Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 0946 – 8633

Convergence analysis of the FEM coupled with Fourier-mode expansion for the electromagnetic scattering by biperiodic structures

Guanghui Hu, Andreas Rathsfeld

submitted: November 26, 2012

Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: guanghui.hu@wias-berlin.de
andreas.rathsfeld@wias-berlin.de

No. 1744
Berlin 2012

2010 *Mathematics Subject Classification.* 78A45, 78M10, 65N30, 35J20.

Key words and phrases. Electromagnetic scattering, diffraction gratings, convergence analysis, finite element methods, mortar technique.

The first author gratefully acknowledges the support by the German Research Foundation (DFG) under Grant No. HU 2111/1-1. Special thanks are given to the Weierstrass Institute and the Research Group "Nonlinear Optimization and Inverse Problems" for the support of the work.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 2044975
E-Mail: `preprint@wias-berlin.de`
World Wide Web: `http://www.wias-berlin.de/`

Abstract

Scattering of time-harmonic electromagnetic plane waves by a doubly periodic surface structure in $\mathbb{R}^3$ can be simulated by a boundary value problem of the time-harmonic curl-curl equation. For a truncated FEM domain, non-local boundary value conditions are required in order to satisfy the radiation conditions for the upper and lower half spaces. Alternatively to boundary integral formulations, to approximate radiation conditions and absorbing boundary methods, Huber et al. [11] have proposed a coupling method based on an idea of Nitsche. In the case of profile gratings with perfectly conducting substrate, the authors have shown previously that a slightly modified variational equation can be proven to be equivalent to the boundary value problem and to be uniquely solvable. Now it is shown that this result can be used to prove convergence for the FEM coupled by truncated wave mode expansion. This result covers transmission gratings and gratings bounded by additional multi-layer systems.

1 Introduction

The diffraction of light by biperiodic gratings, e.g., by doubly periodic surface structures can be simulated by the time-harmonic Maxwell equations. Eliminating the magnetic field, the electric field is the solution of a boundary value problem for the time-harmonic curl-curl equation. For finite element methods (FEM), this problem is reduced to a finite domain, where quasi-periodic lateral boundary conditions and non-local boundary conditions over the upper and lower boundary face are required. The first idea for the solution of the boundary value problem is to express the non-local boundary conditions by integral operators and to couple FEM with boundary elements (cf. [7, 14]). With this approach, for the solution of the boundary value problem, either the case of wave modes propagating parallel to the surface is to be excluded or standard methods for integral operators with non-trivial null space are to be applied. Alternatively to integral operators, a saddle point type formulation (cf. e.g. [1]) or absorbing boundary conditions (cf. e.g. [20]) can be used.

On the other hand, the radiation conditions mean that the functions can be extended in form of a Rayleigh series expansion of upward resp. downward radiating Fourier modes. So the idea to couple finite elements and Rayleigh expansions is natural. Huber et al. [11] propose such a method, where the finite elements and the Rayleigh series are coupled employing a mortar technique of Nitsche (cf. [16, 23]). In [10], the case of perfectly conducting profile gratings has been considered and the coupling terms of [11] have been slightly modified. It has been proved that the variational equation for the coupling of FEM and Rayleigh expansions is equivalent to the boundary value problem for the scattering by gratings. If the last problem is uniquely solvable, then the operator of the variational equation is uniquely solvable too.

The subject of the present paper is first to generalize the results on the variational formulation of [10] to the case of transmission gratings and, second, to analyze the corresponding numerical scheme, i.e., the coupling of FEM with truncated Rayleigh series expansions. We formulate the boundary value problem and some solvability results in Sect. 3. In Sect. 4 we define the variational form and derive the Fredholm property for the operator corresponding to this form. The numerical discretization of the variational equation is introduced in Sect. 5. The stability and convergence of this method is proved. Of course, edge elements (cf. e.g. [13]) are employed for the FEM. In Sect. 6 we discuss the case of multi-layer systems beneath the grating structure. Instead of an extention of the FEM domain by the layers of the multi-layer system, we replace the downgoing Fourier modes by special wave modes of the multi-layer system. Note that this idea goes back to the authors of [11]. The convergence analysis of Sect. 5 can be generalized to the multi-layer case too. Finally, we add a simple test showing that our method converges to the same solution as the 2D FEM for periodic 2D gratings and to the same solution as the method of [11].

2 Preliminaries

Throughout the paper, the symbols e_j $(j = 1, 2, 3)$ denote the unit coordinate vectors in the three dimensional Cartesian coordinate system. The symbol $(\cdot)^\top$ denotes the transpose of a vector in $\mathbb{C}^2$ or $\mathbb{C}^3$, while the symbol **a**$\perp$**b** means the orthogonality of the vectors **a** $= (a_1, a_2, a_3)$, **b** $= (b_1, b_2, b_3) \in \mathbb{C}^3$ in the sense that $\sum_{j=1}^{3} a_j b_j = 0$. Denote the unit sphere by $\mathbb{S}^2 := \{x = (x_1, x_2, x_3)^\top \in \mathbb{R}^3 : ||x|| = 1\}$, and define $x' := (x_1, x_2)$ for $x \in \mathbb{R}^3$. The branch of the square root $\sqrt{a}$ is chosen such that the imaginary part of $\sqrt{a}$ is always positive, i.e. $\sqrt{a} = i\sqrt{-a}$ if $a < 0$.

3 Diffraction problem

Consider the scattering of a time-harmonic electromagnetic plane wave by a biperiodic structure (diffraction grating) which consists of at least two optical materials. By biperiodic or doubly periodic structure (cf. Fig. 1), we mean that the structure is periodic in two orthogonal directions x_1 and x_2 and bounded in x_3. The optical material inside the grating can be completely characterized by its dielectric coefficient and its magnetic permeability. For simplicity we assume that the medium is nonmagnetic with a constant magnetic permeability $\mu(x) = \mu_0 > 0$ in $\mathbb{R}^3$. However, our arguments can be adapted to the case where $\mu(x)$ is a periodic and piecewise constant function. The electric permittivity $\epsilon(x)$ and the conductivity $\sigma(x)$ are supposed to be Λ_j-periodic in x_j $(j = 1, 2)$ inside the grating and are homogeneous above and below the grating structure. More precisely, we assume that there exists a constant $b > 0$ such that

$$\begin{aligned} \epsilon(x_1 + n_1\Lambda_1, x_2 + n_2\Lambda_2, x_3) &= \epsilon(x_1, x_2, x_3), \\ \sigma(x_1 + n_1\Lambda_1, x_2 + n_2\Lambda_2, x_3) &= \sigma(x_1, x_2, x_3) \end{aligned}$$

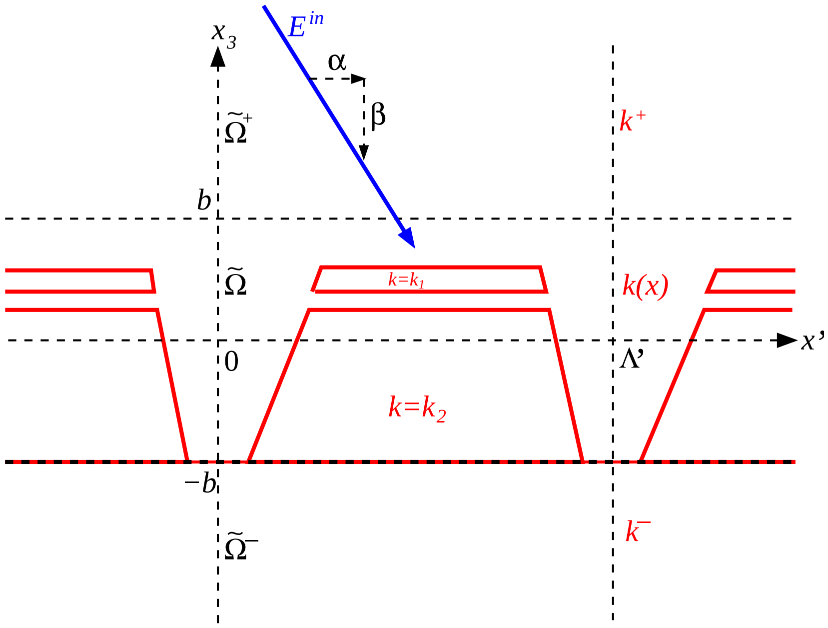

Figure 1: Geometry of grating.

in $\widetilde{\Omega} := \{x : |x_3| < b\}$ for any $n = (n_1, n_2) \in \mathbb{Z}^2$, and

$$\begin{aligned} \epsilon(x) &= \epsilon_0^+ > 0,\ \sigma(x) = 0 \quad \text{in} \quad x_3 > b,\\ \epsilon(x) &= \mathrm{Re}\,\epsilon_0^-,\ \sigma(x) = \omega\mathrm{Im}\,\epsilon_0^- > 0 \quad \text{in} \quad x_3 < -b \end{aligned}$$

with the circular frequence $\omega > 0$. Further, we restrict ourselves to the mostly used gratings, where $\epsilon(x)$ and $\sigma(x)$ are piecewise constant functions satisfying

$$0 < \epsilon_0 < \epsilon(x) < \infty, \quad 0 \le \sigma(x) < \infty \quad \text{in } \mathbb{R}^3. \tag{3.1}$$

Set $\widetilde{\Omega}^\pm := \{x : x_3 \gtrless \pm b\}$. Suppose that a time-harmonic electromagnetic plane wave $E^{in}(x)e^{-i\omega t}$ with E^{in} of the form

$$E^{in}(x) := q\exp(ik^+ x \cdot \hat{\theta}) = q\exp\Big(i(x' \cdot \alpha - \beta x_3)\Big), \quad i := \sqrt{-1} \tag{3.2}$$

is incident on the grating from $\widetilde{\Omega}^+$. Here $k^+ := \omega\sqrt{\epsilon_0^+ \mu_0}$ (resp. $k^- := \omega\sqrt{\epsilon_0^- \mu_0}$) is defined as the wavenumber characterizing the homogenous medium in $\widetilde{\Omega}^+$ (resp. $\widetilde{\Omega}^-$). In (3.2), the symbol $\hat{\theta}$ denotes the direction of incidence

$$\hat{\theta} \quad := \quad (\sin\theta_1\cos\theta_2, \sin\theta_1\sin\theta_2, -\cos\theta_1)^\top \in \mathbb{S}^2,$$

with the incident angles $\theta_1 \in [0, \pi/2)$, $\theta_2 \in [0, 2\pi)$. Further, in (3.2), the three dimensional vector $q = (q_1, q_2, q_3)^\top \in \mathbb{S}^2$ stands for the direction of polarization satisfying $q \perp \hat{\theta}$, and

$$\alpha = (\alpha_1, \alpha_2)^\top := k(\sin\theta_1\cos\theta_2, \sin\theta_1\sin\theta_2)^\top \in \mathbb{R}^2,\ \beta := k\cos\theta_1.$$

Eliminating the magnetic field from the reduced time-harmonic Maxwell's equations, we end up with the electric curl-curl equation

$$\mathrm{curl}\,\mathrm{curl}\,E(x) - k^2(x)E(x) = 0 \quad \text{for} \quad x \in \mathbb{R}^3, \tag{3.3}$$

where $k^2(x) := \omega^2\mu_0(\epsilon(x) + i\sigma(x)/\omega)$ and the electric field E in $\widetilde{\Omega}^+$ is the sum of the incident field E^{in} and the scattered field E^{sc}. The periodicity of the grating together with the form of E^{in} motivates us to look for α-quasiperiodic solutions in the sense that $E(x)\exp(-i\alpha \cdot x')$ is (Λ_1, Λ_2)-periodic in x'. In other words, it is required that

$$\begin{aligned} E(x_1 + \Lambda_1, x_2, x_3) &= \exp(i\Lambda_1\alpha_1)E(x_1, x_2, x_3), \\ E(x_1, x_2 + \Lambda_2, x_3) &= \exp(i\Lambda_2\alpha_2)E(x_1, x_2, x_3) \end{aligned} \tag{3.4}$$

for all $x \in \mathbb{R}^3$. Since the domain is unbounded in the x_3-direction, a radiation condition must be imposed. Noting that $k(x) = k^\pm$ in $\widetilde{\Omega}^\pm$, we suppose that the scattered field E^{sc} in $\widetilde{\Omega}^+$ and the electric field E in $\widetilde{\Omega}^-$ are composed of bounded outgoing plane waves in the form of

$$\begin{aligned} E^{sc}(x) &= \sum_{n\in\mathbb{Z}^2} E_n^+ \exp\Big(i(\alpha_n \cdot x' + \beta_n^+ x_3)\Big) \quad \text{for } x_3 > b, \;\; E_n^+ \perp (\alpha_n, \beta_n^+)^\top, \\ E(x) &= \sum_{n\in\mathbb{Z}^2} E_n^- \exp\Big(i(\alpha_n \cdot x' - \beta_n^- x_3)\Big) \quad \text{for } x_3 < -b, \;\; E_n^- \perp (\alpha_n, -\beta_n^-)^\top, \end{aligned} \tag{3.5}$$

where $\alpha_n := (\alpha_n^{(1)}, \alpha_n^{(2)}) \in \mathbb{R}^2$, with $\alpha_n^{(j)} = \alpha_j + 2\pi n_j/\Lambda_j, j = 1, 2$ for $n = (n_1, n_2)^\top \in \mathbb{Z}^2$, and

$$\beta_n^\pm = \beta_n^\pm(k^\pm, \alpha) := \sqrt{(k^\pm)^2 - |\alpha_n|^2}.$$

We say that the scattered fields satisfy the radiation condition if expansions of the form (3.5) exist. These expansions are also referred to as the Rayleigh series expansions. The constant vectors $E_n^\pm$ are called Rayleigh coefficients. Since $\beta_n^\pm$ are real-valued only for finitely many indices n, we observe that only a finite number of wave modes in (3.5) propagate into the far field, while the remaining part consists of evanescent (or surface) waves decaying exponentially as $x_3 \to \pm\infty$. Thus, the above expansion for E^{sc} resp. E converges uniformly with all derivatives in the half space $\{x_3 > a\}$ resp. $\{x_3 < -a\}$ for any $a > b$.

Since the squared wavenumber $k^2(x)$ is (Λ_1, Λ_2)-periodic in x' and both the incident and scattered fields are quasiperiodic, we can reduce the scattering problem to a single periodic cell. To this end, we introduce the following notation

$$\begin{aligned} \tilde{\Gamma}_b^\pm &:= \Big\{(x_1, x_2, x_3)^\top \in \mathbb{R}^3 \colon x_3 = \pm b\Big\}, \\ \Gamma_b^\pm &:= \Big\{(x_1, x_2, x_3)^\top \in \tilde{\Gamma}_b^\pm \colon 0 < x_j < \Lambda_j, j = 1, 2, \Big\}, \\ \Omega^\pm &:= \Big\{(x_1, x_2, x_3)^\top \in \widetilde{\Omega}^\pm \colon 0 < x_j < \Lambda_j, j = 1, 2\Big\}, \\ \Omega &:= \Big\{x \in \widetilde{\Omega} \colon 0 < x_j < \Lambda_j, j = 1, 2\}. \end{aligned}$$

We next introduce some scalar and vector valued α-quasiperiodic Sobolev spaces. Let $H^s(\tilde{\Gamma}_b^\pm)$ be the complex valued L^2-based Sobolev spaces of order s over $\tilde{\Gamma}_b^\pm$. Write

$$\begin{aligned}
H_{loc}(\mathrm{curl}\,,\widetilde{\Omega}) &:= \Big\{G\colon \chi G, \mathrm{curl}\,(\chi G) \in L^2(\widetilde{\Omega})^3, \forall \chi \in C_0^\infty(\mathbb{R}^3)\Big\},\\
H^s_{loc}(\tilde{\Gamma}_b^\pm) &:= \Big\{G\colon \chi G \in H^s(\tilde{\Gamma}_b^\pm), \forall \chi \in C_0^\infty(\tilde{\Gamma}_b^\pm)\Big\},\\
H^s_{t,loc}(\tilde{\Gamma}_b^\pm) &:= \Big\{G \in H^s_{loc}(\tilde{\Gamma}_b^\pm)^3\colon e_3 \cdot G = 0\Big\},\\
H^s_{t,loc}(\mathrm{Div}\,,\tilde{\Gamma}_b^\pm) &:= \Big\{G\colon G \in H^s_{t,loc}(\tilde{\Gamma}_b^\pm),\ \mathrm{Div}\,G \in H^s_{t,loc}(\tilde{\Gamma}_b^\pm)\Big\},\\
H^s_{t,loc}(\mathrm{Curl}\,,\tilde{\Gamma}_b^\pm) &:= \Big\{G\colon G \in H^s_{t,loc}(\tilde{\Gamma}_b^\pm),\ \mathrm{Curl}\,G \in H^s_{t,loc}(\tilde{\Gamma}_b^\pm)\Big\},
\end{aligned}$$

and

$$\begin{aligned}
H(\mathrm{curl}\,,\Omega) &:= \Big\{G|_\Omega\colon G \in H_{loc}(\mathrm{curl}\,,\widetilde{\Omega}), G \text{ is } \alpha\text{-quasiperiodic}\Big\},\\
H^s_{qp}(\Omega) &:= \Big\{g|_\Omega\colon g \in H^s_{loc}(\widetilde{\Omega}), g \text{ is } \alpha\text{-quasiperiodic}\Big\},\\
H^s_t(\Gamma_b^\pm) &:= \Big\{G|_{\Gamma_b^\pm}\colon G \in H^s_{t,loc}(\tilde{\Gamma}_b^\pm),\ G \text{ is } \alpha\text{-quasiperiodic}\Big\},\\
H^s_t(\mathrm{Div}\,,\Gamma_b^\pm) &:= \Big\{G|_{\Gamma_b^\pm}\colon G \in H^s_{t,loc}(\mathrm{Div}\,,\tilde{\Gamma}_b^\pm),\ G \text{ is } \alpha\text{-quasiperiodic}\Big\},\\
H^s_t(\mathrm{Curl}\,,\Gamma_b^\pm) &:= \Big\{G|_{\Gamma_b^\pm}\colon G \in H^s_{t,loc}(\mathrm{Curl}\,,\tilde{\Gamma}_b^\pm),\ G \text{ is } \alpha\text{-quasiperiodic}\Big\},
\end{aligned}$$

where $\mathrm{Div}\,(\cdot)$ and $\mathrm{Curl}\,(\cdot)$ stand for the surface divergence and the surface scalar rotational operators, respectively. Note that, for $x' \mapsto E(x', \pm b)$ in $H^s_t(\Gamma_b^\pm)$, $s \in \mathbb{R}$, we have the Fourier series expansion

$$\begin{aligned}
E(x', \pm b) &= \sum_{n\in\mathbb{Z}^2} E_n^\pm \exp(i\alpha_n \cdot x'),\\
E_n^\pm &:= (\Lambda_1\Lambda_2)^{-1} \int_0^{\Lambda_1}\int_0^{\Lambda_2} E(x', \pm x_3) \exp(-i\alpha_n \cdot x')\mathrm{d}x_1\mathrm{d}x_2 \in \mathbb{C}^3.
\end{aligned}$$

Then, the spaces $H^s_t(\Gamma_b^\pm)$, $H^s_t(\mathrm{Div}\,,\Gamma_b^\pm)$ and $H^s_t(\mathrm{Curl}\,,\Gamma_b^\pm)$ can be equipped with the following equivalent Sobolev norms

$$\begin{aligned}
||E||_{H^s_t(\Gamma_b^\pm)} &= \left(\sum_{n\in\mathbb{Z}^2} |E_n^\pm|^2\left(1+|\alpha_n|^2\right)^s\right)^{1/2},\\
||E||_{H^s_t(\mathrm{Div}\,,\Gamma_b^\pm)} &= \left(\sum_{n\in\mathbb{Z}^2} \left(1+|\alpha_n|^2\right)^s\left(|E_n^\pm|^2 + |E_n^\pm\cdot(\alpha_n,0)^\top|^2\right)\right)^{1/2},\\
||E||_{H^s_t(\mathrm{Curl}\,,\Gamma_b^\pm)} &= \left(\sum_{n\in\mathbb{Z}^2} \left(1+|\alpha_n|^2\right)^s\left(|E_n^\pm|^2 + |E_n^\pm\times(\alpha_n,0)^\top|^2\right)\right)^{1/2}.
\end{aligned}$$

Recall that the space dual to $H_t^s(\mathrm{Div}\,,\Gamma_b^\pm)$ w.r.t. the L^2-scalar product is $H_t^s(\mathrm{Div}\,,\Gamma_b^\pm)' = H_t^{-s-1}(\mathrm{Curl}\,,\Gamma_b^\pm)$, and that, for $s=-1/2$,

$$\begin{aligned} H_t^{-1/2}(\mathrm{Div}\,,\Gamma_b^\pm) &= \Big\{(e_3\times E)|_{\Gamma_b^\pm}\colon\ E\in H(\mathrm{curl}\,,\Omega)\Big\},\\ H_t^{-1/2}(\mathrm{Curl}\,,\Gamma_b^\pm) &= \Big\{(e_3\times E)|_{\Gamma_b^\pm}\times e_3\colon\ E\in H(\mathrm{curl}\,,\Omega)\Big\}. \end{aligned}$$

Further, the trace mappings from $H(\mathrm{curl}\,,\Omega)$ to the tangential spaces $H_t^{-1/2}(\mathrm{Div}\,,\Gamma_b^\pm)$ and $H_t^{-1/2}\,(\mathrm{Curl}\,,\Gamma_b^\pm)$ are continuous and surjective (see [6, 13] and the references there). Finally, define our variational space

$$X = X_b := \Big\{E:\Omega\to\mathbb{C}^3\colon E\in H(\mathrm{curl}\,,\Omega)\Big\}$$

endowed with the norm

$$||E||_X := ||E||_{H(\mathrm{curl}\,,\Omega)} = \Big(||E||^2_{L^2(\Omega)^3} + ||\mathrm{curl}\,E||^2_{L^2(\Omega)^3}\Big)^{1/2}.$$

The boundary value problem for our scattering problem can be stated as follows.

(BVP): Given an incident electric field E^{in}, determine the quasiperiodic total electric field $E\in H_{loc}(\mathrm{curl}\,,\mathbb{R}^3)$ such that $E(x)|_\Omega$ satisfies the curl-curl equation (3.3) in Ω in the distributional sense and that the scattered field $E^{sc}=E-E^{in}$ in $x_3>b$ as well as the transmitted field E in $x_3<-b$ admit a Rayleigh expansion of the form (3.5).

Introduce the set

$$\Upsilon_{\mathrm{res}} := \Upsilon^+_{\mathrm{res}}\cup\Upsilon^-_{\mathrm{res}},\quad \Upsilon^\pm_{\mathrm{res}} := \Big\{n\in\mathbb{Z}^2\colon\ \beta_n^\pm(k^\pm,\alpha)=0\Big\}. \tag{3.6}$$

An incident angular frequency ω with $\Upsilon_{\mathrm{res}}\neq\emptyset$ is called Rayleigh frequency. Note that the set $\mathcal{F}$ of all Rayleigh frequencies depends on $k^\pm$, Λ_1 and Λ_2 but not on the shape of Γ.

Below we collect some uniqueness and existence results of (BVP) for a broad class of incident plane waves. Assume that the incident electric wave takes the form

$$E^{in}_{gen} := \sum_{n\colon\beta_n>0} Q_n\exp\Big(\alpha_n\cdot x' - \beta_n x_3\Big), \tag{3.7}$$

where $Q_n\in\mathbb{C}^3$ satisfies $Q_n\perp(\alpha_n,-\beta_n)^\top$. Note that E^{in} of (3.2) is of the form (3.7), where $Q_n=q$ for $n=(0,0)^\top$ and $Q_n=(0,0,0)^\top$ else.

Theorem 3.1. *Consider the scattering problem (BVP) with E^{in} replaced by E^{in}_{gen}.*

(i) *There exists a unique solution to (BVP) for all $\omega\in\mathbb{R}^+\backslash\mathcal{D}$, where $\mathcal{D}\supset\mathcal{F}$ is a discrete set with the only accumulating point at infinity.*

(ii) *The problem (BVP) admits at least one solution for any $\omega\in\mathbb{R}^+$. Moreover, the far-field part of the solution scattered into the half space $x_3\gtrless\pm b$ is unique, i.e., the Rayleigh coefficients of the plane wave modes propagating into the half space $x_3\gtrless\pm b$ (namely, those $E_n^\pm$ with $\beta_n^\pm>0$) are unique.*

(iii) *There exists a small frequency $\omega_0 > 0$ such that the problem (BVP) admits a unique solution for all $\omega \in (0, \omega_0]$.*

The assertions (i) and (ii) follow from the existence and uniqueness of the magnetic field in the space $H^1(\Omega)^3$ (see [4, 5, 8, 21, 22]). Note that the constant magnetic permeability implies the piecewise H^1-regularity of the magnetic field, which is not true for the electric field. In the non-resonance case (i.e. $\Upsilon_{\rm res} = \emptyset$), (i) and (ii) can also be proved by studying the following variational formulation for the electric field E in Ω: find $E \in X$ such that

$$\begin{aligned}\int_\Omega \left[\operatorname{curl} E \cdot \operatorname{curl} \overline{\varphi} - k^2(x) E \cdot \overline{\varphi}\right] \mathrm{d}x - \int_{\Gamma_b^+} \mathcal{R}^+(e_3 \times E) \cdot (e_3 \times \overline{\varphi}) \mathrm{ds} \\ + \int_{\Gamma_b^-} \mathcal{R}^-(e_3 \times E) \cdot (e_3 \times \overline{\varphi}) \mathrm{ds} \\ = \int_{\Gamma_b^+} \left[(\operatorname{curl} E^{in})_T - \mathcal{R}^+(e_3 \times E^{in})\right] \cdot (e_3 \times \overline{\varphi}) \mathrm{ds} \qquad (3.8)\end{aligned}$$

for all $\varphi \in X$, where $(\cdot)_T := [e_3 \times (\cdot)]|_{\Gamma_b^+} \times e_3$ and the operators $\mathcal{R}^\pm$: $H_t^{-1/2}(\mathrm{Div}\,, \Gamma_b^\pm) \to H_t^{-1/2}(\mathrm{Curl}\,, \Gamma_b^\pm)$ are the Dirichlet-to-Neumann maps defined by

$$(\mathcal{R}^\pm \widetilde{E})(x') = \mp \sum_{n \in \mathbb{Z}^2} \frac{1}{i\beta_n^\pm} \left[k^2 \widetilde{E}_n^\pm - (\alpha_n \cdot \widetilde{E}_n^\pm)\alpha_n\right] \exp(i\alpha_n \cdot x'), \qquad (3.9)$$

for $\widetilde{E}(x') = \sum_{n \in \mathbb{Z}^2} \widetilde{E}_n^\pm \exp(i\alpha_n \cdot x') \in H_t^{-1/2}(\mathrm{Div}\,, \Gamma_b^\pm)$, $\tilde{E}_n^\pm \in \mathbb{C}^2$; see [1, 2]. Note that the operator $\mathcal{R}^+$ maps $e_3 \times E^{sc}$ to $(\operatorname{curl} E^{sc})_T$ on Γ_b^+ and that $\mathcal{R}^-$ maps $-e_3 \times E$ to the trace $(e_3 \times \operatorname{curl} E) \times e_3$ on Γ_b^-. If the incident frequency ω is sufficiently small, then the set $\Upsilon_{\rm res}$ is always empty and one can prove that the sesquilinear form generated by the left-hand side of (3.8) is positive coercive over $X \times X$ under the assumption (3.1). We refer to [10, Lemma 6.1] for the proof of the third assertion for perfectly conducting grating profiles using a variational formulation analogously to (3.8) but posed only in the upper half space. These results can be easily extended to transmission gratings.

There are two drawbacks in using (3.8) to compute the electric field. First, the transparent boundary operators $\mathcal{R}^\pm$ do not make sense if $\beta_n^\pm = 0$ (i.e. in the resonance case). Thus Rayleigh frequencies must be excluded. Second, in practice, $\mathcal{R}^\pm$ cannot be computed straightforwardly from (3.8). Instead, they must be approximated by taking sufficiently many terms in the expansions (see [5, Section 6] for the error estimates). Motivated by the variational formulations proposed in [11, 19] and based on the mortar technique of Nitsche (see Nitsche [16] and Sternberg [23]), we employ a consistent coupling of the electric field E on the interfaces $\Gamma_b^\pm$ as a replacement of the Dirichlet-to-Neumann maps. This way we propose a more general variational formulation than (3.8) for the electric field, which allows us not only to handle (BVP) in the resonance case but also to approximate the non-local boundary operators on $\Gamma_b^\pm$. Numerical experiments and convergence rate for a similar variational formulation were already reported in [11]. The goals of this paper are to provide a theoretical justification of the modified Nitsche's method and to prove the convergence of its numerical discretization using Nédélec's finite elements.

4 Variational formulation based on a coupling method

In this section we propose a variational formulation equivalent to (BVP). We begin with the fact that any column vector $E_n^+ \in \mathbb{C}^3$ satisfying $(\alpha_n, \beta_n^+)^\top \perp E_n^+$ for some $n = (n_1, n_2)^\top \in \mathbb{Z}^2$ can be represented as a linear combination of two vectors $E_{n,0}^+, E_{n,1}^+ \in \mathbb{C}^3$:

$$E_n^+ = C_{n,0}^+ \, E_{n,0}^+ + C_{n,1}^+ \, E_{n,1}^+, \quad C_{n,0}^+, \; C_{n,1}^+ \in \mathbb{C},$$

where

$$E_{n,0}^+ := \begin{cases} (-\alpha_n^{(2)}, \alpha_n^{(1)}, 0)^\top / |\alpha_n| \in \mathbb{S}^2 & \text{if } |\alpha_n| \neq 0 \\ (0,1,0)^\top & \text{else,} \end{cases} \tag{4.1}$$

$$E_{n,1}^+ := \begin{cases} \frac{|\alpha_n|}{h_n^+} (\alpha_n, \beta_n^+)^\top \times E_{n,0}^+ = (-\alpha_n^{(1)} \beta_n^+, -\alpha_n^{(2)} \beta_n^+, |\alpha_n|^2)^\top / h_n^+ & \text{if } |\alpha_n| \neq 0 \\ (-1,0,0)^\top & \text{else,} \end{cases} \tag{4.2}$$

with $h_n^+ := |\alpha_n| \sqrt{|\alpha_n|^2 + |\beta_n^+|^2}$. Obviously, it holds that $(\alpha_n, \beta_n^+)^\top \perp E_{n,l}^+$, $|E_{n,l}^+| = 1$ for $l = 0, 1, n \in \mathbb{Z}^2$. One can observe further that $E_{n,1}^+ \in \mathbb{S}^2$ if $\beta_n^+ \in \mathbb{R}$, and that $E_{n,1}^+ = e_3$ if $\beta_n^+ = 0$. The above decomposition of E_n^+ allows us to rewrite the Rayleigh expansion (3.5) for E^{sc} as (see also [19, Section 2.5])

$$E^{sc}(x) = \sum_{n \in \mathbb{Z}^2, \, l=1,2} C_{n,l}^+ \, U_{n,l}^+(x), \; U_{n,l}^+ := E_{n,l}^+ \exp\Big(i[\alpha_n \cdot x' + \beta_n^+ x_3]\Big), \; C_{n,l}^+ \in \mathbb{C} \tag{4.3}$$

for $x_3 > b$. Analogously, there holds

$$E(x) = \sum_{n \in \mathbb{Z}^2, \, l=1,2} C_{n,l}^- \, U_{n,l}^-(x), \; U_{n,l}^- := E_{n,l}^- \exp\Big(i[\alpha_n \cdot x' - \beta_n^- x_3]\Big), \; C_{n,l}^- \in \mathbb{C} \tag{4.4}$$

for $x < -b$, where $E_{n,0}^- = E_{n,0}^+$ and

$$E_{n,1}^- := \begin{cases} \frac{|\alpha_n|}{h_n^-} (\alpha_n, -\beta_n^-)^\top \times E_{n,0}^- = (\alpha_n^{(1)} \beta_n^-, \alpha_n^{(2)} \beta_n^-, |\alpha_n|^2)^\top / h_n^- & \text{if } |\alpha_n| \neq 0 \\ (-1,0,0)^\top & \text{else,} \end{cases}$$

with $h_n^- := |\alpha_n| \sqrt{|\alpha_n|^2 + |\beta_n^-|^2}$. Define the layers $D^\pm$ of height one above Γ_b^+ and below Γ_b^- by

$$\begin{aligned} D^+ &:= \Big\{ x \in \mathbb{R}^3 \colon 0 < x_j < \Lambda_j, \; j = 1,2, \; b < x_3 < b+1 \Big\}, \\ D^- &:= \Big\{ x \in \mathbb{R}^3 \colon 0 < x_j < \Lambda_j, \; j = 1,2, \; -b-1 < x_3 < -b \Big\}. \end{aligned}$$

Now we introduce the Sobolev spaces $Y_l^\pm$ as follows:

$$Y_l^\pm := \Big\{ U \in H(\mathrm{curl}\,, D^\pm) \colon U(x) = \sum_{n \in \mathbb{Z}^2} C_{n,l}^\pm \, U_{n,l}^\pm(x), \; C_{n,l}^\pm \in \mathbb{C} \Big\}, \; l = 0, 1. \tag{4.5}$$

Then we see that the function $E^+(x) := E^{sc}|_{D^+}$ belongs to the space $Y^+ := Y_0^+ \oplus Y_1^+$, and that $E^-(x) := E|_{D^-}$ belongs to the space $Y^- := Y_0^- \oplus Y_1^-$. Hence, the following problem is equivalent to (BVP):

(BVP'): Given an incident electric field E^{in}, find the α-quasiperiodic fields $(E, E^+, E^-) \in \mathbb{H} := X \times Y^+ \times Y^-$ such that E satisfies the curl-curl equation (3.3) in Ω in a distributional sense and the transmission conditions

$$\begin{aligned} &e_3 \times (E - E^{in} - E^+) = 0, \quad e_3 \times \operatorname{curl}(E - E^{in} - E^+) = 0 \quad \text{on} \quad \Gamma_b^+, \\ &e_3 \times (E - E^-) \qquad\quad\;\; = 0, \quad e_3 \times \operatorname{curl}(E - E^-) \qquad\quad\;\; = 0 \quad \text{on} \quad \Gamma_b^-. \end{aligned} \tag{4.6}$$

Motivated by the arguments in [19, Section 3.2] and the variational formulations in [10, 11], we propose a new variational formulation that is equivalent to (BVP'). For the triples of functions (E, E^+, E^-),$(V, V^+, V^-) \in \mathbb{H}$, define the sesquilinear form $a(\cdot, \cdot) : \mathbb{H} \times \mathbb{H} \to \mathbb{C}$ by

$$\begin{aligned} &a\Big((E, E^+, E^-), (V, V^+, V^-)\Big) \qquad\qquad (4.7)\\ &:= \int_\Omega \{\operatorname{curl} E \cdot \operatorname{curl} \overline{V} - k^2(x) E \cdot \overline{V}\} \,\mathrm{d}x \\ &\quad - \int_{\Gamma_b^+} \Big\{\operatorname{curl} E^+ \cdot e_3 \times \overline{V} - e_3 \times (E - E^+) \cdot \operatorname{curl} \overline{V}^+\Big\} \,\mathrm{d}s \\ &\quad + \int_{\Gamma_b^-} \Big\{\operatorname{curl} E^- \cdot e_3 \times \overline{V} - e_3 \times (E - E^-) \cdot \operatorname{curl} \overline{V}^-\Big\} \,\mathrm{d}s \\ &\quad - \eta^+ \sum_{n \in \Upsilon^+} \left[\int_{\Gamma_b^+} e_3 \times (E - E^+) \cdot (e_3 \times \overline{U}_{n,0}^+) \,\mathrm{d}s \; \overline{\int_{\Gamma_b^+} e_3 \times V^+ \cdot (e_3 \times \overline{U}_{n,0}^+) \mathrm{d}s}\right] \\ &\quad - \eta^- \sum_{n \in \Upsilon^-} \left[\int_{\Gamma_b^-} e_3 \times (E - E^-) \cdot (e_3 \times \overline{U}_{n,0}^-) \,\mathrm{d}s \; \overline{\int_{\Gamma_b^-} e_3 \times V^- \cdot (e_3 \times \overline{U}_{n,0}^-) \mathrm{d}s}\right], \end{aligned}$$

where $\eta^\pm > 0$ are constant factors. The set $\Upsilon^\pm$ is a finite fixed subset of $\mathbb{Z}^2$ with $\Upsilon_{\mathrm{res}}^\pm \subseteq \Upsilon^\pm$ (cf. (3.6)). Our variational formulation is to find $(E, E^+, E^-) \in \mathbb{H}$ such that

$$a\Big((E, E^+, E^-), (V, V^+, V^-)\Big) = -a\Big((0, E^{in}, 0), (V, V^+, V^-)\Big) \tag{4.8}$$

for all $(V, V^+, V^-) \in \mathbb{H}$. Note that terms like $\int_{\Gamma_b^\pm} \operatorname{curl} E^\pm \cdot e_3 \times \overline{V} \,\mathrm{d}s$ are bounded. Indeed, since $E^\pm$ is the solution of the curl-curl equation in $D^\pm$, we get $\operatorname{curl} E^\pm \in H(\operatorname{curl}, D^\pm)$ and $(\operatorname{curl} E^\pm)|_{\Gamma_b^\pm} \in H^{-1/2}(\operatorname{Curl}, \Gamma_b^\pm)$. Further, note that the second part of the second and third terms on the right-hand side of (4.7) both have opposite signs than the corresponding terms in [11]. Moreover, the integrals with factor $\eta^\pm$ in (4.7) are modifications of the following terms involved in the variational equation of [11]:

$$\eta^\pm \int_{\Gamma_b^\pm} e_3 \times (E - E^\pm) \cdot e_3 \times (\overline{V - V^\pm}) \mathrm{d}s. \tag{4.9}$$

The expressions in (4.9) are not meaningful for general $(E, E^+, E^-), (V, V^+, V^-) \in \mathbb{H}$, since both $e_3 \times (E - E^\pm)$ and $e_3 \times (\overline{V - V^\pm})$ belong to the space $H_t^{-1/2}(\operatorname{Div}, \Gamma_b^\pm)$. Integrals like $\eta \int_{\Gamma_b^\pm} e_3 \times u \cdot e_3 \times \overline{v} \,\mathrm{d}s$ in the mortar approach make sense for finite element methods, where u

and v are finite element functions and η tends to zero with the meshsize. The idea employed in [19] is to replace the integral (4.9) by the Galerkin approximation

$$\sum_{\substack{n,l:|n|^2<N \\ \beta_n^\pm\neq 0 \,\mathrm{or}\, l=0}} \left[\eta^\pm \int_{\Gamma_b^\pm} e_3\times(E-E^\pm)\cdot e_3\times\overline{U}_{n,l}^\pm \mathrm{d}s \,\overline{\int_{\Gamma_b^\pm} e_3\times(V-V^\pm)\cdot e_3\times\overline{U}_{n,l}^\pm \mathrm{d}s}\right] \tag{4.10}$$

$$+\eta^\pm \sum_{n:\beta_n^\pm=0} \left[\int_{\Gamma_b^\pm} e_3\times(E-E^\pm)\cdot\overline{U}_{n,0}^\pm \mathrm{d}s \,\overline{\int_{\Gamma_b^\pm} e_3\times(V-V^\pm)\cdot\overline{U}_{n,0}^\pm \mathrm{d}s}\right] \tag{4.11}$$

with a sufficiently large number $N>0$. It is also mentioned in [19] that the summation in (4.10) and (4.11) can even be restricted to all $n\in\mathbb{Z}^2$ with $\beta_n^\pm=0$. In the present paper, we only use the terms of (4.10) with $n\in\Upsilon^\pm$ and simplify them to get the last two terms in (4.7). Note that choosing $\Upsilon^\pm$ larger than $\Upsilon_{\mathrm{res}}^\pm$ makes the numerical scheme more stable in the near-resonance case.

Arguing similarly to [10, Lemma 3.3], we can prove the equivalence of the variational formulation (4.8) and the problem (BVP'). Moreover, in the non-resonance case, i.e. $\Upsilon_{\mathrm{res}}=\emptyset$, and for $\Upsilon=\Upsilon_{\mathrm{res}}$, the variational formulations (4.8) and (3.8) are equivalent (see [10, Remark 3.4]). Thus, the variational formulation (4.8) is indeed more general than (3.8). It is worth to mention that, using (4.8), we can also prove the solvability results in Theorem 3.1, since the arguments in [10] for perfectly conducting grating profiles can be easily adapted to transmission gratings. To prepare the convergence analysis of the finite element discretization, in this paper we only check the Fredholm property of the operator $A:\mathbb{H}\to\mathbb{H}'$ generated by the bounded sesquilinear form $a(\cdot,\cdot)$ defined in Section 4, i.e. A is given by

$$a\Big((E,E^+,E^-),(V,V^+,V^-)\Big)=\Big\langle A(E,E^+,E^-),(V,V^+,V^-)\Big\rangle. \tag{4.12}$$

Here $\mathbb{H}'$ denotes the space dual to $\mathbb{H}$ with respect to the duality $\langle\cdot,\cdot\rangle$ extending the scalar product in $L^2(\Omega)^3\times L^2(D^+)^3\times L^2(D^-)^3$. The rest of this section is devoted to verify

Theorem 4.1. *The operator A defined by (4.12) is a Fredholm operator with index zero.*

First we recall

Definition 4.2. *A bounded sesquilinear form $l(\cdot,\cdot)$ given on some Hilbert space Y is called strongly elliptic if there exists a compact form $\tilde{l}(\cdot,\cdot)$ and a constant $c>0$ such that*

$$\mathrm{Re}\, l(u,u)\geq c\,||u||_Y^2-\tilde{l}(u,u),\quad \forall\ u\in Y.$$

To prove Theorem 4.1, we need a periodic analogue of the Hodge decomposition of X.

Lemma 4.3. (i) *We have $X=X_0\oplus X_1$, where*

$$\begin{aligned} X_1 &:= \Big\{\nabla p\colon p\in H^1_{qp}(\Omega)\Big\}\subset X,\\ X_0 &:= \Big\{E_0\in X\colon \int_\Omega k^2(x)\nabla p\cdot\overline{E}_0\,\mathrm{d}x=0 \text{ for all } \nabla p\in X_1\Big\}. \end{aligned}$$

and the space X_0 is compactly embedded into $L^2(\Omega)^3$.

(ii) *We have* $\operatorname{div}(k^2(x)E_0) = 0$ *in* Ω *and* $e_3 \cdot E_0 = 0$ *on* $\Gamma_b^\pm$ *for any* $E_0 \in X_0$.

Proof. See e.g. [3, Sect. 3.1] for the proof of the first assertion in more general periodic chiral structures and [13, Sect. 4.4] in the case of non-periodic structures where $k^2(x)$ is allowed to be a complex valued function. Using integration by parts, it follows from the definition of X_0 that $\operatorname{div}(k^2(x)E_0) = 0$ in Ω and $e_3 \cdot k^2(x)E_0 = 0$ on $\Gamma_b^\pm$. Since $k^2(x)$ is a non-vanishing piecewise constant function in $\overline{\Omega}$, we obtain $e_3 \cdot E_0 = 0$ on $\Gamma_b^\pm$. □

By Lemma 4.3 and the definitions of $Y_l^\pm$, we can decompose our space $\mathbb{H}$ into six subspaces

$$\mathbb{H} = (X_0 \oplus X_1) \times (Y_0^+ \oplus Y_1^+) \times (Y_0^- \oplus Y_1^-).$$

For $(E, E^+, E^-), (V, V^+, V^-) \in \mathbb{H}$, we may assume that

$$\begin{aligned} E = \nabla p + E_0,\ E^\pm = E_0^\pm + E_1^\pm, \quad &\text{where } \nabla p \in X_1,\ E_0 \in X_0,\ E_l^\pm \in Y_l^\pm,\ l{=}0,1, \\ V = \nabla \xi + V_0,\ V^\pm = V_0^\pm + V_1^\pm, \quad &\text{where } \nabla \xi \in X_1,\ V_0 \in X_0,\ V_l^\pm \in Y_l^\pm,\ l{=}0,1. \end{aligned}$$

For the analysis of form a, we define several sesquilinear forms as follows. Let

$$\begin{aligned} a_1(\nabla p, \nabla \xi) \ &:= \ \int_\Omega k^2(x) \nabla p \cdot \nabla \overline{\xi}\, \mathrm{d}x, \quad \forall\, \nabla p, \nabla \xi \in X_1, \\ a_2(E_0, V_0) \ &:= \ \int_\Omega \left\{ \operatorname{curl} E_0 \cdot \operatorname{curl} \overline{V_0} - k^2(x) E_0 \cdot \overline{V_0} \right\} \mathrm{d}x, \quad \forall\, E_0, V_0 \in X_0, \\ a_3^\pm(E_0^\pm, V_0^\pm) \ &:= \ \pm \int_{\Gamma_b^\pm} e_3 \times E_0^\pm \cdot \operatorname{curl} \overline{V}_0^\pm \, \mathrm{d}s, \quad \forall\, E_0^\pm, V_0^\pm \in Y_0^\pm, \\ a_4^\pm(E_1^\pm, V_1^\pm) \ &:= \ \pm \int_{\Gamma_b^\pm} e_3 \times E_1^\pm \cdot \operatorname{curl} \overline{V}_1^\pm \, \mathrm{d}s, \quad \forall\, E_1^\pm, V_1^\pm \in Y_1^\pm, \end{aligned}$$

and let

$$\begin{aligned} & a_5^\pm\Big((E, E^+, E^-), (V, V^+, V^-)\Big) := \pm \int_{\Gamma_b^\pm} e_3 \times E \cdot \operatorname{curl} \overline{V}^\pm \mathrm{d}s, \\ & a_6^\pm\Big((E, E^+, E^-), (V, V^+, V^-)\Big) \\ & \quad := \ -\eta^\pm \sum_{n \in \Upsilon^\pm} \left\{ \int_{\Gamma_b^\pm} e_3 \times (E - E^\pm) \cdot (e_3 \times \overline{U}_{n,0}^\pm)\, \mathrm{d}s \ \overline{\int_{\Gamma_b^\pm} (e_3 \times V^\pm) \cdot (e_3 \times \overline{U}_{n,0}^\pm)\, \mathrm{d}s} \right\} \end{aligned}$$

for any $(E, E^+, E^-), (V, V^+, V^-) \in \mathbb{H}$. For brevity we write

$$a_5^\pm\Big((E, E^+, E^-), (V, V^+, V^-)\Big) = a_5^\pm(E, V^\pm), \quad \forall\, E \in X,\ V^\pm \in Y^\pm. \tag{4.13}$$

Lemma 4.4. *For any* $\nabla \xi \in X_1$ *and* $V_0^\pm \in Y_0^\pm$*, we have* $a_5^\pm(\nabla \xi, V_0^\pm) = 0$.

Proof. Assume that $\nabla \xi \in X_1$ and $V_0^+ \in Y_0^+$. Without loss of generality ξ can be assumed to be smooth. We can expand the function $\xi(x)$ into the series

$$\xi(x) = \sum_{n \in \mathbb{Z}^2} f_n(x_3) \exp(i \alpha_n \cdot x'), \quad f_n \in C^2(\mathbb{R}^+),$$

in a sufficiently small neighborhood of Γ_b^+. This implies that

$$(e_3 \times \nabla\xi)|_{\Gamma_b^+} = \sum_{n\in\mathbb{Z}^2} i f_n(b)(-\alpha_n^{(2)}, \alpha_n^{(1)}, 0)^\top \exp(i\alpha_n \cdot x'). \tag{4.14}$$

Making use of $\operatorname{curl} U_{n,0}^+ = iU_{n,1}^+ \sqrt{|\alpha_n|^2 + |\beta_n^+|^2}$ (see [10, Lemma 3.1]), and recalling the definition of $U_{n,1}^+$ and the sesquilinear form a_5^+, we end up with the identity

$$a_5^+(\nabla\xi, V_0^+) = \int_{\Gamma_b^+} (e_3 \times \nabla\xi) \cdot \operatorname{curl} \overline{V_0^+} \mathrm{d}s = 0.$$

The proof for a_5^- can be carried out analogously. □

Note the last proof is a new and simpler proof for [10, Lemma 4.3]. Using Lemmas 4.3 and 4.4, the definition of a and a simple calculation imply (see Table 1)

$$\begin{aligned}
& a\Big((E, E^+, E^-), (V, V^+, V^-)\Big) \\
&= a\Big((\nabla p + E_0, E_0^+ + E_1^+, E_0^- + E_1^-), (\nabla\xi + V_0, V_0^+ + V_1^+, V_0^- + V_1^-)\Big) \\
&= -a_1(\nabla p, \nabla\xi) + a_2(E_0, V_0) - a_3^+(E_0^+, V_0^+) - a_4^+(E_1^+, V_1^+) + a_5^+(E_0, V_0^+) \\
&\quad -\overline{a_5^+(V_0, E_0^+)} + a_5^+(E_0, V_1^+) - \overline{a_5^+(V_0, E_1^+)} + a_5^+(\nabla p, V_1^+) - \overline{a_5^+(\nabla\xi, E_1^+)} \\
&\quad + a_6^+\Big((E, E^+, E^-), (V, V^+, V^-)\Big) + a_3^-(E_0^-, V_0^-) + a_4^-(E_1^-, V_1^-) - a_5^-(E_0, V_0^-) \\
&\quad + \overline{a_5^-(V_0, E_0^-)} - a_5^-(E_0, V_1^-) + \overline{a_5^-(V_0, E_1^-)} - a_5^-(\nabla p, V_1^-) + \overline{a_5^-(\nabla\xi, E_1^-)} \\
&\quad + a_6^-\Big((E, E^+, E^-), (V, V^+, V^-)\Big).
\end{aligned} \tag{4.15}$$

Proof of Theorem 4.1. Obviously, we have

- a_1 is coercive on X_1, i.e. there exists some constant $C > 0$ such that

  $$\operatorname{Re}[a_1(\nabla p, \nabla p)] \geq C||\nabla p||_X, \quad \forall\, \nabla p \in X_1.$$

- a_2 is strongly elliptic over X_0, due to the estimate

  $$\operatorname{Re}[a_2(E_0, E_0)] \geq ||E_0||_X - [1 + ||k^2||_{L^\infty(\Omega)}]\, ||E_0||^2_{L^2(\Omega)^3}$$

  for any $E_0 \in X_0$ and the compact imbedding of X_0 into $L^2(\Omega)^3$ (see Lemma 4.3).

- $a_6^\pm$ are compact forms over $\mathbb{H}$, since each of them corresponds to a finite rank operator over $\mathbb{H}$.

To demonstrate the Fredholm property of the sesquilinear form a, we now need to study the other forms $a_3^\pm$, $a_4^\pm$ and $a_5^\pm$. Concerning a_3^+ and a_4^+, it is shown in [10, Lemma 4.5] that, there exist compact forms $\tilde{a}_3^+ : Y_0^+ \times Y_0^+ \to \mathbb{C}$ and $\tilde{a}_4^+ : Y_1^+ \times Y_1^+ \to \mathbb{C}$ such that

$$\begin{aligned}
-\operatorname{Re} a_3^+(E_0^+, E_0^+) &\geq C_3^+\, ||E_0^+||^2_{H(\mathrm{curl}, D^+)} - \tilde{a}_3^+(E_0^+, E_0^+), \ \forall\, E_0^+ \in Y_0^+, \\
\operatorname{Re} a_4^+(E_1^+, E_1^+) &\geq C_4^+\, ||E_1^+||^2_{H(\mathrm{curl}, D^+)} - \tilde{a}_4^+(E_1^+, E_1^+), \ \forall\, E_1^+ \in Y_1^+
\end{aligned} \tag{4.16}$$

for some constants $C_3^+, C_4^+ > 0$, i.e., the sesquilinear forms $-a_3^+$ and a_4^+ are strongly elliptic over Y_0^+ and Y_1^+, respectively. The proof of the estimates in (4.16) can be easily extended to the sesquilinear forms a_3^- and a_4^-. That is, we can find compact forms $\tilde{a}_3^- : Y_0^- \times Y_0^- \to \mathbb{C}$ and $\tilde{a}_4^- : Y_1^- \times Y_1^- \to \mathbb{C}$ such that

$$\begin{aligned} \operatorname{Re} a_3^-(E_0^-, E_0^-) &\geq C_3^- \, ||E_0^-||^2_{H(\mathrm{curl}, D^-)} - \tilde{a}_3^-(E_0^-, E_0^-), \ \forall\, E_0^- \in Y_0^-, \\ -\operatorname{Re} a_4^-(E_1^-, E_1^-) &\geq C_4^- \, ||E_1^-||^2_{H(\mathrm{curl}, D^-)} - \tilde{a}_4^-(E_1^-, E_1^-), \ \forall\, E_1^- \in Y_1^- \end{aligned} \tag{4.17}$$

for some constants $C_3^-, C_4^- > 0$. Hence the strong ellipticity of a_3^- and $-a_4^-$ follows. Finally, in view of [10, Lemma 4.7] we have

- a_5^+ is compact over $X_0 \times Y_1^+$,

and analogously

- a_5^- is compact over $X_0 \times Y_1^-$.

To prove the Fredholm property of the variational formulation (4.8), it suffices to verify that the operator corresponding to the sesquilinear form $a - a_6^+ - a_6^-$ is Fredholm over $\mathbb{H}$ with index zero. For this purpose, we define the spaces $\mathbb{H}_j = X_j \times Y_j^+ \times Y_j^-$ for $j = 0, 1$, so that we can rewrite $\mathbb{H} = X \times Y^+ \times Y^- = \mathbb{H}_0 \oplus \mathbb{H}_1$. Define the sesquilinear forms

$$\begin{aligned} & b_0\Big((E_0, E_0^+, E_0^-), (V_0, V_0^+, V_0^-)\Big) \\ & \quad := a_2(E_0, V_0) - a_3^+(E_0^+, V_0^+) + a_3^-(E_0^-, V_0^-) \\ & \qquad + a_5^+(E_0, V_0^+) - \overline{a_5^+(V_0, E_0^+)} - a_5^-(E_0, V_0^-) + \overline{a_5^-(V_0, E_0^-)} \end{aligned}$$

for all $(E_0, E_0^+, E_0^-), (V_0, V_0^+, V_0^-) \in \mathbb{H}_0$, and

$$\begin{aligned} & b_1\Big((\nabla p, E_1^+, E_1^-), (\nabla \xi, V_1^+, V_1^-)\Big) \\ & \quad := -a_1(\nabla p, \nabla \xi) - a_4^+(E_1^+, V_1^+) + a_4^-(E_1^-, V_1^-) \\ & \qquad + a_5^+(\nabla p, V_1^+) - \overline{a_5^+(\nabla \xi, E_1^+)} - a_5^-(\nabla p, V_1^-) + \overline{a_5^-(\nabla \xi, E_1^-)} \end{aligned}$$

for all $(\nabla p, E_1^+, E_1^-), (\nabla \xi, V_1^+, V_1^-) \in \mathbb{H}_1$. Now split the form in Table 1 in blocks corresponding to the splitting $\mathbb{H} = \mathbb{H}_1 \times \mathbb{H}_2$. Then the restriction to $\mathbb{H}_1$ is the form b_0 with the strongly elliptic quadratic form

$$\begin{aligned} & \operatorname{Re} b_0\Big((E_0, E_0^+, E_0^-), (E_0, E_0^+, E_0^-)\Big) \\ & \quad = \operatorname{Re} a_2(E_0, E_0) - \operatorname{Re} a_3^+(E_0^+, E_0^+) + \operatorname{Re} a_3^-(E_0^-, E_0^-). \end{aligned}$$

The restriction to $\mathbb{H}_1$ is the form b_1, and $-b_1$ has the strongly elliptic quadratic form

$$\begin{aligned} & -\operatorname{Re} b_1\Big((\nabla p, E_1^+, E_1^-), (\nabla p, E_1^+, E_1^-)\Big) \\ & \quad = \operatorname{Re} a_1(\nabla p, \nabla p) + \operatorname{Re} a_4^+(E_1^+, E_1^+) - \operatorname{Re} a_4^-(E_1^-, E_1^-). \end{aligned}$$

Consequently, the diagonal blocks of the splitting into 2×2 blocks of size 3×3 correspond to Fredholm operators with index zero. On the other hand, the full form in Table 1 differs from the diagonal block matrix only by compact terms. Hence the form a generates a Fredholm operator with index zero. □

		$\mathbb{H}_0 := X_0 \times Y_0^+ \times Y_0^-$			$\mathbb{H}_1 := X_1 \times Y_1^+ \times Y_1^-$		
		$X_0(E_0)$	$Y_0^+(E_0^+)$	$Y_0^-(E_0^-)$	$X_1(\nabla p)$	$Y_1^+(E_1^+)$	$Y_1^-(E_1^-)$
$\mathbb{H}_0$	$X_0(V_0)$	$a_2(E_0, V_0)$	$-\overline{a_5^+(V_0, E_0^+)}$	$\overline{a_5^-(V_0, E_0^-)}$	0	$-\overline{a_5^+(V_0, E_1^+)}$	$\overline{a_5^-(V_0, E_1^-)}$
	$Y_0^+(V_0^+)$	$a_5^+(E_0, V_0^+)$	$-a_3^+(E_0^+, V_0^+)$	0	0	0	0
	$Y_0^-(V_0^-)$	$-\,a_5^-(E_0, V_0^-)$	0	$a_3^-(E_0^-, V_0^-)$	0	0	0
$\mathbb{H}_1$	$X_1(\nabla\xi)$	0	0	0	$-a_1(\nabla p, \nabla\xi)$	$-\,\overline{a_5^+(\nabla\xi, E_1^+)}$	$\overline{a_5^-(\nabla\xi, E_1^-)}$
	$Y_1^+(V_1^+)$	$a_5^+(E_0, V_1^+)$	0	0	$a_5^+(\nabla p, V_1^+)$	$-a_4^+(E_1^+, V_1^+)$	0
	$Y_1^-(V_1^-)$	$-\,a_5^-(E_0, V_1^-)$	0	0	$-\,a_5^-(\nabla p, V_1^-)$	0	$a_4^-(E_1^-, V_1^-)$

Table 1: The diagram for the sesquilinear form $a - a_6^+ - a_6^-$ over $\mathbb{H} \times \mathbb{H}$, where $\mathbb{H} = X \times Y^+ \times Y^-$.

5 Numerical analysis of Finite Element Method

5.1 Finite element space and FEM

As mentioned in the introduction, we assume that the optical medium in $\mathbb{R}^3$ is piecewise smooth. For the convergence analysis, we suppose that the interface between any two different materials is a polyhedral surface. Let $\tau_h = \tau_h(\Omega)$ be a partition of $\overline{\Omega}$ by tetrahedrons K of diameter h_K, i.e. $\overline{\Omega} = \cup_{K\in\tau_h}\overline{K}$, where h denotes the maximum diameter of the elements in τ_h. Of course, we suppose that ϵ and k are constant over each $K \in \tau_h$. We will use standard Nédélec's edge elements (cf. [13]) and analyze convergence for $h \to 0$. For each element $K \in \tau_h$ and $\mathsf{k} > 1$, denote by P_{k} the polynomials of maximal total degree k and by $\widetilde{P}_{\mathsf{k}}$ the homogeneous polynomials of total degree k. Define the subspace $\mathcal{S}_{\mathsf{k}}$ of homogeneous vector polynomials of degree k by $\mathcal{S}_{\mathsf{k}} := \{\mathbf{p} \in (\widetilde{P}_{\mathsf{k}})^3 |\ x \cdot \mathbf{p}(x) = 0\}$. The curl conforming edge elements of Nédélec rely on the use of the vector polynomial space $R_K := (P_{\mathsf{k}-1})^3 \oplus \mathcal{S}_{\mathsf{k}}$. More precisely, the Nédélec finite element space of edge elements of degree k are defined by

Definition 5.1. *Let $X_h \subset X$ be the set of functions $E_h : \Omega \to \mathbb{C}^3$ such that:*

(i) For any $K \in \tau_h$, we have $E_h|_K \in R_K$.

(ii) For any edge e of the FE partition and for any $K, K' \in \tau_h$ s.t. $e \subseteq \overline{K} \cap \overline{K}'$, we have $\int_e (E_h|_K) \cdot \tau\ q\ \mathrm{d}e = \int_e (E_h|_{K'}) \cdot \tau\ q\ \mathrm{d}e$ for any $q \in P_{\mathsf{k}-1}$. Here τ is the unit vector pointing into the direction of e.

(iii) For any face f of the FE partition and for any $K, K' \in \tau_h$ such that $f \subseteq K \cap K'$, there holds $\int_f (E_h|_K) \cdot \mathbf{q}\ \mathrm{d}s = \int_f (E_h|_{K'}) \cdot \mathbf{q}\ \mathrm{d}s$ for any $\mathbf{q} \in (P_{\mathsf{k}-2})^3$ with $\mathbf{q} \cdot \nu_f = 0$. Here ν_f denotes the normal to the face f.

To define the discretized spaces for $Y_l^{\pm}$, we introduce the finite set $\Upsilon_h := \{n \in \mathbb{Z}^2 : |n| \le C/h\}$ for some constant $C > 0$. Then, set

$$Y_h^{\pm} := Y_{h,0}^{\pm} \oplus Y_{h,1}^{\pm}, \quad Y_{h,l}^{\pm} := \mathrm{span}\Big\{U_{n,l}^{\pm} :\ n \in \Upsilon_h\Big\},\ l = 0, 1.$$

The discretized full space is defined as $\mathbb{H}_h := X_h \times Y_h^+ \times Y_h^-$. Now the finite element approximation associated to (4.8) can be formulated as follows: find $(E_h, E_h^+, E_h^-) \in \mathbb{H}_h$ such that

$$a\Big((E_h, E_h^+, E_h^-), (V_h, V_h^+, V_h^-)\Big) = -a\Big((0, E^{in}, 0), (V_h, V_h^+, V_h^-)\Big) \tag{5.1}$$

for all $(V_h, V_h^+, V_h^-) \in \mathbb{H}_h$.

5.2 Auxiliary notation and facts

Let $(F, F^+, F^-) \in \mathbb{H}_h'$ be defined as the right-hand side of Equation (5.1), and let $P_h := (P^{X_h}, P^{Y_h^+}, P^{Y_h^-})$ be the orthogonal projection of $\mathbb{H}$ onto $\mathbb{H}_h$. Then we obtain the operator equation of the FEM

$$A_h(E_h, E_h^+, E_h^-) = (P_h)^*(F, F^+, F^-), \quad A_h := (P_h)^* A|_{\mathbb{H}_h}, \tag{5.2}$$

where $A\colon \mathbb{H} \to \mathbb{H}'$ is given in (4.12). Note that the operators $A_h\colon \mathbb{H}_h \to \mathbb{H}'_h$ are uniformly bounded in $h > 0$. It follows from [13, Lemma 10.10] that $P^{X_h} \longrightarrow I$ in X, and by the definitions of $Y_h^\pm$ we see $P^{Y_h^\pm} \longrightarrow I$ in $Y_h^\pm$. This implies the strong convergence of P_h to I in $\mathbb{H}$. Consequently, the convergence $(P_h)^* \longrightarrow I$ holds in $\mathbb{H}'$ and $A_h P_h \longrightarrow A$.

Definition 5.2. *The operators $A_h : \mathbb{H}_h \longrightarrow \mathbb{H}'_h$ are called stable if there exists an $h_0 > 0$ such that A_h is invertible for all $h \le h_0$ and if $||A_h^{-1}|| \le c$ for some constant $c > 0$ independent of $h \in (0, h_0)$.*

Note that the operator norm of the inverse operator A_h^{-1} can be computed as

$$||A_h^{-1}|| = \inf_{(0,0,0)\neq(E_h,E_h^+,E_h^-)\in\mathbb{H}_h} \sup_{(0,0,0)\neq(V_h,V_h^+,V_h^-)\in\mathbb{H}_h} \frac{\left|a\Big((E_h, E_h^+, E_h^-), (V_h, V_h^+, V_h^-)\Big)\right|}{\|(E_h, E_h^+, E_h^-)\|_{\mathbb{H}}\ \|(V_h, V_h^+, V_h^-)\|_{\mathbb{H}}}\,.$$

Definition 5.3. *We say that the FEM for (4.8) is convergent if, for any $(F, F^+, F^-) \in \mathbb{H}'$ and for all $h < h_0$, the approximate solution (E_h, E_h^+, E_h^-) to*

$$a\Big((E_h, E_h^+, E_h^-), (V_h, V_h^+, V_h^-)\Big) = \Big\langle (F, F^+, F^-), (V_h, V_h^+, V_h^-)\Big\rangle \tag{5.3}$$

for all $(V_h, V_h^+, V_h^-) \in \mathbb{H}_h$ exists and is unique, and if (E_h, E_h^+, E_h^-) converges strongly in $\mathbb{H}$ to the exact solution (E, E^+, E^-) of the continuous variational problem $A(E, E^+, E^-) = (F, F^+, F^-)$.

Now we recall two well-known results on the convergence and perturbations (cf. e.g. [17, Chapter 1], [18]), which are our main tools for analyzing the discrete variational problem (5.1). Lemma 5.4 is a simple consequence of the Banach-Steinhaus theorem and, for the reader's convenience, we provide a short proof of Lemma 5.5.

Lemma 5.4. *Suppose the strong convergence $P_h \longrightarrow I$. Then the finite element scheme (5.3) is convergent if and only if the operators A_h defined in (5.2) are stable.*

Lemma 5.5. *Suppose $P_h \to I$. Furthermore, suppose that the operators $B_h\colon \mathbb{H}_h \longrightarrow \mathbb{H}'_h$ are stable, and that the convergence $B_h P_h \longrightarrow B$ holds as $h \to 0$ with some operator $B\colon \mathbb{H} \longrightarrow \mathbb{H}'$. Moreover, let $T\colon \mathbb{H} \to \mathbb{H}'$ be a compact operator such that $C := B + T$ is invertible. Let the operators $C_h : \mathbb{H}_h \longrightarrow \mathbb{H}'_h$ be small perturbations of $B_h + (P_h)^* T|_{\mathbb{H}_h}$, i.e.,*

$$C_h = B_h + (P_h)^* T|_{\mathbb{H}_h} + D_h, \quad ||D_h|| \to 0 \quad \text{as } h \to 0.$$

Then the C_h are stable.

Proof. The small perturbations D_h can be treated by the usual Neumann series argument. Hence, it suffices to prove that the operators $B_h + (P_h)^* T|_{\mathbb{H}_h}\colon \mathbb{H}_h \longrightarrow \mathbb{H}'_h$ are stable, i.e. that the inverse operators of the $B_h + (P_h)^* T|_{\mathbb{H}_h}$ exist and are uniformly bounded.

We first show that B^{-1} exists. Since C is invertible and T is compact, B is a Fredholm operator with index zero. Hence we only need to show that $\mathrm{Ker} B = \{0\}$. Noting that the B_h are stable, we get, for any $u \in \mathbb{H}$, $||P_h u||_{\mathbb{H}} = ||B_h^{-1} B_h P_h u||_{\mathbb{H}} \leq c\,||B_h P_h u||_{\mathbb{H}'}$ with a constant $c > 0$ independent of h. Letting $h \to 0$, we obtain $||u||_{\mathbb{H}} \leq c||Bu||_{\mathbb{H}'}$ which implies $\mathrm{Ker} B = \{0\}$.

Now the pointwise convergence $B_h^{-1}(P_h)^* \to B^{-1}$ is easy to see, and thus the norm convergence $||[B_h^{-1}(P_h)^* - B^{-1}]T|| \to 0$ as $h \to 0$ follows. A simple calculation shows

$$\begin{aligned} B_h + (P_h)^* T|_{\mathbb{H}_h} &= B_h[I|_{\mathbb{H}_h} + B_h^{-1}(P_h)^* T|_{\mathbb{H}_h}] \\ &= B_h\left\{P_h(I + B^{-1}T)|_{\mathbb{H}_h} + P_h[B_h^{-1}(P_h)^* - B^{-1}]T|_{\mathbb{H}_h}\right\}. \end{aligned}$$

To prove the stability of $B_h + (P_h)^* T|_{\mathbb{H}_h}$, we only need to prove that of $P_h(I + B^{-1}T)|_{\mathbb{H}_h}$, because the second term in the curly brackets of the previous identity tends to zero as $h \to 0$. From the invertibility of C it follows the existence of $(I + B^{-1}T)^{-1}$. Then, we can check that

$$\begin{aligned} &[P_h(I + B^{-1}T)^{-1}|_{\mathbb{H}_h}][P_h(I + B^{-1}T)|_{\mathbb{H}_h}] \\ &\quad = [P_h(I + B^{-1}T)^{-1}(P_h - I)(I + B^{-1}T)|_{\mathbb{H}_h}] + I|_{\mathbb{H}_h} \\ &\quad = [P_h(I + B^{-1}T)^{-1}(P_h - I)B^{-1}T|_{\mathbb{H}_h}] + I|_{\mathbb{H}_h}, \end{aligned}$$

where $||P_h(I+B^{-1}T)^{-1}(P_h-I)B^{-1}T|_{\mathbb{H}_h}|| \leq c\,||(P_h-I)B^{-1}T|| \to 0$. Hence, the product of $([P_h(I+B^{-1}T)^{-1}(P_h-I)B^{-1}T|_{\mathbb{H}_h}]+I|_{\mathbb{H}_h})^{-1}$ and $[P_h(I+B^{-1}T)^{-1}|_{\mathbb{H}_h}]$ is the uniformly bounded inverse of $P_h(I + B^{-1}T)|_{\mathbb{H}_h}$. $\square$

Remark 5.6. *The projection P_h in Lemma 5.5 can even be replaced by operators which are not projections. If the P_h are orthogonal projections and if $B_h = (P_h)^* B|_{\mathbb{H}_h}$, then Lemma 5.5 reduces to the classical stability property of projection methods (see e.g. [12, Theorem 13. 7]).*

5.3 Convergence analysis of FEM

To prove the convergence of the FEM, we need the Hodge decomposition of the discrete functions in X_h. Define $\mathcal{S}_h := \{p_h \in H^1_{qp}(\Omega_b)\colon\ p_h|_K \in P_{\mathsf{k}} \text{ for all } K \in \tau_h\}$. We have the discrete Hodge decomposition $X_h = X_{h,0} \oplus X_{h,1}$ analogously to Lemma 4.3, where

$$\begin{aligned} X_{h,1} &:= \Big\{\nabla p_h\colon\ p_h \in \mathcal{S}_h\Big\} \subseteq X_1, \\ X_{h,0} &:= \Big\{E_h \in X_h\colon\ 0 = \int_\Omega k^2(x) E_h \cdot \nabla p_h \,\mathrm{d}x \text{ for all } \nabla p_h \in X_{h,1}\Big\}. \end{aligned}$$

Unfortunately, it is not true that $X_{h,0} \subset X_0$. This causes difficulties in our convergence analysis. The following property of discrete compactness will help us to overcome these difficulties.

Definition 5.7. *We say the $X_{h,0}$ have the discrete compactness property if, for any sequence $E_{n,0} \in X_{h_n,0},\ n = 1, 2, \cdots$ such that $\|E_{n,0}\|_X < c$ with some c independent of index n, there is an element $E_0 \in X_0$ and a subsequence of $E_{n,0}$ converging in $L^2(\Omega)^3$ to E_0.*

Definition 5.8. *Let ρ_K denote the diameter of the largest sphere inscribed in the tetrahedron K. We say that the partitions τ_h are regular as $h \to 0$ if there exist constants $c, h_0 > 0$ such that $\max_{K\in\tau_h}(h_K/\rho_K) \leq c$ for all $h \in (0, h_0)$.*

Analogously to [13, Theorems 7.17, 7.18 and 11.11], we can prove

Lemma 5.9. *Suppose that the partitions τ_h are regular. Then the subspaces $X_{h,0}$ posses the property of discrete compactness.*

Finally, the main convergence result is

Theorem 5.10. *Suppose that there only exists the trivial solution to the homogeneous variational equation (4.8) and that the partitions τ_h of Ω are regular. Then the finite element method (5.1) with Nédélec's edge elements coupled to truncated Rayleigh series expansions converges.*

Proof. Define the discrete subspaces $\mathbb{H}_{h,l} := X_{h,l} \times Y^{+}_{h,l} \times Y^{-}_{h,l} \subset \mathbb{H}_h$ for $l = 0, 1$. Let $P^{Y^{\pm}_{h,l}}\colon \mathbb{H} \longrightarrow Y^{\pm}_{h,l}$, $P^{X_{h,l}}\colon \mathbb{H} \longrightarrow X_{h,l}$ and $P^{\mathbb{H}_{h,l}}\colon \mathbb{H} \longrightarrow \mathbb{H}_{h,l}$ be the orthogonal projections. Note that $P^{X_{h,1}} \longrightarrow P^{X_1}$ as $h \to 0$, where P^{X_1} is the orthogonal projection from $\mathbb{H}$ to X_1. Indeed, for $\nabla p \in X_1$, the problem of finding $\nabla p_h \in X_{h,1}$ such that $\langle \nabla p_h, \nabla q_h\rangle = \langle \nabla p, \nabla q_h\rangle$ for all $\nabla q_h \in X_{h,1}$ corresponds to the finite element scheme for the quasi-periodic boundary value problem of finding $f \in H^1_{qp}(\Omega)$ such that

$$\Delta f = \Delta p \quad \text{in } \Omega, \quad e_3 \cdot \nabla f = e_3 \cdot \nabla p, \quad \text{on } \Gamma^{\pm}_b.$$

This boundary value problem only admits the unique quasiperiodic solution $f = p$ if X does not contain constant functions. If X contains constants, i.e., if the direction of incidence is $\hat{\theta} = (0, 0, -1)^{\top}$, then the finite element scheme $\langle \nabla p_h, \nabla q_h\rangle = \langle \nabla p, \nabla q_h\rangle$ can be considered in the factor space $H^1_{qp}(\Omega)/\mathbb{C}$. In any case, we have $\nabla p_h \to \nabla p$ in $L^2(\Omega)^3$ and $P^{X_{h,1}} \longrightarrow P^{X_1}$ as $h \to 0$. This together with $P^X_h \longrightarrow P^X$ implies the convergence $P^{X_{h,0}} \longrightarrow P^{X_0}$ as $h \to 0$. It is easy to see that $P^{Y^{\pm}_{h,l}} \longrightarrow P^{Y^{\pm}_l}$.

Let operator $A\colon \mathbb{H} \to \mathbb{H}'$ be given as in (4.12). To prove the convergence of the FEM, by Lemma 5.4 we only need to prove the stability of $(P_h)^* A|_{\mathbb{H}_h}$. For clarity, we divide our proof into five steps by introducing several auxiliary operators and then applying Lemma 5.5.

Step 1. Introduce a new operator $B_1\colon \mathbb{H}_1 \to \mathbb{H}_1'$ as

$$\begin{aligned}
&\Big\langle B_1(\nabla p, E_1^+, E_1^-), (\nabla \xi, V_1^+, V_1^-)\Big\rangle \\
&\quad := \Big\langle A|_{\mathbb{H}_1}(\nabla p, E_1^+, E_1^-), (\nabla \xi, V_1^+, V_1^-)\Big\rangle - \tilde{a}_4^+(E_1^+, V_1^+) - \tilde{a}_4^-(E_1^-, V_1^-) \\
&\qquad - a_6^+\big((\nabla p, E_1^+, E_1^-), (\nabla \xi, V_1^+, V_1^-)\big) \; - a_6^-\big((\nabla p, E_1^+, E_1^-), (\nabla \xi, V_1^+, V_1^-)\big).
\end{aligned}$$

where the sesquilinear forms $\tilde{a}_4^{\pm}$ are given in (4.17). Obviously, $-B_1$ is positively coercive over $\mathbb{H}_1$, i.e.

$$\begin{aligned}
&-\mathrm{Re}\Big\langle B_1(\nabla p, E_1^+, E_1^-), (\nabla p, E_1^+, E_1^-)\Big\rangle \\
&\quad = a_1(\nabla p, \nabla p) + [a_4^+(E_1^+, E_1^+) + \tilde{a}_4^+(E_1^+, E_1^+)] + [-a_4^-(E_1^-, E_1^-) + \tilde{a}_4^-(E_1^-, E_1^-)] \\
&\quad \geq c\, \Big(||\nabla p||^2_{H(\mathrm{curl},\Omega)} + ||E_1^+||^2_{H(\mathrm{curl},D^+)} + ||E_1^-||^2_{H(\mathrm{curl},D^-)}\Big)
\end{aligned}$$

for some constant $c > 0$. Thus the operators $[(P_{h,1})^* B_1|_{\mathbb{H}_{h,1}}]$ are stable as the Galerkin approximations of B_1.

Define the operator $B_0 : Z \to Z'$, $Z := \mathbb{H}_0 \times X_1$ by

$$\begin{aligned}&\Big\langle B_0(E_0+\nabla p, E_0^+, E_0^-), (V_0+\nabla\xi, V_0^+, V_0^-)\Big\rangle\\&\qquad = -a_3^+(E_0^+,V_0^+)+\tilde{a}_3^+(E_0^+,V_0^+)+a_5^+(E_0,V_0^+)-\overline{a_5^+(V_0,E_0^+)}\\&\qquad\quad -a_5^-(E_0,V_0^-)+\overline{a_5^-(V_0,E_0^-)}+a_3^-(E_0^-,V_0^-)+\tilde{a}_3^-(E_0^-,V_0^-)\\&\qquad\quad +\int_\Omega\left[\operatorname{curl}E_0\cdot\operatorname{curl}\overline{V}_0+k^2(x)E_0\cdot\overline{V}_0+k^2(x)\nabla p\cdot\nabla\overline{\xi}\right]\,\mathrm{d}x,\end{aligned}$$

with the sesquilinear forms $\tilde{a}_3^\pm$ given in (4.16). From the proof of Theorem 4.1, B_0 is positively coercive over Z, i.e.

$$\begin{aligned}&\operatorname{Re}\Big\langle B_0(E,E_0^+,E_0^-),(E,E_0^+,E_0^-)\Big\rangle\\&\qquad\geq c\,\Big(\|E\|^2_{H(\operatorname{curl},\Omega)}+\|E_0^+\|^2_{H(\operatorname{curl},D^+)}+\|E_0^-\|^2_{H(\operatorname{curl},D^-)}\Big)\end{aligned}\tag{5.4}$$

where $E = E_0 + \nabla p$. Consequently, the operators $(P^{\mathbb{H}_{h,0}})^* B_0|_{\mathbb{H}_{h,0}}$ inherit the coercivity of B_0 in (5.4). Note that, although $\mathbb{H}_{h,0} \subset \mathbb{H}_0$ does not hold in general, we have $\mathbb{H}_{h,0} \subset Z$. Therefore, the $(P^{\mathbb{H}_{h,0}})^* B_0|_{\mathbb{H}_{h,0}} : \mathbb{H}_{h,0} \to \mathbb{H}'_{h,0}$ are stable.

Next, we define the operators $B\colon \mathbb{H} \to \mathbb{H}'$ and $B_h\colon \mathbb{H}_h \to \mathbb{H}'_h$ as follows:

$$\begin{aligned}&\Big\langle B(E,E^+,E^-),(E,E^+,E^-)\Big\rangle\\&\qquad:=\Big\langle B_0(E_0,E_0^+,E_0^-),(V_0,V_0^+,V_0^-)\Big\rangle+\Big\langle B_1(\nabla p,E_1^+,E_1^-),(\nabla p,E_1^+,E_1^-)\Big\rangle,\end{aligned}\tag{5.5}$$

$$B_h(E_h,E_h^+,E_h^-) := \begin{pmatrix}(P^{\mathbb{H}_{h,0}})^*B_0|_{\mathbb{H}_{h,0}} & 0\\ 0 & (P^{\mathbb{H}_{h,1}})^*B_1|_{\mathbb{H}_{h,1}}\end{pmatrix}\begin{pmatrix}(E_{h,0},E_{h,0}^+,E_{h,0}^-)\\(\nabla p_h,E_{h,1}^+,E_{h,1}^-)\end{pmatrix}.$$

Obviously, the B_h are stable and the limit operator $\lim_{h\to 0} B_h P_h$ is equal to B. If we introduce the operators $T_j : \mathbb{H} \to \mathbb{H}'$, $j = 0, 1$, by

$$\begin{aligned}\Big\langle T_0(E,E^+,E^-),(V,V^+,V^-)\Big\rangle &:= -2\int_\Omega k^2(x)E\cdot\overline{V}\mathrm{d}x,\\ \Big\langle T_1(E,E^+,E^-),(V,V^+,V^-)\Big\rangle &:=\\ &-\tilde{a}_3^+(E_0^+,V_0^+)-\tilde{a}_3^-(E_0^-,V_0^-)+\tilde{a}_4^+(E_1^+,V_1^+)+\tilde{a}_4^-(E_1^-,V_1^-)\\ &+a_5^+(E_0,V_1^+)-a_5^-(E_0,V_1^-)-\overline{a_5^+(V_0,E_1^+)}+\overline{a_5^-(V_0,E_1^-)}\\ &+a_6^+\big((E,E^+,E^-),(V,V^+,V^-)\big)+a_6^-\big((E,E^+,E^-),(V,V^+,V^-)\big),\end{aligned}$$

then we arrive at

$$(P_h)^*A|_{\mathbb{H}_h} = B_h + (P_h)^*T_1|_{\mathbb{H}_h} + \begin{pmatrix}(P^{\mathbb{H}_{h,0}})^*T_0|_{\mathbb{H}_{h,0}} & (P_h^{\mathbb{H}_0})^*(A-T_1)|_{\mathbb{H}_{h,1}}\\ (P^{\mathbb{H}_{h,1}})^*(A-T_1)|_{\mathbb{H}_{h,0}} & 0\end{pmatrix}.\tag{5.6}$$

Step 2. It is easy to see that T_1 is compact over $\mathbb{H}$, and the term $(P_h)^*T_1|_{\mathbb{H}_h}$ can be treated by Lemma 5.5. Next we show that $(P^{\mathbb{H}_{h,0}})^*T_0|_{\mathbb{H}_{h,0}}$ can be treated by Lemma 5.5 as well. Denote by Π the orthogonal projection from the space X into X_0 with respect to the inner product $\langle E, V\rangle_X = \int_\Omega \{\operatorname{curl} E \cdot \overline{\operatorname{curl} V} + E \cdot \overline{V}\}\mathrm{d}x$. Then, Π is also an orthogonal projection in the $L^2(\Omega)^3$ sense. Moreover, by the proof of Lemma 4.3, the operator $I - \Pi\colon X \longrightarrow X_1$ is an orthogonal projection too. By the definitions of T_0 and Π,

$$\begin{aligned} \left[(P^{\mathbb{H}_{h,0}})^* T_0|_{\mathbb{H}_{h,0}}\right] P^{\mathbb{H}_{h,0}}|_{\mathbb{H}_h} &= (P_h)^* T_2|_{\mathbb{H}_h} + D_h^{(0)} + D_h^{(1)}, \\ D_h^{(0)} &:= -2(P^{X_{h,0}})^*[k^2(x)(I-\Pi)]|_{X_{h,0}}\, P^{X_{h,0}}|_{\mathbb{H}_h}, \\ D_h^{(1)} &:= -2(P_h)^*(P^{X_{h,0}} - P^{X_0})^*[k^2(x)\Pi]|_{X_{h,0}}\, P^{X_{h,0}}|_{\mathbb{H}_h} \\ &\quad\; -2(P_h)^*(P^{X_0})^*[k^2(x)\Pi]|_{X_{h,0}}(P^{X_{h,0}} - P^{X_0})|_{\mathbb{H}_h}, \\ T_2 &:= -2(P^{X_0})^*[k^2(x)\Pi]P^{X_0}. \end{aligned} \tag{5.7}$$

Here T_2 is compact due to Lemma 4.3. Again $(P_h)^* T_2|_{\mathbb{H}_h}$ can be treated by Lemma 5.5 and it remains to show $\|D_h^{(j)}\|_{\mathbb{H}_h \to \mathbb{H}_h'} \to 0$ for $j = 0, 1$.

The convergence $\|D_h^{(1)}\| \to 0$ follows easily since $[k^2(x)\Pi]\colon X \to X'$ is compact and since $(P^{X_{h,0}} - P^{X_0}) \to 0$. Consequently, it remains to prove that $\|D_h^{(0)}\| \to 0$ as $h \to 0$. It suffices to show that $\|(I-\Pi)|_{X_{h,0}}\|_{X_{h,0}\to X'} \to 0$ with $h \to 0$, i.e., that, for any sequence $\|(I-\Pi)|_{X_{h_n,0}}\|_{X_{h_n,0}\to X'}$ with $h_n \to 0$, there is a subsequence tending to zero. Choose $E_{h_n,0} \in X_{h_n,0}$ such that $\|E_{h_n,0}\|_X = 1$ and $\|(I-\Pi)E_{h_n,0}\|_{X'} = \|(I-\Pi)|_{X_{h_n,0}}\|_{X_{h_n,0}\to X'}$. Recalling Lemma 5.9, without loss of generality we can assume the convergence $E_{h_n,0} \longrightarrow E_0 \in X_0$ in $L^2(\Omega)^3$. Since Π is bounded in L^2, we have $(I-\Pi)E_{h_n,0} \to (I-\Pi)E_0 = 0$ in $L^2(\Omega)^3$. Noting that $X \subseteq L^2(\Omega)^3$ and $L^2(\Omega)^3 \subseteq X'$, we finally conclude

$$\|(I-\Pi)E_{h_n,0}\|_{X'} \le \|(I-\Pi)E_{h_n,0}\|_{L^2(\Omega)^3} \to \|(I-\Pi)E_0\|_{L^2(\Omega)^3} = 0.$$

This gives $||D_h^{(0)}|| \to 0$ as $h \to 0$.

Step 3. For $E_h, V_h \in X_h$, recall the decompositions

$$\begin{aligned} E_h &= E_{h,0} + \nabla p_h = \Pi(E_{h,0}) + (I-\Pi)(E_{h,0}) + \nabla p_h, \\ V_h &= V_{h,0} + \nabla \xi_h = \Pi(V_{h,0}) + (I-\Pi)(V_{h,0}) + \nabla \xi_h, \end{aligned}$$

with $E_{h,0}, V_{h,0} \in X_{h,0}$ and $\nabla p_h, \nabla\xi_h \in X_{h,1}$. We set $T := T_0 + T_1$ and claim that

$$(P_h)^* A|_{\mathbb{H}_h} = B_h + (P_h)^* T|_{\mathbb{H}_h} + D_h^{(0)} + D_h^{(1)} + D_h^{(2)}, \tag{5.8}$$

where $D_h^{(2)} : \mathbb{H}_h \to \mathbb{H}_h'$ is defined by

$$\begin{aligned} &\left\langle D_h^{(2)}(E_h, E_h^+, E_h^-), (V_h, V_h^+, V_h^-)\right\rangle \\ &\qquad := \overline{a_5^+\left((I-\Pi)(E_{h,0}), V_{h,1}^+\right)} - \overline{a_5^-\left((I-\Pi)(E_{h,0}), V_{h,1}^-\right)} \\ &\qquad\quad\; - a_5^+\left((I-\Pi)(V_{h,0}), E_{h,1}^+\right) + a_5^-\left((I-\Pi)(V_{h,0}), E_{h,1}^-\right). \end{aligned}$$

In fact, the formulas (5.6) and (5.7) imply (5.8) if we can show that the operator $D_h^{(2)}$ is the off-diagonal part of the matrix on the right-hand side of (5.6). Hence, it suffices to prove $D_h^{(2)} =$

$[(P^{\mathbb{H}_{h,0}})^*(A-T_1)|_{\mathbb{H}_{h,1}}]P^{\mathbb{H}_{h,1}} + [(P^{\mathbb{H}_{h,1}})^*(A-T_1)|_{\mathbb{H}_{h,0}}]P^{\mathbb{H}_{h,0}}$. We conclude

$$\begin{aligned}
&\Big\langle (P^{\mathbb{H}_{h,1}})^*(A-T_1)|_{\mathbb{H}_{h,0}}(E_h, E_h^+, E_h^-), (V_h, V_h^+, V_h^-)\Big\rangle \\
&\quad = -a_1\left((I-\Pi)(E_{h,0}), \nabla\xi_h\right) + a_5^+\left((I-\Pi)(E_{h,0}), V_{h,1}^+\right) - a_5^-\left((I-\Pi)(E_{h,0}), V_{h,1}^-\right) \\
&\quad = \ a_5^+\left((I-\Pi)(E_{h,0}), V_{h,1}^+\right) - a_5^-\left((I-\Pi)(E_{h,0}), V_{h,1}^-\right),
\end{aligned} \tag{5.9}$$

where we have used the identity

$$a_1\left((I-\Pi)(E_{h,0}), \nabla\xi_h\right) = \int_\Omega k^2(x) E_{h,0}\cdot\overline{\nabla\xi_h}\mathrm{d}x - \int_\Omega k^2(x)\Pi(E_{h,0})\cdot\overline{\nabla\xi_h}\mathrm{d}x = 0.$$

Analogously, it can be seen that

$$\begin{aligned}
&\Big\langle (P^{\mathbb{H}_{h,0}})^*(A-T_1)|_{\mathbb{H}_{h,1}}(E_h, E_h^+, E_h^-), (V_h, V_h^+, V_h^-)\Big\rangle \\
&\qquad = -\overline{a_5^+\left((I-\Pi)(V_{h,0}), E_{h,1}^+\right)} + \overline{a_5^-\left((I-\Pi)(V_{h,0}), E_{h,1}^-\right)}.
\end{aligned} \tag{5.10}$$

Eqns. (5.9) and (5.10) imply that $[(P^{\mathbb{H}_{h,0}})^*(A-T_1)|_{\mathbb{H}_{h,1}}]P^{\mathbb{H}_{h,1}} + [(P^{\mathbb{H}_{h,1}})^*(A-T_1)|_{\mathbb{H}_{h,0}}]P^{\mathbb{H}_{h,0}}$ coincides with $D_h^{(2)}$. Formula (5.8) is thus proven.

Step 4. We shall prove $||D_h^{(2)}|| \to 0$. First we derive $||(P^{\mathbb{H}_{h,1}})^*(A-T_1)|_{\mathbb{H}_{h,0}}|| \to 0$. By (5.9), we choose functions $E_{h,0}$ and $V_{h,1}^\pm$ with $||E_{h,0}||_{H(\mathrm{curl}\,,\Omega)} = 1$, $||V_{h,1}^\pm||_{H(\mathrm{curl}\,,D^\pm)} = 1$ such that

$$||(P^{\mathbb{H}_{h,1}})^*(A-T_1)|_{\mathbb{H}_{h,0}}|| = a_5^+(\nabla q_h, V_{h,1}^+) - a_5^-(\nabla q_h, V_{h,1}^-), \quad \nabla q_h := (I-\Pi)E_{h,0}.$$

Using the definition of a_5^+, we get

$$\begin{aligned}
a_5^+(\nabla q_h, V_{h,1}^+) &= -\int_{\Gamma_b^+} e_3\times\nabla q_h \ \cdot\ \overline{\mathrm{curl}\, V_{h,1}^+}\,\mathrm{ds}, \\
|a_5^+(\nabla q_h, V_{h,1}^+)| &\le ||e_3\times\nabla q_h||_{H_t^{-1/2}(\Gamma_b^+)}\, ||\mathrm{curl}\, V_{h,1}^+||_{H_t^{1/2}(\Gamma_b^+)}.
\end{aligned} \tag{5.11}$$

On the one hand, we have for any $a\in\mathbb{C}$,

$$\begin{aligned}
||e_3\times\nabla q_h||_{H_t^{-1/2}(\Gamma_b^+)} &= ||\nabla_{\Gamma_b^+}(q_h+a)||_{H_t^{-1/2}(\Gamma_b^+)} \le c\,||q_h+a||_{H^{1/2}(\Gamma_b^+)} \\
&\le c\,||q_h+a||_{H^1(\Omega)},
\end{aligned}$$

where $\nabla_{\Gamma_b^+}$ denotes the surface gradient operator over Γ_b^+. Hence,

$$||e_3\times\nabla q_h||_{H_t^{-1/2}(\Gamma_b^+)} \le c\,\inf_{a\in\mathbb{C}} ||q_h+a||_{H^1(\Omega)} \le c\,||\nabla q_h||_{L^2(\Omega)^3}. \tag{5.12}$$

On the other hand, for $V_{h,1}^+ = \sum_{n:\,|n|\le C/|h|} c_n U_{n,1}^+ \in Y_1^+$, there holds

$$||\mathrm{curl}\, V_{h,1}^+||_{H_t^{1/2}(\Gamma_b^+)} \le ||\mathrm{curl}\, V_{h,1}^+||_{H^1(D^+)^3} = \Big\|\sum_{n:\,|n|\le C/|h|} c_n\,\mathrm{curl}\, U_{n,1}^+\Big\|_{H^1(D^+)^3}. \tag{5.13}$$

In view of the identity $\operatorname{curl} U^+_{n,1} = -i(k^+)^2/\sqrt{|\alpha_n|^2+|\beta^+_n|^2}\; U^+_{n,0}$ (see [10, Lemma 3.1]) and the relation $\sqrt{|\alpha_n|^2+|\beta^+_n|^2} = \mathcal{O}(|n|)$ as $|n|\to\infty$, we get

$$\begin{aligned}
\Big\| \sum_{n:\,|n|\le C/|h|} c_n \operatorname{curl} U^+_{n,1} \Big\|_{H^1(D^+)^3} &\le c \left(\sum_{n:\,|n|\le C/|h|} |c_n|^2 |n|^{-2} ||U^+_{n,0}||_{H^1(D^+)^3} \right)^{1/2} \\
&\le c \left(\sum_{n:\,|n|\le C/|h|} |c_n|^2\, ||U^+_{n,0}||_{L^2(D^+)^3} \right)^{1/2} \\
&\le c \left(\sum_{n:\,|n|\le C/|h|} |c_n|^2\, ||U^+_{n,1}||_{L^2(D^+)^3} \right)^{1/2},
\end{aligned}$$

where the last two equalities follow from the estimates derived in the proof of [10, Lemma 4.5]. Recalling (5.13) and the representation of $V^+_{h,1}$ as an expansion with respect to the basis functions $U^+_{n,1}$, we obtain

$$||\operatorname{curl} V^+_{h,1}||_{H^{1/2}_t(\Gamma^+_b)} \le c\, ||V^+_{h,1}||_{L^2(D^+)^3} \le c\, ||V^+_{h,1}||_{H(\operatorname{curl},D^+)} \le c. \tag{5.14}$$

Inserting the estimates (5.12) and (5.14) into (5.11) yields $|a^+_5(\nabla q_h, V^+_{h,1})| \le c^+_5 ||\nabla q_h||_{L^2(\Omega)^3}$ for some $c^+_5 > 0$, and analogously, there exists another non-negative constant c^-_5 such that $|a^-_5(\nabla q_h, V^-_{h,1})| \le c^-_5\, ||\nabla q_h||_{L^2(\Omega)^3}$. Thus, to prove $||(P^{\mathbb{H}_{h,1}})^*(A-T_1)|_{\mathbb{H}_{h,0}}|| \to 0$, we only need to verify $||\nabla q_h||_{L^2(\Omega)^3} \to 0$ as $h \to 0$. However, we can choose $E_{h,0}$ with $\|E_{h,0}\|_X = 1$ such that $||(P^{\mathbb{H}_{h,1}})^*(A-T_1)E_{h,0}|| = ||(P^{\mathbb{H}_{h,1}})^*(A-T_1)|_{\mathbb{H}_{h,0}}||$. From the discrete compactness of the space $X_{h,0}$ in Lemma 5.9, for any sequence $E_{h_n,0}$, we can always find a subsequence converging in $L^2(\Omega)^3$ to an $E_0 \in X_0$. We denote this subsequence again by $E_{h_n,0}$. Then $||\nabla q_{h_n}||_{L^2(\Omega)^3} = ||(I-\Pi)E_{h_n,0}||_{L^2(\Omega)^3} \to ||(I-\Pi)E_0||_{L^2(\Omega)^3} = 0$. In other words, any sequence $||(P^{\mathbb{H}_{h_n,1}})^*(A-T_1)E_{h_n,0}||$ has a subsequence tending to zero. Consequently, $||(P^{\mathbb{H}_{h,1}})^*(A-T_1)E_{h,0}||$ converges to zero.

Arguing analogously, one can prove the convergence $||(P^{\mathbb{H}_{h,0}})^*(A-T_1)|_{\mathbb{H}_{h,1}}|| \to 0$ as $h \to 0$ via the identity (5.10). Hence, it holds that $||D^{(2)}_h|| \to 0$.

Step 5. Setting $D_h := D^{(0)}_h + D^{(1)}_h + D^{(2)}_h$, Equ. (5.8) is the representation of Lemma 5.5. It can be concluded from Steps 1-4 that the B_h are stable operators, T is compact and that D_h is only a small perturbation. By the uniqueness assumption in Theorem 5.10, we see from Theorem 3.1 that A is invertible. Now, applying Lemma 5.5 yields the stability of $(P_h)^*A|_{\mathbb{H}_h}$. The proof of the convergence of the FEM is thus completed. ☐

6 Multi-layer system beneath the grating structure

In many applications there is an adjacent multi-layer system beneath the lower face $x_3 = -b$ of the grating. More precisely (cf. Fig. 2), for a sequence b_{k}, $\mathsf{k} = 0,\dots,K$ of x_3-coordinates such that $-b = b_0 > b_1 > \dots > b_K$, the function $\epsilon(x) + i\sigma(x)/\omega$ in the layer $b_{\mathsf{k}-1} > x_3 > b_{\mathsf{k}}$ takes

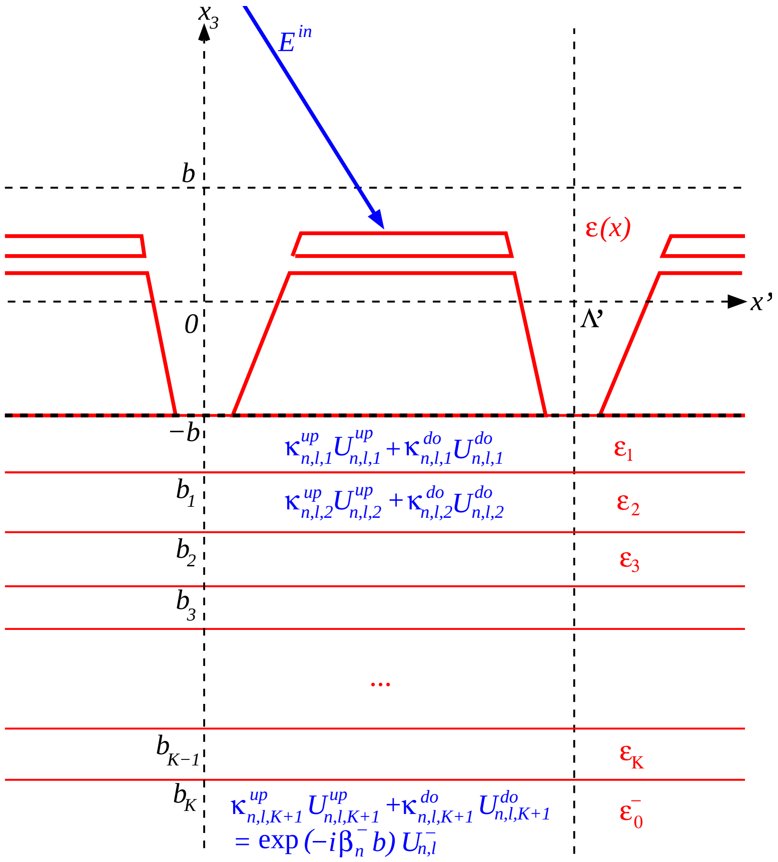

Figure 2: Grating with multi-layer system.

the constant value ϵ_{k} with $\operatorname{Im}\epsilon_{\mathsf{k}} \geq 0$ such that $\operatorname{Re}\epsilon_{\mathsf{k}} > 0$ for $\operatorname{Im}\epsilon_{\mathsf{k}} = 0$. Of course, in the lower half space $b_K > x_3$ we suppose $\epsilon(x) + i\sigma(x)/\omega = \epsilon_0^-$.

For a variational formulation adapted to the multi-layer system, we need modified spaces Y_l^-, $l=0,1$. Clearly, the tangential traces of E and $\operatorname{curl} E$ are continuous over the interfaces $x_3 = b_{\mathsf{k}}$. Solving these transmission problems, each downward propagating $E = \exp(-i\beta_n^- b)U_{n,l}^-$ in the half space $b_K > x_3$ corresponds to an extended field E in $b_0 > x_3$ such that $E(x) = \kappa^{\mathrm{up}}_{n,l,\mathsf{k}}U^{\mathrm{up}}_{n,l,\mathsf{k}}(x) + \kappa^{\mathrm{do}}_{n,l,\mathsf{k}}U^{\mathrm{do}}_{n,l,\mathsf{k}}(x)$ for $b_{\mathsf{k}-1} > x_3 > b_{\mathsf{k}},\ \mathsf{k} = 1, 2, \ldots, K$, where $\kappa^{\mathrm{up}}_{n,l,\mathsf{k}}, \kappa^{\mathrm{do}}_{n,l,\mathsf{k}} \in \mathbb{C}$ and

$$U^{\mathrm{up}}_{n,0,\mathsf{k}}(x) := e^{i[\alpha_n\cdot x' + \beta_{n,\mathsf{k}}(x_3+b)]} \begin{cases} (0,-1,0)^\top & \text{if } |\alpha_n| = 0 \\ \frac{1+i(x_3+b)}{|\alpha_n|}(-\alpha_n^{(2)}, \alpha_n^{(1)}, 0)^\top & \text{if } \beta_{n,\mathsf{k}} = 0 \\ \frac{1}{|\alpha_n|}(-\alpha_n^{(2)}, \alpha_n^{(1)}, 0)^\top & \text{else} \end{cases},$$

$$U^{\mathrm{up}}_{n,1,\mathsf{k}}(x) := e^{i[\alpha_n\cdot x' + \beta_{n,\mathsf{k}}(x_3+b)]} \begin{cases} (1,0,0)^\top & \text{if } |\alpha_n| = 0 \\ \frac{1}{\sqrt{|\alpha_n|^2+|\alpha_n|^4}}\Big(-\alpha_n, |\alpha_n|^2\big(1+i(x_3+b)\big)\Big)^\top & \text{if } \beta_{n,\mathsf{k}} = 0 \\ \frac{1}{|\alpha_n|\sqrt{|\alpha_n|^2+|\beta_{n,\mathsf{k}}|^2}}(-\beta_{n,\mathsf{k}}\alpha_n, |\alpha_n|^2)^\top & \text{else} \end{cases},$$

$$U^{\mathrm{do}}_{n,0,\mathsf{k}}(x) := e^{i[\alpha_n\cdot x'-\beta_{n,\mathsf{k}}(x_3+b)]}\begin{cases}(0,1,0)^\top & \text{if } |\alpha_n|=0\\ \frac{1-i(x_3+b)}{|\alpha_n|}(\alpha_n^{(2)},-\alpha_n^{(1)},0)^\top & \text{if } \beta_{n,\mathsf{k}}=0\\ \frac{1}{|\alpha_n|}(\alpha_n^{(2)},-\alpha_n^{(1)},0)^\top & \text{else}\end{cases},$$

$$U^{\mathrm{do}}_{n,1,\mathsf{k}}(x) := e^{i[\alpha_n\cdot x'-\beta_{n,\mathsf{k}}(x_3+b)]}\begin{cases}(-1,0,0)^\top & \text{if } |\alpha_n|=0\\ \frac{1}{\sqrt{|\alpha_n|^2+|\alpha_n|^4}}\Big(\alpha_n,|\alpha_n|^2\big(1-i(x_3+b)\big)\Big)^\top & \text{if } \beta_{n,\mathsf{k}}=0\\ \frac{1}{|\alpha_n|\sqrt{|\alpha_n|^2+|\beta_{n,\mathsf{k}}|^2}}(\beta_{n,\mathsf{k}}\alpha_n,|\alpha_n|^2)^\top & \text{else}\end{cases},$$

$$\beta_{n,\mathsf{k}} := \sqrt{\omega^2\mu_0\epsilon_{\mathsf{k}}-|\alpha_n|^2}, \quad \beta_{n,K+1} := \beta_n^-.$$

Fix n and l. It is not hard to see (cf. [15, Sect. III.4]) that, for each linear combination of $U^{\mathrm{up}}_{n,l,K+1}$ and $U^{\mathrm{do}}_{n,l,K+1}$ in the half space $x_3<b_K$, there exist unique linear combinations of the $U^{\mathrm{up}}_{n,l,\mathsf{k}}$ and $U^{\mathrm{do}}_{n,l,\mathsf{k}}$ in the layers $b_{\mathsf{k}}<x_3<b_{\mathsf{k}-1},\ \mathsf{k}=1,\dots,K$ such that the tangential traces over the interfaces $x_3=b_{\mathsf{k}},\ \mathsf{k}=1,\dots,K$ of the functions and of their curls in the adjacent layers coincide. Similarly, to each linear combination of $U^{\mathrm{up}}_{n,l,1}$ and $U^{\mathrm{do}}_{n,l,1}$ in the layer $b_1<x_3<b_0$ there exist unique linear combinations of the $U^{\mathrm{up}}_{n,l,\mathsf{k}}$ and $U^{\mathrm{do}}_{n,l,\mathsf{k}}$ in the layers $b_{\mathsf{k}}<x_3<b_{\mathsf{k}-1},\ \mathsf{k}=2,\dots,K$ and in the half space $x_3<b_K$ such that the tangential traces of the functions and of their curls in adjacent layers coincide. Hence, the coefficients $\kappa^{\mathrm{up}}_{n,l,\mathsf{k}},\kappa^{\mathrm{do}}_{n,l,\mathsf{k}}$ are uniquely determined. For instance, if all the $\beta_{n,\mathsf{k}}$ are non-zero and $|\alpha_n|\neq 0$, then

$$\begin{pmatrix}\kappa^{\mathrm{up}}_{n,l,1}\\ \kappa^{\mathrm{do}}_{n,l,1}\end{pmatrix} = \mathcal{M}_{n,l,1}\,\mathcal{M}_{n,l,2}\,\dots\,\mathcal{M}_{n,l,K}\begin{pmatrix}0\\1\end{pmatrix}, \tag{6.1}$$

$$\mathcal{M}_{n,0,\mathsf{k}} := \begin{pmatrix}\frac{\beta_{n,\mathsf{k}+1}+\beta_{n,\mathsf{k}}}{2\beta_{n,\mathsf{k}}}e^{i[\beta_{n,\mathsf{k}+1}-\beta_{n,\mathsf{k}}]b_{\mathsf{k}}} & \frac{\beta_{n,\mathsf{k}+1}-\beta_{n,\mathsf{k}}}{2\beta_{n,\mathsf{k}}}e^{-i[\beta_{n,\mathsf{k}+1}+\beta_{n,\mathsf{k}}]b_{\mathsf{k}}}\\ \frac{\beta_{n,\mathsf{k}+1}-\beta_{n,\mathsf{k}}}{2\beta_{n,\mathsf{k}}}e^{i[\beta_{n,\mathsf{k}+1}+\beta_{n,\mathsf{k}}]b_{\mathsf{k}}} & \frac{\beta_{n,\mathsf{k}+1}+\beta_{n,\mathsf{k}}}{2\beta_{n,\mathsf{k}}}e^{-i[\beta_{n,\mathsf{k}+1}-\beta_{n,\mathsf{k}}]b_{\mathsf{k}}}\end{pmatrix},$$

$$\mathcal{M}_{n,1,\mathsf{k}} := \sqrt{\frac{|\alpha_n|^2+|\beta_{n,\mathsf{k}}|^2}{|\alpha_n|^2+|\beta_{n,\mathsf{k}+1}|^2}}$$

$$\begin{pmatrix}\left[\frac{|\alpha_n|^2+\beta^2_{n,\mathsf{k}+1}}{|\alpha_n|^2+\beta^2_{n,\mathsf{k}}}+\frac{\beta_{n,\mathsf{k}+1}}{\beta_{n,\mathsf{k}}}\right]e^{i[\beta_{n,\mathsf{k}+1}-\beta_{n,\mathsf{k}}]b_{\mathsf{k}}} & \left[\frac{|\alpha_n|^2+\beta^2_{n,\mathsf{k}+1}}{|\alpha_n|^2+\beta^2_{n,\mathsf{k}}}-\frac{\beta_{n,\mathsf{k}+1}}{\beta_{n,\mathsf{k}}}\right]e^{-i[\beta_{n,\mathsf{k}+1}+\beta_{n,\mathsf{k}}]b_{\mathsf{k}}}\\ \left[\frac{|\alpha_n|^2+\beta^2_{n,\mathsf{k}+1}}{|\alpha_n|^2+\beta^2_{n,\mathsf{k}}}-\frac{\beta_{n,\mathsf{k}+1}}{\beta_{n,\mathsf{k}}}\right]e^{i[\beta_{n,\mathsf{k}+1}+\beta_{n,\mathsf{k}}]b_{\mathsf{k}}} & \left[\frac{|\alpha_n|^2+\beta^2_{n,\mathsf{k}+1}}{|\alpha_n|^2+\beta^2_{n,\mathsf{k}}}+\frac{\beta_{n,\mathsf{k}+1}}{\beta_{n,\mathsf{k}}}\right]e^{-i[\beta_{n,\mathsf{k}+1}-\beta_{n,\mathsf{k}}]b_{\mathsf{k}}}\end{pmatrix}.$$

Note that the coefficients $\kappa^{\mathrm{up}}_{n,l,1}$ and $\kappa^{\mathrm{do}}_{n,l,1}$ can be computed by numerically stable algorithms (cf. e.g. [15, Sect. III.6]).

Setting $\widetilde{U}^-_{n,l} := \kappa^{\mathrm{do}}_{n,l,1}U^{\mathrm{do}}_{n,l,1}+\kappa^{\mathrm{up}}_{n,l,1}U^{\mathrm{up}}_{n,l,1}$, we define the modified spaces Y^-_l by (4.5) but with $U^-_{n,l}$ replaced by $\widetilde{U}^-_{n,l}$. Now the new variational formulation for the transmission problem is just (4.8) with a modified (4.7) defined over $\mathbb{H} := X\times Y^+\times(Y^-_0\oplus Y^-_1)$ including the modified spaces Y^-_l. The modified sesquilinear form is the sum of (4.7) and the additional term

$$-\eta^-\sum_{l=0}^{1}\ \sum_{n:\,e_3\times\widetilde{U}^-_{n,l}=0}\left[\int_{\Gamma^-_b}e_3\times(E-E^-)\cdot(e_3\times\overline{U}^-_{n,l})\,\mathrm{d}s\ \overline{\int_{\Gamma^-_b}(\operatorname{curl}V^-)\cdot(\operatorname{curl}\overline{U}^-_{n,l})\,\mathrm{d}s}\right].$$

.

Remark 6.1. *All the results for the variational formulation and for the FEM coupled by the wave modes remain true for the case of multi-layer systems beneath the grating structure and the new variational form.*

Indeed, we sketch the proof now. From the definitions of the $U^{\mathrm{up}}_{n,l,1}$ and $U^{\mathrm{do}}_{n,l,1}$, we observe $e_3 \times U^{\mathrm{up}}_{n,l,1} = -e_3 \times U^{\mathrm{do}}_{n,l,1}$ and $(\operatorname{curl} U^{\mathrm{up}}_{n,l,1})_T = (\operatorname{curl} U^{\mathrm{do}}_{n,l,1})_T$ over the curve Γ^-_b. Consequently, the traces entering the sesquilinear forms satisfy

$$\begin{aligned} e_3 \times \widetilde{U}^-_{n,l} &= [\kappa^{\mathrm{do}}_{n,l,1} - \kappa^{\mathrm{up}}_{n,l,1}] \;\; e_3 \times U^{\mathrm{do}}_{n,l,1}, \\ (\operatorname{curl} \widetilde{U}^-_{n,l})_T &= [\kappa^{\mathrm{do}}_{n,l,1} + \kappa^{\mathrm{up}}_{n,l,1}] \; (\operatorname{curl} U^{\mathrm{do}}_{n,l,1})_T. \end{aligned} \tag{6.2}$$

If $\beta_{n,1} = 0$, then $[\kappa^{\mathrm{do}}_{n,1,1} - \kappa^{\mathrm{up}}_{n,1,1}] \neq 0$ since otherwise $e_3 \times \widetilde{U}^-_{n,1} = 0$, which together with $(\operatorname{curl} \widetilde{U}^-_{n,1})_T = 0$ would contradict to the one-to-one mapping between the linear combinations of wave modes mentioned above. This fact and the special choice of the additional term in the modified sesquilinear form guarantee (cf. [10, proof of Lemma 3.3]) the equivalence of the boundary value problem and the variational equation in the case of multi-layer systems.

Fredholm property with index zero for the variational operator and convergence of the FEM coupled by wave modes follow from the fact that the operator corresponding to the modified variational form is a compact perturbation of that of the original form. To see this fact, we observe $\beta_{n,\mathsf{k}}/|n| \to i$ for $|n| \to \infty$ and $\beta_{n,\mathsf{k}} - \beta_{n,\mathsf{k}+1} = (k^2_{\mathsf{k}} - k^2_{\mathsf{k}+1})/(\beta_{n,\mathsf{k}} + \beta_{n,\mathsf{k}+1}) \sim |n|^{-1}$ with $k_{\mathsf{k}} := \omega\sqrt{\epsilon_{\mathsf{k}}\mu_0}$. Consequently, Equ. (6.1) implies $\kappa^{\mathrm{do}}_{n,l,1} \to 1$, $\kappa^{\mathrm{up}}_{n,l,1} \to 0$ and $[\kappa^{\mathrm{do}}_{n,l,1} \pm \kappa^{\mathrm{up}}_{n,l,1}] \to 1$ for the factors in (6.2). In other words, the difference between the modified operator and the original is the multiplication by operators represented with respect to the wave mode basis by the diagonal matrices $([\kappa^{\mathrm{do}}_{n,l,1} \pm \kappa^{\mathrm{up}}_{n,l,1}]\delta_{n,n'})_{n,n'}$. In view of $U^{\mathrm{do}}_{n,l,1} = \exp(-i\beta^-_n b) U^-_{n,l}$ and

$$\left\| \sum_{n \in \mathbb{Z}^2} \sum_{l=0}^{1} c_{n,l} U^-_{n,l} \right\|_{H(\operatorname{curl}, D^-)} \sim \left(\sum_{n \in \mathbb{Z}^2} \sum_{l=0}^{1} e^{-2|n|b} \, \frac{1 + |n|^{2l}}{1 + |n|} \, |c_{n,l}|^2 \right)^{1/2}$$

(cf. [10, Lemma 3.1]) such a diagonal operator is a compact perturbation of the identity.

7 Numerical example

For a simple numerical test we consider two profile gratings on the surface of a SiO$_2$ body. The echelle grating (cf. the left of Fig. 3) is designed to deflect light into the direction specular with respect to the inclined upper faces. The idea of blaces (cf. the right of Fig. 3) with the width b less and the length l larger than the wavelength of light λ, is to provide a similar effective medium distribution and to function like an echelle grating. Hopefully, such blaces are of better stability (cf. [9]).

In Table 2 we compare the new 3D coupling algorithm (4.8) of Sect. 5.1 applied to the 2D echelle grating with the reliable results of the 2D FEM code solving the Helmholtz equation. The efficiencies

$$e^+_n := \frac{\beta^+_n}{\beta^+_{(0,0)}} |E^+_n|^2 \,, \qquad e^+_n := \frac{(k^+)^2}{(k^-)^2} \frac{\beta^-_n}{\beta^+_{(0,0)}} |E^-_n|^2$$

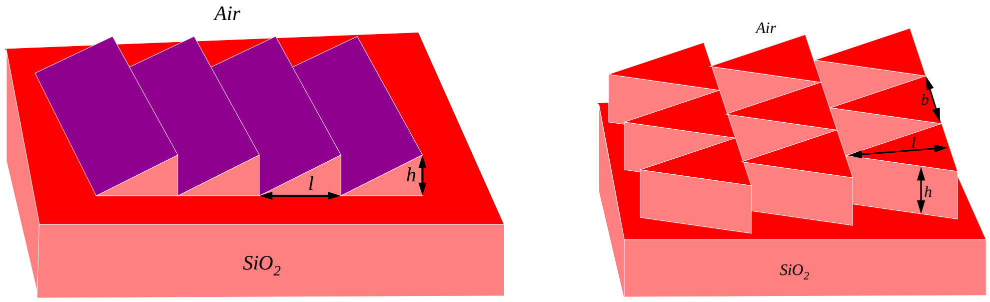

Figure 3: Geometry of grating: left - echelle grating, right - blaces.

meshsize	$e^+_{-2,0}$	$e^+_{0,0}$	$e^-_{1,0}$	$e^-_{2,0}$
125.0 nm	4.82	0.0027	43.23	3.78
62.5 nm	4.530	0.0022	45.0080	4.1289
31.2 nm	4.5039	0.0019	45.0559	4.1142
2D code	4.5025	0.0019	45.0630	4.1145

Table 2: Computation of efficiencies for echelle grating. Comparison of FEM from Sect. 5.1 with two-dimensional FEM simulation.

of the electric field solution are computed for wavelength $\lambda = 500$ nm, period $l = 10\,\mu$m, and height $h = 0.5\,\mu$m. The grating is illuminated exactly from above under TE polarization. The FEM of Sect. 5.1 is applied with quadratic edge elements. The upper coupling modes $n = (n_1, n_2)$ are restricted to $|n_1| \leq 22$ and $|n_2| \leq 2$, the lower modes to $|n_1| \leq 32$ and $|n_2| \leq 2$ Moreover, the coupling parameters $\eta^{\pm}$ are set to zero. For the mesh-size tending to zero, the 3D results converge to those of the 2D simulation. Adding more coupling modes does not improve the accuracy.

Next we apply the same 3D algorithm to the blaces and compare the results with those obtained by the algorithm of Huber et al. (cf. [11]). Here the periods are chosen as $\Lambda_1 = l = 10\,\mu$m and $\Lambda_2 = b = \lambda/2$ and the other parameters like for the echelle grating. The resulting efficiencies coincide upto numerical errors.

meshsize	$e^+_{0,0}$	$e^+_{0,0}$	$e^+_{1,0}$	$e^+_{1,0}$	$e^-_{0,0}$	$e^-_{0,0}$	$e^-_{1,0}$	$e^-_{1,0}$
125.0 nm	2.8328	3.0985	0.1661	0.1661	75.2800	76.289	10.1503	10.1465
62.5 nm	2.8172	2.8333	0.1918	0.1918	75.5412	75.553	10.7248	10.7197
31.2 nm	2.8119	2.8136	0.1944	0.1944	75.4717	75.490	10.7787	10.7711

Table 3: Computation of efficiencies for blaces. Comparison of FEM from Sect. 5.1 (left numbers in column) with FEM of [11] (right numbers).

References

[1] Abboud, T., 1993, Formulation variationnelle des équations de Maxwell dans un réseau bipériodique de $\mathbb{R}^3$, *C.R. Acad. Sci. Paris* **317**, 245–248.

[2] Ammari, H., 1995, Uniqueness theorems for an inverse problem in a doubly periodic structure, *Inverse Problems* **11**, 823–833.

[3] Ammari, H. and Bao G., 2008, Coupling of finite element and boundary element methods for the scattering by periodic chiral structures, *Journal of Computational Mathematics* **3**, 261-283.

[4] Bao G. and Dobson D.C., 2000, On the scattering by biperiodic structures, *Proc. Am. Math. Soc.* **128**, 2715–2723.

[5] Bao G., 1997, Variational approximation of Maxwell's equation in biperiodic structures, *SIAM J. Appl. Math.* **57**, 364–381.

[6] Buffa, A., Costabel, M. and Sheen, D., 2002, On traces for H$(\mathrm{curl}\,,\Omega)$ in Lipschitz domains, *J. Math. Anal. Appl.* **276**, 847-867.

[7] Chen, X. and Friedman, A., 1991, Maxwell's equations in a periodic structure, *Transactions AMS.* **323**, 465–507.

[8] Dobson, D.C., 1994, A variational method for electromagnetic diffraction in biperiodic structures, *RAIRO Model. Math. Anal. Numer.* **28**, 419–439.

[9] Elfström, H., Kuittinen, M., Vallius, T., Kleemann, B.H., Ruoff, J., and Arnold, R., 2006, Fabrication of blased gratings by area-coded effective medium structures, *Optics Communications* **266**, 697–703.

[10] Hu, G. and Rathsfeld, A., 2012, Scattering of time-harmonic electromagnetic plane waves by perfectly conducting diffraction gratings, *WIAS Preprint No.* **1694**, Berlin.

[11] Huber, M., Schoeberl, J., Sinwel, A. and Zaglmayr, S., 2009, Simulation of diffraction in periodic media with a coupled finite element and plane wave approach, *SIAM J. Sci. Comput.* **31**, 1500–1517.

[12] Kress, R., 1989, *Linear Integral Equations*, New York, Springer-Verlag.

[13] Monk P., 2003, *Finite Element Method for Maxwell's Equations* Oxford, Oxford University Press.

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

[15] Nevière, M. and Popov, E., 2003 , *Light Propagation in Periodic Media*, New-York, Basel, Marcel Dekker, Inc.

[16] Nitsche, J., 1970, Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind, *Abh. Math. Sem. Univ. Hamburg,* **36**, 9–15.

[17] Prössdorf S. and Silbermann B., 1991, *Numerical Analysis for Integral and Related Operator Equations*, Berlin, Akademie-Verlag.

[18] Prössdorf S. and Silbermann B., 1977, *Projektionsverfahren und die näherungsweise Lösung singulärer Gleichungen*, Leipzig, Teubner.

[19] Rathsfeld, A., 2011, Shape derivatives for the scattering by biperiodic gratings. *WIAS Preprint No.* **1640**, Berlin.

[20] Schaedle, A., Zschiedrich, L., Burger, S., Klose, R. and Schmidt, F., 2007, Domain decomposition method for Maxwell's equations: Scattering of periodic structures, J. Comput. Physics **226**, 477–493.

[21] Schmidt, G., 2003, On the diffraction by biperiodic anisotropic structures, *Applicable Analysis* **82**, 75–92.

[22] Schmidt, G., 2004, Electromagnetic scattering by periodic structures, *Journal of Mathematical Sciences* **124**, 5390–5405.

[23] Sternberg, R., 1988, Mortaring by a method of J. A. Nitsche, *Computational Mechanics, New trends and Applications,* eds.: S. Idelsohn, E. Onate, E. Dvorkin, Barcelona, CIMNE.