CONTROLLED SCATTERING OF LIGHT WAVES: OPTIMAL DESIGN OF DIFFRACTIVE OPTICS

DAVID C. DOBSON*

1. Introduction. Diffractive optics is a vigorously growing technology in which optical components are micromachined with tiny features. Exploiting diffraction effects, devices can be fabricated which perform functions unattainable with conventional optics. These devices have great advantages in terms of size and weight, and can be mass produced at low cost [14]. The current applications are far-reaching, including head-mounted displays, projection displays, flat panel displays, sensors, image acquisition systems, anti-glare surfaces, and optical communications components.

In this article, we summarize some recent progress in the optimal design of diffractive structures. The optimal design of diffractive optics is generally more difficult than the design of conventional (refractive or reflective) optics. The added difficulty is primarily due to the more subtle wave behavior—accurately predicting wave propagation through a diffractive device often requires the numerical solution of Maxwell's equations or a related PDE model, while conventional devices are usually well-modeled by ray-tracing. Optimal design techniques for conventional optics are thus inadequate for diffractive optics.

One of the most common geometrical configurations for diffractive optical structures is a periodic pattern etched into the surface of an optical substrate, such as the diffraction grating shown in Figure 1. The pattern is often created with a sequential photolithographic mask-etch process, sometimes combined with the deposition of additional material layers. The techniques are similar to fabrication processes used in the semiconductor industry.

Most applications of diffractive optics involve narrow-band radiation, so models are generally based on time-harmonic waves. Scattering of time-harmonic waves from infinite periodic structures is a classical problem, dating back to Rayleigh, Floquet and Bloch. A fundamental feature of the problem is that in homogeneous regions, the scattered field can be

* Department of Mathematics and Institute for Scientific Computation, Texas A&M University, College Station, TX 77843-3368. dobson@math.tamu.edu

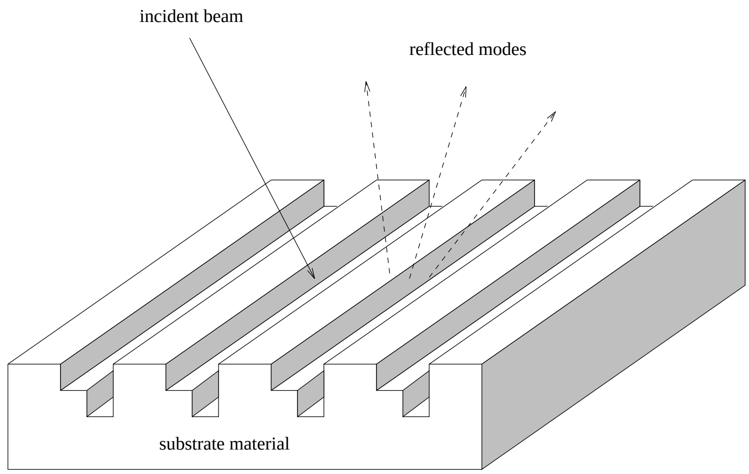


FIG. 1. *Diffraction grating. The structure is assumed to be of infinite extent. The period of the grating is generally comparable to the wavelength of the incident beam.*

expanded as an infinite sum of plane waves. In its simplest form, say for a reflected field u of two variables, the expansion can be written

$$u(x_1, x_2) = \sum_{n=-\infty}^{\infty} a_n e^{i(nx_1 + \beta_n x_2)}, \tag{1}$$

where the a_n are unknown coefficients. This is often called the *Rayleigh expansion*. In the case of a medium with real refractive index $k > 0$, the coefficients β_n in the sum (1) are defined by

$$\beta_n = \begin{cases} \sqrt{k^2 - n^2} & \text{if } k \geq |n|, \\ i\sqrt{n^2 - k^2} & \text{if } k < |n|. \end{cases}$$

Since β_n is real for at most a *finite* number of indices n, we see from (1) that only a finite number of plane waves in the sum propagate into the far field, with the remaining *evanescent* modes decaying exponentially as $x_2 \to +\infty$. The number of propagating modes and the direction of propagation for each mode is determined by the frequency of the incident wave, the refractive index of the material, and the period (cell dimension) of the structure. In the optics literature, the ratio of the energy of a given propagating mode to the energy of the incoming wave is called the *efficiency* of the mode. From an engineering point of view, the key feature of any grating is the efficiency of each propagating mode.

A good introduction to the periodic scattering problem can be found for example in [12, 33]. Descriptions of some additional mathematical problems which arise in diffractive optics and industrial applications may be found in the books [22]–[24].

The specific problem we discuss here is that of designing the periodic interface in such a way that the propagating modes have a specified phase/intensity pattern. The foundation of this problem is the analysis of an underlying wave propagation model. In Section 2 we briefly review a variational formulation of that model: the Helmholtz equation in a periodic geometry. In Section 3 the approximation of the variational problem by finite elements and some resulting computational issues are discussed. We then outline formulation of the optimal design problem as a minimization, and describe two fundamentally different cases which arise depending on assumptions about the admissible class of designs. In the first case, a large class of profiles is taken as the admissible class, and a "relaxed" formulation results. In the second case, the admissible class is restricted, resulting in a "shape design" problem. In both cases, manufacturing considerations must be taken into account in the problem formulation. In Section 5 we outline a simple minimization scheme for the shape design problem, and present some numerical results in Section 6. We conclude with some possible directions for further research in this area. Most of the material presented here is a summary of previous work. The discussion on preconditioners in Section 3 is new, as is much of the material on the shape design problem.

2. The direct diffraction problem. In this section we outline a simple variational formulation for the direct diffraction problem. More detailed accounts can be found for example in [3, 16]. All of the work described here concerns the infinite periodic geometry. We note that if one solves the infinite periodic problem then periodic structures of finite extent can be analyzed using a field-matching technique due to Kriegsmann [31].

We restrict our attention to the simplest geometrical setting, where the diffractive structure is constant in one direction, as in Figure 1. Denote points in $\mathbb{R}^3$ by $x = (x_1, x_2, x_3)$ and for convenience assume that the direction of the grating "grooves" is $\vec{x}_3$. Then the dielectric coefficient $\epsilon(x)$ is a function of (x_1, x_2) only. The periodicity implies that

$$\epsilon(x_1 + nL, x_2) = \epsilon(x_1, x_2), \quad \text{for all } (x_1, x_2)$$

where L is the period, and n is any integer. By rescaling the problem if necessary, the period L can be taken to be 2π.

The direct diffraction problem is then to solve the time-harmonic (time dependence $e^{-i\omega t}$) Maxwell equations

$$\nabla \times E - i\omega\mu H = 0 , \tag{2}$$
$$\nabla \times H + i\omega\epsilon E = 0 , \tag{3}$$

where E and H are the electric and magnetic field vectors, respectively, when a plane wave is incident on the grating from above. In the present "two-dimensional" geometry, there are three fundamentally different situations which arise, depending upon the direction and polarization of the incident wave:

1. *TE (transverse electric) polarization:* the incidence vector is orthogonal to the x_3-axis and the E field is parallel to x_3,
2. *TM (transverse magnetic) polarization:* the incidence vector is orthogonal to the x_3-axis and the H field is parallel to x_3,
3. *Conical diffraction:* the incidence vector is not orthogonal to x_3.

These problems are essentially listed in order of increasing difficulty. In the first case, Maxwell's equations reduce to a simple scalar Helmholtz equation model for the E field. In the second case, Maxwell's equations also reduce to a simple scalar model (this time for the H field), but the regularity of solutions is reduced. For the third problem, a full vector model generally needs to be retained, although the geometry is still "two-dimensional". All three cases, as well as the case of the full Maxwell equations in three dimensions have been studied extensively in the engineering and mathematical literature; see for example [1, 2, 6, 11, 13, 17, 20, 25, 27, 32] and the references therein. For simplicity, here we only consider the first and simplest case, TE polarization.

Let $E = u\,\vec{x}_3$ where $u = u(x_1, x_2)$ is a scalar function. The Maxwell equations then reduce to the Helmholtz equation:

$$(\triangle + k^2)u = 0 , \quad \text{in } \mathbb{R}^2 , \tag{4}$$

where k is the index of refraction: $k^2 = \omega^2\epsilon\mu$. Let $k_1^2 = \omega^2\epsilon_1\mu$ and $k_2^2 = \omega^2\epsilon_2\mu$, where ω is the frequency, μ is the magnetic permeability (assumed to be constant), and ϵ_1 and ϵ_2 are the constant dielectric coefficients in the transmission medium and substrate material, respectively. The constants k_1, k_2 can be complex in general (nonzero imaginary part corresponds to absorbing material). We assume that $Im\ k_1 = 0$, i.e., the transmission medium is non-absorbing.

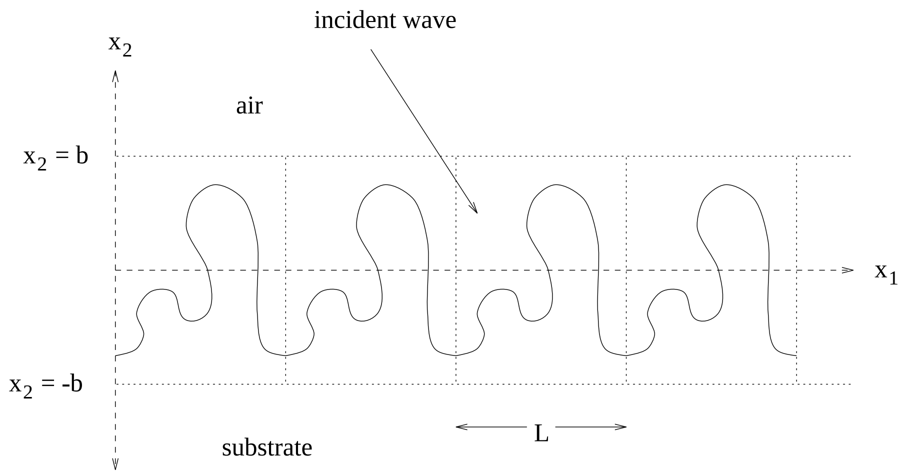


FIG. 2. *Geometrical configuration for TE design. The interface profile is not assumed to be smooth. For convenience, the problem is scaled so that the grating period* $L = 2\pi$.

Let S be a simple curve in $\mathbb{R}^2$, 2π-periodic in x_1 and bounded in the x_2 direction. The curve S represents an interface between the transmission medium and the substrate material. See Figure 2. We define the refractive index $k(x)$ in $\mathbb{R}^2$ by

$$k(x) = \begin{cases} k_1 & \text{if } x \text{ lies above } S, \\ k_2 & \text{if } x \text{ lies below } S \end{cases}$$

We wish to solve the Helmholtz equation (4) when an incoming plane wave

$$u_* = e^{i\alpha x_1 - i\beta_1 x_2}$$

is incident on the interface from above ($x_2 = +\infty$). Here

$$\alpha = k_1 \sin\theta, \quad \beta_1 = k_1 \cos\theta,$$

and $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$ is the angle of incidence. It is difficult to solve (4) directly in the unbounded domain $\mathbb{R}^2$ without additional regularity assumptions on S, which facilitate integral equation formulations. Our approach is to instead reformulate the problem in variational form. To do so requires reducing the original problem over $\mathbb{R}^2$ to an equivalent problem on a bounded domain.

We are interested in "quasiperiodic" solutions, that is, solutions u such that $u_\alpha = ue^{-i\alpha x_1}$ is 2π-periodic. It is easily seen that if u satisfies (4) then u_α satisfies

$$(\triangle_\alpha + k^2)u_\alpha = 0 \quad \text{in } \mathbb{R}^2, \tag{5}$$

where the operator $\triangle_\alpha$ is defined by

$$\triangle_\alpha = \triangle + 2i\alpha\partial_1 - |\alpha|^2.$$

We will henceforth deal only with solutions of equation (5), and so we drop the subscript α. Since both u and k are now 2π-periodic in the x_1 variable, the problem is reduced to solving (5) with periodic boundary conditions in x_1. Equivalently, we consider (5) on the quotient space $Q = \mathbb{R}^2/\{2\pi Z, 0\}$, where $Z = \{0, \pm 1, \pm 2, \ldots\}$. Let $b > \max\{|x_2| : (x_1, x_2) \in S\}$ and define the periodic strip $\Omega = \{x \in Q : -b < x_2 < b\}$, and the two infinite regions $\Omega_1 = \{x \in Q : x_2 > b\}$ and $\Omega_2 = \{x \in Q : x_2 < -b\}$, above and below Ω, respectively. Define the boundaries $,_1 = \partial\Omega_1$, and $,_2 = \partial\Omega_2$.

Our intention is to find u on the domain Ω. This requires appropriate boundary conditions on $,_1$ and $,_2$, which are derived next. We make use of nonlocal "exact" boundary operators. First u is expanded in a Fourier series:

$$u(x_1, x_2) = \sum_{n \in Z} u_n(x_2) e^{inx_1}, \tag{6}$$

where $u_n(x_2) = \frac{1}{2\pi}\int_0^{2\pi} u(x_1, x_2) e^{-inx_1} dx_1$. Define for $j = 1, 2$ the coefficients

$$\beta_j^n(\alpha) = e^{i\gamma_j/2} \left|\omega^2 k_j^2 - (n+\alpha)^2\right|^{1/2}, \quad n \in Z,$$

where

$$\gamma_j = \arg(\omega^2 k_j^2 - (n+\alpha)^2), \quad 0 \le \gamma_j < 2\pi.$$

We henceforth assume that $k_j^2 \neq (n+\alpha)^2$ for all $n \in Z$, $j = 1, 2$. This condition excludes "resonant" cases—where waves can propagate along the x_1-axis—and ensures that a fundamental solution for (5) exists inside Ω_1 and Ω_2. It then follows from knowledge of the fundamental solution [13] that u can be expressed as a sum of plane waves:

$$u|_{\Omega_j} = \sum_{n \in Z} a_j^n e^{\pm i\beta_j^n(\alpha)x_2 + inx_1}, \quad j = 1, 2, \tag{7}$$

where the a_j^n are complex scalars. Imposing a radiation condition on (7) ensures that u is composed of bounded, outgoing plane waves in Ω_1 and Ω_2, plus the incident incoming wave u_* in Ω_1. Matching (6) and (7) then leads to an explicit expression for $u|_{\Omega_j}$, in terms of the Fourier coefficients of the traces $u|_{\Gamma_j}$:

$$u_n(x_2) = \begin{cases} u_n(b)e^{i\beta_1^n(\alpha)(x_2-b)}, & n \neq 0, \quad \text{in } \Omega_1, \\ u_0(b)e^{i\beta_1(x_2-b)} + e^{-i\beta_1 x_2} - e^{i\beta_1(x_2-2b)}, & n = 0, \quad \text{in } \Omega_1, \\ u_n(-b)e^{-i\beta_2^n(\alpha)(x_2+b)}, & \quad\quad\quad\;\; \text{in } \Omega_2. \end{cases}$$

This expression can be differentiated with respect to x_2 and summed to obtain

$$\begin{aligned} \left.\frac{\partial u}{\partial x_2}\right|_{\Gamma_1} &= \sum_{n\in Z} i\beta_1^n(\alpha)u^n(b)e^{inx_1} - 2i\beta_1 e^{-i\beta_1 b}, \\ \left.\frac{\partial u}{\partial x_2}\right|_{\Gamma_2} &= -\sum_{n\in Z} i\beta_2^n(\alpha)u^n(-b)e^{inx_1}. \end{aligned}$$

Define the operator T_j^α by

$$(T_j^\alpha f)(x_1) = (-1)^{j+1}\sum_{n\in Z} i\beta_j^n(\alpha)f_n e^{inx_1}, \quad j = 1, 2, \tag{8}$$

where f_n are the Fourier coefficients of f. Since the coefficients β_j^n grow roughly like $|n|$, it can be easily established that each $T_j^\alpha : H^{\frac{1}{2}}(\,,_j) \to H^{-\frac{1}{2}}(\,,_j)$ is continuous. We have

$$T_j^\alpha(u|_{\Gamma_j}) = \frac{\partial u}{\partial x_2}|_{\Gamma_j}, \quad j = 1, 2,$$

in other words, T_j^α is a Dirichlet–Neumann map. The operators T_j^α define "transparent" boundary conditions on $\partial\Omega$.

The scattering problem can be formulated as follows: find $u \in H^1(\Omega)$ such that

$$(\triangle_\alpha + k^2)u = 0 \text{ in } \Omega, \tag{9}$$

$$(T_1^\alpha - \frac{\partial}{\partial x_2})u = 2i\beta_1 e^{-i\beta_1 b} \text{ on } ,_1, \tag{10}$$

$$(T_2^\alpha - \frac{\partial}{\partial x_2})u = 0 \text{ on } ,_2. \tag{11}$$

Notice that conditions (10), (11) already incorporate an "outgoing wave condition" by the construction of the T_j^α operators. The formulation (9)–(11) admits a variational form: $u \in H^1(\Omega)$ satisfies

$$\begin{aligned} \int_{\Omega_0} \nabla u \cdot \nabla \bar{v} \;&-\; \int_{\Omega_0} (\omega^2 k^2 - \alpha^2)u\bar{v} - 2i\alpha\int_{\Omega_0}(\partial_1 u)\bar{v} \\ &-\; \int_{\Gamma_1}(T_1^\alpha u)\bar{v} - \int_{\Gamma_2}(T_2^\alpha u)\bar{v} = -\int_{\Gamma_1} 2i\beta_1 e^{-i\beta_1 b}\bar{v} \end{aligned} \tag{12}$$

for all $v \in H^1(\Omega)$. Here $\int_{\Gamma_j}$ represents the dual pairing of $H^{-\frac{1}{2}}(,_j)$ with $H^{\frac{1}{2}}(,_j)$. It can be shown that (9)–(11) admits a unique weak solution $u \in H^2(\Omega)$ for all $k \in L^\infty$ sufficiently small, and more generally for all but a discrete set of k_1, k_2.

The TM polarization case can be modeled similarly. Assuming that the fields and geometry are invariant in the x_3 direction, the magnetic field vector H is pointed along the x_3 direction, *i.e.*, $H = u\ \vec{x}_3$, where $u(x_1, x_2)$ is a scalar function. The Maxwell equations (2), (3) reduce to

$$\nabla_\alpha \cdot (\frac{1}{k^2} \nabla_\alpha u) + u = 0, \quad \text{in } \mathbb{R}^2$$

where the operator ∇_α is defined to be $\nabla + i(\alpha, 0)$. Since k is a fixed constant in Ω_j $(j = 1, 2)$, one may derive the boundary conditions for u on $,_j$ using the same recipe as above, which leads to a corresponding weak form. Details may be found in [6].

3. Computational issues.

3.1. Finite element method. As outlined in the previous section, the TE diffraction problem can be posed in variational form: find $u \in H^1(\Omega)$ such that

$$a(u, v) = (f, v), \quad \text{for all } v \in H^1(\Omega), \tag{13}$$

where a is the sesquilinear form defined in (12) and (f, v) is the functional defined by $(f, v) = -2i\beta_1 e^{-i\beta_1 b} \int_{\Gamma_1} \bar{v}$.

In the standard fashion, let $\{S^h : h \in (0, 1]\}$ denote a family of finite dimensional subspaces of H^1, usually taken to be a space of piecewise polynomials which satisfies some approximation assumptions. The parameter h stands for the maximum mesh size after partitioning Ω into simple subdomains. The usual finite element approximation u^h to the solution u_α of (13) would be defined by

$$a(u^h, v^h) = (f, v^h), \quad \text{for all } v^h \in S^h.$$

However, in any real implementation, the nonlocal boundary operators T_j must be approximated. Perhaps the simplest approximation is by truncation, i.e. T_j is replaced by

$$(T_j^N f)(x_1) = \sum_{|n|<N} i\beta_j^n(\alpha) f^{(n)} e^{i\alpha_n x_1}.$$

This leads to a slightly different finite element problem

$$a^N(u_N^h, v^h) = (f, v^h), \quad \text{for all } v^h \in S^h, \tag{14}$$

where a^N is the form arising from the truncated operators. The convergence of the approximations u_N^h is not obvious. However, Bao [5] was able to prove the following.

THEOREM 3.1. *There exist $h_0 \in (0,1]$ and an integer N_0 independent of h_0, such that for $h \in (0, h_0)$ and $N \geq N_0$, the problem (14) attains a unique solution u_N^h. Moreover, the following estimates hold*

$$\begin{aligned} \|u - u_N^h\|_{L^2} &\leq Ch(h + N^{-1/2}), \\ \|u - u_N^h\|_{H^1} &\leq Ch, \end{aligned}$$

where C is independent of h and N.

Thus "optimal" error estimates are obtained provided N is on the order of h^{-2}.

Bao [6] has also proved finite element convergence results for the TM polarization case. The techniques developed in [6] should extend to the analysis of the convergence properties of the finite element method for the case of Maxwell's equations in biperiodic structures. Variational approaches have already been applied to solve 3D biperiodic diffraction problems for Maxwell's equations [1, 17].

3.2. Preconditioning. The finite element discretization from above tends to generate linear systems that are sparse but non positive definite, and non-Hermitian. Standard iterative linear systems solvers often exhibit slow (or no) convergence when confronted with these systems. Still, iterative solvers can be very efficient when combined with an effective preconditioner. Efficiency is extremely important, since the direct problem may need to be solved many times in the design optimization process.

One preconditioner for the Helmholtz equation (9)–(11) can be developed as follows. For simplicity, assume that the incident wave is normal. Simple estimates indicate that the diffracted field is relatively insensitive to *locally averaged* perturbations in the refractive index. Thus a rough approximation to the inverse of the scattering operator could be obtained by averaging the refractive index in one direction and solving the resulting direct problem, i.e., solve

$$(\triangle + \tilde{a})u = f \quad \text{in } \Omega, \tag{15}$$

$$(T_1 - \frac{\partial}{\partial x_2})u \;=\; g_1 \quad \text{on } ,_{1}, \tag{16}$$

$$(T_2 - \frac{\partial}{\partial x_2})u \;=\; g_2 \quad \text{on } ,_{2}, \tag{17}$$

where $\tilde{a}(x_2) = \frac{1}{2\pi}\int_0^{2\pi} k^2(t, x_2)dt$. Since $\tilde{a}$ is constant in the x_1-direction, one can separate variables with the Fourier transform to obtain a sequence of ODEs:

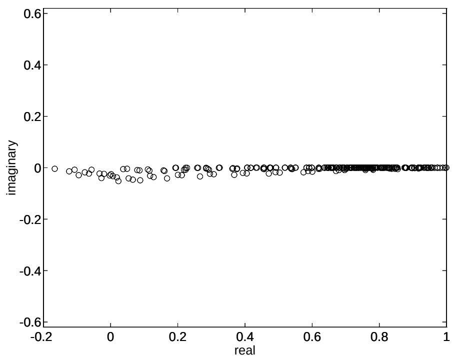
$$\begin{aligned}
(\frac{d}{dx_2} - n^2 + \tilde{a})u_n &= \hat{f}_n \quad \text{in } (-b, b), \\
\frac{du_n}{dx_2}(b) &= i\beta_1^n - \hat{g}_{1n}, \\
\frac{du_n}{dx_2}(-b) &= i\beta_2^n - \hat{g}_{2n},
\end{aligned}$$

$n = 0, \pm 1, \pm 2, \ldots$, where $u_n(x_2)$ are the Fourier components of u in the x_1 variable.

Now suppose that the original scattering problem has been discretized, say with piecewise bilinear finite elements on an $N_1 \times N_2$ grid. Denote the resulting finite element matrix by A. Using the fast Fourier transform in the x_1-variable and a finite-element discretization in x_2 of the system above, one can approximately solve the system (15)–(17) in $\mathcal{O}(N_2 N_1 \log N_1)$ time. Denote this solution operator (the preconditioner) by B.

Plots of the normalized eigenvalues for a prototypical finite element matrix A, and of the preconditioned matrix BA are shown in Figure 3. For this example, the incident wavelength is slightly less than the period of the structure. Notice that A has eigenvalues clustered around the origin, indicating poor conditioning. On the other hand, none of the eigenvalues of BA are near the origin and most of its eigenvalues are closely grouped. Computational tests with the Orthomin iterative solver have indicated that for incident wavelengths on the order of the grating period and substrate refractive indexes which are comparable to the index of the transmission medium, this preconditioner can increase efficiency on the order of 10-fold. We anticipate that similar preconditioners can be developed for the TM polarization case and the full Maxwell equations.

4. Approaches to the optimal design problem. The general form of the design problem we consider is: determine a diffractive structure which generates a specified "output" for a given incident beam, or range



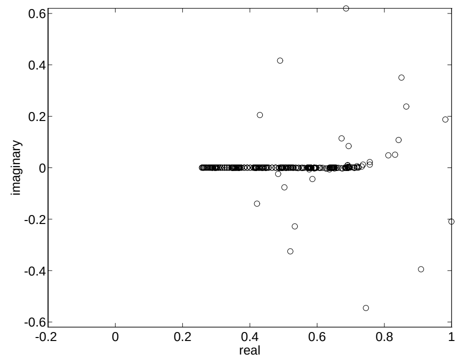


FIG. 3. *(a) Normalized eigenvalues for unpreconditioned matrix A, (b) normalized eigenvalues for preconditioned matrix BA.*

of incident beams. The "output" usually consists of the far-field intensity pattern of the scattered field.

The grating design problem can be approached in various ways. One commonly used approach is based on the *Fraunhofer approximation* [10] for the diffracted field. Under this approximation, optimal design of the diffractive structure can be formulated as a "phase reconstruction" problem. The phase problem is often solved computationally with methods related to the Gerchberg-Saxton algorithm [21, 26], or optimization-based methods [15]. Other design methods based on asymptotic approximations have been proposed [28].

Although advantageous in terms of simplicity and computational efficiency, design methods based on approximations to the underlying PDE model can yield sub-optimal results. A more accurate approach is to retain the full PDE. In [16] and [18], the variational formulation was employed to model the direct problem, and the design problem was solved by a relaxation approach. This approach is summarized in Section 4.1. The variational approach was introduced in [3] for solving optimal photocell design problems; while the relaxation optimization method for general design problems was first proposed in [30].

An alternative to the relaxed design approach is to view the problem as a shape optimization. The formulation of such a problem is given in Section 4.2. A minimization method for solving the shape optimization is outlined in Section 5, and some numerical results are given in Section 6. Details of this work will appear elsewhere [19].

We note that several articles carrying out analytical work for determining periodic interfaces in the context of inverse problems have recently appeared. See [4, 9, 29] for uniqueness and stability results.

4.1. Relaxed optimal design. Recall from Section 2 the geometrical configuration: we are given a periodic curve $S \subset \Omega$ which defines the grating profile. The material above S has refractive index k_1 and the material below S has index k_2. To make explicit the dependence of the refractive index on S, define

$$a_S(x) = \begin{cases} k_1^2 & \text{if } x \text{ is above } S, \\ k_2^2 & \text{if } x \text{ is below } S. \end{cases}$$

We then consider the problem

$$(\triangle_\alpha + a_S)u = 0 \quad \text{in } \Omega, \tag{18}$$

$$(T_1^\alpha - \frac{\partial}{\partial x_2})u = 2i\beta_1 e^{-i\beta_1 b} \quad \text{on } ,_1, \tag{19}$$

$$(T_2^\alpha - \frac{\partial}{\partial x_2})u = 0 \quad \text{on } ,_2. \tag{20}$$

Suppose that the materials, the period of the structure, and the frequency of the incoming waves are fixed. There are then a fixed number of propagating modes, each of which corresponds to an index n for which the propagation constant β_j^n is real-valued. Let us define two sets of indices of propagating modes

$$\Lambda_j = \{n \text{ integer} : \beta_j^n \text{ is real}\}, \quad j = 1, 2.$$

The set Λ_1 contains the indices of the reflected propagating modes; Λ_2 corresponds to the transmitted modes. The coefficients of each propagating reflected mode are determined by the Fourier components of the trace of u on the boundary $,_1$:

$$\begin{aligned} r_n &= u_n(b)e^{-i\beta_1 b} && \text{for } n \neq 0, \quad n \in \Lambda_1, \\ r_0 &= u_0(b)e^{-i\beta_1 b} - const. && \text{for } n = 0. \end{aligned} \tag{21}$$

Similarly, the coefficients of the propagating transmitted modes are

$$t_m = u_m(-b)e^{-i\beta_2 b} \quad \text{for } m \in \Lambda_2. \tag{22}$$

Writing the reflection and transmission coefficients as vectors

$$r = (r_n)_{n\in\Lambda_1}, \quad t = (t_m)_{m\in\Lambda_2},$$

denote the pair $(r,t)=F$. The coefficients r_n and t_m, and hence F, are functions of the interface profile S. Denote this dependence by $F(a_S)$. A general optimal design problem is to find a profile S such that $F(a_S)$ is as close as possible to some specified diffraction pattern q. Asking that $F(a_S)$ is close to q in a least-square sense, one obtains the problem

$$\min_{a_S\in\mathcal{A}} J(a_S)=\frac{1}{2}\|F(a_S)-q\|_2^2.$$

The choice of the admissible set of coefficients $\mathcal{A}$ is important. To achieve a well-posed optimization problem, there are two general routes. The first is to choose a relatively small admissible set, compact with respect to the topology induced by the map $J(a_S)$ and thus ensuring that the problem has a solution. This has the possible side-effect of introducing "artificial constraints", which could result in sub-optimal designs.

The other route is to begin with a large class of admissible curves, and "relax" the problem, enlarging the admissible set to include appropriate "mixtures" of materials. This can be accomplished as follows. Take the set of all continuous simple curves $S\subset\Omega$ as admissible. Denote this set of profiles by $\mathcal{S}$. Define the set of admissible coefficients:

$$\tilde{\mathcal{A}}=\{a_S : S\in\mathcal{S}\}.$$

We want to find the closure of $\tilde{\mathcal{A}}$ with respect to the functional J. Consider

$$\mathcal{A}=\{a=k_2^2\gamma+k_1^2(1-\gamma):\gamma\in L^\infty(\Omega),\ \ 0\le\gamma\le 1,\},$$

which could be described as the set of all mixtures of the two materials.

Under the assumption of low-frequency waves, it can be shown that problem (18)–(20) has weak solutions for any refractive index $a\in\mathcal{A}$. Furthermore, one can bound $\|u\|_{H^1(\Omega)}$ independent of the particular mixture $a\in\mathcal{A}$ [16]. We can then define for each mixture $a\in\mathcal{A}$ corresponding reflection and transmission vectors $r(a)$, $t(a)$. Using weak convergence arguments, it can be shown that for each $a\in\mathcal{A}$, there exists a sequence $a_n\in\tilde{\mathcal{A}}$ such that $r(a_n)\to r(a)$, and similarly for $t(a)$. In this sense, $\mathcal{A}$ is the closure of $\tilde{\mathcal{A}}$ with respect to $F(a)$.

We can then formulate the "relaxed" minimization problem

$$\min_{a\in\mathcal{A}} J(a)=\frac{1}{2}\|F(a)-q\|_2^2. \tag{23}$$

Such problems are studied in [18], where numerical results are also presented.

One could of course generalize and specify a range of incidence angles, or a range of frequencies (or both). One such problem involving the design of antireflective structures was studied in [16], where existence of solutions to the minimization analogous to (23) is proved, again under the assumption of low-frequency waves. Several examples of optimal antireflective structures are also illustrated.

4.2. Shape design. A practical difficulty with the relaxed optimal design approach is that it can generate structures which are prohibitively expensive or difficult to fabricate. One remedy is to use total variation constraints in the minimization in an effort to generate "simple" designs [18]. This approach can generate designs composed of homogeneous "blocks" of material. Still, this is not the most satisfactory approach in all cases.

Recall that the relaxed formulation was obtained by taking a relatively large class $\mathcal{S}$ of admissible interface profiles. Another natural route is to instead restrict the admissible interfaces S to those given by the graph of a bounded function $s(x_1)$. This can be done within the general framework of the variational approach.

This suggests the problem

$$\min J(s) = \frac{1}{2}\|F(s) - q\|_2^2, \quad \text{subject to: } \|s\|_{L^\infty} \leq b' < b \tag{24}$$

where $F(s)$ is the diffraction pattern corresponding to the grating profile S determined by the graph of s. Without further constraints on s, one would not expect that a solution to (24) would exist in general, due to the possibility of oscillatory minimizing sequences. Unfortunately, "smoothness" constraints on s would not be appropriate, since mask-etch type manufacturing processes generally produce profiles with corners. Thus requiring that s is smooth eliminates manufacturable designs from the admissible class.

One convenient quantity which measures oscillations but allows corners is the total variation seminorm

$$TV(s) = \int |s'(x_1)|\, dx_1.$$

Thus we are led to consider the problem

$$\begin{aligned} \min J(s) &= \frac{1}{2}\|F(s) - q\|_2^2, \\ \text{subject to:} \quad & TV(s) \leq M, \\ & \|s\|_{L^\infty} \leq b' < b. \end{aligned} \tag{25}$$

The following result is proved in [19].

THEOREM 4.1. *For sufficiently low-frequency incident waves, the constrained minimization problem (25) admits a solution $s \in BV$.*

5. Minimization by gradient descent. A simple numerical approach to the shape design problem is to discretize the interface profile as the graph of a sum of step functions

$$s(x_1) = \sum_{j=1}^{N} s_j \chi_j \tag{26}$$

where χ_j is the indicator function in the interval $[(j-1)h, jh)$ and h is the cell width in the x_1 direction. Any s in the form (26) with $|s_j| \leq b'$ automatically satisfies $TV(s) \leq 2Nb'$, so one could proceed with the numerical minimization without explicitly enforcing a total variation constraint. Of course, for N large, minimizing sequences could still be "too oscillatory" to be useful in any practical design. In this case $TV(s) \leq M$ should be explicitly enforced.

Consider application of the gradient descent method to find local minima of problem (25). For the sake of simplicity, assume that the total variation constraint is not enforced. The gradient of $J(s)$ can be found using an adjoint-state calculation.

Let us view $J(s)$ as a map over some subset D of $L^2(0, 2\pi)$. Let δs be a small perturbation to a continuous "background" function s, and consider the linearized response $DJ(s)(\delta s)$ of $J(s)$ to δs. Formally, $DJ(s)(\delta s) = Re\{DF(s)(\delta s) \cdot \overline{(F(s) - q)}\}$, where $DF(s)(\delta s)$ denotes the linearization of F. The components of DF are

$$\begin{aligned} Dr_n(s)(\delta s) &= \frac{e^{-i\beta_1 b}}{2\pi} \int_{\Gamma_1} \delta u \, e^{-inx_1}, \\ Dt_m(s)(\delta s) &= \frac{e^{-i\beta_2 b}}{2\pi} \int_{\Gamma_2} \delta u \, e^{-imx_1}, \end{aligned}$$

where δu solves the linearized problem

$$\begin{aligned} (\triangle_\alpha + a_s)\delta u &= -\delta s \, (k_2^2 - k_1^2)\mu_s u \quad \text{in } \Omega, \\ (T_j - \frac{\partial}{\partial \nu})\delta u &= 0 \qquad \text{on } ,_j, \quad j = 1, 2. \end{aligned}$$

Here μ_s is the measure defined by

$$\int_\Omega f \mu_s = \int_0^{2\pi} f(x_1, s(x_1)) \, dx_1.$$

for functions f on Ω. Viewing the domain of DF as $L^2(0, 2\pi)$, the adjoint $DF^*(s)(\cdot)$ is defined by

$$DF(s)(\delta s) \cdot \bar{\delta q} = \int_0^{2\pi} \delta s \cdot \overline{DF^*(s)(\delta q)}\, dx_1,$$

for $\delta q = (\psi, \phi)$ with $\psi = (\psi_n)_{n\in\Lambda_1}$ and $\phi = (\phi_m)_{m\in\Lambda_2}$. Let $w \in H^1(\Omega)$ solve

$$(\triangle_\alpha + a_s)w = 0 \quad \text{in } \Omega, \tag{27}$$

$$(T_1^* - \frac{\partial}{\partial\nu})w = -\frac{e^{i\beta_1 b}}{2\pi} \sum_{n\in\Lambda_1} \psi_n e^{inx_1} \quad \text{on } \Gamma_1, \tag{28}$$

$$(T_2^* - \frac{\partial}{\partial\nu})w = -\frac{e^{i\beta_2 b}}{2\pi} \sum_{m\in\Lambda_2} \phi_m e^{imx_1} \quad \text{on } \Gamma_2, \tag{29}$$

where $T_j^* f = -\sum i\bar{\beta}_j^n f_n e^{inx}$. Notice that this adjoint problem for w represents waves propagating *into* Ω. With an integration by parts calculation, one finds that

$$DF(s)(\delta s) \cdot \overline{\delta q} = (k_2^2 - k_1^2) \int_0^{2\pi} \delta s(x_1)\ (\overline{w}u)(x_1, s(x_1))\ dx_1$$

We then make the identification $\overline{DF^*(s)(\delta q)}(x_1) = (k_2^2 - k_1^2)\overline{w}u(x_1, s(x_1))$, and the gradient of $J(s)$ is given by $G(s) = Re\ \{DF^*(s)(F(s) - q)\}$, or

$$G(s)(x_1) = Re\{(k_2^2 - k_1^2)\overline{w}u\}(x_1, s(x_1)),$$

where w solves (27)–(29) with $(\psi, \phi) = F(s) - q$. Since both w and u are in H^2 and hence have continuous representatives, $G(s)(x_1)$ is well-defined pointwise.

Unfortunately a gradient descent step defined by $s_1 = s_0 - tG(s_0)$, $t > 0$ does not necessarily lie in the computational domain Ω. Thus we define a projection operator P by

$$(Pf)(x_1) = \begin{cases} \min\{f(x_1), b'\} & \text{if } f(x_1) \geq 0, \\ \max\{f(x_1), -b'\} & \text{if } f(x_1) < 0, \end{cases}$$

where $b' < b$.

Straightforward gradient descent would then proceed as follows:

1. Choose an initial guess s_0.
2. For $j = 0, \ldots$convergence, set $s_{j+1} = P(s_j - t_j G(s_j))$ for a suitably chosen step parameter t_j.

Computationally, this algorithm is slow but generally effective. Techniques to improve the efficiency of the basic algorithm have been developed for the relaxed design problem [18]. The general idea is to take advantage of the underlying PDE model by viewing it as a constraint and performing inexact solves following infeasible point techniques from constrained optimization. The same ideas can be applied to the shape optimization problem. Further details, as well as a description of the incorporation of the TV constraint into the algorithm, and modifications to allow for intensity-only diffraction patterns can be found in [19].

6. Numerical experiments. In this section we describe some numerical results for the shape design problem. Some earlier results were presented in [8].

Consider first the design of an "ideal array generator", a diffractive structure which splits a single normally incident plane wave into several equal intensity transmitted modes with 100% efficiency. Such a devices would be useful, for example, in optical communications interconnections. We choose $k_2/k_1 = 1.5$ (approximately a glass/air interface) and take the incoming wavelength such that the grating supports seven transmitted orders. We specify that all transmitted orders have equal energy and all reflected orders have zero energy. Using a flat profile as an initial guess, and *no total variation penalty*, the shape optimization method produced the profile shown in Figure 4. The structure is essentially 100% efficient (up to discretization and roundoff error). This profile is simple enough to be approximated well by a stepped grating using the "mask-etch" fabrication process. A comparison of relaxed and profile designs for (five-order) array generators can be found in [8]. The relaxed designs have the disadvantage that they specify regions with intermediate-index materials, seriously complicating the fabrication process.

It is interesting that some problems converge well to reasonable profiles without explicit regularization, as with the problem above, while others appear to have nonconvergent minimizing sequences. As an illustration, consider the problem of maximizing the energy in the $+1$ transmitted order, given a normally incident plane wave on a grating which supports nine transmitted orders. Such structures are useful for example in waveguide couplers. The material parameters are again taken such that $k_2/k_1 = 1.5$. Figure 5 shows a design for such a structure obtained from a phase reconstruction method. This design is very intuitive and it directs 70% of the incident energy into the $+1$ transmitted order. Using the pro-

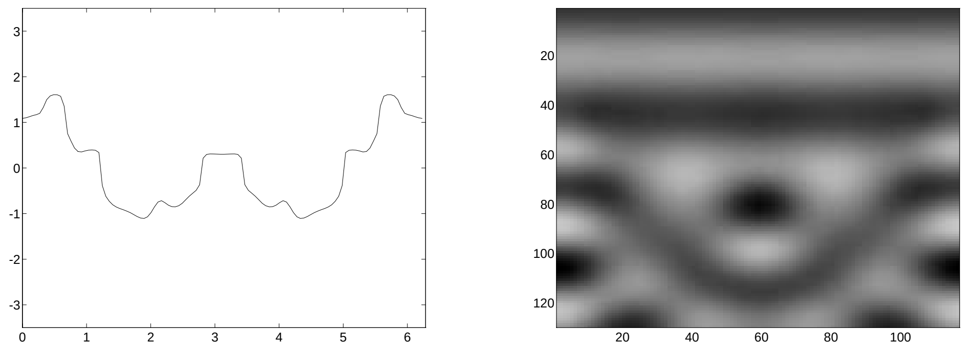


FIG. 4. *Surface profile for seven-order beam splitter and real part of diffracted field.*

file in Figure 5 as an initial guess, after approximately 200 steps of the minimization algorithm, the profile in Figure 6 was obtained. This profile directs approximately 99.9% of the incident energy into the +1 transmitted order. One can clearly see the nearly undisturbed plane wavefront exiting the bottom of the structure at an angle. However, the interface is very complicated and undesirable from the point of view of fabrication. An intuitive explanation of the complicated shape of the profile is that the design is seeking a relaxed "mixture" by oscillating rapidly: the underlying problem may have a minimizing sequence which does not converge.

The oscillatory behavior can be corrected, possibly at the expense of some efficiency, by applying the total variation constraint. Figure 7 shows a profile obtained solving the same problem with the constraint $TV(s) \leq 12.9$. The resulting diffraction efficiency is 95%. This profile is sufficiently "nice" to allow approximation by a "stepped" profile which can be directly fabricated using a mask-etch process. Such an approximation is shown in Figure 8; note that the efficiency decreases to 87%, but that the structure is still much more efficient than the ramp profile of Figure 5.

7. Directions for further research. The work described here was limited to the two-dimensional model in TE polarization mode. This is restrictive, as real devices often use other polarization modes (as described in Section 2) and other geometries. Extension of the results described here to other models is not immediate, primarily because the behavior of the surface profile-to-electromagnetic field map is fundamentally different (and perhaps more interesting) in these cases. In addition, three-dimensional

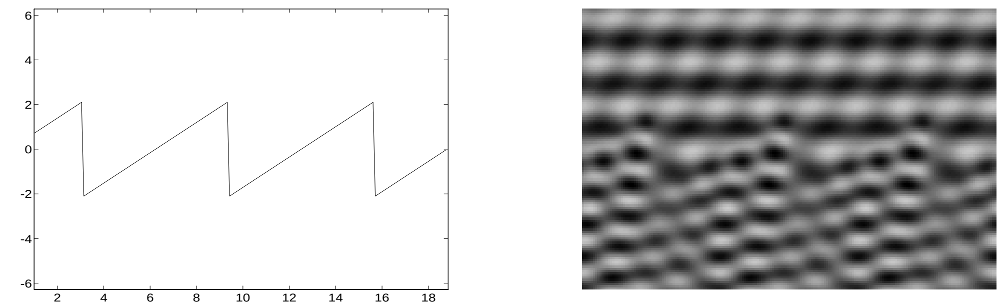


FIG. 5. *Optimal ramp profile. Gray-scale plot shows the real part of the diffracted field. The structure directs 70% of the incoming energy into the* $+1$ *transmitted order.*

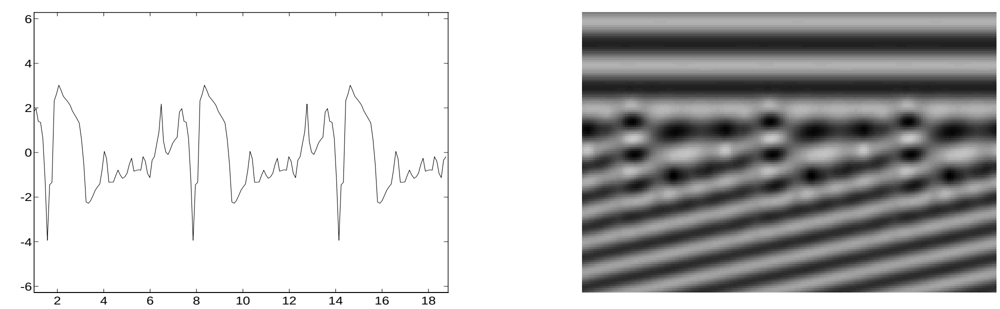


FIG. 6. *A 99.9% efficient near-optimal profile, obtained with unconstrained minimization, and real part of the diffracted field. The total variation of the profile is* 33.9.

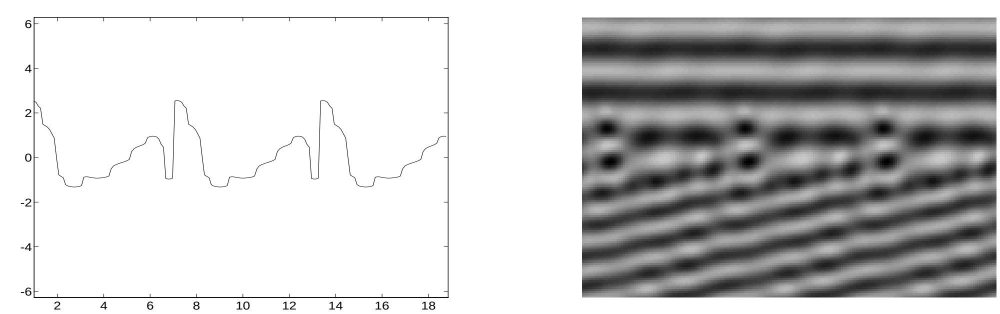


FIG. 7. *A 95.0% efficient approximate profile, obtained by applying a total variation penalty, and real part of the diffracted field. The total variation of the profile is* 12.8.

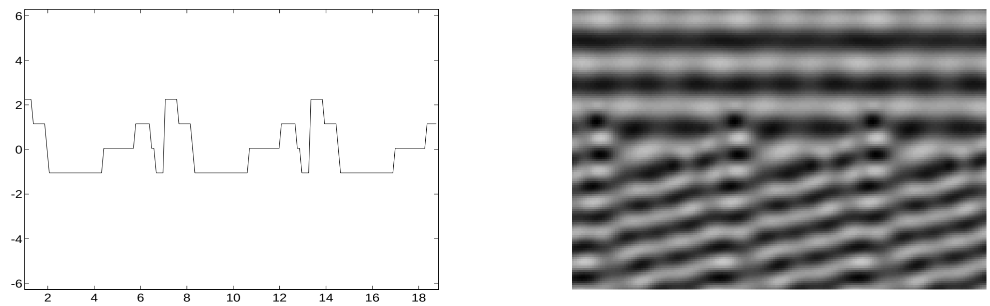


FIG. 8. *An 87.0% efficient "4-level approximation", and real part of the diffracted field.*

geometries present a computational challenge, requiring the development of very efficient solution procedures.

In a slightly different direction, the need exists for optimal design techniques for certain *nonlinear* structures. For example, one particularly interesting structure is based on the phenomenon of "grating enhanced second-harmonic generation". In these devices, a nonlinear optical material with relatively large second-order susceptibility tensor is either applied over a diffraction grating, or a grating structure is etched into the surface of a nonlinear optical crystal. When an intense "pump" beam is applied, a second-harmonic field (at twice the pump frequency) is generated. Thus for example, coherent blue light can be generated from a red pump laser. The nonlinear susceptibility is generally very small, and thus the intensity of the second-harmonic field is weak. It has been found that the grating structure can significantly enhance the second-harmonic conversion efficiency [34]. Presumably the enhancement is due to increased pump field intensity within the nonlinear material near grating resonances. This effect has recently been studied analytically and computationally [7]. The optimal design question arises: which nonlinear grating structures maximize second-harmonic conversion efficiency?

An enormous variety of other structures and devices—both linear and nonlinear—are currently being developed in the optical engineering community, presenting a wealth of important optimal design problems. Understanding the mathematical properties of these problems and developing effective computational techniques to solve them will be an essential, enabling component of future technological development in the area.

Acknowledgements. I wish to thank the organizers of the Symposium, Irena Lasiecka and Blaise Morton, for the opportunity to present this material, and SIAM for its support. Much of the work described here is either due to, or the result of joint work with Gang Bao; his efforts are gratefully acknowledged. Many thanks are due to Dr. J. Allen Cox of the Honeywell Technology Center, for his continuing collaboration on this research. This work was partially supported by an NSF Mathematical Sciences Postdoctoral Fellowship.

REFERENCES

[1] T. Abboud, *Etude mathématique et numérique de quelques problémes de diffraction d'ondes électromagnétiques*, PhD thesis, Ecole Polytechnique, Palaiseau, 1991.

[2] T. Abboud, *Electromagnetic waves in periodic media*, in *Second International Conference on Mathematical and Numerical Aspects of Wave Propagation*, edited by R.Kleinman et al, SIAM, Philadelphia, 1993, 1–9.

[3] Y. Achdou, *Optimization of a photocell*, Opt. Control Appl. Math., 12 (1991), 221–246.

[4] G. Bao, *A uniqueness theorem for an inverse problem in periodic diffractive optics*, Inverse Problems **10** (1994), 335-340.

[5] G. Bao, *Finite elements approximation of time harmonic waves in periodic structures*, SIAM J. on Numerical Analysis, to appear.

[6] G. Bao, *Diffractive optics in periodic structures: the TM polarization*, preprint.

[7] G. Bao and D. Dobson, *Diffractive optics in nonlinear media with periodic structure*, (1993) preprint.

[8] G. Bao, D. Dobson, and J. A. Cox, *Mathematical studies in rigorous grating theory,* J. Opt. Soc. Am. **A**, to appear.

[9] G. Bao and A. Friedman, *Inverse problems for scattering by periodic structures*, preprint.

[10] Born, M. and Wolf, E., *Principles of Optics,* sixth edition, Pergamon Press, Oxford, 1980.

[11] O. Bruno and F. Reitich, *Numerical solution of diffraction problems: a method of variation of boundaries; II. Dielectric gratings, Pade approximants and singularities; III. Doubly-periodic gratings*, J. Opt. Soc. Amer. **A.**, **10** (1993), 1168–1175, 2307–2316, 2551–2562.

[12] M. Cadilhac, *Rigorous vector theories of diffraction gratings*, in Progress in Optics **21**, E. Wolf, ed., North-Holland, New York, 1984.

[13] X. Chen and A. Friedman, *Maxwell's equations in a periodic structure*, Trans. Amer. Math. Soc., **323** (1991), 465-507.

[14] J. A. Cox, *Replicated hybrid optics for display applications*, "1994 SID International Symposium Digest of Technical Papers", Vol. XXV, to appear.

[15] D. C. Dobson, *Phase reconstruction via nonlinear least-squares*, Inverse Problems, 8 (1992), 541–557.

[16] D.C. Dobson, *Optimal design of periodic antireflective structures for the Helmholtz equation*, Euro. J. Appl. Math., **4** (1993), 321-340.

[17] D.C. Dobson, *A variational method for electromagnetic diffraction in biperiodic structures*, RAIRO Modél. Math. Anal. Numér., **28** (1994), 419–439.

[18] D.C. Dobson, *Exploiting ill-posedness in the design of diffractive optical structures,* in "Mathematics in Smart Structures", edited by H. T. Banks, SPIE Proc. vol. 1919 (1993), pp. 248–257.

[19] D.C. Dobson, *Optimal shape design of blazed diffraction gratings: the TE case,* in preparation.

[20] D.C. Dobson and A. Friedman, *The time-harmonic Maxwell equations in a doubly periodic structure,* J. Math. Anal. Appl., **165** (1992), pp. 507-528.

[21] M.W. Farn, *New iterative algorithm for the design of phase-only gratings*, Computer and Optically Generated Holographic Optics, I.N. Cindrich and S. Lee, eds., Proc. SPIE 1555 (1991), pp. 34-42.

[22] A. Friedman, *Mathematics in Industrial Problems*, IMA Volume 16, Springer–Verlag, New York (1988).

[23] A. Friedman, *Mathematics in Industrial Problems, Part 3*, IMA Volume 38, Springer–Verlag, New York (1991).

[24] A. Friedman, *Mathematics in Industrial Problems, Part 7*, Springer–Verlag, New York (1994).

[25] T. K. Gaylord and M. G. Moharam, *Rigorous coupled-wave analysis and applications of grating diffraction*, in "Diffractive and Miniaturized Optics (Critical Reviews)" Sing Lee, ed., SPIE vol. CR-49 (1994), pp. 3–31.

[26] R.W. Gerchberg and W.O. Saxton, *A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures*, Optik (1972), pp. 237-46.

[27] C. Goldstein, *A finite element method for solving Helmholtz type equations in waveguides and other unbounded domains*, Math. Comp., **39** (1982), 309-324.

[28] Y. Han and C.A. Delisle, *Exact surface relief profile of kinoform lenses from a given phase function,* Proc. SPIE **2152** (1994), to appear.

[29] A. Kirsch, *Uniqueness theorems in inverse scattering theory for periodic structures*, Inverse Problems **10** (1994), 145-152.

[30] R. Kohn and G. Strang, *Optimal design and relaxation of variational problems I, II, III*, Comm. Pure Appl. Math. **39**(1986), 113-137, 139-182, 353-377.

[31] G.A. Kriegsmann and C.L. Scandrett, *Large membrane array scattering*, J. Acoust. Soc. Am., to appear.

[32] J.C. Nédélec and F. Starling, *Integral equation methods in a quasi-periodic diffraction problem for the time-harmonic Maxwell's equations,* SIAM J. Math. Anal., **22** (1991), 1679-1701.

[33] *Electromagnetic Theory of Gratings,* Topics in Current Physics, Vol. 22, edited by R. Petit, Springer-Verlag, Heidelberg, 1980.

[34] R. Reinisch, M. Nevière, H. Akhouayri, J. Coutaz, D. Maystre, and E. Pic, *Grating enhanced second harmonic generation through electromagnetic resonances*, Opt. Eng., Vol. 27, No. 11 (1988), 961-971.