

B-spline modal method: A polynomial approach compared to the Fourier modal method

Michael Walz,¹ Thomas Zebrowski,¹ Jens Küchenmeister,¹ and Kurt Busch^{2,*}

¹ *Institut für Theoretische Festkörperphysik (TFP) and DFG-Center for Functional Nanostructures (CFN), Karlsruhe Institute of Technology (KIT), Wolfgang-Gaede-Str. 1, 76131 Karlsruhe, Germany*

² *Humboldt-Universität zu Berlin, Institut für Physik, AG Theoretische Optik & Photonik, Newtonstr. 15, 12489 Berlin, and Max-Born-Institut, Max-Born-Str. 2A, 12489 Berlin, Germany*

[*kurt.busch@physik.hu-berlin.de](mailto:kurt.busch@physik.hu-berlin.de)

Abstract: A detailed analysis of the B-spline Modal Method (BMM) for one- and two-dimensional diffraction gratings and a comparison to the Fourier Modal Method (FMM) is presented. Owing to its intrinsic capability to accurately resolve discontinuities, BMM avoids the notorious problems of FMM that are associated with the Gibbs phenomenon. As a result, BMM facilitates significantly more efficient eigenmode computations. With regard to BMM-based transmission and reflection computations, it is demonstrated that a novel Galerkin approach (in conjunction with a scattering-matrix algorithm) allows for an improved field matching between different layers. This approach is superior relative to the traditional point-wise field matching. Moreover, only this novel Galerkin approach allows for a competitive extension of BMM to the case of two-dimensional diffraction gratings. These improvements will be very useful for high-accuracy grating computations in general and for the analysis of associated electromagnetic field profiles in particular.

© 2013 Optical Society of America

OCIS codes: (050.1970) Diffractive optics; (050.1755) Computational electromagnetic methods; (160.5298) Photonic crystals.

References and links

1. K. Busch, G. von Freymann, S. Linden, S. F. Mingaleev, L. Tkeshelashvili, and M. Wegener, "Periodic nanostructures for photonics," *Phys. Rep.* **444**, 101–202 (2007).
2. L. Li, "New formulation of the Fourier modal method for crossed surface-relief gratings," *J. Opt. Soc. Am. A* **14**, 2758–2767 (1997).
3. L. Li, "Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings," *J. Opt. Soc. Am. A* **13**, 1024–1035 (1996).
4. K. Edee, P. Schiavone, and G. Granet, "Analysis of defect in extreme UV lithography mask using a modal method based on nodal B-spline expansion," *Jpn. J. Appl. Phys.* **44**, 6458–6462 (2005).
5. P. Bouchon, F. Pardo, R. Härdar, and J.-L. Pelouard, "Fast modal method for subwavelength gratings based on B-spline formulation," *J. Opt. Soc. Am. A* **27**, 696–702 (2010).
6. A. Buffa, G. Sangalli, and R. Vazquez, "Isogeometric analysis in electromagnetics: B-splines approximation," *Comput. Method. Appl. M.* **199**, 1143–1152 (2010).
7. C. de Boor, "On calculating with B-splines," *J. Approx. Theory* **6**, 50–62 (1972).
8. M. G. Cox, "The numerical evaluation of B-Splines," *IMA J. Appl. Math.* **10**, 134–149 (1972).

9. C. de Boor, *A Practical Guide to Splines* (Springer, 2001).
10. P. Lalanne, and G. M. Morris "Highly improved convergence of the coupled-wave method for TM polarization," *J. Opt. Soc. Am. A* **13**, 779–784 (1996).
11. G. Granet and B. Guizal, "Efficient implementation of the coupled-wave method for metallic lamellar gratings in TM polarization," *J. Opt. Soc. Am. A* **13**, 1019–1023 (1996).
12. Z. Sacks, D. Kingsland, R. Lee, and J.-F. Lee, "A perfectly matched anisotropic absorber for use as an absorbing boundary condition," *IEEE T. Antenn. Propag.* **43**, 1460–1463 (1995).
13. T. Weiss, G. Granet, N. A. Gippius, S. G. Tikhodeev, and H. Giessen, "Matched coordinates and adaptive spatial resolution in the Fourier modal method," *Opt. Express* **17**, 8051–8061 (2009).
14. S. Essig and K. Busch, "Generation of adaptive coordinates and their use in the Fourier Modal Method," *Opt. Express* **18**, 23258 (2010).
15. J. Küchenmeister, T. Zebrowski, and K. Busch, "A construction guide to analytically generated meshes for the Fourier Modal Method," *Opt. Express* **20**, 17319–17347 (2012).
16. E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide* (SIAM, Philadelphia, 1999).
17. R. B. Lehoucq, *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods* (SIAM, Philadelphia, 1998).
18. A. W. Snyder and J. D. Love, *Optical Waveguide Theory* (Chapman and Hall, 1983).
19. E. H. Moore, "On the reciprocal of the general algebraic matrix," *Bull. Amer. Math. Soc.* **26**, 394–395 (1920).
20. R. Penrose, "A generalized inverse for matrices," *Math. Proc. Cambridge* **51**, 406–413 (1955).
21. E. Noponen and J. Turunen, "Eigenmode method for electromagnetic synthesis of diffractive elements with three-dimensional profiles," *J. Opt. Soc. Am. A* **11**, 2494–2502 (1994).

1. Introduction

Transmittance and reflectance computations from periodic photonic structures such as diffraction gratings and photonic crystals are of considerable interest as appropriately designed structures facilitate a far-reaching control over light propagation and light-matter interaction [1]. The Fourier Modal Method (FMM) represents a standard tool for such type of computations [2].

In this work, we analyze the B-spline Modal Method (BMM) which uses the general idea that underlies FMM but instead of a Fourier basis it uses a B-spline basis. As a matter of fact, the basic approach of modal methods such as FMM or BMM is that, in propagation direction, the structure of interest is approximated by several layers, where each layer is homogeneous along the stacking direction, e.g., the z -direction. This allows, in each layer, a plane wave ansatz $e^{i\lambda z}$ for the z -dependence that transforms Maxwell equations in frequency domain into an eigenvalue problem for the corresponding propagation constants λ and associated eigenmodes. This eigenvalue problem is solved for all eigenmodes and the electromagnetic field is expanded into these eigenmodes. Different layers are subsequently connected via a scattering-matrix algorithm that utilizes the continuity conditions for the tangential \mathbf{E} - and \mathbf{H} -field across the interface between adjacent layers [3]. In essence, this procedure reconstructs the electromagnetic field in the entire structure.

The main difference between FMM and BMM is that the FMM uses a Fourier basis for discretizing the eigenvalue problem while the BMM uses B-splines. The use of a finite Fourier basis in FMM is problematic because a finite Fourier series will always be infinitely many times continuously differentiable, and therefore, the FMM is unable to accurately represent cusps or discontinuities in the fields. Away from material interfaces, this is not a problem because there the fields themselves are infinitely many times continuously differentiable but at interfaces the fields naturally become discontinuous or at least are no longer continuously differentiable. To overcome this problem, we use a finite B-spline basis which is smooth away from any interface but can represent cusps and discontinuities at interfaces as detailed below. Naturally, B-splines have already been used successfully for modelling electromagnetic fields and their properties at interfaces in other works, e.g., see [4–6]. In particular, Bouchon *et al.* [5] have recently introduced B-splines for use in one-dimensional grating computations based on a point-wise field matching within the scattering-matrix algorithm. In this work, we show that a novel Galerkin

approach to the field matching considerably improves accuracy and efficiency of BMM computations for one-dimensional gratings. We further show how this Galerkin approach facilitates the extension of BMM to two-dimensional gratings. A detailed comparison with corresponding FMM computations suggests that BMM with Galerkin-based field matching is a competitive alternative to FMM.

The paper is organized as follows: In Sec. 2, we introduce the B-spline basis and develop in Sec. 3 the theoretical framework of BMM for both one- and two-dimensional systems along with corresponding numerical results. We close with a discussion and an outlook in Sec. 4.

2. B-splines

In a nutshell, B-splines are localized piecewise polynomial functions with conveniently adjustable (non)-continuous derivatives. In our implementation, we use B-splines as defined by the recurrence relation of de Boor and Cox [7, 8]. An overview of B-spline techniques can, for example, be found in [9].

Specifically, a one-dimensional B-spline of degree n is a function that is defined piecewise by polynomial segments of degree n between adjacent knots t_i . At single knots, the B-splines segments are connected so that the first $(n - 1)$ derivatives are continuous. The general advantage of B-splines is that one can place m knots on top of each other, i.e., a single knot may exhibit a multiplicity m . Then, only the first $(n - m)$ derivatives are continuous. The idea here is to place several knots at material interfaces and to use only single knots away from interfaces. Then, the B-splines can model cusps or jumps at interfaces while still being many times differentiable away from interfaces.

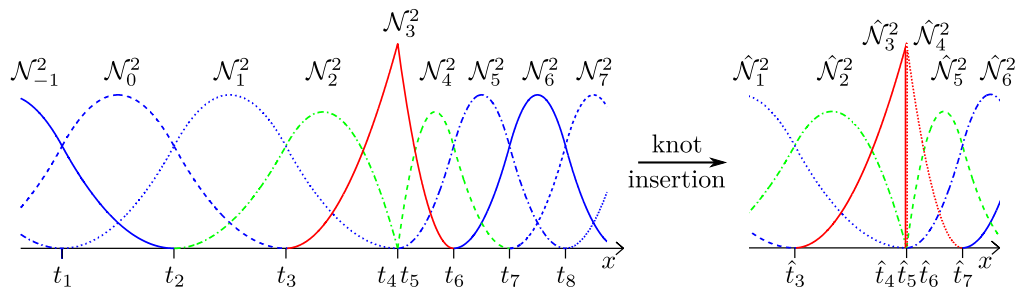


Fig. 1. B-splines \mathcal{N}_i^n of degree $n = 2$ defined over a knot sequence t_i with a degenerate knot at $x = t_4 = t_5$. The blue B-splines are continuously differentiable everywhere, whereas the green B-splines, \mathcal{N}_2^2 and \mathcal{N}_4^2 , have discontinuous derivatives at the degenerate knot. The red B-spline \mathcal{N}_3^2 is continuous but not differentiable at the $(m = n)$ -fold knot (left picture) and if an additional knot \hat{t}_6 is inserted (right panel) splits into two discontinuous B-splines, $\hat{\mathcal{N}}_3^2$ and $\hat{\mathcal{N}}_4^2$, featuring a jump at the additionally inserted knot.

The left of Fig. 1 displays B-splines of degree $n = 2$ with a knot of multiplicity $m = n$, i.e., already the first derivative has a discontinuity at $x = t_4 = t_5$ —and this is the position at which we assume a material interface to be located. As mentioned above, these discontinuities are the major difference to a Fourier basis which is infinitely many times differentiable. Together, the set of all B-splines of degree n forms a basis for all piecewise polynomial functions of degree less or equal n which exhibit the same smoothness properties at the knots, i.e., they are at least $(n - m)$ times continuously differentiable. Such a function $f(x)$ is given by the expansion $f(x) = \sum_{i=1}^N c_i \mathcal{N}_i^n(x)$ where c_i are the expansion coefficients and N denotes the number of B-splines.

Since the B-splines form a basis for all piecewise polynomial functions that comply with

the smoothness properties at the knots, adding an additional knot extends the function space so that the original function space is fully contained in the extended space. Hence, adding a knot to an already existing knot sequence generates new B-splines $\hat{\mathcal{N}}_i^n$ which (i) can still exactly represent all the original B-splines and (ii) acquire the extra freedom that one further derivative may be discontinuous at the newly inserted knot. Therefore, it is possible to add knots into an already existing knot sequence and calculate new coefficients \hat{c}_i in terms of the original ones by using a knot insertion algorithm [9, Chap. XI]. Then, the new and the old coefficients represent exactly the same function, e.g.,

$$f(x) = \sum_{i=1}^N c_i \mathcal{N}_i^n(x) = \sum_{i=1}^{N+1} \hat{c}_i \hat{\mathcal{N}}_i^n(x). \quad (1)$$

In this work, we only need to add one additional knot to an already n -fold knot, creating an $(n+1)$ -fold knot as depicted in Fig. 1. This is a simplified case where the knot insertion simply splits a B-spline \mathcal{N}_k into two new B-splines $\hat{\mathcal{N}}_k$ and $\hat{\mathcal{N}}_{k+1}$ ($k=3$ in the example depicted in Fig. 1). All the other B-splines remain unchanged—with the exception of a shift of index by one. The new coefficients are then given by

$$\hat{c}_i = \begin{cases} c_i & \text{for } i \leq k, \\ c_{i-1} & \text{for } i \geq k+1. \end{cases} \quad (2)$$

In other words: The old coefficient c_k is used twice since its corresponding B-spline \mathcal{N}_k is split into $\hat{\mathcal{N}}_k$ and $\hat{\mathcal{N}}_{k+1}$. In the general case, adding a knot would alter all $(n+1)$ B-splines which reach over the new knot and recalculating the coefficients would be more complicated.

In our case, we use the knot insertion algorithm because we have to deal with fields that exhibit different smoothness properties. For instance, the H_y -field in Sec. 3.1 exhibits a cusp at material interfaces for constant values of x but the E_x -field which is computed from the H_y -field in Sec. 3.2 features a discontinuity. Therefore, an additional knot has to be inserted into the B-spline basis.

3. B-spline modal method for one-dimensional gratings

Within BMM, we use B-splines as basis functions with multiple knots placed at every interface. This ensures that the smoothness conditions of the fields can be modeled. In Sec. 3.1, we first analyze a single one-dimensional layer and show how to discretize the eigenvalue problem using a Galerkin choice. We compare the convergence of our calculated guided eigenmodes with the convergence of the FMM eigenmodes. Next, in Sec. 3.2, a scattering-matrix algorithm is used to match the fields across adjacent layers. Following our Galerkin approach, this leads to non-square matrices which are inverted using special techniques. Again, a numerical comparison to the FMM is given. Finally, in Sec. 3.3, we present the eigenvalue problem for a two-dimensional layer which differs qualitatively from the FMM since in the BMM a product of operators cannot be discretized by discretizing both operators separately.

3.1. One-dimensional layers

For a one-dimensional problem neither the material distribution nor the resulting fields depend on the y -coordinate. Further, the system is homogeneous in z -direction. As a result, all derivatives with respect to y disappear from the Maxwell equations. For linear constitutive relations, $\mathbf{D} = \underline{\epsilon}\mathbf{E}$ and $\mathbf{B} = \underline{\mu}\mathbf{H}$, the Maxwell curl equations in dimensionless units are

$$\nabla \times \mathbf{E} = i\omega \underline{\mu}\mathbf{H}, \quad \nabla \times \mathbf{H} = -i\omega \underline{\epsilon}\mathbf{E}. \quad (3)$$

Here, and throughout the entire manuscript, we use a harmonic time-dependence of $e^{-i\omega t}$. For diagonal permittivity tensor $\underline{\epsilon}$ and diagonal permeability tensor $\underline{\mu}$, Eqs. (3) decouple into the so-called TE and TM polarization problems. Using a plane wave ansatz $e^{i\lambda z}$ for the z -dependence, we obtain a generalized eigenvalue problem with eigenvalue λ^2 . Historically, the TM case has been considerably more challenging because it involves discontinuities in the fields [10, 11]. Therefore, we only focus on this case but would like to note that the TE case follows from the TM case by interchanging $\mathbf{E} \leftrightarrow -\mathbf{H}$ and $\epsilon \leftrightarrow \mu$. Explicitly, the generalized eigenvalue problem for the TM case reads

$$\underbrace{\left[\mu_y(x) + \frac{\partial_x}{\omega} \frac{1}{\epsilon_z(x)} \frac{\partial_x}{\omega} \right]}_{=: \mathcal{A}} H_y(x) = \frac{\lambda^2}{\omega^2} \underbrace{\left[\frac{1}{\epsilon_x(x)} \right]}_{=: \mathcal{B}} H_y(x). \quad (4)$$

In the above equation, we have introduced the operators \mathcal{A} and \mathcal{B} which have to be discretized by a B-spline expansion. The remaining non-vanishing fields E_x and E_z can be determined from knowledge of the H_y -field via

$$E_x = \frac{\lambda}{\omega \epsilon_x} H_y, \quad \text{and} \quad E_z = \frac{i}{\omega \epsilon_z} \partial_x H_y. \quad (5)$$

For the TM-polarized case, the fields E_y, H_x, H_z are zero. We expand the H_y -field into B-splines, $H_y(x) = \sum_{i=1}^N c_i \mathcal{N}_i^n(x)$. So far, we have followed Bouchon *et al.* [5]. However, instead of proceeding with their point collocation method, we employ below a Galerkin method to discretize the operators \mathcal{A} and \mathcal{B} , i.e., $A_{ji} = \int \mathcal{N}_j^{n*} \mathcal{A} \mathcal{N}_i^n dx$ and $B_{ji} = \int \mathcal{N}_j^{n*} \mathcal{B} \mathcal{N}_i^n dx$, where the asterisk * stands for complex conjugation. Formally, we multiply Eq. (4) by $\mathcal{N}_j^{n*}(x)$, insert the expansion for H_y , and integrate over the computational domain, i.e., the unit cell of the one-dimensional grating. Thus, the B-splines are used as basis as well as test functions, i.e., a Galerkin approach. The resulting discretized generalized eigenvalue problem may be cast in matrix-vector form and reads

$$\mathbf{A} \mathbf{c} = \frac{\lambda^2}{\omega^2} \mathbf{B} \mathbf{c}, \quad (6)$$

with discretized forms of the operators \mathcal{A} and \mathcal{B} given by

$$A_{ji} = \int \mathcal{N}_j^{n*} \mu_y \mathcal{N}_i^n dx - \int \left[\frac{\partial_x}{\omega} \mathcal{N}_j^{n*} \right] \frac{1}{\epsilon_z} \left[\frac{\partial_x}{\omega} \mathcal{N}_i^n \right] dx, \quad \text{and} \quad B_{ji} = \int \mathcal{N}_j^{n*} \frac{1}{\epsilon_x} \mathcal{N}_i^n dx. \quad (7)$$

At this point, we would like to note that we have used integration by parts to avoid having to compute $\partial_x \epsilon_z^{-1}$ in the operator \mathcal{A} which—for the very important case of piece-wise constant material parameters—is a sum of delta peaks located at the material interfaces. Relative to a collocation method, this represents a significant advantage as this complete treatment of $\partial_x \epsilon_z^{-1}$ naturally leads to the continuity of $E_z \propto \epsilon_z^{-1} \partial_x H_y$. In the collocation method, this continuity has instead to be enforced "by hand" because it is rather difficult to completely include the term $\partial_x \epsilon_z^{-1}$ at material interfaces (see the discussion in [5]). Further, completely including $\partial_x \epsilon_z^{-1}$ allows us also to treat spatially nonconstant permittivities, arising either naturally or artificially. For instance, the latter includes employing perfectly matched layers (PMLs) such as uniaxial PMLs as first discussed by Sacks *et al.* [12] or the entire framework of coordinate transformations developed within FMM [13–15]. We return to the use of coordinate transformation techniques when extending BMM to two-dimensional systems in Sec. 3.3. Finally, we want to note that due to the finite support of the B-splines, no boundary terms arise when using integration by parts. We compute the required overlap integrals numerically using appropriate Gaussian quadrature schemes.

The resulting eigenvalue problems can be solved by standard linear algebra tools, i.e., LAPACK [16] or—since the matrices are sparse due to the finite support of the B-splines—by ARPACK [17]. As mentioned above, we have to solve the eigenvalue problem in all layers. For each layer (p), this gives a set of N eigenvalues $(\lambda_m^{(p)})^2$ and associated eigenvectors $\mathbf{c}_m^{(p)}$, $m = 1, \dots, N$. We arrange the eigenvectors $\mathbf{c}_m^{(p)}$ as the columns of the eigenmode matrix $\check{\mathbf{H}}_{im}^{(p)}$. Then, the m^{th} eigenmode $\mathbf{H}_{y,m}^{(p)}(x)$ in layer (p) can be written as

$$\mathbf{H}_{y,m}^{(p)}(x) = \sum_{i=1}^N \check{\mathbf{H}}_{im}^{(p)} \mathcal{N}_i^n(x). \quad (8)$$

Here, we use of a sans-serif typeface for the numerically computed eigenmode. We will use the entire eigenmode matrix $\check{\mathbf{H}}$ in Sec. 3.2 when matching the fields by means of the scattering-matrix algorithm. Before we do so, an accuracy analysis of the eigenmode computations is in order.

3.1.1. Accuracy analysis of eigenmode computations

As a test system, we investigate a one-dimensional grating layer with a periodically repeated unit cell of $a = 10 \mu\text{m}$ as depicted in layer (2) in Fig. 2(a). The unit cell is divided into two regions, an air part with isotropic permittivity $\varepsilon_1 = 1$ of $9 \mu\text{m}$ width and a material part with isotropic permittivity $\varepsilon_2 = 5$ of $1 \mu\text{m}$ width. The permeability is assumed to be unity, $\mu = 1$.

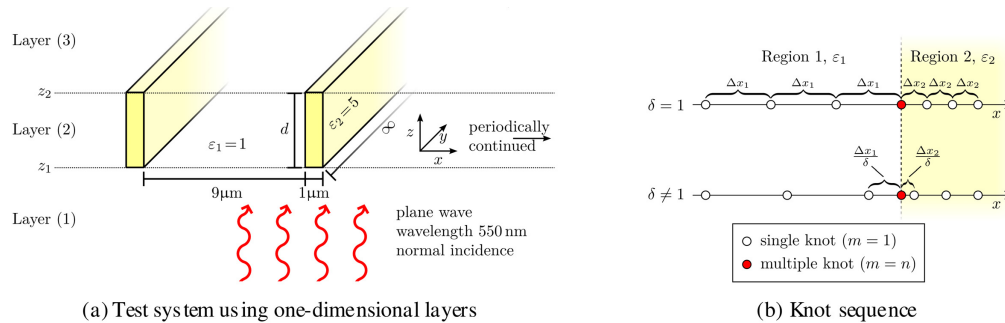


Fig. 2. (a) A periodic three-layer system invariant in y -direction with a unit cell size in x -direction of $a = 10 \mu\text{m}$. The unit cell of the second layer is divided into two regions, an air part with $\varepsilon = 1$ of $9 \mu\text{m}$ width and a waveguide part with $\varepsilon = 5$ of $1 \mu\text{m}$ width. Layers (1) and (3) denote homogeneous half spaces filled with air ($\varepsilon = 1$). (b) A knot sequence as used in the calculation. In general, we place n -fold knots at every material interface and distribute the remaining knots so that each region contains the same number of knots. For $\delta = 1$ the knots are spaced equidistantly in each region. The *equidistant* spacing is called Δx_i . For $\delta \neq 1$, the distance of the first knot next to the interface is divided by δ as compared to an equidistant spacing. The remaining knots are again spaced equidistantly but, of course, with a slightly larger spacing (slightly larger than Δx_i).

We compute the eigenmodes (propagation constants and field distributions) for a value of the vacuum wavelength of 550 nm . For B-splines of degree $n = 4, 7, 10$, we employ a knot sequence with multiple knots, $m = n$, at each air-dielectric interface (cf. Fig. 2(b) with $\delta = 1$). First, the knots are equidistantly spaced in each region and this is indicated with $\delta = 1$ in Fig. 2(b). Below, we will also discuss the more general case of non-equidistant knots within a given material region ($\delta \neq 1$) which exhibits certain advantages near interfaces. In both cases, we extend the B-splines periodically so that our computation corresponds to the case of normal incidence.

More general Bloch conditions for oblique incidence can be included in a straightforward manner. The main point at this stage of the presentation is that we ensure the correct smoothness conditions for the B-splines at the material interfaces, i.e., by placing $m = n$ knots at every interface the B-splines are still continuous but their derivatives have a finite discontinuity as depicted in Fig. 1 at $x = t_4 = t_5$. Consequently, the continuity of the tangential H_y -field at the interface is enforced by design while its derivative $\partial_x H_y \propto \epsilon_z E_z$ may acquire a finite discontinuity which is required by the fact that since E_z is continuous across the interface but the material distribution ϵ_z is not.

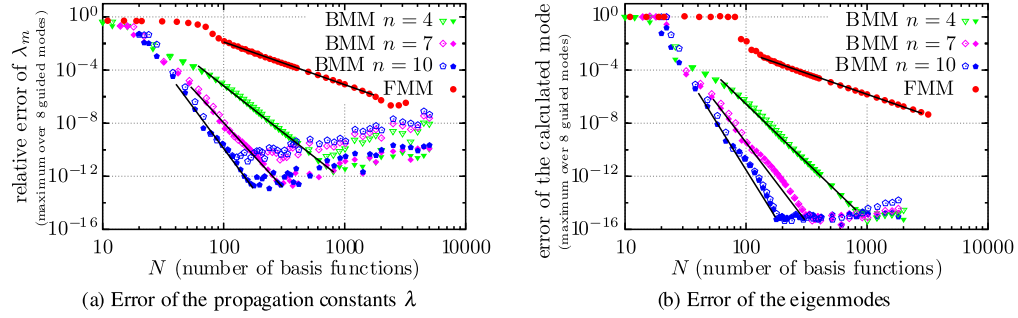


Fig. 3. Convergence plot of the errors of the propagation constant λ and the mode profiles. The convergence behaviour is fitted as N^{-r} which is shown as a straight line in the double logarithmic plot. The filled markers refer to a B-spline knot sequence with $\delta = 1$ whereas the open markers refer to $\delta = 10$; see the text and Fig. 2(b). (a) Convergence rates: $r_{\text{BMM}_4} = 7.1$, $r_{\text{BMM}_7} = 9.7$, $r_{\text{BMM}_{10}} = 11.9$, $r_{\text{FMM}} = 3.2$. (b) Convergence rates: $r_{\text{BMM}_4} = 8.7$, $r_{\text{BMM}_7} = 10.9$, $r_{\text{BMM}_{10}} = 12.9$, $r_{\text{FMM}} = 3.2$.

In Fig. 3, we compare the numerical results to an analytic solution, e.g., along the lines of [18, Sec. 12-15]. Specifically, we analyze the errors of both the propagation constants λ_m and the corresponding field distributions $H_{y,m}(x)$. We first consider equally spaced knots (filled markers in Fig. 3, $\delta = 1$). In Fig. 3(a), we present the dependence on the number of basis functions of the maximal relative error of the propagation constant for all guided modes. In this context, we define a mode being a guided mode if its propagation constant λ exceeds the vacuum wave vector (which, in dimensionless units equals the angular frequency ω). The error of the corresponding eigenmode field distributions is displayed in Fig. 3(b). Here, we measure the error of the eigenmode field distribution $H_y(x)$, via the deviation from the analytic solution $H_y(x)$. With both modes normalized, i.e. $\int dx |H_y(x)|^2 = 1$, the mode error thus reads

$$\text{error} = \frac{1}{2} \int dx |H_y(x) - H_y(x)|^2 = 1 - \text{Re} \int dx H_y^*(x) H_y(x). \quad (9)$$

The convergence rates of the BMM are far superior relative to the rates for FMM and they can be adjusted by the choice of the B-spline degree n . With increasing number N of basis functions, the relative error can easily be reduced to machine precision. Beyond this number, the error exhibits numerical artifacts due to the inaccuracies of the diagonalization routines. Further, for small N , we observe a plateau in the mode error which, however, is much more pronounced in the FMM than in the BMM. In this region, the FMM is unable to resolve the last mode ($\lambda \approx 1.013\omega$) above the guiding cut-off at $\lambda = \omega$.

We now consider an unequally spaced knot sequence (cf. Fig. 2(b) with $\delta \neq 1$; open markers in Fig. 3). As before, we use multiple knots at the material interfaces but the distance of the first knots on either side of the interfaces is divided by δ relative to an equidistant spacing. For the computations, we choose $\delta = 10$ as an illustrative example. With regard to the error

of the eigenvalues and field distributions of the eigenmodes, the impact of the modified knot sequence is moderately disappointing. Basically, an additional numerical error is introduced since the support of the B-splines exactly at the material interfaces is much smaller than without the modification. Since the matrix of the eigenproblem essentially consists of overlap integrals, this worsens the condition number of the eigenvalue problem. This is the reason for the increase of the numerical error for large N -values when using the modified knot sequence. However, in Sec. 3.2, we will demonstrate that the modified knot sequence features a less-obvious but rather important advantage with regards to the field matching within the S-matrix algorithm. For computing the eigenmode structure alone, the modified knot sequence, as described through a knot modification factor $\delta \neq 1$, is not helpful (but also not harmful).

3.2. Scattering-matrix algorithm

In this section, we demonstrate that our above-introduced Galerkin approach leads to rectangular, i.e., non-square matrices in the scattering-matrix algorithm. This represents non-trivial complications notably with regard to the interface matrices that are used to match the tangential fields between adjacent layers. Clearly, we cannot provide all the intricacies of the scattering-matrix algorithm and refer instead to the lucid exposition of Li [3]. More precisely, we will present our development up to the point where we reestablish Eqs. (2a) and (7) of [3], which represent the starting point for the scattering-matrix algorithm, i.e., from this point on the results of Li can be utilized without any change. In addition, our notation will closely follow his notation with the only exception that we have interchanged the meaning of the coordinates z and y .

In the case of one-dimensional layers in TM-polarization, the tangential fields consist only of H_y and E_x . Thus, our first task is to derive a representation for the electric field distributions of the eigenmodes $E_{x,m}^{(p)}(x)$ in terms of B-splines similar to Eq. (8), e.g.,

$$E_{x,m}^{(p)}(x) = \sum_{i=1}^{\hat{N}} \check{E}_{im}^{(p)} \mathcal{N}_i^{\hat{n}}(x). \quad (10)$$

Owing to the different continuity conditions for the electric field relative to the magnetic field (discontinuities vs. cusps) at the in-layer material interfaces (the interfaces at constant x), the B-splines for the electric fields cannot be the same as those for the magnetic fields $H_{y,m}^{(p)}(x)$. In order to distinguish the quantities, we use $\mathcal{N}_i^{\hat{n}}$ to denote the B-splines used as an expansion basis for the electric eigenmode and \hat{N} to denote their number. With the help of Eq. (5), we calculate the fields E_x from the fields H_y

$$E_{x,m}^{(p)}(x) \stackrel{(5)}{\propto} \frac{\lambda_m^{(p)}}{\epsilon_x^{(p)}(x)} H_{y,m}^{(p)}(x) \stackrel{(8)}{=} \sum_{i=1}^N \frac{\lambda_m^{(p)}}{\epsilon_x^{(p)}(x)} \check{H}_{im}^{(p)} \mathcal{N}_i(x). \quad (11)$$

This is not yet of the form of Eq. (10). We could easily define $\lambda_m^{(p)} \check{H}_{im}^{(p)}$ as $\check{E}_{im}^{(p)}$ but the permittivity $\epsilon_x^{(p)}(x)$ cannot be included because it is still x -dependent and, therefore, it has to be incorporated into the basis. However, this is a bit tricky since the permittivity features jumps at every interface while the B-splines \mathcal{N}_i have been chosen such that they are still continuous at the material interface (so that they can represent the magnetic field). Hence, we add an additional knot at every in-layer material interface so that the new B-splines $\mathcal{N}_i^{\hat{n}}$ can also represent jumps at the material interfaces. Then, each B-spline only contributes in a region with smooth permittivity. In our simple example, the permittivity is even piecewise constant and, therefore, strictly constant over the support of each new B-spline $\mathcal{N}_i^{\hat{n}}$. Consequently, the permittivity can simply be multiplied to the coefficients. By doing so, in analogy Eq. (2), we obtain the coefficients $\check{E}_{im}^{(p)}$ for the electric field distributions. For a spatially continuously varying permittivity,

an interpolation scheme has to be applied. This is possible since the B-splines form a basis for all piecewise polynomial functions that exhibit the correct smoothness properties at the knots. Further, we want to note that there is no need to include the angular frequency ω in Eq. (11) because it is constant over all modes and all layers. This may be regarded as matching $E_x \omega$ across the layers instead of E_x . At this point, we would like to emphasize that the above construction is facilitated by the special properties of the B-splines as described in Sec. 2.

Using the expansion matrices $\check{\mathbf{H}}^{(p)}$ and $\check{\mathbf{E}}^{(p)}$, we can combine the tangential fields H_y and E_x into a two-component tangential field vector \mathbf{F} . Its expansion into the eigenmodes can then be written in matrix notation as

$$\begin{aligned} \mathbf{F}^{(p)}(x, z) &:= \begin{pmatrix} H_y^{(p)}(x, z) \\ E_x^{(p)}(x, z) \end{pmatrix} = \underbrace{\begin{pmatrix} \mathcal{N}_1(x) \cdots \mathcal{N}_N(x) & 0 \\ 0 & \hat{\mathcal{N}}_1(x) \cdots \hat{\mathcal{N}}_{\hat{N}}(x) \end{pmatrix}}_{=: \mathbf{N}(x)} \underbrace{\begin{pmatrix} \check{\mathbf{H}}^{(p)} & \check{\mathbf{H}}^{(p)} \\ \check{\mathbf{E}}^{(p)} & -\check{\mathbf{E}}^{(p)} \end{pmatrix}}_{=: \mathbf{W}^{(p)}} \begin{pmatrix} \mathbf{u}^{(p)}(z) \\ \mathbf{d}^{(p)}(z) \end{pmatrix} \\ &= \mathbf{N}(x) \mathbf{W}^{(p)} \begin{pmatrix} \mathbf{u}^{(p)}(z) \\ \mathbf{d}^{(p)}(z) \end{pmatrix}. \end{aligned} \quad (12)$$

Here, we have introduced a shorthand notation for the B-spline basis, $\mathbf{N}(x)$, which is a $2 \times (N + \hat{N})$ -matrix that depends on the coordinate x . The B-splines are assumed to be the same in every layer to allow the subsequent matching procedure. Hence, $\mathbf{N}(x)$ does not carry a layer index. This is similar to the case of FMM computations where the same plane wave basis is used in every layer. Note, however, that for BMM this also means that the knot sequence is the same in every layer and must be adjusted to all interfaces in all layers. The new coefficients $u_m^{(p)}(z + \Delta z) = u_m^{(p)}(z) e^{+i\lambda_m^{(p)} \Delta z}$ and $d_m^{(p)}(z + \Delta z) = d_m^{(p)}(z) e^{-i\lambda_m^{(p)} \Delta z}$ denote upward and downward propagating or evanescently decaying eigenmodes, respectively, and are written as column vectors in Eq. (12). The matrix $\mathbf{W}^{(p)}$ consists of the eigenmode expansion coefficients. The minus sign stems from a phase difference between the downward travelling electric and magnetic modes. Its position inside the matrix $\mathbf{W}^{(p)}$ is arbitrary.

Finally, we are in a position to match the tangential fields \mathbf{F} at the layer interface at $z = z_p$ across layers (p) and $(p + 1)$. Since the B-splines $\mathbf{N}(x)$ are linearly independent (and the same in every layer), their coefficients must be the same:

$$\mathbf{F}^{(p+1)}(x, z_p + 0) \stackrel{!}{=} \mathbf{F}^{(p)}(x, z_p - 0) \quad \Rightarrow \quad \mathbf{W}^{(p+1)} \begin{pmatrix} \mathbf{u}^{(p+1)}(z_p + 0) \\ \mathbf{d}^{(p+1)}(z_p + 0) \end{pmatrix} \stackrel{!}{=} \mathbf{W}^{(p)} \begin{pmatrix} \mathbf{u}^{(p)}(z_p - 0) \\ \mathbf{d}^{(p)}(z_p - 0) \end{pmatrix} \quad (13) \text{ (2a in [3])}$$

In the next step, we now have to calculate an interface matrix $\mathbf{t}^{(p)}$ by inverting the matrix $\mathbf{W}^{(p+1)}$,

$$\mathbf{t}^{(p)} = [\mathbf{W}^{(p+1)}]^{-1} \mathbf{W}^{(p)}. \quad (14) \text{ (7 in [3])}$$

As indicated next to the equation numbering, Eqs. (13) and (14) are taken from [3]. However, in contrast to [3], our matrix $\mathbf{W}^{(p+1)}$ is not a square matrix but of dimensions $(2N + n_{\text{int}}) \times 2N$, where n_{int} denotes the number of in-layer material interfaces. We have used N basis functions for the magnetic field expansion but have used $\hat{N} = N + n_{\text{int}}$ basis functions for representing E_x accurately, i.e., we have added n_{int} knots, one for each in-layer interface. Hence, we have to elaborate on the notion of "matrix inversion"—clearly a rectangular matrix can only be inverted in an approximate sense. As a matter of fact, we have tried a number of different approaches in terms of accuracy and efficiency and have found that for our situation, this rectangular (non-square) matrix is best to be approximately inverted by means of a least-squares inverse. More

precisely, we call the matrix $[A]_{\text{ls}}^{-1}$ a least-squares inverse of the matrix A if $\hat{\mathbf{x}} = [A]_{\text{ls}}^{-1} \mathbf{b}$ approximately solves the linear equation $A\mathbf{x} = \mathbf{b}$ by minimizing the weighted squared deviation

$$S = (\mathbf{Ax} - \mathbf{b})^\dagger \mathbf{w} (\mathbf{Ax} - \mathbf{b}), \quad (15)$$

where \mathbf{w} denotes a Hermitian positive-definite weight matrix and the dagger symbol \dagger refers to Hermitian conjugation. Such a least-squares inverse is given by

$$[A]_{\text{ls}}^{-1} := (\mathbf{A}^\dagger \mathbf{w} \mathbf{A})^{-1} \mathbf{A}^\dagger \mathbf{w}. \quad (16)$$

This matrix only approximately inverts A because the matrix A has more rows than columns. However, we want to note that the least-squares inverse does not depend on \mathbf{b} . This is required for our computations (and excludes a number of other approaches for defining approximate inverse matrices for rectangular matrices) because an explicit left-hand-side vector for the scattering-matrix algorithm is not available. Rather, the scattering-matrix algorithm multiplies an entire matrix with the above inverse.

We utilize the above least-squares inverse to calculate the interface matrix $\mathbf{t}^{(p)}$ according to Eq. (14) by least-squares inverting the matrix $\mathbf{W}^{(p+1)}$,

$$\mathbf{t}^{(p)} = [\mathbf{W}^{(p+1)}]_{\text{ls}}^{-1} \mathbf{W}^{(p)}. \quad (17)$$

Here, we employ the basis function matrix \mathbf{N} (introduced in Eq. (12)) as the weight matrix

$$\mathbf{w} = \int \mathbf{N}^\dagger(x) \mathbf{N}(x) dx. \quad (18)$$

This particular weight matrix has the distinct advantage of a clear physical interpretation. As a matter of fact, the squared deviation S , that is minimized by the least-squares inverse, equals the squared deviation S' of the tangential fields integrated over the computational domain. This is seen as follows

$$\begin{aligned} S' &= \int \left| \mathbf{F}^{(p+1)}(x) - \mathbf{F}^{(p)}(x) \right|^2 dx = \int \left| \mathbf{N}(x) \left[\mathbf{W}^{(p+1)} \begin{pmatrix} \mathbf{u}^{(p+1)} \\ \mathbf{d}^{(p+1)} \end{pmatrix} - \mathbf{W}^{(p)} \begin{pmatrix} \mathbf{u}^{(p)} \\ \mathbf{d}^{(p)} \end{pmatrix} \right] \right|^2 dx \\ &= \left(\underbrace{\mathbf{W}^{(p+1)}}_A \underbrace{\begin{pmatrix} \mathbf{u}^{(p+1)} \\ \mathbf{d}^{(p+1)} \end{pmatrix}}_x - \underbrace{\mathbf{W}^{(p)}}_B \underbrace{\begin{pmatrix} \mathbf{u}^{(p)} \\ \mathbf{d}^{(p)} \end{pmatrix}}_x \right)^\dagger \underbrace{\int \mathbf{N}^\dagger(x) \mathbf{N}(x) dx}_w \left(\underbrace{\mathbf{W}^{(p+1)}}_A \underbrace{\begin{pmatrix} \mathbf{u}^{(p+1)} \\ \mathbf{d}^{(p+1)} \end{pmatrix}}_x - \underbrace{\mathbf{W}^{(p)}}_B \underbrace{\begin{pmatrix} \mathbf{u}^{(p)} \\ \mathbf{d}^{(p)} \end{pmatrix}}_x \right) \\ &= (\mathbf{Ax} - \mathbf{b})^\dagger \mathbf{w} (\mathbf{Ax} - \mathbf{b}) = S. \end{aligned} \quad (19)$$

Instead of a least-squares inverse, we have also employed a Moore-Penrose pseudoinverse [19, 20] which can be regarded as a least-squares inverse with a unit weight matrix, $\mathbf{w} = \mathbf{1}$. This does not minimize the tangential field deviations but minimizes the deviation of the field expansion coefficients which themselves do not have a clear physical meaning. In our computations, the use of a Moore-Penrose pseudoinverse always led to inferior results.

3.2.1. Validation of the Galerkin approach

In order to validate our Galerkin approach, we utilize the same system as depicted in Fig. 2(a) with a single layer thickness of $d = 70\text{nm}$ that is sandwiched between two half-spaces of air. The transmittance and reflectance can then be computed using our matching technique for rectangular matrices described above in combination with the scattering-matrix algorithm [3].

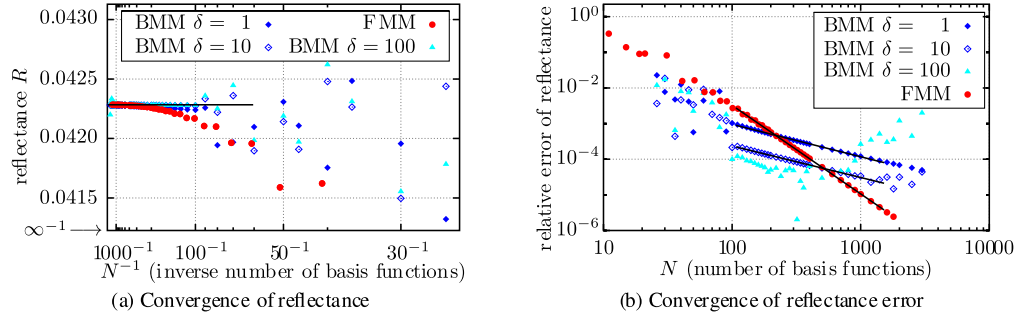


Fig. 4. Convergence plot of the reflectance calculation for the FMM and the BMM using B-splines of degree $n = 10$ with different knot sequences identified by the values of δ . (a) The black horizontal line at $R = 0.04228344$ is the fitted limiting value of the FMM calculation which we assume to converge to the correct value. (b) The relative error is calculated against the limit value of the FMM. Convergence rates: $r_{\text{BMM}_1} = 0.9$, $r_{\text{BMM}_{10}} = 0.9$, $r_{\text{FMM}} = 2.5$.

In Fig. 4, we depict the convergence characteristics of reflectance computations with regard to the number of basis functions. In the left panel, Fig. 4(a), we observe that the FMM and the different BMM computations converge to the same reflectance value $R = 0.04228344$ for $N \rightarrow \infty$ (which is at the left of the plot due to the use of N^{-1} for the abscissa). This allows us to take this limiting value and to determine the relative errors which we show in Fig. 4(b). For increasing values of the knot modification parameter δ the error for intermediate numbers of basis functions ($N \approx 100 - 500$) decreases. In other words, while the knot modification parameter does not increase the accuracy of the eigenmode computations it does decrease the errors resulting from the use of the least-square inverse for rectangular matrices. For larger values of N , the numerical error in the eigenvalue problem grows (cf. Fig. 3) and this leads to errors in the scattering-matrix algorithm. Asymptotically, the FMM shows a better convergence behavior but in the intermediate regime (relative accuracies down to 10^{-3}), the BMM features smaller errors. This is an important result since the amount of basis functions is severely limited when moving from one-dimensional to two-dimensional layers later on. For instance, taking 500 basis functions per direction, i.e., 25000 basis function for a two-dimensional computation, is currently impractical within FMM. Here, the BMM provides a possibility of using less basis functions while still computing highly accurate results.

3.3. B-spline modal method for two-dimensional gratings

For two-dimensional diffraction gratings the computation of eigenmodes for two-dimensional layers is required. This is tantamount to solving the full three-dimensional Maxwell equations, Eq. (3), with a plane-wave ansatz (propagation constant λ) in propagation direction (the z -direction in which the system is homogeneous) and a two-dimensional lateral unit cell with periodic (or Bloch-type) boundary conditions. In view of the recent progress in FMM computations using in-layer coordinate transformations [13–15], it is of special interest to consider materials with in-layer anisotropy even if one would "only" be interested in isotropic materials. For such in-layer anisotropic systems, the permittivity and permeability are of the form

$$\underline{\epsilon} = \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & 0 \\ \epsilon_{yx} & \epsilon_{yy} & 0 \\ 0 & 0 & \epsilon_{zz} \end{pmatrix} \quad \text{and} \quad \underline{\mu} = \begin{pmatrix} \mu_{xx} & \mu_{xy} & 0 \\ \mu_{yx} & \mu_{yy} & 0 \\ 0 & 0 & \mu_{zz} \end{pmatrix}. \quad (20)$$

The basic calculation is the same as that for the FMM (cf. Appx. A in [21]) but there is one crucial difference. For the BMM, we cannot separately discretize two operators \mathcal{F} and \mathcal{G} and subsequently multiply them. Rather, we have to discretize the entire operator product $\mathcal{F} \cdot \mathcal{G}$, i.e., within BMM we require an analytical expression for $\mathcal{F} \cdot \mathcal{G}$.

To obtain the corresponding eigenvalue problem, we start by eliminating the z -components E_z and H_z from the Maxwell equations, Eq. (3). Using Eq. (20) we arrive at an eigenvalue equation for the remaining four field components E_x , E_y , H_x and H_y that exhibits block anti-diagonal form

$$\underbrace{\begin{pmatrix} 0 & 0 & -\frac{\partial_x}{\omega} \frac{1}{\epsilon_{zz}} \frac{\partial_y}{\omega} + \mu_{yx} & +\frac{\partial_x}{\omega} \frac{1}{\epsilon_{zz}} \frac{\partial_x}{\omega} + \mu_{yy} \\ 0 & 0 & -\frac{\partial_y}{\omega} \frac{1}{\epsilon_{zz}} \frac{\partial_y}{\omega} - \mu_{xx} & +\frac{\partial_y}{\omega} \frac{1}{\epsilon_{zz}} \frac{\partial_x}{\omega} - \mu_{xy} \\ +\frac{\partial_x}{\omega} \frac{1}{\mu_{zz}} \frac{\partial_y}{\omega} - \epsilon_{yx} & -\frac{\partial_x}{\omega} \frac{1}{\mu_{zz}} \frac{\partial_x}{\omega} - \epsilon_{yy} & 0 & 0 \\ +\frac{\partial_y}{\omega} \frac{1}{\mu_{zz}} \frac{\partial_y}{\omega} + \epsilon_{xx} & -\frac{\partial_y}{\omega} \frac{1}{\mu_{zz}} \frac{\partial_x}{\omega} + \epsilon_{xy} & 0 & 0 \end{pmatrix}}_{=:\begin{pmatrix} 0 & \mathcal{F} \\ \mathcal{G} & 0 \end{pmatrix}} \begin{pmatrix} E_x \\ E_y \\ H_x \\ H_y \end{pmatrix} = \frac{\lambda}{\omega} \begin{pmatrix} E_x \\ E_y \\ H_x \\ H_y \end{pmatrix}. \quad (21)$$

Here, we have already replaced the z -derivative by $i\lambda$ due to the plane wave ansatz $e^{i\lambda z}$ for the z -dependence. In addition, we have introduced the 2×2 -matrix differential operators \mathcal{F} and \mathcal{G} . Upon rewriting Eq. (21) as

$$\mathcal{F} \cdot \mathcal{G} \begin{pmatrix} E_x \\ E_y \end{pmatrix} = \frac{\lambda^2}{\omega^2} \begin{pmatrix} E_x \\ E_y \end{pmatrix}, \quad (22)$$

we obtain the two-dimensional extension of the one-dimensional eigenvalue problem, Eq. (4). We discretize Eq. (22) via our Galerkin approach in complete analogy to the one-dimensional case as described in Sec. 3.1. A straightforward but laborious calculation yields

$$(\mathcal{F} \cdot \mathcal{G})_{11} = \frac{\partial_x}{\omega} \frac{1}{\epsilon_{zz}} \left[\frac{\partial_x}{\omega} \epsilon_{xx} + \frac{\partial_y}{\omega} \epsilon_{yx} \right] + \mu_{yx} \left[\frac{\partial_x}{\omega} \frac{1}{\mu_{zz}} \frac{\partial_y}{\omega} - \epsilon_{yx} \right] + \mu_{yy} \left[\frac{\partial_y}{\omega} \frac{1}{\mu_{zz}} \frac{\partial_y}{\omega} + \epsilon_{xx} \right], \quad (23a)$$

$$(\mathcal{F} \cdot \mathcal{G})_{12} = \frac{\partial_x}{\omega} \frac{1}{\epsilon_{zz}} \left[\frac{\partial_x}{\omega} \epsilon_{xy} + \frac{\partial_y}{\omega} \epsilon_{yy} \right] - \mu_{yx} \left[\frac{\partial_x}{\omega} \frac{1}{\mu_{zz}} \frac{\partial_x}{\omega} + \epsilon_{yy} \right] + \mu_{yy} \left[-\frac{\partial_y}{\omega} \frac{1}{\mu_{zz}} \frac{\partial_x}{\omega} + \epsilon_{xy} \right], \quad (23b)$$

$$(\mathcal{F} \cdot \mathcal{G})_{21} = \frac{\partial_y}{\omega} \frac{1}{\epsilon_{zz}} \left[\frac{\partial_x}{\omega} \epsilon_{xx} + \frac{\partial_y}{\omega} \epsilon_{yx} \right] + \mu_{xx} \left[-\frac{\partial_x}{\omega} \frac{1}{\mu_{zz}} \frac{\partial_y}{\omega} + \epsilon_{yx} \right] - \mu_{xy} \left[\frac{\partial_y}{\omega} \frac{1}{\mu_{zz}} \frac{\partial_y}{\omega} + \epsilon_{xx} \right], \quad (23c)$$

$$(\mathcal{F} \cdot \mathcal{G})_{22} = \frac{\partial_y}{\omega} \frac{1}{\epsilon_{zz}} \left[\frac{\partial_x}{\omega} \epsilon_{xy} + \frac{\partial_y}{\omega} \epsilon_{yy} \right] + \mu_{xx} \left[\frac{\partial_x}{\omega} \frac{1}{\mu_{zz}} \frac{\partial_x}{\omega} + \epsilon_{yy} \right] + \mu_{xy} \left[\frac{\partial_y}{\omega} \frac{1}{\mu_{zz}} \frac{\partial_x}{\omega} - \epsilon_{xy} \right]. \quad (23d)$$

Here, we want to note that although \mathcal{F} and \mathcal{G} are both second-order differential operators, their product is not of fourth but only of second order. This is the result of certain non-trivial cancellations.

Eq. (22) represents an eigenvalue problem for the electric fields. The corresponding magnetic fields can then be obtained via the relation $(H_x \ H_y)^T = \frac{\omega}{\lambda} \mathcal{G} (E_x \ E_y)^T$, (cf. Eq. (21)). Of course, it is also possible to formulate an eigenvalue problem for the magnetic fields. In this case, the relevant operator for the eigenvalue problem is $\mathcal{G} \cdot \mathcal{F}$ which—due to the symmetry of the Maxwell equations—can be easily obtained by interchanging $\underline{\epsilon}$ and $\underline{\mu}$ in the above expressions, Eqs. (23), for the product $\mathcal{F} \cdot \mathcal{G}$.

3.3.1. Validation and Accuracy Analysis

As a test system for two-dimensional layers, we investigate a circular dielectric waveguide with radius $r = 800\text{ nm}$ and isotropic permittivity $\varepsilon = 2$ placed in free space using a vacuum operation wavelength of 800 nm corresponding to an operation frequency of $f = \frac{\omega}{2\pi} = 375\text{ THz}$. The permeability is assumed to be unity everywhere, $\mu = 1$. We choose a square unit cell with a lattice constant $a = 4\text{ }\mu\text{m}$ that is periodically extended in x - and y -direction. Therefore, we are actually investigating a periodic array of circular waveguides. However, we have chosen the unit cell size sufficiently large (in our case $r = 0.2a$) so that the field distributions of the waveguide's lower guided modes are essentially zero at the cell boundaries so that the actual boundary conditions do not matter. In other words, we perform a supercell computation that allows us to compare with the analytical reference solution for a single waveguide as described by Snyder and Love [18, Sec. 12-8ff.].

As basis functions, we use two-dimensional B-splines $\mathcal{N}_{i,j}^n(x,y)$ formed by a tensor-product of one-dimensional B-splines,

$$\mathcal{N}_{i,j}^n(x,y) := \mathcal{N}_i^n(x) \cdot \mathcal{N}_j^n(y). \quad (24)$$

As before, we extend the B-splines periodically to implement periodic boundary conditions.

Owing to the choice of a tensor-product basis of B-splines, we are restricted to a tensor-product like knot sequence. Specifically, we employ an equidistant Cartesian mesh without any multiple knot lines so that we expect roughly the same characteristics as the standard FMM without adaptive coordinates or adaptive spatial resolution. Having N B-splines, \sqrt{N} per direction, the distance between two knots in x - or y -direction is given by $\frac{a}{\sqrt{N}}$.

Since we utilize in our test system only lossless and isotropic materials and do not employ in-layer coordinate transformations in the computations (which would be required when using adaptive coordinates and adaptive resolution) certain additional simplifications arise. For instance, upon using

$$\frac{1}{\varepsilon} \partial_i \varepsilon = \partial_i \ln(|\varepsilon|) + (1 - \ln(|\varepsilon|)) \partial_i, \quad (25)$$

the (11)-component of the operator product $\mathcal{F} \cdot \mathcal{G}$ in Eq. (23a) can be rewritten as

$$(\mathcal{F} \cdot \mathcal{G})_{11} = \mu \varepsilon + \frac{\partial_x^2}{\omega^2} \ln(|\varepsilon|) + \frac{\partial_x}{\omega} (1 - \ln(|\varepsilon|)) \frac{\partial_x}{\omega} + \ln(|\mu|) \frac{\partial_y^2}{\omega^2} + \frac{\partial_y}{\omega} (1 - \ln(|\mu|)) \frac{\partial_y}{\omega}. \quad (26)$$

The remaining components of $\mathcal{F} \cdot \mathcal{G}$ in Eqs. (23) can be processed in an analogous manner.

Transformed expressions of the type displayed in Eq. (26) are more convenient for discretization via a Galerkin choice. In particular, the derivative operators only appear on the right hand side or the left hand side of the operator $\mathcal{F} \cdot \mathcal{G}$. Therefore, after an integration by parts they only act on the B-splines similar to the one-dimensional case.

However, we would like to emphasize that for the general in-layer anisotropic case the formulas given in Eqs. (23) need to be applied. This would require to deal with derivatives of ε and μ directly. Because of discontinuities in the material parameters, additional care is indispensable. One approach is to dismantle the discontinuous part as a sum of Heaviside step functions which are centered at the in-layer interfaces,

$$\varepsilon_{ab}(x,y) = \sum_{ij} \Theta(x-x_i) \Theta(y-y_j) \tilde{\varepsilon}_{ab}^{ij}(x,y), \quad \mu_{ab} \text{ analogous.} \quad (27)$$

The derivatives of the Heaviside step functions are Dirac delta functions which can be handled analytically when performing the integration in the Galerkin discretization scheme. The

remaining parts $\tilde{\epsilon}_{ab}^{ij}$ and $\tilde{\mu}_{ab}^{ij}$ are continuously differentiable and can therefore be treated numerically. Clearly, a decomposition as in Eq. (27) requires that all in-layer interfaces are aligned to the coordinate lines and would thus require implementing a coordinate transformation as it is routinely employed in the FMM. We further discuss this issue in the outlook, Sec. 4, and display a possible coordinate transformation in Fig. 5(b).

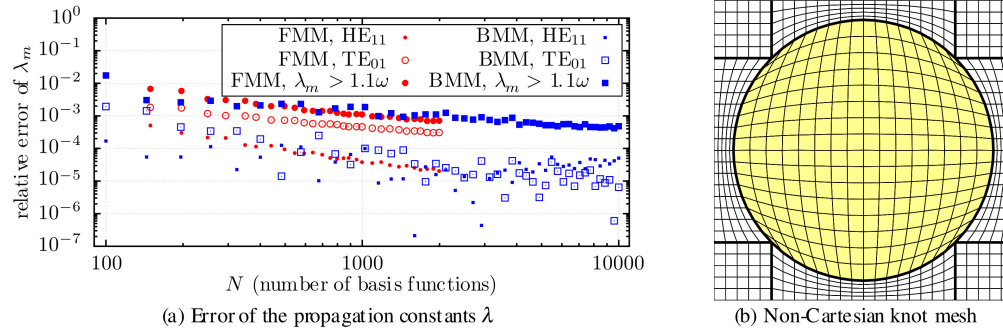


Fig. 5. (a) Comparison of the error of the propagation constants using an FMM and a BMM calculation using B-splines of degree $n = 7$. (b) An illustration of a non-Cartesian knot mesh with multiple knot lines (bold lines) directly at the material interfaces.

In Fig. 5(a), we compare the error of the propagation constants λ_m for BMM computations outlined above (square symbols) with the corresponding standard FMM computations (circular symbols). Specifically, we depict the relative error of the mode with the largest and second largest propagation constants λ_m , i.e., of the HE_{11} (small filled symbols) and TE_{01} mode (open symbols), respectively. In addition, we also depict the maximal error of all guided modes with $\lambda_m > 1.1\omega$ (large filled symbols). Here, we refrain from using all available guided modes up to the guiding cut-off, i.e., all modes with $\lambda = \omega$. The reason is that for our particular choice of geometrical parameters, the analytically available reference modes for a single waveguide near cut-off exhibit spatial profiles that are too extended for the numerical supercell computations to deliver accurate results.

From Fig. 5(a) we infer similar convergence characteristics for BMM and standard FMM upon increasing the number of basis functions. As described above, we have anticipated this behavior because in the absence of multiple knot lines in the BMM computations, the source of error is essentially the same for BMM and standard FMM: Neither the BMM nor the standard FMM is capable of accurately resolving cusps or discontinuities of the fields at material interfaces. However, the TE_{01} -mode does neither exhibit a discontinuity or cusp in the relevant field components—and for this mode the BMM computations provide results that, for the same number of basis functions, are about a factor of 10 more accurate than those of standard FMM computations. The main advantage that BMM offers is that, for the same computational effort, we can employ more basis functions N than standard FMM. Solving only the dense eigenvalue problem for FMM took 19 min 07 sec at $N = 1997$ on a quad core computer with 2.66GHz and 8 GByte RAM (Intel[®] Core[™] 2 Quad CPU Q9400@2.66GHz). The entire BMM eigenmode computation $N = 10000$ (which includes computing the B-splines, the integrals for discretizing $\mathcal{F} \cdot \mathcal{G}$, and solving the sparse eigenvalue problem for all guided eigenmodes) required only 12 min 51 sec on the same computer. Similarly, the memory requirements are such that at $N = 10000$, already the storage of the discretized operator $\mathcal{F} \cdot \mathcal{G}$ in a dense matrix—as used by the FMM—would require 3.0 GByte whereas 137 MByte suffice for the BMM exploiting its sparse matrix structure since only 2.3% of the entries are nonzero (for $n = 7$ as used in Fig. 5(a)). The sparseness obviously depends on the B-spline degree since the latter influences

the number of overlapping B-splines.

4. Conclusions and outlook

In summary, we have shown that BMM computations offer significant advantages over FMM computations. Specifically, we have shown that in one-dimensional gratings the placement of multiple knots at material interfaces allows to accurately treat all discontinuities and cusps of the corresponding electric and magnetic fields. Together with our novel Galerkin approach, this allows for very efficient eigenmode computations. As a result, this Galerkin approach leads to rectangular, i.e., non-square matrices when matching the fields between different layers within the scattering-matrix algorithm. We have further shown how to deal with these rectangular matrices by way of a least-square inverse and that this facilitates highly accurate and efficient transmittance and reflectance computations. In addition, we have demonstrated how to extend our Galerkin approach to the case of two-dimensional gratings. Even when using simple tensor-product B-splines without multiple knots as basis sets, the resulting BMM computations are considerably more efficient than standard FMM computations, while showing the same accuracy.

Clearly, realizing the full power of BMM computations for general two-dimensional gratings requires the use of adaptive coordinates that allow for placing multiple knot lines at curved material interfaces. In complete analogy to FMM, these coordinate transformations may be constructed as shown in [13, 15] and we depict an example of such curvilinear coordinates with multiple knot lines in Fig. 5(b). Implementing such coordinate transformations is equivalent to treating in-plane anisotropic materials and we have derived the corresponding eigenvalue problem in Eqs. (23). As noted above, solving the general eigenvalue problem, Eqs. (23), requires the consistent handling of the derivatives of the material parameters. This, together with a full convergence analysis of BMM using adaptive coordinates and multiple knot lines is outside the scope of the present manuscript and we leave it for future work.

Acknowledgments

We acknowledge support by the Deutsche Forschungsgemeinschaft (DFG) and the State of Baden-Württemberg through the DFG-Center for Functional Nanostructures (CFN) within sub-project A1.1. T.Z. acknowledges funding by the Carl Zeiss foundation and the Karlsruhe House of Young Scientists (KHYS) and his research is embedded in the Karlsruhe School of Optics & Photonics (KSOP). Furthermore, we acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of the Karlsruhe Institute of Technology (KIT).