Shape optimization for the strong routing of light in periodic diffraction gratings

Juan C. Araújo C. [a,b,*], Christian Engström [c], Eddie Wadbro [b,d]

[a] *Department of Science and Mathematics Education, Umeå University, Sweden*
[b] *Department of Computing Science, Umeå University, Sweden*
[c] *Department of Mathematics, Linnaeus University, Sweden*
[d] *Department of Mathematics and Computer Science, Karlstad University, Sweden*

ARTICLE INFO

ABSTRACT

In the quest for the development of faster and more reliable technologies, the ability to control the propagation, confinement, and emission of light has become crucial. The design of guide mode resonators and perfect absorbers has proven to be of fundamental importance. In this project, we consider the shape optimization of a periodic dielectric slab aiming at efficient directional routing of light to reproduce similar features of a guide mode resonator. For this, the design objective is to maximize the routing efficiency of an incoming wave. That is, the goal is to promote wave propagation along the periodic slab. A Helmholtz problem with a piecewise constant and periodic refractive index medium models the wave propagation, and an accurate Robin-to-Robin map models an exterior domain. We propose an optimal design strategy that consists of representing the dielectric interface by a finite Fourier formula and using its coefficients as the design variables. Moreover, we use a high order finite element (FE) discretization combined with a bilinear Transfinite Interpolation formula. This setting admits explicit differentiation with respect to the design variables, from where an exact discrete adjoint method computes the sensitivities. We show in detail how the sensitivities are obtained in the quasi-periodic discrete setting. The design strategy employs gradient-based numerical optimization, which consists of a BFGS quasi-Newton method with backtracking line search. As a test case example, we present results for the optimization of a so-called single port perfect absorber. We test our strategy for a variety of incoming wave angles and different polarizations. In all cases, we efficiently reach designs featuring high routing efficiencies that satisfy the required criteria.

1. Introduction

The scattering of waves by periodic layered linear media [1,2] in two and three dimensions arises in many engineering and physics applications. Examples of the so-called *diffraction gratings* [1] exist in material's science [3], geophysics [4,5], imaging [6], and acoustics [7], to name a few. Additionally, important examples are found in electromagnetic and photonic applications [8–11], as well as in the field of plasmonics [12,13].

* Corresponding author at: Department of Science and Mathematics Education, Umeå University, Sweden.
  E-mail addresses: juan.araujo@umu.se (J.C. Araújo C.), christian.engstrom@lnu.se (C. Engström), eddiew@cs.umu.se (E. Wadbro).

The guided resonances of layered electromagnetic systems have been extensively researched aiming at the design of efficient devices [14–17]. Moreover, the so-called Guided-Mode Resonators (GMR) and Perfect Absorbers are of particular interest for their high degree of optical tunability (wavelength, polarization, and intensity) [18], which is particularly attractive for applications requiring waveguide coupling, electromagnetic field enhancement, and the filtering and focusing of light. Examples are found in bio-sensing [19] and photovoltaics [8,20]. Furthermore, GMRs have application in the design of efficient polarizers, spectrometers, optical switches, modulators, light-electricity transducers [21–23], and light-trapping layers for solar cells [8,24]. For an introduction to the concepts of GMR, the reader is referred to Quaranta et al. [18] and Guise et al. [25].

Diffraction gratings consisting of a single periodic layer (slab) with 1D periodicity can be fine-tuned [26] to exhibit remarkable resonant absorption (high quality factor). The so-called Yablonovitch limit [27,28] gives an estimate that characterizes the absorption efficiency in a diffraction grating with a single layer in terms of its refractive index and thickness. The estimate is derived from ray optics, where the thickness of the layer is assumed greater than the wavelength of the incoming wave. Design techniques for the assemblies of optical devices with efficiencies beyond the Yablonovitch limit have been proposed by employing thin coatings [8,29], photonic crystal mirrors [30,31], or plasmonic gratings [32].

Alternatively, it has been reported that slabs whose thickness is smaller than the wavelength of the incoming wave have the potential to be more efficient than what is predicted by the Yablonovitch limit [24]. For this reason, the design of thin slabs is of particular interest. In such a case, the assumptions from the ray-optics approximation are no longer applicable, and the Maxwell equations are necessary to accurately describe the physical phenomena. The diffraction grating's efficiency can then be assessed by looking at the guided resonances of the open system [18]. The groove depth of a periodic slab with a modulated surface is the maximum deviation from the surface reference point. Gratings with shallower groove depth typically exhibit higher quality factors, or larger and sharper resonance peaks [18,33]. In general, an increase in the groove depth causes the imaginary part of the guided resonance to increase, thus decreasing the resonance peak and lowering the quality factor of the device.

Diffraction gratings can be modeled by using integral equation based formulations [34]; namely, by using volume [10] or boundary integral formulations [7,9,11,35–37]. In these formulations, quasi-periodicity and outgoing conditions are imposed naturally by the use of the outgoing quasi-periodic Green's function. Alternatively to integral equations, it is common to use a partial differential equation (PDE) subject to quasi-periodic and outgoing conditions [2,38,39]. In the PDE based formulation, quasi-periodicity is imposed via wall matching conditions, and outgoing conditions are set up by using Rayleigh–Bloch expansions. Two important situations that require special attention arise in the analysis of periodic diffraction gratings. The first case refers to the guided resonances [18] of the system. The second case pertains to the so-called Rayleigh frequencies arising from the Rayleigh representation of outgoing waves. The Rayleigh frequencies depend on the angle of incidence of the incoming plane wave. For historical reasons, both situations are discussed in the literature as Wood's anomalies [33].

At Rayleigh frequencies, the quasi-periodic Green function is not defined [11,40]. However, several techniques can be employed to overcome this difficulty [11]. For example, by tailoring special Green-functions that are well-defined at Rayleigh frequencies. In discretizations based on boundary integral equations, it is possible to avoid the non-existence of solutions by using an explicit method for imposing periodicity [41]. Additionally, Gillman et al. [35] proposed a strategy that uses locally periodic and non-periodic Green's functions.

In the PDE-based formulation, it is common to use Perfectly Matched Layers (PML) [38,42] and Dirichlet-to-Neumann maps (DtN) to set up outgoing conditions. Discretizations applied to quasi-periodic diffraction gratings employing a PML suffer from solutions having low accuracy near Rayleigh frequencies. A possible remedy to this problem is described by Zhou et al. [42]. Alternatively to the PML method, DtN formulations have been used for solving diffraction grating problems. The DtN is set up by employing the Rayleigh representation of outgoing waves, and hence presents non-uniqueness of solutions at Rayleigh frequencies [43]. A possible fix is to use so-called Robin-to-Robin (RtR) transparent boundary conditions, which exhibit better stability properties compared to DtN boundary conditions [44].

The RtR formulation allows us to solve the direct problem also at Rayleigh frequencies and characterizes the set where the formulation delivers a non-unique solution.

1.1. Overview of optimization techniques for diffraction gratings

In this work, we are particularly interested in designing efficient absorbers: Devices that when illuminated with an incident plane wave delivers most of its energy into the grating. Typical optimization techniques for diffraction gratings with GMR in mind are based on the theoretical framework of Moharam et al. [45]. The latter consists of building the scattering matrix for an array of straight slabs using the Maxwell equations and coupled-wave analysis [46,47] on simple geometries. In the scattering matrix formalism, from a given geometry and given incoming waves, we assemble the corresponding scattering matrix of the system. Perfect absorption, or absence of scattering, occurs at the complex guided resonances of the system. For a given dielectric structure, scattering near perfect absorption requires the supply of the right incident waveform.

Alternatively, an optimization problem can be set up for finding the closest shape of a slab that resembles the properties of a perfect absorber for a given frequency and incoming wave. Common optimization schemes applied to diffraction gratings are set to find few geometrical parameters for the slab. A first order parameterization (straight interfaces) along with piecewise permittivity constants forms the basis of the most commonly used optimization strategy for diffraction gratings.
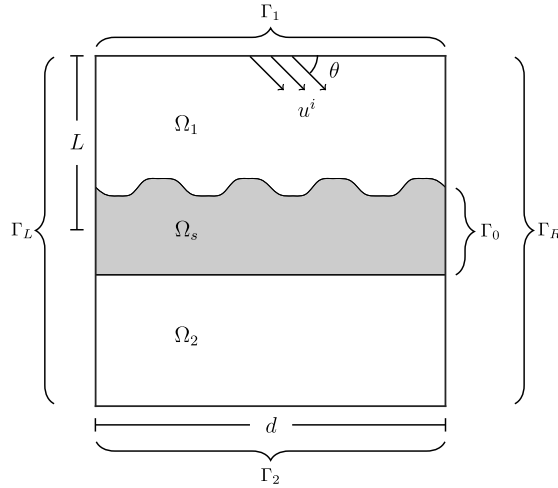
Fig. 1. Geometry of the diffraction grating problem.

Examples of the latter can be found in studies using the so-called particle swarm optimization [46–48], or the work by Giese et al. [25].

Alternatively, a Nelder–Mead simplex optimization scheme has been also used for similar problems [49]. More recently the so-called optimized figure of merit has been presented [50], for searching for optimal groove-depths and grating periods for diffraction grating problems. An important limitation of these approaches is the low dimensionality of the design space.

Instead of the optimization methods mentioned above, we discretize the PDE of the problem and use a discrete adjoint method for computing the sensitivities. In this work, we present a novel way of optimizing designs for the perfect absorption of incoming planar waves into a quasi-periodic slab. The problem falls under the class of PDE-constrained shape optimization problems, where the sensitivity of the shape of the metasurface is driven by the scattered wave, which is constrained to satisfy Helmholtz equation subject to quasi-periodic constraints and appropriate outgoing wave conditions. The proposed method is flexible and allows for a simple parameterization of the surface. The PDE is discretized by a high order polynomial version of the Finite Element Method, and the outgoing wave conditions are modeled by Robin-to-Robin maps (RtR) based on the Rayleigh expansion of the quasi-periodic solution. We use a fixed number of elements, which has the advantage of avoiding difficulties related to re-meshing. The objective function is evaluated using the numerical solution to the PDE, and the sensitivities of the design variables are obtained by using a discrete adjoint method. The numerical optimization consists of the BFGS method with backtracking line search.

A similar approach from Aylwin et al. [36] has recently appeared for the computation and optimization of periodic diffraction gratings consisting of metallic surfaces by using boundary integral formulation.

2. Optimization problem and parameterization

We consider a single layer GMR device consisting of a dielectric slab with a parametrically shaped upper interface and a flat base. The slab's thickness is in the subwavelength regime. Let the domains Ω_s, Ω_1, and Ω_2 be the regions depicted in Fig. 1. We define the computational domain $\Omega := \Omega_1 \cup \Omega_s \cup \Omega_2$ as the rectangle of length d and height $2L$. The top and bottom boundaries of Ω are denoted Γ_1 and Γ_2, respectively. Similarly, Γ_L and Γ_R denote the left and right boundaries of Ω. In the optimization strategy, we also make use of the right boundary of Ω_s, which is denoted Γ_0.

The objective of the optimization is to find the shape of the slab that best promotes the routing of light through a dielectric slab. We consider slabs with a modulated upper surface of small groove depths (imposed by the box constraints), in order to reduce losses by scattering. We search optimal designs with low frequency modulation of the shape of the GMR. For this, we impose additional constraints.

2.1. Governing equations

In the following, we consider non-magnetic, locally homogeneous and isotropic materials, and assume the time harmonic ansatz $e^{-i\omega t}$, with $\omega \in \mathbb{R}$ a given frequency. From the Maxwell equations, we obtain decoupled wave equations [51,52] for the electric field **E**

$$\nabla \times \nabla \times \mathbf{E} - \left(\frac{\omega}{c}\right)^2 \epsilon \, \mathbf{E} = 0, \tag{1}$$

and for the magnetic field

$$\nabla \times \left(\frac{1}{\epsilon}\nabla \times \mathbf{H}\right) - \left(\frac{\omega}{c}\right)^2 \mathbf{H} = 0, \tag{2}$$

where c is the speed of light in vacuum, and the relative permittivity ϵ accounts for the description of dielectrics. In this work ϵ is real, positive, and piecewise constant.

Dielectric interfaces with infinite length along the z-axis are represented by using $\epsilon = \epsilon(x, y)$, which is independent of z. This particular symmetry suggests the following explicit separation of variables

$$\mathbf{E} := \tilde{\mathbf{E}}(x, y)e^{ik_3 z}, \text{ and } \mathbf{H} := \tilde{\mathbf{H}}(x, y)e^{ik_3 z}. \tag{3}$$

We let $\boldsymbol{x} := (x, y)$ and we assume that the incoming plane wave propagates in one of the following special polarizations:

Transverse Magnetic (TM) polarized waves: In this case the fields have the explicit form $\tilde{\mathbf{E}} := (0, 0, E_3)^\top$ and $\tilde{\mathbf{H}} := (H_1, H_2, 0)^\top$. From equations (1) and (3), we obtain the equivalent problem

$$-\Delta E_3 - \omega^2 \epsilon E_3 = 0, \text{ for } \boldsymbol{x} \in \mathbb{R}^2. \tag{4}$$

Transverse Electric (TE) polarized waves: In this case the fields have the explicit form $\tilde{\mathbf{E}} := (E_1, E_2, 0)^\top$ and $\tilde{\mathbf{H}} := (0, 0, H_3)^\top$. From equations (2) and (3), we obtain the equivalent problem

$$-\nabla \cdot \left(\frac{1}{\epsilon}\nabla H_3\right) - \omega^2 H_3 = 0, \text{ for } \boldsymbol{x} \in \mathbb{R}^2. \tag{5}$$

From the suggested symmetry, the three dimensional vector problems (1) and (2) are reduced to the equivalent two dimensional scalar problems (4) and (5) that are written in compact notation as

$$-\nabla \cdot (\rho \nabla u) - \omega^2 \eta u = 0, \text{ for } \boldsymbol{x} \in \mathbb{R}^2, \tag{6}$$

where $u := E_3$, $\rho := 1$, $\eta = \epsilon$ for the TM-polarization and $u := H_3$, $\rho := 1/\epsilon$, $\eta := 1$ for the TE polarization.

2.2. Description of the scattering problem

The solution u that satisfies Helmholtz equation (6) is denoted the total wave. For a piecewise constant relative permittivity $\epsilon \in L^\infty(\Omega)$, a dielectric slab can be defined as $\Omega_s := \text{supp}(\epsilon - 1)$, which is assumed to be periodic in the x-axis and of thickness which is compact in the y-axis. A sketch of the problem is presented in Fig. 1.

For future reference, we introduce the slab's refractive index as n_s. In the case of illumination by an incoming plane wave with wave vector $\boldsymbol{k} := \omega(\cos\theta, -\sin\theta)^\top$, we define the functions

$$u^i := \begin{cases} e^{i\boldsymbol{k}\cdot\boldsymbol{x}}, & \boldsymbol{x} \in \Omega_1 \\ 0, & \boldsymbol{x} \in \Omega_s \\ 0, & \boldsymbol{x} \in \Omega_2 \end{cases} \text{ and } \epsilon := \begin{cases} 1, & \boldsymbol{x} \in \Omega_1 \\ n_s^2, & \boldsymbol{x} \in \Omega_s \\ 1, & \boldsymbol{x} \in \Omega_2 \end{cases}. \tag{7}$$

The strong formulation for the total wave reads

$$\begin{aligned} -\nabla \cdot (\rho \nabla u) - \omega^2 \eta u &= 0, & \boldsymbol{x} \in \Omega, \\ \nabla(u - u^i) \cdot \mathbf{n}_1 - i\sigma(u - u^i) &= R_1(u - u^i), & \boldsymbol{x} \in \Gamma_1, \\ \nabla u \cdot \mathbf{n}_2 - i\sigma u &= R_2 u, & \boldsymbol{x} \in \Gamma_2, \\ u(\boldsymbol{x} + d\mathbf{e}_1) &= \psi u(\boldsymbol{x}), & \boldsymbol{x} \in \Gamma_L, \\ \nabla u(\boldsymbol{x} + d\mathbf{e}_1) \cdot \mathbf{n}_L &= \psi \nabla u(\boldsymbol{x}) \cdot \mathbf{n}_L, & \boldsymbol{x} \in \Gamma_L, \end{aligned} \tag{8}$$

where $\sigma > 0$, $\mathbf{n}_j$ is the outward unit vector with respect to Γ_j, $\psi := e^{iad}$, $a := \omega\cos\theta$ and, exact outgoing wave conditions are defined by the use of Robin-to-Robin operators R_m on Γ_m, $m = 1, 2$.

In order to introduce periodicity in the model problem and to establish it's effect in the application of outgoing conditions, we make use of the following definitions and properties.

Property: The functions $\{e^{ia_n x}\}$, with $a_n := 2\pi n/d$, $n = 0, 1, 2, 3, \ldots$, form an orthogonal basis in the interval $[0, d]$. Particularly, the following identity hold

$$\frac{1}{d}\int_0^d e^{ia_n x}e^{-ia_m x}\,dx = \delta_{n,m} \text{ with } \delta_{n,m} := \begin{cases} 1 & \text{if} \quad n = m \\ 0 & \text{else.} \end{cases} \tag{9}$$

As indicated in equation (8), the scattered wave $u^s := (u - u^i)$ satisfies the outgoing boundary conditions that are imposed on Γ_m, $m = 1, 2$. Outgoing waves in the exterior domain ($|x_2| \geq L$) are written in terms of the so-called Rayleigh expansions [33,43]

$$u^{\pm} := \sum_{n=-\infty}^{\infty} A_n^{\pm} e^{i(a_n+a)x_1 \pm ib_n x_2}, \quad \text{with} \quad A_n^{\pm} := \frac{1}{d} e^{-ib_n L} \int_0^d u^{\pm}(s, \pm L) e^{-i(a_n+a)s} \, ds, \tag{10}$$

where

$$c_n := \omega^2 - (a_n + a)^2 \quad \text{and} \quad b_n := \begin{cases} \sqrt{c_n}, & c_n \geq 0, \\ i\sqrt{-c_n}, & c_n < 0. \end{cases} \tag{11}$$

The coefficients $A_n^{\pm}$ in expansion (10) are obtained by assuming that $u^{\pm}(\boldsymbol{x})$ is known for $\boldsymbol{x} \in \Gamma^{\pm}$ and using the orthogonality property.

Let $H_{\sharp}^1(\Omega)$ denote the space of quasi-periodic functions defined as

$$H_{\sharp}^1(\Omega) := \{u \in H^1(\Omega) : u(\boldsymbol{x} + d\mathbf{e}_1) = \psi u(\boldsymbol{x}) \text{ for } \boldsymbol{x} \in \Gamma_L\}, \quad \text{with} \quad \psi := e^{iad}. \tag{12}$$

We define $H_{\sharp}^{1/2}(\Gamma_m)$ as the trace of $H_{\sharp}^1(\Omega)$ on the boundary Γ_m and $H_{\sharp}^{-1/2}(\Gamma_m)$ as its dual space.

To require that the solution is outgoing, we employ the so-called Robin-to-Robin (RtR) maps, which map Robin conditions for $u \in H_{\sharp}^1(\Omega)$ with Robin conditions for the Rayleigh expansion (10) that are matched on Γ_m. In this way, the operators R_m are defined as the application of Robin boundary conditions on the Rayleigh expansion (10)

$$\begin{aligned} R_m : \quad & H_{\sharp}^{1/2}(\Gamma_m) & \rightarrow \quad & H_{\sharp}^{-1/2}(\Gamma_m), \\ & \sum_{n=-\infty}^{\infty} A_n^{\pm} e^{i(a_n+a)x_1 \pm ib_n L} & \rightarrow \quad & \sum_{n=-\infty}^{\infty} \frac{i(b_n - \sigma)}{d} A_n^{\pm} e^{i(a_n+a)x_1 \pm ib_n L}. \end{aligned} \tag{13}$$

Therefore, the effect of the operators R_m applied to u is given by

$$R_m u := \sum_{n=-\infty}^{\infty} \frac{i(b_n - \sigma)}{d} e^{i(a_n+a)x_1} \int_0^d u(s, \pm L) e^{-i(a_n+a)s} \, ds, \quad m = 1, 2. \tag{14}$$

In practice, a truncated operator R_m^l is used instead of the full series (14). Similarly as for Dirichlet-to-Newmann formulations [53,54], the accuracy of the FE computation depends critically on the number of terms used in the truncation of the series.

The so-called Poynting's vector $\mathbf{S} := \mathbf{E} \times \overline{\mathbf{H}}$ specifies the pointwise direction of wave propagation and its magnitude is equivalent to the electromagnetic energy flux. The energy of the electromagnetic scattered wave leaving Ω through the boundary Γ_k is proportional to the functional

$$\text{Re} \int_{\Gamma_k} \mathbf{S} \cdot \mathbf{n}_k \, d\boldsymbol{x} = \frac{1}{\omega} \text{Im} \int_{\Gamma_k} u^s \nabla \bar{u}^s \cdot \mathbf{n}_k \, d\boldsymbol{x}, \tag{15}$$

where $\mathbf{n}_k$ is the outer normal vector to Γ_k.

3. Variational formulation and discretization

We introduce smooth quasi-periodic functions u and v and use Green identities on the first equation in (8) to obtain

$$\int_{\Omega} \rho(\boldsymbol{x}) \nabla u \cdot \nabla \bar{v} \, d\boldsymbol{x} - \omega^2 \int_{\Omega} \eta(\boldsymbol{x}) u \bar{v} \, d\boldsymbol{x} - \sum_{m=1}^{2} \int_{\Gamma_m} \nabla u \cdot \mathbf{n}_j \bar{v} \, d\boldsymbol{x} = 0. \tag{16}$$

The contributions on Γ_L and Γ_R cancel each other since $\boldsymbol{n}_L = -\boldsymbol{n}_R$ and since u and v are quasi-periodic functions.

The variational formulation to problem (8) reads: For given $f \in L^2(\Gamma)$, $\omega \in \mathbb{R}$, and $\rho, \eta \in L^{\infty}(\Omega)$, find $u \in H_{\sharp}^1(\Omega)$ such that

$$\int_{\Omega} \rho(\boldsymbol{x}) \nabla u \cdot \nabla \bar{v} \, d\boldsymbol{x} - \omega^2 \int_{\Omega} \eta(\boldsymbol{x}) u \bar{v} \, d\boldsymbol{x} - \sum_{m=1}^{2} \int_{\Gamma_m} (R_m + i\sigma) u \bar{v} \, d\boldsymbol{x} = \int_{\Gamma_1} f \bar{v} \, d\boldsymbol{x}, \tag{17}$$

for all $v \in H_{\sharp}^1(\Omega)$. In the case of illumination with a plane wave u^i as in equation (7), we obtain $f = -2i\omega \sin\theta \, e^{i\mathbf{k} \cdot \boldsymbol{x}}$ corresponding to the source term in equation (17).

3.1. Stability using the RtR map

In this section, we cite articles that use a different but equivalent formulation of the problem. Therefore, we will reformulate the problem in terms of a shifted gradient operator $\nabla_a := \nabla + ia\mathbf{e}_1$, where $\mathbf{e}_1$ is the unit vector in the x-direction. The operator ∇_a is in the alternative variational formulation applied to x-periodic functions.

The function u is quasi-periodic and $u_a = ue^{-iax}$ is a x-periodic function with period d. Let $H^1_{\text{per}}(\Omega)$ denote the space of x-periodic functions with period d. The trace of u_a on Γ_m has the Fourier series expansion

$$\hat{u}(x) := \sum_{n \in \mathbb{Z}} u_n^m e^{ia_n x}, \quad a_n = 2n\pi/d, \tag{18}$$

where the Fourier coefficients u_n^m, $m = 1, 2$ indicate that the trace is on Γ_m. For each $s \in \mathbb{R}$, we let $\mathcal{H}^s(\Gamma_m)$ denote the space

$$\mathcal{H}^s(\Gamma_m) := \left\{ \hat{v} = \sum_{n \in \mathbb{Z}} v_n^m e^{ia_n x} : \sum_{n \in \mathbb{Z}} \left(1 + a_n^2\right)^s |v_n^m|^2 < \infty \right\}. \tag{19}$$

Set $\hat{a}_n = a + a_n$, $c_n = \omega^2 - \hat{a}_n^2$, and

$$b_n = \begin{cases} \sqrt{c_n}, & \text{for } |\hat{a}_n| \le \omega, \\ i\sqrt{-c_n}, & \text{for } |\hat{a}_n| > \omega. \end{cases} \tag{20}$$

Let $C^\infty_{\text{per}}(\Omega)$ denote the space of smooth x-periodic functions with period d. The space $H^1_{\text{per}}(\Omega)$ is then defined as the completion of $C^\infty_{\text{per}}(\Omega)$ in the H^1-norm. Define for $u_a, v_a \in H^1_{\text{per}}(\Omega)$ the sesquilinear form

$$t(\omega)[u_a, v_a] := t_1[u_a, v_a] - (\omega^2 + 1)t_2[u_a, v_a], \tag{21}$$

where

$$t_1[u_a, v_a] := \int_\Omega \rho \nabla_a u_a \cdot \overline{\nabla_a v_a} + \eta u_a \bar{v}_a \, d\boldsymbol{x} - \sum_{m=1}^{2} \int_{\tilde{\Gamma}_m} (R_m + i\sigma) u_a \bar{v}_a \, d\boldsymbol{x} \tag{22}$$

and

$$t_2[u_a, v_a] := \int_\Omega \eta u_a \bar{v}_a \, d\boldsymbol{x}. \tag{23}$$

An equivalent formulation of (17) is then: Find $u_a \in H^1_{\text{per}}(\Omega)$ such that

$$t(\omega)[u_a, v_a] = \int_{\Gamma_1} f \bar{v}_a \, d\boldsymbol{x} \quad \text{for all} \quad v_a \in H^1_{\text{per}}(\Omega). \tag{24}$$

We have that $\text{Re } t_1[u_a, u_a] \ge \min\{\rho, \eta\} \|u_a\|^2_{H^1_{\text{per}}}$ and t_2 is compact [55], i.e. the sesquilinear form can be represented by a compact operator on $H^1_{\text{per}}(\Omega)$. Hence, $t(\omega)$ fulfills a Gårding inequality, which implies that the Fredholm alternative holds. Hence, the original problem has an unique solution if the problem: Find $u_a \in H^1_{\text{per}}(\Omega)$ such that $t(\omega)[u_a, u_a] = 0$ only has the trivial solution.

The sum over $n \in \mathbb{Z}$ in the Fourier series can be decomposed into a sum over the sets

$$S^- := \{n \in \mathbb{Z} : c_n < 0\}, \quad S^0 := \{n \in \mathbb{Z} : c_n = 0\}, \quad S^+ := \{n \in \mathbb{Z} : c_n > 0\}, \tag{25}$$

where S^-, S^0, and S^+ correspond to the exponentially decaying, constant, and propagating waves, respectively. The sets S^+ and S^0 are by definition finite and the set S^- is for non-trivial cases an infinite set.

The truncated RtR operator $R_m^N : \mathcal{H}^{1/2}(\Gamma_m) \to \mathcal{H}^{-1/2}(\Gamma_m)$ is

$$R_m^N \hat{u} := \sum_{|n| \le N} i(b_n - \sigma) u_n^m e^{ia_n x}, \quad \hat{u}(x) = \sum_{n \in \mathbb{Z}} u_n^m e^{ia_n x}. \tag{26}$$

Let $\sigma > 0$. Then we have

$$\text{Im } ((R_m^N + i\sigma)u_a, u_a)_{\Gamma_m} = \sum_{n \in S^+, |n| \le N} \sqrt{c_n} |u_n^m|^2 + \sigma \sum_{|n| > N} |u_n^m|^2, \tag{27}$$

which implies that $t(\omega)[u_a, u_a] = 0$ only is possible if $u_n^m = 0$ for $n \in S^+$, $|n| \le N$ and $u_n^m = 0$ for all $|n| > N$. This improves the corresponding result for the DtN map [43, Lemma 3.2].

3.2. Energy balance

The output energy over a boundary segment Γ_k is proportional to

$$J_k := \operatorname{Im} \int_{\Gamma_k} u^s \, \nabla \bar{u}^s \cdot \mathbf{n}_k \, \mathrm{d}\boldsymbol{x} = \frac{1}{2\mathrm{i}} \int_{\Gamma_k} (u^s \, \nabla \bar{u}^s - \bar{u}^s \, \nabla u^s) \cdot \mathbf{n}_k \, \mathrm{d}\boldsymbol{x}, \; k = 0, 1, 2, \tag{28}$$

where $u^s := u - u^i$, with u^i as defined in expression (7), and u solves governing equation (8).

Under the assumption of non-dispersive dielectrics, we work with the real valued piecewise constants ρ and η. By setting $v \equiv u$, and taking the imaginary part of equation (16), we obtain the total flux

$$\Phi_T := \sum_{j=1}^{2} \operatorname{Im} \int_{\Gamma_j} \nabla u \cdot \boldsymbol{n}_j \bar{u} \, \mathrm{d}\boldsymbol{x} = 0. \tag{29}$$

We define the net scattered flux as

$$\Phi_S := J_1 + J_2. \tag{30}$$

Lemma. *For the given incoming wave* (7)*, the net scattered flux is* $\Phi_S = b_0 d$.

To prove the claim, we first replace $u = u^s + u^i$ on Γ_1 in expression (29). Using definition (28) and expanding, we obtain

$$J_1 + J_2 + \operatorname{Im} \left\{ \int_{\Gamma_1} \left(\nabla u^s \cdot \boldsymbol{n}_1 \bar{u}^i \, \mathrm{d}\boldsymbol{x} + \nabla u^i \cdot \boldsymbol{n}_1 \bar{u}^s \right) \mathrm{d}\boldsymbol{x} + \int_{\Gamma_1} \nabla u^i \cdot \boldsymbol{n}_1 \bar{u}^i \, \mathrm{d}\boldsymbol{x} \right\} = 0. \tag{31}$$

Next, we use $u^i := \mathrm{e}^{\mathrm{i}(ax - b_0 y)}$ and $\|u^i\| = 1$, in order to evaluate equation (31). From where we reach

$$\operatorname{Im} \int_{\Gamma_1} \nabla u^i \cdot \boldsymbol{n}_1 \bar{u}^i \, \mathrm{d}\boldsymbol{x} = -b_0 d. \tag{32}$$

Next, we show that the mixed terms in (31) vanish. Notice that for $\boldsymbol{x} \in \Gamma_1$ we have $\boldsymbol{x} := (x, L)^{\top}$, $u^s := u^+$, $\mathbf{n}_1 := (0, 1)^{\top}$ and $\nabla v \cdot \mathbf{n}_1 \equiv \partial_y v$. Additionally, by definition $\nabla u^i \cdot \mathbf{n}_1 = -\mathrm{i} b_0 u^i$ and from the Rayleigh expansion (10), we write for $\boldsymbol{x} \in \Gamma_1$

$$u^s = \sum_{n=-\infty}^{\infty} A_n \mathrm{e}^{\mathrm{i}(a_n + a)x + \mathrm{i} b_n L} \text{ and } \nabla u^s \cdot \mathbf{n}_1 = \sum_{n=-\infty}^{\infty} \mathrm{i} b_n A_n \mathrm{e}^{\mathrm{i}(a_n + a)x + \mathrm{i} b_n L}. \tag{33}$$

Thus,

$$\int_0^d \partial_y u^s \, \bar{u}^i \, \mathrm{d}s = \sum_{n=-\infty}^{\infty} \mathrm{i} b_n A_n \int_0^d \mathrm{e}^{\mathrm{i}(a_n + a)s + \mathrm{i} b_n L} \mathrm{e}^{-\mathrm{i}(as - b_0 L)} \, \mathrm{d}s = \mathrm{i} d \, b_0 \, A_0 \mathrm{e}^{2\mathrm{i} b_0 L}, \tag{34}$$

by using the orthogonality property. Similarly, we compute

$$\int_0^d \partial_y u^i \, \bar{u}^s \, \mathrm{d}s = -\mathrm{i} d \, b_0 \, \bar{A}_0 \mathrm{e}^{-2\mathrm{i} b_0 L}. \tag{35}$$

Then,

$$\operatorname{Im} \int_{\Gamma_1} (\nabla u^s \cdot \boldsymbol{n}_1 \bar{u}^i \, \mathrm{d}\boldsymbol{x} + \nabla u^i \cdot \boldsymbol{n}_1 \bar{u}^s) \, \mathrm{d}\boldsymbol{x} = \operatorname{Re} \left(d \, b_0 \, A_0 \mathrm{e}^{2\mathrm{i} b_0 L} - d \, b_0 \overline{A_0 \mathrm{e}^{2\mathrm{i} b_0 L}} \right) = 0, \tag{36}$$

where we take the imaginary part and use identities of complex numbers to complete the proof.

Equations (29) and (30) are used in order to measure the quality of our numerical solution strategy that is introduced in the following section.

3.3. Objective function

We parameterize the upper *interface* of the dielectric slab Ω_s as the curve $Y(x)$ as depicted in Fig. 1. The function $Y(x)$ is bounded by the fixed values y_1 and y_2. That is, the designs satisfy $0 < y_1 < Y(x) < y_2 < L$, and we set $y_m := (y_1 + y_2)/2$. In this setting, it is natural to use the following Fourier representation of the interface of the slab

$$Y(x) = y_m + \sum_{j=1}^{N} \alpha_j \sin \frac{2\pi j}{d} x, \tag{37}$$

from where the parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^\top$ are chosen as the design variables.

Routing efficiency: To measure the efficiency of the designs, we use the ratio between the routed wave energy entering Γ_0 and the losses due to scattering away from the slab. That is, we consider

$$Q_e(\boldsymbol{\alpha}) := \left| \frac{J_0}{J_1 + J_2} \right| = \frac{|J_0|}{b_0 d}. \tag{38}$$

The absolute value indicates that the direction of routing is not important. Definition (38) naturally motivates the choice of J_0^2, as the functional to be maximized. Thus, we introduce the so-called *routing objective* as

$$F(\boldsymbol{\alpha}) := \log J_0^2. \tag{39}$$

We cast the constrained optimization problem as an unconstrained optimization problem by introducing a penalization that increases when $Y(\boldsymbol{\alpha}, x)$ oscillates away from y_m and grows unbounded as $Y(\boldsymbol{\alpha}, x)$ tends to y_1 or y_2. In order to enforce the bound constraints, we introduce the penalty function

$$F_c(\boldsymbol{\alpha}) := -\int_0^d \log |Y(\boldsymbol{\alpha}, x) - y_1| + \log |y_2 - Y(\boldsymbol{\alpha}, x)| \, dx. \tag{40}$$

This penalty function is an integral version including logarithmic barrier functions, which are routinely used to enforce inequality constraints in non-linear optimization. Moreover, to promote optimal designs with less oscillatory shapes, we use the following penalization term

$$F_p(\boldsymbol{\alpha}) := \int_0^d \left(\frac{d^2 Y}{dx^2} \right)^2 dx = \frac{1}{2} \left(\frac{2\pi}{d} \right)^4 \sum_{j=1}^{N} j^4 \alpha_j^2, \tag{41}$$

where we have used the orthogonality of trigonometric functions and representation (37).

We define the objective function as

$$\min_{\boldsymbol{\alpha}} \; -F(\boldsymbol{\alpha}) + \tau F_c(\boldsymbol{\alpha}) + \gamma F_p(\boldsymbol{\alpha}), \tag{42}$$

for given penalty parameters $\tau > 0$ and $\gamma > 0$.

3.4. Discretization

Let the domain $\Omega \subset \mathbb{R}^2$ be covered with a regular and quasi-uniform finite element mesh $\mathcal{T}(\Omega)$ consisting of N_K quadrilateral elements $\{K_j\}_{j=1}^{N_K}$. The mesh is *shape regular* [56, Sec. 4.3] and designed such that the permittivity ϵ is constant in each K_j. Let h_j be the length of the largest diagonal of the non-curved primitive K_j and denote by h the maximum mesh size $h := \max_j h_j$. Additionally, $\mathcal{P}_p$ denotes the space of polynomials on $\mathbb{R}^2$ of degree $\leq p$ in each space coordinate. Finally, we define the N_h dimensional finite element space as

$$V_h(\Omega) := \{u \in H^1_\sharp(\Omega) : u|_{K_j} \in \mathcal{P}_p(K_j) \text{ for } K_j \in \mathcal{T}\}. \tag{43}$$

We introduce the shape functions $\{\varphi_j\}_{j=1}^{N_h}$, the solution vector $\mathbf{u} := (u_1, u_2, \ldots, u_{N_h})^\top$, and the FE representation $u_h = \sum_1^{N_h} u_j \varphi_j$.

To use the exact parameterization of shapes, we employ curvilinear elements [57] and bend the edges of an initial mesh of good quality. Curved boundaries or interfaces are implemented by a bilinear transfinite interpolant [58,59], where the parameterization of curved edges are available. For reference, computational details and application to FE assembly can be found in the paper by Gordon and Hall [58] as well as in the reference book by Solin [60, Sec. 3.2].

3.5. Details on the FE discretization

The outgoing condition follows definition (14), and the RtR is implemented in a similar way as it was done for the DtN in Araújo et al. [54]. We compute along Γ_m

$$
\begin{aligned}
(R_m \varphi_j, \varphi_i)_{\Gamma_m} &= \sum_{n=-l}^{l} \frac{i(b_n-\sigma)}{d} \left(\int_0^d \varphi_j(x, \pm L)\, e^{-i(a_n+a)x}\, dx \right) \left(\int_0^d \bar{\varphi}_i(x', \pm L)\, e^{i(a_n+a)x'}\, dx' \right) \\
&= \sum_{n=-l}^{l} \frac{i(b_n-\sigma)}{d} \hat{\varphi}_j^n\, \bar{\hat{\varphi}}_i^n,
\end{aligned}
\tag{44}
$$

where we have used that $\hat{\varphi}_j^n = \int_0^d \varphi_j\, e^{-i(a_n+a)x}\, dx$.

From variational form (17), we obtain our FE matrices $\mathbf{A}, \mathbf{M}$, and $\mathbf{Q}^n$ and load vector $\mathbf{b}$ with entries

$$
A_{ij} = (\rho \nabla\varphi_j, \nabla\varphi_i)_\Omega,\ M_{ij} = (\eta\varphi_j, \varphi_i)_\Omega,\ Q_{ij}^n = \hat{\varphi}_j^n\, \bar{\hat{\varphi}}_i^n,\ G_{ij}^m = (\varphi_j, \varphi_i)_{\Gamma_m},\ \text{and } b_i = (f, \varphi_i)_{\Gamma_1},
\tag{45}
$$

respectively. The truncated RtR matrix contribution is

$$
\mathbf{R}_m^l := \sum_{n=-l}^{l} \frac{i(b_n-\sigma)}{d} \mathbf{Q}^n + i\sigma\, \mathbf{G}_m.
\tag{46}
$$

Thus the state equation, or discrete scattering problem, becomes

$$
\mathbf{S}(\omega)\, \mathbf{u} := \left(\mathbf{A} - \omega^2 \mathbf{M} - \sum_{m=1}^{2} \mathbf{R}_m^l \right) \mathbf{u} = \mathbf{b}.
\tag{47}
$$

In the discrete case, functional (28) is evaluated as

$$
J_k(\boldsymbol{\alpha}) := \mathrm{Im}\{\overline{(\mathbf{u} - \mathbf{u}^i)}^\top \mathbf{W}_k(\mathbf{u} - \mathbf{u}^i)\}, \ \text{for } k = 0, 1, 2,
\tag{48}
$$

where $\mathbf{u}^i$ stands for the FE interpolant of u^i, and $\mathbf{W}_k$ are the *window* matrices with entries

$$
\{\mathbf{W}_k\}_{ij} := \int_{\Gamma_k} (\varphi_i \nabla\varphi_j - \varphi_j \nabla\varphi_i) \cdot \mathbf{n}_k\, d\boldsymbol{x}, \ \text{for } k = 0, 1, 2.
\tag{49}
$$

3.6. Quasi-periodicity conditions

Consider support points $\boldsymbol{x}_L \in \Gamma_L$, the FE mesh has been set up such that Γ_L and Γ_R have matching support points satisfying $\boldsymbol{x}_L + d\mathbf{e}_1 =: \boldsymbol{x}_R \in \Gamma_R$. The quasi-periodic conditions are imposed in a similar way as done in Zolla et al. [61, Sec. 4.5]. Assume that we have arranged the FE nodal values in the following fashion $\mathbf{u} := (\mathbf{u_C}^\top, \mathbf{u_L}^\top, \mathbf{u_R}^\top)^\top$, where $\mathbf{u_L}$ and $\mathbf{u_R}$ hold nodal values on support points on Γ_L and Γ_R respectively, and $\mathbf{u_C}$ holds the remaining nodal values. Let $\psi = e^{i\omega d \cos\theta}$, $\mathbf{v} := (\mathbf{u_C}^\top, \mathbf{u_L}^\top)^\top$ and $\mathbf{P}$ be the matrix such that $\mathbf{Pv} := (\mathbf{u_C}^\top, \mathbf{u_L}^\top, \psi \mathbf{u_L}^\top)^\top$. Simple inspection gives

$$
\mathbf{P} := \begin{bmatrix} \mathbf{I} & \\ & \mathbf{I} \\ & \psi\mathbf{I} \end{bmatrix} \text{ and } \mathbf{Pv} := \begin{bmatrix} \mathbf{u_C} \\ \mathbf{u_L} \\ \psi\mathbf{u_L} \end{bmatrix}.
\tag{50}
$$

In order to impose the periodicity condition $\mathbf{u_R} = \psi\mathbf{u_L}$ at matrix level, consider the 3×3 block matrix $\mathbf{S}$. Then, we compute explicitly

$$
\bar{\mathbf{P}}^\top \mathbf{SP} := \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} + \psi\mathbf{S}_{13} \\ \mathbf{S}_{21} + \bar{\psi}\mathbf{S}_{31} & \mathbf{S}_{22} + \mathbf{S}_{33} + \psi\mathbf{S}_{23} + \bar{\psi}\mathbf{S}_{32} \end{bmatrix} \text{ and } \bar{\mathbf{P}}^\top \mathbf{b} := \begin{bmatrix} \mathbf{b_C} \\ \mathbf{b_L} + \bar{\psi}\mathbf{b_R} \end{bmatrix}.
\tag{51}
$$

Let $\tilde{\mathbf{S}} := (\bar{\mathbf{P}}^\top \mathbf{SP})$ and $\tilde{\mathbf{b}} := \bar{\mathbf{P}}^\top \mathbf{b}$. Then, we solve for $\mathbf{v}$ from the reduced system

$$
\tilde{\mathbf{S}}\mathbf{v} = \tilde{\mathbf{b}}
\tag{52}
$$

and recover $\mathbf{u_R} = \psi\mathbf{u_L}$.

3.7. Numerical sensitivities

For simplicity of the presentation the derivatives with respect to design variables are referred to as *sensitivities*. The sensitivities of objective function (42) are

$$\nabla f := \left(\frac{\partial f}{\partial \alpha_1}, \frac{\partial f}{\partial \alpha_2}, \ldots, \frac{\partial f}{\partial \alpha_N} \right)^{\top} \text{ with } \frac{\partial f}{\partial \alpha_j} := -\frac{\partial F}{\partial \alpha_j} + \tau \frac{\partial F_c}{\partial \alpha_j} + \gamma \frac{\partial F_p}{\partial \alpha_j}, \; j = 1, 2, \ldots, N. \tag{53}$$

The sensitivities corresponding to the penalization terms can be written as follows. For the penalty terms, we write

$$\frac{\partial F_c}{\partial \alpha_j} = -\int_0^d \left(\frac{1}{Y(\boldsymbol{\alpha}, x) - y_1} - \frac{1}{y_2 - Y(\boldsymbol{\alpha}, x)} \right) \frac{\partial Y}{\partial \alpha_j} \, dx \text{ and } \frac{\partial F_p}{\partial \alpha_j} = \left(\frac{2\pi j}{d} \right)^4 \alpha_j. \tag{54}$$

Next, we briefly describe how to obtain an exact expression that allows us to compute the sensitivities (53) by using an exact discrete adjoint method [62, Sec. 6.2.1]. We remark that the computation of sensitivities for the present case is more challenging than what was presented in our previous work [63] since we need to include the influence of the quasi-periodic conditions in the adjoint equation.

The state $\mathbf{v}(\boldsymbol{\alpha})$ solves the system $\tilde{\mathbf{S}}(\boldsymbol{\alpha})\mathbf{v}(\boldsymbol{\alpha}) = \tilde{\mathbf{b}}(\boldsymbol{\alpha})$, from where $u(\boldsymbol{\alpha})$ is recovered and used to evaluate $F(\boldsymbol{\alpha})$ using expression (39). The sensitivities of $J_k(\boldsymbol{\alpha})$ are obtained [63] by employing the chain rule for complex valued problems:

$$\frac{\partial J_k}{\partial \alpha_j} = \frac{1}{2i} \left((\mathbf{W}_k \mathbf{u})^{\top} \frac{\partial \mathbf{u}}{\partial \alpha_j} + (\mathbf{W}_k^{\top} \bar{\mathbf{u}})^{\top} \frac{\partial \bar{\mathbf{u}}}{\partial \alpha_j} \right) \equiv \text{Im} \left\{ (\mathbf{W}_k \mathbf{u})^{\top} \frac{\partial \mathbf{u}}{\partial \alpha_j} \right\}, \tag{55}$$

for $k = 0, 1, 2$, where we have used the anti-symmetric property $\mathbf{W}_k^{\top} = -\mathbf{W}_k$.

By differentiating state equation (52), we obtain

$$\frac{\partial \tilde{\mathbf{S}}}{\partial \alpha_j} \mathbf{v} + \tilde{\mathbf{S}} \frac{\partial \mathbf{v}}{\partial \alpha_j} = \mathbf{0}, \text{ which we reorder as } \tilde{\mathbf{S}} \frac{\partial \mathbf{v}}{\partial \alpha_j} = -\frac{\partial \tilde{\mathbf{S}}}{\partial \alpha_j} \mathbf{v}. \tag{56}$$

The periodicity constraints are enforced into the state sensitivities in the following way

$$\frac{\partial \mathbf{u}}{\partial \alpha_j} = \mathbf{P} \frac{\partial \mathbf{v}}{\partial \alpha_j} = -\mathbf{P}\tilde{\mathbf{S}}(\boldsymbol{\alpha})^{-1} \left[\frac{\partial \tilde{\mathbf{S}}(\boldsymbol{\alpha})}{\partial \alpha_j} \mathbf{v}(\boldsymbol{\alpha}) \right] = -\mathbf{P}\tilde{\mathbf{S}}(\boldsymbol{\alpha})^{-1} \left[\bar{\mathbf{P}}^{\top} \frac{\partial \mathbf{S}(\boldsymbol{\alpha})}{\partial \alpha_j} \mathbf{u}(\boldsymbol{\alpha}) \right]. \tag{57}$$

Let the adjoint vector $\boldsymbol{\lambda} \in \mathbb{C}^{N_h}$ be the solution to the so-called adjoint equation

$$\tilde{\mathbf{S}}^{\top}(\boldsymbol{\alpha})\boldsymbol{\lambda} = \mathbf{g}, \quad \text{with} \quad \mathbf{g} := -\mathbf{P}^{\top}\mathbf{W}_0\mathbf{u}. \tag{58}$$

Then, we obtain the sensitivities

$$\frac{\partial J_0}{\partial \alpha_j} = \text{Im} \left\{ \boldsymbol{\lambda}^{\top} \left(-\bar{\mathbf{P}}^{\top} \frac{\partial \mathbf{S}(\boldsymbol{\alpha})}{\partial \alpha_j} \mathbf{u}(\boldsymbol{\alpha}) \right) \right\}. \tag{59}$$

We describe in Section 3.7.1 how the sensitivities for the system matrix and load vector, used in (59), are computed.

3.7.1. Explicit computation of matrix sensitivities

In this section, we briefly present how the sensitivities from the state equation are computed. Let $\mathcal{K} := (-1, 1)^2$ denote the reference element, where chart points $\boldsymbol{\xi} \in \mathcal{K}$ are assigned the coordinates $\boldsymbol{\xi} = (\xi, \eta)^{\top}$. Consider a point $\boldsymbol{x} = (x, y)^{\top}$ in a physical element K. Let $X_K : (\xi, \eta) \to (x, y)$ be the mapping transforming coordinates from the reference to physical element. In the case where K is a quadrilateral, the transformation is a bilinear mapping. Furthermore, when K has curved edges, then we represent $X_K(\mathcal{K})$ by a bilinear Transfinite Interpolation.

Consider the physical (quadrilateral) element K with primitive edges $e_1, \ldots, e_4$ and vertices $\mathbf{x}_1, \ldots, \mathbf{x}_4$. Curved scatterers are represented on a FE cell K by enforcing its edges to follow smooth curves. These are defined by the parameterizations $X^{e_1}(\zeta), \ldots, X^{e_4}(\zeta) \subset \mathbb{R}^2$, $\zeta \in [-1, 1]$. The vector valued bilinear transfinite interpolation formula is given by

$$
\begin{aligned}
X_K(\boldsymbol{\xi}) &:= \begin{pmatrix} X_{K,1}(\boldsymbol{\xi}) \\ X_{K,2}(\boldsymbol{\xi}) \end{pmatrix} \\
&= \frac{1-\xi}{2} X^{e_1}(\eta) + \frac{1+\xi}{2} X^{e_2}(\eta) + \frac{1-\eta}{2} X^{e_3}(\xi) + \frac{1+\eta}{2} X^{e_4}(\xi) \\
&\quad - \frac{(1-\xi)}{2} \frac{(1-\eta)}{2} \mathbf{x}_1 - \frac{(1-\xi)}{2} \frac{(1+\eta)}{2} \mathbf{x}_4 \\
&\quad - \frac{(1+\xi)}{2} \frac{(1-\eta)}{2} \mathbf{x}_2 - \frac{(1+\xi)}{2} \frac{(1+\eta)}{2} \mathbf{x}_3,
\end{aligned}
\tag{60}
$$

satisfying $X_K(-1,-1) = \mathbf{x}_1$, $X_K(1,-1) = \mathbf{x}_2$, $X_K(1,1) = \mathbf{x}_3$, $X_K(-1,1) = \mathbf{x}_4$ (see [58,59] and [60, Sec. 3.3]). This is, at corners the mapping X_K corresponds to the value of the physical nodes.

In the matrix assembly process, integrals over K are mapped to the reference element by using the X_K's Jacobian matrix $\mathbf{J}$ and its determinant $|\mathbf{J}| = \det \mathbf{J}$ [60, Sec. 3.4]. At element K, we have

$$\mathbf{J} := \begin{bmatrix} \dfrac{\partial X_{K,1}}{\partial \xi} & \dfrac{\partial X_{K,1}}{\partial \eta} \\[3mm] \dfrac{\partial X_{K,2}}{\partial \xi} & \dfrac{\partial X_{K,2}}{\partial \eta} \end{bmatrix}, \quad \mathbf{Q} := \begin{bmatrix} \dfrac{\partial X_{K,2}}{\partial \eta} & -\dfrac{\partial X_{K,2}}{\partial \xi} \\[3mm] -\dfrac{\partial X_{K,1}}{\partial \eta} & \dfrac{\partial X_{K,1}}{\partial \xi} \end{bmatrix}, \quad \mathbf{J}^{-1} = \frac{1}{|\mathbf{J}|}\mathbf{Q}, \text{ and } |\mathbf{J}| := \det \mathbf{J}. \tag{61}$$

In this setting, we denote by φ_j and $\nabla\varphi_j$ the shape functions and shape gradients corresponding to the physical mesh. Similarly, ϕ_j and $\nabla_\xi \phi_j$ denote the Lagrange basis functions and gradients corresponding to $\mathcal{K}$. The assembly of the FE matrices given in Section 3.5 is performed by adding local FE matrices computed at each element. At element K, the local mass matrix is computed as

$$M_{ij}^K = \int_K \eta_K \varphi_j(\mathbf{x})\varphi_i(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_\mathcal{K} \eta_K \phi_j(\boldsymbol{\xi})\phi_i(\boldsymbol{\xi})\,|\mathbf{J}(\boldsymbol{\xi})|\,\mathrm{d}\boldsymbol{\xi}, \tag{62}$$

and the sensitivities are computed as

$$\frac{\partial M_{ij}^K}{\partial \alpha_m} = \frac{\partial}{\partial \alpha_m}\int_K \eta_K \varphi_j(\mathbf{x})\varphi_i(\mathbf{x})\,\mathrm{d}\mathbf{x} = \eta_K \int_\mathcal{K} \phi_j(\boldsymbol{\xi})\phi_i(\boldsymbol{\xi})\frac{\partial |\mathbf{J}|}{\partial \alpha_m}\,\mathrm{d}\boldsymbol{\xi}, \tag{63}$$

where we have used that η is an elementwise constant function. We let η_K and ρ_K denote the values of η and ρ in element K. Notice that since $\mathcal{K}$ is fixed then, ϕ_j and $\nabla_\xi \phi_j$ are not sensitive to variations of $\boldsymbol{\alpha}$.

Similarly, the stiffness matrix is computed as

$$A_{ij}^K = \int_K \rho_K \nabla\varphi_j(\mathbf{x}) \cdot \nabla\varphi_i(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_\mathcal{K} \rho_K \mathbf{J}^{-1}\nabla_\xi \phi_j(\boldsymbol{\xi}) \cdot \mathbf{J}^{-1}\nabla_\xi \phi_i(\boldsymbol{\xi})\,|\mathbf{J}(\boldsymbol{\xi})|\,\mathrm{d}\boldsymbol{\xi}, \tag{64}$$

where we have used the properties $\phi_j(\boldsymbol{\xi}) = \varphi_j \circ X_K(\boldsymbol{\xi})$ and $\nabla_\xi \phi_j = \mathbf{J}\nabla\varphi_j$. The corresponding computation of the sensitivities is
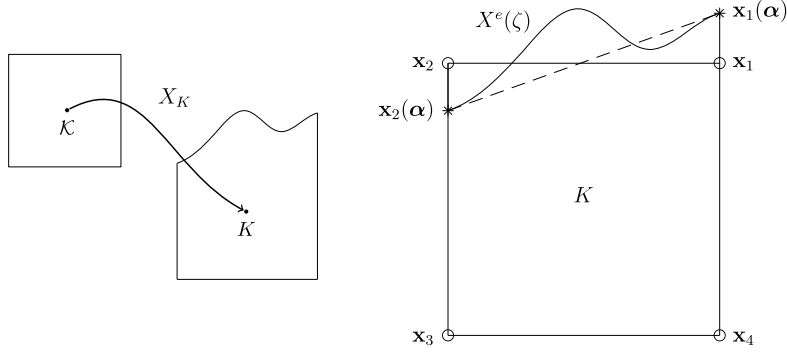
$$\begin{aligned} \frac{\partial A_{ij}^K}{\partial \alpha_m} &= \frac{\partial}{\partial \alpha_m}\int_K \rho_K \nabla\varphi_j(\mathbf{x}) \cdot \nabla\varphi_i(\mathbf{x})\,\mathrm{d}\mathbf{x} \\[2mm] &= \rho_K \frac{\partial}{\partial \alpha_m}\int_\mathcal{K} \mathbf{J}^{-1}\nabla_\xi \phi_j(\boldsymbol{\xi}) \cdot \mathbf{J}^{-1}\nabla_\xi \phi_i(\boldsymbol{\xi})\,|\mathbf{J}(\boldsymbol{\xi})|\,\mathrm{d}\boldsymbol{\xi} \\[2mm] &= \rho_K \frac{\partial}{\partial \alpha_m}\int_\mathcal{K} \mathbf{Q}\nabla_\xi \phi_j \cdot \mathbf{Q}\nabla_\xi \phi_i \frac{1}{|\mathbf{J}(\boldsymbol{\xi})|}\,\mathrm{d}\boldsymbol{\xi} \\[2mm] &= \rho_K \frac{\partial}{\partial \alpha_m}\int_\mathcal{K} (\nabla_\xi \phi_j)^\top \mathbf{Q}^\top \mathbf{Q}\nabla_\xi \phi_i \frac{1}{|\mathbf{J}(\boldsymbol{\xi})|}\,\mathrm{d}\boldsymbol{\xi} \\[2mm] &= \rho_K \int_\mathcal{K} (\nabla_\xi \phi_j)^\top \left(\frac{\partial \mathbf{Q}^\top}{\partial \alpha_m}\mathbf{Q} + \mathbf{Q}^\top \frac{\partial \mathbf{Q}}{\partial \alpha_m} - \frac{1}{|\mathbf{J}(\boldsymbol{\xi})|}\frac{\partial |\mathbf{J}(\boldsymbol{\xi})|}{\partial \alpha_m}\mathbf{Q}^\top \mathbf{Q}\right)\nabla_\xi \phi_i \frac{1}{|\mathbf{J}(\boldsymbol{\xi})|}\,\mathrm{d}\boldsymbol{\xi}. \end{aligned} \tag{65}$$

Finally, after *local-to-global* assembly, the sensitivities of the system matrix read

$$\frac{\partial \mathbf{S}}{\partial \alpha_m} = \frac{\partial \mathbf{A}}{\partial \alpha_m} - \omega^2 \frac{\partial \mathbf{M}}{\partial \alpha_m}. \tag{66}$$

Remark 1. The cells containing edges on the boundaries Γ_k, $k = 1,2$ and the edges on Γ_0 do not depend on the design variables $\boldsymbol{\alpha}$. This implies that the source vector as well as the RtR representation and window matrices $\mathbf{W}_k$ are not sensitive to variations of the shape. That is, for any j the following holds

$$\frac{\partial \mathbf{b}}{\partial \alpha_j} \equiv 0, \quad \frac{\partial \mathbf{R}_m^l(\omega)}{\partial \alpha_j} \equiv 0 \text{ and } \frac{\partial \mathbf{W}_k}{\partial \alpha_j} \equiv 0 \text{ for } k = 0,1,2, \ m = 1,2. \tag{67}$$

Fig. 2. Illustration of (left) the action of the mapping $X_K : \mathcal{K} \to K$ and (right) the corresponding parameterization used for representing the bending of the edge that belongs to the shape $\partial \Omega_s(\boldsymbol{\alpha})$.

Remark 2. In a numerical optimization strategy, Remark 1 implies that the matrices $\mathbf{R}_m^l$ and $\mathbf{W}_k$ only need to be assembled once and should not be updated in the iterative scheme. Additionally, the assembly of the sensitivities (66) may be performed only over cells that have an edge following the curve $Y(x)$. For this, at the creation of the mesh $\mathcal{T}(\Omega)$, we label cells according to their corresponding domain.

3.7.2. Sensitivities of the Jacobian matrix

To derive the sensitivities of the local FE matrices, we consider the blended formula (60). The parameterizations $X^{e_1}, \ldots, X^{e_4}$ are explicit functions of the design variables $\boldsymbol{\alpha}$, so X_K has an explicit form depending on ξ, η, and $\boldsymbol{\alpha}$. To compute the Jacobian matrix $\mathbf{J}$, we need

$$\frac{\partial X_K}{\partial \xi} = -\frac{1}{2} X^{e_1}(\eta) + \frac{1}{2} X^{e_2}(\eta) + \frac{1-\eta}{2} \frac{\partial X^{e_3}}{\partial \xi} + \frac{1+\eta}{2} \frac{\partial X^{e_4}}{\partial \xi}$$
$$+ \frac{1}{4} [(1-\eta)(\mathbf{x}_1 - \mathbf{x}_2) + (1+\eta)(\mathbf{x}_4 - \mathbf{x}_3)] \tag{68}$$

and

$$\frac{\partial X_K}{\partial \eta} = \frac{1-\xi}{2} \frac{\partial X^{e_1}}{\partial \eta} + \frac{1+\xi}{2} \frac{\partial X^{e_2}}{\partial \eta} - \frac{1}{2} X^{e_3}(\xi) + \frac{1}{2} X^{e_4}(\xi)$$
$$+ \frac{1}{4} [(1-\xi)(\mathbf{x}_1 - \mathbf{x}_4) + (1+\xi)(\mathbf{x}_2 - \mathbf{x}_3)]. \tag{69}$$

In this work, we exploit that the transfinite interpolation formula (60) is explicit. This allows for the straightforward differentiation with respect to the design variables once the parameterizations $X^{e_1}(\boldsymbol{\alpha}), \ldots, X^{e_4}(\boldsymbol{\alpha})$ are known. For convenience, Fig. 2 illustrates the situation where an element K with vertices $\mathbf{x}_1, \ldots, \mathbf{x}_4$ is bent following a shape represented by X^e. In this case, the vertices $\mathbf{x}_1 = (x_1, y_1)^\top$ and $\mathbf{x}_2 = (x_2, y_2)^\top$ are constrained to follow the shape described by $\boldsymbol{\alpha}$. As described in Section 3.7.1, we obtain exact sensitivities for the FE matrices and vectors by computing

$$\frac{\partial \mathbf{J}}{\partial \alpha_m} \text{ and } \frac{\partial |\mathbf{J}|}{\partial \alpha_m}, \text{ which require } \frac{\partial}{\partial \alpha_m} \left(\frac{\partial X_K}{\partial \xi} \right) \text{ and } \frac{\partial}{\partial \alpha_m} \left(\frac{\partial X_K}{\partial \eta} \right). \tag{70}$$

The sensitivities (70) of the Jacobian matrix and its determinant (61), require the assembly of each term in expressions (68) and (69).

In our implementation, we have used the horizontal coordinate x as the global parameter describing our curved interfaces $Y(x)$. For the computation of the required derivatives (70), we write the edge function

$$X^e = \begin{bmatrix} x \\ Y(x) \end{bmatrix}, \text{ hence } \frac{\partial X^e}{\partial x} = \begin{bmatrix} 1 \\ \partial Y / \partial x \end{bmatrix}. \tag{71}$$

We introduce the parameterization for the horizontal coordinate as $x(\zeta) := \frac{1}{2}(1+\zeta)x_2 + \frac{1}{2}(1-\zeta)x_1$ and mention that in the current setting x_1 and x_2 do not vary with α_m. Notice that since we assume an arbitrary orientation of the FE cell K, then $\zeta \in [-1, 1]$ may become ξ or η on the edge following the curved shape. The following expressions hold

$$\frac{\partial X^e}{\partial \zeta} = \frac{\partial X^e}{\partial x} \frac{\partial x}{\partial \zeta} \quad \text{and} \quad \frac{\partial}{\partial \alpha_m} \left(\frac{\partial X^e}{\partial \zeta} \right) = \frac{\partial}{\partial \alpha_m} \left(\frac{\partial X^e}{\partial x} \right) \frac{\partial x}{\partial \zeta}. \tag{72}$$

Direct differentiation of representation (37) results in explicit expressions for

$$\frac{\partial Y}{\partial x}, \quad \frac{\partial Y}{\partial \alpha_m}, \quad \text{and} \quad \frac{\partial}{\partial \alpha_m}\left(\frac{\partial Y}{\partial x}\right). \tag{73}$$

In this section, we have derived all the necessary information required in the formulas for the sensitivities (70) of the Jacobian matrix (61), which in turn, give formulas for the explicit computation of the sensitivities of the local system matrix and local load vector as described in Section 3.7.1. The reader is referred to our previous work, Araújo et al. [63], for further implementation details.

3.8. Numerical optimization strategy (BFGS)

The chosen numerical optimization strategy, which we describe below, is a variant of the so-called BFGS method [64], where a correction is added when an update leads to an unfeasible design.

Assume that we are at iteration n and let $f_n := f(\boldsymbol{\alpha}_n)$ be the corresponding evaluation of the objective function, and $\mathbf{H}_n$ be a (low rank) positive definite approximation to the inverse of the Hessian matrix $\nabla^2 f(\boldsymbol{\alpha}_n)$. The proposed numerical optimization strategy consists in performing the following steps

$$\mathbf{p}_n = -\mathbf{H}_n \nabla f_n, \quad \boldsymbol{\alpha}_{n+1} = \boldsymbol{\alpha}_n + \mu_n \mathbf{p}_n, \tag{74}$$

where μ_n is the step length in the search direction $\mathbf{p}_n$. The inverse Hessian approximation is updated as

$$\mathbf{H}_{n+1} = \mathbf{H}_n + \left(1 + \frac{\mathbf{y}_n^\top \mathbf{H}_n \mathbf{y}_n}{\mathbf{s}_n^\top \mathbf{y}_n}\right) \frac{\mathbf{s}_n \mathbf{s}_n^\top}{\mathbf{s}_n^\top \mathbf{y}_n} - \left(\frac{\mathbf{s}_n \mathbf{y}_n^\top \mathbf{H}_n + \mathbf{H}_n \mathbf{y}_n \mathbf{s}_n^\top}{\mathbf{s}_n^\top \mathbf{y}_n}\right) \tag{75}$$

with $\mathbf{s}_n = \boldsymbol{\alpha}_{n+1} - \boldsymbol{\alpha}_n$, $\mathbf{y}_n = \nabla f_{n+1} - \nabla f_n$, and with $\mathbf{H}_0 = \mathbf{I}$ the identity matrix.

To determine the step length μ_n, we apply an inexact line search using backtracking. So we require μ_n to satisfy the Armijo condition

$$f(\boldsymbol{\alpha}_n + \mu_n \mathbf{p}_n) \leq f(\boldsymbol{\alpha}_n) + c_1 \mu_n \mathbf{p}_n^\top \nabla f_n \text{ for } c_1 > 0. \tag{76}$$

The standard approach is to use as initial guess $\mu_n = 1$ and verify the condition (76). However, in our case this may lead to unfeasible shapes. So, if $y_1 < Y(\boldsymbol{\alpha}_n + \mathbf{p}_n, x) < y_2 \,\forall x$ we use $\mu_n = 1$ as initial guess, else we set the initial μ_n to the largest constant $c \in (0, 1]$ such that $y_1 \leq Y(\boldsymbol{\alpha}_n + c\mathbf{p}_n, x) \leq y_2$. In our setting the feasibility check is computationally inexpensive compared to solving the governing equation (52). For the Armijo condition (76) we use the recommended [64] value $c_1 = 10^{-4}$.

We employ a so-called continuation approach for the penalization parameters τ and γ. That is, instead of solving a single problem we solve for designs in a sequence of problems that approach in each level the original problem. The starting penalization is chosen as $\tau_0 = 1$, $\gamma_0 = 1$, where a design is computed from a problem penalized with τ_j, γ_j and it is used as an initial guess for a new problem with τ_{j+1}, γ_{j+1}.

The penalization is updated as

$$\tau_{j+1} = \tau_j/2, \ \gamma_{j+1} = \gamma_j/2 \quad \text{when} \quad \|\nabla_{\boldsymbol{\alpha}} f\| < \text{TOL}_\tau. \tag{77}$$

The updating parameter is chosen as $\text{TOL}_\tau = 10^{-2}$, for levels satisfying $\tau_j \geq 10^{-2}$ and for $\tau_j < 10^{-2}$ we reach the final level and set $\text{TOL}_\tau = 10^{-5}$.

3.9. FE approximation properties and error estimation

In this section, we give an a-priori estimation for FE approximation properties and for the quality of the polynomial shape representation of our implementation. These estimations give important guidance on how to construct the computational setting in order to achieve reliable results. The estimates used in this work do not account for a thorough FE pre-asymptotic error estimation for Helmholtz problems [65]. However, they provide good guidance on how to set up our FE discretization.

Namely, convergence of the FE error for Helmholtz problems deteriorates with increasing wave number k, and that the use of high polynomial order is advantageous to improve convergence rates [53,65–67]. Additionally, for uniform meshes, with largest mesh size h, and polynomial order p, it has been shown that under sufficient regularity assumptions the conditions $p \approx \log(k)$ and $kh/p \leq 1$ ensure quasi-optimality of the Galerkin method [65]. The derivation of reliable error estimates can be done thoroughly for one dimensional Helmholtz problems [67], but this is a difficult task for higher dimensions. Nonetheless, it is still possible to draw a sufficiently accurate picture of the behavior of the FE approximation for the two dimensional problem that will suffice for the purpose of the current work.

The estimation presented here is motivated by the results presented Araújo et al. [68] for resonance problems, where we estimate the requirements for quasi-optimality of our FE discretization individually on each element in the mesh. The refractive index is a piecewise constant function that has its maximum magnitude in the elements defining Ω_s. The spatial frequency of oscillation of solutions is quantified by the wavenumber $k = |n_s|\omega$. The dispersive error for a polynomial
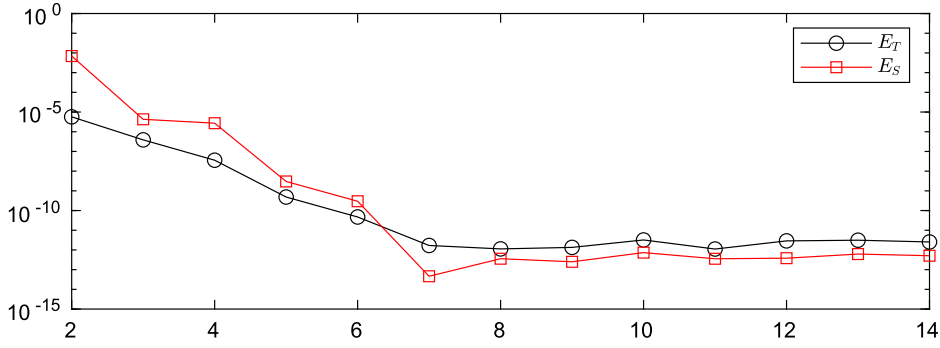
Fig. 3. Convergence of $E_T(p)$ and $E_S(p)$, defined in (80), with respect to polynomial degree p. For the study we use a shape with $N = 20$, we fix $n_s = 3$ and perform the computations for frequency $\omega = 3$.

approximation is expected to be worse in elements with high wavenumber. To have sufficiently small dispersive error in the FE solution, we require that

$$\left(\frac{\omega |n_s| h}{\sigma p}\right)^p < 1, \quad \text{for a given constant } \sigma > 0. \tag{78}$$

From this, we obtain the criterion $p > (\omega |n_s| h)/\sigma$.

For the error estimation for the shape representation, consider a mesh with fixed h. Then, as the number of Fourier terms N increases, it is expected that a higher polynomial degree is needed to obtain an accurate FE solution. Each component of our Fourier representation (37) satisfies a one dimensional Helmholtz equation with periodic boundary conditions. We define a polynomial approximation Y_h and let $h = d/N_e$, with N_e the number of FE edges that define the curved interface of $\partial\Omega_s$. For the highest frequency $j = N$, we estimate the FE requirements for the decay of the error $\|Y - Y_h\|$ in the pre-asymptotic regime. We use standard pre-asymptotic FE error estimation [53, Sec. 4.7.6] corresponding to one dimensional wave problems and obtain

$$\left(\frac{Nh}{2p}\right)^p < 1, \tag{79}$$

which tells us that the FE pollution effect is small provided that $N < 2p/h$ is satisfied. The FE mesh used for this project features $N_e = 10$, from where estimate (79) gives $N < 20p/d$.

4. Results

In this section, we describe the results from numerical computations performed with our implementation developed by using the FE library `deal.II` [69]. In all test cases, $\omega = 3$ is used as the incoming plane wave frequency and $n_s = 3$ as the refractive index of the slab. The box constraints have the fixed values $y_1 = 0.4$ and $y_2 = 0.6$, so $y_m = 0.5$. Unless otherwise stated, we employ $N = 20$ for the Fourier representation (37), and the initial design is set to is set to $\alpha_j = 0$ for $j = 1, 2, \ldots, N$.

Convergence of the discretization error: To test the quality of our discretization, we perform a convergence study on the polynomial order used for the FE. We set $N = 20$ and $d = 4$ and create a random feasible design. Let the residuals arising from the total and scattered fluxes introduced in Section 3.2 be defined as

$$E_T := |\Phi_T|, \quad \text{and} \quad E_S := |\Phi_S - b_0 d|. \tag{80}$$

As it is common in the literature [11,35], we assess the reliability of our computations by checking the convergence of the flux residuals as we increase the polynomial order of our FE discretization while keeping the mesh fixed. In Fig. 3, we see that $p = 10$ is a good choice since both the residuals E_T and E_S have converged to an absolute value below 10^{-10}.

Convergence of the optimization strategy: Fig. 4 shows a typical convergence of the proposed numerical optimization scheme. From left to right, we show plots for a) the resulting optimal designs (not at scale), and b) the iteration history for $f(\boldsymbol{\alpha})$ drawn with blue stars and $\log \|\nabla_{\boldsymbol{\alpha}} f\|$ drawn with red circles. The upper panels show results for TM, $\theta = 75°$, $d = 2$, and $N = 20$; the lower panels show results for TE, $\theta = 15°$, $d = 5$ and $N = 10$.

From the history plots, we observe the convergence of the numerical optimization strategy. Particularly, the echelon pattern in blue reveals the effect of the dynamic penalization from formula (40), which corresponds of the levels in the continuation approach.
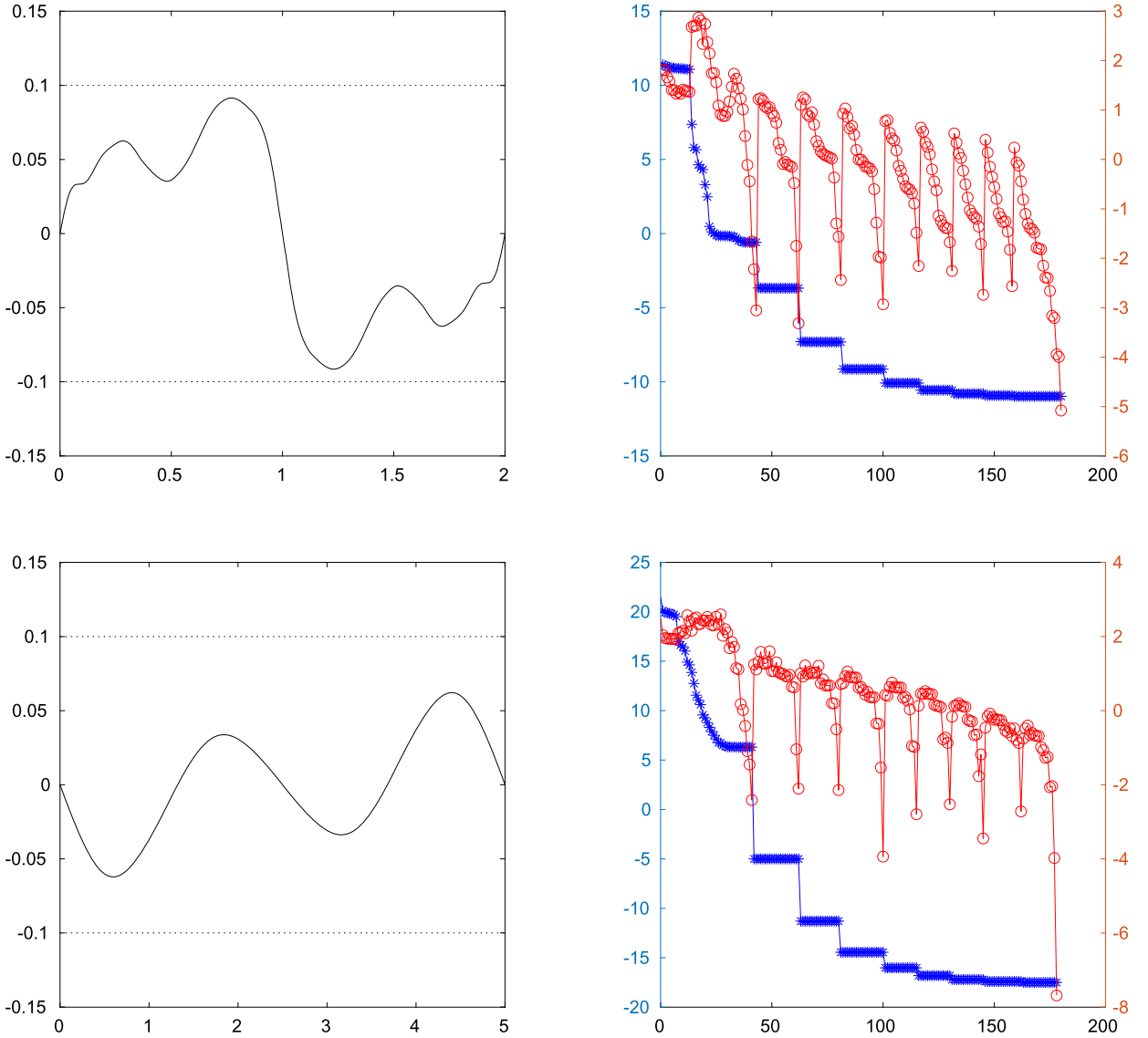
Fig. 4. Convergence of the numerical optimization scheme: we show results for $\omega = 3$ and $n_s = 3$. In the upper panels we show results for TM, $d = 2$, $\theta_c = 75°$ and $N = 20$, in the lower panels for TE, $d = 5$, $\theta_c = 15°$ and $N = 10$. From left to right we present a) centered optimal shapes $Y(x) - y_m$, and b) history of $f(\boldsymbol{\alpha})$, $\log_{10} \|\nabla_{\boldsymbol{\alpha}} f\|$ vs. iteration. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

We observe that the number of iterations corresponding to the initial level in the continuation ($k = 0$) and the following levels ($k > 0$) differ. The optimal design from the initial level is reached with more iterations than those optimal designs computed in the following continuation levels. This is explained by acknowledging that the flat slab is not a very good initial guess, and that the optimal design for penalty τ_k is a very good approximation for the numerical optimization problem with penalty τ_{k+1}.

Numerical computations indicate that the convergence of the gradient is very fast in each of the steps in the continuation approach. The norm of the gradient is decreasing in each continuation level and converges very quickly in around 20 to 50 iterations.

Finally, from the history of computations performed with $N = 10$ and $N = 20$, we obtain very similar plots for two different angles of incidence and different polarizations. This behavior illustrates the robustness of the numerical scheme.

Influence of the number of Fourier terms: In this section, we report numerical results by computing the numerical optimization for varying $N = 0, 1, 2 \ldots$ and using fixed values ω, θ, and d. We promote results consisting of shapes with low oscillatory behavior by using the penalization term F_p. This penalty is justified by the fact that simpler designs are easier to manufacture.

Case 1: TM, $d = 2,\ \theta = 75°$ Case 2: TE, $d = 5,\ \theta = 15°$
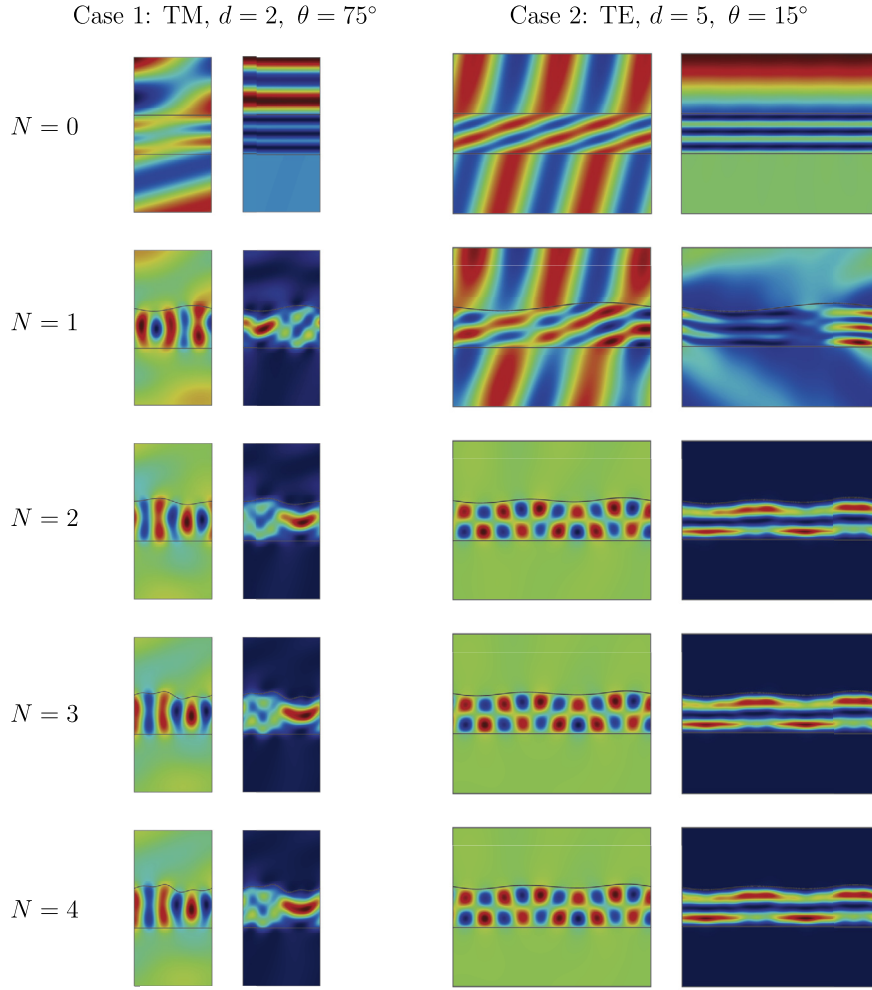
Fig. 5. Optimal designs resulting from computations with $N = 0, 1, 2, 3, 4$. In the left column we present Case 1: TM, $d = 2,\ \theta = 75°$, and in the right column Case 2: TE, $d = 5,\ \theta = 15°$. In each column we present plots for Re u followed by $|u|$.

In Fig. 5, we present two columns. On the left column, we present case 1 (C1) corresponding to TM, with $\theta = 75°$ and $d = 2$. On the right column, we present case 2 (C2) corresponding to TE, with $\theta = 15°$ and $d = 5$. In each case, we plot horizontally Re u and $|u|$. Vertically, we show for each case the computed optimal designs for increasing N, in the sequence $N = 0, 1, 2, 3$. Notice that $N = 0$ corresponds to a flat slab. The color scale for each plot in the sequence is not the same. The reason is that as N increases, the amplitude of waves inside the dielectric increases very rapidly, and fixing the maximum amplitude value would not allow us to see the patterns of the plots with low N.

In Fig. 6, we plot the resulting routing efficiency $Q_e(N)$ for a sequence of experiments with increasing number of Fourier terms N. For the study we use the examples: case 1 (C1) in black lines, and case 2 (C2) in red lines. For reference, we give below the parameterization for the shape optimized for TE with $\theta = 15°$, $\omega = 3$, $n_s = 3$, $d = 5$ and $N = 10$. The resulting parameterization $\boldsymbol{\alpha}$ from the representation (37) is

$$\alpha_1 = -0.013959, \quad \alpha_2 = -0.047569, \quad \alpha_3 = -0.006526, \quad \alpha_4 = -0.000176, \quad \alpha_5 = -0.000641,$$
$$\alpha_6 = 0.000152, \quad \alpha_7 = 0.000207, \quad \alpha_8 = -0.000069, \quad \alpha_9 = -0.000053, \quad \alpha_{10} = -0.000234.$$

We observe from the sequences in Fig. 5, that already when $N = 2$ the optimized designs are very efficient. It is observed that optimal designs mainly exhibit propagation of waves along the slab and low amplitude of scattered waves. Moreover, we see from Fig. 6, that $Q_e(N)$ is almost converged for these cases at $N > 3$.

Numerical computations suggest that our optimization problem has many local minima. In several cases, shapes with pronounced high modulation frequency produce more efficient designs than simple shapes with lower modulation frequency. We experience that, for different N, we may get locally optimal designs that exhibit very different shapes. That is, for each N, there may exist different local minima. In Fig. 6, the results of experiments with a flat slab as initial guess are drawn in dotted lines. These results suggest that there are, at least two, different accumulation values for the routing efficiency from the optimized designs.
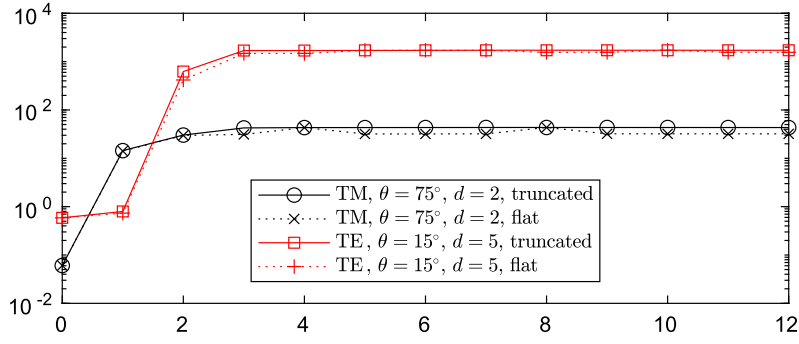
Fig. 6. Convergence of the routing efficiency $Q_e(N)$ with respect to the number of design variables N. For the study we use the examples: TM, $\theta = 75°$ and $d = 2$ in black lines, and TE, $\theta = 15°$ and $d = 5$ in red lines. In dotted lines we give the results of experiments with a flat slab as initial guess, and in solid lines the results from taking a truncation of an optimal design as initial guess.
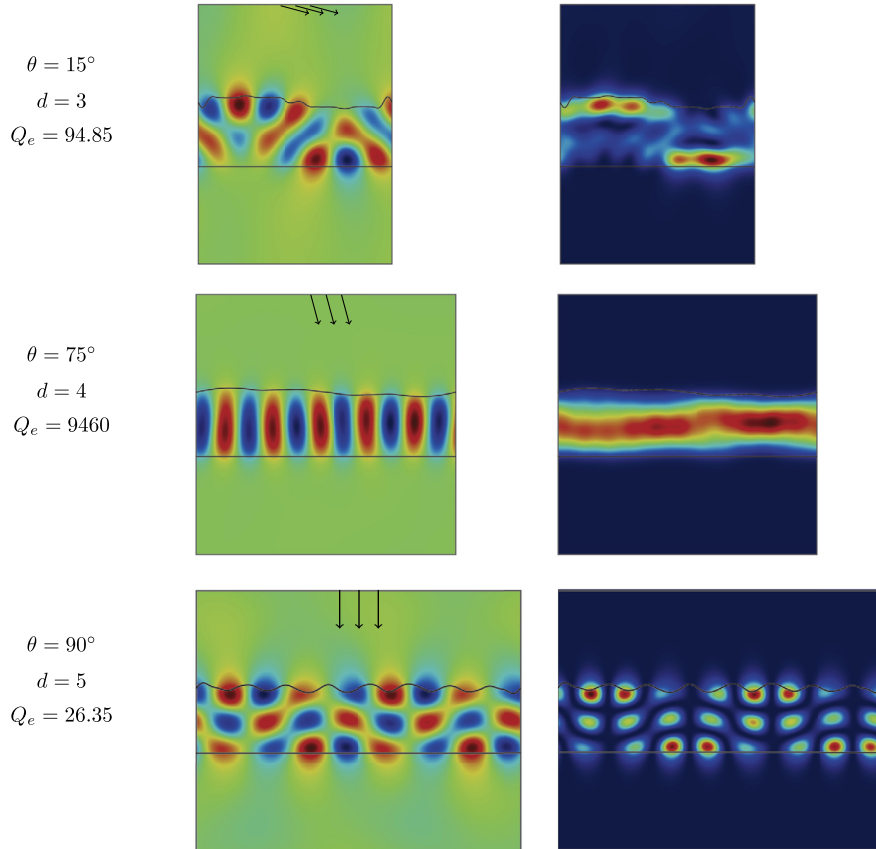
Fig. 7. Results for TM polarization.

For a large enough N^*, we use the Fourier coefficients of the design that exhibits larger routing efficiency ($N^* \approx 20$), and we denote it the high-fidelity design. We use its coefficients as initial guess for computing sequences of designs with different N. Then, we compute for $N = 1, 2, \ldots, 10 < N^*$ an optimal design using as initial guess the truncation to $N < N^*$ of the high-fidelity design. Since we assume that we are close to the optimal, we set $\tau_0 = \gamma_0 = 10^{-3}$. The result of this process is illustrated in Fig. 6, where we show in dotted lines optimal designs computed from taking as initial guess the flat slab. In solid lines we show optimal designs computed from taking as initial guess a truncation of an optimal design with high-fidelity. We observe a monotonic increase of $Q_e(N)$ in solid lines instead of the erratic pattern presented in dotted lines. This study allows us to trace a family of designs with different N once an efficient high-fidelity design is available.
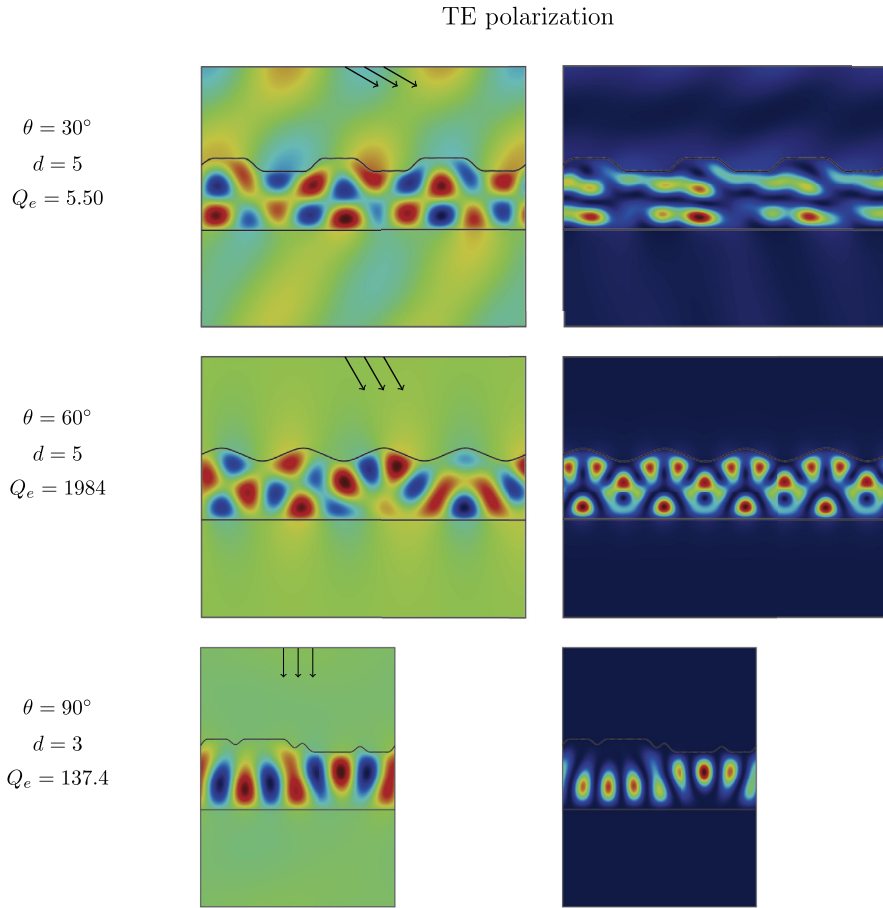
TE polarization

Fig. 8. Results for TE polarization.

Finally, we present results for various angles of incidence, periodicity lengths d and TM as well as TE polarizations. We gather results for TM polarization in Fig. 7, and for TE polarization in Fig. 8. In the figures, we horizontally present plots for $\mathrm{Re}\,u$ followed by $|u|$. For quick reference, we add on the left column information for the incoming angle of incidence θ, the periodicity length d and the corresponding routing efficiency Q_e. The direction of incidence of the incoming wave is marked by black arrows. Finally, the optimal shape corresponding to the parameterization of the model problem is drawn using a solid black line.

We observe from the figures that the resulting optimal design achieves very efficient routing of light. This is, the optimal shape promotes propagation through the slab and minimizes the scattering of waves propagating through Γ_1 and Γ_2.

We observe that some parameter-combinations yield the numerical optimization scheme to reach optimal designs that are extremely efficient at light routing. For example, the case for TM polarization, $\omega = 3$, $n_s = 3$, $d = 4$, and $\theta = 30°, 45°, 75°$. The examples suggest that in the neighborhood of the selected ω, there might be guided mode resonances for the model parameterization, polarization and optimal shape.

5. Conclusions

We have proposed a shape optimization strategy for the efficient routing of light in applications involving diffraction gratings. Moreover, the high order finite element method and the explicit differentiation scheme for the sensitivities lead to accurate computations at a low computational cost. Numerical experiments indicate that the proposed numerical optimization algorithm can produce slab designs with outstanding routing efficiencies for any combination of the parameters ω, θ, d, n_s, and polarization TM or TE.

CRediT authorship contribution statement

Juan C. Araújo C.: Conceptualization, Data curation, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Christian Engström:** Conceptualization, Funding acquisition, Investigation,

Methodology, Writing – review & editing. **Eddie Wadbro:** Conceptualization, Funding acquisition, Investigation, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

References

[1] R. Petit (Ed.), Electromagnetic Theory of Gratings, Topics in Current Physics, vol. 22, Springer-Verlag, Heidelberg, 1980.
[2] G. Bao, D.C. Dobson, J.A. Cox, Mathematical studies in rigorous grating theory, J. Opt. Soc. Am. A 12 (5) (1995) 1029–1042.
[3] C. Godreche, Solids Far from Equilibrium, C. Godreche (Eds.), Cambridge University Press, Cambridge, New York, 1991.
[4] J. Virieux, S. Operto, An overview of full-waveform inversion in exploration geophysics, Geophysics 74 (6) (2009) WCC1–WCC26.
[5] F. Bleibinhaus, S. Rondenay, Effects of surface scattering in full-waveform inversion, Geophysics 74 (6) (2009) WCC69–WCC77.
[6] F. Natterer, F. Wubbeling, Mathematical Methods in Image Reconstruction, Society for Industrial and Applied Mathematics, USA, 2001.
[7] M.B. Samaneh, H. Peters, N. Kessissoglou, S. Marburg, Three-dimensional analysis of a noise barrier using a quasi-periodic boundary element method, J. Acoust. Soc. Am. 137 (6) (2015) 3107–3114.
[8] N.P. Sergeant, M. Agrawal, P. Peumans, High performance solar-selective absorbers using coated sub-wavelength gratings, Opt. Express 18 (6) (2010) 5525–5540.
[9] J. Lai, L. Greengard, M. O'Neil, Robust integral formulations for electromagnetic scattering from three-dimensional cavities, J. Comput. Phys. 345 (2017) 1–16.
[10] D-L. Nguyen, A volume integral equation method for periodic scattering problems for anisotropic Maxwell's equations, Appl. Numer. Math. 98 (2015) 59–78.
[11] O.P. Bruno, A.G. Fernandez-Lado, On the evaluation of quasi-periodic Green functions and wave-scattering at and around Rayleigh-Wood anomalies, J. Comput. Phys. 410 (2020) 109352.
[12] H. Raether, Surface Plasmons on Smooth and Rough Surfaces and on Gratings, vol. 111, 1988.
[13] Stefan Alexander Maier, Plasmonics: Fundamentals and Applications, Springer Science, 2007.
[14] P. Vincent, M. Nevière, Corrugated dielectric waveguides: a numerical study of the second-order stop bands, Appl. Phys. 20 (4) (1979) 345–351.
[15] L. Mashev, E. Popov, Zero order anomaly of dielectric coated gratings, Opt. Commun. 55 (6) (1985) 377–380.
[16] I.A. Avrutsky, V.A. Sychugov, Reflection of a beam of finite size from a corrugated waveguide, J. Mod. Opt. 36 (11) (1989) 1527–1539.
[17] S.S. Wang, R. Magnusson, Theory and applications of guided-mode resonance filters, Appl. Opt. 32 (14) (1993) 2606–2613.
[18] G. Quaranta, G. Basset, O. Martin, B. Gallinet, Recent advances in resonant waveguide gratings, Laser Photonics Rev. 12 (2018) 1800017.
[19] Y-K. Tu, M-Z. Tsai, I-C. Lee, H-Y. Hsu, C-S. Huang, Integration of a guided-mode resonance filter with microposts for in-cell protein detection, Analyst 141 (2016) 4189–4195.
[20] T. Khaleque, R. Magnusson, Light management through guided-mode resonances in thin-film silicon solar cells, J. Nanophotonics 8 (1) (2014) 1–14.
[21] J. Zhang, K.F. MacDonald, N.I. Zheludev, Controlling light-with-light without nonlinearity, Light Sci. Appl. 1 (7) (2012) e18.
[22] W. Wan, Y. Chong, L. Ge, H. Noh, A.D. Stone, H. Cao, Time-reversed lasing and interferometric control of absorption, Science 331 (6019) (2011) 889–892.
[23] J.W. Yoon, G.M. Koh, S.H. Song, R. Magnusson, Measurement and modeling of a complete optical absorption and scattering by coherent surface plasmon-polariton excitation using a silver thin-film grating, Phys. Rev. Lett. 109 (2012) 257402.
[24] Z. Yu, A. Raman, S. Fan, Fundamental limit of nanophotonic light trapping in solar cells, Proc. Natl. Acad. Sci. 107 (41) (2010) 17491–17496.
[25] J.A. Giese, J.W. Yoon, B.R. Wenner, J.W. Allen, M.S. Allen, R. Magnusson, Guided-mode resonant coherent light absorbers 39 (3) (2014) 486–488.
[26] Y. Ding, R. Magnusson, Resonant leaky-mode spectral-band engineering and device applications, Opt. Express 12 (23) (2004) 5661–5674.
[27] E. Yablonovitch, Statistical ray optics, J. Opt. Soc. Am. 72 (7) (1982) 899–907.
[28] R.A. Pala, J.S.Q. Liu, E.S. Barnard, D. Askarov, E.C. Garnett, S. Fan, M.L. Brongersma, Optimization of non-periodic plasmonic light-trapping layers for thin-film solar cells, Nat. Commun. 4 (1) (2013) 2095.
[29] M. Agrawal, P. Peumans, Broadband optical absorption enhancement through coherent light trapping in thin-film photovoltaic cells, Opt. Express 16 (8) (2008) 5385–5396.
[30] P. Bermel, C. Luo, L. Zeng, L.C. Kimerling, J.D. Joannopoulos, Improving thin-film crystalline silicon solar cell efficiencies with photonic crystals, Opt. Express 15 (25) (2007) 16986–17000.
[31] C. Alongkarn, J. Sajeev, Light trapping and absorption optimization in certain thin-film photonic crystal architectures, Phys. Rev. A 78 (2008) 023825.
[32] R.A. Pala, J. White, E. Barnard, J. Liu, M.L. Brongersma, Design of plasmonic thin-film solar cells with broadband absorption enhancements, Adv. Mater. 21 (34) (2009) 3504–3509.
[33] A. Hessel, A.A. Oliner, A new theory of wood's anomalies on optical gratings, Appl. Opt. 4 (10) (1965) 1275–1297.
[34] B. Gallinet, J. Butet, O.J.F. Martin, Numerical methods for nanophotonics: standard problems and future challenges, Laser Photonics Rev. 9 (6) (2015) 577–603.
[35] A. Gillman, A. Barnett, A fast direct solver for quasi-periodic scattering problems, J. Comput. Phys. 248 (2013) 309–322.
[36] R. Aylwin, G. Silva-Oelker, C. Jerez-Hanckes, P. Fay, Optimization methods for achieving high diffraction efficiency with perfect electric conducting gratings, J. Opt. Soc. Am. A 37 (8) (2020) 1316–1326.

[37] Y. Zhang, A. Gillman, A fast direct solver for two dimensional quasi-periodic multilayered media scattering problems, BIT Numer. Math. 61 (1) (2021) 141–171.

[38] Z. Chen, H. Wu, An adaptive finite element method with perfectly matched absorbing layers for the wave scattering by periodic structures, SIAM J. Numer. Anal. 41 (3) (2004) 799–826.

[39] Z. Xia, K. Du, A tensor product finite element method for the diffraction grating problem with transparent boundary conditions, Comput. Math. Appl. 73 (4) (2017) 628–639.

[40] O.P. Bruno, S.P. Shipman, C. Turc, V. Stephanos, Three-dimensional quasi-periodic shifted Green function throughout the spectrum, including wood anomalies, Proc. R. Soc. A, Math. Phys. Eng. Sci. 473 (2207) (2017) 20170242.

[41] Y. Wu, Y.Y. Lu, Analyzing diffraction gratings by a boundary integral equation Neumann-to-Dirichlet map method, J. Opt. Soc. Am. A 26 (11) (2009) 2444–2451.

[42] W. Zhou, Haijun Wu, An adaptive finite element method for the diffraction grating problem with PML and few-mode DtN truncations, J. Sci. Comput. 76 (3) (2018) 1813–1838.

[43] A-S. Bonnet-Bendhia, F. Starling, Guided waves by electromagnetic gratings and non-uniqueness examples for the diffraction problem, Math. Methods Appl. Sci. 17 (5) (1994) 305–338.

[44] S. Fliss, D. Klindworth, K. Schmidt, Robin-to-Robin transparent boundary conditions for the computation of guided modes in photonic crystal waveguides, BIT Numer. Math. 55 (1) (2015) 81–115.

[45] M.G. Moharam, T.K. Gaylord, Rigorous coupled-wave analysis of planar-grating diffraction, J. Opt. Soc. Am. 71 (7) (1981) 811–818.

[46] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95 - International Conference on Neural Networks, vol. 4, 1995, pp. 1942–1948.

[47] M. Shokooh-Saremi, R. Magnusson, Particle swarm optimization and its application to the design of diffraction grating filters, Opt. Lett. 32 (8) (2007) 894–896.

[48] C. Forestiere, M. Donelli, G.F. Walsh, E. Zeni, G. Miano, L. Dal Negro, Particle-swarm optimization of broadband nanoplasmonic arrays, Opt. Lett. 35 (2) (2010) 133–135.

[49] D.L. Brundrett, E.N. Glytsis, T.K. Gaylord, Subwavelength transmission grating retarders for use at 10.6 μm, Appl. Opt. 35 (31) (1996) 6195–6202.

[50] Y. Zhou, B. Wang, Z. Guo, X. Wu, Guided mode resonance sensors with optimized figure of merit, Nanomaterials 9 (6) (2019).

[51] M. Cessenat, Mathematical Methods in Electromagnetism, Series on Advances in Mathematics for Applied Sciences, vol. 41, World Scientific Publisher, Singapore, 1996.

[52] J.D. Jackson, Classical Electrodynamics, third edition, John Wiley & Sons, New York, 1999.

[53] F. Ihlenburg, Finite Element Analysis of Acoustic Scattering. Applied Mathematical Sciences, Springer, New York, 1998.

[54] J.C. Araújo, C. Engström, E. Jarlebring, Efficient resonance computations for Helmholtz problems based on a Dirichlet-to-Neumann map, J. Comput. Appl. Math. 330 (2018) 177–192.

[55] L. Demkowicz, Asymptotic convergence in finite and boundary element methods: part 1: theoretical results, Comput. Math. Appl. 27 (12) (1994) 69–84.

[56] C. Schwab, P- and Hp-Finite Element Methods: Theory and Applications in Solid and Fluid Mechanics, Oxford University Press, 1998.

[57] I. Babuška, B.Q. Guo, The h, p and h-p version of the finite element method: basis theory and applications, Adv. Eng. Softw. 15 (3–4) (November 1992) 159–174.

[58] W.J. Gordon, C.A. Hall, Construction of curvilinear co-ordinate systems and applications to mesh generation, Int. J. Numer. Methods Eng. 7 (4) (1973) 461–477.

[59] W.J. Gordon, C.A. Hall, Transfinite element methods: blending-function interpolation over arbitrary curved element domains, Numer. Math. 21 (2) (Apr 1973) 109–129.

[60] P. Solin, K. Segeth, D. Ivo, Higher-Order Finite Element Methods. Studies in Advanced Mathematics, Chapman &Hall/CRC, Boca Raton, London, 2004.

[61] F. Zolla, G. Renversez, A. Nicolet, B. Kuhlmey, S. Guenneau, D. Felbacq, Foundations of Photonic Crystal Fibres, 2nd ed., Imperial College Press, 2012, Distributed by World Scientific London: Singapore, Hackensack, NJ.

[62] P.W. Christensen, A. Klarbring, An Introduction to Structural Optimization. Solid Mechanics and Its Applications, Springer, Netherlands, 2008.

[63] J.C. Araújo, E. Wadbro, Shape optimization for the strong directional scattering of dielectric nanorods, Int. J. Numer. Methods Eng. 122 (15) (2021) 3683–3704.

[64] J. Nocedal, S.J. Wright, Numerical Optimization, second edition, Springer, New York, NY, USA, 2006.

[65] J.M. Melenk, S. Sauter, Wave-number explicit convergence analysis for Galerkin discretizations of the Helmholtz equation, SIAM J. Numer. Anal. 49 (3/4) (2011) 1210–1243.

[66] L. Thompson, P.M. Pinsky, Complex wavenumber Fourier analysis of the p-version finite element method, Comput. Mech. 13 (4) (Jul 1994) 255–275.

[67] M. Ainsworth, Discrete dispersion relation for hp-version finite element approximation at high wave number, SIAM J. Numer. Anal. 42 (2) (2005) 553–575.

[68] J.C. Araújo, C. Campos, C. Engström, J.E. Roman, Computation of scattering resonances in absorptive and dispersive media with applications to metal-dielectric nano-structures, J. Comput. Phys. 407 (2020) 109220.

[69] G. Alzetta, D. Arndt, W. Bangerth, V. Boddu, B. Brands, D. Davydov, R. Gassmoeller, T. Heister, L. Heltai, K. Kormann, M. Kronbichler, M. Maier, J.-P. Pelteret, B. Turcksin, D. Wells, The `deal.II` library, version 9.0, J. Numer. Math. 26 (4) (2018) 173–183.