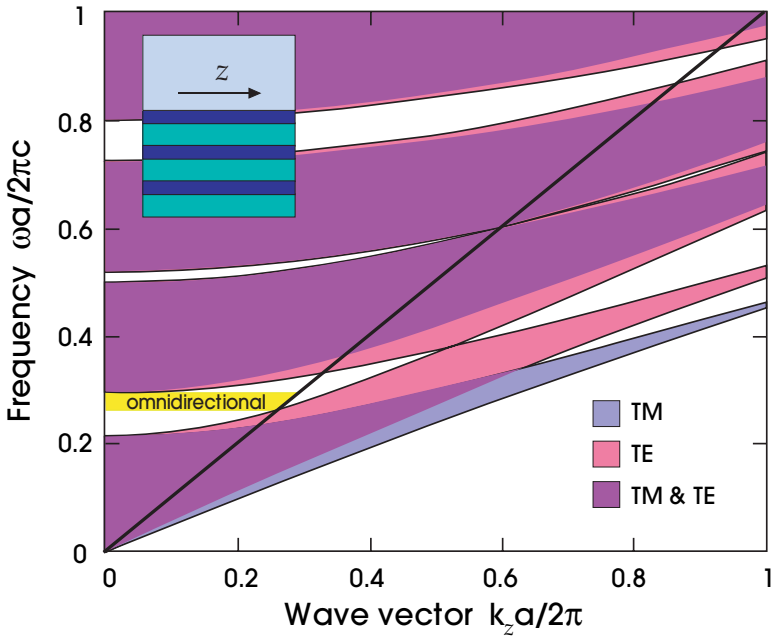


Figure 14: Projected band diagram for glancing-angle quarter-wave stack of indices $n_1 = 1.6$ and $n_2 = 2.7$, plotted versus wave vector k_z parallel to layers (inset). Regions where propagating modes exist in the stack are shaded: blue for TM (**E** out of the incidence plane) and red for TE (**E** in the incidence plane), or purple for both. The frequency range of omnidirectional reflection from an air medium (corresponding to the black light line $\omega = ck_z$) is shaded yellow. (Note that this "TM/TE" terminology is the *reverse* of the "te/tm" terminology used later in the chapter.)

the ingredients for omnidirectional reflection.[28] Additionally, an omnidirectional gap has potential advantages when one wants to trap light incident from many different directions.[29]

It may seem that we are forgetting something important by considering only modes propagating in the rz plane. In fact, if we include modes propagating *out* of this plane, corresponding to light travelling in the *azimuthal* direction, we would find no band gaps at all. Fortunately, such modes are excluded: for any finite angular mode number m, as $r \to \infty$ the corresponding azimuthal wave vector $k_\varphi = m/r \to 0$. Put another way, the conservation of angular momentum prevents a mode in the waveguide core from escaping into a mode travelling in the azimuthal direction infinitely far away. This conservation law makes gaps,

[28] For the less-favorable tm polarization, as defined later in this chapter, it has been argued that a low-contrast, non-omnidirectional regime also works well (Skorobogatiy, 2005).

[29] For example, omnidirectional gaps are beneficial for collecting and guiding light emitted by fluorescent sources within the hollow core (Bermel et al., 2004).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

especially gaps at k_z near 0, much easier to create in Bragg fibers than in fibers with two-dimensional periodicity such as the holey fibers of the previous section. On the other hand, holey fibers can be constructed from a single solid material (e.g., silica), whereas Bragg fibers require two different solid materials to be found that can be drawn together.

Guided modes of Bragg fibers

Given the multilayer mirror with the parameters of the previous section, we form a hollow-core fiber by wrapping the mirror around an air core. This involves two important decisions: what core radius R do we employ, and how do we terminate the crystal? The choice of R generally involves a tradeoff between different loss mechanisms, which we will discuss in the section Losses in Hollow-Core Fibers. For now, we will simply choose $R = 3a$. As for the termination, we will end the crystal at R with *half* of a *high*-index layer: using a high-index layer on the inner surface confines modes in the core more strongly, and using half of a layer eliminates a troublesome surface state.

The resulting band diagram, focusing on the first band gap above the air light line, is shown in figure 15(a). As might be expected from the fact that the core is several wavelengths ($6a$) in diameter, a number of modes are supported. It turns out, however, that these modes can be directly related to the modes of a much simpler waveguide: a hollow perfect-metal cylinder. The band diagram of a hollow metallic waveguide with the same $R = 3a$ is shown in figure 15(b), with the Bragg-mirror gaps superimposed for comparison. Comparing the two band diagrams, we see that there is a one-to-one correspondence between the modes. The modes of the Bragg fiber are essentially given by the metallic modes that fall inside the band gap.

A hollow metallic waveguide is also a mirror wrapped around a hollow core. With this in mind, it is perhaps not surprising that its mode structure is similar to that of a Bragg fiber, at least for large R or in the omnidirectional regime where the multilayer mirror acts much like a metal.[30] Both structures have modes described by an angular index m, but a metallic waveguide (with a homogeneous interior) has an additional special property: all of the modes can be divided into two polarization states. Unfortunately, the terms "TE" and "TM" in the fiber literature have a somewhat different meaning than our usage in the previous chapters, and so we use lower-case letters to distinguish the meanings here: **te** means modes with electric fields purely in the xy plane but the magnetic field may be in any direction, while **tm** means modes with magnetic fields purely in the xy plane but electric fields in any direction. The modes of the cylindrical metallic waveguide are then labelled as $\mathrm{te}_{m\ell}$ and $\mathrm{tm}_{m\ell}$ where m is the angular index and $\ell = 1, 2, \ldots$ is a radial order (a number of field nodes in $r > 0$). From figure 15(b), we see

[30] The analogy between Bragg fibers and hollow metal waveguides is analyzed in more detail by Ibanescu et al. (2003).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

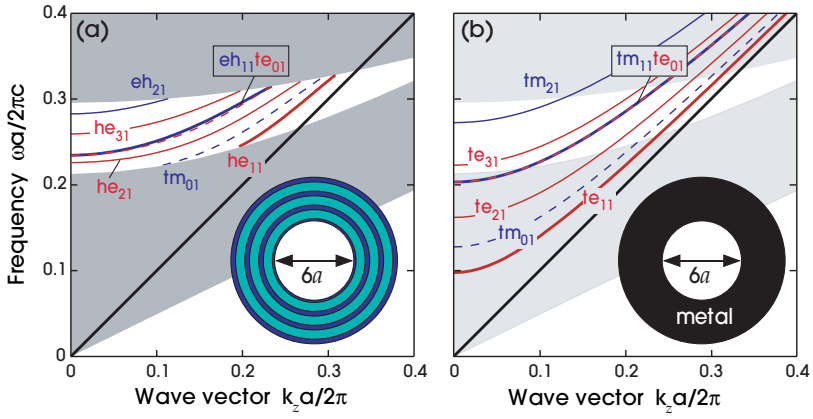


Figure 15: (a) Band diagram of hollow-core ($R = 3a$) Bragg fiber with layer indices of 2.7/1.6 (blue/green layers in inset); gray regions indicate extended modes that propagate in the multilayer mirrors. (b) Band diagram showing the first few bands of a hollow perfect-metal waveguide of the same radius; light gray regions indicate extended modes of Bragg mirrors from (a), for comparison. Modes are labelled by polarization and radial/angular mode numbers as described in the text.

that the lowest-ω metallic mode is te_{11}, followed by tm_{01}, te_{21}, te_{01}, tm_{11}, te_{31}, and so on. Of course, any $m \neq 0$ mode is doubly degenerate with the $-m$ mode. Also, te_{01} and tm_{11} happen to be exactly degenerate.

The modes of the Bragg fiber are analogous, but there are a few important differences. Most obviously, the frequencies are slightly shifted compared to the metallic structure, and as a side effect the modes analogous to te_{01} and tm_{11} are no longer exactly degenerate.

Of greater importance, however, is the fact that only the $m = 0$ modes remain purely polarized. As a consequence of mirror symmetry in φ, the $\mathrm{te}_{0\ell}$ modes have their electric fields purely in the φ direction and the $\mathrm{tm}_{0\ell}$ have electric fields purely in the rz plane. All of the $m \neq 0$ modes, however, are changed from te and tm into "hybrid" polarizations labelled **he** and **eh** depending on whether they are mostly te-like or mostly tm-like, respectively. So, for example, the te_{11} mode of the metal waveguide becomes the he_{11} mode of the Bragg fiber, and so on. It turns out that the most important two modes for many practical applications are he_{11} and te_{01}: te_{01} is usually the *lowest-loss* mode (in both the metallic waveguide and Bragg fibers) for reasons discussed below, while he_{11} is usually the lowest loss $m = 1$ mode (the only m that can directly couple to planewave input light). The field and intensity patterns of these two modes are depicted in figure 16; they closely resemble the corresponding metallic-waveguide modes.

You may recall from chapter 4 that we can classify the modes propagating in a multilayer structure by whether **E** is parallel (TE) or perpendicular (TM) to the plane of propagation. Here, we can do the same thing for $m = 0$ because in this

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

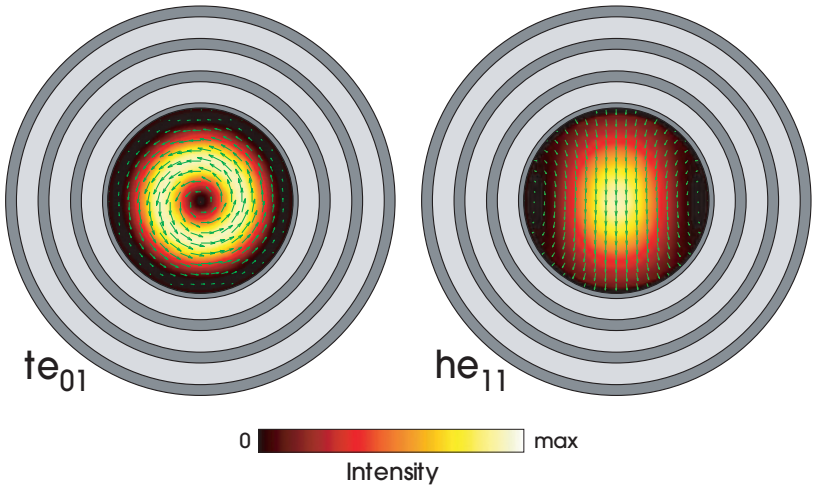


Figure 16: Intensity patterns for two core-guided modes in hollow-core Bragg fiber of figure 15 at the quarter-wave frequency. Power intensity is shown by color scale, and transverse (xy) electric-field pattern is shown by green arrows.

case rz is a mirror plane for every φ, but the labelling convention is *reversed*. The plane of propagation, far away, is the rz plane: $\mathbf{E}$ parallel to the rz plane is what we are now calling $\mathrm{tm}_{0\ell}$, and $\mathbf{E}$ perpendicular to rz is what we are now calling $\mathrm{te}_{0\ell}$. In chapter 4 we saw that the te (previously TM or s-polarized) band gap is usually larger than the tm (previously TE or p-polarized) gap, because the latter closes entirely at Brewster's angle. As a consequence, every mode except for the $\mathrm{te}_{0\ell}$ modes is generally limited by the smaller tm gap, which is what is shown in figure 15.

The $\mathrm{te}_{0\ell}$ modes, on the other hand, can have both larger bandwidth and stronger confinement because they see only the larger te gap. Partly as a consequence of this, as we discuss in the next section, the te_{01} mode has the lowest predicted losses for sufficiently large R. An analogous result is well known for the cylindrical metallic waveguide, albeit for a somewhat different reason. For a perfect metal, the boundary condition at $r = R$ is that E_z and E_φ must be zero while E_r can be nonzero, and thus the $\mathrm{te}_{0\ell}$ modes (for which $E_r = 0$ everywhere) are the *only* modes that have $\mathbf{E} = 0$ at $r = R$. For an *imperfect* metal, the field penetrates into the region $r > R$ and causes ohmic losses. The te_{01} mode will have the smallest penetration into the metal (since its electric fields are nearly zero at the interface) and thus the lowest ohmic losses. In fact, a similar field node at $r \approx R$ is apparent in the Bragg-fiber te_{01} mode of figure 16, due to the system's close resemblance to the metallic waveguide.

Losses in Hollow-Core Fibers

Optical fibers are used to transport light over distances ranging from meters to thousands of kilometers. Over such distances, even small imperfections can lead

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

to substantial effects. Conventional silica fibers have attained such an amazing degree of perfection that their losses (only about 0.2 dB/km at 1.55 μm) are limited by a combination of intrinsic material absorption and scattering from microscopic density fluctuations. At longer wavelengths, on the other hand, such as the 10.6 μm high-power lasers used for many industrial and medical applications, silica and other common fiber materials are not transparent at all.

One of the promises of hollow-core photonic-bandgap fibers is that they may allow for lower losses than are possible with solid-core fibers, by relaxing the fundamental limitations imposed by solid material properties. In order to design fibers for such a goal, however, one must comprehend the different loss mechanisms in realistic fibers. Although a detailed study of these losses is outside the scope of this book, a broad understanding may be gained by dividing fiber losses into two categories depending on how they scale with the radius R of the air core. There is a tradeoff between losses that *decrease* with R (losses associated with field penetration into the cladding), and losses that *increase* with R (losses associated with coupling between modes). This makes the choice of R a delicate balancing act that is crucial for fiber performance.

Interestingly, not all losses are bad. As we have seen, most of the proposed hollow-fiber designs have been multi-mode. They support multiple guided modes that propagate at different speeds. Unchecked, this results in **modal dispersion**: since it is impossible to avoid exciting multiple modes, the differing velocities cause pulses to spread and information transmission to be scrambled. However, this problem is reduced in a hollow-core fiber by *differential attenuation*: some modes (typically the lower-order modes) have much lower losses than others, and thus transmission in everything but the lowest-loss mode will be filtered out after propagation over a long distance.[31]

Cladding losses

Three important loss mechanisms are associated with the amount of field penetration into the cladding: material absorption, radiative leakage due to the finite crystal size, and scattering from disorder. All of these will tend to decrease as the core radius R increases. We will show that they typically decrease asymptotically as $1/R^3$.

Of these three loss mechanisms, the simplest one to analyze is material absorption. This can be described by a small imaginary part $i\kappa$ that is added to the real refractive index n (κ is called the **extinction coefficient**). Because $\kappa \ll n$ for transparent materials, one can obtain essentially exact results for the loss by starting with the eigenmode of the lossless structure and employing perturbation theory, as in the section The Effect of Small Perturbations of chapter 2. Equation (29) of chapter 2 tells us the imaginary change $\Delta\omega$ in the frequency due to κ, and to obtain the loss rate per unit distance we compute

[31] This filtering is discussed in more detail for Bragg fibers by Johnson et al. (2001*b*).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

$\Delta k_z = -\Delta\omega/v_g$ via the group velocity $v_g = d\omega/dk_z$. Letting $\alpha \triangleq 2\,\mathrm{Im}\,\Delta k_z$, this describes a decay $e^{-\alpha z/2}$ in the fields and $e^{-\alpha z}$ in the intensity. Combining these equations, the decay rate α due to a single absorbing material with a complex refractive index $n + i\kappa$ is

$$\alpha = \frac{2\omega\kappa}{v_g n} \cdot (\text{fraction of } \int \varepsilon|\mathbf{E}|^2 \text{ in absorbing material }). \tag{8}$$

(For the case of multiple materials, one simply adds the α from each material.) As a special case, if the field energy propagates entirely within the material with a group velocity $v_g = c/n$ as for a plane wave (neglecting material dispersion), then one obtains a **bulk absorption** loss $\alpha_0 \triangleq 2\omega\kappa/c = 4\pi\kappa/\lambda$. Therefore, a useful dimensionless figure of merit for a hollow-core fiber mode is the ratio α/α_0. This is called the *absorption suppression* factor, the factor by which loss is decreased due to the portion of light in air.

For example, let us consider the hollow-core Bragg fiber from the previous section, and suppose that the *low*-index material ($n_1 = 1.6$) has some absorption. This is motivated by the experimental fiber from figure 13, which is designed to operate at 10.6 μm. Its absorption loss is dominated by the effect of the low-index polymer, which has a bulk absorption of around 50,000 dB/m ($\kappa = 0.01$). In figure 17, we plot the absorption suppression factor α/α_0 of four modes for the core radius $R = 3a$ of figure 15a. Even for this small radius, we see that absorption losses can be suppressed by more than a factor of 10. (Notice also that the absorption losses *diverge* as the zero–group-velocity band-edge of the te_{01} or eh_{11} mode is approached.) This radius, however, is much smaller than that of the experimental structure, which has a dramatic effect on the loss. In figure 18, we plot the absorption suppression factors versus R at the quarter-wave $\omega a/2\pi c \approx 0.30$. Regardless of polarization, the mode absorptions all approach a $1/R^3$ dependence, with the te_{01} mode having the lowest asymptotic loss. At the experimental radius $R \approx 80a$, the he_{11} suppression factor is almost 10^{-5}, and indeed losses less than 1 dB/m were observed experimentally, representing a suppression of the polymer absorption by over four orders of magnitude.

What is the source of this $1/R^3$ power law? The key fact from equation (8) is that the contribution to the loss from a particular absorbing material is proportional to the fraction of the electric-field energy in the material. One might jump to the conclusion that, for a core-guided mode (not a surface state), the losses will scale as $1/R$: if the field penetrates a certain distance d_p into the cladding, then the fraction of field in the cladding goes as the penetration area $2\pi R d_p$ divided by the core area πR^2, yielding $\sim 1/R$. Such an argument, however, assumes that the field *amplitude* $|\mathbf{E}|$ in the cladding compared to the core is independent of R, and in fact this is not the case. This is easiest to see for the te_{01} mode, which by analogy with the metal waveguide has a node in $\mathbf{E}$ near $r = R$ (as in figure 16). As a consequence, the te_{01}'s cladding $|\mathbf{E}|$ is proportional not to the field at $r = R$ (which is ≈ 0) but rather to the *slope* $d|\mathbf{E}|/dr$ at $r = R$, which scales as $1/R$ for a fixed max $|\mathbf{E}|$ in the core. As

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

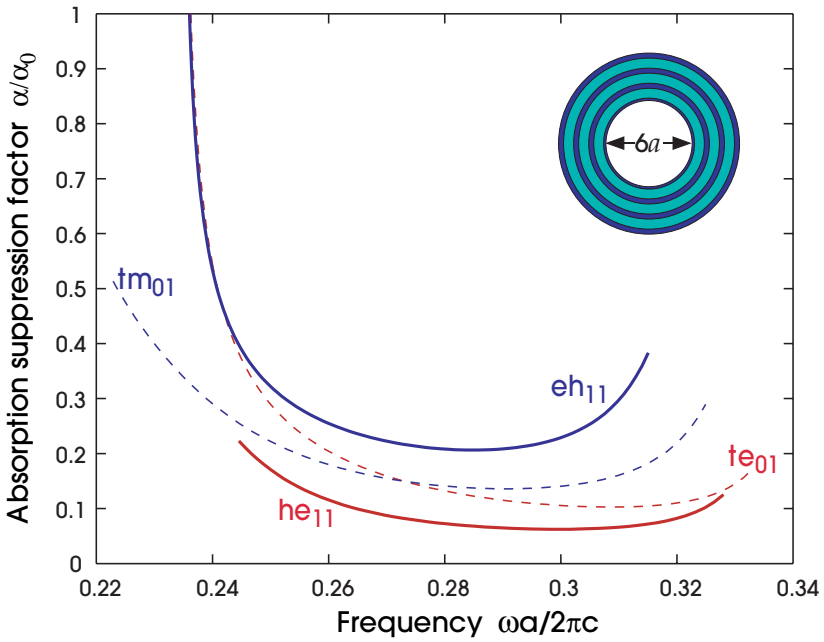


Figure 17: Absorption suppression factor α/α_0 of $R = 3a$ Bragg fiber from figure 15(a) versus frequency for four modes. α/α_0 is the radio of the modes absorption losses to the absorption losses for the bulk material, assuming that the absorption losses are dominated by those of the *low*-index material ($n = 1.6$, shaded green in inset). Here, the he_{11} (fundamental) mode is the lowest loss, while for larger R the te_{01} mode becomes the lowest.

a result, the te_{01}'s $|\mathbf{E}|^2$ in the cladding picks up an additional $1/R^2$ factor, and the absorption losses scale as $1/R^3$.

In fact, a similar argument holds for all core modes, because of the scalar limit described in the subsection The scalar limit and LP modes. For any given mode, in the limit of large R the mode becomes more and more similar to a plane wave propagating along the z axis. Its dispersion relation approaches the air light line, and its penetration depth into the cladding becomes negligible compared to the scale of the transverse oscillations. These were precisely the conditions in which the scalar limit applies. In this limit, we can describe the mode as a linear polarization multiplied by a scalar amplitude $\psi(x, y)$ that is zero in the cladding. In reality, there is some small nonzero amplitude in the cladding, but because of the approximate zero boundary condition at $r = R$, the amplitude of the cladding field goes as $1/R$ just as we explained for te_{01}. Thus, all modes approach a $1/R^3$ scaling.

Another cladding-related loss is radiative leakage. Because a real photonic-crystal fiber cannot have an infinite number of crystal periods, the fields will have a small exponential tail beyond the edge of the crystal, which will couple

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

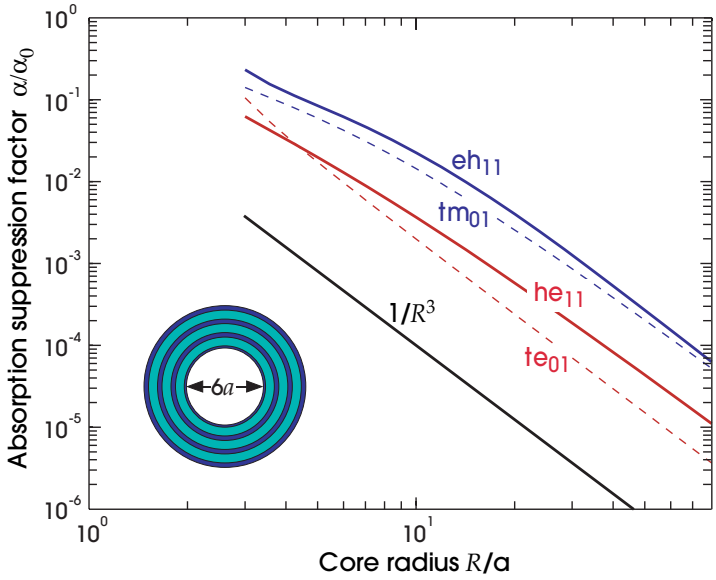


Figure 18: Scaling of the absorption suppression factor α/α_0 versus core radius R, at quarter-wave frequency, for several modes of the hollow-core Bragg fiber with indices 2.7/1.6 (blue/green in inset). The absorption loss is assumed to be dominated by that of the low-index material. All modes asymptote to a $1/R^3$ scaling (shown as a black line, for reference).

to radiating modes. Again, this loss scales as $1/R^3$: $1/R$ from the surface-area/volume ratio, and $1/R^2$ from the field amplitude scaling in the scalar limit. As a practical matter, however, such radiation can be more easily reduced by simply increasing the number of periods. For a high-contrast band-gap fiber such as the ones discussed in this chapter, the radiative leakage typically decreases by a factor of ten for every period or two that is added to the cladding. As a result, even the most stringent loss requirements can be met by including a few dozen periods at most.

Finally, one can also have losses from disorder, which causes light to scatter and radiate by breaking translational symmetry. Because photonic-crystal fibers typically have many high-contrast interfaces, the most serious problem seems to be due to surface roughness, especially in silica-based fibers at wavelengths where the absorption is small. A detailed analysis of disorder is complicated,[32] but a few general statements can be made for the usual case in which the length scale of the roughness is much smaller than the wavelength. In this case, known as **Rayleigh scattering**, the scattered (lost) power is roughly proportional to

[32] See, for example, Johnson et al. (2005) for some of the difficulties that arise in modelling surface perturbations for high-contrast materials. Roberts et al. (2005) analyze roughness due to inevitable surface-tension effects in hollow-core fibers.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

the $|\mathbf{E}|^2$ at the scattering location and to the square of the scatterer's volume.[33] The $|\mathbf{E}|^2$ dependence produces the same $1/R^3$ scaling for disorder-induced loss as for absorption loss, since disorder again affects only the small fraction of the field inside the cladding where the interfaces or materials are.

What about two-dimensionally periodic photonic-crystal fibers, such as the hollow-core holey structure in figure 1(b)? Overall, the same asymptotic $1/R^3$ scaling should apply: the core interface/area ratio goes as $1/R$ and there is an additional $1/R^2$ factor from the cladding field amplitude in the scalar limit. However, an additional wrinkle is provided by the proliferation of surface states. Unless a crystal termination is chosen that eliminates surface states as in figure 12, as the core size is increased we will get more and more surface states. These surface states cross the guided band and chop up its usable bandwidth (West et al., 2004). Precisely such a phenomenon was observed experimentally when Mangan et al. (2004) replaced the air core of figure 9 with one of about 2.2 times the diameter: the losses were reduced by a factor of eight (from 13 dB/km to 1.6 dB/km), but the bandwidth was reduced by a factor of five because the surface states were not eliminated. (The surface states below the light line do *not* have absorption/leakage/scattering losses that decrease with R, because they remain localized at the cladding surface regardless of R.)

Inter-modal coupling

Given the $1/R^3$ dependencies of the loss mechanisms described above, it may seem that increasing the core radius R is always a winning strategy. This is not the case. As R grows, we worsen losses and other problems due to inter-modal coupling: the transfer of energy from one mode to another at the same ω but different k_z. This is caused by fiber nonuniformities that break the translational symmetry in z. This is a problem because higher-order modes will typically have higher losses (e.g., stronger penetration into the cladding), and will also produce modal dispersion. As we mentioned above, the differential losses of higher-order modes will suppress modal dispersion, but not if we couple into them faster than they are filtered out. Inter-modal coupling tends to worsen with increasing core radius R, for two reasons.

First, the number of core-guided modes increases. (The number of modes scales with the area $\sim R^2$.) Correspondingly, the mode spacing Δk_z decreases. This makes it easier for a nonuniformity to couple different modes. Roughly speaking, $\pi/\Delta k_z$ is a minimum length scale for nonuniformities: if the fiber changes shape (e.g. due

[33] The scattered power also scales as ω^4 in homogeneous nondispersive media (Jackson, 1998), but this is modified here by the photonic crystal's effect on the local density of states and on the mode's variation with frequency. The ω^4 dependence of Rayleigh scattering is famous as the reason why the sky is blue: air molecules scatter short (blue) wavelengths more strongly than long (red) ones.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

to ellipticity or stresses) over distances of this length scale or shorter, then coupling between modes may be substantial.[34]

Second, even for a fixed Δk_z between modes, the inter-modal coupling due to fiber *bends* will worsen as R increases. Intuitively, as R grows, the difference in path length between the part of the core on the inside of the bend and the part on the outside will also grow. The result is a bigger "centrifugal force" that distorts the modes. The mathematical treatment of bending is rather complicated, and involves a coordinate transformation of the bent waveguide to a straight waveguide.[35] However, the final result is elegant, and we summarize it here. If we take x to be the direction away from the center of the bend and $x = 0$ to be the center of the fiber at the bend radius R_b, the bend effectively adds a perturbation $\Delta\varepsilon$ and $\Delta\mu$ proportional to x/R_b. This acts like a potential ramp, analogous to a Stark effect in quantum mechanics, that pushes the fields towards the outside of the bend. Fiber performance then deteriorates for two reasons. First, far enough away from the core, the exponential tails of the guided modes will "see" a perturbation so large that the band gap is shifted to a different frequency. This results in radiation losses that increase exponentially with $1/R_b$. Second, for sufficiently large R_b, the bend radiation is negligible, and instead the losses in high-contrast photonic-crystal fibers are dominated by coupling between guided modes. This coupling varies as $(R/R_b)^2$ for large R_b (since R is the maximum value of x in the core), not including changes in Δk_z.

Finally, another important form of inter-modal coupling is known as **polarization-mode dispersion** (PMD). PMD arises in an ordinary fiber because the operating mode is doubly degenerate, with two orthogonal polarizations. Any imperfection or stress in the fiber, however, can break the symmetry and split these two polarizations into modes that travel at different speeds. This produces a form of modal dispersion, where pulses spread due to random imperfections in the fiber. The same thing can happen in a photonic-crystal fiber *if* we operate in a doubly degenerate mode.[36] The differential losses of a hollow-core fiber, however, allow one to operate in a low-loss higher-order mode like te_{01} that is *non*-degenerate and hence immune to PMD (no perturbation can split the mode into two). Alternatively, one can design a core that is so asymmetric that it supports only a single nondegenerate mode (Kubota et al., 2004; Li et al., 2005).

[34] The $\pi/\Delta k_z$ length scale can be justified via the coupled-mode/coupled-power theory of Marcuse (1991), as in e.g. Johnson et al. (2001*b*).

[35] Bending losses were derived by Marcatilli (1969) for dielectric waveguides. For further discussion, see Katsenelenbaum et al. (1998) for metallic waveguides and Johnson et al. (2001*b*) for Bragg fibers. A general result expressing coordinate transformations as simple change in ε and μ was given by Ward and Pendry (1996).

[36] In fact, PMD effects may be worse in hollow-core fibers than in conventional silica fibers because of the increased index contrast, or equivalently because of the large group-velocity dispersion (Skorobogatiy et al., 2002).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Further Reading

Russell (2003) offers a review article on photonic-crystal fibers, and a broad overview of unusual fiber technologies can be found in Mendez and Morse (2006). The literature on Bragg fibers, and the analysis of their losses, was reviewed by Johnson et al. (2001*b*). Two textbooks focusing especially on photonic-crystal fibers (mainly the holey-fiber varieties) are Bjarklev et al. (2003) and Zolla et al. (2005). A useful general textbook on conventional fiber optics is Ramaswami and Sivarajan (1998), and nonlinear fiber phenomena are covered by Agrawal (2001).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

10

Designing Photonic Crystals for Applications

In the first three chapters, we assembled a number of theoretical tools to help us understand the properties of photonic crystals. The next six chapters focused on the questions: *Which* structures have interesting properties, and *why?* This chapter will be concerned with a different question altogether: How can we *design* a device made from photonic-crystal components?

Overview

We have expended a great deal of effort to understand the different ways in which photonic crystals can reflect and trap light, thereby forming mirrors, waveguides, and resonant cavities. These three components are themselves very useful, especially because they can have unusual properties that are not shared by their predecessors made from unstructured materials. Now, however, we will examine some useful ways in which these components can be *combined*. We will see that there are simple universal behaviors that result from such combinations, regardless of the specific geometric structure, which are captured by the formalism of **temporal coupled-mode theory**. This allows us to design devices easily from first principles, and only afterwards determine the quantitative details from a small number of variables: the symmetries, frequencies, and decay rates of the resonant cavities. We will provide examples of **filters**, which only transmit light within a specified frequency band; **bends**, which guide light around a sharp corner; and **splitters**, which divide a waveguide into two. Finally, we will consider further the applications of nonlinear materials (a topic we touched upon in the chapter 9). With a suitable nonlinear material, the photonic-crystal filter can act as an optical "transistor."

For simplicity, most of our examples will be drawn from two-dimensional systems. The ideas generalize easily to the cases of one- and three-dimensional

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

crystals. However, it is useful to consider three-dimensional structures when dealing with the impact of losses on device performance; for this purpose we will consider hybrid structures such as those of chapter 7, where radiative losses inevitably arise for resonant cavities.

Finally, although most of this chapter and this book are concentrated on structures in which light is confined in waveguides and cavities, we will close with a brief review of phenomena associated with free propagation of light in the crystal: reflection, refraction, and diffraction.

A Mirror, a Waveguide, and a Cavity

Long ago, engineers solved the problem of controlling light propagation in the microwave regime[1] by using metallic components to guide, reflect, and trap light. These components rely on the high conductivity of metals, a rather complicated electronic property that may depend strongly on frequency. Unfortunately, for light of higher frequency (like visible light, for instance), metallic components suffer from high dissipative losses. In contrast, the dielectric materials of photonic crystals can have a much simpler behavior, with weak frequency dependence. The only demand we make on our dielectric materials is that, for the frequency range of interest (which is often a narrow band), they should be essentially lossless. Such materials are widely available all the way from the ultraviolet regime to the microwave. We saw in chapter 2 that the photonic properties scale easily with frequency and ε, so devices made at one scale are sure to work at other scales, if appropriate materials can be found.

Designing a mirror

Since the heart of so many devices is reflectivity, our first task is to design a two-dimensional crystal that reflects all in-plane TM waves within some specified frequency band, without appreciable absorption. Once finished, we could use this crystal in a band-stop filter. Or, since the band structures of two-dimensional photonic crystals are different for TE and TM light, we could employ it as a polarizer. Or, as we will see later in this chapter, we could use it as a building block in a waveguide, resonator, or other device. For concreteness, we choose a particular wavelength of light in vacuum: $\lambda = 1.5\,\mu$m, the wavelength that is most commonly used in telecommunications. Our choice of material will be gallium arsenide (GaAs), a material widely used in optoelectronics, for which the dielectric constant is 11.4 at $\lambda = 1.5\,\mu$m.[2]

[1] The "microwave regime" includes light with wavelengths in the range from about one millimeter to about ten centimeters.

[2] See, e.g., Palik (1998). The dielectric constant of GaAs ranges from 10.7 at 10 μm to 12.3 at 1 μm.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

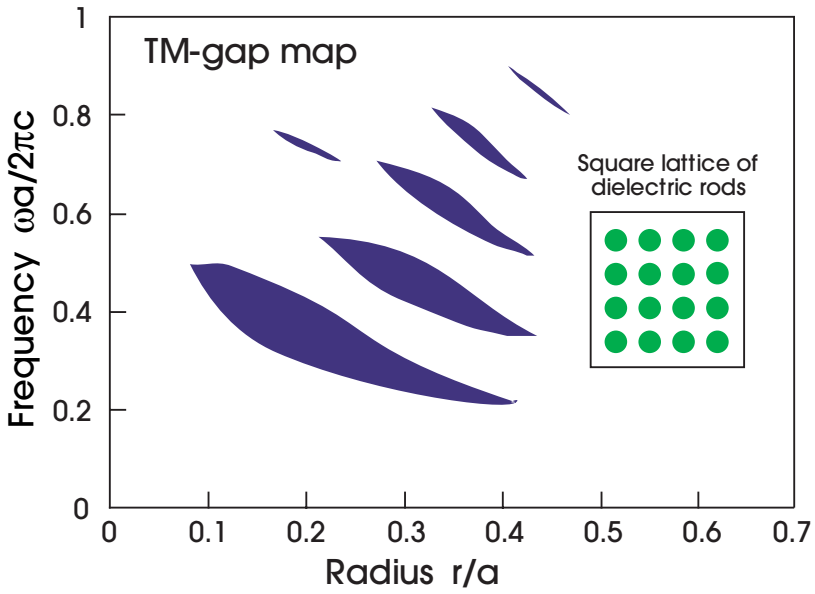


Figure 1: "Gap map" showing TM gap locations versus rod radius for a square lattice of dielectric rods ($\varepsilon = 11.4$) in air (inset).

For maximum reflectivity, we should choose a crystal geometry that has a photonic band gap, and that is not too difficult to fabricate (although in this conceptual two-dimensional example, we are not concerned with the details of fabrication). After consulting an atlas of gap maps, such as the abbreviated one provided in appendix C, we notice a particularly simple geometry with those characteristics: the square lattice of dielectric rods, which has large TM band gaps. The gap map is reproduced in figure 1, and exhibits a large TM gap for a rod radius of $r = 0.2a$, where a is the lattice constant. The band diagram is similar to figure 2 of chapter 5. There is a gap between the first and second TM bands, from $\omega a/2\pi c = 0.287$ to $\omega a/2\pi c = 0.422$, as shown in figure 1. Expressed as a fraction of the midgap frequency $\omega a/2\pi c = 0.355$, the band-gap width is 38%. As we described in the section The Size of the Band Gap of chapter 4, converting these dimensionless quantities into physical units is a simple matter. If we want our operating wavelength $\lambda = 2\pi c/\omega = 1.5\,\mu$m to lie at the middle of the gap, we need

$$\frac{\omega a}{2\pi c} = \frac{a}{\lambda} = \frac{a}{1.5\ \mu\text{m}} = 0.355 \implies a = 0.533\ \mu\text{m}. \tag{1}$$

Given a, we can calculate the rod radius $r = 0.2a = 0.107\,\mu$m, and the structure is completely determined. The TM gap ranges from a wavelength of

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

$a/0.422 = 1.26\,\mu$m to $a/0.287 = 1.86\,\mu$m, which is much wider than the entire bandwidth of a typical optical communication channel.[3]

Designing a waveguide

A waveguide is generally intended to transport waves of a particular frequency from place to place along a one-dimensional (possibly curved) path. At microwave frequencies, this can be accomplished with a hollow metallic tube or the familiar coaxial cable. At optical and infrared frequencies, dielectric waveguides in common use employ index-guiding in a high-ε core or strip (see the subsection Index guiding of chapter 3). In optoelectronic devices, light is index-guided from one end of a microchip to another. In a fiber-optic network for telecommunications, light is index-guided from one end of a continent to another. Waveguides that confine light via photonic band gaps are a newer development. Their unique properties are already seeing commercial application in hollow fibers, as discussed in chapter 9.

As we first described in the section Linear Defects and Waveguides of chapter 5, the existence of a two-dimensional gap allows us to design a waveguide by forming a linear defect in the crystal. For example, we can remove a single row of rods from the crystal, resulting in the guided mode shown in figure 2. Also shown is the projected band diagram, in which we can see that the guided mode (red, in the yellow-shaded gap) covers a range of frequencies described by its dispersion relation $\omega(k)$. In previous chapters, we already remarked on one obvious feature of this waveguide: the light is guided primarily within the *air*, much like a hollow metallic waveguide, but very unlike a traditional index-guiding structure. This property can be exploited to reduce the interaction between the light and the material (e.g., to reduce absorption or nonlinearities), as discussed in chapter 9.

Alternatively, when one wants to *enhance* the interaction between the light and the material, the periodicity of the structure is also helpful. The trick is that the slope of $\omega(k)$ goes to zero at the edges of the Brillouin zones ($k = 0$, in this case). As described in the section Bloch-Wave Propagation Velocity of chapter 3, the slope can be interpreted as the *group velocity*, the velocity at which pulses and energy propagate along the waveguide. As it goes to zero, the interaction time between the field energy and the material is increased, which can be used to enhance the gain of a distributed-feedback laser or to manipulate nonlinear phenomena.[4] This can happen even in the very simple periodic index-guided waveguides described in chapter 7, because they have a zero-slope band edge.

For transmission of information, on the other hand, one typically operates at a frequency away from the zero-slope band-edge, both to decrease losses[5] and

[3] As described in Ramaswami and Sivarajan (1998), the usable bandwidth of standard silica fiber around $\lambda = 1.5\,\mu$m is about 180 nm; in practice, a much narrower bandwidth is typically used. A single 100-GHz signal channel occupies a bandwidth of less than 0.1%.

[4] See, for example, Xu et al. (2000) or Soljačić et al. (2002*b*).

[5] See Hughes et al. (2005) and Johnson et al. (2005).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

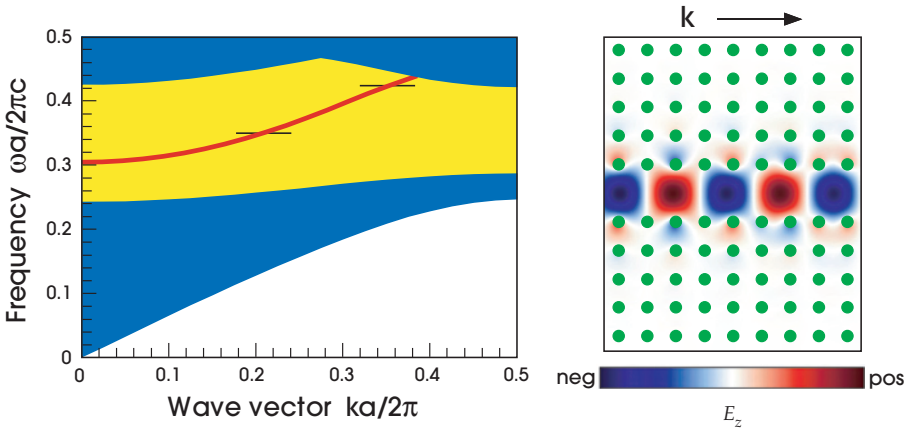


Figure 2: Line-defect waveguide formed by missing row of rods in square lattice of radius $0.2a$ dielectric rods ($\varepsilon = 11.4$) in air. *Left:* projected band diagram, with extended modes of crystal shaded blue, TM gap shaded yellow, and guided mode (red line in gap). Horizontal black lines indicate low-dispersion range where guided band is nearly flat. *Right:* Electric field E_z of guided mode at $\omega a/2\pi c = 0.38$, with dielectric rods in green.

to minimize distortion of the signal. It is usually desirable to operate where the slope (velocity) is nearly *constant*, such as the 20% range $\omega a/2\pi c = 0.35$ to 0.42 indicated by horizontal black lines in figure 2. The reason is that the variation of the slope ($d^2\omega/dk^2 \neq 0$) causes **group-velocity dispersion**. Pulses will spread and distort, and the pulse-spreading rate diverges as the zero-slope band-edge is approached.[6] Here, the center of our low-dispersion bandwidth is $\omega a/2\pi c = 0.38$, and we can easily rescale our structure so that this lies at our operating frequency $\lambda = 1.5\,\mu$m: using a calculation similar to equation (1), we choose $a = 0.57\,\mu$m and $r = 0.2a = 0.114\,\mu$m.

From the perspective of device design, however, perhaps the most valuable novelty of a photonic-bandgap waveguide is that it is effectively a *one-dimensional system*. In a conventional index-guiding dielectric wave guide, any defect that breaks the translational symmetry will cause light to be scattered out and be lost. This is why there are radiation losses at a sharp bend, at the locations of imperfections, in the transition zones from one waveguide to another, and at the interfaces with other devices. These losses can be eliminated in a photonic-crystal waveguide because the band gap forbids the light from going in any direction except along a particular line. The only loss mechanism (besides material absorption) is reflection, and we will see that in many cases reflection can be eliminated by simple symmetry rules.

[6] See footnote 10 on page 199, and Ramaswami and Sivarajan (1998). In some cases, as in Povinelli et al. (2005), the diverging dispersion is actually useful: it allows one to use a small change in the structure to achieve a large change in signal delay.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

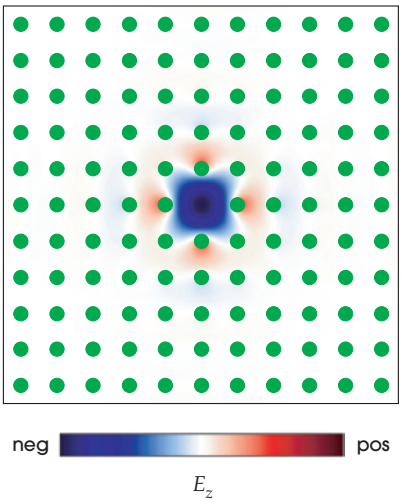


Figure 3: Point-defect cavity formed by a single missing rod in a square lattice of radius $0.2a$ dielectric rods ($\varepsilon = 11.4$) in air. The cavity supports a single mode of frequency $\omega a/2\pi c = 0.38$, inside the TM band gap, whose electric field E_z is shown.

Designing a cavity

Just as we designed a waveguide by making a linear defect, we can design an electromagnetic cavity by making a defect around a single point (as we saw in the section Localization of Light by Point Defects of chapter 5). For example, by removing a single dielectric rod, we trap a localized state at a frequency $\omega_0 a/2\pi c = 0.38$, as shown in figure 3. Intuitively, no matter which direction the light turns, it has no way to escape because of the band gap. Of course, our structure will confine light only within the plane of periodicity. To prevent it from escaping in the third direction, another method is needed. One might sandwich the triangular lattice between two metallic plates, or use a fully three-dimensional photonic crystal. Alternatively, one can employ approximate index-confinement in the third direction, as we discussed in chapter 8. For pedagogical purposes, we will remain focused on the two-dimensional case.

Such a cavity is useful whenever one would like to control light within a narrow frequency range or, equivalently, for a long time (since lifetime and bandwidth are inversely related by the Fourier transform). Later in this chapter, for example, we will use a cavity to make a narrow-bandwidth filter, and we will also show how to make a nonlinear switch where the long cavity lifetime increases nonlinear effects.

Cavities are also used to affect the rate of atomic transitions, which are accompanied by the emission or absorption of a photon at a particular frequency ω corresponding to the transition energy $\hbar\omega$. Such transitions can be suppressed by placing the atom in a photonic crystal in which there are no appropriate photon

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

states, or they can be enhanced by placing the atom in a cavity in which there is a tightly concentrated photon state at precisely the transition frequency. This topic falls under the heading of **cavity quantum electrodynamics** (**cavity QED**).[7]

There is not much point in trapping light if we cannot get it out *eventually*. For this reason, we must generally arrange for the cavity oscillations to have a specified lifetime: we must provide a way for the light to escape. The most controlled way to do this is to place a photonic-crystal waveguide in close proximity to the cavity, as in figure 4. The exponential tails of the cavity mode slowly leak into the waveguide, and energy escapes only into the desired channel (or channels). We will discuss this type of cavity–waveguide coupling in the next section.

Three important properties of a cavity mode are its frequency, its lifetime, and its symmetry. The frequency can be controlled by changing the geometry of the cavity. For example, we can change the size or ε of the rods in figure 16 of chapter 5. The lifetime τ is defined such that the field in the cavity decays as $e^{-t/\tau}$. It is often convenient to characterize the lifetime by the dimensionless quality factor $Q = \omega_0\tau/2$, as in the section Quality Factors of Lossy Cavities of chapter 7. The Q value can be controlled by changing the number of crystal periods between the cavity and the waveguide. It can also be tuned by altering the size of the band gap (e.g., by changing the rod radii) or by changing the frequency of the cavity mode with respect to the gap. The symmetry of the mode refers to whether its field pattern is monopole, dipole, and so forth as in figure 17 of chapter 5. The symmetry is important because it is possible that a mode of one symmetry will not couple at all to a waveguide (or an excitation, or atomic transition) with a different symmetry.

A Narrow-Band Filter

Consider the structure shown in the topmost inset of figure 4. The background photonic crystal is the same $\varepsilon = 11.4$ crystal that was discussed in the previous section. The cavity, formed by the absence of a single rod, is adjacent to two waveguides, each of which is formed by the absence of a row of rods. If we were to excite the cavity mode directly (by, say, a current source or atomic transition within the cavity), then the energy in the cavity would leak out slowly. In this case, $Q \approx 410$. However, something more interesting happens when we excite the cavity from one of the waveguides.

The top panel of figure 4 shows the **transmission spectrum**: the transmitted power in the output waveguide (bottom panels, right) as a fraction of the incident power from the input waveguide (left), as a function of frequency ω. This spectrum exhibits several important features, most notably a sharp peak precisely centered at the cavity frequency $\omega_0 a/2\pi c = 0.38$.

[7] See, for example, Berman (1994).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

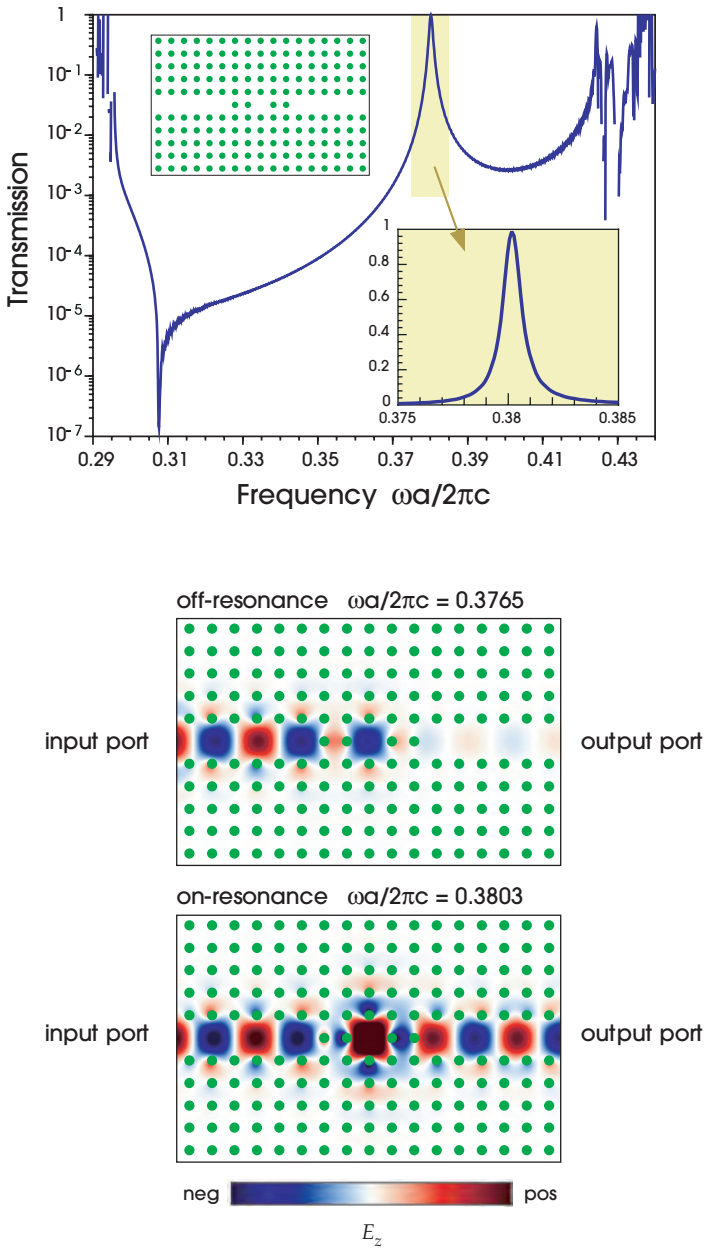


Figure 4: Waveguide–cavity–waveguide filter in rod crystal (inset, top). *Top:* Transmission spectrum, showing 100% peak at cavity resonance frequency ($\omega a/2\pi c = 0.3803$) with $Q = 410$; inset shows enlarged peak. Oscillations at low and high frequencies correspond to propagation outside the band gap, and sharp dip near $\omega a/2\pi c = 0.308$ corresponds to the zero-slope guided-band edge. *Bottom:* E_z field for transmission at a frequency 1% *below* resonance peak (upper), and exactly *at* resonance peak (lower).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

This sharp peak means that the device acts as a *narrow-band filter*. The light is transmitted for frequencies near the resonant frequency of the cavity, and is reflected for somewhat lower or higher frequencies. The existence of the resonance peak conforms with intuition: near the resonant frequency, light from the input waveguide can couple into the cavity, and the cavity in turn can couple into the output waveguide. What may be surprising, however, is that the peak transmission is precisely 100%. The field pattern for transmission at resonance is shown in the bottom panel of figure 4. If we shift the frequency by only 1%, the transmission drops to less than 2%, corresponding to the fields in the middle panel. The fractional width $\Delta\omega/\omega_0$ at half-maximum (50% transmission) is precisely equal to $1/Q$, where Q is the quality factor of the cavity mode when excited internally. These and other properties of the transmission peak will be explained in the next section, Temporal Coupled-Mode Theory, in a more general setting.

Before we analyze the resonance peak, however, it is worth commenting on the other features of the transmission spectrum in figure 4. The sharp dip in the transmission at around $\omega a/2\pi c = 0.3075$ corresponds to the zero-slope band edge of the waveguide mode, where coupling light through the device is especially difficult. The oscillations at high and low frequencies correspond to frequencies outside the band gap, where energy propagates through the crystal instead of being confined to the waveguide and cavity. In a truly infinite system, light outside the gap would escape, but because we simulate this structure within a finite computational box, some light returns to the output waveguide, where interference results in an oscillating spectrum.

Temporal Coupled-Mode Theory

In order to analyze a broad range of devices, including the one in figure 4, we can exploit a powerful theoretical framework that falls into a general class of methods known as **coupled-mode theories**: one describes a system in terms of a set of idealized components (e.g. isolated waveguides and cavities) that are perturbed, or *coupled*, in some fashion. These methods are analogous to time-dependent perturbation theory in quantum mechanics, and they take many forms. Often, they are formulated as an expansion in the exactly computed eigenmodes of the idealized systems, providing a numerical result for a particular geometry.[8] The method we will discuss, **temporal coupled-mode theory**,[9] uses a more abstract formulation.

[8] A classic presentation of this idea can be found in Marcuse (1991). Extensions to photonic crystals and further references can be found in Johnson et al. (2002*b*) and Povinelli et al. (2004).

[9] See Haus (1984, ch. 7), or Suh et al. (2004) for a generalization. By *temporal coupled-mode theory*, we refer not merely to any time-dependent description, but to the more abstract formalism here, which was developed by Pierce (1954), Haus, and others as described in Haus and Huang (1991) and Louisell (1960). Related ideas also appear in the Breit–Wigner scattering theory of quantum mechanics (Landau and Lifshitz, 1977).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

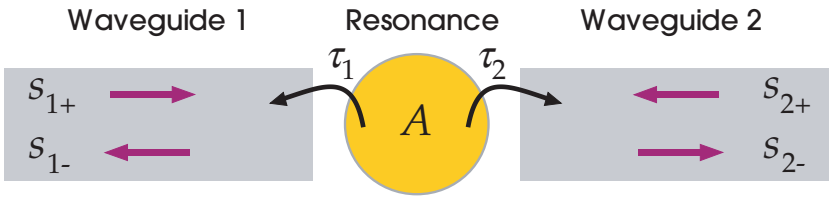


Figure 5: Abstract diagram showing the essential features of the filter from figure 4: a single-mode input waveguide 1, with input/output field amplitudes s_{1+}/s_{1-}; a single-mode output waveguide 2 with input/output field amplitudes s_{2+}/s_{2-}; and a single resonant mode of field amplitude A and frequency ω_0, coupled to waveguides 1 and 2 with lifetimes τ_1 and τ_2 ($\tau_1 = \tau_2$ in figure 4). The $s_{\ell\pm}$ are normalized so that $|s_{\ell\pm}|^2$ is power in the waveguide, and A is normalized so that $|A|^2$ is energy in the cavity.

In temporal coupled-mode theory, the system is considered as a set of essential components that are analyzed using only very general principles such as conservation of energy. Our building blocks will be *localized modes* (resonant cavities) and *propagating modes* (in waveguides). The result is a *universal* description of a certain class of devices. To obtain a quantitative result, the description is parameterized by a small number of unknowns such as the frequencies and decay rates of the resonant modes, which depend on the specific geometry and must be determined by a separate calculation.

This rather abstract idea is best understood by example. The structure of figure 4 is described in temporal coupled-mode theory as a resonant cavity connected to two single-mode waveguides (labelled 1 and 2), as depicted schematically in figure 5. There are no other places for the light to go; the rest of the crystal is ignored. The cavity mode has some resonant frequency ω_0 and decays with lifetimes τ_1 and τ_2 (defined more precisely below) into the two waveguides. In our structure, by symmetry, we must have $\tau_1 = \tau_2$, and it will turn out that this is the condition for 100% transmission on resonance. The key assumption of temporal coupled-mode theory (as with many other approximate methods) is that the coupling between the various elements is weak. In figure 5, for instance, we assume that the cavity energy leaks only slowly into the waveguides. We can imagine guaranteeing weak coupling by (for example) surrounding the cavity with a sufficient number of periods of the photonic crystal.

The temporal coupled-mode equations

We will now derive a set of equations describing the coupling of the cavity to the waveguides, in terms of the field amplitudes in those components. To do this, we will rely on five very general assumptions: weak coupling, linearity, time-invariance (i.e., the materials/geometry don't change over time), conservation of energy, and time-reversal invariance.[10] The most important of these

[10] Recall the section Time-Reversal Invariance of chapter 3.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

is weak coupling; the latter four can actually be relaxed, as we will see in later sections.

We suppose that the fields in the cavity are proportional to some variable A. That is, since the eigenequation of the cavity completely determines the electric and magnetic fields up to some overall complex amplitude, we call this overall amplitude A (giving both magnitude and phase). Since the units of A can be chosen arbitrarily, we make the convenient choice that $|A|^2$ is the electromagnetic *energy* stored in the cavity.

We express the fields in the waveguide as the sum of incoming and outgoing waveguide modes, which are again defined up to an arbitrary complex amplitude $s_{\ell\pm}$ ($\ell = 1, 2$). Here, $s_{\ell+}$ is the amplitude of the mode in waveguide ℓ going *towards* the cavity, and $s_{\ell-}$ is the amplitude of the mode going *away from* the cavity. Again, since the units are arbitrary, we choose to make $|s_{\ell\pm}|^2$ the incoming (or outgoing) *power* in the waveguide modes.[11]

What are the equations governing these quantities? To begin with, consider the cavity mode by itself, with no incident power from the waveguides. Because the coupling is weak, it is safe to assume that the mode will decay exponentially over time with some lifetime τ. This is justified intuitively as follows.[12] If the mode hardly decays at all over one optical period, then the solution is approximately that of the lossless cavity. There is a fixed field pattern proportional to A, and the outgoing Poynting flux $\mathrm{Re}[\mathbf{E}^* \times \mathbf{H}]/2$ must therefore be proportional to $|A|^2$, the energy; since the rate of energy loss is proportional to the energy, an exponential decay ensues. Quantitatively, we require $\tau \gg 2\pi/\omega_0$, or $Q = \omega_0\tau/2 \gg \pi$. (In practice, we typically find temporal coupled-mode theory to be nearly exact for $Q > 30$, and often qualitatively accurate even for smaller Q.) If the cavity has two loss mechanisms, with decay constants τ_1 and τ_2, then the net lifetime is given by $1/\tau = 1/\tau_1 + 1/\tau_2$. The amplitude A satisfies a differential equation $dA/dt = -i\omega_0 A - A/\tau$, for which the solution is $A(t) = A(0)e^{-i\omega_0 t - t/\tau}$.

Now we include the waveguides. Input energy from $s_{\ell+}$ can couple into the cavity, or it can be reflected into $s_{\ell-}$ (or both). Energy from the cavity must also flow into $s_{\ell-}$. The most general linear, time-invariant equations relating these quantities, assuming weak coupling,[13] are

$$\frac{dA}{dt} = -i\omega_0 A - A/\tau_1 - A/\tau_2 + \alpha_1 s_{1+} + \alpha_2 s_{2+} \tag{2}$$

$$s_{\ell-} = \beta_\ell s_{\ell+} + \gamma_\ell A, \tag{3}$$

[11] More precisely, the time-dependent function $s_{\ell\pm}(t)$ is normalized so that its Fourier transform $\tilde{s}_{\ell\pm}(\omega)$ gives $|\tilde{s}_{\ell\pm}(\omega)|^2$ as the power per unit frequency at ω. This is because the waveguide field is not determined by a scalar amplitude at a single time: the field pattern is ω-dependent. However, we can largely ignore this subtlety because we are mostly interested in the response at either a single ω or for ω near the cavity resonance.

[12] A more formal argument can be based on the framework of leaky modes (Snyder and Love, 1983).

[13] Weak coupling is assumed here not only in that A is exponentially decaying, but also in that dA/dt depends simply on $s_{\ell+}$ multiplied by a constant. More generally, one could imagine a

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

for some proportionality constants α_ℓ, β_ℓ, and γ_ℓ. The constants α_ℓ and γ_ℓ represent the strength of the cavity–waveguide coupling, and β_ℓ is a reflection coefficient. It may seem that there are too many unknowns for this approach to be useful, but in fact we can eliminate *all* of the unknowns except for ω_0 and τ_ℓ.

The constants γ_1 and γ_2 can be determined using the conservation of energy. Consider the simplified case where $\tau_2 \to \infty$, so that the cavity is decoupled from waveguide 2, and suppose $s_{1+} = s_{2+} = 0$, so that there is no input energy. In this case, the cavity mode decays exponentially as $A(t) = A(0)e^{-i\omega_0 t - t/\tau_1}$, and thus the energy $|A|^2$ is decreasing. The only place for this energy to go is into the outgoing power $|s_{1-}|^2$. Thus, we must have

$$-\frac{d|A|^2}{dt} = \frac{2}{\tau_1}|A|^2 = |s_{1-}|^2 = |\gamma_1|^2|A|^2. \tag{4}$$

Therefore, $|\gamma_1|^2 = 2/\tau_1$, and since the phase of s_{1-} is arbitrary (it could represent the field amplitudes anywhere along the waveguide) we can choose $\gamma_1 = \sqrt{2/\tau_1}$. Similarly, if we let $\tau_1 \to \infty$, we find $\gamma_2 = \sqrt{2/\tau_2}$. But when both τ_1 and τ_2 are finite, does the decay into waveguide 2 affect γ_1 or vice versa? No, not if the decay rates are weak. The quantity γ_1 is already small; any change in γ_1 due to $1/\tau_2$ (another small quantity) is a second-order effect, which we will neglect.[14]

The constants α_ℓ and β_ℓ can be determined by time-reversal symmetry. We just saw that for $s_{\ell+} = 0$, the cavity mode decays and the output fields are given by $s_{\ell-} = \sqrt{2/\tau_\ell}A$. Time-reversal symmetry tells us we can obtain another valid solution of the equations by running the original solution backwards in time, and conjugating in order to retain an $e^{-i\omega_0 t}$ time dependence. That is, we must have a solution to equation (2) of the form $A(t) = A(0)e^{-i\omega_0 t + t/\tau}$ (exponentially growing) with input fields $s_{\ell+} = \sqrt{2/\tau_\ell}A$ and zero output fields $s_{\ell-} = 0$. Plugging this into equation (3), we immediately conclude that $\beta_\ell = -1$. (Thus, for $\tau_\ell \to \infty$ we get 100% reflection, $s_{\ell-} = -s_{\ell+}$, as we might expect. The minus sign is an artifact of our phase choice for γ_ℓ earlier and is not physically significant.) To determine α_1, we again employ the trick of taking $\tau_2 \to \infty$, in which case plugging $A(t)$ in equation (2) immediately gives $\alpha_1\sqrt{2/\tau_1}A = 2A/\tau_1$. Thus, $\alpha_\ell = \sqrt{2/\tau_\ell} = \gamma_\ell$, and we can again neglect higher-order effects of τ_2 on α_1 or vice versa, thanks to weak coupling.[15]

---

convolution of $s_{\ell+}$ at different times, or equivalently a frequency-dependent $\alpha_\ell(\omega)$. Under weak coupling, however, only frequencies near ω_0 matter for coupling, in which case we can approximate $\alpha_\ell(\omega) \approx \alpha_\ell(\omega_0)$. Similarly for γ_ℓ.

[14] It is useful to see in this way that τ_1 and τ_2 are essentially independent quantities in the weak-coupling limit. Alternatively, conservation of energy implies that $2/\tau = |\gamma_1|^2 + |\gamma_2|^2$, and some authors simply *define* the individual decay rates by $2/\tau_\ell \triangleq |\gamma_\ell|^2$.

[15] This argument, while simple, is not strictly necessary. One can instead show that $\alpha_\ell = \gamma_\ell$ by exploiting an additional consequence of time-reversal symmetry and energy conservation, that the scattering matrix relating $s_{\ell+}$ and $s_{\ell-}$ is symmetric (Fan et al., 2003). We could also use reciprocity (Landau et al., 1984).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Finally, we have obtained the **temporal coupled-mode equations** for the system of figure 5:

$$\frac{dA}{dt} = -i\omega_0 A - \sum_{\ell=1}^{2} A/\tau_\ell + \sum_{\ell=1}^{2} \sqrt{\frac{2}{\tau_\ell}} s_{\ell+} \tag{5}$$

$$s_{\ell-} = -s_{\ell+} + \sqrt{\frac{2}{\tau_\ell}} A. \tag{6}$$

Note that we made no reference to the particular geometry of figure 5 in deriving these equations. They are valid for *any* filter satisfying our assumptions; the details matter only in determining the values of ω_0 and τ_ℓ. This approach is easily generalized to include more than two wave guides, radiative losses, and so on, as we will see. For now, however, we will leave the equations as they are and complete our analysis of the filter from figure 4.

The filter transmission

Given the coupled-mode equations (5) and (6), we can predict the transmission spectrum of any weakly-coupled waveguide-cavity-waveguide system. The transmission spectrum is simply the fractional output power $T(\omega) \triangleq |s_{2-}|^2/|s_{1+}|^2$ when $s_{2+} = 0$ (no input power from the right), as a function of the frequency ω.

Since frequency is conserved in a linear system, if the input oscillates at a fixed frequency ω, then the field everywhere must oscillate as $e^{-i\omega t}$, and $dA/dt = -i\omega A$. Plugging this, and $s_{2+} = 0$, into equations (5) and (6), we obtain:

$$-i\omega A = -i\omega_0 A - \frac{A}{\tau_1} - \frac{A}{\tau_2} + \sqrt{\frac{2}{\tau_1}} s_{1+} \tag{7}$$

$$s_{1-} = -s_{1+} + \sqrt{\frac{2}{\tau_1}} A \tag{8}$$

$$s_{2-} = \sqrt{\frac{2}{\tau_2}} A. \tag{9}$$

To solve for the transmission spectrum, divide equation (9) by s_{1+} and then solve for A/s_{1+} from equation (7). This gives:

$$T(\omega) = \frac{|s_{2-}|^2}{|s_{1+}|^2} = \frac{\frac{2}{\tau_2}|A|^2}{|s_{1+}|^2} = \frac{\frac{4}{\tau_1\tau_2}}{(\omega-\omega_0)^2 + \left(\frac{1}{\tau_1}+\frac{1}{\tau_2}\right)^2}. \tag{10}$$

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

This is the equation of a **Lorentzian peak** with a maximum at $\omega = \omega_0$. In the same way, we can derive the reflection spectrum:

$$R(\omega) = \frac{|s_{1-}|^2}{|s_{1+}|^2} = \frac{(\omega-\omega_0)^2 + \left(\frac{1}{\tau_1} - \frac{1}{\tau_2}\right)^2}{(\omega-\omega_0)^2 + \left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right)^2}. \tag{11}$$

It is easy to verify that $R(\omega) + T(\omega) = 1$ everywhere (energy is conserved), and that the reflection approaches 100% far from ω_0.

By inspecting equation (11) or equation (10), we see that $T(\omega_0) = 1$ *only* if $\tau_1 = \tau_2$, that is, when the cavity decays into the two waveguides at equal rates. In our photonic-crystal structure of figure 4, this equality is guaranteed by symmetry. The resonant reflection $R(\omega_0)$ is then zero. There are actually *two* sources of reflection—the direct reflection, and the light decaying backwards from the cavity—and at the resonance frequency these two reflections cancel exactly by destructive interference.

It is sometimes useful to write the transmission spectrum in terms of the quality factor Q instead of τ. The total lifetime is $1/\tau = 1/\tau_1 + 1/\tau_2 = 2/\tau_1$, and so $Q = \omega_0\tau/2$ implies $1/\tau_1 = 1/\tau_2 = \omega_0/4Q$. In this case, equation (10) becomes

$$T(\omega) = \frac{\frac{1}{4Q^2}}{\left(\frac{\omega-\omega_0}{\omega_0}\right)^2 + \frac{1}{4Q^2}}. \tag{12}$$

From equation (12), it follows that the fractional width $\Delta\omega/\omega_0$ at half-maximum ($T = 0.5$) is $1/Q$, as we observed in figure 4. In fact, if we were to plot equation (12) in figure 4, plugging in ω_0 and Q as determined by a small numerical computation (see appendix D), it would be nearly indistinguishable from the computed resonant peak shown in the inset.

To summarize, we have derived sufficient conditions for us to achieve a narrow-band filter with 100% transmission. We should have (i) a *symmetric* waveguide–cavity–waveguide system that is (ii) *single-mode* with (iii) *no other loss* mechanisms (such as radiation or absorption). Interestingly, the condition that our system be weakly coupled and amenable to the temporal coupled-mode theory is not really necessary, as we see in the next section. A photonic crystal provides the ideal situation for (iii), because it forbids all other radiative modes, whereas the losses that arise for an incomplete gap are analyzed in the section A Three-Dimensional Filter with Losses.

A Waveguide Bend

The applicability of temporal coupled-mode theory and figure 5 to the photonic-crystal filter of figure 4 is clear. Similar ideas can help us to understand

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

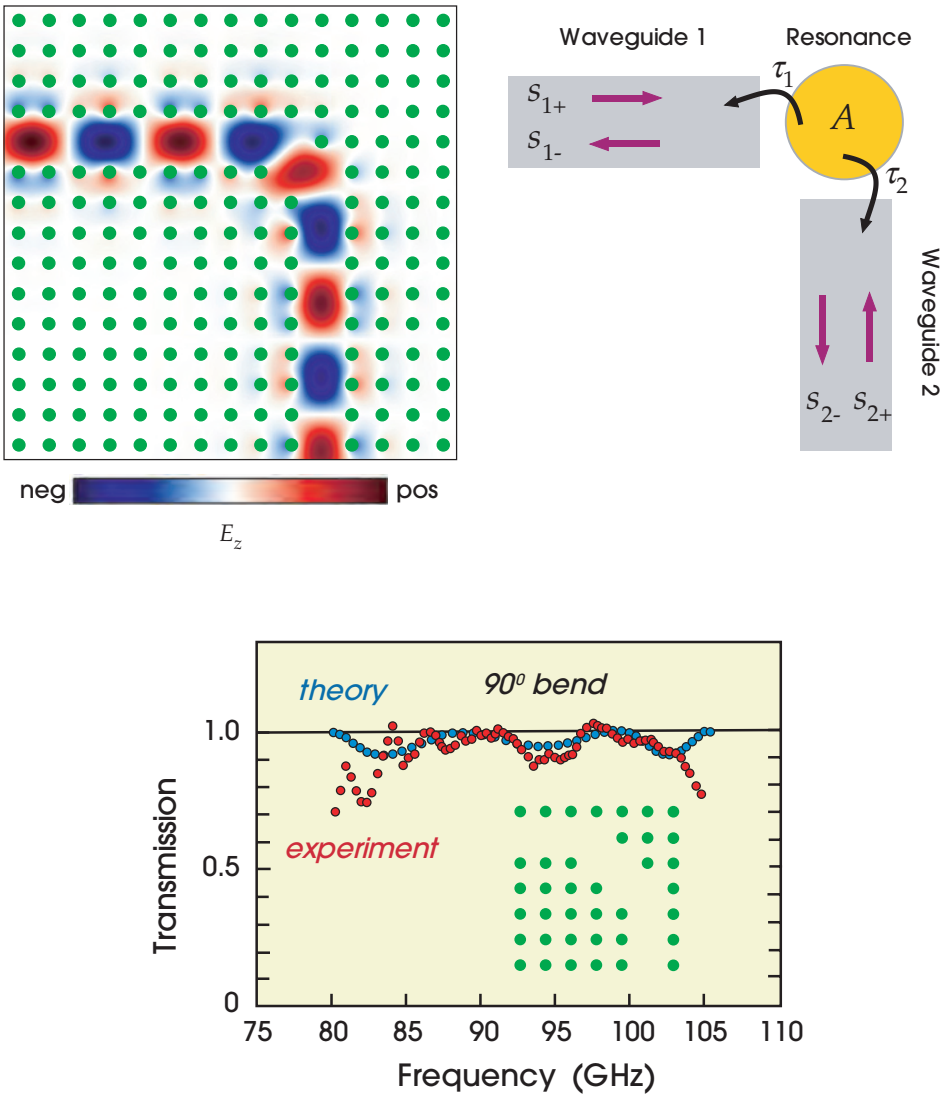


Figure 6: *Top left:* E_z field in right-angle bend, for waveguide of figure 2, showing 100% transmission at $\omega a/2\pi c = 0.35$. *Top right:* Roughly, the bend can be thought of as weakly resonant filter as in figure 5, with $\tau_1 = \tau_2$ by symmetry and thus 100% transmission with a broad (low-Q) resonance. *Bottom:* Experimental transmission spectra for 90° bend, along with theoretical prediction from simple 1D model, in a square lattice of $\varepsilon = 8.9$ alumina rods with $a = 1.27$ mm (Lin et al., 1998*a*).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

situations that seem at first to be quite different. Our first example of this is shown in figure 6, which depicts a sharp 90° bend in our missing-rod waveguide of figure 2.

When one bends an ordinary dielectric waveguide, two things happen: some light is reflected, and some light is radiated away. In general, the sharper the bend, the greater the radiation loss. In a low-contrast optical fiber, a bend radius of less than a few centimeters will result in nearly complete radiation loss, whereas for a high-contrast waveguide on a chip, the radiation loss can be quite small even for a wavelength-scale bend.[16] The situation changes in a photonic-crystal waveguide because the band gap prohibits radiation losses. One need only cope with reflection losses. And, remarkably, it is possible to arrange for the reflection loss to be zero, as illustrated in the top-left panel of figure 6. At certain frequencies, a photonic-crystal bend can exhibit 100% transmission even when the bend "radius" is smaller than the wavelength.

We can gain an immediate intuitive understanding of this transmission resonance by applying the coupled-mode results of the previous section. Let us picture the corner of the bend to be a weak (low-Q) resonant "cavity." The bend couples this "cavity" to two waveguides, as represented schematically by the upper-right diagram in figure 6. The fact that the geometry is bent is irrelevant to the analysis of the previous section. All that matters is that, by symmetry, the corner-resonator must decay at equal rates into the horizontal and vertical waveguides, and that there are no other radiation channels. If this system were weakly coupled, we would be able to conclude that the transmission peaks at 100% on resonance (although the resonance would be broad due to the low Q).

Of course, the flaw in this analysis is that the bend is *not* weakly coupled to the waveguides, and does *not* trap light for a long time ($Q < 10$). For these reasons, we cannot expect coupled-mode theory to be quantitatively accurate, but its qualitative predictions turn out to be valid. A more accurate theoretical model of the bend may be constructed by exploiting the fact that the problem is essentially one-dimensional: the light can go only forward or backward at every point. The bend may therefore be mapped onto the classic quantum-mechanical model of scattering from a symmetric one-dimensional potential well, in which it is known that 100% transmission resonances occur. The bottom panel of figure 6 shows the predicted transmission spectrum for such a model (Mekis et al., 1996) along with an experimental measurement at microwave frequencies (Lin et al., 1998*a*).

Like the filter, the single-modedness of the waveguide and the symmetry of the bend are the key enabling factors for the high transmission. Unlike the filter, the low Q in this case is actually a good thing, because it means that high transmission is achieved over a broad bandwidth (although the low Q admittedly makes a quantitative analysis more difficult).

[16] The radiation increases exponentially as the bend radius decreases (Marcatilli, 1969), but efficient sharp bends in high-contrast media were demonstrated by, for example, Manolatou et al. (1999*b*).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

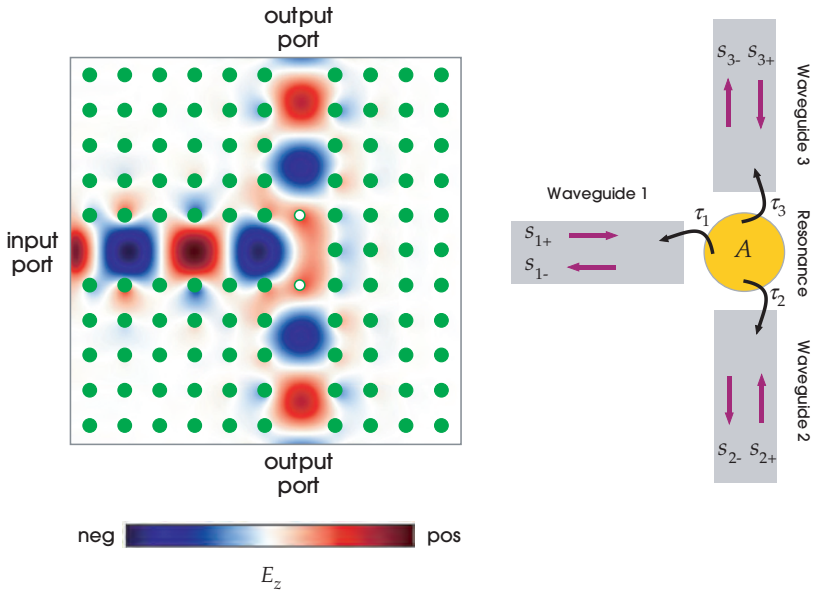


Figure 7: *Left:* E_z field in "T" splitter for waveguide of figure 2, showing essentially 100% transmission (50% in each branch) at $\omega a/2\pi c \approx 0.4$. *Right:* Abstract model, treating junction as weak resonance, predicting 100% transmission when $1/\tau_1 = 1/\tau_2 + 1/\tau_3$. Thus, to get 100% transmission, we *cannot* use a symmetric junction, and must insert rods (white rods with $\varepsilon = 3.5$) that *obstruct* the output waveguides in order to *increase* transmission.

A Waveguide Splitter

Another useful waveguide device is a *splitter*, which divides the power in an input waveguide equally between two output waveguides. Like the bend, the photonic band gap eliminates radiation loss and we need only deal with the possibility of reflection. *Unlike* the bend, it turns out that we *cannot* eliminate reflections by a symmetry argument, and must do something counterintuitive. We need to *obstruct* the output waveguides in order to *increase* transmission.[17]

An example of a T-shaped splitter structure for our missing-rod photonic-crystal waveguide is shown in figure 7(left). To analyze it qualitatively, as for the bend, we again treat the junction point as a low-Q resonant cavity and apply temporal coupled-mode theory, as depicted schematically in figure 7(right). Since this is now a three-port system, the coupled-mode equation (5) is modified to become

$$\frac{dA}{dt} = -i\omega_0 A - \sum_{\ell=1}^{3} A/\tau_\ell + \sum_{\ell=1}^{3} \sqrt{\frac{2}{\tau_\ell}} s_{\ell+}. \tag{13}$$

[17] The concepts in this section are discussed in more detail by Fan et al. (2001*b*).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Using equations (13) and (6) with $s_{2+} = s_{3+} = 0$ and solving for the reflection and transmission spectra as before, gives

$$R(\omega) = \frac{|s_{1-}|^2}{|s_{1+}|^2} = \frac{(\omega-\omega_0)^2 + \left(\frac{1}{\tau_1} - \frac{1}{\tau_2} - \frac{1}{\tau_3}\right)^2}{(\omega-\omega_0)^2 + \left(\frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3}\right)^2} \tag{14}$$

for reflection back into waveguide 1,

$$T_{1\to 2}(\omega) = \frac{|s_{2-}|^2}{|s_{1+}|^2} = \frac{\frac{4}{\tau_1\tau_2}}{(\omega-\omega_0)^2 + \left(\frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3}\right)^2}. \tag{15}$$

for transmission into waveguide 2, and

$$T_{1\to 3}(\omega) = \frac{|s_{3-}|^2}{|s_{1+}|^2} = \frac{\frac{4}{\tau_1\tau_3}}{(\omega-\omega_0)^2 + \left(\frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3}\right)^2}. \tag{16}$$

for transmission into waveguide 3.

From equation (14) it is clear that zero reflection can be achieved at $\omega = \omega_0$, and thus 100% transmission from waveguide 1 to waveguides 2 and 3, if

$$\frac{1}{\tau_1} = \frac{1}{\tau_2} + \frac{1}{\tau_3}. \tag{17}$$

This relation is very interesting for two reasons. First, it can never be satisfied in a 120° rotationally symmetric junction with $1/\tau_1 = 1/\tau_2 = 1/\tau_3$ (and in fact, the transmission cannot exceed $8/9$ in this case). Conceptually, equation (17) is the same as our condition for the filter: the decay rate to the input must equal the *total* decay rate to the output. As for the bend, this analysis is not quantitatively accurate when there is no strongly confined mode at the junction, but more careful analysis shows that the qualitative predictions are borne out.

Second, if we cannot satisfy equation (17) purely by symmetry, then we must force it "manually." In particular, (17) tells us that in a T splitter, where $\tau_2 = \tau_3 = \tau$, we must have τ larger than τ_1: the junction must be coupled twice as strongly to the input than to either of the outputs individually. Since even non-symmetric junctions such as figure 7(left) will usually couple similarly to the three output ports, we are led to a counterintuitive conclusion: we must generally add *obstructions* between the junction and the outputs to weaken their coupling in order to *increase* transmission!

In the case of figure 7(left), we accomplish this obstruction by adding a single rod (white) in before each of the output waveguides. Since we cannot satisfy equation (17) *a priori*, we must adjust the strength of this obstruction (varying the

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

radius or ε of the rod) until numerical simulations yield maximum transmission at a desired frequency. In this case, a rod ε of 3.5 creates a broad region of high transmission (>95%) from mid-gap towards a peak near the upper band edge, exceeding 99% for a roughly 4% bandwidth around $\omega a/2\pi c \approx 0.4$.

A Three-Dimensional Filter with Losses

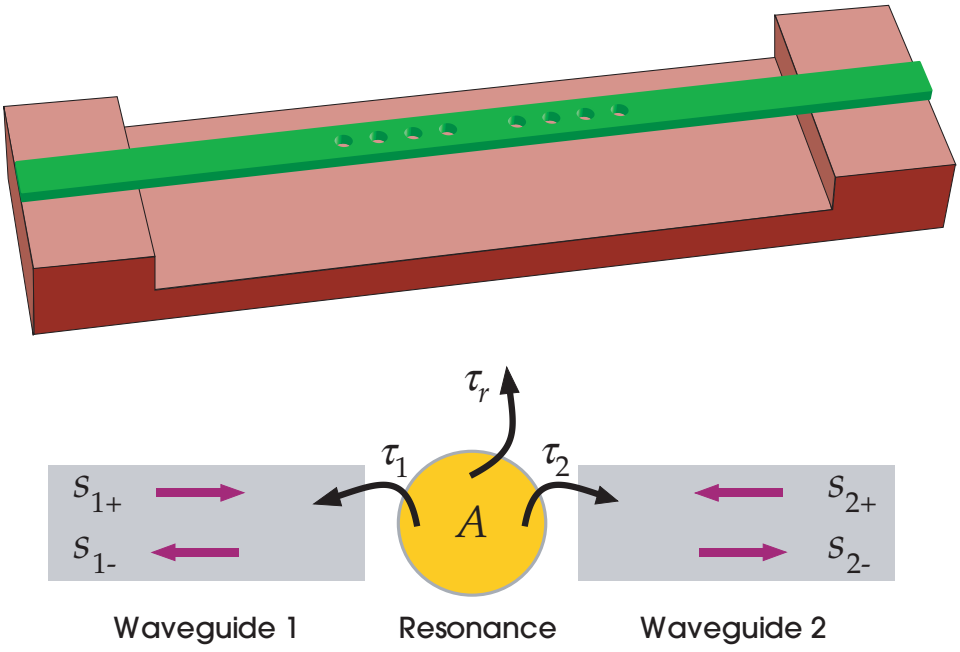
To fabricate a narrow-band filter in three dimensions, analogous to the design of figure 4, we would ideally employ a three-dimensional photonic crystal with a complete band gap. For example, almost exactly the same missing-rod geometry could be used with the crystal of the subsection A stack of two-dimensional crystals of chapter 6. However, fabrication would be greatly simplified by employing a hybrid structure combining band gaps with index guiding, as in chapters 7 and 8. In that case, we saw that point-defect cavities have intrinsic radiation losses. If we are to describe such structures with coupled-mode equations, we must add appropriate terms for this loss.

Again, our description will apply to a large class of devices, but we will illustrate it with the specific example shown in figure 8. This structure is similar to the one analyzed in the section Point Defects in Periodic Dielectric Waveguides of chapter 7: a $1.2a \times 0.43a$ dielectric waveguide suspended in air and pierced by a sequence of holes (radius $0.23a$) with period a, with four holes on either side of a defect with spacing $d = 1.26a$. This defect supports a resonant mode with a frequency $\omega_0 a/2\pi c = 0.3$ and a quality factor $Q = 350$. A very similar structure was fabricated to operate at wavelengths around $1.5\,\mu$m (1500 nm) in GaAs ($\varepsilon \approx 11.4$), corresponding here to $a = 460$ nm (Ripin et al., 1999).

A schematic model of this system is shown in the middle panel, which is identical to figure 5 except that we have added a new decay mechanism: the cavity mode radiates into the surrounding air with some lifetime τ_r. This is related to the radiative quality factor Q_r, defined in the section Quality Factors of Lossy Cavities of chapter 7, by $Q_r = \omega_0 \tau_r/2$. This decay is included into equation (5) for dA/dt by adding a $-A/\tau_r$ term. How does this new term affect the other decay rates and coupling constants?

There are two ways to approach cavity radiation loss in coupled-mode theory, both of which lead to the same result. The first way is to treat the radiation loss as simply another output port coupled to the cavity (similar to the splitter). Because a given cavity mode radiates with a fixed radiation pattern, the radiated fields can be characterized by a single amplitude, much like a single-mode waveguide. The other way is to rely on the weak-coupling assumption. This way has the advantage that it can be more easily extended to include material absorption losses (as in the next section), which breaks time-reversal symmetry and conservation of energy in the Maxwell equations. In particular, since the radiation loss is small ($Q_r \gg 1$), as required by the weak-coupling assumption, we can neglect the effect of τ_r on τ_ℓ or on the coupling constants between A and $s_{\ell\pm}$. This is essentially the same as the

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS



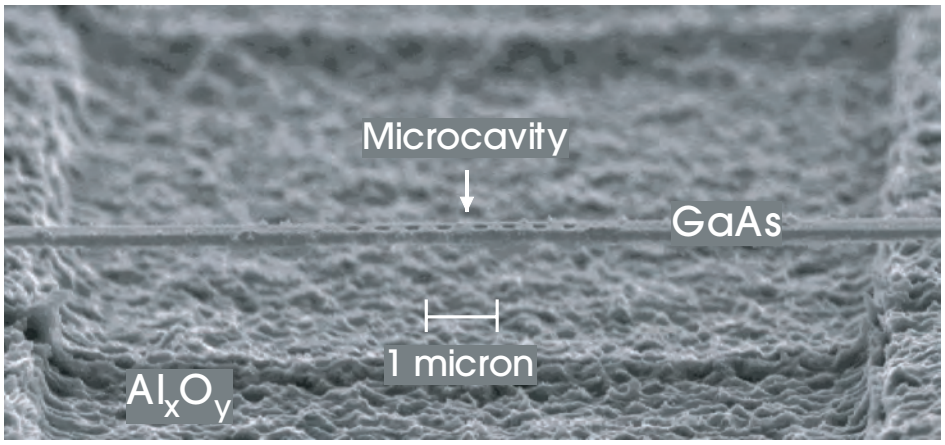


Figure 8: *Top:* Waveguide–cavity–waveguide filter formed in a suspended ("air bridge") dielectric waveguide by a periodic sequence (period a) of four air holes surrounding a defect (separation $d > a$). *Middle:* Schematic diagram of this structure, similar to figure 5 but with the addition of a radiative lifetime τ_r. *Bottom:* SEM image of fabricated structure in GaAs with period $a \approx 0.5\ \mu\mathrm{m}$ (Ripin et al., 1999).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

approximation we already made in the subsection The temporal coupled-mode equations of chapter: the effect of one cavity loss rate on another is a second-order effect. Therefore, equation (6) for $s_{\ell-}$ is unaffected, and equation (5) becomes

$$\frac{dA}{dt} = -i\omega_0 A - A/\tau_r - \sum_{\ell=1}^{2} A/\tau_\ell + \sum_{\ell=1}^{2} \sqrt{\frac{2}{\tau_\ell}} s_{\ell+}. \tag{18}$$

These equations can then be solved to find the transmission spectrum, as in equation (12). Again, $\tau_1 = \tau_2$ by symmetry, and we denote the total decay rate into the input/output waveguides by $1/\tau_w = 1/\tau_1 + 1/\tau_2$ (with $Q_w = \omega_0 \tau_w/2$). Thus, the total decay rate $1/\tau = 1/\tau_1 + 1/\tau_2 + 1/\tau_r$ corresponds to a total quality factor Q:

$$Q = \left(\frac{1}{Q_w} + \frac{1}{Q_r}\right)^{-1} = \frac{Q_w Q_r}{Q_w + Q_r}. \tag{19}$$

In terms of these quantities, the transmission spectrum is

$$T(\omega) = \frac{\frac{1}{4Q_w^2}}{\left(\frac{\omega-\omega_0}{\omega_0}\right)^2 + \frac{1}{4Q^2}}. \tag{20}$$

We can now evaluate the effect of the radiation loss Q_r. We still have a Lorentzian peak at the resonance frequency ω_0, with fractional width at half-maximum given by $1/Q$. The peak transmission $T(\omega_0)$ is no longer 100%, however. Now, the maximum is

$$T(\omega_0) = \left(\frac{Q}{Q_w}\right)^2 = \left(\frac{Q_r}{Q_w + Q_r}\right)^2 \approx 1 - \frac{2Q}{Q_r} + O([Q_w/Q_r]^2), \tag{21}$$

where the rightmost expression is for the usual case of $Q_r \gg Q_w$. From this equation, we see that the transmission approaches 100% for $Q_r \gg Q_w$, that is, when the cavity mode decays much more quickly into the waveguide than into the air. That confirms the intuition we developed in chapters 7 and 8, when we argued that making Q_r as large as possible is important for practical devices.

For example, suppose we want a filter with a bandwidth of 0.1% (a single 100 GHz channel at a wavelength of 1.5 μm), corresponding to a total $Q \approx 1000$. If we are willing to tolerate 2% losses, then we must have $Q_r \approx 10^5$ (so that $2Q/Q_r \approx 0.02$).

In the specific structure of figure 8, the radiation loss of an isolated cavity (surrounded by infinitely many holes) is computed to be $Q_r \approx 3000$. We therefore expect about 22% radiation loss from equation (21). In the computed transmission spectrum shown in figure 9(left), the peak transmission is about 74%. This is close to our prediction, and the reason why the loss is slightly greater than expected is explained at the end of this section. In comparison, the experimental transmission spectrum in arbitrary units is shown in figure 9(right) for two

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

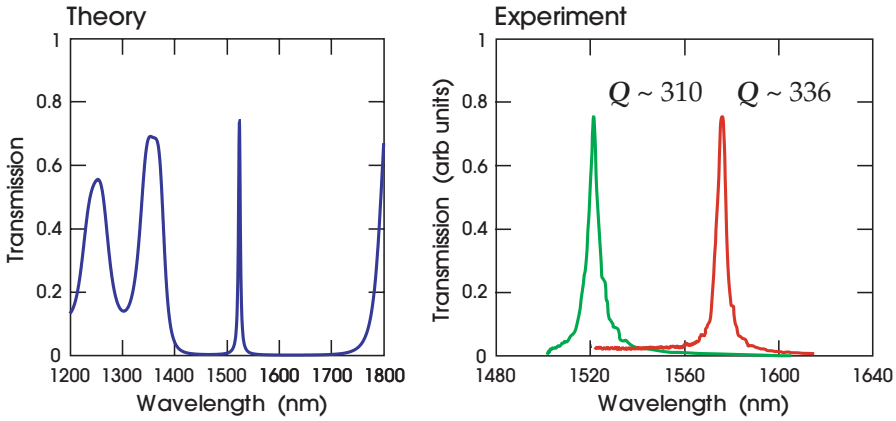


Figure 9: Theoretical and experimental transmission spectra (versus vacuum wavelength) for air-bridge filter of figure 8. *Left:* Theory, showing transmission peak in gap and broad Fabry–Perot oscillations outside gap. *Right:* Experiment, showing transmission for two defect separations $d = 1.26a$ (left, $Q = 310$) and $d = 1.41a$ (right, $Q = 336$).

different defects: d=1.26a and d=1.41a. (Measuring the absolute transmission is much more difficult.) As we might expect, the larger d corresponds to a larger resonance wavelength, whereas the Q value is almost unchanged (both resonances are deep inside the gap). Much higher transmission can be obtained by altering the structure to optimize Q_r; for example, 94% transmission was achieved in Ripin et al. (1999).

What happens to the energy that is not transmitted? Although most of it is radiated, part of it is reflected because the radiation loss spoils our zero-reflection condition. In particular, if we solve for the reflection spectrum $R(\omega)$ similar to equation (11), we find that the reflection $R(\omega_0)$ at resonance is Q^2/Q_r^2. So, the dominant effect of a small radiation loss ($Q_r \gg Q_w$) is radiation loss, not reflection. On the other hand, in the opposite limit of $Q_w \gg Q_r$, the reverse occurs: almost all of the light is reflected and the radiation loss goes as $2Q_r/Q_w$, while the transmission is roughly Q_r^2/Q_w^2.

In fact, the large-Q_w limit highlights a loss mechanism that we have not yet taken into account. In the equations so far, if we let $Q_w \to \infty$ (by adding many periods of the crystal) the result is 100% reflection ($s_{\ell-} = -s_{\ell+}$). In the real structure, however, some radiation loss occurs when light reflects from a semi-infinite crystal. The reason is that the broken translational symmetry and the incomplete gap allow us to couple to radiating modes inside the light cone.

It is possible to directly include this reflection–radiation loss in the temporal coupled-mode theory.[18] It is simpler, however, to realize that this loss will take place regardless of the value of Q for the cavity. Therefore, if we require that the

[18] In the formalism of Suh et al. (2004), we add a new radiation "port" that is directly coupled to the waveguide.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

losses be (say) 1% at most, then we must make sure that the reflection–radiation loss from the semi-infinite crystal is well under 1%, in addition to the $2Q/Q_r$ cavity losses above.

In the structure of figure 8, the reflection–radiation loss from a semi-infinite crystal is computed to be about 10% at ω_0, and the contribution of this loss is why the peak transmission in figure 9(left) was slightly lower than that predicted from the radiation losses of an isolated cavity. This loss could be decreased by, for example, "turning on" the crystal with a slow (adiabatic) transition rather than abruptly. In general, the problem of designing a low-loss reflection is related to the problem of designing a high-Q_r cavity that we discussed in chapter 8, and the use of a gradual transition is another example of the principle of trading off size for loss.[19]

Resonant Absorption and Radiation

In the previous section we incorporated radiation losses into our formalism by including a small radiative loss rate $1/\tau_r$, or equivalently, a large radiative quality factor Q_r. We can include any other external loss mechanisms in the same way. For example, material absorption (a small imaginary part of ε) would be included by adding a small absorption loss rate $1/\tau_a$ (corresponding to $Q_a = \omega_0\tau_a/2$).[20] In general, equation (5) becomes

$$\frac{dA}{dt} = -i\omega_0 A - A/\tau_x - \sum_{\ell=1}^{2} A/\tau_\ell + \sum_{\ell=1}^{2} \sqrt{\frac{2}{\tau_\ell}} s_{\ell+}, \tag{22}$$

where we have included the external loss rate $1/\tau_x = 1/\tau_r + 1/\tau_a + \cdots$.

Although external losses are usually undesirable, this is not always the case. In a photodetector, for example, the absorbed light may be converted into a useful current. Similarly, the radiated light may be fed into a camera or other sensor.[21] The narrow resonance bandwidth could then be used to detect the presence of certain frequencies (perhaps fluoresced in the presence of particular chemicals). If the absorption can be turned on and off, one can also use it as a switching mechanism.[22] In such cases, our design criterion is changed: we *want* 100% absorption or radiation at resonance.

[19] However, a gradual transition between the waveguide and crystal does not affect the cavity's modal volume, unlike tradeoffs in Q_r. A quantitative model relating Q_r to the reflectivity *inside* a cavity was analyzed in Sauvan et al. (2005).

[20] Given the cavity field pattern for the lossless (real-ε) materials, a small absorption loss is easily found by plugging the imaginary part of ε into the perturbation formula of the section The Effect of Small Perturbations of chapter 2. This yields a small imaginary part $\mathrm{Im}\,\Delta\omega$, which is precisely $1/\tau_a$.

[21] See, for example, Chutinan et al. (2001) for an analysis of one such design.

[22] See, for example, Wakita (1998) for absorption-based optical modulators and Fan et al. (2001*a*) for absorption-based switching in a channel-drop filter.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

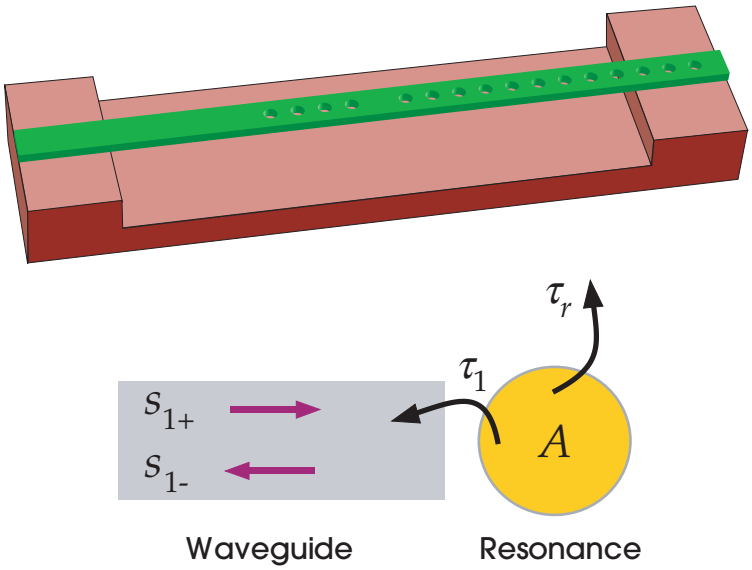


Figure 10: Schematic of air-bridge structure designed to radiate rather than transmit light on resonance: transmission is blocked by the band gap of many periods of holes on the right side of the cavity. This is modelled abstractly (bottom), similar to figure 5, by a resonant cavity coupled only to a single input waveguide and to radiation, but no output waveguide. The result is that $\tau_1 = \tau_r$ in order to maximize radiation on resonance, ideally achieving 100% radiation.

For example, suppose that we want to modify our filter structure of figure 8 to radiate 100% of incident light at ω_0, and to reflect other frequencies. In that case, transmission to the output waveguide is considered a loss. We can eliminate it by using a structure similar to figure 10. That is, we add many periods of crystal on the other side of the cavity, so that $\tau_2 \to \infty$. Furthermore, let us assume that absorption loss is negligible ($\tau_a \to \infty$). Then, solving the coupled-mode equations (6) and (22) for the radiation loss $1 - R(\omega)$, we find that it is precisely a Lorentzian peak as in equation (10), but with τ_r replacing τ_2.

Therefore, just as in our original filter design, we find that the condition for 100% radiation loss on resonance is that the two decay rates must be matched: $\tau_r = \tau_1$. Intuitively, if τ_1 is too big then the coupling to the cavity mode is too weak (everything is reflected), and if τ_1 is too small then the light does not persist for long enough in the cavity to radiate. Unlike the filter, however, we cannot satisfy this condition by symmetry. It must be arranged by varying the cavity parameters (e.g., the hole sizes) and thereby tuning τ_r and τ_1 until they are matched.

Similarly, if we want to attain 100% absorption, we should choose $\tau_a = \tau_1$ with $\tau_2 \to \infty$ and $\tau_r \to \infty$. If τ_r is finite, then just as in equation (21) it will lead to a radiation loss of $\approx 2Q/Q_r$.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

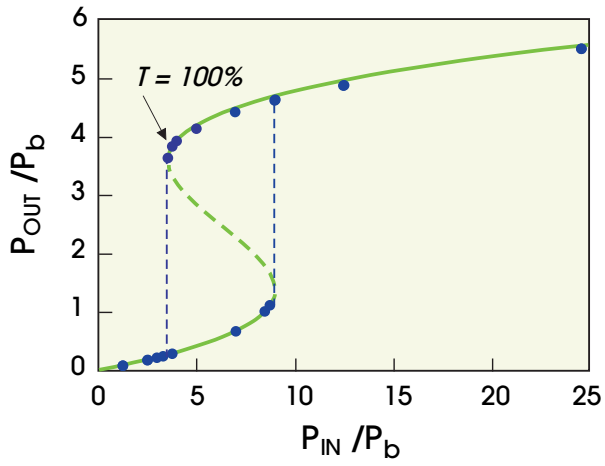


Figure 11: Output versus input power, in units of a characteristic power P_b, for the filter of figure 12 when a **Kerr nonlinearity** is included in the cavity: the frequency of the cavity shifts proportionally to the power. When we operate at a frequency just below the cavity resonance, the power follows the hysteresis curve shown, jumping between two bistable states (dashed green line is unstable solution). Dots are exact numerical calculation, and lines give theoretical curve predicted from coupled-mode theory with a nonlinear perturbation.

Nonlinear Filters and Bistability

We have shown that a wide variety of useful devices can be designed with simple, passive, linear materials. Nonlinear optical effects are typically so weak that we have been well justified in ignoring them for most of this book. However, similar to our discussion in the previous chapter, a cavity combines two features that make it easier for one to observe—and exploit—nonlinear phenomena. In a resonant cavity, light is stored for (i) a long time in (ii) a small volume. Both of these features lead to high field intensities and a greater sensitivity to small changes. Moreover, a photonic crystal is arguably the ideal environment for nonlinearity, because it allows us to combine tight confinement (on the wavelength scale) with long lifetimes, without the inevitable tradeoffs of the index-confined systems of chapters 7 and 8. In this section, we explore one important example of a nonlinear cavity effect: **optical bistability**.[23]

Optical bistability refers to the nonlinear relation between the output and input power in a filter with a nonlinear cavity, and in particular to the *S*-shaped curve as seen in the example of figure 11. (Contrast this with a linear device, in which the output power is always a linear function of the input power.) Because the middle (dashed) branch of the *S* is unstable, the power follows either the upper or lower

[23] Kerr-nonlinearity–based bistability in microcavity filters was originally proposed by Felber and Marburger (1976).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

branch of the S, jumping discontinuously when a branch ends. The output power depends on its past values (following the lower branch when starting from low power, and the upper branch when starting from high power), making this a type of **hysteresis effect**.

A simple qualitative picture to understand the source of this behavior is the following. Suppose that we have a cavity with a resonance at ω_0, and have input power at a frequency ω that is slightly below ω_0. Let us now imagine what happens as we slowly increase this power. (In the linear regime, the output power is simply proportional to the input power.) As the input power grows, ε will increase due to the nonlinearity, and this will shift ω_0 to lower frequencies, and thus pull down the resonant peak through ω. Consequently, one might expect a rise and fall in transmission. However, this simple picture is modified by feedback: as one moves into the resonance, coupling to the cavity is enhanced (positive feedback), creating a sharper "on" transition; and as one moves out of the resonance, the coupling is reduced (negative feedback), causing a delayed "off" transition.

This type of nonlinear transmission allows us to create an *all-optical transistor*: by adding or removing power from the input waveguide, we can switch from a low-transmission to a high-transmission state, and vice versa. Everything that can be done with an electronic transistor can also be accomplished with our optical device: switching, logic gates, signal rectification, amplification, and many other functions. The key question is how much power is required for the onset of the bistability, which is determined by the characteristic power P_b discussed below.

The simplest way to achieve this bistable phenomenon is to use the waveguide–cavity–waveguide filter of figure 5, where now we will include a **Kerr nonlinearity**: we suppose that the dielectric constant ε is proportional to $|\mathbf{E}|^2$.[24] Although the same nonlinear materials will typically be used everywhere in our device, the nonlinear effect is appreciable only in the cavity, where the field intensity is greatest. There, the shift in ε causes a shift $\Delta\omega_0$ in the *frequency* of our mode, with $\Delta\omega_0 \sim |A|^2$. A Kerr nonlinearity can also couple different frequencies, and most famously it can result in **third-harmonic generation** where a frequency ω is converted into light at a frequency 3ω. Unless there happens to be another cavity mode at $3\omega_0$, however, the frequency shift in ω_0 will be the strongest nonlinear effect in our cavity.

As before, we will develop a universal description for nonlinear waveguide–cavity–waveguide devices, but will illustrate it with a specific example. In particular, we will use the filter structure of figure 12, following a paper by Soljačić et al. (2002*a*). This structure is similar to that of figure 4, except that it

[24] There is a subtlety here, because in a nonlinear system we cannot write the field as a complex quantity and take the real part at the end of the calculation; we must take the real part from the beginning. After carefully developing the correct coupled-mode equations in terms of the real fields, however, it is possible to "put back" the imaginary parts in order to employ the convenient $e^{-i\omega t}$ time dependence (see, e.g., Soljačić et al., 2002*a*).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

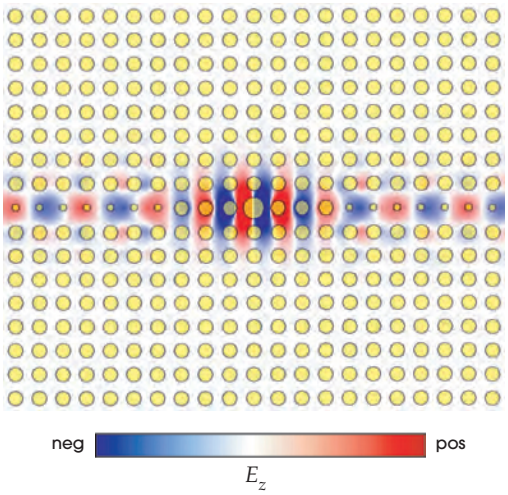


Figure 12: Waveguide–cavity–waveguide filter in square lattice of rods, similar to figure 4 but with slightly different parameters. Crystal: $\varepsilon = 5.4$ rods with radius $r = 0.2a$ and an 18% TM gap. Waveguide: row of reduced-radius ($r = 0.08a$) rods. Cavity: increased-radius ($r = 0.42a$) rod with a dipole-like state at $\omega_0\, a/2\,\pi c = 0.25814$.

happens to use slightly different parameters as given in the figure caption.[25] Thus, we can treat the problem just as we did for figure 4, and the linear spectrum (at low powers) is a Lorentzian peak centered on ω_0 with $Q = 557$.

The nonlinearity is included by assuming that the dielectric material of the rods has an ε that increases in proportion to $|\mathbf{E}|^2$, resulting in a cavity frequency that *decreases* in proportion to $|A|^2$. As we saw in the qualitative picture earlier, for such a nonlinear cavity in which ε increases with power, we will wish to operate at a frequency ω slightly below the linear resonant frequency ω_0. How far below? The important question is, how big is $\omega_0 - \omega$ relative to the half-width $\omega_0/2Q$ of the Lorentzian peak? This tells us how small the linear transmission will be, and leads us to define a **detuning parameter** δ by

$$\delta \triangleq 2Q\frac{\omega_0 - \omega}{\omega_0}. \tag{23}$$

Using equation (12), we can write the *linear* transmission at ω in terms of δ as $T(\delta) = 1/(\delta^2 + 1)$. For our specific example, we choose $\omega a/2\pi c = 0.25726$, yielding $\delta = 3.8$ and a linear transmission of 6.5%.

We now want to determine how the transmission will change with $|A|^2$ after we include our nonlinear material. But, from a practical point of view, it is more useful to write our transmission equation in terms of the input and output

[25] Although there are two dipole-like modes, only one of them (shown in figure 12) has the right symmetry to couple with the waveguide, since the other cavity mode is odd with respect to the waveguide mirror plane.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

powers, rather than $|A|^2$. To do this, we exploit the fact that the energy $|A|^2$ in the cavity is proportional to the power in the output waveguide, P_{out}. This is certainly true in the linear regime, and it turns out to be true even in the nonlinear regime since the output power must always come directly from the cavity energy.[26] That means that we can write our nonlinear effect as a decrease in ω_0 proportional to P_{out} instead of $|A|^2$, so that δ is replaced by $\delta - P_{out}/P_b$ for some proportionality constant $1/P_b$. Since δ is dimensionless, P_b must have units of power: it is a *characteristic power* that sets the scale for nonlinear effects to occur. (We will explain how to predict P_b below.)

As δ decreases, the transmission spectrum changes dramatically, because we are shifting the Lorentzian peak closer to our operating frequency ω. By substituting the shifted δ into the $T(\delta)$ equation, we find that the transmission becomes

$$\frac{P_{\text{out}}}{P_{\text{in}}} = \frac{1}{1 + (\delta - P_{\text{out}}/P_b)^2}. \tag{24}$$

Equation (24) is essentially exact as long as the nonlinear frequency shift $\Delta\omega_0/\omega_0$ is small. This assumption is always valid in practice, because in realistic materials the Kerr effect only changes the refractive index, and thus the frequency, by $< 1\%$. Equation (24) is *not* a Lorentzian. It is a cubic equation for P_{out}, and when solved at $\delta = 3.8$ it yields the S-shaped curve $P_{\text{out}}(P_{\text{in}})$ plotted in figure 11. The dots in figure 11 show the results of exact numerical solutions of the nonlinear Maxwell equations for figure 12, which confirm the prediction of equation (24).

A closer examination of equation (24) reveals that bistability (i.e., the presence of three real solutions and thus an S-shaped transmission) is attained for $\delta > \sqrt{3}$. Even without bistability, the transmission still becomes very nonlinear at powers near P_b, since 100% transmission is possible when $P_{\text{in}} = P_{\text{out}} = P_b\delta$. But what determines P_b?

The power scale P_b for nonlinear effects is determined by the power required to shift the cavity frequency by its bandwidth, which in turn depends upon the nonlinearity of the materials and the field strength $|\mathbf{E}|^2$ inside the cavity for a given input power. This field strength is inversely proportional to a measure of the modal volume V and is proportional to the lifetime Q (over which time the field builds up in the cavity). On the other hand, the required frequency shift of the cavity is proportional to the frequency width $1/Q$ of the Lorentzian transmission spectrum. Ultimately, therefore, $P_b \sim V/Q^2$. These simple arguments are confirmed by a more detailed analytical theory summarized below. Unlike traditional cavities such as ring resonators, photonic crystals with complete gaps impose no tradeoff between V and Q. The quality factor Q

[26] That is, $P_{\text{out}} = |s_{2-}|^2 = |A|^2 \cdot 2/\tau_2$, where any slight change in τ_2 from the nonlinearity is a higher-order effect. In contrast, $|A|^2$ is *not* proportional to P_{in} since the resonance frequency changes with power, as we shall see.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

can be increased arbitrarily (up to the required signal bandwidth) while V is maintained near its minimum of $\sim(\lambda/2n)^3$, where n is the index of refraction. Indeed, in the example system here, assuming reasonable semiconductor material parameters and the requirement $Q = 4000$ set by the telecommunications bandwidth, one obtains a theoretical operating power of only a few milliwatts. Experimentally, bistability was observed in a photonic-crystal slab cavity using a narrower bandwidth ($Q = 30,000$) with a power threshold of only $40\,\mu$W (Notomi et al., 2005).

We conclude by presenting an analytical formula for the characteristic power P_b in terms of the linear cavity mode.[27] The derivation, based on the perturbation theory of the section The Effect of Small Perturbations of chapter 2, is similar to the nonlinear effective-area calculation in the subsection Enhancement of nonlinear effects of chapter 9, and we use the same definition of the **Kerr coefficient** n_2 (the index shift per unit power in the average field intensity). The characteristic power P_b, is

$$P_b = \frac{V_{\text{kerr}}}{Q^2 \max(n_2/\varepsilon)} \cdot \frac{\omega_0}{4c}. \tag{25}$$

Here, V_{kerr} is a measure of the modal volume precisely defined to make equation (25) correct, and is given by[28]

$$V_{\text{kerr}} \triangleq \frac{\left(\int d^3\mathbf{r}\,\varepsilon|\mathbf{E}|^2\right)^2 \max(n_2/\varepsilon)}{\int d^3\mathbf{r}\,\frac{n_2\varepsilon}{3}\left(|\mathbf{E}\cdot\mathbf{E}|^2 + 2|\mathbf{E}|^4\right)}. \tag{26}$$

This definition quantifies the concentration of the electric field in the nonlinear material(s), similar to the effective area defined in equation (6) of chapter 9. The volume would normally be reported in units of $(\lambda/2n)^3$ to relate it to the wavelength in the material. In the two-dimensional structure of figure 12, we have a modal *area* ($d^3\mathbf{r} \to d^2\mathbf{r}$) $V_{\text{kerr}} = 1.7(\lambda/2n)^2 = (1.3\lambda/2n)^2$, which is close to a half-wavelength diameter.

Some Other Possibilities

In this chapter, we have analyzed only a few systems in detail, to illustrate the types of optical engineering that are made possible by an understanding of photonic crystals and coupled-mode theory. To give a sense of the wider scope

[27] This result is based on the expression derived in Soljačić et al. (2002*a*), cast into a slightly more convenient form. We use the conventional definition of n_2 here, which differs by a factor of $4/3$ from the one in the reference.

[28] Different physical processes lead to different measures of the modal volume; for example, a different V was used for spontaneous emission in the subsection Delocalization of chapter 8. The physical mechanism determines V's definition only up to a dimensionless scale factor; this scale factor is chosen so that V is the ordinary geometric volume for a *constant* field in some region of constant n_2 and ε.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

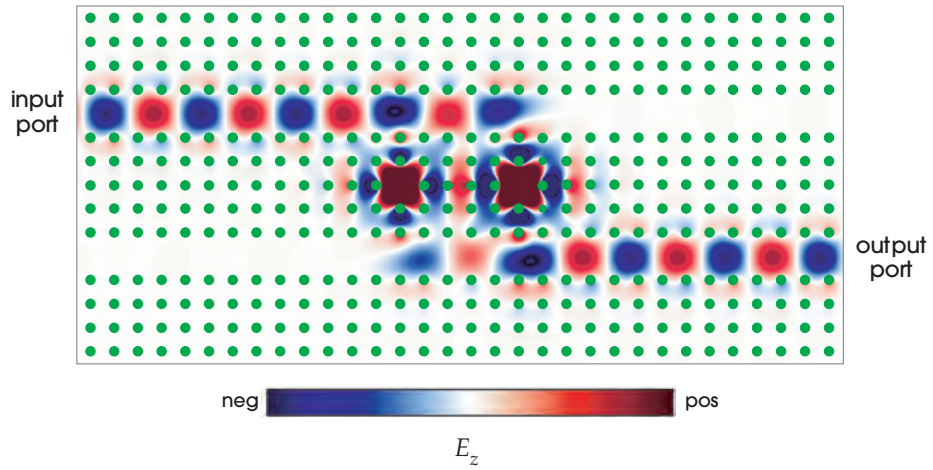


Figure 13: A **channel-drop filter** redirects 100% of light from the input port to the output port, but only around a single resonant frequency. Other frequencies propagate along the top waveguide unimpeded. E_z field at resonance is shown. This relies on a degenerate *pair* of resonant modes (here created by two missing-rod cavities), with appropriate symmetries, that couple the two waveguides.

of possibilities, we devote this brief section to a few other examples that will be discussed in less detail.

Our filter structure in figure 5 interrupted a waveguide with a cavity. What if, instead, we put a single-mode cavity *alongside* an unbroken waveguide? In this case, it turns out that we will get a Lorentzian *dip* in transmission, with nearly zero transmission on resonance (Haus and Lai, 1991). In general, when the input and output ports are coupled both directly and through a cavity, the resulting phenomenon is known as a **Fano resonance**. If the direct coupling is not perfect (i.e., partially reflecting even without the cavity mode), then one obtains an asymmetric spectrum of a sharp peak followed or preceded by a sharp dip.[29]

The transmission dip of the side-coupled cavity is interesting, but it just means that a narrow bandwidth is reflected. An even more useful device is a **channel-drop filter**, in which a narrow bandwidth is *redirected* to another "drop" waveguide while other frequencies are unaffected. An example channel-drop filter in our two-dimensional photonic-crystal system is shown in figure 13. To make an ideal channel-drop filter, it turns out that one always needs *two* cavity modes in between the two waveguides, which here are created simply by using two missing-rod cavities. If we only had one cavity mode, then nothing would prevent it from decaying in both directions in the drop waveguide; for a similar reason,

[29] See Fano (1961), or Fan et al. (2003) for a review and analysis of Fano resonances by temporal coupled-mode theory.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

partial reflection back into the input port would be unavoidable. With two modes, however, by careful design, one can arrange for their decaying components to interfere and cancel in unwanted output directions. In particular, if the two modes are forced to be degenerate and satisfy certain symmetries, it is possible to achieve 100% dropping in one direction of the output waveguide.[30]

In a channel-drop filter, transferring power from one waveguide to another is desirable. In other contexts, however, such transfer is called **crosstalk** and is to be avoided. If waveguides are kept far enough away from one another, of course, their crosstalk is exponentially small, but a complicated optical circuit may need to cross one waveguide over or through another at some point. In this case, crosstalk can be avoided by mediating the intersection with an appropriately designed cavity. If the cavity supports a pair of dipole-like modes, then the crosstalk is almost zero by symmetry.[31]

In deriving the coupled-mode equations, we assumed time-reversal invariance, but this symmetry is violated in an interesting case: magneto-optic materials.[32] Because they break time-reversal symmetry without relying on losses, magneto-optic materials can be used to design ideal **optical isolators**, which allow light to pass one way but not the other, and **circulators**, which direct light in different directions depending upon which way it is going. Such devices are crucial in large-scale photonic circuits in order to suppress reflections. Just as tiny filters can be constructed from cavities in photonic crystals, it turns out that tiny isolators and circulators can be designed using cavities made of magneto-optic materials, and temporal coupled-mode theory can be extended to this case as in Wang and Fan (2005).

In this chapter we showed how to design filters with a Lorentzian spectrum, but oftentimes a Lorentzian peak is not ideal. One would rather have a square band-pass response, with 100% transmission within a given frequency range and zero elsewhere. Such an ideal box-like spectrum can be approximated to any desired accuracy, in principle, simply by cascading the effects of several resonant cavities (Little et al., 1997).

Finally, let us mention one interesting twist on the coupled-mode device concept. All of our previous examples employed combinations of waveguides and cavities. Instead, one can couple *only* cavities together, in a periodic sequence (i.e., a linear defect), to *make* a waveguide. This type of linear defect has been termed a **coupled-cavity waveguide** by Yariv et al. (1999), and can be analyzed simply by coupled-mode theory. The cavities can form a waveguide for the same reason as in any other periodic structure: there is a conserved wave vector along the direction of periodicity, and so modes propagate from one cavity to the next

[30] For a derivation of these conditions, and further explanation of figure 13, see Fan et al. (1998) and Manolatou et al. (1999*a*), or Soljačić et al. (2003) for a nonlinear version. The two cavity modes are conventionally achieved by the counterpropagating modes of a ring resonator (as in, e.g., Little et al., 1997), which turns out to be a special case of the more general criteria.

[31] See Johnson et al. (1998), and experimental realization in (Roh et al., 2004). This geometry was proposed for nonlinear switching in Yanik et al. (2003).

[32] See footnote 8 on page 198.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

without reflecting. If the cavities are weakly coupled, however, light leaks from one cavity to the next very slowly, and this results in a low group velocity. Although such **slow light** has many potential applications,[33] the low velocity is unfortunately accompanied by a narrow bandwidth, reflecting an intrinsic delay–bandwidth tradeoff in passive, finite, linear systems. An intriguing proposal to overcome the bandwidth limitation by using a *time-varying* structure can be found in Yanik and Fan (2004).

Reflection, Refraction, and Diffraction

Although much of this book is concerned with *confining* light, we would be remiss if we did not at least mention the many interesting phenomena associated with *free* propagation of waves in and around photonic crystals. In this section, we digress to briefly review a few of these phenomena, and to relate their underlying principles to the foundations established in the preceding chapters.

Consider the case where an incident plane wave strikes an interface of a photonic crystal, as depicted in figure 14(left). Some of the light will be *reflected*, at an angle equal to the incident angle. Outside of a photonic band gap, some of the light can be transmitted or *refracted*, propagating at some angle (of the group velocity) within the crystal. And, depending upon the frequency, the interface periodicity, and the band structure, there may also be a finite number of additional reflected and/or refracted waves, a *Bragg-diffraction* phenomenon.[34]

All of these processes can be understood with the help of Bloch's theorem, via a corollary we pointed out in chapter 3: in a linear system with discrete translational symmetry, the Bloch wave vector $\mathbf{k}$ is conserved as light propagates, up to addition of reciprocal lattice vectors. Here, there is only translational symmetry along directions parallel to the interface (xz), and so only the wave vector $\mathbf{k}_\parallel$ *parallel* to the interface is conserved. That is, suppose that the periodicity parallel to the interface is Λ, and the incident plane wave has wave vector $(k_\parallel, k_\perp)$ and frequency ω. Then, any reflected or refracted wave must also have frequency ω and a wave vector $(k_\parallel + 2\pi\ell/\Lambda, k'_\perp)$ for any integer ℓ and some $k_\perp{}'$. Below, we illustrate this principle by examining several cases in more detail, and introduce the useful tool of an *isofrequency diagram* to analyze refracted waves in the crystal.

[33] For example, to enhance nonlinear effects (Xu et al., 2000; Soljačić et al., 2002*b*) or to achieve tunable time delays (Povinelli et al., 2005).

[34] *Diffraction* is a confusing word because it refers to phenomena in two very different limits. In the limit where the wavelength λ is small relative to the structure, it refers to deviations from geometric optics due to the fact that $\lambda > 0$ (Jackson, 1998). In the context of scattering from periodic structures, on the other hand, it refers to unusual reflected/refracted waves that arise because $\lambda < \infty$, and especially in the case where $\lambda/2$ is comparable to or smaller than the periodicity. Here, we are using it in the latter sense, specifically for the phenomenon of multiple reflected/refracted waves; to quote the original (1671) definition of the word *diffraction*: "when the parts of Light, separated by a manifold dissection, do in *the same medium* proceed in different ways"(see, e.g., Hall, 1990).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

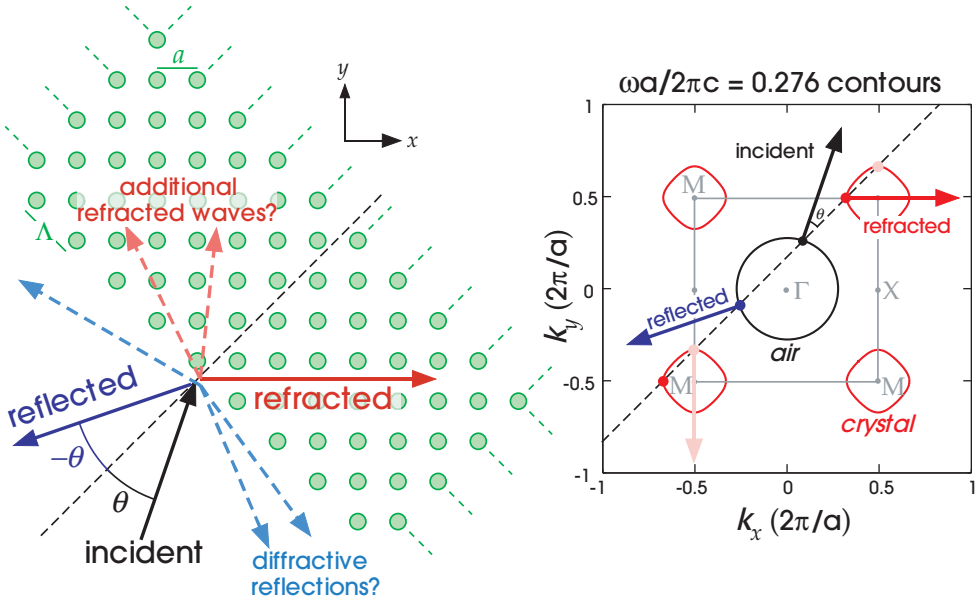


Figure 14: *Left:* Schematic of reflection (blue) and refraction (red) of a plane wave incident (black) on a square lattice of dielectric rods (green) in air, for an interface with period $\Lambda = a\sqrt{2}$ in the diagonal (110) direction. Depending on the frequency, there may also be additional reflected and/or refracted waves due to Bragg diffraction. *Right:* Isofrequency contours in $\mathbf{k}$ space at $\omega a/2\pi c = 0.276$ for air (black circle) and crystal (red contours), with the Brillouin zone in gray. The group velocity direction at various $\mathbf{k}$ points is shown as arrows (black/blue/red for incident/reflected/refracted waves). Because the wave vector component parallel to the interface is conserved, all reflected and refracted solutions (blue and red dots) must lie along the dashed line (running perpendicular to the interface).

Note that the periodicity Λ of the interface need not be the same as the lattice constant of the crystal. For the example of figure 14(left), a square lattice with period a, the diagonal (110) interface has period $\Lambda = a\sqrt{2}$. Indeed, the interface need not be periodic at all,[35] but we restrict ourselves here to periodic interfaces because those are the cases where the behavior can be predicted most simply.

Reflection

The familiar **specular** reflected wave corresponds to $\ell = 0$, that is, a plane wave with $\mathbf{k} = (k_\parallel, k'_\perp)$. Moreover, suppose that the incident medium has a refractive

[35] The periodic interface orientations are the *lattice planes*, where every lattice plane is normal to some reciprocal lattice vector. The lattice planes are typically described by Miller indices, as defined in appendix B.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

index n_i, so that $n_i^2\omega^2/c^2 = k_\parallel^2 + k_\perp^2 = k_\parallel^2 + (k'_\perp)^2$. It follows that $k'_\perp = \pm k_\perp$ and, because the reflected wave must be propagating *away* from the interface, we obtain $k'_\perp = -k_\perp$. This reflected wave, as shown in figure 14(left), thus satisfies the well-known "law" of equal incident and reflected angles.

The existence of $\ell \neq 0$ diffractive reflections depends on the frequency. Conservation of ω means that, in general,

$$k'_\perp = -\sqrt{n_i^2\omega^2/c^2 - (k_\parallel + 2\pi\ell/\Lambda)^2}. \tag{27}$$

From this equation, however, it is clear that if ω is too small, or if ℓ is too large, then $k'_\perp$ for $\ell \neq 0$ will be *imaginary*, corresponding to an *evanescent* field that decays exponentially away from the interface. We will obtain a non-evanescent diffractive reflection at ℓ only for $\omega > c|k_\parallel + 2\pi\ell/\Lambda|/n_i$. If we write $k_\parallel = \omega\sin(\theta)n_i/c$ for an incident angle $\theta \geq 0$, then the first diffractive reflection will occur at $\ell = -1$, for

$$\frac{\omega\Lambda}{2\pi c} = \frac{\Lambda}{\lambda} > \frac{1}{n_i(1+\sin\theta)}. \tag{28}$$

For the common case of $n_i = 1$ (air), this means we can have *no* diffractive reflections if $\omega\Lambda/2\pi c = \Lambda/\lambda \leq 0.5$, which includes the frequencies around most band gaps in this book if $\Lambda = a$. When diffractive reflections do occur, each diffracted order (ℓ) starts at glancing angles ($k'_\perp = 0$) and moves towards the specular reflection angle $-\theta$ as ω increases.

Refraction and isofrequency diagrams

The analysis of the refracted wave(s) is similar to that of reflection: we look for propagating waves in the crystal with frequency ω, surface-parallel wave vector $\mathbf{k}_\parallel + 2\pi\ell/\Lambda$, and group velocity pointing *away* from the interface. Now, however, the set of available states $\omega(\mathbf{k})$ is more complicated, given by the band structure of the crystal.

For most of this book, we have used a band diagram as a "map" of the available states $\omega(\mathbf{k})$, but this is insufficient here: an ordinary band diagram shows the states only around the boundaries of the Brillouin zone. Instead, we make a contour plot of $\omega(k_x, k_y)$ in the (k_x, k_y) plane, showing the (periodic) curves of constant ω. This contour plot, sometimes called an **isofrequency diagram** or a **wavevector diagram**, is shown in figure 15 for the first two TM bands of our square lattice of dielectric rods from the beginning of this chapter. As we saw in chapter 5, the first band has a maximum at the M point(s) in the Brillouin zone, and the second band has a minimum at the X point(s) and a maximum at Γ. Moreover, by a well-known theorem of vector calculus, the gradient $\nabla_\mathbf{k}\omega$—the group velocity, from chapter 3—is perpendicular to the ω contours and points in the direction of increasing ω.

Given the isofrequency diagram, it is straightforward to determine the number of refracted waves (if any), and in what states/directions they propagate.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

(The *amplitudes* of the refracted and reflected waves, on the other hand, depend on the surface termination and require a more detailed solution of the Maxwell equations.) The procedure is depicted in figure 14(right). For a chosen frequency, here $\omega a/2\pi c = 0.276$, we superimpose the (black) contour for the incident medium (air, a circle $\omega = c|\mathbf{k}|$) and the (red) contours for the transmitted medium (the photonic crystal). The incident angle, in this plot, corresponds to a particular wave vector $\mathbf{k}$ (black dot) and its group velocity (black arrow) on the incident contour. Then, to select modes with the same $k_\parallel$, we draw a dashed line through the incident $\mathbf{k}$ and perpendicular to the interface (here, along the Γ–M direction). The place(s) where this dashed line intersects the photonic-crystal contours determines the refracted wave(s).

Although the fixed-$k_\parallel$ line may intersect the photonic-crystal contours in several places, not all of these correspond to distinct refracted waves. First, we can eliminate any intersections (pink dots) corresponding to a group velocity (pink arrow) *towards* the interface from the crystal, as these violate our boundary conditions (the only incoming power is from the incident medium). Second, points that differ by a reciprocal lattice vector (i.e. equivalent points in the periodic $\mathbf{k}$-space) correspond to the same eigenstate as we saw in chapter 3, so we need only keep one of them.

In the example of figure 14(right), therefore, we happen to only have a single refracted wave (red dot/arrow), which in this case lies on the same side of the normal as the incident wave (the opposite of the usual Snell refraction). There are two ways to get multiple refracted waves. First, we could intersect multiple bands of the band structure, which would involve superimposing multiple contours on the isofrequency diagram. (This doesn't happen here because we are operating at an ω below the band gap, where there is only one band.) Second, our fixed-$k_\parallel$ line could intersect a given contour at inequivalent points in different periodic unit cells. For example, if we had an interface at a different angle, the fixed-$k_\parallel$ line would not be at 45° and could easily intersect the same contour at many different points.

As for reflections, there will always be some cutoff frequency below which at most a single refracted wave is obtained—indeed, in the large-λ limit the behavior approaches the ordinary Snell's law, corresponding to an effective medium with some "average" refractive index. Correspondingly, the isofrequency diagram in figure 15, for small $\omega \ll \pi c/a$, approaches circular contours with a constant group velocity.

Conversely, there will be some cases for which there is *no* refracted wave. This happens even for Snell's law, in the case of total internal reflection, but here we have more opportunities for reflection thanks to the photonic band gaps. For example, in figure 14(right), we can see that for a larger incident angle θ our dashed line would not intersect the red photonic-crystal contour, leading to total reflection—even though this frequency lies outside the photonic band gap for other angles. This is simply another perspective on the same bandgap-confinement mechanisms that we have explored for most of this book.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

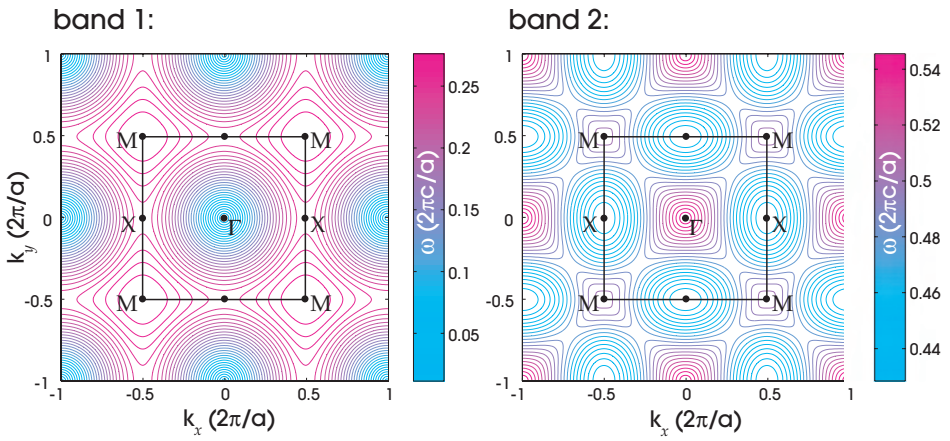


Figure 15: Isofrequency diagrams: contour plots of $\omega(k_x, k_y)$ for the first two TM bands of a square lattice of radius $0.2a$ dielectric rods ($\varepsilon = 11.4$) in air. The first Brillouin zone is shown as black squares.

It may seem that we have forgotten something—shouldn't we also draw dashed lines for $k_\parallel + 2\pi\ell/\Lambda$ for all ℓ? This is not necessary, however, because the periodic replication of the photonic-crystal contour in **k**-space is equivalent to adding $2\pi\ell/\Lambda$ to $k_\parallel$. We could also solve for the diffractive reflections with the same diagram, if we wished, by periodically replicating the black air contour to lie around every reciprocal lattice point. (Here, however, our frequency is too low to have diffractive reflections for this geometry.)

Unusual refraction and diffraction effects

Because the band structure and isofrequency diagram of a photonic crystal are so different from the simple circles or ellipsoids of a homogeneous dielectric medium, a number of interesting effects can be obtained.

For instance, notice that many of the contours in figure 15 exhibit fairly sharp corners (such as the fourth contour around the M point in the band 1). For refracted waves near these corners, as the angle or frequency is slightly changed, we can quickly switch from one side of the corner to the other. This results a large (possibly over 90°) change in the angle of the group velocity. The possibility of an enormous change in refracted angle for a small change in incident angle or frequency has been termed a **superprism** effect.[36] Recall that an ordinary prism splits different wavelengths into different angles, resulting in a rainbow, due to the

[36] The group-velocity superprism effect was proposed by Kosaka et al. (1998), following a phase-velocity effect proposed by Lin et al. (1996), and its applications were subsequently investigated by many authors (e.g., Kosaka et al., 1999*b*; Wu et al., 2003; Serbin and Gu, 2005).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

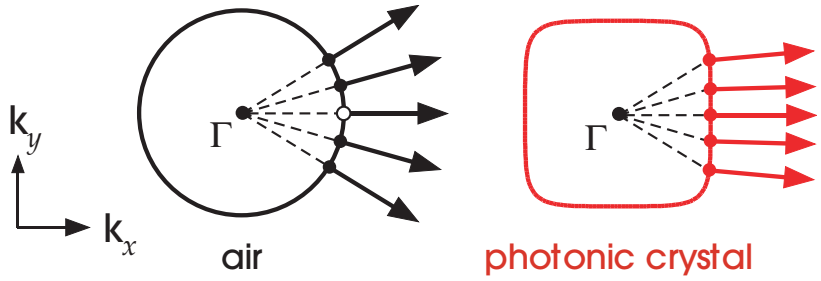


Figure 16: Isofrequency contours in air (left, black) and in the photonic crystal (right, red) of figure 15 (the $\omega a/2\pi c = 0.496$ contour of the second band). Arrows show the group velocity directions for a variety of $\mathbf{k}$ vectors (dots) with different k_y components. The nearly flat contour of the photonic crystal means that a finite-width beam made up of these wave vectors will spread (diffract) very slowly compared to air, an effect called *supercollimation*.

small material dispersion. Here, a much smaller range of wavelengths can be split over a much wider range of angles, leading to potential applications in frequency demultiplexing and related areas. A similar phenomenon can also be achieved using very *flat* contours, based on the observation that a flat contour makes it easy to achieve a large change in *phase* velocity.[37]

Another unusual characteristic of the photonic-crystal isofrequency diagram is that some of the contours are nearly flat in figure 14 (almost square). This flatness can almost negate the effects of another form of diffraction, which normally causes spreading of a narrow beam in a uniform medium. Suppose that we have a finite-width beam of light propagating in the $\hat{\mathbf{x}}$ direction, and consider its decomposition into different k_y components (i.e., its Fourier transform along the y dimension). An infinitely wide beam (a plane wave) consists of the single $k_y = 0$ component (corresponding to the hollow dot in the air contour of figure 16), but as the beam becomes narrower the Fourier uncertainty principle implies that it has more and more nonnegligible k_y components. In a homogeneous medium of index n where the frequency contour is a circle, these different k_y components all have different angles $\theta = \sin^{-1}(ck_y/n\omega)$, as shown by the black arrows in figure 16. And because the different Fourier components propagate at different angles, the beam spreads out—a classic *diffraction* effect of the nonzero wavelength, where the spreading angle is proportional to the wavelength divided by the beam width. If the frequency contour is nearly flat and perpendicular to $\hat{\mathbf{x}}$, however (e.g. some of the contours around the Γ point of band 2 in figure 15), then the group velocities of many different k_y values all point approximately in the $\hat{\mathbf{x}}$ direction. This is depicted by the red arrows in figure 16, which are nearly parallel

[37] A large phase-velocity superprism effect and its advantages were proposed by Luo et al. (2004), Matsumoto and Baba (2004), and Bakhtazad and Kirk (2005).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

despite their differing **k** vectors (red dots). A beam whose Fourier components lie within such a flat contour, therefore, spreads out very slowly, an effect called **supercollimation**.[38] Experimentally, beams only a few infrared wavelengths wide have been observed to propagate for nearly a centimeter with almost no spreading. This was accomplished by using a photonic-crystal slab (a square lattice of $\sim 10^9$ air holes in silicon) designed to operate at $\lambda = 1.5\,\mu$m (Rakich et al., 2006).

Finally, we already observed in figure 14 a behavior quite different from ordinary refraction: **negative refraction**, where a single incident beam couples to a single refracted beam on the *same* side of the normal. By appropriately designing the crystal, one can obtain single-beam negative refraction for *all* incident angles in a certain frequency range.[39] This phenomenon has excited interest because it is analogous (but not identical) to the behavior of a homogeneous material with a negative index of refraction, first studied by Veselago (1968). In a homogeneous material, a negative index arises when both ε and μ are negative, and gives rise to a number of interesting effects such as near-field imaging through a flat "lens." Although no natural material has a negative index, proposals by Pendry et al. (1999) and Smith et al. (2000) demonstrated how to approximately construct one at microwave frequencies by a composite of tiny metallic resonant structures. Pendry's design involves a periodic structure, but it operates at a wavelength much larger than the periodicity, allowing it to be accurately approximated by an effective homogeneous medium with appropriately "averaged" ε and μ.[40] All-dielectric negative refraction, such as in figure 14, is conceptually quite different: here, the wavelength is comparable to the periodicity and so we cannot accurately describe *all* the behavior of the medium by a *single* effective "index."[41] Nevertheless, one can emulate selected behaviors, such as negative refaction, of a true negative-index medium—perhaps the most tantalizing of which is the possibility of subwavelength imaging, which in an all-dielectric structure requires careful engineering of surface states in addition to the refracted waves.[42] The ultimate degree to which such behaviors of a theoretical negative-index medium can be emulated in practice, taking into account limitations such as the finite size and material absorption, remains an intriguing open question.

[38] Supercollimation was proposed by Kosaka et al. (1999*a*), and has been subsequently realized in two and three dimensions by numerous groups (Wu et al., 2003; Prather et al., 2004; Shin and Fan, 2005; Lu et al., 2006; Rakich et al., 2006).

[39] Early examples of such designs include Notomi (2000) and Luo et al. (2002*b*).

[40] This and subsequent work is reviewed in Smith et al. (2004).

[41] One can, of course, *define* an effective index, for example by using the refraction angle or a phase velocity (e.g., Dowling and Bowden, 1994), or from transmission/reflection spectra (as reviewed in, e.g., Smith et al., 2005). (Using the phase velocity requires a choice of reciprocal lattice vector, as discussed in the section Bloch-Wave Propagation Velocity of chapter 3.) Such definitions, while often useful, must be employed with care outside the long-wavelength limit—even if an effective index is defined so as to correctly predict one behavior, such as the refraction angle, it will typically not simultaneously predict other behaviors, such as reflection coefficients.

[42] The ideal realization of such subwavelength imaging, with theoretical negative-index materials, was described in Pendry (2000) and subsequently generalized in Pendry and Ramakrishna (2003). The role of surface states in all-dielectric photonic-crystal imaging was analyzed in Luo et al. (2003).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Further Reading

We have provided pointers to the literature for many of the examples used in this chapter. For a general introduction to traditional optoelectronics, see Yariv (1997). The classic reference on temporal coupled mode theory is the book by Hermann Haus (1984, ch. 7), which applies it mainly in the context of microwave devices. A generalization, with a somewhat more abstract approach, can be found in Suh et al. (2004).

Epilogue

Throughout most of the text, we have emphasized the basic physical principles underlying photonic crystals. By examining a few fundamental components in detail, we also hoped to convey the practical importance and versatility of photonic-crystal technology. We believe that these examples are only the tip of an iceberg of possibilities that are yet to be discovered. Our goal in writing this text was to stimulate the imaginations of researchers in diverse fields who might benefit from the possibilities afforded by photonic crystals and turn them into realities.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Appendix A

Comparisons with Quantum Mechanics

THROUGHOUT THE TEXT, especially in chapters 2 and 3, we make several comparisons between our formalism and the equations of quantum mechanics and solid-state physics. In this appendix, we present an extensive listing of these comparisons. It will hopefully serve as both a brief summary of the phenomena surrounding photonic crystals, and a way to relate them to (perhaps) familiar concepts in other fields.

The heart of the subject of photonic crystals is the propagation of electromagnetic waves in a periodic dielectric medium. In a sense, quantum mechanics is also the study of wave propagation, although the waves are a bit more abstract. At the atomic scale, particles (like the electron) begin to display wavelike properties, including interference and nonlocalization. The function that contains the information about the particle obeys the Schrödinger equation, which bears some resemblance to a familiar wave equation.

It therefore comes as no surprise that the study of quantum mechanics in a periodic potential contains direct parallels to our study of electromagnetism in a periodic dielectric. Since the quantum mechanics of periodic potentials is the basic theory of solid-state physics, the field of photonic crystals can also inherit some of the theorems and terminology of solid-state physics, in slightly modified form. Table 1 lists some of these correspondences.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

November 13, 2007 Time: 05:17pm appendixa.tex

Table 1

	Quantum Mechanics in a Periodic Potential (Crystal)	*Electromagnetism in a Periodic Dielectric (Photonic Crystal)*
What is the "key function" that contains all of the information?	The scalar wave function, $\Psi(\mathbf{r},t)$.	The magnetic vector field $\mathbf{H}(\mathbf{r},t)$.
How do we separate the time dependence of the function from the spatial dependence?	Expand in a set of energy eigenstates: $\Psi(\mathbf{r},t)=\sum_E c_E\Psi_E(\mathbf{r})e^{-iEt/\hbar}$.	Expand in a set of harmonic modes. (frequency eigenstates): $\mathbf{H}(\mathbf{r},t)=\sum_\omega c_\omega\mathbf{H}_\omega(\mathbf{r})e^{-i\omega t}$.
What is the "master equation" that determines the eigenstates of the system?	The Schrödinger equation: $\left[-\frac{\hbar^2}{2m}\nabla^2+V(\mathbf{r})\right]\Psi_E(\mathbf{r})=E\Psi_E(\mathbf{r})$.	The Maxwell equations: $\nabla\times\frac{1}{\varepsilon(\mathbf{r})}\nabla\times\mathbf{H}_\omega(\mathbf{r})=\frac{\omega^2}{c^2}\mathbf{H}_\omega(\mathbf{r})$.
Are there any other conditions on the key function?	Yes, the scalar field must be normalizable.	Yes, the vector field must be both normalizable and transverse: $\nabla\cdot\mathbf{H}=0$.
Where does the periodicity of the system enter?	The potential: $V(\mathbf{r})=V(\mathbf{r}+\mathbf{R})$, for all lattice vectors $\mathbf{R}$.	The dielectric function: $\varepsilon(\mathbf{r})=\varepsilon(\mathbf{r}+\mathbf{R})$, for all lattice vectors $\mathbf{R}$.
Is there any interaction between normal modes?	Yes, there is an electron–electron repulsive interaction that makes large-scale computation difficult.	In the linear regime, electromagnetic modes do not interact, and can be calculated independently.
What important properties do the normal modes have in common?	Eigenstates with different energies are orthogonal, have real eigenvalues, and can be found through a variational principle.	Modes with different frequencies are orthogonal, have nonnegative real eigenvalues, and can be found through a variational principle.
What are the properties of the master equation that guarantee these properties of the normal modes?	The Hamiltonian, $\hat{H}$, is a linear Hermitian operator.	The Maxwell operator, $\hat{\Theta}$, is a linear positive-semidefinite Hermitian operator.

Extended comparison between quantum mechanics in a periodic potential and electromagnetism in a periodic dielectric

What is the variational theorem that is used to determine the normal modes and frequencies?	$E_{\text{var}} = \frac{(\Psi,\hat{H}\Psi)}{(\Psi,\Psi)}$ is minimized when Ψ is an eigenstate of $\hat{H}$.	$U_{\text{var}} = \frac{(\mathbf{H},\hat{\Theta}\mathbf{H})}{(\mathbf{H},\mathbf{H})}$ is minimized when $\mathbf{H}$ is an eigenstate of $\hat{\Theta}$.
What is the heuristic that goes along with the variational theorem?	The wave function concentrates in potential wells, without oscillating too fast, while remaining orthogonal to lower-energy states.	The electromagnetic fields concentrate their energy in high-ε regions, without oscillating too fast, while remaining orthogonal to lower-frequency modes.
What is the physical energy of the system?	The eigenvalue E of the Hamiltonian.	The time-average electromagnetic energy: $U = \frac{1}{4}\int d^3\mathbf{r}\,\left(\varepsilon_0\varepsilon|\mathbf{E}|^2 + \mu_0|\mathbf{H}|^2\right)$.
Is there a natural length scale to the system?	Usually. Physical constants such as the Bohr radius set the length scale.	No. Solutions are generally scale-free.
What is the mathematical statement that says "A is a symmetry of the system"?	$\hat{A}$ commutes with the Hamiltonian: $[\hat{A},\hat{H}] = 0$.	$\hat{A}$ commutes with the Maxwell operator: $[\hat{A},\hat{\Theta}] = 0$.
How do we use the symmetries of a system to classify the eigenstates?	Distinguish them by how they transform under a symmetry operation $\hat{A}$.	Distinguish them by how they transform under a symmetry operation $\hat{A}$.
If a system has discrete translational symmetry (as a crystal does), then how can the modes be classified?	By wave vector $\mathbf{k}$. Write the wave function in Bloch form: $\Psi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$.	By wave vector $\mathbf{k}$. Write the harmonic modes in Bloch form: $\mathbf{H}_{\mathbf{k}}(\mathbf{r}) = \mathbf{u}_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$.
What are the nonredundant values for the wave vector $\mathbf{k}$?	They lie in the irreducible Brillouin zone in reciprocal space.	They lie in the irreducible Brillouin zone in reciprocal space.

continued on next page

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

November 13, 2007 Time: 05:17pm appendixa.tex

Table 1

	Quantum Mechanics in a Periodic Potential (Crystal)	*Electromagnetism in a Periodic Dielectric (Photonic Crystal)*
What do we mean by the term "band structure?"	The functions $E_n(\mathbf{k})$, a set of continuous functions that specify the energies of the eigenstates.	The functions $\omega_n(\mathbf{k})$, a set of continuous functions that specify the frequencies of the harmonic modes.
What is the physical origin of the band structure?	The electron wave scatters coherently from the different potential regions.	The electromagnetic fields scatter coherently at the interfaces between different dielectric regions.
What property characterizes a "gap" in the band structure?	Within that range of energies, there are no propagating electron states, regardless of wave vector.	Within that range of frequencies, there are no propagating electromagnetic modes, regardless of wave vector or polarization.
What are the terms for the bands that are immediately above and below a gap?	The band above the gap is the **conduction band**. The band below the gap is the **valence band**.	The band above the gap is the **air band**. The band below the gap is the **dielectric band**.
How are defects introduced into the system?	By adding foreign atoms to the crystal, thereby breaking the translational symmetry of the atomic potential.	By changing the dielectric constant at particular locations, thereby breaking the translational symmetry of the dielectric function.
What is a possible result of introducing a defect?	It may create an allowed state within a band gap, thereby permitting a localized electron state to exist in the vicinity of the defect.	It may create an allowed state within a band gap, thereby permitting a localized electromagnetic mode to exist in the vicinity of the defect.
How do we classify different types of defects?	*Donor atoms* pull states from the conduction band down into the gap. *Acceptor atoms* push states from the valence band up into the gap.	*Dielectric defects* pull states from the air band down into the gap. *Air defects* push states from the dielectric band up into the gap.
In short, why is studying the physics of the system important?	We can tailor the *electronic* properties of materials to our needs.	We can tailor the *optical* properties of materials to our needs.

Continued from previous page

Appendix B

The Reciprocal Lattice and the Brillouin Zone

BEGINNING IN CHAPTER 4, we used Bloch's theorem to express an electromagnetic mode as a plane wave that is modulated by a periodic function $\mathbf{u}(\mathbf{r})$. The function $\mathbf{u}$ shares the same periodicity as the crystal. We also argued that we need only consider wave vectors $\mathbf{k}$ that lie in a certain region of the reciprocal lattice known as the *Brillouin zone*.

This is not news to anyone who has studied solid-state physics or other fields in which lattices play a large role. But for readers who have never encountered the Brillouin zone, this appendix will provide enough information to completely understand the material in this text. Specifically, we will introduce the reciprocal lattice and identify the Brillouin zone for some simple lattices that we use throughout the text. In addition, we describe the Miller index notation we use in chapter 6 to refer to crystal planes. For more details, it is best to consult the first few chapters of a solid-state physics text, such as Kittel (1996) or Ashcroft and Mermin (1976).

The Reciprocal Lattice

Suppose we have a function $f(\mathbf{r})$ that is periodic on a lattice; that is, suppose $f(\mathbf{r}) = f(\mathbf{r}+\mathbf{R})$ for all vectors $\mathbf{R}$ that translate the lattice into itself (i.e., connect one lattice point to the next). Our dielectric function $\varepsilon(\mathbf{r})$ is an example of such a function. The vectors $\mathbf{R}$ are called the **lattice vectors**.

A natural thing to do when analyzing periodic functions is to take the **Fourier transform**; that is, we build the periodic function $f(\mathbf{r})$ out of plane waves with various wave vectors. The expansion looks like this:

$$f(\mathbf{r}) = \int d^3\mathbf{q}\, g(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}}. \tag{1}$$

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Here $g(\mathbf{q})$ is the coefficient on the plane wave with wave vector $\mathbf{q}$. An expansion like this can be performed on *any* well-behaved function. But our function f is periodic on the lattice—what information does this tell us about the expansion? Requiring $f(\mathbf{r}) = f(\mathbf{r}+\mathbf{R})$ in the expansion yields

$$f(\mathbf{r}+\mathbf{R}) = \int d^3\mathbf{q}\, g(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}} e^{i\mathbf{q}\cdot\mathbf{R}} = f(\mathbf{r}) = \int d^3\mathbf{q}\, g(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}}. \tag{2}$$

The periodicity of f tells us that its Fourier transform $g(\mathbf{q})$ has the special property $g(\mathbf{q}) = g(\mathbf{q})\exp(i\mathbf{q}\cdot\mathbf{R})$. But this is impossible, unless either $g(\mathbf{q}) = 0$ or $\exp(i\mathbf{q}\cdot\mathbf{R}) = 1$. In other words, the transform $g(\mathbf{q})$ is zero everywhere, except for spikes at the values of $\mathbf{q}$ such that $\exp(i\mathbf{q}\cdot\mathbf{R}) = 1$ for all $\mathbf{R}$.

What we have just discovered is that if we are building a lattice-periodic function f out of plane waves, we need only use those plane waves with wave vectors $\mathbf{q}$ such that $\exp(i\mathbf{q}\cdot\mathbf{R}) = 1$ for all of the lattice vectors $\mathbf{R}$. This statement has a simple analogue in one dimension: if we are building a function $f(x)$ with period τ out of sinusoids, then we need only use the "fundamental" sinusoid with period τ and its "harmonics" with periods $\tau/2$, $\tau/3$, $\tau/4$ and so on, forming a **Fourier series**.

Those vectors $\mathbf{q}$ such that $\exp(i\mathbf{q}\cdot\mathbf{R}) = 1$, or, equivalently, $\mathbf{q}\cdot\mathbf{R} = 2\pi N$ (N an integer), are called **reciprocal lattice vectors** and are usually designated by the letter $\mathbf{G}$. They form a lattice of their own; for example, adding two reciprocal lattice vectors $\mathbf{G}_1$ and $\mathbf{G}_2$ yields another reciprocal lattice vector, as you can easily verify. We can build our function $f(\mathbf{r})$ with an appropriate weighted sum over all of the reciprocal lattice vectors, as follows:

$$f(\mathbf{r}) = \sum_{\mathbf{G}} f_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}. \tag{3}$$

Constructing the Reciprocal Lattice Vectors

Given a lattice with a set of lattice vectors $\mathbf{R}$, how can we determine all of the reciprocal lattice vectors $\mathbf{G}$? We need to find all $\mathbf{G}$ such that $\mathbf{G}\cdot\mathbf{R}$ is some integer multiple of 2π for every $\mathbf{R}$.

We know that every lattice vector $\mathbf{R}$ can be written in terms of the **primitive lattice vectors**, which are the smallest vectors pointing from one lattice point to another. For example, on a simple cubic lattice with spacing a, the vectors $\mathbf{R}$ would all be of the form $\mathbf{R} = \ell a\hat{\mathbf{x}} + ma\hat{\mathbf{y}} + na\hat{\mathbf{z}}$, where (ℓ, m, n) are integers. In general, we call the primitive lattice vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$. They need not be of unit length.

We have already mentioned that the reciprocal lattice vectors $\{\mathbf{G}\}$ form a lattice of their own. In fact, the reciprocal lattice has a set of primitive vectors $\mathbf{b}_i$ as well, so that every reciprocal lattice vector $\mathbf{G}$ can be written as $\mathbf{G} = \ell\mathbf{b}_1 + m\mathbf{b}_2 + n\mathbf{b}_3$.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Our requirement that $\mathbf{G} \cdot \mathbf{R} = 2\pi N$ boils down to the primitive requirement

$$\mathbf{G} \cdot \mathbf{R} = (\ell \mathbf{a}_1 + m\mathbf{a}_2 + n\mathbf{a}_3) \cdot (\ell' \mathbf{b}_1 + m' \mathbf{b}_2 + n' \mathbf{b}_3) = 2\pi N. \tag{4}$$

For all choices of (ℓ, m, n), the above must hold for some N. A little thought will suggest that we could satisfy the above if we construct the $\mathbf{b}_i$ so that $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi$ if $i = j$, and 0 if $i \neq j$. More compactly, we write $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$. Given the set $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$, our task is to find the corresponding set $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ such that $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$.

One way to do this is to exploit a feature of the cross product. Remembering that $\mathbf{x} \cdot (\mathbf{x} \times \mathbf{y}) = 0$ for any vectors $\mathbf{x}$ and $\mathbf{y}$, we can construct the primitive reciprocal lattice vectors with the following recipe:

$$\mathbf{b}_1 = \frac{2\pi\, \mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}, \quad \mathbf{b}_2 = \frac{2\pi\, \mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}, \quad \mathbf{b}_3 = \frac{2\pi\, \mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}. \tag{5}$$

More compactly, if A is the 3×3 matrix whose columns are $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ and B is the 3×3 matrix whose columns are $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$, then B is 2π times the inverse of the transpose of A.

In summary, when we take the Fourier transform of a function that is periodic on a lattice, we need only include terms with wave vectors that are reciprocal lattice vectors. To construct the reciprocal lattice vectors, we take the primitive lattice vectors and perform the operations of equation (5).

The Brillouin Zone

In chapter 3, we found that the discrete translational symmetry of a photonic crystal allows us to classify the electromagnetic modes with a wave vector $\mathbf{k}$. The modes can be written in "Bloch form," consisting of a plane wave modulated by a function that shares the periodicity of the lattice:

$$\mathbf{H}_\mathbf{k}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{u}_\mathbf{k}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{u}_\mathbf{k}(\mathbf{r} + \mathbf{R}). \tag{6}$$

One important feature of the Bloch states is that different values of $\mathbf{k}$ do not necessarily lead to different modes. Specifically, a mode with wave vector $\mathbf{k}$ and a mode with wave vector $\mathbf{k} + \mathbf{G}$ are the same mode, if $\mathbf{G}$ is a reciprocal lattice vector. The wave vector $\mathbf{k}$ serves to specify the phase relationship between the various cells that are described by $\mathbf{u}$. If $\mathbf{k}$ is incremented by $\mathbf{G}$, then the phase between cells is incremented by $\mathbf{G} \cdot \mathbf{R}$, which we know is $2\pi N$ and not really a phase difference at all. So incrementing $\mathbf{k}$ by $\mathbf{G}$ results in the same physical mode.

This means that there is a lot of redundancy in the label $\mathbf{k}$. We can restrict our attention to a finite zone in reciprocal space in which you *cannot* get from one part of the volume to another by adding any $\mathbf{G}$. All values of $\mathbf{k}$ that lie outside of this zone, by definition, can be reached from within the zone by adding $\mathbf{G}$, and are

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

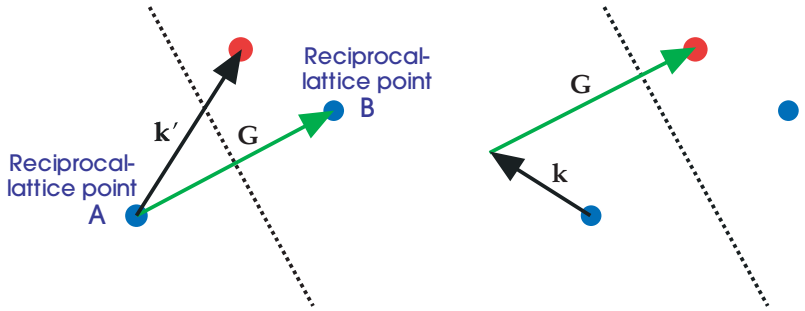


Figure 1: Characterization of the Brillouin zone. The dotted line is the perpendicular bisector of the line joining two reciprocal lattice points (blue). If we choose the left point as the origin, any lattice vector (such as $\mathbf{k}'$) that reaches to an arbitrary point on the other side (red) can be expressed as the sum of a same-side vector (such as $\mathbf{k}$) plus a reciprocal lattice vector $\mathbf{G}$.

therefore redundant labels. There are actually many such zones, but we focus on the region that is closest to $\mathbf{k} = 0$.

This zone is the (first) **Brillouin zone**.[1] A more visual way to characterize it is the following: around any lattice point in reciprocal space, highlight the volume that is closer to *that* lattice point than to *any other* lattice point. If we call the original lattice point the origin, then the highlighted region is the Brillouin zone.

The two definitions are equivalent. If a particular $\mathbf{k}$ is closer to a neighboring lattice point, you can always reach it by staying close to the original lattice vector and then translating by the $\mathbf{G}$ that reaches from one lattice point to the other. This situation is depicted in figure 1.

The next few sections will be devoted to a study of the reciprocal lattice vectors and Brillouin zones of some of the particular lattices that appear in this text.

Two-Dimensional Lattices

In chapter 5, we worked extensively with photonic crystals that are based on a square or triangular lattice. What are the reciprocal lattice vectors and Brillouin zones of each?

For a **square lattice** with spacing a, the lattice vectors are $\mathbf{a}_1 = a\hat{\mathbf{x}}$ and $\mathbf{a}_2 = a\hat{\mathbf{y}}$. To use our prescription (5), we can use a third basis vector $\mathbf{a}_3$, in the z-direction of any length, since the crystal is homogeneous in that direction. The results are $\mathbf{b}_1 = (2\pi/a)\hat{\mathbf{x}}$ and $\mathbf{b}_2 = (2\pi/a)\hat{\mathbf{y}}$. The reciprocal lattice is also a square lattice, but with spacing $2\pi/a$ instead of a. The name "reciprocal lattice" suits this fact well.

[1] The second, third, etc., Brillouin zones are regions farther away from the $\mathbf{k} = 0$ origin.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

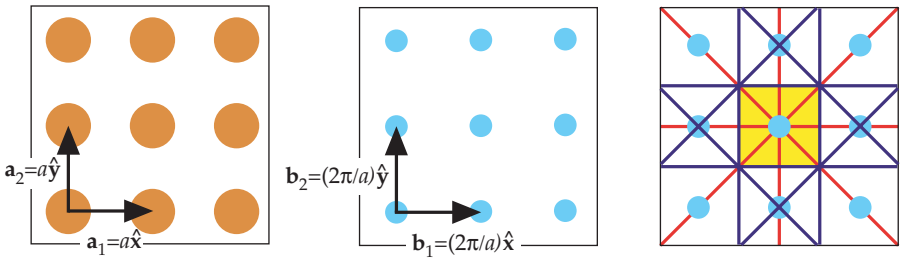


Figure 2: The square lattice. On the left is the network of lattice points in real space. In the middle is the corresponding reciprocal lattice. On the right is the construction of the first Brillouin zone: taking the center point as the origin, we draw the lines connecting the origin to the other lattice points (red), their perpendicular bisectors (blue), and highlight the square boundary of the Brillouin zone (yellow).

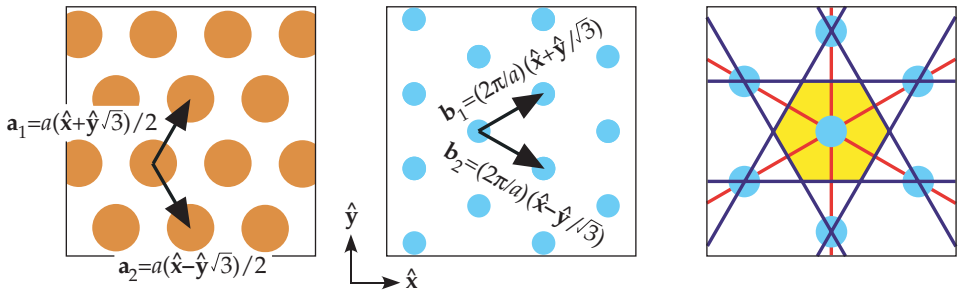


Figure 3: The triangular lattice. On the left is the network of lattice points in real space. In the middle is the corresponding reciprocal lattice, which in this case is a rotated version of the original. On the right is the Brillouin zone construction. In this case, the first Brillouin zone is a hexagon centered around the origin.

To determine the Brillouin zone, we fix our attention on a particular lattice point (the origin) and shade the area that is closer to that point than to any other. Geometrically, we draw perpendicular bisectors of every lattice vector that starts at the origin. Each bisector divides the lattice into two half-planes, one of which contains our lattice point. The intersection of all the half-planes that contain our lattice point is the Brillouin zone. The square lattice vectors, its reciprocal lattice vectors, and its Brillouin zone are shown in figure 2.

We can handle the **triangular lattice** in a similar fashion. The lattice vectors are $(\hat{\mathbf{x}}+\hat{\mathbf{y}}\sqrt{3})a/2$ and $(\hat{\mathbf{x}}-\hat{\mathbf{y}}\sqrt{3})a/2$, as shown in figure 3. Using our prescription (5), we obtain reciprocal lattice vectors $(2\pi/a)(\hat{\mathbf{x}}+\hat{\mathbf{y}}/\sqrt{3})$ and $(2\pi/a)(\hat{\mathbf{x}}-\hat{\mathbf{y}}/\sqrt{3})$. This is again a triangular lattice, but rotated by 90° with respect to the first, and with

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

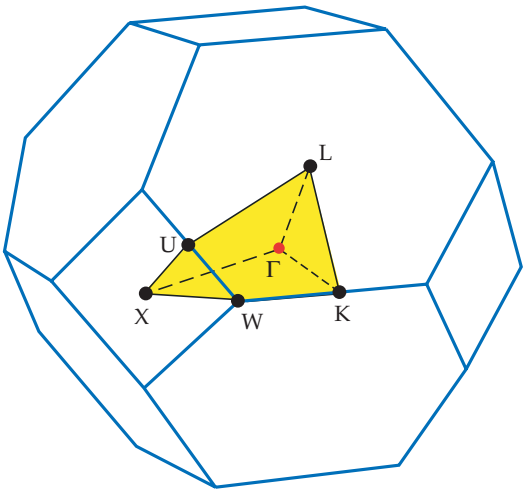


Figure 4: The Brillouin zone for the face-centered cubic (fcc) lattice. The reciprocal lattice is a body-centered cubic (bcc) lattice, and the Brillouin zone is a truncated octahedron with center at Γ. Also shown are some of the labels that are traditionally given to the special points in the zone. The irreducible Brillouin zone is the yellow polyhedron with corners at Γ, X, U, L, W, and K.

spacing $4\pi/a\sqrt{3}$. The Brillouin zone, as determined by the construction outlined above, is a hexagon.

Three-Dimensional Lattices

Our construction works in three dimensions as well, with the proper extensions. Instead of drawing perpendicular bisectors to the lattice vectors, draw perpendicular *planes* that bisect the line. Then, the Brillouin zone is the intersection of all of the half-spaces that contain the origin.

The one lattice we devote the most attention to in the text is the **face-centered cubic** (fcc) lattice, as in figure 1 of chapter 6. This is like a cubic lattice, but with an additional lattice point at each face of the cubes. The primitive lattice vectors are $\mathbf{a}_1 = (\hat{\mathbf{x}} + \hat{\mathbf{y}})a/2$, $\mathbf{a}_2 = (\hat{\mathbf{y}} + \hat{\mathbf{z}})a/2$ and $\mathbf{a}_3 = (\hat{\mathbf{x}} + \hat{\mathbf{z}})a/2$.

Using our recipe (5), we find that the reciprocal lattice vectors form a *body*-centered cubic (bcc) lattice, which is like a cubic lattice with an additional lattice point at the *center* of every cube. The primitive reciprocal lattice vectors are $\mathbf{b}_1 = (2\pi/a)(\hat{\mathbf{x}} + \hat{\mathbf{y}} - \hat{\mathbf{z}})$, $\mathbf{b}_2 = (2\pi/a)(-\hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}})$, and $\mathbf{b}_3 = (2\pi/a)(\hat{\mathbf{x}} - \hat{\mathbf{y}} + \hat{\mathbf{z}})$.

The Brillouin zone, which can be determined in the usual fashion, is a truncated octahedron. It is depicted in figure 4, along with the traditional names for the special points in the irreducible Brillouin zone.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

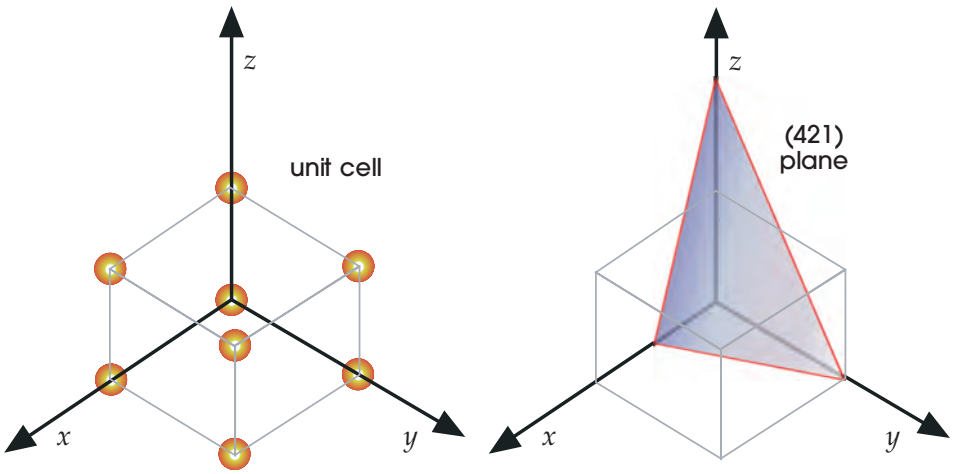


Figure 5: The Miller indices for a crystal plane. *Left:* A plot of the unit cell. The length of the unit cell in each direction is the lattice constant for that direction. *Right:* The shaded plane is named by locating the intercepts of the plane with the axes, as described in the text.

Miller Indices

We return to the real lattice, as opposed to the reciprocal lattice, for one last miscellaneous topic. It is often convenient to have a systematic way of referring to directions and planes in a crystal. For example, in chapter 6 we refer to various cross sections and directions in fcc lattices. The traditional way to refer to planes in a crystal lattice is the system of **Miller indices** (see, e.g., Ashcroft and Mermin, 1976; Kittel, 1996). To characterize a plane, we need to specify three noncollinear points in the plane. The Miller indices of a crystal plane are integers that give the location of three such points, relative to the lattice vectors of the crystal. (For the fcc, bcc, and diamond lattices, however, it is conventional to define the Miller indices with respect to the lattice vectors of the simple-cubic supercell as in figure 4 of chapter 6.)

Suppose the crystal has lattice vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$. Then the Miller indices are three integers ℓ, m, and n, written as (ℓmn). They denote a a plane that intersects the three points: $\mathbf{a}_1/\ell$, $\mathbf{a}_2/m$, and $\mathbf{a}_3/n$ (or some multiple thereof). That is, the Miller indices are proportional to the *inverses* of the *intercepts* of the plane with the lattice vectors. Let us consider an example.

The unit cell of the crystal can be plotted with $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ along its edges. To name a particular plane in the crystal, we draw the plane on such a plot; in general, the plane will intersect each of the axes $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$. An example of such a plot is shown in figure 5, for the special case $\mathbf{a}_1 = a\hat{\mathbf{x}}$, $\mathbf{a}_2 = a\hat{\mathbf{y}}$, $\mathbf{a}_3 = a\hat{\mathbf{z}}$. We would like to name a surface that intersects the x-axis at one-half lattice constant

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

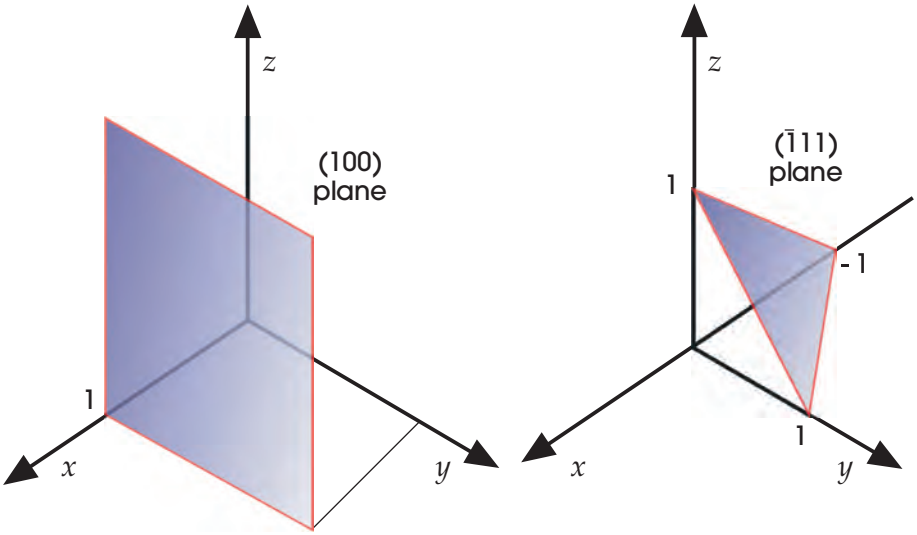


Figure 6: Special cases of the Miller index system. *Left:* The (100) plane of a simple-cubic lattice. The plane never intersects the y or z axes, so the corresponding indices are zero. *Right:* The $(\bar{1}11)$ plane of the simple-cubic lattice. The plane intersects the x axis at -1, so a bar is placed over the corresponding index.

away from the origin, the y-axis one lattice constant away, and the z-axis two lattice constants away.

First, we take the reciprocals of these intercepts: 2, 1, and $1/2$. Next, we find the smallest integers that have the same ratios as these three: 4, 2, and 1. These integers are the Miller indices, and we refer to this plane as the (421) plane.

There are two common complications. First, if the surface is parallel to a particular lattice vector, then it will never intercept that axis. The intercept is infinite, so to speak, and the reciprocal is zero. The Miller index for that direction is zero. Figure 6 shows an example of this case, for the (100) plane of a cubic lattice.

Second, if the plane intercepts an axis at a negative value, the Miller index for that direction is negative. Traditionally, this is indicated by placing a bar over a positive integer instead of using a minus sign. For example, the $(\bar{1}11)$ plane of the cubic lattice is shown in figure 6.

There is another, completely equivalent, interpretation of the Miller indices that is sometimes more convenient. The indices (ℓmn) denote the planes *perpendicular* to the direction $\ell\mathbf{b}_1 + m\mathbf{b}_2 + n\mathbf{b}_3$, where $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$ are the primitive reciprocal lattice vectors. That is, the Miller indices give the components of the normal to the planes in the basis of the reciprocal lattice vectors. By convention, we always divide the Miller indices by any common factor, so they correspond to the shortest reciprocal lattice vector normal to the planes.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

In a related notation, the expression $[\ell mn]$ (square brackets) denotes the *direction* $\ell\mathbf{a}_1 + m\mathbf{a}_2 + n\mathbf{a}_3$ (i.e., in the basis of the lattice vectors). The direction $[\ell mn]$ is *not* generally perpendicular to the (ℓmn) plane, except in an important special case: lattices with cubic symmetry.

For lattices with cubic symmetry, including fcc and bcc as well as the simple cubic lattice, the Miller indices are conventionally defined in terms of the cubic lattice vectors. In this case, both $[\ell mn]$ and the normal to (ℓmn) are simply the (ℓ, m, n) direction in ordinary Cartesian coordinates.

The families of parallel planes defined by integer Miller indices are known as **lattice planes**, and are the only planes whose intersection with the crystal is (two-dimensionally) periodic. It may seem that we have left out many possible planes by restricting the indices (ℓmn) to be integers—what if they are real numbers with *irrational* ratios? For such irrational Miller indices, the resulting intersection with the crystal is a non-periodic pattern known as a **quasicrystal** (**?**). Quasicrystals are especially interesting when they occupy the whole physical space (not just a crystal termination). For example, a particular irrational two-dimensional slice of a *five*-dimensional hypercubic lattice yields a famous pattern known as a **Penrose tiling**.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Appendix C

Atlas of Band Gaps

In this appendix, we will chart the locations and sizes of the photonic band gaps in several different two and three-dimensional photonic crystals. We hope this will serve two purposes. First, to demonstrate that the engineering possibilities with photonic crystals are enormous; the class of possible photonic crystals is so large that designing crystals with gaps in desirable locations is a real possibility. Second, to serve as a convenient atlas of easily fabricated crystals, in case one of these crystals suits a particular application.

A plot of the locations of the photonic band gaps of a crystal, as one or more of the parameters of the crystal are varied, is what we call a **gap map**. In this appendix we will present the gap maps for the square and triangular lattices of cylindrical columns as the radius of the columns varies. We will do this both for dielectric columns in air, and for air columns in dielectric, both with a dielectric contrast of 11.4.[1] We will plot the TE and TM polarizations together.

Along the horizontal axis of a gap map is the radius of the columns; along the vertical axis is the frequency (in dimensionless units). The locations of the band gaps are outlined for both the TE and TM polarizations. To use a gap map to find a crystal for a specific application, the frequency and the lattice constant must be scaled to the desired levels, using the scaling laws developed in chapter 3.

The gap maps show the gaps as a function of radius for a fixed ε ratio of 11.4:1, or an index contrast of 3.38:1. One can also ask the converse question, however: what is the *optimal* radius that *maximizes* the gap size as a *function* of the index contrast? Also, what is the *minimum* index contrast required to obtain a gap for a given geometry? We answer these questions with a second sort of plot, which we call a **gap recipe**: a plot of the optimal radius and the corresponding gap size as a function of index contrast. This appendix presents gap recipes for the most common two-dimensional crystals, for TE, TM, and complete gaps. We also show

[1] This is the dielectric contrast between gallium arsenide (GaAs) and air, for wavelengths around 1.5 μm (Palik, 1998).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

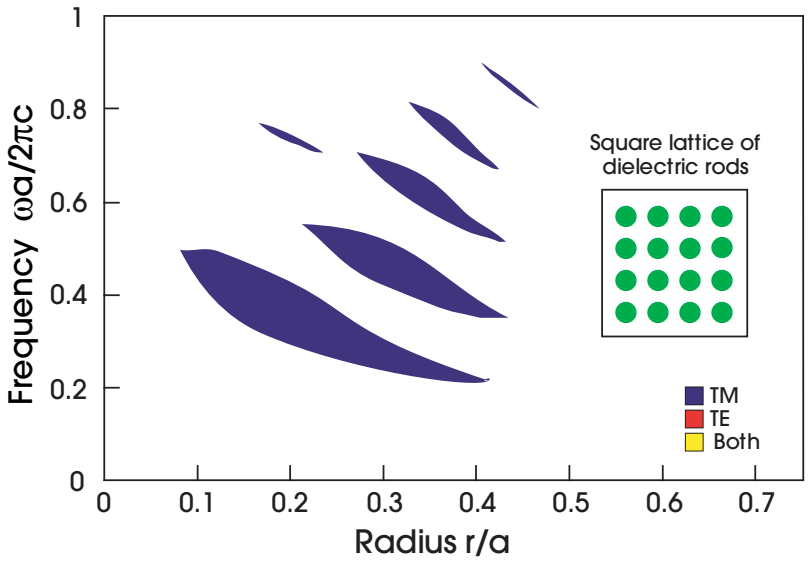


Figure 1: Gap map for a square lattice of dielectric rods, $\varepsilon = 11.4$.

gap recipes for the simplest (although not the most practical) three-dimensional crystals, a diamond lattice of dielectric spheres or holes (as in the subsection Spheres in a diamond lattice of chapter 6).

A Guided Tour of Two-Dimensional Gaps

The first two-dimensional photonic crystal that we will consider consists of parallel columns arranged in a square lattice. The columns have radius r, and the lattice constant is a. We first consider the case of dielectric $\varepsilon = 11.4$ columns in air. The gap map is shown in figure 1.

At a glance, the gap map reveals some interesting regularities. First, the gaps all decrease in frequency as r/a increases. This is an expected feature, since the frequency scales as $1/\sqrt{\varepsilon}$ in a medium of dielectric constant ε and, as r/a increases, the average dielectric constant of the medium steadily increases. Second, even though the plot extends all the way to $r/a = 0.70$, for which the dielectric columns fill space, all of the gaps seal up by the time $r/a = 0.50$. At that value, the dielectric columns begin to touch one another. The third and perhaps most remarkable feature is the repetition of the lowest, largest gap at higher frequencies; progressively smaller copies of the lowest gap are stacked above it at roughly equal intervals. For $r/a = 0.38$, there are four TM gaps in the band structure!

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

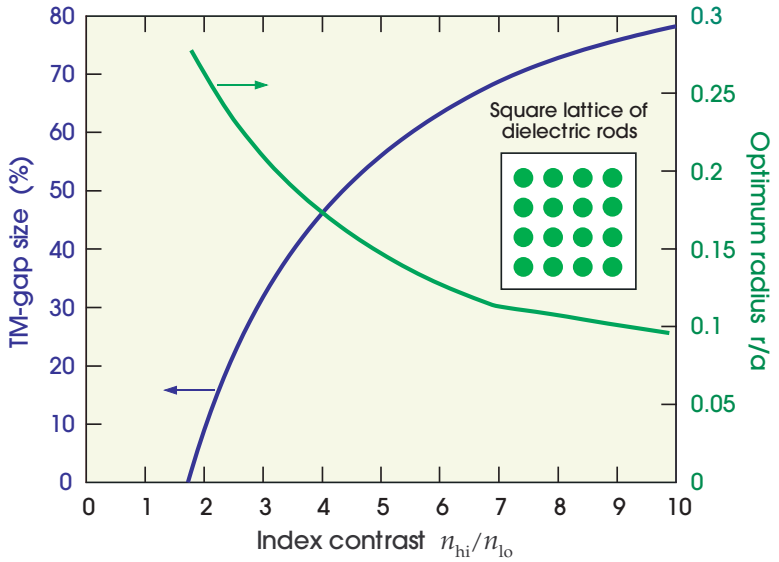


Figure 2: Gap recipe for the lowest TM gap of the square lattice of dielectric rods.

Since the lowest gap is the largest one, in most cases this is the gap that is used. In figure 2, therefore, we plot the gap recipe for the lowest gap of this structure as a function of index contrast. As we might expect, the gap size increases monotonically with index contrast. The optimal radius, on the other hand, decreases with index contrast. This makes sense because the diameter of the rods should be on the order of a half wavelength in the dielectric material, and the relative wavelength in the material decreases proportional to the index contrast.

Moreover, there is a *minimum index contrast* of about 1.72:1 to obtain a TM gap with the square lattice of rods. (Compare with the one-dimensional gaps of chapter 4, where even an infinitesimal contrast opened a gap.) This was explained in the subsection Spheres in a diamond lattice of chapter 6: since a two-dimensional (or three-dimensional) crystal has different periodicity in different directions, the gaps in different directions tend not to overlap unless they are large enough, which requires a minimum index contrast. The more equal we can make the periodicity in different directions, corresponding to a *more circular* Brillouin zone, the easier it generally becomes to create a gap. We return to this issue with triangular lattices, below.

For the TE polarization, however, figure 1 shows a much more barren terrain. There are no significant TE gaps at all for the square lattice in the frequency range displayed. Our results are in agreement with the heuristic of chapter 5, which holds that connectivity of high-ε regions is conducive to TE gaps, and isolated patches of high-ε regions lead to TM gaps.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

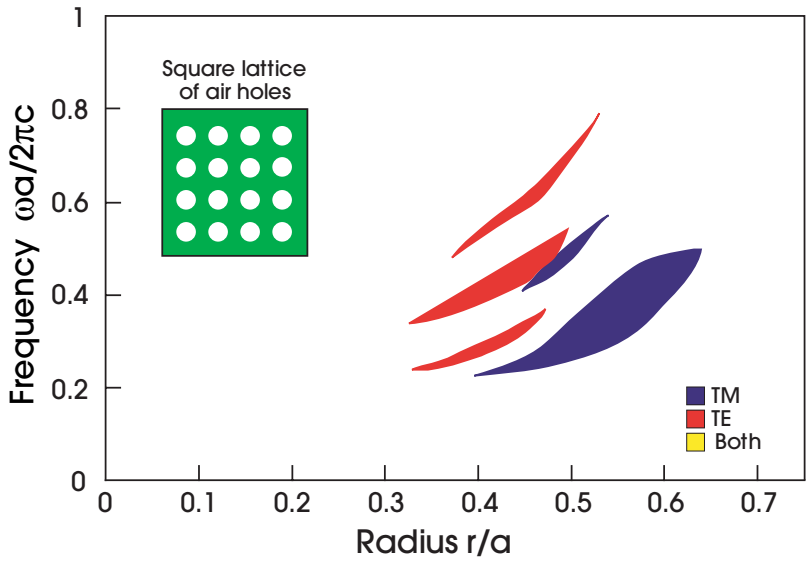


Figure 3: Gap map for a square lattice of air holes in dielectric ($\varepsilon = 11.4$).

Now we reverse the dielectric configuration, so that the columns of the square lattice have $\varepsilon = 1$, and the surrounding medium has $\varepsilon = 11.4$. The gap map for both polarizations is shown in figure 3. Immediately we see that frequencies increase with r/a this time, since the average dielectric constant decreases as the air columns grow. The gap structure for the TM modes seems to open up around $r/a = 0.45$, in contrast with the sharp cutoff seen for the dielectric columns. Apparently, connectivity between the air columns is of importance, since figure 3 displays a significant change of behavior around $r/a = 0.5$.

For the TE polarization, the square lattice of air columns fares a bit better than that of dielectric columns, as figure 3 attests. Several thin gaps may be noted. However, none of these gaps overlaps a gap for the TM polarization, so there is no *complete* band gap for the square lattice for this value of the dielectric contrast. We shall see that our next structure, the triangular lattice of columns, does possess such a complete band gap.

Our second structure, the triangular lattice of columns, is depicted in the inset in figure 4. Here the columns begin touching one another at $r/a = 0.50$, and fill space at $r/a = 0.58$. Again we can distinguish the cases of dielectric $(\varepsilon = 11.4)$ columns in air, and air columns in dielectric. We begin with dielectric columns in air.

Figure 4 displays the gap map. The remarkable self-similarity of the figure 1, which was for the TM polarization of the square lattice of dielectric columns, is mirrored here. The successive gaps are similar in shape and

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

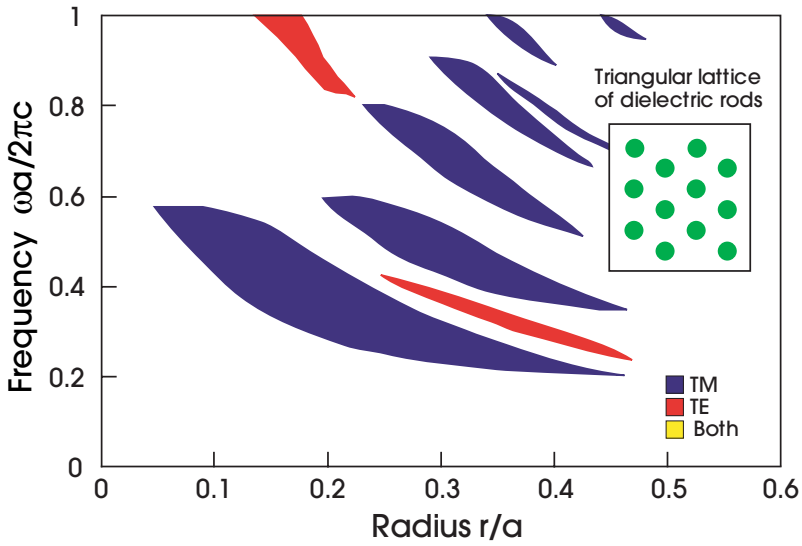


Figure 4: Gap map for a triangular lattice of dielectric rods ($\varepsilon = 11.4$).

orientation, and stack regularly upon one another. The cutoff at $r/a = 0.45$ is once again near the column-touching condition.

The gap map for the TE polarization is almost as sparse as the corresponding case of the square lattice. Only a few slivers are noticeable. The gross properties of the map (decrease in ω with r/a, transition at $r/a = 0.5$) follow the same trends as those already discussed.

On the other hand, the TM gaps for the triangular lattice of rods are visibly larger than those of the square lattice. The reason for this is that the triangular lattice has more symmetry, and in particular its Brillouin zone (a hexagon) is more circular than the Brillouin zone of a square lattice (a square). As discussed above, this makes it easier to create gaps. This is seen more quantitatively in figure 5, which shows the gap recipe for the lowest gap in this structure. Its band gaps are larger than those of the square lattice for the same dielectric contrast, and the minimum index contrast for a TM gap is now only 1.32:1.

Finally, we turn to the case of air columns in dielectric. The gap map is shown in figure 6. Although the TM gaps are hardly comparable in size to the gaps for dielectric columns, the lowest gap should be noted. It happens to occur at the same location as the lowest TE gap, thereby forming a complete band gap.

Just as the lattice of rods creates large TM gaps, the lattice of holes creates large TE gaps; this is made quantitative in figure 7, which gives the gap recipe for the lowest TE gap of the triangular hole lattice. The minimum index contrast to create a TE gap is only 1.39:1 in this structure. The resulting enormous TE gap at $\varepsilon = 11.4$, between $r/a = 0.20$ and 0.50 in figure 6, provides ample space for an overlap with the TM gap noted above. Thus, for r/a around 0.45, the triangular lattice of

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

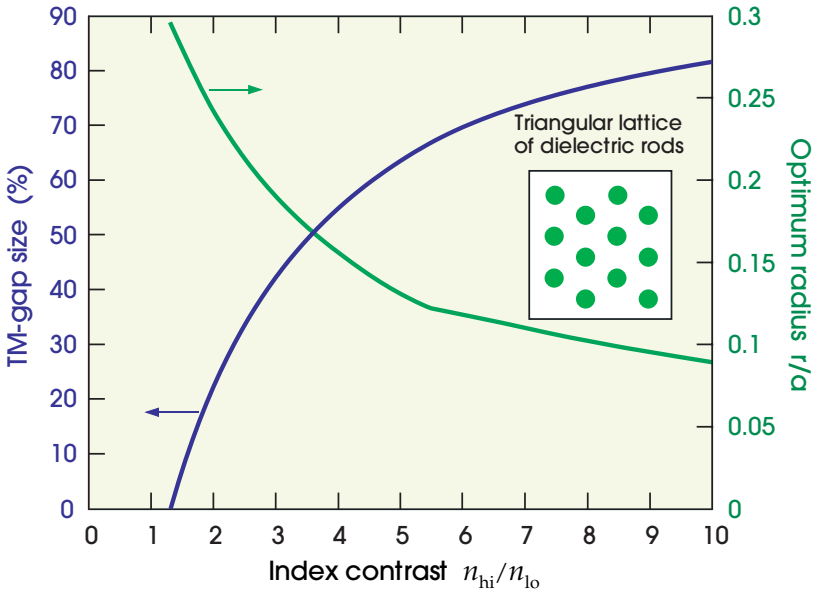


Figure 5: Gap recipe for the lowest TM gap of the triangular lattice of dielectric rods.

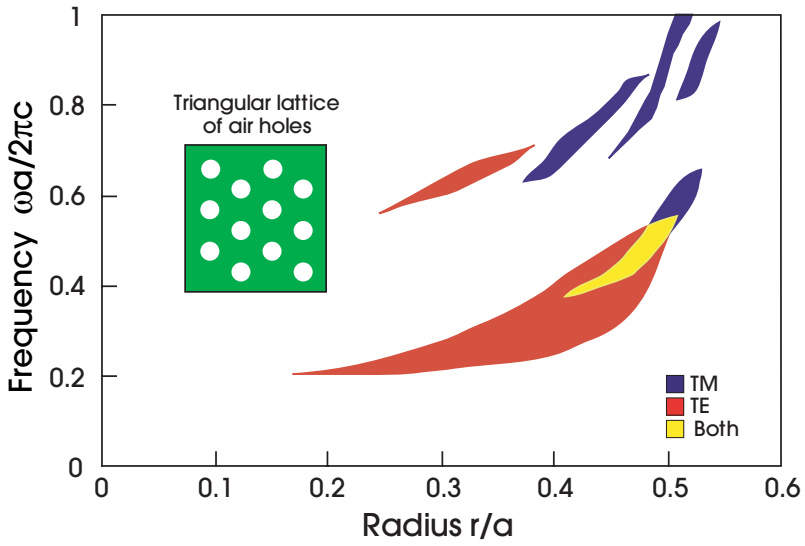


Figure 6: Gap map for a triangular lattice of air holes in dielectric ($\varepsilon = 11.4$).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

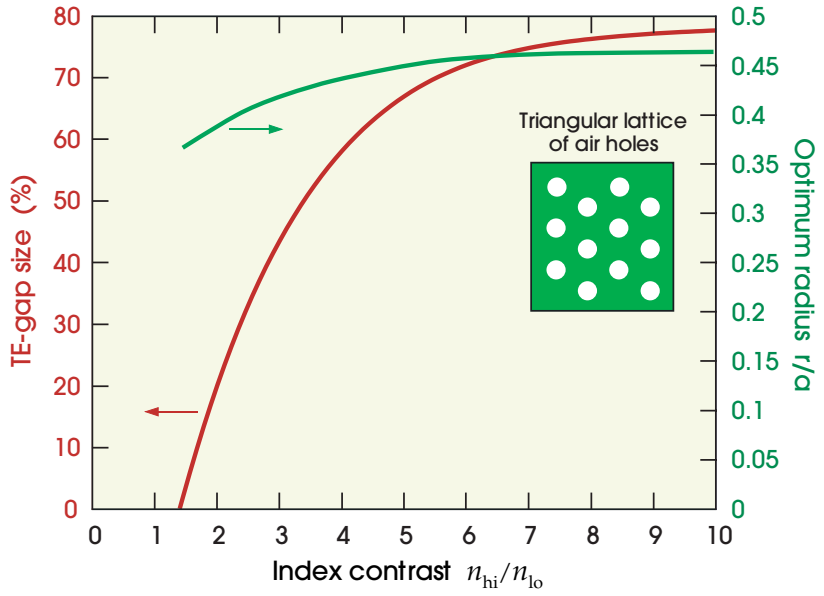


Figure 7: Gap recipe for the lowest TE gap of the triangular lattice of air holes in dielectric.

air columns possesses a complete band gap for all polarizations for frequencies around $0.45(2\pi c/a)$. This discovery was first reported in Meade et al. (1992) and Villeneuve and Piché (1992). Again, we quantify this complete gap by giving its gap recipe, plotting the optimal size and radius of the complete gap versus index contrast in figure 8. For the complete gap, the minimum index contrast is much larger: 2.63:1 in this structure.

Now that the survey is complete, we can assemble the highlights from our atlas of the square and triangular lattices. For dielectric columns in air, band gaps are most abundant for the TM polarization, as long as the columns are not large enough to touch one another. For air columns, the TE polarization has more band gaps, and there is even a complete band gap for the triangular lattice.

These are not the only interesting structures that exhibit gaps, of course. In particular, it may be desirable to have a two-dimensional structure that has an overlapping band gap for both polarizations and is easier to fabricate. Although we saw in figure 6 that the triangular lattice of air columns has such a gap, it occurs at a diameter of $d = 0.95a$, at a midgap frequency of $\omega a/2\pi c = 0.48$. Thus, this structure has very thin dielectric veins of width $0.05a$ between the air columns. In fact, to fabricate such a structure with a band gap at $\lambda = 1.5\,\mu$m would require a feature size of $0.035\,\mu$m. While such fine feature sizes may be fabricated using e-beam lithography, this is a relatively difficult procedure. Luckily, there are so many possibilities for the geometry of the lattice that suitable structures can be found. Consider, for instance, the "honeycomb lattice." The gap map for this structure is presented in figure 9. This figure shows a large overlap around

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

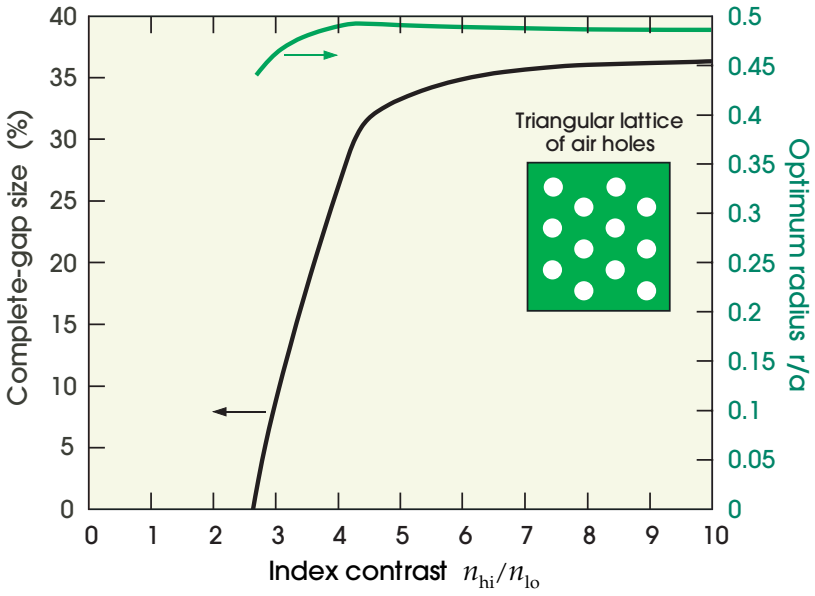


Figure 8: Gap recipe for the lowest *complete* gap of the triangular lattice of air holes in dielectric.

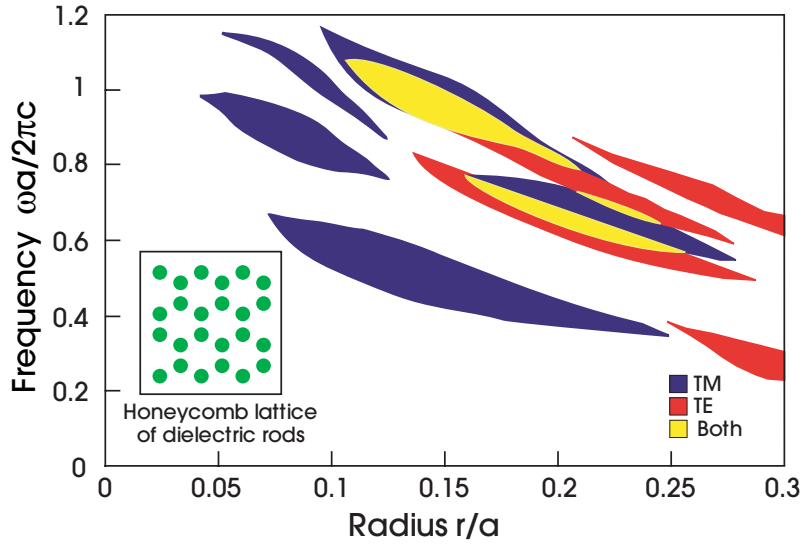


Figure 9: Gap map for a honeycomb lattice of dielectric rods ($\varepsilon = 11.4$).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

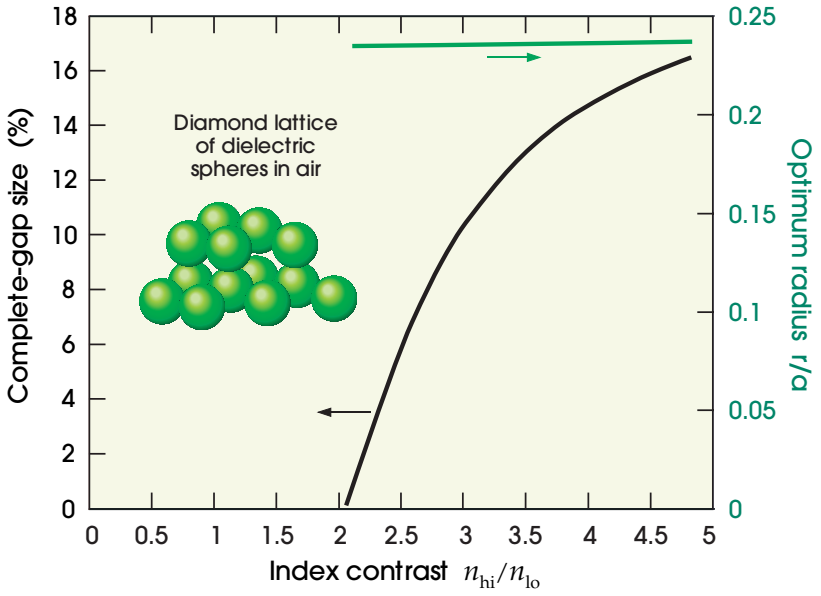


Figure 10: Gap recipe for the lowest *complete* gap of the diamond lattice of dielectric spheres in air.

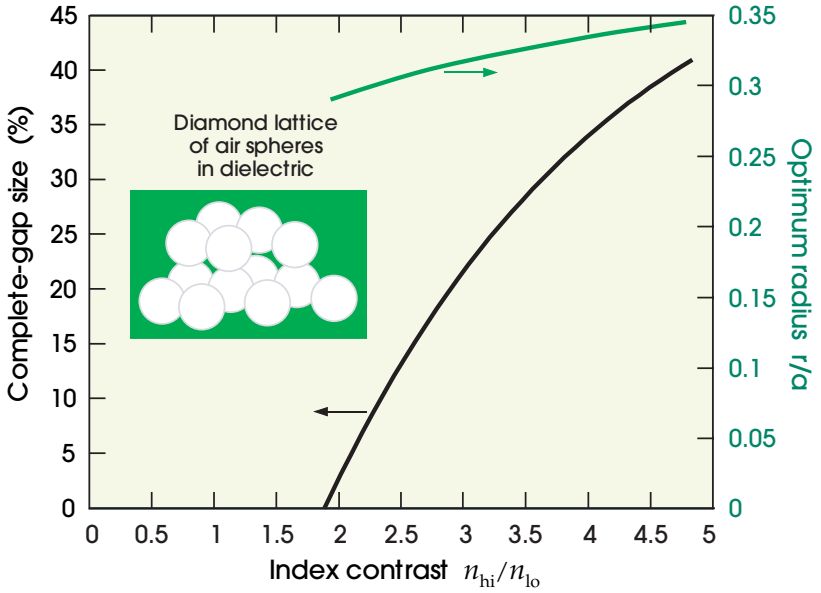


Figure 11: Gap recipe for the lowest *complete* gap of the diamond lattice of spherical air holes in dielectric.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

$r/a = 0.14$ and $\omega a/2\pi c \sim 1.0$, which is of much larger extent than the complete band gap of the triangular lattice. To fabricate such a structure with a band gap at $\lambda = 1.5\,\mu$m would require a feature size of 0.45 μm. This improvement should make the production of such two-dimensional lattices less formidable.

On the other hand, in discussing practical fabrication, it is important to recall the lessons of chapter 8. The two-dimensional band diagrams presented here are accurate in three dimensions only if the pattern can be made very thick, so that the finite height can be neglected. For structures of a finite thickness, the projected band diagrams of chapter 8 are more appropriate. The same qualitative features appear, of course: rods favor TM-like gaps, and holes favor TE-like gaps, and even quantitatively the mismatch is not too great. We also saw that, as long as one has an approximate horizontal mirror symmetry plane, a complete gap for both polarizations is not required.

Three-Dimensional Gaps

If we want *perfect* dielectric confinement of light in all three dimensions, a three-dimensional photonic crystal is usually required,[2] as in chapter 6. For comparison with the previous section, we also present gap recipes for the simplest three-dimensional photonic crystal (Ho et al., 1990): a diamond lattice of dielectric spheres in air or of spherical holes in dielectric, shown respectively in figures 10 and 11.

Both of these graphs display the same monotonic increase of gap size with index contrast, but we see that it is easier to create a gap with air holes than with dielectric spheres. Dielectric spheres in air require an index contrast of about 2.05:1 to create a gap, whereas air spheres in dielectric need only a contrast of 1.88:1 to obtain a complete three-dimensional gap.

Notice also that the spheres are always overlapping, which matches our expectation in chapter 6 that gaps favor connected structures. In particular, just tangent spheres would correspond to a radius of $r/a = \sqrt{3}/8 \approx 0.21651$. The optimal radii are always larger than this ($r/a \approx 0.235$ for the dielectric spheres and $r/a \geq 0.29$ for the air holes).

[2] Two exceptions are described in Watts et al. (2002) and Xu et al. (2003).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Appendix D

Computational Photonics

In chapters 2 and 3, 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.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

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 eigenproblems**: 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 responses**: 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 10). There is, of course, some overlap between the problems; 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

[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*b*; Johnson et al., 2002*b*).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

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 occurs only if all of the singularities are accounted for

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

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: **weighted 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) restricted to structures that are slowly varying in at least one direction.

Frequency-Domain Eigenproblems

A frequency-domain eigensolver solves the Maxwell eigenproblem for the frequencies of a periodic system (or nonperiodic, as described below), as given by equation (11) of chapter 3:

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

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

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 need only consider the finite unit cell of the structure. In addition to the eigenequation, $\mathbf{u}_\mathbf{k}(\mathbf{r})$ must satisfy the transversality constraint:

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

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** $Ax = \omega^2 Bx$, 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 needed only 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 quickly to the true eigenvector. In this way, any desired accuracy can be obtained in a small number of steps. Second, they merely 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.

[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 $\omega \neq 0$.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

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 Electromagnetic Energy and the Variational Principle of chapter 2. 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 2): $x^\dagger Bx_0 = 0$. This process continues for ω_2, ω_3, and so on.

This suffices for periodic structures, but what about nonperiodic 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 $\mathbf{k}$ 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 nonperiodic structures, such as the "leaky" cavity modes of chapters 7 and 8, 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.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Frequency-Domain Responses

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 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

[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).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

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 response of the previous section. Unlike solvers for a frequency-domain response, 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 $\mathrm{Re}\int \mathbf{E}^* \times \mathbf{H}/2$ through a filter structure like those of chapter 10, 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

[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 8 on page 40). 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.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

which the flux is obtained at each ω. As for frequency-domain methods above, PML absorbing layers are used to simulate open boundaries.

Why, then, would anyone ever directly compute the frequency-domain response? 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 nonlocal 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).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

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 $c_n(k) = \frac{1}{a}\int_0^a dx\, e^{-i\frac{2\pi n}{a}x} u_k(x)$. 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, $-\frac{N}{2}$ to $\frac{N}{2}-1$). 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 multidimensional 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 B), and $\mathbf{c_G} = \frac{1}{V}\int d^3\mathbf{r}\, e^{-i\mathbf{G}\cdot\mathbf{r}}\mathbf{u_k}(\mathbf{r})$ 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

[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).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

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:

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

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

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

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

Equation (8) 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 (8) 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 $\varepsilon_\mathbf{G}^{-1}$ 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 $\varepsilon_\mathbf{G}^{-1}$ 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 (8) is a *finite* matrix eigenequation of the form $Ax=\omega^2x$, 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 nonzero for all $\mathbf{G}'$ and $\mathbf{G}$, our matrix A is *dense* (mostly nonzero), 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

[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 $Ax=\omega^2Bx$, or rather why B here is the identity, is that the planewave basis functions are orthogonal to one another.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

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 (8) via a three-step process. First, we take the cross-product $(\mathbf{k}+\mathbf{G}) \times \mathbf{c}_{\mathbf{G}}$, which takes $O(N)$ time. Then, we 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 2 (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

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

[13] Put another way, we are taking advantage of the fact that equation (8) 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).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

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 (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 catalogues. 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 finds the frequency-domain response by a transfer-matrix method that is especially efficient for structures that can be subdivided into a sequence of uniform cross sections along a given direction.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

BIBLIOGRAPHY

Agrawal, Govind P. 2001. *Nonlinear Fiber Optics*. 3rd ed. San Diego: Academic Press.

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.

Aoki, Kanna, Hideki T. Miyazaki, Hideki Hirayama, Kyoji Inoshita, Toshihiko Baba, Norio Shinya, and Yoshinobu Aoyagi. 2002. "Three-dimensional photonic crystals for optical wavelengths assembled by micromanipulation." *Appl. Phys. Lett.* 81: 3122–3124.

Argyros, A., T. A. Birks, S. G. Leon-Saval, C. M. B. Cordeiro, F. Luan, and P. St. J. Russell. 2005. "Photonic bandgap with an index step of one percent." *Opt. Express* 13(1): 309–314.

Ashcroft, N. W., and N. D. Mermin. 1976. *Solid State Physics*. Philadelphia: Holt Saunders.

Aspnes, D. E. 1982. "Local-field effects and effective medium theory: A microscopic perspective." *Am. J. Phys.* 50:704–709.

Assefa, Solomon, Peter T. Rakich, Peter Bienstman, Steven G. Johnson, Gale S. Petrich, John D. Joannopoulos, Leslie A. Kolodziejski, Erich P. Ippen, and Henry I. Smith. 2004. "Guiding 1.5 μm light in photonic crystals based on dielectric rods.'' *Appl. Phys. Lett.* 85(25):6110–6112.

Atkin, D. M., P. St. J. Russell, and T. A. Birks. 1996. "Photonic band structure of guided Bloch modes in high index films fully etched through with periodic microstructure." *J. Mod. Opt.* 43(5):1035–1053.

Axmann, W., P. Kuchment, and L. Kunyansky. 1999. "Asymptotic methods for thin high-contrast two-dimensional PBG materials." *J. Lightwave Tech.* 17(11):1996–2007.

Baba, T., N. Fukaya, and J. Yonekura. 1999. "Observation of light propagation in photonic crystal optical waveguides with bends." *Electron. Lett.* 35(8):654–655.

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.

Bakhtazad, Aref, and Andrew G. Kirk. 2005. "1-D slab photonic crystal *k*-vector superprism demultiplexer: Analysis, and design." *Opt. Express* 13(14):5472–5482.

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.

Benisty, H., D. Labilloy, C. Weisbuch, C. J. M. Smith, T. F. Krauss, D. Cassagne, A. Béraud, and C. Jouanin. 2000. "Radiation losses of waveguide-based two-dimensional photonic crystals: Positive role of the substrate." *Appl. Phys. Lett.* 76(5):532–534.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Berman, Paul R., ed. 1994. *Cavity Quantum Electrodynamics*. San Diego: Academic Press.

Bermel, Peter, J. D. Joannopoulos, Yoel Fink, Paul A. Lane, and Charles Tapalian. 2004. "Properties of radiating pointlike sources in cylindrical omnidirectionally reflecting waveguides." *Phys. Rev. B* 69:035316.

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

Birks, T. A., P. J. Roberts, P. St. J. Russell, D. M. Atkin, and T. J. Shepherd. 1995. "Full 2-D photonic bandgaps in silica/air structures." *Electron. Lett.* 31(22):1941–1943.

Birks, T. A., J. C. Knight, and P. St. J. Russell. 1997. "Endlessly single-mode photonic crystal fiber." *Opt. Lett.* 22(13):961–963.

Birks, T. A., W. J. Wadsworth, and P. St. J. Russell. 2000. "Supercontinuum generation in tapered fibers." *Opt. Lett.* 25(19):1415–1417.

Biró, L. P., Zs. Bálint, K. Kertész, Z. Vértesy, G. I. Márk, Z. E. Horváth, J. Balázs, D. Méhn, I. Kiricsi, V. Lousse, and J.-P. Vigneron. 2003. "Role of photonic-crystal-type structures in the thermal regulation of a Lycaenid butterfly sister species pair." *Phys. Rev. E* 67:021907.

Bjarklev, Anders, Jes Broeng, and Araceli Sanchez Bjarklev. 2003. *Photonic Crystal Fibres*. New York: Springer.

Bloch, Felix. 1928. "Über die Quantenmechanik der Electronen in Kristallgittern." *Z. Physik* 52:555–600.

Bloembergen, N. 1965. *Nonlinear Optics*. New York: W. A. Benjamin.

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.

Brillouin, Léon. 1946. *Wave Propagation in Periodic Structures*. New York: McGraw-Hill.

Brillouin, Léon, ed. 1960. *Wave Propagation and Group Velocity*. New York: Academic Press.

Brown, E. R., and O. B. McMahon. 1995. "Large electromagnetic stop bands in metallodielectric photonic crystals." *Appl. Phys. Lett.* 67(15):2138–2140.

Busch, Kurt, and J. Sajeev. 1998. "Photonic band gap formation in certain self-organizing systems." *Phys. Rev. E* 58:3896–3908.

Chan, C. T., K. M. Ho, and C. M. Soukoulis. 1991. "Photonic band gaps in experimentally realizable periodic dielectric structures." *Europhys. Lett.* 16:563–568.

Chan, Y. S., C. T. Chan, and Z. Y. Liu. 1998. "Photonic band gaps in two dimensional photonic quasicrystals." *Phys. Rev. Lett.* 80(5):956–959.

Chen, Chin-Lin. 1981. "Transverse electric fields guided by doubly-periodic structures." *J. Appl. Phys.* 52(8):4926–4937.

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

Chow, Edmond, S.-Y. Lin, J. R. Wendt, S. G. Johnson, and J. D. Joannopoulos. 2001. "Quantitative analysis of bending efficiency in photonic-crystal waveguide bends at $\lambda = 1.55\,\mu$m wavelengths." *Opt. Lett.* 26(5):286–288.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Chutinan, Alongkarn, and Susumu Noda. 1998. "Spiral three-dimensional photonic-band-gap structure." *Phys. Rev. B* 57:2006–2008.

Chutinan, Alongkarn, Masamitsu Mochizuki, Masahiro Imada, and Susumu Noda. 2001. "Surface-emitting channel drop filters using single defects in two-dimensional photonic crystal slabs." *Appl. Phys. Lett.* 79(17):2690–2692.

Coccioli, R., M. Boroditsky, K. W. Kim, Y. Rahmat-Samii, and E. Yablonovitch. 1998. "Smallest possible electromagnetic mode volume in a dielectric cavity." *IEE Proc. Optoelectron.* 145(6):391–397.

Courant, R., and D. Hilbert. 1953. *Methods of Mathematical Physics*. 2nd ed. New York: Wiley.

Cregan, R. F., B. J. Mangan, J. C. Knight, T. A. Birks, P. St.-J. Russell, and P. J. Roberts. 1999. "Single-mode photonic band gap guidance of light in air." *Science* 285:1537–1539.

Čtyroký, Jiřii. 2001. "Photonic bandgap structures in planar waveguides." *J. Opt. Soc. Am. A* 18(2):435–441.

Dowling, Jonathan P., and Charles M. Bowden. 1994. "Anomalous index of refraction in photonic bandgap materials." *J. Mod. Opt.* 41(2):345–351.

Drikis, I., S. Y. Yang, H. E. Horng, Chin-Yih Hong, and H. C. Yang. 2004. "Modified frequency-domain method for simulating the electromagnetic properties in periodic magnetoactive systems." *J. Appl. Phys.* 95(10):5876–5881.

Eisenhart, L. P. 1948. "Enumeration of potentials for which one-particle Schrödinger equations are separable." *Phys. Rev.* 74:87–89.

Elachi, Charles. 1976. "Waves in active and passive periodic structures: A review." *Proc. IEEE* 64(12):1666–1698.

Fan, Shanhui, Joshua N. Winn, Adrian Devenyi, J. C. Chen, Robert D. Meade, and J. D. Joannopoulos. 1995*a*. "Guided and defect modes in periodic dielectric waveguides." *J. Opt. Soc. Am. B* 12(7):1267–1272.

Fan, Shanhui, P. R. Villeneuve, and J. D. Joannopoulos. 1995*b*. "Theoretical investigation of fabrication-related disorder on the properties of photonic crystals." *J. Appl. Phys.* 78:1415–1418.

Fan, Shanhui, Pierre R. Villeneuve, and J. D. Joannopoulos. 1995*c*. "Large omnidirectional band gaps in metallodielectric photonic crystals." *Phys. Rev. B* 54:11245–11251.

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.

Fan, Shanhui, Pierre R. Villeneuve, J. D. Joannopoulos, and H. A. Haus. 1998. "Channel drop tunneling through localized states." *Phys. Rev. Lett.* 80(5):960–963.

Fan, Shanhui, Pierre R. Villeneuve, J. D. Joannopoulos, and H. A. Haus. 2001*a*. "Loss-induced on/off switching in a channel add/drop filter." *Phys. Rev. B* 64:245302.

Fan, Shanhui, Steven G. Johnson, J. D. Joannopoulos, C. Manolatou, and H. A. Haus. 2001*b*. "Waveguide branches in photonic crystals." *J. Opt. Soc. Am. B* 18(2):162–165.

Fan, Shanhui, and J. D. Joannopoulos. 2002. "Analysis of guided resonances in photonic crystal slabs." *Phys. Rev. B.* 65: 235112.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Fan, Shanhui, Wonjuoo Suh, and J. D. Joannopoulos. 2003. "Temporal coupled-mode theory for the Fano resonance in optical resonators." *J. Opt. Soc. Am. A* 20(3):569–572.

Fano, U. 1961. "Effects of configuration interaction on intensities and phase shifts." *Phys. Rev.* 124(6):1866–1878.

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.

Felber, F. S., and J. H. Marburger. 1976. "Theory of nonresonant multistable optical devices." *Appl. Phys. Lett.* 28(12):731–733.

Fink, Yoel, Joshua N. Winn, Shanhui Fan, Chiping Chen, Jurgen Michel, J. D. Joannopoulos, and Edwin L. Thomas. 1998. "A dielectric omnidirectional reflector." *Science* 282:1679–1682.

Fink, Yoel, Augustine M. Urbas, Moungi G. Bawendi, John D. Joannopoulos, and Edwin L. Thomas. 1999*a*. "Block copolymers as photonic bandgap materials." *J. Lightwave Tech.* 17(11):1963–1969.

Fink, Yoel, Daniel J. Ripin, Shanhui Fan, Chiping Chen, J. D. Joannopoulos, and Edwin L. Thomas. 1999*b*. "Guiding optical light in air using an all-dielectric structure." *J. Lightwave Tech.* 17(11):2039–2041.

Floquet, Gaston. 1883. "Sur les équations différentielles linéaires à coefficients périodiques." *Ann. École Norm. Sup.* 12:47–88.

Foresi, J. S., P. R. Villeneuve, J. Ferrera, E. R. Thoen, G. Steinmeyer, S. Fan, J. D. Joannopoulos, L. C. Kimerling, Henry I. Smith, and E. P. Ippen. 1997. "Photonic-bandgap microcavities in optical waveguides." *Nature* 390(13):143–145.

Fowles, Grant R. 1975. *Introduction to Modern Optics*. New York: Dover.

Garcia-Santamaria, Florencio, Hideki T. Miyazaki, Alfonso Urquia, Marta Ibisate, Manuel Belmonte, Norio Shinya, Francisco Meseguer, and Cefe Lopez. 2002. "Nanorobotic manipulation of microspheres for on-chip diamond architectures." *Adv. Materials* 14(16):1144–1147.

Gloge, D. 1971. "Weakly guiding fibers." *Appl. Opt.* 10(10):2252–2258.

Gohberg, Israel, Sehmour Goldberg, and Marinus A. Kaashoek. 2000. *Basic Classes of Linear Operators*. Basel: Birkhäuser.

Graetsch, Heribert. 1994. "Structural characteristics of opaline and microcrystalline silica minerals." *Reviews in Mineralogy* 29:209–232.

Gralak, Boris, Gerard Tayeb, and Stefan Enoch. 2001. "Morpho butterflies wings color modeled with lamellar grating theory." *Opt. Express* 9(11):567–578.

Griffiths, D. J. 1989. *Introduction to Electrodynamics*. Englewood Cliffs, NJ: Prentice Hall.

Hall, A. Rupert. 1990. "Beyond the fringe: Diffraction as seen by Grimaldi, Fabri, Hooke and Newton." *Notes and Records of the Royal Soc. London* 44(1):13–23.

Harrison, W. A. 1980. *Electronic Structure and the Properties of Solids*. San Francisco: Freeman.

Haus, Hermann A. 1984. *Waves and Fields in Optoelectronics*. Englewood Cliffs, NJ: Prentice-Hall.

Haus, Hermann A., and Weiping Huang. 1991. "Coupled-mode theory." *Proc. IEEE* 79(10):1505–1518.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Haus, Hermann A., and Y. Lai. 1991. "Narrow-band distributed feedback reflector design." *J. Lightwave Tech.* 9(6):754–760.

Hecht, Eugene, and Alfred Zajac. 1997. *Optics*. 3rd ed. Reading, MA: Addison-Wesley.

Hill, George William. 1886. "On the part of the motion of the lunar perigee which is a function of the mean motions of the sun and moon." *Acta Math.* 8:1–36. This work was initially published and distributed privately in 1877.

Hill, K. O., Y. Fujii, D. C. Johnson, and B. S. Kawasaki. 1978. "Photosensitivity in optical fiber waveguides: Application to reflection filter fabrication." *Appl. Phys. Lett.* 32(10):647–649.

Ho, K. M., C. T. Chan, and C. M. Soukoulis. 1990. "Existence of a photonic gap in periodic dielectric structures." *Phys. Rev. Lett.* 65:3152–3155.

Ho, K. M., C. T. Chan, C. M. Soukoulis, R. Biswas, and M. Sigalas. 1994. "Photonic band gaps in three dimensions: New layer-by-layer periodic structures." *Solid State Comm.* 89:413–416.

Hughes, S., L. Ramunno, Jeff F. Young, and J. E. Sipe. 2005. "Extrinsic optical scattering loss in photonic crystal waveguides: Role of fabrication disorder and photon group velocity." *Phys. Rev. Lett.* 94:033903.

Ibanescu, Mihai, Steven G. Johnson, Marin Soljačić, J. D. Joannopoulos, Yoel Fink, Ori Weisberg, Torkel D. Engeness, Steven A. Jacobs, and M. Skorobogatiy. 2003. "Analysis of mode structure in hollow dielectric waveguide fibers." *Phys. Rev. E* 67:046608.

Ibanescu, Mihai, Evan J. Reed, and J. D. Joannopoulos. 2006. "Enhanced photonic band-gap confinement via Van Hove saddle point singularities." *Phys. Rev. Lett.* 96(3):033904.

Inui, T., Y. Tanabe, and Y. Onodera. 1996. *Group Theory and Its Applications in Physics*. Heidelberg: Springer.

Istrate, Emanuel, and Edward H. Sargent. 2006. "Photonic crystal heterostructures and interfaces." *Rev. Mod. Phys.* 78:455–481.

Jackson, J. D. 1998. *Classical Electrodynamics*. 3rd ed. New York: Wiley.

John, Sajeev. 1984. "Electromagnetic absorption in a disordered medium near a photon mobility edge." *Phys. Rev. Lett.* 53:2169–2172.

John, Sajeev. 1987. "Strong localization of photons in certain disordered dielectric superlattices." *Phys. Rev. Lett.* 58:2486–2489.

Johnson, Steven G., and J. D. Joannopoulos. 2000. "Three-dimensionally periodic dielectric layered structure with omnidirectional photonic band gap." *Appl. Phys. Lett.* 77:3490–3492.

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., C. Manolatau, Shanhui Fan, Pierre R. Villeneuve, and J. D. Joannopoulos. 1998. "Elimination of cross talk in waveguide intersections." *Opt. Lett.* 23(23):1855–1857.

Johnson, Steven G., Shanhui Fan, P. R. Villeneuve, J. D. Joannopoulos, and L. A. Kolodziejski. 1999. "Guided modes in photonic crystal slabs." *Phys. Rev. B* 60:5751–5758.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Johnson, Steven G., Shanhui Fan, P. R. Villeneuve, and J. D. Joannopoulos. 2000. "Linear waveguides in photonic-crystal slabs." *Phys. Rev. B* 62:8212–8222.

Johnson, Steven G., Attila Mekis, Shanhui Fan, and J. D. Joannopoulos. 2001*a*. "Molding the flow of light." *Computing Sci. Eng.* 3(6):38–47.

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

Johnson, Steven G., Shanhui Fan, Attila Mekis, and J. D. Joannopoulos. 2001*c*. "Multipole-cancellation mechanism for high-Q cavities in the absence of a complete photonic band gap." *Appl. Phys. Lett.* 78(22):3388–3390.

Johnson, Steven G., M. Ibanescu, M. A. Skorobogatiy, O. Weisberg, J. D. Joannopoulos, and Y. Fink. 2002*a*. "Perturbation theory for Maxwell's equations with shifting material boundaries." *Phys. Rev. E* 65:066611.

Johnson, Steven G., Peter Bienstman, M. Skorobogatiy, Mihai Ibanescu, Elefterios Lidorikis, and J. D. Joannopoulos. 2002*b*. "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.

Kanskar, M., P. Paddon, V. Pacradouni, R. Morin, A. Busch, Jeff F. Young, S. R. Johnson, Jim MacKenzie, and T. Tiedje. 1997. "Observation of leaky slab modes in an air-bridged semiconductor waveguide with a two-dimensional photonic lattice." *Appl. Phys. Lett.* 70(11):1438–1440.

Karalis, Aristeidis, Steven G. Johnson, and J. D. Joannopoulos. 2004. "Discrete-mode cancellation mechanism for high-Q integrated optical cavities with small modal volume." *Opt. Lett.* 29(19):2309–2311.

Katsenelenbaum, B. Z., L. Mercader del Río, M. Pereyaslavets, M. Sorolla Ayza, and M. Thumm. 1998. *Theory of Nonuniform Waveguides: The Cross-Section Method*. London: Inst. of Electrical Engineers.

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

Kawakami, Shojiro. 2002. "Analytically solvable model of photonic crystal structures and novel phenomena." *J. Lightwave Tech.* 20(8):1644–1650.

Kim, Hyang Kyun, Michel J. F. Digonnet, Gordon S. Kino, Jonghwa Shin, and Shanhui Fan. 2004. "Simulations of the effect of the core ring on surface and air-core modes in photonic bandgap fibers." *Opt. Express* 12(15):3436–3442.

Kittel, Charles. 1996. *Introduction to Solid State Physics*. New York: Wiley.

Knight, J. C., T. A. Birks, P. St. J. Russell, and D. M. Atkin. 1996. "All-silica single-mode optical fiber with photonic crystal cladding." *Opt. Lett.* 21(19):1547–1549.

Knight, J. C., J. Broeng, T. A. Birks, and P. St.-J. Russell. 1998. "Photonic band gap guidance in optical fibers." *Science* 282:1476–1478.

Körner, T. W. 1988. *Fourier Analysis*. Cambridge: Cambridge University Press.

Kosaka, Hideo, Takayuki Kawashima, Akihisa Tomita, Masaya Notomi, Toshiaki Tamamura, Takashi Sato, and Shojiro Kawakami. 1998. "Superprism phenomena in photonic crystals." *Phys. Rev. B* 58(16):10096–10099.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Kosaka, Hideo, Takayuki Kawashima, Akihisa Tomita, Masaya Notomi, Toshiaki Tamamura, Takashi Sato, and Shojiro Kawakami. 1999*a*. "Self-collimating phenomena in photonic crystals." *Appl. Phys. Lett.* 74(9):1212–1214.

Kosaka, Hideo, Takayuki Kawashima, Akihisa Tomita, Masaya Notomi, Toshiaki Tamamura, Takashi Sato, and Shojiro Kawakami. 1999*b*. "Superprism phenomena in photonic crystals: Toward microscale lightwave circuits." *J. Lightwave Tech.* 17(11):2032–2038.

Krauss, Thomas F., Richard M. De La Rue, and Stuart Brand. 1996. "Two-dimensional photonic-bandgap structures operating at near-infrared wavelengths." *Nature* 383:699–702.

Kubota, Hiirokazu, Satoki Kawanishi, Shigeki Koyanagi, Masatoshi Tanaka, and Shyunichiro Yamaguchi. 2004. "Absolutely single polarization photonic crystal fiber." *IEEE Photon. Tech. Lett.* 16(1):182–184.

Kuchinsky, S., D. C. Allan, N. F. Borrelli, and J.-C. Cotteverte. 2000. "3d localization in a channel waveguide in a photonic crystal with 2d periodicity." *Optics Commun.* 175:147–152.

Kuzmiak, V., A. A. Maradudin, and F. Pincemin. 1994. "Photonic band structures of two-dimensional systems containing metallic components." *Phys. Rev. B* 50:16835–16844.

Landau, L. D., and E. M. Lifshitz. 1977. *Quantum Mechanics*. 3rd ed. Oxford: Butterworth-Heinemann.

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

Lau, Wah Tung, and Shanhui Fan. 2002. "Creating large bandwidth line defects by embedding dielectric waveguides into photonic crystal slabs." *Appl. Phys. Lett.* 81:3915–3917.

Lemaire, P. J., R. M. Atkins, V. Mizrahi, and W. A. Reed. 1993. "High pressure H_2 loading as a technique for achieving ultrahigh UV photosensitivity and thermal sensitivity in GeO_2 doped optical fibres." *Electron. Lett.* 29(13):1191–1193.

Leung, K. M., and Y. F. Liu. 1990. "Full vector wave calculation of photonic band structures in face-centered-cubic dielectric media." *Phys. Rev. Lett.* 65:2646–2649.

Li, Ming-Jun, Xin Chen, Daniel A. Nolan, George E. Berkey, Ji Wang, William A. Wood, and Luis A. Zenteno. 2005. "High bandwidth single polarization fiber with elliptical central air hole." *J. Lightwave Tech.* 23(11):3454–3460.

Liboff, R. L. 1992. *Introductory Quantum Mechanics*. 2nd ed. Reading, MA: Addison-Wesley.

Lidorikis, Eleftherios, M. Soljačić, A. Karalis, Mihai Ibanescu, Yoel Fink, and J. D. Joannopoulos. 2004. "Cutoff solitons in axially uniform systems." *Opt. Lett.* 29(8):851–853.

Lin, Shawn-Yu, and J. G. Fleming. 1999. "A three-dimensional optical photonic crystal." *J. Lightwave Tech.* 17(11):1944–1947.

Lin, Shawn-Yu, V. M. Hietala, Li Wang, and E. D. Jones. 1996. "Highly dispersive photonic band-gap prism." *Opt. Lett.* 21(21):1771–1773.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Lin, Shawn-Yu, Edmund Chow, Vince Hietala, Pierre R. Villeneuve, and J. D. Joannopoulos. 1998*a*. "Experimental demonstration of guiding and bending of electromagnetic waves in a photonic crystal." *Science* 282:274–276.

Lin, Shawn-Yu, J. G. Fleming, D. L. Hetherington, B. K. Smith, R. Biswas, K. M. Ho, M. M. Sigalas, W. Zubrzycki, S. R. Kurtz, and Jim Bur. 1998*b*. "A three-dimensional photonic crystal operating at infrared wavelengths." *Nature* 394:251–253.

Lin, Shawn-Yu, E. Chow, S. G. Johnson, and J. D. Joannopoulos. 2000. "Demonstration of highly efficient waveguiding in a photonic crystal slab at the 1.5-μm wavelength." *Opt. Lett.* 25(17):1297–1299.

Lin, Shawn-Yu, E. Chow, S. G. Johnson, and J. D. Joannopoulos. 2001. "Direct measurement of the quality factor in a two-dimensional photonic-crystal microcavity." *Opt. Lett.* 26(23):1903–1905.

Litchinitser, Natalia M., Steven C. Dunn, Brian Usner, Benjamin J. Eggleton, Thomas P. White, Ross C. McPhedran, and C. Martijn de Sterke. 2003. "Resonances in microstructured optical waveguides." *Opt. Express* 11(10):1243–1251.

Little, B. E., S. T. Chu, H. A. Haus, J. Foresi, and J.-P. Laine. 1997. "Microring resonator channel dropping filters." *J. Lightwave Tech.* 15(6):998–1005.

Lodahl, Peter, A. Floris van Driel, Ivan S. Nikolaev, Arie Irman, Karin Overgaag, Daniël Vanmaekelbergh, and Willem L. Vos. 2004. "Controlling the dynamics of spontaneous emission from quantum dots by photonic crystals." *Nature* 430:654–657.

Lončar, Marko, Dušan Nedeljković, Theodor Doll, Jelena Vučković, Axel Scherer, and Thomas P. Pearsall. 2000. "Waveguiding in planar photonic crystals." *Appl. Phys. Lett.* 77(13):1937–1939.

Lončar, Marco, Tomoyuki Yoshie, Axel Scherer, Pawan Gogna, and Yueming Qiu. 2002. "Low-threshold photonic crystal laser." *Appl. Phys. Lett.* 81(15):2680–2682.

Louisell, William H. 1960. *Coupled Mode and Parametric Electronics*. New York: Wiley.

Lu, Zhaolin, Shouyuan Shi, Janusz A. Murakowski, Garrett J. Schneider, Christopher A. Schuetz, and Dennis W. Prather. 2006. "Experimental demonstration of self-collimation inside a three-dimensional photonic crystal." *Phys. Rev. Lett.* 96: 173902.

Luo, Chiyan, Steven G. Johnson, and J. D. Joannopoulos. 2002*a*. "All-angle negative refraction in a three-dimensionally periodic photonic crystal." *Appl. Phys. Lett.* 81:2352–2354.

Luo, Chiyan, Steven G. Johnson, J. D. Joannopoulos, and J. B. Pendry. 2002*b*. "All-angle negative refraction without negative effective index." *Phys. Rev. B* 65:201104.

Luo, Chiyan, Steven G. Johnson, J. D. Joannopoulos, and J. B. Pendry. 2003. "Subwavelength imaging in photonic crystals." *Phys. Rev. B* 68:045115.

Luo, Chiyan, Marin Soljačić, and J. D. Joannopoulos. 2004. "Superprism effect based on phase velocities." *Opt. Lett.* 29(7):745–747.

Lyapunov, Alexander Mihailovich. 1992. *The General Problem of the Stability of Motion*. London: Taylor and Francis. Translated by A. T. Fuller from Edouard Davaux's French translation (1907) of the original Russian dissertation (1892).

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Maldovan, Martin, and Edwin L. Thomas. 2004. "Diamond-structured photonic crystals." *Nature Materials* 3:593–600.

Maldovan, M., A. M. Urbas, N. Yufa, W. C. Carter, and E. L. Thomas. 2002. "Photonic properties of bicontinuous cubic microphases." *Phys. Rev. B* 65:165123.

Malitson, I. H. 1965. "Interspecimen comparison of the refractive index of fused silica." *J. Opt. Soc. Am.* 55(10):1205–1209.

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

Mangan, B. J., L. Farr, A. Langford, P. J. Roberts, D. P. Williams, F. Couny, M. Lawman, M. Mason, S. Coupland, R. Flea, and H. Sabert. 2004. Low loss (1.7 dB/km) hollow core photonic bandgap fiber. In *Proc. Opt. Fiber Commun. Conf. (OFC)*. Los Angeles. Paper PDP24.

Manolatou, C., M. J. Khan, Shanhui Fan, Pierre R. Villeneuve, H. A. Haus, and J. D. Joannopoulos. 1999*a*. "Coupling of modes analysis of resonant channel add-drop filters." *IEEE J. Quantum Electron.* 35(9):1322–1331.

Manolatou, C., Steven G. Johnson, S. Fan, P. R. Villeneuve, H. A. Haus, and J. D. Joannopoulos. 1999*b*. "High-density integrated optics." *J. Lightwave Tech.* 17(9):1682–1692.

Marcatilli, E. A. J. 1969. "Bends in optical dielectric waveguides." *Bell Syst. Tech. J.* 48:2103–2132.

Marcuse, D. 1991. *Theory of Dielectric Optical Waveguides*. 2nd ed. San Diego: Academic Press.

Martorell, Jordi, and N. M. Lawandy. 1990. "Observation of inhibited spontaneous emission in a periodic dielectric structure." *Phys. Rev. Lett.* 65:1877–1880.

Mathews, J., and R. Walker. 1964. *Mathematical Methods of Physics*. Redwood City, CA: Addison-Wesley.

Matsumoto, T., and T. Baba. 2004. "Photonic crystal *k*-vector superprism." *J. Lightwave Tech.* 22(3):917–922.

McCall, S. L., P. M. Platzman, R. Dalichaouch, David Smith, and S. Schultz. 1991. "Microwave propagation in two-dimensional dielectric lattices." *Phys. Rev. Lett.* 67:2017–2020.

McGurn, Arthur R., and Alexei A. Maradudin. 1993. "Photonic band structures of two- and three-dimensional periodic metal or semiconductor arrays." *Phys. Rev. B* 48:17576–17579.

McPhedran, R. C., N. A. Nicorovici, D. R. McKenzie, G. W. Rouse, L. C. Botton, V. Welch, A. R. Parker, M. Wohlgennant, and V. Vardeny. 2003. "Structural colours through photonic crystals." *Physica B: Cond. Matter* 338:182–185.

Meade, Robert D., Karl D. Brommer, Andrew M. Rappe, and J. D. Joannopoulos. 1991*a*. "Electromagnetic Bloch waves at the surface of a photonic crystal." *Phys. Rev. B* 44:10961–10964.

Meade, Robert D., Karl D. Brommer, Andrew M. Rappe, and J. D. Joannopoulos. 1991*b*. "Photonic bound states in periodic dielectric materials." *Phys. Rev. B* 44:13772–13774.

Meade, Robert D., Karl D. Brommer, Andrew M. Rappe, and J. D. Joannopoulos. 1992. "Existence of a photonic band gap in two dimensions." *Appl. Phys. Lett.* 61:495–497.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

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).

Meade, Robert D., A. Devenyi, J. D. Joannopoulos, O. L. Alerhand, D. A. Smith, and K. Kash. 1994. "Novel applications of photonic band gap materials: Low-loss bends and high *Q* cavities." *J. Appl. Phys.* 75(9):4753–4755.

Mekis, Attila, J. C. Chen, I. Kurland, Shanhui Fan, Pierre R. Villeneuve, and J. D. Joannopoulos. 1996. "High transmission through sharp bends in photonic crystal waveguides." *Phys. Rev. Lett.* 77(18):3787–3790.

Melekhin, V. N., and A. B. Manenkov. 1968. "Dielectric tube as a low-loss waveguide." *Zhurnal Tekhnicheskoi Fiziki* 38(12):2113–2115.

Mendez, Alexis, and T. Morse, eds. 2006 *Specialty Optical Fibers Handbook.* New York: Academic Press.

Merzbacher, E. 1961. *Quantum Mechanics*. New York: Wiley.

Norris, David J., Erin G. Arlinghaus, Linli Meng, Ruth Heiny, and L. E. Scriven. 2004. "Opalline photonic crystals: How does self-assembly work?" *Adv. Materials* 16(16):1393–1399.

Notomi, M. 2000. "Theory of light propagation in strongly modulated photonic crystals: Refractionlike behavior in the vicinity of the photonic band gap." *Phys. Rev. B* 62:10696–10705.

Notomi, M., K. Yamada, A. Shinya, J. Takahashi, C. Takahashi, and I. Yokohama. 2001. "Extremely large group-velocity dispersion of line-defect waveguides in photonic crystal slabs." *Phys. Rev. Lett.* 87(25):253902.

Notomi, Masaya, Akihiko Shinya, Satoshi Mitsugi, Goh Kira, Eiichi Kuramochi, and Takasumi Tanabe. 2005. "Optical bistable switching action of Si high-*Q* photonic-crystal nanocavities." *Opt. Express* 13(7):2678–2687.

Ochiai, T., and K. Sakoda. 2001. "Dispersion relation and optical transmittance of a hexagonal photonic crystal slab." *Phys. Rev. B* 63:125107.

Okuno, Toshiaki, Masashi Onishi, Tomonori Kashiwada, Shinji Ishikawa, and Masayuki Nishimura. 1999. "Silica-based functional fibers with enhanced nonlinearity and their applications." *IEEE J. Sel. Top. Quant. Elec.* 5(5):1385–1391.

Oliner, A. A., and A. Hessel. 1959. "Guided waves on sinusoidally-modulated reactance surfaces." *IRE Trans. Antennas and Propagation* 7(5):S201–S208.

Olivier, S., M. Rattier, H. Benisty, C. Weisbuch, C. J. M. Smith, R. M. De La Rue, T. F. Krauss, U. Oesterle, and R. Houdré. 2001. "Mini-stopbands of a one-dimensional system: The channel waveguide in a two-dimensional photonic crystal." *Phys. Rev. B* 63:113311.

Paddon, P., and Jeff F. Young. 2000. "Two-dimensional vector-coupled-mode theory for textured planar waveguides." *Phys. Rev. B* 61: 2090–2101.

Painter, O. J., A. Husain, A. Scherer, J. D. O'Brien, I. Kim, and P. D. Dapkus. 1999. "Room temperature photonic crystal defect lasers at near-infrared wavelengths in InGaAsP." *J. Lightwave Tech.* 17(11):2082–2088.

Palik, Edward D., ed. 1998. *Handbook of Optical Constants of Solids*. London: Academic Press.

Pantelides, Sokrates T. 1978. "The electronic structure of impurities and other point defects in semiconductors." *Rev. Mod. Phys.* 50:797–858.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Parker, Andrew R., Victoria L. Welch, Dominique Driver, and Natalia Martini. 2003. "Structural colour: Opal analogue discovered in a weevil." *Nature* 426:786–787.

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.

Pendry, J. B. 2000. "Negative refraction makes a perfect lens." *Phys. Rev. Lett.* 85:3966–3969.

Pendry, J. B., and S. A. Ramakrishna. 2003. "Focusing light using negative refraction." *J. Phys.: Cond. Matter* 15:6345–6364.

Pendry, J. B., A. J. Holden, D. J. Robbins, and W. J. Stewart. 1999. "Magnetism from conductors and enhanced nonlinear phenomena." *IEEE Trans. Microwave Theory Tech.* 47(11):2075–2084.

Peng, S. T., Theodor Tamir, and Henry L. Bertoni. 1975. "Theory of periodic dielectric waveguides." *IEEE Trans. Microwave Theory Tech.* 23(1):123–133.

Petrov, A. Yu., and M. Eich. 2004. "Zero dispersion at small group velocities in photonic crystal waveguides." *Appl. Phys. Lett.* 85(21):4866–4868.

Petrov, E. P., V. N. Bogomolov, I. I. Kalosha, and S. V. Gaponenko. 1998. "Spontaneous emission of organic molecules embedded in a photonic crystal." *Phys. Rev. Lett.* 81:77–80.

Phillips, P. L., J. C. Knight, B. J. Mangan, and P. Russell. 1999. "Near-field optical microscopy of thin photonic crystal films." *J. Appl. Phys.* 85(9):6337–6342.

Pierce, J. R. 1954. "Coupling of modes of propagation." *J. Appl. Phys.* 25:179–183.

Plihal, M., and A. A. Maradudin. 1991. "Photonic band structure of two-dimensional systems: The triangular lattice." *Phys. Rev. B* 44:8565–8571.

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

Povinelli, M. L., Steven G. Johnson, Shanhui Fan, and J. D. Joannopoulos. 2001. "Emulation of two-dimensional photonic crystal defect modes in a photonic crystal with a three-dimensional photonic band gap." *Phys. Rev. B* 64:075313.

Povinelli, M. L., Steven G. Johnson, Elefterios Lidorikis, J. D. Joannopoulos, and Marin Soljačić. 2004. "Effect of a photonic band gap on scattering from waveguide disorder." *Appl. Phys. Lett.* 84(18):3639–3641.

Povinelli, M. L., Steven G. Johnson, and J. D. Joannopoulos. 2005. "Slow-light, band-edge waveguides for tunable time delays." *Opt. Express* 13:7145–7159.

Prather, Dennis W., Shouyuan Shi, David M. Pustai, Caihua Chen, Sriram Venkataraman, Ahmed Sharkawy, Garrett J. Schneider, and Janusz Murakowski. 2004. "Dispersion-based optical routing in photonic crystals." *Opt. Lett.* 29(1):50–52.

Purcell, E. M. 1946. "Spontaneous emission probabilities at radio frequencies." *Phys. Rev.* 69:681–686.

Qi, Minghao, Eleftherios Lidorikis, Peter T. Rakich, Steven G. Johnson, J. D. Joannopoulos, Erich P. Ippen, and Henry I. Smith. 2004. "A three-dimensional optical photonic crystal with designed point defects." *Nature* 429:538–542.

Rakich, Peter T., Marcus S. Dahlem, Sheila Tandon, Mihai Ibanescu, Marin Soljačić, Gale S. Petrich, J. D. Joannopoulos, Leslie A. Kolodziejski, and Erich P. Ippen. 2006. "Achieving centimetre-scale supercollimation in a large-area two-dimensional photonic crystal." *Nature Materials* 5:93–96.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Ramaswami, Rajiv, and Kumar N. Sivarajan. 1998. *Optical Networks: A Practical Perspective*. London: Academic Press.

Ranka, Jinendra K., Robert S. Windeler, and Andrew J. Stentz. 2000. "Visible continuum generation in air-silica microstructure optical fibers with anomalous dispersion at 800 nm." *Opt. Lett.* 25(1):25–27.

Lord Rayleigh. 1887. "On the maintenance of vibrations by forces of double frequency, and on the propagation of waves through a medium endowed with a periodic structure." *Philosophical Magazine* 24:145–159.

Lord Rayleigh. 1917. "On the reflection of light from a regularly stratified medium." *Proc. Royal Society of London* 93:565–577.

Ripin, Daniel J., Kuo-Yi Lim, G. S. Petrich, Pierre R. Villeneuve, Shanhui Fan, E. R. Thoen, J. D. Joannopoulos, E. P. Ippen, and L. A. Kolodziejski. 1999. "One-dimensional photonic bandgap microcavities for strong optical confinement in GaAs and GaAs/Al_xO_y semiconductor waveguides." *J. Lightwave Tech.* 17(11):2152–2160.

Roberts, P., F. Couny, H. Sabert, B. Mangan, D. Williams, L. Farr, M. Mason, A. Tomlinson, T. Birks, J. Knight, and P. St. J. Russell. 2005. "Ultimate low loss of hollow-core photonic crystal fibres." *Opt. Express* 13(1):236–244.

Robertson, W. M., G. Arjavalingam, R. D. Meade, K. D. Brommer, A. M. Rappe, and J. D. Joannopoulos. 1992. "Measurement of photonic band structure in a two-dimensional periodic dielectric array." *Phys. Rev. Lett.* 68:2023–2026.

Robertson, W. M., G. Arjavalingam, R. D. Meade, K. D. Brommer, A. M. Rappe, and J. D. Joannopoulos. 1993. "Observation of surface photons on periodic dielectric arrays." *Opt. Lett.* 18(7):528–530.

Robinson, Jacob T., Christina Manolatou, Long Chen, and Michal Lipson. 2005. "Ultrasmall mode volumes in dielectric optical microcavities." *Phys. Rev. Lett.* 95:143901.

Rodriguez, Alejandro, M. Ibanescu, J. D. Joannopoulos, , and Steven G. Johnson. 2005. "Disorder-immune confinement of light in photonic-crystal cavities." *Opt. Lett.* 30:3192–3194.

Roh, Young-Geun, Sungjoon Yoon, Heonsu Jeon, Seung-Ho Han, and Q-Han Park. 2004. "Experimental verification of cross talk reduction in photonic crystal waveguide crossings." *Appl. Phys. Lett.* 85(16):3351–3353.

Rudin, Walter. 1964. *Principles of Mathematical Analysis*. New York: McGraw-Hill.

Russell, Philip. 2003. "Photonic crystal fibers." *Science* 299(5605):358–362.

Saitoh, K., N. A. Mortensen, and M. Koshiba. 2004. "Air-core photonic band-gap fibers: The impact of surface modes." *Opt. Express* 12(3):394–400.

Sakurai, J. J. 1994. *Modern Quantum Mechanics*. Rev. ed. Reading, MA: Addison-Wesley.

Sanders, J. V. 1964. "Colour of precious opal." *Nature* 204:1151–1153.

Satpathy, S., Ze Zhang, and M. R. Salehpour. 1990. "Theory of photon bands in three-dimensional periodic dielectric structures." *Phys. Rev. Lett.* 64:1239–1242.

Sauvan, C., P. Lalanne, and J. P. Hugonin. 2005. "Slow-wave effect and mode-profile matching in photonic crystal microcavities." *Phys. Rev. B* 71:165118.

Scherer, A., O. Painter, B. D'Urso, R. Lee, and A. Yariv. 1998. "InGaAsP photonic band gap crystal membrane microresonators." *J. Vac. Sci. Tech. B* 16(6):3906–3910.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Serbin, Jesper, and Min Gu. 2005. "Superprism phenomena in polymeric woodpile structures." *J. Appl. Phys.* 98:123101.

Shankar, R. 1982. *Principles of Quantum Mechanics*. New York: Plenum Press.

Sharp, D. N., M. Campbell, E. R. Dedman, M. T. Harrison, R. G. Denning, and A. J. Turberfield. 2002. "Photonic crystals for the visible spectrum by holographic lithography." *Optical and Quantum Electron.* 34:3–12.

Shepherd, T. J., and P. J. Roberts. 1995. "Scattering in a two-dimensional photonic crystal: An analytical model." *Phys. Rev. E* 51(5):5158–5161.

Shin, Jonghwa, and Shanhui Fan. 2005. "Conditions for self-collimation in three-dimensional photonic crystals." *Opt. Lett.* 30(18):2397–2399.

Sievenpiper, D. F., M. E. Sickmiller, and E. Yablonovitch. 1996. "3D wire mesh photonic crystals." *Phys. Rev. Lett.* 76:2480–2483.

Sigalas, M. M., C. T. Chan, K. M. Ho, and C. M. Soukoulis. 1995. "Metallic photonic band-gap materials." *Phys. Rev. B* 52:11744–11751.

Sigalas, M. M., C. M. Soukoulis, R. Biswas, and K. M. Ho. 1997. "Effect of the magnetic permeability on photonic band gaps." *Phys. Rev. B* 56(3):959–962.

Simon, Barry. 1976. "The bound state of weakly coupled Schrödinger operators in one and two dimensions." *Ann. Phys.* 97(2):279–288.

Sipe, J. E. 2000. "Vector $\mathbf{k} \cdot \mathbf{p}$ approach for photonic band structures." *Phys. Rev. E* 62:5672–5677.

Skorobogatiy, M. 2005. "Efficient antiguiding of TE and TM polarizations in low-index core waveguides without the need for an omnidirectional reflector." *Opt. Lett.* 30(22):2991–2993.

Skorobogatiy, M., Mihai Ibanescu, Steven G. Johnson, Ori Weisberg, Torkel D. Engeness, Marin Soljačić, Steven A. Jacobs, and Yoel Fink. 2002. "Analysis of general geometric scaling perturbations in a transmitting waveguide. The fundamental connection between polarization mode dispersion and group-velocity dispersion." *J. Opt. Soc. Am. B* 19(12):2867–2875.

Smith, Charlene M., Natesan Venkataraman, Michael T. Gallagher, Dirk Müller, James A. West, Nicholas F. Borrelli, Douglas C. Allan, and Karl W. Koch. 2003. "Low-loss hollow-core silica/air photonic bandgap fibre." *Nature* 424:657–659.

Smith, D. R., R. Dalichaouch, N. Kroll, S. Schultz, S. L. McCall, and P. M. Platzman. 1993. "Photonic band structure and defects in one and two dimensions." *J. Opt. Soc. Am. B* 10(2):314–321.

Smith, D. R., Willie J. Padilla, D. C. Vier, S. C. Nemat-Nasser, and S. Schultz. 2000. "Composite medium with simultaneously negative permeability and permittivity." *Phys. Rev. Lett.* 84:4184–4187.

Smith, D. R., J. B. Pendry, and M. C. K. Wiltshire. 2004. "Metamaterials and negative refractive index." *Science* 305:788–792.

Smith, D. R., D. C. Vier, Th. Koschny, and C. M Soukoulis. 2005. "Electromagnetic parameter retrieval from inhomogeneous materials." *Phys. Rev. E* 71:036617.

Snyder, A. W., and J. D. Love. 1983. *Optical Waveguide Theory*. London: Chapman and Hall.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Soljačić, Marin, Mihai Ibanescu, Steven G. Johnson, Yoel Fink, and J. D. Joannopoulos. 2002*a*. "Optimal bistable switching in non-linear photonic crystals." *Phys. Rev. E Rapid Commun.* 66:055601(R).

Soljačić, Marin, Steven G. Johnson, Shanhui Fan, Mihai Ibanescu, Erich Ippen, and J. D. Joannopoulos. 2002*b*. "Photonic-crystal slow-light enhancement of non-linear phase sensitivity." *J. Opt. Soc. Am. B* 19:2052–2059.

Soljačić, Marin, Chiyan Luo, J. D. Joannopoulos, and Shanhui Fan. 2003. "Nonlinear photonic crystal microdevices for optical integration." *Opt. Lett.* 28(8):637–639.

Song, Bong-Shik, Susumu Noda, Takashi Asano, and Yoshihiro Akahane. 2005. "Ultra-high-*Q* photonic double-heterostructure nanocavity." *Nature Materials* 4(3):207–210.

Sözüer, H. S., and J. P. Dowling. 1994. "Photonic band calculations for woodpile structures." *J. Mod. Opt.* 41(2):231–239.

Sözüer, H. S., and J. W. Haus. 1993. "Photonic bands: Simple-cubic lattice." *J. Opt. Soc. Am. B* 10(2):296–302.

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.

Srinivasan, Kartik, and Oskar Painter. 2002. "Momentum space design of high-*Q* photonic crystal optical cavities." *Opt. Express* 10(15):670–684.

Srinivasan, Kartik, Paul Barclay, and Oskar Painter. 2004. "Fabrication-tolerant high quality factor photonic crystal microcavities." *Opt. Express* 12:1458–1463.

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

Suck, J.-B., M. Schreiber, and P. Häussler, eds. 2004. *Quasicrystals*. Berlin: Springer.

Sugimoto, Yoshimasa, Yu Tanaka, Naoki Ikeda, Yusui Nakamura, and Kiyoshi Asakawa. 2004. "Low propagation loss of 0.76 dB/mm in GaAs-based single-line-defect two-dimensional photonic crystal slab waveguides up to 1 cm in length." *Opt. Express* 12(6):1090–1096.

Suh, Wonjoo, Zheng Wang, and Shanhui Fan. 2004. "Temporal coupled-mode theory and the presence of non-orthogonal modes in lossless multimode cavities." *IEEE J. Quantum Electron.* 40(10):1511–1518.

Sze, S. M. 1981. *Physics of Semiconductor Devices*. New York: Wiley.

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

Temelkuran, Burak, Shandon D. Hart, Gilles Benoit, John D. Joannopoulos, and Yoel Fink. 2002. "Wavelength-scalable hollow optical fibres with large photonic bandgaps for CO_2 laser transmission." *Nature* 420:650–653.

Tinkham, Michael. 2003. *Group Theory and Quantum Mechanics*. New York: Dover.

Toader, Ovidiu, and Sajeev John. 2001. "Proposed square spiral microfabrication architecture for large three-dimensional photonic band gap crystals." *Science* 292:1133–1135.

Tokushima, Masatoshi, Hideo Kosaka, Akihisa Tomita, and Hirohito Yamada. 2000. "Lightwave propagation through a 120° sharply bent single-line-defect photonic crystal waveguide." *Appl. Phys. Lett.* 76(8):952–954.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Tokushima, Masatoshi, Hirohito Yamada, and Yasuhiko Arakawa. 2004. "1.5-μm-wavelength light guiding in waveguides in square-lattice-of-rod photonic crystal slab." *Appl. Phys. Lett.* 84:4298–4300.

Tong, Limin, Rafael R. Gattass, Jonathan B. Ashcom, Sailing He, Jingyi Lou, Mengyan Shen, Iva Maxwell, and Eric Mazur. 2003. "Subwavelength-diameter silica wires for low-loss optical wave guiding." *Nature* 426:816–819.

Torres, D., O. Weisberg, G. Shapira, C. Anastassiou, B. Temelkuran, M. Shurgalin, S. A. Jacobs, R. U. Ahmad, T. Wang, U. Kolodny, S. M. Shapshay, Z. Wang, A. K. Devaiah, U. D. Upadhyay, and J. A. Koufman. 2005. "OmniGuide photonic bandgap fibers for flexible delivery of CO_2 laser energy for laryngeal and airway surgery." *Proc. SPIE* 5686(1):310–321.

Tzolov, Velko P., Marie Fontaine, Nicolas Godbout, and Suzanne Lacroix. 1995. "Nonlinear self-phase-modulation effects: A vectorial first-order perturbation approach." *Opt. Lett.* 20(5):456–458.

Ulrich, R., and M. Tacke. 1973. "Submillimeter waveguiding on periodic metal structure." *Appl. Phys. Lett.* 22(5):251–253.

Veselago, Victor G. 1968. "The electrodynamics of substances with simultaneously negative values of ε and μ." *Sov. Phys. Uspekhi* 10:509–514.

Villeneuve, Pierre R., and Michel Piché. 1992. "Photonic band gaps in two-dimensional square and hexagonal lattices." *Phys. Rev. B* 46:4969–4972.

Villeneuve, Pierre R., Shanhui Fan, J. D. Joannopoulos, Kuo-Yi Lim, G. S. Petrich, L. A. Kolodziejski, and Rafael Reif. 1995. "Air-bridge microcavities." *Appl. Phys. Lett.* 67(2):167–169.

Villeneuve, P. R., S. Fan, S. G. Johnson, and J. D. Joannopoulos. 1998. "Three-dimensional photon confinement in photonic crystals of low-dimensional periodicity." *IEE Proc. Optoelectron.* 145(6):384–390.

Vlasov, Yurii A., Xiang-Zheng Bo, James C. Sturm, and David J. Norris. 2001. "On-chip natural assembly of silicon photonic bandgap crystals." *Nature* 414:289–293.

Vlasov, Yurii A., N. Moll, and S. J. McNab. 2004. "Mode mixing in asymmetric double-trench photonic crystal waveguides." *J. Appl. Phys.* 95(9):4538–4544.

Vučković, Jelena, Marco Lončar, Hideo Mabuchi, and Axel Scherer. 2002. "Design of photonic crystal microcavities for cavity QED." *Phys. Rev. E* 65:016608.

Wadsworth, William J., Arturo Ortigosa-Blanch, Jonathan C. Knight, Tim A. Birks, T. -P. Martin Man, and Phillip St. J. Russell. 2002. "Supercontinuum generation in photonic crystal fibers and optical fiber tapers: A novel light source." *J. Opt. Soc. Am. B* 19(9):2148–2155.

Wakita, Koichi. 1997. *Semiconductor Optical Modulators*. New York: Springer.

Wang, Zheng, and Shanhui Fan. 2005. "Magneto-optical defects in two-dimensional photonic crystals." *Appl. Phys. B* 81:369–375.

Ward, A. J., and J. B. Pendry. 1996. "Refraction and geometry in Maxwell's equations." *J. Mod. Opt.* 43(4):773–793.

Watts, M. R., S. G. Johnson, H. A. Haus, and J. D. Joannopoulos. 2002. "Electromagnetic cavity with arbitrary Q and small modal volume without a complete photonic bandgap." *Opt. Lett.* 27(20):1785–1787.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

West, James, Charlene Smith, Nicholas Borrelli, Douglas Allan, and Karl Koch. 2004. "Surface modes in air-core photonic band-gap fibers." *Opt. Express* 12(8):1485–1496.

Winn, Joshua N., Robert D. Meade, and J. D. Joannopoulos. 1994. "Two-dimensional photonic band-gap materials." *J. Mod. Opt.* 41(2):257–273.

Winn, Joshua N., Yoel Fink, Shanhui Fan, and J. D. Joannopoulos. 1998. "Omnidirectional reflection from a one-dimensional photonic crystal." *Opt. Lett.* 23(20):1573–1575.

Wu, Lijun, Michael Mazilu, and Thomas F. Krauss. 2003. "Beam steering in planar-photonic crystals: From superprism to supercollimator." *J. Lightwave Tech.* 21(2): 561–566.

Xu, Youg, Reginald K. Lee, and Amnon Yariv. 2000. "Propagation and second-harmonic generation of electromagnetic waves in a coupled-resonator optical waveguide." *J. Opt. Soc. Am. B* 17(3):387–400.

Xu, Yong, Wei Liang, Amnon Yariv, J. G. Fleming, and Shawn-Yu Lin. 2003. "High-quality-factor Bragg onion resonators with omnidirectional reflector cladding." *Opt. Lett.* 28(22):2144–2146.

Yablonovitch, E. 1987. "Inhibited spontaneous emission in solid-state physics and electronics." *Phys. Rev. Lett.* 58:2059–2062.

Yablonovitch, E., and T. J. Gmitter. 1989. "Photonic band structure: The face-centered-cubic case." *Phys. Rev. Lett.* 63:1950–1953.

Yablonovitch, E., T. J. Gmitter, and K. M. Leung. 1991*a*. "Photonic band structure: The face-centered-cubic case employing nonspherical atoms." *Phys. Rev. Lett.* 67:2295–2298.

Yablonovitch, E., T. J. Gmitter, R. D. Meade, A. M. Rappe, K. D. Brommer, and J. D. Joannopoulos. 1991*b*. "Donor and acceptor modes in photonic band structure." *Phys. Rev. Lett.* 67:3380–3383.

Yang, K., and M. de Llano. 1989. "Simple variational proof that any two-dimensional potential well supports at least one bound state." *Am. J. Phys.* 57(1):85–86.

Yanik, Mehmet Fatih, and Shanhui Fan. 2004. "Stopping light all optically." *Phys. Rev. Lett.* 92(8):083901.

Yanik, Mehmet Fatih, Shanhui Fan, Marin Soljačić, and J. D. Joannopoulos. 2003. "All-optical transistor action with bistable switching in a photonic crystal cross-waveguide geometry." *Opt. Lett.* 28(24):2506–2508.

Yariv, A., Y. Xu, R. K. Lee, and A. Scherer. 1999. "Coupled-resonator optical waveguide: A proposal and analysis." *Opt. Lett.* 24:711–713.

Yariv, Amnon. 1997. *Optical Electronics in Modern Communications*. 5th ed. Oxford: Oxford University Press.

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

Yeh, P., and A. Yariv. 1976. "Bragg reflection waveguides." *Opt. Commun.* 19:427–430.

Yeh, P., A. Yariv, and E. Marom. 1978. "Theory of Bragg fiber." *J. Opt. Soc. Am.* 68:1196–1201.

Yeh, P. 1979. "Electromagnetic propagation in birefringent layered media." *J. Opt. Soc. Am.* 69:742–756.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

Yeh, Pochi. 1988. *Optical Waves in Layered Media*. New York: Wiley.

Zengerle, R. 1987. "Light propagation in singly and doubly periodic planar waveguides." *J. Mod. Opt.* 34(12):1589–1617.

Zhang, Ze, and Sashi Satpathy. 1990. "Electromagnetic wave propagation in periodic structures: Bloch wave solution of Maxwell's equations." *Phys. Rev. Lett.* 65:2650–2653.

Zi, Jian, Xindi Yu, Yizhou Li, Xinhua Hu, Chun Xu, Xingjun Wang, Xiaohan Liu, and Rongtang Fu. 2003. "Coloration strategies in peacock feathers." *Proc. Nat. Acad. Sci. USA* 100(22):12576–12578.

Zolla, Frederic, Gilles Renversez, André Nicolet, Boris Kuhlmey, Sebastien Guenneau, and Didier Felbacq. 2005. *Foundations of Photonic Crystal Fibres*. London: Imperial College Press.

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

INDEX

Page-nimber typeface:
bold = introductions/definitions
italic = footnotes

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

November 13, 2007 Time: 06:57pm index.tex

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS

November 13, 2007 Time: 06:57pm index.tex

COPYRIGHT 2008, PRINCETON UNIVERSITY PRESS