

# A multiscale hp-FEM for 2D photonic crystal band

H. Brandsmeier, K. Schmidt\* and Ch. Schwab†

Research Report No. 2010-12  
April 2010

Seminar für Angewandte Mathematik  
Eidgenössische Technische Hochschule  
CH-8092 Zürich  
Switzerland

---

\*Project POEMS, INRIA Paris-Rocquencourt, 78150 Rocquencourt, France

†The research of this author was supported by the European Research Council under grant 247277

# A Multiscale hp-FEM for 2D Photonic Crystal Bands

Holger Brandsmeier<sup>a</sup>, Kersten Schmidt<sup>b,c,a</sup>, Christoph Schwab<sup>a,1</sup>

<sup>a</sup>*Seminar for Applied Mathematics, ETH Zürich, 8092 Zürich, Switzerland*

<sup>b</sup>*Project POEMS, INRIA Paris-Rocquencourt, 78150 Rocquencourt, France*

<sup>c</sup>*Hausdorff Center for Mathematics, University of Bonn, 53115 Bonn, Germany*

---

## Abstract

A Multiscale generalized *hp*-Finite Element Method (MSFEM) for time harmonic wave propagation in bands of locally periodic media of large, but finite extent, *e.g.*, photonic crystal (PhC) bands, is presented. The method distinguishes itself by its size robustness, *i.e.*, to achieve a prescribed error its computational effort does not depend on the number of periods. The proposed method shows this property for general incident fields, including plane waves incident at a certain angle to the infinite crystal surface, and at frequencies in and outside of the bandgap of the PhC. The proposed MSFEM is based on a precomputed problem adapted multiscale basis. This basis incorporates a set of complex Bloch modes, the eigenfunctions of the infinite PhC, which are modulated by macroscopic piecewise polynomials on a macroscopic FE mesh. The multiscale basis is shown to be efficient for finite PhC bands of any size, provided that boundary effects are resolved with a simple macroscopic boundary layer mesh. The MSFEM, constructed by combing the multiscale basis inside the crystal with some exterior discretisation, is a special case of the Generalised Finite Element Method (g-FEM). For the rapid evaluation of the matrix entries we introduce a size robust algorithm for integrals of quasi-periodic micro functions and polynomial macro functions. Size robustness of the present MSFEM in both, the number of basis functions and the computation time, is verified in extensive numerical experiments.

*Key words:* Finite Photonic Crystals, Multiscale FEM, Generalised FEM, Scattering, Helmholtz equation, Fast Quadrature of quasi-periodic functions.

*2010 MSC:* 35J05, 35J20, 35J25, 65N30, 78M10, 78M30, 65D30

---

## 1. Introduction

Photonic crystals (PhC) structures [40] are dielectric materials with a periodic fine structure, *i.e.*, locally the material turns out to be the periodic repetition of the same dielectric pattern. Light injected into the PhC is diffracted and refracted by the many dielectric scatterers arranged in the periodic arrays. The superposition of diffraction and refraction may lead to exceptional properties of the propagation of light, which occur especially at wavelengths comparable to the periodicity length. The light in PhC structures is dispersive and its propagation properties depend additionally on the wavelength. Such special properties are the localisation of light in PhC waveguides, its lossless bending around corners, its slowing down resulting in high intensities [4, 46] or the negative refraction [62, 57], to name but a few. Several examples of applications of PhCs can be found in [39, 75]. Periodic optical material are also used at wavelengths much larger or much smaller than the periodicity length [16] to exploit different physical properties; however, this also poses different challenges for numerical methods which will not be studied in the present work.

The properties of a particular PhC structure are rarely predictable with experience alone, and therefore numerical simulations are frequently used. Those simulations have to meet the issue of resolving the waves

---

*Email addresses:* `holger.brandsmeier@sam.math.ethz.ch` (Holger Brandsmeier), `kersten.schmidt@inria.fr` (Kersten Schmidt), `schwab@math.ethz.ch` (Christoph Schwab)

<sup>1</sup>The research of this author was supported by the European Research Council under grant 247277

in the small periodic and homogeneous parts of the devices. Most of the PhC devices consist of a large number of dielectric scatterers each of size comparable to the wavelength. Hence, their simulation by direct application of standard discretisation schemes like the finite element method (FEM) or the finite difference time domain method (FDTD) results in particularly high computational costs as the crystal's cell scale has to be resolved by the discretisation scheme. However, simplifications of the model allow to predict several properties of engineering interest without fully scale-resolving FE computations.

A significant simplification is to study the wave propagation in an infinite crystal with the same dielectric pattern as the finite crystal. In this model, the Floquet transform [47] is used, resulting in parametrised eigenvalue problems on the unit cell which can be efficiently solved by various methods [11], from plane wave expansion [41] to *hp*-FEM [68] and to adaptive eigenvalue solvers [25]. The solutions, called *Bloch modes*, are eigenmodes in the infinite crystal and can be assembled to the well-known *band diagrams*. Even with this simplified model, already quite some properties of the light propagation inside large finite PhCs and on its interface to an homogeneous medium can be deduced from the band diagram. Predictive simulations increase in accuracy the larger the finite crystal is.

A step further is to model the interface of an homogeneous material in one half-plane and a periodic material in the other one. For rectangular periodicity the Floquet-Bloch transform in the direction of the interface can be applied leading to a family of problems in an infinite strip with quasi-periodic boundary conditions [22], where the semi-infinite periodic part can be approximated by *transparent boundary conditions* for periodic waveguides [42, 17, 77]. These boundary conditions have been extended to the case of localised perturbations of the otherwise fully periodic space where only the area of the perturbation and the unit cell have to be discretised by a FEM [21].

Eigenmodes in localised perturbations or line “defects” inside infinite PhCs can be modelled by the *supercell approach* [76, 67] leading to the same eigenvalue problems and numerical methods as for the infinite crystal, just on a larger computational domain. An extension to PhC waveguides of finite extent in one direction, the so called open supercell approach, has been studied in [72, 67].

With the progress in the development of photonic crystal devices, more advanced structures are studied [78] which include bends and branchings. These can be combined to form complex PhC circuits like directional couplers [45] or Mach-Zehnder interferometers [74]. Such demanding PhC structures consist of a finite, but large number of scatterers in which the wave propagation can hardly be predicted by model reductions like an infinite crystal or a line defect inside an infinite crystal.

For these advanced structures, efficient high resolution algorithms to accurately simulate large *finite* PhC structures are needed. A measure to compare algorithms is their *efficiency*, which we define in the present work as quotient of the achieved accuracy divided by the required computational effort. Algorithms with an efficiency independent of the number of scatterers in the PhC are desirable. We call such an algorithm *size robust* and an algorithm without this property *size dependent*. The computational effort will be measured either in the number of degrees of freedom (nDOFs), *e.g.*, the number of unknowns in a linear system that is solved in an algorithm, or in the algorithm's CPU time. Direct application of discretisation schemes like FEM or FDTD are unsuited for these simulations as they are *fully* size dependent, *i.e.*, their computational effort is at least proportional to the number  $n$  of scatterers. With the progress in the development of larger and more complex PhC structures it becomes essential to have more efficient simulation methods that are ideally size robust or are *weakly* size dependent with a computational effort strictly less than  $O(n)$ .

In elliptic systems, for example in heat conduction or elastodynamics, the high complexity of large, locally-periodic structures can be avoided by using homogenisation techniques [7, 59] which amounts to a size robust discretisation method in our terminology. Applying these techniques the periodic structures can be accurately modelled by a homogeneous material with an effective material coefficient. Similar techniques have been applied for wave-propagation in periodic media if the periodicity lengths are much smaller than the wavelength [30, 18]. For photonic crystals where the wavelength is of the order of the size of the periodicity length the crystal can no longer be accurately replaced by an effective material. An effective material can, for instance, hardly describe specific effects like bandgaps or slow light in finite PhC structures.

For the scattering on finite PhC structures, two different approaches have been proposed for numerical simulations that are more effective compared to the direct application of FEM or FDTD. The first, originally the scattering or S-matrix approach, applies to PhC circuits [56], that is a network of waveguides and

localized defects like PhC bends or junctions inside a PhC background. Inside the waveguides, the solution is represented by a finite number of (guided) modes of the infinite waveguide. The solution in each defect area is solved independently using the waveguide modes as the incoming and out-going radiation conditions. The efficiency of this approach can be further enhanced by using only a few Wannier functions per period [12] or an hierarchy of Dirichlet-to-Neumann operators [34]. This (approximative) mode decomposition is intended to gain a size-robust method; however, justifying investigations for large circuits have not been shown yet. The second approach is to study reflection and transmission of plane waves or Gaussian beams on a finite PhC by directly coupling plane waves to propagating and decaying Bloch modes at the surface of the crystal [36, 38]. The method works by approximately conserving the reflection and transmission coefficients on the surface. To the knowledge of the authors; however, it has not yet been studied if the discretization error remains at the same order regardless of the number of scatterers in the PhC and thus if the method is indeed size robust. We refer to Istrate and Sargent [37] for a review of methods to calculate the eigenstates in so called PhC heterostructures including the above mentioned PhC waveguides and finite PhCs.

In this article we will propose a numerical method in the framework of the generalised FEM (g-FEM). The g-FEM was introduced by Babuška *et. al.* [79, 6] as a combination of the classical Finite Element Method (FEM) and the Partition of Unity Method (PUFEM [54]). In contrast to many mesh-free methods with special basis functions, the g-FEM keeps the standard FEM basis functions and uses standard meshes (in 2D triangular or quadrilateral) to form the partition of unity. We will propose a Multiscale generalised *hp*-Finite Element Method (MSFEM) for photonic crystal bands with two-scale basis functions. The fine scale basis consists of Bloch modes whereas the coarse basis are continuous, piecewise polynomials on a macroscopic mesh. The piecewise polynomials of maximal degree  $p_{\text{mac}} \geq 1$  localise the fine scale basis functions. The two scale basis functions obtained by modulating Bloch-modes of infinite PhC's which resolve the microscopic structure of the PhC with macroscopic *hp*-Finite Element shape functions constitute a partition of unity FEM in the sense of [54].

Beyond the continuous, piecewise linear macroscopic FE spaces required for PUFEM the two scale basis contains further basis functions for  $p_{\text{mac}} > 1$  and the method will converge for  $p_{\text{mac}} \rightarrow \infty$  similar to p-FEM, regardless of the choice of the microscopic FE-space. The strength of the method; however, is its problem adapted fine scale basis which seems to imply as shown in the present work an error reduction that is independent of the size of the crystal, *i.e.*, *size robustness*.

The fine scale basis is specifically adapted to the local structure of the PhC. It contains decaying and non-decaying Bloch modes at the frequency of the incident wave. This macroscopic space can be an *h*-, *p*- or *hp*-FE space in which case we will call the resulting method a *multiscale h-, p- or hp-FEM*.

The presently proposed method extends ideas from the generalised FEM for elliptic homogenization problems with periodic micro-structure by Matache *et. al.* [52, 51] and Rügge [64] to PhC's. Related work for numerical homogenization in the non-periodic setting is [32, 33, 1]. The multiscale FEM (MSFEM) is introduced here for the scattering of incident waves by a model photonic crystal band which is of infinite extent in one direction, but is of finite and possibly large extent in the other direction. Numerical results for plane waves at different incident angles are reported which strongly indicate the size robustness of the presently method.

The outline of this paper is as follows. The exact model will be introduced in Section 2, which we reduce to a family of problems on a thin and finite strip by (i) applying a Floquet-Bloch transform in the infinite direction and (ii) using a transparent boundary condition. In Section 3 we study the accuracy of a basis consisting of only a few Bloch modes multiplied with macroscopic polynomials in the crystal, both for frequencies inside and outside the bandgap. The multiscale FEM with Bloch modes and polynomials in macroscopic cells will be introduced in Section 4, including the coupling to an exterior discretisation (*e.g.*, high order p-FEM) in the homogeneous ambient outer space. In Sec. 5 a size-robust numerical integration procedure for such multiscale basis functions is presented. Finally, we show the accuracy of the method for several configurations of incident plane waves in Section 6 including convergence results for a *p*-refinement scheme. In all the experiments we computed the “exact” solution with a high order FE discretisation on a mesh resolving the fine scale of the crystal and high order transparent boundary conditions, *i.e.*, we compare to a fully size dependent method.

## 2. Model problem

### 2.1. Governing equations

The time-harmonic transverse magnetic (TM) and transverse electric (TE) fields in 2D PhC structures are described by (see, *e.g.*, [40])

$$-\nabla \cdot \nabla e(\mathbf{x}) = \left(\frac{\omega}{c_0}\right)^2 \varepsilon(\mathbf{x}) e(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^2, \quad (1\text{-TM})$$

$$-\nabla \cdot (\varepsilon^{-1}(\mathbf{x}) \nabla h(\mathbf{x})) = \left(\frac{\omega}{c_0}\right)^2 h(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^2, \quad (1\text{-TE})$$

where  $\omega \in \mathbb{R}$  is the frequency,  $e(\mathbf{x})$  and  $h(\mathbf{x})$  are the out-of-plane electric and magnetic fields,  $c_0$  is the vacuum speed of light, and  $\varepsilon(\mathbf{x})$  is the relative dielectricity. We assume  $\varepsilon(\mathbf{x})$  to be real, strictly positive and bounded, *i.e.*, there exists lower and upper bounds  $\varepsilon_{\min}, \varepsilon_{\max} \in \mathbb{R}_{\geq 0}$  such that  $0 < \varepsilon_{\min} \leq \varepsilon(\mathbf{x}) \leq \varepsilon_{\max} < \infty$  for almost all  $\mathbf{x} \in \mathbb{R}^2$ .

For notational convenience we collect the two scalar valued equations (1) in the generalised Helmholtz equation

$$\nabla \cdot (a(\mathbf{x}) \nabla u(\mathbf{x})) + b^2(\mathbf{x}, \omega) u(\mathbf{x}) = 0, \quad (2)$$

where the presumed properties of  $\varepsilon(\mathbf{x})$  transfer to the coefficient functions  $a(\mathbf{x})$  and  $b^2(\mathbf{x}, \omega)$ , *i.e.*,  $0 < a_{\min} \leq a(\mathbf{x}) \leq a_{\max} < \infty$  and  $0 < b_{\min}^2 \leq b^2(\mathbf{x}, \omega) \leq b_{\max}^2 < \infty$  for almost all  $\mathbf{x} \in \mathbb{R}^2$  and some  $a_{\min}, a_{\max}, b_{\min}^2, b_{\max}^2 \in \mathbb{R}_{\geq 0}$ . The coefficient functions are  $a(\mathbf{x}) = 1$ ,  $b^2(\mathbf{x}, \omega) = (\frac{\omega}{c_0})^2 \varepsilon(\mathbf{x})$  for the TM mode and  $a(\mathbf{x}) = \varepsilon^{-1}(\mathbf{x})$ ,  $b^2(\mathbf{x}, \omega) = (\frac{\omega}{c_0})^2$  for the TE mode.

To admit a unique solution the generalised Helmholtz equation has to be completed by appropriate radiation conditions (RC's for short) at infinity:

$$\text{RC}(u - u^{\text{inc}}) = 0, \quad (3)$$

where  $u^{\text{inc}}$  is the field incident from infinity, which might be a plane wave or a Gaussian beam. The radiation condition of Sommerfeld is adequate for finite scatterers, but not for scatterers of semi infinite extent, which are studied in the present work. We will detail the radiation condition RC (3) for such an infinite scatterer in Sec. 2.3.

The two equations (1) together with the radiation condition (3) not only define the out-of-plane fields, rather they define the magnetic and electric fields completely, since their in-plane components can be obtained using the equations

$$-\mu_0 \mathbf{h}(\mathbf{x}) = \nabla^\perp e(\mathbf{x}), \quad \varepsilon_0 \varepsilon(\mathbf{x}) \mathbf{e}(\mathbf{x}) = \nabla^\perp h(\mathbf{x}),$$

where  $\mu_0$  and  $\varepsilon_0$  are the vacuum permittivity and dielectricity constants and  $\mathbf{v}^\perp$  denotes the vector  $\mathbf{v}$  rotated by  $90^\circ$ .

### 2.2. Geometric setting

In the present work we will study the *infinite PhC barrier* which is a band made by photonic crystals  $\Omega_\infty^{\text{cr}}$  with infinite extent in one direction, which we fix to  $\mathbf{e}_1 = (1, 0)^\top$ , and finite extent in all other directions, especially in the direction  $\mathbf{e}_2 = (0, 1)^\top$  (see Fig. 1a). In the PhC crystal  $\Omega_\infty^{\text{cr}}$  the dielectricity is locally periodic, meaning that

$$\varepsilon(\mathbf{x} + \mathbf{a}_i) = \varepsilon(\mathbf{x}) \quad \text{for all } \mathbf{x} \in (\Omega_\infty^{\text{cr}} \cap (\Omega_\infty^{\text{cr}} - \mathbf{a}_i)) \text{ and } i = 1, 2, \quad (4)$$

where  $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^2$  are the *principal directions of periodicity* [40]. The crystal is globally periodic in direction  $\mathbf{a}_1 = |\mathbf{a}_1| \mathbf{e}_1$  but only locally periodic in direction  $\mathbf{a}_2$ . The fact that a periodicity direction  $\mathbf{a}_1$  exists which

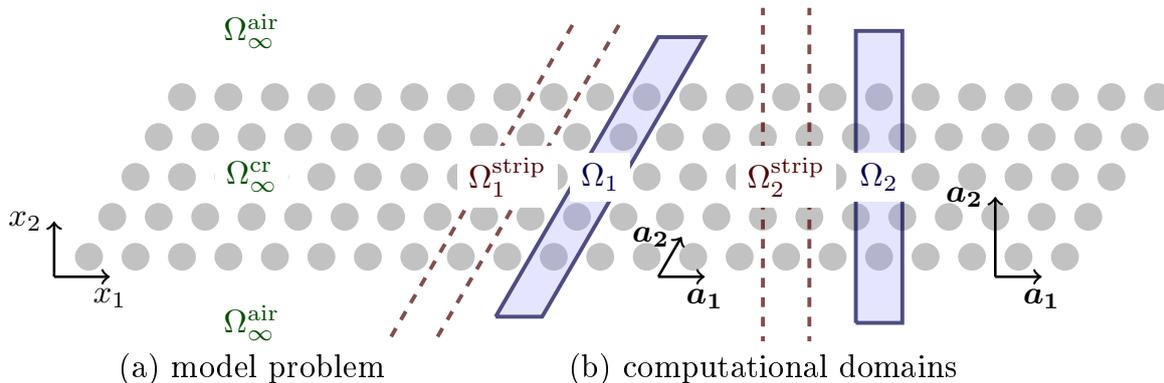


Figure 1: PhC crystal barrier model problem for a hexagonal PhC. (a) The crystal  $\Omega_\infty^{\text{cr}}$  has infinite extent in  $x_1$  direction and is embedded in homogeneous space  $\Omega_\infty^{\text{air}}$ . (b) Two possibilities  $\Omega_1$  and  $\Omega_2$  for computational domains are shown. They are defined by different directions of periodicity  $\mathbf{a}_1$  and  $\mathbf{a}_2$ . In an intermediate step the strips  $\Omega_1^{\text{strip}}$  and  $\Omega_2^{\text{strip}}$  are used.

coincides with the direction of the crystal's surface ensures that the whole geometry is periodic and that a global periodicity cell  $\Omega^{\text{strip}}$ —with width  $|\mathbf{a}_1|$  in the direction  $\mathbf{e}_1$ —exists (see Fig. 1b).

The parallelogram-shaped domain formed by the vectors  $\mathbf{a}_1$ ,  $\mathbf{a}_2$  is the *fundamental periodicity domain*

$$\widehat{\Omega} := \{\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 : \alpha_i \in (0, 1), i = 1, 2\},$$

which we will refer to as the *unit cell*. The choice of the periodicity directions  $\mathbf{a}_1$ ,  $\mathbf{a}_2$  for the unit cell  $\widehat{\Omega}$  is not unique, as shown in Fig. 1b for the example of a hexagonal lattice.

The crystal may be surrounded by a  $\mathbf{a}_1$ -periodic dielectric material on both sides, *e.g.*, a special surface structure or several homogeneous layers of constant thickness, and homogeneous material outside of the band  $-L^- \leq x_2 \leq L^+$  for sufficiently large  $L^-, L^+ > 0$ . If the surrounding dielectric is  $n\mathbf{a}_1$ -periodic for some  $n \in \mathbb{N}$  one may replace  $\mathbf{a}_1$  by  $n\mathbf{a}_1$  and continue with the scaled unit cell. The structured domain between the homogeneous exterior  $\Omega_\infty^{\text{air}}$  and the crystal  $\Omega_\infty^{\text{cr}}$  is denoted by  $\Omega_\infty^{\text{str}}$ , and it holds  $\overline{\Omega_\infty^{\text{cr}}} \cup \overline{\Omega_\infty^{\text{str}}} \cup \overline{\Omega_\infty^{\text{air}}} = \mathbb{R}^2$ . We call  $k(\mathbf{x}, \omega) := a^{-1}(\mathbf{x})\sqrt{b^2(\mathbf{x}, \omega)}$  the wavenumber which takes constant values  $k_+$  and  $k_-$  in the two half-planes  $x_2 > L^+$  and  $x_2 < -L^-$  of  $\Omega_\infty^{\text{air}}$ .

### 2.3. Radiation condition for the infinite scatterer

A radiation condition is used to ensure uniqueness, but it also limits the model to yield only physically meaningful solutions. An infinite scatterer, as studied here, is not a physical model. We consider the infinite scatterer as the limit problem  $n_1 \rightarrow \infty$  of a physical model of scattering on a finite crystal with an  $n_1 \times n_2$  array of scatterers embedded in free space. The radiation condition for the finite crystal is the Sommerfeld radiation condition. In both models the response to external excitations  $u^{\text{inc}}$  is searched, where the excitation may contain a geometric optics part, *i.e.*, plane waves. As is customary in scattering problems, the response is described by the total field  $u = u^{\text{sc}} + u^{\text{inc}}$ , the sum of the scattered field  $u^{\text{sc}}$  and the incident field  $u^{\text{inc}}$ . We assume the incident field to consist of a finite number of incoming plane waves and a wave optics part vanishing for  $|x_1| \rightarrow \infty$ .

After reflection and refraction by the infinite dielectric band  $\Omega_\infty^{\text{cr}} \cup \Omega_\infty^{\text{str}}$  the solution  $u$  may contain a finite number of out-going plane waves and a scattered geometric optics part  $u^{\text{go}}$ . In contrast, for the finite scatterer the scattered field  $u^{\text{sc}}$  does not contain a geometric optics part. Therefore, for dielectric scatters which extends at least in one direction towards infinity the Sommerfeld radiation condition in its integral form is only applied to the so called *diffracted field* (see [5, 50])

$$u^{\text{d}} := u^{\text{sc}} - u^{\text{go}}.$$

We apply this condition to the half-planes  $x_2 \geq L^+$ ,  $x_2 \leq -L^-$  of  $\Omega_\infty^{\text{air}}$  separately, *i.e.*, we require the fields to fulfill for  $\mathbf{x}(r, \varphi) = (r \cos(\varphi), \pm r \sin(\varphi) \pm L^\pm)^\top$  that

$$\lim_{R \rightarrow \infty} \int_0^\pi |\partial_r u^d(\mathbf{x}) - ik_\pm u^d(\mathbf{x})|^2 R d\varphi = 0, \quad (5)$$

where  $k_\pm$  are the wave-numbers in the two half-planes.

The geometric optics part of the incident field  $u^{\text{inc}}$ , consisting of a finite number of plane waves, leads to reflected and transmitted waves  $u^{\text{go}}$  on the two sides of the band and a wave inside the band propagating along the band which is incoming from one side and out-going to the other [28].

**Remark 1** (On guided modes). *We will show in Sec. 2.9, that with condition (5) the solution is uniquely defined except for a finite number of linearly independent solutions to the homogeneous Helmholtz equation (2). These solutions, the guided modes  $u_j^g$  (also called trapped modes), propagate along the band and decay in all other directions. The guided modes have been investigated for homogeneous dielectric bands [9] as well as PhC bands [67]. As we consider the band  $\Omega_\infty^{\text{cr}}$  as a limit problem of the finite crystal where guided modes do not exist, it is reasonable to exclude them for the limit problem. Hence, only the geometric optics part  $u^{\text{go}}$  determines the behaviour at infinity in all directions ( $|\mathbf{x}| \rightarrow \infty$ ).*

To summarise, the radiation condition RC consists of the Sommerfeld radiation condition in integral form (5) in the two half-planes outside the infinite PhC band applied to the diffracted field  $u^d$ . Moreover, RC excludes a finite number of guided modes  $u_j^g$ . With this condition the solution is now well-defined, see Sec. 2.9. The handling of the scattered geometric optics part  $u^{\text{go}}$  for an incident plane wave will be explained in Sec. 2.7.

#### 2.4. The strong formulation

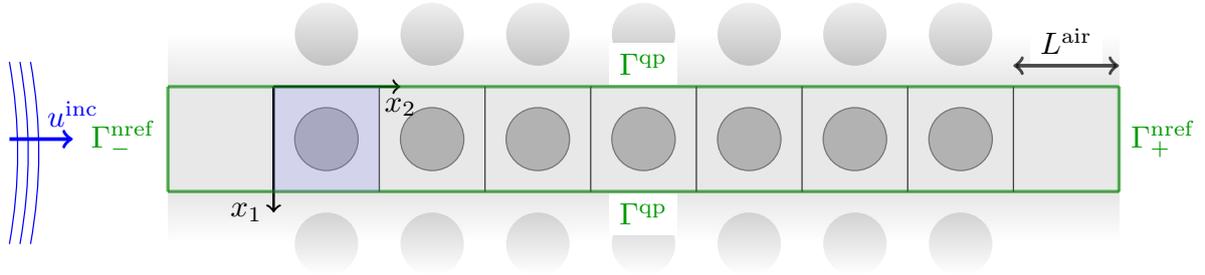


Figure 2: Infinite photonic barrier model problem. An incident field  $u^{\text{inc}}$  in  $x_2$  directions induces an electromagnetic field inside a PhC, here with  $n_p = 7$  periods. This can be seen as an example of a finite crystal with dielectric rods (dark shaded), *e.g.*,  $\varepsilon_r = 8.9$  (aluminium oxide) and radius  $r = 0.3$ , inside air (light shaded) with  $\varepsilon_r = 1$ . The blue highlighted area is the unit cell  $\Omega$ .

The whole space,  $\mathbb{R}^2$ , is the natural domain for the wave equation (2), but needs to be reduced to a finite computational domain  $\Omega$  for simulations. Due to the global periodicity in the direction  $\mathbf{e}_1$  we can apply the *Floquet transform*  $\mathcal{F}$  [47] to the solution  $u(\mathbf{x})$ . The transform introduces a *Floquet parameter*  $k_1 \in [-\pi/|\mathbf{a}_1|, \pi/|\mathbf{a}_1|] =: \mathcal{B}_1$  in the *one-dimensional Brillouin zone*  $\mathcal{B}_1$ . For every Floquet parameter  $k_1$ , the transformed functions  $u_{k_1}(\cdot) := (\mathcal{F}u)(k_1, \cdot)$  are quasi-periodic solutions to problems on a reduced domain with data  $u_{k_1}^{\text{inc}}(\cdot) := (\mathcal{F}u^{\text{inc}})(k_1, \cdot)$ . This reduced domain is the strip  $\Omega^{\text{strip}}$  (see Fig. 1b) which we have defined in Sec. 2.2. The real solution  $u$  can then be recovered by the inverse Floquet transformation which requires all the solutions  $u_{k_1}$  for  $k_1 \in \mathcal{B}_1$ . In a second step the problems on the strip  $\Omega^{\text{strip}}$  are reduced to problems on a finite computational domain  $\Omega$ , a parallelogram with sides parallel to  $\mathbf{a}_1$  and  $\mathbf{a}_2$ . The boundary  $\partial\Omega$  of  $\Omega$  is made of the parts  $\Gamma^{\text{qP}}$  and  $\Gamma^{\text{nref}} := \Gamma_-^{\text{nref}} \cup \Gamma_+^{\text{nref}}$ , where  $\Gamma^{\text{qP}}$  is parallel to  $\mathbf{a}_2$  and  $\Gamma^{\text{nref}}$  to  $\mathbf{a}_1$ , see Fig. 2. The computational domain  $\Omega$  contains (i) a strip of  $n_p$  crystal cells in a domain

$\Omega^{\text{cr}} = \Omega_{\infty}^{\text{cr}} \cap \Omega$ , (ii) possibly two surrounding sub-domains of dielectric material  $\Omega^{\text{str}} = \Omega_{\infty}^{\text{str}} \cap \Omega$  and (iii) an exterior homogeneous layer  $\Omega^{\text{air}} = \Omega_{\infty}^{\text{air}} \cap \Omega$ , that extends into  $\Omega_{\infty}^{\text{air}}$  with only a finite depth  $L^{\text{air}} \in \mathbb{R}^+$ . The solution outside the computational domain  $\Omega$  is represented by Dirichlet-to-Neumann (DtN) operators  $M_{k_1}$  acting on  $\Gamma^{\text{ref}}$ .

For each given *Floquet parameter*  $k_1 \in \mathcal{B}_1$ , we are searching for solutions  $u_{k_1}$  that satisfy (see Fig. 2)

$$\nabla \cdot (a(\mathbf{x}) \nabla u_{k_1}(\mathbf{x})) + b^2(\mathbf{x}, \omega) u_{k_1}(\mathbf{x}) = 0 \quad \text{for all } \mathbf{x} \text{ in } \Omega, \quad (6a)$$

$$u_{k_1}(\mathbf{x} + \mathbf{a}_1) = e^{ik_1|\mathbf{a}_1|} u_{k_1}(\mathbf{x}) \quad \text{for all } \mathbf{x} \text{ on } \Gamma^{\text{qp}}, \quad (6b)$$

$$(\partial_{\mathbf{n}} - M_{k_1})u_{k_1}(\mathbf{x}) = (\partial_{\mathbf{n}} - M_{k_1})u_{k_1}^{\text{inc}}(\mathbf{x}) \quad \text{for all } \mathbf{x} \text{ on } \Gamma^{\text{ref}}, \quad (6c)$$

where  $\partial_{\mathbf{n}} = \mathbf{n} \cdot \nabla$  and  $\mathbf{n}$  the outward pointing normal on a point  $\mathbf{x} \in \Gamma^{\text{ref}}$ .

Details about the Floquet transform and quasi-periodicity condition (6b) are given in Sec. 2.6, while the DtN operators  $M_{k_1}$  are described in Sec. 2.8. In Section 2.7 we will discuss incident plane waves and it will be explained how to treat the Floquet parameters  $k_1$  for this incident field.

### 2.5. The variational formulation

In the following, we discretise the model (6) with the Galerkin method. To this end, the strong formulation will be replaced by a variational one. The associated Sobolev space  $H_{k_1}^1(\Omega)$  on the bounded computational domain  $\Omega \subset \mathbb{R}^2$  is given by

$$H_{k_1}^1(\Omega) := \{u_{k_1} \in H^1(\Omega) : u_{k_1}(\mathbf{x} + \mathbf{a}_1) = e^{ik_1|\mathbf{a}_1|} u_{k_1}(\mathbf{x}) \quad \forall \mathbf{x} \in \Gamma^{\text{qp}}\}.$$

Then the variational formulation of (6) reads:

$$\text{Find } u_{k_1} \in H_{k_1}^1(\Omega), \text{ such that} \quad \Phi_{k_1}(u_{k_1}, v) = f(v) \quad \text{for all } v \in H_{k_1}^1(\Omega), \quad (7a)$$

with the sesquilinear form  $\Phi_{k_1}(u, v)$  and the (anti-)linear form  $f(v)$  defined by

$$\begin{aligned} \Phi_{k_1}(u, v) := & \int_{\Omega} a(\mathbf{x}) \nabla u(\mathbf{x}) \cdot \nabla \bar{v}(\mathbf{x}) \, d\mathbf{x} - \int_{\Omega} b^2(\mathbf{x}, \omega) u(\mathbf{x}) \bar{v}(\mathbf{x}) \, d\mathbf{x} \\ & + \int_{\Gamma^{\text{ref}}} M_{k_1} u(\mathbf{x}) \bar{v}(\mathbf{x}) \, dS, \end{aligned} \quad (7b)$$

$$f(v) := \int_{\Gamma^{\text{ref}}} (\partial_{\mathbf{n}} + M_{k_1}) u^{\text{inc}}(\mathbf{x}) \bar{v}(\mathbf{x}) \, dS. \quad (7c)$$

### 2.6. Formulation on the infinite strip

The problem on  $\mathbb{R}^2$  is reduced to a family of problems on  $\Omega^{\text{strip}}$  using the 1-dimensional Floquet transform and its inverse [47, 48, 44], as defined by

$$u_{k_1}(\mathbf{x}) = (\mathcal{F}u)(k_1, \mathbf{x}) = \frac{1}{|\mathcal{B}_1|} \sum_{m \in \mathbb{Z}} u(x_1 - m|\mathbf{a}_1|, x_2) e^{ik_1 m |\mathbf{a}_1|}, \quad (8a)$$

$$u(\mathbf{x}) = (\mathcal{F}^{-1}u_{k_1})(\mathbf{x}) = \int_{\mathcal{B}_1} u_{k_1}(\mathbf{x}) \, dk_1, \quad (8b)$$

where we consider  $u_{k_1}(\mathbf{x})$  as a function in  $k_1$  and  $\mathbf{x}$  which is  $\mathbf{a}_1$ -quasi-periodic in  $\mathbf{x}$ , *i.e.*,  $u_{k_1}(x_1, x_2) = e^{ik_1|\mathbf{a}_1|} u_{k_1}(x_1 + |\mathbf{a}_1|, x_2)$ .

The Floquet transform  $\mathcal{F}$  is applied to (2) which decouples the equation and results in (6a), a family of problems with Floquet parameter  $k_1$  posed only on  $\Omega^{\text{strip}}$ . The quasi-periodicity of  $u_{k_1}$  transforms into the boundary condition (6b).

For general incident fields like a Gaussian beam,  $u_{k_1}$  is non-zero for all  $k_1 \in \mathcal{B}$ . However, the parameter  $k_1$  can be sampled and the integral in (8b) efficiently approximated with the composite trapezoidal rule. For incident plane waves only one  $k_1$  contributes which will be seen in the following section.

### 2.7. Scattering of plane waves

Even though plane waves carry infinite energy and are not physical, the scattering of plane waves is commonly studied, as the field emitted from any localised electromagnetic source which is far away from the scatterer locally approaches that of a plane wave. An incident plane wave with wave vector  $\mathbf{k}^{\text{inc}} \in \mathbb{R}^2$  and amplitude 1 is given by

$$u^{\text{inc}}(\mathbf{x}) = e^{i\mathbf{k}^{\text{inc}} \cdot \mathbf{x}}. \quad (9)$$

The incident field satisfies the Helmholtz equation in the homogeneous space, *i.e.*, it holds that  $|\mathbf{k}^{\text{inc}}|^2 = k_-^2$  (or  $k_+^2$ ). We define the Floquet transform of  $u^{\text{inc}}$  as

$$u_{k_1}^{\text{inc}}(\mathbf{x}) = (\mathcal{F}u^{\text{inc}})(k_1, \cdot) := \delta(Tk_1^{\text{inc}} - k_1)u^{\text{inc}}(\mathbf{x}), \quad (10)$$

with Dirac's  $\delta$ -distribution and  $T$  being the projection into the Brillouin zone  $\mathcal{B}_1 = [-\pi/|\mathbf{a}_1|, \pi/|\mathbf{a}_1|]$  such that  $Tk_1^{\text{inc}} \in \mathcal{B}_1$  and  $k_1^{\text{inc}} - Tk_1^{\text{inc}} = j2\pi/|\mathbf{a}_1|$  for some  $j \in \mathbb{Z}$ . It can be easily seen that the inverse Floquet transform (8b) of  $u_{k_1}^{\text{inc}}$  recovers  $u^{\text{inc}}$ . This special definition for plane waves is necessary since the formal Floquet transform (8a) does not determine  $u_{k_1}^{\text{inc}}(\cdot)$  for  $Tk_1^{\text{inc}} \neq k_1$  (does not converge). However, the quasi-periodicity of the Floquet transform and its formal commutativity with the differential operators leading to (6a) and (6b) are recovered with the definition (10).

Due to the  $\delta$ -distribution in (10) we will compute the solution  $u_{k_1}(\mathbf{x})$  of (6) only for the single Floquet parameter  $k_1 = Tk_1^{\text{inc}}$ . However, note that  $u_{k_1}$  has an infinite amplitude, and we will therefore not compute  $u_{k_1}(\mathbf{x})$  itself, but rather its shape  $u_{k_1, S}$  where  $u_{k_1} = u_{k_1, S}\delta(Tk_1^{\text{inc}} - k_1)$ . Applying the inverse transform (8b) is then equivalent to extending the shape  $u_{k_1, S}$  quasi-periodically in  $\mathbf{a}_1$ .

### 2.8. The DtN operator

The strip  $\Omega^{\text{strip}}$  still has infinite extent in  $\mathbf{a}_2$ -direction (cf. Fig. 1). However, most of the space is occupied by homogeneous material, namely by two semi-infinite sub-domains in  $\Omega^{\text{air}}$ . The solutions of (2) in a homogeneous medium are well known to be superpositions of plane waves [27].

This allows to restrict (2) to a finite computational domain  $\Omega$  with the additional boundary  $\Gamma^{\text{href}}$  and an approximated *transparent boundary condition* on  $\Gamma^{\text{href}}$ . We refer to [26] for a survey of transparent boundary conditions. If this boundary condition can be constructed so that the solution for the problem posed on  $\Omega$  is identical to the solution of the Floquet transformed Helmholtz equation (2) in the infinite strip  $\Omega^{\text{strip}}$  (restricted to  $\Omega$ ), then it is called *exact*. A transparent boundary condition that is not exact introduces an error which is inherent to the model (6). In this section we introduce an exact transmission condition by means of DtN operators  $M_{\text{ex}, k_1}^{\pm}$  which have been derived by an expansion in plane waves in the homogeneous space  $\Omega_{\infty}^{\text{air}}$ . The DtN operator  $M_{\text{ex}, k_1}$  for the exact transmission condition can be approximated to any accuracy by the truncated DtN operator  $M_{k_1}$  [61].

The two boundaries  $\Gamma_+^{\text{href}}$  and  $\Gamma_-^{\text{href}}$  are the lines  $[x_1^+, x_1^+ + |\mathbf{a}_1|] \times \{x_2^+\}$  and  $[x_1^-, x_1^- + |\mathbf{a}_1|] \times \{x_2^-\}$ , both parallel to the  $x_1$ -axis. For  $\Gamma_+^{\text{href}}$  and  $\Gamma_-^{\text{href}}$  we have the DtN operators  $M_{k_1}^+ = M_{k_1}|_{\Gamma_+^{\text{href}}}$  and  $M_{k_1}^- = M_{k_1}|_{\Gamma_-^{\text{href}}}$ , respectively, which will be an approximation to the exact DtN maps  $M_{\text{ex}, k_1}^{\pm}$  given by

$$\begin{aligned} M_{\text{ex}, k_1}^{\pm} u_{k_1}^{\text{sc}}(x_1, x_2^{\pm}) &= -\nabla u_{k_1}^{\text{sc}}(x_1, x_2^{\pm}) \cdot \mathbf{n} \\ &= -\frac{1}{|\mathbf{a}_1|} \sum_{j \in \mathbb{Z}} iK_2^{\pm, (j)} \int_{x_1^{\pm}}^{x_1^{\pm} + |\mathbf{a}_1|} u_{k_1}^{\text{sc}}(x'_1, x_2^{\pm}) e^{ij\frac{2\pi}{|\mathbf{a}_1|}(x_1 - x'_1)} e^{ik_1(x_1 - x'_1)} dx'_1, \end{aligned} \quad (11)$$

with  $K_2^{\pm, (j)} := \sqrt[+]{(k_{\pm})^2 - (k_1 + 2\pi j/|\mathbf{a}_1|)^2}$  depending on the wave-numbers  $k_-$  and  $k_+$  of the homogeneous regions (cf. Sec. 2.2) and where we use the positive square root  $\sqrt[+]{r} := \sqrt{r}$  and  $\sqrt[-]{r} := i\sqrt{r}$  for any  $r \in \mathbb{R}_{\geq 0}$ . The positive square root selects only out-going or decaying waves. Real  $K_2^{\pm, (j)}$  correspond to propagating plane waves which are out-going and purely imaginary  $K_2^{\pm, (j)}$  to evanescent plane waves which

decay exponentially for  $x_2 \rightarrow \pm\infty$ . Hence, the exact transparent boundary condition given by (11) is in accordance with the radiation condition defined in Sec. 2.3.

As we assume the boundary  $\Gamma^{\text{nref}}$  to be located in the homogeneous domain  $\Omega^{\text{air}}$ , the normal derivative of  $u$  and therefore  $u_{k_1}^{\text{sc}}$  are continuous. Hence, (11) holds for the interior or exterior normal derivative on  $\Gamma^{\text{nref}}$ .

The evanescent plane waves decay the faster the larger their values of  $j$  and it is possible to get accurate approximations to  $M_{\text{ex},k_1}^{\pm}$  by truncation, *i.e.*, taking only the finite sum with  $|j| \leq n^{\text{DtN}}$  in (11) for some  $n^{\text{DtN}} \in \mathbb{N}$ . The well-posedness of the variational formulation may be lost in a direct truncation, *i.e.*, artificial modes belonging to higher Fourier coefficients may appear. We use a *modified DtN condition* [29], which preserves well-posedness of the original problem by replacing the coefficients  $K_2^{\pm,(j)}$  for  $|j| > n^{\text{DtN}}$  in (11) by the  $j$ -independent values  $k_{\pm} \in \mathbb{R}^+$ . With the abbreviation  $(\mathcal{G}_j u_{k_1}^{\text{sc}})(x_1, x_2) := \frac{1}{|\mathbf{a}_1|} \int_{x_1^{\pm}}^{x_1^{\pm} + |\mathbf{a}_1|} u_{k_1}^{\text{sc}}(x'_1, x_2) e^{ij \frac{2\pi}{|\mathbf{a}_1|} (x_1 - x'_1)} e^{ik_1(x_1 - x'_1)} dx'_1$  we define the approximate DtN maps by

$$\begin{aligned} M_{k_1}^{\pm} u_{k_1}^{\text{sc}}(x_1, x_2^{\pm}) &= - \sum_{|j| \leq n^{\text{DtN}}} iK_2^{\pm,(j)} (\mathcal{G}_j u_{k_1}^{\text{sc}})(x_1, x_2^{\pm}) - \sum_{|j| > n^{\text{DtN}}} ik_{\pm} (\mathcal{G}_j u_{k_1}^{\text{sc}})(x_1, x_2^{\pm}) \\ &= - \sum_{|j| \leq n^{\text{DtN}}} i(K_2^{\pm,(j)} - k_{\pm}) (\mathcal{G}_j u_{k_1}^{\text{sc}})(x_1, x_2^{\pm}) - ik_{\pm} u_{k_1}^{\text{sc}}(x_1, x_2^{\pm}), \end{aligned} \quad (12)$$

where we used the definition of the Fourier expansion and its inverse. For sufficiently large  $n^{\text{DtN}}$  the induced modelling error decays exponentially in  $n^{\text{DtN}}$  and  $L^{\text{air}}$  [61], and already moderate values of  $L^{\text{air}}$  and  $n^{\text{DtN}}$  yield a negligible modelling error.

Equation (6) is posed in terms of the total field  $u_{k_1}$ , but  $M_{k_1}$  acts on the scattered part  $u_{k_1}^{\text{sc}}$ . Using the identity  $u_{k_1} = u_{k_1}^{\text{sc}} + u_{k_1}^{\text{inc}}$  the boundary condition (6c) is obtained exactly. The solution  $u_{k_1}$  on  $\Omega^{\text{strip}}$  (see Sec. 2.6) is identical to the solution  $u_{k_1}$  of (6) on  $\Omega$  if the exact DtN map  $M_{\text{ex},k_1}$  is used. But we use the truncated DtN map  $M_{k_1}$  in (6) and thus a modelling error occurs. The overall error of our method is the sum of this modelling error and the discretisation error (see below). However, this article focuses on an efficient discretisation for  $\Omega^{\text{cr}}$  and not on transparent boundary conditions, so we fix some large values for  $L^{\text{air}}$  and  $n^{\text{DtN}}$  and do not distinguish between the approximated  $u_{k_1}$  and the exact  $u_{k_1}$ , both are simply denoted by  $u_{k_1}$ .

As a final remark on computational aspects, inserting  $M_{k_1}$  (see (11)) into (7) reveals that certain double integrals need to be evaluated numerically. These integrals are separable and can be rewritten as sums  $\sum_{|j| \leq n^{\text{DtN}}}$  of products of one dimensional integrals in  $x_1$  and  $x'_1$  respectively.

### 2.9. Well-posedness of the variational formulation

In this section we will explain why the solution  $u_{k_1}(\mathbf{x})$  of (7) is well-defined for a.e.  $k_1$ . For a finite number of Floquet parameters  $k_1$  the formulation has a non-trivial null space spanned by the guided modes  $u_j^g$ . As they do not contribute to the solution (see Rem. 1) we could exclude them by Lagrange multipliers or by using linear system solvers ignoring the null space (*e.g.*, Krylov subspace solvers). In the present work we will assume for simplicity that (7) is not simulated for these values of  $k_1$ , *i.e.*, we assume that the kernel  $\ker(\Phi_{k_1})$  is trivial.

The procedure to show uniqueness for the PhC band on the bounded domain  $\Omega$  with quasi-periodic boundary conditions in  $\mathbf{a}_1$ -direction and a truncated DtN operator  $M_{k_1}$  in  $\mathbf{a}_2$ -direction is very similar to uniqueness proofs for the wave equation in bounded domains, *e.g.*, in [35], except that the guided modes need to be excluded. We will therefore strongly follow the discussion in [35, Sec. 2.4.3].

**Lemma 2.** *The sesquilinear form  $\Phi_{k_1}(\cdot, \cdot)$  is  $H_{k_1}^1(\Omega)$ -coercive.*

*Proof.* First we note that  $H_{k_1}^1(\Omega) \subset L^2(\Omega) \subset H_{k_1}^{-1}(\Omega)$  is a Gelfand triple (see, *e.g.*, [31]), where  $H_{k_1}^{-1}(\Omega)$  is the dual space of  $H_{k_1}^1(\Omega)$ . For  $H_{k_1}^1(\Omega)$ -coercivity we have to show that some positive constants  $C, \alpha > 0$  exist so that for every  $u \in H_{k_1}^1(\Omega)$  a Gårding inequality

$$|\Phi_{k_1}(u, u) + C\|u\|_{L^2(\Omega)}^2| \geq \alpha\|u\|_{H^1(\Omega)}^2, \quad (13)$$

holds. In particular it is sufficient to show that  $\operatorname{Re}(\Phi_{k_1}(u, u)) + C\|u\|_{L^2(\Omega)}^2 \geq \alpha\|u\|_{H_{k_1}^1(\Omega)}^2$ . First we note, that

$$\operatorname{Re} \int_{\Gamma^{\text{href}}} M_{k_1} u(\mathbf{x}) \bar{u}(\mathbf{x}) \, dS \geq 0$$

holds, which can be shown using the Fourier series expansion of the trace of  $u \in H_{k_1}^1(\Omega)$  on  $\Gamma^{\text{href}}$  and the fact that factors  $K_2^{\pm, (j)}$  in (11) are purely imaginary or real and positive. The remaining part of  $\Phi_{k_1}$  is  $H_{k_1}^1(\Omega)$ -coercive with  $C = b_{\max}^2 + a_{\min}$  and  $\alpha = a_{\min}$ , and (13) is proved.  $\square$

**Proposition 3** (Well-posedness of (7)). *or all Floquet parameter  $k_1 \in \mathbb{R}$  such that  $\ker \Phi_{k_1} = \emptyset$ , then the variational formulation (7) admits a unique solution  $u_{k_1} \in H_{k_1}^1(\Omega)$  for every right hand side  $f \in H_{k_1}^{-1}(\Omega)$ .*

*Proof.* As the domain  $\Omega$  is bounded and its boundary  $\partial\Omega$  is Lipschitz continuous the embedding  $H_{k_1}^1(\Omega) \subset L_{k_1}^2(\Omega)$  is compact. Together with the just proven  $H_{k_1}^1(\Omega)$ -coercivity the Fredholm alternative applies. As we assumed  $\ker \Phi_{k_1} = \emptyset$  the Fredholm alternative gives the desired uniqueness of  $u_{k_1}$ .  $\square$

In particular  $f \in H_{k_1}^{-1}(\Omega)$  holds if  $u^{\text{inc}} \in H_{\text{loc}}^1(\mathbb{R}^2)$ , *i.e.*, incident plane waves (cf. Sec. 2.7) are a valid choice for  $u^{\text{inc}}$ .

### 3. Bloch modes and best-approximation of finite crystals solutions

In this section several approximations of solutions  $u$  for the finite crystals (7) will be investigated; first, using the eigenmodes in the corresponding infinite crystal, the Bloch modes, and second using the Bloch modes multiplied with macro-polynomials, the functions which are potentially Floquet modes [47].

Infinite crystals are fully periodic crystals, *i.e.*,  $\Omega^{\text{cr}} = \mathbb{R}^2$  in (4), and the 2-dimensional Floquet transform, a generalisation of (8), can be applied in two dimensions, see *e.g.*, [68, 20]. For practical reasons, which will be explained later, we will use the Bloch transform instead. The Bloch transform is the Floquet transform multiplied by  $e^{-i\mathbf{k}\cdot\mathbf{x}}$ . It transforms (2) into a family of problems posed only on  $\widehat{\Omega} \subset \mathbb{R}^2$ , to which we shall refer to as *unit cell problems*. It reads as follows: Find all triples  $(\tilde{u}_{\mathbf{k}} \in H^1(\widehat{\Omega})/\{0\}, \mathbf{k} \in \mathcal{B} + i\mathbb{R}^2 \subset \mathbb{C}^2, \omega \in \mathbb{R}_{\geq 0})$  so that

$$(\nabla + i\mathbf{k})^\top (a(\mathbf{x})(\nabla + i\mathbf{k})) \tilde{u}_{\mathbf{k}}(\mathbf{x}) + b^2(\mathbf{x}, \omega) \tilde{u}_{\mathbf{k}}(\mathbf{x}) = 0. \quad \text{for all } \mathbf{x} \text{ in } \widehat{\Omega}, \quad (14a)$$

$$\tilde{u}_{\mathbf{k}}(\mathbf{x} \pm \mathbf{a}_j) = \tilde{u}_{\mathbf{k}}(\mathbf{x}) \quad \text{for all } j = 1, 2 \text{ and } \mathbf{x} \text{ on } \partial\widehat{\Omega}, \quad (14b)$$

where the frequency  $\omega$  is, by definition in (2), part of  $b^2(\mathbf{x}, \omega)$  and  $\mathcal{B}$  is the two-dimensional *Brillouin zone* [47]. The Brillouin zone arises from the  $2\pi$ -periodicity of  $e^{i\mathbf{x}}$ . For  $\widehat{\Omega} = [0, a]^2$  we have  $\mathcal{B} = [-\pi/a, \pi/a]^2$ . The solutions  $\tilde{u}_{\mathbf{k}}(\mathbf{x})$  of the unit cell problem are periodically extended to the whole space  $\mathbb{R}^2$ . The functions  $\tilde{u}_{\mathbf{k}}(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}$  are then called *Bloch modes*.

The combinations  $(\mathbf{k}, \omega)$  for which a non-vanishing solution  $\tilde{u}_{\mathbf{k}}$  exists with real wave vectors  $\mathbf{k} = (0, k) \subset \mathbb{R}^2$  are shown in Fig. 3(a), where we used the notion of the *scale independent frequency*  $\hat{\omega} := \frac{\omega a}{2\pi c_0}$ . Such a plot is called a *band diagram*. Bloch modes with real wave vectors  $\mathbf{k} \in \mathbb{R}^2$  are *propagating* modes which are bounded and carry finite energy. Bloch modes with complex wave vectors  $\mathbf{k} \in \mathbb{C}^2 \setminus \mathbb{R}^2$  are called *evanescent* modes. They are unbounded and carry infinite energy in  $\mathbb{R}^2$ . The evanescent modes are shown in Fig. 3(b). Both figures Figs. 3(a) and 3(b) have to be imagined mirror symmetric with respect to the  $\hat{\omega}$  axis. Additionally the evanescent modes in Fig. 3(b) occur in quadruples with the wave vectors  $\mathbf{k}, \bar{\mathbf{k}}, -\mathbf{k}, -\bar{\mathbf{k}}$  (see, *e.g.*, [20]). The real part  $\operatorname{Re} \mathbf{k}$  is furthermore  $2\pi/|\mathbf{a}_2|$ -periodic.

Photonic crystals guide the light differently depending on the frequency  $\omega$ . For certain frequencies no Bloch modes with real wave vectors exist. These frequencies are called *bandgap* frequencies and intervals of bandgap frequencies are called the *bandgaps*. At such frequencies, light can not propagate in the crystal and is reflected out. Inside the crystal the fields decay from the both surfaces of the crystal  $\Omega^{\text{cr}}$  and the

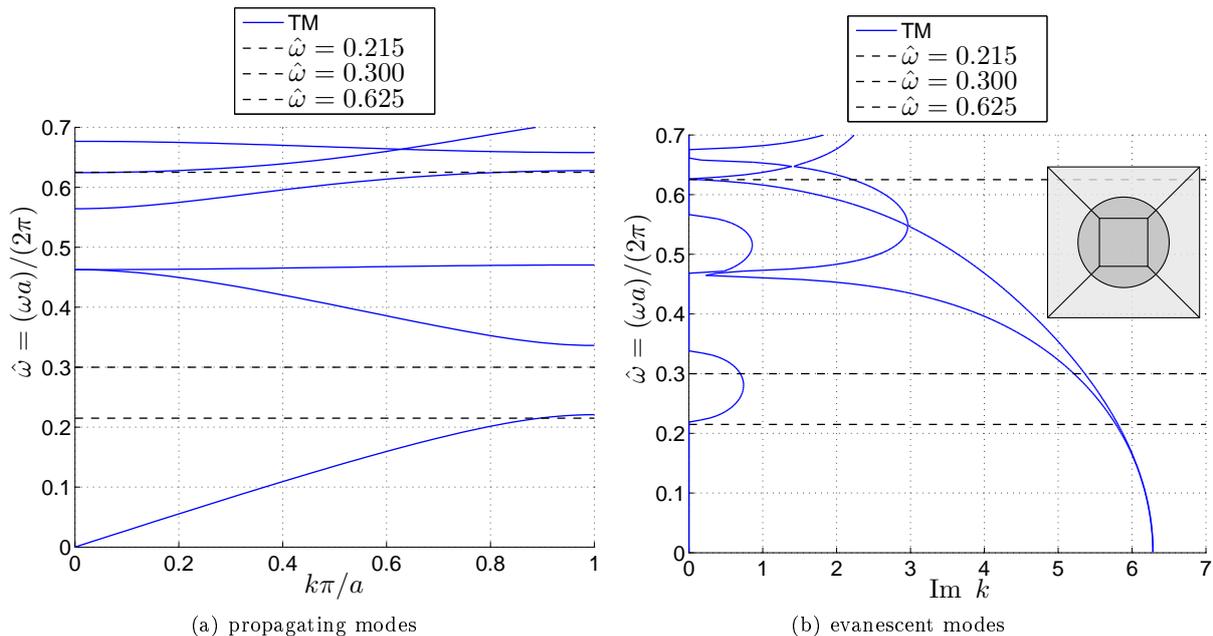


Figure 3: Band diagram for waves in  $x_2$ -direction ( $\mathbf{k} = (0, k)^\top$ ,  $a = |\mathbf{a}_2|$ ). (a) Shows pairs  $(\hat{\omega}, \mathbf{k})$  for which propagating Bloch modes exist. (b) The imaginary part  $\text{Im } k$  shows how fast non-propagating Bloch modes decay.

decay constant is given by  $\pm \text{Im } k$ , see Fig. 3(b). The bandgaps are numbered from the lowest frequency to the highest. For example  $\hat{\omega} = 0.3$  (see Fig. 3(a)) is a bandgap frequency in the first bandgap.

The frequencies for which Bloch modes with real wave vectors exist are called *propagating* frequencies and the corresponding intervals are the *bands* of the crystal. For example  $\hat{\omega} = 0.215$  is a propagating frequency where two Bloch modes with real wave vectors exist, they have  $k/a = \pm 2.81$ . The second solution is due to the symmetry of the Helmholtz operator and is always the complex conjugate of the first solution. Another example is the frequency  $\hat{\omega} = 0.625$  where four Bloch modes with real wave vectors exist, see Figs. 4(a) and 4(b). These three frequencies describe different phenomena and will serve as examples to illustrate the size-robustness of our MSFEM in all following sections.

An immediate consequence of the inverse Floquet-Bloch transform is that bounded solutions  $u$  of (2) for infinite crystals are superpositions of the solutions  $\tilde{u}_{\mathbf{k}}$  of the unit cell problem (14)

$$u(\mathbf{x}) = \int_{\mathcal{B}} \tilde{u}_{\mathbf{k}}(\mathbf{x}) e^{i\mathbf{k} \cdot \mathbf{x}} d\mathbf{k}, \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^2. \quad (15)$$

In this representation formula, the integration is only over propagating Bloch modes, as evanescent modes are unbounded and irrelevant for physical solutions on infinite crystals.

At a few frequencies, at which the bands possess a stationary point, two or more Bloch modes degenerate. The resulting modes, which take the form of polynomials multiplied with Bloch modes, are called Floquet modes [47] and have a vanishing group velocity [77, Proposition 6.4].

In our development of the generalized *hp*-FEM as a tensorization of hierarchic macroscopic and microscopic FE spaces a hierarchical ordering of Bloch modes will be crucial. Therefore, we adopt the following ordering principle.

### 3.1. Ordering of Bloch modes

We define the set of all Bloch modes as

$$\mathcal{B}_{k_1} := \{ \tilde{u}_{\mathbf{k}}(\mathbf{x}) e^{i\mathbf{k} \cdot \mathbf{x}} : (\tilde{u}_{\mathbf{k}}, \mathbf{k}) \text{ solution of (14) with } \text{Re } k_2 \in (-\pi, \pi] \},$$

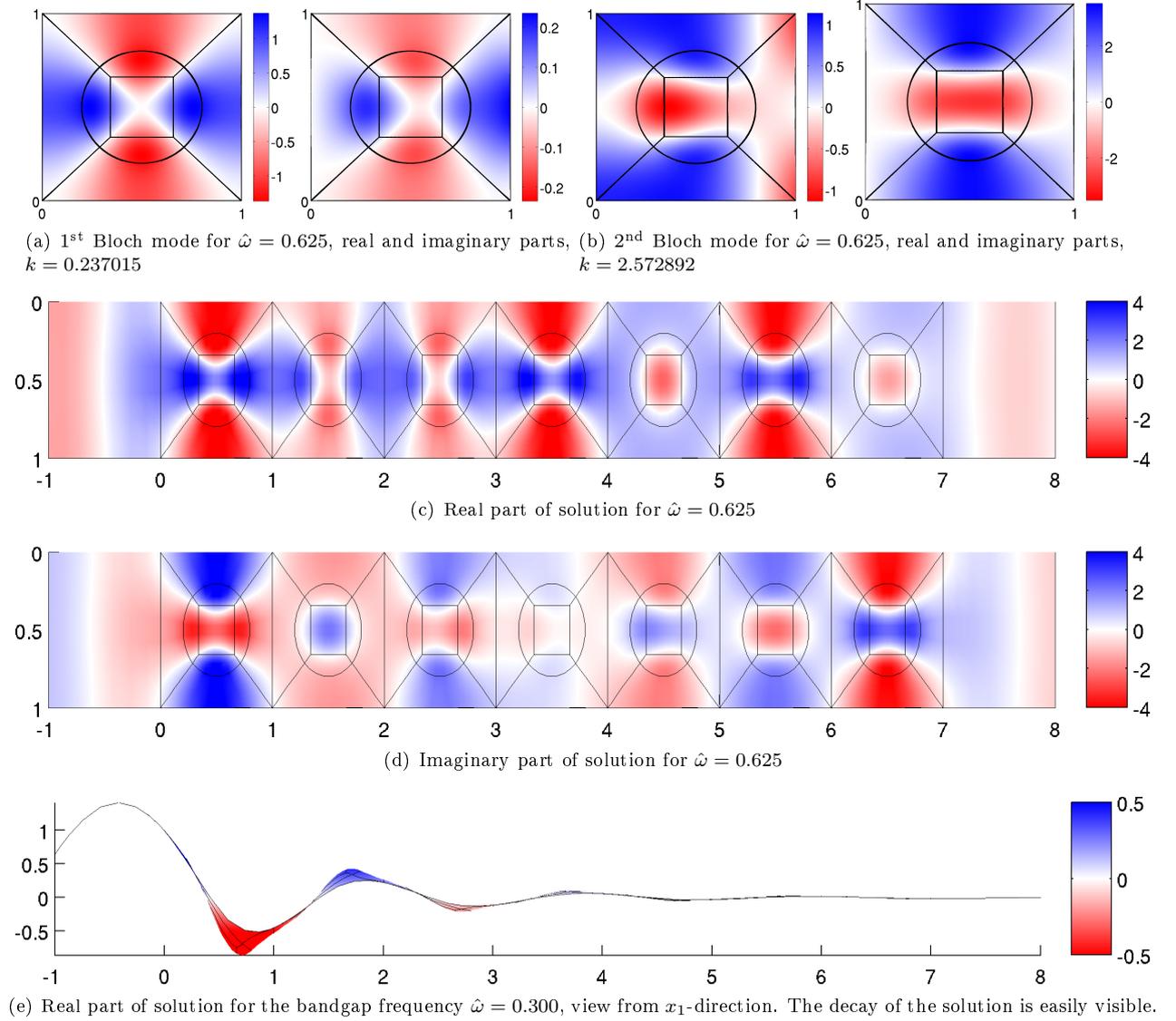


Figure 4: Field plots for solutions of the finite crystal with  $n_p = 7$  in comparison to the Bloch modes of the infinite crystal. There are four propagating Bloch modes for  $\hat{\omega} = 0.625$  the two functions of (a) and (b) and their complex conjugates with the wave vectors  $-k$ .

for a fixed Floquet parameter  $k_1$  and a fixed frequency  $\omega$ . We equip  $\mathcal{B}_{k_1}$  with an ordering of the elements  $\tilde{u}_k^{(i)}(\mathbf{x})e^{i\mathbf{k}^{(i)}\mathbf{x}} \in \mathcal{B}_{k_1}$  by their degree of exponential growth or decay, *i.e.*,  $|\text{Im } k_2^{(1)}| \geq |\text{Im } k_2^{(2)}| \geq \dots$ . The ordering of modes with equal  $|\text{Im } k_2|$  is arbitrary. Moreover, we define  $\mathcal{B}_{k_1}(n_{\text{bloch}})$  to be the subspace of  $\mathcal{B}_{k_1}$  containing the first  $n_{\text{bloch}}$  modes of  $\mathcal{B}_{k_1}$  according to this order. As we will practically use numerical approximations to (14) we abuse notation slightly and denote by  $\mathcal{B}_{k_1}$  also sets of approximative Bloch modes where their accuracy will be clear from the context. We partition  $\mathcal{B}_{k_1}$  into the set of propagating Bloch modes

$$\mathcal{B}_{k_1}^{\text{prop}} := \{ \tilde{u}_k(\mathbf{x})e^{i\mathbf{k}\mathbf{x}} \in \mathcal{B}_{k_1} : \text{Im } k_2 = 0 \},$$

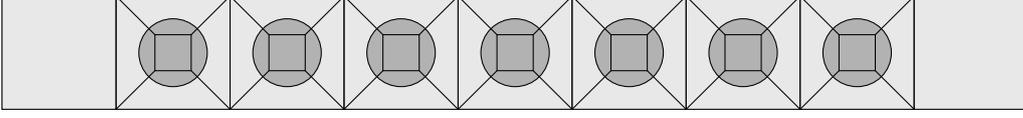


Figure 5: A size dependent mesh for a finite crystal with  $n_p = 7$  periods. Per period of the crystal 9 curvilinear quadrilateral cells are used, plus two elements in the exterior domain  $\Omega^{\text{air}} \cup \Omega^{\text{str}}$ .

and the sets of exponentially decaying  $\mathcal{B}_{k_1}^{+\text{dec}}$  and exponentially growing  $\mathcal{B}_{k_1}^{-\text{dec}}$  Bloch modes

$$\mathcal{B}_{k_1}^{+\text{dec}} := \{\tilde{u}_{\mathbf{k}}(\mathbf{x})e^{i\mathbf{k}\mathbf{x}} \in \mathcal{B}_{k_1} : \text{Im } k_2 > 0\}, \quad \mathcal{B}_{k_1}^{-\text{dec}} := \{\tilde{u}_{\mathbf{k}}(\mathbf{x})e^{i\mathbf{k}\mathbf{x}} \in \mathcal{B}_{k_1} : \text{Im } k_2 < 0\}. \quad (16a)$$

In this work we will study the approximation quality of set of dominant Bloch modes

$$\mathcal{B}_{k_1}^{\text{dom}} := \{\tilde{u}_{\mathbf{k}}(\mathbf{x})e^{i\mathbf{k}\mathbf{x}} \in \mathcal{B}_{k_1} : \text{Im } k_2 = \text{Im } k_2^{(1)}\},$$

which contains only the propagating Bloch modes  $\mathcal{B}_{k_1}^{\text{dom}} = \mathcal{B}_{k_1}^{\text{prop}}$  for frequencies  $\omega$  outside the bandgap and only the weakest decaying and growing Bloch modes for bandgap frequencies  $\mathcal{B}_{k_1}^{\text{dom}} \subset \mathcal{B}_{k_1}^{+\text{dec}} \cup \mathcal{B}_{k_1}^{-\text{dec}}$  (cf. (16a)).

### 3.2. Solving the unit cell problem

The Bloch modes can be approximated by considering (14) as an eigenvalue problem with eigenfunction  $\tilde{u}$  and eigenvalues  $(\mathbf{k}, \omega)$  and solving it using a FEM discretisation and a numerical eigenvalue solver. There are two alternatives to pose (14) as an eigenvalue problem. For a fixed  $\mathbf{k} \in \mathbb{C}^2$ , (14) is a linear eigenvalue problem in  $\omega^2 \in \mathbb{R}_{\geq 0}$ . Fig. 3(a) has been created by sampling  $\mathbf{k} = (0, k)$  with  $k \in [0, \pi/a]$ . ARPACK [49] was used as eigenvalue solver. For the discretisation the FEM solver Concepts [15] with polynomial degree  $p = 10$  in all 9 curvilinear cells of the unit cell  $\hat{\Omega} = [0, 1]^2$  has been used.

The alternative is to fix a frequency  $\omega \in \mathbb{R}_{\geq 0}$  and search for  $\mathbf{k} \in \mathcal{B} + i\mathbb{R}^2 \subset \mathbb{C}^2$ , as used in [20, 19]. With the Bloch transform the eigenvalue  $\mathbf{k}$  occurs as a linear or quadratic factor in the PDE instead of a non-linear factor  $e^{i\mathbf{k}\cdot\mathbf{a}_j}$  in the boundary condition for the Floquet transform.

As eigenvalues have to be scalar, we used the representation

$$\mathbf{k} = \lambda(\sin(\theta), \cos(\theta))^\top + \mathbf{k}^{\text{off}} \quad \text{for } \lambda \in \mathbb{C}, \theta \in [0, 2\pi], \mathbf{k}^{\text{off}} \in \mathbb{C}^2$$

with offset  $\mathbf{k}^{\text{off}}$ , direction  $\theta$  and length  $\lambda$ . We fix all parameters except  $\lambda$  and obtain a quadratic eigenvalue problem in  $\lambda$ , *i.e.*, the eigenvalue problem is of the form  $(\lambda^2 A + \lambda B + C)\tilde{u}_{\mathbf{k}}(\mathbf{x}) = 0$  for some operators  $A, B, C$ . The quadratic eigenvalue problem is linearised as described in [80, Sec. 3.10.2]. Then, the linearised problem is discretised and solved as described for Fig. 3(a). Fig. 3(b) has been created by sampling  $\omega \in \mathbb{R}_{\geq 0}$  and fixing  $\theta = 0$  and  $\mathbf{k}^{\text{off}} = \mathbf{0}$ .

### 3.3. Best-approximation of finite crystal solutions by Bloch modes and macro-polynomials

This section examines how well the solutions  $u$  for the problem (6) for a finite PhC with  $n_p$  periods, the *finite crystal solutions*, can be matched by a few Bloch modes  $\tilde{u}_{\mathbf{k}}(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}$  modulated by macroscopic polynomials for a given frequency  $\omega$ . We investigate in particular size robustness, *i.e.*, how the approximation accuracy behaves using a fixed number of degrees of freedom as the number of periods in the finite PhC increases.

#### 3.3.1. Perpendicular incidence

The finite crystal solutions  $u$  for the finite crystals (6) have been computed with  $p$ -FEM on a size dependent mesh (see Fig. 5) and are sufficiently well discretised (uniform polynomial degree  $p = 23$  in all cells). As incident fields  $u^{\text{inc}}$  we choose plane waves  $e^{i\mathbf{k}^{\text{inc}}\cdot\mathbf{x}}$  with  $\mathbf{k}^{\text{inc}} = (0, \omega)^\top$  (see 2.7), *i.e.*, incident

perpendicularly from left in view of Fig. 4. The solutions for  $\hat{\omega} = 0.625$ ,  $n_p = 7$  is illustrated in Figs. 4(c) and 4(d). Although not obvious, a linear combination of the four Bloch modes of Figs. 4(a) and 4(b) can accurately represent the finite crystal solution in Figs. 4(c) and 4(d). We will show this hereafter by means of an  $L^2$ -projection. The solutions for frequencies in the bandgap behave differently, they decay inside the crystal. This can be clearly seen in the example shown in Fig. 4(e) for the frequency  $\hat{\omega} = 0.300$  lying near the centre of the first bandgap. Here, already 7 periods are sufficient to reflect most of the field.

*Best-approximation by highly resolved pure Bloch modes.* In a first step in Fig. 6(b) we approximated for  $\hat{\omega} = 0.625$  and varying number of periods  $n_p$  the finite crystal solution  $u$  with a linear combination of Bloch modes  $\tilde{u}_{\mathbf{k}}(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}} \in \mathcal{B}_{k_1}^{\text{dom}}$ . The Bloch modes have been computed numerically by solving the unit cell problem (14) with  $p$ -FEM (cf. previous section) for different polynomial degrees  $p_{\text{bloch}}$ . Fig. 6(b) shows the relative  $L^2$ -error for the  $L^2$ -projection of the quasi-periodically extended Bloch modes (see Figs. 4(a) and 4(b)) onto  $u$ . Each curve corresponds to a finite crystal solution  $u$  for a fixed number of periods  $n_p$ , e.g., for  $n_p = 7$  the function  $u$  is displayed in Figs. 4(c) and 4(d).

The approximation by the  $L^2$ -projection was done away from the crystal's surfaces, more precisely in the *domain of inner cells*  $\Omega^{\text{inner}}$ . The domain  $\Omega^{\text{inner}} \subset \Omega^{\text{cr}}$  is the crystal  $\Omega^{\text{cr}}$  without the first and last period of the crystal, see Fig. 6(a). It is observed that for each  $n_p$  a saturation level is obtained when the error level does not decrease further. Approximated Bloch modes with polynomial degrees  $p_{\text{bloch}}$  inside this saturation level sufficiently resolve the analytic solution of (14). This saturation level thus expresses to which accuracy the exact Bloch modes—the solutions for infinite crystals—approximate the solution for finite crystals with  $n_p$  periods. The four Bloch modes (nDOF = 4) at  $\hat{\omega} = 0.625$  are sufficient to approximate the solution for finite PhC's with up to  $n_p = 200$  periods at a relative  $L^2(\Omega^{\text{inner}})$ -error of about  $10^{-4}$ . This result indicates that a size robust method based on Bloch modes exists.

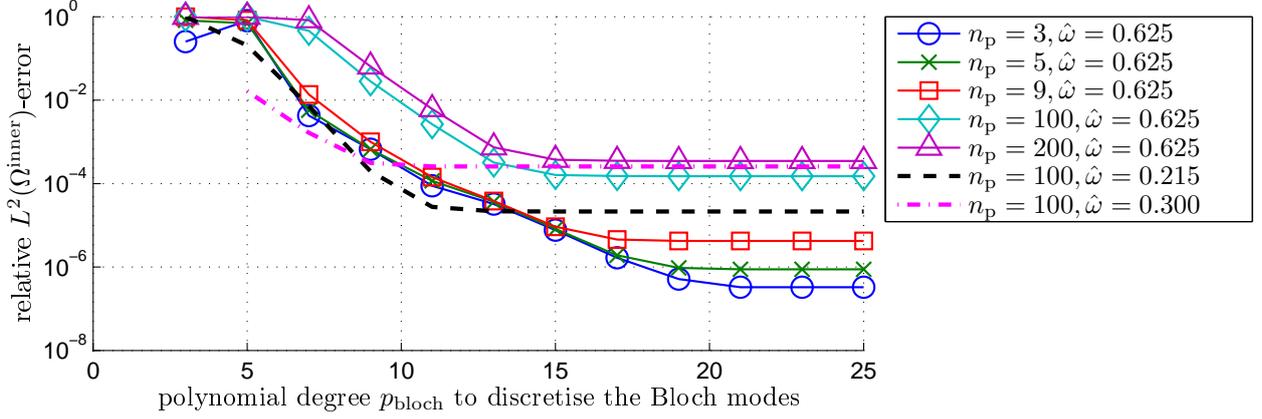
*Best-approximation by Bloch modes multiplied with macro-polynomials—first example:*  $\hat{\omega} = 0.625$ . We match Bloch modes  $\tilde{u}_{\mathbf{k}}(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}} \in \mathcal{B}_{k_1}^{\text{dom}}$  modulated by macroscopic polynomials to the finite crystal solutions (see Fig. 6(c)). The polynomials have a maximal polynomial degree  $p_{\text{mac}}$ , which will be varied. We used for the matching again the  $L^2$ -projection in  $\Omega^{\text{inner}}$ . In this case we fixed the Bloch mode discretisation to  $p_{\text{bloch}} = 25$ , which gives a saturation level in terms of Fig. 6(b). By adding macroscopic polynomials with  $p_{\text{mac}} = 1$  or  $p_{\text{mac}} = 2$  the  $L^2(\Omega^{\text{inner}})$ -best approximation error decreases by about two magnitudes to approximately  $10^{-6}$ . Small crystals can even be approximated with an error as low as  $10^{-8}$ . A further increase of the polynomial degree  $p_{\text{mac}}$  leads to exponential convergence, but with a base notably depending on the number of periods  $n_p$ . Thus for large  $n_p$  a lower error than  $10^{-6}$  is only expected for impractically large  $p_{\text{mac}}$ . Moreover, the convergence already breaks down at  $p_{\text{mac}} = 5$  for  $n_p = 3$  periods due to redundancies in the basis. This break down may even occur earlier, depending on conditioning of the polynomial basis.

The results shown in Fig. 6(d) express why we used the domain of inner cells  $\Omega^{\text{inner}}$  in the previous experiments. This figure illustrates the convergence of the  $L^2$ -projection on the whole crystal  $\Omega^{\text{cr}}$ . The error levels of about  $10^{-4}$  are clearly higher than for Fig. 6(c) even for larger  $p_{\text{mac}}$ . This indicates that the Bloch modes, the solutions of infinite crystals, do not accurately describe the solutions for finite crystals up to their surfaces.

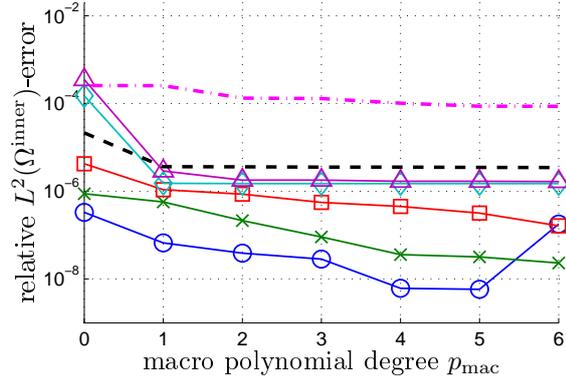
This observation is the motivation to propose a special mesh, the so called *boundary layer mesh*, for the macroscopic scale. This boundary layer mesh consists of three cells, two small cells on the crystal's surface and a large cell in the interior. In view of Fig. 6(c) in the inner cells, already a polynomial degree  $p_{\text{mac}}$  as small as 1 is a very good approximation. Higher values of  $p_{\text{mac}}$  improve the approximation of the solution in the two cells on the surface. This is illustrated in Fig. 6(e) in which we show the relative error of the  $L^2$ -projection to only the first cell of the crystal. Already pure Bloch modes ( $p_{\text{mac}} = 0$ ) achieve error levels below  $10^{-3}$  and  $p_{\text{mac}} = 4$  is enough to reach  $10^{-6}$  the error level comparable to the  $L^2$ -projection on  $\Omega^{\text{inner}}$  in Fig. 6(c). This shows that a basis of Bloch modes multiplied with piecewise polynomials converges very fast on the surface of the crystal. Moreover, we observed that this basis is superior to a pure polynomial basis on  $\Omega^{\text{first}}$ , not resolving the material discontinuities, in the sense that fewer functions are required to achieve a particular error.



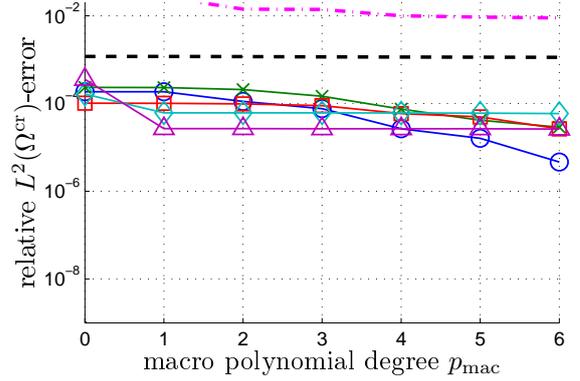
(a) The framed area is the domain containing all cells (all of  $\Omega^{\text{cr}}$ ) and the highlighted area the domain of inner cells  $\Omega^{\text{inner}}$  excluding the first period  $\Omega^{\text{first}}$  and the last period of the crystal. The single cell in the centre of the crystal is denoted by  $\Omega^{\text{center}}$ .



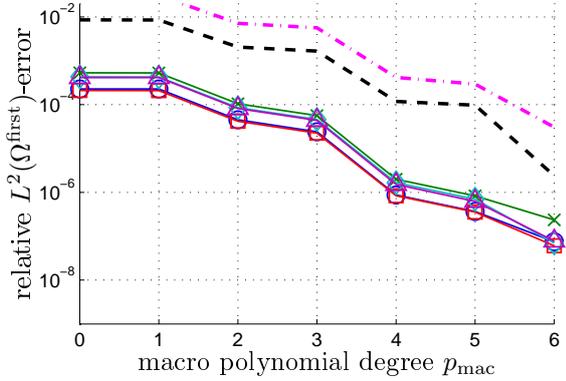
(b)  $L^2$ -projection of pure Bloch modes ( $p_{\text{mac}} = 0$ ) for the inner cells  $\Omega^{\text{inner}}$ .



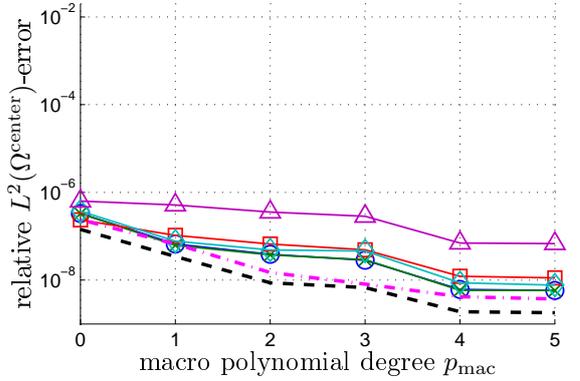
(c)  $L^2$ -projection of polynomials multiplied with Bloch modes for the inner cells  $\Omega^{\text{inner}}$  and fixed  $p_{\text{Bloch}} = 25$ .



(d)  $L^2$ -projection of polynomials multiplied with Bloch modes for the whole crystal  $\Omega^{\text{cr}}$  and fixed  $p_{\text{Bloch}} = 25$ .



(e)  $L^2$ -projection of polynomials multiplied with Bloch modes for the first cell of the crystal with fixed  $p_{\text{Bloch}} = 25$ .



(f)  $L^2$ -projection of polynomials multiplied with Bloch modes for the center cell  $\Omega^{\text{center}}$  of the crystal with fixed  $p_{\text{Bloch}} = 25$ .

Figure 6: Best-approximation error by  $L^2$ -projection of polynomials multiplied with Bloch modes in varying matching domains onto finite crystal solution  $u$  for  $n_p$  number of periods. The matching domains are for **b**), **c**) the inner cells  $\Omega^{\text{inner}}$ , for **d**) the whole crystal  $\Omega^{\text{cr}}$ , for **e**) the first cell  $\Omega^{\text{first}}$  and for **f**) the center cell  $\Omega^{\text{center}}$ . The Bloch modes (e.g. Figs. 4(a) and 4(b) for  $\hat{\omega} = 0.625$ ) have been discretised with p-FEM with maximal polynomial degree  $p_{\text{Bloch}}$ . In **c**)-**f**) the Bloch modes have been modulated with piecewise polynomials with maximal polynomial degree  $p_{\text{mac}}$ . All the solid lines correspond to the propagating frequency  $\hat{\omega} = 0.625$  ( $n_{\text{Bloch}} = 4$ ) and the dashed and dash-dotted lines are for the propagating frequency  $\hat{\omega} = 0.215$  ( $n_{\text{Bloch}} = 2$ ) and the bandgap frequency  $\hat{\omega} = 0.300$  ( $n_{\text{Bloch}} = 2$ ).

The number of periods  $n_{\text{outer}}$  in the two macro cells on the crystal surface does not have to be one. A boundary layer mesh with larger boundary cells might reduce the error in the interior but increase the error in the first cells. The best approximation in a minimal interior cell, the single centre cell  $\Omega^{\text{center}} = (0, a) \times (\lfloor \frac{n_p-1}{2} \rfloor, \lfloor \frac{n_p+1}{2} \rfloor)$  (see Fig. 6(a)), can be seen as a lower bound for larger interior cells. In Fig. 6(f) the relative error for the centre cell  $\Omega^{\text{center}}$  is shown for  $\hat{\omega} = 0.625$ . Already for  $p_{\text{mac}} = 0$  the relative error is lower than  $10^{-6}$ . This error level correspond approximately to the saturation level for the much larger inner cell  $\Omega^{\text{inner}}$  (see Fig. 6(c)). Decreasing the size of the inner cells would not lead to Thus, we may conclude for the frequency  $\hat{\omega} = 0.625$  that it is efficient to use a boundary layer mesh with maximally sized inner cells having  $n_p - 2$  periods and in which a single crystal period on each side of the surface is resolved by a macro cell ( $n_{\text{outer}} = 1$ ).

*Best-approximation by Bloch modes multiplied with macro-polynomials—second example:*  $\hat{\omega} = 0.215$  and  $\hat{\omega} = 0.300$ . The frequency  $\hat{\omega} = 0.625$  studied above is a propagating frequency with  $n_{\text{bloch}} = |\mathcal{B}_{k_1}^{\text{dom}}| = 4$  linearly independent Bloch modes  $\tilde{u}_{\mathbf{k}}(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}} \in \mathcal{B}_{k_1}^{\text{prop}}$ . We will also compare the approximation quality for the propagating frequency  $\hat{\omega} = 0.215$  with  $n_{\text{bloch}} = 2$  linearly independent Bloch modes  $\tilde{u}_{\mathbf{k}}(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}} \in \mathcal{B}_{k_1}^{\text{prop}}$  and for the bandgap frequency  $\hat{\omega} = 0.215$  with  $n_{\text{bloch}} = 2$  linearly independent evanescent Bloch modes  $\tilde{u}_{\mathbf{k}}(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}} \in \mathcal{B}_{k_1}^{\text{dec}} \cup \mathcal{B}_{k_1}^{+\text{dec}}$  with  $|\text{Im } k_2| = 0.70$ . For  $\hat{\omega} = 0.215$  and  $\hat{\omega} = 0.300$ , the approximation results of the single example crystal with  $n_p = 100$  periods have been included in Figs. 6(b)–6(f). The same properties are observed as for  $\hat{\omega} = 0.625$ . Both frequencies have only half as many Bloch modes as  $\hat{\omega} = 0.625$  and we see that the error levels for the matching domains  $\Omega^{\text{cr}}$  and  $\Omega^{\text{first}}$  are higher for the same  $p_{\text{mac}}$ . Moreover, in order to minimize a global error on  $\Omega^{\text{cr}}$  for propagating frequencies the error in  $\Omega^{\text{inner}}$  is more important than that in  $\Omega^{\text{first}}$ . However, for bandgap frequencies which decay inside the crystal the error in  $\Omega^{\text{first}}$  can sometimes be more important than the error in  $\Omega^{\text{inner}}$ . In this sense, simulations at bandgap frequencies are in general easier. Already heuristic methods like the simulation only for a certain number of periods and extension of the solution by zero, lead to size robust methods for large finite PhC's.

### 3.3.2. Oblique incidence

We now investigate the approximation quality of the basis based on the dominant Bloch modes and macroscopic polynomials for oblique incidence. We use the example frequency  $\hat{\omega} = 0.230$  which has a bandstructure in dependence of  $k_1$  as depicted in Figs. 7(a) and 7(b). The bandstructures have been computed as solutions of the quadratic eigenvalue problem described in Sec. 3.2 by sampling  $k_1$  and specifying  $\mathbf{k}^{\text{off}} = (k_1, 0)^\top$  and  $\theta = \frac{\pi}{2}$ . The chosen frequency  $\hat{\omega} = 0.230$  is particularly interesting as there are only propagating modes for  $\varphi > 47^\circ$  and otherwise the frequency is in the bandgap, *i.e.*, changing incident angle results in a notably change of the crystal behaviour.

*Selection of dominant Bloch modes for bandgap frequencies.* We compare three different sets of Bloch modes based on the set  $\mathcal{B}_{k_1}^{\text{dom}}$ , this set always contains the set  $\mathcal{B}_{k_1}^{\text{prop}}$  for the propagating regime and in the bandgap regime it contains

- (D1) the dominant decaying Bloch modes  $\mathcal{B}_{k_1}^{+\text{dec}} \cap \mathcal{B}_{k_1}^{\text{dom}}$  and dominant growing Bloch modes  $\mathcal{B}_{k_1}^{\text{dec}} \cap \mathcal{B}_{k_1}^{\text{dom}}$
- (D2) only the dominant decaying Bloch modes  $\mathcal{B}_{k_1}^{+\text{dec}} \cap \mathcal{B}_{k_1}^{\text{dom}}$ ,
- (D3) the dominant decaying Bloch modes  $\mathcal{B}_{k_1}^{+\text{dec}} \cap \mathcal{B}_{k_1}^{\text{dom}}$  and dominant artificially propagating Bloch modes  $\widehat{\mathcal{B}}_{k_1}^{\text{dec}}$ , where

$$\widehat{\mathcal{B}}_{k_1}^{\text{dec}} := \left\{ \tilde{u}_{\mathbf{k}}(\mathbf{x})e^{i(k_1x_1 + \text{Re } k_2x_2)} : \tilde{u}_{\mathbf{k}}(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}} \in \mathcal{B}_{k_1}^{\text{dec}} \cap \mathcal{B}_{k_1}^{\text{dom}} \right\}.$$

The set  $\widehat{\mathcal{B}}_{k_1}^{\text{dec}}$  is a set of propagating modes which has the same periodic part as the Bloch modes  $\mathcal{B}_{k_1}^{\text{dec}}$ , but which discards the exponential growth of the modes  $\mathcal{B}_{k_1}^{\text{dec}}$  in direction  $x_2$ .

Exemplary for  $p_{\text{mac}} = 2$ , the best approximation results for the selection rules (D1)–(D3) are displayed in Fig. 7(c). For (D1), an error of about  $10^{-4}$  is obtained independent of  $\varphi$  for both small crystals ( $n_p = 3$ )

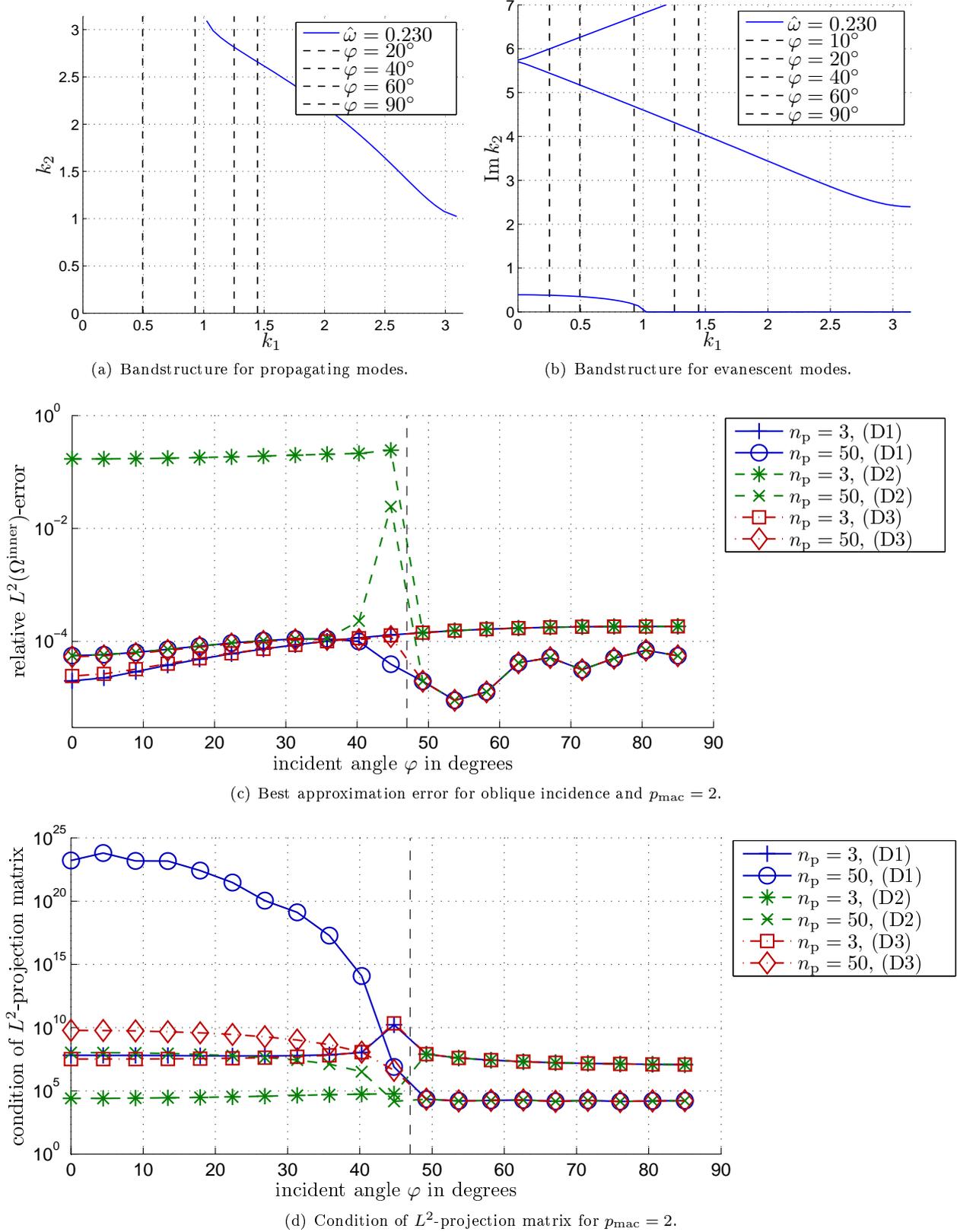


Figure 7: Best-approximation for oblique incidence for the example frequency  $\hat{\omega} = 0.230$  ( $n_{\text{block}} = 2$  for all  $\varphi$ ) which is a bandgap frequency for incident angles  $\varphi < 47^\circ$  and propagating otherwise. The bandstructure for  $\hat{\omega} = 0.230$  is displayed in a) and b). Specifying an incident angle  $\varphi$  is equivalent to specifying the Floquet parameter  $k_1$  of the model problem (6). The  $L^2$ -best approximation result is displayed in c) exemplary for a crystal with very few periods ( $n_p = 3$ ) and many periods ( $n_p = 50$ ). We compared the approximation quality in c) and the condition number in d) for different selection rules (D1), (D2), (D3) for the sets of Bloch modes.

and large crystals ( $n_p = 50$ ). This concludes that the method is size and incident angle robust. Moreover, it is robust with respect to the decay strength  $\text{Im} k_2$  as no degeneration is observed close to  $\varphi = 47^\circ$ . However, in Fig. 7(d) we see that the condition numbers for the selection rule (D1) can become huge for large  $n_p$  and away from the propagating regime. On the other hand selection rule (D2) in Figs. 7(c) and 7(d) shows that in this problematic case (large  $|\text{Im} k_2|$  and  $n_p$ ) the modes  $\mathcal{B}_{k_1}^{\text{dec}} \cap \mathcal{B}_{k_1}^{\text{dom}}$  are not necessary. This observation suggests, that there exists a selection rule which has reasonable condition numbers independent of  $\varphi$  and  $n_p$  and achieves an approximation error of  $10^{-4}$ . For small crystals it behaves like (D1), close to the propagating regime like (D1) and for large crystals and away from the propagating regime like (D2). Indeed such a selection rule is given by (D3). The selection rules (D1) and (D3) both also work equally well for the first cell  $\Omega^{\text{first}}$ , whereas (D2) does not work for the first cell. For (D2) and  $\hat{\omega} = 0.230$  or  $\hat{\omega} = 0.300$  only  $n_{\text{bloch}} = 1$  Bloch mode  $\tilde{u}_k(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}$  is used which even has a linearly dependent real part  $\text{Re} \tilde{u}_k(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}$  and imaginary part  $\text{Im} \tilde{u}_k(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}$ . This dependence is due to the wave vector being on the boundary of the Brillouin zone ( $\text{Re} k_2 = \pi$ ) and the symmetry properties of Bloch modes.

#### 4. The Multiscale FEM

The multiscale FEM has many similarities to other FEM variants. We will especially contrast it to *standard* FEM, by which we mean all FEM variants that use piecewise polynomial basis functions on a mesh  $\mathcal{M}$ . For example the  $h$ -,  $p$ -FEM [10, 73] and  $hp$ -FEM [69] variants or spectral elements [13, 60] belong to this category. The main difference of the multiscale FEM to other FEMs is the use of basis functions which have variations on multiple, clearly defined scales. Hence, the multiscale FEM belongs to the large class of generalised FEMs (g-FEM) with possibly non-polynomial basis functions. A multiscale FEM for homogenisation problems with local periodicity has been introduced in [52, 64]. Our work extends this approach to the simulation of PhC structures at resonance frequencies for which homogenisation does not apply. For these structures we will use a two-scale basis with Bloch modes (see Section 3) as microscopic basis functions. Rather than restricting ourselves to the specific case of Bloch modes we will introduce hereafter how to construct a two-scale multiscale FEM for a more general basis of quasi-periodic micro functions and  $C^0$ -piecewise polynomial macro functions.

Our multiscale FEM is designed to be efficient for the particular locally periodic structure of a finite PhC structure in  $\Omega^{\text{cr}}$ . As this structure is generally not apparent in the exterior domain  $\Omega^{\text{ext}} = \Omega^{\text{str}} \cup \Omega^{\text{air}}$ , our presentation focuses on combining multiscale FEM with an arbitrary *exterior* discretisation in  $\Omega^{\text{ext}}$ . The combined method is thus a hybrid FEM of the interior and exterior discretisations. The method approximates the variational formulation (7) in a finite-dimensional subspace  $V_{k_1}^{\text{gfem}}$  of the original space  $H_{k_1}^1(\Omega)$  of the variational formulation:

$$\text{find } u_N^{\text{gfem}} \in V_{k_1}^{\text{gfem}} : \quad \Phi_{k_1}(u_N^{\text{gfem}}, v_N) = f(v_N) \quad \text{for all } v_N \in V_{k_1}^{\text{gfem}}, \quad (17)$$

where  $\Phi_{k_1}$  and  $f$  are given in (7). The solution  $u_N^{\text{gfem}}$  is a linear combination of the basis functions  $b^{\text{gfem}}$  of  $V_{k_1}^{\text{gfem}}$

$$u_N^{\text{gfem}}(\mathbf{x}) = \sum_{\ell=1}^{n^{\text{gfem}}} \alpha_\ell^{\text{gfem}} b_\ell^{\text{gfem}}(\mathbf{x}) \quad \text{for all } \mathbf{x} \text{ in } \Omega, \quad (18)$$

with the coefficients  $\alpha_\ell^{\text{gfem}} \in \mathbb{C}$  and where  $n^{\text{gfem}}$  is the dimension of  $V_{k_1}^{\text{gfem}}$ .

The space  $V_{k_1}^{\text{gfem}}$  on the computational domain  $\Omega$  is constructed from two spaces: the space  $V_{k_1}^{\text{ext}} \subset H_{k_1}^1(\Omega^{\text{ext}})$  on the exterior domain  $\Omega^{\text{ext}}$  and the MSFEM space  $V_{k_1}^{\text{multi}} \subset H_{k_1}^1(\Omega^{\text{cr}})$  on the crystal domain  $\Omega^{\text{cr}}$ . The two spaces are each suitable for solving the Helmholtz equation (2) on the subdomains  $\Omega^{\text{ext}}$  and  $\Omega^{\text{cr}}$ , with suitable boundary conditions.

In Sec. 4.1 we define the space  $V_{k_1}^{\text{multi}}$  and its multiscale basis functions. Then we explain in Sec. 4.2 how to combine the multiscale space  $V_{k_1}^{\text{multi}}$  with some exterior FE space  $V_{k_1}^{\text{ext}}$  to form the hybrid space  $V_{k_1}^{\text{gfem}}$ . The hybridisation for the concrete example of standard FEM as exterior discretisation  $V_{k_1}^{\text{ext}}$  will be given in Sec. 4.3.

#### 4.1. Multiscale basis functions

In the finite crystal  $\Omega^{\text{cr}}$  two scales are apparent: a microscopic scale with oscillations on the size of a single unit cell and a macroscopic scale with oscillations on the size of the whole crystal. The multiscale space  $V_{k_1}^{\text{multi}}$  takes advantage of this structure by introducing basis functions with good approximation properties at these microscopic and macroscopic scales.

Accordingly, the basis functions  $b^{\text{multi}}$  are the two-scale functions

$$b_{\mathbf{i}}^{\text{multi}}(\mathbf{x}) = b_{i_1}^{\text{mac}}(\mathbf{x}) b_{i_2}^{\text{mic}}(\mathbf{x}), \quad \text{for } \mathbf{i} = (i_1, i_2) \text{ and } 1 \leq i_1 \leq n_{\text{mac}}, 1 \leq i_2 \leq n_{\text{mic}} \quad (19a)$$

spanning the space

$$V_{k_1}^{\text{multi}} = \text{span} \{b_{\mathbf{i}}^{\text{multi}} : 1 \leq i_1 \leq n_{\text{mac}}, 1 \leq i_2 \leq n_{\text{mic}}\} = V_{k_1}^{\text{mic}} \otimes V_{\text{per}}^{\text{mac}}, \quad (19b)$$

where  $n_{\text{mic}}$  and  $n_{\text{mac}}$  are the number of micro and macro functions. The space has tensor product structure generated by the below defined spaces  $V_{k_1}^{\text{mic}}$  and  $V_{\text{per}}^{\text{mac}}$ . Every micro function  $b^{\text{mic}}$  will be combined with every macro function  $b^{\text{mac}}$  and the dimension  $n^{\text{multi}}$  of  $V_{k_1}^{\text{multi}}$  is  $n_{\text{mic}}n_{\text{mac}}$ . It is sometimes convenient to linearly enumerate the basis functions as  $b_{\ell}^{\text{multi}} = b_{i,j}^{\text{multi}}$  with the *packing operator*  $\mathbb{N}^2 \rightarrow \mathbb{N} : (i_1, i_2) \mapsto \ell := i_1 + (i_2 - 1)n_{\text{mic}}$ .

The Figures 8(b) and 8(c) illustrate the multiscale structure. The macro polynomials  $b^{\text{mac}}$  are piecewise polynomials on a *macroscopic mesh*  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  (see Fig. 8(a)). The cells  $K^{\text{mac}} \in \mathcal{M}_{\text{cr}}^{\text{mac}}$  of this mesh are parallelograms which are aligned with the periodic pattern and contain  $1 \times n$  full repetitions of the unit cell for some  $n \in \mathbb{N}$ . The micro basis functions are quasi-periodic functions on  $\mathbb{R}^2$  spanning the space  $V_{k_1}^{\text{mic}}$ ,

$$b_{i_2}^{\text{mic}}(\mathbf{x}) = e^{i\mathbf{k}_{i_2} \cdot \mathbf{x}} b_{i_2}^{\text{per}}(\mathbf{x}), \quad \mathcal{E}_{k_1}^{\text{mic}} = \{b_{i_2}^{\text{mic}} : 1 \leq i_2 \leq n_{\text{mic}}\}, \quad V_{k_1}^{\text{mic}} = V_{k_1}^{\text{mic}}(n_{\text{mic}}) = \text{span} \mathcal{E}_{k_1}^{\text{mic}}.$$

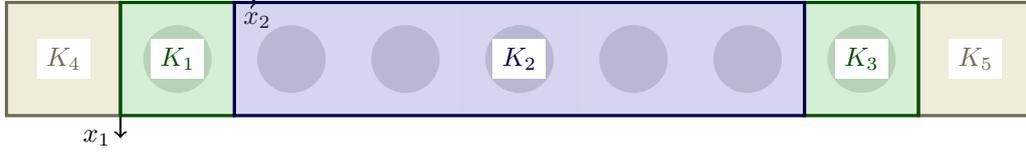
The micro functions have periodic parts  $b_{i_2}^{\text{per}} \in H_{\text{per}}^1(\widehat{\Omega})$  and wave vectors  $\mathbf{k}_{i_2} \in \mathbb{C}^2$  where  $k_{1,i_2} = k_1$  so that  $b_{i_2}^{\text{mic}}$  satisfies the boundary condition (6b). For example the functions  $b^{\text{per}}$  can be solutions of the unit cell problems (14) for wave vectors  $\mathbf{k}_{i_2}$ , i.e. the first  $n_{\text{bloch}}$  Bloch modes are a valid choice for  $b^{\text{mic}}$  where  $V_{k_1}^{\text{mic}} = V_{k_1}^{\text{mic}}(n_{\text{bloch}}) = \text{span} \mathcal{B}_{k_1}(n_{\text{bloch}})$  (cf. Sec. 3.1). The spaces  $V_{k_1}^{\text{mic}}(n_{\text{bloch}})$  obtained by successively adding micro basis functions are hierarchical, i.e.,  $V_{k_1}^{\text{mic}}(1) \subset V_{k_1}^{\text{mic}}(2) \subset \dots$

*Macro basis functions.* The macro basis functions  $b^{\text{mac}}$  can be any polynomial basis functions which are  $\mathbf{a}_1$ -periodic on the mesh  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  (see Fig. 8(a)) and have tensor product structure on the reference cell of each macro cell  $K^{\text{mac}} \in \mathcal{M}_{\text{cr}}^{\text{mac}}$ . We use tensor product basis functions due to Karniadakis and Sherwin [43] which are basically integrated Legendre polynomials. Such a basis is known to result in well-conditioned system matrices for standard FEM, see [24]. The macro functions  $b^{\text{mac}}$  span the space

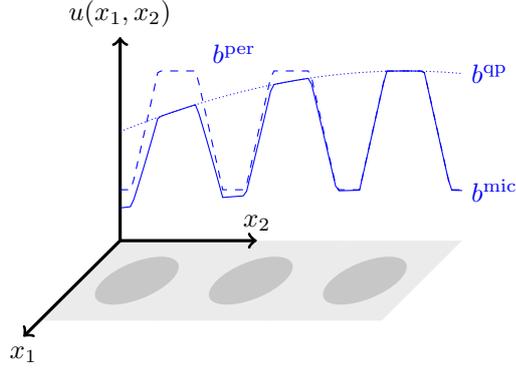
$$V_{\text{per}}^{\text{mac}} = V_{\text{per}}^{\text{mac}}(\mathcal{M}_{\text{cr}}^{\text{mac}}, \mathbf{p}_1^{\text{mac}}, \mathbf{p}_2^{\text{mac}}) = \text{span} \{b_{i_1}^{\text{mac}} : 1 \leq i_1 \leq n_{\text{mac}}\}, \quad (20)$$

of dimension  $n_{\text{mac}}$  which is determined by the mesh  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  and the maximal polynomial degree distributions  $\mathbf{p}_1^{\text{mac}}, \mathbf{p}_2^{\text{mac}} \in \mathbb{N}^{\ell}$  in the two local directions in each of the  $\ell$  cells  $K^{\text{mac}} \in \mathcal{M}_{\text{cr}}^{\text{mac}}$ . In the present work we restrict ourself to a fixed mesh  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  as displayed in Fig. 8(a) and a uniform macro polynomial degree  $p_{\text{mac}}$  on each cell  $K^{\text{mac}} \in \mathcal{M}_{\text{cr}}^{\text{mac}}$  and in both  $x_1$ - and  $x_2$ -directions. The corresponding spaces  $V_{\text{per}}^{\text{mac}}(p_{\text{mac}})$  are hierarchical with  $V_{\text{per}}^{\text{mac}}(1) \subset V_{\text{per}}^{\text{mac}}(2) \subset \dots$

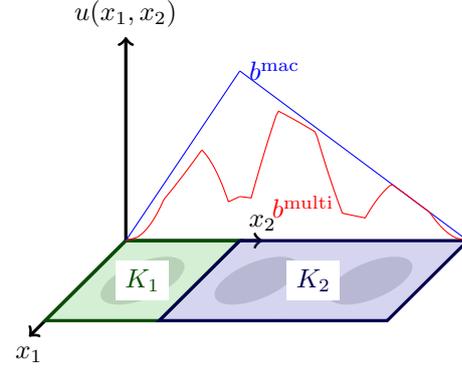
To build the spaces  $V_{\text{per}}^{\text{mac}}(p_{\text{mac}})$  it is convenient to use so called *shape functions*  $s_{\hat{\mathbf{i}}}^{\text{mac}} : [0, 1]^2 \rightarrow \mathbb{R}$ , which are restrictions of macro functions  $b^{\text{mac}}$  pushed-back onto the reference cell, i.e.,  $s_{\hat{\mathbf{i}}}^{\text{mac}}(F_{K^{\text{mac}}}^{-1}\mathbf{x}) = b_{i_1}^{\text{mac}}(\mathbf{x})|_{K^{\text{mac}}}$ . Here  $F_{K^{\text{mac}}} : [0, 1]^2 \rightarrow K^{\text{mac}}$  is the element map of the macro cell  $K^{\text{mac}}$ . Moreover, the shape functions have tensor product structure  $s_{\hat{\mathbf{i}}}^{\text{mac}}(\hat{\mathbf{x}}) = s_{\hat{i}_1}^{\text{mac}}(\hat{x}_1) s_{\hat{i}_2}^{\text{mac}}(\hat{x}_2)$  and there are  $p_{\text{mac}} + 1$  polynomials in each direction, i.e.,  $0 \leq \hat{i}_1, \hat{i}_2 \leq p_{\text{mac}}$ . As the macro mesh in the crystal  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  consists of parallelograms only, the element maps are affine and both the macro basis functions and the shape functions are polynomials. In most FEMs it is convenient for the evaluation of the bilinear forms to work with the shape functions  $s^{\text{mac}}$  rather than to work directly with the macro basis functions  $b^{\text{mac}}$ . The bilinear forms can be written as sums of integrals over  $K^{\text{mac}}$  and each  $K^{\text{mac}}$  can be considered individually with its own set of shape functions  $s_{\hat{\mathbf{i}}}^{\text{mac}}$ , whereas  $b_{\ell}^{\text{mac}}$  possibly has a support of several cells  $K^{\text{mac}}$ . The relation between the shape function index  $\hat{\mathbf{i}}$  and the macro index  $i_1$  is given via the local to global mapping  $m_{K^{\text{mac}}}^{\text{glob}} : (\hat{\mathbf{i}}, K^{\text{mac}}) \rightarrow i_1$ .



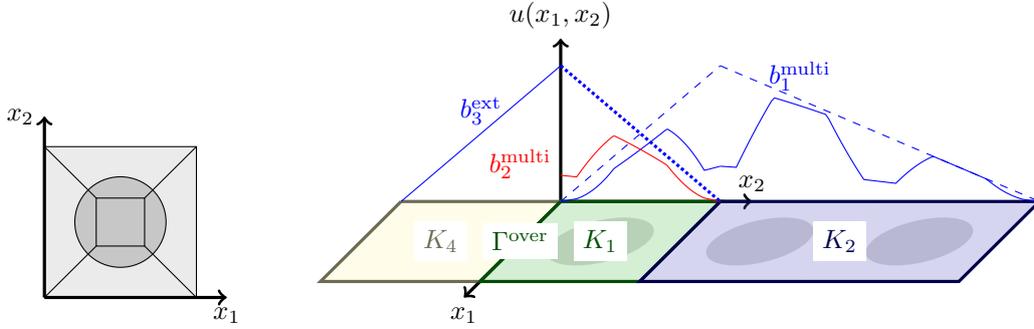
(a) Global macro cell discretisation  $\mathcal{M}^{\text{mac}}$  of  $\Omega$  containing a crystal with  $n_p = 7$  periods. The three multiscale cells  $K_1, K_2, K_3$  form a mesh  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  of  $\Omega^{\text{cr}}$  on which the macro functions  $b^{\text{mac}}$  are defined. Similarly, the two standard-FEM cells  $K_4, K_5$  form a mesh  $\mathcal{M}_{\text{ext}}^{\text{mac}}$  of  $\Omega^{\text{ext}}$  and are used for the exterior discretisation.



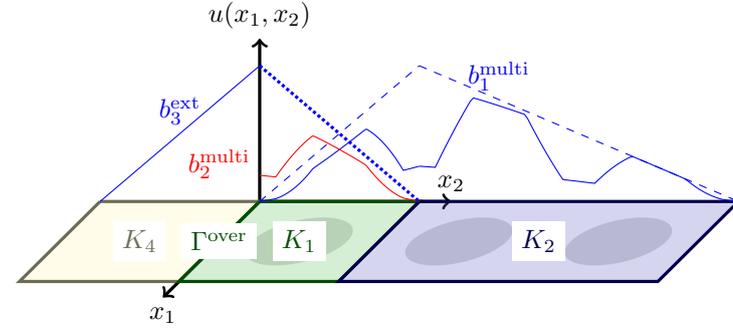
(b) The micro basis function  $b^{\text{mic}}$  is the product of a periodic function  $b^{\text{per}}$  with a quasi-periodic function  $b^{\text{qp}}$ .  $b_k^{\text{qp}} = e^{ikx}$ .



(c) The macro basis function  $b^{\text{mac}}$  is a polynomial in the macro cells  $K_1$  and  $K_2$ . The multi-scale basis function  $b^{\text{multi}}$  is the product of the micro and macro basis functions.



(d) Mesh  $\mathcal{M}^{\text{mic}}$  of the unit cell made out of 9 curvilinear quadrilateral *micro cells*. The macro basis functions are in each micro cell a tensor product of polynomials on the reference cell. The unit cell is discretised with a uniform polynomial degree.



(e) Illustration of the overlap cell handling to achieve a globally continuous hybrid space  $V_{k_1}^{\text{gfem}} \subset H^1(\Omega)$ . Basis functions like  $b_1^{\text{multi}}$  are supported purely inside the crystal and are not affected. Functions that are supported on the boundary  $\Gamma^{\text{over}}$  between multiscale and exterior cells need special attention. The shape function  $b_2^{\text{multi}}$  cannot be matched continuously on  $\Gamma^{\text{over}}$  with polynomials on  $K_4$  and is dropped. To match the shape function  $b_3^{\text{ext}}$  on  $K_4$  the special constant micro function  $b^{\text{mic}} \equiv 1$  is introduced which is in our illustration multiplied with a linear macro function (dotted).

Figure 8: Illustration of the multiscale basis functions (b), (c), the macro and micro meshes (a), (d) and the special handling of basis functions on  $\Gamma^{\text{over}}$  (e).

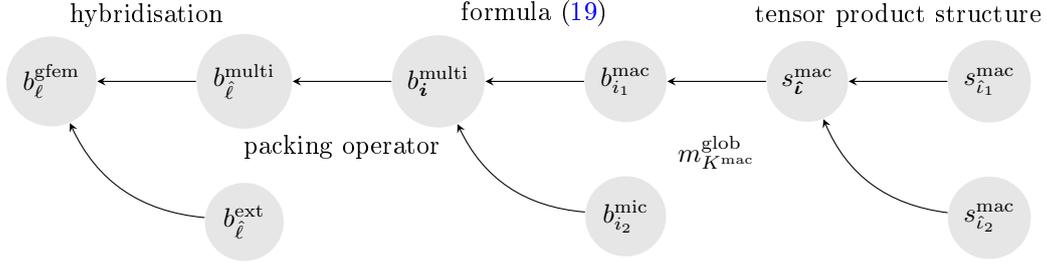


Figure 9: Dependencies, naming scheme and index conventions for basis functions used in Multiscale FEM. Next to the arrows the rules are displayed according to which the indices are related. The indices are the shape function indices  $\hat{\ell}$ , the multiscale indices  $\hat{i}$  and the scalar indices  $\ell, \hat{\ell}$  of the spaces  $V_{k_1}^{\text{gfem}}, V_{k_1}^{\text{multi}}, V_{k_1}^{\text{ext}}$ .

*Hierarchical structure of MSFEM.* The above mentioned hierarchy of the spaces  $V_{k_1}^{\text{mic}}(n_{\text{mic}})$  and  $V_{\text{per}}^{\text{mac}}(p_{\text{mac}})$  implies a hierarchy of two-scale space  $V_{k_1}^{\text{multi}} = V_{k_1}^{\text{multi}}(n_{\text{mic}}, p_{\text{mac}}) = V_{k_1}^{\text{mic}}(n_{\text{mic}}) \otimes V_{\text{per}}^{\text{mac}}(p_{\text{mac}})$  in both  $n_{\text{mic}}$  and  $p_{\text{mac}}$ . The symbol  $\otimes$  stands for the tensor product in the sense of the representation formula (19). A better approximation in MSFEM can be achieved by increasing  $n_{\text{mic}}$  or  $p_{\text{mac}}$  or a combination of both. Many powerful techniques for standard FEM are also applicable to MSFEM. The space  $V_{\text{per}}^{\text{mac}}(p_{\text{mac}})$  can be generalised by using  $h$ -,  $p$ - or  $hp$ -refinements of  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  and the dimension  $n^{\text{multi}}$  of  $V_{k_1}^{\text{multi}}$  can be reduced by sparse tensor approximation techniques [70].

*Basis function naming schemes and dependencies.* Figure 9 illustrates the naming scheme of basis and shape functions as used in this article as well as their indices and relation.

*Unit cell problem.* The micro functions  $b^{\text{mic}}$  are usually not known analytically, they are FEM solutions of special *unit cell problems* on some micro mesh  $\mathcal{M}^{\text{mic}}$ . In all our tests the micro functions  $b^{\text{mic}}$  are approximations to the Bloch modes and solutions to the unit cell problem (14). The micro mesh  $\mathcal{M}^{\text{mic}}$  used for the approximation is shown in Fig. 8(d). The unit cell problem (14) was solved as described in Sec. 3.2. The polynomial degree  $p_{\text{bloch}}$  used to solve (14) on  $\mathcal{M}^{\text{mic}}$  was always chosen sufficiently high so that a saturation level was reached, see Fig. 6.

*Conformity of  $V_{k_1}^{\text{multi}}$  in  $H_{k_1}^1(\Omega^{\text{cr}})$ .* The micro functions  $b^{\text{mic}}$  are constructed to be  $\mathbf{a}_1$ -quasi-periodic with wave number  $k_1$  and they are globally continuous, *i.e.*, they are in  $H_{k_1}^1(\Omega^{\text{cr}})$ . The macro functions are  $\mathbf{a}_1$ -periodic and continuous on  $\Omega^{\text{cr}}$ . Hence, each multiscale basis function  $b^{\text{gfem}}$  is in the space  $H_{k_1}^1(\Omega^{\text{cr}})$ , as it is as a product of a micro and a macro function. Thus the space  $V_{k_1}^{\text{multi}}$  spanned by the functions  $b^{\text{gfem}}$  is  $H_{k_1}^1(\Omega^{\text{cr}})$ -conforming.

**Remark 4.** *Plane wave methods [3, 14] could be considered as a special case where no macro basis exists, *i.e.*,  $b^{\text{mac}} \equiv 1$ , and the micro functions are plane waves ( $b^{\text{per}} \equiv 1$ ). In plane wave methods, the reduction of the discretisation error is achieved by increasing the number of wave vectors  $\mathbf{k}$ . On the other hand,  $h$ -,  $p$ - or  $hp$ -FEM can be considered as a special case where the micro basis is trivial, *i.e.*,  $b^{\text{mic}} \equiv 1$ , and the macro basis are polynomials. Here, a better approximation is achieved by refining the mesh  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  or increasing the maximal polynomial degree  $p_{\text{mac}}$  respectively. Since  $b^{\text{mac}}$  form in particular a partition of unity, the present proposed method can be considered as a special case of PUFEM [54].*

**Remark 5** (Dispersion and pollution error). *The dispersion error [2] in a homogeneous medium is the (best-approximation) error in the wave-vector  $\mathbf{k} \in \mathbb{R}^2$  of a plane wave  $e^{i\mathbf{k} \cdot \mathbf{x}}$  when approximated in the discrete space. Thus, it represents the error in the phase shift  $i(\mathbf{k} - \mathbf{k}_{\text{appr}}) \cdot \mathbf{x}$  for some (long) distance  $\mathbf{x}$ . The analog for the photonic crystal is the error to correctly approximate a propagating Bloch mode. In the proposed MSFEM for photonic crystal bands, a quasi 1D setting, the Bloch modes are included in the basis leading to an absence of a dispersion error for a single Floquet parameter  $k_1$  (as for plane-wave methods in 1D). For*

non-plane incoming waves the Brillouin zone for the Floquet transform in direction  $x_1$  has to be sampled (see Sec. 2.6) and a dispersion error occurs which decays with the number of sampling points.

A pollution error may occur additionally as a loss of stability of the discrete approximation of the sesquilinear forms [58]. For homogeneous media, this is the case for large wave-numbers  $|\mathbf{k}|$ , and so for large frequencies  $\omega$ . The interesting phenomena in PhCs occur in a certain frequency range (the incident wave and periodicity length are of the same order), and the pollution error of our method is consequently equivalent to the discretisation error.

#### 4.2. Hybridisation of the multiscale FEM—general case

The two spaces  $V_{k_1}^{\text{ext}}$  and  $V_{k_1}^{\text{multi}}$  need to be combined to the g-FEM space  $V_{k_1}^{\text{gfem}} \subset H_{k_1}^1(\Omega)$ . For this we will define how the basis functions  $b^{\text{gfem}}$  of  $V_{k_1}^{\text{gfem}}$  are constructed using the basis functions  $b^{\text{ext}}$  and  $b^{\text{multi}}$  of  $V_{k_1}^{\text{ext}}$  and  $V_{k_1}^{\text{multi}}$ . The functions  $b^{\text{gfem}}$  belong to one of the following three categories:

- (B1) the exterior basis functions  $b^{\text{ext}}$  supported inside  $\overline{\Omega^{\text{air}}} \setminus \Gamma^{\text{over}}$ ,
- (B2) the multiscale basis functions  $b^{\text{multi}}$  supported inside  $\overline{\Omega^{\text{cr}}} \setminus \Gamma^{\text{over}}$  and
- (B3) mixed basis functions  $b^{\text{over}}$  supported both in  $\overline{\Omega^{\text{air}}}$  and  $\overline{\Omega^{\text{cr}}}$  (overlapping functions).

The basis functions  $b^{\text{ext}}$  and  $b^{\text{multi}}$  without support on  $\Gamma^{\text{over}}$  can be taken directly as basis functions  $b^{\text{gfem}}$  as they are also continuous on the whole domain  $\Omega$  if extended by zero. Only functions  $b^{\text{ext}}$  and  $b^{\text{multi}}$  with support on  $\Gamma^{\text{over}}$  need to be matched on  $\Gamma^{\text{over}}$  in order to create continuous functions in  $\Omega$ . If a linear combination of exterior functions  $b^{\text{ext}}$  can be matched continuously with a linear combination of multiscale functions  $b^{\text{multi}}$  on  $\Gamma^{\text{over}}$ , then these linear combinations form a function  $b^{\text{over}}$ . However, to find non-trivial linear combinations might not be possible (see Fig. 8(e)). To match two spaces  $V_m$  and  $V_s$ , where  $V_m = V_{k_1}^{\text{multi}}$  and  $V_s = V_{k_1}^{\text{ext}}$  or vice versa, two operations are possible for each basis function  $b_m$  of the master space  $V_m$  that does not vanish on  $\Gamma^{\text{over}}$ :

- (M1) The function  $b_m$  is extended continuously into the domain of  $V_s$  and this extension is added as a new basis function in  $V_s$  (**enrichment**). The extension and  $b_m$  form an overlap function  $b_\ell^{\text{over}}$  which can be selected as a basis function of  $V_{k_1}^{\text{gfem}}$  (B3).
- (M2) The function  $b_m$  is removed from  $V_s$  and thus will not appear as basis function of  $V_{k_1}^{\text{gfem}}$  (**omission**).

One of the two operations is applied to each non-matching basis function in  $V_{k_1}^{\text{multi}}$  and  $V_{k_1}^{\text{ext}}$ . Let the modified spaces be  $\tilde{V}_{k_1}^{\text{multi}}$  and  $\tilde{V}_{k_1}^{\text{ext}}$ . The modification procedure assures that the space  $V_{k_1}^{\text{gfem}}$  can be constructed by simply inserting basis functions of  $\tilde{V}_{k_1}^{\text{multi}}$  and  $\tilde{V}_{k_1}^{\text{ext}}$  that vanish on  $\Gamma^{\text{over}}$  (categories (B1),(B2)), and by pairwise matching of two respective functions  $b_1 \in \tilde{V}_{k_1}^{\text{multi}}$ ,  $b_2 \in \tilde{V}_{k_1}^{\text{ext}}$  with support on  $\Gamma^{\text{over}}$  (category (B3)). If the number of basis functions in  $\tilde{V}_{k_1}^{\text{ext}}$  and  $\tilde{V}_{k_1}^{\text{multi}}$  are  $\tilde{n}^{\text{ext}}$  and  $\tilde{n}^{\text{multi}}$ , then the number of basis functions in  $V_{k_1}^{\text{gfem}}$  is  $n^{\text{gfem}} = \tilde{n}^{\text{multi}} + \tilde{n}^{\text{ext}} - n^{\text{over}}$  where  $n^{\text{over}}$  is the number of functions in (B3).

The challenge is to select the spaces  $V_{k_1}^{\text{ext}}$  and  $V_{k_1}^{\text{multi}}$  and to perform the hybridisation (M1) and (M2) so that the hybrid space  $V_{k_1}^{\text{gfem}}$  has good properties. Enrichment of the space (M1) can help for good approximation properties but omission (M2) leads to less nDOFs, reduced computational cost and possibly simpler implementations. Although the matching affects the approximation quality for a constant  $n_p$  it does not affect the asymptotic behaviour for large  $n_p$  very much as the extension in the enrichment process can be limited to a certain depth into  $\Omega^{\text{cr}}$  for all  $n_p$  and the effect of the functions  $b^{\text{over}}$  is thus concentrated close to  $\Gamma^{\text{over}}$ . Depending on the choice of the matching and the exterior discretisation  $V_{k_1}^{\text{ext}}$ , a certain nDOF is required to discretise  $u_N^{\text{gfem}}$  sufficiently accurate in  $\Omega^{\text{air}}$  and  $\Omega^{\text{cr}}$ . But these nDOFs are constant for all  $n_p$ . In particular, if the method is size robust for one exterior discretisation it is size robust for all exterior discretisations. Thus it is sufficient to analyse the multiscale basis for a single exterior discretisation.

**Remark 6.** An alternative approach to create the hybrid space  $V_{k_1}^{\text{gfem}}$  is to use a non-conforming method where  $V_{k_1}^{\text{gfem}}$  is no longer a subspace of  $H_{k_1}^1(\Omega)$  and to use mortar elements [8] near the boundary  $\Gamma^{\text{over}}$ . For mortar elements the variational formulation (7) changes and terms are introduced which penalise discontinuities on  $\Gamma^{\text{over}}$ . It is harder to analyse such a non-conforming method but it has advantages for implementations as neither  $V_{k_1}^{\text{multi}}$  nor  $V_{k_1}^{\text{ext}}$  need to be altered, while retaining good approximation properties for an appropriate penalisation.

#### 4.3. Hybridisation of multiscale FEM with standard FEM

In this section we describe a concrete realisation of the hybridisation for the case that the exterior discretisation is standard FEM.

*Exterior discretisation with standard FEM.* A possible discretisation for the exterior domain is the standard FEM (e.g., p-FEM) on a mesh  $\mathcal{M}_{\text{ext}}^{\text{mac}}$  of  $\Omega^{\text{ext}} = \Omega^{\text{air}} \cup \Omega^{\text{str}}$  (see Fig. 8(a)). The mesh can be chosen sufficiently fine to resolve the material interfaces in the structured layer  $\Omega^{\text{str}}$  but can contain large cells with high polynomial degrees for an efficient discretisation of the homogeneous parts in  $\Omega^{\text{air}}$ . The  $H_{k_1}^1(\Omega^{\text{ext}})$ -conforming basis functions  $b^{\text{ext}}$  are constructed by combining shape functions to globally continuous and  $\mathbf{a}_1$ -periodic functions on  $\Omega^{\text{ext}}$  and then multiplying them by  $e^{ik_1x_1}$  to fulfil the quasi-periodicity condition of  $H_{k_1}^1(\Omega^{\text{ext}})$ . For the sake of the hybridisation the mesh  $\mathcal{M}_{\text{ext}}^{\text{mac}}$  is assumed to be conforming to  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  so that no *hanging nodes* occur.

*Hybridisation.* It is possible to perform the hybridisation by exclusively enriching both spaces  $V_{k_1}^{\text{multi}}$  and  $V_{k_1}^{\text{ext}}$  (M1) so that no basis function  $b^{\text{multi}}$  or  $b^{\text{ext}}$  need to be dropped. Another approach has been used in the multiscale FEM proposition [64], where only  $V_{k_1}^{\text{ext}}$  was enriched by extending the multiscale functions into the homogeneous exterior domain, but  $V_{k_1}^{\text{multi}}$  was not enriched at all. In contrast we prefer not to enrich the space  $V_{k_1}^{\text{ext}}$  and not to modify the exterior discretisation scheme for implementation reasons. The hybridisation of  $V_{k_1}^{\text{multi}}, V_{k_1}^{\text{ext}}$  into  $\tilde{V}_{k_1}^{\text{multi}}, \tilde{V}_{k_1}^{\text{ext}}$  is performed three steps (see Fig. 8(e))

- all functions  $b^{\text{multi}}$  with support on  $\Gamma^{\text{over}}$  are omitted (M2),
- all functions  $b^{\text{ext}}$  with polynomial degree  $p \leq p_{\text{mac}}$  are extended into  $\Omega^{\text{cr}}$  (M1),
- the functions  $b^{\text{ext}}$  with polynomial degree  $p > p_{\text{mac}}$  are omitted (M2)—in analogy to the minimum rule in the p-FEM.

The extension of the polynomial functions to  $\Omega^{\text{cr}}$  is performed by adding a special micro function  $b_{\star}^{\text{mic}}(\mathbf{x}) = e^{ik_1x_1}$  with periodic part  $b^{\text{per}} \equiv 1$  and wave vector  $\mathbf{k} = (k_1, 0)^\top$  and by matching the periodic part  $b^{\text{ext}}(\mathbf{x})e^{-ik_1x_1}$  of  $b^{\text{ext}}(\mathbf{x})$  with  $b^{\text{mac}}$ . The addition of  $b_{\star}^{\text{mic}}$  is only necessary in the overlapping cells tangent to the interface  $\Gamma^{\text{over}}$  (the cells  $K_2, K_4 \in \mathcal{M}^{\text{mac}}$  in Fig. 8(a)). Additionally, we propose to add  $b_{\star}^{\text{mic}}$  globally on the whole of  $\Omega^{\text{cr}}$ . Then the g-FEM space  $V_{k_1}^{\text{gfem}}$  becomes a true generalisation of the standard FEM as  $V_{k_1}^{\text{gfem}}$  is a superset of a standard FEM space on the mesh  $\mathcal{M}^{\text{mac}}$ .

A disadvantage of this hybridisation is that some multiscale functions  $b^{\text{multi}}$  are dropped. Especially the vertex DOFs on  $\Gamma^{\text{over}}$  are dropped, i.e., these micro functions  $b^{\text{mic}}$  are not part of  $\tilde{V}_{k_1}^{\text{multi}}$  any more. To relax this restriction we propose to use small single-period multiscale cells close to  $\Gamma^{\text{over}}$ , e.g.,  $K_2, K_4$  in Fig. 8(a). Then, the functions  $b^{\text{mic}}$  are at least contained in  $K_3$ . The discretisation error in  $K_2, K_4$  can be selectively controlled by taking higher  $p_{\text{mac}}$  in these cells (enrichment) which does not affect the goal to achieve size robustness as  $K_2$  and  $K_4$  only contain one period of the crystal. The addition of the small cells  $K_2$  and  $K_4$  is also essential to resolve the boundary layer effect discussed in Sec. 3.3.

**Remark 7.** A similar hybridisation procedure could be applied for an exterior discretisation with global plane waves. For this several new micro functions  $b^{\text{mic}}$  (with varying wave vectors  $\mathbf{k}$  and constant periodic part  $b^{\text{per}} \equiv 1$ ) need to be added (M1) and matched with the exterior plane waves. Only the macro functions  $b^{\text{mac}}$  that are linear on  $\Gamma^{\text{over}}$  need to be combined (in the sense of (19)) with the newly added micro functions.

#### 4.4. Well-posedness of the discrete problem

In contrast to elliptic problems (*e.g.* heat conduction), the well-posedness of a Galerkin method for indefinite problems (*e.g.*, for Helmholtz equation (2)) does not directly follow from the well-posedness of the variational formulation (7) on the whole space  $H_{k_1}^1(\Omega)$ , even for conforming discretisations with an  $N$ -dimensional subspace  $V_N \subset H_{\text{per}}^1(\Omega)$ . It is usually not guaranteed that the discrete problem (17) has a unique solution  $u_N^{\text{gfem}}$ , especially when the discretisation is very coarse. But at least uniqueness can often be shown for a sufficiently fine discretisation. This also applies in the present case of g-FEM spaces  $V_{p_{\text{mac}}}^{\text{gfem}}$  and predicts a unique solution when the polynomial degree  $p_{\text{mac}}$  exceeds some value  $p_{\text{mac}}^* \in \mathbb{N}$ . Here, we used the g-FEM space  $V_{p_{\text{mac}}}^{\text{gfem}}$  with macro polynomial degree  $p_{\text{mac}}$  and an exterior standard FEM discretisation with  $p_{\text{ext}} \geq p_{\text{mac}}$ .

**Proposition 8** (Well-posedness of the Multiscale FEM discretisation). *Given an infinite sequence  $(V_{p_{\text{mac}}}^{\text{gfem}})_{p_{\text{mac}}=1}^\infty$  of discrete g-FEM spaces  $V_{p_{\text{mac}}}^{\text{gfem}} \subset H_{k_1}^1(\Omega)$  with macroscopic polynomial degree  $p_{\text{mac}}$ , then there exists some  $p_{\text{mac}}^* \in \mathbb{N}$  so that (17) has a unique solution  $u_{k_1}^{(p_{\text{mac}})}$  in  $V_{p_{\text{mac}}}^{\text{gfem}}$  for  $p_{\text{mac}} \geq p_{\text{mac}}^*$ . In this case, the quasi-optimal error estimate*

$$\|u_{k_1} - u_{k_1}^{(p_{\text{mac}})}\|_{H^1(\Omega)} \leq C \min_{v \in V_{p_{\text{mac}}}^{\text{gfem}}} \|u_{k_1} - v\|_{H^1(\Omega)} \quad (21)$$

holds, where  $C > 0$  is a  $p_{\text{mac}}$ -independent constant and  $u_{k_1}$  is the solution of (7).

*Proof.* This proposition is a direct application of [66, Satz 4.2.9], we check here its prerequisites. The sesquilinear form  $\Phi(\cdot, \cdot)$  is the sum of a  $H_{k_1}^1(\Omega)$ -coercive sesquilinear form  $a(u, v) = \int_{\Omega} a(\mathbf{x}) \nabla u(\mathbf{x}) \cdot \nabla \bar{v}(\mathbf{x}) \, d\mathbf{x}$  and a compact perturbation  $t(u, v) = \Phi(u, v) - a(u, v)$ . The variational formulation has a unique solution  $u_{k_1}$  (7) as shown in Prop. 3 and the only solution with zero right hand side is  $u_{k_1} \equiv 0$ . Finally, we note that  $(V_{p_{\text{mac}}}^{\text{gfem}})_{p_{\text{mac}}}$  is a dense sequence of finite dimensional subspaces of  $H_{k_1}^1(\Omega)$ .  $\square$

Note, that the proposition applies also to p-FEM spaces  $V_p$  on the mesh  $\mathcal{M}^{\text{mac}}$  with uniform polynomial degree  $p = p_{\text{mac}}$  as no special information of the functions  $b^{\text{mic}}$  has been used. It can also be extended to g-FEM spaces with a non-uniform polynomial degree distribution, where the minimal polynomial degree has to exceed  $p_{\text{mac}}^*$ .

The underlying proof in [66, Satz 4.2.9] does neither determine  $p_{\text{mac}}^*$  nor on which parameters it depends. We cannot exclude, that it even depends on  $n_p$ . However, in all our numerical experiments we observed  $p_{\text{mac}}^* \leq 2$  for both p-FEM and MSFEM, and that  $p_{\text{mac}}^*$  did not depend on  $n_p$  at all. In numerical experiments, the non-uniqueness would express itself in (almost) singular matrices or in an error for a particular  $p_{\text{mac}}$  which is larger than the error for  $p_{\text{mac}} - 1$ . The increase in the error is an artifact of the discretisation, that may occur if the quasi-optimality of FEM does not yet hold. We observed only an increase in the error when increasing  $p_{\text{mac}}$  from 1 to 2. This observation leads us to the following conjecture.

**Conjecture 9** (Size-independent stability). *Let the periodic function 1 with wave-vector  $\mathbf{k} = (k_1, 0)^\top$  and a finite set of Bloch modes at least containing the dominant Bloch modes  $\mathcal{B}_{k_1}^{\text{dom}}$  be included in the micro space  $V_{k_1}^{\text{mic}}$  of the g-FEM space  $V_{p_{\text{mac}}}^{\text{gfem}}$ . Then, the minimal polynomial degree  $p_{\text{mac}}^*$  for uniqueness (see Prop. 8) does not depend on  $n_p$ .*

**Remark 10.** *With Proposition 8 the well-posedness is guaranteed for sufficiently refined MSFEM spaces, if no guided mode is present at the investigated frequency  $\omega$  and investigated wavenumber  $k_1$ , which was already assumed in Sec. 2.3. Nevertheless, the required resolution  $p_{\text{mac}}^*$  in Prop. 8 for well-posedness might depend on the distance of the wavenumber  $k_1$  to the wavenumber  $k_1^g$  of a guided mode  $u^g$  and (more generally) on eigenmodes of the continuous problem for frequencies in a complex neighbourhood of the simulation frequency  $\omega \in \mathbb{R}$ . In such resonance situations usually very high computational effort (high polynomial degrees) is needed to obtain a stable discretisation. Near resonance situations are usually indicated by high condition numbers.*

## 5. Computation of matrix entries by fast quadrature

This section is concerned with the linear system that has to be solved to determine the discrete solution  $u_N$  in the hybrid space  $V_{k_1}^{\text{gfem}}$ . The linear system will be introduced in Sec. 5.1. The main difficulty is the fast evaluation of integrals of the two-scale basis function  $b^{\text{multi}}$  in the macro cells  $K^{\text{mac}} \in \mathcal{M}_{\text{cr}}^{\text{mac}}$  of the crystal. The integration rule has to take the microscopic structure  $b^{\text{mic}}$  of  $b^{\text{multi}}$  into account. A simple integration rule that satisfies this requirement, the size dependent quadrature, is introduced in Sec. 5.2. The computational effort of the size dependent quadrature scales linearly with the number of periods  $n_p$ . In the remainder of this section we introduce an efficient integration rule, the *fast quadrature*, which keeps the accuracy of the size dependent quadrature, but which is size robust in terms of computational time, *i.e.*, independent of  $n_p$ .

This algorithm is not limited to the here proposed multiscale FEM for the PhC bands in two space dimensions with  $1 \times n$ -period macro cells. Rather it applies to any space dimension  $d$  and parallelepiped macro cells with  $n_1 \times n_2 \times \dots \times n_d$ -periods for some number of periods  $n_1, \dots, n_d \in \mathbb{N}$ . On such macro cells the the algorithm can numerically integrate two-scale functions which are products of  $\widehat{\Omega}$ -quasi-periodic micro functions and polynomial macro functions. We will introduce the fast quadrature exemplary for a 2-dimensional finite crystal with  $n_{p,1} \times n_{p,2}$  periods. Furthermore, we will optimise the algorithm with respect to  $p_{\text{mac}}$  and give in Sec. 5.5 an explicit algorithm for fast quadrature whose computational effort is equivalent to that of  $p$ -FEM on the coarse mesh  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  with polynomial degree  $p = p_{\text{mac}}$ . This means the assembly of system matrices for the two-scale basis functions  $b^{\text{multi}}$  with their highly oscillatory microscopic parts is only as expensive as the integration of only the much smoother macroscopic part, *i.e.*, two-scale integration with fast quadrature is only as expensive as one-scale integration.

### 5.1. Computation of matrix entries

To solve (17) for the discrete solution  $u_N^{\text{gfem}}$  a linear system  $\mathbf{A}\boldsymbol{\alpha} = \mathbf{f}$  is assembled and solved for the coefficient vector  $\boldsymbol{\alpha} = (\alpha_{\ell_1}^{\text{gfem}})_{\ell_1=1}^N$ . Here, the system matrix and the vector of the right hand side are given by

$$\mathbf{A} = \left( \Phi_{k_1}(b_{\ell_1}^{\text{gfem}}, b_{\ell_2}^{\text{gfem}}) \right)_{\ell_1, \ell_2=1}^N, \quad \mathbf{f} = \left( f(b_{\ell_2}^{\text{gfem}}) \right)_{\ell_2=1}^N, \quad (22)$$

where  $\Phi_{k_1}$  is the sesquilinear form defined in (7b) and  $f$  the linear form defined in (7c).

The linear form (7c) involves only traces on the interface  $\Gamma^{\text{nref}}$  which is not adjacent to any cells  $K^{\text{mac}} \in \mathcal{M}_{\text{cr}}^{\text{mac}}$  so  $\mathbf{f}$  can be evaluated independent of  $n_p$ . Thus, we can restrict us in the following to the sesquilinear form  $\Phi_{k_1}$ .

The sesquilinear form  $\Phi_{k_1}$ , defined in (7b), can be split into contributions for each macro cell  $K^{\text{mac}} \in \mathcal{M}^{\text{mac}}$  by restricting the domain of integration from  $\Omega$  (resp.  $\Gamma^{\text{nref}}$ ) to  $K^{\text{mac}}$  (resp.  $\Gamma^{\text{nref}} \cap K^{\text{mac}}$ ) and the matrix entries are

$$\mathbf{A}_{\ell_2 \ell_1} = \Phi_{k_1}(b_{\ell_1}^{\text{gfem}}, b_{\ell_2}^{\text{gfem}}) = \sum_{K^{\text{mac}} \in \mathcal{M}^{\text{mac}}} \Phi_{K^{\text{mac}}}(b_{\ell_1}^{\text{gfem}}, b_{\ell_2}^{\text{gfem}}). \quad (23)$$

Clearly, the evaluation of  $\Phi_{K^{\text{mac}}}$  for exterior macro cells  $K^{\text{mac}} \in \mathcal{M}_{\text{ext}}^{\text{mac}}$  does not depend on the number of periods  $n_p$ . For  $K^{\text{mac}} \in \mathcal{M}_{\text{cr}}^{\text{mac}}$  the functions  $b^{\text{gfem}}$  are two-scale functions  $b^{\text{multi}}$ , as defined in (19), and we have to evaluate the sesquilinear forms

$$\Phi_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}}) := \mathbf{a}_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}}) + \mathbf{b}_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}}), \quad (24a)$$

with

$$\mathbf{a}_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}}) := \int_{K^{\text{mac}}} a(\mathbf{x}) \nabla b_i^{\text{multi}}(\mathbf{x}) \cdot \nabla \overline{b_j^{\text{multi}}}(\mathbf{x}) \, d\mathbf{x}, \quad (24b)$$

$$\mathbf{b}_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}}) := \int_{K^{\text{mac}}} b^2(\mathbf{x}, \omega) b_i^{\text{multi}}(\mathbf{x}) \overline{b_j^{\text{multi}}}(\mathbf{x}) \, d\mathbf{x}. \quad (24c)$$

See Fig. 9 for the index convention of  $b^{\text{multi}}$ .

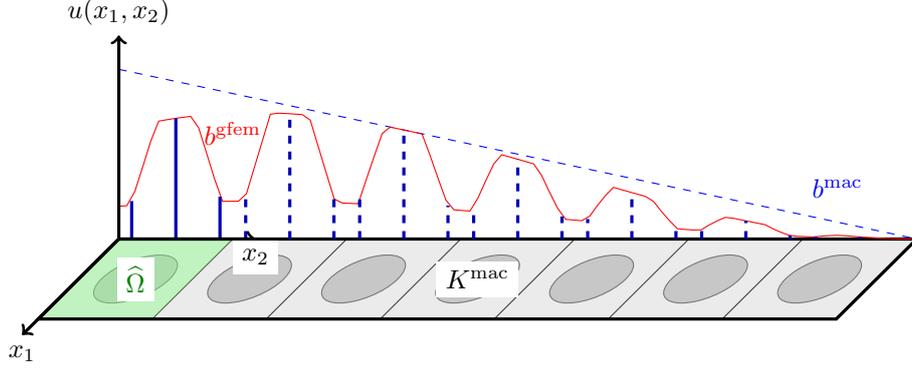


Figure 10: Sketch of a macro cell  $K^{\text{mac}}$  over which the sesquilinear forms need to be evaluated. The basis function  $b^{\text{multi}}$  oscillates on the length scale of one period. To resolve the oscillations the quadrature points from the unit cell  $\hat{\Omega}$  are used, here  $n_{\text{quad}} = 3$  Gauss-Legendre points (solid lines). The computational time depends only on  $n_{\text{quad}}$  but the approximation quality is as good as replicating the  $n_{\text{quad}}$  points on all  $n_p$  periods (dashed lines).

### 5.2. A size dependent quadrature

This section explains a size dependent integration rule for periodic structures with  $n_{p,1} \times n_{p,2}$ -periods. According to the discussion in the previous section (5.1) it is sufficient to describe a rule for one multiscale cell  $K^{\text{mac}} \in \mathcal{M}_{\text{cr}}^{\text{mac}}$  with  $n_1 \times n_2$  periods. For the PhC band only the special case of a  $1 \times n_{p,2}$ -period crystal and a  $1 \times n_2$ -period multiscale cell  $K^{\text{mac}} \in \mathcal{M}_{\text{cr}}^{\text{mac}}$  will be required. The integration rule works on two-scale functions  $b^{\text{multi}}$  and does not only have to resolve the macroscopic-part, the polynomials  $b^{\text{mac}}$ , but also the microscopic part, the functions  $b^{\text{mic}}$ , which oscillate on the level of a single unit cell  $\hat{\Omega}$ . So in general, as the number of periods  $n = n_1 n_2$  contained in  $K^{\text{mac}}$  grows, more quadrature points are needed to integrate up to machine precision. One way to compute the integrals in  $\Phi_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}})$  is to choose an integration rule on the unit cell  $\hat{\Omega}$  and replicate the quadrature points on each repetition of  $\hat{\Omega}$  contained in  $K^{\text{mac}}$  (see Fig. 10). The integration of the unit cell can use the microscopic mesh  $\mathcal{M}^{\text{mic}}$  (see Fig. 8(d)) and, e.g., Gauss quadrature points on the reference element of each micro cell  $K^{\text{mic}} \in \mathcal{M}^{\text{mic}}$ . For a micro function  $b^{\text{mic}}$  a constant number of points is needed per unit cell  $\hat{\Omega}$ , as  $\hat{\Omega}$  is the size of the smallest oscillation of  $b^{\text{mic}}$ . So after replication  $O(n)$  points are needed. To integrate only the much smoother macro function  $b^{\text{mac}}$  only  $O(p_{\text{mac}}^2)$  points on  $K^{\text{mac}}$  are required. For a multiscale function  $b^{\text{multi}}$  the integration rules have to be combined and the total number of points depends on  $p_{\text{mac}}^2$  and  $n$ . We propose to use  $O(p_{\text{mac}}^2)$  points per unit cell which is sufficient to evaluate  $\Phi_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}})$  up any precision. Such a quadrature with  $\hat{\Omega}$ -periodic points we call an  $\hat{\Omega}$ -periodic quadrature.

We note that the  $O(p_{\text{mac}}^2)$  points per unit cell are a worst-case estimate and for large number of periods fewer points might be sufficient. Moreover, a general size dependent quadrature with non- $\hat{\Omega}$ -periodic quadrature points would still require  $O(n)$  points and would therefore be asymptotically equivalent to the  $\hat{\Omega}$ -periodic one.

### 5.3. An algorithm for fast quadrature of two-scale functions with quasi-periodic micro structure

In this section we will show how to approximate the integrals to any accuracy, but with computational complexity independent of the number of periods  $n$  of a macro cell  $K^{\text{mac}} \in \mathcal{M}_{\text{cr}}^{\text{mac}}$  inside the crystal. The obtained accuracy is up to round-off errors equivalent to the size dependent quadrature presented in the previous section. It will take advantage of the structure in the system matrix  $\mathbf{A}$  induced by (i) the multiscale functions  $b^{\text{multi}}$ , (ii) the periodicity of the dielectric pattern and so of  $a(\mathbf{x})$  and  $b^2(\mathbf{x})$ , (iii) the quasi-periodicity of the micro functions  $b^{\text{mic}}$  and (iv) the fact that the macro functions are polynomials. The presented approach is an extension of [53, 65], where multiscale functions exhibit a tensor product structure with purely periodic micro functions.

The remainder of this section is organised as follows. First, we state Lemma 12, which allows to rearrange the terms occurring in  $\Phi_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}})$  so that instead of an integral on  $K^{\text{mac}}$  only a weighted sum of  $n$ -independent integrals on  $\widehat{\Omega}$  needs to be computed. The weights incorporate some sums  $S^{(r)} = \sum_{\ell=0}^{n_r} a^\ell \ell^k$ ,  $r = 1, 2$  which can be evaluated by Lemma 13 with computational costs independent of  $n$ . The main result of this section, the size robust assembling of the system matrix  $\mathbf{A}$ , is then given in Theorem 15. All proofs are constructive and thus the resulting formulas describe an algorithm for fast quadrature. The resulting optimised algorithm will be presented in Sec. 5.5.

**Remark 11.** *In this section we will use that powers  $a^n$  for  $a \in \mathbb{C}$  and  $n \in \mathbb{N}$  can be computed with  $O(1)$  computational time up to machine precision.*

**Lemma 12.** *Consider a macro cell  $K^{\text{mac}} \subset \Omega^{\text{cr}}$  with  $n_1 \times n_2$ -periods, an  $\mathbf{a}_1$ - and  $\mathbf{a}_2$ -periodic function  $f(\mathbf{x})$  (e.g.  $a(\mathbf{x})$  or  $b^2(\mathbf{x})$ ), the two-scale functions  $b_{i_1, i_2}^{\text{multi}}$  and  $b_{j_1, j_2}^{\text{multi}}$  with wave vectors  $\mathbf{k}_{i_2}, \mathbf{k}_{j_2} \in \mathbb{C}^2$  (in the form of (19)), a component selector  $\eta \in \{1, 2\}$  and some derivative selectors  $\mathbf{d}_i, \mathbf{d}_j$  with  $d_{i_1}, d_{i_2}, d_{j_1}, d_{j_2} \in \{0, 1\}$ . Then the integral*

$$T = \int_{K^{\text{mac}}} f(\mathbf{x}) D_\eta^{d_{i_1}} b_{i_1}^{\text{mac}}(\mathbf{x}) D_\eta^{d_{j_1}} b_{j_1}^{\text{mac}}(\mathbf{x}) D_\eta^{d_{i_2}} (b_{i_2}^{\text{per}}(\mathbf{x}) e^{i\mathbf{k}_{i_2} \cdot \mathbf{x}}) D_\eta^{d_{j_2}} (b_{j_2}^{\text{per}}(\mathbf{x}) e^{-i\bar{\mathbf{k}}_{j_2} \cdot \mathbf{x}}) d\mathbf{x}, \quad (25)$$

with the differential operator

$$D_\eta^{d_i} u := \begin{cases} u & \text{for } d_i = 0, \quad \text{independent of } \eta \\ \partial_{x_\eta} u & \text{for } d_i = 1, \end{cases}$$

can be rewritten as

$$T = \sum_{i \in \mathcal{I}} \alpha_i S_i^{(1)} S_i^{(2)} Q_i, \quad (26)$$

in which  $\mathcal{I}$  is a finite index set independent of  $n_1$  and  $n_2$ ;  $\alpha_i \in \mathbb{C}$  are some coefficients; the two sums  $S_i^{(r)}$ ,  $r = 1, 2$  take the form  $S_i^{(r)} = S_{k, n_r}(a) = \sum_{\ell=0}^{n_r-1} a^\ell \ell^k$  for some  $a \in \mathbb{C}$ ,  $k \in \mathbb{N}_0$ ; and the terms  $Q_i$  are integrals over  $\widehat{\Omega}$ , which do not depend on  $n_1, n_2$ .

**Lemma 13.** *For any  $n \in \mathbb{N}$ ,  $a \in \mathbb{C}$ ,  $k \in \mathbb{N}_0$  the sum*

$$S_{k, n}(a) = \sum_{\ell=0}^{n-1} a^\ell \ell^k \quad (27)$$

can be evaluated with  $O(k^2)$  arithmetic operations, where the constant in  $O(k^2)$  is independent of  $n$ .

The precise formulas for  $S_{k, n}(a)$  in (27) and  $T$  in (25) will be given in the proofs of the respective lemmas, see particularly (31) for  $T$  and (33), (34), (35) for  $S_{k, n}(a)$ . With these lemmas, we are in the position to give the main result of a size robust assembly time of each entry of the element matrices  $\Phi_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}})$  and thus of the whole linear system.

**Proposition 14** (Scale independent complexity of element matrix assembly). *Let  $K^{\text{mac}}$  be a macro cell in  $\Omega^{\text{cr}}$  with  $n_1 \times n_2$  periods. Then, the entries of the element matrix  $\Phi_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}})$  defined in (24a) can be computed up to machine precision with complexity independent of  $n_1$  and  $n_2$ .*

*Proof.* We split the terms  $\mathbf{a}_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}})$  and  $\mathbf{b}_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}})$  of  $\Phi_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}})$  into single terms  $T$  of the form (25). To split  $\mathbf{a}_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}})$  the chain rule is applied and the dot product is decomposed as a sum of two component-wise multiplications. This leads to eight terms  $T$  of the form (25) for  $\mathbf{a}_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}})$  and one for  $\mathbf{b}_{K^{\text{mac}}}(b_i^{\text{multi}}, b_j^{\text{multi}})$ . Thus the assumptions of Lemma 12 are satisfied. The application of Lemmas 12 and 13 completes the proof.  $\square$

**Corollary 15** (Scale independent complexity of the assembly of the linear system). *Given a crystal with  $n_{p_1} \times n_{p_2}$  periods and macroscopic mesh  $\mathcal{M}^{\text{mac}}$  consisting of  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  in the domain of the crystal  $\Omega^{\text{cr}}$  and  $\mathcal{M}_{\text{ext}}^{\text{mac}}$  in the remaining exterior domain  $\Omega^{\text{ext}}$  of  $\Omega$ . Let  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  consisting of macro cells  $K^{\text{mac}}$  with  $n_1(K^{\text{mac}}) \times n_2(K^{\text{mac}})$  periods where the number of macro cells is independent of  $n_{p_1}$  and  $n_{p_2}$ . Then, the system matrix  $\mathbf{A}$  given in (22), can be evaluated accurately up to machine precision with complexity independent of the number of periods  $n_{p,1}$ ,  $n_{p,2}$ .*

*Proof.* The corollary follows directly from Theorem 14 and the derivations and discussion in Sec. 5.1.  $\square$

**Remark 16.** *The approach also works for 3D finite crystals and  $n_1 \times n_2 \times n_3$ -period macro cells  $K^{\text{mac}}$ . In this case instead of the TE and TM modes (1) we have the full vector valued Maxwell's equations and this changes the  $\nabla$  in  $\mathbf{a}(\cdot, \cdot)$  into **curl** operators and the basis functions  $b^{\text{multi}}$  are vector valued. However, we can still write  $\mathbf{a}_{K^{\text{mac}}}(\cdot, \cdot)$  as a sum of the terms (25).*

#### 5.4. Derivation of the fast quadrature

This section is devoted to the proofs of Lemmas 12 and 13.

*Proof of Lemma 12.* The proof consists in the decomposition of  $T$  into a sum of integrals  $Q_i$  over the unit cell  $\tilde{\Omega}$  only, each multiplied with a weight. These weights will be derived as products of two separate sums  $S_i^{(1)}$  and  $S_i^{(2)}$  only depending on  $n_1$  or  $n_2$ , respectively.

To achieve this separation of  $S_i^{(1)}$  and  $S_i^{(2)}$  we transform the parallelogram-shaped macro cell  $K^{\text{mac}}$  onto the rectangular cell  $\tilde{K} = [0, n_1] \times [0, n_2]$  by an affine transformation. On  $\tilde{K}$  we perform a monomial expansion of  $b^{\text{mac}}$  and apply the binomial theorem.

The affine mapping  $F_{\mathbf{c}} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$

$$F_{\mathbf{c}}(\tilde{\mathbf{x}}) := F_0(\tilde{\mathbf{x}}) + \mathbf{c}, \quad F_0(\tilde{\mathbf{x}}) := \mathbf{a}_1 \tilde{x}_1 + \mathbf{a}_2 \tilde{x}_2, \quad (28)$$

transforms  $\tilde{K}$  onto the parallelogram-shaped macro cell  $K^{\text{mac}}$  with the corners  $\mathbf{c}$ ,  $\mathbf{c} + n_1 \mathbf{a}_1$ ,  $\mathbf{c} + n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2$  and  $\mathbf{c} + n_2 \mathbf{a}_2$ . The unit cell of the transformed cell  $\tilde{K}$  is the *standard unit cell*  $\tilde{\Omega} := [0, 1]^2 = F_0^{-1}(\tilde{\Omega})$ . The determinant of the Jacobian of  $F_{\mathbf{c}}$  or  $F_0$  is  $|\tilde{\Omega}|$ .

Using the transformation  $F_{\mathbf{c}}$  we represent  $T$  as an integral over  $\tilde{K}$  and decompose the integral as sums over translated standard unit cells  $\tilde{\Omega}$

$$\begin{aligned} T &= |\tilde{\Omega}| \int_{\tilde{K}} f(F_{\mathbf{c}} \tilde{\mathbf{x}}) (D_{\eta}^{di_1} b_{i_1}^{\text{mac}})(F_{\mathbf{c}} \tilde{\mathbf{x}}) (D_{\eta}^{dj_1} b_{j_1}^{\text{mac}})(F_{\mathbf{c}} \tilde{\mathbf{x}}) (D_{\eta}^{di_2} b_{i_2}^{\text{mic}})(F_{\mathbf{c}} \tilde{\mathbf{x}}) (D_{\eta}^{dj_2} \overline{b_{j_2}^{\text{mic}}})(F_{\mathbf{c}} \tilde{\mathbf{x}}) d\tilde{\mathbf{x}} \\ &= |\tilde{\Omega}| \sum_{\boldsymbol{\ell} \in \mathcal{A}} \int_{\tilde{\Omega}} f(F_{\mathbf{c}} \tilde{\mathbf{x}}) (D_{\eta}^{di_1} b_{i_1}^{\text{mac}})(F_{\mathbf{c}}(\tilde{\mathbf{x}} + \boldsymbol{\ell})) (D_{\eta}^{dj_1} b_{j_1}^{\text{mac}})(F_{\mathbf{c}}(\tilde{\mathbf{x}} + \boldsymbol{\ell})) \\ &\quad (D_{\eta}^{di_2} b_{i_2}^{\text{mic}})(F_{\mathbf{c}}(\tilde{\mathbf{x}} + \boldsymbol{\ell})) (D_{\eta}^{dj_2} \overline{b_{j_2}^{\text{mic}}})(F_{\mathbf{c}}(\tilde{\mathbf{x}} + \boldsymbol{\ell})) d\tilde{\mathbf{x}} \end{aligned} \quad (29)$$

where we use the offset vector  $\boldsymbol{\ell}$  and the index set  $\mathcal{A} := [0, n_1] \times [0, n_2] \subset \mathbb{N}^2$ .

The macro functions  $b^{\text{mac}}$  are constructed as polynomials of maximal degrees  $p_1$  and  $p_2$  on a reference cell  $\hat{K} = [0, 1]^2$ . They remain polynomials (of the same maximal degrees) on  $K^{\text{mac}}$  and on  $\tilde{K}$  as the mappings  $F_{K^{\text{mac}}} : \hat{K} \rightarrow K^{\text{mac}}$  and  $F_{\mathbf{c}} : K^{\text{mac}} \rightarrow \tilde{K}$  are affine. The same holds for the derivatives of a macro function  $b^{\text{mac}}$ . Let the monomial decomposition of  $(D_{\eta}^{di_1} b_{i_1}^{\text{mac}})(F_{\mathbf{c}} \tilde{\mathbf{x}})$  for some derivative selector  $di_1$ , component selector  $\eta$  and macro index  $i_1$  be given by

$$(D_{\eta}^{di_1} b_{i_1}^{\text{mac}})(F_{\mathbf{c}} \tilde{\mathbf{x}}) = \sum_{\boldsymbol{\nu} \in P} \alpha_{\boldsymbol{\nu}}^{(i_1, di_1, \eta)} \tilde{\mathbf{x}}^{\boldsymbol{\nu}}. \quad (30)$$

Here  $\alpha_{\boldsymbol{\nu}}^{(i_1, di_1, \eta)} \in \mathbb{C}$  are the monomial coefficients,  $P = [0, p_1] \times [0, p_2]$  is the set of polynomials degrees up to orders  $p_1$ ,  $p_2$  and  $\boldsymbol{\nu}$  is the multi-index notation for  $x_1^{\nu_1} x_2^{\nu_2}$ .

The micro functions  $(b_{i_2}^{\text{mic}})(F_c \tilde{\mathbf{x}})$  and their derivatives  $(D_\eta^{di_2} b_{i_2}^{\text{mic}})(F_c \tilde{\mathbf{x}})$  are  $\mathbf{e}_1$ - and  $\mathbf{e}_2$ -quasi-periodic which enables us to write

$$(D_\eta^{di_2} b_{i_2}^{\text{mic}})(F_c(\tilde{\mathbf{x}} + \boldsymbol{\ell})) = (D_\eta^{di_2} b_{i_2}^{\text{mic}})(F_c \tilde{\mathbf{x}} + F_0 \boldsymbol{\ell}) = e^{i \mathbf{k}_{i_2} \cdot F_0 \boldsymbol{\ell}} (D_\eta^{di_2} b_{i_2}^{\text{mic}})(F_c \tilde{\mathbf{x}}).$$

Inserting this equality and the monomial decomposition (30) of the macro functions  $b^{\text{mac}}$  into (29) we get

$$T = |\widehat{\Omega}| \sum_{\nu \in P} \sum_{\mu \in P} \alpha_\nu^{(i_1, di_1, \eta)} \alpha_\mu^{(j_1, dj_1, \eta)} \sum_{\boldsymbol{\ell} \in \mathcal{A}} e^{i(\mathbf{k}_{i_2} - \bar{\mathbf{k}}_{j_2}) \cdot F_0 \boldsymbol{\ell}} \int_{\widehat{\Omega}} f(F_c \tilde{\mathbf{x}}) (\tilde{\mathbf{x}} + \boldsymbol{\ell})^{\nu + \mu} (D_\eta^{di_2} b_{i_2}^{\text{mic}})(F_c \tilde{\mathbf{x}}) (D_\eta^{dj_2} \overline{b_{j_2}^{\text{mic}}})(F_c \tilde{\mathbf{x}}) d\tilde{\mathbf{x}}.$$

Now, we apply the binomial theorem to  $(\tilde{\mathbf{x}} + \boldsymbol{\ell})^{\nu + \mu}$ , transform  $\tilde{\Omega}$  back to  $\widehat{\Omega}$  and obtain

$$T = \sum_{\nu \in P} \sum_{\mu \in P} \alpha_\nu^{(i_1, di_1, \eta)} \alpha_\mu^{(j_1, dj_1, \eta)} \sum_{\mathbf{m}=\mathbf{0}}^{\nu + \mu} \binom{\nu_1 + \mu_1}{m_1} \binom{\nu_2 + \mu_2}{m_2} S_{\mathcal{A}, i_2, j_2}^{\nu + \mu - \mathbf{m}} Q_{f, i_2, j_2}^{\mathbf{m}, di_2, dj_2} \quad (31)$$

with

$$S_{\mathcal{A}, i_2, j_2}^{\boldsymbol{\kappa}} := \sum_{\boldsymbol{\ell} \in \mathcal{A}} e^{i(\mathbf{k}_{i_2} - \bar{\mathbf{k}}_{j_2}) \cdot F_0 \boldsymbol{\ell}} \boldsymbol{\ell}^{\boldsymbol{\kappa}}, \quad Q_{f, i_2, j_2}^{\mathbf{m}, di_2, dj_2} := \int_{\widehat{\Omega}} f(\mathbf{x}) (F_0^{-1} \mathbf{x})^{\mathbf{m}} D_\eta^{di_2} b_{i_2}^{\text{mic}}(\mathbf{x}) D_\eta^{dj_2} \overline{b_{j_2}^{\text{mic}}}(\mathbf{x}) d\mathbf{x}.$$

The index set  $\mathcal{I}$  in (26) stems from the three multi-index sums in (31) and  $Q_i$  corresponds to  $Q_{f, i_2, j_2}^{\mathbf{m}, di_2, dj_2}$  which is an integral over  $\widehat{\Omega}$  of micro functions or their derivatives with the transformed monomials  $(F_0^{-1} \mathbf{x})^{\mathbf{m}}$  of bounded polynomial degree  $\mathbf{m} \in P$ .

It is left to show that the term  $S_{\mathcal{A}, i_2, j_2}^{\boldsymbol{\kappa}}$  can be decomposed into  $S_i^{(1)}$  and  $S_i^{(2)}$ . Inserting the definition of the vector  $\boldsymbol{\ell} = \ell_1 \mathbf{e}_1 + \ell_2 \mathbf{e}_2$  and of  $\boldsymbol{\ell}^{\boldsymbol{\kappa}} = \ell_1^{\kappa_1} \ell_2^{\kappa_2}$  the summands separate in products of terms in  $\ell_1$  and  $\ell_2$  and with the definition (27) we obtain

$$S_{\mathcal{A}, i_2, j_2}^{\boldsymbol{\kappa}} = \sum_{\boldsymbol{\ell} \in \mathcal{A}} \left( e^{-i \ell_1 (\mathbf{k}_{i_2} - \bar{\mathbf{k}}_{j_2}) \cdot F_0 \mathbf{e}_1} \ell_1^{\kappa_1} \right) \left( e^{-i \ell_2 (\mathbf{k}_{i_2} - \bar{\mathbf{k}}_{j_2}) \cdot F_0 \mathbf{e}_2} \ell_2^{\kappa_2} \right) = S_{\kappa_1, n_1}(a_1) S_{\kappa_2, n_2}(a_2)$$

with  $a_r = e^{-i(\mathbf{k}_{i_2} - \bar{\mathbf{k}}_{j_2}) \cdot F_0 \mathbf{e}_r}$  for  $r = 1, 2$ , and the proof is complete.  $\square$

Note, that the separation of  $S_i^{(1)}$  and  $S_i^{(2)}$  at the end of the proof would not have been possible if the monomial expansion (30) would have been performed on the unit cell  $\widehat{\Omega}$  with the periodicity directions  $\mathbf{a}_1$  and  $\mathbf{a}_2$  and an offset vector  $\tilde{\mathbf{x}}_{\text{off}}(\boldsymbol{\ell}) = \ell_1 \mathbf{a}_1 + \ell_2 \mathbf{a}_2$  instead of  $\boldsymbol{\ell}$ .

For the optimised algorithm in Sec. 5.5 it is important to remark, that the tensor product structure of the shape functions  $s_i^{\text{mac}}$  (cf. Sec. 4.1) allows to express the coefficients  $\alpha_\nu^{(i_1, di_1, \eta)}$  in (30) in terms of the coefficients  $\alpha_\nu^{(i)}$  of the monomial expansion of the shape function  $s_i^{\text{mac}}(\hat{\mathbf{x}}) = \sum_{\nu=0}^p \alpha_\nu^{(i)} \hat{\mathbf{x}}^\nu$ . For  $di_1 = 0$  the coefficients  $\alpha_\nu^{(i_1, 0, \eta)}$  are given by

$$\alpha_\nu^{(i_1, 0, \eta)} = \left( \left( \frac{1}{n_1} \right)^{\nu_1} \alpha_{\nu_1}^{(i_1)} \right) \left( \left( \frac{1}{n_2} \right)^{\nu_2} \alpha_{\nu_2}^{(i_2)} \right) =: \alpha_{\nu_1}^{(i_1, n_1)} \alpha_{\nu_2}^{(i_2, n_2)} \quad (32a)$$

which follows from the identity

$$b_{i_1}^{\text{mac}}(F_c \tilde{\mathbf{x}}) = s_{i_1}^{\text{mac}}\left(\frac{\tilde{\mathbf{x}}_1}{n_1}\right) s_{i_2}^{\text{mac}}\left(\frac{\tilde{\mathbf{x}}_2}{n_2}\right). \quad (32b)$$

Similarly for  $di_1 = 1$  we derive expressions for the coefficients of derivatives of  $b^{\text{mac}}$ , which are derivatives with respect to the coordinate  $\mathbf{x} \in K^{\text{mac}}$ . Using the transformation of the gradient we obtain

$$\begin{aligned} (D_\eta^1 b_{i_1}^{\text{mac}})(F_c \tilde{\mathbf{x}}) &= (\nabla_{\mathbf{x}, \eta} b_{i_1}^{\text{mac}})(F_c \tilde{\mathbf{x}}) = (\tilde{J}_{\cdot, \eta}^{-\top} \nabla_{\tilde{\mathbf{x}}} b_{i_1}^{\text{mac}})(F_c \tilde{\mathbf{x}}) \\ &= (\tilde{J}_{1, \eta}^{-\top} \nabla_{\tilde{\mathbf{x}}, 1} b_{i_1}^{\text{mac}})(F_c \tilde{\mathbf{x}}) + (\tilde{J}_{2, \eta}^{-\top} \nabla_{\tilde{\mathbf{x}}, 2} b_{i_1}^{\text{mac}})(F_c \tilde{\mathbf{x}}), \end{aligned}$$

where  $\tilde{J}_{\cdot,\eta}^{-\top}$  is the  $\eta$ -th column of the transpose inverse of the Jacobian  $\tilde{J}$ . In view of (32b) we can now express the coefficient  $\alpha_{\nu}^{(i_1,1,\eta)}$  as

$$\alpha_{\nu}^{(i_1,1,\eta)} = \tilde{J}_{1,\eta}^{-\top}(\nu_1 + 1)\alpha_{\nu_1+1}^{(\hat{i}_1,n_1)}\alpha_{\nu_2}^{(\hat{i}_2,n_2)} + \tilde{J}_{2,\eta}^{-\top}(\nu_2 + 1)\alpha_{\nu_1}^{(\hat{i}_1,n_1)}\alpha_{\nu_2+1}^{(\hat{i}_2,n_2)}. \quad (32c)$$

The entries of  $\tilde{J}^{-\top}$  can be given by the simple formula

$$\tilde{J}_{ij}^{-\top} = \frac{1}{|\hat{\Omega}|} (-1)^{i+j} a_{3-i,3-j} \quad \text{for } 1 \leq i, j \leq 3, \quad (32d)$$

for the transposed inverse of an  $2 \times 2$ -matrix. Here  $a_{i,j}$  is the  $j$ -th component of the periodicity direction  $\mathbf{a}_i$ .

*Proof of Lemma 13.* The proof falls into three parts, the proof of the two special cases  $a = 1$  and  $k = 0$  and the more general case  $a \neq 1$ ,  $k > 0$ .

For  $a = 1$  pre-computed Bernoulli polynomials  $B_k(n)$  can be used to compute the sum in  $O(1)$  operations

$$S_{k,n}(1) = \sum_{\ell=0}^{n-1} \ell^k = \frac{B_{k+1}(n) - B_{k+1}(0)}{k+1}. \quad (33)$$

The Bernoulli polynomials can be pre-computed according to the formula

$$B_k(x) = \sum_{n=0}^k \frac{1}{n+1} \sum_{\ell=0}^n (-1)^\ell \binom{n}{\ell} (x+\ell)^k,$$

with computational effort  $O(k^2)$ .

For  $k = 0$  and  $a \neq 1$  we have the geometric series

$$S_{0,n}(a) = \sum_{\ell=0}^{n-1} a^\ell = \frac{a^n - 1}{a - 1}, \quad (34)$$

which computes the sum in  $O(1)$  operations.

In all other cases, *i.e.*,  $k > 0$  and  $a \neq 1$ , we claim that

$$S_{k,n}(a) = \frac{a}{(1-a)^{k+1}} \left( \sum_{j=0}^{k-1} \alpha_{k,j} a^j + a^{n-1} \sum_{j=0}^k \beta_{k,j}(n) a^j \right), \quad (35a)$$

where the constants  $\alpha_{k,j}$  and  $\beta_{k,j}(n)$  are recursively defined by

$$\alpha_{k,j} := \begin{cases} 1, & j = 0, k-1 \\ (j+1)\alpha_{k-1,j} + (k-j)\alpha_{k-1,j-1}, & 1 < j < k-1, \end{cases} \quad (35b)$$

$$\beta_{k,j}(n) := \begin{cases} -n, & k=1, j=0, \\ n-1, & k=1, j=1, \\ n\beta_{k-1,0}(n), & k > 1, j=0, \\ (n+j)\beta_{k-1,j}(n) - (k-j-n+1)\beta_{k-1,j-1}(n), & k > 1, 0 < j < k, \\ (1-n)\beta_{k-1,k-1}(n), & k > 1, j=k. \end{cases} \quad (35c)$$

The proof of (35) is by induction in  $k$  and is based on the observation, that  $S_{k,n}(a)$  is the Z-transform of the sequence  $\{x_\ell^{n,k}\}$  with  $x_\ell^{n,k} = \ell^k$  for  $0 \leq \ell < n$  and  $x_\ell^{n,k} = 0$  otherwise. We define the Z-transform equivalently with  $a^\ell$  instead of  $a^{-\ell}$  as

$$S_{k,n}(a) = \mathcal{Z}(x_\ell^{n,k}) = \sum_{\ell=-\infty}^{+\infty} a^\ell x_\ell^{n,k}.$$

Using the rule for differentiation of the Z-transform we deduce

$$S_{k+1,n}(a) = a S'_{k,n}(a). \quad (36)$$

Differentiating (34) and multiplying with  $a$  we get

$$S_{1,n}(a) = \frac{a}{(a-1)^2} (1 + a^{n-1} (-n + (n-1)a)),$$

which proves (35) for  $k = 1$ . Assuming (35) to hold for  $k$ , we will prove it for  $k + 1$ .

Differentiating (35) and multiplying by  $a$  we conclude in view of (36) that

$$\begin{aligned} S_{k+1,n}(a) &= a \frac{\sum_{j=0}^{k-1} (j+1) \alpha_{k,j} a^j + a^{n-1} \sum_{j=0}^k (n+j) \beta_{k,j}(n) a^j}{(1-a)^{k+1}} \\ &\quad + a^2 (k+1) \frac{\sum_{j=0}^{k-1} \alpha_{k,j} a^j + a^{n-1} \sum_{j=0}^k \beta_{k,j}(n) a^j}{(1-a)^{k+2}} \\ &\stackrel{(35)}{=} \frac{a}{(1-a)^{k+2}} \left( \underbrace{\alpha_{k,0}}_{\alpha_{k+1,0}} + \sum_{j=1}^{k-1} \underbrace{\left( (j+1) \alpha_{k,j} + (k+1-j) \alpha_{k,j-1} \right)}_{\alpha_{k+1,j}} a^j \right. \\ &\quad \left. + \underbrace{\alpha_{k,k-1}}_{\alpha_{k+1,k}} a^k + \underbrace{n \beta_{k,0}(n)}_{\beta_{k+1,0}(n)} a^{n-1} + \underbrace{(1-n) \beta_{k,k}(n)}_{\beta_{k+1,k+1}(n)} a^{n+k} \right. \\ &\quad \left. + a^{n-1} \sum_{j=1}^k \underbrace{\left( (n+j) \beta_{k,j}(n) - (k-j-n+2) \beta_{k,j-1}(n) \right)}_{\beta_{k+1,j}(n)} a^j \right), \end{aligned}$$

which proves (35) for  $k + 1$ . The constants  $\alpha_{k,j}$  and  $\beta_{k,j}(n)$  in (35) and so  $S_{k,n}(a)$  can be computed with  $O(k^2)$  operations, and the proof is complete.  $\square$

**Remark 17.** *Apart from the Bernoulli polynomials for the special case  $a = 1$  also the powers  $a^n$  can be pre-computed (see Remark 11). The polynomials  $\sum_{j=1}^{k-1} \alpha_{k,j} a^j$  and  $\sum_{j=0}^k \beta_{k,j}(n) a^j$  can be stably evaluated via the Horner scheme. The coefficients of the Horner scheme can be likewise pre-computed.*

### 5.5. An optimised algorithm for fast quadrature

It has already been shown in Cor. 15 that the fast quadrature algorithm has computational complexity independent of  $n_{p,1} \times n_{p,2}$  the size of the crystal. This means that the computational cost per DOF is constant independent of  $n_{p,1} \times n_{p,2}$ . Thus the proposed multiscale FEM is indeed size robust in terms of computational time if we can show that it is size robust in terms of  $n_{\text{dof}}$ . However, the algorithm contains several sums and the computational costs for each DOF can be large for large  $p_{\text{mac}}$ . In this section a closer investigation of the constants in the formulas reveals this dependence on  $p_{\text{mac}}$  and shows how to optimize it to a complexity of  $O(p_{\text{mac}}^5)$  using sum factorisation techniques [55]. The complexity of the optimised algorithm is up to a constant equivalent to the assembly time for tensor product polynomials of degree  $p_{\text{mac}}$ .

An implementation of the fast quadrature needs to evaluate formula (31) for each pair of multiscale functions  $b_i^{\text{multi}}$  and  $b_j^{\text{multi}}$ . The total number of evaluations per pair is  $O(1)$ , there are 8 evaluations for the sesquilinear form  $a(b_i^{\text{multi}}, b_j^{\text{multi}})$  and one evaluation for the sesquilinear form  $b(b_i^{\text{multi}}, b_j^{\text{multi}})$ . The number of basis functions  $b^{\text{multi}}$  is  $O(p_{\text{mac}}^2)$  due to the tensor product structure and assumed constant number of micro functions  $n_{\text{mic}}$ . The complexity of the algorithm is thus  $O(p_{\text{mac}}^4)$  times the complexity to evaluate (31). In (31) the unit cell integrals  $Q_i$  and the Z-transform of powers  $S_i^{(1)}$ ,  $S_i^{(2)}$  need to be evaluated. The former requires a unit cell quadrature of at most a complexity of  $O(p_{\text{mac}}^2)$  (see Sec. 5.2 with  $n = 1$ ). The latter requires to evaluate one of the cases (33), (34) or (35). Only a time of  $O(1)$  is required to evaluate

**Algorithm 1:** Fast quadrature algorithm

---

```

input :  $V_{k_1}^{\text{gfem}}$  with  $\mathcal{M}^{\text{mic}}$ ,  $p_{\text{bloch}}$ ,  $n_{\text{mic}}$ ,  $\mathcal{M}^{\text{mac}}$ ,  $p_{\text{mac}}$  and exterior discretisation parameters, an
          $\widehat{\Omega}$ -periodic material function  $f(\mathbf{x})$  (here  $f(\mathbf{x}) = a(\mathbf{x})$ )
output: Stiffness part  $\mathbf{A}^{\text{stiff}}$  of system matrix

1 allocate space for  $\mathbf{A}_{\text{ext}}^{\text{stiff}}$ 
2 for  $K^{\text{mac}} \in \mathcal{M}_{\text{ext}}^{\text{mac}}$  do //all exterior elements
3    $\lfloor$  standard assembly process for  $K^{\text{mac}}$  ; time: e.g.,  $O(p^5)$  for p-FEM

4 allocate space for  $\mathbf{A}_{\text{cr}}^{\text{stiff}}(i_1, j_1, i_2, j_2)$  ; space:  $O(p_{\text{mac}}^4 n_{\text{mic}}^2)$ 
5 for  $K^{\text{mac}} \in \mathcal{M}_{\text{cr}}^{\text{mac}}$  do //all multiscale elements
6   let  $K^{\text{mac}}$  be a multiscale cell with  $n_1 \times n_2$  periods and the index set  $\mathcal{A} := [0, n_1] \times [0, n_2]$ .
7   for  $d\mathbf{i}, d\mathbf{j} \in \{(0, 1), (1, 0)\}$ ,  $\eta \in \{1, 2\}$  do //8 combinations
8     for  $\hat{\eta}_1 \in \{0, 1\}$  if  $d i_1 = 1$ , else  $\hat{\eta}_1 = 1$  do //2 combinations for  $d i_1 = 1$  in (32c)
9       if  $d i_1 = 1$  then  $J_1 \leftarrow \tilde{J}_{\hat{\eta}_1+1, \eta}^{-\top}$  else  $J_1 \leftarrow 1$  ; //see equation (32)
10      for  $\hat{\eta}_2 \in \{0, 1\}$  if  $d i_2 = 1$ , else  $\hat{\eta}_2 = 0$  do //2 combinations for  $d i_2 = 1$ 
11        if  $d i_2 = 1$  then  $J_2 \leftarrow \tilde{J}_{\hat{\eta}_2+1, \eta}^{-\top}$  else  $J_2 \leftarrow 1$  ; //see equation (32)
12        for  $i_2, j_2 = 0, \dots, n_{\text{mic}} - 1$  do //all micro functions
13          allocate space for  $S_{\mathcal{A}, i_2, j_2}^{\kappa_1, (1)}, S_{\mathcal{A}, i_2, j_2}^{\kappa_2, (2)}$  (only  $\kappa_1, \kappa_2$  variable) ; space:  $O(p_{\text{mac}})$ 
14          allocate space for  $Q_{f, i_2, j_2}^{\mathbf{m}, d i_2, d j_2}$  (only  $\mathbf{m}$  variable) ; space:  $O(p_{\text{mac}}^2)$ 
15          for  $\kappa_1 = 0, \dots, p_{\text{mac}}$  do //all values  $\kappa_1 = \nu_1 + \mu_1 - m_1$ 
16             $\lfloor$  compute  $S_{\mathcal{A}, i_2, j_2}^{\kappa_1, (1)}$  using Lem. 13 ; time:  $O(p_{\text{mac}})$ 
17            for  $\kappa_2 = 0, \dots, p_{\text{mac}}$  do //all values  $\kappa_2 = \nu_2 + \mu_2 - m_2$ 
18               $\lfloor$  compute  $S_{\mathcal{A}, i_2, j_2}^{\kappa_2, (2)}$  using Lem. 13 ; time:  $O(p_{\text{mac}})$ 
19              for  $m_1, m_2 = 0, \dots, p_{\text{mac}}$  do //all monomial powers
20                 $\lfloor$  compute  $Q_{f, i_2, j_2}^{\mathbf{m}, d i_2, d j_2}$  using unit cell quadrature ; time:  $O(p_{\text{mac}}^2)$ 
21                allocate space for  $H_1(\hat{i}_2, \hat{j}_1, \hat{j}_2, \nu_1)$  ; space:  $O(p_{\text{mac}}^4)$ 
22                allocate space for  $H_2(\hat{i}_2, \hat{j}_2, \kappa_1)$  ; space:  $O(p_{\text{mac}}^3)$ 
23                allocate space for  $H_4(\kappa_1, \kappa_2)$  ; space:  $O(p_{\text{mac}}^2)$ 
24                for  $\kappa_1, \kappa_2 = 0, \dots, p_{\text{mac}}$  do
25                   $\lfloor$   $H_4(\kappa_1, \kappa_2) \leftarrow \sum_{\mathbf{m}=0}^{\kappa} \binom{\kappa_1}{m_1} \binom{\kappa_2}{m_2} S_{\mathcal{A}, i_2, j_2}^{\kappa_1, (1)} S_{\mathcal{A}, i_2, j_2}^{\kappa_2, (2)} Q_{f, i_2, j_2}^{\mathbf{m}, d i_2, d j_2}$  ; time:  $O(p_{\text{mac}}^2)$ 
26                  for  $\hat{i}_2, \hat{j}_2, \kappa_1 = 0, \dots, p_{\text{mac}}$  do
27                     $\lfloor$   $H_2(\hat{i}_2, \hat{j}_2, \kappa_1) \leftarrow$ 
28                       $\sum_{\nu_2=0}^{p_{\text{mac}}} \sum_{\mu_2=0}^{p_{\text{mac}}} (\nu_2 + 1)^{d i_1 \hat{\eta}_1} \alpha_{\nu_2 + d i_1 \hat{\eta}_1}^{(\hat{i}_2, n_2)} (\mu_2 + 1)^{d i_2 \hat{\eta}_2} \alpha_{\mu_2 + d i_2 \hat{\eta}_2}^{(\hat{j}_2, n_2)} H_4(\kappa_1, \nu_2 + \mu_2)$ 
29                      time:  $O(p_{\text{mac}}^2)$ 
30                    for  $\hat{i}_2, \hat{j}_1, \hat{j}_2, \nu_1 = 0, \dots, p_{\text{mac}}$  do
31                       $\lfloor$   $H_1(\hat{i}_2, \hat{j}_1, \hat{j}_2, \nu_1) \leftarrow \sum_{\mu_1=0}^{p_{\text{mac}}} (\mu_1 + 1)^{d i_2 (1 - \hat{\eta}_2)} \alpha_{\mu_1 + d i_2 (1 - \hat{\eta}_2)}^{(\hat{j}_1, n_1)} H_2(\hat{i}_2, \hat{j}_2, \nu_1 + \mu_1)$ 
32                      time:  $O(p_{\text{mac}})$ 
33                    for  $\hat{i}_1, \hat{i}_2, \hat{j}_1, \hat{j}_2 = 0, \dots, p_{\text{mac}}$  do
34                       $\lfloor$   $\mathbf{A}_{\text{cr}}^{\text{stiff}}(i_1, j_1, i_2, j_2) \leftarrow \mathbf{A}_{\text{cr}}^{\text{stiff}}(i_1, j_1, i_2, j_2)$ 
35                         $+ J_1 J_2 \sum_{\nu_1=0}^{p_{\text{mac}}} (\nu_1 + 1)^{d i_1 (1 - \hat{\eta}_1)} \alpha_{\nu_1 + d i_1 (1 - \hat{\eta}_1)}^{(\hat{i}_1, n_1)} H_1(i_2, \hat{j}_1, \hat{j}_2, \nu_1)$ 
36                        time:  $O(p_{\text{mac}})$ 
37                     $\rfloor$ 
38                   $\rfloor$ 
39                 $\rfloor$ 
40               $\rfloor$ 
41             $\rfloor$ 
42           $\rfloor$ 
43         $\rfloor$ 
44       $\rfloor$ 
45     $\rfloor$ 
46   $\rfloor$ 
47  $\mathbf{A}^{\text{stiff}} \leftarrow$  combine  $(\mathbf{A}_{\text{ext}}^{\text{stiff}}, \mathbf{A}_{\text{cr}}^{\text{stiff}})$  //respect overlap handling
48 return  $\mathbf{A}^{\text{stiff}}$ 

```

---

(33) and (34), a time of  $O(p_{\text{mac}})$  to evaluate (35a), and a time of  $O(p_{\text{mac}}^2)$  to set up  $\alpha_{k,j}$  and  $\beta_{k,j}(n)$  by (35b) and (35c). So the complexity to evaluate the terms  $S_i^{(1)}$ ,  $S_i^{(2)}$  and  $Q_i$  for a fixed  $i$  is  $O(p_{\text{mac}}^2)$ . There are  $O(p_{\text{mac}}^6)$  of these terms to be summed, which leads to a total time of  $O(p_{\text{mac}}^8)$  for a tuple  $(i, j)$ . Such an implementation thus has an overall complexity of  $O(p_{\text{mac}}^{12})$ .

Fortunately, some optimisations are possible. The optimized algorithm for the sesquilinear form  $a(\cdot, \cdot)$  is Alg. 1 (see Fig. 9 for the notation of the indices). A similar algorithm works for the sesquilinear form  $b(\cdot, \cdot)$  with the material function  $f(\mathbf{x}) = b^2(\mathbf{x}, \omega)$  and with a single combination instead of the 8 combinations in Alg. 1. The combination is given by the derivative selectors  $\mathbf{di} = \mathbf{dj} = \mathbf{0}$  and where the component selector  $\eta$  remains arbitrary. The algorithm uses memorisation and sum factorisation as in [55] and has the overall complexity of  $O(p_{\text{mac}}^5)$  in  $p_{\text{mac}}$ . This is up to a constant equal to the time for an assembly for the standard FEM on the coarse macro mesh  $\mathcal{M}_{\text{cr}}^{\text{mac}}$ ; however, it resolves variations on the micro scale.

## 6. Numerical results

This section presents results of the multiscale FEM presented in the previous sections for the model problem described in Sec. 2. We will give results for the same test cases as studied for the  $L^2$ -best approximation of the multiscale basis in Sec. 3.3 and supplement the results with an experiment for oblique incidence. The multiscale FEM is always compared to a scale-dependent p-FEM simulations of a high polynomial degree. The p-FEM simulation is known to produce good approximations but with high computational costs in time and memory for large  $n_p$ .

A C++ Implementation of the multiscale FEM solver of Section 4 has been integrated into the FEM solver Concepts [15]. A tool chain to compute a finite crystal solution  $u^{\text{gfem}}$  for one test case is explained in Alg. 2. Experiments for varying parameters of Alg. 2 will be shown in the remainder of this section. Especially we vary the modelling parameters  $\hat{\omega}$ ,  $\varphi$  and  $n_p$  but also the discretisation parameters  $p_{\text{mac}}$ ,  $\mathcal{M}^{\text{mac}}$ . We will always fix a set of micro functions  $\mathcal{E}_{k_1}^{\text{mic}}(n_{\text{mic}})$  based on  $\mathcal{B}_{k_1}^{\text{dom}}$  depending on  $\hat{\omega}$  and  $\varphi$ . The unit cell  $\hat{\Omega} = [0, 1]^2$  and material parameters (see Fig. 2), the micro mesh  $\mathcal{M}^{\text{mic}}$  (see Fig. 8(d)), the polynomial degree  $p_{\text{bloch}} = 25$  (see discussion in Sec. 3.3), the exterior polynomial degree  $p_{\text{ext}} = 15$ , the size of the air layer  $L^{\text{air}} = 6$  and the number of terms in the DtN operator ( $n^{\text{DtN}} = 7$ ) have been fixed in all experiments. The constants have been chosen so that the modelling error of the non-reflecting boundary condition is negligible and so that the over-all discretisation error is dominated by the error in  $\Omega^{\text{cr}}$ . We always compare the method in the  $H^1(\Omega^{\text{cr}})$ -norms and the nDOFs with support in  $\Omega^{\text{cr}}$  which have been used for the simulation. The nDOFs completely supported in  $\Omega^{\text{ext}}$  and the error in  $\Omega^{\text{air}}$  have not been taken into the analysis as the focus of the present work is the discretisation of  $\Omega^{\text{cr}}$ .

---

### Algorithm 2: Toolchain for the multiscale FEM solver

---

```

define normalised frequency  $\hat{\omega}$  ; //compare to the bandstructure in Fig. 3
1 define incidence angle  $\varphi$  ; // $\varphi = 0$  defines perpendicular incidence (default)
2 setup incident plane wave (default amplitude 1) ; //see Sec. 2.7
3 compute Floquet parameter  $k_1 = T(\sin(\varphi)\omega^2)$  ;

4 define micro mesh  $\mathcal{M}^{\text{mic}}$  of  $\hat{\Omega}$  ; //see Fig. 8(d)
5 define polynomial degree  $p_{\text{bloch}}$  to solve the unit cell problem (14) ;
6 compute set of the set of first  $n_{\text{bloch}}$  micro functions  $\mathcal{B}_{k_1}(n_{\text{bloch}})$  ; //see Sec. 3.2
7 define exterior macro mesh  $\mathcal{M}_{\text{ext}}^{\text{mac}}$  of width  $L^{\text{air}}$  ; //see Fig. 8(a)
8 define number of periods  $n_p$  of  $\Omega^{\text{cr}}$  and multiscale macro mesh  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  ; //see Fig. 8(a)
9 define macro polynomial degree  $p_{\text{mac}}$  for  $V_{k_1}^{\text{multi}}$  and polynomial degree  $p_{\text{ext}}$  for  $V_{k_1}^{\text{ext}}$  ;
10 compute finite crystal solution  $u$  using g-FEM solver ;
11 postprocess  $u$  ; //e.g., compute  $H^1(\Omega^{\text{cr}})$ -error to a p-FEM reference simulation.

```

---

The reference finite crystal solution  $u^{\text{ref}}$  has been computed with p-FEM on a size dependent mesh (see Fig. 5) with uniform polynomial degree  $p^{\text{ref}} = 20$ . Since the PhC scatterer in the investigated example is made of smooth dielectric circular rods, the p-FEM with curved cells ensures a high accuracy of the reference solution [68]. We observed that the relative  $H^1(\Omega^{\text{cr}})$  error of  $u^{\text{ref}}$  is approximately constant for a fixed  $p^{\text{ref}}$  independent of  $n_{\text{p}}$ . By computing a finite crystal solution for small  $n_{\text{p}}$  with a polynomial degree  $p > p^{\text{ref}}$  we assure that the error of  $u^{\text{sr}}$  is below  $10^{-6}$  with small variations depending on  $\hat{\omega}$ . All  $H^1(\Omega^{\text{cr}})$ -errors are measured with respect to  $u^{\text{ref}}$ .

The efficiency of the g-FEM solutions  $u^{\text{gfem}}$  are compared to the efficiency of solutions  $u^{\text{sr}}$  of a size dependent method, where we used again p-FEM with uniform polynomial degrees  $p^{\text{sr}}$  well below the reference degree  $p^{\text{ref}}$ . The nDOFs completely determines the computational time used by p-FEM or multiscale FEM. Note, that each DOF for multiscale FEM is up to a constant as expensive as a DOF for p-FEM (cf. Sec. 5), and thus a comparison of nDOFs is sufficient. Using an adaptive  $hp$ -FEM instead of p-FEM, a smaller nDOFs may be obtained for a certain discretisation error, but no essentially different numerical results are expected.

The macro mesh  $\mathcal{M}_{\text{cr}}^{\text{mac}}$  was always fixed to the one presented in Fig. 8(a) and the polynomial degree  $p_{\text{mac}}$  was varied. The set of micro functions  $\mathcal{E}_{k_1}^{\text{mic}}$  was chosen as the set of dominant Bloch modes  $\mathcal{B}_{k_1}^{\text{dom}}$  completed by the special micro function  $b_{\star}^{\text{mic}}$  (see Sec. 4.3), consequently we always have  $n_{\text{mic}} = n_{\text{bloch}} + 1$ . In the present work we will not show results for larger micro spaces  $V_{k_1}^{\text{mic}}$  whose generators  $\mathcal{E}_{k_1}^{\text{mic}}$  contain higher Bloch modes. For bandgap frequencies we modified  $\mathcal{E}_{k_1}^{\text{mic}}$  according to the selection rule (D3) described in Sec. 3.3.

### 6.1. Experiments for perpendicular incidence

First, we present the results for the test cases of Sec. 3.3 with  $\hat{\omega} \in \{0.215, 0.300, 0.625\}$  and perpendicular incidence ( $\varphi = 0$ ) for several  $n_{\text{p}}$ .

*Convergence in  $p_{\text{mac}}$ .* In Fig. 11(a) the convergence w.r.t.  $p_{\text{mac}}$  is shown for  $\hat{\omega} = 0.625$ . The error decreases as  $p_{\text{mac}}$  is increased which is in accordance to the best approximation error for the first cell (see Fig. 6(e)). The final relative  $H^1(\Omega^{\text{cr}})$ -error is below  $10^{-3}$  and even one magnitude lower for larger  $n_{\text{p}}$ . In Fig. 11(b) all three frequencies  $\hat{\omega} = 0.215, 0.300, 0.625$  are compared, for an easier presentation for the extreme cases  $n_{\text{p}} = 3$  and  $n_{\text{p}} = 100$ . The computations have been done for the same values of  $p_{\text{mac}}$  as in Fig. 11(a), where we now present a plot in dependence of nDOFs instead of  $p_{\text{mac}}$ . We observe in all three cases, for the propagating frequency  $\hat{\omega} = 0.215$  with  $n_{\text{bloch}} = 2$ , the bandgap frequency  $\hat{\omega} = 0.300$  with  $n_{\text{bloch}} = 2$  and the propagating frequency  $\hat{\omega} = 0.625$  with  $n_{\text{bloch}} = 4$ , a relative  $H^1(\Omega^{\text{cr}})$ -error lower than  $10^{-3}$  with less than 1000 multiscale DOFs. For the propagating frequencies  $\hat{\omega} = 0.215$  and  $\hat{\omega} = 0.625$  the error decays for larger  $n_{\text{p}}$  as then the error in the first cells becomes less and less important. For bandgap frequencies this is not the case, as the fields themselves decay and the first cells will always be most important.

*Comparison to p-FEM.* The results for a similar but more extensive experiment are shown in Fig. 12 for the frequency  $\hat{\omega} = 0.215$ . In this experiment all the solutions for  $n_{\text{p}} = 3$  up to  $n_{\text{p}} = 50$  have been computed for  $p_{\text{mac}} = 1, \dots, 9$ . The error for  $n_{\text{dof}} = 8$  ( $p_{\text{mac}} = 1$ ),  $n_{\text{dof}} = 78$  ( $p_{\text{mac}} = 3$ ) and  $n_{\text{dof}} = 434$  ( $p_{\text{mac}} = 7$ ) is displayed. The plot shows that the error of multiscale FEM is almost independent of  $n_{\text{p}}$  and the error even decreases for larger  $n_{\text{p}}$ . Additionally to the error of MSFEM, the plot also shows the error for a p-FEM simulation using the size dependent mesh displayed in Fig. 5 that resolves the material interfaces. The superiority of a size robust method over a size dependent method is clearly visible. For  $n_{\text{p}} = 50$  the multiscale FEM achieves about the same error with 434 DOFs for which p-FEM needs 50'000 DOFs. And for larger  $n_{\text{p}}$  the difference in DOFs is even more pronounced. The nDOFs for p-FEM grow linearly with  $n_{\text{p}}$  which leads to a linearly growing assembly time, a more than linear increase in the solver time and linearly growing memory usage. In contrast to this, for multiscale FEM, the assembly time, solver time and memory usage is practically constant. Already for  $n_{\text{p}} = 3$  the multiscale FEM requires less DOFs as p-FEM, but as the constant in the assembly time per DOF of multiscale FEM is larger than for p-FEM, the total runtime of the solver is larger for very small  $n_{\text{p}}$ . We note that the size dependent mesh for p-FEM is necessary to achieve exponential convergence. If the material interfaces would not have been resolved p-FEM would

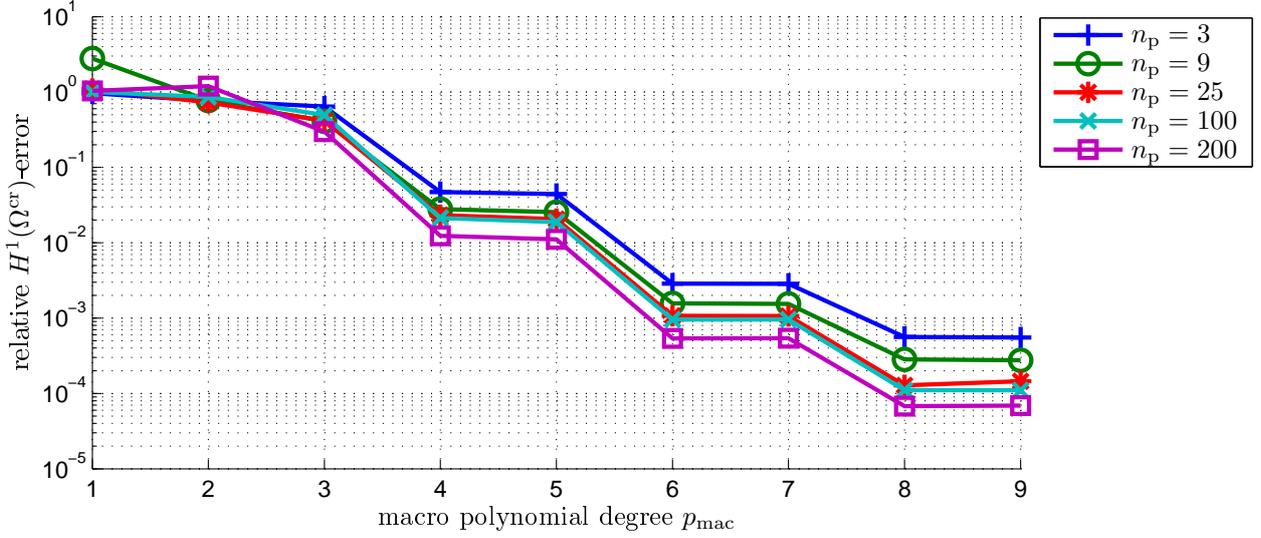
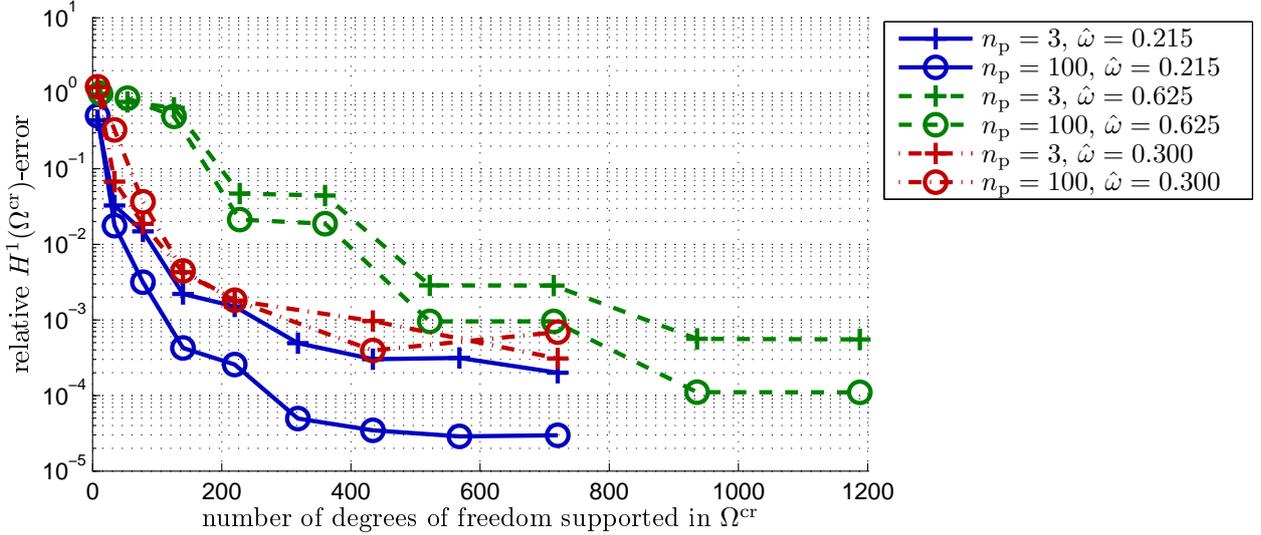
(a) Multiscale FEM for  $\hat{\omega} = 0.625$  ( $n_{\text{bloch}} = 4$ )(b) Multiscale FEM for  $\hat{\omega} = 0.215$  ( $n_{\text{bloch}} = 2$ ),  $\hat{\omega} = 0.625$  ( $n_{\text{bloch}} = 4$ ) and  $\hat{\omega} = 0.300$  ( $n_{\text{bloch}} = 2$ , bandgap). All simulations have been done for  $p_{\text{mac}} = 1, 2, \dots, 9$  and for each  $p_{\text{mac}}$  the used nDOFs and the achieved error are displayed.

Figure 11: Multiscale FEM for perpendicular incidence and the three frequencies studied in Sec. 3.3, see Fig. 3 for the corresponding bandstructure. In all three cases the MSFEM is size robust for constant  $p_{\text{mac}}$  and constant nDOFs. The convergence graphs for  $\hat{\omega} = 0.625$  includes larger nDOFs due to the two additional micro functions.

reach only algebraic convergence which is confirmed in practise. Apparently, p-FEM has to use more DOFs for growing  $n_p$  to retain a particular error level.

### 6.2. MSFEM for oblique incidence

We also computed the error of MSFEM for oblique incidence for  $\hat{\omega} = 0.230$  as studied for the best-approximation in Sec. 3.3. The results are shown exemplary for  $n_p = 20$  in Fig. 13(a). The MSFEM accuracy is not affected by the incident angle  $\varphi$  and thus the method works uniformly well deep inside the bandgap, close to the propagating regime and inside the propagating regime. We note that this result relies on the selection rule (D3). With the selection rule (D1) the condition number of the system matrix is even

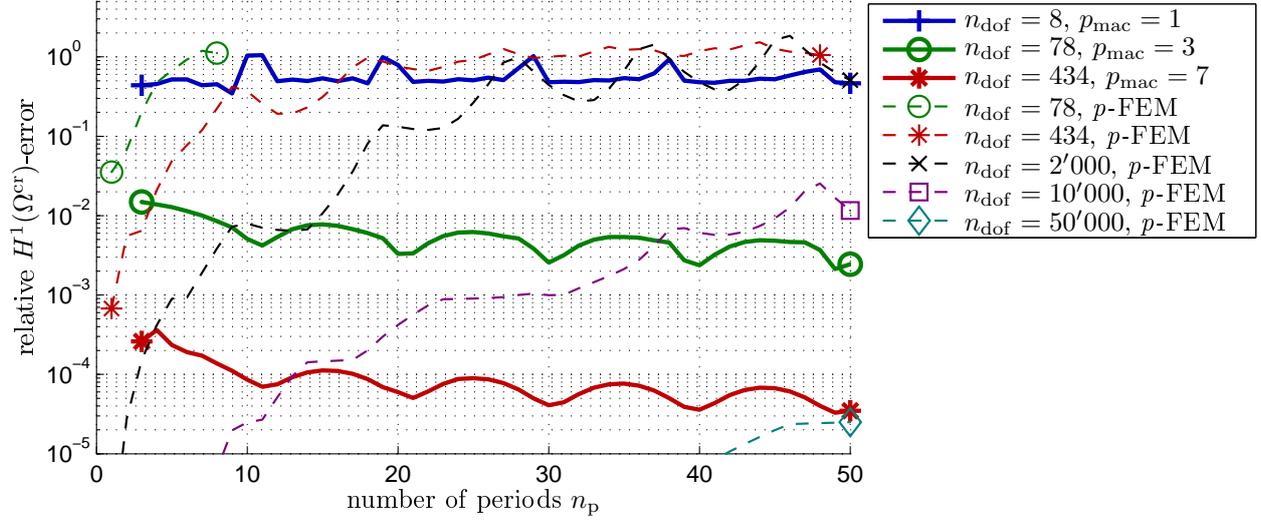


Figure 12: Multiscale FEM for perpendicular incidence and the frequency  $\hat{\omega} = 0.215$  in comparison to  $p$ -FEM.  $p$ -FEM is size dependent and the error for a fixed nDOF increases with  $n_p$ , whereas MSFEM is size robust and achieves for a fixed nDOF a constant error independent of  $n_p$ .

larger than that of the  $L^2$ -best projection and the method fails before reaching even an error level of  $10^{-2}$ . The reduction of the condition numbers is subject of future research.

*Experimental investigation of the transmission coefficient.* An important quantity for manufacturing a PhC structure is its transmission coefficient  $T$ , it is desired to control  $T$ . If a PhC is used to shield a waveguide it is desired to have a very small value for  $T$ , so that most of the field is transported along the waveguide. To couple a field into the waveguide a high value of  $T \approx 1$  is desired. Often the transmission coefficient of a complicated PhC device is deduced by the bandstructure of the corresponding infinite PhC. The bandstructure of the infinite crystal give the general idea of the value of  $T$ , but there are some effects unique to finite PhCs.

As a generalisation of reflection and transmission coefficients for wave optics [63], we define  $T$  as the power transmission coefficient [71]

$$T = \frac{\text{en}(u_{k_1}, n_p + L^{\text{air}})}{\text{en}(u_{k_1}^{\text{inc}}, n_p + L^{\text{air}})},$$

where the power flux  $\text{en}(u, y)$  of a field  $u$  over an interface line  $x_2 = y$  is given by

$$\text{en}(u, y) = \text{Im} \int_0^a \frac{\partial u}{\partial x_2}(x_1, y) \bar{u}(x_1, y) dx_1.$$

The transmission coefficient for  $\hat{\omega} = 0.230$ , varying incident angles  $\varphi$  and several sizes of the crystal is displayed in Fig. 13(b). We displayed the transmission coefficient as computed by  $p$ -FEM with polynomial degree  $p = 20$  and by MSFEM with  $p_{\text{mac}} = 3$ . The  $p$ -FEM computation for  $p = 20$  has resolved the field  $u_{k_1}$  with high precision and thus the computed transmission coefficient is expected to be very accurate. The MSFEM for  $p_{\text{mac}} = 3$  with its only 78 DOFs approximates the transmission coefficient so accurate, that in Fig. 13(b) no difference between the two computations can be seen. As predicted by the infinite PhC, the transmission coefficient is very small for large crystals and away from the propagating regime. Parts of the field are reflected on both interfaces of the crystal. Depending on the number of periods  $n_p$  and the angle  $\varphi$  there can be destructive or constructive interference between the two reflected fields. This explains why the reflection coefficient can vary strongly in the propagating regime, which cannot be predicted by knowing the

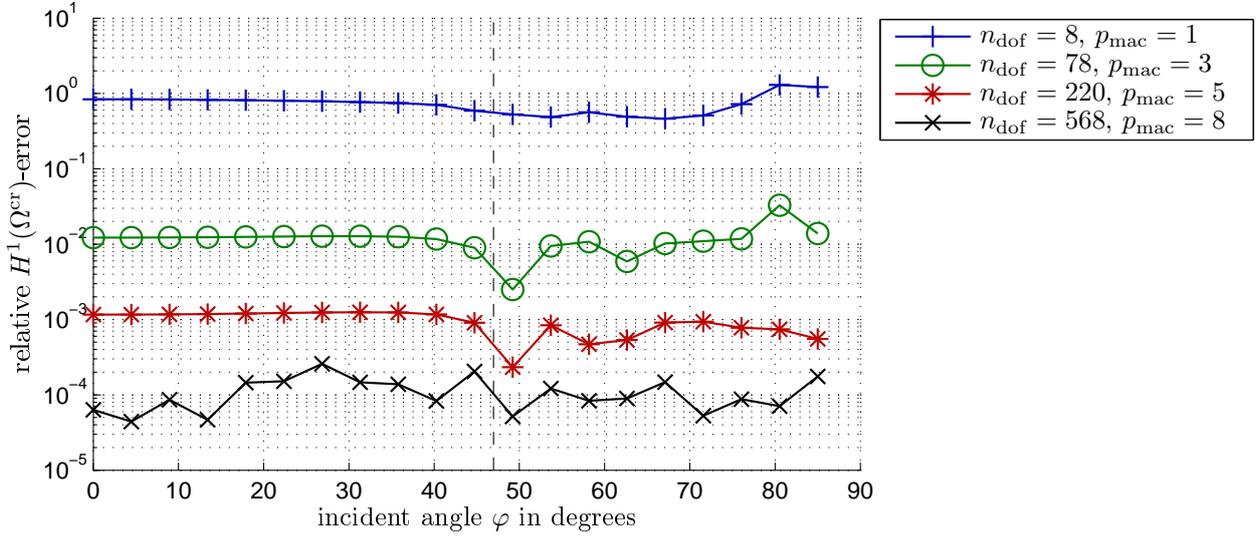
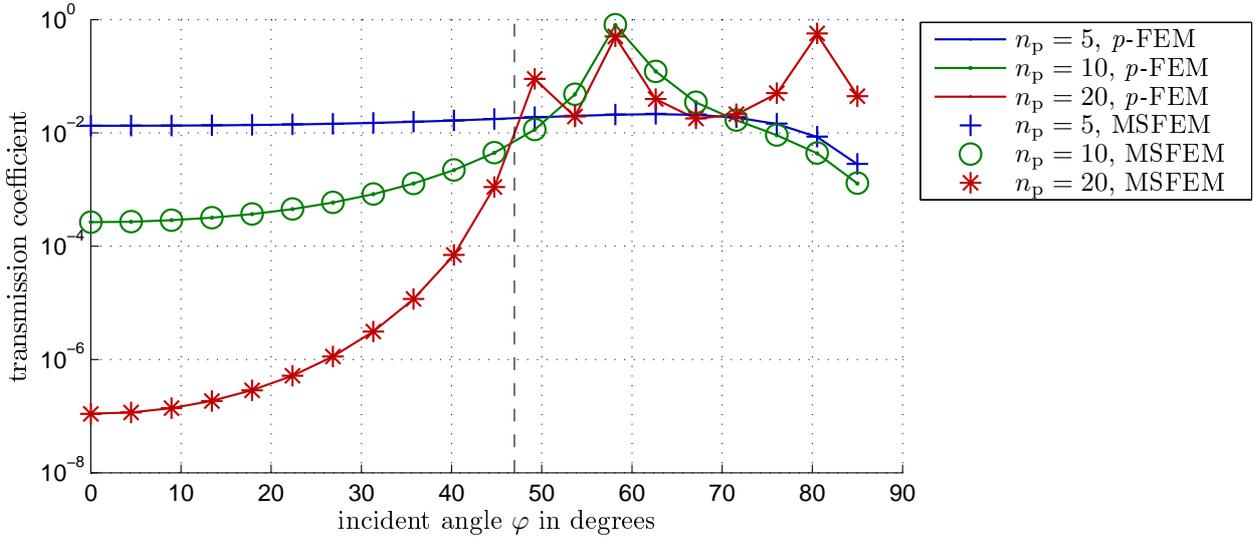
(a) MSFEM error for  $n_p = 20$ (b) Transmission coefficient as computed by  $p$ -FEM with  $p = 20$  and MSFEM with  $p_{\text{mac}} = 3$ .

Figure 13: Multiscale FEM for  $n_p = 20$  and oblique incidence for the frequency  $\hat{\omega} = 0.230$  ( $n_{\text{block}} = 2$ ) which is a bandgap frequency for incident angles  $\varphi < 47^\circ$  and propagating otherwise. Compare this result to the bandstructure, reflection coefficient and best-approximation result in Fig. 7. In **a**) we observe the MSFEM accuracy is incident angle robust, it works inside the bandgap (fast decaying modes) on the boundary of the bandgap (slow decaying modes) and in the propagating regime (no decay in modes). In **b**) the power transmission coefficient is displayed for varying  $n_p$ . No difference can be seen between the transmission coefficient computed by  $p$ -FEM with more than 50'000 DOFs and MSFEM with only 78 DOFs.

behaviour of waves in the infinite crystal. Note that for small crystals ( $n_p = 5$ ), the transmission coefficient can actually be smaller in the propagating regime than in the bandgap.

The results of this section show that the dominant Bloch modes  $\mathcal{B}_{k_1}^{\text{dom}}$  modulated by macroscopic continuous, piecewise polynomials of degree  $p_{\text{mac}} \geq 1$  are a very promising basis to numerically approximate finite crystal solutions. In the next section we will extend this result and propose a FEM which has this basis build-in into its basis functions. For  $p_{\text{mac}} = 1$  the Bloch modes modulated by macro polynomials form a partition of unity, *i.e.*, the Bloch modes themselves are contained in the space. Moreover, the modulation by macroscopic polynomials localises the Bloch modes and allows the finite crystal effects. Higher values for  $p_{\text{mac}}$  are useful close to the boundary of the crystal for the transition of the solution in the exterior space into solutions of the infinite crystal.

## 7. Conclusion

In this article, we introduced a size robust multiscale basis and multiscale FEM for PhC structures, as a special case of the Generalised Finite Element Method. The multiscale basis consists of quasi-periodic micro functions and piecewise polynomial macro functions. The micro functions can be automatically computed as numerical solutions of a unit cell problem. In particular, we used the unit cell problem (14) and solved for the dominant Bloch modes  $\mathcal{B}_{k_1}^{\text{dom}}$  (cf. Sec. 3). Together with a boundary layer mesh (see Fig. 8(a)) this basis can approximate solutions inside the PhC band size robustly. Using this multiscale basis we constructed a MSFEM which can be coupled to an arbitrary exterior discretisation. The efficiency of the MSFEM can be controlled by changing the set of micro functions  $\mathcal{E}_{k_1}^{\text{mic}}$ , the macroscopic mesh  $\mathcal{M}^{\text{mac}}$ , and the polynomial degree  $p_{\text{mac}}$ . Moreover, the MSFEM allows *h*-, *p*- or *hp*-refinements on the macroscopic part and sparsification techniques of the tensor product  $V_{k_1}^{\text{multi}} = V_{k_1}^{\text{mic}} \otimes V_{\text{per}}^{\text{mac}}$ . The MSFEM is size robust for the PhC barrier and to achieve a particular error it only requires a few DOFs independent of the size of the crystal, the angle of the incident field or the frequency  $\hat{\omega}$ . In particular, the method works inside and outside the bandgap of the corresponding infinite crystal. For large PhC structures, the method is much faster than size dependent standard discretisation schemes like FDTD or *h*-, *p*- and *hp*-FEM but it is already useful for smaller crystals. For example for  $n_p = 3$  it requires only the same nDOFs than *p*-FEM. For each DOF the MSFEM only requires a constant computational time which is not more expensive than a constant multiple of the time for a *p*-FEM DOF. This is due to the efficient integration rule for two-scale functions with quasi-periodic micro structure described in Sec. 5.

Various simulation techniques have been proposed for PhC crystals of different sizes. Very small PhC structures with only a few periods ( $n_p \ll \infty$ ) can be efficiently simulated by direct discretisation schemes like *p*-FEM [67]. Infinitely large PhC structures can be reduced by a Floquet- or Bloch-transformation to a family of computationally manageable problems where each can be simulated by direct discretisation schemes. An infinitely large PhC with a localised perturbation or an infinitely PhC in a half space can be treated by special DtN operators as proposed by Fliss *et al.* [23]. For the practical case of large PhC structures present methods are still not satisfactory. The proposed MSFEM for PhC bands can be extended to 2D structures where the techniques presented in this article are building blocks. This is work in progress and will be reported elsewhere.

Moreover, further work has to be done in the numerical analysis of the MSFEM. The conjecture of *size-independent stability*, Conj. 9, is not yet proven and the size-independence results are only numerically confirmed.

## References

- [1] ABDULLE, A., AND WEINAN, E. Finite difference heterogeneous multi-scale method for homogenization problems. *J. Comput. Phys.* 191, 1 (2003), 18–39.
- [2] AINSWORTH, M. Discrete dispersion relation for *hp*-version finite element approximation at high wave number. *SIAM Journal on Numerical Analysis* 42, 2 (2005), 553–575.
- [3] ALVES, C. J., AND VALTCHEV, S. S. Numerical comparison of two meshfree methods for acoustic wave scattering. *Eng. Anal. Bound. Elem.* 29, 4 (2005), 371 – 382.
- [4] BABA, T. Slow light in photonic crystals. *Nat. Photonics* 2, 8 (2008), 465–473.

- [5] BABICH, V., LYALINOV, M., AND GRIKUROV, V. *Diffraction theory: The Sommerfeld-Malyuzhinets technique*. Alpha Science, Oxford, UK, 2007.
- [6] BABUŠKA, I., BANERJEE, U., AND OSBORN, J. Generalized finite element methods – main ideas, results and perspective. *Int. J. Comput. Methods* 1, 1 (2004), 67–103.
- [7] BENSOUSSAN, A., LIONS, J. L., AND PAPANICOLAOU, G. *Asymptotic analysis for periodic structures*, vol. 5 of *Studies in Mathematics and its Applications*. North-Holland, Amsterdam, Netherlands, 1978.
- [8] BERNARDI, C., MADAY, Y., AND RAPETTI, F. Basics and some applications of the mortar element method. *GAMM Mitt.* 28, 2 (2005), 97–123.
- [9] BONNET-BENDHIA, A.-S., AND JOLY, P. *Mathematical analysis and numerical approximation of optical waveguides.*, vol. 22 of *Frontiers Appl. Math.* SIAM, Philadelphia, USA, 2001, pp. 273–324.
- [10] BRAESS, D. *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*, 3th ed. Cambridge University Press, Cambridge, UK, 2007.
- [11] BUSCH, K. Photonic band structure theory: assessment and perspectives. *C. R. Physique* 3, 53 (2002), 53–66.
- [12] BUSCH, K., MINGALEEV, S., GARCIA-MARTIN, A., SCHILLINGER, M., AND HERMANN, D. The Wannier function approach to photonic crystal circuits. *Journal of Physics: Condensed Matter* 15 (2003), R1233.
- [13] CANUTO, C., HUSSAINI, M. Y., QUARTERONI, A., AND ZANG, T. A. *Spectral methods: Fundamentals in single domains*. Springer Verlag, Heidelberg, Germany, 2006.
- [14] COLTON, D., AND KRESS, R. *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer-Verlag, Heidelberg, Germany, 1998.
- [15] CONCEPTS DEVELOPMENT TEAM. *Webpage of Numerical C++ Library Concepts 2*. <http://www.concepts.math.ethz.ch>, 2009.
- [16] EHRHARDT, M., Ed. *Wave Propagation in Periodic Media - Analysis, Numerical Techniques and practical Applications*, vol. 1 of *Progress in Computational Physics*. Bentham Science Publishers Ltd., 2010.
- [17] EHRHARDT, M., AND ZHENG, C. Exact artificial boundary conditions for problems with periodic structures. *J. Comput. Phys.* 227, 14 (2008), 6877–6894.
- [18] ENGSTRÖM, C. *Effective materials in electromagnetics*. PhD thesis, University Lund, Lund, Sweden, 2005.
- [19] ENGSTRÖM, C., HAFNER, C., AND SCHMIDT, K. Computations of lossy bloch waves in two-dimensional photonic crystals. *J. Comput. Theor. Nanosci.* 6 (2009), 775–783.
- [20] ENGSTRÖM, C., AND RICHTER, M. On the spectrum of an operator pencil with applications to wave propagation in periodic and frequency dependent materials. *SIAM J. Appl. Math.* 70, 1 (2009), 231–247.
- [21] FLISS, S. *Etude mathématique et numérique de la propagation des ondes dans un milieu périodique présentant un défaut*. PhD thesis, INRIA Paris-Rocquencourt, France, May 2009.
- [22] FLISS, S., CASSAN, E., AND BERNIER, D. New approach to describe light refraction at the surface of a photonic crystal. *J. Opt. Soc. Amer. B* (submitted).
- [23] FLISS, S., AND JOLY, P. Exact boundary conditions for time-harmonic wave propagation in locally perturbed periodic media. *Applied Numerical Mathematics* 59, 9 (2009), 2155 – 2178. Second Chilean Workshop on Numerical Analysis of Partial Differential Equations (WONAPDE 2007).
- [24] FRAUENFELDER, P. *hp-Finite Element Methods on Anisotropically, Locally Refined Meshes in Three Dimensions with Stochastic Data*. PhD thesis, ETH Zurich, Zürich, Switzerland, 2004.
- [25] GIANI, S., AND GRAHAM, I. G. Adaptive finite element methods for computing band gaps in photonic crystals. BICS Preprint 6-10, University of Bath, 2010.
- [26] GIVOLI, D. Non-reflecting boundary conditions. *J. Comput. Phys.* 94, 1 (1991), 1–29.
- [27] GIVOLI, D. Recent advances in the DtN FE method. *Arch. Comput. Methods Eng.* 6, 2 (1999), 71–116.
- [28] GOLDSTEIN, C. A finite element method for solving Helmholtz type equations in waveguides and other unbounded domains. *Math. Comp.* 39, 160 (1982), 309–324.
- [29] GROTE, M., AND KELLER, J. On nonreflecting boundary conditions. *Journal of Computational Physics* 122 (1995), 231–243.
- [30] GUENNEAU, S., AND ZOLLA, F. Homogenization of three-dimensional finite photonic crystals. *Prog. Electromag. Res.* 27 (2000), 91–127.
- [31] HACKBUSCH, W., FADIMAN, R., AND ION, P. *Elliptic differential equations: theory and numerical treatment*. Springer, 1992.
- [32] HOU, T., AND WU, X. A multiscale finite element method for elliptic problems in composite materials and porous media. *Journal of Computational Physics* 134, 1 (1997), 169–189.
- [33] HOU, T., WU, X., AND CAI, Z. Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients. *Math. Comp.* 68, 227 (1999), 913–943.
- [34] HU, Z., AND LU, Y. Improved Dirichlet-to-Neumann map method for modeling extended photonic crystal devices. *Optical and quantum electronics* 40, 11 (2008), 921–932.
- [35] IHLENBURG, F. *Finite Element Analysis of Acoustic Scattering*. Applied Mathematical Sciences. Springer Verlag, 1998.
- [36] ISTRATE, E., GREEN, A. A., AND SARGENT, E. H. Behavior of light at photonic crystal interfaces. *Phys. Rev. B* 71, 19 (2005), 195122.
- [37] ISTRATE, E., AND SARGENT, E. H. Photonic crystal heterostructures and interfaces. *Rev. Mod. Phys.* 78, 2 (May 2005), 455–481.
- [38] JIANG, W., CHEN, R., AND LU, X. Theory of light refraction at the surface of a photonic crystal. *Phys. Rev. B* 71 (2005), 245115.
- [39] JOANNOPOULOS, J., VILLENEUVE, P., AND FAN, S. Photonic crystals: putting a new twist on light. *Nature* 386, 6621

- (1997), 143–149.
- [40] JOANNOPOULOS, J. D., JOHNSON, S. G., WINN, J. N., AND MEADE, R. D. *Photonic Crystals*, 2nd ed. Princeton University Press, Princeton, USA, 2008.
- [41] JOHNSON, S. G., AND JOANNOPOULOS, J. D. Block-iterative frequency-domain methods for Maxwell’s equations in a planewave basis. *Optics Express* 8, 3 (2001), 173–190.
- [42] JOLY, P., LI, J., AND FLISS, S. Exact boundary conditions for periodic waveguides containing a local perturbation. *Comm. Comp. Phys* 1, 6 (2006), 945–973.
- [43] KARNIADAKIS, G. E., AND SHERWIN, S. J. *Spectral/hp Element Methods for Computational Fluid Dynamics*. Oxford University Press, Oxford, UK, 2005.
- [44] KITTEL, C. *Introduction to Solid State Physics*, 8th ed. John Wiley and Sons, New York, USA, 2004.
- [45] KOSHIBA, M. Wavelength division multiplexing and demultiplexing with photonic crystal waveguide couplers. *J. Lightwave Technol.* 19, 12 (2001), 1970–1975.
- [46] KRAUSS, T. Why do we need slow light? *Nat. Photonics* 2, 8 (2008), 448–450.
- [47] KUCHMENT, P. *Floquet Theory for Partial Differential Equations*. Birkhäuser Verlag, Basel, Switzerland, 1993.
- [48] KUCHMENT, P. The mathematics of photonic crystals. In *Mathematical modeling in optical science*, vol. 22 of *Frontiers Appl. Math.* SIAM, Philadelphia, USA, 2001, pp. 207–272.
- [49] LEHOUCQ, R. B., SORENSEN, D. C., AND YANG, C. *ARPACK Users’ Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. SIAM, Philadelphia, USA, 1998.
- [50] LYALINOV, M. Generalized Sommerfeld integral and diffraction in an angle-shaped domain with a radial perturbation. *J. Phys. A: Math. Gen.* 38, 43 (2005), L707–L714.
- [51] MATAACHE, A. Sparse Two-Scale FEM for Homogenization Problems. *J. Sci. Comput.* 17, 1 (2002), 659–669.
- [52] MATAACHE, A., BABUŠKA, I., AND SCHWAB, C. Generalized p-FEM in homogenization. *Numer. Math.* 86, 2 (2000), 319–375.
- [53] MATAACHE, A.-M. *Spectral and p- Finite Elements for problems with microstructure*. PhD thesis, ETH Zurich, Zürich, Switzerland, 2000.
- [54] MELENK, J., AND BABUŠKA, I. The partition of unity finite element method: basic theory and applications. *Comput. Methods Appl. Mech. Engrg.* 139, 1-4 (1996), 289–314.
- [55] MELENK, J., GERDES, K., AND SCHWAB, C. Fully discrete hp-finite elements: Fast quadrature. *Computer Methods in Applied Mechanics and Engineering* 190, 32-33 (2001), 4339–4364.
- [56] MINGALEEV, S., AND BUSCH, K. Scattering matrix approach to large-scale photonic crystal circuits. *Optics letters* 28, 8 (2003), 619–621.
- [57] MOUSSA, R., FOTEINOPOULOU, S., ZHANG, L., TUTTLE, G., GUVEN, K., OZBAY, E., AND SOUKOULIS, C. Negative refraction and superlens behavior in a two-dimensional photonic crystal. *Phys. Rev. B* 71 (2005), 085106.
- [58] ODEN, J., PRUDHOMME, S., AND DEMKOWICZ, L. A posteriori error estimation for acoustic wave propagation problems. *Archives of Computational Methods in Engineering* 12, 4 (2005), 343–389.
- [59] OLEINIK, O., SHAMAEV, A., AND YOSIFIAN, G. *Mathematical Problems in Elasticity and Homogenization*. North-Holland, Amsterdam, Netherlands, 1992.
- [60] PASQUETTI, R., AND RAPETTI, F. Spectral element methods on triangles and quadrilaterals: comparisons and applications. *J. Comput. Phys.* 198, 1 (2004), 349–362.
- [61] PEDREIRA, D. G., AND JOLY, P. A Method for Computing Guided Waves in Integrated Optics. Part II: Numerical Approximation and Error Analysis. *SIAM J. Numer. Anal.* 39, 5 (2002), 1684–1711.
- [62] PENDRY, J. B. Negative refraction makes a perfect lens. *Phys. Rev. Lett.* 85, 18 (2000), 3966–3969.
- [63] ROTHWELL, E. Efficient computation of the time-domain TM plane-wave reflection coefficient. *IEEE Transactions on Antennas and Propagation* 53, 10 (2005), 3417–3419.
- [64] RÜEGG, A. W. Implementation of Generalized Finite Element Methods for Homogenization Problems. *J. Sci. Comput.* 17 (2002), 671–681.
- [65] RÜEGG, A. W., SCHNEEBELI, A., AND LAUPER, R. Generalized hp-FEM for lattice structures. Research report no. 2002-23, ETH Zurich, Seminar of Applied Mathematics, 2002.
- [66] SAUTER, S., AND SCHWAB, C. *Randelementmethoden*. Teubner Stuttgart, 2004.
- [67] SCHMIDT, K., AND KAPPELER, R. Efficient computation of photonic crystal waveguide modes with dispersive material. *Optics Express* 18, 7 (2010), 7307–7322.
- [68] SCHMIDT, K., AND KAUF, P. Computation of the band structure of two-dimensional Photonic Crystals with hp Finite Elements. *Comput. Methods Appl. Mech. Engrg.* 198 (2009), 1249–1259.
- [69] SCHWAB, C. *p- and hp- Finite Element Methods: Theory and Applications in Solid and Fluid Mechanics*. Oxford University Press, Oxford, UK, 1998.
- [70] SCHWAB, C., AND TODOR, R. Sparse finite elements for elliptic problems with stochastic loading. *Numer. Math.* 95, 4 (2003), 707–734.
- [71] SMAJIC, J., HAFNER, C., AND ERNI, D. Optimization of photonic crystal structures. *Journal of the Optical Society of America A* 21, 11 (2004), 2223–2232.
- [72] SMAJIC, J., HAFNER, C., RAUSCHER, K., AND ERNI, D. Computation of Radiation Leakage in Photonic Crystal Waveguides. In *Proc. Progr. Electromagn. Res. Symp. 2004, Pisa, Italy* (2004), pp. 21–24.
- [73] ŠOLÍN, P. *Partial Differential Equations and the Finite Element Method*. Wiley-Interscience, Hoboken, USA, 2006.
- [74] SOLJAČIĆ, M., JOHNSON, S., FAN, S., IBANESCU, M., IPPEN, E., AND JOANNOPOULOS, J. Photonic-crystal slow-light enhancement of nonlinear phase sensitivity. *J. Opt. Soc. Amer. B* 19, 9 (2002), 2052–2059.
- [75] SOUKOULIS, C. M. *Photonic crystals and light localization in the 21st century*. NATO Science Series C. Springer-Verlag, Heidelberg, Germany, 2001.

- 
- [76] SOUSSI, S. Convergence of the supercell method for defect modes calculations in photonic crystals. *SIAM J. Numer. Anal.* *43* (2005), 1175–1201.
- [77] SOUSSI, S., AND HOHAGE, T. Riesz bases and Jordan form of the translation operator in semi-infinite periodic waveguides. Tech. Rep. 2009-06, Universität Göttingen, Germany, 2009.
- [78] STRASSER, P. *The development of a fabrication process for passive photonic crystal devices in InP / InGaAsP*. PhD thesis, ETH Zürich, Zurich, Switzerland, 2008.
- [79] T., S., BABUSKA, I., AND COPPS, K. The design and analysis of the generalized finite element method. *Comput. Methods Appl. Mech. Engrg.* *181* (2000), 43–69.
- [80] TISSEUR, F., AND MEERBERGEN, K. The quadratic eigenvalue problem. *SIAM Rev.* *43*, 2 (2001), 235–286.

# Research Reports

No.	Authors/Title
10-12	<i>H. Brandsmeier, K. Schmidt and Ch. Schwab</i> A multiscale hp-FEM for 2D photonic crystal band
10-11	<i>V.H. Hoang and C. Schwab</i> Sparse tensor Galerkin discretizations for parametric and random parabolic PDEs. I: Analytic regularity and gpc-approximation
10-10	<i>V. Gradinaru, G.A. Hagedorn, A. Joye</i> Exponentially accurate semiclassical tunneling wave functions in one dimension
10-09	<i>B. Pentenrieder and C. Schwab</i> hp-FEM for second moments of elliptic PDEs with stochastic data. Part 2: Exponential convergence
10-08	<i>B. Pentenrieder and C. Schwab</i> hp-FEM for second moments of elliptic PDEs with stochastic data. Part 1: Analytic regularity
10-07	<i>C. Jerez-Hanckes and J.-C. Nédélec</i> Asymptotics for Helmholtz and Maxwell solutions in 3-D open waveguides
10-06	<i>C. Schwab and O. Reichmann</i> Numerical analysis of additive, Lévy and Feller processes with applications to option pricing
10-05	<i>C. Schwab and R. Stevenson</i> Fast evaluation of nonlinear functionals of tensor product wavelet expansions
10-04	<i>B.N. Khoromskij and C. Schwab</i> Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs
10-03	<i>A. Cohen, R. DeVore and C. Schwab</i> Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs
10-02	<i>V. Gradinaru, G.A. Hagedorn, A. Joye</i> Tunneling dynamics and spawning with adaptive semi-classical wave-packets