

## Calculating photonic band structure

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1996 J. Phys.: Condens. Matter 8 1085

(<http://iopscience.iop.org/0953-8984/8/9/003>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.208

The article was downloaded on 13/05/2010 at 16:18

Please note that [terms and conditions apply](#).

REVIEW ARTICLE

## Calculating photonic band structure

J B Pendry

The Blackett Laboratory, Imperial College, London SW7 2BZ, UK

Received 20 November 1995

**Abstract.** Photonic materials structured on the scale of the wavelength of light have become the subject of an active field of research fed by the hope of creating novel properties. Theory plays a central role reinforced by the difficulty of manufacturing photonic materials: unusually we are better able to design a photonic material than to build it. In this review we explore the application of scattering theory to Maxwell's equations that has enabled theory to make such a central contribution: implementation of Maxwell's equations on a discrete mesh, development of the electromagnetic transfer matrix, order- $N$  methods, and adaptive meshes. At the same time we present applications to a few key problems by way of illustration, and discuss the special circumstances of metallic photonic structures and their unique properties.

### 1. Introduction

As optics merges with electronics to fuel a revolution in computing and communication, it is natural to compare and contrast electrons and photons. The former are for the moment the more active members of the partnership, but for future progress we look increasingly to photons. Electrons have enjoyed the advantage in this contest because of the ease with which they can be controlled: piped down wires, stored in memories, or switched from one channel to another. Semiconductors are the medium in which they operate, and semiconductors enjoy their control of electronic properties because of diffraction effects in the atomic lattice. Structure and function operate harmoniously together in these materials to deliver the properties we need.

Photons also have a wave-like nature but on too large a scale to be diffracted by atomic lattices. Therefore to a large extent they have escaped detailed engineering of their properties. However, there is no reason that we should not artificially structure materials on the scale of the wavelength of light to produce band gaps, defect states, delay lines and other novel properties by analogy with the electron case. This agenda of reworking semiconductor physics for the photon has been seized upon, largely by condensed-matter theorists well versed in the intricacies of band theory.

It is a field for the moment chiefly populated by theorists as experiment struggles to develop the complex technology required to give three-dimensional structure to a material on the micron scale, even though there are some interesting optical structures prepared by 'self-assembly' methods [1]. Theorists have been very active over the past five years in developing the methodology of photonic band theory, and applying it to a variety of systems. In this short review we shall scrutinize the new techniques from the standpoint of a condensed-matter theorist, following an agenda very much inspired by electronic band theory.

We begin by discussing the peculiarities of Maxwell's equations viewed from a computational standpoint, and present some of the early solutions to this problem. Next we show how to make a consistent adaptation of Maxwell's equations to a discrete mesh and develop perhaps the most powerful approach to the problem. Finite-difference equations can be solved in several different ways. Transfer-matrix methods are widely used and are efficient and accurate. More recently 'order- $N$ ' methods have been developed and are extremely efficient in some circumstances. One problem for photonic calculations is that materials can be structured in a wide variety of different configurations, in contrast to atomic lattices all of which have the same basic geometry. This leads us to considering adaptive-mesh calculations in which the real-space mesh is adjusted to the geometry of the material. Finally we discuss the special case of metallic photonic structures which have some remarkable properties.

## 2. Maxwell's equations on a computer

We are fortunate in photonic band theory in that the equations governing the photon's behaviour are somewhat better defined than for electrons. Maxwell's equations in SI units are (see Panofsky and Phillips [2], or Jackson [3])

$$\begin{aligned}\nabla \cdot \mathbf{D} &= \rho & \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\partial \mathbf{B} / \partial t & \nabla \times \mathbf{H} &= +\partial \mathbf{D} / \partial t + \mathbf{j}\end{aligned}\quad (1)$$

where  $\mathbf{D}$  is the electric displacement vector,  $\mathbf{E}$  the electric field vector,  $\mathbf{B}$  the magnetic flux density,  $\mathbf{H}$  the magnetic field intensity,  $\rho$  the charge density, and  $\mathbf{j}$  the electric current density. Most commonly we work at a fixed frequency, so

$$\begin{aligned}\mathbf{D}(\mathbf{r}, t) &= \mathbf{D}(\mathbf{r}) \exp(-i\omega t) & \mathbf{E}(\mathbf{r}, t) &= \mathbf{E}(\mathbf{r}) \exp(-i\omega t) \\ \mathbf{B}(\mathbf{r}, t) &= \mathbf{B}(\mathbf{r}) \exp(-i\omega t) & \mathbf{H}(\mathbf{r}, t) &= \mathbf{H}(\mathbf{r}) \exp(-i\omega t) \\ \rho(\mathbf{r}, t) &= \rho(\mathbf{r}) \exp(-i\omega t) & \mathbf{j}(\mathbf{r}, t) &= \mathbf{j}(\mathbf{r}) \exp(-i\omega t).\end{aligned}\quad (2)$$

Note the sign in the exponential: a common source of confusion between physicists and engineers. We adopt the physicists' notation here, of course. Substituting into Maxwell's equations gives

$$\begin{aligned}\nabla \cdot \mathbf{D} &= \rho & \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= +i\omega \mathbf{B} & \nabla \times \mathbf{H} &= -i\omega \mathbf{D} + \mathbf{j}.\end{aligned}\quad (3)$$

These equations taken together with the constitutive equations

$$\mathbf{D} = \varepsilon \varepsilon_0 \mathbf{E} \quad \mathbf{B} = \mu \mu_0 \mathbf{H} \quad \mathbf{j} = \sigma \mathbf{E} \quad (4)$$

define the problem.

All the physics is contained in  $\varepsilon$ ,  $\mu$ ,  $\sigma$ , and for nearly all the problems of interest these can be taken as *local* functions; that is to say the most general linear form reduces to

$$\mathbf{D}(\mathbf{r}) = \int \varepsilon(\mathbf{r}, \mathbf{r}') \varepsilon_0 \mathbf{E}(\mathbf{r}') \, d^3 \mathbf{r}' = \varepsilon(\mathbf{r}) \varepsilon_0 \mathbf{E}(\mathbf{r}) \quad \text{etc.} \quad (5)$$

Non-locality results from atomic-scale processes and is only important when photonic materials have relevant structure on an atomic scale. There are exceptions to this rule, for example when excitonic effects are important, but these specialist topics are beyond the scope of this review.

For some materials such as metals, the dielectric constant,  $\varepsilon$ , depends strongly on the frequency in all frequency ranges. For many insulators  $\varepsilon$  is constant in the optical region

of the spectrum. However, it is an inevitable consequence of the properties of matter that all materials show dispersion in some range of frequencies: commonly in the ultraviolet.

The first calculations of photonic band structure [4–8] addressed the problem of finding the dispersion relationship  $\omega(\mathbf{k})$  in a periodic structure often with the specific aim of finding a ‘photonic insulator’ as first proposed in [9] (an excellent summary of some of these early calculations is to be found in [10, 11]). It is usual to eliminate either the electric or the magnetic component from Maxwell’s equations. For example, on the common assumption that the permeability,  $\mu$ , is everywhere unity, we can derive

$$\nabla \times \nabla \times \mathbf{E} = -\nabla^2 \mathbf{E} + \nabla(\nabla \cdot \mathbf{E}) = +\omega^2 c_0^{-2} \varepsilon \mathbf{E} \quad (6)$$

where

$$c_0 = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \quad (7)$$

is the velocity of light in free space.

For condensed-matter theorists the resemblance of (6) to the Schrödinger equation is illusory. Certainly the familiar ‘del-squared’ term is there, but the additional presence of  $\nabla(\nabla \cdot \mathbf{E})$  transforms the equation into a decidedly awkward customer from the computational point of view. This second term is responsible for the elimination of the longitudinal mode from the spectrum and ensures that any wave of the form

$$\mathbf{E}_\ell = A \mathbf{k} \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (8)$$

is always a solution of (6) provided that

$$\omega_\ell(\mathbf{k}) = 0 \quad (9)$$

which thus removes all longitudinal components from the finite-frequency spectrum. However, under certain circumstances this longitudinal mode can return to haunt an ill constructed computation.

The next step was to observe that the solutions for a periodic medium must take the Bloch wave form, and therefore the electric field can be expanded in a discrete Fourier series:

$$\mathbf{E}_\mathbf{k}(\mathbf{r}) = \sum_{\mathbf{g}} \{E_{s\mathbf{g}} \mathbf{e}_s + E_{p\mathbf{g}} \mathbf{e}_p\} \exp[i(\mathbf{k} + \mathbf{g}) \cdot \mathbf{r}] \quad (10)$$

where  $\mathbf{g}$  is a reciprocal-lattice vector and the two unit transverse vectors are defined by

$$\mathbf{e}_s = \frac{[(\mathbf{k} + \mathbf{g}) \times \mathbf{n}]}{\sqrt{[(\mathbf{k} + \mathbf{g}) \times \mathbf{n}] \cdot [(\mathbf{k} + \mathbf{g}) \times \mathbf{n}]}} \quad (11)$$

$$\mathbf{e}_p = \frac{[(\mathbf{k} + \mathbf{g}) \times (\mathbf{k} + \mathbf{g}) \times \mathbf{n}]}{\sqrt{[(\mathbf{k} + \mathbf{g}) \times (\mathbf{k} + \mathbf{g}) \times \mathbf{n}] \cdot [(\mathbf{k} + \mathbf{g}) \times (\mathbf{k} + \mathbf{g}) \times \mathbf{n}]}}. \quad (12)$$

Each of the two components of the field, ‘ $s$ ’ and ‘ $p$ ’, is constructed to be transverse to the relevant wave vector. The Fourier expansion contains an infinite summation but, as for electronic band structure, is truncated at some finite value of  $(\mathbf{k} + \mathbf{g})$  chosen large enough to ensure convergence. Conventionally the unit vector,  $\mathbf{n}$ , is the normal to a surface of the material, but could be chosen to be any arbitrary vector.

In this basis equation (6) becomes

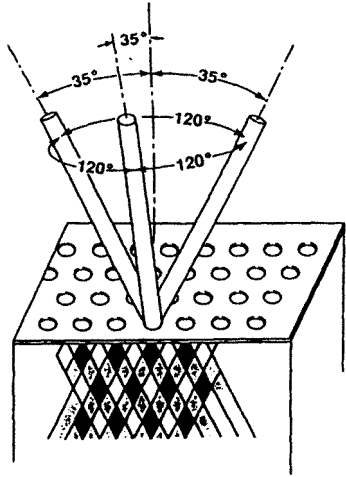
$$|\mathbf{k} + \mathbf{g}|^2 E_{t\mathbf{g}} = +\omega^2 c_0^{-2} \sum_{t'\mathbf{g}'} \varepsilon_{t\mathbf{g};t'\mathbf{g}'} E_{t'\mathbf{g}'} \quad (13)$$

where

$$\varepsilon_{t\mathbf{g},t'\mathbf{g}'} = \mathbf{e}_t(\mathbf{k} + \mathbf{g}) \cdot \mathbf{e}_{t'}(\mathbf{k} + \mathbf{g}') \Omega^{-1} \int \varepsilon(\mathbf{r}) \exp[i(\mathbf{g} - \mathbf{g}') \cdot \mathbf{r}] d^3\mathbf{r} \quad (14)$$

and the integration is over the volume,  $\Omega$ , of the unit cell.

Provided that  $\varepsilon$  does not depend on  $\omega$ , equation (13) is a generalized eigenvalue equation and can be solved for  $\omega(\mathbf{k})$  by conventional techniques.



**Figure 1.** ‘Yablonovite’: a slab of material is covered by a mask consisting of a triangular array of holes. Each hole is drilled through three times, at an angle  $35.26^\circ$  away from normal, and spread out  $120^\circ$  on the azimuth. The resulting criss-cross of holes below the surface of the slab, suggested by the cross-hatching shown here, produces a fully three-dimensionally periodic fcc structure.

This technique and closely related ones were used to calculate the first photonic band structures. For example we see in figure 2 the band structure of the ‘Yablonovite’ structure (figure 1) calculated by this technique using several thousand plane waves in the expansion [6].

This was the first structure to show an absolute band gap and gave great impetus to the subject. So far the structure has not been realized experimentally at visible wavelengths, but a microwave version has been built and measured. This is already sufficient confirmation of the validity of the theory because Maxwell’s equations obey a scaling law. If a structure  $\varepsilon(\mathbf{r})$  sustains a solution of Maxwell’s equations,  $\mathbf{E}_k(\mathbf{r})$ ,

$$-\nabla^2 \mathbf{E}(\mathbf{r}) + \nabla(\nabla \cdot \mathbf{E}(\mathbf{r})) = +\omega^2 c_0^{-2} \varepsilon(\mathbf{r}) \mathbf{E}(\mathbf{r}) \quad (15)$$

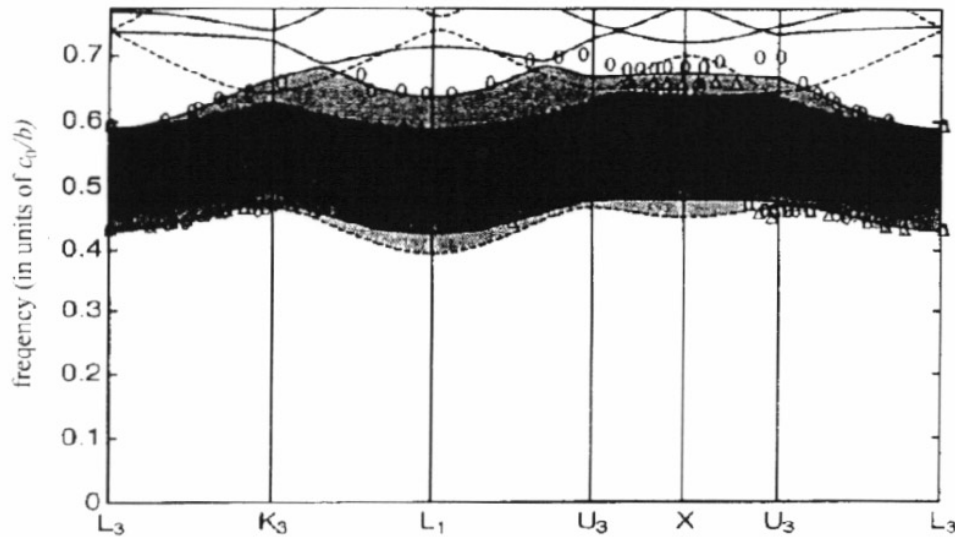
then we can deduce a solution to a different problem by making the substitution

$$\mathbf{r}' = \mathbf{r}/a \quad (16)$$

so that

$$-\nabla'^2 \mathbf{E}(a\mathbf{r}') + \nabla'(\nabla' \cdot \mathbf{E}(a\mathbf{r}')) = +a^2 \omega^2 c_0^{-2} \varepsilon(a\mathbf{r}') \mathbf{E}(a\mathbf{r}'). \quad (17)$$

The transformed equation tells us that a new photonic structure in which all the length scales have been expanded by  $a$  sustains a solution whose frequency is less than the original by the same factor  $a$ . So if an experimental structure has an absolute gap in the GHz band,



**Figure 2.** Frequency versus wave-vector dispersion along various directions in fcc  $k$ -space, where  $c_0/b$  is the speed of light divided by the fcc cube length. The ovals and triangles are the experimental points for s and p polarization respectively. The full and broken curves are the calculations for s and p polarization respectively. The dark-shaded band is the totally forbidden band gap. The lighter-shaded stripes above and below the dark band are forbidden only for s and p polarization respectively.

the frequency of that gap will scale upwards as the lattice spacing of the structure is scaled down. Our argument assumes that  $\varepsilon$  is not a function of frequency.

This method of solution has a number of advantages.

- It gives a stable and reliable algorithm for computing  $\omega(\mathbf{k})$ .
- The computer programming required is minimal and can draw extensively on standard methodologies such as Fourier transformation and matrix diagonalization.
- The longitudinal mode causes no numerical problems because it is projected out of the equations at an early stage.
- An arbitrary  $\varepsilon(\mathbf{r})$  can be handled since no assumptions are made about the shape of objects.

Therefore it was an excellent way to start a serious study of the subject. However, there are some situations in which the method runs into difficulties. Some examples are given below.

- If  $\varepsilon$  happens to depend on  $\omega$  then equation (13) is no longer an eigenvalue equation, but something much more complicated and difficult to handle. The problem stems from the methodology of fixing  $\mathbf{k}$ , then calculating  $\omega$ .
- Experiments commonly measure quantities such as reflection coefficients at a fixed frequency and to calculate these we need all the Bloch waves at that frequency, i.e. again we need to fix  $\omega$  and calculate  $\mathbf{k}$ , not the other way round.
- A plane-wave expansion of the wave field is rather inefficient and renders calculations for the more complex structures impossibly time consuming. (Matrix diagonalization scales as the cube of the number of plane waves.)

- The latter problem is particularly acute in situations where several different length scales are forced upon us. For example when an electromagnetic wave is incident on a metal surface the scale of the penetration depth and the wavelength in vacuum can differ by many orders of magnitude, and yet the Fourier components have to reproduce both to good accuracy.

These are difficulties which are also encountered in electronic band structure and have been countered with techniques which can also be applied to the photonic case.

### 3. Maxwell's equations on a lattice

The two classic models of electronic band structure are the 'nearly-free-electron approximation' and the 'tight-binding model'. They start from completely different standpoints: the former assumes that plane waves are a good approximation to the electron wave function, the latter that electrons interact strongly with the atoms and only occasionally hop across the space between one atom and the next. Both have been developed as the basis for more accurate and sophisticated models. The plane-wave expansions discussed above are the obvious counterparts of the nearly-free-electron model. But it is also possible to find a tight-binding analogy for the photon case. We do this by choosing to represent the photon wave field on a discrete lattice of points in real space. This has several pitfalls and has to be done carefully, but when successfully completed it has many rewards in terms of speed of computation.

The original derivation by MacKinnon and Pendry (see [12, 13]) was made with a view to extending MacKinnon's successful tight-binding models of disordered electronic systems to the photon case, but has in fact had much wider application to ordered systems. Of course, the main problem in representing differential equations on a lattice is in approximating the derivatives. For Maxwell's equations we have another problem: they have the property that all longitudinal modes are 'dead modes', appearing only as zero-frequency solutions. This fundamental property, which has at its heart the conservation of charge, must be preserved in a valid scheme of approximations.

We start from equation (3) and Fourier transform to give

$$\mathbf{k} \times \mathbf{E} = +\omega\mu_0\mathbf{H} \quad \mathbf{k} \times \mathbf{H} = -\omega\varepsilon_0 \int \varepsilon(\mathbf{k}, \mathbf{k}')\mathbf{E}(\mathbf{k}') d^3\mathbf{k}' \quad (18)$$

where we have assumed that we are inside a dielectric medium where

$$\mu = 1 \quad \rho = \mathbf{j} = 0. \quad (19)$$

Note that the 'dead-mode' property holds for a more general class of equations:

$$\boldsymbol{\kappa}_E(\mathbf{k}) \times \mathbf{E} = +\omega\mu_0\mathbf{H} \quad \boldsymbol{\kappa}_H(\mathbf{k}) \times \mathbf{H} = -\omega\varepsilon_0 \int \varepsilon(\mathbf{k}, \mathbf{k}')\mathbf{E}(\mathbf{k}') d^3\mathbf{k}' \quad (20)$$

since, from above,

$$(\boldsymbol{\kappa}_H(\mathbf{k}) \cdot \boldsymbol{\kappa}_E(\mathbf{k}))\mathbf{E}(\mathbf{k}) - \boldsymbol{\kappa}_H(\mathbf{k})[\boldsymbol{\kappa}_E(\mathbf{k}) \cdot \mathbf{E}(\mathbf{k})] = +\omega^2 c_0^{-2} \int \varepsilon(\mathbf{k}, \mathbf{k}')\mathbf{E}(\mathbf{k}') d^3\mathbf{k}' \quad (21)$$

and choosing

$$\mathbf{E}_\ell(\mathbf{k}) = A\boldsymbol{\kappa}_H(\mathbf{k}) \quad (22)$$

always gives the result

$$\omega_\ell(\mathbf{k}) = 0. \quad (23)$$

The idea is that if we choose  $\kappa_E(\mathbf{k})$  and  $\kappa_H(\mathbf{k})$  so that they have simple representations in real space in terms of difference equations, and so that they approximate the exact equations, we shall have obtained a real-space approximation that exactly fulfils the ‘dead-mode’ requirement. In  $k$ -space the differential operators transform exactly to

$$\frac{\partial \mathbf{E}(\mathbf{r})}{\partial x} \leftrightarrow ik_x \mathbf{E}(\mathbf{k}) \quad \text{etc.} \quad (24)$$

However, there is a closely related transformation

$$a^{-1}[\mathbf{E}(\mathbf{r} + a\hat{x}) - \mathbf{E}(\mathbf{r})] \leftrightarrow (ia)^{-1}[\exp(ik_x a) - 1]\mathbf{E}(\mathbf{k}) \quad \text{etc.} \quad (25)$$

So if we choose

$$\begin{aligned} \kappa_{E_x}(\mathbf{k}) &= (+ia)^{-1}[\exp(+ik_x a) - 1] \approx k_x + O(k_x^2 a) \\ \kappa_{H_x}(\mathbf{k}) &= (-ia)^{-1}[\exp(-ik_x a) - 1] \approx k_x + O(k_x^2 a) \end{aligned} \quad (26)$$

and substitute into equation (20), and transform back into real space, we have

$$\begin{aligned} +[E_z(\mathbf{r} + \mathbf{b}) - E_z(\mathbf{r})] - [E_y(\mathbf{r} + \mathbf{c}) - E_y(\mathbf{r})] &= +ia\omega\mu_0 H_x(\mathbf{r}) \\ +[E_x(\mathbf{r} + \mathbf{c}) - E_x(\mathbf{r})] - [E_z(\mathbf{r} + \mathbf{a}) - E_z(\mathbf{r})] &= +ia\omega\mu_0 H_y(\mathbf{r}) \\ +[E_y(\mathbf{r} + \mathbf{a}) - E_y(\mathbf{r})] - [E_x(\mathbf{r} + \mathbf{b}) - E_x(\mathbf{r})] &= +ia\omega\mu_0 H_z(\mathbf{r}) \end{aligned} \quad (27a)$$

$$\begin{aligned} -[H_z(\mathbf{r} - \mathbf{b}) - H_z(\mathbf{r})] + [H_y(\mathbf{r} - \mathbf{c}) - H_y(\mathbf{r})] &= -ia\omega\varepsilon_0 \varepsilon(\mathbf{r}) E_x(\mathbf{r}) \\ -[H_x(\mathbf{r} - \mathbf{c}) - H_x(\mathbf{r})] + [H_z(\mathbf{r} - \mathbf{a}) - H_z(\mathbf{r})] &= -ia\omega\varepsilon_0 \varepsilon(\mathbf{r}) E_y(\mathbf{r}) \\ -[H_y(\mathbf{r} - \mathbf{a}) - H_y(\mathbf{r})] + [H_x(\mathbf{r} - \mathbf{b}) - H_x(\mathbf{r})] &= -ia\omega\varepsilon_0 \varepsilon(\mathbf{r}) E_z(\mathbf{r}) \end{aligned} \quad (27b)$$

and we retrieve a set of finite-difference equations. In these equations  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  represent displacements through a distance  $a$  along the  $x$ ,  $y$ ,  $z$  axes respectively. Further inspection will reveal that the lattice on which we sample the fields is a simple cubic lattice: fields on each site are coupled to the six neighbouring sites of this lattice.

Equations (27), (28) are the counterparts of the nearest-neighbour tight-binding approximation familiar in quantum mechanics, but derived from Maxwell’s equations. In fact we can formulate them to resemble a Hamiltonian even more closely:

$$\sum_{t't'} H_{tt'}(\mathbf{r}, \mathbf{r}') F_{t'}(\mathbf{r}') = \omega F_t(\mathbf{r}) \quad (28)$$

where

$$\mathbf{F}(\mathbf{r}) = \begin{bmatrix} E_x(\mathbf{r}) \\ E_y(\mathbf{r}) \\ E_z(\mathbf{r}) \\ H_x(\mathbf{r}) \\ H_y(\mathbf{r}) \\ H_z(\mathbf{r}) \end{bmatrix}. \quad (29)$$

In the limit of very small  $a$ , relative to the wavelength of radiation, they are an exact representation. Note that the approximations have all to do with the differential operators; therefore we can get an impression of how accurate the equations are by asking how well they represent free space, where

$$\varepsilon = 1. \quad (30)$$



In this case equation (21) gives

$$\begin{aligned}
 a^{-2} \begin{bmatrix} \kappa_{E_y} \kappa_{H_y} + \kappa_{E_z} \kappa_{H_z} & -\kappa_{E_x} \kappa_{H_y} & -\kappa_{E_x} \kappa_{H_z} \\ -\kappa_{E_y} \kappa_{H_x} & \kappa_{E_x} \kappa_{H_x} + \kappa_{E_z} \kappa_{H_z} & -\kappa_{E_y} \kappa_{H_z} \\ -\kappa_{E_z} \kappa_{H_x} & -\kappa_{E_z} \kappa_{H_y} & \kappa_{E_x} \kappa_{H_x} + \kappa_{E_y} \kappa_{H_y} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \\
 = \omega^2 c_0^{-2} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}.
 \end{aligned} \tag{31}$$

If we calculate the dispersion for  $\mathbf{k}$  directed along the  $x$  axis,

$$\begin{aligned}
 \kappa_{E_x}(\mathbf{k}) &= (+ia)^{-1} [\exp(+ik_x a) - 1] & \kappa_{E_y}(\mathbf{k}) &= 0 & \kappa_{E_z}(\mathbf{k}) &= 0 \\
 \kappa_{H_x}(\mathbf{k}) &= (-ia)^{-1} [\exp(-ik_x a) - 1] & \kappa_{H_y}(\mathbf{k}) &= 0 & \kappa_{H_z}(\mathbf{k}) &= 0
 \end{aligned} \tag{32}$$

we get

$$\omega^2 = c_0^2 a^{-2} 4 \sin^2(\frac{1}{2} k_x a). \tag{33}$$

Obviously this reduces to the correct limit when  $a$  is small:

$$\omega^2 = c_0^2 k_x^2 [1 + \frac{1}{2} k_x^2 a^2 + \dots]. \tag{34}$$

This real-space formulation of the problem as set out in (27), (28) can be used as an alternative to the  $k$ -space version described in the last section when computing the band structure. However, this is not the best way forward as the real-space formulation offers some efficient alternatives to straightforward matrix diagonalization. The efficiency of these new methods is owing to the sparse nature of equations (27), (28).

As we point out above, the accuracy of the real-space formulation is limited by the mesh size. The situation can be improved, at the expense of simplicity, by devising a more accurate approximation to  $\mathbf{k}$ . For example we can improve on (26) as follows:

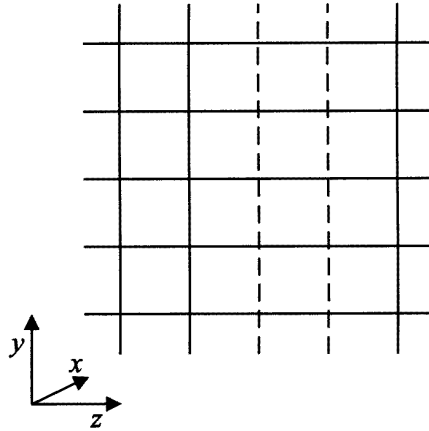
$$\begin{aligned}
 \kappa'_{E_x}(\mathbf{k}) &= (+ia)^{-1} [-\frac{1}{2} \exp(+ik_x a) + 2 \exp(+2ik_x a) - \frac{3}{2}] \approx k_x + O(k_x^3 a^2) \\
 \kappa'_{H_x}(\mathbf{k}) &= (-ia)^{-1} [-\frac{1}{2} \exp(-ik_x a) + 2 \exp(-2ik_x a) - \frac{3}{2}] \approx k_x + O(k_x^3 a^2).
 \end{aligned} \tag{35}$$

Transformation into real space of the resulting equations gives difference equations coupling beyond the nearest neighbours to second-nearest neighbours of a simple cubic lattice. Taking higher orders of the expansion of  $k_x$  in terms of  $[\exp(+ik_x a) - 1]$  gives more accurate approximations and more complex formulae. Of course the computer has no objection to complex formulae, but the benefits of a more accurate approximation that allows us to use fewer sampling points must be balanced against the disadvantage that the matrix grows less sparse as the number of nearest-neighbour couplings rises.

Finally in this section, a word about the type of lattice on which we sample the fields. It might be imagined that we could make a more accurate approximation by choosing a more complex lattice of sampling points: fcc instead of sc perhaps? Although it is true that we can do better in terms of improved approximations to  $k_x, k_y, k_z$ , the more complex lattices usually have some subtle but fatal disadvantages. A detailed discussion is beyond the scope of this review, but has to do with introduction of spurious solutions in addition to the ones we seek. These mix with the desired solutions to introduce numerical instabilities. In this instance a simple cubic lattice has virtues of stability and robustness. It should not be abandoned lightly.

#### 4. How to solve the discrete Maxwell equations using transfer matrices

Many of the experiments with which we shall be concerned were made at fixed frequency. This certainly makes life simpler in materials where the dielectric function depends on  $\omega$ :  $\varepsilon(\omega)$  can immediately be specified. And there is a further advantage in that we can employ one of the efficient ‘on-shell’ methodologies. In this context the term ‘on-shell’ means that the calculation is confined to a surface or shell in configuration space on which all the states have the same frequency. On-shell methods have commonly been employed in low-energy electron diffraction theory [14]. The analogy with LEED has also been used by Stefanou *et al* [15], and the transfer-matrix approach has been addressed in a different context by Russell *et al* [16].



**Figure 3.** A plan view of the mesh on which the fields are sampled showing the orientation of the axes. Two planes of points are marked by dashed lines: at a fixed frequency the fields on these two sets of planes are related by Maxwell’s equations, so if we know the fields on one of the planes, we can calculate the fields on the other.

Looking back to equations (27), (28): if  $\omega$  is fixed the equations specify a relationship between fields on different sites of the sampling mesh. For example in figure 3 we see a plan view of the sampling mesh. Two sets of planes are marked by dashed lines and if we know the fields on one of these planes we can infer the fields on the other by virtue of equations (27), (28). In fact if we consider a slab of material, then specifying fields on a single plane on one side of the sample infers the fields at all points within the sample. At fixed frequency, given the fields on one side of a sample we can *transfer* the fields throughout the whole sample by successively applying Maxwell’s equations. This result is no more than the discrete equivalent of the well known result for differential equations: if we specify the boundary conditions, then we can integrate the equations throughout the sample.

The detailed derivation of this result from (27), (28) is elementary but tedious and the reader is referred to [12] for details. The result is

$$\begin{aligned}
 E_x(\mathbf{r} + \mathbf{c}) = & E_x(\mathbf{r}) + i c \omega \mu_0 \mu(\mathbf{r}) H_y(\mathbf{r}) + \frac{i}{a \omega \varepsilon_0 \varepsilon(\mathbf{r})} \left[ \begin{array}{l} +\{H_y(\mathbf{r} - \mathbf{a}) - H_y(\mathbf{r})\} \\ -\{H_x(\mathbf{r} - \mathbf{b}) - H_x(\mathbf{r})\} \end{array} \right] \\
 & - \frac{i}{a \omega \varepsilon_0 \varepsilon(\mathbf{r} + \mathbf{a})} \left[ \begin{array}{l} +\{H_y(\mathbf{r}) - H_y(\mathbf{r} + \mathbf{a})\} \\ -\{H_x(\mathbf{r} + \mathbf{a} - \mathbf{b}) - H_x(\mathbf{r} + \mathbf{a})\} \end{array} \right] \quad (36a)
 \end{aligned}$$

$$E_y(\mathbf{r} + \mathbf{c}) = E_y(\mathbf{r}) - i\omega\mu_0\mu(\mathbf{r})H_x(\mathbf{r}) + \frac{i}{a\omega\varepsilon_0\varepsilon(\mathbf{r})} \begin{bmatrix} +\{H_y(\mathbf{r} - \mathbf{a}) - H_y(\mathbf{r})\} \\ -\{H_x(\mathbf{r} - \mathbf{b}) - H_x(\mathbf{r})\} \end{bmatrix} \\ - \frac{i}{a\omega\varepsilon_0\varepsilon(\mathbf{r} + \mathbf{b})} \begin{bmatrix} +\{H_y(\mathbf{r} - \mathbf{a} + \mathbf{b}) - H_y(\mathbf{r} + \mathbf{b})\} \\ -\{H_x(\mathbf{r}) - H_x(\mathbf{r} + \mathbf{b})\} \end{bmatrix} \quad (36b)$$

$$H_x(\mathbf{r} + \mathbf{c}) = H_x(\mathbf{r}) - i\omega\varepsilon_0\varepsilon(\mathbf{r} + \mathbf{c})E_y(\mathbf{r} + \mathbf{c}) \\ + \frac{i}{a\omega\mu_0\mu(\mathbf{r} - \mathbf{a} + \mathbf{c})} \begin{bmatrix} +\{E_y(\mathbf{r} + \mathbf{c}) - E_y(\mathbf{r} - \mathbf{a} + \mathbf{c})\} \\ -\{E_x(\mathbf{r} - \mathbf{a} + \mathbf{b} + \mathbf{c}) - E_x(\mathbf{r} - \mathbf{a} + \mathbf{c})\} \end{bmatrix} \\ - \frac{i}{a\omega\mu_0\mu(\mathbf{r} + \mathbf{c})} \begin{bmatrix} +\{E_y(\mathbf{r} + \mathbf{a} + \mathbf{c}) - E_y(\mathbf{r} + \mathbf{c})\} \\ -\{E_x(\mathbf{r} + \mathbf{b} + \mathbf{c}) - E_x(\mathbf{r} + \mathbf{c})\} \end{bmatrix} \quad (36c)$$

$$H_y(\mathbf{r} + \mathbf{c}) = H_y(\mathbf{r}) + i\omega\varepsilon_0\varepsilon(\mathbf{r} + \mathbf{c})E_x(\mathbf{r} + \mathbf{c}) \\ + \frac{i}{a\omega\mu_0\mu(\mathbf{r} - \mathbf{b} + \mathbf{c})} \begin{bmatrix} +\{E_y(\mathbf{r} + \mathbf{a} - \mathbf{b} + \mathbf{c}) - E_y(\mathbf{r} - \mathbf{b} + \mathbf{c})\} \\ -\{E_x(\mathbf{r} + \mathbf{c}) - E_x(\mathbf{r} - \mathbf{b} + \mathbf{c})\} \end{bmatrix} \\ + \frac{i}{a\omega\mu_0\mu(\mathbf{r} + \mathbf{c})} \begin{bmatrix} +\{E_y(\mathbf{r} + \mathbf{a} + \mathbf{c}) - E_y(\mathbf{r} + \mathbf{c})\} \\ -\{E_x(\mathbf{r} + \mathbf{b} + \mathbf{c}) - E_x(\mathbf{r} + \mathbf{c})\} \end{bmatrix}. \quad (36d)$$

Notice that we work with only two components of each of the fields. The third component is inferred from the 'dead longitudinal mode' which we carefully built into our equations:

$$-(ia)^{-1}[H_y(\mathbf{r} - \mathbf{a}) - H_y(\mathbf{r})] + (ia)^{-1}[H_x(\mathbf{r} - \mathbf{b}) - H_x(\mathbf{r})] = -\omega\varepsilon_0\varepsilon(\mathbf{r})E_z(\mathbf{r}) \\ + (ia)^{-1}[E_y(\mathbf{r} + \mathbf{a}) - E_y(\mathbf{r})] - (ia)^{-1}[E_x(\mathbf{r} + \mathbf{b}) - E_x(\mathbf{r})] = +\omega\mu_0\mu(\mathbf{r})H_z(\mathbf{r}). \quad (37)$$

Equations (36) define the transfer matrix:

$$\begin{bmatrix} E_x(\mathbf{r} + \mathbf{c}) \\ E_y(\mathbf{r} + \mathbf{c}) \\ H_x(\mathbf{r} + \mathbf{c}) \\ H_y(\mathbf{r} + \mathbf{c}) \end{bmatrix} = \sum_{\mathbf{r}'} \begin{bmatrix} \mathbf{T}_{11}(\mathbf{r}, \mathbf{r}') & \mathbf{T}_{12}(\mathbf{r}, \mathbf{r}') & \mathbf{T}_{13}(\mathbf{r}, \mathbf{r}') & \mathbf{T}_{14}(\mathbf{r}, \mathbf{r}') \\ \mathbf{T}_{21}(\mathbf{r}, \mathbf{r}') & \mathbf{T}_{22}(\mathbf{r}, \mathbf{r}') & \mathbf{T}_{23}(\mathbf{r}, \mathbf{r}') & \mathbf{T}_{24}(\mathbf{r}, \mathbf{r}') \\ \mathbf{T}_{31}(\mathbf{r}, \mathbf{r}') & \mathbf{T}_{32}(\mathbf{r}, \mathbf{r}') & \mathbf{T}_{33}(\mathbf{r}, \mathbf{r}') & \mathbf{T}_{34}(\mathbf{r}, \mathbf{r}') \\ \mathbf{T}_{41}(\mathbf{r}, \mathbf{r}') & \mathbf{T}_{42}(\mathbf{r}, \mathbf{r}') & \mathbf{T}_{43}(\mathbf{r}, \mathbf{r}') & \mathbf{T}_{44}(\mathbf{r}, \mathbf{r}') \end{bmatrix} \begin{bmatrix} E_x(\mathbf{r}') \\ E_y(\mathbf{r}') \\ H_x(\mathbf{r}') \\ H_y(\mathbf{r}') \end{bmatrix} \quad (38)$$

where the summation is over all points in a plane.

In the case where we have a periodic structure which repeats after  $L_z$  cells we can relate the fields on either side of the unit cell and exploit this to calculate the band structure:

$$\begin{bmatrix} E_x(\mathbf{r} + L_z\mathbf{c}) \\ E_y(\mathbf{r} + L_z\mathbf{c}) \\ H_x(\mathbf{r} + L_z\mathbf{c}) \\ H_y(\mathbf{r} + L_z\mathbf{c}) \end{bmatrix} = \sum_{\mathbf{r}'} \mathbf{T}(L_z, 0) \begin{bmatrix} E_x(\mathbf{r}') \\ E_y(\mathbf{r}') \\ H_x(\mathbf{r}') \\ H_y(\mathbf{r}') \end{bmatrix} \quad (39)$$

where

$$\mathbf{T}(L_z, 0) = \prod_{j=1}^{L_z} \mathbf{T}(j, j-1). \quad (40)$$

Hence on applying Bloch's theorem to the periodic structure,

$$\begin{bmatrix} E_x(\mathbf{r} + L_z\mathbf{c}) \\ E_y(\mathbf{r} + L_z\mathbf{c}) \\ H_x(\mathbf{r} + L_z\mathbf{c}) \\ H_y(\mathbf{r} + L_z\mathbf{c}) \end{bmatrix} = \exp(iK_z L_z a) \begin{bmatrix} E_x(\mathbf{r}) \\ E_y(\mathbf{r}) \\ H_x(\mathbf{r}) \\ H_y(\mathbf{r}) \end{bmatrix} \quad (41)$$

we identify the Bloch waves as eigenvectors of the transfer matrix:

$$\sum_{\mathbf{r}'} \mathbf{T}(L_z, 0) \begin{bmatrix} E_x(\mathbf{r}') \\ E_y(\mathbf{r}') \\ H_x(\mathbf{r}') \\ H_y(\mathbf{r}') \end{bmatrix} = \exp(iK_z L_z a) \begin{bmatrix} E_x(\mathbf{r}) \\ E_y(\mathbf{r}) \\ H_x(\mathbf{r}) \\ H_y(\mathbf{r}) \end{bmatrix}. \quad (42)$$

Thus we can find all the Bloch waves at a given *frequency*. Contrast this with equation (13) which gives all the Bloch waves at a fixed *wave vector*. Obviously (42) has the advantage if  $\varepsilon$  depends on  $\omega$  since the eigenvalue equation remains unchanged because  $\omega$  has already been specified, so we know what value of  $\varepsilon$  to write in the equations.

Transfer-matrix methods have another advantage in that they generally work with smaller matrices than would otherwise be required. Suppose that we have a periodic system in which the unit cell is subdivided into a mesh of  $10 \times 10 \times 10$  sampling points. We could diagonalize equations (27), (28) to find the eigenfrequencies of the system. The relevant matrices would have dimensions  $6000 \times 6000$  reflecting the 6000 independent field components of the system. This would be a huge numerical calculation requiring the services of a supercomputer. Transforming into  $k$ -space would bring no relief because to give a similar level of precision (13) would have to employ a similar number of plane waves in the Fourier expansion. In contrast, transfer matrices for this system have dimensions  $400 \times 400$  and can easily be diagonalized on a workstation or even a highly specified personal computer.

We can calculate the scaling of computing times with size of system. For off-shell methods

$$\tau_{off\ shell} \sim (L_x L_y L_z)^3 \quad (43)$$

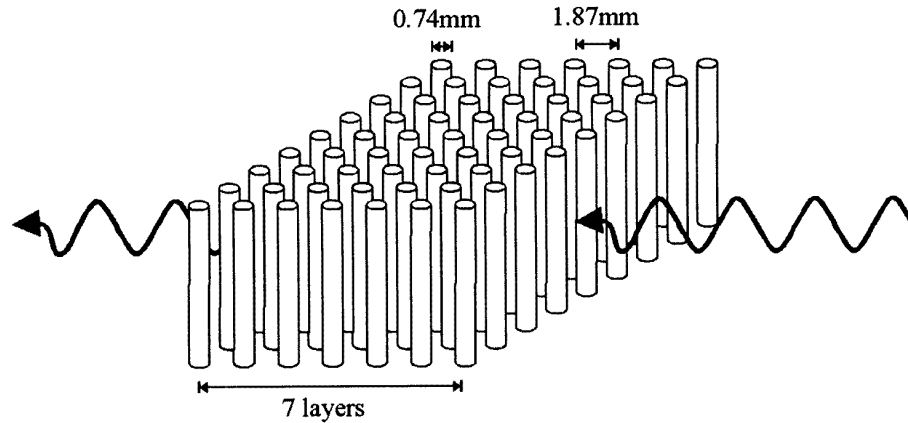
whereas for on-shell methods

$$\tau_{on\ shell} \sim (L_x L_y)^3 L_z \quad (44)$$

representing a considerable saving of time in systems which require many layers of mesh points to describe their structure. In our example above the saving is approximately a factor of  $10^2 = 100$ .

Because they calculate all the states at a given frequency, transfer-matrix methods are very well suited to calculating the scattered wave field from a complex object, and it was transfer-matrix methods which opened up this aspect of the subject. Robertson *et al* [17] have measured the response of a photonic system in the microwave band. Their structure consisted of an array of cylinders and is shown in figure 4.

In [13] the unit cell of the system was divided into a  $10 \times 10 \times 1$  mesh: for each cell an average was taken over the dielectric constant within that cell. The transfer matrix was found by multiplying the matrices for each of the ten slices. Its eigenvalues give  $K_z(\omega, K_x, K_y)$ . For the transmission coefficient a similar division of the cell was made, but multiplying transfer matrices for the seven layers of unit cells would have led to numerical instability, so instead multiplication was halted after integrating through one unit cell at which point the transmission and reflection coefficients were calculated for a slab one cell thick. Slabs were then stacked together using the multiple-scattering formula familiar in the theory of low-energy electron diffraction [14]. Convergence was tested by repeating the calculation for a  $20 \times 20 \times 1$  mesh: changes of the order of 1% at 80 GHz were found in the band structure. The results are shown compared to the experiment in figure 5. The computer codes used in this calculation have been published [18].



**Figure 4.** A schematic version of the microwave experiment performed by Robertson *et al.* Microwaves are incident on a simple cubic array of cylinders, with  $\epsilon = 8.9$ , seven layers deep but effectively infinite in the directions perpendicular to the beam. The amplitude of the transmitted beam is measured.

### 5. How to solve the discrete Maxwell equations using ‘order- $N$ ’ methods

When handling very large systems, the way in which computing times scale with system size is the determining factor in their efficiency. Obviously the optimum scaling of a system with  $N$  independent components must be proportional to  $N$  because it would take at least this time to write down the problem. As computers have become more powerful and able to treat very large systems the importance of achieving an ‘order- $N$ ’ methodology has been realized.

In the electron band-structure problem such methodologies have been realized for tight-binding systems in which the Hamiltonian reduces to a sparse matrix. Formally containing  $N^2$  elements, the Hamiltonian would appear to lead to computation times at least as large as  $N^2$  but is saved from this fate by only order- $N$  components being non-zero. Therefore any method based on a finite number of matrix  $\times$  vector multiplications, which are now of order  $N$ , gives the desired optimum scaling.

In the context of photonic band structure the formulation of order- $N$  methodologies awaited construction of a tight-binding formalism. The real-space formulation described in section 3 gives rise to sparse matrices in exactly the desired manner. Each matrix defined by (27), (28) contains components where  $N$  is the number of mesh points in the system. These possibilities were first exploited by Chan, Yu, and Ho [19].

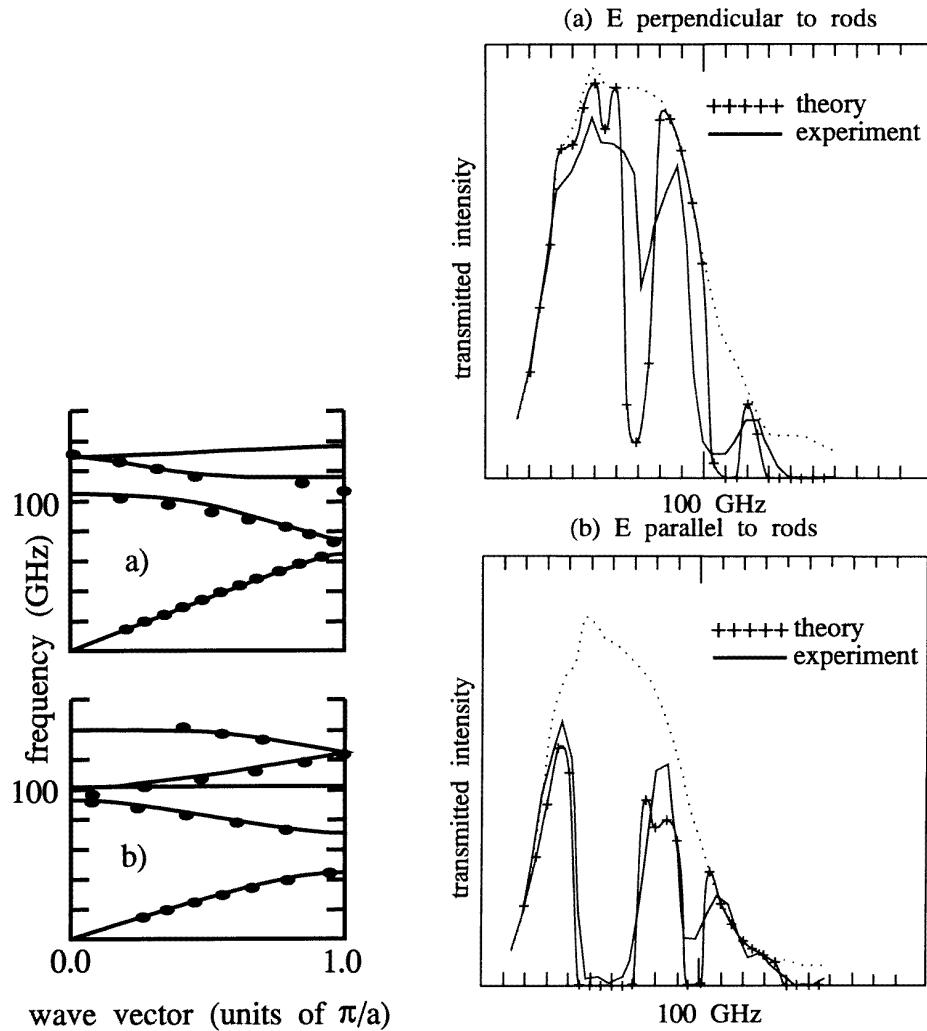
The idea is as follows: if we start with a set of fields, arbitrary except in that they obey the Bloch condition for wave vector  $\mathbf{K}$ :

$$\mathbf{E}_{\mathbf{K}}(t=0) \quad \mathbf{H}_{\mathbf{K}}(t=0) \quad (45)$$

and use Maxwell’s equations to calculate their time evolution, we can in effect find the band structure of the system. Since we know that we can expand the arbitrary fields in terms of the Bloch waves:

$$\mathbf{E}_{\mathbf{K}}(t) = \sum_j b_{\mathbf{K}j} \mathbf{E}_j \exp(-i\omega_{\mathbf{K}j}t) \quad \mathbf{H}_{\mathbf{K}}(t) = \sum_j b_{\mathbf{K}j} \mathbf{H}_j \exp(-i\omega_{\mathbf{K}j}t). \quad (46)$$

Hence Fourier transforming the fields with respect to time picks out a series of delta



**Figure 5.** Left: the dispersion relation for propagation of electromagnetic waves along the (10) direction of a 2D array of dielectric cylinders. (a)  $E$  perpendicular to the cylinders; (b)  $E$  parallel to the cylinders. The experimental results are shown as black dots and have a resolution of 5 GHz. Note that very narrow bands are not resolved experimentally. Right: the transmitted power for an array of seven rows of dielectric cylinders. The dotted curve shows the instrument response in the absence of the cylinders. Upper curves (a):  $E$  perpendicular to the cylinders; lower curves (b):  $E$  parallel to the cylinders.

functions located at the Bloch wave frequencies. The resolution of these frequencies depends on the time interval over which integration proceeds. One must be careful that the starting fields contain a finite component of the desired Bloch wave.

We need to reformulate Maxwell's equations so as to represent them not only on a mesh which is discrete in space, but also on one that is discrete in time. To do this we go back to (27), (28) and further approximate these equations so that when transformed back into the time domain, they result in finite-difference equations. We use the same tricks that we

used in the spatial case. The approximation

$$w_E = (-iT)^{-1}[\exp(-i\omega T) - 1] \approx \omega \quad w_H = (+iT)^{-1}[\exp(+i\omega T) - 1] \approx \omega \quad (47)$$

gives a set of equations that conserve energy as before:

$$\begin{aligned} +[E_z(\mathbf{r} + \mathbf{b}) - E_z(\mathbf{r})] - [E_y(\mathbf{r} + \mathbf{c}) - E_y(\mathbf{r})] &= +ia w_H \mu_0 H_x(\mathbf{r}) \\ +[E_x(\mathbf{r} + \mathbf{c}) - E_x(\mathbf{r})] - [E_z(\mathbf{r} + \mathbf{a}) - E_z(\mathbf{r})] &= +ia w_H \mu_0 H_y(\mathbf{r}) \\ +[E_y(\mathbf{r} + \mathbf{a}) - E_y(\mathbf{r})] - [E_x(\mathbf{r} + \mathbf{b}) - E_x(\mathbf{r})] &= +ia w_H \mu_0 H_z(\mathbf{r}) \end{aligned} \quad (48a)$$

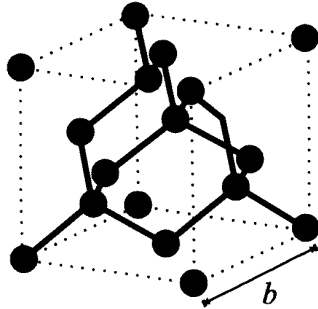
$$\begin{aligned} -[H_z(\mathbf{r} - \mathbf{b}) - H_z(\mathbf{r})] + [H_y(\mathbf{r} - \mathbf{c}) - H_y(\mathbf{r})] &= -ia w_E \varepsilon_0 \varepsilon(\mathbf{r}) E_x(\mathbf{r}) \\ -[H_x(\mathbf{r} - \mathbf{c}) - H_x(\mathbf{r})] + [H_z(\mathbf{r} - \mathbf{a}) - H_z(\mathbf{r})] &= -ia w_E \varepsilon_0 \varepsilon(\mathbf{r}) E_y(\mathbf{r}) \\ -[H_y(\mathbf{r} - \mathbf{a}) - H_y(\mathbf{r})] + [H_x(\mathbf{r} - \mathbf{b}) - H_x(\mathbf{r})] &= -ia w_E \varepsilon_0 \varepsilon(\mathbf{r}) E_z(\mathbf{r}) \end{aligned} \quad (48b)$$

and in the time domain

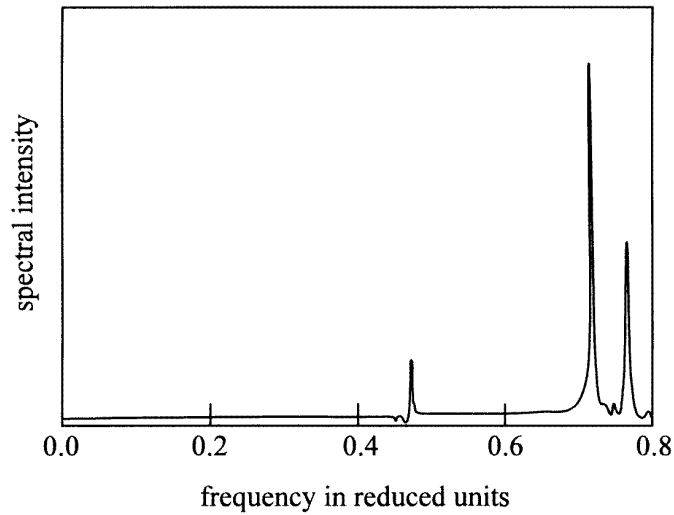
$$\begin{aligned} +[E_z(\mathbf{r} + \mathbf{b}, t) - E_z(\mathbf{r}, t)] - [E_y(\mathbf{r} + \mathbf{c}, t) - E_y(\mathbf{r}, t)] \\ = +aT^{-1} \mu_0 H_x(\mathbf{r}, t - T) - H_x(\mathbf{r}, t) \\ +[E_x(\mathbf{r} + \mathbf{c}, t) - E_x(\mathbf{r}, t)] - [E_z(\mathbf{r} + \mathbf{a}, t) - E_z(\mathbf{r}, t)] \\ = +aT^{-1} \mu_0 [H_y(\mathbf{r}, t - T) - H_y(\mathbf{r}, t)] \\ +[E_y(\mathbf{r} + \mathbf{a}, t) - E_y(\mathbf{r}, t)] - [E_x(\mathbf{r} + \mathbf{b}, t) - E_x(\mathbf{r}, t)] \\ = +aT^{-1} \mu_0 [H_z(\mathbf{r}, t - T) - H_z(\mathbf{r}, t)] \end{aligned} \quad (49a)$$

$$\begin{aligned} -[H_z(\mathbf{r} - \mathbf{b}, t) - H_z(\mathbf{r}, t)] + [H_y(\mathbf{r} - \mathbf{c}, t) - H_y(\mathbf{r}, t)] \\ = +aT^{-1} \varepsilon_0 \varepsilon(\mathbf{r}) [E_x(\mathbf{r}, t + T) - E_x(\mathbf{r}, t)] \\ -[H_x(\mathbf{r} - \mathbf{c}, t) - H_x(\mathbf{r}, t)] + [H_z(\mathbf{r} - \mathbf{a}, t) - H_z(\mathbf{r}, t)] \\ = +aT^{-1} \varepsilon_0 \varepsilon(\mathbf{r}) [E_y(\mathbf{r}, t + T) - E_y(\mathbf{r}, t)] \\ -[H_y(\mathbf{r} - \mathbf{a}, t) - H_y(\mathbf{r}, t)] + [H_x(\mathbf{r} - \mathbf{b}, t) - H_x(\mathbf{r}, t)] \\ = +aT^{-1} \varepsilon_0 \varepsilon(\mathbf{r}) [E_z(\mathbf{r}, t + T) - E_z(\mathbf{r}, t)]. \end{aligned} \quad (49b)$$

These equations can be used to advance the fields in time steps of  $T$ .

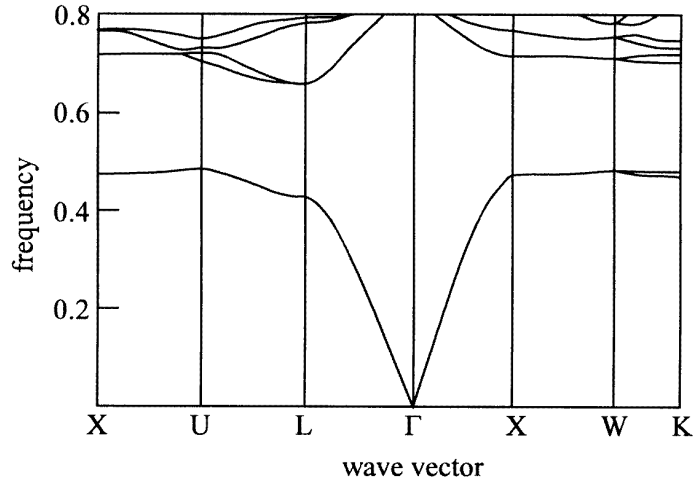


**Figure 6.** The unit cell of the photonic structure consisting of a series of cylinders, with  $\varepsilon = 12.96$ , connecting the points of a diamond lattice. The diameter of the cylinders is adjusted so that they occupy 20% of the volume. To simplify the figure the lattice points are shown as circles, and the cylinders are not drawn to scale.



**Figure 7.** The spectral intensity at the X point of the diamond structure. The frequency is given in units of  $c_0/b$  where  $b$  is the lattice constant.

Using methods similar to the one described above, Chan *et al* made calculations for a structure shown in figure 6: an array of cylinders packed in a diamond lattice with periodic boundary conditions imposed on the starting wavefunction. The sampling grid was a  $48 \times 48 \times 48$  mesh. The time interval used was  $T = 0.003 b/c_0$  where  $b$  is the lattice constant, and the integration continued through  $10^5$  steps.



**Figure 8.** The photonic band structure of a diamond structure generated with the spectral method. The frequency is given in units of  $c_0/b$  where  $b$  is the lattice constant.

The time sequence was the Fourier analysed to give the spectrum shown in figure 7 where the boundary conditions are set for the X point in the Brillouin zone. Repeating the exercise for other  $K$ -points results in the band structure shown in figure 8.



These results were checked against calculations made using a plane-wave expansion and agreed very well. The main difference was the length of time taken to make the calculation.

The time evolution method has a further advantage over the plane-wave expansion methods, one which it shares with transfer-matrix methods: it can be used to calculate reflection coefficients of objects by choosing the initial condition to be a pulse of radiation incident on the object in question, allowing the pulse to evolve with time until it has finished interacting with the object, then Fourier transforming the result to get the response at each frequency.

The time taken for calculations using the time evolution operator is proportional to

$$\tau_{time\ evol} = (\text{number of mesh points}) \times (\text{number of time steps}) \quad (50)$$

and hence is ‘order  $N$ ’ if  $N$  is taken to be the number of mesh points and the number of time steps is independent of  $N$ . The latter condition is fulfilled for a wide class of systems:

- when absorption is finite, so any incident radiation is removed after a finite interval of time;
- when the system scatters only weakly, so radiation escapes from the system after a finite length of time.

However, there are situations where the number of time steps does depend on the number of mesh points. For example in a large system in which no absorption is present, radiation can scatter to and fro within the system-defining structure on an increasingly fine frequency scale, and requiring many time steps before it can be resolved. This happens in so-called ‘Anderson localized’ systems where radiation can only escape from a disordered system in a time which increases exponentially with system size. It also is the case for ‘defect states’ enclosed within a photonic band-gap material—at least if we wish to estimate the lifetime of the defect which grows exponentially with system size. For this class of system, methods based on the time evolution operator have a disastrous scaling law:

$$\tau_{time\ evol}(\text{localized system}) = (\text{number of mesh points}) \times \exp(\text{number of mesh points}). \quad (51)$$

Not surprisingly these methods are not used to study Anderson localization where transfer-matrix techniques are the order of the day.

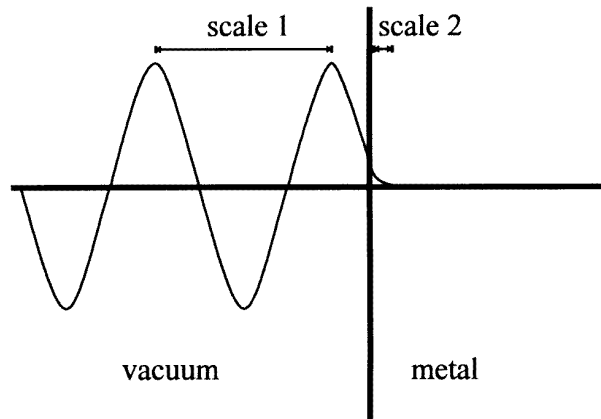
In a system which is not localized, in which there are no photonic band gaps and radiation diffuses around a complex structure, the time to escape will be determined by the diffusion equation and be proportional to the square of the linear dimensions:

$$\begin{aligned} \tau_{time\ evol}(\text{delocalized}) &= (\text{number of mesh points}) \times (\text{number of mesh points})^{2/3} \\ &= (\text{number of mesh points})^{4/3}. \end{aligned} \quad (52)$$

Although not ‘order  $N$ ’, this is still a very favourable scaling law.

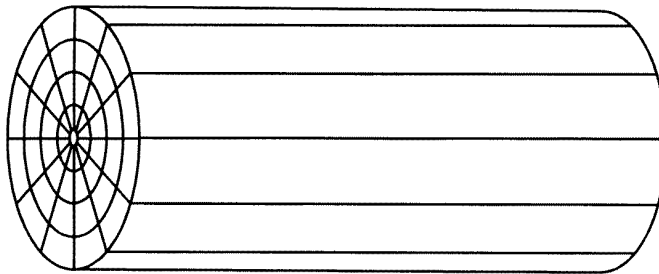
## 6. Maxwell’s equations on a non-uniform mesh

By and large, electronic structure calculations all deal with the same basic structure: an assembly of atomic potentials. Although this presents its own difficulties, it sits in contrast to the very wide variety of structures and of materials with which we work in photonic band structure. Most of the methods described above work well in situations where there is only one length scale. For example in the array of dielectric cylinders shown in figure 4, electric and magnetic fields can be described accurately by sampling on a uniform mesh.



**Figure 9.** At a vacuum–metal interface there is a scale change in the wave field. In the vacuum, the scale is determined by the wavelength, and in the metal by the skin depth.

In contrast, systems which contain metals with finite conductivity have two length scales: the wavelength of radiation outside the metal, and the skin depth inside, often differing by many orders of magnitude: see figure 9. Unless the wave field is correctly described inside the metal, estimates of energy dissipation will be in error. This requires either an impossibly dense uniform mesh, or a non-uniform mesh adapted to the shape of the metal.



**Figure 10.** An adaptation of the simple cubic mesh to the geometry of a fibre. The  $z$  component of the mesh can be taken to be uniform.

Another instance where a uniform mesh is inappropriate occurs when the material has some symmetry which is relevant to its function but which is incompatible with the symmetry of a uniform simple cubic mesh. For example, in an optical fibre, cylindrical symmetry is a central concept. Here we may wish to choose a cylindrical mesh adapted to this symmetry as shown in figure 10.

At first sight it may seem that we have a very tiresome task ahead of us: that of reworking all our formalism for every different mesh we encounter. In fact we are saved by a most elegant result: if Maxwell's equations are rewritten in a new coordinate system they take exactly the same form as in the old system provided that we renormalize, and according to a simple prescription [20]. This affords a huge saving in programming effort because it reduces what appears to be a new problem to an old one: that of solving for the wave field on a uniform mesh in a non-uniform medium.

Consider a general coordinate transformation:

$$q_1(x, y, z) \quad q_2(x, y, z) \quad q_3(x, y, z). \quad (53)$$

After some algebra we can rewrite Maxwell's equations in the new system:

$$(\nabla_q \times \hat{\mathbf{E}})^i = -\mu_0 \sum_{j=1}^3 \hat{\mu}^{ij} \frac{\partial \hat{H}_j}{\partial t} \quad (54a)$$

$$(\nabla_q \times \hat{\mathbf{H}})^i = +\varepsilon_0 \sum_{j=1}^3 \hat{\varepsilon}^{ij} \frac{\partial \hat{E}_j}{\partial t} \quad (54b)$$

which as promised are identical to the familiar equations written in a Cartesian system of coordinates. The renormalized quantities are

$$\hat{\varepsilon}^{ij} = \varepsilon g^{ij} |\mathbf{u}_1 \cdot (\mathbf{u}_2 \times \mathbf{u}_3)| Q_1 Q_2 Q_3 (Q_i Q_j)^{-1} \quad (55a)$$

$$\hat{\mu}^{ij} = \mu g^{ij} |\mathbf{u}_1 \cdot (\mathbf{u}_2 \times \mathbf{u}_3)| Q_1 Q_2 Q_3 (Q_i Q_j)^{-1} \quad (55b)$$

where  $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$  are three unit vectors pointing along the  $q_1, q_2, q_3$  axes respectively, and

$$\mathbf{g}^{-1} = \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \mathbf{u}_1 \cdot \mathbf{u}_2 & \mathbf{u}_1 \cdot \mathbf{u}_3 \\ \mathbf{u}_2 \cdot \mathbf{u}_1 & \mathbf{u}_2 \cdot \mathbf{u}_2 & \mathbf{u}_2 \cdot \mathbf{u}_3 \\ \mathbf{u}_3 \cdot \mathbf{u}_1 & \mathbf{u}_3 \cdot \mathbf{u}_2 & \mathbf{u}_3 \cdot \mathbf{u}_3 \end{bmatrix} \quad (56)$$

$$Q_{ij} = \frac{\partial x}{\partial q_i} \frac{\partial x}{\partial q_j} + \frac{\partial y}{\partial q_i} \frac{\partial y}{\partial q_j} + \frac{\partial z}{\partial q_i} \frac{\partial z}{\partial q_j} \quad (57)$$

$$Q_i^2 = Q_{ii}. \quad (58)$$

This result has great value in calculations based on real-space methods as it enables all the original computer programs written for a uniform mesh to be exploited without change when an adaptive mesh is employed. Less obviously it is also of value in the plane-wave expansion methods too: all that is necessary is to apply the plane-wave expansion in  $q$ -space to the new version of Maxwell's equations and all the benefits of the adaptive mesh accumulate in the enhanced convergence of the Fourier expansion.

We remark in passing that similar transformations can be used in the case of the Schrödinger equation which also retains its form under a coordinate transformation. However, in this case one must consider the most general form which includes a vector potential as well as a scalar potential because curvature of the coordinate system implies an effective magnetic field.

Equations (54) raise an intriguing possibility. Suppose that we set up a coordinate system in free space that is non-uniform in some finite region of space, but far away from this region reverts to a uniform Cartesian system. Obviously the results of any scattering experiment in which radiation enters the region in question from infinity cannot depend on the choice of coordinate system. Whatever system we choose, must return the result that radiation is not scattered by what after all remains free space.

Now let us change our viewpoint. Suppose that the 'real system' contains a material for which

$$\begin{aligned} \hat{\varepsilon}^{ij} &= g^{ij} |\mathbf{u}_1 \cdot (\mathbf{u}_2 \times \mathbf{u}_3)| Q_1 Q_2 Q_3 (Q_i Q_j)^{-1} \\ \hat{\mu}^{ij} &= g^{ij} |\mathbf{u}_1 \cdot (\mathbf{u}_2 \times \mathbf{u}_3)| Q_1 Q_2 Q_3 (Q_i Q_j)^{-1}. \end{aligned} \quad (59)$$

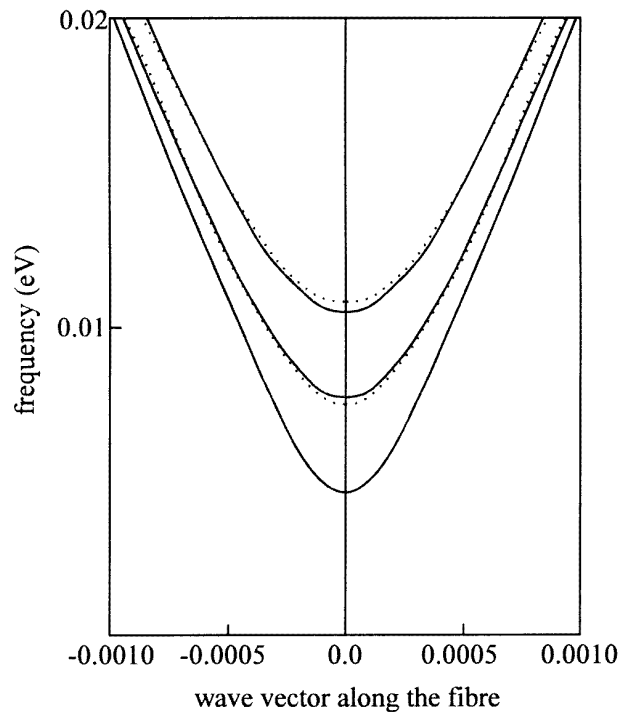
Then the electromagnetic influence of the medium can be undone by the reverse coordinate transformation: we can find a system in which  $\varepsilon$  and  $\mu$  are everywhere unity and which has

no observable influence on electromagnetic radiation. The object in question is therefore invisible.

In fact the restriction that the coordinate transformation be embedded in a flat 3-space is a considerable one, and for most objects we cannot completely undo their scattering in this way even if the other condition we require:

$$\varepsilon = \mu \quad (60)$$

is met. However, there are instances where this is possible: a sheet of material is perfectly non-reflecting to normally incident radiation provided that condition (60) is true. This observation is the basis of much stealth technology since only radar signals at normal incidence have a chance of returning to the transmitter.



**Figure 11.** The lowest three modes of a  $10^{-4}$  m radius cylindrical metal waveguide for angular momentum  $m = 0$ , plotted against wave vector [20]. The full lines are the results of a numerical calculation using a cylindrical mesh; the dotted lines are the exact results of analytic theory.

Finally we show in figure 11 this scheme in action for a cylindrical waveguide with perfect metal boundaries. One advantage of the cylindrical mesh is that angular momentum is now a good quantum number in the computation, and in addition cylindrical symmetry can be exploited to speed up the calculation. The computations are compared to the analytic result and show that the distorted mesh shows the same excellent convergence as the original uniform version.

## 7. Metallic photonic structures

Metals are different. Although their properties can be represented by an effective dielectric function they are unique amongst materials in that this function can be essentially negative. For example, nearly-free-electron metals such as aluminium can be modelled by

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega + i\gamma)} \quad (61)$$

where  $\omega_p$  is the plasma frequency, and  $\gamma$  is related to the conductivity,  $\sigma$ , by

$$\gamma = \frac{\omega_p^2 \varepsilon_0}{\sigma}. \quad (62)$$

If the conductivity is large, and  $\omega$  is below the plasma frequency, then  $\varepsilon(\omega)$  is essentially negative. This seemingly innocent property of a metal has a profound impact on its electro-magnetic properties.

Let us work in the electrostatic limit, ignoring retardation for the moment. We have the requirement that

$$\nabla \cdot \mathbf{D}(\mathbf{r}) = \nabla \cdot [\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] = 0 \quad (63)$$

where  $\mathbf{D}$  is the displacement field and  $\phi$  is the scalar potential. Normally this implies that there are no localized electrostatic modes in the absence of external charge, and that in the presence of external charges the fields are unique: see p 42 of [2]. Consider a system containing no external charges, but which contains a localized region where  $\varepsilon$  is non-unity. Enclose the region by a surface  $\Gamma$  which lies arbitrarily far from the region of interest. Normally we can show that if the potential and its gradient on this surface are zero, then the fields everywhere are zero. In other words, fields inside the surface have to be induced by sources outside the surface, whose fields will pass through the surface itself. Consider

$$\int_{\Gamma} \phi(\varepsilon_0 \varepsilon \nabla \phi) \cdot d\mathbf{S} = 0 \quad (64)$$

which is zero in the limit in which the surface is infinitely far from the region containing the electrostatic activity. From Gauss's theorem this implies

$$\begin{aligned} 0 &= \int_V \phi \nabla \cdot (\varepsilon_0 \varepsilon \nabla \phi) + \varepsilon_0 \varepsilon (\nabla \phi) \cdot (\nabla \phi) dV \\ &= \int_V \phi \nabla \cdot \mathbf{D} + \varepsilon_0 \varepsilon \mathbf{E} \cdot \mathbf{E} dV = \int_V \varepsilon_0 \varepsilon \mathbf{E} \cdot \mathbf{E} dV. \end{aligned} \quad (65)$$

If  $\varepsilon$  is everywhere positive, then the implication is that

$$\mathbf{E} = 0 \quad \text{everywhere.} \quad (66)$$

However, if the surface  $\Gamma$  encloses a photonic metallic structure within which  $\varepsilon$  is alternately negative and positive in different parts of the structure, then the theorem no longer holds. Metallic systems may support localized resonances whose fields are entirely confined to the photonic region, at least in the electrostatic limit. In fact these modes can be thought of as originating in surface plasmons [21] decorating the surfaces of the photonic structure, often shifted in frequency by strong interaction between one surface and another. These localized resonances are a dominant feature of structured metallic systems as discussed by Lamb *et al* [22].

One further general result is of interest. It concerns the frequencies of these localized resonances. If we have a non-zero solution to equation (63), then by making the substitution

$$\mathbf{r} = \alpha \mathbf{r}' \quad (67)$$

we can find another solution:

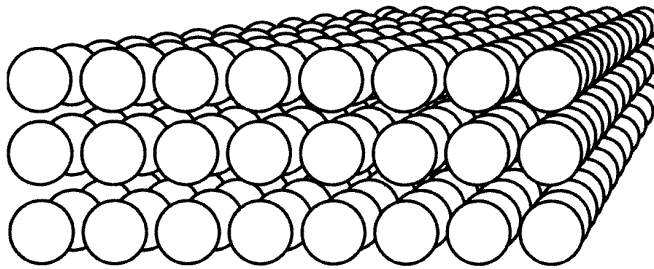
$$0 = \nabla \cdot \mathbf{D}(\mathbf{r}) = \nabla \cdot [\varepsilon(\omega, \mathbf{r}) \nabla \phi(\mathbf{r})] = \alpha^{-1} \nabla' \cdot [\varepsilon(\omega, \alpha \mathbf{r}') \alpha^{-1} \nabla' \phi(\alpha \mathbf{r}')] \quad (68)$$

and therefore

$$\nabla \cdot [\varepsilon(\omega, \alpha \mathbf{r}) \nabla \phi(\alpha \mathbf{r})] = 0 \quad (69)$$

so having found a localized resonance for one structure at a frequency  $\omega$ , we know that all structures with the same shape have a resonance at exactly the same frequency. This is because Laplace's equation contains no natural length scale.

If we now generalize away from the electrostatic limit, the localized resonances will in general couple to externally incident electromagnetic radiation. If the photonic structure is on a scale much finer than the wavelength, the coupling will be relatively small and therefore the size of the structure does in fact change the nature of the resonances if we go beyond the electrostatic limit: it alters the external coupling of the localized modes.



**Figure 12.** An ordered simple cubic structure composed of  $16 \text{ \AA}$  diameter aluminium spheres, with lattice spacing  $52.9 \text{ \AA}$ .

A good illustration of these effects is seen in an 'ordered colloid' as shown in figure 12.

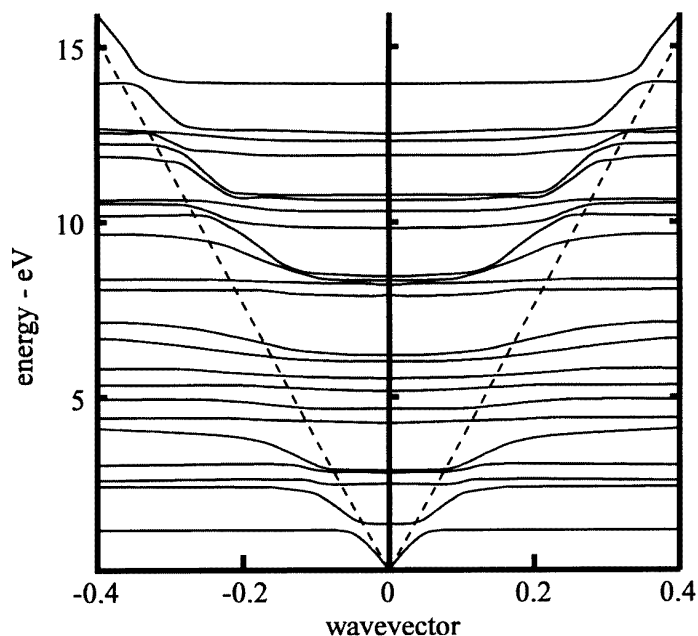
Aluminium has a plasma frequency of  $15 \text{ eV}$  and in the first instance we shall assume no dissipation from electrical resistance so that we can substitute into (61)

$$\omega_p = 15 \text{ eV} \quad \gamma = 0. \quad (70)$$

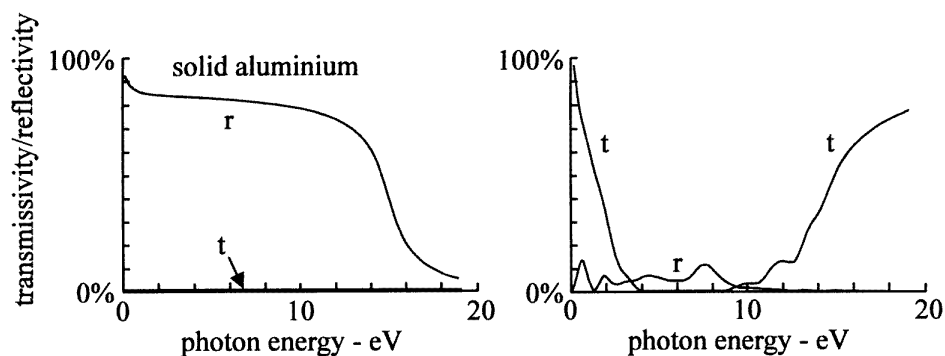
We are now in a position to calculate the photonic band structure of this object which is shown in figure 13, taken from [12].

As we predicted, the presence of structured metal introduces a host of resonances into the system, here manifesting themselves as very flat bands which, again as predicted, hybridize weakly with the transverse modes which intersect them. A structure with twice the lattice spacing and twice the sphere diameter would show very similar structure but would have stronger interaction with the transverse modes.

Reintroducing the electrical resistance of the metal leads to the localized resonances now being able to dissipate energy and, since there are a great number of resonances, this will lead to strong absorption of incident light. In fact metallic colloids, such as the colloidal silver familiar from photographic plates, are extremely black in appearance. We can show this quantitatively by calculating the reflection and transmission coefficients of a thin layer of this colloidal structure. The results are shown in figure 14. Clearly the colloid, in contrast to the metal, neither transmits nor reflects in a wide band of frequencies from the near infra-red into the ultraviolet. This is a direct consequence of the new resonant modes introduced by the colloidal structure.

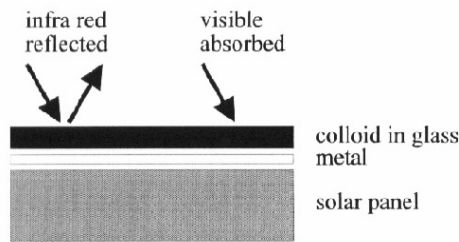


**Figure 13.** The electromagnetic band structure of a simple cubic array of aluminium spheres calculated for loss-free metal with a dielectric constant given by (61). Note the characteristic structure which consists of a large number of extremely flat bands: these are surface plasma modes of the metal spheres, found in the range  $15/\sqrt{3}$  eV to  $15/\sqrt{2}$  eV for isolated spheres, but in this instance spread to a lower range of energies by interaction between spheres. The free-space dispersion relation is shown as a dashed line, and clearly interacts strongly with the surface modes.

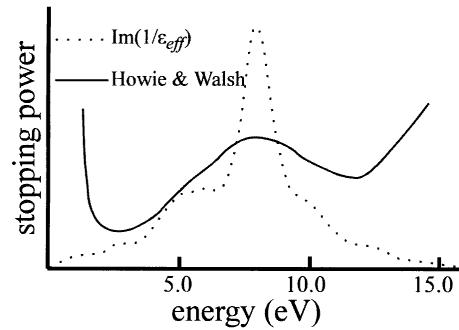


**Figure 14.** Reflectivity and transmissivity of solid aluminium, left, and that of a colloidal array of 16 Å radius aluminium spheres filling 12% of the sample volume. The thickness of material considered is 3385 Å.

These resonances are also believed to be responsible for surface-enhanced Raman scattering observed at rough silver surfaces [23, 24].



**Figure 15.** A schematic figure of a solar panel showing how the strong absorption in the visible, and weak absorption in the infra-red, of a metal colloid is exploited to give an efficient solar panel.



**Figure 16.** The stopping power, as a function of the energy quantum lost, observed by Howie and Walsh [25] for 50 keV electrons passing through an aluminium colloid, compared to our theoretical estimate based on the band structure shown in figure 13.

An interesting aside is that this effect has been exploited for some time in solar heating panels. Coating a layer of polished metal with metal colloid suspended in glass gives a composite which is strongly absorbing in the visible and near infra-red, but reflecting in the far infra-red. Hence the panel absorbs solar radiation efficiently, but retains the heat because it is a bad emitter in the far infra-red. Figure 15 shows a schematic arrangement.

This high density of states means that colloids are very effective at extracting energy from fast charged particles [26, 27]. Essentially this can be regarded as Čerenkov radiation. Reference to figure 13 shows that the group velocity of the resonant modes, as defined by  $d\omega/dk$ , is much less than the velocity of light, given by the slope of the dotted lines. Therefore a moderately fast particle such as a 50 keV electron will strongly Čerenkov radiate. We show in figure 16 some electron energy-loss data obtained by Howie and Walsh [25] in experiments on aluminium colloids compared to theory given by Martín-Moreno, and Pendry [28, 29]. There is very strong absorption in the region of the resonant modes.

## 8. Conclusions

I hope that this review has shown photonic theory to be in good shape, despite its rapid development. Certainly we have an excellent capability for calculation of band structure in periodic dielectrics and the main challenge in photonic band-gap materials is their manufacture and integration with electronic technology. There are many promising developments in this respect and it cannot be long before the first structures having band gaps at visible frequencies emerge.

For the moment the microwave region of the spectrum is the most accessible experimentally and provides a testing ground for theory. However, there are many potentially valuable applications to microwave technology particularly of complex metallic structures, and this area of application should be recognized as valuable in its own right.

Another experimental tool which has much to offer is the electron microscope. Long used to study microstructured materials, electron microscopes have also developed sophisticated methods for measuring electron energy loss of the order of 10 eV or 20 eV and provide a valuable tool for studying the density of electromagnetic states in structured materials. A community of theorists concerned with the impact of radiation on solids [26]



has for some time been studying what are essentially photonic effects, and have much to contribute to the debate.

For the future, theory still has many challenges. This is particularly true for metallic structures, which as we have indicated show unique and complex effects. The computational methodology for handling such systems has only recently become available through the adaptive-mesh approach [20] and will be applied in the next few years to probing new structures, especially those of microwave significance.

Finally there is the link between electromagnetic properties and forces acting within photonic structures. These vary from the relatively weak Van der Waals forces acting between dielectric spheres, which may play a role in self-assembly, to the much more powerful dispersion forces acting between metallic surfaces separated by distances of only a few ångströms. This electromagnetic analogue of the electronic total energy calculation remains to be investigated in detail. The computational techniques are all in place and await application.

## References

- [1] Vos W L, Sprik R, van Blaarderen A, Imhof A, Lagendijk A and Wegdam G H 1996 *Phys. Rev. Lett.* at press
- [2] Panofsky W K H and Phillips M 1962 *Classical Electricity and Magnetism* (Reading, MA: Addison-Wesley)
- [3] Jackson J D 1962 *Classical Electrodynamics* (New York: Wiley)
- [4] Ho K M, Chan C T and Soukoulis C M 1991 *Phys. Rev. Lett.* **65** 3152
- [5] Leung K M and Liu Y F 1991 *Phys. Rev. Lett.* **65** 2646
- [6] Yablonovitch E, Gmitter T J and Leung K M 1991 *Phys. Rev. Lett.* **67** 2295
- [7] Zhang Z and Satpathy S 1991 *Phys. Rev. Lett.* **65** 2650
- [8] Haus J W 1994 *J. Mod. Opt.* **41** 195
- [9] Yablonovitch E 1987 *Phys. Rev. Lett.* **58** 2059
- [10] Joannopoulos J D, Meade R D and Winn J N 1995 *Photonic Crystals* (Princeton, NJ: Princeton University Press)
- [11] Yablonovitch E 1993 *J. Phys.: Condens. Matter* **5** 2443
- [12] Pendry J B 1994 *J. Mod. Opt.* **41** 209
- [13] Pendry J B and MacKinnon A 1992 *Phys. Rev. Lett.* **69** 2772
- [14] Pendry J B 1974 *Low Energy Electron Diffraction* (London: Academic)
- [15] Stefanou N, Karathanos V and Modinos A 1992 *J. Phys.: Condens. Matter* **4** 7389
- [16] Russell P St J, Birks T A and Lloyd-Lucas F D 1995 *Confined Electrons and Photons* ed E Burstein and C Weisbuch (New York: Plenum)
- [17] Robertson W M, Arjavalingham G, Meade R D, Brommer K D, Rappe A M and Joannopoulos J D 1992 *Phys. Rev. Lett.* **68** 2023
- [18] Bell P M, Pendry J B, Martín-Moreno L and Ward A J 1995 *Comput. Phys. Commun.* **85** 306
- [19] Chan C T, Yu Q L and Ho K M 1996 *Phys. Rev. Lett.* at press
- [20] Ward A J and Pendry J B 1995 *J. Mod. Opt.* submitted
- [21] Ritchie R H 1957 *Phys. Rev.* **106** 874
- [22] Lamb W, Wood D M and Ashcroft N 1980 *Phys. Rev. B* **21** 2248
- [23] Fleischmann, M, Hendra P J and Mcquillan A J 1974 *Chem. Phys. Lett.* **26** 163
- [24] Garcia-Vidal F J and Pendry J B 1996 *Prog. Surf. Sci.* at press
- [25] Howie A and Walsh C A 1991 *Microsc. Microanal. Microstruct.* **2** 171
- [26] Ritchie R H and Howie A 1988 *Phil. Mag. A* **58** 753
- [27] Schmeits M and Danby L 1991 *Phys. Rev. B* **44** 12 706
- [28] Pendry J B and Martín-Moreno L 1994 *Phys. Rev. B* **50** 5062
- [29] Martín-Moreno L and Pendry J B 1995 *Nucl. Instrum. Methods Phys. Res. B* **96** 565