# Robust fast direct integral equation solver for quasi-periodic scattering problems with a large number of layers

**Min Hyung Cho**[1,*] **and Alex H. Barnett**[1]

[1]*Department of Mathematics, Dartmouth College, Hanover, NH 03755, USA*
[*]*mhcho@math.dartmouth.edu*

**Abstract:** We present a new boundary integral formulation for time-harmonic wave diffraction from two-dimensional structures with many layers of arbitrary periodic shape, such as multilayer dielectric gratings in TM polarization. Our scheme is robust at all scattering parameters, unlike the conventional quasi-periodic Green's function method which fails whenever any of the layers approaches a Wood anomaly. We achieve this by a decomposition into near- and far-field contributions. The former uses the free-space Green's function in a second-kind integral equation on one period of the material interfaces and their immediate left and right neighbors; the latter uses proxy point sources and small least-squares solves (Schur complements) to represent the remaining contribution from distant copies. By using high-order discretization on interfaces (including those with corners), the number of unknowns per layer is kept small. We achieve overall linear complexity in the number of layers, by direct solution of the resulting block tridiagonal system. For device characterization we present an efficient method to sweep over multiple incident angles, and show a $25\times$ speedup over solving each angle independently. We solve the scattering from a 1000-layer structure with $3 \times 10^5$ unknowns to 9-digit accuracy in 2.5 minutes on a desktop workstation.

**OCIS codes:** (050.1755) Computational electromagnetic methods; (050.1950) Diffraction gratings; (160.2710) Inhomogeneous optical media.

---

## References and links

1. M. D. Perry, R. D. Boyd, J. A. Britten, D. Decker, B. W. Shore, C. Shannon, and E. Shults, "High-efficiency multilayer dielectric diffraction gratings," Opt. Lett. **20**, 940–942 (1995).
2. C. P. J. Barty, M. Key, J. Britten, R. Beach, G. Beer, C. Brown, S. Bryan, J. Caird, T. Carlson, J. Crane, J. Dawson, A. Erlandson, D. Fittinghoff, M. Hermann, C. Hoaglan, A. Iyer, L. J. II, I. Jovanovic, A. Komashko, O. Landen, Z. Liao, W. Molander, S. Mitchell, E. Moses, N. Nielsen, H.-H. Nguyen, J. Nissen, S. Payne, D. Pennington, L. Risinger, M. Rushford, K. Skulina, M. Spaeth, B. Stuart, G. Tietbohl, and B. Wattellier, "An overview of LLNL high-energy short-pulse technology for advanced radiography of laser fusion experiments," Nuclear Fusion **44**, S266 (2004).
3. G. A. Kalinchenko and A. M. Lerer, "Wideband all-dielectric diffraction grating on chirped mirror," J. Lightwave Technology **28**, 2743–49 (2010).
4. H. A. Atwater and A. Polman, "Plasmonics for improved photovoltaic devices," Nature Materials **9**, 205–213 (2010).
5. M. D. Kelzenberg, S. W. Boettcher, J. A. Petykiewicz, D. B. Turner-Evans, M. C. Putnam, E. L. Warren, J. M. Spurgeon, R. M. Briggs, N. S. Lewis, and H. A. Atwater, "Enhanced absorption and carrier collection in Si wire arrays for photovoltaic applications," Nature Materials **9**, 239–244 (2010).

6. N. P. Sergeant, M. Agrawal, and P. Peumans, "High performance solar-selective absorbers using coated sub-wavelength gratings," Opt. Express **18**, 5525–5540 (2010).

7. J. D. Joannopoulos, S. G. Johnson, R. D. Meade, and J. N. Winn, *Photonic Crystals: Molding the Flow of Light* (Princeton University, 2008), 2nd ed.

8. R. Model, A. Rathsfeld, H. Gross, M. Wurm, and B. Bodermann, "A scatterometry inverse problem in optical mask metrology," J. Phys.: Conf. Ser. **135**, 012071 (2008).

9. D. Colton and R. Kress, *Inverse acoustic and electromagnetic scattering theory*, vol. 93 of *Applied Mathematical Sciences* (Springer-Verlag, 1998), 2nd ed.

10. Bonnet-BenDhia, A.-S. and F. Starling, "Guided waves by electromagnetic gratings and non-uniqueness examples for the diffraction problem," Math. Meth. Appl. Sci. **17**, 305–338 (1994).

11. J. A. Stratton, *Electromagnetic Theory* (John Wiley & Sons, 2007).

12. J. D. Jackson, *Classical Electrodynamics* (Wiley, 1998), 3rd ed.

13. R. W. Wood, "On a remarkable case of uneven distribution of light in a diffraction grating spectrum," Philos. Mag. **4**, 396–408 (1902).

14. C. M. Linton and I. Thompson, "Resonant effects in scattering by periodic arrays," Wave Motion **44**, 165–175 (2007).

15. A. H. Barnett and L. Greengard, "A new integral representation for quasi-periodic scattering problems in two dimensions," BIT Numer. Math. **51**, 67–90 (2011).

16. S. Shipman, *Resonant scattering by open periodic waveguides* (Bentham Science Publishers, 2010), vol. 1 of *Progress in Computational Physics (PiCP)*, pp. 7–50.

17. G. L. Wojcik, J. Mould Jr., E. Marx, and M. P. Davidson, "Numerical reference models for optical metrology simulation," in "SPIE Microlithography 92: IC Metrology, Inspection, and Process Control VI, #1673-06," (1992).

18. A. Taflove, *Computational Electrodynamics: The Finite-Difference Time-Domain Method* (Artech House, 1995).

19. H. Holter and H. Steyskal, "Some experiences from FDTD analysis of infinite and finite multi-octave phased arrays," IEEE Trans. Antennae Prop. **50**, 1725–1731 (2002).

20. G. Bao and D. C. Dobson, "Modeling and optimal design of diffractive optical structures," Surv. Math. Ind. **8**, 37–62 (1998).

21. J. Elschner, R. Hinder, and G. Schmidt, "Finite element solution of conical diffraction problems," Adv. Comput. Math. **16**, 139–156 (2002).

22. I. M. Babuska and S. A. Sauter, "Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers?" SIAM J. Numer. Anal. **34**, 2392–2423 (1997).

23. M. G. Moharam and T. G. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," J. Opt. Soc. Am. **71**, 811–818 (1981).

24. L. Li, "Use of Fourier series in the analysis of discontinuous periodic structures," J. Opt. Soc. Am. A **13**, 1870–1876 (1996).

25. L. Li, "Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings," J. Opt. Soc. Am. A **13**, 1024–1035 (1996).

26. A. Lechleiter and D.-L. Nguyen, "A trigonometric Galerkin method for volume integral equations arising in TM grating scattering," Adv. Comput. Math. **40**, 1–25 (2014).

27. M. H. Cho and W. Cai, "A parallel fast algorithm for computing the Helmholtz integral operator in 3-D layered media," J. Comput. Phys. **231**, 5910–25 (2012).

28. D. Y. K. Ko and J. R. Sambles, "Scattering matrix method for propagation of radiation in stratified media: attenuated total reflection studies of liquid crystals," J. Opt. Soc. Am. A **5**, 1863–1866 (1988).

29. J. C. Nédélec and F. Starling, "Integral equation methods in a quasi-periodic diffraction problem for the time-harmonic Maxwell's equations," SIAM J. Math. Anal. **22**, 1679–1701 (1991).

30. A. Meier, T. Arens, S. N. Chandler-Wilde, and A. Kirsch, "A Nyström method for a class of integral equations on the real line with applications to scattering by diffraction gratings and rough surfaces," J. Integral Equations Appl. **12**, 281–321 (2000).

31. R. Kress, "Boundary integral equations in time-harmonic acoustic scattering," Mathl. Comput. Modelling **15**, 229–243 (1991).

32. S. Hao, A. H. Barnett, P. G. Martinsson, and P. Young, "High-order accurate Nyström discretization of integral equations with weakly singular kernels on smooth curves in the plane," Adv. Comput. Math. **40**, 245–272 (2014).

33. T. Arens, "Scattering by biperiodic layered media: The integral equation approach," Habilitation thesis, Karlsruhe (2010).

34. M. J. Nicholas, "A higher order numerical method for 3-D doubly periodic electromagnetic scattering problems," Commun. Math. Sci. **6**, 669–694 (2008).

35. Y. Otani and N. Nishimura, "A periodic FMM for Maxwell's equations in 3D and its applications to problems related to photonic crystals," J. Comput. Phys. **227**, 4630–52 (2008).

36. O. P. Bruno and M. C. Haslam, "Efficient high-order evaluation of scattering by periodic surfaces: deep gratings, high frequencies, and glancing incidences," J. Opt. Soc. Am. A **26**, 658–668 (2009).

37. A. Gillman and A. Barnett, "A fast direct solver for quasiperiodic scattering problems," J. Comput. Phys. **248**, 309–322 (2013).

38. L. Greengard, K. L. Ho, and J.-Y. Lee, "A fast direct solver for scattering from periodic structures with multiple material interfaces in two dimensions," J. Comput. Phys. **258**, 738–751 (2014).

39. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, 1964), 10th ed.

40. C. M. Linton, "Lattice sums for the Helmholtz equation," SIAM Review **52**, 603–674 (2010).

41. K. V. Horoshenkov and S. N. Chandler-Wilde, "Efficient calculation of two-dimensional periodic and waveguide acoustic Green's functions," J. Acoust. Soc. Amer. **111**, 1610–1622 (2002).

42. O. P. Bruno, S. Shipman, C. Turc, and S. Venakides, "Efficient evaluation of doubly periodic Green functions in 3D scattering, including Wood anomaly frequencies," (2013). Preprint, `arXiv:1307.1176v1`.

43. O. P. Bruno and B. Delourme, "Rapidly convergent two-dimensional quasi-periodic Green function throughout the spectrum—including Wood anomalies," J. Comput. Phys. **262**, 262–290 (2014).

44. A. H. Barnett and L. Greengard, "A new integral representation for quasi-periodic fields and its application to two-dimensional band structure calculations," J. Comput. Phys. **229**, 6898–6914 (2010).

45. A. Bogomolny, "Fundamental solutions method for elliptic boundary value problems," SIAM Journal on Numerical Analysis **22**, 644–669 (1985).

46. A. H. Barnett and T. Betcke, "Stability and convergence of the Method of Fundamental Solutions for Helmholtz problems on analytic domains," J. Comput. Phys. **227**, 7003–7026 (2008).

47. P. Martinsson and V. Rokhlin, "A fast direct solver for boundary integral equations in two dimensions," J. Comp. Phys. **205**, 1–23 (2005).

48. N. A. Gumerov and R. Duraiswami, "A method to compute periodic sums," J. Comput. Phys. **272**, 307–326 (2014).

49. C. Hafner, *The Generalized Multipole Technique for Computational Electromagnetics* (Artech House Books, 1990).

50. L. Zhao and A. H. Barnett, "Robust and efficient solution of the drum problem via Nyström approximation of the Fredholm determinant," (2014). `arxiv:1406.5252`, submitted to *J. Comput. Phys.*

51. C. Müller, *Foundations of the Mathematical Theory of Electromagnetic Waves* (Springer-Verlag, 1969).

52. V. Rokhlin, "Solution of acoustic scattering problems by means of second kind integral equations," Wave Motion **5**, 257–272 (1983).

53. P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration* (Academic, 1984).

54. R. Kress, *Linear Integral Equations*, vol. 82 of *Appl. Math. Sci.* (Springer, 1999), 2nd ed.

55. B. K. Alpert, "Hybrid Gauss-trapezoidal quadrature rules," SIAM J. Sci. Comput. **20**, 1551–1584 (1999).

56. J. Lai, M. Kobayashi, and A. H. Barnett, "A fast and robust solver for the scattering from a layered periodic structure with multi-particle inclusions," (2014). In preparation.

57. G. H. Golub and C. F. Van Loan, *Matrix computations*, Johns Hopkins Studies in the Mathematical Sciences (Johns Hopkins University, 1996), 3rd ed.

58. W. Y. Crutchfield, Z. Gimbutas, G. L., J. Huang, V. Rokhlin, N. Yarvin, and J. Zhao, *Remarks on the implementation of the wideband FMM for the Helmholtz equation in two dimensions* (Amer. Math. Soc., 2006), vol. 408 of *Contemp. Math.*, pp. 99–110.

59. A. H. Barnett and T. Betcke, "`MPSpack`: A MATLAB toolbox to solve Helmholtz PDE, wave scattering, and eigenvalue problems," (2008–2014). `http://code.google.com/p/mpspack`.

## 1. Introduction

Periodic geometries (such as diffraction gratings and antennae) and multilayered media (such as dielectric mirrors) are both essential for the manipulation of waves in modern optical and electromagnetic devices. In an increasing number of applications these features occur in tandem: for instance, performance of a grating with transverse periodicity will be enhanced by using many layers of different indices. Dielectric gratings for high-powered laser [1, 2] or wideband [3] applications rely on structures with up to 50 dielectric layers. Efficient thin-film photovoltaic cells exploit multiple layers of silicon, transparent conductors, and dielectrics, which are patterned to enhance absorption [4, 5]. For solar thermal power, efficient visible-light absorbers which reflect in the infrared require patterning of several layers [6]. Related wave scattering problems appear in photonic crystals [7], process control in semiconductor lithography [8], in the electromagnetic characterization of increasingly multilayered integrated circuits, and in models for underwater acoustic wave propagation. In most of these settings, it is common to solve for the scattering for a large number of incident angles and/or wavelengths, then repeat this inside a design optimization loop. Thus, a robust and efficient solver is crucial. We present such a solver,
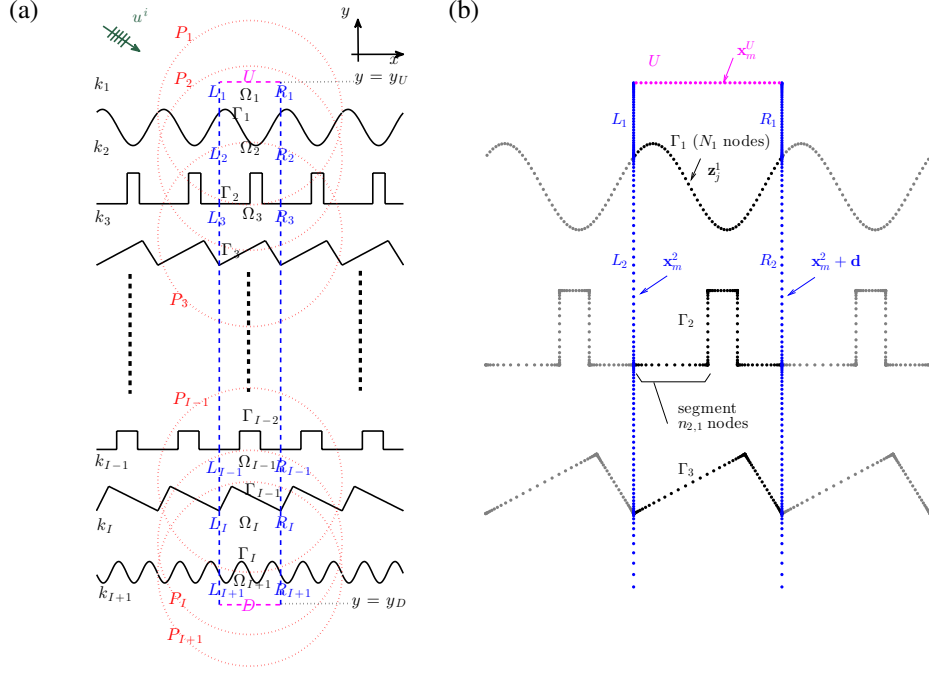
Fig. 1. (a) Geometry of scattering problem. The periodicity is in the horizontal direction, with one period lying between the vertical blue dotted lines. $\Gamma_i$ for $i = 1, \ldots, I$ are the material interfaces. The medium is uniform in the $i$th layer, which lies above the $i$th interface and has wavenumber $k_i$. Our algorithm also uses: $L_i$ and $R_i$ which are the left and right walls of one period $\Omega_i$ of the $i$th layer, $P_i$ the proxy circle for this layer, and $U, D$ the upper and lower fictitious interfaces (at $y = y_U$ and $y = y_D$) where the radiation condition is applied. (b) Zoom of the top part of the geometry, showing quadrature nodes for Nyström method and collocation (for clarity, less nodes are shown than actually used), including the near-field neighboring copies.

which scales optimally with respect to the number of layers.

Let us describe the geometry of the problem (Fig. 1(a)). Consider $I$ interfaces $\Gamma_i$, each of which has the same periodicity $d$ in the horizontal ($x$) direction. The interfaces lie between $I + 1$ homogeneous material layers, each filling a domain $\Omega_i \subset \mathbb{R}^2$, $i = 1 \ldots, I + 1$. The $i$th layer lies between $\Gamma_i$ and $\Gamma_{i-1}$, whilst the top and bottom layers are semi-infinite. The wavenumber will be $k_i$ in the $i$th layer. A plane wave is incident in the uppermost layer,

$$u^{\text{inc}}(\mathbf{r}) = \begin{cases} e^{i\mathbf{k}\cdot\mathbf{r}}, & \mathbf{r} \in \Omega_1 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

with wavevector $\mathbf{k} = (k_1 \cos\theta^{\text{inc}}, k_1 \sin\theta^{\text{inc}})$, at angle $-\pi < \theta^{\text{inc}} < 0$. The incident wave is *quasi-periodic* (periodic up to a phase), meaning $u^{\text{inc}}(x+d, y) = \alpha u^{\text{inc}}(x, y)$ for all $(x, y) \in \mathbb{R}^2$, where the *Bloch phase* (phase factor associated with translation by one unit cell) is

$$\alpha := e^{idk_1 \cos\theta^{\text{inc}}}. \tag{2}$$

Note that $\alpha$ is controlled by the period and indicent wave alone. We will seek a solution sharing this quasi-periodic symmetry.

As is standard for scattering theory [9], the incident wave causes a scattered wave $u$ to be generated, and the physical wave is their total $u^{\text{inc}} + u$. The scattered wave is given by solving the following boundary value problem (BVP). We have the Helmholtz equation in each layer,

$$\Delta u_i(\mathbf{r}) + k_i^2 u_i(\mathbf{r}) = 0, \qquad \mathbf{r} = (x, y) \in \Omega_i \tag{3}$$

where we write $u_i$ for the scattered wave in the $i$th layer, and the following interface, boundary, and radiation conditions:

- Continuity of the value and derivative the total wave on each interface, i.e.

$$u_1 - u_2 \;\; = \;\; -u^{\text{inc}} \qquad \text{and} \qquad \frac{\partial u_1}{\partial \mathbf{n}} - \frac{\partial u_2}{\partial \mathbf{n}} = -\frac{\partial u^{\text{inc}}}{\partial \mathbf{n}} \qquad \text{on } \Gamma_1, \tag{4}$$

$$u_i - u_{i+1} \;\; = \;\; 0 \qquad \text{and} \qquad \frac{\partial u_i}{\partial \mathbf{n}} - \frac{\partial u_{i+1}}{\partial \mathbf{n}} = 0 \qquad \text{on } \Gamma_i, \; i = 2, 3, \cdots I . \tag{5}$$

- Quasi-periodicity in all layers, i.e. for all $i = 1, \ldots, I+1$,

$$u_i(x + d, y) = \alpha u_i(x, y) , \qquad \text{for all } (x, y) \in \mathbb{R}^2 . \tag{6}$$

- Outgoing radiation conditions in $u_1$ and $u_{I+1}$, namely the uniform convergence of Rayleigh–Bloch expansions [10] in the upper and lower half-spaces,

$$u_1(x, y) \;\; = \;\; \sum_{n \in \mathbb{Z}} a_n^U e^{i \kappa_n x} e^{i k_n^U (y - y_U)} \qquad \text{for } y \geq y_U , \tag{7}$$

$$u_{I+1}(x, y) \;\; = \;\; \sum_{n \in \mathbb{Z}} a_n^D e^{i \kappa_n x} e^{i k_n^D (-y + y_D)} \qquad \text{for } y \leq y_D , \tag{8}$$

where the horizontal wavenumbers in the modal expansion are

$$\kappa_n := k_1 \cos \theta^{\text{inc}} + \frac{2 \pi n}{d} ,$$

and the upper and lower vertical wavenumbers are $k_n^U = \sqrt{k_1^2 - \kappa_n^2}$ and $k_n^D = \sqrt{k_{I+1}^2 - \kappa_n^2}$ respectively, with all signs taken as positive real or imaginary. The complex coefficients $a_n^U$ and $a_n^D$ are the Bragg diffraction amplitudes of the reflected and transmitted $n$th order; only the orders for which $k_n^U > 0$ or $k_n^D > 0$ carry power away from the system, and they characterize the far-field diffraction pattern.

The above BVP describes time-harmonic acoustics with layers of different sound speeds (but the same densities) [9], or the time-harmonic full Maxwell equations in the case of a $z$-invariant multilayer dielectric in TM polarization, as we now briefly review; see [11, 12]. (The modification to the interface conditions for differing densities or TE polarization are simple, and we will not present them.) Letting the time dependence be $e^{-i \omega t}$, Maxwell's equations state that the divergence-free vector fields $\mathbf{E}$ and $\mathbf{H}$ satisfy

$$\nabla \times \mathbf{E} \;\; = \;\; i \omega \mu \mathbf{H} , \tag{9}$$

$$\nabla \times \mathbf{H} \;\; = \;\; -i \omega \varepsilon \mathbf{E} . \tag{10}$$

Writing $\mathbf{E} = (0, 0, u)$ and $\mathbf{H} = (H_x, H_y, 0)$, and eliminating $\mathbf{H}$ gives in the $i$th layer the Helmholtz PDE (3) with wavenumber $k_i = \omega \sqrt{\varepsilon_i \mu_i}$, where $\varepsilon_i$ and $\mu_i$ are the permittivity and permeability

of the layer. Continuity of transverse **E** and **H** then gives the interface conditions (4)-(5). Note that all of the layer wavenumbers $k_i$ are scaled linearly by the overall frequency $\omega$.

The BVP (3)–(8) has a solution for all parameters [10, Thm. 9.2]. The solution is unique at all but a discrete set of frequencies $\omega$ when $\theta^{\text{inc}}$ is fixed [10, Thm. 9.4]; these frequencies correspond to guided modes of the dielectric structure, where resonance makes the physical problem ill-posed. They are distinct from (but in the literature sometimes confused with) *Wood anomalies* [13], which are scattering parameters ($\theta^{\text{inc}}, \omega$) for which $k_U^n = 0$ or $k_D^n = 0$ for some $n$, making the upper or lower $n$th Rayleigh–Bloch mode a horizontally traveling plane wave. A Wood anomaly does *not* prevent the solution from being unique, although it does become arbitrarily sensitive to changes in $\theta^{\text{inc}}$ or $\omega$. For more detail see [14, 15], and the extensive review in the three-dimensional (3D) case by Shipman [16].

There are many low-order numerical methods used to solve multilayer scattering problems, which in test problems may only agree to 1 digit of accuracy [17]. Finite difference time-domain (FDTD) [18, 19] is easy to code but has dispersion errors, and requires artificial absorbing boundary conditions and arbitrarily long settling times near resonances. Direct discretization in the frequency-domain, as in finite difference (FD) or finite element methods, is also possible [20, 21, 8], although they require a large number of unknowns, and, as the frequency grows, "pollution" error means that an increasing number of degrees of freedom per wavelength is needed [22]. They also demand artificial absorbing boundary conditions (perfectly matched layers). The rigorous-coupled wave analysis (RCWA) or Fourier Modal Method is specially designed for multilayer gratings [23]; to overcome slow convergence a Fourier factorization method is needed [24, 25]. However, RCWA is not easy to apply for arbitrary shapes (relying on an intrinsically low-order "staircase" approximation of layer shapes), nor to generalize to 3D. Other methods include volume integral equations [3, 26], for which it is hard to exceed low-order convergence. In general, when the layers are strictly planar, the problem becomes 1D [27] and the (unstable) transfer matrix and (stable) scattering matrix approaches are standard [28]. However, we are concerned with interfaces of arbitrary shape.

Since the medium is piecewise constant, boundary integral equations (BIE) formulated on the interfaces are natural and mathematically rigorous [9, 29, 30]. By exploiting the reduced dimensionality, the number of unknowns is much reduced, and high-order quadratures exist, in 2D [31, 32] but also in 3D. Combined with fast algorithms for handling the resulting linear systems, this leads to much higher efficiency and accuracy than FD methods, and has started to be used effectively in periodic problems [30, 33, 34, 35, 36, 37, 38]. In the setting of quasi-periodic scattering, the interfaces are unbounded and the usual approach is to replace the 2D free-space kernel (Green's function) for waves in layer $i$,

$$G^i(\mathbf{r}, \mathbf{r}') := \frac{i}{4} H_0^{(1)}(k_i \|\mathbf{r} - \mathbf{r}'\|) \tag{11}$$

(where $H_0^{(1)}$ is the Hankel function of order zero [39]), by the quasi-periodic one obeying (6), in which case the problem may be formulated on a *single period* of the interface. The usual quasi-periodic Green's function for layer $i$ is

$$G_{\text{qp}}^i(\mathbf{r}, \mathbf{r}') := \sum_{l \in \mathbb{Z}} \alpha^l G^i(\mathbf{r}, \mathbf{r}' + l\mathbf{d}), \qquad \text{where } \mathbf{d} := (d, 0), \tag{12}$$

a sum whose slow convergence renders it computationally useless. Thus a large industry has been built around efficient evaluation of $G_{\text{qp}}$ using convergence acceleration, Ewald's method, or lattice sums [40]. It is convenient to expand slightly the definition of Wood anomaly, as follows.

**Definition 1.** *We say that layer i is at a Wood anomaly if either $\kappa_n = k_i$ or $\kappa_n = -k_i$ (or both) for some $n \in \mathbb{Z}$.*

The problem then with $G_{\mathrm{qp}}$-based methods is that (12) does not exist (the sum diverges) whenever the $i$th layer is at a Wood anomaly. As the number of different materials increases in a structure, the chances of some layer hitting (or being close to) a Wood anomaly, and thus of failure, increases.

Two classes of solutions to this non-robustness problem have recently been introduced:

I. Replace (12) by the quasi-periodic Green's function for the Dirichlet [30] or impedance half-plane problems [41], or their generalization to larger numbers of images [42, 43].

II. Return to the free-space Green's function (11) for the central unit cell plus immediate neighbors, plus a new representation of far-field contributions, imposing the quasi-periodicity condition in the least-squares sense via additional rows in the linear system [44, 15, 37].

In the multilayer dielectric setting robustness using class I would require the impedance Green's function, which is difficult to evaluate; while the class II contour-integral approach of Barnett–Greengard [15] does not generalize well to multiple layers.

In this paper we introduce a simpler class II BIE method which combines the free-space Green's function for the unit cell and neighbors, with a ring of *proxy point* sources (i.e. the method of fundamental solutions, or MFS [45, 46]) to represent the far-field contributions which "periodize" the field. This combines ideas in [44, Sec. 3.2] and the fast direct solver community [47, Sec. 5], and has been independently proposed recently for Laplace problems by Gumerov–Duraiswami [48]. Related representations have been used for some time in the engineering community [49]. A modern interpretation of the key idea is that the far-field contribution is smooth—the interaction between distant periodic copies and the central unit cell is *low rank*; eg see [37]—and hence only a small number of proxy points is needed, at least if the frequency is not too high.

**Remark 1.** *Conveniently, in our new formulation we can take Schur complements to eliminate the proxy strength unknowns for each layer* without *recreating* (12) *and its associated Wood anomaly problem, as would happen in prior methods [44, 15] (see [44, Remark 8]). The difference is that in [15] both upward and downward radiation conditions (of the type* (7) *and* (8)*) are imposed on the Green's function, making it equivalent to* (12)*, whereas we do not impose* any *radiation condition in the finite-thickness layers, and impose outgoing conditions only in the semi-infinite layers 1 and $I+1$. Since non-divergent Green's functions do exist which satisfy these minimal radiation conditions, they are selected by the least-squares linear algebra in the Schur complement (see Sec. 3.2).*

We present our new representation and its discretization in Sec. 2, then combine it in Sec. 3 with a direct solver which has two steps: Schur complements to eliminate the proxy unknowns, followed by direct block-tridiagonal factorization. The tridiagonal structure arises simply because layer $i$ couples only with layers $i-1$ and $i+1$. The overall scaling is $\mathcal{O}(IN^3)$, i.e. linear in the number of layers and cubic in $N$ the number of unknowns per layer. This allows our solver to tackle problems with $NI$ of order $10^6$ unknowns in only a few minutes. Since the solution is direct, as explained in Sec. 3.4 we can solve new incident waves that share the same $\alpha$ without extra effort, and, by reusing matrix blocks, handle other $\alpha$ values efficiently. We test the solver's error and speed performance with a variety of interface shapes, with up to $I = 1000$ layers, with random or periodic permittivities, and multiple incident angles including a Wood anomaly, in Sec. 4. We conclude with a summary and implications for future work in Sec. 5.

## 2.   Boundary integral formulation, periodizing scheme, and its discretization

We now reformulate the BVP as a system of linear second-kind integral equations on the interfaces $\Gamma_i$, $i = 1, \ldots, I$ which lie in a single unit cell, coupled with linear conditions on fictitious unit cell walls; the complete system will be summarized by (53) below. A little extra geometry notation is needed, as shown in Fig. 1. Let us define the (central) unit cell as the vertical strip of width $d$ lying between $x = -d/2$ and $x = d/2$; of course its horizontal displacement is arbitrary. The blue dashed vertical lines $\{L_i\}_{i=1}^{I+1}$ and $\{R_i\}_{i=1}^{I+1}$ are the left and right boundaries of the layer domains $\{\Omega_i\}_{i=1}^{I+1}$ lying inside the unit cell. The proxy points for layer $i$ lie on the circle $P_i$ (shown by red dotted lines). The magenta dashed lines $U$ and $D$ are fictitious interfaces for radiation conditions located at $y = y_U$ and $y = y_D$, touching $\Omega_1$ and $\Omega_{I+1}$, respectively.

### 2.1.   Representation of the scattered wave

Using (11) we define standard potentials for the Helmholtz equation, the single- and double-layer representations [9] lying on a general curve $W$, at wavenumber $k_i$ for the $i$th layer,

$$(\mathscr{S}_W^i \sigma)(\mathbf{r}) := \int_W G^i(\mathbf{r}, \mathbf{r}')\sigma(\mathbf{r}')\, ds_{\mathbf{r}'}\,, \quad (\mathscr{D}_W^i \tau)(\mathbf{r}) := \int_W \frac{\partial G^i}{\partial \mathbf{n}'}(\mathbf{r}, \mathbf{r}')\tau(\mathbf{r}')\, ds_{\mathbf{r}'}\,, \quad \mathbf{r} \in \mathbb{R}^2 \quad (13)$$

where $\mathbf{n}'$ is the unit normal on the curve $W$ at $\mathbf{r}'$, and $ds$ the arclength element. Shortly we will set $W$ to be either $\Gamma_{i-1}$ or $\Gamma_i$, with the normals pointing down (into the layer below the interface). Integral representations which include phased contributions from the nearest neighbors are indicated with a tilde,

$$(\tilde{\mathscr{S}}_W^i \sigma)(\mathbf{r}) := \sum_{l=-1}^{1} \alpha^l \int_W G^i(\mathbf{r}, \mathbf{r}' + l\mathbf{d})\sigma(\mathbf{r}')\, ds_{\mathbf{r}'}\,, \quad (14)$$

$$(\tilde{\mathscr{D}}_W^i \tau)(\mathbf{r}) := \sum_{l=-1}^{1} \alpha^l \int_W \frac{\partial G^i}{\partial \mathbf{n}'}(\mathbf{r}, \mathbf{r}' + l\mathbf{d})\tau(\mathbf{r}')\, ds_{\mathbf{r}'}\,. \quad (15)$$

Let the proxy points $\{\mathbf{y}_p^i\}_{p=1}^{P} \in \mathbb{R}^2$ lie uniformly on the circle $P_i$ of radius $R$ which is centered on the domain $\Omega_i$. As is well known in MFS theory, increasing $R$ allows a higher convergence rate with respect to $P$ [46, Thm. 3]; however, since the proxy points are representing the contributions from far periodic interface copies $\{-\infty, \ldots, -3, -2\}$ and $\{2, 3, \ldots, \infty\}$, which thus have singularities at $|x| > 3d/2$, the proxy charge strengths will turn out to be exponentially large [46, Thm. 7] if $R$ exceeds $3d/2$ by much. In practice we choose $R \in [3d/2, 2d]$. Note that, should a layer $i$ be very tall (high aspect ratio), its proxy points should instead be chosen on a vertical oval to retain the "shielding" of $\Omega_i$ from its far periodic copies. In order to make the proxy representation more robust at high wavenumbers we use a "combined field" approach, choosing the proxy basis functions for $i$th layer,

$$\phi_p^i(\mathbf{r}) := \frac{\partial G^i}{\partial \mathbf{n}_p}(\mathbf{r}, \mathbf{y}_p^i) + ik_i G^i(\mathbf{r}, \mathbf{y}_p^i)\,, \qquad \mathbf{r} \in \Omega_i\,, \qquad p = 1, \ldots, P \quad (16)$$

where $\mathbf{n}_p$ is the outwards-directed unit normal to the circle $P_i$ at the $p$th proxy point. This results in smaller coefficients than if monopoles or dipoles alone were used (which can be justified by treating the proxy points as a discrete approximation to a layer potential on $P_i$, and considering arguments in [50, Sec. 7.1]).

Combining the near-field single- and double-layer potentials and proxy representations in

each layer we have, recalling the notation $u_i$ for $u$ in $\Omega_i$,

$$u_1 = \tilde{\mathscr{D}}^1_{\Gamma_1}\tau_1 + \tilde{\mathscr{S}}^1_{\Gamma_1}\sigma_1 + \sum_{p=1}^{P} c_p^1 \phi_p^1 \tag{17}$$

$$u_i = \tilde{\mathscr{D}}^i_{\Gamma_{i-1}}\tau_{i-1} + \tilde{\mathscr{S}}^i_{\Gamma_{i-1}}\sigma_{i-1} + \tilde{\mathscr{D}}^i_{\Gamma_i}\tau_i + \tilde{\mathscr{S}}^i_{\Gamma_i}\sigma_i + \sum_{p=1}^{P} c_p^i \phi_p^i , \qquad i = 2,3,\cdots,I \tag{18}$$

$$u_{I+1} = \tilde{\mathscr{D}}^{I+1}_{\Gamma_I}\tau_I + \tilde{\mathscr{S}}^{I+1}_{\Gamma_I}\sigma_I + \sum_{p=1}^{P} c_p^{I+1} \phi_p^{I+1} \tag{19}$$

By construction, for all layers $i = 1,\ldots,I+1$, and for all density functions $\sigma_i$ and $\tau_i$ and proxy unknown vectors $c^i := \{c_p^i\}_{p=1}^P$, this representation satisfies the Helmholtz equations (3). In the following subsections, we describe in turn how to enforce the interface matching, quasi-periodicity, and radiation conditions. Each of these three conditions will comprise a block row of the final linear system (53) that enables us to solve for the densities and proxy unknowns.

## 2.2. Matching conditions at material interfaces

In this subsection, matching conditions (4) and (5) will be enforced at all material interfaces in a standard Müller–Rokhlin [51, 52] scheme.

In the indirect approach, boundary integral operators arise from the restriction of representations (13) to curves [9]. Following [15] we use notation $S^i_{V,W}$ to indicate the single-layer operator at wavenumber $k_i$ from a source curve $W$ to target curve $V$. Similarly we use $D^i_{V,W}$ for the double-layer operator, $D^{i,*}_{V,W}$ for the target-normal derivative of the single-layer operator, and $T^i_{V,W}$ for the target-normal derivative of the double-layer operator. As before, we use a tilde to indicate summation over the source curve and its phased nearest neighbors, thus, for a target point $\mathbf{x} \in V$,

$$(\tilde{S}^i_{V,W}\sigma)(\mathbf{r}) := \sum_{l=-1}^{1} \alpha^l \int_W G^i(\mathbf{r},\mathbf{r}'+l\mathbf{d})\sigma(\mathbf{r}') \, ds_{\mathbf{r}'} , \tag{20}$$

$$(\tilde{D}^i_{V,W}\tau)(\mathbf{r}) := \sum_{l=-1}^{1} \alpha^l \int_W \frac{\partial G^i}{\partial \mathbf{n}'}(\mathbf{r},\mathbf{r}'+l\mathbf{d})\tau(\mathbf{r}') \, ds_{\mathbf{r}'} , \tag{21}$$

$$(\tilde{D}^{i,*}_{V,W}\sigma)(\mathbf{r}) := \sum_{l=-1}^{1} \alpha^l \int_W \frac{\partial G^i}{\partial \mathbf{n}}(\mathbf{r},\mathbf{r}'+l\mathbf{d})\sigma(\mathbf{r}') \, ds_{\mathbf{r}'} , \tag{22}$$

$$(\tilde{T}^i_{V,W}\tau)(\mathbf{r}) := \sum_{l=-1}^{1} \alpha^l \int_W \frac{\partial^2 G^i}{\partial \mathbf{n}\partial \mathbf{n}'}(\mathbf{r},\mathbf{r}'+l\mathbf{d})\tau(\mathbf{r}') \, ds_{\mathbf{r}'} . \tag{23}$$

When the target curve is the same as the source ($V = W$), we note that the single-layer operator is a weakly singular integral operator, that the action of the double-layer and its transpose must be taken in their principal value sense, and that the $T$ operator is hypersingular.

At the first interface $\Gamma_1$, $u_1$ and $u_2$ are coupled. The functions $u_1$, $u_2$, $\frac{\partial u_1}{\partial \mathbf{n}}$, and $\frac{\partial u_2}{\partial \mathbf{n}}$ at $\Gamma_1$ can be found by letting $\mathbf{r}$ in (17)–(18) approach $\Gamma_1$ from the respective side, and using the standard *jump relations* [9, Thm. 3.1 and p.66] which introduce terms of one half times the density to

each $D$ and $D^*$ term, giving

$$u_1 = -\frac{1}{2}\tau_1 + \tilde{D}^1_{\Gamma_1,\Gamma_1}\tau_1 + \tilde{S}^1_{\Gamma_1,\Gamma_1}\sigma_1 + \sum_{p=1}^{P} c^1_p \phi^1_p, \quad \text{on } \Gamma_1 \tag{24}$$

$$u_2 = \frac{1}{2}\tau_1 + \tilde{D}^2_{\Gamma_1,\Gamma_1}\tau_1 + \tilde{S}^2_{\Gamma_1,\Gamma_1}\sigma_1 + \tilde{D}^2_{\Gamma_1,\Gamma_2}\tau_2 + \tilde{S}^2_{\Gamma_1,\Gamma_2}\sigma_2 + \sum_{p=1}^{P} c^2_p \phi^2_p, \quad \text{on } \Gamma_1 \tag{25}$$

$$\frac{\partial u_1}{\partial \mathbf{n}} = \tilde{T}^1_{\Gamma_1,\Gamma_1}\tau_1 + \frac{1}{2}\sigma_1 + \tilde{D}^{1,*}_{\Gamma_1,\Gamma_1}\sigma_1 + \sum_{p=1}^{P} c^1_p \frac{\partial \phi^1_p}{\partial \mathbf{n}}, \quad \text{on } \Gamma_1 \tag{26}$$

$$\frac{\partial u_2}{\partial \mathbf{n}} = \tilde{T}^2_{\Gamma_1,\Gamma_1}\tau_1 - \frac{1}{2}\sigma_1 + \tilde{D}^{2,*}_{\Gamma_1,\Gamma_1}\sigma_1 + \tilde{T}^2_{\Gamma_1,\Gamma_2}\tau_2 + \tilde{D}^{2,*}_{\Gamma_1,\Gamma_2}\sigma_2 + \sum_{p=1}^{P} c^2_p \frac{\partial \phi^2_p}{\partial \mathbf{n}}, \quad \text{on } \Gamma_1. \tag{27}$$

On this interface only, the matching conditions (4) include the indicent wave, and enforcing them using the above gives the inhomogeneous coupled BIE and functional equations,

$$-\tau_1 + (\tilde{D}^1_{\Gamma_1,\Gamma_1} - \tilde{D}^2_{\Gamma_1,\Gamma_1})\tau_1 + (\tilde{S}^1_{\Gamma_1,\Gamma_1} - \tilde{S}^2_{\Gamma_1,\Gamma_1})\sigma_1$$

$$-\tilde{D}^2_{\Gamma_1,\Gamma_2}\tau_2 - \tilde{S}^2_{\Gamma_1,\Gamma_2}\sigma_2 + \sum_{p=1}^{P}(c^1_p \phi^1_p - c^2_p \phi^2_p)|_{\Gamma_1} = -u^{\text{inc}}|_{\Gamma_1}, \tag{28}$$

$$(\tilde{T}^1_{\Gamma_1,\Gamma_1} - \tilde{T}^2_{\Gamma_1,\Gamma_1})\tau_1 + \sigma_1 + (\tilde{D}^{1,*}_{\Gamma_1,\Gamma_1} - \tilde{D}^{2,*}_{\Gamma_1,\Gamma_1})\sigma_1$$

$$-\tilde{T}^2_{\Gamma_1,\Gamma_2}\tau_2 - \tilde{D}^{2,*}_{\Gamma_1,\Gamma_2}\sigma_2 + \sum_{p=1}^{P}\left(c^1_p \frac{\partial \phi^1_p}{\partial \mathbf{n}} - c^2_p \frac{\partial \phi^2_p}{\partial \mathbf{n}}\right)\Big|_{\Gamma_1} = -\frac{\partial u^{\text{inc}}}{\partial \mathbf{n}}\Big|_{\Gamma_1}. \tag{29}$$

Note that the half density terms added to give a $-\tau_1$ in the first equation and a $+\sigma_1$ in the second; these terms appear for every layer and will cause the BIE to be of Fredholm second kind.

On the middle interfaces $\Gamma_i$, $i = 2, \ldots, I-1$, we similarly match $u_i$ and $u_{i+1}$ and their normal derivatives, noting that now there is coupling to both the above and below interfaces, but no effect of the incident wave, to get

$$-\tau_i + (\tilde{D}^i_{\Gamma_i,\Gamma_i} - \tilde{D}^{i+1}_{\Gamma_i,\Gamma_i})\tau_i + (\tilde{S}^i_{\Gamma_i,\Gamma_i} - \tilde{S}^{i+1}_{\Gamma_i,\Gamma_i})\sigma_i + \tilde{D}^i_{\Gamma_i,\Gamma_{i-1}}\tau_{i-1} + \tilde{S}^i_{\Gamma_i,\Gamma_{i-1}}\sigma_{i-1}$$

$$-\tilde{D}^{i+1}_{\Gamma_i,\Gamma_{i+1}}\tau_{i+1} - \tilde{S}^{i+1}_{\Gamma_i,\Gamma_{i+1}}\sigma_{i+1} + \sum_{p=1}^{P}(c^i_p \phi^i_p - c^{i+1}_p \phi^{i+1}_p)|_{\Gamma_i} = 0, \tag{30}$$

$$(\tilde{T}^i_{\Gamma_i,\Gamma_i} - \tilde{T}^{i+1}_{\Gamma_i,\Gamma_i})\tau_i + \sigma_i + (\tilde{D}^{i,*}_{\Gamma_i,\Gamma_i} - \tilde{D}^{i+1,*}_{\Gamma_i,\Gamma_i})\sigma_i + \tilde{T}^i_{\Gamma_i,\Gamma_{i-1}}\tau_{i-1} + \tilde{D}^{i,*}_{\Gamma_i,\Gamma_{i-1}}\sigma_{i-1}$$

$$-\tilde{T}^{i+1}_{\Gamma_i,\Gamma_{i+1}}\tau_{i+1} - \tilde{D}^{i+1,*}_{\Gamma_i,\Gamma_{i+1}}\sigma_{i+1} + \sum_{p=1}^{P}\left(c^i_p \frac{\partial \phi^i_p}{\partial \mathbf{n}} - c^{i+1}_p \frac{\partial \phi^{i+1}_p}{\partial \mathbf{n}}\right)\Big|_{\Gamma_i} = 0. \tag{31}$$

On the bottom interface $\Gamma_I$, the only change is the absence of coupling from any lower interface, so,

$$-\tau_I + (\tilde{D}^I_{\Gamma_I,\Gamma_I} - \tilde{D}^{I+1}_{\Gamma_I,\Gamma_I})\tau_I + (\tilde{S}^I_{\Gamma_I,\Gamma_I} - \tilde{S}^{I+1}_{\Gamma_I,\Gamma_I})\sigma_I$$

$$+\tilde{D}^I_{\Gamma_I,\Gamma_{I-1}}\tau_{I-1} + \tilde{S}^I_{\Gamma_I,\Gamma_{I-1}}\sigma_{I-1} + \sum_{p=1}^{P}(c^I_p \phi^I_p - c^{I+1}_p \phi^{I+1}_p)|_{\Gamma_I} = 0, \tag{32}$$

$$(\tilde{T}^I_{\Gamma_I,\Gamma_I} - \tilde{T}^{I+1}_{\Gamma_I,\Gamma_I})\tau_I + \sigma_I + (\tilde{D}^{I,*}_{\Gamma_I,\Gamma_I} - \tilde{D}^{I+1,*}_{\Gamma_I,\Gamma_I})\sigma_I$$

$$+\tilde{T}^I_{\Gamma_I,\Gamma_{I-1}}\tau_{I-1} + \tilde{D}^{I,*}_{\Gamma_I,\Gamma_{I-1}}\sigma_{I-1} + \sum_{p=1}^{P}\left(c^I_p \frac{\partial \phi^I_p}{\partial \mathbf{n}} - c^{I+1}_p \frac{\partial \phi^{I+1}_p}{\partial \mathbf{n}}\right)\Big|_{\Gamma_I} = 0. \tag{33}$$

We wish to write these in a more compact form, hence we pair up double- and single-layer densities, then stack them into a single column vector,

$$\boldsymbol{\eta} := \begin{bmatrix} \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_I \end{bmatrix}^T, \qquad \text{where} \quad \boldsymbol{\eta}_i := \begin{bmatrix} \tau_i \\ \sigma_i \end{bmatrix}, \quad i = 1, 2, \cdots, I. \tag{34}$$

Similarly we stack the proxy strength vectors $c^i = \{c_p^i\}_{p=1}^P$, and form a vector of right-hand side functions,

$$\mathbf{c} = \begin{bmatrix} c^1, c^2, \cdots, c^{I+1} \end{bmatrix}^T, \qquad \mathbf{f} = \begin{bmatrix} -u^{\text{inc}}|_{\Gamma_1}, -\frac{\partial u^{\text{inc}}}{\partial \mathbf{n}}|_{\Gamma_1}, 0, \cdots, 0 \end{bmatrix}^T. \tag{35}$$

Then all of the coupled BIEs and functional equations (28)-(33) can be compactly grouped into the matrix-type notation,

$$\mathbf{A}\boldsymbol{\eta} + \mathbf{B}\mathbf{c} = \mathbf{f}, \tag{36}$$

where $\mathbf{A}$ is a $I$-by-$I$ matrix, each of whose entries $\mathbf{A}_{i,j}$ is a $2 \times 2$ block of operators which maps $\eta_j$ to a pair of functions (i.e. values then normal derivatives) on $\Gamma_i$. Every block of $\mathbf{A}$ is zero apart from the following tridiagonal entries,

$$\mathbf{A}_{i,i} = \begin{bmatrix} -\mathbf{I} + (\tilde{D}_{\Gamma_i,\Gamma_i}^i - \tilde{D}_{\Gamma_i,\Gamma_i}^{i+1}) & (\tilde{S}_{\Gamma_i,\Gamma_i}^i - \tilde{S}_{\Gamma_i,\Gamma_i}^{i+1}) \\ (\tilde{T}_{\Gamma_i,\Gamma_i}^i - \tilde{T}_{\Gamma_i,\Gamma_i}^{i+1}) & \mathbf{I} + (\tilde{D}_{\Gamma_i,\Gamma_i}^{i,*} - \tilde{D}_{\Gamma_i,\Gamma_i}^{i+1,*}) \end{bmatrix}, \quad i = 1, 2, \cdots, I,$$

$$\mathbf{A}_{i,i+1} = \begin{bmatrix} -\tilde{D}_{\Gamma_i,\Gamma_{i+1}}^{i+1} & -\tilde{S}_{\Gamma_i,\Gamma_{i+1}}^{i+1} \\ -\tilde{T}_{\Gamma_i,\Gamma_{i+1}}^{i+1} & -\tilde{D}_{\Gamma_i,\Gamma_{i+1}}^{i+1,*} \end{bmatrix}, \qquad i = 1, 2, \cdots, I-1,$$

$$\mathbf{A}_{i,i-1} = \begin{bmatrix} \tilde{D}_{\Gamma_i,\Gamma_{i-1}}^i & \tilde{S}_{\Gamma_i,\Gamma_{i-1}}^i \\ \tilde{T}_{\Gamma_i,\Gamma_{i-1}}^i & \tilde{D}_{\Gamma_i,\Gamma_{i-1}}^{i,*} \end{bmatrix}, \qquad i = 2, 3, \cdots, I, \tag{37}$$

where $\mathbf{I}$ is the identity operator. $\mathbf{B}$ is an $I$-by-$(I+1)$ matrix, each of whose entries $\mathbf{B}_{i,j}$ is a stack of $P$ continous function columns (sometimes called a *quasi-matrix*) expressing the effect of each proxy point strength $c_p^j$ on the value and normal derivative functions on $\Gamma_i$. The only nonzero blocks of $\mathbf{B}$ are

$$\mathbf{B}_{i,i} = \begin{bmatrix} \phi_1^i|_{\Gamma_i}, & \cdots, & \phi_P^i|_{\Gamma_i} \\ \frac{\partial \phi_1^i}{\partial \mathbf{n}}|_{\Gamma_i}, & \cdots, & \frac{\partial \phi_P^i}{\partial \mathbf{n}}|_{\Gamma_i} \end{bmatrix}, \quad \mathbf{B}_{i,i+1} = \begin{bmatrix} -\phi_1^{i+1}|_{\Gamma_i}, & \cdots, & -\phi_P^{i+1}|_{\Gamma_i} \\ -\frac{\partial \phi_1^{i+1}}{\partial \mathbf{n}}|_{\Gamma_i}, & \cdots, & -\frac{\partial \phi_P^{i+1}}{\partial \mathbf{n}}|_{\Gamma_i} \end{bmatrix}, \, i = 1, 2, \cdots, I \tag{38}$$

The term $\mathbf{A}\boldsymbol{\eta}$ in (36) is precisely (barring the summation over neighbors) the Müller–Rokhlin formulation [51, 52] for multiple material interfaces. This is of Fredholm second kind since the off-diagonal blocks in (37) have continuous kernels, and cancellation of the leading singularities occurs in the pairs in parentheses in (37), leaving the diagonal operators at most weakly singular, hence compact.

## 2.3. Imposing the quasi-periodicity conditions

Quasi-periodicity (6) will be enforced in each layer by matching both values and normal derivatives between the left $L_i$ and right $R_i = L_i + \mathbf{d}$ walls. Since the PDE is second-order, matching two functions (values and normal derivatives) is sufficient Cauchy data to guarantee extension as a quasi-periodic solution.

We evaluate the first layer representation (17) on the walls, and exploit the following simplification due to translational symmetry (as in [15, 44]) which cancels six terms (three from each

near-field sum) down to two,

$$
\begin{aligned}
\alpha^{-1} & u_1|_{R_1} - u_1|_{L_1} \\
&= \alpha^{-1}\left(\tilde{D}^1_{R_1,\Gamma_1}\tau_1 + \tilde{S}^1_{R_1,\Gamma_1}\sigma_1 + \sum_{p=1}^{P}c^1_p\phi^1_p|_{R_1}\right) - \left(\tilde{D}^1_{L_1,\Gamma_1}\tau_1 + \tilde{S}^1_{L_1,\Gamma_1}\sigma_1 + \sum_{p=1}^{P}c^1_p\phi^1_p|_{L_1}\right) \\
&= \left(\alpha^{-2}D^1_{R_1+\mathbf{d},\Gamma_1} - \alpha D^1_{L_1-\mathbf{d},\Gamma_1}\right)\tau_1 + \left(\alpha^{-2}S^1_{R_1+\mathbf{d},\Gamma_1} - \alpha S^1_{L_1-\mathbf{d},\Gamma_1}\right)\sigma_1 \\
&\quad + \sum_{p=1}^{P}\left(\alpha^{-1}\phi^1_p|_{R_1} - \phi^1_p|_{L_1}\right)c^1_p
\end{aligned}
\tag{39}
$$

For quasi-periodicity we wish this function to vanish, so we make it the first operator block row of a homogeneous linear system. Doing the same for the normal derivatives on the $L_1$ and $R_1$ walls, and then for similar conditions for all other layers $i = 2,\dots,I+1$, gives equations that can be written compactly with a matrix notation as follows:

$$
\mathbf{C}\boldsymbol{\eta} + \mathbf{Q}\mathbf{c} = \mathbf{0} ,
\tag{40}
$$

where $\mathbf{C}$ is an $(I+1)$-by-$(I)$ matrix, each entry of which is a $2 \times 2$ block of operators mapping interface densities to wall values and normal derivatives. Every block of $\mathbf{C}$ is zero apart from the bidiagonal blocks,

$$
\mathbf{C}_{i,i} = \begin{bmatrix} \alpha^{-2}D^i_{R_i+\mathbf{d},\Gamma_i} - \alpha D^i_{L_i-\mathbf{d},\Gamma_i} & \alpha^{-2}S^i_{R_i+\mathbf{d},\Gamma_i} - \alpha S^i_{L_i-\mathbf{d},\Gamma_i} \\ \alpha^{-2}T^i_{R_i+\mathbf{d},\Gamma_i} - \alpha T^i_{L_i-\mathbf{d},\Gamma_i} & \alpha^{-2}D^{i,*}_{R_i+\mathbf{d},\Gamma_i} - \alpha D^{i,*}_{L_i-\mathbf{d},\Gamma_i} \end{bmatrix}
\tag{41}
$$

$$
\mathbf{C}_{i,i-1} = \begin{bmatrix} \alpha^{-2}D^i_{R_i+\mathbf{d},\Gamma_{i-1}} - \alpha D^i_{L_i-\mathbf{d},\Gamma_{i-1}} & \alpha^{-2}S^i_{R_i+\mathbf{d},\Gamma_{i-1}} - \alpha S^i_{L_i-\mathbf{d},\Gamma_{i-1}} \\ \alpha^{-2}T^i_{R_i+\mathbf{d},\Gamma_{i-1}} - \alpha T^i_{L_i-\mathbf{d},\Gamma_{i-1}} & \alpha^{-2}D^{i,*}_{R_i+\mathbf{d},\Gamma_{i-1}} - \alpha D^{i,*}_{L_i-\mathbf{d},\Gamma_{i-1}} \end{bmatrix}
\tag{42}
$$

for $i = 1,2,\cdots,I$ and $i = 2,3,\cdots,I+1$, respectively. $\mathbf{Q}$ is an $(I+1)$-by-$(I+1)$ matrix, each entry of which is a stack of $P$ function columns (as with $\mathbf{B}_{i,j}$), but only the diagonal entries are nonzero,

$$
\mathbf{Q}_{i,i} =: \mathbf{Q}_i = \begin{bmatrix} \alpha^{-1}\phi^i_1|_{R_i} - \phi^i_1|_{L_i}, & \cdots, & \alpha^{-1}\phi^i_P|_{R_i} - \phi^i_P|_{L_i} \\ \alpha^{-1}\frac{\partial\phi^i_1}{\partial\mathbf{n}}\big|_{R_i} - \frac{\partial\phi^i_1}{\partial\mathbf{n}}\big|_{L_i}, & \cdots, & \alpha^{-1}\frac{\partial\phi^i_P}{\partial\mathbf{n}}\big|_{R_i} - \frac{\partial\phi^i_P}{\partial\mathbf{n}}\big|_{L_i} \end{bmatrix} \quad \text{for } i = 1,2,\cdots,I+1 .
\tag{43}
$$

## 2.4. Imposing the radiation conditions

First we enforce the upward radiation condition (7) at the artificial interface $U$ (with upward-pointing normal), substituting the layer-1 representation (17) to get,

$$
\tilde{D}^1_{U,\Gamma_1}\tau_1 + \tilde{S}^1_{U,\Gamma_1}\sigma_1 + \sum_{p=1}^{P}\phi^1_p|_U c^1_p - \sum_{n\in\mathbb{Z}}a^U_n e^{i\kappa_n x} = 0 .
\tag{44}
$$

Matching values at $U$ is not enough: we also need to match normal ($y$) derivatives, to ensure that the second-order PDE solution continues smoothly through $U$, thus,

$$
\tilde{T}^1_{U,\Gamma_1}\tau_1 + \tilde{D}^{1,*}_{U,\Gamma_1}\sigma_1 + \sum_{p=1}^{P}\frac{\partial\phi^1_p}{\partial\mathbf{n}}\bigg|_U c^1_p - \sum_{n\in\mathbb{Z}}a^U_n ik^U_n e^{i\kappa_n x} = 0 .
\tag{45}
$$

We will truncate the Rayleigh–Bloch expansion to $2K+1$ terms, from $n = -K$ to $K$, since it is exponentially convergent once $|\kappa_n|$ exceeds $k_1$ (in the upper layer) and $k_{I+1}$ (lower layer). We

also apply the downward radiation condition (8) at $D$ to the representation (19), giving a second set of homogeneous linear conditions. We choose the normals of $U$ and $D$ both to point in the upward sense. As with $\boldsymbol{\eta}$ and $\mathbf{c}$, we stack all coefficients into a single vector,

$$\mathbf{a} = \left[\mathbf{a}^U, \mathbf{a}^D\right]^T = \left[a^U_{-K}, \cdots, a^U_K, a^D_{-K}, \cdots, a^D_K\right]^T .$$ (46)

The resulting conditions can again be written in a simple matrix form:

$$\mathbf{Z}\boldsymbol{\eta} + \mathbf{V}\mathbf{c} + \mathbf{W}\mathbf{a} = \mathbf{0} ,$$ (47)

where

$$\mathbf{Z} = \begin{bmatrix} Z_U & 0 & \cdots & 0 \\ 0 & \cdots & 0 & Z_D \end{bmatrix}, \mathbf{V} = \begin{bmatrix} V_U & 0 & \cdots & 0 \\ 0 & \cdots & 0 & V_D \end{bmatrix}, \mathbf{W} = \begin{bmatrix} W_U & 0 \\ 0 & W_D \end{bmatrix},$$ (48)

in which

$$\mathbf{Z}_U = \begin{bmatrix} \tilde{D}^1_{U,\Gamma_1} & \tilde{S}^1_{U,\Gamma_1} \\ \tilde{T}^1_{U,\Gamma_1} & \tilde{D}^{1,*}_{U,\Gamma_1} \end{bmatrix}, \mathbf{Z}_D = \begin{bmatrix} \tilde{D}^{I+1}_{D,\Gamma_I} & \tilde{S}^{I+1}_{D,\Gamma_I} \\ \tilde{T}^{I+1}_{D,\Gamma_I} & \tilde{D}^{I+1,*}_{D,\Gamma_I} \end{bmatrix},$$ (49)

$$\mathbf{V}_U = \begin{bmatrix} \phi^1_1|_U, & \cdots, & \phi^1_P|_U \\ \frac{\partial \phi^1_1}{\partial \mathbf{n}}|_U, & \cdots, & \frac{\partial \phi^1_P}{\partial \mathbf{n}}|_U \end{bmatrix}, \mathbf{V}_D = \begin{bmatrix} \phi^{I+1}_1|_D, & \cdots, & \phi^{I+1}_P|_D \\ \frac{\partial \phi^{I+1}_1}{\partial \mathbf{n}}|_D, & \cdots, & \frac{\partial \phi^{I+1}_P}{\partial \mathbf{n}}|_D \end{bmatrix},$$ (50)

$$\mathbf{W}_U = \begin{bmatrix} -e^{i\kappa_{-K}x}|_U, & \cdots, & -e^{i\kappa_K x}|_U \\ -ik^U_{-K}e^{i\kappa_{-K}x}|_U, & \cdots, & -ik^U_K e^{i\kappa_K x}|_U \end{bmatrix},$$ (51)

$$\mathbf{W}_D = \begin{bmatrix} -e^{i\kappa_{-K}x}|_D, & \cdots, & -e^{i\kappa_K x}|_D \\ ik^D_{-K}e^{i\kappa_{-K}x}|_D, & \cdots, & ik^D_K e^{i\kappa_K x}|_D \end{bmatrix}.$$ (52)

To clarify, in $\mathbf{W}_U$ and $\mathbf{W}_D$, the $2K+1$ columns are pairs of Fourier functions evaluated over $x \in (-d/2, d/2)$, the $x$-coordinate extent of the lines $U$ and $D$.

### 2.5. Discretization of functions and operators

Finally, combining the linear conditions from the previous three subsections, we have the coupled BIE and functional equations,

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{0} \\ \mathbf{C} & \mathbf{Q} & \mathbf{0} \\ \mathbf{Z} & \mathbf{V} & \mathbf{W} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ \mathbf{c} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix},$$ (53)

Recall that $\boldsymbol{\eta}$ contains unknown density functions, while $\mathbf{c}$ and $\mathbf{a}$ are discrete coefficient vectors. On the right hand side, $\mathbf{f}$ involves functions (from the incident wave), and each $\mathbf{0}$ is a stack of zero functions. Thus $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{Z}$ are blocks of operators, while the other six matrix blocks involve quasi-matrices (stacks of function columns).

For numerical computation the continuous variables must be discretized, turning each function into a discrete set of values, and each operator block into a matrix. This is simple for the functional conditions in the 2nd and 3rd block rows of (53): we just sample at a discrete set of collocation points. For the 2nd block row, we use $M_w$ nodes $\{\mathbf{x}^i_m\}^{M_w}_{m=1}$ of a Gauss–Legendre quadrature living on the left wall $L_i$ for the $i$th layer. See Fig. 1(b). Thus each diagonal block of $\mathbf{Q}$ (43) is replaced by a $2M_w$-by-$2P$ matrix $Q_i$ with elements

$$(Q_i)_{mp} = \begin{cases} \alpha^{-1}\phi^i_p(\mathbf{x}^i_m + \mathbf{d}) - \phi^i_p(\mathbf{x}^i_m), & m = 1, \ldots, M_w, \quad p = 1, \ldots, P \\ \alpha^{-1}\frac{\partial \phi^i_p}{\partial \mathbf{n}}(\mathbf{x}^i_{m-M_w} + \mathbf{d}) - \frac{\partial \phi^i_p}{\partial \mathbf{n}}(\mathbf{x}^i_{m-M_w}), & m = M_w+1, \ldots, 2M_w, \quad p = 1, \ldots, P \end{cases}$$ (54)

Similarly for the 3rd row, we use $M$ equally-spaced (trapezoid rule) nodes $\{\mathbf{x}_m^U\}_{m=1}^M$ on $U$, and $\{\mathbf{x}_m^D\}_{m=1}^M$ on $D$. The trapezoid rule is appropriate here since functions will be periodic. Inserting these nodes, the formulae for the matrices $V_U, V_D, W_U$, and $W_D$ discretizing (50)-(52) are similar to (54) above.

To discretize the remaining blocks $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{Z}$, we need to fix a set of quadrature nodes $\{\mathbf{z}_j^i\}_{j=1}^{N_i}$ on each interface $\Gamma_i$. These nodes have associated weights $\{w_j^i\}_{j=1}^{N_i}$, such that for any smooth $d$-periodic function $f$ on $\Gamma_i$, the quadrature rule

$$\int_{\Gamma_i} f(\mathbf{r}) ds_{\mathbf{r}} \approx \sum_{j=1}^{N_i} f(\mathbf{z}_j^i) w_j^i \tag{55}$$

holds to high accuracy. To choose these nodes and weights, we first consider the case of $\Gamma_i$ a smooth interface (e.g. $\Gamma_1$ in Fig. 1(b)). Let one period of the interface be parametrized by a vector function $Z(s)$ for $0 \le s < 2\pi$. By changing variable, the periodic trapezoid rule $s_j = 2\pi(j-1/2)/N_i$, $j = 1,\ldots,N_i$ in parameter $s$ gives a quadrature rule $\mathbf{z}_j^i = Z(s_j)$ and $w_j^i = (2\pi/N_i)|Z'(s_j)|$. Then for ($C^\infty$) smooth $d$-periodic integrands on $\Gamma_i$ the error in (55) will be superalgebraically convergent [53, (2.9.16)]. On the other hand, if $\Gamma_i$ has corners, it breaks into one or more "segments" (e.g. $\Gamma_2$ in Fig. 1(b)). These need not be straight lines, merely smooth. Say a segment is again parametrized by a function $Z(s)$ for $0 \le s < 2\pi$. Then we *reparametrize* it via $Z(w(s))$ using the corner grading function suggested by Kress [31, (6.9)],

$$w(s) = 2\pi \frac{v(s)^q}{v(s)^q + v(2\pi - s)^q} \ , \ \text{where } v(s) = \left(\frac{1}{q} - \frac{1}{2}\right)\left(\frac{\pi - s}{\pi}\right)^3 + \frac{1}{q}\frac{s - \pi}{\pi} + \frac{1}{2} \ , \ 0 \le s < 2\pi \ ,$$

where $q$ controls the grading at endpoints. Higher $q$ will cause more nodes to be close to the endpoints; typically we choose $q = 6$ or higher. Let $n_{i,l}$ be the number of nodes used for the $l$th segment of $\Gamma_i$. Then a separate trapezoid rule $s_j = 2\pi j/n_{i,l}$, $j = 1,\ldots,n_{i,l}$ is used on each segment, making $N_i = \sum_l n_{i,l}$ nodes in total. The formula for the nodes and weights (we implement this in the `MPSpack` command `segment(...,'pc')`) are the same as in the smooth case, with the proviso that $Z$ is replaced by the composed function $Z \circ w$. The grading function, since its derivative vanishes to high order at the endpoints, insures that the trapezoid rule achieves high order accuracy (typically order $q$). We find this more efficient for up to 10-digit accuracy than dyadically-refined panel quadratures [38].

With interface quadratures defined, blocks $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{Z}$ are easy to discretize by restriction of the continuous variable to the set of nodes. For example, each block $\mathbf{B}_{i,i}$ in (38), describing the interaction of the $i$th proxy basis with $\Gamma_i$, is replaced by a $2N_i$-by-$P$ matrix $B_{i,i}$ with elements

$$(B_{i,i})_{jp} = \begin{cases} \phi_p^i(\mathbf{z}_j^i) \ , & j = 1,\ldots,N_i, \quad p = 1,\ldots,P \\ \frac{\partial \phi_p^i}{\partial \mathbf{n}}(\mathbf{z}_{j-N_i}^i) \ , & j = N_i + 1,\ldots,2N_i, \quad p = 1,\ldots,P \end{cases} \tag{56}$$

The matrix for $\mathbf{B}_{i,i+1}$ is similar. Operator blocks $\mathbf{C}$ and $\mathbf{Z}$ involve boundary integral representations over interfaces: each integral is replaced by a sum according to (55). For example, the $M_w$-by-$N_i$ matrix discretizing the upper-right block of $\mathbf{C}_{i,i}$ in (41) has elements $\left(\alpha^{-2} G^i(\mathbf{x}_m^i + \mathbf{d}, \mathbf{z}_j^i) - \alpha G^i(\mathbf{x}_m^i - \mathbf{d}, \mathbf{z}_j^i)\right) w_j^i$, for $m = 1,\ldots,M_w$ and $j = 1,\ldots,N_i$. Other blocks of $\mathbf{C}$ and $\mathbf{Z}$ are discretized similarly; for the reader's sanity we refrain from giving all formulae.

Finally we discretize $\mathbf{A}$. The operators in the blocks $\mathbf{A}_{i,j}$ for $i \ne j$, and for the neighboring terms $l = \pm 1$ in the local sums (20)–(23) even when $i = j$, involve only interactions between differing interfaces, and thus may be replaced simply by substitution of the native quadrature rule (55) for the sources, and evaluation at discrete target nodes, as above. This method of discretizing an integral operator is called Nyström's method [54, Sec. 12.2]. This leaves only the

*self-interaction* terms $l = 0$ in $\mathbf{A}_{i,i}$, which involve operators that are logarithmically singular. To achieve high-order accuracy for these operators, we use the standard "plain" Nyström matrix entries of the form $A(\mathbf{z}_m^i, \mathbf{z}_j^i) w_j^i$, for $m, j = 1, \ldots, N_i$ (here $A$ symbolizes a generic operator block) for entries far from the diagonal. Near-diagonal entries are adjusted by local Lagrange interpolation of the smooth density from the existing periodic trapezoid nodes onto a set of auxiliary nodes special to the logarithmic singularity, due to Alpert [55]; see [32, Sec. 4] for the full recipe. We use 30 auxiliary nodes per target node, which achieves high-order convergence with error $\mathcal{O}(N_i^{-16} \log N_i)$. With this Nyström matrix, the non-zero right-hand side terms in $\mathbf{f}$ become the samples at the first interface nodes $\mathbf{z}_j^1$ and all the unknown vectors $\eta_i$ become samples of the densities $\tau_i$ and $\sigma_i$ at the nodes of all interfaces.

The size of the resulting matrix $\mathbf{A}$ is $N_{\text{den}}$-by-$N_{\text{den}}$, where the total number of density unknowns is

$$N_{\text{den}} := 2 \sum_{i=1}^{I} N_i , \tag{57}$$

the factor of two coming from the two types of layer potential per interface.

**Remark 2.** *The Alpert correction to the periodic trapezoid rule [55] has become a standard option for closed curves [32], on which the kernel and densities are of course periodic. However, our interfaces $\Gamma_i$ do not close on each other, moreover the solution density $\eta_i$ is* quasi-periodic *with Bloch phase $\alpha$. This means that, when $\Gamma_i$ is smooth and hence has a single periodic quadrature rule, we must modify the Alpert correction entries in the northeast and southwest corners of the matrix, to account for the continuation of the interface into the next unit cell, and the phase factors $\alpha$ and $\alpha^{-1}$ (this periodic-segment adjustment of the Alpert correction is in the* `MPSpack` *code* `quadr.alpertizeselfmatrix`*).*

*Kress' periodic logarithmic correction would also be a slightly more accurate option [31] [32, Sec. 6]; however, we found that it was less convenient to adjust this scheme for quasi-periodic densities and open interface segments. (See Meier et al. [30] for an example of this; extra cut-off functions are required.)*

We will see that, due to the high-order convergence, the numbers of collocation nodes $N_i$, $n_{i,l}$, $M_w$, and $M$ can be small, of order a hundred, even for 10-digit accuracies. Note that for smooth interfaces the periodic trapezoid rule is in fact inaccurate for the interactions from neighboring interfaces, e.g. terms like $S^i_{\Gamma_i \pm \mathbf{d}, \Gamma_i}$, due to "dangling" ends of these interfaces, but that this is handled by the periodizing scheme, retaining high-order convergence.

To summarize, the linear algebraic system which discretizes (53) has identical structure to (53), but with all of its blocks matrices as constructed above. We notate these blocks using standard (non-bold) font. It will now be rearranged in order to solve it in a fast direct fashion, exploiting its tridiagonal structure.

## 3.  Rearrangement of equations, Schur complement, and fast solver

The discretized BIEs of the previous section require a few hundred to a couple of thousand unknowns per layer, to represent typical geometries as shown in Fig. 1 to accuracies of around 10 digits. The total number of unknowns includes the densities, proxy strengths, and scattered amplitudes, namely

$$\mathcal{N} = N_{\text{den}} + IP + 2(2K+1) .$$

We will find $P \sim 10^2$, thus there is an order of magnitude less proxy unknowns than density unknowns. In any case, a device with dozens or more layers leads to linear systems that are too large for direct $\mathcal{O}(\mathcal{N}^3)$ inversion or solution. On the other hand, due to the proxy point (MFS) representation, the full linear system is exponentially ill-conditioned [46], so iterative

solution of the full system is impossible. Therefore, for robustness, we describe in this section a direct solution technique, that will be "fast" (optimally scaling in $I$ the number of layers), by exploiting the algebraic structure. This also will make it easier to solve for the technologically-important case of multiple incident angles $\theta^{\text{inc}}$ at the same frequency $\omega$.

### 3.1. Rearrangement

The first step is to rearrange the unknowns, in a form amenable to elimination of the proxy and scattering amplitude unknowns. We reorder our vector of all unknowns to be

$$\mathbf{x} = \left[\eta, c_1, \mathbf{a}^U, c_2, c_3, \cdots c_{I-1}, c_I, c_{I+1}, \mathbf{a}^D\right]^T = \left[\eta, \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{I+1}\right]^T \tag{58}$$

where

$$\mathbf{x}_1 := \left[c_1, \mathbf{a}^U\right]^T, \qquad \mathbf{x}_i := c_i, \quad i = 2, 3, \cdots, I, \qquad \mathbf{x}_{I+1} := \left[c_{I+1}, \mathbf{a}^D\right]^T. \tag{59}$$

Now similarly rearranging the (discretized) blocks of (53) we get the full linear system,

$$
\left[
\begin{array}{ccccccccccc}
 & & & & & & B'_{1,1} & B_{1,2} & 0 & \cdots & 0 \\
 & & & & & & 0 & B_{2,2} & B_{2,3} & \cdots & 0 \\
 & & & A & & & 0 & 0 & B_{3,3} & \cdots & 0 \\
 & & & & & & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & & & & & & 0 & 0 & 0 & \cdots & B_{I-1,I+1} \\
 & & & & & & 0 & 0 & 0 & \cdots & B'_{I,I+1} \\
C'_{1,1} & 0 & 0 & \cdots & 0 & 0 & Q'_1 & 0 & 0 & \cdots & 0 \\
C_{2,1} & C_{2,2} & 0 & \cdots & 0 & 0 & 0 & Q_2 & 0 & \cdots & 0 \\
0 & C_{3,2} & C_{3,3} & \cdots & 0 & 0 & 0 & 0 & Q_3 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & C_{I,I-1} & C_{I,I} & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 & C'_{I+1,I} & 0 & 0 & 0 & \cdots & Q'_{I+1}
\end{array}
\right]
\mathbf{x} =
\left[
\begin{array}{c}
\mathbf{f} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0}
\end{array}
\right].
\tag{60}
$$

Because the amplitudes $\mathbf{a}$ were separated into $\mathbf{a}^U$ and $\mathbf{a}^D$ then merged into $c_1$ and $c_{I+1}$ to make $\mathbf{x}_1$ and $\mathbf{x}_{I+1}$, respectively, the first and last blocks of $B$, $C$ and $Q$ matrix had to be expanded to

$$
\begin{aligned}
B'_{1,1} &= \begin{bmatrix} B_{1,1} & \mathbf{0} \end{bmatrix}, & B'_{I,I+1} &= \begin{bmatrix} B_{I,I+1} & \mathbf{0} \end{bmatrix}, \\
C'_{1,1} &= \begin{bmatrix} C_{1,1} \\ Z_U \end{bmatrix}, & C'_{I+1,I} &= \begin{bmatrix} C_{I+1,I} \\ Z_U \end{bmatrix}, \\
Q'_1 &= \begin{bmatrix} Q_1 & \mathbf{0} \\ V_U & W_U \end{bmatrix}, & Q'_{I+1} &= \begin{bmatrix} Q_{I+1} & \mathbf{0} \\ V_D & W_D \end{bmatrix}.
\end{aligned}
$$

### 3.2. Schur complements

The rearrangement in subsection 3.1 enables us to use $I+1$ independent Schur complements to eliminate all the unknown vectors $\mathbf{x}_i$; this corresponds to periodizing all of the Green's functions. For example, consider the equation in the first row,

$$A_{1,1}\eta_1 + A_{1,2}\eta_2 + B'_{1,1}\mathbf{x}_1 + B_{1,2}\mathbf{x}_2 = \mathbf{f}.$$

The first two equation rows starting the **C** block are

$$C'_{1,1}\eta_1 + Q'_1\mathbf{x}_1 = \mathbf{0}, \qquad C_{2,1}\eta_1 + C_{2,2}\eta_2 + Q_2\mathbf{x_2} = 0.$$

With the first of these $\mathbf{x}_1$ can be eliminated, and with the second $\mathbf{x}_2$ can, giving

$$(A_{1,1} - B'_{1,1}Q_1^{'\dagger}C'_{1,1} - B_{1,2}Q_2^{\dagger}C_{2,1})\eta_1 + (A_{1,2} - B_{1,2}Q_2^{\dagger}C_{2,2})\eta_2 = \mathbf{f}, \tag{61}$$

where $Q_i^{'\dagger}$ denotes the *pseudo-inverse* of the rectangular matrix $Q'_i$, for $i = 1, \ldots, I+1$. Filling the matrix $Q_i^{'\dagger}$ and then using matrix-matrix multiplication with the $C$ blocks to fill matrices in (61) would lose accuracy, because the exponentially large matrix entries in the pseudo-inverses would cause unacceptably amplified round-off error. To retain full accuracy, we do not in fact ever evaluate the pseudo-inverse matrices. Rather, for product matrices such as $X = Q_1^{'\dagger}C'_{1,1}$ appearing above, we solve the (ill-conditioned) linear system $Q'_1 X = C'_{1,1}$ with a standard backward-stable direct dense solver. For all such dense solves we use the "backslash" or `mldivide` command in MATLAB. (Another option that retains full accuracy would be to apply $Q_1^{'\dagger}$ in two multiplication stages using its SVD factored form; for details see [56, Sec. 5]).

Similar computations eliminate $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ from the second equation to give

$$(A_{2,1} - B_{2,2}Q_2^{\dagger}C_{2,1})\eta_1 + (A_{2,2} - B_{2,2}Q_2^{\dagger}C_{2,2} - B_{2,3}Q_3^{\dagger}C_{3,2})\eta_2 + (A_{2,3} - B_{2,3}Q_3^{\dagger}C_{3,3})\eta_3 = \mathbf{0}.$$

By repeating the same computation for all the equations, all the $\mathbf{x}_j$ are eliminated and a block tridiagonal system for $\eta$ is obtained as

$$\begin{bmatrix} A'_{1,1} & A'_{1,2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ A'_{2,1} & A'_{2,2} & A'_{2,3} & 0 & \cdots & 0 & 0 & 0 \\ 0 & A'_{3,2} & A'_{3,3} & A'_{3,4} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & A'_{I-1,I-2} & A'_{I-1,I-1} & A'_{I-1,I} \\ 0 & 0 & 0 & 0 & \cdots & 0 & A'_{I,I-1} & A'_{I,I} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_{I-1} \\ \eta_I \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \tag{62}$$

with new interaction matrices for the density unknowns,

$$A'_{1,1} = A_{1,1} - B'_{1,1}Q_1^{'\dagger}C'_{1,1} - B_{1,2}Q_2^{\dagger}C_{2,1}, \tag{63}$$

$$A'_{i,i} = A_{i,i} - B_{i,i}Q_i^{\dagger}C_{i,i} - B_{i,i+1}Q_{i+1}^{\dagger}C_{i+1,i}, \qquad i = 2, 3, \cdots, I-1, \tag{64}$$

$$A'_{I,I} = A_{I,I} - B_{I,I}Q_I^{\dagger}C_{I,I} - B'_{I,I+1}Q_{I+1}^{'\dagger}C'_{I+1,I}, \tag{65}$$

$$A'_{i,i+1} = A_{i,i+1} - B_{i,i+1}Q_{i+1}^{\dagger}C_{i+1,i+1}, \qquad i = 1, 2, \cdots, I-1, \tag{66}$$

$$A'_{i,i-1} = A_{i,i-1} - B_{i,i}Q_i^{\dagger}C_{i,i-1}, \qquad i = 2, 3, \cdots, I. \tag{67}$$

**Remark 3.** *The Schur complements* (63)-(67) *replacing matrix blocks* $A_{i,j}$ *by* $A'_{i,j}$ *correspond to replacing each layer free-space Green's function* $G^i$ (11) *by an arbitrary quasi-periodic Green's function* $G'^i$ *that obeys the layer's correct wall boundary conditions* $G'^i(\mathbf{x}+\mathbf{d}, \mathbf{y}) = \alpha G'^i(\mathbf{x}, \mathbf{y})$, $\mathbf{x} \in L_i$, $\mathbf{y} \in \Omega_i$. *Crucially, since upward and downward radiation conditions are never imposed together in any single layer i, this is* not *the standard quasi-periodic* $G^i_{qp}$ *of* (12), *which diverges at that layer's Wood anomalies. The* $G'^i$ *selected by the backward-stable solves are always non-divergent; intuitively, since proxy strength vectors of small norm are possible, hence these are selected.*

### 3.3. Block tridiagonal solve and evaluation of scattered wave

The block tridiagonal system (62) can be efficiently solved with a block LU decomposition [57, Sec. 4.5.1], at a cost of dense direct inversion of diagonal blocks. Since they derive from a

second-kind integral equation, these diagonal blocks are all well conditioned, and no significant rounding error occurs when their inverses are multiplied as matrices. We write $f_i$ for the block vectors of the right-hand side of (62). The algorithm is initialized by setting $\tilde{f}_1 = f_1$ and $\tilde{A}_{1,1} = A'_{1,1}$, then the *forward sweep*, for $i = 2$ to $I$ in order,

$$\begin{aligned} \tilde{A}_{i,i} &= A'_{i,i} - A'_{i,i-1}(\tilde{A}_{i-1,i-1})^{-1}A'_{i-1,i} \,, \\ \tilde{f}_i &= f_i - (\tilde{A}_{i-1,i-1})^{-1}A'_{i-1,i}\tilde{f}_i \,. \end{aligned}$$

To save RAM, it is possible to have $\tilde{A}_{i,i}$ overwrite $A'_{i,i}$. The *backward sweep* starts with solving $\tilde{A}_{I,I}\eta_I = \tilde{f}_I$, then for $i = I - 1$ down to 1 in order, solve for $\eta_i$ in

$$\tilde{A}_{i,i}\eta_i = \tilde{f}_i - A'_{i,i+1}\eta_{i+1} \,.$$

If each $N_i \approx N$, for some constant $N$, the cost is $\mathcal{O}(N^3 I)$ due to two block inversions per layer. This is roughly $I^2$ times faster than naive inversion of the whole system. Note that the matrix filling time is only $\mathcal{O}(N^2 I)$, but dominates for the small $N$ in our settings, due the large prefactor in evaluating special functions and applying Alpert corrections.

Once all of the density vectors $\eta_i$ are known, the proxy and scattering amplitude vectors $\mathbf{x}_i$ are easily recovered by

$$\mathbf{x}_1 = -Q'^{\dagger}_1 C'_{1,1}\eta_1 \,, \tag{68}$$

$$\mathbf{x}_i = -Q^{\dagger}_i \begin{bmatrix} C_{i,i-1} & C_{i,i} \end{bmatrix} \begin{bmatrix} \eta_{i-1} \\ \eta_i \end{bmatrix} \,, \qquad i = 2, 3, \cdots, I \,, \tag{69}$$

$$\mathbf{x}_{I+1} = -Q'^{\dagger}_{I+1} C'_{I+1,I}\eta_I \,. \tag{70}$$

Here the products of the type $Q^{\dagger}_i C_{i,j}$ can be reused from their prior computation in the Schur complements (63)-(67).

Finally, the scattered wave solution $u_i$ in each layer can be evaluated from their representations (17)-(19) by applying the interface quadrature rules to the single- and double-layer potentials, using the discrete density vectors $\eta_i$. The evaluation involves only free-space Green's functions and monopole/dipole sources, which would be compatible with a standard FMM [58]. Above $y = y_U$ and below $y = y_D$, the solution is evaluated from the Rayleigh–Bloch expansions (truncated versions of (7)-(8)), using the Bragg amplitudes $\mathbf{a}^U$ and $\mathbf{a}^D$.

### 3.4.   *Accelerated sweep over multiple incident angles at one frequency*

Modeling many real-world devices requires characterizing transmission and reflection over a wide range of incident angles $\theta^{\text{inc}}$ at one frequency $\omega$. In general, changing $\theta^{\text{inc}}$ changes both the operators and the right-hand side in (53), thus naively an independent solution is required for each angle, making the task very expensive. Here we show how to exploit two independent types of structure that speeds this up by typically an order of magnitude.

Firstly, we exploit the fact that the operators in (53), and hence the entire system matrix, depends on $\theta^{\text{inc}}$ only through $\alpha$ in (2). Thus we may group together multiple incident angles that share $\alpha$ (as in [37]), and solve them together at a cost that is essentially the same as a single angle, using the precomputed inverse blocks in the tridiagonal solve of Sec. 3.3. Roughly there are $k_1 d/\pi$ such angles, hence this is the speed-up factor. It grows in proportion to the incident wavenumber. For example, if $k_1 = 40$ and $d = 1$, the speedup is around a factor of 6. To set up a sweep of all incident angles, we choose a uniform grid in $\theta^{\text{inc}}$ with spacing $2\pi/(ndk_1)$ for some integer $n$, insuring that only $n$ independent $\alpha$-value solves are needed.
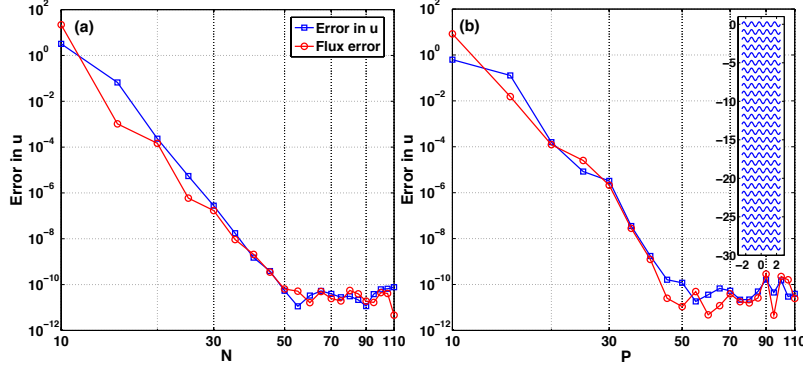
Fig. 2. Convergence of $u(0.15, 0.6)$ and flux error for 30 sine interfaces (see the inset in (b)) with $\omega = 10$ and random $\varepsilon_i$ between 1 and 2 . **(a)** Convergence in $N$, number of nodes per sine interface (blue square) and flux error (red triangle) while $P = 110$. **(b)** Convergence in $P$ (blue square) and flux error (red triangle), fixing $N = 70$ per sine interface. All other parameters are fixed at $M_w = 110$, $M = 100$, $K = 10$, and $R = 2$. (Color online.)

Secondly, we exploit the fact that filling (rather than solving) the matrix blocks in (60) often accounts for the bulk of the solution time. Consider one of the integral operators used in the BIE matrix (defined in (20)),

$$(\mathscr{D}_{V,W}^i \tau)(\mathbf{r}) = \sum_{l=-1}^{1} \alpha^l \int_W \frac{\partial G_i}{\partial \mathbf{n}'}(\mathbf{r}, \mathbf{r}' + l\mathbf{d})\tau(\mathbf{r}') \, ds_{\mathbf{r}'}. \tag{71}$$

The integral part is independent of the incident angle or $\alpha$. Therefore

$$\int_W \frac{\partial G_i}{\partial \mathbf{n}'}(\mathbf{r}, \mathbf{r}' + l\mathbf{d})\tau(\mathbf{r}') \, ds_{\mathbf{r}'} \tag{72}$$

can be precomputed for $l \in \{-1, 0, 1\}$ then used to assemble $(\tilde{D}_{V,W}^i \tau)(\mathbf{r})$ whenever a new $\alpha$ is given. Exactly the same argument applies to $(\tilde{S}_{V,W}^i \sigma)(\mathbf{r})$, $(\tilde{T}_{V,W}^i \tau)(\mathbf{r})$, and $(\tilde{D}_{V,W}^{i,*} \sigma)(\mathbf{r})$. Therefore, the $A$, $B$, $C$, and $Q$ matrices can be assembled for any $\alpha$ simply by adding and subtracting precomputed integral operators; this speeds up the solution at multiple $\alpha$ values at the expense of using extra RAM.

## 4. Numerical results

In all numerical examples, we let the first layer be "vacuum" with $\varepsilon_1 = 1$, set $\mu_i = 1$ for all layers, and choose periodicity $d = 1$. All the computations are conducted using MATLAB 2014a running on a workstation with two Intel Xeon E5-2687W processors (total 16 cores) with 256 GB memory and CentOS 6.5. For filling of the Nyström matrix blocks we use `MPSpack` [59], which has an interface to Fortran codes for Hankel function evaluations by Vladimir Rokhlin. The `parfor` command in MATLAB's Parallel Computing Toolbox is also used to fill the $A$ and $C$ matrices, using only 12 threads. In the following we present: numerical tests on convergence, scaling of timing and memory with frequency $\omega$ and the number of layers $I$, solution plots for a 1000-interface case, and accelerated computation of transmission and reflection spectra at multiple angles.

**Remark 4.** *The Bragg coefficients $\{a_n^U\}$ and $\{a_n^D\}$ will be used as an independent measure of the accuracy of our numerical scheme based on conservation of the flux (energy) [14, 16],*
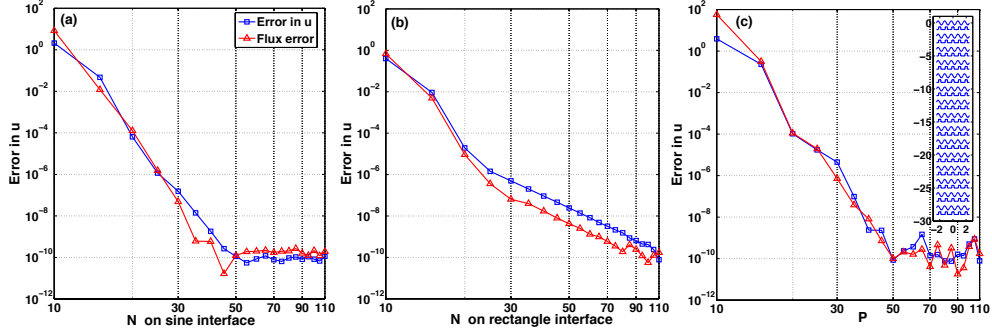
Fig. 3. Convergence of $u(0.15, 0.6)$ and flux error for 30 mixed sine and rectangle interfaces (see the inset in (c)) with $\omega = 10$ and random $\varepsilon_i$ between 1 and 2. **(a)** Convergence in $N$ on sine interface (blue square) and flux error (red triangle), while $N = 110$ on each line segment of rectangle interfaces, and $P = 110$. **(b)** Convergence in $N$ on each line segment of rectangle interfaces (blue square) and flux error (red triangle), while $N = 70$ on sine interfaces and $P = 110$. **(c)** Convergence in $P$ (blue square) and flux error (red triangle), while $N = 70$ on sine, $N = 110$ on rectangle interfaces. All other parameters are fixed at $M_w = 110$, $M = 120$, $K = 20$, and $R = 2$. (Color online.)

*namely*

$$\sum_{k_n^U > 0} k_n^U |a_n^U|^2 + \sum_{k_n^D > 0} k_n^D |a_n^D|^2 = k_1 \cos \theta^{\text{inc}} . \tag{73}$$

*This holds when all the material properties $\varepsilon_i$ and $\mu_i$ are real. Therefore, we will define the relative flux error as*

$$\text{Flux error} := \left| \frac{\sum_{k_n^U > 0} k_n^U |a_n^U|^2 + \sum_{k_n^D > 0} k_n^D |a_n^D|^2 - k_1 \cos \theta^{\text{inc}}}{k_1 \cos \theta^{\text{inc}}} \right| . \tag{74}$$

### 4.1. Convergence

We study the convergence of the scattered field at a fixed point $(0.15, 0.6)$ in the first layer, relative to its converged value, and convergence of flux errors defined in (74). The frequency $\omega = 10$ corresponds to a period of about 1.6 wavelengths in the first layer, and larger numbers of wavelengths in the other layers. Thirty sine interfaces with random $\varepsilon_i$ chosen between 1 and 2 are considered in Fig. 2. The actual geometry is depicted in the inset of Fig. 2(b). All the interfaces have a periodic trapezoid rule with the same number of quadrature points $N_i = N$, $i = 1, 2, \cdots 30$. Fig. 2(a) shows convergence of the scattered field error (blue square) and flux error (red circle) as a function of $N$: the expected 16th-order rate is observed, and the fact that the two types of error are essentially equivalent, at least not near any material interfaces. The same convergence test is conducted as a function of the number of proxy points $P$ per layer, in Fig. 2(b). The slight downwards curvature on the log-log axes is consistent with super-algebraic convergence. The fact that 10-digit accuracy results with only $N = P = 50$ is a testament to the extremely rapid convergence of the method.

Secondly, in order to study convergence for non-smooth interfaces, every other sine interface is replaced by a rectangular-ridge interface consisting of five line segments (see the inset in Fig. 3(c)), with Kress grading parameter $q = 6$. We study the convergence of the two types of interfaces independently. Figure 3(a) shows convergence in $N$ on sine interfaces, where $N_i = N$, $i = 1, 3, \cdots, 29$, while the quadrature points on rectangle is fixed at $N_i = 110 \times 5$, $i = 2, 4, \cdots, 30$

Table 1. CPU time, memory, and flux error: $\omega = 5$ (period is $0.8\lambda$ in vacuum), $\varepsilon_1 = 1$ and all other $\varepsilon_i$ are random between 1 and 2, $\theta^{\text{inc}} = -\pi/5$, $N_i = 70$ on sine, $N_i = 100 \times 2$ on triangle, and $N_i = 100 \times 5$ on rectangle interfaces, $M_w = 120$, $M = 60$, $P = 60$, $K = 20$, and $R = 2$.

| Number of interfaces | 1 | 3 | 10 | 30 | 100 | 300 | 1000 |
|---|---|---|---|---|---|---|---|
| Matrix Filling (sec) | 0.518 | 1.860 | 4.200 | 5.600 | 12.384 | 32.332 | 103.331 |
| Schur Complement (sec) | 0.028 | 0.058 | 0.299 | 0.644 | 2.263 | 6.525 | 21.037 |
| Block Solve (sec) | 0.003 | 0.041 | 0.398 | 0.898 | 2.805 | 8.626 | 26.655 |
| Memory (MB) | 18 | 41 | 83 | 183 | 608 | 1753 | 5830 |
| Flux Error | 4.8e-12 | 3.1e-11 | 2.4e-11 | 4.0e-11 | 2.2e-11 | 1.3e-10 | 9.1e-10 |

Table 2. CPU time, memory, and flux error: $\omega = 40$ (period is $6.4\lambda$ in vacuum), $\varepsilon_1 = 1$ and all other $\varepsilon_i$ are random between 1 and 2, $\theta^{\text{inc}} = -\pi/5$, $N_i = 180$ on sine, $N_i = 150 \times 2$ on triangle, and $N_i = 340 \times 5$ on rectangle interfaces, $M_w = 120$, $M = 60$, $P = 160$, $K = 20$, and $R = 2$.

| Number of interfaces | 1 | 3 | 10 | 30 | 100 | 300 | 1000 |
|---|---|---|---|---|---|---|---|
| Matrix Filling (sec) | 0.813 | 2.739 | 13.295 | 16.312 | 30.960 | 75.915 | 247.597 |
| Schur Complement (sec) | 0.080 | 0.231 | 1.366 | 3.361 | 10.553 | 33.097 | 108.384 |
| Block Solve (sec) | 0.018 | 0.118 | 3.879 | 8.637 | 28.180 | 90.187 | 277.933 |
| Memory (MB) | 58 | 112 | 496 | 1069 | 3576 | 10319 | 34347 |
| Flux Error | 8.3e-13 | 5.3e-12 | 8.9e-10 | 3.4e-10 | 3.6e-09 | 5.0e-09 | 4.7e-08 |

(i.e. $n_{i,l} = 110$ for all $l = 1, \ldots, 5$). It shows the same convergence as before. Then, $N_i$ is fixed at 70 on all sine interfaces ($i = 1, 3, \cdots, 29$), and $N_i = N \times 5$ on all rectangle interfaces ($i = 2, 4, \cdots, 30$), is increased. Both the scattered field and flux error appear to converge as $\mathscr{O}(N^{-6})$ in Fig. 3(b), the order expected from the $q = 6$ grading. Figure 3(c) shows at least very high-order convergence in $P$.

These tests confirm that flux error is a good indicator of pointwise error in $u$, at least not close to the interfaces; from now on we quote flux error.

### 4.2. Performance

Tables 1 and 2 present the CPU time, memory usage, and flux error for periodic dielectric structures with $I = 1, 3, 10, 30, 100, 300,$ and 1000 mixed sine, triangle, and rectangle interfaces. The first is at frequency $\omega = 5$ (0.8 wavelengths per period), the second $\omega = 40$ (6.4 wavelengths per period). Triangle interfaces are placed every 8th interface starting from the 4th interface and rectangle interfaces are placed every 15th interface starting from the 4th interface. All other interfaces are sine. All other parameters used for computations are specified in the table captions. In both tables, the computation time to solve the matrix (Schur complement and block matrix solve) and memory usage increases linearly as number of interfaces increases as expected in Sec. 3.3. For 1000 interfaces, it took 151 sec and 634 sec for $\omega = 5$ and 40, respectively; note that more quadrature nodes are needed to accurately discretize the more oscillatory functions for higher $\omega$. The sizes of the full matrix (60) for 1000 interfaces are $468260 \times 287922$ and $832000 \times 751762$, respectively. At $\omega = 40$ the structure is around $8000\lambda$ tall.

In order to present a performance of the numerical method in some extreme cases, three numerical examples are presented. First, we considered 100 interfaces chosen randomly as sine, triangle, or rectangle type, with random heights, phases, layer thicknesses (while preventing collisions), and layer permittivities; see Fig. 4(a). We chose a frequency corresponding to a period of 4.5 wavelengths in the top layer, and an incident angle making the top layer precisely at a Wood anomaly. Due to space limitations, the real part of total field in only the first 10 and last 10 layers (regions enclosed by rectangles in Fig.4(a)) is plotted in Fig. 4(b). Matrix filling time was 19 sec, the Schur complements 9 sec, and the tridiagonal solve 8 sec. We achieve
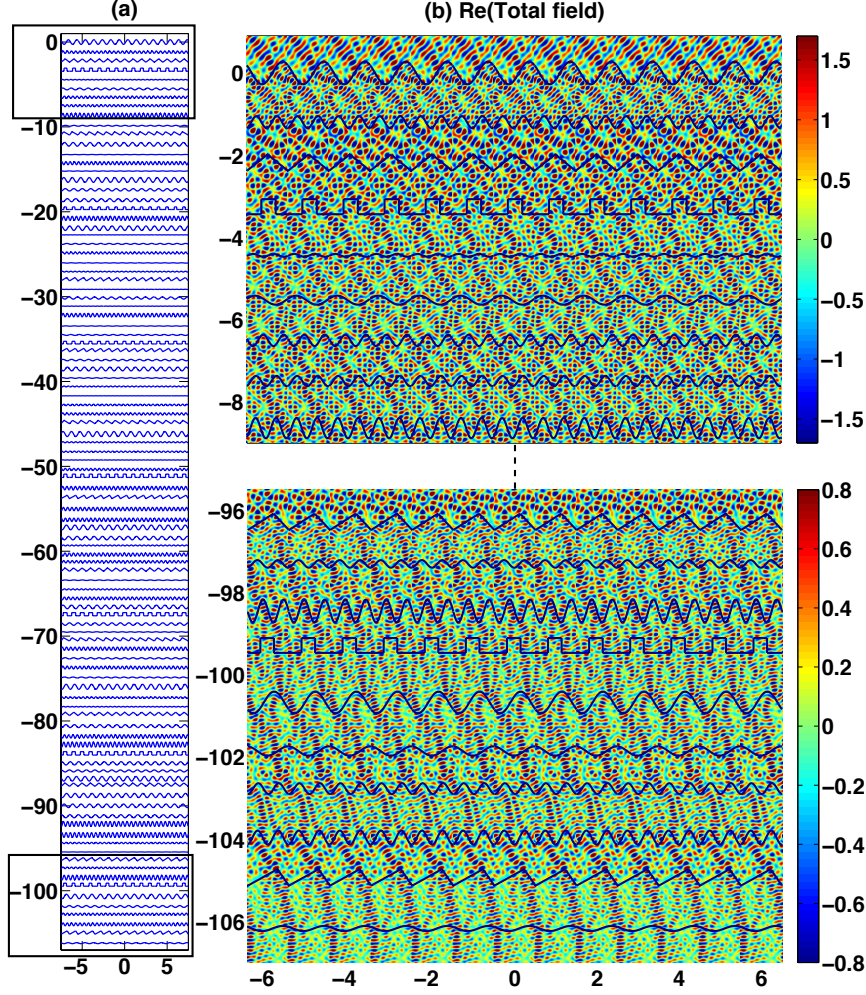
Fig. 4. **(a)** The 100-interface structure tested at a Wood anomaly for the top layer. **(b)** Real part of total field $u + u^{\text{inc}}$ in the rectangles drawn in (a), for $\omega = 9\pi$, $\theta^{\text{inc}} = -\cos^{-1}(1 - 2\pi/\omega)$, $\varepsilon_1 = 1$ and all other $\varepsilon_i$ are randomly chosen between 1 and 2. $N_i = 260$ on sines, $100 \times 2$ on triangles, $90 \times 5$ on rectangles, $M_w = 120$, $M = 60$, $P = 120$, and $K = 10$. Flux error is $5 \times 10^{-10}$, total solution time 35 sec (not including field evaluation). (Color online.)

9-digit accuracy in flux; in general we find that the flux error is no worse at Wood anomalies than at other angles.

The second and third examples are for 1000 interfaces. The real part of total field $u + u^{\text{inc}}$ in only the top 10 layers is displayed due to figure resolution limitations. Figure 5(a) and (b) show the top 30 interfaces of the structure and the real part of the total field from the 1000-interface case in the last column of Table 2, respectively. The third example is presented to highlight the geometric flexibility: we considered 1000 interfaces consisting of seven complex interface shapes on top of 993 sine interfaces (Fig. 6(a)). Here $\varepsilon_i$, $i = 2, \ldots, 1000$ are chosen randomly between 1 and 3. All other parameters are given in the figure caption. The total field is computed and displayed in Fig. 6(b). The CPU time for the solve was about 400 sec with $7 \times 10^{-9}$ flux error; evaluation of the solution at 1 million points took a similar time.
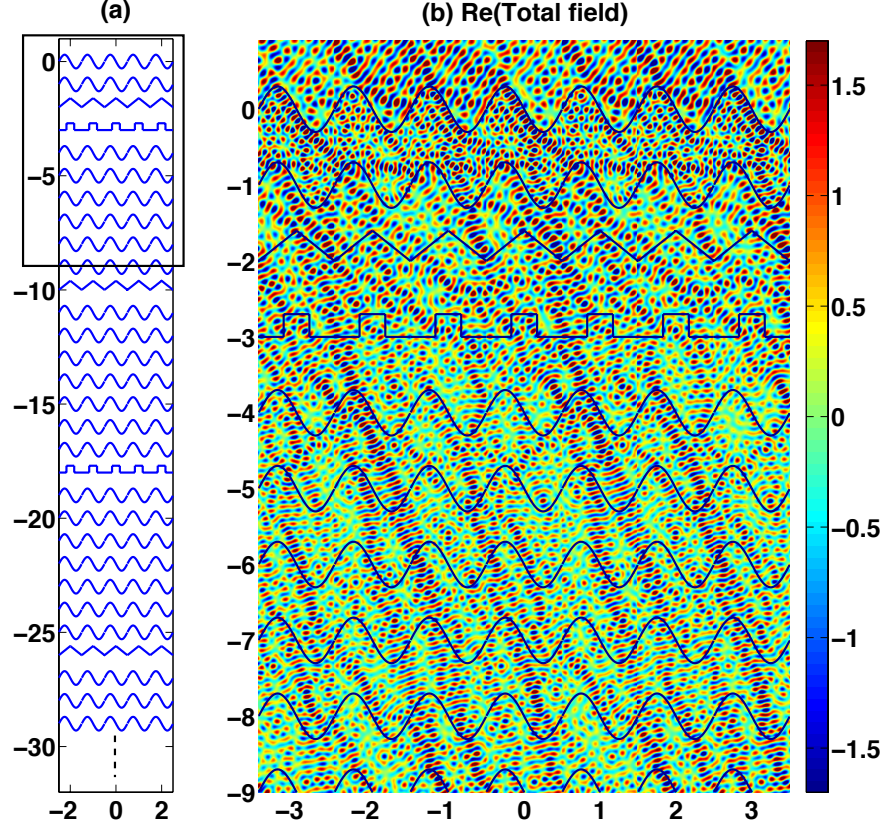
Fig. 5. **(a)** First 30 interfaces of the 1000-interface structure used in Tables 1 and 2. **(b)** Real part of the total field $u + u^{\text{inc}}$ in the rectangle drawn in (a): $\omega = 40$, $\theta^{\text{inc}} = -\pi/5$, $\varepsilon_1 = 1$, and all other $\varepsilon_i$ randomly chosen between 1 and 2. Flux error is $4.7 \times 10^{-8}$. (Color online.)

### 4.3. Transmission and reflection spectrum

We now compute transmission $(T)$ and reflection $(R)$ spectra,

$$T(\theta^{\text{inc}}) := \frac{\sum_{k_n^D > 0} k_n^D |a_n^D|^2}{k_1 \cos \theta^{\text{inc}}}, \qquad R(\theta^{\text{inc}}) := \frac{\sum_{k_n^U > 0} k_n^U |a_n^U|^2}{k_1 \cos \theta^{\text{inc}}}, \tag{75}$$

respectively, as a function of incident angle, $-\pi < \theta^{\text{inc}} < 0$, and benchmark the acceleration technique of Sec. 3.4. The 30-interface structure shown in Fig. 7(a) is used, the same as in Tables 1 and 2.

First, $\omega = 2$ (0.3 wavelengths per period) and periodic permittivities $\varepsilon = \{1, 4, 1, 4, \cdots, 4, 1\}$ are considered. The spectrum clearly shows Bragg mirror or Fabry-Perot characteristics (there are ranges of incident angles that have total reflection), and symmetry, in Fig. 7(b). Because the wavelength is larger than the geometric features, the interface shape does not play an important role in determining the scattering. However, when $\varepsilon_i$ are set to random numbers between 1 and 4, the total reflection regime disappears in Fig. 7(c). Since $\omega d/\pi < 1$, essentially no benefit comes from multiple angles sharing $\alpha$ values, but precomputation of matrix blocks does help. For 200 incident angles, the computation took 50 sec to produce with acceleration but 352 sec without, a $7\times$ speed-up.
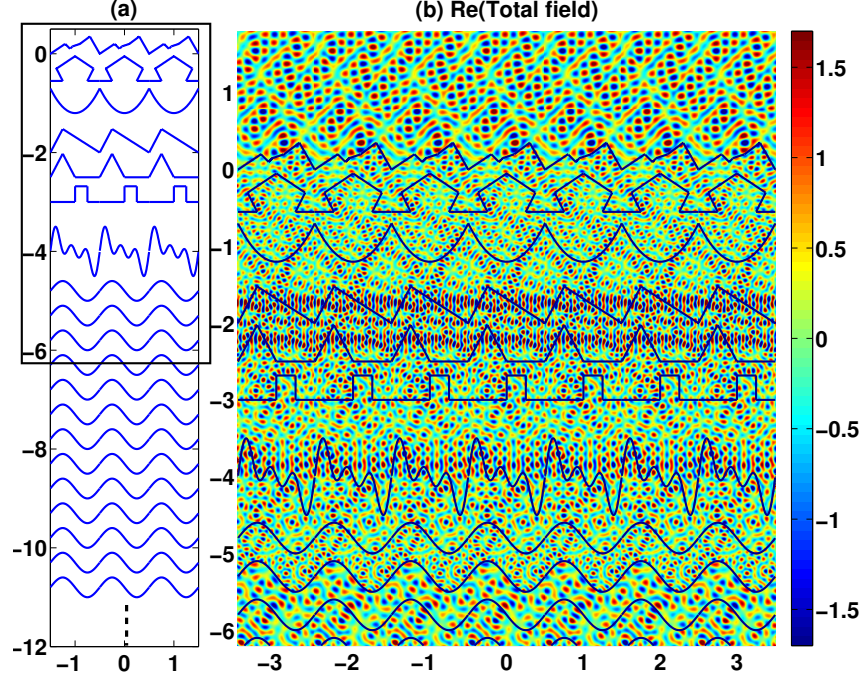
Fig. 6. **(a)** 1000-interface structure consisting of 7 complex-shaped interfaces on top of 993 sine interfaces. **(b)** Real part of total field $u + u^{\text{inc}}$ in the rectangular region drawn in (a): $\omega = 40$, $\theta^{\text{inc}} = -\pi/4$, $\varepsilon_1 = 1$ and all other $\varepsilon_i$ are chosen randomly between 1 and 3. $N_1 = 160 \times 6$, $N_2 = 160 \times 6$, $N_3 = 250$, $N_4 = 180 \times 2$, $N_5 = 160 \times 3$, $N_6 = 160 \times 5$, and $N_7 = 300$ on the first 7 interfaces. $N_i = 300$ for the rest of the sine interfaces. $M_w = 130$, $M = 80$, $P = 150$, $K = 20$, $R = 1.5$. Flux error is $7 \times 10^{-9}$, total matrix filling time 192 sec, Schur complement 107 sec, block matrix solve 103 sec, total memory used 28 GB. Field evaluation ($1000 \times 1000$ grid points) took 446 sec. (Color online.)

The frequency is now increased to $\omega = 10$ (1.6 wavelengths per period), with 641 incident angles. Regardless of the $\varepsilon$ distribution, the transmission-reflection spectrum behaves very abruptly in both periodic (Fig. 7(d)) and random (Fig. 7(e)) $\varepsilon$ cases. Also notice that the symmetry of the spectrum is broken because the layered structure has rectangle and triangle shaped interfaces which are now *resolved* by the wavelength. The computation took 121 sec with acceleration, and 1619 sec without, a speed-up of $13\times$. Finally, at $\omega = 20$ (3.2 wavelengths per period), using 1279 incident angles, the speed-up was about $25\times$ (we don't show these spectra since they do not show any new phenomena). This is consistent with the acceleration factor growing linearly with $\omega$. Thus for $\omega = 40$ as in Table 2 a speedup of $50\times$ is expected.

## 5. Conclusion

We presented a new robust and fast integral equation method for 2D scattering from a periodic dielectric grating with an arbitrary number of layers of general shape. There are three main features: (1) The computational cost of the new method scales optimally (linearly) in the number of layers, allowing 1000-layer structures to be solved rapidly. (2) The method is stable for *all* scattering parameters including Wood anomalies, since it is based on free-space rather than quasi-periodic Green's functions. (3) The periodizing scheme is simple, high-order accurate, and largely supersedes the Sommerfeld integral method of the second author and Greengard
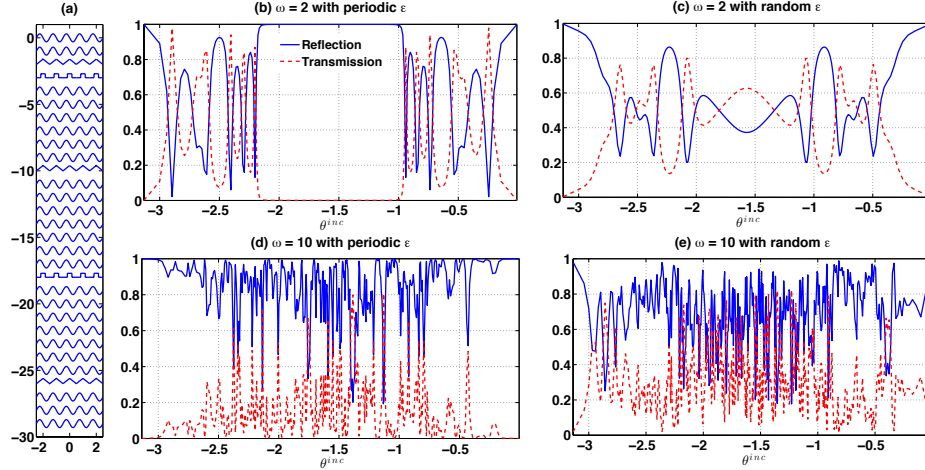
Fig. 7. **(a)** 30-interface structure. Reflection (blue solid line) and transmission (red dashed line) as a function of incident angle from $-\pi$ to 0 for: **(b)** $\omega = 2$ with periodic $\varepsilon = \{1, 4, 1, 4, \cdots, 4, 1, 4, 1\}$, average flux error $8.3 \times 10^{-9}$; **(c)** $\omega = 2$ with $\varepsilon_1 = 1$ and all other $\varepsilon_i$ random between 1 and 4, average flux error $1.3 \times 10^{-10}$; **(d)** same structure as (b) but $\omega = 10$, average flux error $4.7 \times 10^{-7}$; and **(e)** same structure as (c) but $\omega = 10$, average flux error $4.1 \times 10^{-10}$. (Color online.)

in [15]. This solver is expected to be useful in variety of wave applications in engineering and experimental physics, including the high accuracy modeling and optimization of optical, electromagnetic, and acoustic devices, and meta-materials.

There are several natural extensions of this work. Allowing dielectric inclusions in the layers (as in [56]), or material triple-junctions (e.g. incorporating robust representations of Lee–Greengard as in [38]), is simply a matter of bookkeeping, as long as the number of unknowns per layer remains small (e.g. less than $10^4$). Since we use high-order quadrature schemes, this would allow significantly more complex unit cell shapes than presented here. Beyond this, an iterative FMM solution of the combined system would be appropriate when the number of layers is small [56]; for robustness with many layers a hierarchical fast direct solver such as in [37, 38] could be used in each layer, combined with our tridiagonal block solve. Our scheme generalizes to 3D without needing new ideas, given a surface quadrature for the integral operators. However, the number of unknowns per layer could then easily exceed $10^4$, demanding something more elaborate than dense direct linear algebra within each layer.

For a production code, matrix filling and evaluation should be implemented in C or Fortran, and a parallel implementation would allow more simultaneous filling of matrix blocks, as well exploiting a parallel tridiagonal solve. In terms of analysis, an extension of the rigorous framework of free-space integral equations to include the presented periodizing scheme is needed.

The MATLAB codes which implement the methods of this paper and generate some of the figures can be downloaded from `http://math.dartmouth.edu/~mhcho/software` These rely on layer-potential quadrature codes in the `MPSpack` toolbox by the second author, which can be downloaded from `http://code.google.com/p/mpspack`

### Acknowledgments