Karlsruhe Institute of Technology

Wave phenomena
analysis and numerics

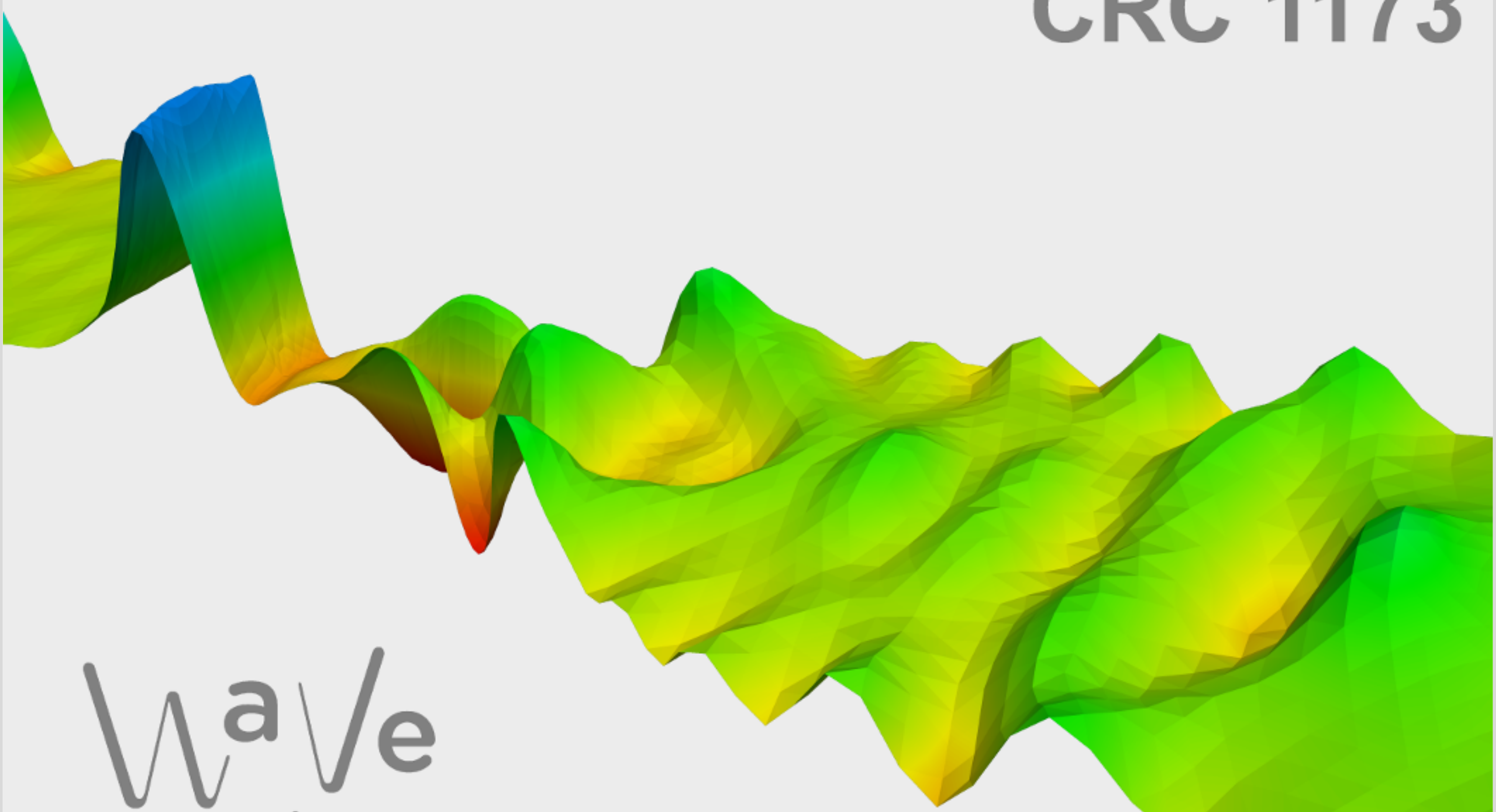
Scattering problems in periodic or locally perturbed periodic domains: Convergence of the perfectly matched layer approximation and efficient numerical solution

Tilo Arens, Nasim Shafieeabyaneh, Ruming Zhang

CRC Preprint 2026/21, June 2026

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



Wave phenomena

Participating universities

UNIVERSITÄT BONN

Universität Stuttgart

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

TU WIEN

Funded by

DFG

ISSN 2365-662X

Scattering Problems in Periodic or Locally Perturbed Periodic Domains: Convergence of the Perfectly Matched Layer Approximation and Efficient Numerical Solution

Tilo Arens*, Nasim Shafieeabyaneh*, and Ruming Zhang†

Abstract

We study scattering of a time-harmonic acoustic wave by a locally perturbed periodic surface, where the incident field need not be periodic. This scattering problem is challenging due to the unboundedness of the domain and the lack of periodicity. Using a coordinate transformation, we remove the perturbation and obtain an equation with non-constant coefficients in a periodic domain. This allows application of the Floquet–Bloch transform, reducing the problem to a coupled family (indexed by the Floquet parameter) of periodic problems in a bounded cell. We approximate the solution with the perfectly matched layer (PML) method and prove the exponential convergence of this approximation on compact sets for an incident spherical wave. Moreover, we show that the Floquet–Bloch transform of the solution of the PML problem is analytic with respect to the Floquet parameter. This enables us to accurately discretize the inverse Floquet–Bloch transform using only a few values of the Floquet parameter. Finally, by exploiting the system structure via the Schur complement, we propose a fast parallel solver and illustrate its efficiency by numerical examples.

Keywords: Scattering problems, Helmholtz equation, perfectly matched layer, Floquet–Bloch transform, error estimation.

AMS subject Classifications: 35P25, 35A35, 65M60.

1 Introduction

In this paper, we consider the scattering of a time-harmonic acoustic wave by a locally perturbed periodic surface. The incident field is not assumed to be periodic with respect to the spatial variables. In the special case of a quasi-periodic incident field scattered by a purely periodic surface, the problem can be reduced directly to a single unit cell of the periodic domain [15, 24]. The resulting reduced problem may then be solved numerically, for instance, using finite element methods [3, 13] or using integral equation methods [23]. However, this

*Institut für Angewandte und Numerische Mathematik, Karlsruher Institut für Technologie, Germany (*tilo.arens@kit.edu, nasim.shafieeabyaneh@kit.edu)*

†Institut für Mathematik, Technische Universität Berlin, Germany (*ruming.zhang@tu-berlin.de*).

reduction is no longer applicable when either the incident field or the surface lacks periodicity. Therefore, new numerical approaches are required to efficiently solve these challenging problems.

Various methods have been used to analyze scattering in locally perturbed media, including the Lippmann-Schwinger equation [10], a volume integral approach [12], a perturbation technique [5, 6] and a numerical scheme specially designed to impose the exact boundary conditions on the vertical segments of a waveguide [11, 14]. These approaches are typically applied for the absorbing case (i.e., for complex wave numbers), which avoids difficulties associated with singularities.

An effective strategy is to employ the Floquet–Bloch (FB) transform (see [18, 20]), which decomposes a problem posed in a periodic domain into a family (indexed by the Floquet parameter) of periodic problems. Since each member of this family involves only periodic fields, the problem can be formulated on a single bounded unit cell and solved using standard numerical methods. To be able to apply the FB transform to the locally perturbed problem, one first removes the perturbation by means of a coordinate transformation. This comes at the cost of dealing with a perturbed Helmholtz equation with non-constant coefficients. Once periodicity of the domain is restored, the FB transform can be applied to derive a coupled family of periodic problems posed on a bounded cell. From the regularity analysis of the non-perturbed case carried out in [1], it is known that the FB transform of the solution has singularities with respect to the Floquet parameter. Therefore, to compute the inverse FB transform, a numerical approach tailored to the structure of singularities requires evaluating the transformed field for a large number of Floquet parameters. Furthermore, the coupling prevents us from solving these problems in parallel, which hence requires a high computational time. Hence, the necessity of proposing a fast solver for such scattering problems becomes more pronounced.

To improve computational efficiency, we approximate the solution by the perfectly matched layer (PML) method as introduced in [9]. When the FB transform is applied to the resulting PML problem, the transformed field depends analytically on the Floquet parameter. This analytic dependence allows the inverse FB transform to be accurately approximated using only a relatively small number of Floquet parameters. Nevertheless, assembling and solving the fully discretized system directly is still time-consuming due to the coupling. To reduce the cost, we recursively employ the Schur complement, which reformulates the resulting linear system in such a way that the matrix-vector multiplications are reduced to sums of terms that can be evaluated independently. Therefore, we can benefit greatly from parallelizing these evaluations by using an iterative method.

This paper is structured as follows. In Section 2, we present the mathematical formulation of the scattering problem in locally perturbed periodic domains. We then restore the periodicity of the domain by means of a coordinate transformation, which enables the application of the FB transform. In addition, we provide a brief overview of the PML method and derive the corresponding PML formulation. In Section 3, we apply the FB transform to the original and PML problems to derive the family of periodic problems in a bounded cell. Section 4 is devoted to the analysis of the PML approximation. For an incident spherical wave, we prove that on every compact set, the PML solution converges exponentially to the solution of the original scattering problem with respect to the PML parameter. In Section 5, we propose a fast iterative solver for computing the PML approximation of the solution. At each iteration, the matrix-vector products corresponding to different Floquet parameters can be evaluated in parallel. In conclusion, by using this technique, we significantly reduce computational time. Finally, Section 6 presents numerical examples that illustrate the accuracy and efficiency of the proposed method.

2 Scattering Problems

In this paper, we are concerned with the scattering of a time-harmonic acoustic field by a periodic or a locally perturbed periodic surface. We consider such problems in two spatial dimensions here. Although this is required for our approach to prove convergence of the PML approximation to the true solution of the scattering problem, we emphasize that the numerical approach works equally well in 3 dimensions as described in [25].

Let us start by introducing some notation used throughout the paper. Given a 2π-periodic Lipschitz continuous function ζ and a further Lipschitz continuous function ζ_{p} such that $\zeta_{\mathrm{p}}-\zeta$ is compactly supported in $(-\pi,\pi)$, we define the surfaces

$$\Gamma=\left\{(x_1,\zeta(x_1))^\top :\ x_1\in\mathbb{R}\right\}, \qquad \Gamma_{\mathrm{p}}=\left\{(x_1,\zeta_{\mathrm{p}}(x_1))^\top :\ x_1\in\mathbb{R}\right\}. \tag{1}$$

The corresponding domains above these surfaces are denoted by

$$\Omega=\left\{x=(x_1,x_2)^\top\in\mathbb{R}^2 :\ x_2>\zeta(x_1)\right\}, \qquad \Omega_{\mathrm{p}}=\left\{x=(x_1,x_2)^\top\in\mathbb{R}^2 :\ x_2>\zeta_{\mathrm{p}}(x_1)\right\}. \tag{2}$$

We will also assume that ζ, $\zeta_{\mathrm{p}}>0$.

Given some number $H\in\mathbb{R}$, we will use the upper half-spaces $U^H=\{x=(x_1,x_2)^\top\in\mathbb{R}^2 : x_2>H\}$ and let $\Gamma^H=\partial U^H$. For $H>\max\{\|\zeta\|_\infty,\|\zeta_{\mathrm{p}}\|_\infty\}$, we also define vertically bounded domains

$$\Omega^H=\Omega\setminus\overline{U^H}, \qquad \Omega^H_{\mathrm{p}}=\Omega_{\mathrm{p}}\setminus\overline{U^H}. \tag{3}$$

The total field u will be assumed to be a solution of the Helmholtz equation,

$$\Delta u+k^2u=0 \qquad \text{in } \Omega \quad (\text{or in } \Omega_{\mathrm{p}}), \tag{4}$$

subject to a Dirichlet boundary condition on the scattering surface,

$$u=0 \qquad \text{on } \Gamma \quad (\text{or on } \Gamma_{\mathrm{p}}). \tag{5}$$

A general existence and uniqueness theory for scattering from such unbounded surfaces has been presented in [7, 8]. Due to the unbounded nature of the scatterers Γ, Γ_{p}, we use weighted Sobolev spaces to obtain well-posed mathematical weak formulations for the scattering problems. Given numbers s, $r\in\mathbb{R}$, these spaces are defined by

$$H^s_r(\Omega^H)=\left\{u\in H^s_{\mathrm{loc}}(\Omega^H) :\ (1+|x_1|^2)^{r/2}\,u\in H^s(\Omega^H)\right\}. \tag{6}$$

Analogous definitions apply to $H^s_r(\Omega^H_{\mathrm{p}})$, $H^s_r(\Gamma)$, $H^s_r(\Gamma_{\mathrm{p}})$ and $H^s_r(\Gamma^H)$. Moreover, the space of functions in $H^s_r(\Omega^H)$ (or $H^s_r(\Omega^H_{\mathrm{p}})$) that satisfy the homogeneous Dirichlet boundary condition given in (5) is denoted by $\widetilde{H}^s_r(\Omega^H)$ (or $\widetilde{H}^s_r(\Omega^H_{\mathrm{p}})$).

As usual in a scattering problem, a radiation condition needs to be satisfied by the scattered field. A function $v \in C^\infty(U^H)$ is said to satisfy the *upward propagating radiation condition* if its trace on Γ^H satisfies $v|_{\Gamma^H} \in H_r^{1/2}(\Gamma^H)$ for some $|r| < 1$ and

$$v(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}x_1\xi + \mathrm{i}\left(k^2-\xi^2\right)^{1/2}(x_2-H)}\, \widehat{v}(\xi, H)\,\mathrm{d}\xi, \qquad x \in U^H. \tag{7}$$

Here, $\widehat{v}(\cdot, H)$ denotes the Fourier transform of $v(\cdot, H)$ with respect to x_1. Condition (7) gives rise to the Dirichlet-to-Neumann (DtN) map $T^+ \colon H_r^{1/2}(\Gamma^H) \to H_r^{-1/2}(\Gamma^H)$ defined by

$$T^+\psi(x) = \frac{\mathrm{i}}{2\pi} \int_{-\infty}^{\infty} (k^2 - \xi^2)^{1/2}\, \mathrm{e}^{\mathrm{i}x_1\xi}\, \widehat{\psi}(\xi, H)\,\mathrm{d}\xi\,, \qquad x \in \Gamma^H. \tag{8}$$

In [7, 8], it has been established that $T^+ \colon H_r^{1/2}(\Gamma^H) \to H_r^{-1/2}(\Gamma^H)$ is well defined and continuous if $|r| < 1$.

In conclusion, we may formulate the scattering problem for a periodic domain as follows: given some number r such that $|r| < 1$ and an incident field $u^i \in H_r^1(\Omega^H)$ find the scattered field $u^s_{\text{per}} \in H_r^1(\Omega^H)$ as a weak solution of

$$\begin{aligned} \Delta u^s_{\text{per}} + k^2 u^s_{\text{per}} &= 0 && \text{in } \Omega^H\,, \\ u^s_{\text{per}} &= -u^i && \text{on } \Gamma\,, \\ \frac{\partial u^s_{\text{per}}}{\partial x_2} &= T^+ u^s_{\text{per}} && \text{on } \Gamma^H\,. \end{aligned} \tag{9}$$

Under the same conditions on r and u^i, the scattering problem for a locally perturbed periodic domain is to find the scattered field $u^s \in H_r^1(\Omega_{\text{p}}^H)$ as the weak solution of

$$\begin{aligned} \Delta u^s + k^2 u^s &= 0 && \text{in } \Omega_{\text{p}}^H\,, \\ u^s &= -u^i && \text{on } \Gamma_{\text{p}}\,, \\ \frac{\partial u^s}{\partial x_2} &= T^+ u^s && \text{on } \Gamma^H\,. \end{aligned} \tag{10}$$

According to [7, Theorem 4.1], both problems (9) and (10) have unique solutions.

Through the Floquet–Bloch transform, there exists a well-established theory to deal with locally perturbed coefficients of a partial differential equation in a periodic domain (see, e.g., [18]). Thus, it is reasonable to transform (10) into a problem posed in the periodic domain Ω^H, introducing variable coefficients on a compact subdomain of Ω^H. Given a number $H' \in (\|\zeta\|_\infty, H)$, we introduce the transform

$$\psi(x) = \begin{cases} x\,, & x_2 \geq H'\,, \\ \left(x_1,\, x_2 + \dfrac{(x_2 - H')^3}{(\zeta(x_1) - H')^3}\,(\zeta_{\text{p}}(x_1) - \zeta(x_1))\right), & x_2 < H'\,. \end{cases}$$

Then ψ is a diffeomorphism such that $\psi(\Omega^H) = \Omega_{\text{p}}^H$. We set $u^s_{\text{tra}} = u^s \circ \psi$ and find that $u^s_{\text{tra}} \in H_r^1(\Omega^H)$ is a weak solution of the problem

$$\begin{aligned} \nabla \cdot (A \nabla u^s_{\text{tra}}) + k^2\, c\, u^s_{\text{tra}} &= 0 && \text{in } \Omega^H\,, \\ u^s_{\text{tra}} &= -u^i \circ \psi && \text{on } \Gamma\,, \\ \frac{\partial u^s_{\text{tra}}}{\partial x_2} &= T^+ u^s_{\text{tra}} && \text{on } \Gamma^H\,, \end{aligned} \tag{11}$$

with the coefficients

$$A(x) = |\det \psi'(x)| \, (\psi'(x))^{-1} (\psi'(x))^{-\top} \quad \in L^\infty(\Omega^H, \mathbb{R}^{2\times 2}), \\ c(x) = |\det \psi'(x)| \quad \in L^\infty(\Omega^H). \tag{12}$$

Note that $A - I$ and $c - 1$ are compactly supported in $\big((-\pi,\pi)\times\mathbb{R}\big) \cap \Omega^H$. In particular, we have $A = I$ and $c = 1$ in a neighborhood of Γ^H.

Our main focus in this paper lies on the numerical solution of problem (11). This requires a discretization of the DtN map T^+. Approaches that are based directly on (the Floquet–Bloch transform of) T^+ and use tailor-made inversion formulas of the inverse Floquet–Bloch transform are possible, but, particularly for the 3D case, require substantial computational effort (see, e.g., [1]). A simpler way to implement a numerical approximation of T^+ is given by the perfectly matched layer (PML) method as shown in [9, Section 2]. To introduce this approach, we extend the domain by an additional horizontal layer of thickness $\lambda > 0$,

$$\Omega^{H+\lambda} = \Omega^H \cup \left(\overline{U^H} \setminus \overline{U^{H+\lambda}}\right).$$

The PML is described by the complex stretched coordinate

$$\widehat{x}_2 = \int_0^{x_2} s(t)\, \mathrm{d}t,$$

where the complex valued function s is defined by

$$s(x_2) = \begin{cases} 1, & \text{if } x_2 < H, \\ 1 + \left(\dfrac{x_2 - H}{\lambda}\right)^2 \rho\, \mathrm{e}^{\mathrm{i}\pi/4}, & \text{if } x_2 \in [H, H+\lambda), \end{cases} \tag{13}$$

where ρ is some positive parameter.

The absorption of the PML depends on the virtual thickness of the layer which is given by

$$\sigma = \int_H^{H+\lambda} s(x_2)\, \mathrm{d}x_2 = \lambda\left(1 + \frac{\rho}{3}\, \mathrm{e}^{\mathrm{i}\pi/4}\right).$$

The PML approximation of the DtN map is obtained by replacing the Helmholtz equation in the layer by the corresponding equation using the stretched complex coordinate, i.e.,

$$\frac{\partial^2 u_\sigma^s}{\partial x_1^2} + \frac{\partial^2 u_\sigma^s}{\partial \widehat{x}_2^2} + k^2 u_\sigma^s = 0 \qquad \text{in } U^H \setminus \overline{U^{H+\lambda}},$$

and a homogeneous Dirichlet boundary condition on $\Gamma^{H+\lambda}$. Changing the variables back from the complex stretched to standard real coordinates and using $\mathrm{d}\widehat{x}_2 / \mathrm{d}x_2 = s(x_2)$ gives the following PML problem

$$\begin{aligned} \nabla\cdot(A\nabla u_\sigma^s) + k^2 c\, u_\sigma^s &= 0 && \text{in } \Omega^{H+\lambda}, \\ u_\sigma^s &= -u^i \circ \psi && \text{on } \Gamma, \\ u_\sigma^s &= 0 && \text{on } \Gamma^{H+\lambda}. \end{aligned} \tag{14}$$

Here, we have extended the coefficients defined in (12) in Ω^H to $\Omega^{H+\lambda}$ by setting

$$A(x) = \operatorname{diag}\left(s(x_2), \frac{1}{s(x_2)}\right), \qquad x \in U^H \setminus \overline{U^{H+\lambda}}$$

and $c = s(x_2)$ in $U^H \setminus \overline{U^{H+\lambda}}$.

According to [9, Section 2], the solution u_σ^s satisfies the following boundary condition on Γ^H

$$\frac{\partial u_\sigma^s}{\partial x_2} = T_\sigma^+ u_\sigma^s\,, \tag{15}$$

where T_σ^+ denotes the PML approximation of the DtN map and is defined by

$$(T_\sigma^+\psi)(x) = \frac{\mathrm{i}}{2\pi}\int_{-\infty}^{\infty} (k^2-\xi^2)^{1/2} \coth\left(\sigma(k^2-\xi^2)^{1/2}\right) \mathrm{e}^{\mathrm{i}x_1\xi}\, \widehat{\psi}(\xi, H)\,\mathrm{d}\xi\ . \tag{16}$$

Therefore, the PML problem (14) is equivalent to

$$\begin{aligned} \nabla\cdot(A\nabla u_\sigma^s) + k^2\, c\, u_\sigma^s &= 0 && \text{in } \Omega^H\,, \\ u_\sigma^s &= -u^i \circ \psi && \text{on } \Gamma\,, \\ \frac{\partial u_\sigma^s}{\partial x_2} &= T_\sigma^+ u_\sigma^s && \text{on } \Gamma^H\,. \end{aligned} \tag{17}$$

The variational formulation of (17) is to find $u_\sigma^s \in H^1(\Omega^H)$ such that

$$\int_{\Omega^H} \left((A\,\nabla u_\sigma^s)\cdot\nabla\overline{v}_\sigma - k^2\, c\, u_\sigma^s\, \overline{v}_\sigma\right)\mathrm{d}x - \left\langle T_\sigma^+ u_\sigma^s, \overline{v}_\sigma\right\rangle_{\Gamma^H} = 0 \quad \text{for all } v_\sigma \in \widetilde{H}^1(\Omega^H)\,, \tag{18}$$

together with the variational form of the Dirichlet boundary condition

$$\langle u_\sigma^s\,, \overline{\eta}\rangle_\Gamma = -\left\langle u^i \circ\psi\,, \overline{\eta}\right\rangle_\Gamma \qquad \text{for all } \eta \in H^{-1/2}(\Gamma)\,. \tag{19}$$

Solvability of (17) is obtained in the following way: Using a suitable extension operator, the boundary values in problems (11) and (17) may be extended to functions in $L^2(\Omega^H)$ with support bounded away from Γ^H. The equivalent formulations of boundary value problems with homogeneous Dirichlet conditions but source terms bounded away from Γ_H are now exactly of the form considered in [9]. Therefore, Theorem 3.3 of that reference implies the unique solvability of the PML problem.

3 Floquet–Bloch Transformed Problems

We first recall the definition of the Floquet–Bloch (FB) transform as in [20, Sec. 2].

Definition 1. *For a given $f \in C_0^\infty(\Omega^H)$, the FB transform is defined by*

$$(\mathcal{J}f)(\alpha; x) = \sum_{j\in\mathbb{Z}} f(x_1 + 2\pi j, x_2)\,\mathrm{e}^{-\mathrm{i}\alpha(x_1+2\pi j)}\,, \quad \alpha\in\mathbb{R}, \quad x \in \Omega^H, \tag{20}$$

where α is called the Floquet parameter.

The summation is well defined since the function f is compactly supported; therefore, the series reduces to a finite sum.

The FB transform $\mathcal{J}f$, for each fixed α, is 2π-periodic with respect to x_1. We denote the fundamental cell of periodicity by

$$\Omega_{2\pi}^H = \{(x_1, x_2) \in \Omega^H : x_1 \in (-\pi, \pi)\}$$

with boundaries given by

$$\Gamma_{2\pi} = \{x \in \Gamma \colon x_1 \in [-\pi, \pi]\} \quad \text{and} \quad \Gamma_{2\pi}^H = \{x \in \Gamma^H \colon x_1 \in [-\pi, \pi]\}\,.$$

Moreover, for each fixed x, the function $\alpha \mapsto \mathrm{e}^{\mathrm{i}\alpha x}(\mathcal{J}f)(\alpha; x)$ is 1-periodic with respect to α with fundamental period $\Lambda = [-1/2, 1/2]$. Therefore, the fundamental cell of periodicity of $\mathcal{J}f$ is $\Lambda \times \Omega_{2\pi}^H$.

Theorem 2. *The FB transform from $C_0^\infty(\Omega^H)$ can be continuously extended to an isometry between $L^2(\Omega^H)$ and $L^2(\Lambda, H^1_{\mathrm{per}}(\Omega_{2\pi}^H))$ and its inverse transform is obtained by*

$$\mathcal{J}^{-1}f(x_1 + 2\pi j, x_2) = \int_\Lambda f(\alpha; x)\mathrm{e}^{\mathrm{i}\alpha(x_1+2\pi j)}\,\mathrm{d}\alpha\,, \quad x \in \Omega_{2\pi}^H \text{ and } j \in \mathbb{Z}. \tag{21}$$

Proof. See [20, Theorem 8]. □

Consider the scattering problem in the purely periodic case. Applying the FB transform to (9) gives a modified weak formulation for $w = \mathcal{J}u^s_{\mathrm{per}} \in L^2(\Lambda; H^1_{\mathrm{per}}(\Omega_{2\pi}^H))$, which satisfies

$$\int_\Lambda a_\alpha(w(\alpha, \cdot), \varphi(\alpha, \cdot))\,\mathrm{d}\alpha = 0\,, \qquad \int_\Lambda \langle w\,,\, \eta\rangle_{\Gamma_{2\pi}}\,\mathrm{d}\alpha = -\int_\Lambda \left\langle \mathcal{J}u^i\,,\, \eta\right\rangle_{\Gamma_{2\pi}}\mathrm{d}\alpha\,, \tag{22}$$

for all $\varphi \in L^2(\Lambda; H^1_{\mathrm{per}}(\Omega_{2\pi}^H))$ and all $\eta \in L^2(\Lambda; H^{1/2}_{\mathrm{per}}(\Gamma_{2\pi}))$. Here, we have used

$$a_\alpha(v, \varphi) = \int_{\Omega_{2\pi}^H} \left(\nabla v \cdot \nabla\overline{\varphi} - 2\mathrm{i}\alpha\,\partial_{x_1} v\,\overline{\varphi} - (k^2 - \alpha^2)\,v\,\overline{\varphi}\right)\mathrm{d}x - \langle T_\alpha^+ v, \varphi\rangle_{\Gamma_{2\pi}^H}\,, \quad v, \varphi \in H^1_{\mathrm{per}}(\Omega_{2\pi}^H)\,,$$

$$T_\alpha^+ v(x_1, H) = \mathrm{i}\sum_{j\in\mathbb{Z}} \gamma_j(\alpha)\,\widehat{v}_j\,\mathrm{e}^{\mathrm{i}jx_1} \quad \text{for} \quad v(x_1, H) = \sum_{j\in\mathbb{Z}} \widehat{v}_j\,\mathrm{e}^{\mathrm{i}jx_1} \in H^{1/2}(\Gamma_{2\pi}^H)\,,$$

$$\gamma_j(\alpha) = (k^2 - (\alpha + j)^2)^{1/2}\,, \quad j \in \mathbb{Z}\,.$$

The problem (22) is equivalent to (9), as proven in [20, Theorem 9]. Moreover, it has been established in [22, Theorem 3.3] that when $u^i \in H_r^1(\Omega^H)$ for some $r \in (1/2, 1)$, then $w \in C(\Lambda, H^1_{\mathrm{per}}(\Omega_{2\pi}^H))$. In this case, it follows that we can replace (22) by the formulation

$$a_\alpha(w(\alpha, \cdot), \varphi) = 0\,, \qquad \langle w(\alpha, \cdot)\,,\, \eta\rangle_{\Gamma_{2\pi}} = \left\langle \mathcal{J}u^i(\alpha, \cdot)\,,\, \eta\right\rangle_{\Gamma_{2\pi}} \tag{23}$$

for all $\varphi \in H^1_{\mathrm{per}}(\Omega_{2\pi}^H)$, $\eta \in H^{1/2}_{\mathrm{per}}(\Gamma_{2\pi})$ and all $\alpha \in \Lambda$.

We will combine techniques from [1, 2, 16, 28] to establish analytic dependence of w on α away from certain exceptional points. This requires the following assumption.

Condition 3. *The wave number k satisfies $k \neq \dfrac{m}{2}$ for all $m \in \mathbb{N}$.*

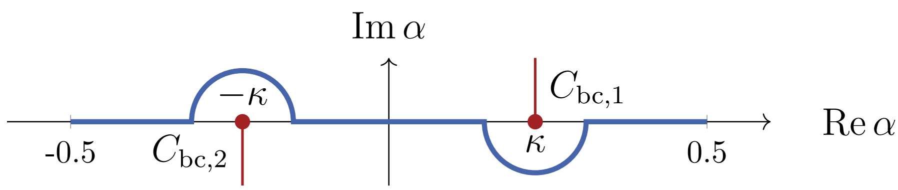


Figure 1: Sketch of the branch cuts (in red) and the integration path $\mathcal{E}$ (in blue) for $\kappa > 0$.

Under Condition 3, there is a nonnegative integer $\widehat{j}$ and a number $\kappa \in (-1/2, 1/2) \setminus \{0\}$ such that $k = \widehat{j} + \kappa$. It follows that $\gamma_{\widehat{j}}(\kappa) = \gamma_{-\widehat{j}}(-\kappa) = 0$. The numbers $\pm\kappa$ are the only roots of $\gamma_{\pm\widehat{j}}$ in Λ and are called *cutoff values.*

Using the Riesz representation theorem, for fixed $\alpha \in \Lambda$ we can formulate (23) as an operator equation

$$K_\alpha w(\alpha, \cdot) = \mathcal{J}\left(u^i|_{\Gamma_{2\pi}}\right)(\alpha, \cdot)\,.$$

The operator K_α depends continuously on α and the same is true for the right-hand side if $r \in (1/2, 1)$. Hence, perturbation theory (see [17, Theorem 10.1]) gives the continuous dependence of w on α as mentioned above. Moreover, K_α depends analytically on α except for cutoff values, so that in a neighborhood of any α_0 that is not a cutoff value and for which the right-hand side also analytically depends on α, the solution w depends analytically on α. In cases that are practically relevant, which we will consider in section 4 below, the incident field satisfies the following more specific assumption.

Condition 4. *A function $f\colon \Lambda \to H^1_{\mathrm{per}}(\Omega^H_{2\pi})$ satisfies this condition if on $\Lambda \setminus \{\pm\kappa\}$, f depends analytically on α, and there exist open neighborhoods $U_\pm$ of $\pm\kappa$ and analytic functions $f_{\pm,1}$, $f_{\pm,2} : U_\pm \to H^1_{\mathrm{per}}(\Omega^H_{2\pi})$, such that*

$$f(\alpha) = f_{\pm,1}(\alpha) + \gamma_{\pm\widehat{j}}(\alpha)\, f_{\pm,2}(\alpha)\,, \quad \alpha \in U_\pm\,.$$

It has been established in [27, Theorem 16] that if $\mathcal{J}u^i$ satisfies Condition 4 then also the solution w of (23) satisfies Condition 4. This allows us to analytically extend w into (parts of) the complex plane. For this purpose, we shift the branch cut of the function $r(z) = z^{1/2}$, $z \in \mathbb{C} \setminus \{0\}$, from the negative real axis to the curve $C_{\mathrm{bc},0} = \{t^2 - 2\mathrm{i}kt : t > 0\}$. Then, the function $\gamma_{\widehat{j}}$ can be analytically extended to $(\Lambda + \mathrm{i}\mathbb{R}) \setminus C_{\mathrm{bc},1}$ and $\gamma_{-\widehat{j}}$ to $(\Lambda + \mathrm{i}\mathbb{R}) \setminus C_{\mathrm{bc},2}$ with the branch cuts

$$C_{\mathrm{bc},1} = \kappa + \mathrm{i}\mathbb{R}_{>0} \qquad \text{and} \qquad C_{\mathrm{bc},2} = -\kappa - \mathrm{i}\mathbb{R}_{>0}\,,$$

respectively. In conclusion, the operator T^+_α and also the operator K_α can be analytically extended to $(\Lambda + \mathrm{i}\mathbb{R}) \setminus (C_{\mathrm{bc},1} \cup C_{\mathrm{bc},2})$. If Condition 4 is satisfied by $\mathcal{J}u^i$, then the same holds for w as shown in [28, 16] that the solution operator induces exactly the same structure of singularities. In particular, by Cauchy's integral theorem, instead of (21), we may compute the inverse FB transform of w by

$$\mathcal{J}^{-1}w(x_1 + 2\pi j, x_2) = \int_{\mathcal{E}} w(\alpha, x)\, \mathrm{e}^{\mathrm{i}\alpha(x_1+2\pi j)}\, \mathrm{d}\alpha, \quad x \in \Omega^H_{2\pi},\ j \in \mathbb{Z}\,, \tag{24}$$

where for sufficiently small $\delta > 0$ the integration path $\mathcal{E}$ is defined by

$$\mathcal{E} = \Lambda \setminus [(\kappa - \delta, \kappa + \delta) \cup (-\kappa - \delta, -\kappa + \delta)] \cup \mathcal{E}_+ \cup \mathcal{E}_-\,,$$
$$\mathcal{E}_\pm = \{\pm\kappa \mp \delta\, \mathrm{e}^{\mathrm{i}\vartheta} :\ \vartheta \in (0, \pi)\}\,.$$

We now turn to the locally perturbed problem. An application of the FB transform to problem (11) requires the additional form

$$b_\alpha(v,\varphi) = \int_{\Omega_{2\pi}^H} \left((A-I)\,\nabla v\cdot\nabla\left(\mathrm{e}^{-\mathrm{i}\alpha x_1}\overline{\varphi}\right) - k^2\,(c-1)\,v\,(\mathrm{e}^{-\mathrm{i}\alpha x_1}\overline{\varphi})\right)\mathrm{d}x\,, \tag{25}$$

for $v,\ \varphi\in H^1(\Omega_{2\pi}^H)$. A discussion of solvability and regularity results for this case is provided in [22, Section 4].

Theorem 5. *For $r\in[0,1)$, the function $u_{\mathrm{tra}}^s\in H_r^1(\Omega^H)$ is a weak solution of (11) if and only if $w=\mathcal{J}u_{\mathrm{tra}}^s\in L^2(\Lambda;H_{\mathrm{per}}^1(\Omega_{2\pi}^H))$ satisfies*

$$\begin{aligned} \int_\Lambda \left(a_\alpha\big(w(\alpha,\cdot),\varphi(\alpha,\cdot)\big)+b_\alpha(\mathcal{J}^{-1}w,\varphi(\alpha,\cdot)\right)\mathrm{d}\alpha = 0\,,\\ \int_\Lambda\int_{\Gamma_{2\pi}}\big(w(\alpha,\cdot)-\mathcal{J}(u^i\circ\psi)(\alpha,\cdot)\big)\,\overline{\eta}\,\mathrm{d}s\,\mathrm{d}\alpha = 0 \end{aligned} \tag{26}$$

for all $\varphi\in L^2(\Lambda,\widetilde{H}_{\mathrm{per}}^1(\Omega_{2\pi}^H))$ and all $\eta\in L^2(\Lambda,H_{\mathrm{per}}^{1/2}(\Gamma_{2\pi}))$. In addition, if $r>1/2$, there holds $w\in C(\Lambda,H_{\mathrm{per}}^1(\Omega_{2\pi}^H))$ and for every $\alpha\in\Lambda$ and every $\varphi\in\widetilde{H}_{\mathrm{per}}^1(\Omega_{2\pi}^H)$, $\eta\in H_{\mathrm{per}}^{1/2}(\Gamma_{2\pi})$,

$$a_\alpha\big(w(\alpha,\cdot),\varphi\big)+b_\alpha(\mathcal{J}^{-1}w,\varphi)=0,\quad \big\langle\big(w(\alpha,\cdot)-\mathcal{J}(u^i\circ\psi)(\alpha,\cdot)\big)\ ,\eta\big\rangle_{\Gamma_{2\pi}}=0\,. \tag{27}$$

We make use of the previous theorem to replace the integration over Λ in (26) by an integration over the path $\mathcal{E}$.

Theorem 6. *Suppose that $\mathcal{J}u^i\circ\psi$ satisfies Condition 4. Then, there exists a unique solution w of (26). Moreover, we may replace Λ by $\mathcal{E}$ in (26) and in the inverse FB transform, i.e.,*

$$u_{\mathrm{tra}}^s(x_1+2\pi j,x_2)=\int_{\mathcal{E}} w(\alpha,x)\,\mathrm{e}^{\mathrm{i}\alpha(x_1+2\pi j)}\,\mathrm{d}\alpha,\quad x\in\Omega_{2\pi}^H,\ j\in\mathbb{Z}\,.$$

Proof. As Problem (11) admits a unique solution, we know from Theorem 5 that there is a unique solution w to (26) and that it satisfies (27). Note that in this last equation $\mathcal{J}^{-1}w=u_{\mathrm{tra}}^s$ is independent of α and $b_\alpha(u_{\mathrm{tra}}^s,\varphi)$ depends analytically on α. It follows that w is the solution to a problem similar to (23) but with an additional analytic source term on the right-hand side of the first equation. Thus, all arguments in the case of a periodic domain apply. □

Applying the FB transform to the PML problem given in (18), we find that $w_\sigma=\mathcal{J}u_\sigma^s$ satisfies

$$\int_\Lambda a_{\alpha,\sigma}\big(w_\sigma(\alpha,\cdot),\varphi(\alpha,\cdot)\big)+b_\alpha(\mathcal{J}^{-1}w_\sigma,\varphi(\alpha,\cdot))\,\mathrm{d}\alpha=0\qquad\text{for all }\varphi\in L^2(\Lambda,\widetilde{H}_{\mathrm{per}}^1(\Omega_{2\pi}^H))\,, \tag{28}$$

where $a_{\alpha,\sigma}$ is obtained from a_α by replacing T_α^+ with its PML approximation defined by

$$T_{\alpha,\sigma}^+v(x_1,H)=\mathrm{i}\sum_{j\in\mathbb{Z}}\gamma_j(\alpha)\,\coth(-\mathrm{i}\gamma_j(\alpha)\,\sigma)\,\widehat{v}_j\,\mathrm{e}^{\mathrm{i}jx_1}\quad\text{for}\quad v(x_1,H)=\sum_{j\in\mathbb{Z}}\widehat{v}_j\,\mathrm{e}^{\mathrm{i}jx_1}\in H^{1/2}(\Gamma_{2\pi}^H)\,.$$

Note that this operator depends analytically on α and hence so does the operator induced by the form b_α. Hence we obtain that the map $\alpha\mapsto w_\sigma(\alpha,\cdot)$ is analytic.

4 Convergence Analysis of the PML Approximation

Recently, there has been a number of results published on the convergence of the PML approximation with respect to the PML parameter σ of the solutions to boundary value problems for the Helmholtz equation for unbounded boundaries (see, e.g., [9, 28]). Instead of a scattering problem, in these works source problems are considered.

We begin by studying the source problem in the periodic domain. Let $g \in L^2(\Omega^H)$ be a compactly supported source in the periodic domain Ω^H. The aim is to find $v \in \widetilde{H}_r^1(\Omega^H)$ such that

$$\begin{aligned} \Delta v + k^2 v &= g && \text{in } \Omega^H\,, \\ v &= 0 && \text{on } \Gamma\,, \\ \frac{\partial v}{\partial x_2} &= T^+ v && \text{on } \Gamma^H\,. \end{aligned} \tag{29}$$

Applying the FB transform to (29), we obtain an equivalent formulation for the function $w_g = \mathcal{J} v \in L^2(\Lambda, \widetilde{H}^1_{\mathrm{per}}(\Omega^H_{2\pi}))$, which

$$\int_\Lambda a_\alpha\big(w_g(\alpha,\cdot),\varphi(\alpha,\cdot)\big)\,\mathrm{d}\alpha = \int_\Lambda \langle \mathcal{J} g(\alpha,\cdot),\varphi(\alpha,\cdot)\rangle_{\Omega^H_{2\pi}}\,\mathrm{d}\alpha \quad \text{for all } \varphi \in L^2(\Lambda, \widetilde{H}^1_{\mathrm{per}}(\Omega^H_{2\pi}))\,. \tag{30}$$

We can simply formulate the source problem corresponding to the PML approximation of v by replacing the boundary condition on Γ^H by a condition with the modified DtN-operator T^+_σ. Thus, we aim to find $v_\sigma \in \widetilde{H}^1(\Omega^H)$ (the modification of the space is justified in [9]) such that

$$\begin{aligned} \Delta v_\sigma + k^2 v_\sigma &= g && \text{in } \Omega^H\,, \\ v_\sigma &= 0 && \text{on } \Gamma\,, \\ \frac{\partial v_\sigma}{\partial x_2} &= T^+_\sigma v_\sigma && \text{on } \Gamma^H\,. \end{aligned} \tag{31}$$

The corresponding variational formulation is to find $w_{g,\sigma} = \mathcal{J} v_\sigma \in L^2(\Lambda, \widetilde{H}^1_{\mathrm{per}}(\Omega^H_{2\pi}))$ such that

$$\int_\Lambda a_{\alpha,\sigma}\big(w_{g,\sigma}(\alpha,\cdot),\varphi(\alpha,\cdot)\big)\,\mathrm{d}\alpha = \int_\Lambda \langle \mathcal{J} g(\alpha,\cdot),\varphi(\alpha,\cdot)\rangle_{\Omega^H_{2\pi}}\,\mathrm{d}\alpha \qquad \text{for all } \varphi \in L^2(\Lambda, \widetilde{H}^1_{\mathrm{per}}(\Omega^H_{2\pi}))\,, \tag{32}$$

where the sesquilinear form $a_{\alpha,\sigma}$ is given as in (28). The sesquilinear forms a_α and $a_{\alpha,\sigma}$ induce bounded linear operators $\mathcal{A}_\alpha$, $\mathcal{A}_{\alpha,\sigma} : \widetilde{H}^1_{\mathrm{per}}(\Omega^H_{2\pi}) \to H^{-1}_{\mathrm{per}}(\Omega^H_{2\pi})$.

It has been shown in [9] that the PML approximation to the true solution of such a source problem exists but cannot be expected to be exponentially convergent on the unbounded domain Ω^H. However, exponential convergence on compact subdomains of a purely periodic domain has been established in [28] for the case of g being a compactly supported L^2-function in Ω^H. Here, we will extend these results to cover the approximation of the solution of (11) by the solution of (17). Let us recall the most important results from [28]:

Theorem 7. *Let $g \in L^2(\Omega^H)$ with compact support. Then there exist constants C, $c > 0$, such that*

$$\|\mathcal{A}_\alpha - \mathcal{A}_{\alpha,\sigma}\| \le C\,\mathrm{e}^{-c|\sigma|} \quad \textit{for all } \alpha \in \mathcal{E}\,.$$

Moreover, for any compact subset $K \subset \Omega^H$, the solutions of v of (29) and v_σ of (31) satisfy

$$\|v - v_\sigma\|_{H^1(K)} \le C\, \mathrm{e}^{-c|\sigma|}\,.$$

Proof. The assertion is proved in Theorems 9 and 11 in [28]. □

Let us now return to the scattering problem for the particular case, where the incident field is Green's function of the Helmholtz equation for a half plane U^0. We denote by

$$\Phi(x,y) = \frac{\mathrm{i}}{4}\, H_0^{(1)}(k|x-y|)\,, \qquad x, y \in \mathbb{R}^2 \quad x \ne y\,,$$

the fundamental solution to the Helmholtz equation in free field conditions. Finally, recalling that $\overline{\Omega^H} \subseteq U^0$, we set

$$y' = (y_1, -y_2)^\top\,, \qquad y \in U^0\,.$$

According to [27, Remark 14], this incident field is in $H_r^{1/2}(\Gamma)$ for some $r > 1/2$.

Theorem 8. *Let $y \in \Omega^H$ and let $u^s_{\mathrm{per}} \in H^1_r(\Omega^H)$ denote a weak solution to (9) for*

$$u^i(x) = \Phi(x,y) - \Phi(x,y')\,, \qquad x \in \Omega^H, \quad x \ne y\,.$$

Moreover, let $u^s_{\mathrm{per},\sigma}$ denote the weak solution of the PML problem (17) in the corresponding case ($A = I$, $c = 1$). Then, for every compact subset $K \subseteq \Omega^H$ and sufficiently large σ, there exist C, $c > 0$ such that

$$\|u^s_{\mathrm{per}} - u^s_{\mathrm{per},\sigma}\|_{H^1(K)} \le C\, \mathrm{e}^{-c|\sigma|}\,.$$

Proof. According to [7, equation (2.8)], there is a compactly supported function $g \in L^2(\Omega^H)$, such that the weak solution $v^i \in H^1_r(U^0 \setminus \overline{U^H})$ to the source problem (29) with Ω^H replaced by $U^0 \setminus \overline{U^H}$ is equal to u^i in $U^a \setminus (\overline{U^H} \cup \operatorname{supp} g)$ for every $a > 0$. In problems (9) and (17), we may hence replace u^i by v^i in the boundary condition on Γ, respectively.

Denote by v^i_σ the PML approximation to v^i, i.e., the solution to (31) with Ω^H replaced by $U^0 \setminus \overline{U^H}$. Let v^s_σ denote the solution to (17) with u^i replaced by v^i_σ. Now, we can estimate

$$\|u^s_{\mathrm{per}} - u^s_{\mathrm{per},\sigma}\|_{H^1(K)} \le \|u^s_{\mathrm{per}} + v^i - (v^s_\sigma + v^i_\sigma)\|_{H^1(K)} + \|v^i - v^i_\sigma\|_{H^1(K)} + \|v^s_\sigma - u^s_{\mathrm{per},\sigma}\|_{H^1(K)}\,.$$

The function $u^s_{\mathrm{per}} + v^i$ is the solution of the source problem (29), while $u^s_\sigma + v^i_\sigma$ is its PML approximation, i.e., the solution of (31). Hence, by Theorem 7, there exist two constants c, C such that

$$\|u^s_{\mathrm{per}} + v^i - (v^s_\sigma + v^i_\sigma)\|_{H^1(K)} \le C\, \mathrm{e}^{-c|\sigma|}$$

for every compact subset $K \subseteq \Omega^H$ and any sufficiently large σ. Similarly, $v^i - v^i_\sigma$ satisfies the same estimate.

Setting $q = v^s_\sigma - u^s_{\mathrm{per},\sigma}$, we see that this function is the weak solution to problem (17) with the boundary values replaced by $h = u^i - v^i_\sigma$. As explained at the end of section 2, for sufficiently large σ, this problem is uniquely solvable and the norms of the solution operators $h \mapsto q$ are uniformly bounded with respect to σ.

In order to obtain an upper bound for $\|q\|_{H^1(K)}$, we apply the FB transform to the above problem. The transformed boundary values are $\mathcal{J}u^i - \mathcal{J}v^i_\sigma$. From [21, Equation (50)], it is known that

$$\mathcal{J}u^i(\alpha, x) = \begin{cases} \frac{1}{2\pi}\sum_{j\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}j(x_1-y_1)+\mathrm{i}\gamma_j(\alpha)y_2}\,\mathrm{sinc}(\gamma_j(\alpha)\,x_2)\,x_2\,, & 0 < x_2 \le y_2\,, \\ \frac{1}{2\pi}\sum_{j\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}j(x_1-y_1)+\mathrm{i}\gamma_j(\alpha)x_2}\,\mathrm{sinc}(\gamma_j(\alpha)\,y_2)\,y_2\,, & x_2 > y_2\,. \end{cases}$$

Similarly, applying the FB transform to [9, Equation (28)] yields

$$\mathcal{J}v^i_\sigma(\alpha, x) = \begin{cases} \frac{1}{2\pi}\sum_{j\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}j(x_1-y_1)}\,\frac{\sin(\gamma_j(\alpha)\,(\sigma+H-y_2))}{\sin(\gamma_j(\alpha)\,(\sigma+H))}\,\mathrm{sinc}(\gamma_j(\alpha)\,x_2)\,x_2\,, & 0 < x_2 \le y_2\,, \\ \frac{1}{2\pi}\sum_{j\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}j(x_1-y_1)}\,\frac{\sin(\gamma_j(\alpha)\,(\sigma+H-x_2))}{\sin(\gamma_j(\alpha)\,(\sigma+H))}\,\mathrm{sinc}(\gamma_j(\alpha)\,y_2)\,y_2\,, & x_2 > y_2\,. \end{cases}$$

Using Euler's formula, it is straightforward to derive the general identity

$$\mathrm{e}^{\mathrm{i}A} - \frac{\sin(B-A)}{\sin(B)} = \mathrm{i}\sin(A)\,(1-\coth(-\mathrm{i}B))\,, \qquad A\in\mathbb{C}, B\in\mathbb{C}\setminus\pi\mathbb{Z}\,.$$

We consider the case $0 < x_2 \le y_2$. Applying the identity with $A = \gamma_j(\alpha)\,y_2$, $B = \gamma_j(\alpha)\,(\sigma+H)$, we obtain

$$\mathcal{J}(u^i - v^i_\sigma)(\alpha, x) = \frac{\mathrm{i}}{2\pi}\sum_{j\in\mathbb{Z}}\Big(\gamma_j(\alpha)\,\big[1-\coth(-\mathrm{i}\gamma_j(\alpha)(\sigma+H))\big] \\ \times\,\mathrm{e}^{\mathrm{i}j(x_1-y_1)}\,\mathrm{sinc}(\gamma_j(\alpha)\,x_2)\,\mathrm{sinc}(\gamma_j(\alpha)\,y_2)\,x_2\,y_2\Big)\,.$$

Since this expression is symmetric with respect to x_2 and y_2, it also holds in the case $x_2 > y_2$, as we obtain the result by interchanging the roles of x_2 and y_2.

Note that the sinc functions are even analytic functions and hence they depend analytically on α. Singularities in the terms of the series with respect to α are thus only contained in the term

$$\gamma_j(\alpha)\,\big[1-\coth(-\mathrm{i}\gamma_j(\alpha)(\sigma+H))\big]\,,$$

which has been thoroughly analyzed in [28, Section 4] and [16, Lemma 18]. Instead of considering $\alpha\in\Lambda$, we consider $\alpha\in\mathcal{E}$. Estimating as in [16], we conclude for sufficiently large σ

$$\|\mathcal{J}q(\alpha,\cdot)\|_{\widetilde{H}^1_{\mathrm{per}}(\Omega^H_{2\pi})} \le C\,\|\mathcal{J}(u^i - v^i_\sigma)(\alpha,\cdot)\|_{H^{1/2}_{\mathrm{per}}(\Gamma_{2\pi})} \le C\,\mathrm{e}^{-c|\sigma|} \quad \text{for all } \alpha\in\mathcal{E}\,. \tag{33}$$

The remainder of the proof is carried out exactly as in the proof of Theorem 11 in [28]: due to Condition 4, we may change the integration path in the inverse FB transform from Λ to $\mathcal{E}$ and then straightforwardly apply (33). □

We now turn to the general case of a locally perturbed periodic surface. Again, we first consider the corresponding source problem: For a given compactly supported $g\in L^2(\Omega^H)$, find $z\in\widetilde{H}^1_r(\Omega^H)$ such that it is a weak solution of

$$\begin{aligned} \nabla\cdot(A\nabla z) + k^2\,c\,z &= g && \text{in}\quad \Omega^H\,, \\ z &= 0 && \text{on}\quad \Gamma\,, \\ \frac{\partial z}{\partial x_2} &= T^+ z && \text{on}\quad \Gamma^H\,. \end{aligned} \tag{34}$$

Here, A and c are defined as in (12). The PML approximation z_σ is obtained by solving the same problem but replacing the boundary condition on Γ^H by

$$\frac{\partial z_\sigma}{\partial x_2} = T_\sigma z_\sigma\,.$$

Both z and z_σ are obtained as transformations of a corresponding rough surface wave propagation problem and its PML approximation. It is shown in [9] that both have unique solutions and it follows that $\|z - z_\sigma\|_{H^1_r(\Omega^H)} \to 0$ as $\sigma \to \infty$. However, convergence is not expected to be faster than linear with respect to σ in the unbounded domain Ω^H.

As before, we apply the FB transform to these problems and find that z is a solution if and only if $w_{g,\mathrm{p}} = \mathcal{J} z$ satisfies

$$\int_\Lambda \big(a_\alpha\big(w_{g,\mathrm{p}}(\alpha,\cdot), \varphi(\alpha,\cdot)\big) + b_\alpha(\mathcal{J}^{-1} w_{g,\mathrm{p}}, \varphi(\alpha,\cdot)\big)\,\mathrm{d}\alpha = \int_\Lambda \langle \mathcal{J} g(\alpha,\cdot), \varphi(\alpha,\cdot)\rangle_{\Omega^H_{2\pi}}\,\mathrm{d}\alpha$$

for all $\varphi \in L^2(\Lambda, \widetilde{H}^1_{\mathrm{per}}(\Omega^H_{2\pi}))$. The PML problem is to find $w_{g,\mathrm{p},\sigma} = \mathcal{J} z_\sigma$ such that

$$\int_\Lambda \big(a_{\alpha,\sigma}\big(w_{g,\mathrm{p},\sigma}(\alpha,\cdot), \varphi(\alpha,\cdot)\big) + b_\alpha(\mathcal{J}^{-1} w_{g,\mathrm{p},\sigma}, \varphi(\alpha,\cdot)\big)\,\mathrm{d}\alpha = \int_\Lambda \langle \mathcal{J} g(\alpha,\cdot), \varphi(\alpha,\cdot)\rangle_{\Omega^H_{2\pi}}\,\mathrm{d}\alpha$$

for all $\varphi \in L^2(\Lambda, \widetilde{H}^1_{\mathrm{per}}(\Omega^H_{2\pi}))$.

The argument in the proof of [22, Theorem 3.3] immediately shows that $\alpha \mapsto w_{g,\mathrm{p}}$ is continuous, so that we may equivalently formulate the first problem as

$$a_\alpha\big(w_{g,\mathrm{p}}(\alpha,\cdot), \varphi\big) + b_\alpha(\mathcal{J}^{-1} w_{g,\mathrm{p}}, \varphi) = \langle \mathcal{J} g(\alpha,\cdot), \varphi\rangle_{\Omega^H_{2\pi}} \quad \text{for all } \varphi \in \widetilde{H}^1_{\mathrm{per}}(\Omega^H_{2\pi})\,,\ \alpha \in \Lambda\,. \tag{35}$$

From the definition of b_α in (25), it is immediately clear that the induced operator $\mathcal{B}_\alpha : H^1(\Omega^H_{2\pi}) \to H^{-1}_{\mathrm{per}}(\Omega^H_{2\pi})$ depends analytically on α. We may now write both the source problem (35) and its PML approximation as operator equations in $C(\Lambda, \widetilde{H}^1_{\mathrm{per}}(\Omega^H_{2\pi}))$,

$$\mathcal{A}_\alpha w_{g,\mathrm{p}} + \mathcal{B}_\alpha \mathcal{J}^{-1} w_{g,\mathrm{p}} = \mathcal{J} g\,, \qquad \mathcal{A}_{\alpha,\sigma} w_{g,\mathrm{p},\sigma} + \mathcal{B}_\alpha \mathcal{J}^{-1} w_{g,\mathrm{p},\sigma} = \mathcal{J} g\,, \qquad \alpha \in \Lambda\,.$$

Invertibility of the operators $\mathcal{A}_\alpha$ and $\mathcal{A}_{\alpha,\sigma}$ has already been shown. Thus we may apply these inverses to both equations, respectively, and argue as in the fully periodic case that we may shift the path of integration in the inverse FB transform to $\mathcal{E}$. This leads to the equations

$$w_{g,\mathrm{p}} + \mathcal{A}_\alpha^{-1} \mathcal{B}_\alpha \mathcal{J}^{-1} w_{g,\mathrm{p}} = \mathcal{A}_\alpha^{-1} \mathcal{J} g\,, \qquad w_{g,\mathrm{p},\sigma} + \mathcal{A}_{\alpha,\sigma}^{-1} \mathcal{B}_\alpha \mathcal{J}^{-1} w_{g,\mathrm{p},\sigma} = \mathcal{A}_{\alpha,\sigma}^{-1} \mathcal{J} g$$

in $C(\mathcal{E}, \widetilde{H}^1_{\mathrm{per}}(\Omega^H_{2\pi}))$. As these equations are equivalent to the corresponding source problem and its PML approximation, respectively, we know that the operators on the left-hand side are boundedly invertible. The perturbation theorem (see [17, Theorem 10.1]) implies that

$$\|w_{g,\mathrm{p}} - w_{g,\mathrm{p},\sigma}\|_{C(\mathcal{E}, \widetilde{H}^1_{\mathrm{per}}(\Omega^H_{2\pi}))} \le C\, \|\mathcal{A}_\alpha^{-1} - \mathcal{A}_{\alpha,\sigma}^{-1}\|\, \|g\|_{H^{-1}_{\mathrm{per}}(\Omega^H_{2\pi})}\,.$$

Applying the inverse FB transform and arguing as in [28], we have proven the following theorem.

Theorem 9. *Let $g \in L^2(\Omega^H)$ with compact support, z be the solution of (34) and z_σ be its PML approximation. Then, for every compact $K \subseteq \Omega^H$ and sufficiently large σ, there exist C, $c > 0$ such that*

$$\|z - z_\sigma\|_{H^1(K)} \leq C\, \mathrm{e}^{-c|\sigma|}\,.$$

Based on Theorem 9, we can now address the case of a scattering problem.

Theorem 10. *Let $y \in \Omega^H$ and $u^s_{\text{tra}} \in H^1_r(\Omega^H)$ denote a weak solution to (11) for*

$$u^i(x) = \Phi(x,y) - \Phi(x,y')\,, \qquad x, y \in \Omega^H_{\text{p}}\,, \quad x \neq y\,.$$

Moreover, let u_σ denote the weak solution of the PML problem (17). Then, for every compact subset $K \subseteq \Omega^H$, there exist C, $c > 0$ such that

$$\|u^s_{\text{tra}} - u_\sigma\|_{H^1(K)} \leq C\, \mathrm{e}^{-c|\sigma|}\,.$$

Proof. We reuse the functions g, v^i and v^i_σ from the proof of Theorem 8. Let v^s_σ again denote the solution of (17) with u^i replaced by v^i_σ. Set $z = u^s_{\text{tra}} + v^i \circ \psi$ and $z_\sigma = v^s_\sigma + v^i_\sigma$. Then z is the solution of (34) with g replaced by $g \circ \psi$ and z_σ its PML approximation. Still, $g \circ \psi$ has compact support so that we can use Theorem 9,

$$\|z - z_\sigma\|_{H^1(K)} \leq C\, \mathrm{e}^{-c|\sigma|}\,.$$

By the same arguments, the corresponding estimate holds for $v^i \circ \psi - v^i_\sigma \circ \psi$.

As in the proof of Theorem 8, it remains to consider $v = v^s_\sigma - u_\sigma$, which is a weak solution to

$$\begin{aligned} \Delta v + k^2 v = 0 \quad & \text{in} \quad \Omega^H\,, \\ v = (u^i - v^i_\sigma) \circ \psi \quad & \text{on} \quad \Gamma\,, \\ \frac{\partial v}{\partial x_2} = T^+_\sigma v \quad & \text{on} \quad \Gamma_H\,. \end{aligned}$$

The boundary values on Γ are equal to $u^i - v^i_\sigma$ plus a compactly supported term. Applying the FB transform gives $\mathcal{J}(u^i - v^i_\sigma)$ plus an analytic term, specifically

$$\begin{aligned} &\mathcal{J}\left((u^i - v^i_\sigma) \circ \psi\right)(\alpha, x) \\ &\quad = \mathcal{J}\left(u^i - v^i_\sigma\right)(\alpha, x) + \left((u^i - v^i_\sigma)(\psi(x)) - (u^i - v^i_\sigma)(x)\right) \mathrm{e}^{-\mathrm{i}\alpha x_1}\,, \quad x \in \Omega^H_{2\pi}\,,\ \alpha \in \mathcal{E}\,. \end{aligned}$$

Thus, the argument to change the integration path in $\mathcal{J}^{-1}$ from Λ to $\mathcal{E}$ remains valid. Moreover, the mapping properties of the FB transform in Sobolev spaces proved in [20, Theorem 4 (b)] as well as standard estimates for Sobolev norms for perturbed domains allow us to estimate

$$\|\left((u^i - v^i_\sigma) \circ \psi - (u^i - v^i_\sigma)\right) \mathrm{e}^{-\mathrm{i}\alpha\cdot}\|_{H^{1/2}_{\text{per}}(\Gamma_{2\pi})} \leq C\, \|\mathcal{J}\left(u^i - v^i_\sigma\right)(\alpha, \cdot)\|_{H^{1/2}_{\text{per}}(\Gamma_{2\pi})}\,, \qquad \alpha \in \mathcal{E}\,.$$

Now the proof is completed by the same arguments as in the proof of Theorem 8. □

5 An Efficient Numerical Scheme

The usual advantage in applying the FB transform to solve a boundary value problem in a periodic domain is that one obtains a decoupled family of periodic problems that can be solved independently of each other. The situation is different for (28): a naive discretization leads to a family of periodic problems that are all coupled due to the presence of the term $\mathcal{J}^{-1}w_\sigma$ and thus to a very large linear system with a prohibitive computational cost. However, in our arguments below, we will show that, by manipulating this system in the right way, we may obtain a formulation that can be solved efficiently using an iterative solver.

To represent the connection between the scattered field u^s_σ and its FB transform $w_\sigma(\alpha,\cdot)$, we will discretize the inverse FB transform (21) using a quadrature formula. Given quadrature points $\alpha_j \in \Lambda = [-1/2, 1/2]$ and weights μ_j, $j = 1,\dots,N$, we have an approximation

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} f(\alpha)\,\mathrm{d}\alpha \approx \sum_{j=1}^{N} \mu_j\, f(\alpha_j)\,.$$

To discretize the variational problem (28), we use the finite element method. For ease of presentation, we restrict ourselves to finite element spaces of piecewise linear polynomials. We discretize the domain $\Omega^H_{2\pi}$ by a mesh supporting a family $\{\varphi_m\}_{m=1}^M$ of M such basis functions and then approximate, for each quadrature point α_j,

$$w_\sigma(\alpha_j, x) \approx w_{j,M}(x) = \sum_{m=1}^{M} W_{j,m}\,\varphi_m(x)\,, \qquad W_{j,m}\in\mathbb{C}, \quad j=1,\dots,N\,, \quad m=1,\dots,M\,,$$

and also the scattered field,

$$u^s_\sigma(x) \approx u_M(x) = \sum_{m=1}^{M} U_m\,\varphi_m(x)\,, \qquad U_m\in\mathbb{C}\,, \quad m=1,\dots,M\,.$$

The Galerkin approximation of (28) then becomes

$$\sum_{m=1}^{M} W_{j,m}\,a_{\alpha_j,\sigma}(\varphi_m,\varphi_\ell) - \sum_{m=1}^{M} U_m\,b_{\alpha_j}(\varphi_m,\varphi_\ell) = 0, \qquad \ell=1,\dots,M, \quad j=1,\dots,N\,, \tag{36}$$

together with the discrete form of the boundary condition,

$$W_{j,m} - \mathcal{J}(u^i\circ\psi)(\alpha_j, x_m) = 0\,, \qquad m=\widetilde{M}+1,\dots,M, \quad j=1,\dots,N$$

where $\widetilde{M}+1,\dots,M$ denotes the indices of the boundary nodes in $\Gamma_{2\pi}$. Additionally, the inverse FB transform (21) is discretized as

$$U_m = \sum_{j=1}^{N} \mu_j\,\mathrm{e}^{\mathrm{i}\alpha_j (x_1)_m}\,W_{j,m}\,, \quad m=1,\dots,M\,, \tag{37}$$

where μ_j denotes the weight of the quadrature rule.

Introducing the vectors of unknowns $W_j = [W_{j,1}, \dots, W_{j,M}]^\top$ and $U = [U_1, \dots, U_M]^\top$, we can formulate the system (36)-(37) in the block vector-matrix form as

$$\begin{bmatrix} A_1 & 0 & \cdots & 0 & B_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & A_N & B_N \\ C_1 & \cdots & \cdots & C_N & I \end{bmatrix} \begin{bmatrix} W_1 \\ \vdots \\ \vdots \\ W_N \\ U \end{bmatrix} = \begin{bmatrix} F_1 \\ \vdots \\ \vdots \\ F_N \\ 0 \end{bmatrix} \in \mathbb{C}^{(N+1)M}, \tag{38}$$

where the block matrices A_j, B_j and $C_j \in \mathbb{C}^{M\times M}$ and the vector $F_j \in \mathbb{C}^M$ for $j = 1, \dots, N$ are defined by

$$(A_j)_{\ell,m} = \begin{cases} a_{\alpha_j,\sigma}(\varphi_m, \varphi_\ell) & \text{if } \ell = 1, \dots, M, m = 1, \dots, \widetilde{M}, \\ \delta_{m\ell} & \text{if } \ell, m = \widetilde{M}+1, \dots, M, \end{cases}$$

$$(B_j)_{\ell,m} = \begin{cases} b_{\alpha_j}(\varphi_m, \varphi_\ell), & \text{if } \ell = 1, \dots, M, m = 1, \dots, \widetilde{M}, \\ 0, & \text{if } \ell, m = \widetilde{M}+1, \dots, M, \end{cases}$$

$$(C_j)_{\ell,m} = \mu_j \mathrm{e}^{\mathrm{i}\alpha_j (x_1)_m} \delta_{m,\ell}, \qquad \text{for } \ell, m = 1, \dots, M$$

and

$$(F_j)_\ell = \big(\mathcal{J} u^i\big)(\alpha_j, x_\ell)\delta_{m,\ell}, \quad \text{for } m = 1, \dots, M, \quad \ell = \widetilde{M}+1, \dots, M.$$

Remark 11. *The coefficient matrix in (38) is known as a permuted square arrowhead matrix, which frequently arises in applications (see, e.g., [4, 19]). Different approaches to explicitly computing the inverse of arrowhead matrices are presented in [26]. However, inverting the coefficient matrix of (38) using these approaches is still computationally expensive. In our arguments below, we will propose an alternative method to solve the linear system (38) without inverting the coefficient matrix of (38).*

In the following theorem, we recursively apply the Schur complement to obtain an equivalent form of the system (38), which can be parallelized more easily.

Theorem 12. *Let A_j, B_j, C_j and F_j for $j = 1, \dots, N$ be defined as above. The linear system (38) is equivalent to*

$$\left(I - \sum_{j=1}^{N} C_j A_j^{-1} B_j \right) U = - \sum_{j=1}^{N} C_j A_j^{-1} F_j. \tag{39}$$

That is, if $[W_1, W_2, \dots, W_N, U]^\top$ solves (38), then U solves (39) and if U solves (39), then $[A_1^{-1}(F_1 - B_1 U), \dots, A_N^{-1}(F_N - B_N U), U]^\top$ solves (38).

Proof. The proof presents an algorithm to reduce (38) to (39), by recursively applying a procedure that removes one unknown vector W_ℓ (for $\ell = 1, \dots, N$). The assertions are followed by an induction on the number of unknowns removed.

For $\ell = 1$, we rewrite the system (38) as follows

$$\begin{bmatrix} A_1 & B_1^{(\mathrm{rem})} \\ C_1^{(\mathrm{rem})} & D_1^{(\mathrm{rem})} \end{bmatrix} \begin{bmatrix} W_1 \\ W_1^{(\mathrm{rem})} \end{bmatrix} = \begin{bmatrix} F_1 \\ F_1^{(\mathrm{rem})} \end{bmatrix} \in \mathbb{C}^{(N+1)M}, \tag{40}$$

where

$$B_1^{(\mathrm{rem})} = \begin{bmatrix} 0 & \cdots & 0 & B_1 \end{bmatrix}, \quad F_1^{(\mathrm{rem})} = \begin{bmatrix} F_2 & \cdots & F_N & 0 \end{bmatrix}^\top,$$
$$C_1^{(\mathrm{rem})} = \begin{bmatrix} 0 & \cdots & 0 & C_1 \end{bmatrix}^\top, W_1^{(\mathrm{rem})} = \begin{bmatrix} W_2 & \cdots & W_N & U \end{bmatrix}^\top,$$

and the block $D_1^{(\mathrm{rem})}$ is given by

$$D_1^{(\mathrm{rem})} = \begin{bmatrix} A_2 & 0 & \cdots & 0 & B_2 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & A_N & B_N \\ C_2 & \cdots & \cdots & C_N & I \end{bmatrix} \in \mathbb{C}^{NM \times NM}.$$

From the first equation in (40), we obtain $W_1 = A_1^{-1}\left(F_1 - B_1^{(\mathrm{rem})} W_1^{(\mathrm{rem})}\right)$. To remove the first unknown, substituting W_1 into the second equation of (40) leads to

$$D_1^{(\mathrm{rem})} W_1^{(\mathrm{rem})} = F_1^{(\mathrm{rem})} - C_1^{(\mathrm{rem})} W_1 = F_1^{(\mathrm{rem})} - C_1^{(\mathrm{rem})} A_1^{-1}\left(F_1 - B_1^{(\mathrm{rem})} W_1^{(\mathrm{rem})}\right).$$

Then, the remaining system is written as

$$\left(D_1^{(\mathrm{rem})} - C_1^{(\mathrm{rem})} A_1^{-1} B_1^{(\mathrm{rem})}\right) W_1^{(\mathrm{rem})} = F_1^{(\mathrm{rem})} - C_1^{(\mathrm{rem})} A_1^{-1} F_1,$$

where the coefficient matrix is given by

$$D_1^{(\mathrm{rem})} - C_1^{(\mathrm{rem})} A_1^{-1} B_1^{(\mathrm{rem})} = \begin{bmatrix} A_2 & 0 & \cdots & 0 & B_2 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & A_N & B_N \\ C_2 & \cdots & \cdots & C_N & I - C_1 A_1^{-1} B_1 \end{bmatrix},$$

and the right-hand side is determined by

$$F_1^{(\mathrm{rem})} - C_1^{(\mathrm{rem})} A_1^{-1} F_1 = \begin{bmatrix} F_2 & \cdots & F_N & -C_1 A_1^{-1} F_1 \end{bmatrix}^\top.$$

So far, the first unknown W_1 has been removed in the initial step. Now, we assume that the theorem holds for $\ell - 1$. To prove that it also holds for ℓ, we need to solve the following linear system

$$\begin{bmatrix} A_\ell & 0 & \cdots & 0 & B_\ell \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & A_N & B_N \\ C_\ell & \cdots & \cdots & C_N & I - \sum_{j=1}^{\ell-1} C_j A_j^{-1} B_j \end{bmatrix} \begin{bmatrix} W_\ell \\ \vdots \\ \vdots \\ W_N \\ U \end{bmatrix} = \begin{bmatrix} F_\ell \\ \vdots \\ \vdots \\ F_N \\ -\sum_{j=1}^{\ell-1} C_j A_j^{-1} F_j \end{bmatrix},$$

which can be written as follows

$$\begin{bmatrix} A_\ell & B_\ell^{(\mathrm{rem})} \\ C_\ell^{(\mathrm{rem})} & D_\ell^{(\mathrm{rem})} \end{bmatrix} \begin{bmatrix} W_\ell \\ W_\ell^{(\mathrm{rem})} \end{bmatrix} = \begin{bmatrix} F_\ell \\ F_\ell^{(\mathrm{rem})} \end{bmatrix} \in \mathbb{C}^{(N+1-(\ell-1))M}, \tag{41}$$

where the block matrices $B_\ell^{(\mathrm{rem})} \in \mathbb{C}^{M\times M(N+1-\ell)}$, $C_\ell^{(\mathrm{rem})} \in \mathbb{C}^{M(N+1-\ell)\times M}$ and the vectors $F_\ell^{(\mathrm{rem})}, W_\ell^{(\mathrm{rem})} \in \mathbb{C}^{M(N+1-\ell)}$ are defined by

$$B_\ell^{(\mathrm{rem})} = \begin{bmatrix} 0 & \cdots & 0 & B_\ell \end{bmatrix}, \quad F_\ell^{(\mathrm{rem})} = \begin{bmatrix} F_{\ell+1} & \cdots & F_N & -\sum_{j=1}^{\ell-1} C_j A_j^{-1} F_j \end{bmatrix}^\top,$$
$$C_\ell^{(\mathrm{rem})} = \begin{bmatrix} 0 & \cdots & 0 & C_\ell \end{bmatrix}^\top, \quad W_\ell^{(\mathrm{rem})} = \begin{bmatrix} W_{\ell+1} & \cdots & W_N & U \end{bmatrix}^\top$$

and the block $D_\ell^{(\mathrm{rem})}$ is given by

$$D_\ell^{(\mathrm{rem})} = \begin{bmatrix} A_{\ell+1} & 0 & \cdots & 0 & B_{\ell+1} \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & A_N & B_N \\ C_{\ell+1} & \cdots & \cdots & C_N & I - \sum_{j=1}^{\ell-1} C_j A_j^{-1} B_j \end{bmatrix} \in \mathbb{C}^{(N+1-\ell)M\times(N+1-\ell)M}.$$

Then, the remaining system is written as

$$\left(D_\ell^{(\mathrm{rem})} - C_\ell^{(\mathrm{rem})} A_\ell^{-1} B_\ell^{(\mathrm{rem})}\right) W_\ell^{(\mathrm{rem})} = F_\ell^{(\mathrm{rem})} - C_\ell^{(\mathrm{rem})} A_\ell^{-1} F_\ell,$$

where

$$C_\ell^{(\mathrm{rem})} A_\ell^{-1} B_\ell^{(\mathrm{rem})} = \begin{bmatrix} 0_{(N-\ell)M\times(N-\ell)M} & 0_{(N-\ell)M\times M} \\ 0_{M\times(N-\ell)M} & C_\ell A_\ell^{-1} B_\ell \end{bmatrix},$$

$$D_\ell^{(\mathrm{rem})} - C_\ell^{(\mathrm{rem})} A_\ell^{-1} B_\ell^{(\mathrm{rem})} = \begin{bmatrix} A_{\ell+1} & \cdots & 0 & B_{\ell+1} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & A_N & B_N \\ C_{\ell+1} & \cdots & C_N & I - \sum_{j=1}^{\ell} C_j A_j^{-1} B_j \end{bmatrix}.$$

At step $\ell = N+1$, the N-th unknown has been removed and this completes the proof. □

Equation (39) can now be solved using an iterative method as described in Algorithm 1. Note that the summands on the right-hand side, i.e., $C_j A_j^{-1} F_j$ and the matrix vector multiplications by the summands on the left-hand side, i.e., $C_j A_j^{-1} B_j U$ are all independent of each other. Hence, they can be performed in parallel.

6 Numerical Results

In this section, we illustrate the efficiency and accuracy of the iterative method described in Algorithm 1 to solve non-periodic scattering problems.

We select the non-periodic incident field as the Dirichlet Green's function in the upper half-space, i.e.,

$$u^i(x) = \frac{\mathrm{i}}{4}\left(H_0^{(1)}(k|x-y|) - H_0^{(1)}(k|x-y'|)\right),$$

Algorithm 1: iterative method for solving (39)

Input: number of quadrature nodes N, initial guess U_0

Compute the Gauss quadrature nodes and weights (α_j, μ_j) for $j \in \{1, \dots, N\}$;

for $j = 1, \dots, N$ **do in parallel**

  Construct the matrices A_j, B_j, C_j and the vector F_j using FEM;

  Compute the LU decomposition of A_j;

  Solve the system $A_j V_j = F_j$ using the above LU decomposition;

  $V_j \leftarrow C_j V_j$;

$V \leftarrow \sum_{j=1}^{N} V_j$;

%To solve the systems on the left-hand side of (39), the following function computes the matrix-vector multiplication for each input.

Define the function *left_hand_side()*

  Input: the vector U

  for $j = 1, \dots, N$ **do in parallel**

    Solve $A_j X_j = B_j U$ using the precomputed LU decomposition of A_j;

    $X_j \leftarrow C_j X_j$;

  return $U - \sum_{j=1}^{N} X_j$;

Solve (39) by GMRES with tolerance 10^{-5} and inputs U_0 and *left_hand_side()*;

return *Numerical solution of* (39)

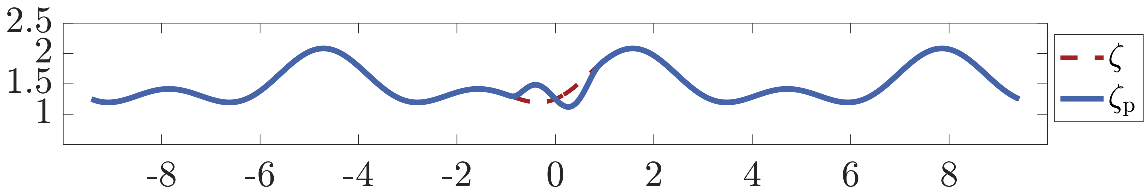


Figure 2: Illustration of the locally perturbed curve Γ_{p}.

where $y = (y_1, y_2)^\top$ is a fixed point source and $y' = (y_1, -y_2)^\top$ is its reflection with respect to $\{x \in \mathbb{R}^2 : x_2 = 0\}$.

We apply our proposed method to compute the scattered field produced by the locally perturbed scatterers described in the following example.

We consider the locally perturbed curve $\Gamma_{\mathrm{p}} = \{(x, \zeta(x) + \delta(x)) : x \in \mathbb{R}\}$, plotted in Figure 2, with the periodic function

$$\zeta(x) = 1.5 + \frac{\sin(x)}{3} - \frac{\cos(2x)}{4}, \quad x \in \mathbb{R}$$

and the perturbation

$$\delta(x) = \exp\left(\frac{1}{x^2 - 1}\right) \sin(\pi(x+1)) \chi_{[-1,1]}(x) .$$

To calculate the error explicitly, we consider the point source y between the flat surface $\mathbb{R} \times \{0\}$ and the locally perturbed scatterers Γ_{p}, since in this case the total field vanishes inside the domain. This means that the exact solution is equal to minus the incident field.

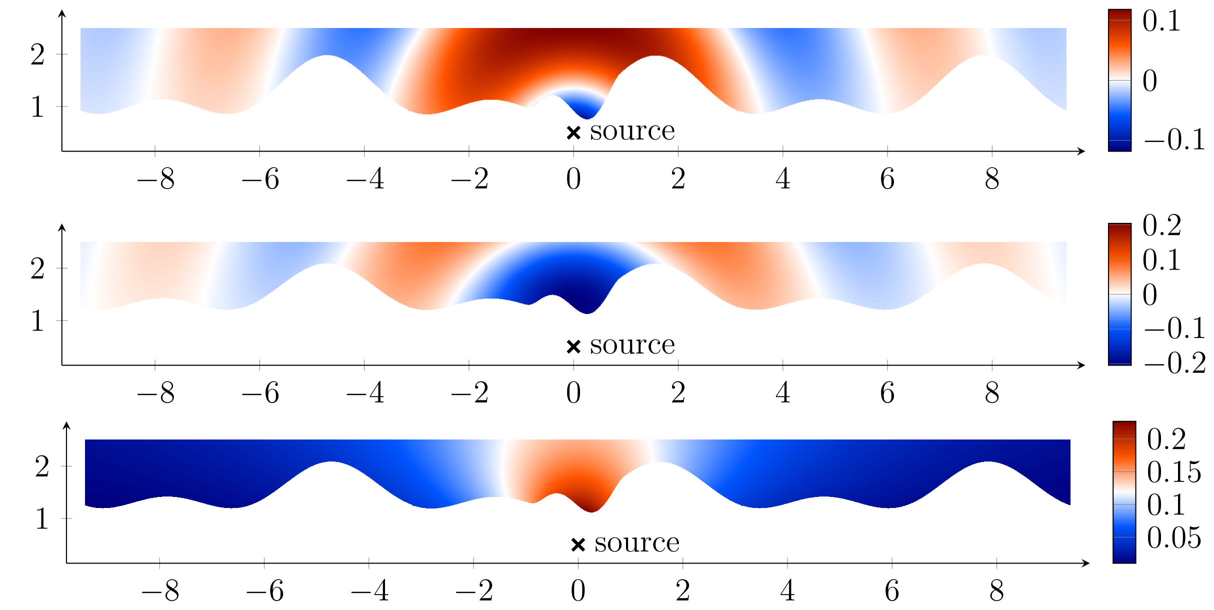


Figure 3: Numerical scattered field with $k = \sqrt{2}$ and $y = (0, 0.5)^\top$ (top: real part, middle: imaginary part, bottom: absolute value).

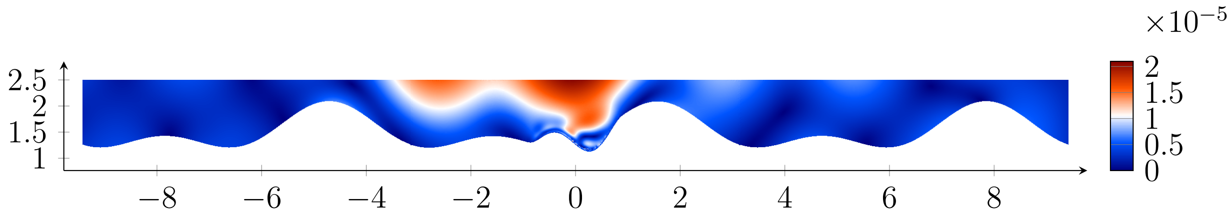


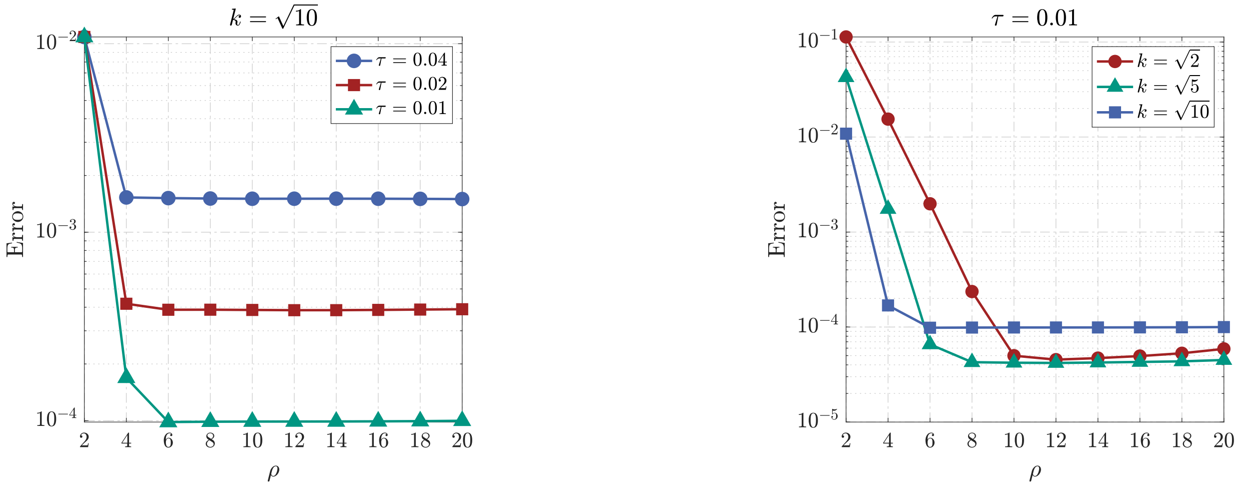
Figure 4: Absolute value of the error with $k = \sqrt{2}$.

In this example, we consider the point source located at $y = (0, 0.5)^\top$ below the surface Γ_p and fix $H = 2.5$. The physical thickness of the PML is $\lambda = 1.5$ and the PML function given in (13) depends only on a positive parameter ρ.

To approximate the scattered field in the main bounded cell $\Omega^H_{2\pi} = \{x \in \Omega^H : x_1 \in (-\pi, \pi)\}$ numerically, we use the iterative solver introduced in Algorithm 1 by setting $N = 20$ and $U_0 = 0$. Once we have the numerical solution for the main cell $\Omega^H_{2\pi}$, we can extend it to the neighboring cells $\Omega^H_{2\pi+j} = \{x \in \Omega^H_\mathrm{p} : x_1 \in (-\pi, \pi) + 2\pi j\}$, for $j = \pm 1$. This extension is obtained by using the discrete inverse FB defined in (21), for $j = \pm 1$.

The behaviour of the numerical scattered field is illustrated in Figure 3 with $k = \sqrt{2}$. These results were obtained using the mesh size of $\tau = 0.01$ and the PML parameter $\rho = 20$. In addition, the absolute value of the numerical error is plotted in Figure 4. They demonstrate that the maximum value of the error is less than 2×10^{-5}, which indicates the accuracy of the proposed method. Moreover, it is evident that the absolute value of the error increases while approaching the PML. This behavior is expected because the PML introduces a numerical error due to the approximation of the DtN map.

In what follows, we analyze the dependence of the relative L^2-error on the PML parameter ρ for various discretization parameters.



(a) for different mesh size τ

(b) for different wave number k

Figure 5: Relative L^2-error with respect to the PML parameter ρ.

In Table 1, we report the relative L^2-error of the proposed method with respect to the PML parameter ρ and the mesh size τ for a fixed wave number k. This result is depicted in Figure 5a for $k = \sqrt{10}$. The error decreases exponentially with increasing ρ up to $\rho = 4$. Beyond these values, the error is dominated by the discretization of the FEM. This behavior is evident from the results shown in Table 1. For $\rho > 4$, exponential convergence ceases and the method exhibits quadratic convergence with respect to the mesh size.

ρ	$\tau = 0.04$	$\tau = 0.02$	$\tau = 0.01$
2	1.0856×10^{-2}	1.0865×10^{-2}	1.0850×10^{-2}
4	1.5296×10^{-3}	4.1735×10^{-4}	1.6889×10^{-4}
6	1.5171×10^{-3}	3.8782×10^{-4}	9.8359×10^{-5}
8	1.5092×10^{-3}	3.8807×10^{-4}	9.8753×10^{-5}
10	1.5054×10^{-3}	3.8687×10^{-4}	9.8921×10^{-5}
12	1.5058×10^{-3}	3.8572×10^{-4}	9.8961×10^{-5}
14	1.5066×10^{-3}	3.8582×10^{-4}	9.9011×10^{-5}
16	1.5058×10^{-3}	3.8707×10^{-4}	9.9185×10^{-5}
18	1.5032×10^{-3}	3.8869×10^{-4}	9.9428×10^{-5}
20	1.4995×10^{-3}	3.9025×10^{-4}	9.9874×10^{-5}

Table 1: Relative L^2-error with respect to the PML parameter ρ and mesh size τ with wave number $k = \sqrt{10}$.

In Table 2, we report the relative L^2-error with respect to the PML parameter ρ and the wave number k for the mesh size $\tau = 0.01$. These results are depicted in Figure 5b for both examples. We again observe an exponential rate of convergence for the wave numbers $k = \sqrt{2}, \sqrt{5}$ and $\sqrt{10}$. Furthermore, the graphs indicate that the damping effect of the PML is more pronounced when the value of $k\rho$ is higher. That is, the convergence is faster and is reached at a lower value of ρ when the wave number k is larger. For each fixed ρ, the error is smaller for larger k unless the spatial discretization error dominates.

ρ	$k=\sqrt{2}$	$k=\sqrt{5}$	$k=\sqrt{10}$
2	1.1309×10^{-1}	4.2618×10^{-2}	1.0850×10^{-2}
4	1.5438×10^{-2}	1.7487×10^{-3}	1.6890×10^{-4}
6	1.9835×10^{-3}	6.5705×10^{-5}	9.8360×10^{-5}
8	2.3638×10^{-4}	4.2657×10^{-5}	9.8753×10^{-5}
10	4.9921×10^{-5}	4.2107×10^{-5}	9.8922×10^{-5}
12	4.5381×10^{-5}	4.1870×10^{-5}	9.8961×10^{-5}
14	4.7108×10^{-5}	4.2330×10^{-5}	9.9012×10^{-5}
16	4.9491×10^{-5}	4.2909×10^{-5}	9.9185×10^{-5}
18	5.2962×10^{-5}	4.3539×10^{-5}	9.9429×10^{-5}
20	5.8908×10^{-5}	4.4983×10^{-5}	9.9875×10^{-5}

Table 2: Relative L^2-error with respect to the PML parameter ρ and wave number k with $\tau = 0.01$.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 258734477 – SFB 1173. The work of N. Shafieeabyaneh and R. Zhang was also supported by DFG Grant Project-ID 433126998.

References

[1] T. Arens, N. Shafieeabyaneh, and R. Zhang. A high-order numerical method for solving non-periodic scattering problems in three-dimensional bi-periodic structures. *ZAMM Z. Angew. Math. Mech.*, 104(9):Paper No. e202300650, 21, 2024. doi:10.1002/zamm.202300650.

[2] T. Arens and R. Zhang. A nonuniform mesh method for wave scattered by periodic surfaces. *Math. Meth. Appl. Sci*, 48:4289–4309, 2025. doi:10.1002/mma.10548.

[3] G. Bao. Finite element approximation of time harmonic waves in periodic structures. *SIAM J. Numer. Anal.*, 32:1155–1169, 1995. doi:10.1137/0732053.

[4] M. Bixon and J. Jortner. Intramolecular radiationless transitions. *J. Chem. Phys.*, 48(2):715–726, 1968. doi:10.1063/1.1668703.

[5] O. P. Bruno and F. Reitich. Solution of a boundary value problem for the Helmholtz equation via variation of the boundary into the complex domain. *Proc. R. Soc. Edinb. A: Math.*, 122(3-4):317–340, 1992. doi:10.1017/S0308210500021132.

[6] O. P. Bruno and F. Reitich. Numerical solution of diffraction problems: a method of variation of boundaries. *J. Opt. Soc. Am. A*, 10(6):1168–1175, 1993. doi:10.1364/JOSAA.10.002307.

[7] S. Chandler-Wilde and J. Elschner. Variational approach in weighted Sobolev spaces to scattering by unbounded rough surfaces. *SIAM J. Math. Anal.*, 42:2554–2580, 2010. doi:10.1137/090776111.

[8] S. Chandler-Wilde and P. Monk. Existence, uniqueness, and variational methods for scattering by unbounded rough surfaces. *SIAM J. Math. Anal.*, 37:598–618, 2005. doi:10.1137/040615523.

[9] S. N. Chandler-Wilde and P. Monk. The PML for rough surface scattering. *Appl. Numer. Math.*, 59(9):2131–2154, 2009. doi:10.1016/j.apnum.2008.12.007.

[10] J. Coatléven. Helmholtz equation in periodic media with a line defect. *J. Comput. Phys.*, page 1675–1704, 2012. doi:10.1016/j.jcp.2011.10.022.

[11] S. Fliss and P. Joly. Wave propagation in locally perturbed periodic media (case with absorption): Numerical aspects. *J. Comput. Phys.*, 231:1244–1271, 2012. doi:10.1016/j.jcp.2011.10.007.

[12] H. Haddar and T. Nguyen. A volume integral method for solving scattering problems from locally perturbed infinite periodic layers. *Appl Anal.*, 96:130–158, 2017. doi:10.1080/00036811.2016.1221942.

[13] G. Hsiao, N. Nigam, J. Pasciak, and X. Liwei. Error analysis of the DtN-FEM for the scattering problem in acoustics via Fourier analysis. *J. Comput. Appl. Math.*, 235:4949–4965, 2011. doi:10.1016/j.cam.2011.04.020.

[14] P. Joly, J.-R. Li, and S. Fliss. Exact boundary conditions for periodic waveguides containing a local perturbation. *Commun. Comput. Phys.*, 1:945–973, 2006. doi:10.4208/cicp.2006.v1.p945.

[15] A. Kirsch. Diffraction by periodic structures. In *Inverse Problems in Mathematical Physics*, pages 87–102. Springer Berlin Heidelberg, 1993.

[16] A. Kirsch and R. Zhang. The PML-method for a scattering problem for a local perturbation of an open periodic waveguide. *Numerische Mathematik*, 157(2):717–748, 2025. doi:10.1007/s00211-025-01456-9.

[17] R. Kress. *Linear Integral Equations*. Springer, New York, 3rd edition edition, 2014. doi:10.1007/978-1-4614-9593-2.

[18] P. Kuchment. *Floquet Theory for Partial Differential Equations*, volume 60 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, Basel, 1993. doi:10.1007/978-3-0348-8573-7.

[19] H. T. Kung and B. W. Suter. A hub matrix theory and applications to wireless communications. *EURASIP J. Adv. Signal Process.*, 2007(1):Art. ID 13659, 2007. doi:10.1155/2007/13659.

[20] A. Lechleiter. The Floquet–Bloch transform and scattering from locally perturbed periodic surfaces. *J. Math. Anal. Appl.*, 446:605–627, 2017. doi:10.1016/j.jmaa.2016.08.055.

[21] A. Lechleiter and R. Zhang. A convergent numerical scheme for scattering of aperiodic waves from periodic surfaces based on the Floquet-Bloch transform. *SIAM J. Numer. Anal.*, 55:713–736, 2017. doi:10.1137/16M1067524.

[22] A. Lechleiter and R. Zhang. A Floquet–Bloch transform based numerical method for scattering from locally perturbed periodic surfaces. *SIAM J. Sci. Comput.*, 39:B819–B839, 2017. doi:10.1137/16M1104111.

[23] J. Nedelec and F. Starling. Integral equation methods in a quasi-periodic diffraction problem for the time-harmonic Maxwell's equations. *SIAM J. Numer. Anal.*, 22:1679–1701, 1991. doi:10.1137/0522104.

[24] G. Schmidt. On the diffraction by biperiodic anisotropic structures. *Appl. Anal.*, 82:75–92, 2003. doi:10.1080/0003681031000068275.

[25] N. Shafieeabyaneh. *Efficient Methods for Solving Direct and Inverse Scattering Problems in Locally Perturbed Periodic Structures.* PhD thesis, Karlsruhe Institute of Technology (KIT), Nov 2025. doi:10.5445/IR/1000189511.

[26] P. S. Stanimirović, V. N. Katsikis, and D. Kolundžija. Inversion and pseudoinversion of block arrowhead matrices. *Appl. Math. Comput.*, 341:379–401, 2019. doi:10.1016/j.amc.2018.09.006.

[27] R. Zhang. A high order numerical method for scattering from locally perturbed periodic surfaces. *SIAM J. Sci. Comput.*, 40:A2286–A2314, 2018. doi:https://10.1137/17M1144945.

[28] R. Zhang. Exponential convergence of perfectly matched layers for scattering with periodic surfaces. *SIAM J. Numer. Anal.*, 60(2):804–823, 2022. doi:10.1137/21M1439043.