HYBRIDIZATION OF THE RIGOROUS COUPLED-WAVE APPROACH WITH TRANSFORMATION OPTICS FOR ELECTROMAGNETIC SCATTERING BY A SURFACE-RELIEF GRATING

B. J. CIVILETTI*, A. LAKHTAKIA†, AND P. B. MONK‡

Abstract. We hybridized the rigorous coupled-wave approach (RCWA) with transformation optics to develop a hybrid coordinate-transform method for solving the time-harmonic Maxwell equations in a 2D domain containing a surface-relief grating. In order to prove that this method converges for the p-polarization state, we studied several different but related scattering problems. The imposition of generalized non-trapping conditions allowed us to prove *a-priori* estimates for these problems. To do this, we proved a Rellich identity and used density arguments to extend the estimates to more general problems. These *a-priori* estimates were then used to analyze the hybrid method. We obtained convergence rates with respect to two different parameters, the first being a slice thickness indicative of spatial discretization in the depth dimension, the second being the number of terms retained in the Rayleigh–Bloch expansions of the electric and magnetic field phasors with respect to the other dimension. Testing with a numerical example revealed faster convergence than our analysis predicted. The hybrid method does not suffer from the Gibbs phenomenon seen with the standard RCWA.

Key words. electromagnetics, error analysis, grating, RCWA, transformation optics, variational method

AMS subject classifications.

1. Introduction. Electromagnetic scattering characteristics of periodic structures are widely researched in physics and engineering communities because of diverse applications including filters, beam splitters, and beam couplers [1, 2, 3, 4]. Design and optimization of these and other devices call for the solutions of scattering problems, thereby creating interest also in the applied mathematics community, especially to settle issues of uniqueness and existence of results for the relevant variational problems [5, 6, 7, 8].

Periodic electromagnetic structures for optical applications are holographic (i.e., volumetric) gratings [9, 10, 11], surface-relief gratings [2, 12], and combinations of both [13, 14]. Our interest here lies in surface-relief gratings that are commonly employed to redirect optical beams [15, 16, 17], for spectroscopic analysis [18, 19], and to enhance photonic absorption in solar cells [20, 21, 22]. A surface-relief grating is a periodically undulating interface of two different media. If the undulations are sufficiently shallow [23, 24], an analytical method due to Rayleigh [25, 26] suffices to predict the scattering characteristics. Semi-analytical methods such as the Rayleigh–Fourier method [27, 28, 29], the T-matrix method [30, 31, 32], the rigorous-coupled wave approach (RCWA) [33, 34, 35], the differential method [36, 37, 38] and perturbation methods [85] are used for moderately deep undulations. Purely numerical techniques such as the finite element method [39, 40, 41], the boundary element method [42, 43], and the finite-difference time-domain method [44, 45, 46], as well as integral-equation methods [47, 48] may be used for deep undulations.

The RCWA is a popular technique for surface-relief gratings, largely because it is mesh-free and its approach to the solution of the time-harmonic Maxwell equations is elegantly intuitive [12, 35, 49, 50]. At its core, the RCWA exploits Floquet theory [51, 52, 53] which shows that the solutions are, in general, quasi-periodic [5, 54]. This fact is used to express the field phasors as well as the periodic relative permittivity in the grating region (containing the undulations) as Fourier series [55], which are appropriately truncated so that only a finite number of terms in the Fourier series are retained. Then, the grating region is decomposed into thin slices and the grating is replaced by a stairstep approximation [12]. Due to this discretization of the spatial domain, the RCWA algorithm requires the solution of a second-order matrix ordinary differential equation in each slice, followed by the

*Department of Mathematical Sciences, University of Delaware, Newark, DE 19716 (bcivilet@udel.edu)

†Department of Engineering Science and Mechanics, Pennsylvania State University, University Park, PA 16802 (akhlesh@psu.edu)

‡Department of Mathematical Sciences, University of Delaware, Newark, DE 19716 (monk@udel.edu).

Funding: US National Science Foundation, Grant Nos. DMS-2011603 and DMS-2011996

enforcement of the appropriate transmission conditions to ensure continuity of the solution and its conormal derivative across the interslice boundaries. Thus, the RCWA approximates the solution in the entire domain.

The RCWA algorithm has been shown to converge with respect to both the number of terms in the truncated Fourier series and the slice thickness, roughly speaking, as long as the relative permittivity is monotonic in the direction perpendicular to the grating [56, 57]. The RCWA yields accurate results for s-polarized incident light, but generally converges more slowly for p-polarized incident light [35, 58, 59].

A drawback of RCWA is that the relative permittivity is replaced in the grating region with a piecewise smooth approximation. Representing such a function with Fourier series results in the Gibbs phenomenon near a discontinuity [60]. This limitation prevent the successful application of RCWA to gratings with deep undulations [50].

The purpose of this paper is to formulate and analyze a numerical method that combines the RCWA and transformation optics to mitigate the Gibbs phenomenon [61, 62, 63]. In order to avoid representing piecewise smooth functions as Fourier series, we first apply a coordinate transformation so that the periodically undulating interface in the grating region is mapped to a flat interface in the new coordinates. Once the solution in the mapped domain is found, the inverse transformation is used to map it back into the original spatial domain. Because we combine two methods, we term our scheme "hybrid". Our hybrid method is motivated by the differential method [36, 37] but we employ a different coordinate transformation and a different solution algorithm (RCWA). The differential method uses a coordinate transform that is not the identity in the two half-spaces above and below the domain, and therefore would require different radiation conditions than what we have considered here. We have chosen a different coordinate transform in order to obtain a method that is amenable to analysis. The differential method is useful in solving a multilayered problem with non-intersecting interfaces of which some interfaces are planar and the others are periodically undulating with the same period [64], and so is our hybrid method.

By using a coordinate transformation, we only have to consider an electromagnetic scattering problem with simple geometry, i.e., a flat interface of infinite extent. The downside is that the Helmholtz equation in the mapped domain has anisotropic coefficient functions, even if the problem in the original domain has isotropic coefficient functions. There exists Rellich theory for such problems [7] with diagonal matrix coefficient functions, but in our case there will be off-diagonal terms in those coefficient functions. Complicating matters even more, we analyze a problem where the anisotropic coefficient functions are only piecewise smooth and the data is not in L^2. In order to prove *a-priori* estimates for this problem, we first consider an easier problem where the coefficient functions are C^∞ and the data is L^2. We derive a Rellich identity for this problem and use it to prove an *a-priori* estimate. Using density arguments similar to those of Graham *et al.* [65], we extend the *a-priori* estimates to the full problem. A generalized Babuška–Brezzi condition [66, 67, 68] is shown to hold for all of our problems, and the existence and uniqueness of the variational solution follows.

This paper is organized as follows. The electromagnetic preliminaries are stated in Sec. 2. In Sec. 3, we define the scattering problem in the original spatial domain. We define the standard Dirichlet-to-Neumann map (DtN) [69, 70] and use it to derive a general form of our variational problem. In Sec. 4, we prove a Rellich identity for a simplified version of the variational problem. Then, assuming that certain non-trapping conditions [65] hold and the right hand side of the Helmholtz equation is in L^2, we show that this Rellich identity implies an *a-priori* estimate for our problem under the assumption that the off-diagonal terms in the matrix coefficient function of the Helmholtz equation are sufficiently small. We then extend this estimate to problems where the right hand side of the Helmholtz equation is only in an appropriate dual space. Section 5 extends these results using density arguments to more general coefficient functions. Using norm-equivalence, we show in Sec. 6 that the transformed problem has a unique solution. We define the discretized form of the transformed problem in Sec. 7 and show that there is a unique solution and determine the convergence rate with respect to slice thickness. We show convergence with respect to the number of retained Fourier terms and derive an order rate in

Sec. 8. Finally, in Sec. 9 we present a numerical example as a test of our convergence theory.

2. Preliminaries. We consider linear optics with an $\exp(-i\omega t)$ dependence on time t, where $i = \sqrt{-1}$ and ω is the angular frequency of light. The electric field phasor is given by $\boldsymbol{E} = E_1\mathbf{e}_1 + E_2\mathbf{e}_2 + E_3\mathbf{e}_3 = (E_1, E_2, E_3)$ and the magnetic field phasor is $\boldsymbol{H} = H_1\mathbf{e}_1 + H_2\mathbf{e}_2 + H_3\mathbf{e}_3 = (H_1, H_2, H_3)$, where the unit vectors $\boldsymbol{e}_1 = (1, 0, 0)$, $\boldsymbol{e}_2 = (0, 1, 0)$, and $\boldsymbol{e}_3 = (0, 0, 1)$. The relative permittivity matrix everywhere can be expressed as

$$\boldsymbol{\varepsilon} = \left(\begin{array}{cc|c} \varepsilon_{11} & \varepsilon_{21} & 0 \\ \varepsilon_{12} & \varepsilon_{22} & 0 \\ \hline 0 & 0 & \varepsilon_{33} \end{array}\right) \tag{2.1}$$

and the relative permeability matrix everywhere as

$$\boldsymbol{\mu} = \left(\begin{array}{cc|c} \mu_{11} & \mu_{21} & 0 \\ \mu_{12} & \mu_{22} & 0 \\ \hline 0 & 0 & \mu_{33} \end{array}\right), \tag{2.2}$$

all ten constitutive scalars in these two matrices being complex-valued functions of the spatial coordinates x_1 and x_2, but not of x_3. The geometry of the scattering problem is invariant along the x_3 axis.

When the electric field phasor and the magnetic field phasor are independent of x_3, then the time-harmonic Maxwell equations decompose into two sets of independent equations. The set of equations involving E_1, E_2, and H_3 refers to the p-polarization state, and the set involving E_3, H_1, and H_2 refers to the s-polarization state [12]. The latter polarization state does not pose any significant problem when the RCWA is implemented [49, 56], but the former does [35, 50, 57]. Therefore, in the following sections, we present the hybrid method only for the p-polarization state, for which the following three partial differential equations emerge from the time-harmonic Faraday and the Ampère–Maxwell equations:

$$\frac{\partial}{\partial x_1} E_2 - \frac{\partial}{\partial x_2} E_1 = i\omega\mu_0\mu_{33} H_3, \tag{2.3}$$

$$\frac{\partial}{\partial x_2} H_3 = -i\omega\varepsilon_0 \left(\varepsilon_{11} E_1 + \varepsilon_{12} E_2\right), \tag{2.4}$$

$$\frac{\partial}{\partial x_1} H_3 = i\omega\varepsilon_0 \left(\varepsilon_{21} E_1 + \varepsilon_{22} E_2\right), \tag{2.5}$$

where $\varepsilon_0 = 8.854 \times 10^{-12}$ F m^{-1} is the permittivity and $\mu_0 = 4\pi \times 10^{-7}$ H m^{-1} is the permeability of vacuum.

Our aim is to obtain a Helmholtz equation for H_3 from (2.3)–(2.5). In order to eliminate E_1, we first multiply (2.4) by ε_{21} and (2.5) by ε_{11} and then add the resulting equations to get

$$i\omega\varepsilon_0 E_2 = |\tilde{\boldsymbol{\varepsilon}}|^{-1} \left(\varepsilon_{21} \frac{\partial}{\partial x_2} H_3 + \varepsilon_{11} \frac{\partial}{\partial x_1} H_3\right), \tag{2.6}$$

where

$$\tilde{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{21} \\ \varepsilon_{12} & \varepsilon_{22} \end{pmatrix} \tag{2.7}$$

is a symmetric matrix and $|\mathbf{A}|$ denotes the determinant of a matrix $\mathbf{A}$. In order to eliminate E_2, we first multiply (2.4) by ε_{22} and (2.5) by ε_{12}, and then add the resulting equations to get

$$i\omega\varepsilon_0 E_1 = -\left|\tilde{\boldsymbol{\varepsilon}}\right|^{-1} \left(\varepsilon_{22} \frac{\partial}{\partial x_2} H_3 + \varepsilon_{12} \frac{\partial}{\partial x_1} H_3\right). \tag{2.8}$$

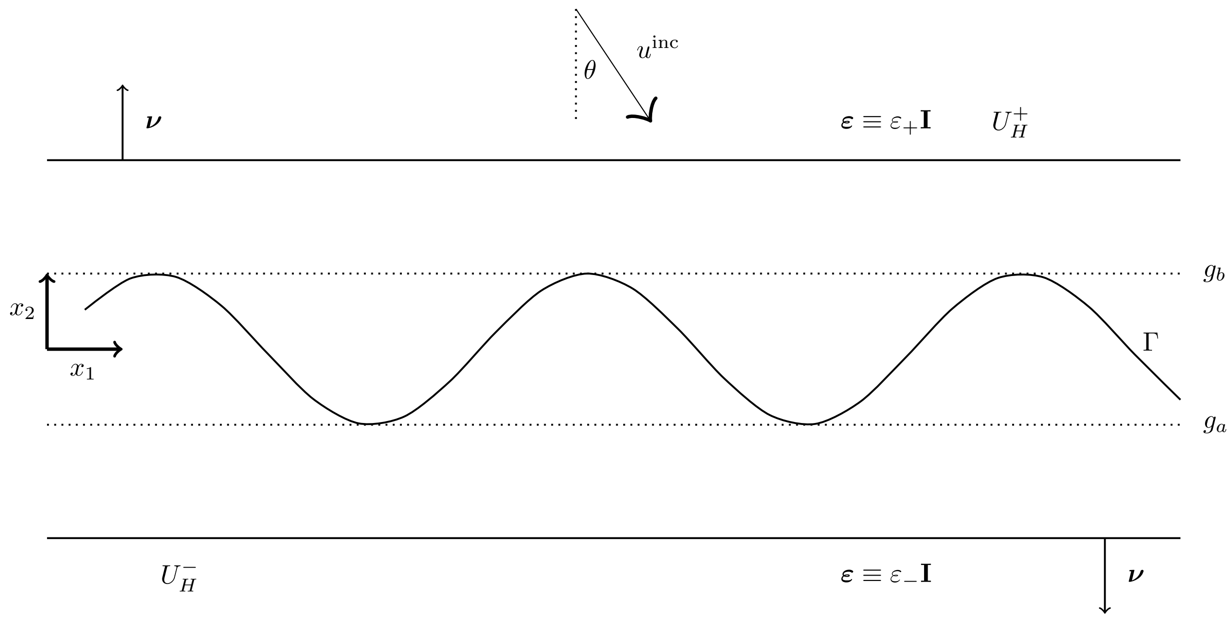


Fig. 3.1: Geometry of the scattering problem.

Substitution of (2.6) and (2.8) in (2.3) yields the Helmholtz equation

$$\nabla \cdot \left(|\tilde{\boldsymbol{\varepsilon}}|^{-1} \, \tilde{\boldsymbol{\varepsilon}}^{\top} \nabla H_3 \right) + \kappa^2 \mu_{33} H_3 = 0 \tag{2.9}$$

satisfied by H_3, where $\kappa = \omega\sqrt{\varepsilon_0 \mu_0}$ is the wavenumber in vacuum and the superscript $\top$ denotes the transpose. We call $|\tilde{\boldsymbol{\varepsilon}}|^{-1} \, \tilde{\boldsymbol{\varepsilon}}^{\top}$ and μ_{33} the coefficient functions of this Helmholtz equation. The hybrid method solves the scattering problem associated with (2.9). Let us note that only the following five constitutive scalars are relevant: ε_{11}, ε_{12}, ε_{21}, ε_{12}, and μ_{33}.

3. Scattering problem and its variational formulation. In this section, we define some notation related to the scattering problem and give the variational formulation to be analyzed in the later sections.

We define $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ in the original coordinate system. As shown in Fig. 3.1, the strip $\widetilde{\Omega} = (-\infty, \infty) \times (-H, H)$ for $H \in \mathbb{R}^+$ contains the grating region. The half-spaces above and below this strip are identified as $U_H^+ = \{\mathbf{x} \,|\, x_2 > H\}$ and $U_H^- = \{\mathbf{x} \,|\, x_2 < -H\}$, respectively. The grating surface is denoted by $\Gamma = \{\mathbf{x} \,|\, x_2 = g(x_1)\}$ for a periodic $g \in C^2(\mathbb{R})$ with period $\Lambda > 0$. Without loss of generality, $g(x_1) \geq 0$ for $x_1 \in \mathbb{R}$. The grating region strictly is the strip $\Omega^\dagger = (-\infty, \infty) \times [g_a, g_b]$, where $g_a = \min\{g(x_1)\}$ and $g_b = \max\{g(x_1)\}$. We choose H large enough so that $\Gamma \subset \Omega^\dagger \subset \widetilde{\Omega}$. The outward unit normal to $\widetilde{\Omega}$ is denoted by $\boldsymbol{\nu}$.

The homogeneous media occupying U_H^+ and U_H^- are isotropic, dielectric, and non-magnetic. Accordingly,

$$\boldsymbol{\varepsilon}(\mathbf{x}) = \begin{cases} \varepsilon_+ \mathbf{I}, & \mathbf{x} \in U_H^+, \\ \varepsilon_- \mathbf{I}, & \mathbf{x} \in U_H^-, \end{cases} \tag{3.1}$$

and $\boldsymbol{\mu}(\mathbf{x}) = \mathbf{I}$ for $\mathbf{x} \in U_H^+ \cup U_H^-$, where $\mathbf{I}$ is the identity matrix. We assume that both $\varepsilon_+ > 0$ and $\varepsilon_- > 0$ are real valued.

Incident on Γ is a downward propagating p-polarized plane wave

$$u^{\mathrm{inc}}(x_1,x_2) = -\exp\left\{i\varepsilon_+^{1/2}\kappa\left[x_1\sin\theta - (x_2-H)\cos\theta\right]\right\}, \tag{3.2}$$

where $\theta \in [0,\pi/2)$ is the incidence angle from the positive x_2-axis. The scattered field $u(\mathbf{x})$ for this scattering problem is quasi-periodic with period Λ [51, 53, 54]. This means that

$$u(x_1+\Lambda,x_2) = \exp(i\alpha_0\Lambda)u(x_1,x_2) \tag{3.3}$$

for $x \in \mathbb{R}$, where $\alpha_0 = \kappa\varepsilon_+^{1/2}\sin\theta$. The multiplicative factor $\Psi = \exp(i\alpha_0\Lambda)$ is called the *phase factor*; $\alpha_0 = 0$ and $\Psi = 1$ for normal incidence (i.e., $\theta = 0$) and in this case u is periodic with period Λ.

Due to u being quasi-periodic, the variational formulation can be written over a bounded region containing only a single period of $\widetilde{\Omega}$. We identify this bounded region as $\Omega = [-\Lambda/2,\Lambda/2]\times(-H,H)$ and its upper and lower boundaries as $\Gamma_{\pm H} = \{\mathbf{x}\,|\,-\Lambda/2 < x_1 < \Lambda/2, x_2 = \pm H\}$. The quasi-periodic boundaries are identified as $\Gamma_{\pm\Lambda/2} = \{\mathbf{x}\,|\,-H < x_2 < H, x_1 = \pm\Lambda/2\}$.

We assume that $\boldsymbol{\varepsilon}(\mathbf{x})$ is a piecewise $C^1(\mathbb{R}^{3\times 3})$ function that may have jumps across the interface Γ, and can be complex valued. Our analysis requires that

$$\lim_{\delta\to 0}\boldsymbol{\varepsilon}(x_1,H-\delta) = \varepsilon_+\mathbf{I} \tag{3.4}$$

for $\delta > 0$, so that $\boldsymbol{\varepsilon}$ does not jump across the top boundary $x_2 = H$, but may jump across the bottom boundary $x_2 = -H$. For the sake of notational simplicity, we assume that $\varepsilon_- = \varepsilon_+$, although it is possible to handle the case where $\varepsilon_- \neq \varepsilon_+$.

Given a source $G = \nabla\cdot\left[(\varepsilon_+^{-1}\mathbf{I} - \mathbf{A})\nabla u^{\mathrm{inc}}\right] + \kappa^2(1-a)u^{\mathrm{inc}}$ with coefficient functions $\mathbf{A} = |\tilde{\boldsymbol{\varepsilon}}|^{-1}\tilde{\boldsymbol{\varepsilon}}$ and $a = \mu_{33}$, the problem is to find a scattered field u such that

$$\nabla\cdot(\mathbf{A}\nabla u) + \kappa^2 a u = G, \qquad \mathbf{x}\in\Omega, \tag{3.5}$$

$$\Psi u(-\Lambda/2,x_2) - u(\Lambda/2,x_2) = 0, \qquad x_2 \in (-H,H), \tag{3.6}$$

$$\Psi\frac{\partial u}{\partial\nu_{\mathbf{A}}}(-\Lambda/2,x_2) - \frac{\partial u}{\partial\nu_{\mathbf{A}}}(\Lambda/2,x_2) = 0, \qquad x_2\in(-H,H), \tag{3.7}$$

together with a suitable radiation condition [71, 72] for u, where $\frac{\partial u}{\partial\nu_{\mathbf{A}}} = \boldsymbol{\nu}^\top\mathbf{A}\nabla u$ is the conormal derivative of u. This problem has been studied previously by us [57] for the special case where $\boldsymbol{\varepsilon} = \varepsilon\mathbf{I}$ and $\mathbf{A} = \varepsilon^{-1}\mathbf{I}$ in Ω by using a Rellich identity to show an *a-priori* estimate. A generalized Babuška–Brezzi condition holds since an *inf-sup* and transposed *inf-sup* condition [7] is satisfied. This is used to obtain a unique variational solution to the problem.

We now define some Sobolev spaces [73] in order to study our subsequent variational formulation. The Hilbert space $H^1_{\mathrm{qp}}(\Omega)$ is defined as the completion of $H^1(\Omega)\cap C^\infty_{\mathrm{qp}}(\Omega)$ in the standard H^1–norm given by

$$\|v\|_{1,\Omega} = \left(\int_\Omega |v|^2 + |\nabla v|^2\right)^{1/2}, \tag{3.8}$$

where $C^\infty_{\mathrm{qp}}(\Omega)$ is the set of quasi-periodic smooth functions in Ω. The space $H^1_{\mathrm{qp}}(\Omega)'$ is the dual space of $H^1_{\mathrm{qp}}(\Omega)$ endowed with the norm

$$\|F\|_{1*,\Omega} = \sup_{0\neq v\in H^1_{\mathrm{qp}}(\Omega)}\frac{|F(v)|}{\|v\|_{1,\Omega}}. \tag{3.9}$$

The trace space $H^k_{\mathrm{qp}}(\Gamma_{\pm H})$ with $k\in\mathbb{N}$ is endowed with the norm

$$\|v\|_{k,\Gamma_{\pm H}} = \left(\sum_{m\in\mathbb{Z}}\left|\kappa^2\varepsilon_+ - \alpha_m^2\right|^k\left|v_m^\pm\right|^2\right)^{1/2} \tag{3.10}$$

with $\alpha_m = \alpha_0 + 2\pi m/\Lambda$ and the Fourier coefficients defined as

$$v_m^{\pm} = \frac{1}{\Lambda}\int_{-\Lambda/2}^{\Lambda/2} v|_{\Gamma_{\pm H}} \exp(-i\alpha_m x_1)\, dx_1. \tag{3.11}$$

In order to state the appropriate radiation conditions and the variational formulation for (3.5)–(3.7), we now define the standard DtN maps on $\Gamma_{\pm H}$ [6, 7, 8]. Since we have assumed that $\boldsymbol{\varepsilon} \equiv \varepsilon_+ \mathbf{I}$ in $U_H^{\pm}$, the Helmholtz equation (3.5) simplifies to

$$\Delta u + \kappa^2 \varepsilon_+ u = 0 \tag{3.12}$$

in $U_H^{\pm}$. In order to avoid Rayleigh–Wood anomalies [2, 74], we assume that $\kappa\varepsilon_+^{1/2} \neq \alpha_m$ for any $m \in \mathbb{Z}$. For $\phi \in H_{\mathrm{qp}}^{1/2}(\Gamma_H)$, we consider $v_\phi \in H^1_{\mathrm{qp,loc}}(\Omega^+)$ satisfying

$$\Delta v_\phi + \kappa^2 \varepsilon_+ v_\phi = 0 \qquad \text{in } \Omega^+, \tag{3.13}$$

$$v_\phi = \phi \qquad \text{on } \Gamma_H, \tag{3.14}$$

with $\Omega^+ = [-\Lambda/2, \Lambda/2] \times (H, \infty)$. Then v_ϕ has the special form

$$v_\phi(\mathbf{x}) = \sum_{m\in\mathbb{Z}} \phi_m^+ \exp\left[i(x_2 - H)\beta_m\right] \exp(i\alpha_m x_1) + \sum_{m\in\mathbb{Z}} \phi_m^+ \exp\left[-i(x_2 - H)\beta_m\right] \exp(i\alpha_m x_1) \tag{3.15}$$

for $\mathbf{x} \in \Omega^+$, where

$$\beta_m = \begin{cases} \sqrt{\kappa^2\varepsilon_+ - \alpha_m^2}, & \alpha_m^2 < \kappa^2\varepsilon_+, \\ i\sqrt{\alpha_m^2 - \kappa^2\varepsilon_+}, & \alpha_m^2 > \kappa^2\varepsilon_+. \end{cases} \tag{3.16}$$

The representation (3.15) consists of two different types of solutions [8]. The first series on the right hand side of (3.15) comprises a finite number of upward propagating plane waves and an infinite number of evanescent waves that decay exponentially as $x_2 \to \infty$. The second series on the right hand side of (3.15) comprises a finite number of downward propagating plane waves and an infinite number of evanescent waves that decay as $x_2 \to -\infty$. The radiation condition [71, 72] enjoins us to choose the first solution type in Ω^+ and the second type in $\Omega^- = [-\Lambda/2, \Lambda/2] \times (-\infty, -H)$.

Using these expansions, we now define the standard DtN maps $T^{\pm} : H_{\mathrm{qp}}^{1/2}(\Gamma_{\pm H}) \to H_{\mathrm{qp}}^{-1/2}(\Gamma_{\pm H})$ as

$$(T^{\pm}\phi)(x_1) = i\varepsilon_+^{-1} \sum_{m\in\mathbb{Z}} \beta_m \phi_n^{\pm} \exp(i\alpha_m x_1) \tag{3.17}$$

for $\phi \in H_{\mathrm{qp}}^{1/2}(\Gamma_{\pm H})$. In the upcoming variational formulation we will replace the conormal derivatives $\frac{\partial w}{\partial \nu_{\mathbf{A}}} = \boldsymbol{\nu}^\top \mathbf{A} \nabla w$ with $\varepsilon_+ T^{\pm}(w)$, and the resulting sesquilinear form will be bounded on $H^1_{\mathrm{qp}}(\Omega) \times H^1_{\mathrm{qp}}(\Omega)$ due to this choice of DtN map. To conclude this discussion on DtN maps we recall that

$$\mathrm{Re} \int_{\Gamma_{\pm H}} \overline{\phi} T^{\pm}(\phi) \leq 0, \tag{3.18}$$

$$\mathrm{Im} \int_{\Gamma_{\pm H}} \overline{\phi} T^{\pm}(\phi) \geq 0, \tag{3.19}$$

for $\phi \in H_{\mathrm{qp}}^{1/2}(\Gamma_{\pm H})$ [7], where the overbar denotes the complex conjugate.

We denote the jump of a function w across an interface Γ as $[[w]]_\Gamma = w|_\Gamma^+ - w|_\Gamma^-$. The matrix coefficient function $\mathbf{A}$ may jump across the interface Γ; while the field w as well as its conormal derivative are assumed to be continuous across Γ so that $[[w]]_\Gamma = 0$ and $\left[\left[\frac{\partial w}{\partial \nu_{\mathbf{A}}}\right]\right]_\Gamma = 0$. Before the coordinate transformation is implemented, the general scattering problem we wish to solve is as follows: Given some data $F \in H^1_{\mathrm{qp}}(\Omega)'$, find a $w \in H^1_{\mathrm{qp}}(\Omega)$ such that

$$\nabla \cdot (\mathbf{A}\nabla w) + \kappa^2 a w = F, \qquad \mathbf{x} \in \Omega, \tag{3.20}$$

$$\Psi w(-\Lambda/2, x_2) - w(\Lambda/2, x_2) = 0, \qquad x_2 \in \mathbb{R}, \tag{3.21}$$

$$\Psi \frac{\partial w}{\partial \nu_{\mathbf{A}}}(-\Lambda/2, x_2) - \frac{\partial w}{\partial \nu_{\mathbf{A}}}(\Lambda/2, x_2) = 0, \qquad x_2 \in \mathbb{R}, \tag{3.22}$$

together with the radiation condition.

The variational formulation of (3.20)–(3.22) is to find a $w \in H^1_{\mathrm{qp}}(\Omega)$ such that

$$\int_\Omega \left[(\nabla v)^* \mathbf{A} \nabla w - \kappa^2 a w \overline{v}\right] - \int_{\Gamma_H} \overline{v} T^+(w) - \int_{\Gamma_{-H}} \overline{v} T^-(w) = -\int_\Omega F \overline{v} \tag{3.23}$$

for $v \in H^1_{\mathrm{qp}}(\Omega)$, which follows in the usual way from the divergence theorem. Here, $*$ denotes conjugate transpose of a vector. There are no boundary terms on the quasi-periodic boundaries $\Gamma_{\pm\Lambda/2}$ since $\overline{v}\frac{\partial w}{\partial \nu_{\mathbf{A}}}$ is periodic with period Λ for all $v \in H^1_{\mathrm{qp}}(\Omega)$.

We finish this section by defining $B(w, v; \mathbf{A}, a, \Omega)$ to be the sesquilinear form on the left hand side of (3.23), so that $B(\cdot, \cdot) : H^1_{\mathrm{qp}}(\Omega) \times H^1_{\mathrm{qp}}(\Omega) \to \mathbb{C}$. The variational problem is to find a $w \in H^1_{\mathrm{qp}}(\Omega)$ such that

$$B(w, v; \mathbf{A}, a, \Omega) = -(F, v)_{0,\Omega} \tag{3.24}$$

for $v \in H^1_{\mathrm{qp}}(\Omega)$, where $(\cdot, \cdot)_{0,\Omega}$ is the L^2 inner product on Ω.

4. A Rellich identity and *a-priori* estimates. Following Lechleiter and Ritterbusch [7], in this section we derive a Rellich identity. We assume that $\mathbf{A}(\mathbf{x})$ and $a(\mathbf{x}) > 0$ are real valued in Ω. Both $\mathbf{A}(\mathbf{x})$ and $a(\mathbf{x})$ are x_1-periodic (with period Λ) such that

$$\mathbf{A}|^-_{\Gamma_H} = \varepsilon_+^{-1}\mathbf{I} \qquad \text{and} \qquad a|^-_{\Gamma_H} = 1. \tag{4.1}$$

Given a $\delta > 0$, the matrix $\mathbf{A}$ is assumed to have the special form

$$\mathbf{A}(\mathbf{x}) = \begin{pmatrix} a_1^2(\mathbf{x}) & \delta a_2(\mathbf{x}) \\ \delta a_2(\mathbf{x}) & a_3^2(\mathbf{x}) \end{pmatrix}, \tag{4.2}$$

with the strictly positive scalar functions $a_1^2(\mathbf{x})$ and $a_3^2(\mathbf{x})$. Finally, we suppose there is constant $c_0 > 0$ such that $\boldsymbol{\xi}^* \mathbf{A}(\mathbf{x}) \boldsymbol{\xi} \geq c_0 |\boldsymbol{\xi}|^2$ for every $\boldsymbol{\xi} \in \mathbb{C}^2$ and $\mathbf{x} \in \Omega$. Our problem differs from that of Lechleiter and Ritterbusch [7] because:

(i) we allow for small off-diagonal terms in $\mathbf{A}$,
(ii) $a_2(\mathbf{x})$ can change sign in Ω, and
(iii) the Helmholtz equation (3.20) may have a non-constant coefficient function $a(\mathbf{x})$.

THEOREM 4.1. *Suppose that $\mathbf{A}$ is uniformly positive definite, $\mathbf{A} \in C^\infty(\mathbb{R}^{2\times 2}, \overline{\Omega})$ and $a \in C^\infty(\overline{\Omega})$ are real valued and periodic in x_1 with period Λ. Let $w \in H^1_{\mathrm{qp}}(\Omega)$ be a variational solution to (3.23) for $F \in L^2(\Omega)$. Then for $\delta > 0$,*

$$\begin{aligned} &\int_\Omega 2\left[a_3^2 \left|\frac{\partial w}{\partial x_2}\right|^2 + \delta a_2 \mathrm{Re}\left(\frac{\partial \overline{w}}{\partial x_1}\frac{\partial w}{\partial x_2}\right)\right] + \kappa^2 \int_\Omega (x_2 + H)\frac{\partial a}{\partial x_2}|w|^2 \\ &\quad + 2H \int_{\Gamma_H}\left[\varepsilon_+^{-1}\left(\left|\frac{\partial w}{\partial x_1}\right|^2 - \left|\frac{\partial w}{\partial x_2}\right|^2\right) - \kappa^2 |w|^2\right] - \int_{\Gamma_H} \overline{w} T^+(w) - \int_{\Gamma_{-H}} \overline{w} T^-(w) \\ &\quad = \int_\Omega (x_2 + H)(\nabla w)^* \frac{\partial \mathbf{A}}{\partial x_2}(\nabla w) - 2\int_\Omega (x_2 + H)\mathrm{Re}\left(\frac{\partial w}{\partial x_2}\overline{F}\right) - \int_\Omega F \overline{w}. \end{aligned} \tag{4.3}$$

Proof. Since w solves (3.23), we have that

$$\int_\Omega (x_2+H)\frac{\partial w}{\partial x_2}\nabla\cdot(\mathbf{A}\nabla\overline{w}) = \int_\Omega (x_2+H)\frac{\partial w}{\partial x_2}\left(\overline{F}-\kappa^2 a\overline{w}\right). \tag{4.4}$$

By taking twice the real part of both sides of (4.4) and using

$$2\mathrm{Re}\left(\frac{\partial w}{\partial x_2}\overline{w}\right) = \frac{\partial |w|^2}{\partial x_2}, \tag{4.5}$$

we get

$$\begin{aligned}2\mathrm{Re}\int_\Omega (x_2+H)\frac{\partial w}{\partial x_2}\nabla\cdot(\mathbf{A}\nabla\overline{w}) &= 2\int_\Omega (x_2+H)\mathrm{Re}\left(\frac{\partial w}{\partial x_2}\overline{F}\right) + \kappa^2\int_\Omega\left[(x_2+H)\frac{\partial a}{\partial x_2}+a\right]|w|^2 \\ &\quad - 2H\kappa^2\int_{\Gamma_H}|w|^2,\end{aligned} \tag{4.6}$$

after integrating by parts in x_2. By virtue of the divergence theorem, we have

$$\int_\Omega \frac{\partial w}{\partial x_2}\nabla\cdot(\mathbf{A}\nabla\overline{w}) = \int_{\partial\Omega}(x_2+H)\frac{\partial w}{\partial x_2}\frac{\partial\overline{w}}{\partial\nu_{\mathbf{A}}} - \int_\Omega \mathbf{A}\nabla\overline{w}\cdot\nabla\left((x_2+H)\frac{\partial w}{\partial x_2}\right). \tag{4.7}$$

We take twice the real part of both sides of (4.7) to obtain

$$\begin{aligned}2\mathrm{Re}\int_\Omega \frac{\partial w}{\partial x_2}\nabla\cdot(\mathbf{A}\nabla\overline{w}) &= 2H\int_{\Gamma_H} 2\mathrm{Re}\left(\frac{\partial w}{\partial x_2}\frac{\partial\overline{w}}{\partial\nu_{\mathbf{A}}}\right) - \int_\Omega 2\left[a_3^2\left|\frac{\partial w}{\partial x_2}\right|^2 + \delta a_2\mathrm{Re}\left(\frac{\partial\overline{w}}{\partial x_1}\frac{\partial w}{\partial x_2}\right)\right] \\ &\quad - \int_\Omega (x_2+H)2\mathrm{Re}\left[\mathbf{A}\nabla\overline{w}\cdot\nabla\left(\frac{\partial w}{\partial x_2}\right)\right].\end{aligned} \tag{4.8}$$

Since

$$2\mathrm{Re}\left[\mathbf{A}\nabla\overline{w}\cdot\nabla\left(\frac{\partial w}{\partial x_2}\right)\right] = \frac{\partial}{\partial x_2}\left[(\nabla w)^*\mathbf{A}\nabla w\right] - (\nabla w)^*\frac{\partial\mathbf{A}}{\partial x_2}\nabla w, \tag{4.9}$$

we get

$$\begin{aligned}\int_\Omega (x_2+H)2\mathrm{Re}\left[\mathbf{A}\nabla\overline{w}\cdot\nabla\left(\frac{\partial w}{\partial x_2}\right)\right] &= 2H\int_{\Gamma_H}(\nabla w)^*\mathbf{A}\nabla w - \int_\Omega (x_2+H)(\nabla w)^*\frac{\partial\mathbf{A}}{\partial x_2}\nabla w \\ &\quad - \int_\Omega (\nabla w)^*\mathbf{A}\nabla w\end{aligned} \tag{4.10}$$

after integrating by parts in x_2. We notice that

$$2\mathrm{Re}\left(\frac{\partial w}{\partial x_2}\frac{\partial\overline{w}}{\partial\nu_{\mathbf{A}}}\right) - (\nabla w)^*\mathbf{A}(\nabla w) = a_3^2\left|\frac{\partial w}{\partial x_2}\right|^2 - a_1^2\left|\frac{\partial w}{\partial x_1}\right|^2. \tag{4.11}$$

We now substitute (4.10) in (4.8) and use (4.11) in the resulting equation. Upon setting $v = w$ in the variational formulation (3.23), we see that

$$\int_\Omega (\nabla w)^*\mathbf{A}(\nabla w) = \kappa^2\int_\Omega a|w|^2 + \int_{\Gamma_H}\overline{w}T^+(w) + \int_{\Gamma_{-H}}\overline{w}T^-(w) - \int_\Omega F\overline{w}, \tag{4.12}$$

which we use on the right hand side of (4.10). Equating the resulting integral equation with (4.6) and rearranging some terms completes the proof. ∎

We now use the Rellich identity to show an *a-priori* estimate for a solution w of the variational problem (3.23). To do this, non-trapping conditions [65] must hold for $\mathbf{A}(\mathbf{x})$ and $a(\mathbf{x})$ wherein their x_2-derivatives have sign conditions (see (4.13)) in Ω. The existence and uniqueness of $w \in H^1_{\mathrm{qp}}(\Omega)$ follows because the *a-priori* estimates imply an *inf-sup* condition for the sesquilinear form $B(\cdot,\cdot)$ on $H^1_{\mathrm{qp}}(\Omega) \times H^1_{\mathrm{qp}}(\Omega)$.

THEOREM 4.2. *In addition to the assumptions of Theorem 4.1, we assume that the non-trapping conditions*

$$\frac{\partial a(\mathbf{x})}{\partial x_2} \geq 0 \qquad \text{and} \qquad \boldsymbol{\xi}^* \frac{\partial \mathbf{A}(\mathbf{x})}{\partial x_2} \boldsymbol{\xi} \leq 0 \tag{4.13}$$

hold for every $\boldsymbol{\xi} \in \mathbb{C}^2$ and $\mathbf{x} \in \Omega$. Then for $\delta > 0$ small enough, the solution $w \in H^1_{\mathrm{qp}}(\Omega)$ to the variational problem (3.23) is unique and there is a constant $C > 0$ such that

$$\|w\|_{1,\Omega} \leq C \|F\|_{0,\Omega}, \tag{4.14}$$

where

$$C = 2\left[\frac{1}{c_0} + 2C_1\left(2H\kappa\varepsilon_+^{1/2} + 2H + 1\right)\right] \tag{4.15}$$

and

$$C_1 = 4H(H+1)\left\{\frac{\kappa^2 \sup_{\mathbf{x}\in\Omega}[a(\mathbf{x})]}{c_0} + 1\right\} \times \left\{\frac{1}{2\inf_{\mathbf{x}\in\Omega}[a_3^2(\mathbf{x})]} + \frac{\varepsilon_+}{\min_{m\in\mathbb{Z}}\left(\sqrt{|\alpha_m^2 - \kappa^2\varepsilon_+|}\right)}\right\}. \tag{4.16}$$

Proof. We take the real part of both sides of the Rellich identity (4.3). From the non-trapping conditions (4.13) and the Cauchy–Schwarz inequality [75], we have

$$\begin{aligned} &2\left\|a_3 \frac{\partial w}{\partial x_2}\right\|_{0,\Omega}^2 - \mathrm{Re}\int_{\Gamma_H} \overline{w} T^+(w) \\ &\leq 2H \int_{\Gamma_H}\left[\varepsilon_+^{-1}\left(\left|\frac{\partial w}{\partial x_2}\right|^2 - \left|\frac{\partial w}{\partial x_1}\right|^2\right) + \kappa^2 |w|^2\right] + 2\delta\left|\int_\Omega a_2 \mathrm{Re}\left(\frac{\partial \overline{w}}{\partial x_1}\frac{\partial w}{\partial x_2}\right)\right| \\ &\qquad + 4H\left\|\frac{\partial w}{\partial x_2}\right\|_{0,\Omega}\|F\|_{0,\Omega} + \|w\|_{0,\Omega}\|F\|_{0,\Omega}. \end{aligned} \tag{4.17}$$

We can control the integral term with the δ factor by recalling that $ab \leq \frac{1}{2}\left(a^2 + b^2\right)$ for all $a \geq 0$ and $b \geq 0$. It follows that

$$2\delta\left|\int_\Omega a_2 \mathrm{Re}\left(\frac{\partial \overline{w}}{\partial x_1}\frac{\partial w}{\partial x_2}\right)\right| \leq \delta \sup_{\mathbf{x}\in\Omega}\left(|a_2(\mathbf{x})|\right)\|\nabla w\|_{0,\Omega}^2. \tag{4.18}$$

We also have the bound [6]

$$2H\int_{\Gamma_H}\left[\varepsilon_+^{-1}\left(\left|\frac{\partial w}{\partial x_2}\right|^2 - \left|\frac{\partial w}{\partial x_1}\right|^2\right) + \kappa^2|w|^2\right] \leq 4H\kappa\varepsilon_+^{1/2}\|w\|_{0,\Omega}\|F\|_{0,\Omega}. \tag{4.19}$$

The use of (4.18) and (4.19) in (4.17) yields

$$\begin{aligned} &2\left\|a_3\frac{\partial w}{\partial x_2}\right\|_{0,\Omega}^2 - \mathrm{Re}\int_{\Gamma_H}\overline{w}T^+(w) \\ &\qquad \leq \delta\sup_{\mathbf{x}\in\Omega}\left(|a_2(\mathbf{x})|\right)\|\nabla w\|_{0,\Omega}^2 + \left(4H\kappa\varepsilon_+^{1/2} + 4H + 1\right)\|w\|_{1,\Omega}\|F\|_{0,\Omega}. \end{aligned} \tag{4.20}$$

On using Parseval's theorem, it follows from the definition of the DtN maps that

$$\operatorname{Im}\int_{\Gamma_H}\overline{w}T^+(w)-\operatorname{Re}\int_{\Gamma_H}\overline{w}T^+(w)\geq\frac{1}{\varepsilon_+}\min_{m\in\mathbb{Z}}\left(\sqrt{|\alpha_m^2-\kappa^2\varepsilon_+|}\right)\|w\|_{0,\Gamma_H}^2. \tag{4.21}$$

Then by taking the imaginary part of both sides of the Rellich identity and using the Cauchy–Schwarz inequality, we have

$$\operatorname{Im}\int_{\Gamma_H}\overline{w}T^+(w)\leq\|w\|_{0,\Omega}\|F\|_{0,\Omega}. \tag{4.22}$$

Combining the last inequality with (4.21) yields

$$-\operatorname{Re}\int_{\Gamma_H}\overline{w}T^+(w)\geq\frac{1}{\varepsilon_+}\min_{m\in\mathbb{Z}}\left(\sqrt{|\alpha_m^2-\kappa^2\varepsilon_+|}\right)\|w\|_{0,\Gamma_H}^2-\|w\|_{0,\Omega}\|F\|_{0,\Omega}. \tag{4.23}$$

We then combine the inequalities (4.20) and (4.23) to obtain

$$\begin{aligned}&2\left\|a_3\frac{\partial w}{\partial x_2}\right\|_{0,\Omega}^2+\frac{1}{\varepsilon_+}\min_{m\in\mathbb{Z}}\left(\sqrt{|\alpha_m^2-\kappa^2\varepsilon_+|}\right)\|w\|_{0,\Gamma_H}^2\\&\leq\delta\sup_{\mathbf{x}\in\Omega}\left(|a_2(\mathbf{x})|\right)\|\nabla w\|_{0,\Omega}^2+2\left(2H\kappa\varepsilon_+^{1/2}+2H+1\right)\|w\|_{1,\Omega}\|F\|_{0,\Omega}.\end{aligned} \tag{4.24}$$

In order to control $\|w\|_{0,\Omega}^2$, we use Lemma 4.3 of Lechleiter and Rittersbusch [7] that delivers

$$\|w\|_{0,\Omega}^2\leq 4H^2\left\|\frac{\partial w}{\partial x_2}\right\|_{0,\Omega}^2+4H\|w\|_{0,\Gamma_H}^2 \tag{4.25}$$

for every $w\in H^1(\Omega)$. Using (4.24) and (4.25) we have

$$\begin{aligned}\|w\|_{0,\Omega}^2&\leq 4H(H+1)\left\{\frac{1}{2\inf_{\mathbf{x}\in\Omega}[a_3^2(\mathbf{x})]}+\frac{\varepsilon_+}{\min_{m\in\mathbb{Z}}\left(\sqrt{|\alpha_m^2-\kappa^2\varepsilon_+|}\right)}\right\}\\&\qquad\times\left[2\left\|a_3\frac{\partial w}{\partial x_2}\right\|_{0,\Omega}^2+\frac{1}{\varepsilon_+}\min_{m\in\mathbb{Z}}\left(\sqrt{|\alpha_m^2-\kappa^2\varepsilon_+|}\right)\|w\|_{0,\Gamma_H}^2\right]\\&\leq 4H(H+1)\left\{\frac{1}{2\inf_{\mathbf{x}\in\Omega}[a_3^2(\mathbf{x})]}+\frac{\varepsilon_+}{\min_{m\in\mathbb{Z}}\left(\sqrt{|\alpha_m^2-\kappa^2\varepsilon_+|}\right)}\right\}\\&\qquad\times\left[\delta\sup_{\mathbf{x}\in\Omega}\left(|a_2(\mathbf{x})|\right)\|\nabla w\|_{0,\Omega}^2+2\left(2H\kappa\varepsilon_+^{1/2}+2H+1\right)\|w\|_{1,\Omega}\|F\|_{0,\Omega}\right].\end{aligned} \tag{4.26}$$

We set $v=w$ in the variational formulation (3.23) and take the real part of both sides. Since $\mathbf{A}$ is uniformly positive definite, there is a constant $c_0>0$ such that

$$\|w\|_{1,\Omega}^2\leq\left\{\frac{\kappa^2\sup_{\mathbf{x}\in\Omega}[a(\mathbf{x})]}{c_0}+1\right\}\|w\|_{0,\Omega}^2+\frac{1}{c_0}\|w\|_{0,\Omega}\|F\|_{0,\Omega}. \tag{4.27}$$

Now we use inequality (4.26) in the last inequality and divide both sides by $\|w\|_{1,\Omega}$. This yields

$$\left[1-\delta C_1\sup_{\mathbf{x}\in\Omega}\left(|a_2(\mathbf{x})|\right)\right]\|w\|_{1,\Omega}\leq\left[\frac{1}{c_0}+2C_1\left(2H\kappa\varepsilon_+^{1/2}+2H+1\right)\right]\|F\|_{0,\Omega}. \tag{4.28}$$

To finish the proof, we can choose $\delta>0$ so that

$$\delta\leq\frac{1}{2C_1\sup_{\mathbf{x}\in\Omega}\left(|a_2(\mathbf{x})|\right)}. \tag{4.29}$$

□

Now we extend the previous *a-priori* estimates to problems where the right hand side $F \in H^1_{\mathrm{qp}}(\Omega)'$, together with the non-trapping conditions (4.13). Instead of $F \in L^2(\Omega)$, we have uniqueness and existence results for the variational problem (3.23) even when the right hand side is a general bounded linear functional in the dual space $H^1_{\mathrm{qp}}(\Omega)'$.

COROLLARY 4.3. *Assume the conditions of Theorems 4.1 and 4.2 hold and $w \in H^1_{\mathrm{qp}}(\Omega)$ is a variational solution to (3.23) with $F \in H^1_{\mathrm{qp}}(\Omega)'$ on the right hand side. Then for $\delta > 0$ small enough, $w \in H^1_{\mathrm{qp}}(\Omega)$ is unique and there is a constant $C > 0$ such that*

$$\|w\|_{1,\Omega} \leq C \|F\|_{1*,\Omega}. \tag{4.30}$$

Proof. Define $B^+(q,v) = B(q,v;\mathbf{A},a,\Omega) + 2\kappa^2 \int_\Omega aq\overline{v}$ for $q, v \in H^1_{\mathrm{qp}}(\Omega)$. Since $\mathbf{A}$ is uniformly positive definite, we have

$$\left|B^+(v,v)\right| \geq \min\left\{c_0, \kappa^2 \inf_{\mathbf{x}\in\Omega}[a(\mathbf{x})]\right\} \|v\|^2_{1,\Omega} \tag{4.31}$$

for $v \in H^1_{\mathrm{qp}}(\Omega)$; therefore, the sesquilinear form $B^+(\cdot,\cdot)$ is coercive. By virtue of the Lax–Milgram lemma [73], there is a unique solution $w^+ \in H^1_{\mathrm{qp}}(\Omega)$ to

$$B^+(w^+, v) = -F(v) \tag{4.32}$$

for $v \in H^1_{\mathrm{qp}}(\Omega)$, and furthermore

$$\left\|w^+\right\|_{1,\Omega} \leq \min\left\{c_0, \kappa^2 \inf_{\mathbf{x}\in\Omega}[a(\mathbf{x})]\right\}^{-1} \|F\|_{1*,\Omega}. \tag{4.33}$$

With w^+ given, $2\kappa^2 a w^+ \in L^2(\Omega)$ and we have by virtue of Theorem 4.2 a unique solution $w_1 \in H^1_{\mathrm{qp}}(\Omega)$ to

$$B(w_1, v; \mathbf{A}, a, \Omega) = 2\kappa^2 \int_\Omega a w^+ \overline{v} \tag{4.34}$$

for $v \in H^1_{\mathrm{qp}}(\Omega)$ and $\delta < 0$ small enough. We therefore have that

$$B(w^+ + w_1, v; \mathbf{A}, a, \Omega) = -F(v) \tag{4.35}$$

for $v \in H^1_{\mathrm{qp}}(\Omega)$, and a constant $C > 0$ such that

$$\begin{aligned}\left\|w^+ + w_1\right\|_{1,\Omega} &\leq \min\left\{c_0, \kappa^2 \inf_{\mathbf{x}\in\Omega}[a(\mathbf{x})]\right\}^{-1} \|F\|_{1*,\Omega} + 2C\kappa^2 \sup_{\mathbf{x}\in\Omega}[a(\mathbf{x})] \left\|w^+\right\|_{0,\Omega} \\ &\leq \min\left\{c_0, \kappa^2 \inf_{\mathbf{x}\in\Omega}[a(\mathbf{x})]\right\}^{-1} \left\{1 + 2C\kappa^2 \sup_{\mathbf{x}\in\Omega}[a(\mathbf{x})]\right\} \|F\|_{1*,\Omega}.\end{aligned} \tag{4.36}$$

The proof follows from $w = w^+ + w_1$. ∎

5. Additional *a-priori* estimates. In this section we consider our variational problem (3.23) where $\mathbf{A} \in L^\infty(\mathbb{R}^{2\times 2}, \Omega)$ and $a \in L^\infty(\Omega)$ are almost everywhere periodic in x_1 with period Λ. In Sec. 4, we derived *a-priori* estimates for the case where the coefficient functions are smooth, but we extend those results to more general coefficient functions in this section.

We begin by defining the function $\psi \in C_0^\infty(\mathbb{R}^2)$ as

$$\psi(\mathbf{x}) = \begin{cases} C\exp\left(\frac{1}{|\mathbf{x}|-1}\right). & |\mathbf{x}| < 1, \\ 0, & |\mathbf{x}| \geq 1, \end{cases} \tag{5.1}$$

for $\mathbf{x} \in \mathbb{R}^2$ with $C > 0$ chosen so that $\int_{\mathbb{R}^2} \psi = 1$. Let $\psi_\zeta(\mathbf{x}) = \zeta^{-2}\psi(\mathbf{x}/\zeta)$ for $\zeta > 0$. We extend the coefficients $\mathbf{A}$ and a by periodicity to the domain $\mathcal{U} = \{\mathbf{x} \in \mathbb{R}^2 \,|\, -\Lambda < x_1 < \Lambda\}$. We can therefore choose $\zeta > 0$ small enough so that $\overline{\Omega} \subset \mathcal{U}_\zeta \subset\subset \mathcal{U}$ where $\mathcal{U}_\zeta = \{\mathbf{x} \in \mathcal{U} \,|\, \mathrm{dist}\,(x, \partial\mathcal{U}) > \zeta\}$.

Next, we define the sequences $\phi_\zeta^\chi \in C^\infty(\mathcal{U}_\zeta)$ as

$$\phi_\zeta^\chi(\mathbf{x}) = \chi \star \psi_\zeta = \int_{\mathbb{R}^2} \chi(\mathbf{x} - \mathbf{y})\psi_\zeta(\mathbf{y})\, d\mathbf{y}, \tag{5.2}$$

where $\star$ denotes convolution and $\chi \in \{a_1^2, a_2, a_3^2, a\}$. This discussion leads us to the following theorem.

THEOREM 5.1. *Suppose that $\mathbf{A} \in L^\infty(\mathbb{R}^{2\times 2}, \Omega)$ and $a \in L^\infty(\Omega)$ are real valued and almost everywhere periodic in x_1 with period Λ. Also, let $\mathbf{A}$ be uniformly positive definite and have the special form*

$$\mathbf{A}(\mathbf{x}) = \begin{pmatrix} a_1^2(\mathbf{x}) & \delta a_2(\mathbf{x}) \\ \delta a_2(\mathbf{x}) & a_3^2(\mathbf{x}) \end{pmatrix} \tag{5.3}$$

and that

$$\operatorname*{ess\,sup}_{\mathbf{x}\in\Omega} \left[\boldsymbol{\xi}^*\mathbf{A}\boldsymbol{\xi}(\mathbf{x} + \tau\mathbf{e}_2) - \boldsymbol{\xi}^*\mathbf{A}\boldsymbol{\xi}(\mathbf{x})\right] \le 0 \tag{5.4}$$

for $\boldsymbol{\xi} \in \mathbb{C}^2$ and $\tau \ge 0$, as well as that

$$\operatorname*{ess\,inf}_{\mathbf{x}\in\Omega} \left[a(\mathbf{x} + \tau\mathbf{e}_2) - a(\mathbf{x})\right] \ge 0 \tag{5.5}$$

for $\tau \ge 0$. Then, given $F \in H^1_{\mathrm{qp}}(\Omega)'$ and $\delta > 0$ small enough, there is a unique solution $w \in H^1_{\mathrm{qp}}(\Omega)$ to (3.23). Furthermore, there is a constant $C > 0$ such that

$$\|w\|_{1,\Omega} \le C\, \|F\|_{1*,\Omega}\, . \tag{5.6}$$

Proof. First, we show that the sequences $\phi_\zeta^\chi \in C^\infty(\overline{\Omega})$ satisfy the conditions of Theorem 4.2. We define the matrix $\mathbf{A}_\zeta \in C^\infty(\mathbb{R}^{2\times 2}, \overline{\Omega})$ as

$$\mathbf{A}_\zeta(\mathbf{x}) = \begin{pmatrix} \phi_\zeta^{a_1^2}(\mathbf{x}) & \delta\phi_\zeta^{a_2}(\mathbf{x}) \\ \delta\phi_\zeta^{a_2}(\mathbf{x}) & \phi_\zeta^{a_3^2}(\mathbf{x}) \end{pmatrix}. \tag{5.7}$$

By virtue of the standard properties of mollifiers [73], we have

$$\left\|\phi_\zeta^\chi - \chi\right\|_{0,\Omega} \to 0 \tag{5.8}$$

as $\zeta \to 0$ for $\chi \in \{a_1^2, a_2, a_3^2, a\}$. Let $\lambda > 0$ be an eigenvalue of $\mathbf{A}$ with eigenvector $v \in \mathbb{R}^2$. Then

$$\|\lambda v - \mathbf{A}_\zeta v\|_{0,\Omega} \le \sqrt{2}\, \|v\|_\infty\, \|\mathbf{A} - \mathbf{A}_\zeta\|_{0,\Omega} \to 0 \tag{5.9}$$

as $\zeta \to 0$. We can therefore choose $\zeta > 0$ small enough so that the eigenvalues of $\mathbf{A}_\zeta$ are strictly positive and $\mathbf{A}_\zeta$ is uniformly positive definite. We notice that

$$\phi_\zeta^\chi(x_1 + \Lambda, x_2) = \int_{\mathbb{R}^2} \chi(\mathbf{x} + \Lambda\mathbf{e}_1 - \mathbf{y})\psi_\zeta(\mathbf{y})\, d\mathbf{y} = \phi_\zeta^\chi(\mathbf{x}), \tag{5.10}$$

since χ is almost everywhere periodic in x_1 with period Λ. We have defined $\mathbf{A}_\zeta$ so that $\boldsymbol{\xi}^*\mathbf{A}_\zeta\boldsymbol{\xi} = (\boldsymbol{\xi}^*\mathbf{A}\boldsymbol{\xi}) \star \psi_\zeta$ for $\boldsymbol{\xi} \in \mathbb{C}^2$. It then follows that

$$\begin{aligned} \left(\boldsymbol{\xi}^*\mathbf{A}_\zeta\boldsymbol{\xi}\right)(\mathbf{x} + \tau\mathbf{e}_2) - \left(\boldsymbol{\xi}^*\mathbf{A}_\zeta\boldsymbol{\xi}\right)(\mathbf{x}) &= \int_{|\mathbf{y}|<\zeta} \left[\boldsymbol{\xi}^*\mathbf{A}\boldsymbol{\xi}(\mathbf{x} + \tau\mathbf{e}_2 - \mathbf{y}) - \boldsymbol{\xi}^*\mathbf{A}\boldsymbol{\xi}\right]\psi_\zeta(\mathbf{y})\, d\mathbf{y} \\ &\le \operatorname*{ess\,sup}_{\mathbf{x}\in\Omega} \left[\boldsymbol{\xi}^*\mathbf{A}\boldsymbol{\xi}(\mathbf{x} + \tau\mathbf{e}_2) - \boldsymbol{\xi}^*\mathbf{A}\boldsymbol{\xi}(\mathbf{x})\right] \int_{|\mathbf{y}|<\zeta} \psi_\zeta(\mathbf{y})\, d\mathbf{y} \\ &\le 0 \end{aligned} \tag{5.11}$$

for $\boldsymbol{\xi} \in \mathbb{C}^2$ and $\tau \geq 0$. Therefore,

$$\boldsymbol{\xi}^* \frac{\partial \mathbf{A}_\zeta(\mathbf{x})}{\partial x_2} \boldsymbol{\xi} \leq 0 \tag{5.12}$$

for $\boldsymbol{\xi} \in \mathbb{C}^2$ and $\mathbf{x} \in \Omega$. On using (5.5), a similar argument to (5.11) shows that $\frac{\partial \phi_\zeta^a(\mathbf{x})}{\partial x_2} \geq 0$ for $\mathbf{x} \in \Omega$. Thus, the coefficients ϕ_ζ^a and $\mathbf{A}_\zeta$ satisfy the conditions of Corollary 4.3 for $\zeta > 0$ and $\delta_\zeta > 0$ sufficiently small. We have that

$$\begin{aligned} B(q, v; \mathbf{A}_\zeta, \phi_\zeta^a, \Omega) = B(q, v; \mathbf{A}, a, \Omega) &+ \int_\Omega (\nabla v)^* \left(\mathbf{A}_\zeta - \mathbf{A}\right) (\nabla q) \\ &+ \kappa^2 \int_\Omega \left(a - \phi_\zeta^a\right) q \overline{v}, \end{aligned} \tag{5.13}$$

for $q \in H^1_{\mathrm{qp}}(\Omega)$ and $v \in H^1_{\mathrm{qp}}(\Omega)$. Furthermore, we find a sequence $(w_\xi)_{\xi > 0}$ with $w_\xi \in C^\infty(\Omega)$ such that $\|w - w_\xi\|_{1,\Omega} \leq \xi$ for $\xi > 0$ and

$$B(w_\xi, v; \mathbf{A}, a, \Omega) = -F(v) - B(w - w_\xi, v; \mathbf{A}, a, \Omega) \tag{5.14}$$

for all $v \in H^1_{\mathrm{qp}}(\Omega)$. Upon setting $q = w_\xi$ in (5.13), we see that

$$\begin{aligned} B(w_\xi, v; \mathbf{A}_\zeta, \phi_\zeta^a, \Omega) = -F(v) - B(w - w_\xi, v; \mathbf{A}, a, \Omega) &+ \int_\Omega (\nabla v)^* \left(\mathbf{A}_\zeta - \mathbf{A}\right) \nabla w_\xi \\ &+ \kappa^2 \int_\Omega \left(a - \phi_\zeta^a\right) w_\xi \overline{v} \end{aligned} \tag{5.15}$$

for $v \in H^1_{\mathrm{qp}}(\Omega)$. Due to Corollary 4.3, with $w - w_\xi \in H^1_{\mathrm{qp}}(\Omega)$ given, we let $w_1 \in H^1_{\mathrm{qp}}(\Omega)$ solve

$$B(w_1, v; \mathbf{A}_\zeta, \phi_\zeta^a, \Omega) = -F(v) - B(w - w_\xi, v; \mathbf{A}, a, \Omega) \tag{5.16}$$

for $v \in H^1_{\mathrm{qp}}(\Omega)$ and $\delta_\zeta > 0$ small enough. We also let $w_2 \in H^1_{\mathrm{qp}}(\Omega)$ solve

$$B(w_2, v; \mathbf{A}_\zeta, \phi_\zeta^a, \Omega) = \int_\Omega (\nabla v)^* \left(\mathbf{A}_\zeta - \mathbf{A}\right) \nabla w_\xi + \kappa^2 \int_\Omega \left(a - \phi_\zeta^a\right) w_\xi \overline{v} \tag{5.17}$$

for $v \in H^1_{\mathrm{qp}}(\Omega)$ and $\delta_\zeta > 0$ small enough. Furthermore, we have a constant $C_\zeta > 0$ such that

$$\begin{aligned} \|w_1\|_{1,\Omega} &\leq C_\zeta \sup_{0 \neq v \in H^1_{\mathrm{qp}}(\Omega)} \frac{|-F(v) - B(w - w_\xi, v; \mathbf{A}, a, \Omega)|}{\|v\|_{1,\Omega}} \\ &\leq C_\zeta \left(\|F\|_{1*,\Omega} + \|w - w_\xi\|_{1,\Omega}\right). \end{aligned} \tag{5.18}$$

We also have

$$\begin{aligned} \|w_2\|_{1,\Omega} &\leq C_\zeta \sup_{0 \neq v \in H^1_{\mathrm{qp}}(\Omega)} \frac{\left|\int_\Omega (\nabla v)^* \left(\mathbf{A}_\zeta - \mathbf{A}\right) \nabla w_\xi + \kappa^2 \int_\Omega \left(a - \phi_\zeta^a\right) w_\xi \overline{v}\right|}{\|v\|_{1,\Omega}} \\ &\leq C_\zeta \left[\|\nabla w_\xi\|_\infty \|\mathbf{A}_\zeta - \mathbf{A}\|_{0,\Omega} + \kappa^2 \|w_\xi\|_\infty \left\|\phi_\zeta^a - a\right\|_{0,\Omega}\right] \end{aligned} \tag{5.19}$$

for the same constant $C_\zeta > 0$.

For each $\chi \in \{a_1^2, a_2, a_3^2, a\}$, we can write

$$\chi = \operatorname*{ess\,inf}_{\mathbf{x} \in \Omega} \chi + \chi', \tag{5.20}$$

where $\chi' \in L^\infty(\Omega)$ and $\operatorname{ess\,inf}_{\mathbf{x}\in\Omega} \chi' = 0$. But then we notice

$$\phi_\zeta^\chi \geq \operatorname*{ess\,inf}_{\mathbf{x}\in\Omega} \chi + \operatorname*{ess\,inf}_{\mathbf{x}\in\Omega} \chi' \int_{\mathbb{R}^2} \psi_\zeta(\mathbf{y})\, d\mathbf{y} = \operatorname*{ess\,inf}_{\mathbf{x}\in\Omega} \chi. \tag{5.21}$$

By writing $\chi = \operatorname{ess\,sup}_\Omega \chi + \chi'$ with $\chi' \in L^\infty(\mathbf{x} \in \Omega)$ and $\operatorname{ess\,sup}_{\mathbf{x}\in\Omega} \chi' = 0$, a similar argument shows that

$$\operatorname*{ess\,inf}_{\mathbf{x}\in\Omega} \chi \leq \phi_\zeta^\chi \leq \operatorname*{ess\,sup}_{\mathbf{x}\in\Omega} \chi. \tag{5.22}$$

Now (5.22) and the explicit form of the constant (4.16) allow us to find a constant $C > 0$ independent of $\zeta > 0$ such that $C_\zeta \leq C$ for all $\zeta > 0$. For the same reason, the sequence $(\delta_\zeta)_{\zeta>0}$ is uniformly bounded away from zero and has a convergence subsequence. Let this subsequence be conveniently denoted as $(\delta_\zeta)_{\zeta>0}$. We therefore find a $\delta > 0$ such that $\delta = \lim_{\zeta\to 0} \delta_\zeta$. By virtue of (5.16) and (5.17), we have that $w_\xi = w_1 + w_2$ and so

$$\begin{aligned} \|w\|_{1,\Omega} &\leq \|w - w_\xi\|_{1,\Omega} + \|w_\xi\|_{1,\Omega} \\ &\leq \xi + \|w_1\|_{1,\Omega} + \|w_2\|_{1,\Omega}. \end{aligned} \tag{5.23}$$

The proof follows by taking $\zeta \to 0$ and $\xi \to 0$ in (5.23). By virtue of (5.18), $\lim_{\xi\to 0} \|w_1\|_{1,\Omega} \leq C\|F\|_{1*,\Omega}$. Finally, (5.19) shows that $\lim_{\zeta\to 0} \|w_2\|_{1,\Omega} = 0$. ∎

6. Problem in the mapped domain. We now study a related scattering problem after a coordinate transformation $\mathcal{G}_\mathcal{S}^{-1}$ is used to map Ω into the domain $\widehat{\Omega} = [-\Lambda/2, \Lambda/2] \times [-H, H]$. The coordinate transformation $\mathcal{G}_\mathcal{S}$ is defined by

$$x_1 = \hat{x}_1, \tag{6.1}$$

$$x_2 = \mathcal{S}(\hat{x}_2) g(\hat{x}_1) + \hat{x}_2, \tag{6.2}$$

for some C^2 function $\mathcal{S}(\hat{x}_2)$; thus, $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2) \in \mathbb{R}^2$.

For our numerical example later on, we chose $\mathcal{S}$ to be piecewise cubic, but there are many possible choices. For the mapping to have the desired properties, we require that $\mathcal{S}(\pm H) = 0$, $\mathcal{S}(0) = 1$, $\mathcal{S}'(\pm H) = 0$ and $\mathcal{S}'(0) = 0$. The grating interface Γ gets mapped to $\hat{x}_2 = 0$ and, therefore, in the transformed coordinate system the interface $\hat{\Gamma} = \{\hat{\mathbf{x}} \,|\, \hat{x}_2 = 0\}$ is flat. Since the interfaces $\Gamma_{\pm H}$ are unchanged by the coordinate transformation, the hybrid method can be extended to multiple smooth interfaces.

For a function w in the original coordinate system, we define $\hat{w}(\hat{\mathbf{x}}) = w(\mathcal{G}_\mathcal{S}(\hat{\mathbf{x}}))$ after mapping. In order to write the equation for the transformed field $\hat{u}$, we denote the Jacobian of $\mathcal{G}_\mathcal{S}$ as $\mathbf{D}$. We seek to find a $\hat{u}$ such that

$$\widehat{\nabla} \cdot \left(|\mathbf{D}|\mathbf{D}^{-1}\widehat{\mathbf{A}}\mathbf{D}^{-\top}\widehat{\nabla}\hat{u}\right) + \kappa^2|\mathbf{D}|\hat{a}\hat{u} = \widehat{\nabla} \cdot \left[|\mathbf{D}|\mathbf{D}^{-1}(\varepsilon_+^{-1}\mathbf{I} - \widehat{\mathbf{A}})\mathbf{D}^{-\top}\widehat{\nabla}\hat{u}^{\text{inc}}\right] + \kappa^2|\mathbf{D}|(1-\hat{a})\hat{u}^{\text{inc}}, \tag{6.3}$$

subject to quasi-periodicity and radiation conditions. The advantage of this problem is that discontinuities of the transformed version $\widehat{\mathbf{A}}$ of $\mathbf{A}$ occur only on flat boundaries. We have chosen the coordinate transform so that (6.3) reduces to (3.12) in the half-spaces $U_H^\pm$ and therefore we use the same radiation conditions as we did with the original problem.

The transformed variational problem is as follows. Given an $\widehat{F} \in H^1_{\text{qp}}(\widehat{\Omega})'$, find a $\hat{w} \in H^1_{\text{qp}}(\widehat{\Omega})$ such that

$$B(\hat{w}, \hat{v}; |\mathbf{D}|\mathbf{D}^{-1}\widehat{\mathbf{A}}\mathbf{D}^{-\top}, |\mathbf{D}|\hat{a}, \widehat{\Omega}) = -(\widehat{F}, \hat{v})_{0,\widehat{\Omega}} \tag{6.4}$$

for $\hat{v} \in H^1_{\text{qp}}(\widehat{\Omega})$.

So far we have shown that the variational problem (3.23) has unique solutions for coefficient functions that are L^∞ as long as, roughly speaking, they satisfy the general non-trapping conditions

(5.4) and (5.5). We now show that there is a unique solution $\hat{w} \in H^1_{\rm qp}(\widehat{\Omega})$ to the transformed problem (6.4). The transformed coefficient functions $\mathbf{C} = |\mathbf{D}|\mathbf{D}^{-1}\widehat{\mathbf{A}}\mathbf{D}^{-\top}$ and $c = |\mathbf{D}|\hat{a}$ do not necessarily satisfy these non-trapping conditions, so we cannot appeal to our theory for the original problem. Instead, we appeal to *norm equivalence* to show an *a-priori* estimate for the transformed problem.

THEOREM 6.1. *Suppose that there exists a constant $c_0 > 0$ such that $|\mathbf{D}(\hat{\mathbf{x}})| \geq c_0 > 0$ for $\hat{\mathbf{x}} \in \widehat{\Omega}$. Then the transformed problem (6.4) has a unique variational solution $\hat{w} \in H^1_{\rm qp}(\widehat{\Omega})$. Furthermore, there is a constant $C > 0$ such that*

$$\|\hat{w}\|_{1,\widehat{\Omega}} \leq C \left\|\widehat{F}\right\|_{1*,\widehat{\Omega}}. \tag{6.5}$$

Proof. We notice that $||\mathbf{D}|\mathbf{D}^{-1}\mathbf{D}^{-\top}| = 1$ and so $\lambda_{\max}(\hat{\mathbf{x}}) = \lambda_{\min}(\hat{\mathbf{x}})^{-1}$. The determinant $|\mathbf{D}(\hat{\mathbf{x}})| = 1 + \mathcal{S}'(\hat{x}_2)g(\hat{x}_1)$ is continuous since we have assumed that $\mathcal{S}(\hat{x}_2)$ and $g(\hat{x}_1)$ are C^2 functions. We therefore have a constant $c_1 > 0$ such that

$$c_1 \geq \operatorname{tr}(|\mathbf{D}|\mathbf{D}^{-1}\mathbf{D}^{-\top}) = |\mathbf{D}|\left(1 + \frac{(\mathcal{S}(\hat{x}_2)g'(\hat{x}_1))^2 + 1}{|\mathbf{D}|^2}\right) \geq c_0. \tag{6.6}$$

Since the trace is the sum of eigenvalues, in particular we can show that $\lambda_{\min}(\hat{\mathbf{x}})$ is positive and bounded away from zero. After multiplying (6.6) by $\lambda_{\min}(\hat{\mathbf{x}})$, we see that

$$c_1 \geq \frac{\lambda^2_{\min}(\hat{\mathbf{x}}) + 1}{\lambda_{\min}(\hat{\mathbf{x}})} \geq c_0 \tag{6.7}$$

for $\hat{\mathbf{x}} \in \widehat{\Omega}$. Since $|\mathbf{D}|\mathbf{D}^{-1}\mathbf{D}^{-\top}$ is real and symmetric, it follows that it is uniformly positive definite and there are constants $C_1 > 0$ and $C_2 > 0$ such that

$$C_1|\boldsymbol{\xi}|^2 \leq \boldsymbol{\xi}^*|\mathbf{D}|\mathbf{D}^{-1}\mathbf{D}^{-\top}(\hat{\mathbf{x}})\boldsymbol{\xi} \leq C_2|\boldsymbol{\xi}|^2 \tag{6.8}$$

for $\boldsymbol{\xi} \in \mathbb{C}^2$ and $\hat{\mathbf{x}} \in \widehat{\Omega}$.

For $w \in C^\infty(\Omega)$, it follows that

$$\begin{aligned} \|w\|^2_{1,\Omega} &= (|\mathbf{D}|\hat{w}, \hat{w})_{0,\widehat{\Omega}} + (|\mathbf{D}|\mathbf{D}^{-\top}\widehat{\nabla}\hat{w}, \mathbf{D}^{-\top}\widehat{\nabla}\hat{w})_{0,\widehat{\Omega}} \\ &\geq \min(c_0, C_1)\|\hat{w}\|^2_{1,\widehat{\Omega}}. \end{aligned} \tag{6.9}$$

On the other hand, we have

$$\|\hat{w}\|^2_{1,\widehat{\Omega}} \geq \min\left(\frac{1}{\sup_{\hat{\mathbf{x}}\in\widehat{\Omega}}(|\mathbf{D}(\hat{\mathbf{x}})|)}, \frac{1}{C_2}\right)\|w\|^2_{1,\Omega}, \tag{6.10}$$

which shows that we have norm equivalence for C^∞ functions. For $w \in H^1_{\rm qp}(\Omega)$ we can find a sequence $(w_\zeta)_\zeta$ of $C^\infty(\Omega)$ functions that converge to w in H^1 using the density of $C^\infty(\Omega) \cap H^1_{\rm qp}(\Omega)$ in $H^1_{\rm qp}(\Omega)$. The mapped sequence $(\hat{w}_\zeta)_\zeta$ defined by $\hat{w}_\zeta = w_\zeta \circ \mathcal{G}_\mathcal{S}$ is a Cauchy sequence, using the norm equivalence for smooth functions. We therefore find a unique $\hat{w} \in H^1_{\rm qp}(\widehat{\Omega})$ and two constants $C_1 > 0$ and $C_2 > 0$ such that

$$C_1\|w\|_{1,\Omega} \leq \|\hat{w}\|_{1,\widehat{\Omega}} \leq C_2\|w\|_{1,\Omega}. \tag{6.11}$$

Given an $\widehat{F} \in H^1_{\rm qp}(\widehat{\Omega})'$, we can find a unique $f_{\widehat{F}} \in H^1_{\rm qp}(\Omega)$ and an $F \in H^1_{\rm qp}(\Omega)'$ such that

$$\widehat{F}(\hat{v}) = (v, |\mathbf{D}(\mathcal{G}^{-1}_\mathcal{S}(\mathbf{x}))|^{-1}f_{\widehat{F}})_{0,\Omega} = F(v). \tag{6.12}$$

For this particular $F(v)$, we have a unique solution $w \in H^1_{\rm qp}(\Omega)$ to the original variational problem (3.23). A similar density argument yields a unique $\hat{w} \in H^1_{\rm qp}(\widehat{\Omega})$ that solves (6.4). To complete the proof, we use norm equivalence again to see that

$$\|\hat{w}\|_{1,\widehat{\Omega}} \leq C_2C\|F\|_{1*,\Omega} \leq C_2^2C\left\|\widehat{F}\right\|_{1*,\widehat{\Omega}}.$$

∎

Our next aim is to show that the solution has some additional regularity, i.e., $w \in H^{1+s}(\Omega)$ for $s \in (0, s_0)$ with $s_0 \in (0, 1/2)$. Furthermore, using this regularity result, we will aim to show that the mapped solution $\hat{w} \in H^{1+s}(\widehat{\Omega})$. The hybrid coordinate-transformation method converges due to this extra regularity, as shown in Secs. 8 and 9.

THEOREM 6.2. *Suppose the conditions of Theorem 5.1 hold. Let $w \in H^1(\Omega)$ be the solution to the variational problem (3.23) with $F \in L^2_{qp}(\Omega)$ on the right hand side. Then there exist constants $C > 0$ and $s_0 \in (0, 1/2)$ such that*

$$\|w\|_{1+s,\Omega} \leq C \|F\|_{0,\Omega} \qquad \text{for } 0 < s \leq s_0. \tag{6.13}$$

Proof. We extend F and w by quasi-periodicity. In U_H^+ and U_H^- we extend F by zero and the solution w using its Rayleigh–Bloch expansion [57]. We denote these extended functions by F° and w°. We define the circular extended domain Ω° as

$$\Omega^\circ = \{\mathbf{x} \mid R > |\mathbf{x}|\}, \tag{6.14}$$

where $R > 0$ is chosen large enough so that $\overline{\Omega} \subset \Omega^\circ$. Also, we let $\chi \in C^\infty$ be a smooth cutoff function such that $\chi \equiv 1$ in Ω° and $\chi \equiv 0$ on $\partial\Omega^\circ$. We define $p \in H^1(\Omega)$ as $p = \chi w^\circ$ that solves the original problem and $p \equiv 0$ on $\partial\Omega^\circ$.

Since

$$\int_{\Omega^\circ} \nabla \cdot (\mathbf{A}\nabla p)\, \overline{v} = \int_{\Omega^\circ} \left[(\nabla\chi)^\top \mathbf{A}\nabla w^\circ + \nabla \cdot (w^\circ \mathbf{A}\nabla\chi) + (F^\circ - \kappa^2 a w^\circ)\chi \right] \overline{v} \tag{6.15}$$

for $v \in H^1(\Omega^\circ)$, p solves the Laplace equation with zero boundary conditions and a right hand side $G = (\nabla\chi)^\top \mathbf{A}\nabla w^\circ + \nabla \cdot (w^\circ \mathbf{A}\nabla\chi) + (F^\circ - \kappa^2 a w^\circ)\chi$. By virtue of Proposition 2.1 of Bonito *et al.* [76], it follows that $G \in H^{s-1}$ for $s_0 \in (0, 1/2)$ and $s \in (0, s_0)$. Also, Theorem 3.1 of Bonito *et al.* [76] shows that we have a constant $C > 0$ such that

$$\|p\|_{1+s,\Omega^\circ} \leq C \|G\|_{s-1,\Omega^\circ}. \tag{6.16}$$

It follows from a predeceesor paper [56] and our *a-priori* bound (5.6) that

$$\begin{aligned} \|w\|_{1+s,\Omega} &\leq C \|G\|_{s-1,\Omega^\circ} \\ &\leq C \left(\|w^\circ\|_{1,\Omega^\circ} + \|F^\circ\|_{0,\Omega^\circ} \right) \\ &\leq C \|F\|_{0,\Omega}. \end{aligned} \tag{6.17}$$

□

COROLLARY 6.3. *Suppose the conditions of Theorems 6.1 and 6.2 hold. Then there exists constants $C > 0$ and $s_0 \in (0, 1/2)$ such that*

$$\|\hat{w}\|_{1+s,\widehat{\Omega}} \leq C \left\|\widehat{F}\right\|_{0,\widehat{\Omega}} \qquad \text{for } 0 < s \leq s_0. \tag{6.18}$$

Proof. The result follows from Theorem 6.2 and Heuer Lemma 2.8 [77]. □

7. Discretized problem. In this section, we define a discretized problem where the transformed coefficient functions $\mathbf{C} = |\mathbf{D}|\mathbf{D}^{-1}\widehat{\mathbf{A}}\mathbf{D}^{-\top}$ and $c = |\mathbf{D}|\hat{a}$ are replaced by piecewise smooth approximations. Since this new problem no longer corresponds to a coordinate transform, we cannot appeal to norm equivalence as we did in Sec. 6 to show existence and uniqueness of solutions. Instead, we use the fact that the discretized problem is a small perturbation of the transformed problem.

For an integer $N \geq 1$, we define the discretization parameter $h = 2H/N$ and the $N+1$ grid points

$$-H = \hat{x}_{2,0} < \hat{x}_{2,1} < \cdots < \hat{x}_{2,N} < \hat{x}_{2,N} = H. \tag{7.1}$$

We assume that $\hat{x}_2 = 0$ is a grid point. The domain $\widehat{\Omega}$ is decomposed into N thin slices

$$S_n = [-\Lambda/2, \Lambda/2] \times [\hat{x}_{2,n}, \hat{x}_{2,n+1}]\,, \quad n \in [0, N-1]\,, \tag{7.2}$$

and the transformed coefficients functions are sampled at the slice midpoints given by

$$\hat{x}^*_{2,n} = \frac{1}{2}\left(\hat{x}_{2,n} + \hat{x}_{2,n+1}\right)\,, \quad n \in [0, N-1]\,. \tag{7.3}$$

On each slice S_n, we define the discretized coefficient functions as

$$\Phi_h(\hat{\mathbf{x}}) = \Phi(\hat{x}_1, \hat{x}^*_{2,n}) \tag{7.4}$$

for $\hat{\mathbf{x}} \in S_n$ and $\Phi \in \{\mathbf{C}, c\}$. The discretized coefficient functions are piecewise constant in $\hat{x}_2$ and only depend on $\hat{x}_1$ in each slice S_n. Since the transformed coefficients only have discontinuities on a flat interface, the discretized coefficients are continuous in every slice. The discretized variational problem is as follows. Given $\hat{F} \in H^1_{\mathrm{qp}}(\widehat{\Omega})'$, find a $\hat{w}_h \in H^1_{\mathrm{qp}}(\widehat{\Omega})$ such that

$$B(\hat{w}_h, \hat{v}, \mathbf{C}_h, c_h, \widehat{\Omega}) = \hat{F}(\hat{v}) \tag{7.5}$$

for $\hat{v} \in H^1_{\mathrm{qp}}(\widehat{\Omega})$.

THEOREM 7.1. *Suppose the conditions of Theorems 6.1 and 6.2 hold. Then for $h > 0$ small enough, there is a unique solution $\hat{w} \in H^1_{\mathrm{qp}}(\widehat{\Omega})$ to the discretized problem (7.5). Furthermore, there exists a constant $C > 0$ such that*

$$\|\hat{u} - \hat{u}_h\|_{1,\widehat{\Omega}} \le Ch \left\|u^{\mathrm{inc}}\right\|_{1,\Omega}. \tag{7.6}$$

Proof. We define the maps $T : H^1_{\mathrm{qp}}(\widehat{\Omega}) \to H^1_{\mathrm{qp}}(\widehat{\Omega})'$ and $T_h : H^1_{\mathrm{qp}}(\widehat{\Omega}) \to H^1_{\mathrm{qp}}(\widehat{\Omega})'$ by $(T\hat{w}, \hat{v}) = B(\hat{w}, \hat{v}, \mathbf{C}, c, \widehat{\Omega})$ and $(T_h\hat{w}, \hat{v}) = B(\hat{w}, \hat{v}, \mathbf{C}_h, c_h, \widehat{\Omega})$, respectively. We see that

$$\begin{aligned} |[(T - T_h)\hat{w}, \hat{v}]| &\le \|\mathbf{C} - \mathbf{C}_h\|_\infty \left\|\widehat{\nabla}\hat{w}\right\|_{0,\widehat{\Omega}} \left\|\widehat{\nabla}\hat{v}\right\|_{0,\widehat{\Omega}} + \kappa^2 \|c - c_h\|_\infty \|\hat{w}\|_{0,\widehat{\Omega}} \|\hat{v}\|_{0,\widehat{\Omega}} \\ &\le Ch \|\hat{w}\|_{1,\widehat{\Omega}} \|\hat{v}\|_{1,\widehat{\Omega}}\,. \end{aligned} \tag{7.7}$$

Then in the operator norm, we have

$$\|T - T_h\| = \sup_{0 \ne \hat{w} \in H^1_{\mathrm{qp}}(\widehat{\Omega})} \frac{\|(T - T_h)\hat{w}\|_{1*,\widehat{\Omega}}}{\|\hat{w}\|_{1,\Omega}} \le Ch. \tag{7.8}$$

It follows from Corollary 10.3 of Kress [78] that, for $h > 0$ small enough, there is a constant $C > 0$ such that

$$\begin{aligned} \|\hat{u} - \hat{u}_h\|_{1,\widehat{\Omega}} &\le C\left(\|(T - T_h)\hat{u}\|_{1*,\widehat{\Omega}} + \left\|\widehat{G} - \widehat{G}_h\right\|_{1*,\widehat{\Omega}}\right) \\ &\le Ch\left(\|\hat{u}\|_{1,\widehat{\Omega}} + \left\|\hat{u}^{\mathrm{inc}}\right\|_{1,\widehat{\Omega}}\right) \\ &\le Ch \left\|u^{\mathrm{inc}}\right\|_{1,\Omega}\,. \end{aligned} \tag{7.9}$$

∎

8. Convergence in the number of retained Fourier terms. In this section, we show that the hybrid coordinate-transformation method converges with respect to the number of terms retained in the Fourier series. We return to the case of a generic orthotropic medium described by (2.1) and (2.2) that occupies Ω, with $\tilde{\boldsymbol{\varepsilon}}$ a symmetric matrix. Later in this section, we show how to choose $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ in order to solve the transformed scattering problem.

In the mapped domain, we consider $\hat{\mathbf{E}}(\hat{\mathbf{x}}) = \hat{E}_1(\hat{\mathbf{x}})\mathbf{e}_1 + \hat{E}_2(\hat{\mathbf{x}})\mathbf{e}_2$ and $\hat{\mathbf{H}}(\hat{\mathbf{x}}) = \hat{H}_3(\hat{\mathbf{x}})\mathbf{e}_3$, where

$$\hat{E}_1 = -\frac{1}{i\omega\varepsilon_0\,|\tilde{\boldsymbol{\varepsilon}}|}\left(\varepsilon_{22}\frac{\partial \hat{H}_3}{\partial \hat{x}_2} + \varepsilon_{12}\frac{\partial \hat{H}_3}{\partial \hat{x}_1}\right), \tag{8.1}$$

$$\hat{E}_2 = \frac{1}{i\omega\varepsilon_0\,|\tilde{\boldsymbol{\varepsilon}}|}\left(\varepsilon_{21}\frac{\partial \hat{H}_3}{\partial \hat{x}_2} + \varepsilon_{11}\frac{\partial \hat{H}_3}{\partial \hat{x}_1}\right), \tag{8.2}$$

and

$$\widehat{\nabla}\cdot\left(|\tilde{\boldsymbol{\varepsilon}}|^{-1}\,\tilde{\boldsymbol{\varepsilon}}^{\top}\widehat{\nabla}\hat{H}_3\right) + \kappa^2\mu_{33}\hat{H}_3 = 0. \tag{8.3}$$

Therefore, in order to solve the correct Helmholtz equation (6.3) we choose the constitutive parameters such that

$$\boldsymbol{\varepsilon} = \left(\begin{array}{c|c} |\widehat{\mathbf{A}}|^{-1}|\mathbf{D}|\mathbf{D}^{-1}\widehat{\mathbf{A}}\mathbf{D}^{-\top} & \mathbf{0} \\ \hline \mathbf{0} & |\widehat{\tilde{\boldsymbol{\mu}}}|^{-1}|\mathbf{D}|\hat{\varepsilon}_{33} \end{array}\right) \quad \text{and} \quad \boldsymbol{\mu} = \left(\begin{array}{c|c} |\mathbf{D}|\mathbf{D}^{-1}\widehat{\tilde{\boldsymbol{\mu}}}\mathbf{D}^{-\top} & \mathbf{0} \\ \hline \mathbf{0} & |\mathbf{D}|\hat{a} \end{array}\right). \tag{8.4}$$

For example, $|\tilde{\boldsymbol{\varepsilon}}|^{-1} = |\widehat{\mathbf{A}}|$ and therefore $|\tilde{\boldsymbol{\varepsilon}}|^{-1}\tilde{\boldsymbol{\varepsilon}}^{\top} = |\mathbf{D}|\mathbf{D}^{-1}\widehat{\mathbf{A}}\mathbf{D}^{-\top}$ since $\widehat{\mathbf{A}}$ is symmetric. The partial differential equation (8.3) is difficult to solve, so we replace the coefficient functions with the piecewise smooth functions $\mathbf{C}_h$ and c_h given by (7.4).

The constitutive parameters are written as the Fourier series

$$\boldsymbol{\varepsilon}(\hat{\mathbf{x}}) = \sum_{m\in\mathbb{Z}}\boldsymbol{\varepsilon}_m(\hat{x}_2)\exp\left(i\frac{2\pi}{\Lambda}m\hat{x}_1\right) \quad \text{and} \quad \boldsymbol{\mu}(\hat{\mathbf{x}}) = \sum_{m\in\mathbb{Z}}\boldsymbol{\mu}_m(\hat{x}_2)\exp\left(i\frac{2\pi}{\Lambda}m\hat{x}_1\right). \tag{8.5}$$

We also expand the field phasors in terms of the Rayleigh–Bloch expansions

$$\hat{\mathbf{E}}(\hat{\mathbf{x}}) = \hat{E}_1(\hat{\mathbf{x}})\mathbf{e}_1 + \hat{E}_2(\hat{\mathbf{x}})\mathbf{e}_2 = \sum_{m\in\mathbb{Z}}\left[\hat{E}_{1,m}(\hat{x}_2)\mathbf{e}_1 + \hat{E}_{2,m}(\hat{x}_2)\mathbf{e}_2\right]\exp\left(i\alpha_m\hat{x}_1\right), \tag{8.6}$$

$$\hat{\mathbf{H}}(\hat{\mathbf{x}}) = \hat{H}_3(\hat{\mathbf{x}})\mathbf{e}_3 = \sum_{m\in\mathbb{Z}}\left[\hat{H}_{3,m}(\hat{x}_2)\mathbf{e}_3\right]\exp\left(i\alpha_m\hat{x}_1\right), \tag{8.7}$$

with the unknown Fourier functions $\hat{E}_{1,m}(\hat{x}_2)$, $\hat{E}_{2,m}(\hat{x}_2)$, and $\hat{H}_{3,m}(\hat{x}_2)$. Substitution of the expansions (8.5)–(8.7) in the time-harmonic Maxwell equations delivers a system of two first-order partial differential equations and an algebraic equations relating the unknown coefficients $\hat{E}_{1,m}(\hat{x}_2)$, $\hat{E}_{2,m}(\hat{x}_2)$, and $\hat{H}_{3,m}(\hat{x}_2)$ [12, 49]. Once this system is solved, the electric and magnetic field phasors can be reconstructed in the mapped domain $\widehat{\Omega}$ using (8.6) and (8.7).

We truncate the series (8.6) and (8.7) to include only $|m| \le M$, $M \ge 0$. In order to write a system of $2M+1$ equations relating the Fourier coefficients, we define the three $(2M+1)\times 1$ matrixes

$$\hat{\mathfrak{E}}_1 = \left(\hat{E}_{1,-M}, \hat{E}_{1,-M+1}, \cdots, \hat{E}_{1,M-1}, \hat{E}_{1,M}\right)^{\top}, \tag{8.8}$$

$$\hat{\mathfrak{E}}_2 = \left(\hat{E}_{2,-M}, \hat{E}_{2,-M+1}, \cdots, \hat{E}_{2,M-1}, \hat{E}_{2,M}\right)^{\top}, \tag{8.9}$$

and

$$\hat{\mathfrak{H}}_3 = \left(\hat{H}_{3,-M}, \hat{H}_{3,-M+1}, \cdots, \hat{H}_{3,M-1}, \hat{H}_{3,M}\right)^{\top}. \tag{8.10}$$

along with the $(2M+1)\times(2M+1)$ matrix

$$\mathfrak{a} = \operatorname{diag}\left(\alpha_{-M}, \alpha_{-M+1}, \cdots, \alpha_{M-1}, \alpha_M\right). \tag{8.11}$$

With $\boldsymbol{\mathcal{T}}(\phi)$ denoting the $(2M+1)\times(2M+1)$ Toeplitz matrix formed from the Fourier coefficients of a periodic function ϕ, the RCWA then requires the solution of the $2M+1$ equations

$$\begin{aligned}\boldsymbol{\mathcal{T}}(C_{h,22})\frac{\partial^2\hat{\boldsymbol{\mathfrak{H}}}_3}{\partial\hat{x}_2^2}+i\left[\mathfrak{a}\boldsymbol{\mathcal{T}}(C_{h,21})+\boldsymbol{\mathcal{T}}(C_{h,12})\mathfrak{a}\right]\frac{\partial\hat{\boldsymbol{\mathfrak{H}}}_3}{\partial\hat{x}_2}\\ +\left[\kappa^2\boldsymbol{\mathcal{T}}(c_h)-\mathfrak{a}\boldsymbol{\mathcal{T}}(C_{h,11})\mathfrak{a}\right]\hat{\boldsymbol{\mathfrak{H}}}_3=\mathbf{0}\end{aligned} \tag{8.12}$$

in each slice S_n, where $C_{h,\ell m}$ denotes the ℓmth component of $\mathbf{C}_h$.

The RCWA enforces transmission conditions for the tangential components of the fields across all interslice boundaries. Also, appropriate boundary conditions are enforced on the top and bottom boundaries of $\widehat{\Omega}$. To this end, we define the reflected and transmitted fields as

$$\hat{\mathbf{H}}^{\mathrm{ref}}(\hat{\mathbf{x}})=\mathbf{e}_3\sum_{m=-M}^{M}\hat{H}_{3,m}^{\mathrm{ref}}\exp[i\beta_m(\hat{x}_2-H)]\exp(i\alpha_m\hat{x}_1),\qquad \hat{x}_2\geq H, \tag{8.13}$$

and

$$\hat{\mathbf{H}}^{\mathrm{tr}}(\hat{\mathbf{x}})=\mathbf{e}_3\sum_{m=-M}^{M}\hat{H}_{3,m}^{\mathrm{tr}}\exp[-i\beta_m(\hat{x}_2+H)]\exp(i\alpha_m\hat{x}_1),\qquad \hat{x}_2\leq -H, \tag{8.14}$$

with unknown coefficients $\hat{H}_{3,m}^{\mathrm{ref}}$ and $\hat{H}_{3,m}^{\mathrm{tr}}$, respectively. Also, the incident field (3.2) can be written as

$$\hat{H}_3^{\mathrm{inc}}(\hat{\mathbf{x}})=-\exp[-i\beta_0(\hat{x}_2-H)]\exp(i\alpha_0\hat{x}_1). \tag{8.15}$$

The boundary conditions across the plane $\hat{x}_2=H$ are

$$\left.\begin{aligned}\hat{H}_{3,m}(\hat{x}_2)&=-\delta_{m0}+\hat{H}_{3,m}^{\mathrm{ref}}\\ \frac{\partial\hat{H}_{3,m}(\hat{x}_2)}{\partial\hat{x}_2}&=i\beta_m\left(\delta_{m0}+\hat{H}_{3,m}^{\mathrm{ref}}\right)\end{aligned}\right\},\quad \hat{x}_2=H\,, \tag{8.16}$$

where $\delta_{mm'}$ is the Kronecker delta. The boundary conditions across the plane $\hat{x}_2=-H$ are

$$\left.\begin{aligned}\hat{H}_{3,m}(\hat{x}_2)&=\hat{H}_{3,m}^{\mathrm{tr}}\\ \frac{\partial\hat{H}_{3,m}(\hat{x}_2)}{\partial\hat{x}_2}&=-i\beta_m\hat{H}_{3,m}^{\mathrm{tr}}\end{aligned}\right\},\quad \hat{x}_2=-H\,. \tag{8.17}$$

In addition, the following transmission conditions hold for every interslice boundary:

$$\left.\begin{aligned}\left[\left[\hat{H}_3\right]\right]_{\hat{x}_2=\hat{x}_{2,n}}&=0\\ \left[\left[\boldsymbol{\nu}^\top\mathbf{C}_h\widehat{\nabla}\hat{H}_3\right]\right]_{\hat{x}_2=\hat{x}_{2,n}}&=0\end{aligned}\right\},\quad n\in[1,N-1]\,. \tag{8.18}$$

Now we can show that the RCWA for this problem is actually a Galerkin method, i.e., the solution solves the appropriate variational problem.

Theorem 8.1. *In the mapped domain,* $\hat{u}_{h,M}^t(\hat{\boldsymbol{x}})=\hat{H}_3(\hat{\boldsymbol{x}})$ *solves the variational problem*

$$B(\hat{u}_{h,M}^t,\hat{v}_M;\mathbf{C}_h,c_h,\widehat{\Omega})=\left\langle\frac{1}{\varepsilon_+}\frac{\partial\hat{u}^{\mathrm{inc}}}{\partial\hat{x}_2}-T^+(\hat{u}^{\mathrm{inc}}),\hat{v}_M\right\rangle_{0,\hat{\Gamma}_H} \tag{8.19}$$

for every $\hat{v}_M\in H^1(-H,H)\otimes S_M$ *with* $S_M=\mathit{span}_{|m|\leq M}[\exp(i\alpha_m\hat{x}_1)]$.

Proof. We let the test function $\hat{v}_M \in H^1(-H,H) \otimes \mathrm{span}_{|m|\le M}[\exp(i\alpha_m \hat{x}_1)]$. We can therefore write

$$\overline{\hat{v}}_M = \sum_{m=-M}^{M} \overline{\hat{v}}_m(\hat{x}_2)\exp(-i\alpha_m \hat{x}_1). \tag{8.20}$$

By virtue of the orthogonality of the basis functions $\exp(i\alpha_m \hat{x}_1)$, it follows from (8.12) that

$$\int_{-\Lambda/2}^{\Lambda/2} \left[\widehat{\nabla}\cdot\left(\mathbf{C}_h \widehat{\nabla}\hat{H}_3\right)\overline{\hat{v}}_M + \kappa^2 c_h \hat{H}_3 \overline{\hat{v}}_M\right] d\hat{x}_1 = 0 \tag{8.21}$$

in each slice. Now we integrate with respect to $\hat{x}_2$ and use the divergence theorem in each slice to obtain

$$\begin{aligned}\int_{S_n} &\left[(\widehat{\nabla}\hat{v}_M)^* \mathbf{C}_h \widehat{\nabla}\hat{H}_3 - \kappa^2 c_h \hat{H}_3 \overline{\hat{v}}_M\right] \\ &\qquad + \int_{\hat{x}_2=\hat{x}_{2,n}} \overline{\hat{v}}_M \boldsymbol{\nu}^\top \mathbf{C}_h \widehat{\nabla}\hat{H}_3 - \int_{\hat{x}_2=\hat{x}_{2,n+1}} \overline{\hat{v}}_M \boldsymbol{\nu}^\top \mathbf{C}_h \widehat{\nabla}\hat{H}_3 = 0.\end{aligned} \tag{8.22}$$

Next, we sum over all the slices and use the transmission conditions to obtain

$$\int_{\widehat{\Omega}} \left[(\widehat{\nabla}\hat{v}_M)^* \mathbf{C}_h \widehat{\nabla}\hat{H}_3 - \kappa^2 c_h \hat{H}_3 \overline{\hat{v}}_M\right] - \int_{\hat{\Gamma}_H} \overline{\hat{v}}_M \frac{1}{\varepsilon_+}\frac{\partial \hat{H}_3}{\partial \hat{x}_2} + \int_{\hat{\Gamma}_{-H}} \overline{\hat{v}}_M \frac{1}{\varepsilon_+}\frac{\partial \hat{H}_3}{\partial \hat{x}_2} = 0. \tag{8.23}$$

By applying the boundary conditions across $\hat{x}_2 = \pm H$, we see that

$$\frac{1}{\varepsilon_+}\frac{\partial \hat{H}_3}{\partial \hat{x}_2}(H) = \frac{1}{\varepsilon_+}\frac{\partial \hat{u}^{\mathrm{inc}}}{\partial x_2} - T^+(\hat{u}^{\mathrm{inc}}) + T^+(\hat{H}_3) \tag{8.24}$$

and

$$\frac{1}{\varepsilon_+}\frac{\partial \hat{H}_3}{\partial \hat{x}_2}(-H) = -T^-(\hat{H}_3). \tag{8.25}$$

The use of (8.24) and (8.25) in (8.23) completes the proof. ∎

To show that we have convergence in the parameter M, we now consider an adjoint problem. To this end, let $B^*(w,v) = \overline{B(v,w)}$ be the adjoint form of the sesquilinear form $B(\cdot,\cdot)$. Given an $\hat{F} \in L^2(\widehat{\Omega})$, the adjoint problem is to seek a $\hat{z}_{\hat{F}} \in H^1_{\mathrm{qp}}(\widehat{\Omega})$ such that

$$B^*(\hat{z}_{\hat{F}}, \hat{v}; \mathbf{C}, c, \widehat{\Omega}) = \left(\hat{F}, \hat{v}\right)_{0,\widehat{\Omega}} \tag{8.26}$$

for $\hat{v} \in H^1_{\mathrm{qp}}(\widehat{\Omega})$. Due to our theory, we know that $\hat{z}_{\hat{F}}$ exists and is unique; furthermore there are constants $C > 0$ and $s_0 \in (0, 1/2)$ such that

$$\left\|\hat{z}_{\hat{F}}\right\|_{1+s,\widehat{\Omega}} \le C\left\|\hat{F}\right\|_{0,\widehat{\Omega}} \tag{8.27}$$

for $s \in (0, s_0)$.

THEOREM 8.2. *Assume that the conditions of Theorem 7.1 hold. For $M > 0$ large enough, there are constants $C > 0$ and $s_0 \in (0, 1/2)$ independent of $h > 0$ such that*

$$\|\hat{u}_h - \hat{u}_{h,M}\|_{0,\widehat{\Omega}} \le C\left(M^{-2s} + hM^{-s}\right) \qquad \textit{for } 0 < s \le s_0. \tag{8.28}$$

Proof. We define the truncation operator $\mathcal{F}_M : H^1_{\text{qp}}(\widehat{\Omega}) \to H^1(-H,H) \otimes S_M$ as

$$\mathcal{F}_M \phi(\hat{\boldsymbol{x}}) = \sum_{m=-M}^{M} \phi_m(\hat{x}_2) \exp(i\alpha_m \hat{x}_1) \tag{8.29}$$

for $\phi \in H^1_{\text{qp}}(\widehat{\Omega})$. We let $\hat{e}_M = \hat{u}_h - \hat{u}_{h,M}$ and set $\hat{F} = \hat{e}_M$ and $\hat{v} = \hat{e}_M$ in the adjoint problem (8.26). By virtue of theorem 8.1 we have the Galerkin orthogonality

$$B(\hat{e}_M, \mathcal{F}_M \hat{z}_{\hat{e}_M}; \mathbf{C}_h, c_h, \widehat{\Omega}) = 0. \tag{8.30}$$

It follows that

$$\begin{aligned}
\|\hat{e}_M\|^2_{0,\widehat{\Omega}} &\overset{(8.30)}{=} B(\hat{e}_M, \hat{z}_{\hat{e}_M} - \mathcal{F}_M \hat{z}_{\hat{e}_M}; \mathbf{C}_h, c_h, \widehat{\Omega}) + B(\hat{e}_M, \hat{z}_{\hat{e}_M}; \mathbf{C} - \mathbf{C}_h, c - c_h, \widehat{\Omega}) \\
&\le \gamma \|\hat{e}_M\|_{1,\widehat{\Omega}} \|\hat{z}_{\hat{e}_M} - \mathcal{F}_M \hat{z}_{\hat{e}_M}\|_{1,\widehat{\Omega}} + \left| B(\hat{e}_M, \hat{z}_{\hat{e}_M}; \mathbf{C} - \mathbf{C}_h, c - c_h, \widehat{\Omega}) \right| \\
&\le \gamma \|\hat{e}_M\|_{1,\widehat{\Omega}} \|\hat{z}_{\hat{e}_M}\|_{1+s,\widehat{\Omega}} M^{-s} + Ch \|\hat{e}_M\|_{1,\widehat{\Omega}} \|\hat{z}_{\hat{e}_M}\|_{1,\widehat{\Omega}} \\
&\overset{(8.27)}{\le} C \left[M^{-s} + h \right] \|\hat{e}_M\|_{0,\widehat{\Omega}} \|\hat{e}_M\|_{1,\widehat{\Omega}}
\end{aligned} \tag{8.31}$$

for $s \in (0, s_0)$. We then divide (8.31) through by $\|\hat{e}_M\|_{0,\widehat{\Omega}}$ to see that

$$\|\hat{e}_M\|_{0,\widehat{\Omega}} \le C \left[M^{-s} + h \right] \|\hat{e}_M\|_{1,\widehat{\Omega}}. \tag{8.32}$$

The proof is complete by applying an argument of Schatz [79] that shows $\|\hat{e}_M\|_{1,\widehat{\Omega}} \le CM^{-s}$ for $M > 0$ large enough. ∎

We conclude this section by combining our convergence theory into a single theorem that shows that the approximate solution $\hat{u}_{h,M}$ in the mapped domain converges to the true solution $\hat{u}$. Generally speaking, we need the relative permittivity $\boldsymbol{\varepsilon}$ and relative permeability $\boldsymbol{\mu}$ from the original scattering problem to be non-trapping.

THEOREM 8.3. *Suppose that the conditions of Theorems 5.1 and 6.1 hold. Then for some $h > 0$ small enough and $M > 0$ large enough, there are constants $C > 0$ and $s_0 \in (0, 1/2)$ independent of h and M such that*

$$\|\hat{u} - \hat{u}_{h,M}\|_{K,\widehat{\Omega}} \le C \left[h + M^{-s(2-K)} + h(1-K) M^{-s(1-K)} \right], \tag{8.33}$$

where $K \in \{0, 1\}$ and $s \in (0, s_0)$.

Proof. We write $\hat{u} - \hat{u}_{h,M} = \hat{u} - \hat{u}_h + \hat{u}_h - \hat{u}_{h,M}$ and use the triangle inequality. The proof follows from Theorems 6.1 and 8.2. ∎

9. Numerical example. We now present a numerical example to test our convergence theory. Let $\mathcal{S}(\hat{x}_2)$ be a piecewise cubic function in the computational domain and identically zero outside; in particular,

$$\mathcal{S}(\hat{x}_2) = \begin{cases} 1 - \frac{3}{H^2}\hat{x}_2^2 + \frac{2}{H^3}\hat{x}_2^3, & \hat{x}_2 \in [0, H], \\ 1 - \frac{3}{H^2}\hat{x}_2^2 - \frac{2}{H^3}\hat{x}_2^3, & \hat{x}_2 \in [-H, 0), \\ 0, & |\hat{x}_2| > H. \end{cases} \tag{9.1}$$

We used Matlab version R2019b to implement the RCWA for anisotropic constitutive parameters.

When the grating is flat and $\boldsymbol{\varepsilon}(\mathbf{x}) = \varepsilon(x_2)\mathbf{I}$ is piecewise uniform, the scattering problem can be solved analytically [80]. However, as analytic solutions are unavailable when $g(x_1)$ is not constant for $x_1 \in [-\Lambda/2, \Lambda/2]$, we compared the results of our hybrid method to those from the finite element method (FEM) [81, 82]. We used NETGEN version 6.2 [83] to implement an FEM solver and $\hat{u}_{h,M}$ was compared to the FEM solution $\hat{u}_{FEM}$ with respect to the relative L^2 norm

$$\|\hat{u}_{h,M} - \hat{u}_{FEM}\| \, \|\hat{u}_{FEM}\|^{-1}. \tag{9.2}$$

After the solution $\hat{u}_{h,M}$ has been calculated, it is possible to map it back into the original spatial domain as

$$u_{h,M} = \hat{u}_{h,M}(\mathcal{G}_{\mathcal{S}}^{-1}(\mathbf{x})). \tag{9.3}$$

Given a point (x_1, x_2) we solved (6.2) for $\hat{x}_2$ using the bisection method. We then found the closest grid point $\hat{x}_{2,n}$ to this solution and set $u_{h,M}(x_1, x_2) = \hat{u}_{h,M}(\hat{x}_1, \hat{x}_{2,n})$. Therefore, we also compared the results of the RCWA solution [57] of the original problem (3.5)–(3.7) to the mapped solution $u_{h,M}$.

We now consider a test case where

$$g(x_1) = \frac{\Lambda}{5} \cos(2\pi x_1/\Lambda) \tag{9.4}$$

with $\Lambda = 500$ nm, $H = 700$ nm, $\kappa = 2\pi/600$ nm^{-1}, $\theta = 0$, $\varepsilon_\pm = 1$, $\boldsymbol{\mu}(\mathbf{x}) = \mathbf{I}$ for $\mathbf{x} \in \Omega$, and

$$\boldsymbol{\varepsilon}(\mathbf{x}) = \begin{cases} (1 + 10^{-9}i)\mathbf{I}, & g(x_1) < x_2 < H, \\ (-15 + 4i)\mathbf{I}, & -H < x_2 < g(x_1). \end{cases} \tag{9.5}$$

The case of complex coefficient functions can be handled using the technique of Lechleiter and Rittersbusch [7]. The small imaginary part in the relative permittivity in the portion of Ω above the grating was done for the sake of numerical stability of the marching method used to implement the RCWA algorithm [57].

When the RCWA algorithm [12, 35, 57] is used, the Gibbs phenomenon [84] arises prominently near the grating [59], especially in the spatial profile of E_1. This is exemplified by the plot of $|E_1(x_1, x_2)|$ vs. x_1/Λ for $x_2/H = 0$ in Fig. 9.1, obtained with $M = 10$ and $h = 1$ nm. Our hybrid coordinate-transformation method implemented with $M = 10$ and $h = 1$ nm delivered a very smooth profile, also shown in the same figure, the relative L^2 error (normalized using the hybrid solution) between the results from the two methods being $\simeq 4 \times 10^{-1}$. Thus, the hybrid method was able to eliminate the Gibbs phenomenon. The relative L^2 error decreases as $|x_2 - g(x_1)|$ increases in $\widetilde{\Omega} - \Omega^\dagger$, and the error becomes insignificant outside $\widetilde{\Omega}$.

For $x_2/H = 0$, the relative L^2 error between the hybrid solution and the FEM solution was approximately 5×10^{-6}. Thus, the FEM can deliver better results for the p-polarization state than the RCWA [59, 57], but the computational resources needed for the FEM solution are much more than for the RCWA since the FEM requires inverting large sparse matrices.

The hybrid method provides the solution in mapped domain $\widehat{\Omega}$ first and the original domain Ω second. $|\hat{H}_3(\hat{x}_1, \hat{x}_2)|$ is plotted vs. $\hat{x}_1/\Lambda$ and $\hat{x}_2/H$ in the left panel of Fig. 9.2, and $|H_3(\hat{x}_1, x_2)|$ is plotted vs. x_1/Λ and x_2/H in the right panel of Fig. 9.2, obtained with $M = 10$ and $h = 1$ nm. The grating is, of course, flat in $\widehat{\Omega}$ but sinusoidally undulating in Ω. The spatial profiles of $|H_3|$ are very different in the central region but become quite similar in the neighborhoods of $\Gamma_{\pm H}$, undoubtedly due to the properties of the function $\mathcal{S}(\hat{x}_2)$ of (9.1) at $\hat{x}_2 \in \{-H, 0, H\}$, as identified in Sec. 6.

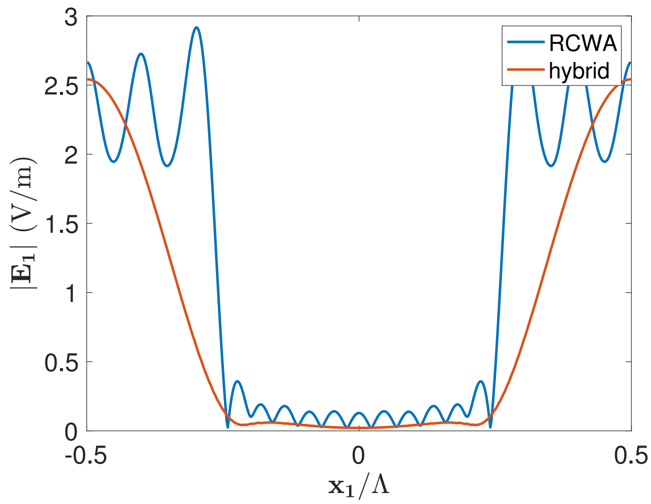


Fig. 9.1: $|E_1(x_1, x_2)|$ vs. x_1/Λ for the example problem when $x_2/H = 0$. Blue curve: RCWA solution calculated with $M = 10$ and $h = 1$ nm. Red curve: Hybrid solution calculated with $M = 10$ and $h = 1$ nm. See the paragraph containing (9.4) for other details.

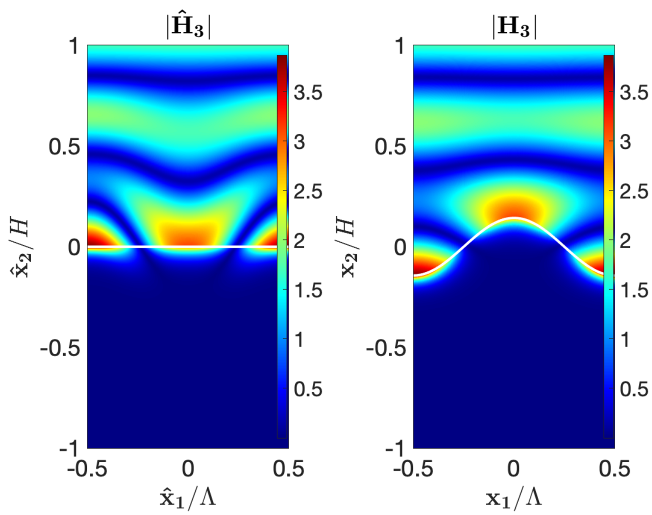
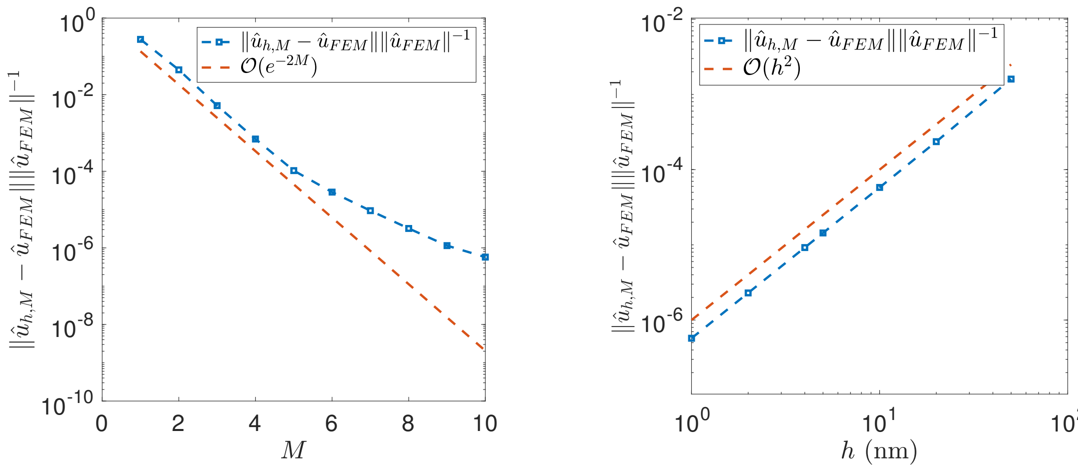


Fig. 9.2: Spatial profiles of (left) $|\hat{H}_3(\hat{x}_1, \hat{x}_2)|$ in $\widehat{\Omega}$ and (right) $|H_3(\hat{x}_1, x_2)|$ in Ω. See the paragraph containing (9.4) for other details. The grating is outlined in white in both domains.

Finally, we studied the convergence of the hybrid method with respect to both M and h, by comparing $\hat{u}_{h,M}$ with the FEM solution $\hat{u}_{FEM}$. The results of this study are shown in Fig. 9.3. Figure 9.3a shows the convergence rate of $\hat{u}_{h,M}$ with respect to the parameter M, $\hat{u}_{h,M}$ being calculated for $M \in \{1, 2, \cdots, 10\}$ with $h = 1$ nm fixed. We observe $\mathcal{O}(e^{-2M})$ convergence for $M \leq 5$ and $\mathcal{O}(e^{-M})$ convergence for $M \geq 6$. Theorem 8.3 predicts faster convergence for small enough M and slower convergence for large enough M, both trends being exemplified numerically in Figure 9.3a. Figure 9.3b shows the convergence rate of $\hat{u}_{h,M}$ with respect to the slice thickness h, calculations having been made for $h \in \{1, 2, 4, 5, 10, 20, 50\}$ nm with $M = 10$ fixed. We observe $\mathcal{O}(h^2)$ convergence.

Previous analysis has predicted slower convergence rates in terms of both M and h for the RCWA [57]. Therefore, a benefit of the hybrid method over the RCWA is the observed exponential convergence rate with respect to the parameter M.



(a) (b)

Fig. 9.3: Convergence plots comparing the hybrid solution with the FEM solution in the mapped domain. See the paragraph containing (9.4) for other details. (a) Semilog plot of relative L^2 error vs. M when $h = 1$ nm. (b) Loglog plot of relative L^2 error vs. h when $M = 10$.

10. Conclusion. In this paper, we formulated and analyzed a hybrid method for solving the time-harmonic Maxwell equations in a spatial domain that contains a grating. This method combines transformation optics with the RCWA. The domain is invariant in one dimension and the chosen constitutive properties allow the reduction of the full Maxwell system to a 2D Helmholtz equation for each linear polarization state, our focus lying on the p-polarization state. We proved the existence of a unique solution to the original scattering problem by establishing integral formulas similar to the Rellich identity to prove *a-priori* estimates. Our analysis required certain non-trapping conditions to hold. After implementing a coordinate transformation from the original domain to a mapped domain, we used norm equivalence to prove the uniqueness and existence of solutions to the transformed scattering problem. The discretized form of the transformed problem was analyzed for convergence analysis with respect to two different parameters: (i) a slice thickness indicative of spatial discretization in one dimension and (ii) the number of terms retained in the Fourier–Bloch expansions of the electric and magnetic field phasors with respect to the other dimension. Testing with a numerical example revealed faster convergence than our analysis predicts, which suggests future work to improve analysis. In contrast to RCWA, the hybrid method does not suffer from the Gibbs phenomenon. However, the hybrid method is applicable only to gratings that are the graphs of functions.

REFERENCES

[1] R. Petit (Ed.), *Electromagnetic Theory of Gratings*, Springer, Heidelberg, Germany (1980).
[2] D. Maystre (Ed.), *Selected Papers on Diffraction Gratings*, SPIE, Bellingham, WA, USA (1993).
[3] E.G. Loewen and E. Popov, *Diffraction Gratings and Applications*, Marcel Dekker, New York, USA (1997).
[4] P.W. Baumeister, *Optical Coating Technology*, SPIE, Bellingham, WA, USA (2004).
[5] V.A. Yakubovich and V.M. Starzhinskii, *Linear Differential Equations with Periodic Coefficients*, Wiley, New

York, NY, USA (1975).
[6] S.N. Chandler-Wilde, P. Monk, and M. Thomas, "The mathematics of scattering by unbounded, rough, inhomogeneous layers," *J. Comput. Appl. Math.* **204**(2), 549–559 (2007).
[7] A. Lechleiter and S. Ritterbusch, "A variational method for wave scattering from penetrable rough layers," *IMA J. Appl. Math.* **75**(3), 366–391 (2010).
[8] W. Dörfler, A. Lechleiter, M. Plum, G. Schneider, and C. Wieners, *Photonic Crystals: Mathematical Analysis and Numerical Approximation*, Birkhäuser, Basel, Switzerland (2011).
[9] K. Ohtaka, "Energy band of photons and low-energy photon diffraction," *Phys. Rev. B* **19**(10), 5057–5067 (1979).
[10] L. Solymar and D.J. Cooke, *Volume Holography and Volume Gratings*, Academic Press, London, United Kingdom (1981).
[11] F. Wang, A. Lakhtakia, and R. Messier, "Coupling of Rayleigh–Wood anomalies and the circular Bragg phenomenon in slanted chiral sculptured thin films," *Eur. Phys. J. Appl. Phys.* **20**(2), 91–103 (2002); errata: **24**(1), 91 (2003).
[12] J.A. Polo Jr., T.G. Mackay, and A. Lakhtakia, Electromagnetic Surface Waves: A Modern Perspective, Elsevier, Waltham, MA, USA (2013).
[13] M.W. McCall and A. Lakhtakia, "Coupling of a surface grating to a structurally chiral volume grating," *Electromagnetics* **23**(1), 1–26 (2003).
[14] A.E. Serebryannikov, T. Magath, and K. Schuenemann, "Bragg transmittance of *s*-polarized waves through finite-thickness photonic crystals with a periodically corrugated interface," *Phys. Rev. E* **74**(6), 066607 (2006).
[15] C. Palmer, *Diffraction Grating Handbook, 7th ed.*, Richardson Gratings, Rochester, NY, USA (2014).
[16] B.H. Kleemann, "Perfect blazing with echelle gratings in TE and TM polarization," *Opt. Lett.* **37**(6), 1002–1004 (2012).
[17] M. Memarian, X. Li, Y. Morimoto, and T. Itoh, "Wide-band/angle blazed surfaces using multiple coupled blazing resonances," *Sci. Rep.* **7**(1), 42285 (2017).
[18] A.A. Michelson, "The echelon spectroscope," *Astrophys. J.* **8**(1), 37–47 (1898).
[19] J.F. James and R.S. Sternburg, *The Design of Optical Spectrometers*, Chapman & Hall, London, United Kingdom (1969).
[20] C. Heine and R. H. Morf, "Submicrometer gratings for solar energy applications," *Appl. Opt.* **34**(14), 2476–2482 (1995).
[21] M. Jalali, H. Nadgaran, and D. Erni, "Semiperiodicity versus periodicity for ultra broadband optical absorption in thin-film solar cells," *J. Nanophoton.* **10**(3), 036018 (2016).
[22] M.V. Shuba and A. Lakhtakia, "Splitting of absorptance peaks in absorbing multilayer backed by a periodically corrugated metallic reflector," *J. Opt. Soc. Amer. A* **33**(4), 779–784 (2016).
[23] R.F. Millar, "The Rayleigh hypothesis and a related least-square solution to scattering problems for periodic surfaces and other scatterers," *Radio Sci.* **8**(8-9), 785–796 (1973).
[24] R.A. Depine and A. Lakhtakia, "Diffraction gratings of isotropic negative-phase velocity materials," *Optik* **116**(1), 31–43 (2005).
[25] Lord Rayleigh, "On the dynamical theory of gratings," *Proc. R. Soc. Lond. A* **79**(531), 399–416 (1907).
[26] R.A. Depine and A. Lakhtakia, "Diffraction by a grating made of a uniaxial dielectric–magnetic medium exhibiting negative refraction," *New J. Phys.* **7**(1), 158 (2005).
[27] A. Wirgin, "On Rayleigh's theory of sinusoidal diffraction gratings," *Optica Acta* **27**(12), 1671–1692 (1980).
[28] E. Popov and L. Mashev, "Convergence of Rayleigh–Fourier method and rigorous differential method for relief diffraction gratings — non-sinusoidal profile," *J. Modern Opt.* **34**(1), 155–158 (1987).
[29] K. Edee, J.P. Plumey, and J. Chandezon, "On the Rayleigh–Fourier method and the Chandezon method: Comparative study," *Opt. Commun.* **286**(1), 31–41 (2013).
[30] P.C. Waterman, "Scattering by periodic surfaces," *J. Acoust. Soc. Amer.* **57**(4), 791–802 (1975).
[31] S.-L. Chuang and J.A. Kong, "Scattering of waves from periodic surfaces," *Proc. IEEE* **69**(9), 1132–1144 (1981).
[32] A. Lakhtakia, V.V. Varadan, and V.K. Varadan, "Scattering by periodic achiral-chiral interfaces," *J. Opt. Soc. Amer. A* **6**(11), 1675–1683 (1989).
[33] M.G. Moharam and T.K. Gaylord, "Rigorous coupled-wave analysis of planar grating diffraction," *J. Opt. Soc. Amer.* **71**(7), 811–818 (1981).
[34] M.G. Moharam, D.A. Pommet, E.B. Grann, and T.K. Gaylord, "Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach," *J. Opt. Soc. Amer. A* **12**(5), 1077–1086 (1995).
[35] P. Lalanne and G.M. Morris, "Highly improved convergence of the coupled-wave method for TM polarization," *J. Opt. Soc. Amer. A* **13**(4), 779–784 (1996).
[36] J. Chandezon, G. Raoult, and D. Maystre, "A new theoretical method for diffraction gratings and its numerical application," *J. Opt. (France)* **11**(4), 235–241 (1980).
[37] J. Chandezon, M.T. Dupuis, G. Cornet, and D. Maystre, "Multicoated gratings: a differential formalism applicable in the entire optical region," *J. Opt. Soc. Amer.* **72**(7), 839–846 (1982).
[38] R.A. Depine, M.E. Inchaussandague, and A. Lakhtakia, "Application of the differential method to uniaxial gratings with an infinite number of refraction channels: Scalar case," *Opt. Commun.* **258**(2), 90–96 (2006).
[39] Z. Chen and 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),799–826 (2003).
[40] D.R. Smith, P.M. Rye, J.J. Mock, D.C. Vier, and A.F. Starr, "Enhanced diffraction from a grating on the surface

of a negative-index metamaterial," *Phys. Rev. Lett.* **93**(13), 137405 (2004).
[41] M.E. Solano, G.D. Barber, A. Lakhtakia, M. Faryad, P.B. Monk, and T.E. Mallouk, "Buffer layer between a planar optical concentrator and a solar cell," *AIP Advances* **5**(9), 097150 (2015).
[42] J. Song, J.-J. He, and S. He, "Polarization performance analysis of etched diffraction grating demultiplexer using boundary element method," *IEEE J. Sel. Top. Quantum Electron.* **11**(1), 224–231 (2005).
[43] S.A. Stewart, S. Moslemi-Tabrizi, T.J. Smy, and S. Gupta, "Scattering field solutions of metasurfaces based on the boundary element method for interconnected regions in 2-D," *IEEE Trans. Antennas Propagat.* **67**(12), 7487–7495 (2019).
[44] H. Ichigawa, "Electromagnetic analysis of diffraction gratings by the finite-difference time-domain method," *J. Opt. Soc. Amer. A* **15**(1), 152–157 (1998).
[45] C. Oh and M.J. Escuti, "Numerical analysis of polarization gratings using the finite-difference time-domain method," *Phys. Rev. A* **76**(4), 043815 (2007).
[46] Z.L. Yang, Q.H. Li, F.X. Ruan, Z.P. Li, B. Ren, H.X. Xu, and Z.Q. Tian, "FDTD for plasmonics: applications in enhanced Raman spectroscopy," *Chin. Sci. Bull.* **55**(24), 2635–2642 (2010).
[47] N.L. Tsitsas, N.K. Uzunoglu, and D.I. Kaklamani, "Diffraction of plane waves incident on a grated dielectric slab: An entire domain integral equation analysis," *Radio Sci. 42*(6), RS6S22 (2007).
[48] N. Chamanara, K. Achouri, and C. Caloz, "Efficient analysis of metasurfaces in terms of spectral-domain GSTC integral equations," *IEEE Trans. Antennas Propagat.* **65**(10), 5340–5347 (2017).
[49] M.G. Moharam, E.B. Grann, D.A. Pommet, and T.K. Gaylord, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," *J. Opt. Soc. Amer. A* **12**(5), 1068–1076 (1995).
[50] L. Li, "Use of Fourier series in the analysis of discontinuous periodic structures," *J. Opt. Soc. Amer. A* **13**(9), 1870–1876 (1996).
[51] F. Bloch, "Über die Quantenmechanik der Elektronen in Kristallgittern," *Z. Phys. A* **52**(7-8), 555–600 (1929).
[52] J. Gazalet, S. Duponta, J.C. Kastelik, Q. Rolland, and B. Djafari-Rouhani, "A tutorial survey on waves propagating in periodic media: Electronic, photonic and phononic crystals. Perception of the Bloch theorem in both real and Fourier domains," *Wave Motion* **50**(3), 619–654 (2013).
[53] A. Lechleiter, "The Floquet–Bloch transform and scattering from locally perturbed periodic surfaces," *J. Math. Anal. Appl.* **446**(1), 605–627 (2017).
[54] G. Floquet, "Sur les équations différentielles à coefficients périodiques," *Ann. Sci. l'Ecole Norm. Super., 2nd Series* **12**, 47–88 (1883).
[55] J.J. Hench and Z. Strakoš, "The RCWA method – a case study with open questions and perspectives of algebraic computations," *Electron. Trans. Numer. Anal.* **31**, 331–357 (2008).
[56] B.J. Civiletti, A. Lakhtakia, and P.B. Monk, "Analysis of the rigorous coupled wave approach for s-polarized light in gratings," *J. Comput. Appl. Math.* **368**(1), 112478 (2020).
[57] B.J. Civiletti, A. Lakhtakia, and P.B. Monk, "Analysis of the rigorous coupled wave approach for p-polarized light in gratings," *J. Comput. Appl. Math.* **386**(1), 113235 (2021).
[58] T. Schuster, J. Ruoff, N. Kerwien, S. Rafler, and W. Osten, "Normal vector method for convergence improvement using the RCWA for crossed gratings," *J. Opt. Soc. Amer. A* **24**(9), 2880–2890 (2007).
[59] M.V. Shuba, M. Faryad, M.E. Solano, P.B. Monk, and A. Lakhtakia, *J. Opt. Soc. Amer. A* **32**(7), 1222–1230 (2015).
[60] E. Hewitt and R.E. Hewitt, "The Gibbs–Wilbraham phenomenon: An episode in Fourier analysis," *Arch. Hist. Exact Sci.* **21**(2), 129–160 (1979).
[61] J. Plebanski, "Electromagnetic waves in gravitational fields," *Phys. rev.* **118**(5), 1396–1408 (1960).
[62] T.G. Mackay, A. Lakhtakia, and S. Setiawan, "Gravitation and electromagnetic wave propagation with negative phase velocity," *New J. Phys.* **7**(1), 75 (2005).
[63] A.V. Kildishev and V.M. Shalaev, "Transformation optics and metamaterials," *Phys.–Usp.* **54**(1), 53–63 (2011).
[64] J. Bischoff, "Improved diffraction computation with a hybrid C-RCWA-method," *Proc. SPIE* **7272**(1), 72723Y (2009).
[65] I.G. Graham, O.R. Pembery, and E.A. Spence, "The Helmholtz equation in heterogeneous media: A priori bounds, well-posedness, and resonances," *J. Diff. Eqns.* **266**(6), 2869–2923 (2019).
[66] I. Babuška, "Error-bounds for finite element method," *Numer. Math.* **16**(4), 322–333 (1971).
[67] F. Brezzi, "On the existence, uniqueness and approximation of saddle-point problems arising from lagrangian multipliers," *Revue Française d'Automatique, Informatique, Recherche Opérationnelle. Analyse Numérique* **8**(2), 129–151 (1974).
[68] T.J.R. Hughes, L.P. Franca, and M. Balestra, "A new finite element formulation for computational fluid dynamics: V. Circumventing the Babuška–Brezzi condition: A stable Petrov–Galerkin formulation of the Stokes problem accommodating equal-order interpolations," *Computer Meth. Appl. Mech. Eng.* **59**(1), 85–99 (1986); errata: **62**(1), 111 (1987).
[69] J.B. Keller and D. Givoli, "Exact non-reflecting boundary conditions," *J. Comput. Phys.* **82**(1), 172–192 (1989).
[70] T.J.R. Hughes, "Multiscale phenomena: Green's functions, the Dirichlet-to- Neumann formulation, subgrid-scale models, bubbles and the origins of stabilized methods," *Computer Meth. Appl. Mech. Eng.* **127**(1-4), 387–401 (1995).
[71] A.E. Heins and S. Silver, "The edge conditions and field representation theorems in the theory of electromagnetic diffraction," *Math. Proc. Cambridge Phil. Soc.* **51**(1), 149–161 (1955).

[72] S.H. Schot, “Eighty years of Sommerfeld’s radiation condition,” *Historia Math.* **19**(4), 385–401 (1992).
[73] L.C. Evans, *Partial Differential Equations*, American Mathematical Society, Providence, RI, USA (1998).
[74] R.W. Wood, “On a remarkable case of uneven distribution of light in a diffraction grating spectrum,” *Proc. Phys. Soc. Lond.* **18**(1), 269?275 (1902).
[75] H.A. Schwarz, “Über ein Flächen kleinsten Flächeninhalts betreffendes Problem der Variationsrechnung,” *Acta Societatis Scientiarum Fennicae* **15**, 318–361 (1888).
[76] A. Bonito, J.-L. Guermond, and F. Luddens, “Regularity of the Maxwell equations in heterogeneous media and Lipschitz domains,” *J. Math. Anal. Appl.* **408**(2), 498–512 (2013).
[77] N. Heuer, “On the equivalence of fractional-order Sobolev semi-norms,” *J. Math. Anal. Appl.* **417**(2), 505–518 (2014).
[78] R. Kress, *Linear Integral Equations*, Springer, New York, NY, USA (1999).
[79] A.H. Schatz, “An observation concerning Ritz–Galerkin methods with indefinite bilinear forms,” *Math. Comp.* **28**(128), 959–962 (1974).
[80] T.G. Mackay and A. Lakhtakia, *The Transfer-Matrix Method in Electromagnetics and Optics*, Morgan & Claypool, San Rafael, CA, USA (2020).
[81] P.B. Monk, *Finite Element Methods for Maxwell’s Equations*, Oxford University Press, Oxford, United Kingdom (2003).
[82] L.R. Scott and S. Brenner, *The Mathematical Theory of Finite Element Methods*, Springer, New York, USA (2008).
[83] J. Schöberl, Netgen/NGSolv, https://ngsolve.org, 5 May 2020.
[84] T.W. Körner, *Fourier Analysis*, Chap. 17, Cambridge University Press, Cambridge, United Kingdom (1988).
[85] D.P. Nicholls, and F.Reitich, “Boundary perturbation methods for high-frequency acoustic scattering: Shallow periodic gratings,” *J. Acoust. Soc. Am.* **123**, 2531–2541 (2008).

Acknowledgments. This work was supported in part by the US National Science Foundation under Grant Nos. DMS-2011603 and DMS-2011996. AL thanks the Charles Godfrey Binder Endowment at Penn State for ongoing support of his research activities.