Accelerated Calderón Preconditioning for Electromagnetic Scattering by Multiple Absorbing Dielectric Objects

Antigoni Kleanthous

A dissertation submitted for the degree of

Doctor of Philosophy

at

UCL.

Department of Mathematics

UCL

September 9, 2021

I, Antigoni Kleanthous, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Antigoni Kleanthous	Date

Abstract

We consider electromagnetic scattering by multiple absorbing dielectric objects using the PMCHWT boundary integral equation formulation. Galerkin discretisation of this formulation leads to ill-conditioned linear systems, and Calderón preconditioning, an operator-based approach, can be used to remedy this. To obtain a stable discretisation of the operator products that arise in this approach, the use of a dual mesh defined on a barycentrically refined grid needs to be considered, increasing memory consumption. Furthermore, to capture the oscillatory solution of the electromagnetic waves, the mesh needs to be refined with respect to frequency, making the simulation of high-frequency problems very expensive.

This thesis presents two complementary approaches to minimising memory cost and computation time (for assembly and solution): modification of the preconditioning operator, and a bi-parametric implementation. The former aims to minimise the number of operators used in the preconditioner to reduce the additional matrix-vector products performed, and the memory cost, while still maintaining a sufficient preconditioning effect. The latter uses two distinct sets of parameters during assembly, to minimise assembly and solution time as well as memory. The operator is assembled with a more expensive set of parameters to obtain an accurate solution. The preconditioner, which is discretised using the expensive dual basis functions, is assembled with a cheaper set of parameters.

The two approaches are explained in the context of a series of model problems, then applied to realistic ice crystal configurations found in cirrus

clouds. They are shown to deliver a reduction of 99% in memory cost and at least 80% in computation time, for the highest frequency considered.

The accelerated formulations have been used at the Met Office to create a new database of the scattering properties of atmospheric ice crystals for future numerical weather prediction. A brief description of that work is also presented in the thesis.

Impact Statement

Numerical experiments presented in this thesis have demonstrated a 99% reduction in memory cost and at least an 80% reduction in total computation time for the highest frequency considered. The open-source boundary element method library Bempp[1] [1] has been used for all simulations presented in this thesis. The open-source nature of the library, allows for a wider use both within and outside academia as it is easily accessible. The accelerating methods used in this thesis are also available on the author's GitHub page[2]. The accelerated methods presented in this thesis have been published in peer reviewed journals [2] and submitted for review [3], with 8 and 2 citations respectively, at the time of writing this thesis.

The accelerated methods of this thesis made it possible to run large scale simulations that were otherwise too expensive (both in terms of memory and total computation time) to consider before. In particular, the methods presented in this thesis were used to run large scale simulations of complex ice crystal configurations found in cirrus clouds at the Met Office. The scattering properties of those ice crystals were computed to generate a new scattering database for the microwave and sub-millimetre regime of the electromagnetic spectrum, to be used for future numerical weather prediction. A paper detailing the methodology and results from those simulations is also under development and the implementation of random orientation used for the database is publicly available[3]. The accelerated numerical methods presented in this

[1] `www.bempp.com`
[2] `www.github.com/ankleanthous/Accelerated_Calderon`
[3] `www.github.com/ankleanthous/EMScattering`

thesis will be used again at the Met Office, to generate new databases for different regimes of the electromagnetic spectrum and for different geometrical configurations.

Acknowledgements

I would like to thank my supervisors Timo Betcke, David Hewett and Anthony Baran for all their support, guidance, advice and feedback during the last four and a half years. I would also like to thank the Met Office for the invaluable opportunity to use my work there and see its impact in real life. Many thanks to the AWS Team at the Met Office for helping us set up our simulations. I would also like to thank NERC and the Met Office for providing funding for this work and for providing further funding for an extension during the Covid-19 pandemic. Many thanks to the Department of Mathematics, the Access and Widening Participation Unit and UCL for the various teaching and outreach opportunities.

I would like to thank Matthew Scroggs for all his help in getting acquainted with Bempp and settling in during the beginning of my PhD, and for the opportunity to co-run residential python schools during the summers. I would also like to thank Xiaoshu Sun for co-running with me one of the summer schools, our first online teaching due to the Covid-19 pandemic(!) and for the numerous chat/lunch breaks from work, making this journey a bit more enjoyable. Many thanks to other colleagues and collaborators during the past few years; Paul Escapil-Inchauspé, Carlos Jerez-Hanckes, Sam Groth.

I would like to thank my friends and family for encouraging my cocktail making hobby and for drinking my (questionable quality) cocktails. Thanks to my friends, Agathi and Elena, for the fun we had throughout the years. I would like to thank my parents, Akis and Katia, for all their support throughout my life. I would like to thank my sister, Chloe, for treating me to food, skincare

treats, for doing my hair and for being my personal pharmacist. I would also like to thank my grandparents, Antigoni, Telemachos, Chloe and Socrates, for all their love, encouragement and support throughout the years. Finally, I would like to thank my partner, Yiannis; for his love and support, enjoying food and cocktails, and going to the theatre and opera with me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Fast and accurate numerical methods that simulate electromagnetic scattering by single or multiple scatterers of arbitrary shape (Figure 1.1) are an active research area in numerical analysis and scientific computing. The scattering problem can be described by Maxwell's time harmonic equation

$$\nabla \times (\nabla \times \mathbf{E}) - k^2 \mathbf{E} = 0, \tag{1.1}$$

for some wavenumber k, electric field $\mathbf{E}$ and subject to suitable boundary conditions. Analytic solutions exist for simplified domains, for example spheres or cylinders, and as such numerical methods are often used to find an approximate solution for more complex scatterers. One such numerical method is the boundary element method (BEM).

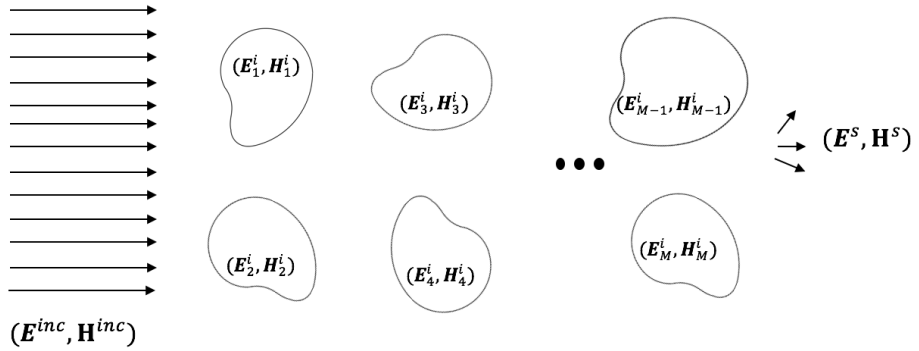


Figure 1.1: Electromagnetic scattering by multiple scatterers of arbitrary shape. Taken from [2].

BEM is a numerical method that allows us to solve certain types of partial differential equations (PDEs) in homogeneous bounded/unbounded media. It involves reformulating the PDE in terms of integral equations on the boundary of the object of interest, the scatterer. The boundary of the scatterer is subsequently discretised and the boundary integral equation (BIE) is transformed into a system of linear equations which can be solved using some direct or iterative method. Depending on the problem of interest, the solution can then be extended from the boundary to the interior, exterior or far field of the scatterer via representation formulae.

There are many advantages to using BEM compared to other numerical methods. By reformulating the PDE as a system of boundary integral equations one reduces the dimensionality of the problem from a 3D scatterer to a 2D manifold. The discretisation of the 2D manifold is faster and easier to achieve and leads to a smaller system of equations that needs to be solved. At the same time, BEM automatically incorporates radiation conditions at infinity, while for example by using a Finite Element Method (FEM) one would have to introduce artificial boundaries. On the other hand, the discretisation of FEM leads to sparse matrices compared to BEM for which matrices are usually dense, requiring fast approximation methods such as $\mathcal{H}$-matrices [13], $\mathcal{H}^2$-matrices [14, 15, 16] or Fast Multipole Methods (FMM) [17, 18, 19]. In addition, the use of BEM requires the knowledge of a fundamental solution, in order to allow for the transformation of the PDEs to BIEs [20], restricting its use to specific problems. However, BEM can handle complicated geometries making it appealing for scattering problems by complex domains, compared to other methods that are restricted to rotationally symmetric particles, for example T-matrix methods [21, 22, 23, 24].

The boundary element method is a popular method used for electromagnetic scattering problems by single or multiple dielectric objects (transmission

problems). The Maxwell scattering problem is reduced to a BIE of the form

$$\boldsymbol{\mathcal{A}}\mathbf{u} = \mathbf{f}, \tag{1.2}$$

where $\boldsymbol{\mathcal{A}}$ is a (collection of) boundary integral operator(s) (BIOs), $\mathbf{u}$ is the unknown quantity of interest (for example traces of the scattered field) and $\mathbf{f}$ some right-hand side data that depends on the source of excitation (for example an incoming plane wave). Examples of such BIE formulations are the PMCHWT formulation due to Poggio, Miller, Chang, Harrington, Wu and Tsai [25, 26, 27, 28], the Müller formulation [29] and its variations [30].

The bounded linear operator $\boldsymbol{\mathcal{A}}$, typically consists of a combination of compact and hypersingular operators which in their discrete versions have eigenvalues accumulating at zero and infinity. Thus upon Galerkin discretisation, (1.2) leads to an ill-conditioned linear system which then requires a large number of iterations to be solved. Preconditioning is therefore required in order to remedy the ill-conditioning of the discrete system.

Different preconditioning methods exist, ranging from algebraic (based for example on $\mathcal{H}$-matrix methods [5, 31] or FMM [32, 33, 34]) to operator-based approaches [35]. In operator-based approaches, one is looking for some operator $\boldsymbol{\mathcal{P}}$ applied to (1.2), giving

$$\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}\mathbf{u} = \boldsymbol{\mathcal{P}}\mathbf{f}, \tag{1.3}$$

where $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$ has better properties than $\boldsymbol{\mathcal{A}}$ and its discrete version is easier to solve numerically.

The overall computational cost of solving (1.3) is affected by both

(i) the choice of preconditioning operator $\boldsymbol{\mathcal{P}}$, and

(ii) the choice of discretisation for the operator product $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$.

With regards to (i), the choice of $\boldsymbol{\mathcal{P}}$ should lead to a smaller number of iterations compared to $\boldsymbol{\mathcal{A}}$. At the same time, the operator product $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$ incurs

additional matrix-vector operations, so the ideal preconditioner should consist of as few operators as possible. Regarding (ii), the discretisation of $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$ should be as cheap (memory and time-wise) as possible, while still producing a sufficiently accurate numerical solution. To achieve a stable discretisation of the operator product $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$, one has to use both a primal and a dual mesh, the latter defined on the barycentrically refined primal mesh leading to a 6-fold increase in the number of elements required. To capture the oscillatory solution of the electromagnetic waves, the mesh needs to be refined with respect to frequency. A dense assembly and storage of the matrix would therefore scale as $\mathcal{O}(N^2)$, where N are the degrees of freedom, or $\mathcal{O}(k^4)$, making the simulation of high-frequency problems very expensive. Fast approximations of the Galerkin matrix, such as $\mathcal{H}$-matrix, $\mathcal{H}^2$-matrix and FMM methods, are therefore of interest to the electromagnetic community dealing with high frequency problems.

We focus our attention to the PMCHWT formulation in this thesis but note that our methods could be adjusted to other formulations. The predominant operator-based preconditioner used for the PMCHWT formulation is the Calderón preconditioner, which leverages the Calderón identities to improve the properties of $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$. The traditional choice of Calderón preconditioner in this case is $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ [36, 37, 38, 39]. This indeed reduces the number of iterations required by iterative solvers but at the expense of an increased computational cost per iteration as a result of additional matrix-vector multiplications performed. A recent study by the author of this thesis [2], showed that even though $\boldsymbol{\mathcal{A}}^2$ has a reduced GMRES count, following a specific implementation detailed later in Chapter 4, it does not perform any better than simple mass-matrix preconditioning if one accounts for the overall matrix-vector products performed.

We contribute to (i) by considering modified versions of the traditional Calderón preconditioner in the context of single and multi-particle electromagnetic scattering. The modified preconditioners include only a portion of

the original operators used in the traditional Calderón preconditioner aiming to reduce assembly time and memory cost. Our work on this is discussed in Chapter 4.

With regards to (ii), we demonstrate that the cost of discretising $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$ can be reduced without sacrificing solution accuracy, by using a cheaper but poorer quality matrix assembly routine for $\boldsymbol{\mathcal{P}}$ than for $\boldsymbol{\mathcal{A}}$. This is based on an adaptation of the bi-parametric implementation proposed in [40] for the electric field integral equation (EFIE) for perfectly conducting scatterers, and later used for Helmholtz problems [41]. By adopting an $\mathcal{H}$-matrix approximation for the assembly and storage of the discrete versions of $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{A}}$, we

a) use a weaker target tolerance for $\boldsymbol{\mathcal{P}}$ than for $\boldsymbol{\mathcal{A}}$ in the approximation of admissible blocks by low-rank approximations; and

b) use lower order quadrature rules for $\boldsymbol{\mathcal{P}}$ than for $\boldsymbol{\mathcal{A}}$.

In addition, going beyond the measures introduced in [40],

c) for $\boldsymbol{\mathcal{P}}$ only, we assemble admissible blocks only if they satisfy sufficient near-field conditions and discard any far-field contributions.

Our contributions to (ii) are presented in Chapter 5.

1.1 Motivation/Application

An application of large scale simulations of electromagnetic scattering by multiple homogeneous dielectric objects where such accelerating techniques are required is that of electromagnetic radiation by ice crystals in cirrus clouds. The scattering properties of atmospheric ice crystals are important in modelling the radiation balance of cirrus clouds [4, 42, 43]. The ice crystals vary in size and shape, and are generally non-spherical, taking complex forms such as hexagonal columns, hexagonal plates, bullet rosettes and aggregates of those [42, 44, 7]. Some examples can be seen in Figure 1.2. The frequency of the incoming radiation also ranges from low to high, requiring methods that are able to handle both extreme cases and those in between.

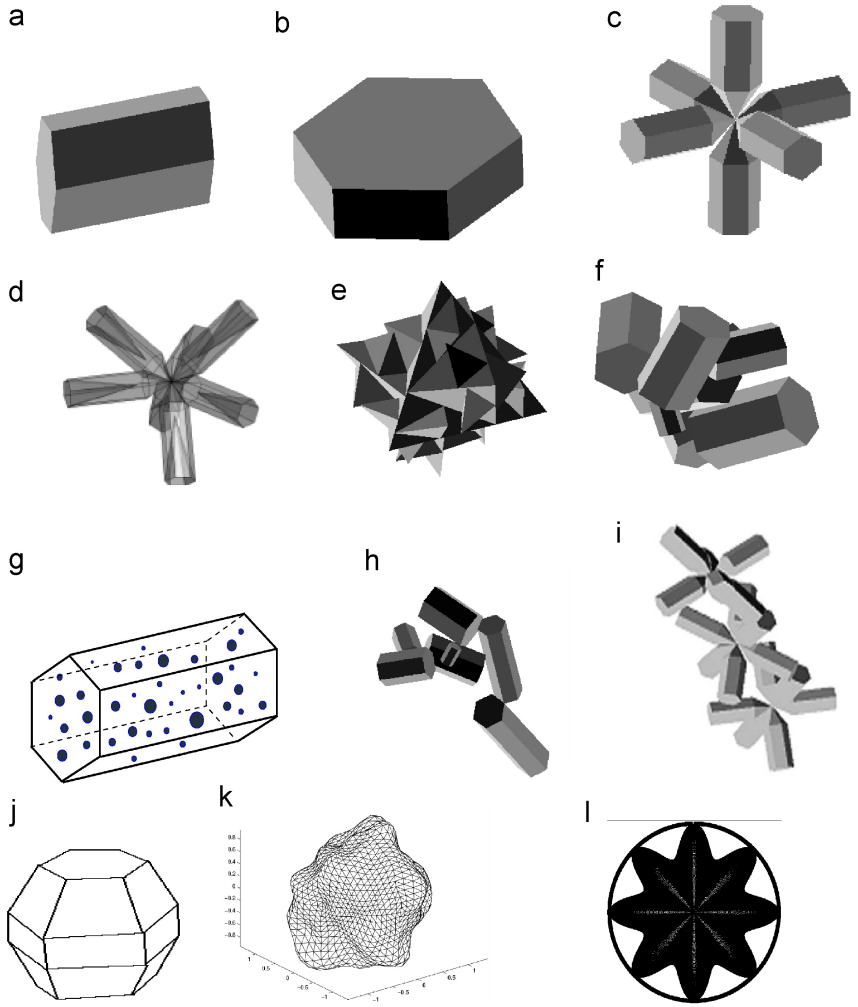


Figure 1.2: Ice crystal models used to predict the light scattering properties of cirrus: (a) a hexagonal column, (b) a hexagonal plate, (c) a six-branched bullet-rosette, (d) a six-branched bullet-rosette with conical air cavities within each branch, (e) a polycrystal, (f) a hexagonal ice aggregate, (g) an Inhomogeneous Hexagonal Monocrystal, (h) a hexagonal chain aggregate, (i) a rosette-aggregate, (j) a droxtal, (k) a Gaussian random sphere and (l) a Chebyshev polynomial. Taken from [4].

A number of different approaches are used by the atmospheric physics community for the simulation of electromagnetic dielectric scattering, depending on the size parameter of the problem, defined as

$$X := \frac{\pi D_{max}}{\lambda_e}, \tag{1.4}$$

where D_{max} is the maximum dimension of the scatterer (diameter of the smallest sphere enclosing the entire scatterer configuration), and $\lambda_e := 2\pi/k_e$ is the

wavelength of the incident wave. For particles of small to moderate size parameter there are "numerically exact" methods [45] such as the Discrete Dipole Approximation (DDA) [46, 47, 48], the Finite-Difference Time-Domain (FDTD) [49, 50, 51] and Pseudo-Spectral Time-Domain (PSTD) [52] methods, and the Extended Boundary Condition [22, 53, 23] and Invariant Imbedded [24, 54] T-matrix methods. For particles of large size parameter one can use "approximate" high-frequency methods such as Geometric Optics/ray tracing and the Kirchhoff approximation (see e.g. [55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66]).

A global method that can therefore handle both low and high frequencies (or small and large size parameter problems respectively) is of interest and the boundary element method is appealing in this case. BEM has no restrictions on the shape of the scatterer (apart from Lipschitz continuity) and by using fast approximations and accelerated preconditioning techniques, such as those detailed earlier, one can also handle high frequency problems.

Even though BEM is well-established in the electrical engineering community it has only recently attracted the attention of the atmospheric physics community. Early applications of BEM to the simulation of light scattering by simple ice crystals include [67] and [68]. More comprehensive studies of complex crystal shapes (including hexagonal columns with conventional and stepped cavities, bullet rosettes and Chebyshev ice particles) have been given recently by Groth *et al.* [5] and Baran and Groth [31], and more recently boundary integral equations were applied to scattering by multiple dielectric particles illuminated by unpolarized high-order Bessel vortex beams in the work by Yu *et al.* [69]. We note that multi-particle scattering is well studied in literature, for example in [70], but without the use of preconditioning and accelerating techniques, applications were restricted to small size parameters or simple domains.

With the use of the accelerating techniques discussed in Chapters 4 and 5, applications of BEM on fixed orientations of complex-shaped ice crystals of larger size parameters were considered in [2] and [3]. We note that complex

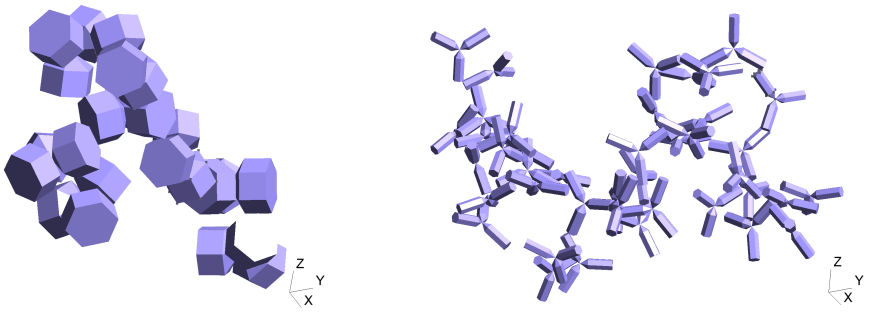


Figure 1.3: Ice crystal models representing a budding rosette of maximum dimension $D_{max} = 1190\mu m$ (left) and a rosette aggregate of maximum dimension $D_{max} = 10235\mu m$ (right).

aggregates consisting of several monomers with little or zero separation touching at single points (for example (c), (f) and (i) in Figure 1.2) were treated as a single particle problem in [5], with the individual monomers being connected by a very small cube to ensure Lipschitz continuity of the boundary. In [2], a different approach was taken by treating this as a multi-particle scattering problem taking advantage of reduced versions of Calderón preconditioners that were better suited for multi-particle configurations. We give more details about the single vs multi-particle approach in Chapter 6. We then apply the methods of Chapters 4 and 5 to such realistic complex shaped ice crystal configurations and investigate their performance (in terms of memory cost, and assembly and solution time).

In addition, the accelerating methods of Chapters 4 and 5 have made it possible to run large scale simulations of electromagnetic scattering by realistic ice crystal configurations that would otherwise be too expensive (both in terms of memory cost and total computation time) before. In particular, the single-scattering properties (SSPs) and phase matrices of randomly-oriented budding rosettes and rosette aggregates (examples of which can be seen in Figure 1.3), are being computed at the Met Office using our accelerated BEM methods for a

range of maximum dimensions and frequencies. The aggregates were generated via Monte Carlo simulations by Westbrook *et al.* [71] and follow observed mass and area dimension power laws [72]. The shapes represent observed ice crystals in cirrus clouds [7] and are assumed to have the density of solid ice. The SSPs of 65 different model aggregates, of maximum dimension between 10 and $10,000\mu m$, at frequencies 50, 183, 243 and 664 GHz, and at temperatures[1] 190, 210, 230, 250 and 270 K are computed. This database is being constructed to take advantage of forthcoming new observations from EUMETSAT's (The European Organisation for the Exploitation of Meteorological Satellites) next generation of polar orbiting satellites that ought to improve numerical weather prediction, and the simulation of airborne radiance observations using the Met Office's International Sub-millimetre Airborne Radiometer (ISMAR, see [43]).

To calculate the SSPs in random orientation, one traditionally fixes the direction of the incident wave and rotates the scatterer. Once many orientations have been considered the SSPs in random orientation can be evaluated. Fixing the direction of the incident wave and rotating the scatterer by some rotation matrix R is equivalent to fixing the orientation of the scatterer and instead rotating the direction of the incident wave by R^{-1}; and this methodology has also been considered in [9]. In this thesis we follow this second approach. This is to avoid transitions between two reference frames that are used to describe the problem (for some details on this see Chapter 7 and for full details Chapter 2.4 in [8]). In this case, we assemble the operator matrix and preconditioner and re-use for all incident wave solutions. With the accelerating techniques of Chapters 4 and 5, storage and assembly times of the operator and preconditioner were minimised while GMRES solves were fast due to the preconditioning effect of P and were distributed to different CPUs.

A total of 4352 orientations have been considered in other databases, such as [73] using the DDA method of [47], in order to simulate random orientation. We find that 14 incident waves are enough to simulate the SSPs in random

[1]We note that the temperature affects the refractive index of the scatterer [11, 12].

orientation with 1% relative error for the smallest aggregates, while we use up to 302 incident waves for the large ones. We discuss the database in detail in Chapter 7.

1.2 Thesis Outline and Contributions

We begin this thesis by reviewing background material that is essential to create the BIE formulations and obtain a stable discretisation of the Calderón product, in Chapters 2 and 3. In particular, in Chapter 2, we describe the scattering problem by multiple dielectric bodies in detail. We recall the function spaces, traces, potential operators and boundary integral operators required to form the BIE formulations. Using those, we present the PMCHWT boundary integral formulation extended to the multi-particle setup.

In Chapter 3, we discuss the transition from the continuous to the discrete setting, the available discrete spaces and the choice(s) we make for a stable discretisation of Calderón preconditioning. We make the choice to discuss the variational forms and discretisation before introducing the operator preconditioner so that we can focus on its impact in the following chapters. The fundamentals of $\mathcal{H}$-matrix and quadrature methods are also discussed as they are later used in Chapter 5.

Our contribution to current literature is presented in Chapters 4, 5, 6 and 7. The results presented in Chapters 4, 5 and 6 have been published in [2] and [3]. We begin Chapter 4, by presenting the fundamentals of traditional Calderón preconditioning in current literature (for example in [36, 38]). Our numerical experiments demonstrate that depending on the discretisation scheme used, traditional Calderón preconditioning is no better than simple mass-matrix preconditioning, regardless of the mesh size h. Our contribution with regards to (i) is presented through reduced versions of the original Calderón preconditioner. Our numerical experiments demonstrate that a 50-75% reduction in memory cost and 60-80% reduction in total computational time is possible with the use of the reduced preconditioners, compared to the

traditional one presented in [36].

In Chapter 5, we present our contribution to existing work with regards to (ii). Our numerical experiments demonstrate that a 60-90% reduction in memory consumption and a 50-60% reduction in total computational time is achieved when a bi-parametric version of the traditional preconditioner of [36] is used. Going further, a bi-parametric implementation of the reduced preconditioners presented in Chapter 4 can achieve more than 95% reduction in memory cost and about 80% reduction in total computation time.

In Chapter 6, we apply the accelerating methods of Chapters 4 and 5 to realistic complex ice crystal configurations such as those presented in [5] and [6]. We demonstrate that for the highest frequency considered, 664 GHz, an estimated 99% reduction in memory consumption and 80% reduction in total computation time can be achieved by a bi-parametric reduced preconditioner, completely alleviating the otherwise prohibitive cost imposed by the barycentrically refined grid.

In Chapter 7, we present the scattering database that is being generated for future numerical weather prediction at the Met Office using our accelerated BEM method. We briefly discuss the microphysical model used for the database, although we note that this was not developed by the author of this thesis but it is included for completeness. An overview of the far field scattering properties and definitions is given along with the basics of random orientation. We present an alternative way of simulating random orientation and test our methodology against a T-matrix method for hexagonal columns (e.g. Figure 1.2 (a)). We note that T-matrix methods are fast and accurate but restricted to rotationally symmetric particles and therefore cannot be applied to the aggregates considered for the database (e.g. Figure 1.3). We discuss the accuracy of our database with respect to mesh refinement, number of incident waves and number of polarisation vectors assumed and briefly comment on the accuracy of other databases. We finish by presenting some early results from our simulations and show how these compare with other databases.

Concluding remarks and future research avenues are discussed in Chapter 8. All numerical experiments were performed using Bempp [2] [1] (Versions 3.3.4 and 3.3.5), an open-source boundary element library. Example notebooks for Chapters 4, 5, 6 and 7 are also available on the author's GitHub page [3].

[2] `www.bempp.com`

[3] Accelerated Calderón: `www.github.com/ankleanthous/Accelerated_Calderon`, Random Orientation: `www.github.com/ankleanthous/EMScattering`

Chapter 2

The scattering problem and the PMCHWT formulation

We begin by describing the scattering problem by multiple isotropic homogeneous dielectric scatterers of arbitrary shape in Section 2.1. Before we present the PMCHWT formulation in the context of multi-particle scattering, it is important that we define the proper function spaces along with their properties (Section 2.2), the potential operators (Section 2.3) and then the boundary integral operators (Section 2.4) that are required for the formulation. We define the function spaces and relevant operators for some generic domain, and in Section 2.5 apply those in the context of the scattering problem presented in Section 2.1. The main reference for the function spaces and operators is the work by Buffa and Hiptmair [74] unless otherwise stated. For the multi-particle scattering setup along with its PMCHWT formulation we refer to our work in [2] with further references made within the following sections. For the original source of the PMCHWT formulation for a single scatterer we refer to [25, 26, 27, 28].

2.1 The scattering problem

We consider M disjoint arbitrary 3D isotropic homogeneous dielectric scatterers occupying bounded domains $\Omega_m^i \subset \mathbb{R}^3$, $m = 1, \ldots, M$, with boundaries $\Gamma_m = \partial\Omega_m^i$, in a homogeneous exterior medium $\Omega^e = \mathbb{R}^3 \backslash \overline{\cup_{m=1}^M \Omega_m^i}$, as in Figure 1.1. The scatterers Ω_m^i, $m = 1, \ldots, M$, are assumed to be Lipschitz continuous. The electric and magnetic fields in the interior domains Ω_m^i, $m = 1, \ldots, M$, and the exterior domain Ω^e, will be denoted $(\mathbf{E}_m^i, \mathbf{H}_m^i)$ and $(\mathbf{E}^e, \mathbf{H}^e)$ respectively. They are assumed to satisfy the time-harmonic Maxwell equations

$$\nabla \times \mathbf{E}_m^i = i\omega\mu_m \mathbf{H}_m^i, \qquad \text{in } \Omega_m^i, m = 1, \ldots, M, \tag{2.1}$$

$$\nabla \times \mathbf{H}_m^i = -i\omega\epsilon_m \mathbf{E}_m^i, \qquad \text{in } \Omega_m^i, m = 1, \ldots, M, \tag{2.2}$$

and

$$\nabla \times \mathbf{E}^e = i\omega\mu_e \mathbf{H}^e, \qquad \text{in } \Omega^e, \tag{2.3}$$

$$\nabla \times \mathbf{H}^e = -i\omega\epsilon_e \mathbf{E}^e, \qquad \text{in } \Omega^e, \tag{2.4}$$

together with the transmission boundary conditions

$$\mathbf{E}_m^i(\mathbf{x}) \times \mathbf{n}_m = \mathbf{E}^e(\mathbf{x}) \times \mathbf{n}_m, \quad \mathbf{x} \in \Gamma_m, \ m = 1, \ldots, M, \tag{2.5}$$

$$\mathbf{H}_m^i(\mathbf{x}) \times \mathbf{n}_m = \mathbf{H}^e(\mathbf{x}) \times \mathbf{n}_m, \quad \mathbf{x} \in \Gamma_m, \ m = 1, \ldots, M. \tag{2.6}$$

Here we assume a time-dependence of the form $e^{-i\omega t}$, with angular frequency $\omega > 0$. The parameters ϵ_m, ϵ_e and μ_m, μ_e, represent respectively the electric permittivity and the magnetic permeability of the domains, and $\mathbf{n}_m$ is the unit normal vector on Γ_m pointing into Ω^e.

In the scattering problem, an incident field $(\mathbf{E}^{inc}, \mathbf{H}^{inc})$ (for instance, a plane wave) gives rise to internal fields $(\mathbf{E}_m^i, \mathbf{H}_m^i)$ in Ω_m^i and a scattered field $(\mathbf{E}^s, \mathbf{H}^s)$ in the exterior domain Ω^e. The latter is assumed to satisfy the Silver-

Müller radiation condition

$$\sqrt{\mu_e}\mathbf{H}^s \times \frac{\mathbf{x}}{|\mathbf{x}|} - \sqrt{\epsilon_e}\mathbf{E}^s = \mathcal{O}\left(\frac{1}{|\mathbf{x}|}\right), \quad \text{as } |\mathbf{x}| \to \infty, \tag{2.7}$$

and the total exterior field is then the sum of incident and scattered fields

$$\mathbf{E}^e = \mathbf{E}^{inc} + \mathbf{E}^s, \qquad \text{in } \Omega^e, \tag{2.8}$$

$$\mathbf{H}^e = \mathbf{H}^{inc} + \mathbf{H}^s, \qquad \text{in } \Omega^e. \tag{2.9}$$

It is sufficient to solve for either the electric or magnetic fields and then recover the remaining fields by (2.1)–(2.2) and (2.3)–(2.4). In what follows, we will solve for the electric fields $\mathbf{E}^i_m$, $\mathbf{E}^e$, which satisfy

$$\nabla \times (\nabla \times \mathbf{E}^i_m) - k^2_m \mathbf{E}^i_m = 0, \qquad \text{in } \Omega^i_m,\ m = 1, \ldots, M, \tag{2.10}$$

$$\nabla \times (\nabla \times \mathbf{E}^e) - k^2_e \mathbf{E}^e = 0, \qquad \text{in } \Omega^e, \tag{2.11}$$

where $k_m = \omega\sqrt{\mu_m \epsilon_m}$ and $k_e = \omega\sqrt{\mu_e \epsilon_e}$ are the wavenumbers in the respective domains.

2.2 Function Spaces

Let $\Omega^i \subset \mathbb{R}^3$, represent the interior of some domain, $\Omega^e = \mathbb{R}^3 \backslash \overline{\Omega^i}$ its exterior and $\Gamma = \partial\Omega^i$ its boundary surface. We use $\Omega \subseteq \mathbb{R}^3$ to denote any of the subsets Ω^i, Ω^e or $\mathbb{R}^3$. In what follows, the subscript $_{\text{loc}}$ should be dropped when Ω refers to a bounded domain, i.e. Ω^i. We follow [74] for the following sections.

2.2.1 Function Spaces in the domain

We denote the space of complex, vector valued, locally square integrable functions by $\mathbf{L}^2_{\text{loc}}(\Omega)$ for some $\mathbf{u} : \Omega \to \mathbb{C}^3$ and the Sobolev space $\mathbf{H}^s_{\text{loc}}(\Omega)$, $s \geq 0$, $s \in \mathbb{R}$. Let $\mathbf{H} := \mathbf{H}^0$ (i.e. $\mathbf{H}^s$ with $s = 0$). Let d and $\mathbf{d}$ be some scalar and vector differential operators. Then for any $s \geq 0$ we can define the 'd' and '$\mathbf{d}$'

spaces by

$$\mathbf{H}^s_{\mathrm{loc}}(d,\Omega) := \{\mathbf{u} \in \mathbf{H}^s_{\mathrm{loc}}(\Omega) : \ d\mathbf{u} \in H^s_{\mathrm{loc}}(\Omega)\}, \tag{2.12}$$

$$\mathbf{H}^s_{\mathrm{loc}}(\mathbf{d},\Omega) := \{\mathbf{u} \in \mathbf{H}^s_{\mathrm{loc}}(\Omega) : \ \mathbf{du} \in \mathbf{H}^s_{\mathrm{loc}}(\Omega)\}, \tag{2.13}$$

and the 'd-free' and '$\mathbf{d}$-free' spaces

$$\mathbf{H}^s_{\mathrm{loc}}(d0,\Omega) := \{\mathbf{u} \in \mathbf{H}^s_{\mathrm{loc}}(\Omega) : \ d\mathbf{u} = 0\}, \tag{2.14}$$

$$\mathbf{H}^s_{\mathrm{loc}}(\mathbf{d}0,\Omega) := \{\mathbf{u} \in \mathbf{H}^s_{\mathrm{loc}}(\Omega) : \ \mathbf{du} = 0\}. \tag{2.15}$$

The above definitions can then be used to define the spaces of div- and **curl**-conforming functions $\mathbf{H}(\mathrm{div},\Omega^i)$ and $\mathbf{H}(\mathbf{curl},\Omega^i)$ and the spaces of div-free and **curl**-free functions $\mathbf{H}(\mathrm{div}0,\Omega^i)$ and $\mathbf{H}(\mathbf{curl}0,\Omega^i)$ on Ω^i. Since $\mathbf{u}$ is a locally square integrable solution satisfying (2.10) or (2.11), then **curl curl u** is also locally square integrable. The solution to the electric/magnetic wave equation with constant coefficients will be contained in the space $\mathbf{H}_{\mathrm{loc}}(\mathbf{curl}^2,\Omega)$.

2.2.2 Function Spaces on the boundary Γ

One can define Sobolev spaces on the manifold Γ denoted by $H^s(\Gamma)$ and $\mathbf{H}^s(\Gamma)$, $s \in [-1,1]$, for scalar and vectors spaces respectively. To map from Sobolev spaces in the domain to Sobolev spaces on the boundary we use trace operators.

The Dirichlet (tangential in [74]), Neumann and normal traces on Γ are defined as follows, for $\mathbf{p} \in \mathbf{H}_{\mathrm{loc}}(\mathbf{curl},\Omega)$, $\mathbf{q} \in \mathbf{H}_{\mathrm{loc}}(\mathbf{curl}^2,\Omega)$ and $\mathbf{r} \in \mathbf{H}_{\mathrm{loc}}(\mathrm{div},\Omega)$

$$\gamma_D^{\pm}\mathbf{p}(\mathbf{x}) := \mathbf{p}(\mathbf{x}) \times \mathbf{n}, \quad \mathbf{x} \in \Gamma, \tag{2.16}$$

$$\gamma_N^{\pm}\mathbf{q}(\mathbf{x}) := \frac{1}{ik}\gamma_D^{\pm}(\mathbf{curl}\ \mathbf{q}(\mathbf{x})), \quad \mathbf{x} \in \Gamma, \tag{2.17}$$

$$\gamma_n^{\pm}\mathbf{r}(\mathbf{x}) := \mathbf{r}(\mathbf{x}) \cdot \mathbf{n}, \quad \mathbf{x} \in \Gamma, \tag{2.18}$$

where the superscripts $^+$ and $^-$ denote the exterior and interior traces respectively and $\mathbf{p}(\mathbf{x})$, $\mathbf{q}(\mathbf{x})$ and $\mathbf{r}(\mathbf{x})$ exist either in Ω^e or Ω^i respectively, and $\mathbf{n}$ is the unit normal vector on Γ pointing into Ω^e. Note that we use a different def-

inition for the Neumann trace than that in [74] which differs by a factor of $1/i$. The interpretation of that is that if we normalise the magnetic permittivity and electric permeability to 1, the definition of (2.17) is the tangential trace of the magnetic field [75]. The jump and mean combinations of the traces are given respectively by

$$[\gamma_\cdot]_\Gamma := \gamma_\cdot^+ - \gamma_\cdot^-, \tag{2.19}$$

$$\{\gamma_\cdot\}_\Gamma := \frac{1}{2}(\gamma_\cdot^+ + \gamma_\cdot^-). \tag{2.20}$$

We define the space of square integrable tangential vector functions on the boundary [75]

$$\mathbf{L}^2_{\mathbf{t}}(\Gamma) := \left\{\mathbf{u} \in \mathbf{L}^2(\Gamma) : \mathbf{u}\cdot\mathbf{n} = 0\right\}. \tag{2.21}$$

We define the tangential trace space

$$\mathbf{H}^{\frac{1}{2}}_{\times}(\Gamma) := \gamma_D^-(\mathbf{H}^1(\Omega^i)) \ = \left\{\gamma_D^-\mathbf{u} \ : \mathbf{u} \in \mathbf{H}^1(\Omega^i)\right\}, \tag{2.22}$$

with its dual $\mathbf{H}^{-\frac{1}{2}}_{\times}(\Gamma)$ with respect to the antisymmetric product

$$\langle \mathbf{a}, \mathbf{b}\rangle_\tau := \int_\Gamma \mathbf{a}\cdot(\mathbf{n}\times\mathbf{b}), \text{ for } \mathbf{a},\ \mathbf{b} \in \mathbf{L}^2_{\mathbf{t}}(\Gamma). \tag{2.23}$$

Remark 1. *The dual space of some Hilbert space* $\mathbf{H}$*, denoted by* $\mathbf{H}^*$*, is the linear space of linear continuous functionals on* $\mathbf{H}$*, which is also a Hilbert space. For practical applications, it is often convenient to work with* ***realisations*** *[76] of the dual space instead of the actual dual space. By a realisation of* $\mathbf{H}^*$*, we mean a pair* $\left(\mathbf{H}^{\text{dual}}, \boldsymbol{\mathcal{L}}\right)$ *where* $\mathbf{H}^{\text{dual}}$ *is a Hilbert space and* $\boldsymbol{\mathcal{L}} : \mathbf{H}^{\text{dual}} \to \mathbf{H}^*$ *is an isomorphism, i.e. a bounded linear operator with a bounded inverse, from which we define*

$$\langle u, v\rangle_{\mathbf{H}^{\text{dual}}\times\mathbf{H}} := (\boldsymbol{\mathcal{L}}u)(v), \quad u \in \mathbf{H}^{\text{dual}},\ v \in \mathbf{H}. \tag{2.24}$$

When we say $\mathbf{H}_{\times}^{-\frac{1}{2}}(\Gamma)$ *is the dual of* $\mathbf{H}_{\times}^{\frac{1}{2}}(\Gamma)$ *with respect to the antisymmetric product of* (2.23)*, this should be interpreted as* $\left(\mathbf{H}_{\times}^{-\frac{1}{2}}(\Gamma), \langle \cdot, \cdot \rangle_{\tau}\right)$ *being a realisation of the actual dual space of* $\mathbf{H}_{\times}^{\frac{1}{2}}(\Gamma)$*. The duality pairing* $\langle \cdot, \cdot \rangle_{\tau}$ *on* $\mathbf{H}_{\times}^{-\frac{1}{2}}(\Gamma) \times \mathbf{H}_{\times}^{\frac{1}{2}}(\Gamma)$ *is extended from the duality pairing of* (2.23) *due to the density of* $\mathbf{H}_{\times}^{\frac{1}{2}}(\Gamma)$ *in* $\mathbf{L}_{\mathbf{t}}^2$*.*

For smooth surfaces, the scalar surface divergence of $\gamma_D^{\pm}\mathbf{u}$, for $\mathbf{u} \in \mathbf{H}_{\mathrm{loc}}(\mathbf{curl}, \Omega)$ is given by

$$\mathrm{div}_{\Gamma}(\gamma_D^{\pm}\mathbf{u}) = \gamma_n^{\pm}(\mathbf{curl}\mathbf{u}). \tag{2.25}$$

In the case of piecewise smooth boundaries this becomes more complicated as the Dirichlet trace is discontinuous across edges of Γ. Assuming that Γ consists of a finite number of smooth faces (as is the case when the boundary is approximated by flat triangles during discretisation), we can express this as

$$\Gamma := \bigcup_{j=1}^{N} \overline{\Gamma^j}, \tag{2.26}$$

for the faces $\Gamma^1, \ldots, \Gamma^N$, $N \in \mathbb{N}$. Then for $\mathbf{u} \in C^{\infty}(\overline{\Omega})$, the scalar surface divergence is given by

$$\mathrm{div}_{\Gamma}\gamma_D^{\pm}\mathbf{u} := \begin{cases} \mathrm{div}_j(\gamma_D^{\pm}\mathbf{u})^j, & \text{on } \Gamma^j, \\ \left((\gamma_D^{\pm}\mathbf{u})^j \cdot \mathbf{n}^{ij} + (\gamma_D^{\pm}\mathbf{u})^i \cdot \mathbf{n}^{ji}\right)\delta_{ij}, & \text{on } \overline{\Gamma^j} \cap \overline{\Gamma^i}, \end{cases} \tag{2.27}$$

where $\mathbf{u}^j$ is the restriction of $\mathbf{u}$ to the face Γ^j, $\mathbf{n}^{ij}$ is the outward pointing tangential normal to Γ^i restricted to the edge $\overline{\Gamma^j} \cap \overline{\Gamma^i}$, div_j is the 2D-divergence computed on the face Γ^j, and δ_{ij} is the delta distribution in local coordinates with support on the edge $\overline{\Gamma^j} \cap \overline{\Gamma^i}$ [74, 75]. By density, this operator can be extended to functionals $\mathbf{u} \in \mathbf{H}_{\times}^{-\frac{1}{2}}(\Gamma)$ [74].

We define the space of surface-div-conforming functions by

$$\mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma}, \Gamma) := \{\mathbf{u} \in \mathbf{H}_{\times}^{-\frac{1}{2}}(\Gamma) : \mathrm{div}_{\Gamma}\mathbf{u} \in \mathbf{H}^{-\frac{1}{2}}(\Gamma)\}. \tag{2.28}$$

The space $\mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_\Gamma,\Gamma)$ is self-dual with respect to the antisymmetric dual form (2.23) [77]. This should be interpreted as a realisation of the dual space, in the sense described in Remark 1. The scalar surface curl is defined as [77]

$$\mathrm{curl}_\Gamma \mathbf{u} := \mathrm{div}_\Gamma(\mathbf{u}\times\mathbf{n}), \qquad \mathbf{u}\in\mathbf{H}_{\times}^{-\frac{1}{2}}(\Gamma). \tag{2.29}$$

and the space of surface-curl-conforming functions by

$$\mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{curl}_\Gamma,\Gamma) := \{\mathbf{u}\in\mathbf{H}_{\times}^{-\frac{1}{2}}(\Gamma) : \mathrm{curl}_\Gamma\mathbf{u}\in H^{-\frac{1}{2}}(\Gamma)\} \tag{2.30}$$

$$= \{\mathbf{u}\times\mathbf{n} : \mathbf{u}\in\mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_\Gamma,\Gamma)\}. \tag{2.31}$$

The space $\mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{curl}_\Gamma,\Gamma)$ can also be seen as dual to $\mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_\Gamma,\Gamma)$ with respect to the standard $\mathbf{L}^2$ product (again, as described in Remark 1).

The above definitions allow us to obtain properties of the trace operators: the Dirichlet and Neumann traces

$$\gamma_D^{\pm} : \mathbf{H}_{\mathrm{loc}}(\mathbf{curl},\Omega)\to\mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_\Gamma,\Gamma), \tag{2.32}$$

$$\gamma_N^{\pm} : \mathbf{H}_{\mathrm{loc}}(\mathbf{curl}^2,\Omega)\to\mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_\Gamma,\Gamma), \tag{2.33}$$

are continuous and surjective [74]. For proofs we refer to [78] for Lipschitz polyhedra and [77] for general Lipschitz domains.

2.3 Potential Operators

Potential operators allow us to map boundary data to analytic functions defined everywhere off the boundary and are later used in the representation formulae allowing us to construct the field off the boundary from the obtained boundary solution. For electromagnetic Maxwell problems, two potential operators are required; the electric and magnetic potential operators $\mathcal{E},\ \mathcal{H} : \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_\Gamma,\Gamma)\to\mathbf{H}_{\mathrm{loc}}(\mathbf{curl}^2,\Omega^i\cup\Omega^e)\cap\mathbf{H}_{\mathrm{loc}}(\mathrm{div}0,\Omega^i\cup\Omega^e)$ defined by

$$\mathcal{E}\mathbf{v}(\mathbf{x}) := ik\int_\Gamma \mathbf{v}(\mathbf{y})G(\mathbf{x},\mathbf{y})d\Gamma(\mathbf{y}) - \frac{1}{ik}\nabla_{\mathbf{x}}\int_\Gamma \nabla_{\mathbf{y}}\cdot\mathbf{v}(\mathbf{y})G(\mathbf{x},\mathbf{y})d\Gamma(\mathbf{y}), \tag{2.34}$$

$$\mathcal{H}\mathbf{v}(\mathbf{x}) := \nabla_{\mathbf{x}} \times \int_{\Gamma} \mathbf{v}(\mathbf{y}) G(\mathbf{x},\mathbf{y}) d\Gamma(\mathbf{y}), \tag{2.35}$$

where $G(\mathbf{x},\mathbf{y})$ is the fundamental solution of the Helmholtz equation given by

$$\mathbf{G}(\mathbf{x},\mathbf{y}) = \frac{\exp(ik|\mathbf{x}-\mathbf{y}|)}{4\pi|\mathbf{x}-\mathbf{y}|}. \tag{2.36}$$

We note that the definition of $\mathcal{E}$ differs from the one in [74] by a factor of i to account for the modified definition of $\gamma_N^{\pm}$ in (2.17).

The potentials satisfy the following symmetry relations on $\mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma},\Gamma)$

$$\mathbf{curl}\ \mathcal{E} = ik\mathcal{H}, \tag{2.37}$$

$$\mathbf{curl}\ \mathcal{H} = -ik\mathcal{E}, \tag{2.38}$$

and both potentials are Maxwell solutions, satisfying

$$\mathbf{curl}\ \mathbf{curl}\ \mathcal{E}(\mathbf{u}) - k^2\mathcal{E}(\mathbf{u}) = 0, \tag{2.39}$$

$$\mathbf{curl}\ \mathbf{curl}\ \mathcal{H}(\mathbf{u}) - k^2\mathcal{H}(\mathbf{u}) = 0, \tag{2.40}$$

for $\mathbf{u} \in \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma},\Gamma)$, while also satisfying the Silver-Müller radiation conditions. The following continuity properties hold: The mappings

$$\mathcal{E} : \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma},\Gamma) \to \mathbf{H}_{\mathrm{loc}}(\mathbf{curl}^2,\Omega^i\cup\Omega^e) \cap \mathbf{H}_{\mathrm{loc}}(\mathrm{div}0,\Omega^i\cup\Omega^e), \tag{2.41}$$

$$\mathcal{H} : \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma},\Gamma) \to \mathbf{H}_{\mathrm{loc}}(\mathbf{curl}^2,\Omega^i\cup\Omega^e) \cap \mathbf{H}_{\mathrm{loc}}(\mathrm{div}0,\Omega^i\cup\Omega^e), \tag{2.42}$$

are continuous.

Given that $\mathcal{E}$ and $\mathcal{H}$ satisfy the Maxwell equations, the Silver-Müller radiation conditions and are continuous mappings, we can use them to express the Maxwell solutions in terms of integral equations. This is done via the Stratton-Chu representation formulae as follows. Any Maxwell solution $\mathbf{u} \in \mathbf{H}(\mathbf{curl}^2,\Omega^i)$

satisfies

$$\mathcal{H}\left(\gamma_D^- \mathbf{u}\right) + \mathcal{E}\left(\gamma_N^- \mathbf{u}\right) = \begin{cases} \mathbf{u}(\mathbf{x}), & \mathbf{x} \in \Omega^i, \\ 0, & \mathbf{x} \notin \overline{\Omega^i}. \end{cases} \tag{2.43}$$

Any Maxwell solution $\mathbf{u} \in \mathbf{H}_{\text{loc}}(\mathbf{curl}^2, \Omega^e)$ that also satisfies the Silver-Müller radiation conditions can be written as

$$-\mathcal{H}\left(\gamma_D^+ \mathbf{u}\right) - \mathcal{E}\left(\gamma_N^+ \mathbf{u}\right) = \begin{cases} \mathbf{u}(\mathbf{x}), & \mathbf{x} \in \Omega^e, \\ 0, & \mathbf{x} \notin \overline{\Omega^e}. \end{cases} \tag{2.44}$$

2.4 Boundary Integral Operators

By taking traces of the potential operators we can define boundary integral operators. The electric and magnetic boundary integral operators on Γ are defined as follows

$$\mathcal{S} := \{\gamma_D\}\mathcal{E}, \tag{2.45}$$

$$\mathcal{C} := \{\gamma_D\}\mathcal{H}, \tag{2.46}$$

for $\mathcal{S}$, $\mathcal{C} : \mathbf{H}_\times^{-\frac{1}{2}}(\text{div}_\Gamma, \Gamma) \to \mathbf{H}_\times^{-\frac{1}{2}}(\text{div}_\Gamma, \Gamma)$. Due to the symmetry relations (2.37)–(2.38) the following properties hold

$$\gamma_N^\pm \mathcal{E} = \gamma_D^\pm \mathcal{H}, \tag{2.47}$$

$$\gamma_N^\pm \mathcal{H} = -\gamma_D^\pm \mathcal{E}, \tag{2.48}$$

allowing us to also express $\mathcal{S}$ and $\mathcal{C}$ through Neumann traces

$$\mathcal{S} := -\{\gamma_N\}\mathcal{H}, \tag{2.49}$$

$$\mathcal{C} := \{\gamma_N\}\mathcal{E}. \tag{2.50}$$

The following continuity properties hold: The mappings

$$\mathcal{S} : \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma}, \Gamma) \to \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma}, \Gamma), \tag{2.51}$$

$$\mathcal{C} : \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma}, \Gamma) \to \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma}, \Gamma), \tag{2.52}$$

are continuous.

To describe the behaviour of the potentials when crossing Γ, we also need jump relations. For the Maxwell setup, the following jump relations hold [74]

$$[\gamma_D]\mathcal{E} = [\gamma_N]\mathcal{H} = 0, \tag{2.53}$$

$$[\gamma_N]\mathcal{E} = [\gamma_D]\mathcal{H} = -\mathcal{I}, \tag{2.54}$$

where $\mathcal{I} : \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma}, \Gamma) \to \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma}, \Gamma)$ is the identity operator mapping every function to itself. Using the definitions (2.45)–(2.46), (2.49)–(2.50) and the jump relations (2.53)–(2.54) the following also hold

$$\mathcal{S} = \gamma_D^{\pm}\mathcal{E} = -\gamma_N^{\pm}\mathcal{H}, \tag{2.55}$$

$$\mathcal{C} = \gamma_N^{\pm}\mathcal{E} \pm \frac{1}{2}\mathcal{I} = \gamma_D^{\pm}\mathcal{H} \pm \frac{1}{2}\mathcal{I}. \tag{2.56}$$

Following the definition of the electric potential operator (2.34) and the relation (2.55), we can express the electric boundary integral operator $\mathcal{S}$ by

$$\mathcal{S} = ik\mathcal{S}^s - \frac{1}{ik}\mathcal{S}^h, \tag{2.57}$$

where $\mathcal{S}^s$ is smooth and $\mathcal{S}^h$ hypersingular. This decomposition is crucial in the need of preconditioning, and is discussed in detail in Chapter 4.

Applying interior Dirichlet and Neumann traces to the interior Stratton Chu formula (2.43) and using (2.55)–(2.56) we can get

$$\gamma_D^-\mathbf{u} = \left(\mathcal{C} + \frac{1}{2}\mathcal{I}\right)(\gamma_D^-\mathbf{u}) + \mathcal{S}(\gamma_N^-\mathbf{u}), \tag{2.58}$$

$$\gamma_N^-\mathbf{u} = -\mathcal{S}(\gamma_D^-\mathbf{u}) + \left(\mathcal{C} + \frac{1}{2}\mathcal{I}\right)(\gamma_N^-\mathbf{u}), \tag{2.59}$$

and similarly by applying exterior traces to the exterior formula (2.44)

$$\gamma_D^+\mathbf{u} = -\left(\mathcal{C} - \frac{1}{2}\mathcal{I}\right)(\gamma_D^+\mathbf{u}) - \mathcal{S}(\gamma_N^+\mathbf{u}), \tag{2.60}$$

$$\gamma_N^+\mathbf{u} = \mathcal{S}(\gamma_D^+\mathbf{u}) - \left(\mathcal{C} - \frac{1}{2}\mathcal{I}\right)(\gamma_N^+\mathbf{u}). \tag{2.61}$$

The above can be written as

$$\left(\frac{1}{2}\boldsymbol{\mathcal{I}} + \tilde{\boldsymbol{\mathcal{A}}}\right)\begin{bmatrix}\gamma_D^-\mathbf{u}\\ \gamma_N^-\mathbf{u}\end{bmatrix} = \begin{bmatrix}\gamma_D^-\mathbf{u}\\ \gamma_N^-\mathbf{u}\end{bmatrix}, \tag{2.62}$$

$$\left(\frac{1}{2}\boldsymbol{\mathcal{I}} - \tilde{\boldsymbol{\mathcal{A}}}\right)\begin{bmatrix}\gamma_D^+\mathbf{u}\\ \gamma_N^+\mathbf{u}\end{bmatrix} = \begin{bmatrix}\gamma_D^+\mathbf{u}\\ \gamma_N^+\mathbf{u}\end{bmatrix}, \tag{2.63}$$

where $\tilde{\boldsymbol{\mathcal{A}}} : \mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_\Gamma,\Gamma)^2 \to \mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_\Gamma,\Gamma)^2$ and $\boldsymbol{\mathcal{I}} : \mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_\Gamma,\Gamma)^2 \to \mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_\Gamma,\Gamma)^2$ are

$$\tilde{\boldsymbol{\mathcal{A}}} = \begin{bmatrix}\mathcal{C} & \mathcal{S}\\ -\mathcal{S} & \mathcal{C}\end{bmatrix}, \qquad \boldsymbol{\mathcal{I}} = \begin{bmatrix}\mathcal{I} & 0\\ 0 & \mathcal{I}\end{bmatrix}. \tag{2.64}$$

The operators $\left(\frac{1}{2}\boldsymbol{\mathcal{I}} + \tilde{\boldsymbol{\mathcal{A}}}\right)$ and $\left(\frac{1}{2}\boldsymbol{\mathcal{I}} - \tilde{\boldsymbol{\mathcal{A}}}\right) : \mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_\Gamma,\Gamma)^2 \to \mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_\Gamma,\Gamma)^2$ are the interior and exterior Calderón projectors and by the Stratton-Chu representation formulae satisfy

$$\left(\frac{1}{2}\boldsymbol{\mathcal{I}} + \tilde{\boldsymbol{\mathcal{A}}}\right)^2 = \left(\frac{1}{2}\boldsymbol{\mathcal{I}} + \tilde{\boldsymbol{\mathcal{A}}}\right), \tag{2.65}$$

$$\left(\frac{1}{2}\boldsymbol{\mathcal{I}} - \tilde{\boldsymbol{\mathcal{A}}}\right)^2 = \left(\frac{1}{2}\boldsymbol{\mathcal{I}} - \tilde{\boldsymbol{\mathcal{A}}}\right) \tag{2.66}$$

Expanding (2.65)–(2.66) leads to the following property

$$\left(\tilde{\boldsymbol{\mathcal{A}}}\right)^2 = \frac{1}{4}\boldsymbol{\mathcal{I}}, \tag{2.67}$$

which can then be expanded to reveal the following Calderón identities

$$\mathcal{S}^2 = -\frac{1}{4}\mathcal{I} + \mathcal{C}^2, \tag{2.68}$$

$$\mathcal{C}\mathcal{S} + \mathcal{S}\mathcal{C} = 0. \tag{2.69}$$

These identities and their properties are fundamental in operator-based Calderón preconditioning, and will be discussed later in Chapter 4. The operator $\mathcal{C}$ is compact on smooth domains [79] and therefore $\mathcal{C}^2$ is also compact. Equation (2.68) implies that $\mathcal{S}^2$ is a second-kind integral operator, i.e. of the form "scaled identity plus compact operator" with eigenvalues accumulating at $-\frac{1}{4}$. This is known as the "self-regularising" property of $\mathcal{S}$ [80], and is a crucial part of Calderón preconditioning discussed later in Chapter 4.

2.5 Boundary Integral Formulations

Now that we have defined the fundamental spaces and operators required we can apply those in the context of scattering by multiple scatterers as was described in Section 2.1. We refer to Table 2.1 for the new notation of the operators adapted to the multi-particle setup. The main reference for this section in the context of multi-particle notation is our work in [2]. We begin with the boundary conditions (2.5)–(2.6) that can be written in terms of Dirichlet traces as follows

$$\gamma_{D,m}^{-}\mathbf{E}_m^i = \gamma_{D,m}^{+}\mathbf{E}^e, \quad m = 1, \ldots, M, \tag{2.70}$$

$$\gamma_{D,m}^{-}\mathbf{H}_m^i = \gamma_{D,m}^{+}\mathbf{H}^e, \quad m = 1, \ldots, M. \tag{2.71}$$

Given that we are interested in solving (2.10)–(2.11) for the electric fields $\mathbf{E}_m^i,\ m = 1, \ldots, M$, and $\mathbf{E}^e$ we can get rid of the magnetic fields in the above traces by using (2.2) and (2.4). Using the definition of the Neumann trace (2.17) and (2.8) we can write the transmission boundary conditions as

$$\mathbf{u}_m^i = \mathbf{u}_m^s + \mathbf{u}_m^{inc}, \quad m = 1, \ldots, M, \tag{2.72}$$

Notation	Interpretation
$(\gamma_{D,m}^-, \gamma_{N,m}^-)$	(γ_D^-, γ_N^-) for $\Gamma = \Gamma_m$ and $k = k_m$, for $m = 1, \ldots, M$
$(\gamma_{D,m}^+, \gamma_{N,m}^+)$	same as above but with $k = k_e$
$(\mathcal{E}_m^i, \mathcal{H}_m^i)$	$(\mathcal{E}, \mathcal{H})$ for $\Gamma = \Gamma_m$ and $k = k_m$, for $m = 1, \ldots, M$
$(\mathcal{E}_m^e, \mathcal{H}_m^e)$	same as above but with $k = k_e$
$(\mathcal{C}_m^i, \mathcal{S}_m^i)$	$(\mathcal{C}, \mathcal{S})$ for $\Gamma = \Gamma_m$ and $k = k_m$, for $m = 1, \ldots, M$
$(\mathcal{C}_m^e, \mathcal{S}_m^e)$	same as above but with $k = k_e$
$(\mathcal{C}_{m\ell}^e, \mathcal{S}_{m\ell}^e)$	map from Γ_ℓ to Γ_m and are defined for a boundary vector field $\mathbf{v}_\ell$ on Γ_l by $\mathcal{C}_{m\ell}^e \mathbf{v}_\ell := (\mathcal{C}^e \tilde{\mathbf{v}}_l)\|_{\Gamma_m}, \quad \mathcal{S}_{m\ell}^e \mathbf{v}_\ell := (\mathcal{S}^e \tilde{\mathbf{v}}_l)\|_{\Gamma_m}$, where $\tilde{\mathbf{v}}_l$ denotes the vector field on $\cup_{j=1}^M \Gamma_j$ which equals $\mathbf{v}_\ell$ on Γ_ℓ and zero on Γ_j, for $j \neq \ell$; $\|_{\Gamma_m}$ its restriction to Γ_m; and $(\mathcal{C}^e, \mathcal{S}^e)$ are $(\mathcal{C}, \mathcal{S})$ for $\Gamma = \cup_{j=1}^M \Gamma_j$ and $k = k_e$
$\mathcal{I}_m$	$\mathcal{I}$ for $\Gamma = \Gamma_m$, for $m = 1, \ldots, M$

Table 2.1: Interpretation of the operators in the context of multi-particle scattering.

where

$$\mathbf{u}_m^i = \begin{bmatrix} \gamma_{D,m}^- \mathbf{E}_m^i \\ \dfrac{k_m}{\mu_m} \gamma_{N,m}^- \mathbf{E}_m^i \end{bmatrix}, \quad \mathbf{u}_m^s = \begin{bmatrix} \gamma_{D,m}^+ \mathbf{E}^s \\ \dfrac{k_e}{\mu_e} \gamma_{N,m}^+ \mathbf{E}^s \end{bmatrix}, \quad \mathbf{u}_m^{inc} = \begin{bmatrix} \gamma_{D,m}^+ \mathbf{E}^{inc} \\ \dfrac{k_e}{\mu_e} \gamma_{N,m}^+ \mathbf{E}^{inc} \end{bmatrix}. \tag{2.73}$$

The next step is to re-write the Stratton Chu representation formulae (2.43)–(2.44) in the context of scattering by multiple particles as follows: The interior and exterior fields $\mathbf{E}_m^i$, $m = 1, \ldots, M$ and $\mathbf{E}^s$ can be represented as

$$\mathcal{H}_m^i \left(\gamma_{D,m}^- \mathbf{E}_m^i \right) + \mathcal{E}_m^i \left(\gamma_{N,m}^- \mathbf{E}_m^i \right) = \begin{cases} \mathbf{E}_m^i(\mathbf{x}), & \mathbf{x} \in \Omega_m^i, \\ 0, & \mathbf{x} \notin \overline{\Omega_m^i}. \end{cases} \tag{2.74}$$

$$-\sum_{m}^{M}\mathcal{H}_m^e\left(\gamma_{D,m}^+\mathbf{E}^s\right)-\sum_{m}^{M}\mathcal{E}_m^e\left(\gamma_{N,m}^+\mathbf{E}^s\right)=\begin{cases}\mathbf{E}^s(\mathbf{x}), & \mathbf{x}\in\Omega^e,\\ 0, & \mathbf{x}\notin\overline{\Omega^e}.\end{cases}\tag{2.75}$$

In a similar manner to the previous section, one can take appropriate interior and exterior Dirichlet and Neumann traces of (2.74)–(2.75) and use (2.55)–(2.56) to show that the boundary traces satisfy

$$\left(\frac{1}{2}\boldsymbol{\mathcal{I}}_m+\boldsymbol{\mathcal{A}}_m^i\right)\mathbf{u}_m^i=\mathbf{u}_m^i,\ m=1,\ldots,M,\tag{2.76}$$

$$\left(\frac{1}{2}\boldsymbol{\mathcal{I}}_m-\boldsymbol{\mathcal{A}}_m^e\right)\mathbf{u}_m^s-\sum_{\ell\neq m}^{M}\boldsymbol{\mathcal{A}}_{m\ell}\mathbf{u}_\ell^s=\mathbf{u}_m^s,\quad m=1,\ldots,M,\tag{2.77}$$

where

$$\boldsymbol{\mathcal{A}}_m^i=\begin{bmatrix}\mathcal{C}_m^i & \frac{\mu_m}{k_m}\mathcal{S}_m^i\\ -\frac{k_m}{\mu_m}\mathcal{S}_m^i & \mathcal{C}_m^i\end{bmatrix},\quad \boldsymbol{\mathcal{A}}_m^e=\begin{bmatrix}\mathcal{C}_m^e & \frac{\mu_e}{k_e}\mathcal{S}_m^e\\ -\frac{k_e}{\mu_e}\mathcal{S}_m^e & \mathcal{C}_m^e\end{bmatrix},\tag{2.78}$$

$$\boldsymbol{\mathcal{A}}_{m\ell}=\begin{bmatrix}\mathcal{C}_{m\ell}^e & \frac{\mu_e}{k_e}\mathcal{S}_{m\ell}^e\\ -\frac{k_e}{\mu_e}\mathcal{S}_{m\ell}^e & \mathcal{C}_{m\ell}^e\end{bmatrix},\quad \boldsymbol{\mathcal{I}}_m=\begin{bmatrix}\mathcal{I}_m & 0\\ 0 & \mathcal{I}_m\end{bmatrix}.\tag{2.79}$$

Here

$$\boldsymbol{\mathcal{A}}_m^i:\mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m},\Gamma_m)^2\to\mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m},\Gamma_m)^2,\tag{2.80}$$

$$\boldsymbol{\mathcal{A}}_m^e:\mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m},\Gamma_m)^2\to\mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m},\Gamma_m)^2,\tag{2.81}$$

$$\boldsymbol{\mathcal{A}}_{m\ell}:\mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_\ell},\Gamma_\ell)^2\to\mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m},\Gamma_m)^2,\tag{2.82}$$

$$\boldsymbol{\mathcal{I}}_m:\mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m},\Gamma_m)^2\to\mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m},\Gamma_m)^2.\tag{2.83}$$

The matrices $\left(\frac{1}{2}\boldsymbol{\mathcal{I}}_m+\boldsymbol{\mathcal{A}}_m^i\right)$ and $\left(\frac{1}{2}\boldsymbol{\mathcal{I}}_m-\boldsymbol{\mathcal{A}}_m^e\right)$ are scaled versions of the Calderón projectors and similar to (2.65)–(2.66) satisfy

$$\left(\frac{1}{2}\boldsymbol{\mathcal{I}}_m+\boldsymbol{\mathcal{A}}_m^i\right)^2=\left(\frac{1}{2}\boldsymbol{\mathcal{I}}_m+\boldsymbol{\mathcal{A}}_m^i\right),\tag{2.84}$$

$$\left(\frac{1}{2}\boldsymbol{\mathcal{I}}_m-\boldsymbol{\mathcal{A}}_m^e\right)^2=\left(\frac{1}{2}\boldsymbol{\mathcal{I}}_m-\boldsymbol{\mathcal{A}}_m^e\right).\tag{2.85}$$

Equations (2.76) and (2.77) hold for any solutions of the Maxwell equations (2.10)–(2.11). To obtain a solution for the specific dielectric scattering problem, one needs to combine (2.76) and (2.77) with the transmission boundary conditions (2.72). One can combine the three systems in different ways to obtain different formulations such as the Müller formulation [29] and its variations [30], or the PMCHWT formulation [25, 26, 27, 28]. We focus here on the well-studied PMCHWT formulation.

2.5.1 The PMCHWT formulation

To obtain the PMCHWT formulation, we subtract (2.77) from (2.76) and eliminate the interior boundary traces $\mathbf{u}_m^i$ via the boundary conditions (2.72) to obtain,

$$\left(\boldsymbol{\mathcal{A}}_m^i + \boldsymbol{\mathcal{A}}_m^e\right)\mathbf{u}_m^s + \sum_{\ell\neq m}^{M} \boldsymbol{\mathcal{A}}_{m\ell}\mathbf{u}_\ell^s = \left(\frac{1}{2}\boldsymbol{\mathcal{I}}_m - \boldsymbol{\mathcal{A}}_m^i\right)\mathbf{u}_m^{inc}, \tag{2.86}$$

for each $m = 1, \ldots, M$. We can combine these M systems into a block system

$$\boldsymbol{\mathcal{A}}\mathbf{u}^s = \left(\frac{1}{2}\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{D}}^i\right)\mathbf{u}^{inc}, \tag{2.87}$$

where

$$\boldsymbol{\mathcal{A}} = \begin{bmatrix} \boldsymbol{\mathcal{A}}_1^e + \boldsymbol{\mathcal{A}}_1^i & \boldsymbol{\mathcal{A}}_{12} & \cdots & \boldsymbol{\mathcal{A}}_{1M} \\ \boldsymbol{\mathcal{A}}_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{\mathcal{A}}_{(M-1)M} \\ \boldsymbol{\mathcal{A}}_{M1} & \cdots & \boldsymbol{\mathcal{A}}_{M(M-1)} & \boldsymbol{\mathcal{A}}_M^e + \boldsymbol{\mathcal{A}}_M^i \end{bmatrix}, \tag{2.88}$$

$$\boldsymbol{\mathcal{D}}^i = \begin{bmatrix} \boldsymbol{\mathcal{A}}_1^i & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{\mathcal{A}}_M^i \end{bmatrix}, \quad \boldsymbol{\mathcal{I}} = \begin{bmatrix} \boldsymbol{\mathcal{I}}_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{\mathcal{I}}_M \end{bmatrix}, \tag{2.89}$$

$$\mathbf{u}^s = \begin{bmatrix} \mathbf{u}_1^s \\ \mathbf{u}_2^s \\ \vdots \\ \mathbf{u}_M^s \end{bmatrix}, \quad \mathbf{u}^{inc} = \begin{bmatrix} \mathbf{u}_1^{inc} \\ \mathbf{u}_2^{inc} \\ \vdots \\ \mathbf{u}_M^{inc} \end{bmatrix}, \tag{2.90}$$

with

$$\boldsymbol{\mathcal{A}} : \bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2 \to \bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2, \tag{2.91}$$

$$\boldsymbol{\mathcal{A}}^i : \bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2 \to \bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2, \tag{2.92}$$

$$\boldsymbol{\mathcal{I}} : \bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2 \to \bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2. \tag{2.93}$$

Equation (2.87) is the PMCHWT formulation, expressed in the context of multi-particle scattering and can be reduced to the traditional PMCHWT formulation for one scatterer by taking $M = 1$. The PMCHWT formulation (2.87) has a unique solution for any frequency ([74] for $M = 1$ and [81] for $M > 1$).

Chapter 3

Galerkin Method, Discretisation and Matrix Assembly

Our goal in this chapter is to describe the Galerkin method used to discretise the PMCHWT formulation (2.87) that was described in Chapter 2 and give information on the methods used for matrix assembly. We begin this section with the variational problem of an abstract equation in the continuous and discrete setup. We work with an abstract form first for three reasons:

- the notation is easier to follow for an abstract operator $\boldsymbol{\mathcal{A}}$, instead of working with the block-structure of the PMCHWT formulation (2.87),
- the variational problem and transition from the continuous to the discrete setting can be used for other BIE formulations, such as the Müller formulation [29] and its variations [30],
- it is easier to understand how and why we pick the specific discrete spaces later.

We extend the variational problem in abstract form to the Maxwell transmission problem, using the PMCHWT formulation. We discuss available discrete spaces and choices for a stable discretisation of the Calderón product. Finally, we detail our matrix assembly methods, namely $\mathcal{H}$-matrix assembly and quadrature rules. The main references for the variational forms are [82, 76], for the choice of spaces for a stable discretisation [36, 75], and for $\mathcal{H}$-matrix assembly and quadrature rules [83, 14, 13, 84, 75].

3.1 Variational forms: abstract formulation

3.1.1 Continuous setting

We consider an abstract equation $\boldsymbol{\mathcal{A}}\mathbf{u} = \mathbf{f}$, where $\boldsymbol{\mathcal{A}} : \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dom}} \to \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}$ is a linear operator mapping between the Hilbert spaces $\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$ and $\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}$. Given $\mathbf{f} \in \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}$, to find a solution one would have to solve a problem of the form: *find* $\mathbf{u} \in \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$ *such that*

$$\boldsymbol{\mathcal{A}}\mathbf{u} = \mathbf{f}. \tag{3.1}$$

Let $\left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\right)^*$ be the dual space of $\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}$, i.e. the linear space of linear functionals on $\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}$, which is itself a Hilbert space. Then the problem (3.1) can be equivalently stated in variational form as: *find* $\mathbf{u} \in \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$ *such that*

$$\ell\left(\boldsymbol{\mathcal{A}}\mathbf{u}\right) = \ell\left(\mathbf{f}\right), \quad \textit{for all } \ell \in \left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\right)^*. \tag{3.2}$$

In practical applications it is often more convenient to work with *realisations* of the dual space [76] instead of the actual dual space. We refer back to Remark 1 from Chapter 2 for the definition of a realisation of a dual space. Let $\left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}, \boldsymbol{\mathcal{L}}\right)$ be a realisation of $\left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\right)^*$, where $\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}$ is a Hilbert space and $\boldsymbol{\mathcal{L}} : \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \to \left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\right)^*$ is an isomorphism, i.e. a bounded linear operator with a bounded inverse, from which we define the duality pairing

$$\langle u, v \rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \times \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}} := \left(\boldsymbol{\mathcal{L}} u\right)(v), \quad u \in \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}},\ v \in \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}. \tag{3.3}$$

The statement that $\left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}, \boldsymbol{\mathcal{L}}\right)$ is a realisation of $\left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\right)^*$ is equivalent to the statement that $\langle \cdot, \cdot \rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \times \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}}$ is bounded and satisfies an inf-sup condition [83].

Given such a realisation of $\left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\right)^*$, we can generate a realisation of $\left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right)^*$ with $\left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}, \boldsymbol{\mathcal{L}}^*\right)$ where $\boldsymbol{\mathcal{L}}^* : \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \to \left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right)^*$ is an isomor-

phism and the adjoint operator of $\boldsymbol{\mathcal{L}}$, i.e.

$$(\boldsymbol{\mathcal{L}}^*\phi)(\psi) = (\boldsymbol{\mathcal{L}}\psi)(\phi), \tag{3.4}$$

giving the duality pairing

$$\langle v, u\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}} = (\boldsymbol{\mathcal{L}}^*v)(u) = (\boldsymbol{\mathcal{L}}u)(v) = \langle u, v\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\times\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}}, \tag{3.5}$$

for $u \in \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}$, $v \in \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}$.

With these realisations, the variational problem of (3.2) is equivalent to the variational problem: *find* $\mathbf{u} \in \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$ *such that*

$$\langle \boldsymbol{\mathcal{A}}\mathbf{u}, \mathbf{q}\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}} = \langle \mathbf{f}, \mathbf{q}\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}}, \ \textit{for all } \mathbf{q} \in \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}. \tag{3.6}$$

The 'weak' form of the problem is given by: *find* $\mathbf{u} \in \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$ *such that*

$$\tilde{\boldsymbol{\mathcal{A}}}\mathbf{u} = \tilde{\mathbf{f}}, \tag{3.7}$$

where $\tilde{\boldsymbol{\mathcal{A}}} := \boldsymbol{\mathcal{L}}^*\boldsymbol{\mathcal{A}} : \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dom}} \to \left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right)^*$ and $\tilde{\mathbf{f}} := \boldsymbol{\mathcal{L}}^*\mathbf{f} \in \left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right)^*$. The 'strong' form of the variational problem is: *find* $\mathbf{u} \in \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$ *such that*

$$(\boldsymbol{\mathcal{L}}^*)^{-1}\tilde{\boldsymbol{\mathcal{A}}}\mathbf{u} = (\boldsymbol{\mathcal{L}}^*)^{-1}\tilde{\mathbf{f}}. \tag{3.8}$$

Note that $(\boldsymbol{\mathcal{L}}^*)^{-1}$ exists by construction of the realisation of $\left(\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right)^*$.

3.1.2 Discrete Setting

In this section we consider a special case of what was presented in Section 3.1.1, by introducing the notation for finite dimensional spaces. We aim to connect the two sections later in Section 3.1.3. Let $\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$, $\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}$ and $\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}$ be finite dimensional spaces with

$$N := \dim\left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}}\right) = \dim\left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\right) = \dim\left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right),\ N \in \mathbb{N}. \tag{3.9}$$

Let $\boldsymbol{\mathcal{A}}_h : \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}} \to \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}$ and $\mathbf{f} \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}$. We now consider the problem of the previous subsection in a discrete setting: *find* $\mathbf{u}_h \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$ *such that*

$$\boldsymbol{\mathcal{A}}_h \mathbf{u}_h = \mathbf{f}_h. \tag{3.10}$$

Let $\left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}, \boldsymbol{\mathcal{L}}_h\right)$ be a realisation of $\left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\right)^*$, with $\boldsymbol{\mathcal{L}}_h : \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \to \left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\right)^*$ being an isomorphism, and let $\left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}, \boldsymbol{\mathcal{L}}_h^*\right)$ be a realisation of $\left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right)^*$, with $\boldsymbol{\mathcal{L}}_h^* : \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \to \left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right)^*$ being an isomorphism and the adjoint operator of $\boldsymbol{\mathcal{L}}_h$. Then the problem of (3.10) is equivalent to the variational problem: *find* $\mathbf{u}_h \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$ *such that*

$$\langle \boldsymbol{\mathcal{A}}_h \mathbf{u}_h, \mathbf{q}_h \rangle_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \times \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}} = \langle \mathbf{f}_h, \mathbf{q}_h \rangle_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \times \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}},\ \textit{for all}\ \mathbf{q}_h \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}. \tag{3.11}$$

The 'weak' form of the problem is given by: *find* $\mathbf{u}_h \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$ *such that*

$$\tilde{\boldsymbol{\mathcal{A}}}_h \mathbf{u}_h = \tilde{\mathbf{f}}_h, \tag{3.12}$$

where $\tilde{\boldsymbol{\mathcal{A}}}_h := \boldsymbol{\mathcal{L}}_h^* \boldsymbol{\mathcal{A}}_h : \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \to \left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right)^*$ and $\tilde{\mathbf{f}}_h := \boldsymbol{\mathcal{L}}_h^* \mathbf{f}_h \in \left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right)^*$. The 'strong' form of the variational problem corresponds to: *find* $\mathbf{u}_h \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$ *such that*

$$\left(\boldsymbol{\mathcal{L}}_h^*\right)^{-1} \tilde{\boldsymbol{\mathcal{A}}}_h \mathbf{u}_h = \left(\boldsymbol{\mathcal{L}}_h^*\right)^{-1} \tilde{\mathbf{f}}_h. \tag{3.13}$$

By choosing the bases

$$\left\{\phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dom}}\right\}_{i=1}^N \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}}, \quad \left\{\phi_i^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\right\}_{i=1}^N \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}, \quad \left\{\phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right\}_{i=1}^N \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}},$$

we can express the matrix of the weak form operator $\tilde{\boldsymbol{\mathcal{A}}}$ by

$$\begin{aligned}
\mathbf{A}_{ij} &= \left(\tilde{\boldsymbol{\mathcal{A}}}_h \phi_j^{\boldsymbol{\mathcal{A}},\mathrm{dom}}\right)(\phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}}) = \left\langle \tilde{\boldsymbol{\mathcal{A}}}_h \phi_j^{\boldsymbol{\mathcal{A}},\mathrm{dom}}, \phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \right\rangle_{\left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right)^* \times \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}} \\
&= \left\langle (\boldsymbol{\mathcal{L}}_h^*)^{-1} \tilde{\boldsymbol{\mathcal{A}}}_h \phi_j^{\boldsymbol{\mathcal{A}},\mathrm{dom}}, \phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \right\rangle_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \times \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}}, \\
&= \left\langle \boldsymbol{\mathcal{A}}_h \phi_j^{\boldsymbol{\mathcal{A}},\mathrm{dom}}, \phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \right\rangle_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \times \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}},
\end{aligned} \tag{3.14}$$

and the matrix form of the isomorphism $\boldsymbol{\mathcal{L}}_h^*$ is given by

$$\begin{aligned}
(\mathbf{M}_{\mathbf{A}})_{ij} &= \left(\boldsymbol{\mathcal{L}}_h^* \phi_j^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\right)(\phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}}) = \left\langle \boldsymbol{\mathcal{L}}_h^* \phi_j^{\boldsymbol{\mathcal{A}},\mathrm{ran}}, \phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \right\rangle_{\left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right)^* \times \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}} \\
&= \left\langle \phi_j^{\boldsymbol{\mathcal{A}},\mathrm{ran}}, \phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \right\rangle_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \times \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}}.
\end{aligned} \tag{3.15}$$

The matrix $\mathbf{M}_{\mathbf{A}}$ is often referred to as the **mass matrix** and is invertible by the assumption that $\boldsymbol{\mathcal{L}}_h^*$ is an isomorphism. Following the above matrix forms, the weak and strong linear systems that correspond to (3.10) are

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{3.16}$$

$$\mathbf{M}_{\mathbf{A}}^{-1}\mathbf{A}\mathbf{x} = \mathbf{M}_{\mathbf{A}}^{-1}\mathbf{b}, \tag{3.17}$$

where $\mathbf{x}$ are the coefficients of the solution in $\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$ and

$$\begin{aligned}
\mathbf{b}_i &= (\tilde{\mathbf{f}})\left(\phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right) = \left\langle \tilde{\mathbf{f}}_h, \phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \right\rangle_{\left(\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}\right)^* \times \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}} \\
&= \left\langle (\boldsymbol{\mathcal{L}}_h^*)^{-1} \tilde{\mathbf{f}}_h, \phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \right\rangle_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \times \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}}, \\
&= \left\langle \mathbf{f}_h, \phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \right\rangle_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \times \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}},
\end{aligned} \tag{3.18}$$

3.1.3 From the continuous to the discrete setting

We now aim to connect the previous two subsections by discussing the transition from the continuous to the discrete setting. We consider the case where

$$\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}} \subset \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dom}}, \quad \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \subset \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}, \quad \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \subset \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}. \tag{3.19}$$

Take $\boldsymbol{\mathcal{L}}_h$ to be the restriction of $\boldsymbol{\mathcal{L}}$ to $\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \subset \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}$ so that $\langle\cdot,\cdot\rangle_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}}$ is the restriction of $\langle\cdot,\cdot\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}}$ to $\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \times \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}$. We define the projection operator $\boldsymbol{\Pi}_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}} : \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}} \to \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}$ to be the solution of the variational problem: *find* $\boldsymbol{\Pi}_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}}\mathbf{w} \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}$ *so that*

$$\left\langle \boldsymbol{\Pi}_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}}\mathbf{w}, \mathbf{q}_h \right\rangle_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}} = \langle \mathbf{w}, \mathbf{q}_h\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}}, \ \textit{for all } \mathbf{q}_h \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}. \tag{3.20}$$

The above is well-posed by the assumption that $\langle\cdot,\cdot\rangle_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}}$ is a stable dual pairing. Then, by taking

$$\boldsymbol{\mathcal{A}}_h := \boldsymbol{\Pi}_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}}\boldsymbol{\mathcal{A}}, \qquad \mathbf{f}_h := \boldsymbol{\Pi}_{\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}}\mathbf{f}, \tag{3.21}$$

the discrete variational form of (3.10) can be written as: *find* $\mathbf{u}_h \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}}$ *such that*

$$\langle \boldsymbol{\mathcal{A}}_h\mathbf{u}_h, \mathbf{q}_h\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}} = \langle \mathbf{f}_h, \mathbf{q}_h\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}}, \ \textit{for all } \mathbf{q}_h \in \mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}, \tag{3.22}$$

and the associated discrete matrices and vectors as

$$\mathbf{A}_{ij} = \left\langle \boldsymbol{\mathcal{A}}\phi_j^{\boldsymbol{\mathcal{A}},\mathrm{dom}}, \phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \right\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}}, \tag{3.23}$$

$$(\mathbf{M}_{\mathbf{A}})_{ij} = \left\langle \phi_j^{\boldsymbol{\mathcal{A}},\mathrm{ran}}, \phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \right\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}}, \tag{3.24}$$

$$\mathbf{b}_i = \left\langle \mathbf{f}, \phi_i^{\boldsymbol{\mathcal{A}},\mathrm{dual}} \right\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}\times\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}}}. \tag{3.25}$$

3.1.4 Operator Products

The strong form allows us to correctly discretise operator products which are essential if we intend to use any form of operator preconditioning. Calderón (operator) preconditioning is discussed in detail in Chapter 4, but for now we consider some abstract operator preconditioner $\boldsymbol{\mathcal{P}} : \mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{dom}} \to \mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{ran}}$ mapping between Hilbert spaces with $\mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{dual}}$ the realisation of the dual of $\mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{ran}}$ in the same sense as was discussed in the earlier subsections. Imposing the restriction $\mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{dom}} = \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}}$ and applying $\boldsymbol{\mathcal{P}}$ to (3.1) gives

$$\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}\mathbf{u} = \boldsymbol{\mathcal{P}}\mathbf{f}. \tag{3.26}$$

Following the same variational approach, with

$$\mathbf{H}_h^{\boldsymbol{\mathcal{P}},\mathrm{dom}} \subset \mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{dom}}, \quad \mathbf{H}_h^{\boldsymbol{\mathcal{P}},\mathrm{ran}} \subset \mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{ran}}, \quad \mathbf{H}_h^{\boldsymbol{\mathcal{P}},\mathrm{dual}} \subset \mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{dual}} \tag{3.27}$$

and basis functions

$$\left\{\phi_i^{\boldsymbol{\mathcal{P}},\mathrm{dom}}\right\}_{i=1}^{N} \in \mathbf{H}_h^{\boldsymbol{\mathcal{P}},\mathrm{dom}}, \quad \left\{\phi_i^{\boldsymbol{\mathcal{P}},\mathrm{ran}}\right\}_{i=1}^{N} \in \mathbf{H}_h^{\boldsymbol{\mathcal{P}},\mathrm{ran}}, \quad \left\{\phi_i^{\boldsymbol{\mathcal{P}},\mathrm{dual}}\right\}_{i=1}^{N} \in \mathbf{H}_h^{\boldsymbol{\mathcal{P}},\mathrm{dual}}, \tag{3.28}$$

one can obtain the weak and strong discrete systems of the preconditioned formulation

$$\mathbf{P}\mathbf{M}_{\mathbf{A}}^{-1}\mathbf{A}\mathbf{x} = \mathbf{P}\mathbf{M}_{\mathbf{A}}^{-1}\mathbf{b}, \tag{3.29}$$

$$\mathbf{M}_{\mathbf{P}}^{-1}\mathbf{P}\mathbf{M}_{\mathbf{A}}^{-1}\mathbf{A}\mathbf{x} = \mathbf{M}_{\mathbf{P}}^{-1}\mathbf{P}\mathbf{M}_{\mathbf{A}}^{-1}\mathbf{b}, \tag{3.30}$$

with

$$\mathbf{P}_{ij} = \left\langle \boldsymbol{\mathcal{P}}\phi_j^{\boldsymbol{\mathcal{P}},\mathrm{dom}}, \phi_i^{\boldsymbol{\mathcal{P}},\mathrm{dual}} \right\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{ran}} \times \mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{dual}}}, \tag{3.31}$$

$$(\mathbf{M}_{\mathbf{P}})_{ij} = \left\langle \phi_j^{\boldsymbol{\mathcal{P}},\mathrm{ran}}, \phi_i^{\boldsymbol{\mathcal{P}},\mathrm{dual}} \right\rangle_{\mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{ran}} \times \mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{dual}}}. \tag{3.32}$$

3.2 Variational forms: the PMCHWT formulation

We are now going to apply the theory from Section 3.1 to the case of the PMCHWT formulation (2.87) where

$$\boldsymbol{\mathcal{A}} : \bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2 \to \bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2.$$

We recall that $\mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)$ is self-dual with respect to the anti-symmetric pairing of (2.23), and therefore $\bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2$ is self-dual with respect to [74]

$$\left\langle \bigoplus_{m=1}^{M} \begin{pmatrix} \mathbf{a}_m \\ \mathbf{b}_m \end{pmatrix}, \bigoplus_{m=1}^{M} \begin{pmatrix} \mathbf{c}_m \\ \mathbf{d}_m \end{pmatrix} \right\rangle_{\tau \times \tau} = \sum_{m=1}^{M} \langle \mathbf{a}_m, \mathbf{d}_m \rangle_{\tau, \Gamma_m} + \langle \mathbf{b}_m, \mathbf{c}_m \rangle_{\tau, \Gamma_m}. \quad (3.33)$$

The above self-duality statements are to be interpreted as realisations of dual spaces as discussed in Remark 1 in Chapter 2. Following the notation of the previous section,

$$\mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dom}} = \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{ran}} = \mathbf{H}^{\boldsymbol{\mathcal{A}},\mathrm{dual}} = \bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2,$$

and if we consider some abstract operator preconditioner

$$\boldsymbol{\mathcal{P}} : \bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2 \to \bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2,$$

then

$$\mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{dom}} = \mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{ran}} = \mathbf{H}^{\boldsymbol{\mathcal{P}},\mathrm{dual}} = \bigoplus_{m=1}^{M} \mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m}, \Gamma_m)^2.$$

3.2.1 Discrete Spaces

We now discuss appropriate discrete spaces, which we later use for the discretisation of the operator products. Let Γ_m^h be a triangulation of Γ_m with $N_m \in \mathbb{N}$ piecewise flat triangular elements of diameter $\leq h$, with

$$\Gamma_m^h := \bigcup_{j=1}^{N_m} \overline{\Gamma_m^j}, \tag{3.34}$$

for the faces (triangles) $\Gamma_m^1, \ldots, \Gamma_m^{N_m}$, for $m = 1, \ldots, M$. We refer to Γ_m^h as the *primal mesh*.

Let $\Gamma_m^{h'}$ be a barycentric refinement of the primal mesh consisting of $6N_m$ elements with diameter $\leq h'$. To barycentrically refine a grid we take the barycentre of each triangle in the primal mesh and connect it with the three original vertices and the three midpoints of its edges, creating 6 new smaller triangles within each of the original one. We refer to $\Gamma_m^{h'}$ as the *barycentric mesh*.

Let Γ_m^d be the *dual* mesh defined on the barycentrically refined grid. The elements of the dual mesh are polygons, consisting of all triangles in the barycentrically refined mesh attached to a vertex of the primal mesh. An example of a primal mesh, its barycentric refinement and the corresponding dual mesh can be seen in Figure 3.1.

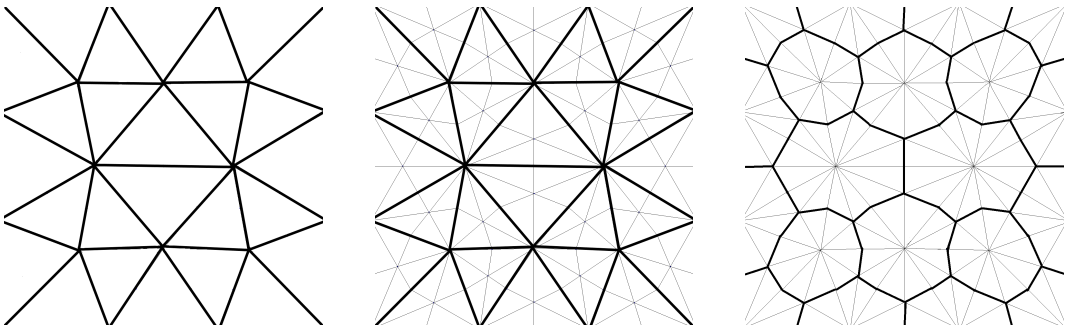


Figure 3.1: Primal mesh (left), barycentric refinement of the primal mesh (centre), and dual mesh (right).

3.2.1.1 Raviart-Thomas and Rao-Wilton-Glisson basis functions

Commonly used finite spaces for $\mathbf{H}_{\times}^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m},\Gamma_m)$ are the Raviart-Thomas (**RT**) [85] and Rao-Wilton-Glisson (**RWG**) basis functions [86]. For the jth edge in the mesh Γ_m^h, between two triangles $\Gamma_m^{j,+}$ and $\Gamma_m^{j,-}$, the 0th order **RT** basis functions are defined by

$$\mathbf{RT}_m^j(\mathbf{r}) := \begin{cases} \dfrac{1}{2A_m^{j,+}}(\mathbf{r}-\mathbf{p}_m^{j,+}), & \mathbf{r}\in\Gamma_m^{j,+}, \\ \dfrac{1}{2A_m^{j,-}}(\mathbf{r}-\mathbf{p}_m^{j,-}), & \mathbf{r}\in\Gamma_m^{j,-}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.35}$$

Here $A_m^{j,+}$ and $A_m^{j,-}$ are the areas of triangles $\Gamma_m^{j,+}$ and $\Gamma_m^{j,-}$ respectively, and $\mathbf{p}_m^{j,+}$, $\mathbf{p}_m^{j,-}$ are vertices on $\Gamma_m^{j,+}$ and $\Gamma_m^{j,-}$ not on the shared edge. A schematic representation is shown in Figure 3.2.

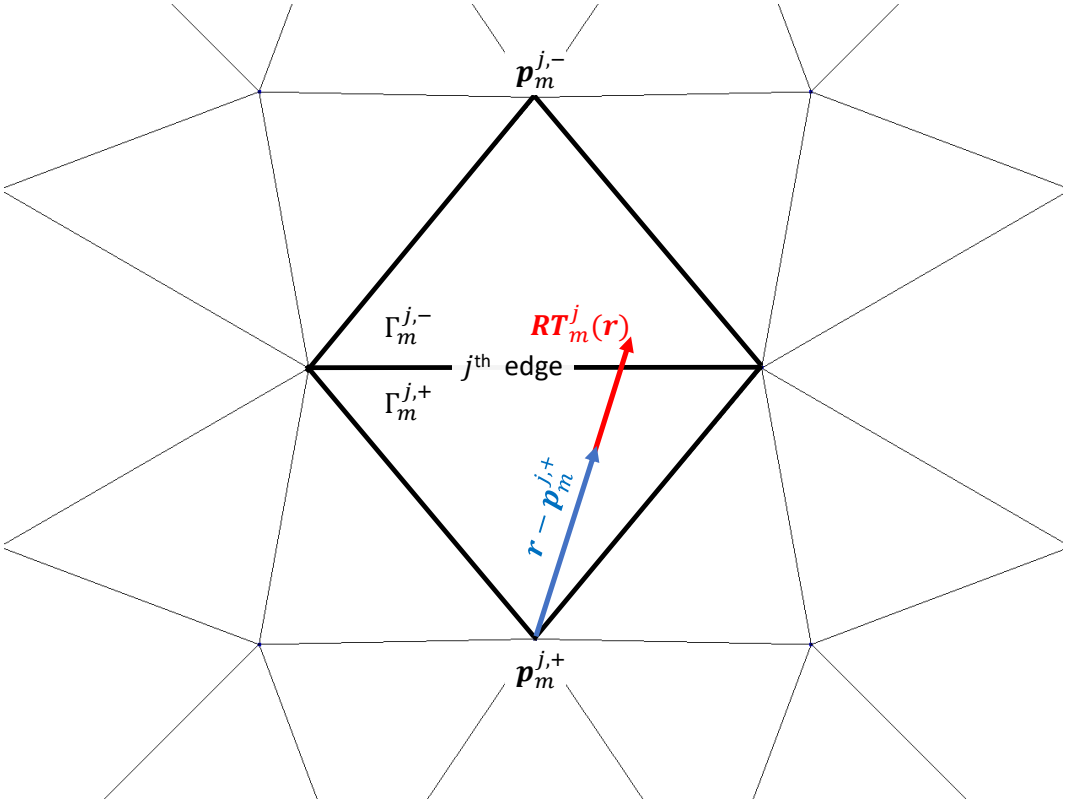


Figure 3.2: The 0th order Raviart-Thomas $\mathbf{RT}_m^j$ basis functions defined on a pair of triangles $\Gamma_m^{j,+}$ and $\Gamma_m^{j,-}$, on the primal mesh Γ_m^h.

On the same mesh Γ_m^h, edge and pair of triangles, the **RWG** basis functions are defined by

$$\mathbf{RWG}_m^j(\mathbf{r}) := l_m^j \mathbf{RT}_m^j(\mathbf{r}), \tag{3.36}$$

where l_m^j is the length of the common edge between $\Gamma_m^{j,+}$ and $\Gamma_m^{j,-}$.

Let $\mathbf{RT}_m$ and $\mathbf{RWG}_m$ be the spaces spanned by $\mathbf{RT}_m^j$ and $\mathbf{RWG}_m^j$, $j = 1, \ldots, N$ respectively. The space of div-conforming basis functions $\mathbf{RT}_m$ (or $\mathbf{RWG}_m$) contains a subspace that is completely orthogonal to itself with respect to the anti-symmetric dual pairing (2.23) [87]. Choosing $\phi_{m,j}^{\boldsymbol{\mathcal{A}},\mathrm{dom}} = \phi_{m,j}^{\boldsymbol{\mathcal{A}},\mathrm{ran}} = \phi_{m,j}^{\boldsymbol{\mathcal{A}},\mathrm{dual}} = \mathbf{RT}_m^j$ (or $\mathbf{RWG}_m^j$) would violate the inf-sup stability conditions making the mass matrix $\mathbf{M_A}$ non-invertible (and similarly for $\mathbf{M_P}$ if the same argument is applied to basis functions of the preconditioner). For this reason, a second set of basis functions defined on the dual mesh is used. A common choice is the Buffa-Christiansen basis functions [88] and these are discussed in the following subsection.

3.2.1.2 Buffa-Christiansen basis functions

Buffa-Christiansen (**BC**) functions [88] are defined on the dual mesh Γ_m^d, based on polygonal pairs, as shown in Figure 3.3. A polygonal pair $P_m^{j,\pm}$ is formed by considering the jth edge of a triangle pair in the primal mesh and taking all the smaller triangles of the barycentrically refined mesh that have as a node either end of the common edge of the triangle pair. The polygon surrounding the starting point of the jth edge is $P_m^{j,+}$ and the polygon surrounding the endpoint is $P_m^{j,-}$. We denote the total number of triangles within polygon $P_m^{j,\pm}$ as $N_{P_m^{j,\pm}} \in \mathbb{N}$. We denote the common edges of the triangles within the polygon pair by $e_m^{ji,\pm}$, $i = 0, \ldots, N_{P_m^{j,\pm}} - 1$, as shown in Figure 3.3.

The Buffa Christiansen basis functions associated with the jth edge $\mathbf{BC}_m^j$ are defined as a combination of **RWG** functions on the barycentrically refined

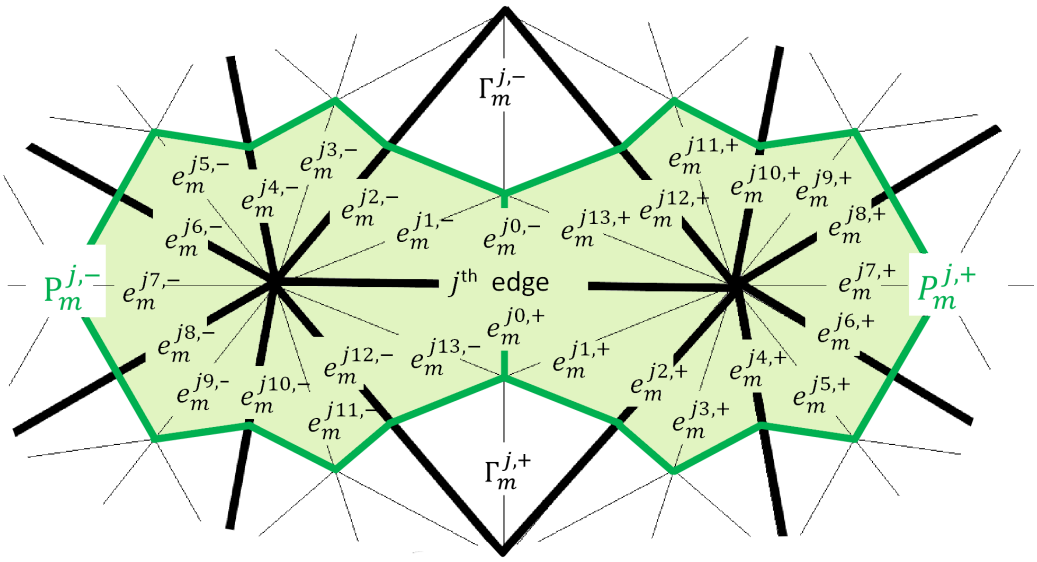


Figure 3.3: A polygonal pair $\mathbf{P}_m^{j,\pm}$ on the barycentrically refined grid $\Gamma_m^{h'}$. The original triangles are marked with bold black edge colour.

mesh [87]

$$\mathbf{BC}_m^j(\mathbf{r}) = \sum_{i=1}^{2N_{P_m^{j,+}}-1} \frac{N_{P_m^{j,+}} - i}{2l_m^{ji,+} N_{P_m^{j,+}}} \mathbf{RWG}_m^{ji,+} - \sum_{i=1}^{2N_{P_m^{j,-}}-1} \frac{N_{P_m^{j,-}} - i}{2l_m^{ji,-} N_{P_m^{j,-}}} \mathbf{RWG}_m^{ji,-} + \frac{1}{2l_m^{j0,+}} \mathbf{RWG}_m^{j0,+} - \frac{1}{2l_m^{j0,-}} \mathbf{RWG}_m^{j0,-}, \tag{3.37}$$

where $l_m^{ji,\pm}$ is the length of edge $e_m^{ji,\pm}$ and $\mathbf{RWG}_m^{ji,\pm}$ is the **RWG** function on the triangle pair associated with edge $e_m^{ji,\pm}$ in polygon $P_m^{j,\pm}$.

Let $\mathbf{BC}_m$ be the space spanned by $\mathbf{BC}_m^j$. Choosing $\phi_{m,j}^{\boldsymbol{\mathcal{A}},\text{ran}} = \mathbf{RWG}_m^j$ and $\phi_{m,j}^{\boldsymbol{\mathcal{A}},\text{dual}} = \mathbf{BC}_m^j$ (and vice versa) gives a well conditioned mass matrix $\mathbf{M_A}$ because inf-sup stability conditions between the two finite spaces are now satisfied. The inf-sup stability condition is as follows and is proved in [88]:

There exists $C > 0$ such that for all h:

$$\inf_{u\in\mathbf{RWG}_m}\sup_{v\in\mathbf{BC}_m}\frac{\langle u,v\rangle_{\tau,\Gamma_m}}{\|u\|_{\mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m},\Gamma_m)}\|v\|_{\mathbf{H}_\times^{-\frac{1}{2}}(\mathrm{div}_{\Gamma_m},\Gamma_m)}}\geq 1/C. \tag{3.38}$$

3.2.2 Stable discretisation of the Calderón product

There are different ways of using $\mathbf{RWG}_m$ and $\mathbf{BC}_m$ functions to take advantage of the inf-sup stability between their pairings. One choice, introduced in [36] and used in our work in [3] is

$$\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}}=\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}=\mathbf{H}_h^{\boldsymbol{\mathcal{P}},\mathrm{ran}}=\bigoplus_{m=1}^{M}\begin{pmatrix}\mathbf{RWG}_m\\ \mathbf{RWG}_m\end{pmatrix}, \tag{3.39}$$

$$\mathbf{H}_h^{\boldsymbol{\mathcal{P}},\mathrm{dom}}=\mathbf{H}_h^{\boldsymbol{\mathcal{P}},\mathrm{dual}}=\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}=\bigoplus_{m=1}^{M}\begin{pmatrix}\mathbf{BC}_m\\ \mathbf{BC}_m\end{pmatrix}, \tag{3.40}$$

while another choice, introduced in [75] and used in [2] is

$$\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dom}}=\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{dual}}=\mathbf{H}_h^{\boldsymbol{\mathcal{A}},\mathrm{ran}}=\mathbf{H}_h^{\boldsymbol{\mathcal{P}},\mathrm{dom}}=\mathbf{H}_h^{\boldsymbol{\mathcal{P}},\mathrm{dual}}=\mathbf{H}_h^{\boldsymbol{\mathcal{P}},\mathrm{ran}}=\bigoplus_{m=1}^{M}\begin{pmatrix}\mathbf{RWG}_m\\ \mathbf{BC}_m\end{pmatrix}. \tag{3.41}$$

Both (3.39)–(3.40) and (3.41) possess symmetries that allow the overall assembly time and memory to be reduced. We have not yet discussed the choice of preconditioner $\boldsymbol{\mathcal{P}}$ (this is discussed in detail in Chapter 4), but Calderón preconditioning consists of some (or all) of the operators used in $\boldsymbol{\mathcal{A}}$. Using discretisation (3.41), where the same discrete spaces are used for both $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{A}}$, allows us to obtain $\mathbf{P}$ for 'free' once $\mathbf{A}$ is assembled, as one can extract the relevant operators from $\mathbf{A}$ to create $\mathbf{P}$. This is not possible with the current implementation using (3.39)–(3.40), since different spaces are used for $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{A}}$ [1]. However, recalling the definition of the operators $\boldsymbol{\mathcal{A}}_m^i$, $\boldsymbol{\mathcal{A}}_m^e$ and $\boldsymbol{\mathcal{A}}_{m\ell}$

[1] We note that one could assemble all operators on the barycentrically refined **RWG** spaces and obtain the standard **RWG** and **BC** through projections. More details on the assembly on the barycentrically refined grid are given later in Section 3.3.4.

from (2.78)–(2.79), we see that each block of 4 operators consists of only two distinct versions of $\mathcal{C}$ and $\mathcal{S}$. This allows us to assemble each of them only once, cutting the assembly cost (time and memory) in half, compared to (3.41) where this is not possible. The same holds for $\boldsymbol{\mathcal{P}}$, with the explicit forms defined in the next chapter. Assembling an operator with $\mathbf{BC}_m$ functions is significantly more expensive compared to discretising with $\mathbf{RWG}_m$, because of the 6-fold increase in the number of elements on the barycentrically refined mesh. Using (3.39)–(3.40) places the burden of this cost on the preconditioner. This cost can then be mitigated by using acceleration techniques, such as

1. reducing the number of operators used in the preconditioner $\boldsymbol{\mathcal{P}}$, which we discuss in Chapter 4,

2. considering a bi-parametric implementation where the assembly routine for $\mathbf{P}$ is cheaper (but of poorer quality) than that of $\mathbf{A}$, as we discuss later in Chapter 5,

or a combination of both (discussed in Chapters 5 and 6).

3.3 Matrix Assembly

A dense assembly of the operators would lead to an $\mathcal{O}(N^2)$ computational cost for N degrees of freedom. In addition, in order to capture the wave solution one needs to refine the mesh size with respect to the wavelength $\lambda = 2\pi/k$, usually at 10 elements per wavelength [5], with the cost therefore scaling as $\mathcal{O}(k^4)$. A dense assembly is therefore impractical for large problems and faster alternatives are needed. Such alternative methods are the $\mathcal{H}$-matrix methods [13, 14] and their variants ($\mathcal{H}^2$ [15]) or the Fast Multipole Methods (FMM) [17, 18, 19]. The complexity using $\mathcal{H}$-matrices reduces to $\mathcal{O}(rN\log N)$ [84], where r is a measure of the local approximation rank required to achieve a prescribed accuracy and to $\mathcal{O}(N\log N)$ for an FMM implementation.

3.3.1 $\mathcal{H}$-matrix assembly

We note that $\mathcal{H}$-matrix compression is a widely used technique and here we focus on its implementation in Bempp [84]. The implementation was not performed by the author of this thesis but we review the basic ideas here so that we can explain our bi-parametric approach later in Chapter 5. In particular, the main ideas required for the bi-parametric approach are in equations (3.45) and (3.46) and Section 3.3.2, but we include more details for completeness. The concepts required for the bi-parametric implementation can be easily translated to FMM or other $\mathcal{H}$-matrix implementations.

We define the index sets

$$\mathcal{I} := \mathcal{I}_1^{(0)} = \{0, 1, \ldots, N-1\},$$
$$\mathcal{J} := \mathcal{J}_1^{(0)} = \{0, 1, \ldots, N-1\},$$

to be the set of degrees of freedom (dofs) in the test and trial spaces respectively. The $\mathcal{H}$-matrix compression is based on a repeated subdivision of those sets into $\mathcal{T}(\mathcal{I})$ and $\mathcal{T}(\mathcal{J})$, referred to as cluster trees. The roots of the trees $\mathcal{T}(\mathcal{I})$ and $\mathcal{T}(\mathcal{J})$, are the index sets $\mathcal{I}$ and $\mathcal{J}$ respectively. On each level, the index sets are divided into two disjoint sets, e.g. $\mathcal{I}_1^{(1)} = \{0, 1, \ldots, \frac{N}{2} - 1\}$ and $\mathcal{I}_2^{(1)} = \{\frac{N}{2}, \ldots, N-1\}$, called nodes, and $\mathcal{J}_1^{(1)} = \{0, 1, \ldots, \frac{N}{2} - 1\}$ and $\mathcal{J}_2^{(1)} = \{\frac{N}{2}, \ldots, N-1\}$ respectively. Each subgroup is then subdivided into two subsubgroups and so on. A schematic representation of the cluster tree $\mathcal{T}(\mathcal{I})$ of depth 3 is shown in Figure 3.4. When the number of dofs in a node is below a specified tolerance n_{min}, the subdivision stops and that node is called a leaf.

Once the cluster trees $\mathcal{T}(\mathcal{I})$ and $\mathcal{T}(\mathcal{J})$ are generated, we proceed to generate a block cluster tree $\mathcal{T}(\mathcal{I} \times \mathcal{J})$ as follows. We set the root of $\mathcal{T}(\mathcal{I} \times \mathcal{J})$ as the index set $\mathcal{I} \times \mathcal{J}$. Consider a node $\mathcal{I}_p^{(q)} \times \mathcal{J}_r^{(s)}$ of $\mathcal{T}(\mathcal{I} \times \mathcal{J})$, where $\mathcal{I}_p^{(q)}$ and $\mathcal{J}_r^{(s)}$ are nodes of $\mathcal{T}(\mathcal{I})$ and $\mathcal{T}(\mathcal{J})$ respectively, at level p (r respectively) and index q (s respectively) and geometric bounding boxes $X\left(\mathcal{I}_p^{(q)}\right)$ and $Y\left(\mathcal{J}_r^{(s)}\right)$

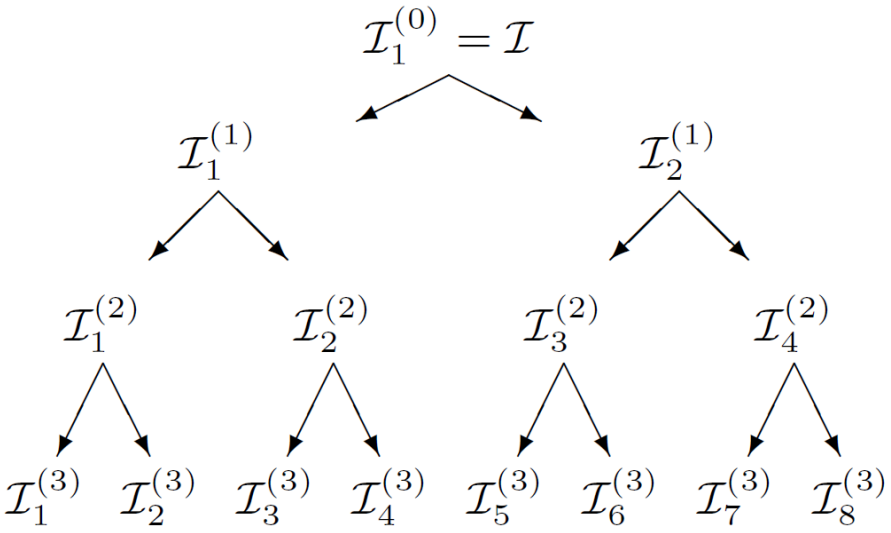


Figure 3.4: Example of the division of dofs of a cluster tree $\mathcal{T}(\mathcal{I})$ with depth 3.

associated with the cluster nodes $\mathcal{I}_p^{(q)}$ and $\mathcal{J}_r^{(s)}$. The pair of nodes $\mathcal{I}_p^{(q)}$ and $\mathcal{J}_r^{(s)}$ is deemed admissible if [2]

$$\operatorname{dist}\left(X\left(\mathcal{I}_p^{(q)}\right), Y\left(\mathcal{J}_r^{(s)}\right)\right) > 0. \tag{3.42}$$

If the admissibility condition fails, the block is further refined by defining children of the node $\mathcal{I}_p^{(q)} \times \mathcal{J}_r^{(s)}$

$$\begin{aligned}\operatorname{children}\left(\mathcal{I}_p^{(q)} \times \mathcal{J}_r^{(s)}\right) = \Big\{&\mathcal{I}_{p_1}^{(q+1)} \times \mathcal{J}_{r_1}^{(s+1)},\ \mathcal{I}_{p_1}^{(q+1)} \times \mathcal{J}_{r_2}^{(s+1)},\\ &\mathcal{I}_{p_2}^{(q+1)} \times \mathcal{J}_{r_1}^{(s+1)},\ \mathcal{I}_{p_2}^{(q+1)} \times \mathcal{J}_{r_2}^{(s+1)}\Big\},\end{aligned} \tag{3.43}$$

where $\mathcal{I}_{p_1}^{(q+1)}$, $\mathcal{I}_{p_2}^{(q+1)}$ are the children of $\mathcal{I}_p^{(q+1)}$ generated by the subdivision

[2] A frequently used condition for admissibility is [84]

$$\min\left\{\operatorname{diam}\left(X\left(\mathcal{I}_p^{(q)}\right)\right), \operatorname{diam}\left(Y\left(\mathcal{J}_r^{(s)}\right)\right)\right\} \leq \alpha \operatorname{dist}\left(X\left(\mathcal{I}_p^{(q)}\right), Y\left(\mathcal{J}_r^{(s)}\right)\right),$$

where diam is the diameter of the bounding box and dist is the distance of the two bounding boxes. The parameter α controls how strong the separation between the two boxes needs to be for the admissibility condition to be satisfied. Bempp uses a weaker condition as shown in (3.42). In practice this works well and usually leads to a fewer number of blocks on the block cluster tree [84].

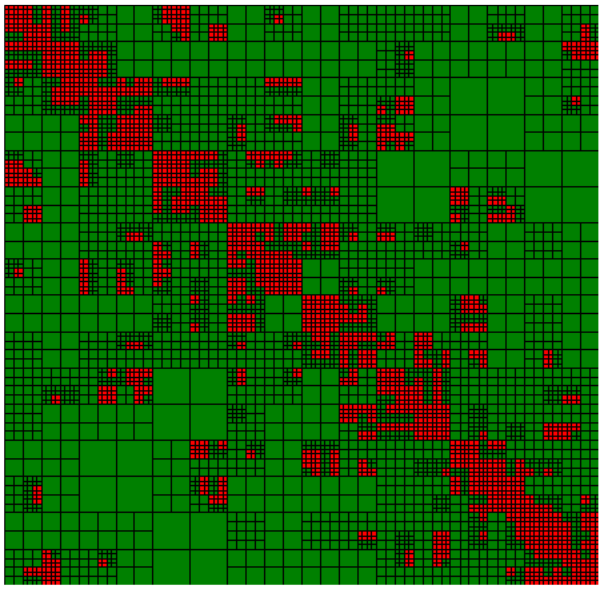


Figure 3.5: An example of a block cluster tree produced by Bempp. The red blocks represent the inadmissible blocks, while the green represent the admissible blocks.

of the index set and $\mathcal{J}_{r_1}^{(s+1)}$, $\mathcal{J}_{r_2}^{(s+1)}$ the children of $\mathcal{J}_r^{(s+1)}$. The sub-blocks generated by the children are then checked for the admissibility condition. The refinement stops until all sub-blocks are deemed admissible or are smaller than some specified minimum block size, in which case they are called inadmissible. An example of a block-cluster tree is given in Figure 3.5.

Once the block cluster tree is generated we can store the Galerkin matrix in $\mathcal{H}$-matrix representation. This consists of storing the sub-matrices corresponding to inadmissible blocks as full matrix (dense) representations and those corresponding to admissible blocks using a low rank approximation, generated by adaptive cross approximation (ACA) [89, 90, 91]. For any admissible block of size $n \times m$ with underlying matrix $\mathbf{B}$ and given a tolerance parameter $\nu > 0$, the ACA algorithm delivers an r-rank approximation

$$\mathbf{B}_\nu = \sum_{i=1}^{r} \mathbf{u}_i \mathbf{v}_i^H, \tag{3.44}$$

such that

$$\|\mathbf{B}_\nu - \mathbf{B}\|_2 \leq \nu \|\mathbf{B}\|_F, \tag{3.45}$$

with $r = r(\nu, \mathbf{B})$. Varying ν controls the quality (accuracy) and cost of the low-rank approximations.

Our $\mathcal{H}$-matrix approximation in Bempp has an additional feature, that saves computational time and memory by assembling only a subset of admissible blocks that represent near-field interactions. Given a user-specified cutoff parameter $\chi \in [0, \infty)$, the admissible blocks that satisfy the near-field condition

$$\operatorname{dist}\left(X\left(\mathcal{I}_p^{(q)}\right), Y\left(\mathcal{J}_r^{(s)}\right)\right) \leq \chi, \tag{3.46}$$

are assembled using ACA, while all other admissible blocks are set to zero. Setting $\chi = \infty$ corresponds to assembling all admissible blocks, as in a standard implementation.

3.3.2 Quadrature

Assembly of both the admissible and inadmissible blocks for $\mathbf{A}$ and $\mathbf{P}$ requires the evaluation of the Galerkin integrals (3.23) and (3.31) respectively. Each of these can be written in terms of integrals of the form

$$\int_{T_1} \int_{T_2} F(\mathbf{x}, \mathbf{y}) \mathrm{d}s(\mathbf{y}) \mathrm{d}s(\mathbf{x}), \tag{3.47}$$

where T_1 and T_2 are triangles in either the primal mesh or the barycentric mesh, and the integrand $F(\mathbf{x}, \mathbf{y})$ involves the fundamental solution G (defined in (2.36)) and a pair of discrete basis functions. When the closures of T_1 and T_2 are disjoint, $F(\mathbf{x}, \mathbf{y})$ is smooth and a standard tensor product Gauss rule based on symmetric Gauss points over triangles is used. The number of quadrature points used depends on whether the integral is classified by Bempp as either near-, medium- or far field (based on the distance between the triangles and their sizes), with corresponding quadrature order parameters q_{near}, q_{medium}

and q_{far} for each case. When T_1 and T_2 share a vertex, edge or are the same triangle, $F(\mathbf{x},\mathbf{y})$ is singular and quadrature routines of the type described in [83] are applied, with a single quadrature order parameter $q_{singular}$ for all singular cases. We use

$$\mathbf{q} = (q_{near}, q_{medium}, q_{far}, q_{singular}) \tag{3.48}$$

to refer to all quadrature order parameters defined in Bempp.

The cost of the assembly of boundary element matrices depends on the number of kernel evaluations for each triangle pair T_1 and T_2. For non-singular integral the default values $(q_{near}, q_{medium}, q_{far}) = (4, 3, 2)$ correspond to $(36, 16, 9)$ integrand evaluations respectively. For singular integrals the default value $q_{singular} = 6$ corresponds to 512 integrand evaluations if T_1 and T_2 share a common vertex, 1280 if T_1 and T_2 share a single common edge, and 1536 if T_1 and T_2 coincide.

3.3.3 Assembly of mass matrices

Computations of the inverse mass matrices $\mathbf{M}_\mathbf{A}^{-1}$ and $\mathbf{M}_\mathbf{P}^{-1}$ are fast compared to the other operators. In particular, the mass-matrices $\mathbf{M_A}$ and $\mathbf{M_P}$ are block diagonal with M sparse blocks. To compute the inverse mass matrices, $2M$ independent sparse LU decompositions are required in total [3]. In the case of 3D scattering, such as the problems considered in this thesis, the mass matrix is formed over a two dimensional manifold in three dimensional space but it is still cheap to assemble compared to the other operators. For more details on the assembly of mass matrices in Bempp we refer to [75].

[3] The discretisation scheme of (3.39)–(3.40) requires M LU decompositions for $\mathbf{M}_\mathbf{A}^{-1}$ and M decompositions for $\mathbf{M}_\mathbf{P}^{-1}$. Using the discretisation scheme of (3.41) gives $\mathbf{M}_\mathbf{A}^{-1} = \mathbf{M}_\mathbf{P}^{-1}$ but still requires $2M$ decompositions in total as each diagonal block now requires 2 LU decompositions. We refer back to Section 3.2.2 for the discussion regarding the two discretisation schemes used.

3.3.4 Assembly on the barycentrically refined grid

Any operators that are defined using the **BC** space (either by using the discretisation of (3.39)–(3.40) or (3.41)), are assembled using **RWG** basis functions on the barycentrically refined grid (via $\mathcal{H}$-matrix assembly as described earlier in Section 3.3.1) and are then projected to the **BC** space through basis transformation matrices. Operators that are defined using **RWG** basis functions are assembled on the coarse mesh without the need for projections. It is possible to disable the default method of projections and assemble the **BC** functions directly on the coarse mesh but this is significantly slower. Given that each diagonal block of the PMCHWT formulation (2.88) consists of 8 operators, it would be very slow to assemble the matrices without projections. Instead, we look at other ways of minimising the memory cost in Chapters 4 and 5. For more details on the implementation in Bempp we refer to [75].

Chapter 4

Accelerated Calderón Preconditioning I: operator-based approach

This chapter discusses one of the author's main contributions of this thesis, with the results presented in this chapter included in [2] and [3]. We recall from Chapter 1 that we are looking for some operator $\boldsymbol{\mathcal{P}}$ applied to the PMCHWT formulation such that $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$ has better properties than $\boldsymbol{\mathcal{A}}$ and therefore its discrete version is easier to solve numerically. The computational cost of solving the preconditioned system is affected by

(i) the choice of preconditioning operator $\boldsymbol{\mathcal{P}}$, and

(ii) the choice of discretisation for the operator product $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$.

In this chapter we focus on (i) by considering a range of different Calderón preconditioners, although the choice of discretisation (ii) is also discussed and justified for our choice of preconditioners.

In Section 4.1, we recall the traditional Calderón preconditioner of the PMCHWT formulation that was presented in [36] for the single-particle set-up ($M = 1$) and then extend it to the multi-particle case ($M > 1$). The spectral properties of the resulting operators are discussed as they provide a heuristic indication of the GMRES performance. Numerical experiments comparing the performance of the traditional Calderón preconditioner are also presented,

demonstrating that for the specific discretisation we use, traditional Calderón preconditioning does not perform any better than simple mass matrix preconditioning when one accounts for the overall matrix-vector products performed.

In an effort to reduce computational cost (including assembly time, memory consumption, number of iterations, matrix-vector products and solve time) we look at minimising the number of operators used in the preconditioner while still maintaining a good preconditioning effect. We present the reduced preconditioners in Section 4.2. This includes a discussion of their spectral properties along with numerical experiments comparing their performance for a range of model problems.

Our numerical experiments show that the reduced preconditioners defined later in (4.27), provide some regularising effect, requiring less memory and assembly time compared to the traditional choice. Some of them incur a longer GMRES time, but due to the reduction in assembly time, the total computational time was shorter. The choice of preconditioner depends on a number of parameters, such as the number of scatterers, the wavenumber, refractive index and how many GMRES solves are to be performed.

The preconditioners $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ (defined later in (4.29)–(4.30)) require only 25% of the memory cost of the traditional preconditioner. For larger wavenumbers, the GMRES solution time is longer but the total computational time is in most cases below 50%. If one GMRES solve is to be performed, then the use of these two preconditioners allows the solution of much higher frequency problems that would otherwise not be possible due to memory and time constraints. If the goal is to simulate random orientation (discussed later in Chapter 7) where multiple GMRES are performed at a time, then our preconditioner of choice would be a preconditioner that does not incur a longer solution time. This can be achieved by preconditioners $\boldsymbol{\mathcal{D}}^i$ and $\boldsymbol{\mathcal{D}}^e$ (defined in (2.89) and later in (4.28) respectively) that achieve a 50% reduction in memory consumption and have a GMRES solution time that is either equal to the traditional preconditioner or up to 50% faster.

4.1 Traditional Calderón preconditioning

The PMCHWT formulation (2.87) is well-posed with a unique solution for every frequency [74, 81] at the continuous setting. Upon discretisation, the resulting linear system of equations is ill-conditioned leading to a large number of iterations when solved with an iterative solver, such as GMRES [92]. To understand the origin of this ill-conditioning we first consider single-particle scattering ($M = 1$).

The PMCHWT operator for $M = 1$ reduces to

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}_1^e + \boldsymbol{\mathcal{A}}_1^i = \begin{bmatrix} \mathcal{C}_1^e + \mathcal{C}_1^i & \frac{\mu_e}{k_e}\mathcal{S}_1^e + \frac{\mu_1}{k_1}\mathcal{S}_1^i \\ -\frac{k_e}{\mu_e}\mathcal{S}_1^e - \frac{k_1}{\mu_1}\mathcal{S}_1^i & \mathcal{C}_1^e + \mathcal{C}_1^i \end{bmatrix}. \tag{4.1}$$

The operators $\mathcal{C}_1^e$ and $\mathcal{C}_1^i$ are compact on smooth domains [79], which implies that their corresponding discrete versions have eigenvalues clustering around zero. The operators $\mathcal{S}_1^e$ and $\mathcal{S}_1^i$ consist of both smooth and hypersingular terms (see (2.57)) that have accumulation points at zero and infinity respectively [36] when discretised. These suggest that the discrete PMCHWT operator will be ill-conditioned. An example of the spectrum of the discrete forms of $\mathcal{C}$ and $\mathcal{S}$ (using the mixed discretisation of (3.41)) is shown in Figure 4.1, which confirms the accumulation points at zero and infinity.

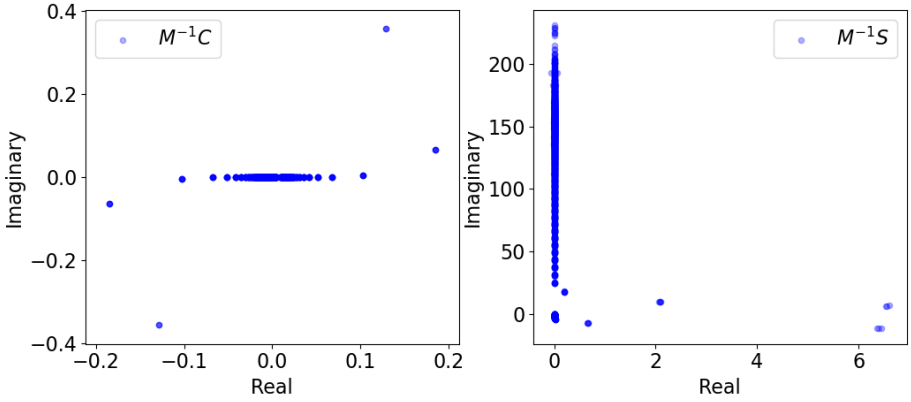


Figure 4.1: Spectrum of the strong discrete operators of $\mathcal{C}$ (left) and $\mathcal{S}$ (right) using the mixed discretisation of (3.41), on the unit sphere with $k = 1.5$ and a mesh width of approximately 0.1 resulting in 3206 elements.

We recall that our goal (with regards to (i)) is to precondition the PM-CHWT operator and minimise the computational costs (assembly time, memory consumption, number of iterations, matrix-vector products and solve time). This can be achieved by obtaining an invertible second kind operator, i.e. of the form of a scaled identity plus a compact operator. An invertible second kind integral operator has a bounded spectrum, i.e. the eigenvalues do not accumulate at infinity, and it is clustered away from zero (accumulating at the scalar which the identity is multiplied with) when discretised. This ensures that the discrete system's condition number, i.e. the ratio of the largest over the smallest singular value is bounded as discretisation becomes finer [36]. To achieve this, one needs to utilise the Calderón identities of (2.68)–(2.69), with the traditional choice in this case being [36]

$$\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}. \tag{4.2}$$

It is well established in literature that the resulting operator $\boldsymbol{\mathcal{A}}^2$ is well-posed [36, 37, 38, 39] (except in the case of very high contrast materials [36]) with accumulation points not clustering around zero or tending to infinity. Niino and Nishimura [38] have remedied the high contrast material problem by pre-multiplying the preconditioner by some scalar operator. For the purposes of this thesis, we do not consider high contrast materials and as such we proceed with the traditional Calderón preconditioner of (4.2). Moving to the multi-particle setup ($M > 1$) the same strategy can be applied by taking $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, where $\boldsymbol{\mathcal{A}}$ is now the full operator of (2.88).

4.1.1 Spectral Properties

For a discrete matrix not far from normality, GMRES converges fast if its eigenvalues are clustered away from the origin [93]. We note that even though our discrete system is non-normal in general, the spectral properties of $\boldsymbol{\mathcal{A}}^2$ are a popular discussion topic when considering Calderón preconditioning as they can give a good heuristic indication of the GMRES convergence, see for

example [36, 38, 94, 95, 39, 96, 97]. In general, the eigenvalue distribution of the discrete system provides an indication of its conditioning and can be used for guidance on whether the methods can perform well in practice.

The resulting operator, $\boldsymbol{\mathcal{A}}^2$, has 2 accumulation points in the case of $M = 1$, at

$$\Lambda_1 = \frac{1}{2} + \frac{1}{4}\frac{\mu_1}{\mu_e} + \frac{1}{4}\frac{\mu_e}{\mu_1}, \tag{4.3}$$

$$\Lambda_2 = \frac{1}{2} + \frac{1}{4}\frac{\mu_1}{\mu_e}\left(\frac{k_e}{k_1}\right)^2 + \frac{1}{4}\frac{\mu_e}{\mu_1}\left(\frac{k_1}{k_e}\right)^2. \tag{4.4}$$

These have been reported in [38] for the case $M = 1$ and were proved using Fourier transforms but one can also use the splitting properties of [94, 95]. We present here derivation of the above accumulation points (4.3)–(4.4) for the special case of a sphere, $M = 1$, as this is not currently available in literature. Derivation of the case $M > 1$ goes beyond the scope of this thesis as it requires expansion of spherical harmonics from different domains. Numerical experiments assuming three spheres, $M = 3$, suggest that the above could be generalised for the multi-particle setting (by adjusting the appropriate wavenumbers and magnetic permeabilities). An example of the spectrum of $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$ together with the theoretical accumulation points is shown in Figure 4.2. We note that results from [38, 94, 95] regarding (4.3)–(4.4) have not been restricted to the unit sphere and should hold for non spherical domains. Numerical experiments, although not present here, have confirmed that the same accumulation points are achieved when the sphere is substituted by a unit cube or a single branch of a rosette aggregate (see Figure 1.2(c)).

Theorem 1. *Consider a sphere of radius 1. Let* $\mathbf{U}_n^p$, $\mathbf{X}_n^p$ *be the vector spherical harmonics of degree* n *and order* p *for the sphere defined as*

$$\mathbf{U}_n^p(\theta,\phi) = \frac{1}{\sqrt{n(n+1)}}\nabla_\Gamma Y_n^p(\theta,\phi), \tag{4.5}$$

$$\mathbf{X}_n^p(\theta,\phi) = \mathbf{n} \times \mathbf{U}_n^p(\theta,\phi), \tag{4.6}$$

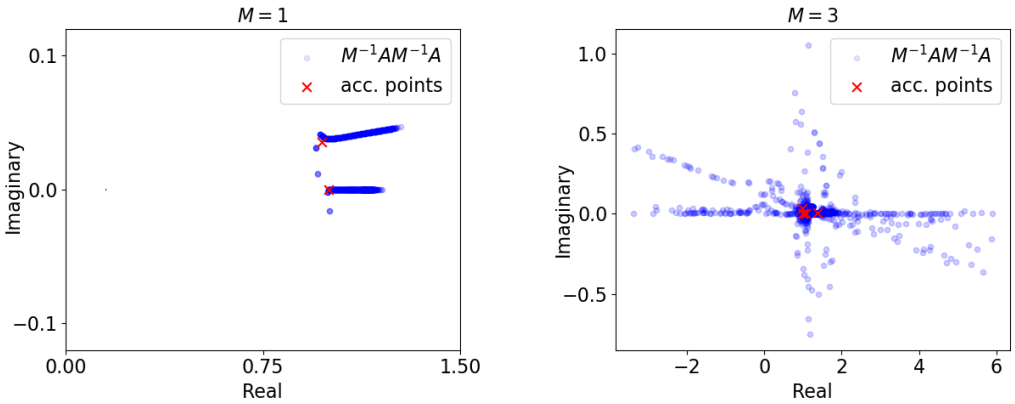


Figure 4.2: Spectrum of the discrete operator $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$ for a single and multi-particle setting. On the left the scatterer is the unit sphere centred at the origin with refractive index $n_1 = 1.0833 + 0.204\mathrm{i}$. On the right the scatterers are three unit spheres centred at $(-1.5,0,0)$, $(0,0,0)$ and $(1.5,0,0)$, with refractive indices $n_1 = 1.311 + 2.289 \times 10^{-9}\mathrm{i}$, $n_2 = 1.7746 + 0.00940\mathrm{i}$ and $n_3 = 1.0833 + 0.204\mathrm{i}$. The exterior wavenumber is $k_e = 1.5$ in both cases. The interior wavenumbers are given by $k_m = n_m k_e$, for $m = 1,2,3$. The spheres were discretised with a mesh width of approximately 20 elements per wavelength.

where Y_n^p *are the scalar spherical harmonics,* (θ,ϕ) *are the spherical polar coordinates and* ∇_Γ *is the surface gradient. Then, for large* n *and for* $M = 1$

$$\boldsymbol{\mathcal{A}}^2 \begin{bmatrix} \mathbf{U}_n^p \\ \mathbf{X}_n^p \end{bmatrix} = \left(\Lambda_1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \begin{bmatrix} \mathbf{U}_n^p \\ \mathbf{X}_n^p \end{bmatrix}, \tag{4.7}$$

$$\boldsymbol{\mathcal{A}}^2 \begin{bmatrix} \mathbf{X}_n^p \\ \mathbf{U}_n^p \end{bmatrix} = \left(\Lambda_2 + \mathcal{O}\left(\frac{1}{n}\right)\right) \begin{bmatrix} \mathbf{X}_n^p \\ \mathbf{U}_n^p \end{bmatrix}. \tag{4.8}$$

Proof. To prove the theorem we use the definitions and properties of the vector spherical harmonics and the properties of boundary integral operators applied to them as presented in [98, 99]. Let $P_n(x)$ denote the standard Legendre polynomial of degree n. The spherical harmonic of degree n and order p is defined by

$$Y_n^p(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}} \sqrt{\frac{(n-|p|)!}{(n+|p|)!}} P_n^{|p|}(\cos\theta)\exp(\mathrm{i}p\phi). \tag{4.9}$$

The spherical Bessel and Hankel functions of the first kind are defined as

$$j_n(x) = \sqrt{\frac{\pi}{2x}} J_{n+\frac{1}{2}}(x), \qquad h_n^{(1)}(x) = \sqrt{\frac{\pi}{2x}} H^{(1)}_{n+\frac{1}{2}}(x), \tag{4.10}$$

where J_n and $H_n^{(1)}$ are the Bessel and Hankel functions of the first kind. The Riccati-Bessel and Riccati-Hankel functions are defined as

$$\mathbb{J}_n(x) = x j_n(x), \qquad \mathbb{H}_n(x) = x h_n^{(1)}(x). \tag{4.11}$$

Applying the magnetic and electric boundary integral operators to the vector spherical harmonics we obtain

$$\begin{aligned}
\mathcal{C}(\mathbf{U}_n^p) &= \frac{1}{2}\left(\mathrm{i}\mathbb{J}_n(k)\mathbb{H}_n'(k) + \mathrm{i}\mathbb{J}_n'(k)\mathbb{H}_n(k)\right)\mathbf{U}_n^p, \qquad \mathcal{S}(\mathbf{U}_n^p) = -\mathbb{J}_n'(k)\mathbb{H}_n'(k)\mathbf{X}_n^p, \\
\mathcal{C}(\mathbf{X}_n^p) &= -\frac{1}{2}\left(\mathrm{i}\mathbb{J}_n(k)\mathbb{H}_n'(k) + \mathrm{i}\mathbb{J}_n'(k)\mathbb{H}_n(k)\right)\mathbf{X}_n^p, \qquad \mathcal{S}(\mathbf{X}_n^p) = \mathbb{J}_n(k)\mathbb{H}_n(k)\mathbf{U}_n^p.
\end{aligned} \tag{4.12}$$

where $\mathbb{J}_n'(k)$ and $\mathbb{H}_n'(k)$ are the derivatives of the functions with respect to their arguments. In the asymptotic limit of $n \to \infty$ the following hold [98]

$$\mathbb{J}_n(x)\mathbb{H}_n(x) \sim -\frac{\mathrm{i}x}{2n} + \mathcal{O}\left(\frac{1}{n^2}\right), \tag{4.13}$$

$$\mathbb{J}_n'(x)\mathbb{H}_n'(x) \sim \frac{n\mathrm{i}}{2x} + \mathcal{O}(1), \tag{4.14}$$

$$\mathrm{i}\mathbb{J}_n(k)\mathbb{H}_n'(k) + \mathrm{i}\mathbb{J}_n'(k)\mathbb{H}_n(k) \sim \mathcal{O}\left(\frac{1}{n}\right). \tag{4.15}$$

We can now apply $\boldsymbol{\mathcal{A}}^2$ to the vector spherical harmonics and obtain their asymptotic forms. One can show that for the diagonal blocks

$$\begin{aligned}
\boldsymbol{\mathcal{A}}_{11}^2(\mathbf{U}_n^p) &= \frac{1}{4}\left(\mathrm{i}\mathbb{J}_n(k_1)\mathbb{H}_n'(k_1) + \mathrm{i}\mathbb{J}_n'(k_1)\mathbb{H}_n(k_1) + \mathrm{i}\mathbb{J}_n(k_e)\mathbb{H}_n'(k_e) + \mathrm{i}\mathbb{J}_n'(k_e)\mathbb{H}_n(k_e)\right)^2 \mathbf{U}_n^p \\
&+ \left(\frac{\mu_1}{k_1}\mathbb{J}_n(k_1)\mathbb{H}_n(k_1) + \frac{\mu_e}{k_e}\mathbb{J}_n(k_e)\mathbb{H}_n(k_e)\right)\left(\frac{k_1}{\mu_1}\mathbb{J}_n'(k_1)\mathbb{H}_n'(k_1) + \frac{k_e}{\mu_e}\mathbb{J}_n'(k_e)\mathbb{H}_n'(k_e)\right)\mathbf{U}_n^p,
\end{aligned} \tag{4.16}$$

$$\boldsymbol{\mathcal{A}}_{11}^2(\mathbf{X}_n^p) = \frac{1}{4}\left(\mathrm{i}\mathbb{J}_n(k_1)\mathbb{H}_n'(k_1) + \mathrm{i}\mathbb{J}_n'(k_1)\mathbb{H}_n(k_1) + \mathrm{i}\mathbb{J}_n(k_e)\mathbb{H}_n'(k_e) + \mathrm{i}\mathbb{J}_n'(k_e)\mathbb{H}_n(k_e)\right)^2 \mathbf{X}_n^p$$

$$+\left(\frac{\mu_1}{k_1}\mathbb{J}'_n(k_1)\mathbb{H}'_n(k_1)+\frac{\mu_e}{k_e}\mathbb{J}'_n(k_1)\mathbb{H}'_n(k_1)\right)\left(\frac{k_1}{\mu_1}\mathbb{J}_n(k_1)\mathbb{H}_n(k_1)+\frac{k_e}{\mu_e}\mathbb{J}_n(k_e)\mathbb{H}_n(k_e)\right)\mathbf{X}_n^p, \tag{4.17}$$

$$\boldsymbol{\mathcal{A}}_{22}^2(\mathbf{U}_n^p)=\frac{1}{4}\left(\mathrm{i}\mathbb{J}_n(k_1)\mathbb{H}'_n(k_1)+\mathrm{i}\mathbb{J}'_n(k_1)\mathbb{H}_n(k_1)+\mathrm{i}\mathbb{J}_n(k_e)\mathbb{H}'_n(k_e)+\mathrm{i}\mathbb{J}'_n(k_e)\mathbb{H}_n(k_e)\right)^2\mathbf{U}_n^p$$
$$+\left(\frac{k_1}{\mu_1}\mathbb{J}_n(k_1)\mathbb{H}_n(k_1)+\frac{k_e}{\mu_e}\mathbb{J}_n(k_e)\mathbb{H}_n(k_e)\right)\left(\frac{\mu_1}{k_1}\mathbb{J}'_n(k_1)\mathbb{H}'_n(k_1)+\frac{\mu_e}{k_e}\mathbb{J}'_n(k_e)\mathbb{H}'_n(k_e)\right)\mathbf{U}_n^p, \tag{4.18}$$

$$\boldsymbol{\mathcal{A}}_{22}^2(\mathbf{X}_n^p)=\frac{1}{4}\left(\mathrm{i}\mathbb{J}_n(k_1)\mathbb{H}'_n(k_1)+\mathrm{i}\mathbb{J}'_n(k_1)\mathbb{H}_n(k_1)+\mathrm{i}\mathbb{J}_n(k_e)\mathbb{H}'_n(k_e)+\mathrm{i}\mathbb{J}'_n(k_e)\mathbb{H}_n(k_e)\right)^2\mathbf{X}_n^p$$
$$+\left(\frac{k_1}{\mu_1}\mathbb{J}'_n(k_1)\mathbb{H}'_n(k_1)+\frac{k_e}{\mu_e}\mathbb{J}'_n(k_e)\mathbb{H}'_n(k_e)\right)\left(\frac{\mu_1}{k_1}\mathbb{J}_n(k)\mathbb{H}_n(k)+\frac{k_e}{\mu_e}\mathbb{J}_n(k_e)\mathbb{H}_n(k_e)\right)\mathbf{X}_n^p, \tag{4.19}$$

and for the off-diagonal blocks

$$\boldsymbol{\mathcal{A}}_{12}^2(\mathbf{U}_n^p)=\frac{1}{2}\left(\mathrm{i}\mathbb{J}_n(k_1)\mathbb{H}'_n(k_1)+\mathrm{i}\mathbb{J}'_n(k_1)\mathbb{H}_n(k_1)+\mathrm{i}\mathbb{J}_n(k_e)\mathbb{H}'_n(k_e)+\mathrm{i}\mathbb{J}'_n(k_e)\mathbb{H}_n(k_e)\right)$$
$$\left(\frac{\mu_1}{k_1}\mathbb{J}'_n(k_1)\mathbb{H}'_n(k_1)+\frac{\mu_e}{k_e}\mathbb{J}'_n(k_e)\mathbb{H}'_n(k_e)\right)\mathbf{X}_n^p$$
$$-\frac{1}{2}\left(\mathrm{i}\mathbb{J}_n(k_1)\mathbb{H}'_n(k_1)+\mathrm{i}\mathbb{J}'_n(k_1)\mathbb{H}_n(k_1)+\mathrm{i}\mathbb{J}_n(k_e)\mathbb{H}'_n(k_e)+\mathrm{i}\mathbb{J}'_n(k_e)\mathbb{H}_n(k_e)\right)$$
$$\left(\frac{\mu_1}{k_1}\mathbb{J}'_n(k_1)\mathbb{H}'_n(k_1)+\frac{\mu_e}{k_e}\mathbb{J}'_n(k_e)\mathbb{H}'_n(k_e)\right)\mathbf{X}_n^p, \tag{4.20}$$

$$\boldsymbol{\mathcal{A}}_{12}^2(\mathbf{X}_n^p)=\frac{1}{2}\left(\mathrm{i}\mathbb{J}_n(k_1)\mathbb{H}'_n(k_1)+\mathrm{i}\mathbb{J}'_n(k_1)\mathbb{H}_n(k_1)+\mathrm{i}\mathbb{J}_n(k_e)\mathbb{H}'_n(k_e)+\mathrm{i}\mathbb{J}'_n(k_e)\mathbb{H}_n(k_e)\right)$$
$$\left(\frac{\mu_1}{k_1}\mathbb{J}_n(k_1)\mathbb{H}_n(k_1)+\frac{\mu_e}{k_e}\mathbb{J}_n(k_e)\mathbb{H}_n(k_e)\right)\mathbf{U}_n^p$$
$$-\frac{1}{2}\left(\mathrm{i}\mathbb{J}_n(k_1)\mathbb{H}'_n(k_1)+\mathrm{i}\mathbb{J}'_n(k_1)\mathbb{H}_n(k_1)+\mathrm{i}\mathbb{J}_n(k_e)\mathbb{H}'_n(k_e)+\mathrm{i}\mathbb{J}'_n(k_e)\mathbb{H}_n(k_e)\right)$$
$$\left(\frac{\mu_1}{k_1}\mathbb{J}_n(k_1)\mathbb{H}_n(k_1)+\frac{\mu_e}{k_e}\mathbb{J}_n(k_e)\mathbb{H}_n(k_e)\right)\mathbf{U}_n^p, \tag{4.21}$$

$$\boldsymbol{\mathcal{A}}_{21}^2(\mathbf{U}_n^p)=\frac{1}{2}\left(\mathrm{i}\mathbb{J}_n(k_1)\mathbb{H}'_n(k_1)+\mathrm{i}\mathbb{J}'_n(k_1)\mathbb{H}_n(k_1)+\mathrm{i}\mathbb{J}_n(k_e)\mathbb{H}'_n(k_e)+\mathrm{i}\mathbb{J}'_n(k_e)\mathbb{H}_n(k_e)\right)$$
$$\left(\frac{k_1}{\mu_1}\mathbb{J}'_n(k_1)\mathbb{H}'_n(k_1)+\frac{k_e}{\mu_e}\mathbb{J}'_n(k_e)\mathbb{H}'_n(k_e)\right)\mathbf{X}_n^p$$
$$-\frac{1}{2}\left(\mathrm{i}\mathbb{J}_n(k_1)\mathbb{H}'_n(k_1)+\mathrm{i}\mathbb{J}'_n(k_1)\mathbb{H}_n(k_1)+\mathrm{i}\mathbb{J}_n(k_e)\mathbb{H}'_n(k_e)+\mathrm{i}\mathbb{J}'_n(k_e)\mathbb{H}_n(k_e)\right)$$
$$\left(\frac{k_1}{\mu_1}\mathbb{J}'_n(k_1)\mathbb{H}'_n(k_1)+\frac{k_e}{\mu_e}\mathbb{J}'_n(k_e)\mathbb{H}'_n(k_e)\right)\mathbf{X}_n^p, \tag{4.22}$$

$$\boldsymbol{\mathcal{A}}_{21}^2(\mathbf{X}_n^p)=\frac{1}{2}\left(\mathrm{i}\mathbb{J}_n(k_1)\mathbb{H}'_n(k_1)+\mathrm{i}\mathbb{J}'_n(k_1)\mathbb{H}_n(k_1)+\mathrm{i}\mathbb{J}_n(k_e)\mathbb{H}'_n(k_e)+\mathrm{i}\mathbb{J}'_n(k_e)\mathbb{H}_n(k_e)\right)$$

$$\left(\frac{k_1}{\mu_1}\mathbb{J}_n(k_1)\mathbb{H}_n(k_1)+\frac{k_e}{\mu_e}\mathbb{J}_n(k_e)\mathbb{H}_n(k_e)\right)(\mathbf{U}_n^p)$$
$$-\frac{1}{2}\left(\mathrm{i}\mathbb{J}_n(k_1)\mathbb{H}_n'(k_1)+\mathrm{i}\mathbb{J}_n'(k_1)\mathbb{H}_n(k_1)+\mathrm{i}\mathbb{J}_n(k_e)\mathbb{H}_n'(k_e)+\mathrm{i}\mathbb{J}_n'(k_e)\mathbb{H}_n(k_e)\right)$$
$$\left(\frac{k_1}{\mu_1}\mathbb{J}_n(k_1)\mathbb{H}_n(k_1)+\frac{k_e}{\mu_e}\mathbb{J}_n(k_e)\mathbb{H}_n(k_e)\right)(\mathbf{U}_n^p). \tag{4.23}$$

The off-diagonal parts are identically zero, while the remaining diagonal terms in asymptotic form give

$$\boldsymbol{\mathcal{A}}_{11}^2(\mathbf{U}_n^p)\sim\Lambda_1\mathbf{U}_n^p, \qquad \boldsymbol{\mathcal{A}}_{11}^2(\mathbf{X}_n^p)\sim\Lambda_2\mathbf{X}_n^p, \tag{4.24}$$
$$\boldsymbol{\mathcal{A}}_{22}^2(\mathbf{U}_n^p)\sim\Lambda_2\mathbf{U}_n^p, \qquad \boldsymbol{\mathcal{A}}_{22}^2(\mathbf{X}_n^p)\sim\Lambda_1\mathbf{X}_n^p. \tag{4.25}$$

Combining the asymptotic forms proves the theorem. □

We now examine whether the accumulation points (4.3)–(4.4) can be zero or tend to infinity. We can observe

- $\Lambda_1=0 \iff \mu_1=-\mu_e$,
- $\Lambda_2=0 \iff \dfrac{\mu_1}{k_1^2}=-\dfrac{\mu_e}{k_e^2}$. We note that for the applications considered here, k_1 is usually complex given by $k_1=n_1k_e$, where n_1 is the complex refractive index of the scatterer. So the condition could be rewritten as $\Lambda_2=0 \iff \dfrac{\mu_1}{\mu_e}=-n_1^2$.

In addition, since all parameter values are finite we do not expect any accumulation points at infinity. We should note though that high contrast materials (for example if one of the permeabilities or one of the wavenumbers is much larger than the other) would result in higher values for Λ_1 and/or Λ_2. For the purposes of this thesis, of electromagnetic scattering by ice crystals, $\mu_e=\mu_1=1$ ensuring the accumulation point $\Lambda_1=1$, i.e. not clustering around zero and not tending to infinity. For the case of high contrast materials, we refer to [38], where a scaled version of the Calderón preconditioner is used, in which the two accumulation points coincide.

4.1.2 Numerical Experiments

We now investigate the performance of the Calderón preconditioner in the context of iterative solvers. For the numerical experiments that follow, we find it helpful to use the mixed discretisation of (3.41), since the preconditioner and operator are identical and one just needs to build the discrete version of $\boldsymbol{\mathcal{A}}$ and reuse it for $\boldsymbol{\mathcal{P}}$ (we refer back to Section 3.2.2 for a more detailed discussion). In this case, $\mathbf{P} = \mathbf{A}$ and $\mathbf{M_P} = \mathbf{M_A} = \mathbf{M}$. For the discretisation and assembly of $\mathbf{A}$ we use $\nu = 0.001$, $\chi = \infty$ and $\mathbf{q} = (4,3,2,6)$, which are the default Bempp parameter values. We recall the definitions of those parameters were given in (3.45), (3.46) and Section 3.3.2 respectively.

To assess the cost of Calderón preconditioning, we count the total number of matrix-vector products (termed matvecs) performed. By a single matvec we mean a single application of one discretised boundary integral operator $\mathcal{C}^i_m$, $\mathcal{C}^e_m$, $\mathcal{C}^e_{m\ell}$ etc. Applications of the inverse mass matrix $\mathbf{M}^{-1}$ are ignored as the cost is negligible compared to that of the other operators (recall Section 3.3.3). Each diagonal block of $\boldsymbol{\mathcal{A}}$ incurs a cost of 8 matvecs, while each off-diagonal block incurs only 4. The overall cost in terms of matvecs of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{A}}^2$ can be seen in the first two lines of Table 4.1 (the remaining preconditioners are discussed later in Section 4.2). The cost shown takes into account the initial pre-multiplication of the right-hand-side (where relevant). We note that such a comparison is valid if the cost of the boundary integral operators is the same for each scatterer Γ_m (or pair of scatterers (Γ_ℓ, Γ_m) in the case of off-diagonal operators), which will be true provided each scatterer is discretised with roughly the same number of degrees of freedom. We also note that the matvec count per GMRES iteration is the same for weak and strong forms (we refer back to Chapter 3 for the weak and strong forms, and in particular to (3.16)–(3.17) and (3.29)–(3.30)), but the value of R (number of GMRES iterations) is expected to be different. Indeed, as we will see in the numerical experiments, strong forms (i.e. mass matrix preconditioning) require a reduced number of GMRES iterations to converge, and hence incur a reduced matvec

Operator	Total matvecs
$\boldsymbol{\mathcal{A}}$	$(4M^2+4M)(R+\lfloor R/\rho \rfloor)$
$\boldsymbol{\mathcal{A}}^2$	$(8M^2+8M)(R+\lfloor R/\rho \rfloor)+4M^2+4M$
$\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{A}}$	$(4M^2+12M)(R+\lfloor R/\rho \rfloor)+8M$
$\boldsymbol{\mathcal{D}}^i\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{D}}^e\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{A}}$	$(4M^2+8M)(R+\lfloor R/\rho \rfloor)+4M$
$\boldsymbol{\mathcal{S}}^i\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{S}}^e\boldsymbol{\mathcal{A}}$	$(4M^2+6M)(R+\lfloor R/\rho \rfloor)+2M$

Table 4.1: Total matvecs for different choices of preconditioning operator $\boldsymbol{\mathcal{P}}$. The preconditioners $\boldsymbol{\mathcal{D}}$, $\boldsymbol{\mathcal{D}}^e$, $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ are defined later in (4.26)–(4.30), while the preconditioner $\boldsymbol{\mathcal{D}}^i$ has been defined in (2.89). Here R is the number of GMRES iterations required to achieve convergence at a specified tolerance (note that R is expected to depend strongly on the choice of $\boldsymbol{\mathcal{P}}$), ρ is the number of iterations per cycle passed as the restart argument in GMRES, $\lfloor \cdot \rfloor$ is the "floor" function, and M is the number of scatterers.

cost.

We remark that the total number of matvecs is not the only metric to assess the computational cost, since one also needs to take into account the size of the discrete system (governed by the mesh resolution) and the assembly method together with the discretisation scheme, as detailed in Chapter 3. While all our experiments are using $\mathcal{H}$-matrix algorithms, reducing the number of matvecs is also advantageous when FMM is used [84], making our findings relevant for other applications or implementations.

We compare the performance of the traditional Calderón preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, in the context of single and multi-particle scattering on benchmark problems, as described in Figure 4.3. In all our experiments the incident wave is a plane wave $\mathbf{E}^{inc} = \mathbf{p}e^{ik_e \mathbf{d}\cdot\mathbf{x}}$ with $\mathbf{p} = (0,0,1)^T$ and $\mathbf{d} = (1,0,0)^T$. In order to capture the wave solution we take the maximum BEM mesh element size to be $2\pi/(10k_e)$, as it was shown in [5] that a discretisation of at least 10 elements per wavelength ($\lambda_e = 2\pi/k_e$), leads to a typical discretisation relative error of approximately 1%, provided that a sufficient resolution of the boundary is achieved. As all our scatterers are polyhedral (except when considering spheres) the scatterer geometry is captured exactly by the BEM mesh. We use

λ_e	k_e	Refractive index	X				
			$M=1$	$M=3$	$M=4$	$M=8$	$M=16$
0.55	11.42	$1.311+2.289\times10^{-9}\mathbf{i}$	3.96	14.08	16.03	17.77	25.08
10.87	0.58	$1.0833+0.204\mathbf{i}$	0.20	0.71	0.81	0.90	1.26

Table 4.2: Wavelengths λ_e (μm), wavenumbers k_e and corresponding refractive indices of ice from [11] and parameter sizes X for the model problems of Figure 4.3

GMRES [92] to solve the discrete linear system, using the Scipy library [100], with parameters as follows: `tol=1E-5`, `restart=200` and `maxiter=2000`.

We present results for two different refractive indices of different absorption: one with weak absorption $n = 1.311 + 2.289 \times 10^{-9}\mathbf{i}$ and one with high absorption $n = 1.0833 + 0.204\mathbf{i}$. These correspond to the measured refractive index of ice at wavelengths $\lambda_e = 0.55\mu m$ ($k_e = 11.42$) and $\lambda_e = 10.87\mu m$ ($k_e = 0.58$) respectively [11]. Using these two refractive indices we present results from numerical experiments for their corresponding wavenumbers first and then apply those for a range of wavenumbers, simulating scattering by fixed geometrical configurations at a range of different size parameters X (defined in (1.4)). The size parameters for the model problems of Figure 4.3 at those two wavelengths are shown in Table 4.2.

We begin with single-particle scattering with the Model Problem $M = 1$ of Figure 4.3. In Table 4.3 we present GMRES iterations, total matvec count and GMRES time (t_{solve}) for the weak and strong forms of the operators $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{A}}^2$ for the two refractive indices and corresponding exterior wavenumbers mentioned above. The ill-conditioning of the weak form $\mathbf{A}$ is clearly visible in the large number of GMRES iterations (and hence matvecs) in the first row. Indeed, for the second case considered, $n = 1.0833 + 0.204\mathbf{i}$, the resulting linear system does not converge to the prescribed tolerance within the maximum number of iterations given. The remaining three formulations, $\mathbf{M}^{-1}\mathbf{A}$, $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$, and $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$ all provide some form of preconditioning indicated by the reduced GMRES count (and hence matvec and solve time). While Calderón preconditioned strong form $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$ achieves the low-

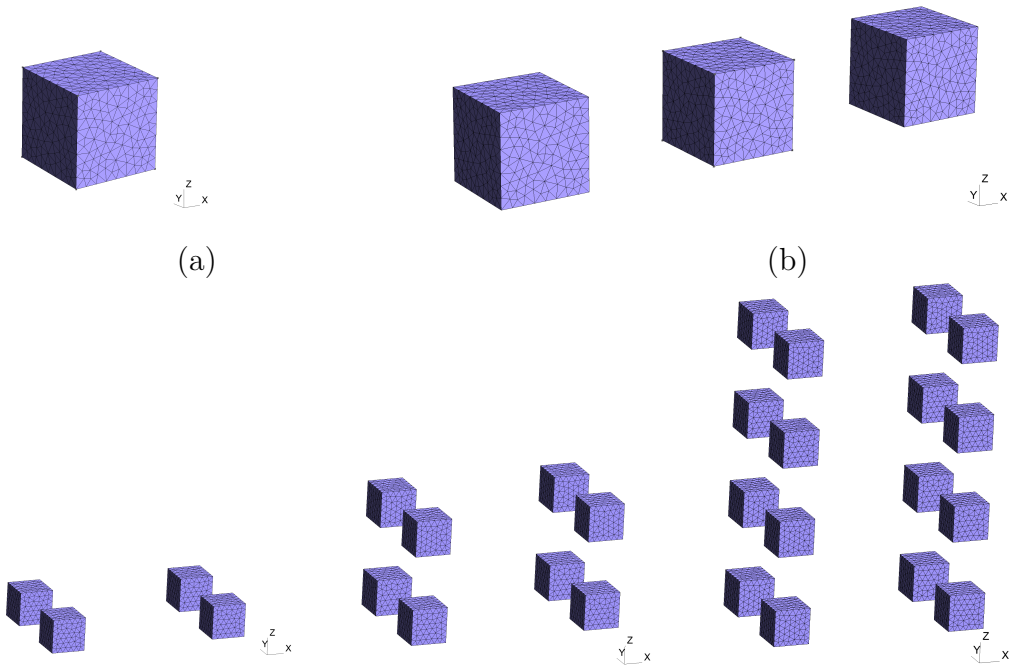


Figure 4.3: Scattering configurations. For the definition of the size parameter X we recall (1.4). We note that the original cube in each configuration is fixed as in (a) and the additional cubes are copies of the original one, with the problem growing as $\mathcal{O}(M)$, where M is the number of scatterers in each case.
(a) **Model Problem** $M = 1$: Cube of side length $0.4\mu m$, aligned with coordinate axes and has the front bottom left vertex at $(0,0,0)$. Size parameter is $X = 2.18/\lambda_e$.
(b) **Model Problem** $M = 3$: Three identical cubes of side length $0.4\mu m$, aligned with coordinate axes that have their bottom left vertices at $(-1,0,0)$, $(0,0,0)$ and $(1,0,0)$ respectively. Size parameter is $X = 7.75/\lambda_e$.
(c) **Model Problem** $M = 4$: Four identical cubes of side length $0.4\mu m$, aligned with coordinate axes that have their bottom left vertices at $(\pm 1,0,0)$ and $(\pm 1,-1,0)$. Size parameter is $X = 8.82/\lambda_e$.
(d) **Model Problem** $M = 8$: Same as above but with $M = 8$ and additional bottom left vertices at $(\pm 1,0,1)$ and $(\pm 1,-1,1)$. Size parameter is $X = 9.77/\lambda_e$.
(e) **Model Problem** $M = 16$: Same as above but with $M = 16$ and additional bottom left vertices at $(\pm 1,0,-2)$, $(\pm 1,-1,-2)$, $(\pm 1,0,-1)$ and $(\pm 1,-1,-1)$. Size parameter is $X = 13.79/\lambda_e$.

	$n = 1.311 + 2.289 \times 10^{-9}\mathrm{i}$			$n = 1.0833 + 0.204\mathrm{i}$		
	Iters	Matvecs	t_{solve}	Iters	Matvecs	t_{solve}
Discrete Operator						
$\mathbf{A}$	142	1136	1	2000	16088	1
$\mathbf{M}^{-1}\mathbf{A}$	11	**88**	**0.080**	8	**64**	**0.004**
$\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$	35	568	0.507	27	440	0.027
$\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$	**6**	104	0.094	**5**	88	0.006

Table 4.3: Number of GMRES iterations, total matvec count and total GMRES time (t_{solve}) for different discrete formulations for the Model Problem $M = 1$ of Figure 4.3, for two refractive indices $n = 1.311 + 2.289 \times 10^{-9}\mathrm{i}$ (left set of columns) and $n = 1.0833 + 0.204\mathrm{i}$ (right set of columns) and their corresponding exterior wavenumbers (shown in Table 4.2, together with size parameters), with the other scattering parameters set to $\mu_e = \mu_1 = 1$ and $k_1 = nk_e$. GMRES times have been normalised relative to that of $\mathbf{A}$. We note that a discretisation of 100 elements per wavelength was prescribed for the results of the second set of columns in order to achieve a sufficient discretisation of the boundary. Minimal values in each column are indicated in bold type.

	$n = 1.311 + 2.289 \times 10^{-9}\mathrm{i}$			$n = 1.0833 + 0.204\mathrm{i}$		
	Iters	Matvecs	t_{solve}	Iters	Matvecs	t_{solve}
Discrete Operator						
$\mathbf{A}$	358	17232	1	2000	96528	1
$\mathbf{M}^{-1}\mathbf{A}$	15	**720**	**0.04**	8	**384**	**0.004**
$\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$	43	4176	0.25	27	2640	0.028
$\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$	**8**	816	0.05	**5**	528	0.006

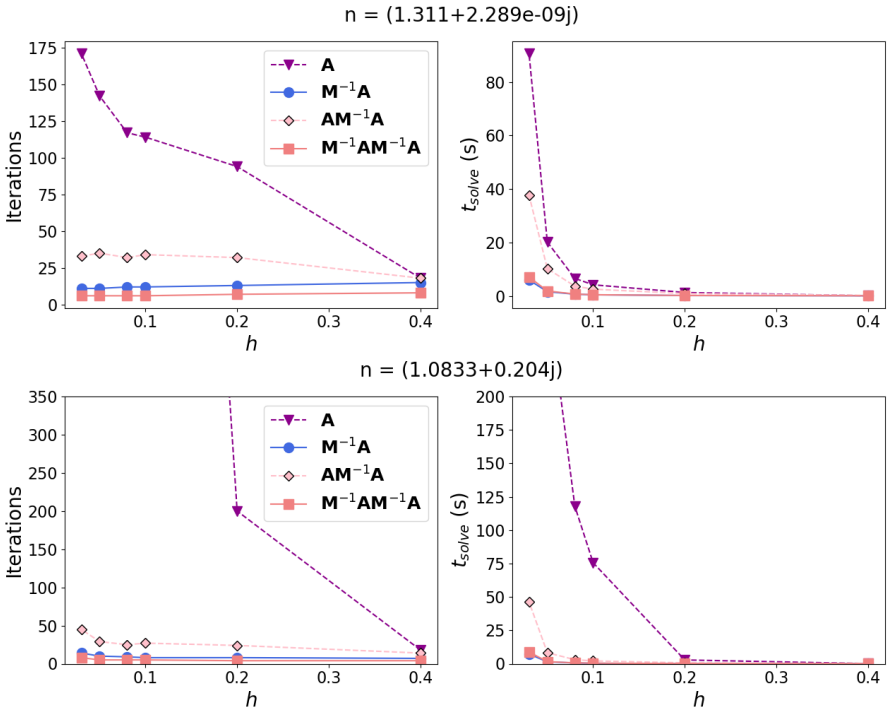
Table 4.4: Same as for Table 4.3 but for $M = 3$ of Figure 4.3.

est number of GMRES iterations, the increased matvec count per iteration means that overall it is more expensive than $\mathbf{M}^{-1}\mathbf{A}$ in terms of matvecs, and somewhat slower. While the Calderón weak form $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$ (which is what is traditionally studied, for example in [36]) provides a preconditioning effect, its performance cannot compete with $\mathbf{M}^{-1}\mathbf{A}$ and $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$.

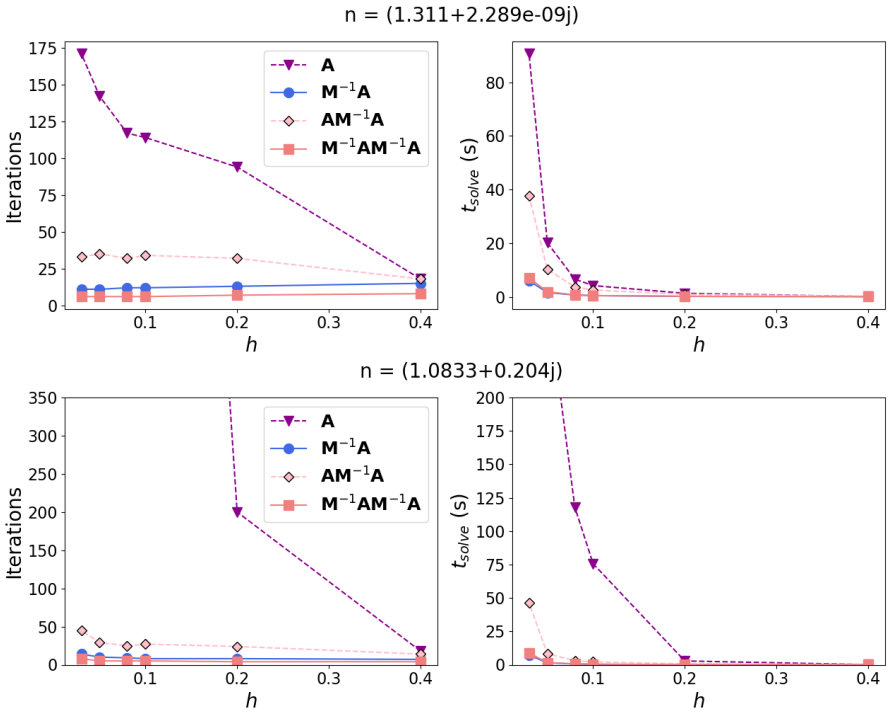
The same behaviour is observed when we move to multi-particle scattering with the Model Problem $M = 3$ of Figure 4.3. In Table 4.4, we repeat the same numerical experiment but for $M = 3$ and report the same output. As with the single-particle case, we can observe the ill-conditioning of $\mathbf{A}$, while the remaining three formulations provide some form of preconditioning. Strong forms $\mathbf{M}^{-1}\mathbf{A}$ and $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$ have the least computational cost either in terms of iterations or matvecs and solver time.

We note that the same behaviour observed in Tables 4.3 and 4.4 is also seen as $h \to 0$. In Figure 4.4, we repeat the simulations of Tables 4.3 and 4.4 but for a decreasing mesh size h. For both $M = 1$ and $M = 3$, for $n = 1.0833 + 0.204\mathrm{i}$, $\mathbf{A}$ converges when $h = 0.4$ and $h = 0.2$ but not for any smaller values. In all problems, $\mathbf{A}$ shows a strong mesh size dependence. The other three operators have a weaker mesh size dependence, with the strong forms $\mathbf{M}^{-1}\mathbf{A}$ and $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$ showing an almost mesh-size-independent behaviour. Consistent with the results from Tables 4.3 and 4.4 mass matrix preconditioning, $\mathbf{M}^{-1}\mathbf{A}$, performs as well as strong Calderón preconditioning, $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$.

Given that $\mathbf{M}^{-1}\mathbf{A}$ and $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$ have performed quite similar in overall matvec count and GMRES time, we wish to compare their behaviour for a larger range of wavenumbers. In Figure 4.5, we compare their performance in terms of GMRES iterations, matvecs and GMRES time as a function of the exterior wavenumber k_e for the Model Problems $M = 1$ and $M = 3$ of Figure 4.3. For both refractive indices the number of GMRES iterations for $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$ is approximately half of that for $\mathbf{M}^{-1}\mathbf{A}$, with faster convergence for high absorption and slower growth with increasing exterior wavenumber k_e, but the overall matvec count for $\mathbf{M}^{-1}\mathbf{A}$ is consistently lower than that of



(a) $M = 1$



(b) $M = 3$

Figure 4.4: Performance of the discrete operators $\mathbf{A}$, $\mathbf{M}^{-1}\mathbf{A}$, $\mathbf{AM}^{-1}\mathbf{A}$ and $\mathbf{M}^{-1}\mathbf{AM}^{-1}\mathbf{A}$ for the Model Problems $M = 1$ and $M = 3$ of Figure 4.3 as a function of decreasing mesh size h. Other parameters are as in Tables 4.3 and 4.4.

$\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$. GMRES time grows with increasing exterior wavenumber k_e because the number of degrees of freedom grows approximately quadratically with k_e in order to maintain a sufficient discretisation of the boundary. Both $\mathbf{M}^{-1}\mathbf{A}$ and $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$ require roughly the same GMRES solver time, with $\mathbf{M}^{-1}\mathbf{A}$ being only somewhat faster when $M = 1$.

All numerical experiments presented in Tables 4.3–4.4 and Figures 4.4–4.5 suggest that even though traditional Calderón preconditioning provides a faster solution with a reduced computational cost compared to the weak PMCHWT operator $\mathbf{A}$, it does not perform any better compared to simple mass matrix preconditioning $\mathbf{M}^{-1}\mathbf{A}$, rendering traditional Calderón preconditioners with the specific implementation used unnecessary. For this reason, we explore reduced versions of Calderón preconditioning in an effort to reduce the overall computational cost.

We note that in the multi-particle case, each aggregate consists of multiple individual scatterers of same size and shape that are placed in different positions (see Figure 4.3). One might consider using block Jacobi to iteratively obtain the inverse for one monomer and attempt to reuse it for the remaining ones. However, one cannot guarantee a priori that Jacobi will provide a robust preconditioning effect in terms of taming the hypersingular component, especially for domains more complex than the cubes considered here. In addition, the performance of $\mathbf{M}^{-1}\mathbf{A}$ and $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$ has shown a very weak mesh dependence, which again cannot be guaranteed for block Jacobi. This was indeed studied in [40] for the EFIE formulation, where Jacobi preconditioning was considered along with other Calderón preconditioners. While it provided a preconditioning effect for the sphere, the experiments showed that this behaviour did not carry forward to more complex domains such as scattering by a destroyer, nor did it exhibit mesh-independent behaviour in contrast to other Calderón preconditioners.

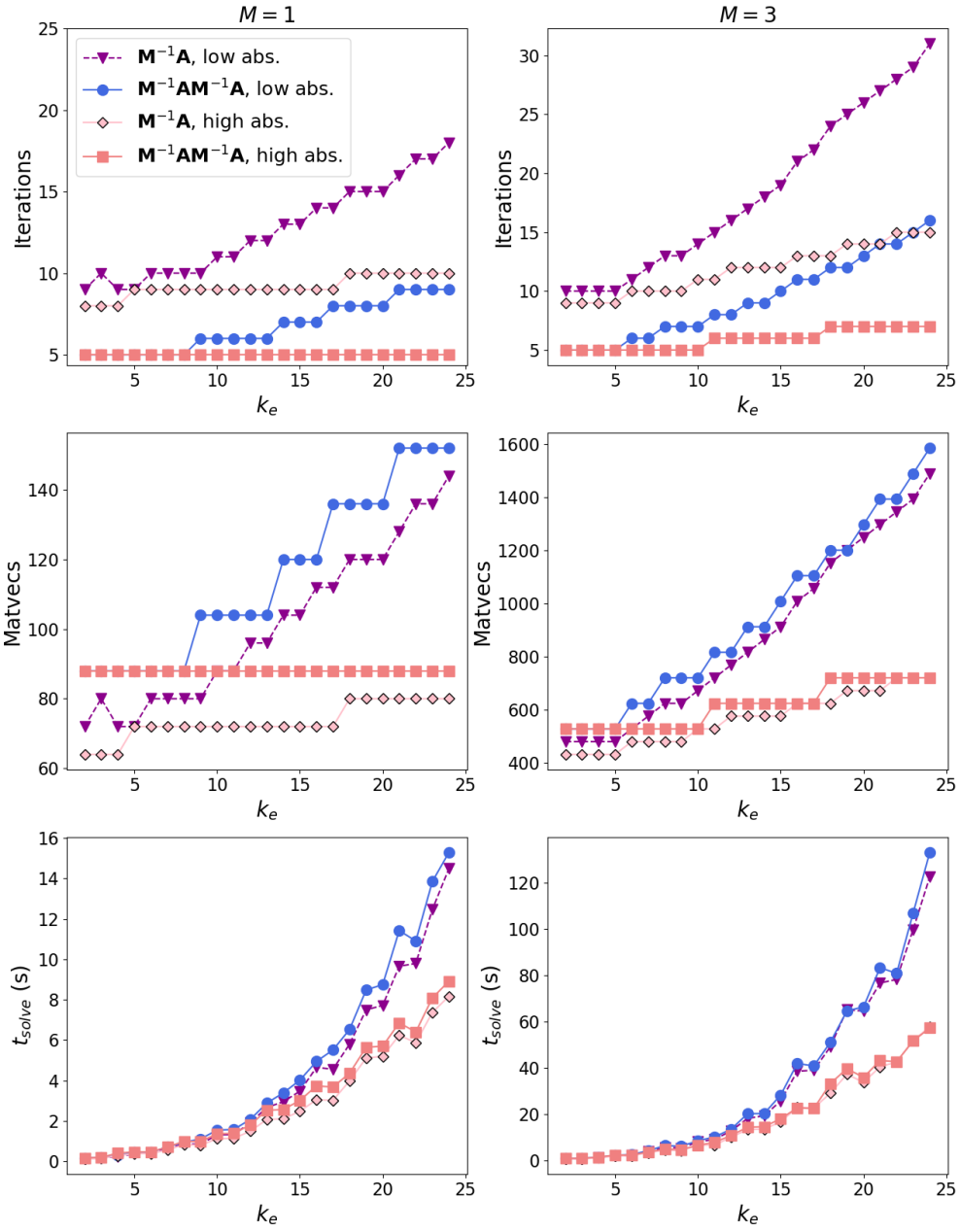


Figure 4.5: Performance of the discrete operators $\mathbf{M}^{-1}\mathbf{A}$ and $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}$ in the case of scattering by a single cube (left) and three cubes (right) (Model Problems $M = 1$ and $M = 3$ of Figure 4.3) as a function of exterior wavenumber k_e. Results are shown for low absorption (refractive index $n = 1.311 + 2.289 \times 10^{-9}\mathrm{i}$) and high absorption (refractive index $n = 1.0833 + 0.204\mathrm{i}$). The number of degrees of freedom range from 126 at $k_e = 2$ to 5544 at $k_e = 24$ with corresponding size parameters $X = 0.69$ and $X = 8.33$ for the case $M = 1$, and 378 at $k_e = 2$ to 16626 at $k_e = 24$ with corresponding size parameters $X = 2.46$ and $X = 29.59$ when $M = 3$. Other parameters are as in Tables 4.3 and 4.4.

4.2 Reduced Calderón preconditioning

The first step in reducing the cost of Calderón preconditioners is noticing that for $M > 1$, the off-diagonal blocks $\boldsymbol{\mathcal{A}}_{m\ell}$ (defined in (2.79)) are actually compact since they map between different boundary components and therefore require no preconditioning. This allows us to introduce the first reduced preconditioner $\boldsymbol{\mathcal{D}}$, which consists of the block-diagonal components of $\boldsymbol{\mathcal{A}}$

$$\boldsymbol{\mathcal{D}} = \begin{bmatrix} \boldsymbol{\mathcal{A}}_1^e + \boldsymbol{\mathcal{A}}_1^i & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{\mathcal{A}}_M^e + \boldsymbol{\mathcal{A}}_M^i \end{bmatrix}. \tag{4.26}$$

Since the off-diagonal blocks do not require any preconditioning, we expect (and confirm later numerically) $\boldsymbol{\mathcal{D}}$ to have the same preconditioning effect as $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, i.e. the number of GMRES iterations should be the same between $\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{A}}^2$. But since $\boldsymbol{\mathcal{D}}$ consists of only diagonal terms, the total matvec cost (see Table 4.1) should reduce by a factor of 2 in the limit $M \to \infty$. The sparsity of $\boldsymbol{\mathcal{D}}$ also implies that applications of $\boldsymbol{\mathcal{D}}$ are faster than those of $\boldsymbol{\mathcal{A}}$, reducing GMRES time as well. We note though, that the off-diagonal blocks have a much cheaper assembly as they map between different domains. Unless the two domains touch, all blocks in the $\mathcal{H}$-matrix are considered admissible and none of the kernels are singular, making their assembly faster and consuming less memory. So even though we have removed the majority of the operators from $\boldsymbol{\mathcal{A}}$ ($4M^2 - 4M$ of the $4M^2 + 4M$ operators), we do not expect to see a proportional decrease in memory consumption and assembly time unless M is large or scatterers are touching each other.

The preconditioner $\boldsymbol{\mathcal{D}}$ reduces to traditional Calderón preconditioning at $M = 1$, but one can reduce the cost even further. The goal is to provide some regularising effect for the electric field boundary operators $\mathcal{S}_m^i$ and $\mathcal{S}_m^e$ in order to accelerate convergence, while keeping the number of operators used

to a minimum in order to minimise assembly time and memory consumption. For this reason we consider the following options as preconditioners:

$$\boldsymbol{\mathcal{D}},\ \boldsymbol{\mathcal{D}}^i,\ \boldsymbol{\mathcal{D}}^e,\ \boldsymbol{\mathcal{S}},\ \boldsymbol{\mathcal{S}}^i,\boldsymbol{\mathcal{S}}^e, \tag{4.27}$$

where $\boldsymbol{\mathcal{D}}$ is defined in (4.26), $\boldsymbol{\mathcal{D}}^i$ in (2.89), $\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{S}}^i + \boldsymbol{\mathcal{S}}^e$ and the remaining preconditioners are defined as

$$\boldsymbol{\mathcal{D}}^e = \begin{bmatrix} \boldsymbol{\mathcal{A}}_1^e & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{\mathcal{A}}_M^e \end{bmatrix}, \tag{4.28}$$

$$\boldsymbol{\mathcal{S}}^e = \begin{bmatrix} \begin{bmatrix} 0 & \frac{\mu_e}{k_e}\mathcal{S}_1^e \\ -\frac{k_e}{\mu_e}\mathcal{S}_1^e & 0 \end{bmatrix} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \begin{bmatrix} 0 & \frac{\mu_e}{k_e}\mathcal{S}_M^e \\ -\frac{k_e}{\mu_e}\mathcal{S}_1^e & 0 \end{bmatrix} \end{bmatrix}, \tag{4.29}$$

$$\boldsymbol{\mathcal{S}}^i = \begin{bmatrix} \begin{bmatrix} 0 & \frac{\mu_1}{k_1}\mathcal{S}_1^i \\ -\frac{k_1}{\mu_1}\mathcal{S}_1^i & 0 \end{bmatrix} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \begin{bmatrix} 0 & \frac{\mu_M}{k_M}\mathcal{S}_M^i \\ -\frac{k_M}{\mu_M}\mathcal{S}_1^i & 0 \end{bmatrix} \end{bmatrix}. \tag{4.30}$$

The accumulation points of $\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{A}}$ are still the same as those of $\boldsymbol{\mathcal{A}}^2$ since the electric field operators that contribute to accumulation points are still present. The magnetic field operators and off-diagonal operators $\boldsymbol{\mathcal{A}}_{m\ell}$

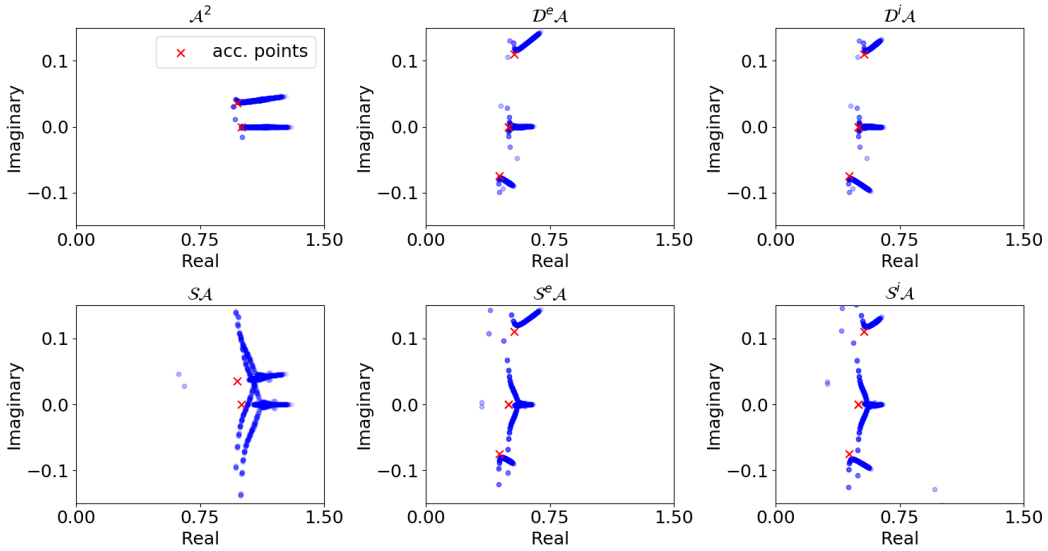


Figure 4.6: Spectrum and accumulation points of the operator products $\boldsymbol{\mathcal{A}}^2$, $\boldsymbol{\mathcal{D}}^e\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{D}}^i\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{S}}^e\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{S}}^i\boldsymbol{\mathcal{A}}$, for a single particle setting. The scattering configurations is the same as that of Figure 4.2. The discretisation of (3.39)–(3.40) was used.

are compact and therefore removing them does not alter the accumulations points. The remaining operator products $\boldsymbol{\mathcal{D}}^e\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{D}}^i\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{S}}^e\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{S}}^i\boldsymbol{\mathcal{A}}$ have 4 accumulation points at

$$\frac{1}{4}+\frac{1}{4}\frac{\mu_1}{\mu_e}, \qquad \frac{1}{4}+\frac{1}{4}\frac{\mu_1}{\mu_e}\left(\frac{k_e}{k_1}\right)^2, \tag{4.31}$$

$$\frac{1}{4}+\frac{1}{4}\frac{\mu_e}{\mu_1}, \qquad \frac{1}{4}+\frac{1}{4}\frac{\mu_e}{\mu_1}\left(\frac{k_1}{k_e}\right)^2, \tag{4.32}$$

for $M = 1$. An example of the spectrum of the operator products, $\boldsymbol{\mathcal{A}}^2$, $\boldsymbol{\mathcal{D}}^e\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{D}}^i\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{S}}^e\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{S}}^i\boldsymbol{\mathcal{A}}$, for a single sphere is shown in Figure 4.6 (same problem considered earlier for Figure 4.2). We note that 3 accumulation points are shown for $\boldsymbol{\mathcal{D}}^e\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{D}}^i\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{S}}^e\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{S}}^i\boldsymbol{\mathcal{A}}$ because for the μ_e, μ_1 values considered the accumulation points $\frac{1}{4}+\frac{1}{4}\frac{\mu_1}{\mu_e}$ and $\frac{1}{4}+\frac{1}{4}\frac{\mu_e}{\mu_1}$ coincide. The arguments used to prove Theorem 1 can also be used to prove the accumulation points of the reduced operator products for the case of a sphere.

The cost of these reduced preconditioners (4.27) in terms of matvecs is given by the bottom three rows of Table 4.1. The preconditioners $\boldsymbol{\mathcal{S}}^i$ and $\boldsymbol{\mathcal{S}}^e$ theoretically result in the lowest matvec count (except when no precondition-

ing is considered), but this does not guarantee *a priori* the lowest GMRES time since R will vary depending on the preconditioning effect of the operator considered. But we expect the memory costs of $\boldsymbol{\mathcal{S}}^i$ and $\boldsymbol{\mathcal{S}}^e$ to be half of $\boldsymbol{\mathcal{D}}^i$, $\boldsymbol{\mathcal{D}}^e$ and $\boldsymbol{\mathcal{S}}$, which in turn should be half of $\boldsymbol{\mathcal{D}}$. Preconditioner $\boldsymbol{\mathcal{D}}$ is much sparser than $\boldsymbol{\mathcal{A}}$, but as mentioned earlier, the off-diagonal blocks that we have removed are fast and cheap to assemble and as such we do not expect a huge improvement in terms of memory consumption and assembly time.

4.2.1 Numerical Experiments

As with traditional Calderón preconditioners, we can use the mixed discretisation of (3.41) to build $\mathbf{A}$ and extract the relevant operators that are needed for the reduced preconditioner, but assembling an operator with **BC** functions as the trial/test space is significantly more expensive than using **RWG**. For these reduced Calderón preconditioners it makes sense to keep **BC** functions for the preconditioners only, where a smaller number of operators are used. For this reason, we use the discretisation of (3.39)–(3.40) for the remaining numerical examples of this section, unless specified otherwise.

We compare the performance of the reduced Calderón preconditioners of (4.27) compared to the traditional one $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, for the Model Problems of Figure 4.3. We only consider their strong forms as we saw earlier that convergence is much faster compared to the corresponding weak forms. As we are only considering strong forms, we refer to the preconditioners by their notation in the continuous set up. We use $\boldsymbol{\mathcal{I}}$ and $\boldsymbol{\mathcal{I}}$-mixed to refer to no Calderón preconditioning which in their discrete strong forms correspond to mass matrix preconditioning; the first one with discretisation (3.39)–(3.40) and the latter with (3.41). All assembly parameters ν, χ and $\mathbf{q}$ are the same as before for both $\mathbf{A}$ and $\mathbf{P}$ and so are the GMRES parameters `tol`, `restart` and `maxiter`.

We begin by repeating the experiments of Tables 4.3 and 4.4, for the reduced preconditioners of (4.27). In addition to GMRES iterations, matvecs and t_{solve}, we also report assembly time, $t_{assembly}$, total time, $t_{total} = t_{assembly} +$

t_{solve} and memory consumption for $\mathbf{P}$ and $\mathbf{A}$. For consistency, we also compare with mass matrix preconditioning using the mixed discretisation (3.41). We note that where $\mathbf{A}$ shows in any legends, this refers to the discrete version of the operator $\boldsymbol{\mathcal{A}}$ (for example 'assembly $\mathbf{A}$' and '$\mathbf{A}$'), and not the preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$. The different preconditioners used are shown in the x-axis.

The results for single-particle scattering ($M = 1$) are reported in Figures 4.7 and 4.8 for low and high absorbing refractive indices respectively. In this case $\boldsymbol{\mathcal{D}}$ reduces to $\boldsymbol{\mathcal{A}}$, and has therefore been omitted from the figures. As expected, the assembly times and memory costs of $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$, are approximately half those for $\boldsymbol{\mathcal{D}}^e$, $\boldsymbol{\mathcal{D}}^i$ and $\boldsymbol{\mathcal{S}}$ which are approximately half those for $\boldsymbol{\mathcal{A}}$. The solve times for $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ are double those of $\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{D}}^e$ and $\boldsymbol{\mathcal{D}}^i$ for the low absorption case (Figure 4.7), due to the increased GMRES iteration and matvec count indicating a weakened preconditioning effect. This is not the case for high absorption (Figure 4.8), where only the solve time of $\boldsymbol{\mathcal{S}}$ is slightly increased with the other preconditioners $\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{D}}^e$, $\boldsymbol{\mathcal{D}}^i$, $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ requiring roughly equivalent solve time. Even though the GMRES iteration and matvec count have slightly increased for $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$, their sparsity implies a reduced time per iteration compensating for that increase of GMRES iterations.

The assembly time for mass matrix preconditioning of the mixed discretisation, $\boldsymbol{\mathcal{I}}$-mixed, is approximately equivalent to the assembly of the operator $\boldsymbol{\mathcal{A}}$ and the preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, but the memory cost is actually double the magnitude (not shown fully in the figures). To understand this, we recall (4.1) and the discussion of Section 3.2.2. The block operator of (4.1) consists of 8 operators, of which only 4 are different at the continuous level: $\mathcal{C}_1^e$, $\mathcal{C}_1^i$, $\mathcal{S}_1^e$ and $\mathcal{S}_1^i$. We recall that using the discretisation of (3.39)–(3.40), enables us to assemble those 4 operators once and reuse them since they are also identical at the discrete level. In this case, this happens twice: 4 operators are assembled with **RWG** functions for the operator and 4 operators are assembled with **BC** functions for the preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$. In the case of the mixed discretisation (3.41), the 8 operators are different at the discrete level with all of them requir-

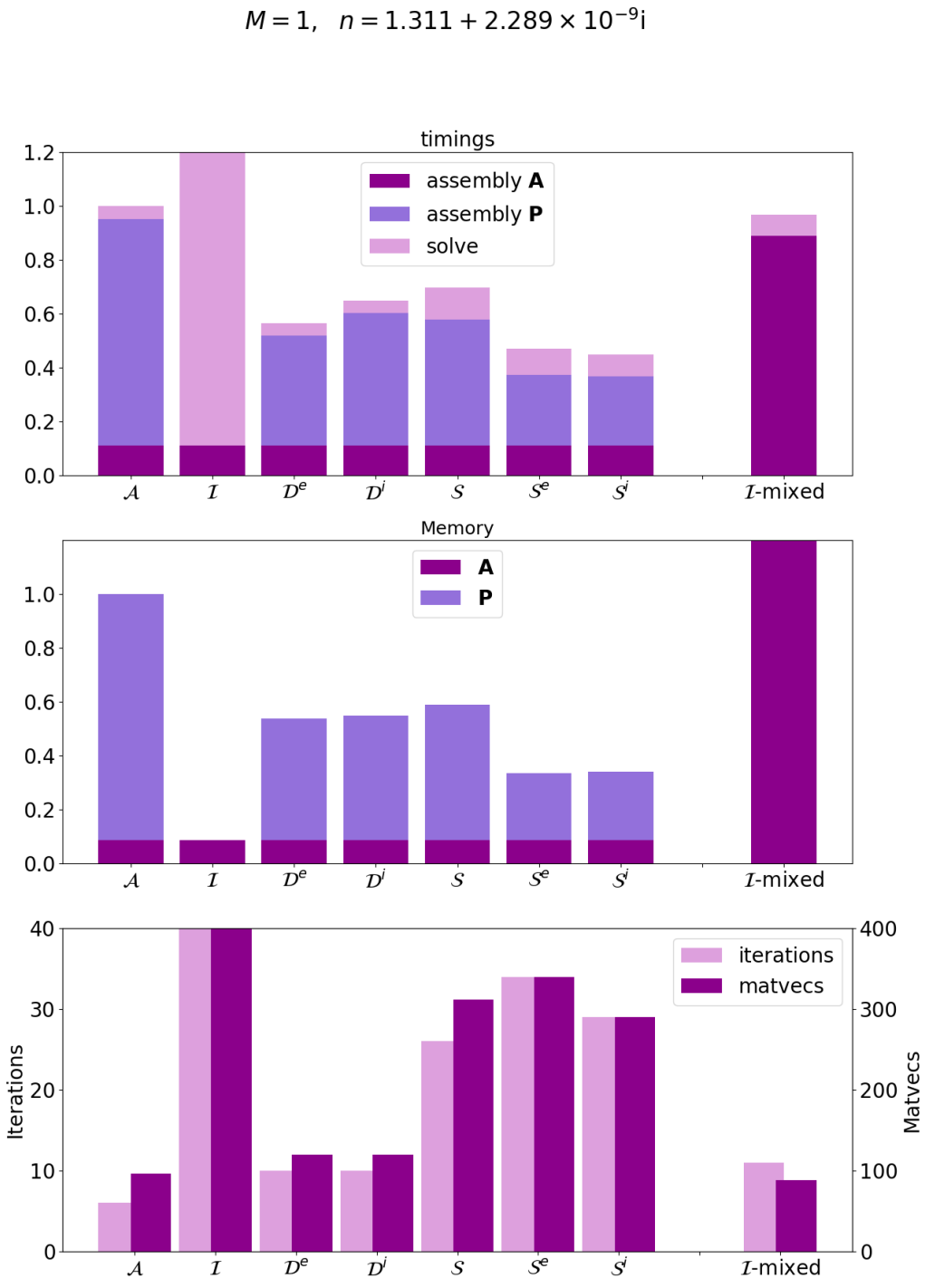


Figure 4.7: Timings and memory costs normalised to the costs for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ and GMRES iteration/matvec count for the strong forms of the reduced preconditioners of (4.27) with discretisation (3.39)–(3.40). $\boldsymbol{\mathcal{I}}$ denotes mass matrix preconditioning using (3.39)–(3.40) and $\boldsymbol{\mathcal{I}}$-mixed denotes mass matrix preconditioning with the mixed discretisation of (3.41). The scattering problem and parameter values are as in Table 4.3, with the refractive index indicated at the top of the figure.

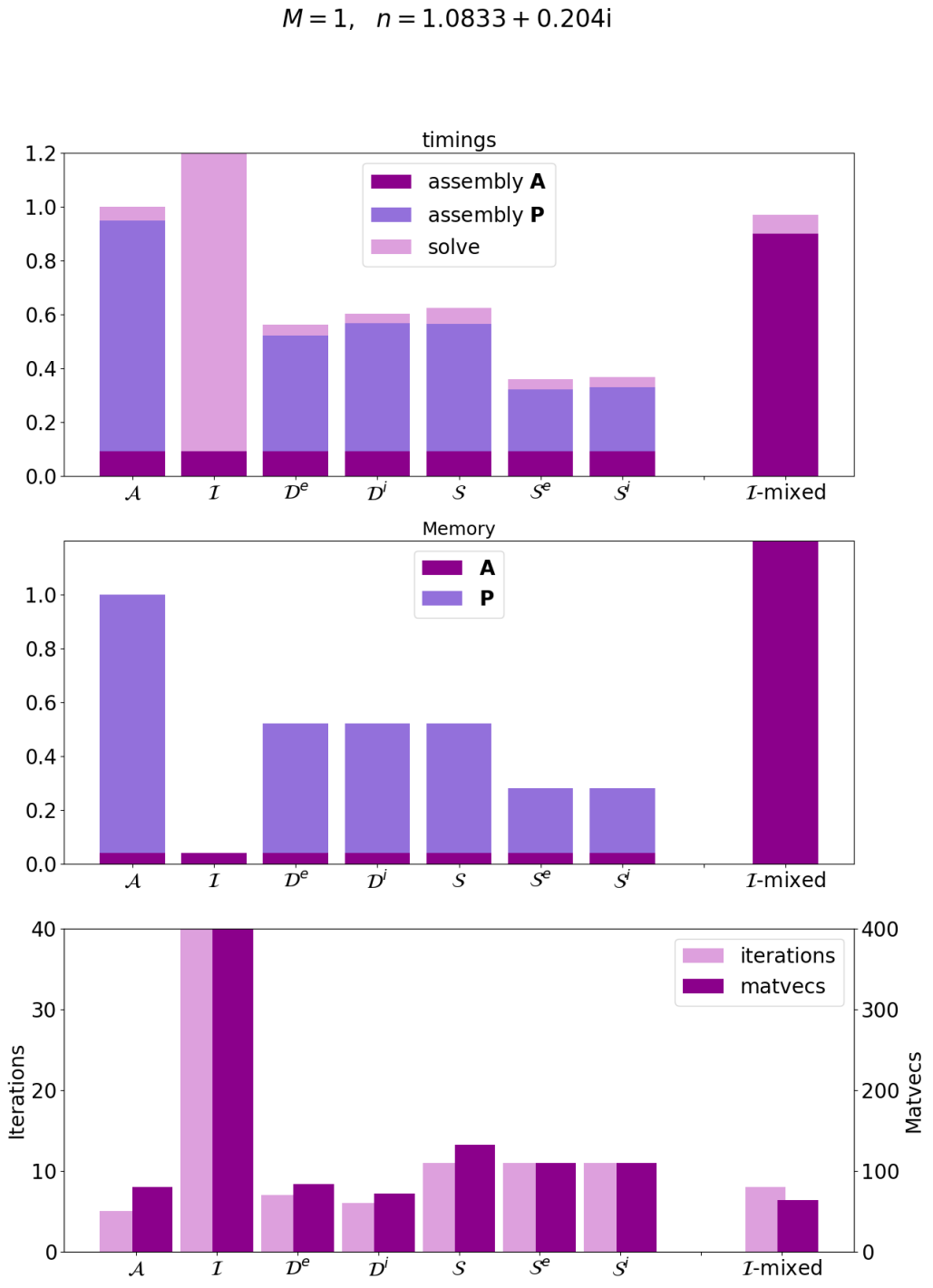


Figure 4.8: Same as in Figure 4.7, but for the refractive index indicated at the top of the figure.

ing a pairing of **RWG** and **BC** functions and needing twice the memory even though no Calderón preconditioning is performed. However, the strong form of $\boldsymbol{\mathcal{I}}$-mixed, is enough to regularise the discrete system and provides the same effect as the traditional Calderón preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ with the discretisation (3.39)–(3.40). This can be observed in the figures, where the total time for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ is roughly equivalent to that of $\boldsymbol{\mathcal{I}}$-mixed. Mass matrix preconditioning with discretisation (3.39)–(3.40), $\boldsymbol{\mathcal{I}}$, does not provide the same preconditioning effect, with the resulting system not converging within the maximum number of GMRES iterations.

A similar behaviour is observed when we move to $M = 3$, in Figures 4.9–4.10. In addition to the preconditioners we have considered for $M = 1$, we now also consider $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{D}}$. As expected, only a small amount of memory savings and assembly time is observed for $\boldsymbol{\mathcal{D}}$. Nevertheless, similar savings are observed for the case of $M = 3$ compared to $M = 1$, with solver time again reducing for high absorption but increasing for low, for the preconditioners $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$. And again, mass matrix preconditioning of the mixed discretisation (3.41), $\boldsymbol{\mathcal{I}}$-mixed, requires a roughly equivalent total time with full Calderón preconditioning of (3.39)–(3.40), but requires double the memory to be assembled. As such, we do not consider the mixed discretisation any further, even though it provides a very nice preconditioning effect in its discrete strong form.

We now wish to compare the behaviour of these reduced preconditioners for a larger range of wavenumbers, similar to what was presented in Figure 4.5. These are shown in Figures 4.11–4.12 for $M = 1$ and low and high absorption respectively, and in Figures 4.13–4.14 for $M = 3$ and low and high absorption respectively. We find it helpful to report memory costs and timings normalised to the values for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ to allow easier comparison between the different preconditioners.

As expected, the memory costs for $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{D}}^e$ and $\boldsymbol{\mathcal{D}}^i$ are approximately half of that for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, and $\boldsymbol{\mathcal{S}}^i$ and $\boldsymbol{\mathcal{S}}^e$ approximately a quarter for $M = 1$. As a reference, we have included the memory of the operator $\boldsymbol{\mathcal{A}}$, which is about

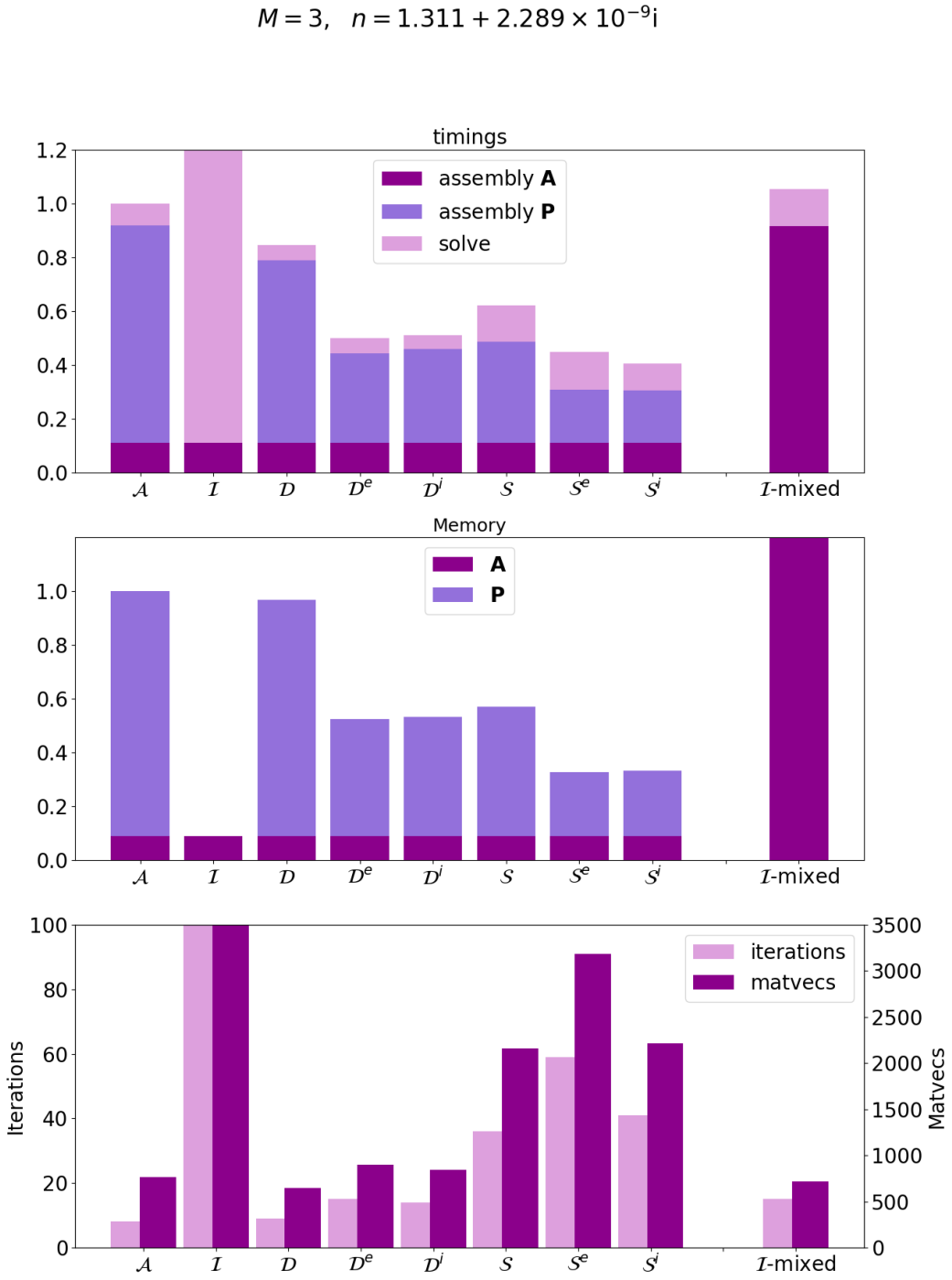


Figure 4.9: Same as in Figure 4.7, but for the scattering problem of Table 4.4 with the refractive index indicated at the top of the figure.

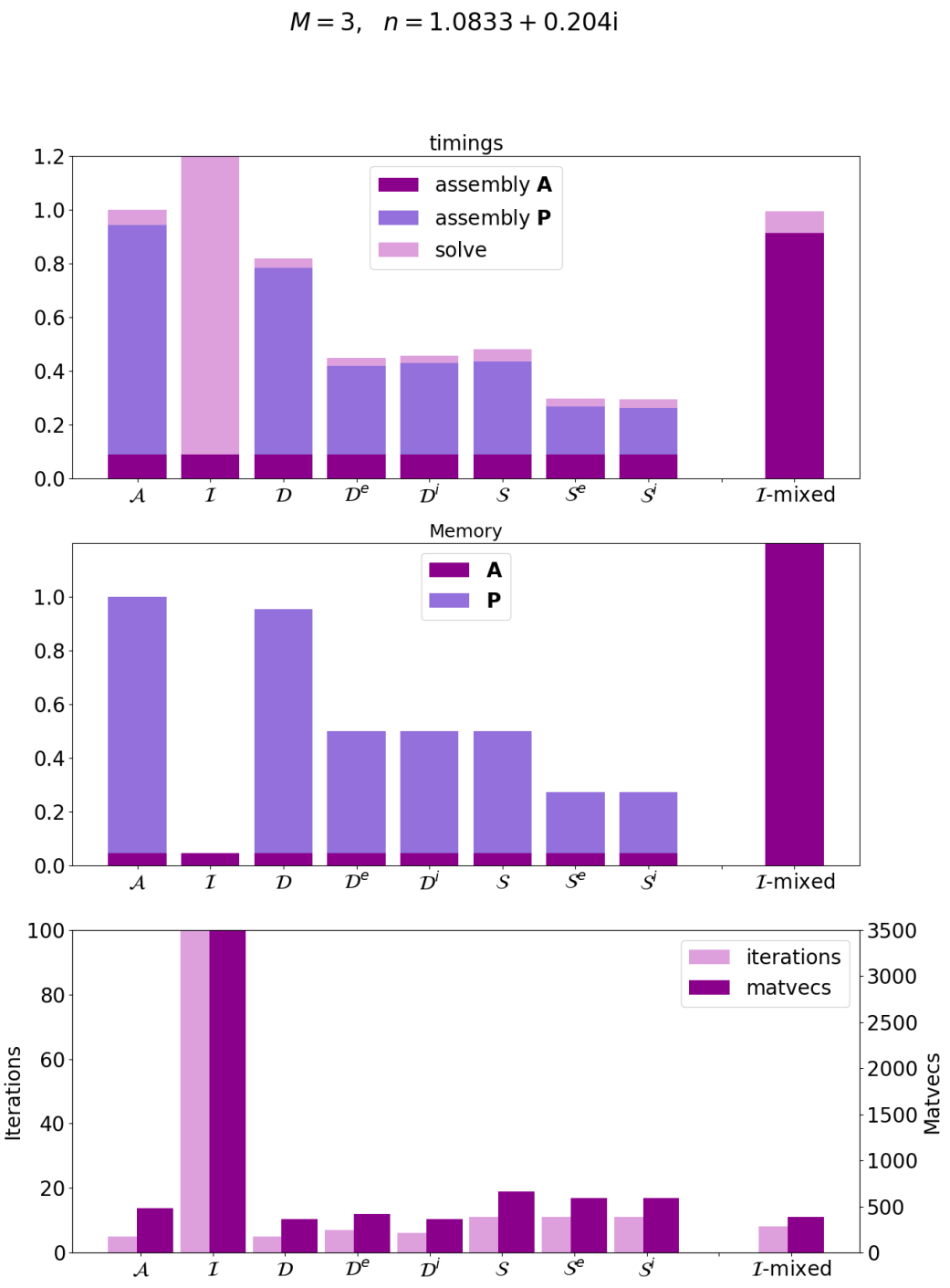


Figure 4.10: Same as in Figure 4.9, but for the refractive index indicated at the top of the figure.

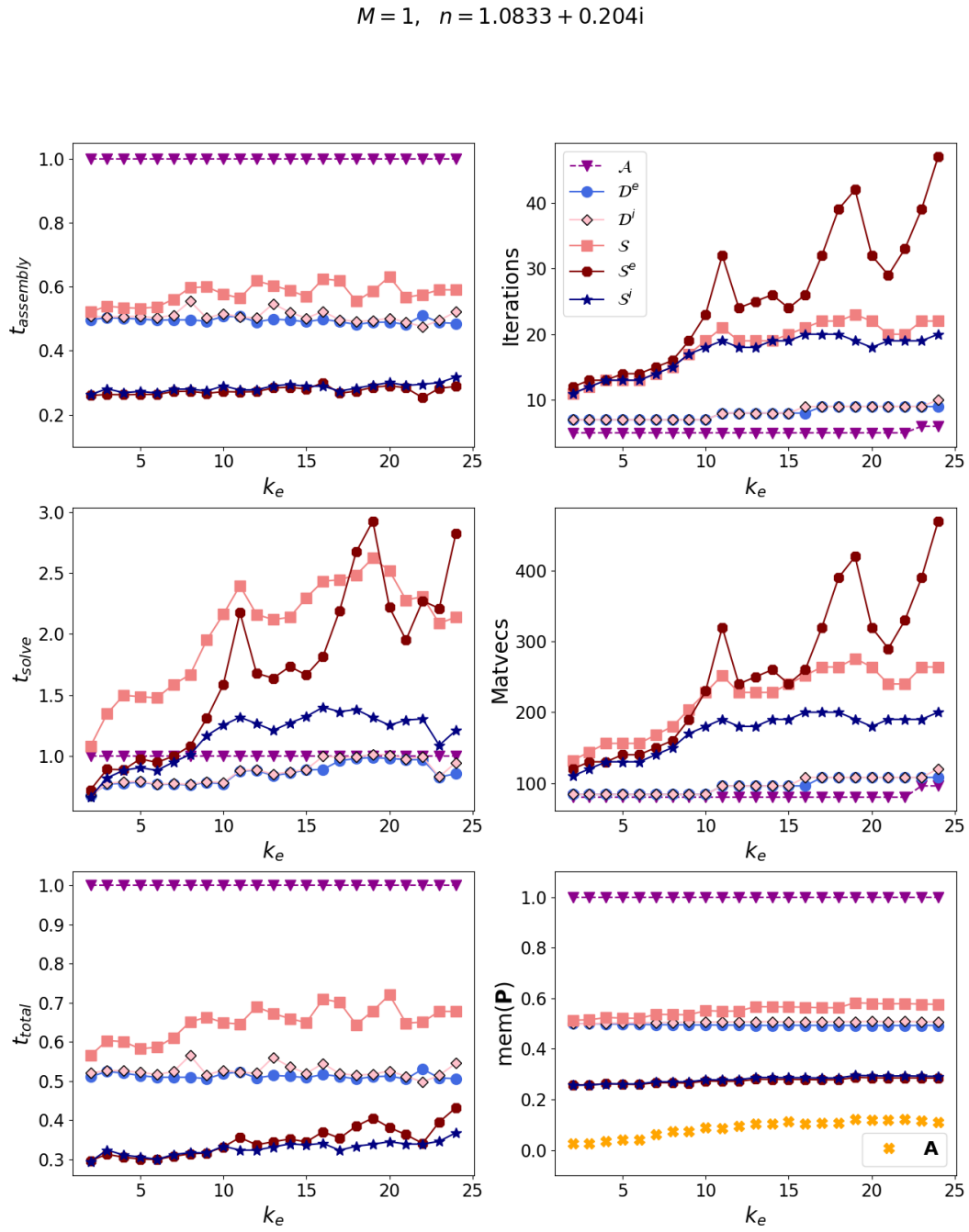


Figure 4.11: Performance of the strong forms of the preconditioners $\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{D}}^e$, $\boldsymbol{\mathcal{D}}^i$, $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ as a function of the exterior wavenumber k_e, for the scattering problem $M = 1$ and the refractive index indicated above. Memory costs and timings are normalised to that of preconditioner $\boldsymbol{\mathcal{A}}$. We also include the memory cost of the operator $\boldsymbol{\mathcal{A}}$ as a reference. All other parameters are as in Figure 4.5.

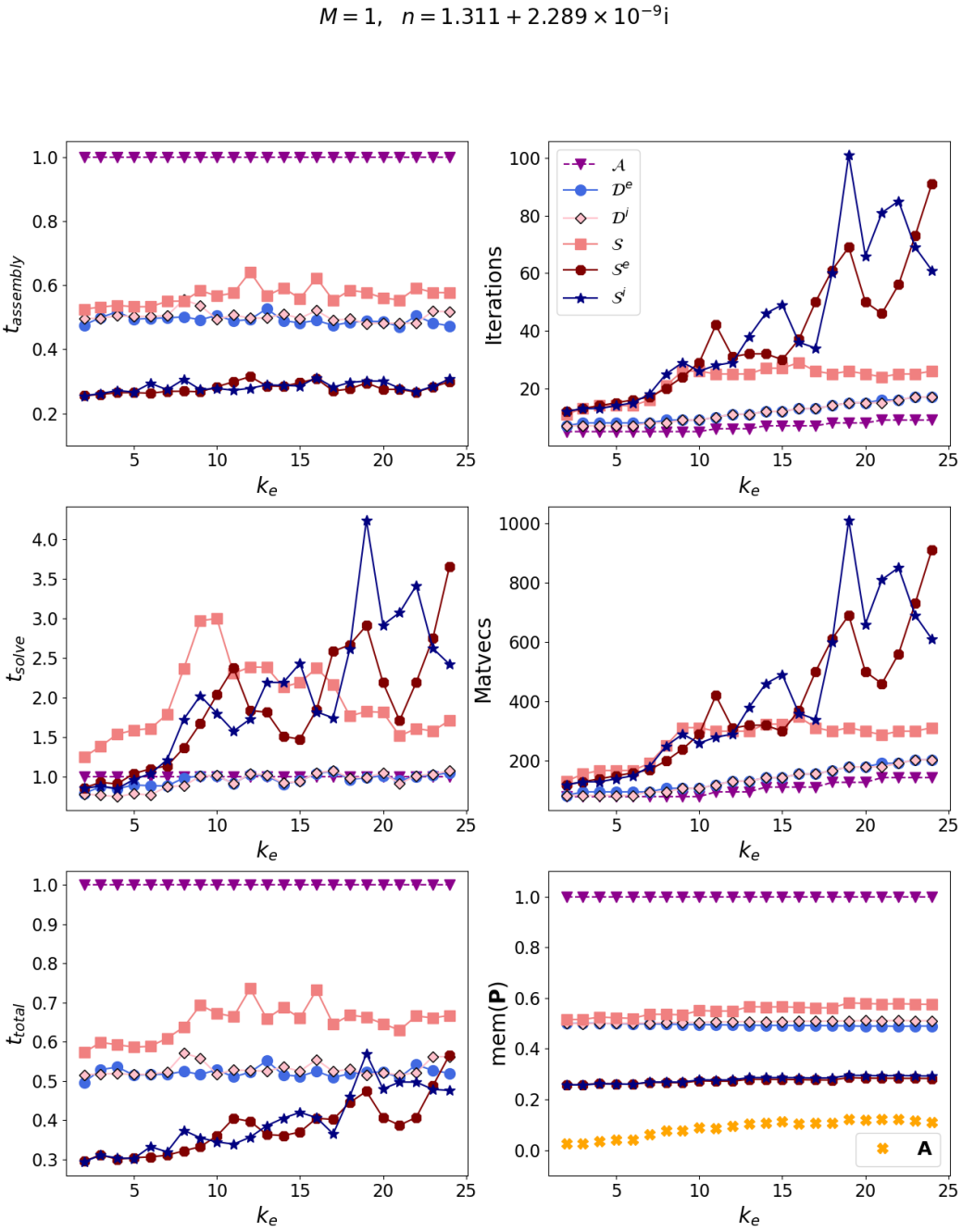


Figure 4.12: As in Figure 4.11 but for the refractive index indicated above.

half of that for $\boldsymbol{\mathcal{S}}^i$ and $\boldsymbol{\mathcal{S}}^e$. Similar trends are observed for the assembly time, although we report assembly time for both the operator and preconditioner, not just for the preconditioner as in Figures 4.7–4.10. The number of GMRES iterations of $\boldsymbol{\mathcal{D}}^e$ and $\boldsymbol{\mathcal{D}}^i$ are roughly double of those for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, with increased matvecs, but t_{solve} is roughly equivalent (for low absorption and larger k_e) and sometimes even less (for high absorption for most of k_e and for low absorption for smaller k_e). Their total time t_{total} is about half that of $\boldsymbol{\mathcal{A}}$, their sparsity again compensating for that GMRES iterations' increase.

The preconditioner $\boldsymbol{\mathcal{S}}$ requires even more GMRES iterations (and matvecs) with t_{solve} being up to 3 times higher and somewhat erratic but again the resulting t_{total} is between 60% - 70% of that for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$. These trends are consistent between the two refractive indices, however this is not the case for preconditioners $\boldsymbol{\mathcal{S}}^i$ and $\boldsymbol{\mathcal{S}}^e$. In the case of low absorption (Figure 4.12) both preconditioners show an erratic behaviour as k_e increases, resulting in increased matvecs and t_{solve}. The total time t_{total} is 30% of that for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ for small k_e, it grows however to about 55% for the larger values of k_e. In the case of high absorption (Figure 4.11) only $\boldsymbol{\mathcal{S}}^e$ exhibits a similar erratic behaviour. The preconditioner $\boldsymbol{\mathcal{S}}^i$, requires a similar number of iterations as $\boldsymbol{\mathcal{S}}$, but fewer matvecs since it contains fewer operators and only up to 1.5 times higher solver time. The total time is only 30 - 40% of that of $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$. We also note that even though $\boldsymbol{\mathcal{S}}^e$ still exhibits an erratic behaviour, it is not as severe as for the weak absorption, resulting in almost the same improvement of t_{total} as $\boldsymbol{\mathcal{S}}^i$, at least for relatively small k_e.

We note that with the use of electric field preconditioners $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ we risk introducing spurious resonances when the wavenumber matches one of the interior electric eigenvalues of the electric field operators (as defined by Buffa and Hiptmair [74]) which could explain the erratic behaviour observed. It seems however that a high absorbing refractive index has a regularising effect, indicated by the smoother behaviour of $\boldsymbol{\mathcal{S}}^i$ in Figure 4.8 compared to that of 4.7. We note that the erratic behaviour might also be related to real

resonances due to the frequencies and observed refractive indices for ice crystals that were used for these simulations. Further investigation would be beneficial in understanding the nature of this behaviour but goes beyond the scope of the thesis.

In the case of $M = 3$ (Figure 4.13–4.14), the memory costs for $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{D}}^e$ and $\boldsymbol{\mathcal{D}}^i$ are approximately half of that for $\boldsymbol{\mathcal{D}}$, and $\boldsymbol{\mathcal{S}}^i$ and $\boldsymbol{\mathcal{S}}^e$ a quarter of that but as expected the memory of $\boldsymbol{\mathcal{D}}$ is only showing a small improvement compared to $\boldsymbol{\mathcal{A}}$. Again, similar trends are observed for the assembly time, with $\boldsymbol{\mathcal{D}}$ now requiring about 70-80% of the assembly time for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$. As we predicted earlier, the preconditioners $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{D}}$ provide the same preconditioning effect and therefore require the same number of GMRES iterations. However, $\boldsymbol{\mathcal{D}}$ results in fewer matvecs and a faster t_{solve}, which in turn together with a faster $t_{assembly}$ imply a faster t_{total}. The preconditioner $\boldsymbol{\mathcal{D}}$ though, provides only a relatively small improvement compared to the other preconditioners. The savings are similar to the single particle case, with $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ exhibiting erratic behaviour again for low absorption, but only $\boldsymbol{\mathcal{S}}^e$ exhibits such behaviour in the case of high absorption.

We have so far considered example cases with 1 or 3 scatterers, and of some significant distance between them. In reality, an ice crystal aggregate consists of multiple monomers with a very small or zero separation between them. For this reason, we consider some final examples where we increase M in which case we expect $\boldsymbol{\mathcal{D}}$ to show greater memory improvement compared to $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$.

In Table 4.5, we compare the reduced preconditioners while increasing M from 4 to 8 and 16 (see Figure 4.3 for the scattering configurations). We omit the number of GMRES iterations and matvecs, as their behaviour is also captured by the GMRES solve time t_{solve}. Confirming our expectations, we see the memory and assembly time savings of $\boldsymbol{\mathcal{D}}$ becoming more significant as M grows. Again to our expectations, the largest memory and assembly time savings are achieved by $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$, with the savings again increasing as

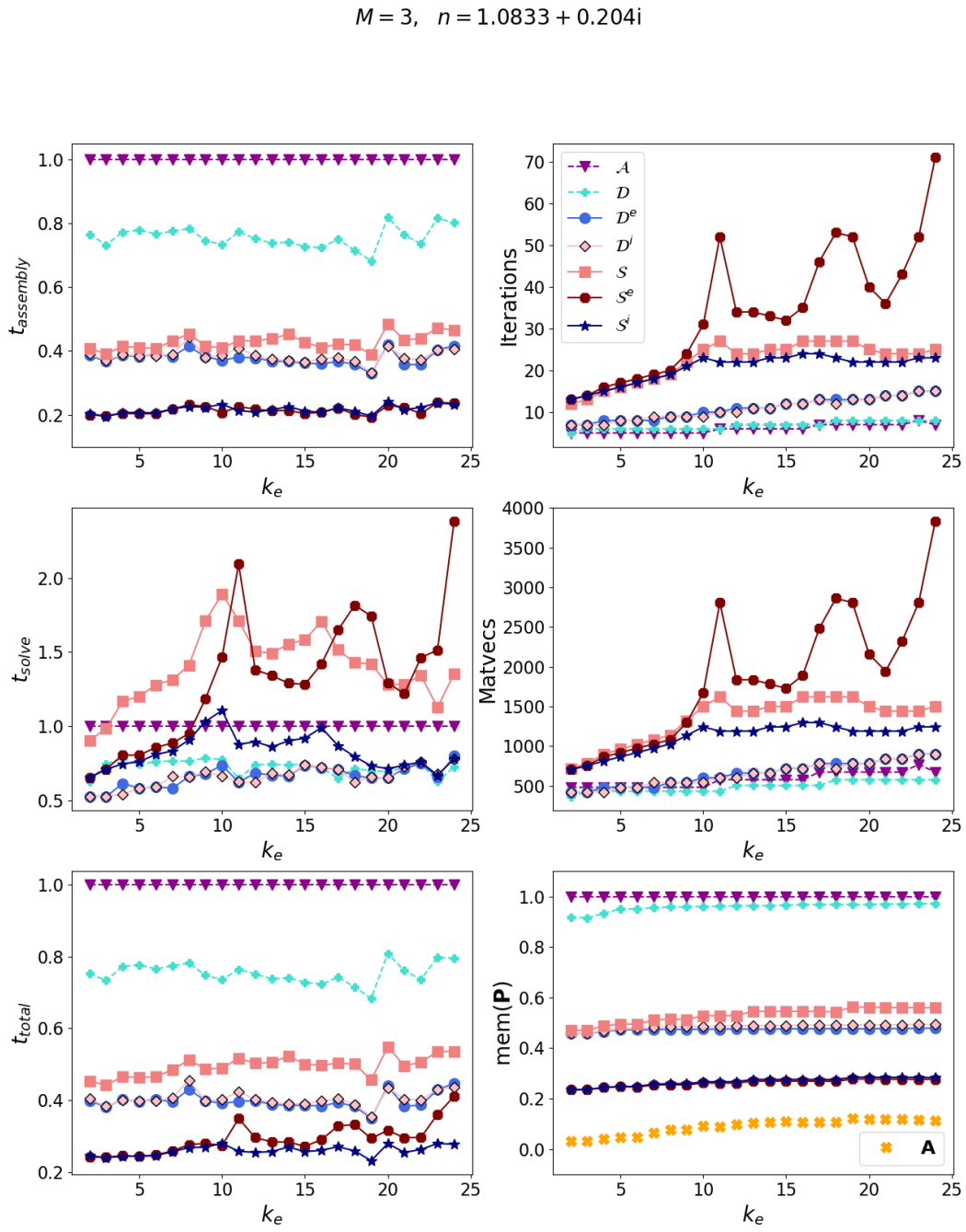


Figure 4.13: As in Figure 4.11, but for the scattering problem $M = 3$, and with the addition of the preconditioner $\boldsymbol{\mathcal{D}}$.

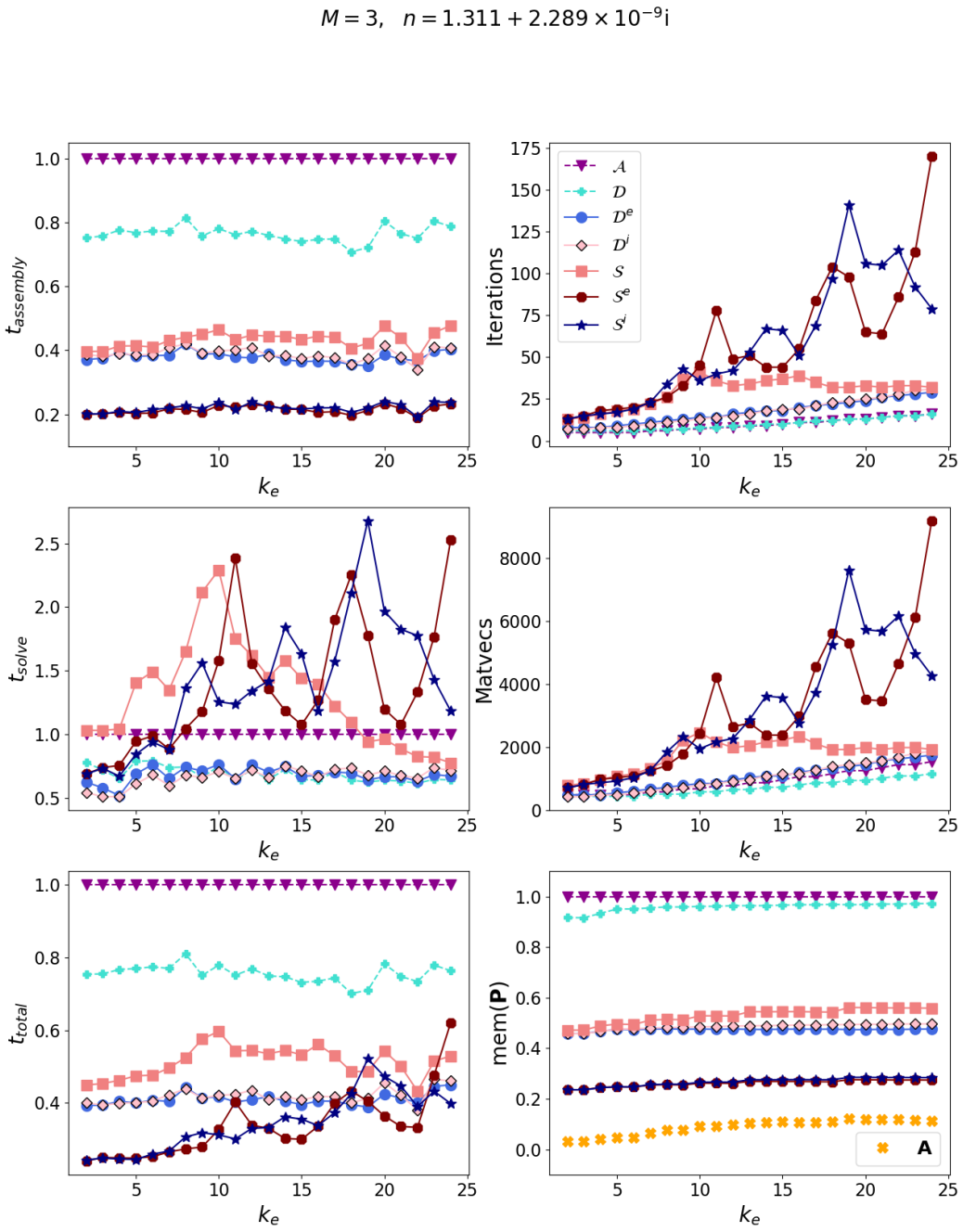


Figure 4.14: As in Figure 4.11, but for the scattering problem $M = 3$ and the refractive index indicated above, and with the addition of the preconditioner $\boldsymbol{\mathcal{D}}$.

	$n = 1.311 + 2.289 \times 10^{-9}\mathrm{i}$											
	Mem($\mathbf{P}$)			$t_{assembly}$			t_{solve}			t_{total}		
$\boldsymbol{\mathcal{P}} \setminus M$	4	8	16	4	8	16	4	8	16	4	8	16
$\boldsymbol{\mathcal{A}}$	1	1	1	1	1	1	1	1	1	1	1	1
$\boldsymbol{\mathcal{D}}$	0.95	0.89	0.80	0.70	0.51	0.31	**0.55**	**0.32**	**0.21**	0.72	0.54	0.37
$\boldsymbol{\mathcal{D}}^e$	0.47	0.44	0.40	0.36	0.25	0.16	0.60	0.34	0.26	0.44	0.34	0.25
$\boldsymbol{\mathcal{D}}^i$	0.48	0.45	0.41	0.37	0.26	0.16	0.60	0.34	0.26	0.45	0.34	0.25
$\boldsymbol{\mathcal{S}}$	0.52	0.49	0.44	0.41	0.29	0.18	1.75	1.15	0.76	0.56	0.45	0.32
$\boldsymbol{\mathcal{S}}^e$	**0.26**	**0.25**	**0.22**	**0.19**	**0.14**	**0.09**	1.72	1.21	0.98	0.39	0.33	0.27
$\boldsymbol{\mathcal{S}}^i$	**0.26**	**0.25**	**0.22**	0.22	**0.14**	**0.09**	1.27	0.82	0.61	**0.37**	**0.30**	**0.24**
	$n = 1.0833 + 0.204\mathrm{i}$											
	Mem($\mathbf{P}$)			$t_{assembly}$			t_{solve}			t_{total}		
$\boldsymbol{\mathcal{P}} \setminus M$	4	8	16	4	8	16	4	8	16	4	8	16
$\boldsymbol{\mathcal{A}}$	1	1	1	1	1	1	1	1	1	1	1	1
$\boldsymbol{\mathcal{D}}$	0.94	0.87	0.77	0.72	0.52	0.34	0.55	0.35	0.22	0.73	0.55	0.39
$\boldsymbol{\mathcal{D}}^e$	0.47	0.43	0.38	0.36	0.26	0.17	0.46	**0.31**	**0.20**	0.43	0.34	0.25
$\boldsymbol{\mathcal{D}}^i$	0.47	0.43	0.38	0.36	0.26	0.17	**0.40**	**0.31**	**0.20**	0.43	0.34	0.25
$\boldsymbol{\mathcal{S}}$	0.47	0.43	0.38	0.37	0.27	0.18	0.72	0.49	0.32	0.46	0.36	0.27
$\boldsymbol{\mathcal{S}}^e$	**0.23**	**0.22**	**0.19**	**0.19**	**0.13**	**0.09**	0.46	0.34	0.24	**0.28**	**0.23**	**0.19**
$\boldsymbol{\mathcal{S}}^i$	**0.23**	**0.22**	**0.19**	**0.19**	**0.14**	**0.09**	0.46	0.35	0.24	**0.28**	**0.23**	**0.19**

Table 4.5: Performance of the strong forms of the preconditioners $\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{D}}^e$, $\boldsymbol{\mathcal{D}}^i$, $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ for an increasing number of scatterers $M = 4$, 8 and 16 from the Model Problems of Figure 4.3. All costs have been normalised to those of the preconditioner $\boldsymbol{\mathcal{A}}$. All other parameters are as in Table 4.2. Minimum values in each column are indicated in bold font.

M grows. The greatest decrease of t_{solve} is achieved by $\boldsymbol{\mathcal{D}}$ in the case of low absorption and $\boldsymbol{\mathcal{D}}^e$ and $\boldsymbol{\mathcal{D}}^i$ in the case of high absorption. But again, the fastest t_{total} is that of $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$, requiring only 24% and 19% at $M = 16$ for low and high absorption respectively.

4.3 Concluding remarks

We conclude this chapter by recalling our goal and emphasising how our main experimental observations have addressed them. We have been looking for some operator $\boldsymbol{\mathcal{P}}$ to precondition the PMCHWT operator so that the discrete matrix system was easier to solve numerically. The preconditioner $\boldsymbol{\mathcal{P}}$ should re-

duce the number of iterations and at the same time have the minimum number of operators required so that it does not incur many additional matrix-vector products, it does not slow down assembly significantly and consumes as little memory as possible. The main observations from our numerical experiments in achieving our goals are:

- Mass matrix preconditioning requires the same computational cost (in terms of matvecs) and GMRES time as traditional Calderón preconditioning when the mixed discretisation of (3.41) is used, making the traditional full Calderón preconditioner unnecessary.

- Strong forms of Calderón preconditioners perform much better (in terms of GMRES iterations, matvecs and solve time) compared to their weak counterparts, so use of weak forms is not recommended.

- Even though the mixed discretisation of (3.41) provides an effective preconditioning effect with its mass matrix, it requires double the memory to assemble the operator compared to the discretisation of (3.39)–(3.40).

- The reduced preconditioners $\boldsymbol{\mathcal{D}}$, $\boldsymbol{\mathcal{D}}^i$, $\boldsymbol{\mathcal{D}}^e$, $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}^i$,$\boldsymbol{\mathcal{S}}^e$ provide some regularising effect, and require less memory and assembly time compared to $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$. Depending on the choice of preconditioner, the solver time was sometimes longer compared to that for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, but the final total time was always shorter. Choosing a preconditioner will depend on different parameters such as the number of scatterers, the wavenumber, refractive index and how many GMRES solves are to be performed. For example

  - $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ result in the biggest savings in memory and assembly time, requiring only 25% compared to those for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$. Their GMRES behaviour though became erratic (large number of iterations and matvecs and longer GMRES time) with increasing k_e. In the case of a high absorbing refractive index, $\boldsymbol{\mathcal{S}}^i$'s behaviour is smoother and therefore should be the preconditioner of choice even for higher

frequency problems, requiring 30-40% of the total time compared to $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$. This potentially allows us to solve higher frequency problems that were not feasible before due to memory constraints.

- When multiple GMRES solves are required (which is the case when simulating random orientation, discussed later in Chapter 7), the priority is to choose a preconditioner that requires the same or even less time t_{solve} compared to that for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$. This is the case when preconditioners $\boldsymbol{\mathcal{D}}^e$ or $\boldsymbol{\mathcal{D}}^i$ are used.
- We do not recommend the use of preconditioner $\boldsymbol{\mathcal{S}}$. Its memory and assembly time savings are equivalent to those for $\boldsymbol{\mathcal{D}}^e$ and $\boldsymbol{\mathcal{D}}^i$, but its GMRES behaviour is unstable, and therefore with the same memory and assembly time savings one can achieve a reduced total time with $\boldsymbol{\mathcal{D}}^e$ and $\boldsymbol{\mathcal{D}}^i$.

In the next chapter, we attempt to reduce computational cost and timings even further by looking at the implementation of the operator and preconditioner.

Chapter 5

Accelerated Calderón Preconditioning II: bi-parametric implementation

We recall again (from Chapter 1) that we are looking for some operator $\boldsymbol{\mathcal{P}}$ applied to the PMCHWT formulation such that $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$ has better properties than $\boldsymbol{\mathcal{A}}$ and therefore its discrete version is easier to solve numerically. The computational cost of solving the preconditioned system is affected by

(i) the choice of preconditioning operator $\boldsymbol{\mathcal{P}}$, and

(ii) the choice of discretisation for the operator product $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$.

In this chapter we focus on (ii) by adjusting its implementation. We recall that $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$ should be as cheap (memory and time-wise) as possible, while still producing a sufficiently accurate numerical solution. We follow a so-called "bi-parametric" implementation to achieve a reduced memory cost, assembly time and overall time (assembly and GMRES solve time).

We discuss the basics of our bi-parametric implementation in Section 5.1, presenting potential savings and the resulting spectrum of the discrete operator product for different bi-parametric implementations. In Section 5.2, we investigate the performance of the bi-parametric implementation in the context of iterative solvers. We first present results for the traditional Calderón

preconditioner, $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, and then for the reduced preconditioners introduced in (4.26)–(4.30) of Chapter 4. Concluding remarks are given in Section 5.3.

To summarise our findings, we demonstrate that a bi-parametric implementation with the discretisation of (3.39)–(3.40), and parameter choice of (5.3)–(5.4) (with the notation defined later in (5.1)–(5.2)), can alleviate the prohibitive memory cost imposed by the **BC** functions. A reduction in memory cost of up to 80% was observed for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, with the memory cost of $\mathbf{P}$ being roughly equivalent with that of the operator $\mathbf{A}$ at higher frequencies. Further, a bi-parametric implementation in combination with the reduced preconditioners of (4.27), resulted in a memory cost of just a few percent of the non bi-parametric traditional Calderón preconditioner, costing even lower than the cost of the operator $\mathbf{A}$. To be precise, a bi-parametric implementation of $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{S}}^i$, with the discretisation of (3.39)–(3.40), and parameter choice of (5.3)–(5.4), required only 2-3% of the memory of the non bi-parametric implementation of $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, with a total time (for the assembly of both the operator and preconditioner and GMRES time) around 20% of that. In addition, the erratic behaviour (increased iterations, matvecs and GMRES time) that was observed for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{S}}^i$ in Chapter 4, is no longer present, making this our preconditioner and implementation of choice for high frequency problems. The results presented in this chapter have been reported in [3].

5.1 Bi-parametric implementation

The idea behind the bi-parametric implementation is that one can use different values for the parameters used in the assembly of $\mathbf{P}$ and $\mathbf{A}$ (for assembly details we refer back to Section 3.3, Chapter 3). For this thesis, these are the accuracy of the ACA algorithm, ν, the near-field cutoff parameter, χ, and the quadrature orders, $\mathbf{q}$ (defined in (3.44)–(3.45), (3.46) and (3.48) respectively). Our expectation is that accuracy of the matrix assembly should be more important for $\mathbf{A}$ than for $\mathbf{P}$, since $\mathbf{A}$ governs overall solution accuracy whereas $\mathbf{P}$ is included merely to accelerate the convergence of the iterative solver. We

adopt the following notation to distinguish the set of parameters:

$$(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}), \tag{5.1}$$

for $\mathbf{P}$, and

$$(\nu_{\mathbf{A}}, \chi_{\mathbf{A}}, \mathbf{q}_{\mathbf{A}}), \tag{5.2}$$

for $\mathbf{A}$. We expect to be able to use a larger value of $\nu_{\mathbf{P}}$, and smaller values of $\chi_{\mathbf{P}}$ and $\mathbf{q}_{\mathbf{P}}$, compared to $\nu_{\mathbf{A}}$, $\chi_{\mathbf{A}}$ and $\mathbf{q}_{\mathbf{A}}$, saving computation time and memory. In combination with the reduced preconditioners introduced in the previous chapter, we will show how this bi-parametric approach can reduce memory cost to a few percent of the traditional Calderón preconditioner with a non bi-parametric implementation.

We now recall the discussion from Section 3.2.2 of Chapter 3 regarding the choice of discretisation for the operator product. Using the mixed discretisation of (3.41), where the same discrete spaces are used for both $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{A}}$, allows us to obtain $\mathbf{P}$ for 'free' once $\mathbf{A}$ is assembled, as one can extract the relevant operators from $\mathbf{A}$ to create $\mathbf{P}$. For (3.39)–(3.40), different spaces are used for $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{A}}$. However, identical occurrences of the same operator at the continuous level, give rise to the same discrete operator at the discrete level, allowing us to reduce assembly cost. This is not possible for (3.41).

For a bi-parametric implementation it makes sense to use the discretisation scheme of (3.39)–(3.40), where we use **RWG** basis functions for the operator and **BC** functions for the preconditioner. The cheaper parameter values for $\mathbf{P}$ can then mitigate the cost of the expensive **BC** functions. Using the mixed discretisation of (3.41) does not make sense in the case of the bi-parametric implementation since both the operator and preconditioner would require the use of **BC** functions, but since they would be assembled with two different sets of parameters, one would no longer be able to take advantage of the 'free' preconditioner offered by this discretisation scheme. Therefore,

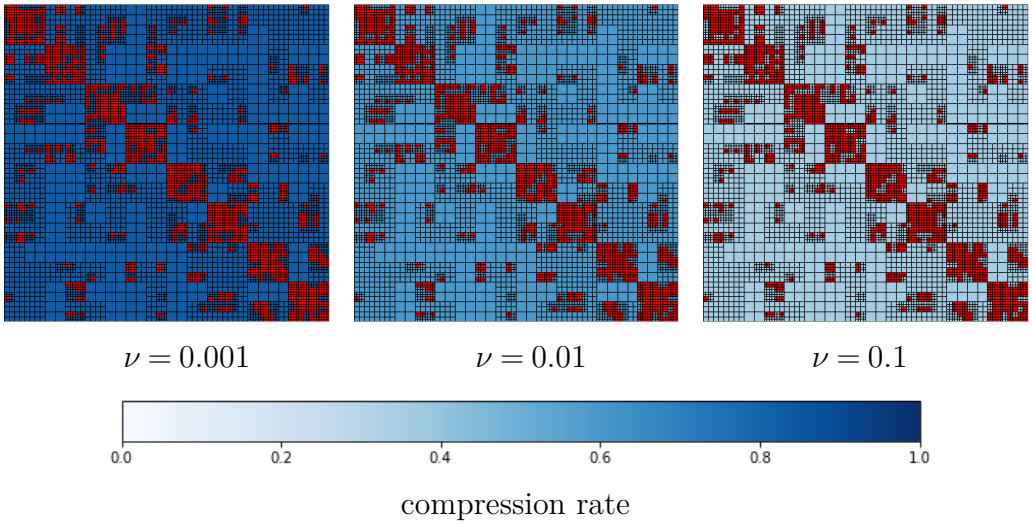
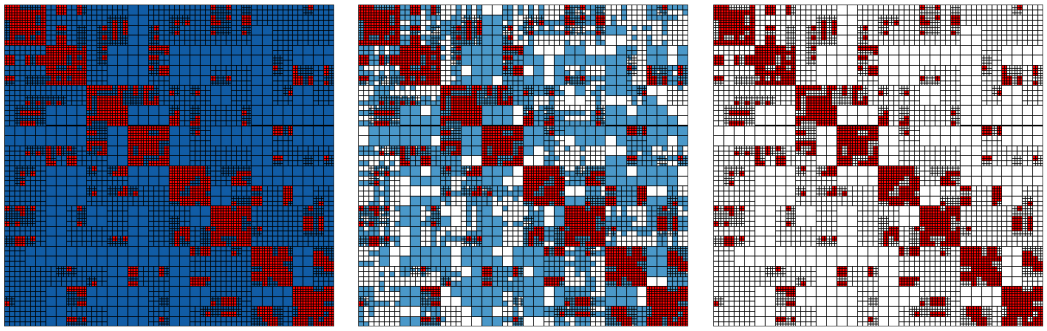


Figure 5.1: Block cluster trees produced by Bempp in the assembly of an electric field operator $\mathcal{S}$ on the unit cube with $k = 5$ and a mesh with approximately 10 elements per wavelength. The red indicates inadmissible blocks, and blue indicates admissible blocks that have been approximated by ACA. The compression rate is indicated by the blue scale for an increasing ACA parameter $\nu = 0.001, 0.01$ and 0.1 (left to right) with a cut-off parameter $\chi = \infty$, with the values of the compression rates being: 0.83, 0.59 and 0.36 respectively.

all numerical experiments presented in this chapter follow the discretisation (3.39)–(3.40), unless specified otherwise.

To illustrate the potential savings of this bi-parametric approach we present Figures 5.1–5.2. Figure 5.1 illustrates the influence of ν on the compression rate, defined as the storage divided by the storage that would be required in a dense assembly. In Figure 5.1, we observe a reduced compression rate as ν gets larger, indicating a reduced memory cost and faster assembly time. We present a relatively low-frequency example, to limit the dofs and make the block structure easier to see; for higher-frequency problems the compression rates improve significantly. Figure 5.2 illustrates the effect of the near-field cut-off parameter χ on the resulting assembled matrix. As expected, the overall compression rate reduces as we remove more and more of the far-field.

Regarding quadrature, for non-singular integrals, reducing the quadrature orders to the minimum possible $(q_{near}, q_{medium}, q_{far}) = (1, 1, 1)$ results in $(1, 1, 1)$ integrand evaluations respectively. For singular integrals, $q_{singular} = 1$ results in 2 integrand evaluations if the two triangles share a single vertex, 5



$\chi = \infty$ $\chi = 0.5$ $\chi = 0$

Figure 5.2: Same as in Figure 5.1 but with ACA parameter $\nu = 0.001$ and decreasing cut-off parameter χ from ∞ to 0.5 and 0. White blocks indicate admissible blocks that do not require assembly. Compression rates from left to right: 0.83, 0.60, and 0.14.

if they share a common edge and 6 if they coincide reducing computational time during kernel evaluations. We recall from Section 3.3.2, Chapter 3, that for non-singular integral the default values $(q_{near}, q_{medium}, q_{far}) = (4, 3, 2)$ correspond to $(36, 16, 9)$ integrand evaluations respectively. For singular integrals the default value $q_{singular} = 6$ corresponds to 512 integrand evaluations if the two triangles share a common vertex, 1280 if they share a single common edge, and 1536 if they coincide. Big savings in assembly time are therefore expected if we reduce the quadrature rules of the preconditioner to the minimum possible.

We note however that the preconditioning effect of $\mathbf{P}$ is expected to weaken as we decrease its accuracy, potentially increasing the number of iterations required for convergence. To see the potential effects of a bi-parametric implementation we present spectrum plots of $\boldsymbol{\mathcal{A}}^2$, for $M = 1$, for different values of the parameters $\nu_{\mathbf{P}}$ in Figure 5.3, $\chi_{\mathbf{P}}$ in Figure 5.4, $\mathbf{q}_{\mathbf{P}}$ in Figure 5.5 and a combination of those in Figure 5.6.

In Figure 5.3, we see no obvious change in the spectrum when $\nu_{\mathbf{P}}$ is increased from 0.001 to 0.01. However, for $\nu_{\mathbf{P}} = 0.1$ and 0.5 the eigenvalues spread further and further away from the accumulation points indicated with red, but we note that the spectrum remains clustered away from zero and does not seem to accumulate at infinity.

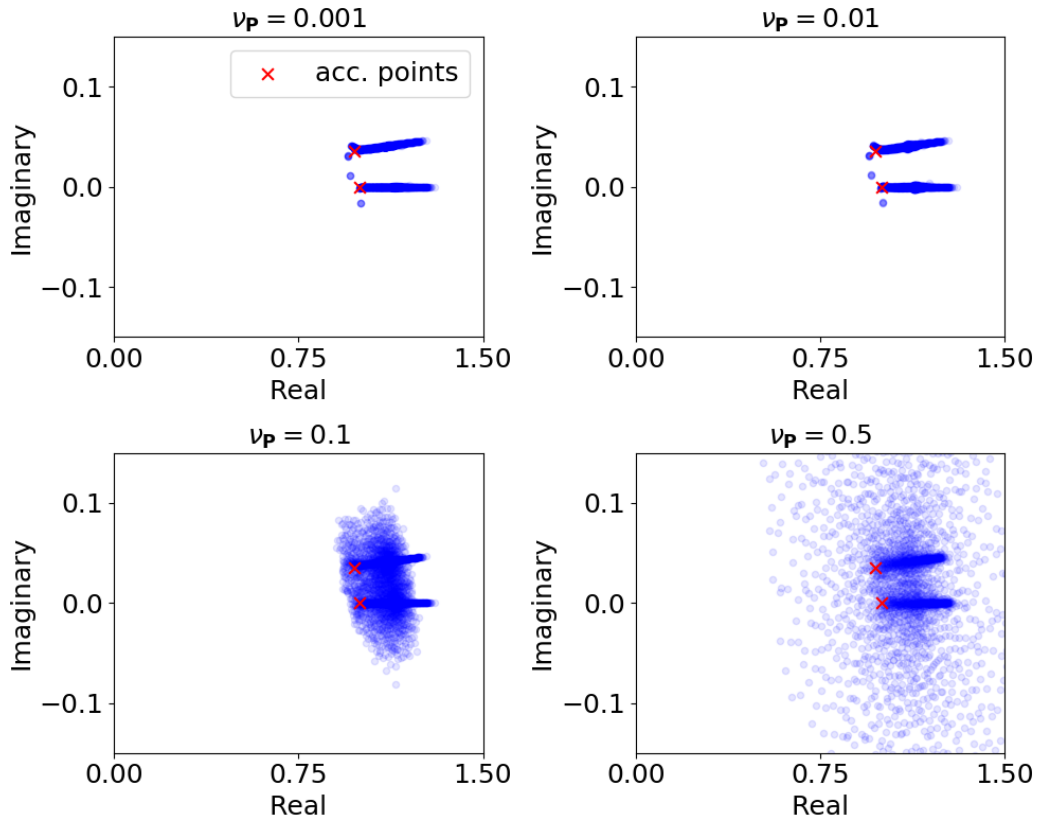


Figure 5.3: Spectrum of $\mathcal{A}^2$ for different $\nu_{\mathbf{P}}$ values indicated at the top of each plot, with $(\chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (\chi_{\mathbf{A}}, \mathbf{q}_{\mathbf{A}}) = (\infty, (4,3,2,6))$ and $\nu_{\mathbf{A}} = 0.001$. The scatterer is a unit sphere centered at the origin with refractive index $n_1 = 1.0833 + 0.204\mathrm{i}$. The exterior wavenumber is $k_e = 1.5$ and the interior is given by $k_1 = n_1 k_1$. The sphere was discretised with a mesh width of approximately 20 elements per wavelength. The red crosses indicate the theoretical accumulation points of (4.3)–(4.4).

In Figure 5.4, all three finite values of $\chi_{\mathbf{P}}$ result in the spectrum spreading away from the accumulation points, but for $\chi_{\mathbf{P}} = 1$ this spread is only mild. Again, the spectrum is clustered away from zero and infinity.

In Figure 5.5, only the change in quadrature order for singular kernels alters the spectrum (middle top and right bottom figures). The remaining cases do not show any obvious changes. Again, the spectrum seems clustered away from zero and infinity.

Finally, we combine all cheap parameters together, with our proposed choice being $(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (0.1, 0, (1,1,1,1))$. This is because we aim to minimise assembly costs and therefore choose the cheapest values possible that have not shown a large spread of the eigenvalues in Figures 5.3–5.5. We plot

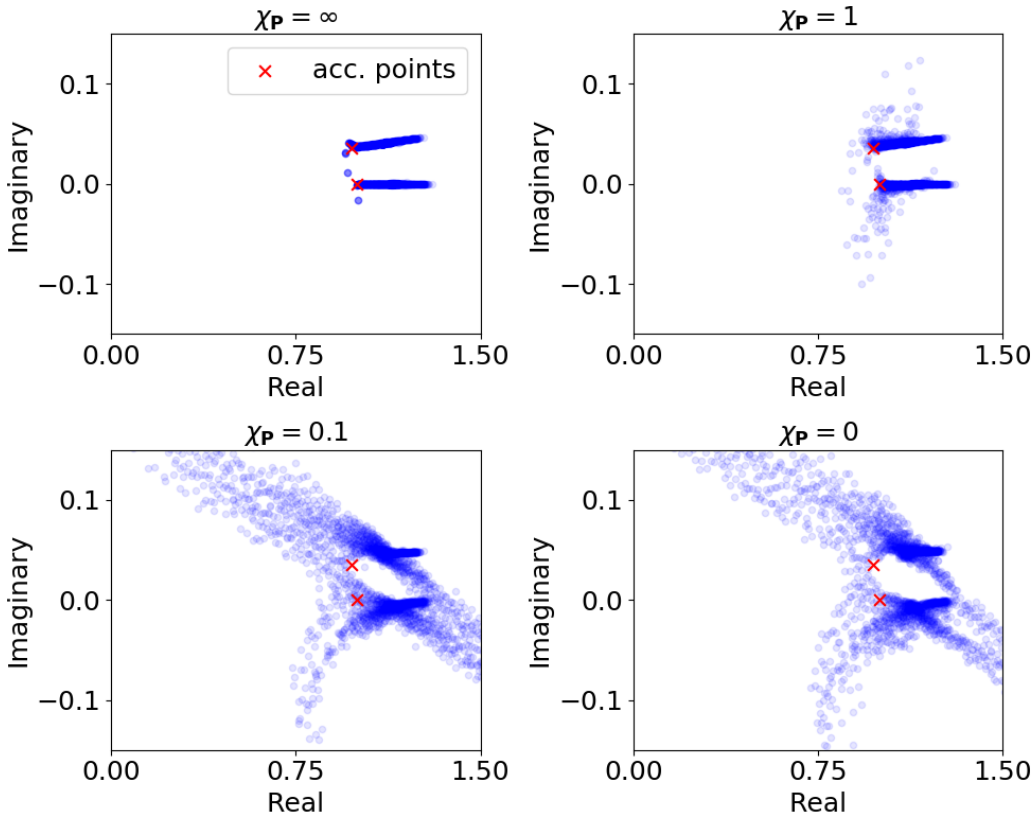


Figure 5.4: As in Figure 5.3 but for different $\chi_{\mathbf{P}}$ values indicated at the top of each plot, with $(\nu_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (\nu_{\mathbf{A}}, \mathbf{q}_{\mathbf{A}}) = (0.001, (4,3,2,6))$ and $\chi_{\mathbf{A}} = \infty$.

the spectrum in Figure 5.6. We notice that the spectrum now shifts further away from the accumulation points but seems to be clustered away from the origin and infinity. The figure suggests a weakened preconditioning effect, since the discretised operator product is now only a cheap approximation of the original operator product at the continuous setting. The weakened preconditioning effect could potentially increase the number of iterations and slow down solution time, which we aim to investigate in the following section.

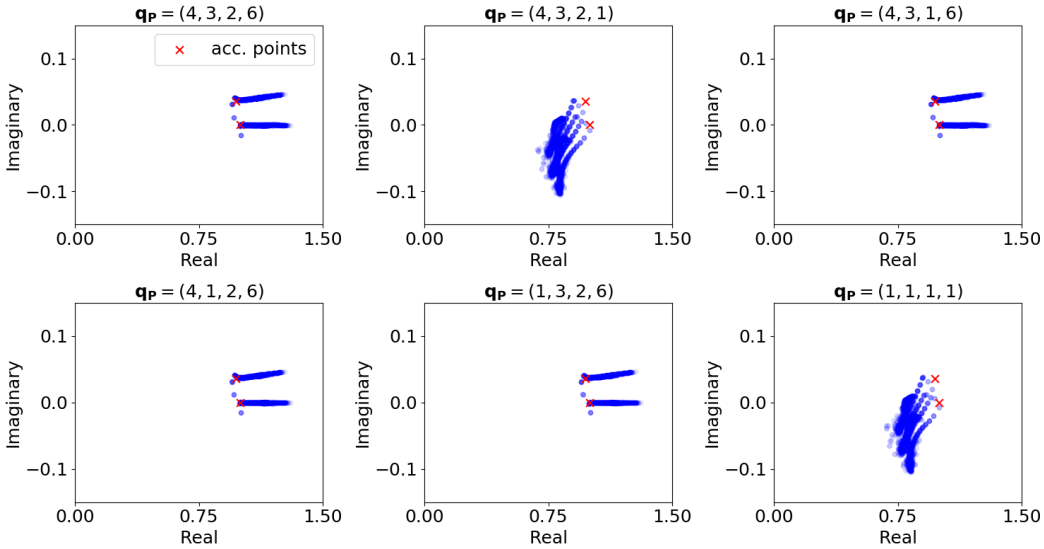


Figure 5.5: As in Figure 5.3 but for different $\mathbf{q_P}$ values indicated at the top of each plot, with $(\nu_\mathbf{P}, \chi_\mathbf{P}) = (\nu_\mathbf{A}, \chi_\mathbf{A}) = (0.001, \infty)$ and $\mathbf{q_A} = (4, 3, 2, 6)$.

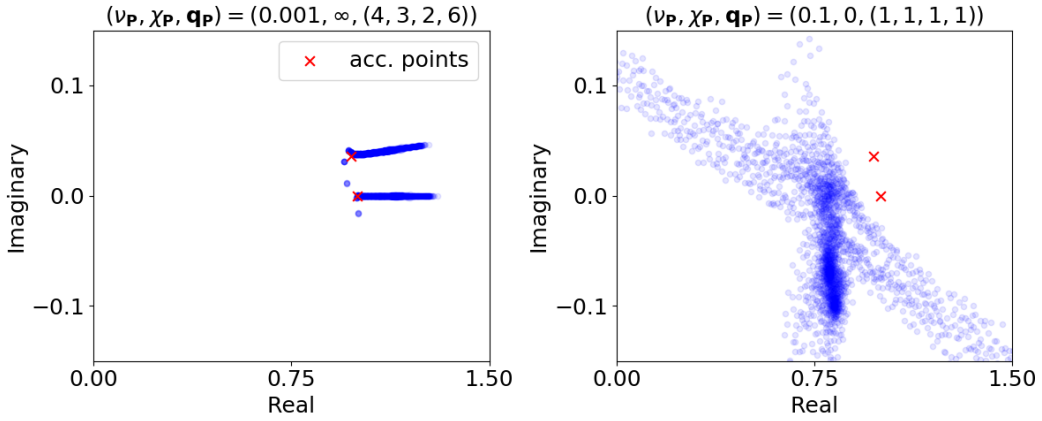


Figure 5.6: As in Figure 5.3 but for different $(\nu_\mathbf{P}, \chi_\mathbf{P}, \mathbf{q_P})$ values indicated at the top of each plot with $(\nu_\mathbf{A}, \chi_\mathbf{A}, \mathbf{q_A}) = (0.001, \infty, (4, 3, 2, 6))$. The discretised product in the right figure is only a cheap approximation of the operator product in the continuous setting explaining the spectrum being further away from the accumulation points indicated in red.

5.2 Numerical Experiments

In the following sections we investigate the performance of the bi-parametric preconditioners in terms of assembly time, memory consumption, solution time, iterations, matvecs and total time (for the assembly of both the operator and preconditioner, and GMRES time). The spectral plots of the previous section suggest a weakened preconditioning effect as we decrease the accuracy of the preconditioner potentially increasing the number of iterations and slowing solution time. We note however, that the examples shown in Figures 5.1–5.2 along with examples mentioned within the text regarding quadrature orders suggest a faster assembly, with a reduced memory cost along with an increased sparseness of the discrete preconditioner. A matrix vector product with a sparser preconditioner is faster, so even if the number of iterations increase, this does not necessarily mean an increased GMRES time. We investigate the performance of the preconditioned operator for a range of cheap parameters $(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}})$ in the following sections, first with the traditional preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, and then with the reduced preconditioners of (4.26)–(4.30), introduced in Chapter 4.

5.2.1 Traditional Calderón preconditioning

We first explore the potential memory and time savings by considering the model problems $M = 1$ and $M = 3$ from Chapter 4, for the two refractive indices and their corresponding wavenumbers (can be found in Table 4.2). We only consider the traditional Calderón preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ for now.

In Table 5.1, we report results for the case $M = 1$ for the two different refractive indices and their corresponding wavenumbers. The table is separated into five different 'areas'. The top row corresponds to a non bi-parametric implementation with

$$(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (\nu_{\mathbf{A}}, \chi_{\mathbf{A}}, \mathbf{q}_{\mathbf{A}}) = (0.001, \infty, (4,3,2,6)).$$

The memory costs and timings of any implementations that follow have been

normalised relative to those of the non bi-parametric implementation.

In the next six rows, we keep $\nu_{\mathbf{P}} = 0.001$ and $\chi_{\mathbf{P}} = \infty$ fixed and reduce the quadrature orders to the minimum, first one by one and then all four of them. We observe that the performance depends on the frequency (and refractive index) considered. In the left sub-columns (which correspond to a low absorption but higher wavenumber) we can see that the number of iterations (and hence matvecs and solve time) are not really affected by the change in quadrature orders. On the other hand, in the right sub-columns (which correspond to a high absorption but a very small wavenumber) we can see that the quadrature order for singular integrals affects the quality of the resulting linear system with a large increase in iterations (and matvecs and solver time). For both wavenumbers, the memory is unaffected while assembly time reduces, with the largest impact observed when one reduces the quadrature orders for singular integrals or integrals in the 'near-field' area. Minimising all quadrature orders results in a total time (which includes assembly time for both $\mathbf{A}$ and $\mathbf{P}$ and GMRES time) that is 50% of the non bi-parametric version for the high wavenumber (left sub-columns), while no savings are observed for the low one (right sub-columns).

In the next three rows we keep $\chi_{\mathbf{P}} = \infty$ and $\mathbf{q_P} = (4, 3, 2, 6)$ fixed and increase $\nu_{\mathbf{P}}$ from 0.001 through 0.01, 0.1 and 0.5. We observe a significant reduction in memory and assembly time, with little or no effect on iteration/matvec count and solver time for the higher frequency considered. This is not the case for the lower frequency, where for $\nu_{\mathbf{P}} = 0.1$ and $\nu_{\mathbf{P}} = 0.5$ the resulting linear system did not converge to the desired tolerance within the maximum number of iterations prescribed.

For the results in the next three rows we keep $\nu_{\mathbf{P}} = 0.001$ and $\mathbf{q_P} = (4, 3, 2, 6)$ fixed and reduce $\chi_{\mathbf{P}}$, which corresponds to neglecting more and more of the far-field behaviour in the preconditioner. As $\chi_{\mathbf{P}}$ reduces to 0.1 and then 0, we see a noticeable reduction in memory cost and assembly time. Unfortunately, this comes at the cost of a significant increase in iteration/matvec

count and hence solve time. This increase appears to be more serious for the low frequency case suggesting that the far-field behaviour is more important for low frequency problems and that neglecting it may be possible only for higher frequencies.

In the final row, we take $\nu_{\mathbf{P}} = 0.1$, $\chi_{\mathbf{P}} = 0$ and $\mathbf{q}_{\mathbf{P}} = (1,1,1,1)$ and again see that the behaviour is dependent on the frequency considered, with the lower frequency not converging. For the higher frequency we see a significant increase in iterations (along with matvecs and solve time) due to completely neglecting the far-field interactions. We achieve a remarkable saving in memory and assembly time bringing them to 17% and 16% of the top row costs respectively. The total time is also reduced to 44% despite the increased solve time, owing to the large decrease in assembly time.

In Table 5.2, we repeat the above but for the case $M = 3$. Similar observations are made compared to the single-particle examples, confirming the frequency dependence of the performance when the far-field is removed or when the parameter $\nu_{\mathbf{P}}$ is increased.

To investigate the possible frequency dependence, we present performance statistics in Figures 5.7–5.8 and 5.9–5.10 for $M = 1$ and $M = 3$ respectively for a range of wavenumbers, and a range of bi-parametric implementations with different far-field cutoff values $\chi_{\mathbf{P}}$. Results are compared to a non bi-parametric implementation with

$$(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (\nu_{\mathbf{A}}, \chi_{\mathbf{A}}, \mathbf{q}_{\mathbf{A}}) = (0.001, \infty, (4,3,2,6)),$$

(presented as purple triangles) and timings and memory costs of the bi-parametric implementations have been normalised relative to that.

The results from Figures 5.7–5.8 confirm that for both refractive indices the significant increase in iteration/matvec count for small values of $\chi_{\mathbf{P}}$ is more important in low frequency cases. This behaviour is also reflected in the total time, where for the lowest k_e considered the total time with a reduced $\chi_{\mathbf{P}}$ is longer than the reference case. For all other wavenumbers, the total time of

$\chi_{\mathbf{P}}$	$\nu_{\mathbf{P}}$	$\mathbf{q_P}$	Iters (Matvecs)		Mem($\mathbf{P}$)		$t_{assembly}$		t_{solve}		t_{total}	
∞	0.001	(4,3,2,6)	6 (104)	5 (88)	1	1	1	1	1	1	1	1
∞	0.001	(4,3,2,1)	7 (120)	97 (1560)	1	1	0.85	0.78	1.14	18.37	0.88	1.69
∞	0.001	(4,3,1,6)	**6 (104)**	**5 (88)**	1	1	0.97	0.97	1.07	1.13	0.98	0.98
∞	0.001	(4,1,2,6)	**6 (104)**	7 (120)	1	1	0.97	0.95	0.98	1.79	0.98	1
∞	0.001	(1,3,2,6)	**6 (104)**	8 (136)	1	1	0.76	0.69	0.99	2.06	0.80	**0.78**
∞	0.001	(1,1,1,1)	7 (120)	62 (1000)	1	1	**0.38**	**0.31**	1.17	13.32	**0.48**	1.03
∞	0.01	(4,3,2,6)	**6 (104)**	**6 (104)**	0.70	0.75	0.90	0.88	0.95	1.32	0.91	0.91
∞	0.1	(4,3,2,6)	**6 (104)**	-	0.44	0.54	0.77	0.76	**0.85**	-	0.80	-
∞	0.5	(4,3,2,6)	8 (136)	-	**0.31**	**0.44**	**0.67**	**0.71**	1.14	-	**0.73**	-
1	0.001	(4,3,2,6)	**6 (104)**	**5 (88)**	1	1	1	1	**1**	**1**	1	**1**
0.1	0.001	(4,3,2,6)	11 (184)	539 (8664)	0.53	0.50	0.84	0.73	1.62	92	0.89	5.36
0	0.001	(4,3,2,6)	26 (424)	1686 (27104)	**0.17**	**0.32**	**0.50**	**0.53**	3.32	272.60	**0.69**	14.30
0	0.1	(1,1,1,1)	32 (520)	-	**0.17**	**0.32**	**0.16**	**0.15**	4.00	-	**0.44**	-

Table 5.1: Performance of various bi-parametric implementations of the strong form of the preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ for the Model Problem $M = 1$ of Figure 4.3. The left sub-columns correspond to $n = 1.311 + 2.289 \times 10^{-9}\mathrm{i}$ while the right sub-columns to $n = 1.0833 + 0.204\mathrm{i}$ (for their corresponding wavenumbers/wavelengths and size parameters we refer to Table 4.2). All other parameters are the same as in Table 4.3. The memory cost and assembly time are for $\mathbf{P}$ alone. The total time includes assembly time for both $\mathbf{P}$ and $\mathbf{A}$ and solver time for the resulting system. The memory costs and timings have been normalised to those for the first row for each refractive index, which correspond to a non bi-parametric implementation. Minimal values in each column are indicated in bold font. The symbol - indicates that convergence was not achieved to the desired tolerance within the maximum number of iterations prescribed.

$\chi_{\mathbf{P}}$	$\nu_{\mathbf{P}}$	$\mathbf{q_P}$	Iters (Matvecs)		Mem($\mathbf{P}$)		$t_{assembly}$		t_{solve}		t_{total}	
∞	0.001	(4,3,2,6)	8 (816)	5 (528)	1	1	1	1	1	1	1	1
∞	0.001	(4,3,2,1)	9 (912)	112 (10800)	1	1	0.94	0.85	1.26	22.60	0.98	2.09
∞	0.001	(4,3,1,6)	**8 (816)**	**5 (528)**	1	1	0.91	0.87	**0.93**	1	0.92	0.90
∞	0.001	(4,1,2,6)	**8 (816)**	7 (720)	1	1	0.96	0.90	0.98	1.42	0.96	0.94
∞	0.001	(1,3,2,6)	**8 (816)**	7 (720)	1	1	0.80	0.73	0.98	1.41	0.84	**0.80**
∞	0.001	(1,1,1,1)	9 (912)	69 (6672)	1	1	**0.44**	**0.41**	1.32	14.81	**0.57**	1.28
∞	0.01	(4,3,2,6)	**8 (816)**	6 (624)	0.71	0.74	0.91	0.84	0.91	1.32	0.92	0.89
∞	0.1	(4,3,2,6)	**8 (816)**	-	0.44	0.52	0.71	0.71	**0.89**	-	0.76	-
∞	0.5	(4,3,2,6)	11 (1104)	-	**0.31**	**0.43**	**0.67**	0.64	1.35	-	**0.76**	-
1	0.001	(4,3,2,6)	9 (912)	**5 (528)**	0.99	0.99	0.99	0.94	1.05	1.23	0.99	**0.96**
0.1	0.001	(4,3,2,6)	16 (1584)	791 (76272)	0.54	0.50	0.75	0.63	1.68	140.51	0.85	8.58
0	0.001	(4,3,2,6)	32 (3120)	-	**0.18**	**0.31**	**0.47**	**0.48**	3.18	-	**0.74**	-
0	0.1	(1,1,1,1)	36 (3504)	-	**0.18**	**0.31**	**0.22**	**0.21**	3.54	-	**0.57**	-

Table 5.2: Same as in Table 5.1 but for the Model Problem $M = 3$.

reduced $\chi_{\mathbf{P}}$ values is between 40-50% of the original method; the increase in solver time is balanced out by the decrease in assembly time which is between 20-40% of the original one. The memory consumption of the preconditioner is also improving with increasing k_e, reaching the same levels of memory required by the operator $\boldsymbol{\mathcal{A}}$, at around 10% of the original reference case. We also note that removing a small portion of the far-field (at $\chi_{\mathbf{P}} = 0.1$) shows a smoother behaviour than the other two ($\chi_{\mathbf{P}} = 0.01$ and 0) in terms of iterations/matvecs and solver time (except for the lowest k_e considered). This is a better choice than the other two cases if memory consumption is not the main concern. The total time is still reduced to 40% of the reference case while the memory consumption is between 20-40%. We also note that a bi-parametric implementation with all far-field interactions (i.e. $\chi_{\mathbf{P}} = \infty$, presented as blue circles) exhibits the smoothest behaviour of all reduced $\chi_{\mathbf{P}}$ values considered in terms of iterations and matvecs. The assembly time is reduced to 40%, the solver time is almost unaffected and as a result the total time is at 40-45%. The memory consumption is somewhat higher than the other cases at 50-60% depending on the wavenumber considered. Analogous results can be observed in Figures 5.9–5.10 for $M = 3$.

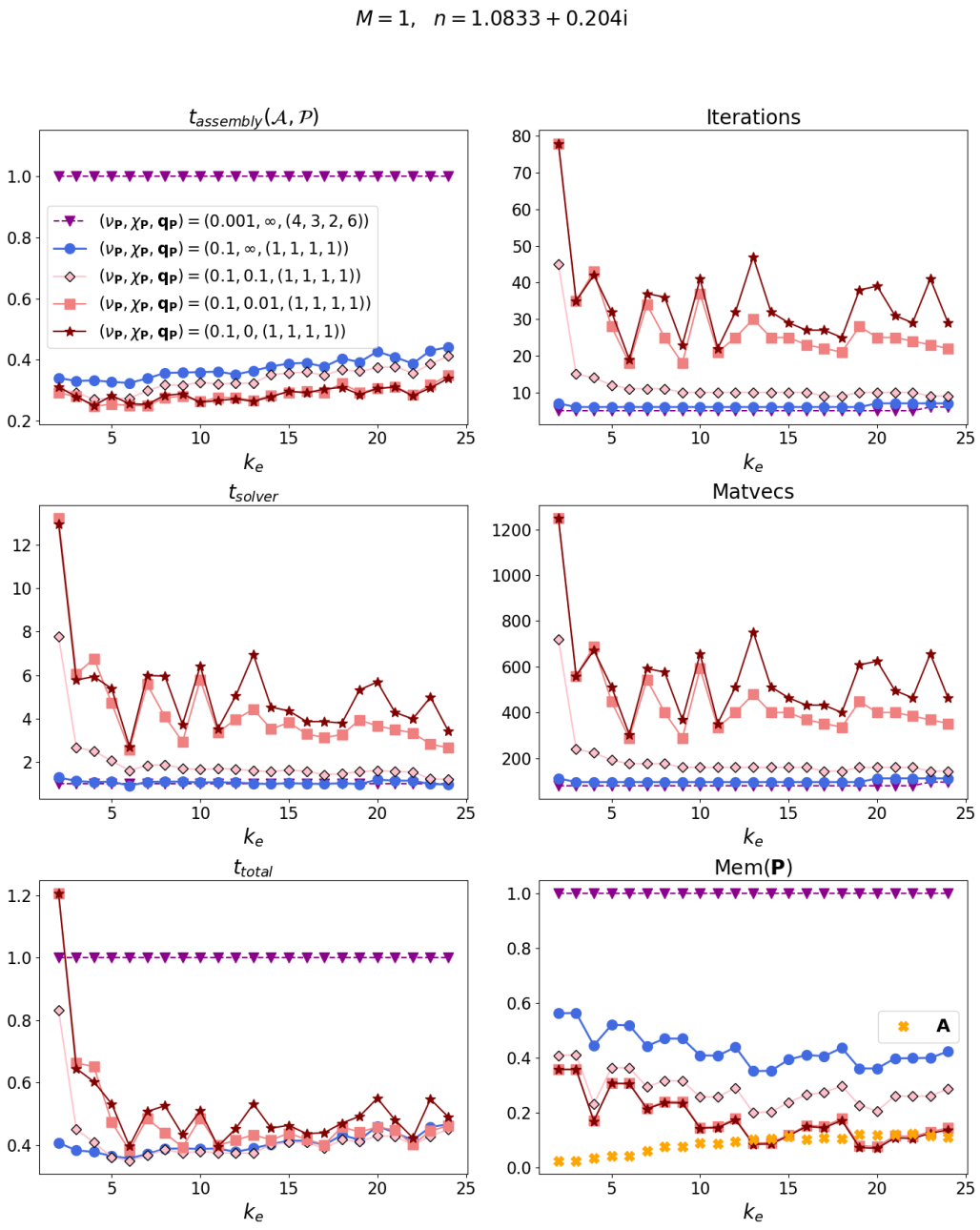


Figure 5.7: Performance of various bi-parametric implementations of the strong form of the preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ as a function of the exterior wavenumber k_e, for the Model Problem $M = 1$ of Figure 4.3, with a low absorbing refractive index ($n = 1.0833 + 0.204\mathrm{i}$). Memory costs and timings have been normalised to those for a non bi-parametric preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ with $(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (\nu_{\mathbf{A}}, \chi_{\mathbf{A}}, \mathbf{q}_{\mathbf{A}}) = (0.001, \infty, (4,3,2,6))$. The memory of the operator $\boldsymbol{\mathcal{A}}$ is also included for reference. For the size parameters and number of degrees of freedom we refer to Figure 4.5.

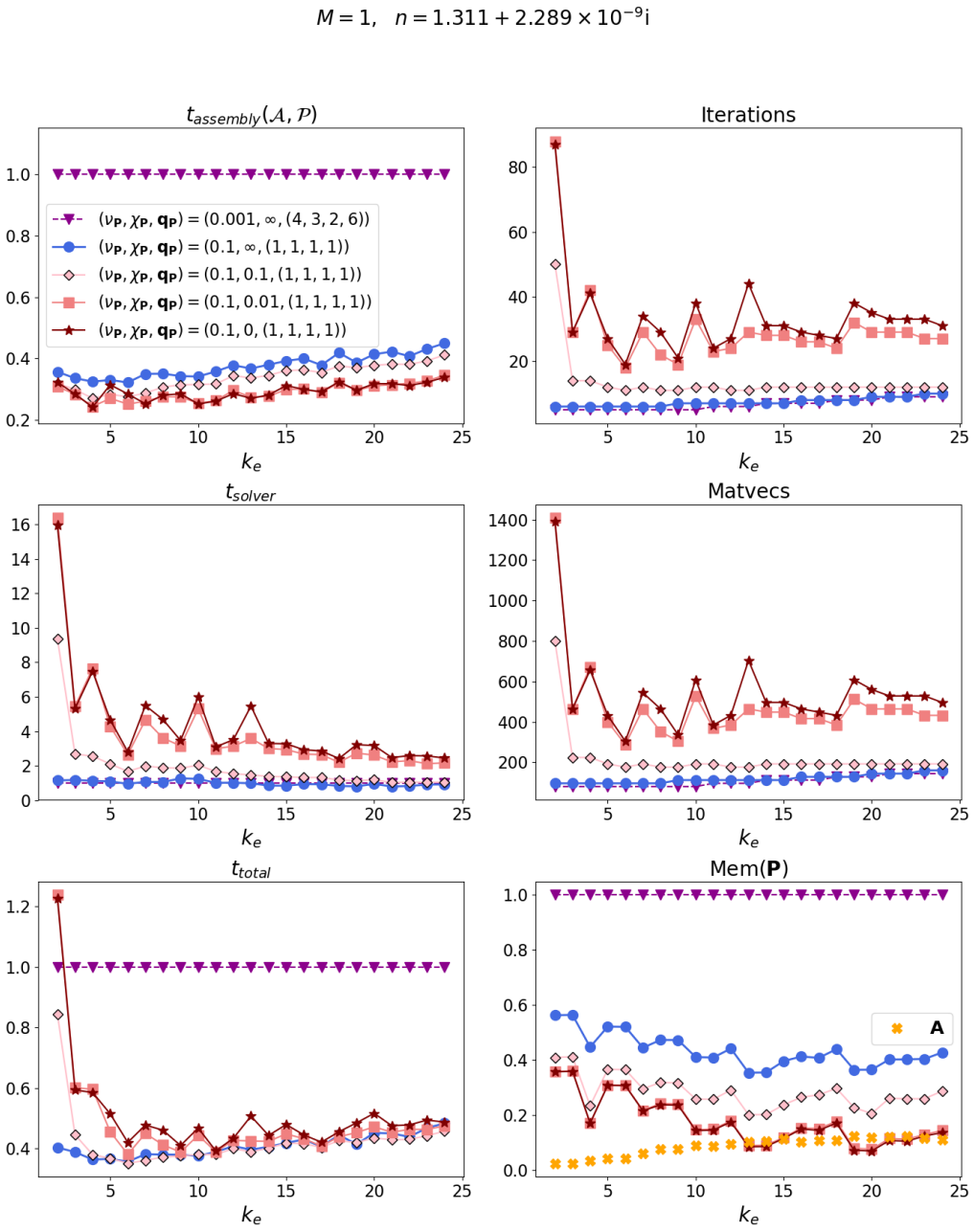


Figure 5.8: Same as in Figure 5.7 but for $n = 1.311 + 2.289 \times 10^{-9}\mathrm{i}$.

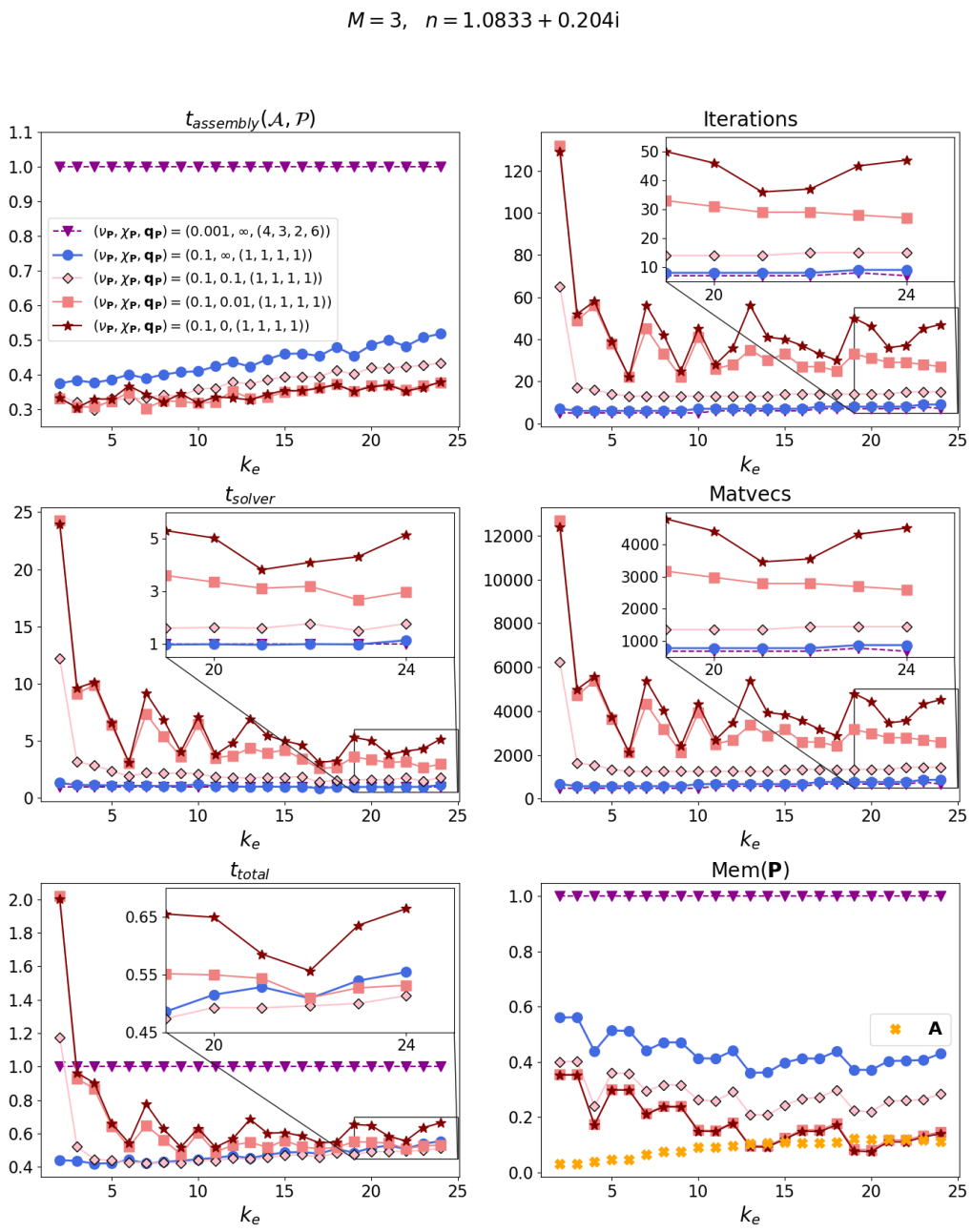


Figure 5.9: Same as in Figure 5.7 but for $M = 3$.

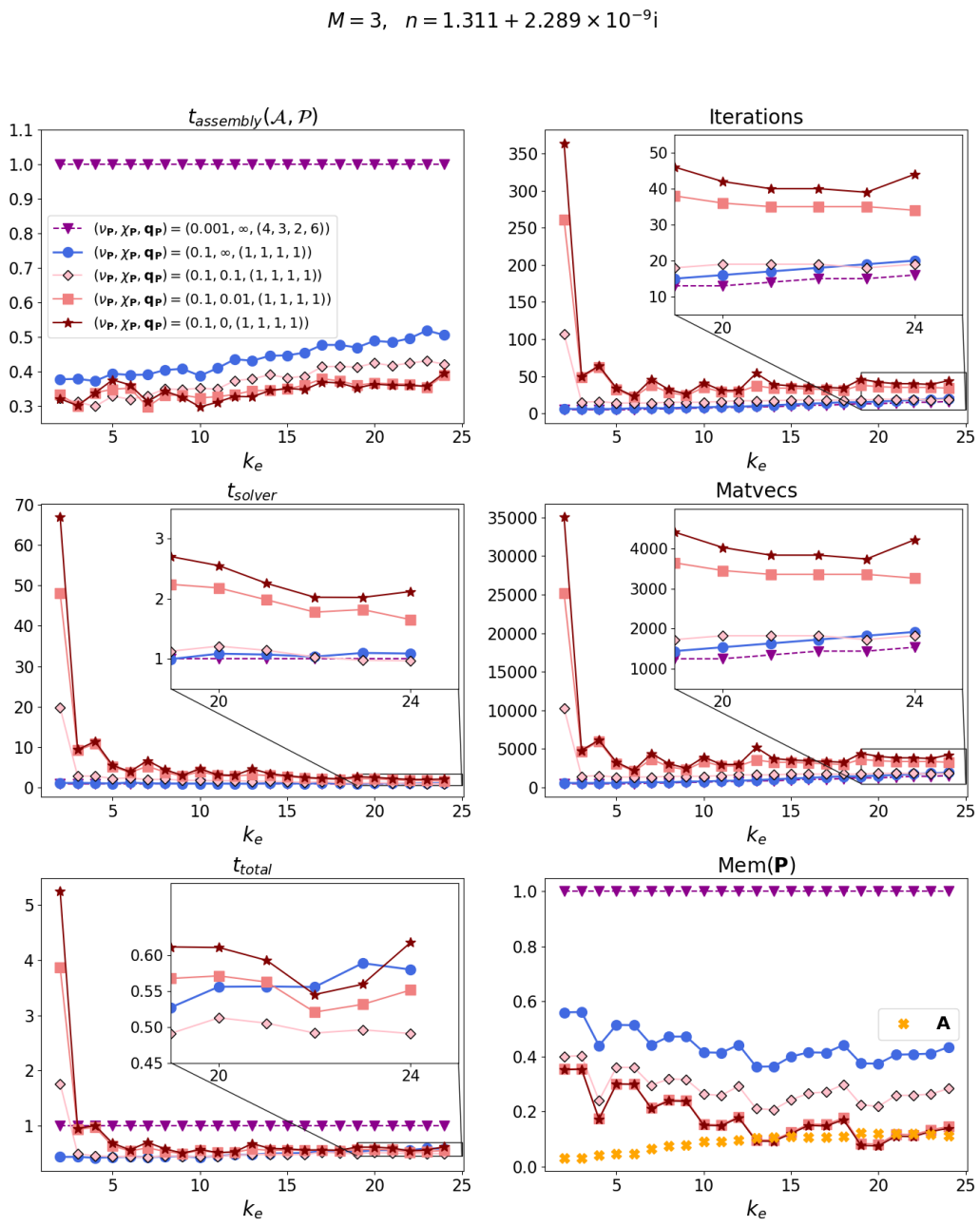


Figure 5.10: Same as in Figure 5.7 but for $n = 1.311 + 2.289 \times 10^{-9}\mathrm{i}$ and $M = 3$.

5.2.2 Reduced Calderón preconditioning

We now present performance statistics for the reduced preconditioners of (4.26)–(4.30) assembled with

$$(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (0.1, 0, (1,1,1,1)),$$

for the Model Problems $M = 1$ and $M = 3$. We make the above choice for preconditioner parameters as they have the potential to significantly reduce assembly time and memory cost, as suggested by the results of the previous section. The results are presented in Figures 5.11–5.12 and 5.13–5.14 for $M = 1$ and $M = 3$ respectively. As before, memory costs and timings are normalised to those for the traditional Calderón preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ assembled with

$$(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (\nu_{\mathbf{A}}, \chi_{\mathbf{A}}, \mathbf{q}_{\mathbf{A}}) = (0.001, \infty, (4,3,2,6)).$$

We first consider $M = 1$, in Figures 5.11–5.12. The assembly times for bi-parametric $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ drop to about 15% of the reference case compared to about 30% for a non bi-parametric implementation (Figures 4.11–4.12), and those of $\boldsymbol{\mathcal{D}}^e$, $\boldsymbol{\mathcal{D}}^i$ and $\boldsymbol{\mathcal{S}}$ to 20-25% compared to 50-60%. Likewise, a remarkable drop in memory cost is observed for all bi-parametric reduced preconditioners at just a few percent of the reference case. As k_e grows the memory cost of the preconditioner becomes smaller than the memory cost of the operator, essentially eliminating the prohibitive cost imposed by the barycentric grid used for the preconditioner. For reference, the memory costs of the non bi-parametric reduced preconditioners (Figures 4.11–4.12) was at 25% for $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ and at about 50% for $\boldsymbol{\mathcal{D}}^e$, $\boldsymbol{\mathcal{D}}^i$ and $\boldsymbol{\mathcal{S}}$.

The cheap assembly comes at a price of a weakened preconditioning effect for all reduced preconditioners, although their behaviour somewhat differs from their non bi-parametric versions. While non bi-parametric $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ showed an erratic behaviour for larger k_e (with $\boldsymbol{\mathcal{S}}^i$ showing a smoother behaviour for high absorption) this is not the case for their bi-parametric versions. Bi-

parametric $\boldsymbol{\mathcal{S}}^i$ only shows a large increase in iterations (and hence matvecs and GMRES time) for the smallest k_e considered but in fact exhibits the smoothest behaviour compared to all other bi-parametric reduced preconditioners. Even though there's an increase in iterations and matvecs for all k_e, t_{solve} is 1-2 times longer compared to the reference case but the total time is consistently between 20-30% of the reference case making it our preferred preconditioner for the problems considered here, ensuring a minimum memory cost at just a few percent with the addition of a reduced total time. Bi-parametric $\boldsymbol{\mathcal{D}}^e$ and $\boldsymbol{\mathcal{S}}^e$ exhibit a rather erratic behaviour and so their use is not recommended.

Analogous observations can be made for $M = 3$, in Figures 5.13–5.14, with the addition of the bi-parametric block-diagonal preconditioner $\boldsymbol{\mathcal{D}}$. While its non bi-parametric implementation achieved a reduction to about 80% of the assembly time (Figures 4.13–4.14), the bi-parametric implementation drops down to about 25%, with the total time dropping from about 80% to 40%, and the memory cost from 95% to just 10% for the largest k_e. It is however outperformed by the other reduced preconditioners in terms of assembly time and memory. It is also outperformed by the reduced preconditioner $\boldsymbol{\mathcal{S}}^i$ in iterations, matvecs, solution time and total time. In fact, $\boldsymbol{\mathcal{S}}^i$ requires about 20% of the total time and just a few percent of the memory of the reference case.

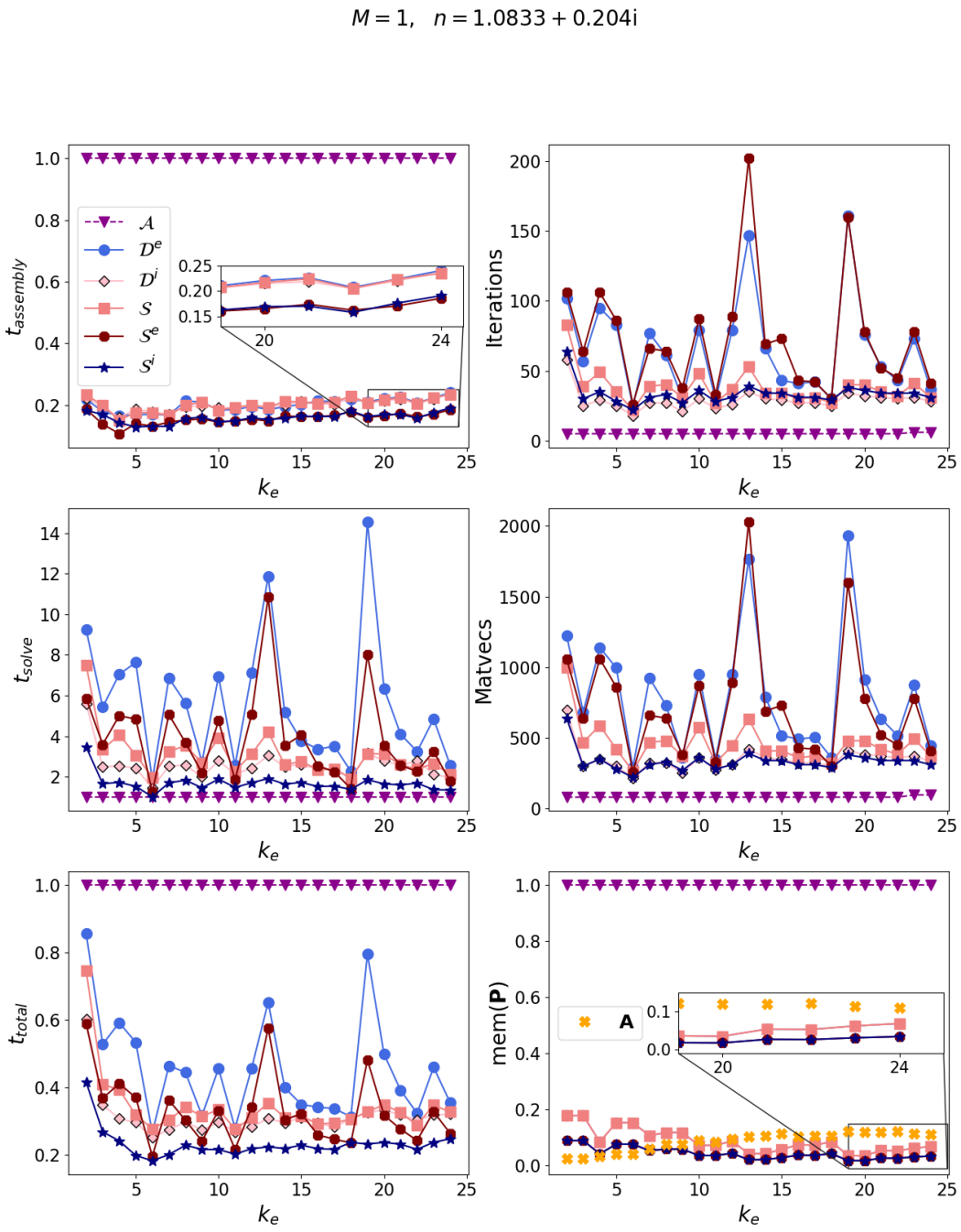


Figure 5.11: Performance of bi-parametric implementations of the strong forms of the reduced preconditioners $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{D}}^e$, $\boldsymbol{\mathcal{D}}^i$, $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{S}}^i$ as a function of the exterior wavenumber k_e, for the Model Problem $M = 1$ of Figure 4.3, with a low absorbing refractive index ($n = 1.0833 + 0.204\mathrm{i}$). Other parameters are as in Figure 5.7.

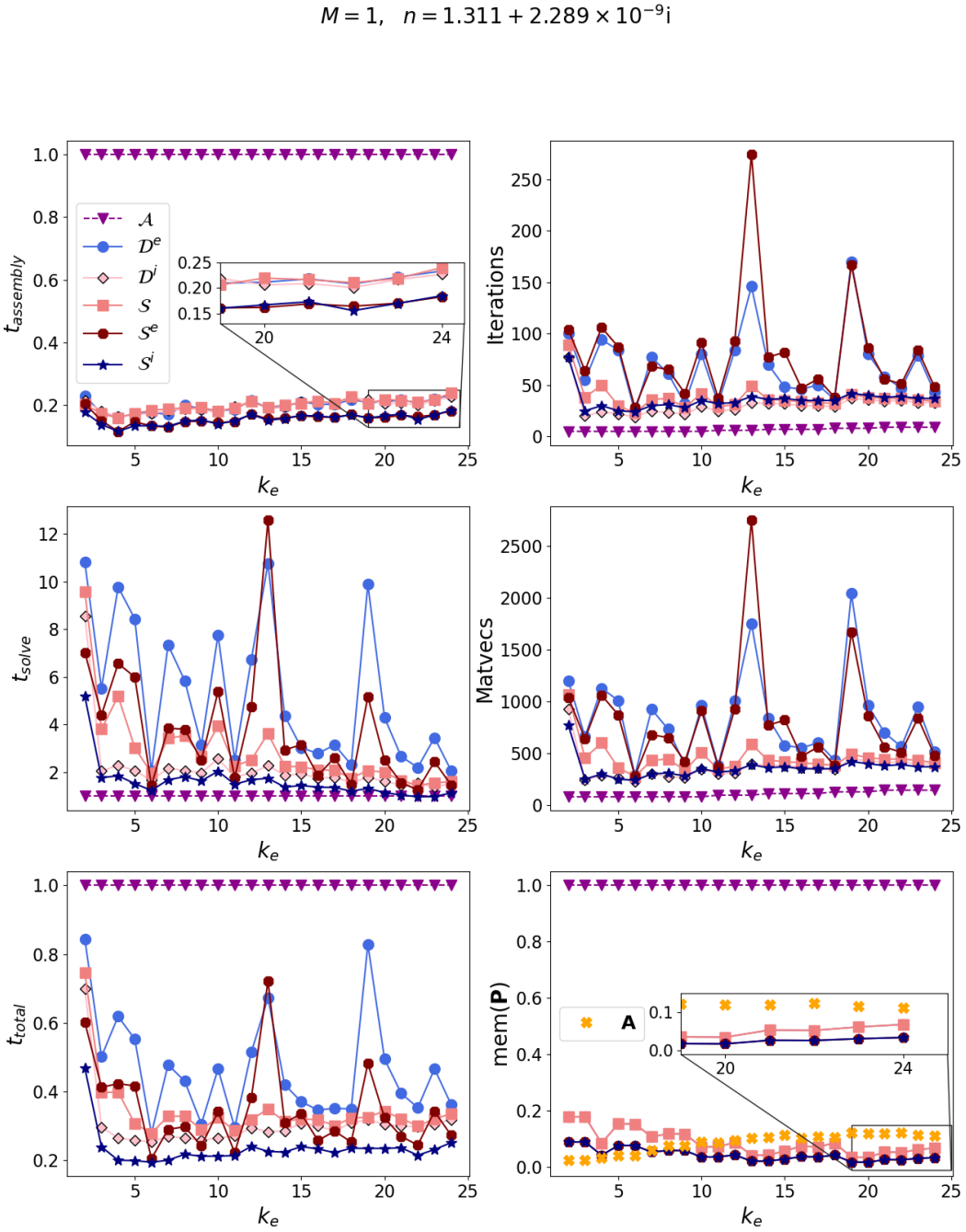


Figure 5.12: Same as in Figure 5.11 but for $n = 1.311 + 2.289 \times 10^{-9}\mathrm{i}$.

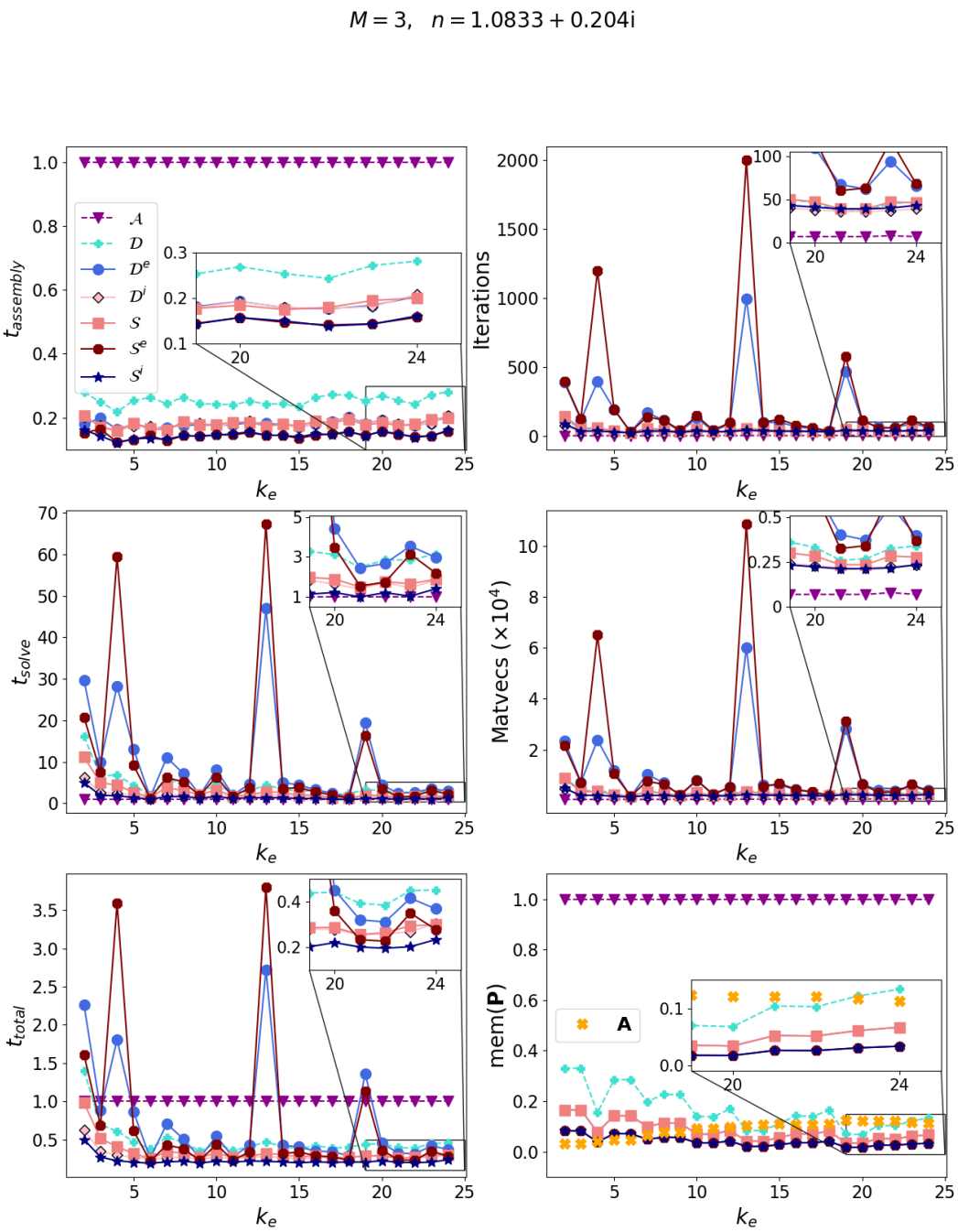


Figure 5.13: Same as in Figure 5.11 but for $M = 3$ and with the addition of $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{D}}$.

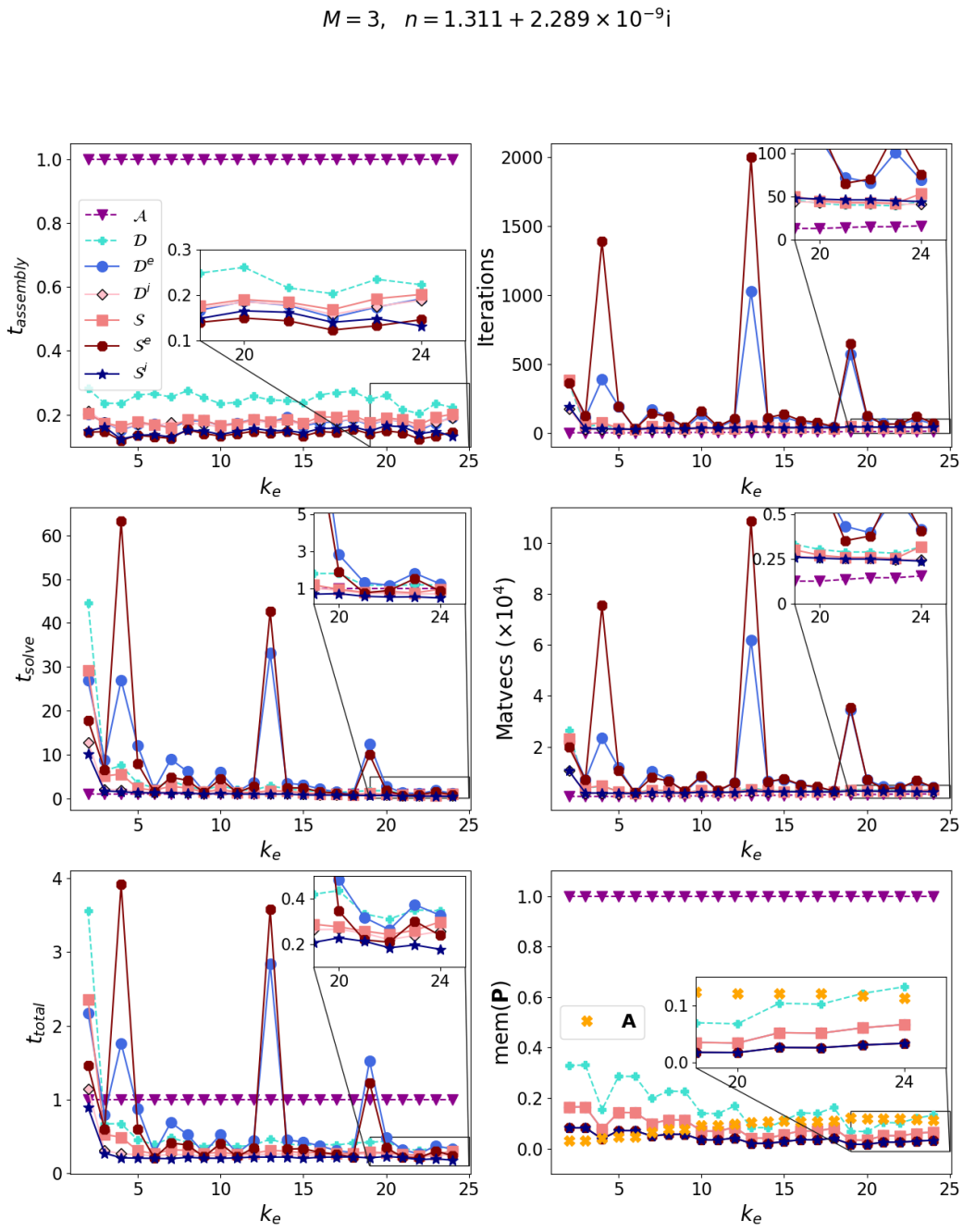


Figure 5.14: Same as in Figure 5.11 but for $n = 1.311 + 2.289 \times 10^{-9}$i and $M = 3$ and with the addition of $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{D}}$.

5.3 Concluding Remarks

We conclude this chapter by recalling our goals and emphasising the main outcomes of our work in addressing them. We recall (from Chapter 1) that both the choice of preconditioner $\boldsymbol{\mathcal{P}}$, and the choice of discretisation of the operator product $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$ affect the computational cost of the preconditioned system. In this chapter we focused on the latter by adjusting its implementation. We followed a so-called "bi-parametric" implementation to achieve a reduced memory cost, assembly time and overall time (assembly and GMRES solve time). By a bi-parametric implementation we mean using two different sets of parameters for the assembly of the operator product; one expensive set of parameters for the accurate assembly of the operator $\boldsymbol{\mathcal{A}}$, and a second cheaper set of parameters for a cheaper approximation of the preconditioner $\boldsymbol{\mathcal{P}}$. The parameters included the accuracy of the ACA algorithm, ν, the near-field cut-off parameter, χ, and the quadrature orders $\mathbf{q}$ (defined in (3.44)–(3.45), (3.46) and (3.48) respectively).

Our numerical experiments have demonstrated that

- A bi-parametric implementation, with

$$(\nu_{\mathbf{A}}, \chi_{\mathbf{A}}, \mathbf{q}_{\mathbf{A}}) = (0.001, \infty, (4,3,2,6)) \tag{5.3}$$

$$(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (0.1, 0, (1,1,1,1)), \tag{5.4}$$

  and with the choice of discretisation (3.39)–(3.40), can alleviate the prohibitive costs imposed by the **BC** basis functions. A reduction in memory cost of up to 80% was observed for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, with the memory cost of $\mathbf{P}$ being roughly the same with that of the operator at higher frequencies.

- A bi-parametric implementation, with (5.3)–(5.4) and with the choice of discretisation (3.39)–(3.40), in combination with the reduced preconditioners of (4.26)–(4.30) results in a memory cost of just a few percent of the original method, below the memory cost of the operator $\mathbf{A}$.

- Completely removing far-field interactions in the preconditioner (by setting $\chi_{\mathbf{P}} = 0$), eliminated the erratic behaviour presented (in Chapter 4) by the non bi-parametric $\boldsymbol{\mathcal{S}}^i$, but results in a worse behaviour for $\boldsymbol{\mathcal{S}}^e$ and $\boldsymbol{\mathcal{D}}^e$.

- For higher frequencies, it is therefore suggested to use $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{S}}^i$ with assembly parameters (5.3)–(5.4) and discretisation of (3.39)–(3.40) which had a memory cost of 2-3% and the total time around 20% compared to the traditional non bi-parametric implementation.

- For lower frequencies, where memory consumption is usually not a concern, the numerical experiments have demonstrated that a bi-parametric implementation with $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, assembly parameters

$$(\nu_{\mathbf{A}}, \chi_{\mathbf{A}}, \mathbf{q}_{\mathbf{A}}) = (0.001, \infty, (4,3,2,6)) \tag{5.5}$$

$$(\nu_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (0.1, (1,1,1,1)), \text{ but } \chi_{\mathbf{P}} = \infty, \tag{5.6}$$

  and the discretisation of (3.39)–(3.40) is best to be used. This choice did not increase GMRES time (compared to other finite values of $\chi_{\mathbf{P}}$) while still achieving a 40% reduction in memory cost and total time. Out of the reduced preconditioners, $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{S}}^i$, with assembly parameters (5.3)–(5.4) and the discretisation of (3.39)–(3.40) had the smallest increase in GMRES time but resulted in a reduction in total time, for lower frequencies.

Chapter 6

Accelerated BEM for atmospheric ice crystals

In this chapter we demonstrate how accelerated Calderón preconditioning can be applied to electromagnetic scattering by atmospheric ice crystals. The ice crystals found in cirrus clouds vary in size and shape, taking forms such as hexagonal columns, hexagonal plates and bullet rosettes [42], and aggregates of those (see Figure 1.2).

Applications of BEM for the calculation of the single scattering properties of such scatterers has been studied by Groth *et al.* [5] but due to BEM being memory intensive without the application of any accelerating techniques, size parameters, X (defined in (1.4)), in terms of the maximum dimension D_{max} have been restricted up to approximately 15. We demonstrate here that with our accelerating techniques, and the computational resources available, we can successfully simulate electromagnetic scattering by monomers and aggregates of much larger size parameters, reducing memory consumption to just 1-2% for the largest problems considered. We demonstrate that the performance observations made in Chapters 4 and 5 carry over to real world applications on domains of complex geometrical structure and not just to the benchmark examples of cube arrays considered earlier.

For single-particle scattering we consider hexagonal columns of increasing geometrical complexity: without cavities, with 'conventional' cavities and with

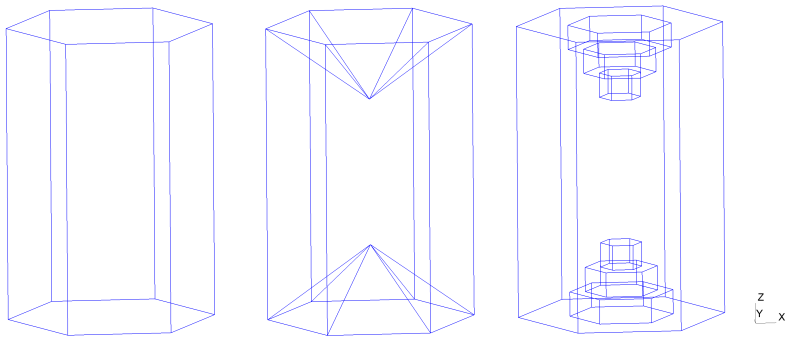


Figure 6.1: Hexagonal columns of [5]: without cavity, with conventional cavity, with stepped cavity. The height of the three columns is the same at $L = 2.86mm$ with the diameter of the hexagonal face being $D = 2mm$. The depth of the conventional cavity is $L/4 = 0.72mm$. Each step of the stepped cavity has depth $0.72/3 = 0.24mm$ and the diameters of the hexagonal faces as we progress down the steps are $D_1 = 0.5D = 1mm$, $D_2 = 0.35D = 0.7mm$ and $D_3 = 0.2D = 0.4mm$.

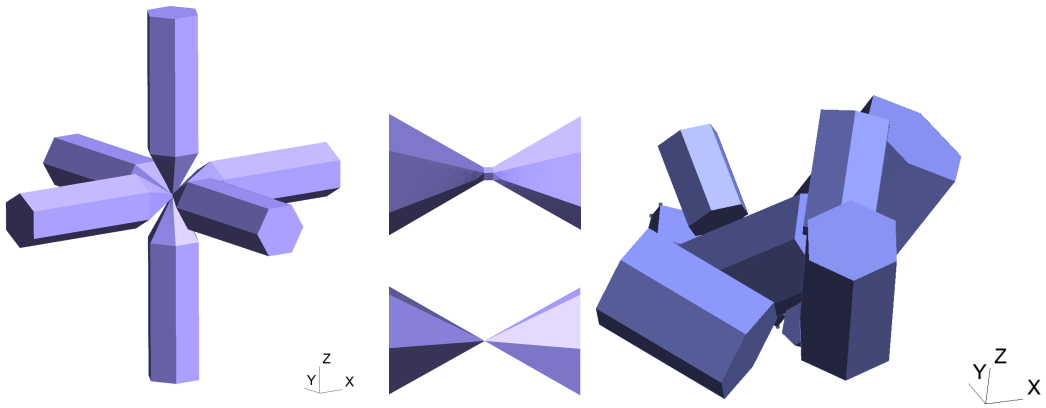


Figure 6.2: Ice crystal aggregates. Left: 6-branched bullet rosette from [5]. Each bullet rosette has a height of $L = 1mm$ with the diameter of the hexagonal face being $D = 0.30mm$. Middle: two branches of a bullet rosette connected via a very small cube (top), or touching at the tip (bottom). Right: 8-branched aggregate from [6].

Frequency f (GHz)	refractive index n	Wavelength λ_e (cm)
50	$1.7754 + 0.00066\mathrm{i}$	0.60
183	$1.7754 + 0.00243\mathrm{i}$	0.16
325	$1.7754 + 0.00440\mathrm{i}$	0.092
664	$1.7754 + 0.00972\mathrm{i}$	0.045

Table 6.1: Refractive indices of ice at $-40°$ C (from [12]) and wavelengths at the four studied frequencies.

stepped cavities. Images of these hexagonal columns can be seen in Figure 6.1. Electromagnetic scattering by such scatterers using BEM has been considered in [5], where their scattering properties have been compared (and validated against results obtained from a T-matrix method in the case of the standard hexagonal column).

For multi-particle scattering we consider the 6-branched bullet rosette from [5] and the 8-branched aggregate from [6], images of which can be seen in Figure 6.2. In [5], the 6 branches of the bullet rosette (and for bullet rosettes of different numbers of branches) were connected through a small cube at the centre where the branches meet to ensure Lipschitz continuity of the boundary (an example can be seen in the middle part of Figure 6.2). We demonstrate that one can treat each branch as a different scatterer resulting in a multi-particle scattering configuration. This allows us to use the reduced preconditioners for $M > 1$ neglecting all off-diagonal blocks and resulting in greater memory and time savings. Theoretical results from [81, 101] confirm the validity of the PMCHWT formulation as the separation between scatterers approaches 0.

In Table 6.1 we give information regarding the frequencies considered, the corresponding refractive index of ice at $-40°$ C from [12] and exterior wavelengths λ_e. Tables 6.2 and 6.3 contain information on the size parameters and number of degrees of freedom for the hexagonal columns of Figure 6.1 and the aggregates of Figure 6.2 respectively. The meshes were generated with the meshing software Gmsh [102]. The incident wave in all cases is a plane wave $\mathbf{E}^{inc} = \mathbf{p}e^{ik_e\mathbf{d}\cdot\mathbf{x}}$ with $\mathbf{p} = (0,1,0)^T$ and $\mathbf{d} = (\sqrt{2}/2, 0, \sqrt{2}/2)$.

f (GHz)	X	# dofs N		
		hexagonal column	hexagonal column conventional cavity	hexagonal column stepped cavity
50	1.5	1170	1314	1674
664	20	45228	48156	53043

Table 6.2: Size parameters and number of dofs, at the frequencies studied for scattering by the hexagonal columns of Figure 6.1. For the refractive indices we refer to Table 6.1.

f (GHz)	X	# dofs N		X	# dofs N
		6-branched rosette, $M = 1$	6-branched rosette, $M = 6$		8-branched aggregate $M = 8$
50	1	804	756	5	2556
183	-	-	-	20	26418
325	-	-	-	34	81318
664	14	10002	9972	70	332523

Table 6.3: Size parameters and number of dofs, at the frequencies studied for scattering by the 6-branched bullet rosettes and the 8-branch ice crystal aggregate from Figure 6.2. For the refractive indices we refer to Table 6.1. The symbol - means we have not considered this case in this thesis.

We first consider the hexagonal columns of Figure 6.1. In Figures 6.3 and 6.4 we report computation times and memory costs for frequencies of 50 and 664 GHz respectively, for six choices of preconditioners: $\boldsymbol{\mathcal{A}}$, the reduced preconditioners $\boldsymbol{\mathcal{D}}^i$ and $\boldsymbol{\mathcal{S}}^i$, and their bi-parametric versions, labelled $\boldsymbol{\mathcal{A}}_{bp}$, $\boldsymbol{\mathcal{D}}^i_{bp}$ and $\boldsymbol{\mathcal{S}}^i_{bp}$. In all cases we assemble the operator $\boldsymbol{\mathcal{A}}$ using the parameters $(\nu_{\mathbf{A}}, \chi_{\mathbf{A}}, \mathbf{q}_{\mathbf{A}}) = (0.001, \infty, (4, 3, 2, 6))$. For the non-bi-parametric versions we take $(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (\nu_{\mathbf{A}}, \chi_{\mathbf{A}}, \mathbf{q}_{\mathbf{A}})$ while for the bi-parametric versions we take $(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (0.1, 0, (1, 1, 1, 1))$. We do not present results for the exterior versions $\boldsymbol{\mathcal{D}}^e$ and $\boldsymbol{\mathcal{S}}^e$, since their behaviour in Chapters 4 and 5 was found to be erratic, leading to longer solver times compared to those of $\boldsymbol{\mathcal{D}}^i$ and $\boldsymbol{\mathcal{S}}^i$, respectively. We note that where $\mathbf{A}$ shows in any legends, this refers to the

discrete version of the operator $\boldsymbol{\mathcal{A}}$ (for example 'assembly $\mathbf{A}$' and 'mem($\mathbf{A}$)'), and not the preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$. The different preconditioners used are shown in the x-axis.

The results of Figures 6.3 and 6.4 are in line with observations made in Chapters 4 and 5. That is, the use of reduced preconditioners $\boldsymbol{\mathcal{D}}^i$ and $\boldsymbol{\mathcal{S}}^i$ significantly reduces the assembly time and memory cost compared to the full preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$. A bi-parametric approach, with the parameters as detailed above, reduces those costs even further, bringing assembly time and memory cost of the preconditioner below that of the operator to just a few percent of the original method.

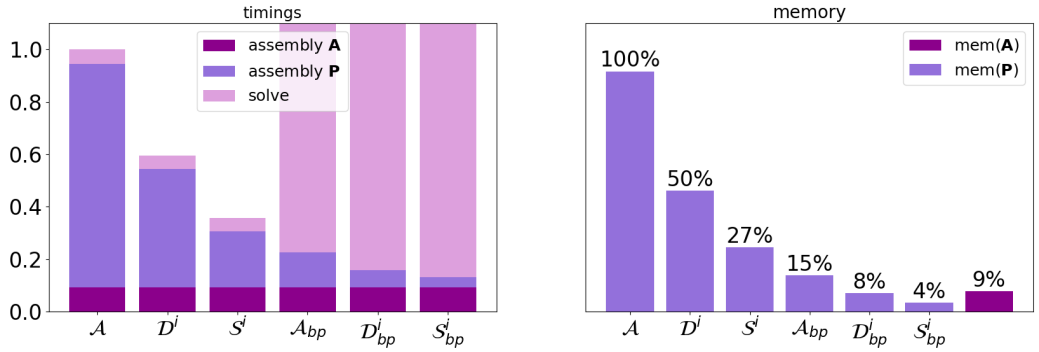
The solution times are different for each frequency, again consistent with observations from Chapters 4 and 5. At the lowest frequency, 50 GHz, which is also associated with a lower absorbing refractive index (see Table 6.1) we see that the solution time is not affected for the reduced non bi-parametric preconditioners $\boldsymbol{\mathcal{D}}^i$ and $\boldsymbol{\mathcal{S}}^i$ (except for $\boldsymbol{\mathcal{S}}^i$ in the hexagonal column with stepped cavity where a slight increase is observed), consistent with observations from Figure 4.12. The number of GMRES iterations for $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{D}}^i$ remains stable at 8 iterations for all three hexagonal columns, with $\boldsymbol{\mathcal{S}}^i$ requiring 13 iterations. The sparseness of $\boldsymbol{\mathcal{S}}^i$ compensates for that increase by reducing the cost per iteration compared to $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{D}}^i$ and keeping the solution time the same between all three methods. This brings the total time for $\boldsymbol{\mathcal{D}}^i$ and $\boldsymbol{\mathcal{S}}^i$ to roughly 55% and 40% respectively. The bi-parametric preconditioners $\boldsymbol{\mathcal{A}}_{bp}$, $\boldsymbol{\mathcal{D}}^i_{bp}$ and $\boldsymbol{\mathcal{S}}^i_{bp}$, do not converge within the maximum number of iterations. This is again consistent with observations from Figure 5.12, where all bi-parametric reduced preconditioners had an increase in solution time for low frequencies and low absorbing refractive indices.

At the highest frequency, 664 GHz, associated with a higher absorbing refractive index (see Table 6.1), we see the solution time decreasing as we move from $\boldsymbol{\mathcal{A}}$ to $\boldsymbol{\mathcal{D}}^i$ and $\boldsymbol{\mathcal{S}}^i$ and then their bi-parametric versions $\boldsymbol{\mathcal{A}}_{bp}$, $\boldsymbol{\mathcal{D}}^i_{bp}$ and $\boldsymbol{\mathcal{S}}^i_{bp}$, bringing the total time to just 20% for $\boldsymbol{\mathcal{S}}^i_{bp}$ compared to the original

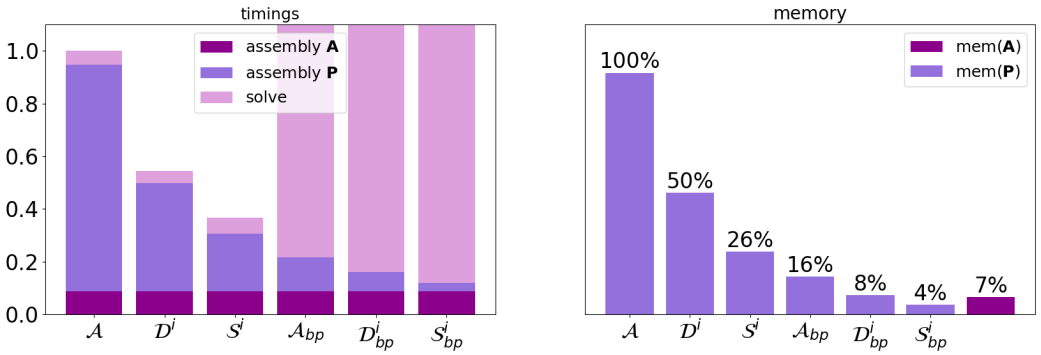
full preconditioner $\boldsymbol{\mathcal{A}}$. The overall behaviour for each frequency is consistent between the three scatterers demonstrating the effectiveness of the methods for increasing geometrical complexities.

We now consider the multi-particle configurations of Figure 6.2. In Figure 6.5 we report computation times and memory costs for the 6-branch bullet rosette at 50 and 664 GHz. The left sub-columns represent results obtained by treating the 6-branch rosette as a single scatterer, i.e. taking $M = 1$. The right sub-columns with the crossed lines correspond to results obtained by treating the rosette as the union of six separate scatterers, i.e taking $M = 6$. The preconditioners considered are the same as before with the addition of the block-diagonal preconditioner $\boldsymbol{\mathcal{D}}$ and its bi-parametric version $\boldsymbol{\mathcal{D}}_{bp}$. We note that for $M = 1$ these just reduce to $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{A}}_{bp}$. The assembly parameters are as for the hexagonal columns. We note that there is a small discrepancy between the number of degrees of freedom between $M = 1$ and $M = 6$ (see Table 6.3); this is to account for the small cube connecting the branches in the case $M = 1$ (see Figure 6.2). The small cube was introduced in [5], in the case $M = 1$, to ensure that the boundary of the scatterer is Lipschitz continuous. Theoretical results from [81, 101] confirm the validity of the PMCHWT formulation when treating the problem as a multi-particle configuration, $M = 6$, with the separation between the scatterers approaching 0. This differences in the number of dofs between $M = 1$ and $M = 6$ is very small and should not affect the comparison of the two cases.

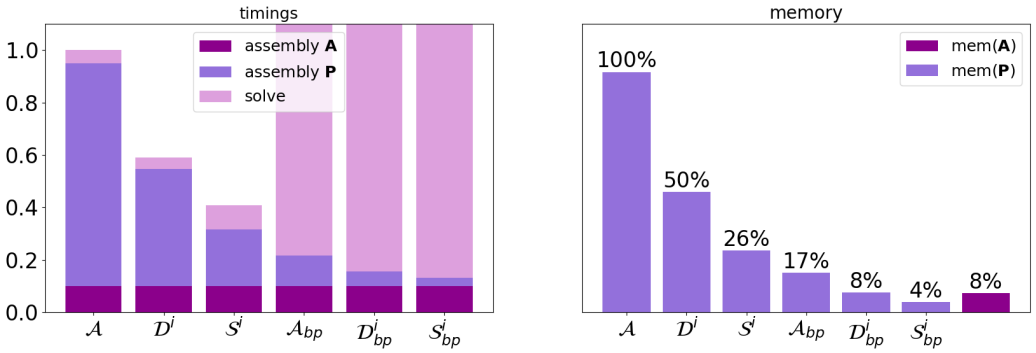
We can see that the assembly and solve times (and consequently total time) between the traditional preconditioners $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ at $M = 1$ and $M = 6$ are roughly equivalent for both frequencies. However, the memory consumption of the same preconditioner at $M = 6$ is only 61% and 65% of that for $M = 1$, at 50 and 664 GHz respectively. This is due to the off-diagonal block elements being cheaper (as discussed in Chapters 4 and 5). In line with earlier observations, reduced bi-parametric preconditioners do not perform well for lower frequencies ($\boldsymbol{\mathcal{D}}_{bp}$ and $\boldsymbol{\mathcal{S}}^i_{bp}$ do converge within the maximum number of iterations but require



(a) hexagonal column

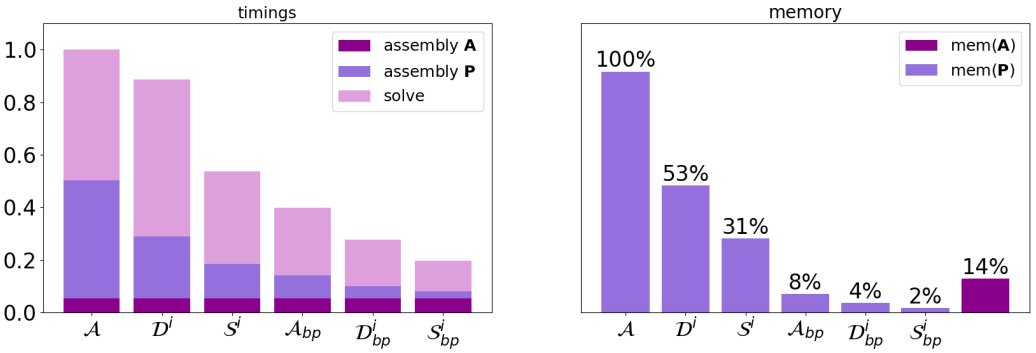


(b) hexagonal column with conventional cavity

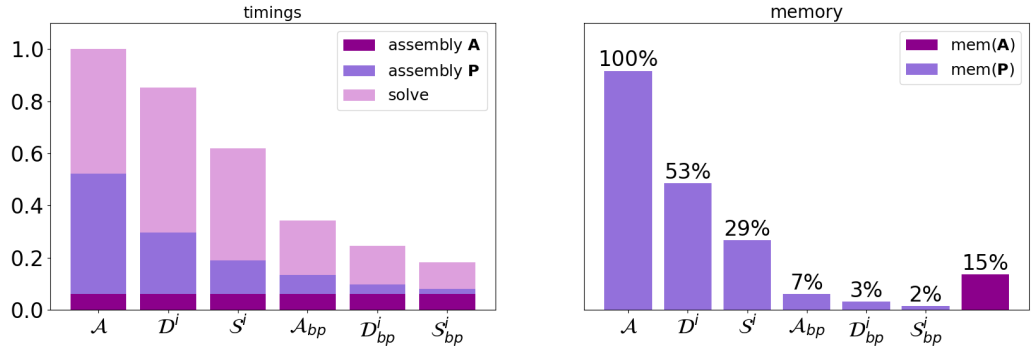


(c) hexagonal column with stepped cavity

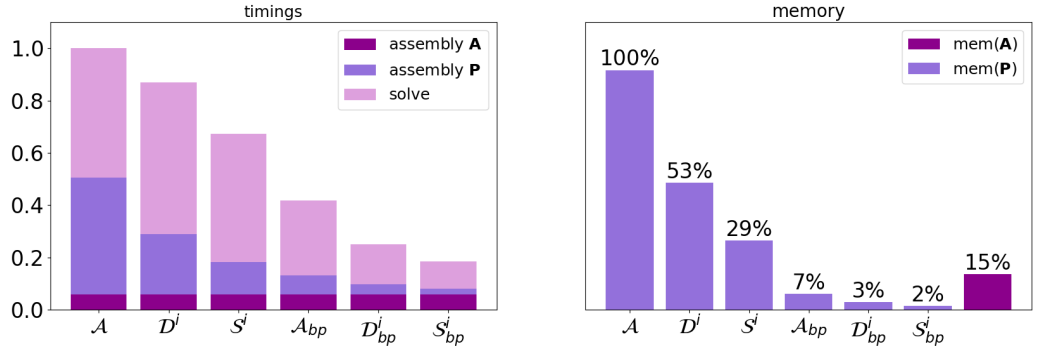
Figure 6.3: Performance of hexagonal columns of Figure 6.1 at 50 GHz. Timings are normalised relative to the total time for $\mathcal{P} = \mathcal{A}$, and memory costs are normalised relative to the memory cost of $\mathcal{P} = \mathcal{A}$. For the refractive indices and size parameters we refer to Tables 6.1 and 6.2. The bi-parametric preconditioners have not converged within the maximum number of iterations.



(a) hexagonal column

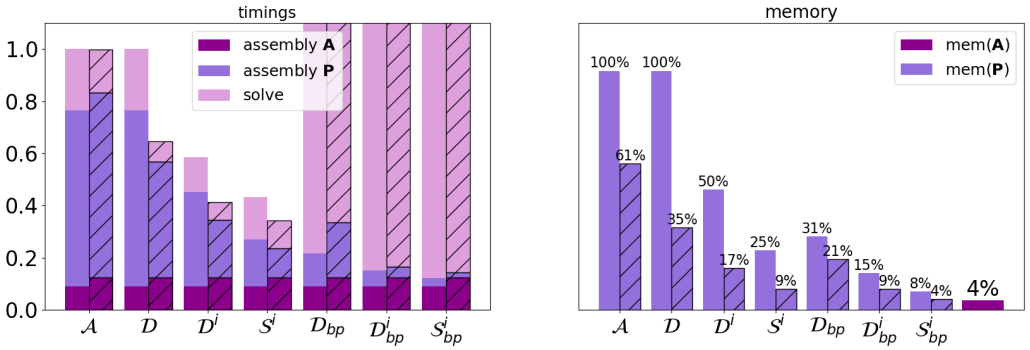


(b) hexagonal column with conventional cavity

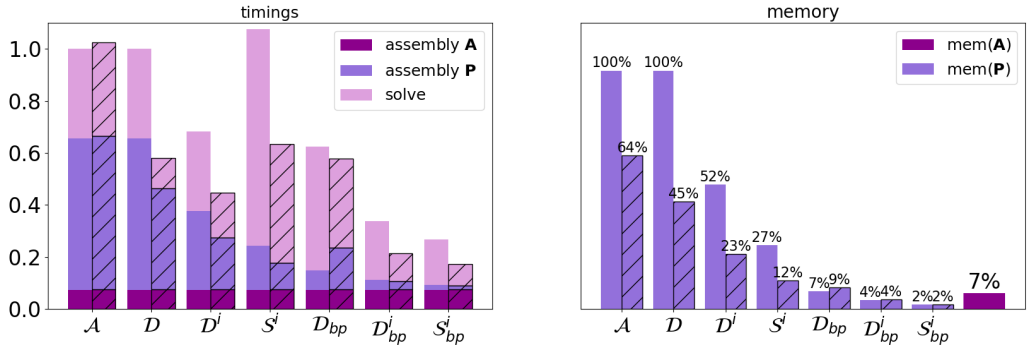


(c) hexagonal column with stepped cavity

Figure 6.4: Same as in Figure 6.3 but at 664GHz.



(a) 50 GHz



(c) 664 GHz

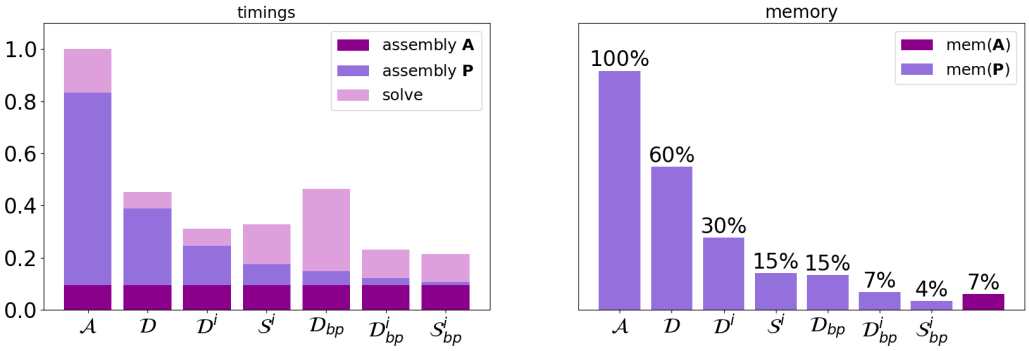
Figure 6.5: Performance of the 6-branched bullet rosette of Figure 6.2. The left sub-plots correspond to $M = 1$, while the right sub-plots with crossed lines correspond to $M = 6$. We note that at $M = 1$, the preconditioner $\boldsymbol{\mathcal{D}}$ reduces to $\boldsymbol{\mathcal{A}}$. Timings are normalised relative to the total time for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ at $M = 1$, and memory costs are normalised relative to the memory cost of $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$ at $M = 1$. For the refractive indices and size parameters we refer to Tables 6.1 and 6.3.

significantly longer time, while $\boldsymbol{\mathcal{D}}^i_{bp}$ does not), but perform much better for 664 GHz. In addition, we observe a performance advantage in assuming $M = 6$ in assembly and solve time as well as memory compared to $M = 1$, however this advantage seems to decrease for the bi-parametric implementations as the frequency increases with equivalent memory consumed either way. This is because assuming a cutoff parameter $\chi_{\mathbf{P}} = 0$ with $M = 1$ will eventually be equivalent to neglecting interactions ($\boldsymbol{\mathcal{A}}_{m\ell}$ terms) between the different branches for $M = 6$.

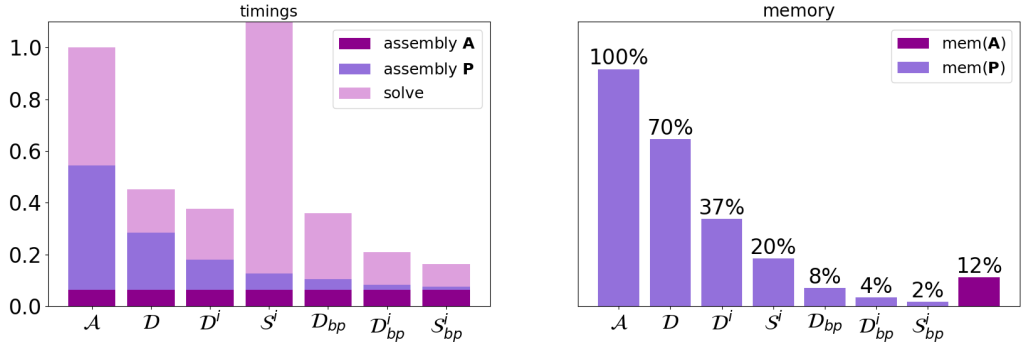
Finally, we consider the 8-branch aggregate of Figure 6.2 at four frequencies: 50, 183, 325 and 664 GHz. For the simulations we treat the problem as

a multi-particle configuration, taking $M = 8$, as the numerical experiments on the 6-branch bullet rosette have demonstrated that memory and solution time savings can be achieved by treating each individual monomer as a scatterer and taking advantage of the block-diagonal preconditioners that are available for $M > 1$. Results for the first three frequencies of Table 6.3 are presented in Figure 6.6. The observations are again in line with those from the previous chapters; the use of the reduced preconditioners significantly reduces assembly time and memory cost compared to the traditional full Calderón preconditioner. In addition, their bi-parametric implementations with only near-field interactions included brings the assembly time and memory cost below that of the operator. At the lowest frequency, 50 GHz, the bi-parametric preconditioner $\boldsymbol{\mathcal{D}}_{bp}$ results in a longer time compared to the original method, but, this is not the case for the higher frequencies. We note that the non bi-parametric reduced preconditioner $\boldsymbol{\mathcal{S}}^i$ did not converge within the maximum number of iterations at 183 and 325 GHz. All other reduced preconditioners (bi-parametric or not) achieve a reduction in solver time, despite an increase in the number of iterations for some choices (number of iterations not reported in the figures); again their sparseness compensates by reducing the cost per iteration.

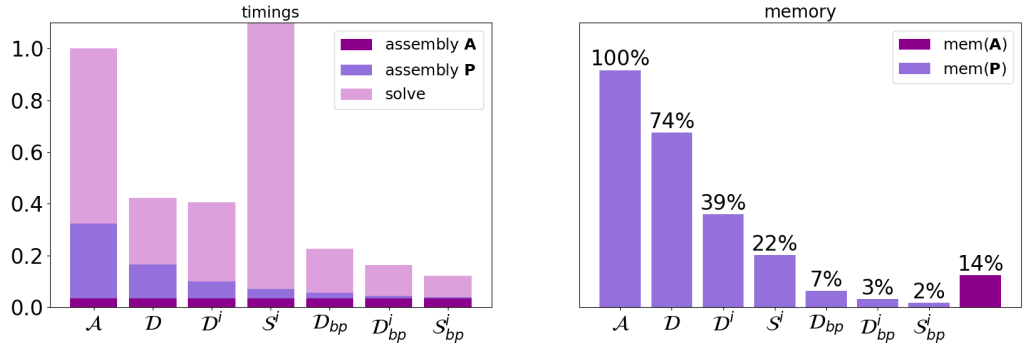
At 664 GHz, memory constraints meant that we were not able to assemble the matrices for any of the methods except for the reduced bi-parametric $\boldsymbol{\mathcal{S}}^i_{bp}$. The assembly time was 10 minutes for the preconditioner, and 32 minutes for the operator, with GMRES converging in 62 minutes with 166 iterations. The memory cost for the preconditioner was approximately 9 GB, just 8% of the memory cost of the operator at 109 GB. Extrapolating from the memory consumption at the lower frequencies, this is estimated to be around 1% of the memory that would be required for the non bi-parametric traditional preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, which is estimated to be approximately 765 GB. Extrapolating from the timings from the lower frequencies, we also expect the total time to be below 20% of the original method. These savings in computational time and memory are significant as they allow us to finally consider large applications



(a) 50 GHz



(b) 183 GHz

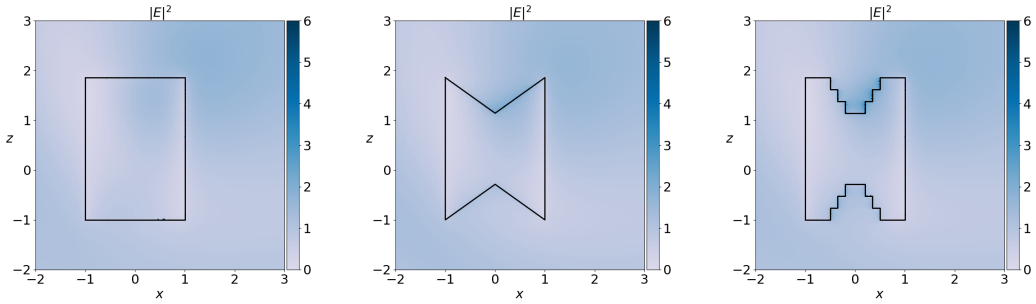


(c) 325 GHz

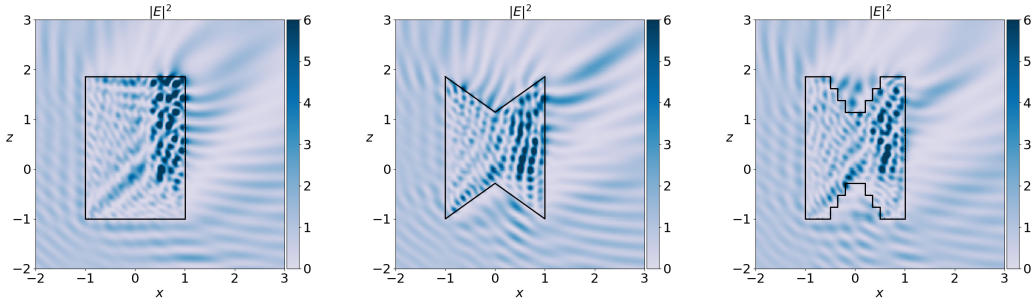
Figure 6.6: Performance of the 8-branched aggregate of Figure 6.2. Timings are normalised relative to the total time for $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$, and memory costs are normalised relative to the memory cost of $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}$. At 183 GHz and 325 GHz the non-bi-parametric version of $\boldsymbol{\mathcal{S}}^i$ (fourth from left in the graphs) did not converge within the maximum number of iterations. For the refractive indices and size parameters we refer to Tables 6.1 and 6.3.

in electromagnetic scattering without the memory constraints that existed before. In particular, applications of these methods in the context of large scale simulations for atmospheric particle scattering are considered next, in Chapter 7.

For completeness, we end this chapter by presenting plots of the near fields produced by some of the ice crystal configurations considered. In Figures 6.7 and 6.8, we present plots of the squared magnitude $|\mathbf{E}|^2$ of the electric field inside and outside the hexagonal columns and the 8-branch aggregate of Figure 6.1 and 6.2, respectively. The plots for the hexagonal columns are restricted to the plane $y = 0$, while those for the 8-branch aggregate to the plane $y = 1$. While the resulting electric fields for the hexagonal columns are very simple at 50 GHz, with their squared magnitudes quite small, we can see far more complex interior and exterior fields as the frequency increases at 664 GHz. The interior pattern between the three hexagonal columns shows a somewhat different behaviour owing to the different structure of the assumed cavities. For the 8-branch aggregate, we can clearly observe the resulting fields changing in complexity as we move to higher frequencies.



(a) 50 GHz



(b) 664 GHz

Figure 6.7: Square magnitude $|\mathbf{E}|^2$ of the electric field for scattering by the hexagonal columns of Figure 6.1 in the plane $y = 0$, at frequencies 50 GHz (top) and 664 GHz (bottom). Computations at 50 GHz were done using the reduced non bi-parametric preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{S}}^i$ and at 664GHz with the reduced bi-parametric preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{S}}^i_{bp}$.

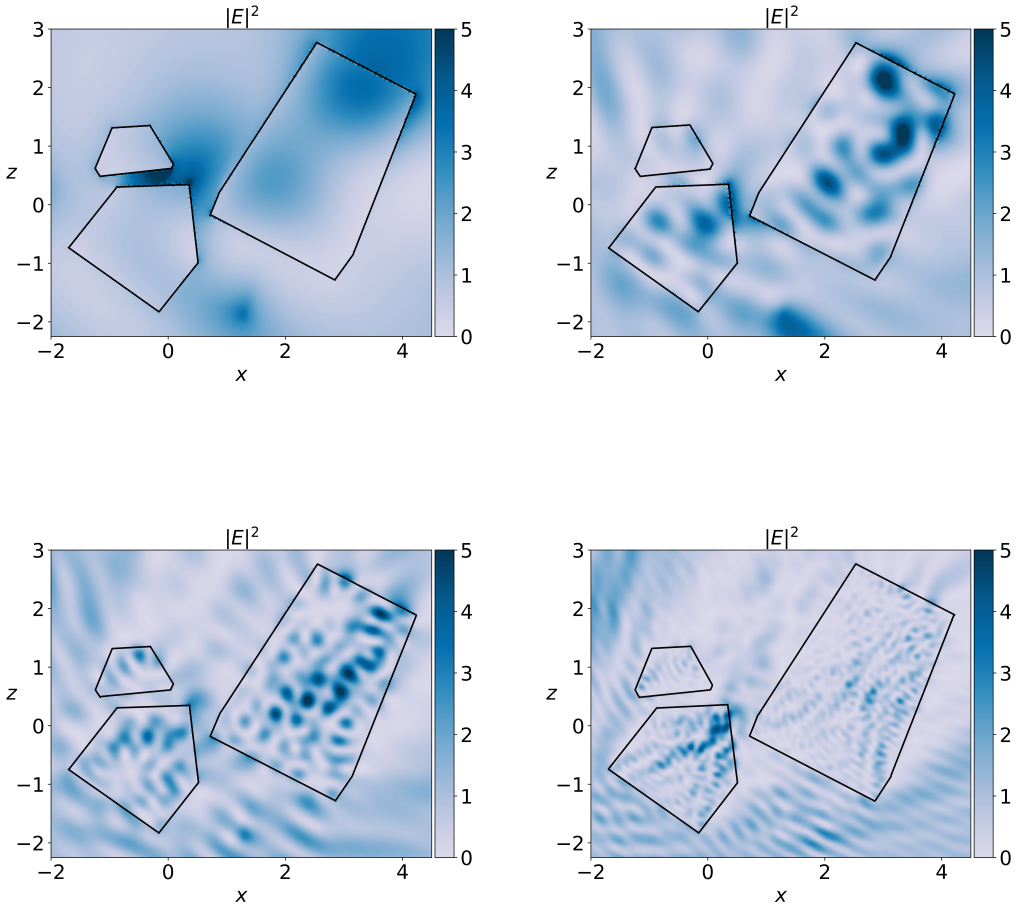


Figure 6.8: Square magnitude $|\mathbf{E}|^2$ of the electric field for scattering by the 8-branch aggregate of Figure 6.2 (right panel) in the plane $y = 1$, at frequencies 50 GHz (top left), 183 GHz (top right), 325 GHz (bottom left) and 664 GHz (bottom right). Computations were done using the reduced bi-parametric preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{S}}^i_{bp}$. Images taken from [3].

Chapter 7

Accelerated BEM for a microwave and sub-millimetre database

In this chapter we demonstrate how our accelerating methods of Chapters 4, 5 and 6 are being used in challenging real world applications. In particular, the methods of accelerating Calderón preconditioning have made it possible to use BEM for the computation of the single-scattering properties (SSPs) of randomly oriented realistic complex ice aggregates found in cirrus clouds, at different frequencies. A new database of the SSPs is being generated at the Met Office for future numerical weather prediction (NWP).

Cirrus clouds are wispy thin clouds that appear high in the Earth's atmosphere, cover a significant percentage of the Earth's surface, and are not confined to a particular latitude or season [4]. The microphysical properties of the constituting ice particles determine the macrophysical properties of the cloud and how much incident solar radiation is transmitted and reflected back to space [4]. This can determine whether there is a positive (warming), neutral or negative (cooling) feedback from cirrus. Understanding the SSPs of cirrus is therefore important in numerical weather prediction and climate modelling [4, 42, 43]. A number of satellites [103, 104, 105] are orbiting the Earth sampling cirrus clouds and measuring their radiative properties.

At the same time, ice crystal models are studied and numerical methods are used to estimate their SSPs so that accurate conclusions can be made from the satellites' data. Several databases have been generated to understand the SSPs of different ice particle shapes in different frequency regimes (some example references include [57, 106, 107, 108, 73, 109, 10] and references within).

A new database of SSPs of ensembles of rosette aggregates is currently being generated using BEM at the frequencies of 50, 183, 243 and 664 GHz, at temperatures of 190, 210, 230, 250 and 270 K. This database is being constructed to take advantage of forthcoming new observations from EUMETSAT's (The European Organisation for the Exploitation of Meteorological Satellites) next generation of polar orbiting satellites that ought to improve numerical weather prediction in the early 2020s, and the simulation of airborne radiance observations using the Met Office's International Sub-millimetre Airborne Radiometer (ISMAR, see [43]).

We begin this chapter by discussing the microphysical model used to generate our database. We note that the microphysical model has not been created by the author of this thesis but we briefly include details for completeness. We then focus on scattering in the far-field zone, giving all the relevant definitions required to compute the scattering properties from [8] along with implementation details. Our implementation of simulating random orientations is also presented and tested against a T-matrix method [22] for simple hexagonal columns. We discuss the accuracy of our implementation for the specific aggregate model based on mesh discretisation, number of incident waves and polarisation rotations and compare with other databases. We finish by presenting some early results from our database and show how these compare with other databases.

7.1 The microphysical model

The ensemble of rosette aggregates is based on an analysis of Cloud Particle Imager (CPI) data from twenty-two campaigns distributed throughout the mid-latitudes, tropics, and southern latitudes analysed by Lawson *et al.* [7]. This included more than ten million CPI images from which representative shape distributions were obtained for four ice cloud regimes. For the purposes of the database we are interested in shape distributions from in-situ generated cirrus, with maximum dimensions greater than 100 μm as these larger sizes are more relevant to the microwave and sub-millimetre region of the electromagnetic spectrum.

In Figures 7.1–7.2 we show typical shape distribution of ice crystals for different temperature regimes and contribution of those ice paricles to the mass and area of in-situ generated cirrus. The most representative ice crystal shapes of relevance to modelling millimetre wave and sub-millimetre wave scattering within in-situ generated cirrus are budding rosettes and aggregates of rosettes, which we use to generate our ensemble model. Examples of the CPI images of budding rosettes and rosette aggregates can be seen in Figures 7.2–7.3.

The aggregates used for the database were generated using the ice aggregation model of Westbrook *et al.*, [71]. We briefly describe the process but refer to [71] for more details. The aggregates are formed by Monte Carlo simulations that sample collisions between particles falling at different speeds. The simulation begins with a population of monomer crystals, the shape of which can be chosen to match the observed monomer crystals in the cloud of interest. These particles collide and stick together to produce an ensemble of complex aggregates. These aggregates are formed by a collection kernel that considers the geometric cross sections for the collisions and the difference in fall speed between the realised ice crystals. Here, to form these ice crystals and aggregates, we consider budding rosettes and aggregates of rosettes. The monomer rosettes are constructed of three-branched rosettes, which are aggregated together to follow the Cotton *et al.*, [72] mass – dimension relationship.

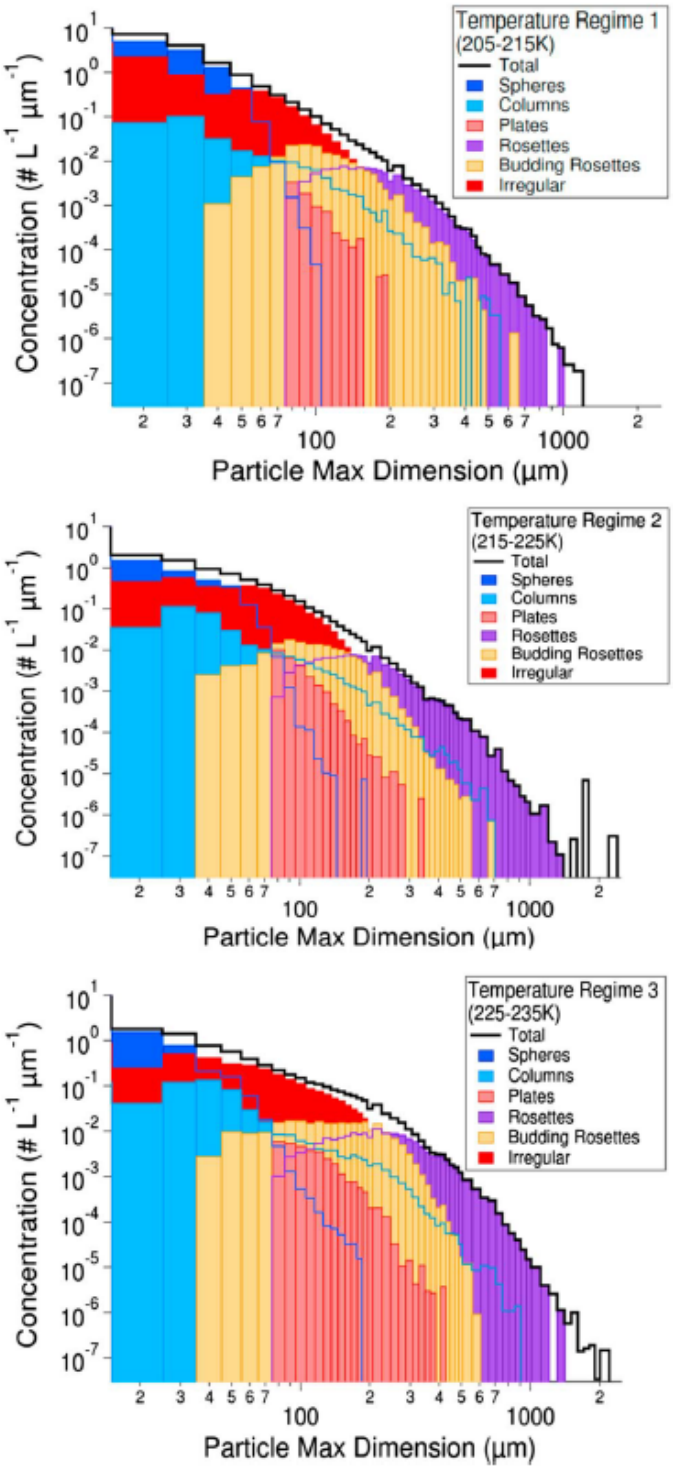


Figure 7.1: Ice crystal shape distribution with respect to the particle's maximum dimension for different temperature regimes obtained during the SPARTICUS campaign for in-situ generated cirrus. Image from [7].

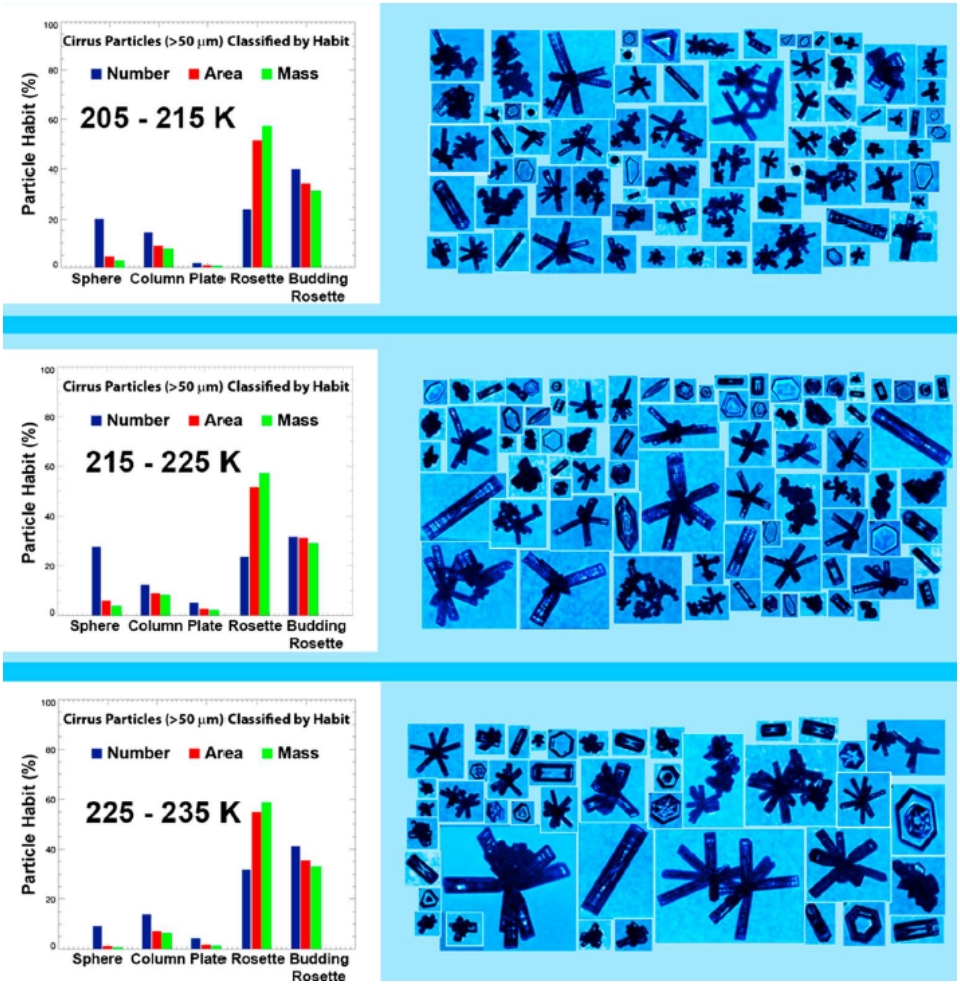
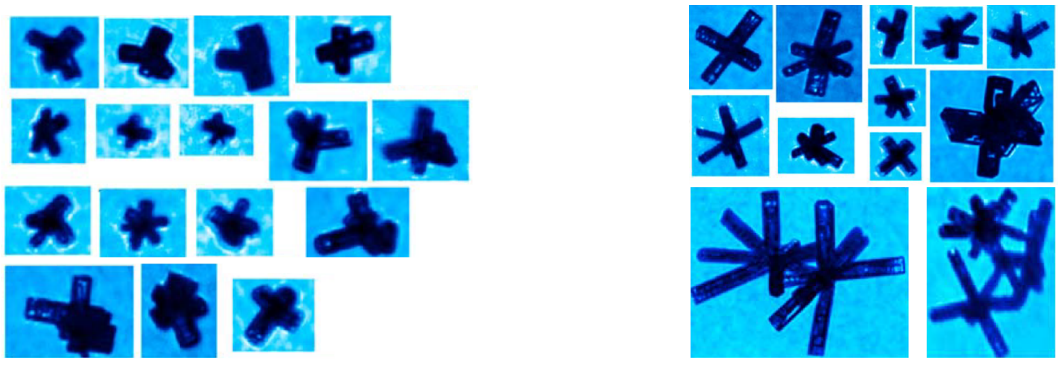


Figure 7.2: Main ice crystal shape contributions of mass and area within in-situ generated cirrus of particles of $D_{max} > 50\ \mu m$ for different temperature regimes (left) and representative examples of CPI images within each temperature range (right). Image from [7].



(a) (b)

Figure 7.3: Examples of CPI images sampled in cirrus from [7]. (a) Typical budding rosettes (b) Typical rosettes and rosette aggregates.

Temp	50 GHz	183 GHz	243 GHz	664 GHz
190	1.7643 + 0.00042i	1.7643 + 0.00154i	1.7643 + 0.00207i	1.7643 + 0.00649i
210	1.7695 + 0.00051i	1.7695 + 0.00188i	1.7695 + 0.00252i	1.7695 + 0.00771i
230	1.7746 + 0.00064i	1.7746 + 0.00235i	1.7746 + 0.00314i	1.7746 + 0.00940i
250	1.7797 + 0.00084i	1.7797 + 0.00309i	1.7795 + 0.00412i	1.7798 + 0.01209i
270	1.7848 + 0.00120i	1.7849 + 0.00442i	1.7849 + 0.00589i	1.7849 + 0.01690i

Table 7.1: Complex refractive index for the five temperatures (K), at the four different frequencies from [12].

This is to be consistent with the Met Office's Unified Model (UM) [110] which assumes the above mass – dimension relation. Each of the monomers that makes up the ensemble of rosette aggregates is assumed to have the density of solid ice, i.e. $917kg/m^3$.

Five different runs of the Monte Carlo simulations were considered, to cover different ranges of maximum dimension, D_{max}. Of those different runs, 52 different aggregates were selected to form our aggregate model to represent 65 different D_{max} values. Images of those aggregates are shown in Figures 7.4–7.6. Aggregates of $D_{max} = 547\mu m - 10235\mu m$ are all distinct. Aggregates of $D_{max} = 10\mu m - 492\mu m$ are rescaled versions of aggregate $D_{max} = 492\mu m$. The aggregates consist of 6 to 168 monomers of different complexity. The meshes used to construct the aggregates have been created using Gmsh [102]. We note that for each configuration the individual monomers are of the same shape and size, assembled together in different orientations. Assuming that the scattering problem for the individual monomer is $\mathcal{O}(1)$, then the scattering problem for the resulting aggregate grows as $\mathcal{O}(M)$, with M being the number of monomers used to construct the aggregate. The single-scattering properties and phase matrix elements for all aggregates are computed at frequencies 50, 183, 243 and 664 GHz, and at temperatures 190, 210, 230, 250 and 270K, with their respective refractive indices taken from [12] and shown in Table 7.1.

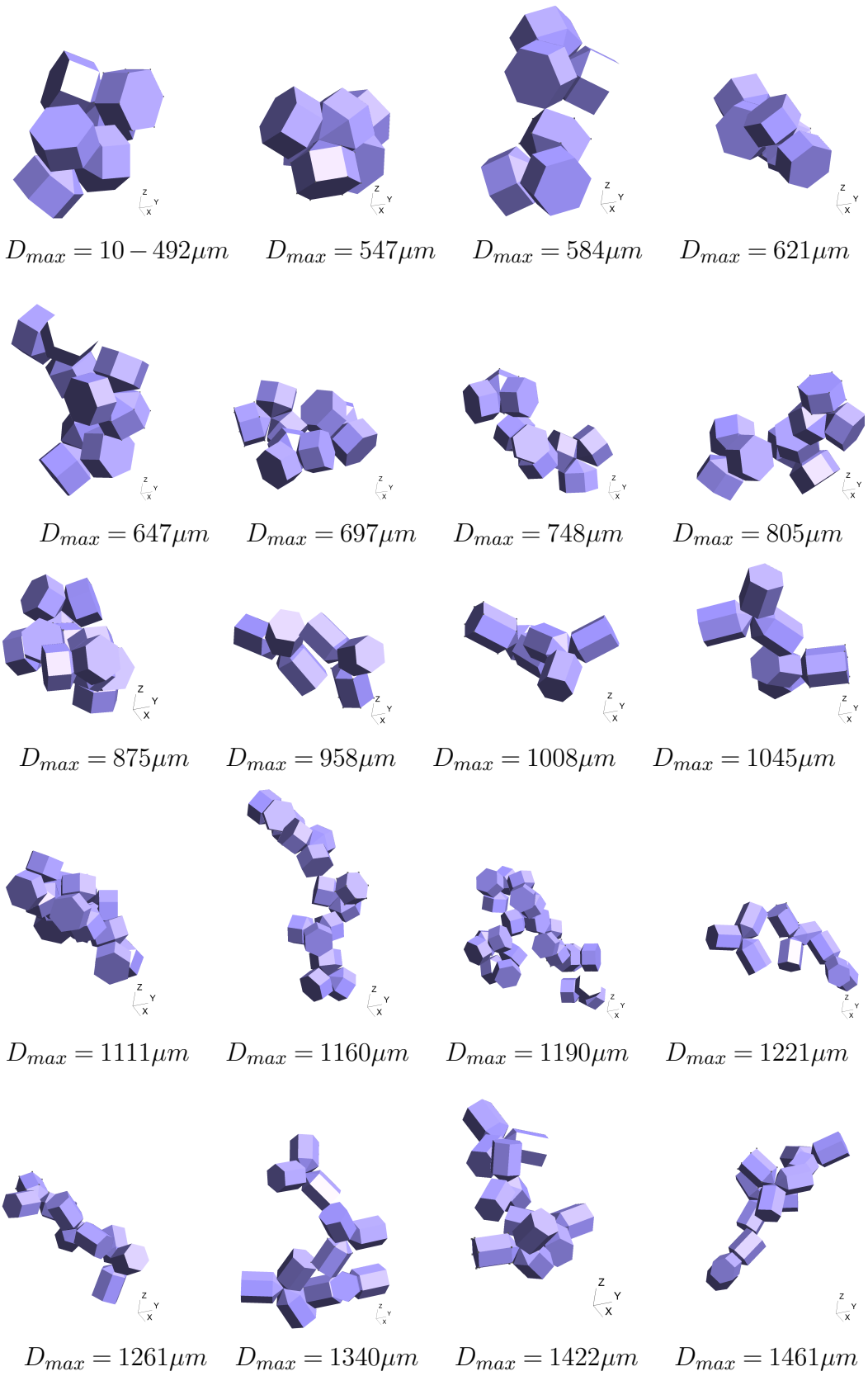


Figure 7.4: Images of the budding rosettes and ice aggregate models used to generate the scattering database. The models are shown as a function of increasing maximum dimension, D_{max}, from 10 to 1461 μm.

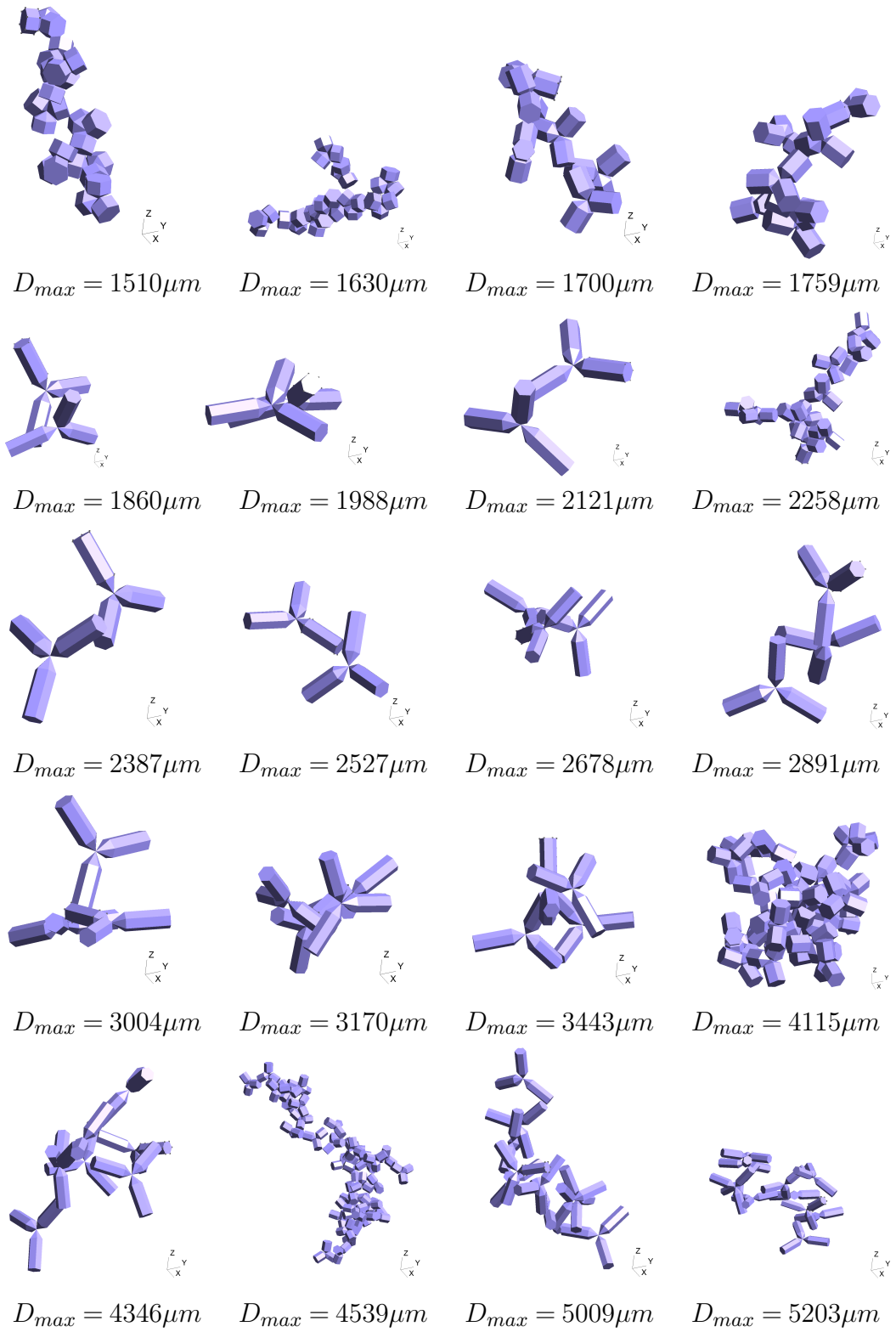
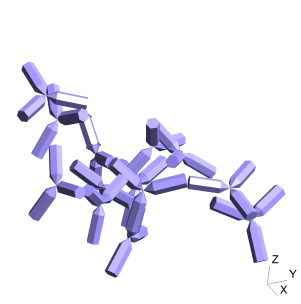
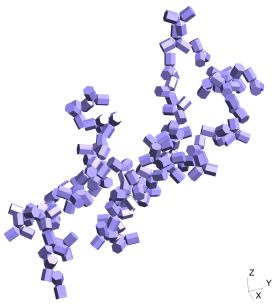


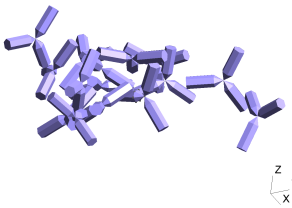
Figure 7.5: Same as in Figure 7.4 but for D_{max} from 1510 to 5203 μm.



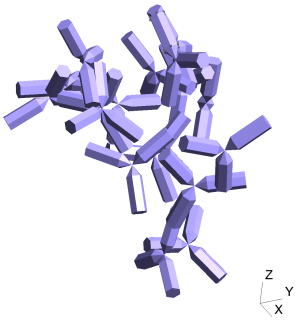
$D_{max} = 5508\mu m$



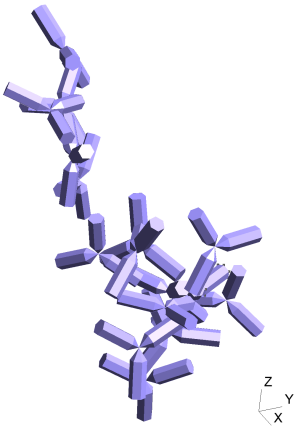
$D_{max} = 5791\mu m$



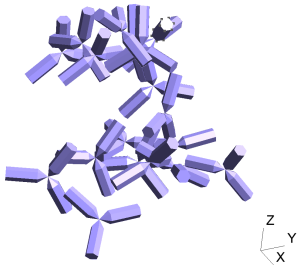
$D_{max} = 6188\mu m$



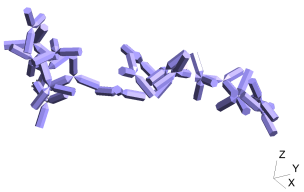
$D_{max} = 6456\mu m$



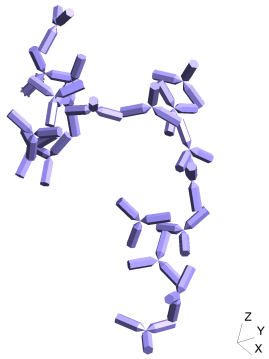
$D_{max} = 6688\mu m$



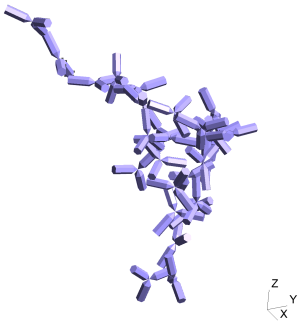
$D_{max} = 6881\mu m$



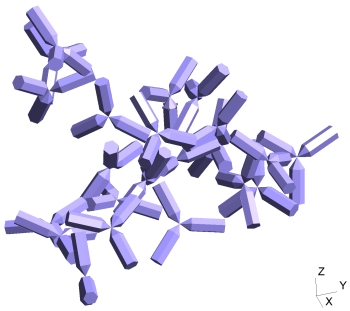
$D_{max} = 7884\mu m$



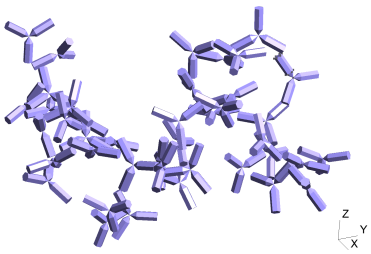
$D_{max} = 8582\mu m$



$D_{max} = 9200\mu m$



$D_{max} = 9594\mu m$



$D_{max} = 10235\mu m$

Figure 7.6: Same as in Figure 7.4 but for D_{max} from 5508 to 10235 μm.

7.2 Scattering in the far-field zone

In the applications considered in this chapter, the scattered far-field[1] is needed to compute the scattering properties of ice crystals. The far-field zone is the area where $k_e r \gg 1$, where r is much bigger than any linear dimension of the scattering object [8]. Before we proceed to define the scattered far-field and its properties we fix some notation. Definitions in this section are taken from [8] unless specified otherwise.

We consider an arbitrary point O close to the geometrical centre of the scattering object as the common origin of all position vectors as in Figure 7.7. We denote the direction of propagation of an incident plane electromagnetic wave by $\hat{\mathbf{n}}^{inc}$, which depends on $(\theta^{inc}, \phi^{inc})$, where $\theta^{inc} \in [0, \pi]$ is the polar (zenith) angle and $\phi^{inc} \in [0, 2\pi)$ is the azimuth angle. The corresponding unit vectors are $\hat{\boldsymbol{\theta}}^{inc}$ and $\hat{\boldsymbol{\phi}}^{inc}$ such that $\hat{\mathbf{n}}^{inc} = \hat{\boldsymbol{\theta}}^{inc} \times \hat{\boldsymbol{\phi}}^{inc}$ and defined as

$$\hat{\mathbf{n}}^{inc} = \begin{bmatrix} \sin\theta^{inc}\cos\phi^{inc} \\ \sin\theta^{inc}\sin\phi^{inc} \\ \cos\theta^{inc} \end{bmatrix}, \ \hat{\boldsymbol{\theta}}^{inc} = \begin{bmatrix} \cos\theta^{inc}\cos\phi^{inc} \\ \cos\theta^{inc}\sin\phi^{inc} \\ -\sin\theta^{inc} \end{bmatrix}, \ \hat{\boldsymbol{\phi}}^{inc} = \begin{bmatrix} -\sin\phi^{inc} \\ \cos\phi^{inc} \\ 0 \end{bmatrix}. \tag{7.1}$$

If the medium of propagation is non absorbing (which it is in our case), the component of the electric field vector along the direction of propagation $\hat{\mathbf{n}}^{inc}$ is zero and the incident field can then be decomposed into components lying in the $\hat{\boldsymbol{\theta}}^{inc}$ and $\hat{\boldsymbol{\phi}}^{inc}$ directions as follows

$$\mathbf{E}^{inc} = (E_\theta^{inc}\hat{\boldsymbol{\theta}}^{inc} + E_\phi^{inc}\hat{\boldsymbol{\phi}}^{inc})\exp(ik_e\hat{\mathbf{n}}^{inc}\cdot\mathbf{r}), \tag{7.2}$$

where $E_\theta^{inc}\hat{\boldsymbol{\theta}}^{inc}$ and $E_\phi^{inc}\hat{\boldsymbol{\phi}}^{inc}$ define the polarisation vectors. In the same manner, one can create a coordinate system $\hat{\mathbf{n}}^{sca}, \hat{\boldsymbol{\theta}}^{sca}, \hat{\boldsymbol{\phi}}^{sca}$ to describe the scattered wave. The linearity of the Maxwell's equations (2.1)–(2.4) means

[1] We note that the far-field mentioned in this chapter refers to the scattered far-field zone and not the far-field interactions included in the $\mathcal{H}$-matrix assembly that depend on the cutoff parameter χ introduced in (5.2). Where the cutoff parameter is used in this chapter, it is mentioned explicitly.

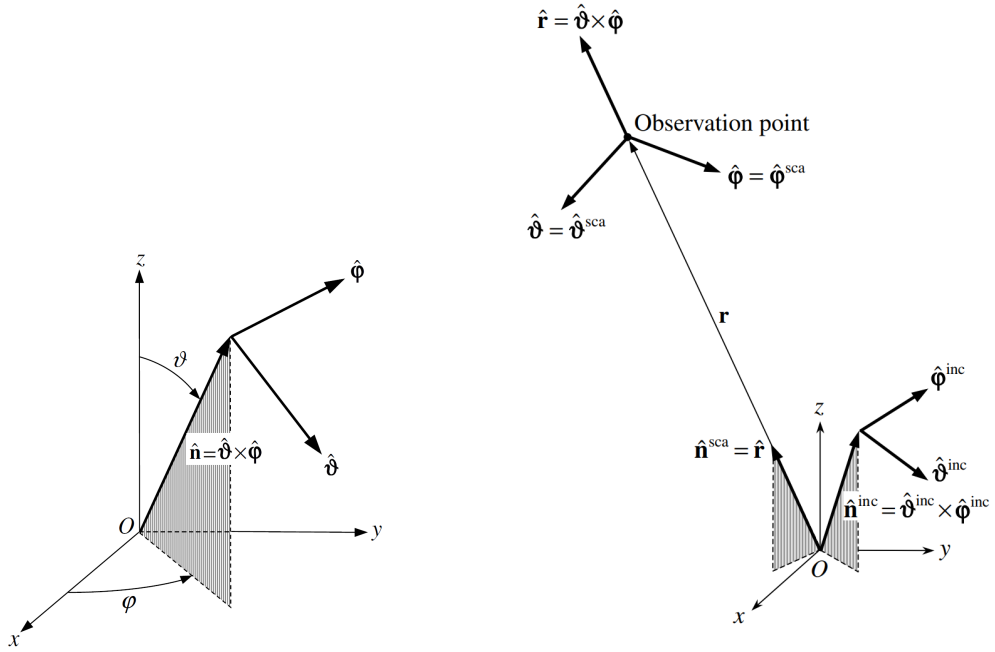


Figure 7.7: (a) Coordinate system used to describe the direction of propagation and the polarisation of a plane electromagnetic wave. (b) Scattering set up. Taken from [8].

that the above decomposition can be used to evaluate the scattered fields of the two incident waves separately and add them to find the total scattered field arising by the sum of the two incident waves.

The scattered far-field, $\mathbf{E}^{s,f}$, decays inversely with distance r from the scattering object

$$\mathbf{E}^{s,f}(\hat{\mathbf{n}}^{sca}) \sim \frac{e^{ik_e r}}{r}\mathbf{F}(\hat{\mathbf{n}}^{sca}), \quad r \to \infty. \tag{7.3}$$

The vector $\mathbf{F}$ is independent of r and describes the angular distribution of the scattered field radiation in the far-field zone. It can be computed via the asymptotic form of Stratton Chu representation formulae

$$\mathbf{F}(\mathbf{x}) = -\sum_{m}^{M} \mathcal{H}_m^{e,f}\left(\gamma_{D,m}^{+}\mathbf{E}_m^s\right) - \sum_{m}^{M} \mathcal{E}_m^{e,f}\left(\gamma_{N,m}^{+}\mathbf{E}_m^s\right), \tag{7.4}$$

where $\mathcal{H}^{e,f}$ and $\mathcal{E}^{e,f}$ are the far-field versions of the magnetic and electric

potential operators given as [5]

$$\mathcal{H}^{e,f}\mathbf{v}(\mathbf{x}) := \frac{ik_e}{4\pi}\int_\Gamma \exp\left(\frac{-ik_e\mathbf{x}\cdot\mathbf{y}}{|\mathbf{x}|}\right)\left(\frac{\mathbf{x}}{|\mathbf{x}|}\times\mathbf{v}(\mathbf{y})\right)d\Gamma(\mathbf{y}), \tag{7.5}$$

$$\begin{aligned}\mathcal{E}^{e,f}\mathbf{v}(\mathbf{x}) :=& \frac{ik_e}{4\pi}\int_\Gamma \exp\left(\frac{-ik_e\mathbf{x}\cdot\mathbf{y}}{|\mathbf{x}|}\right)\mathbf{v}(\mathbf{y})d\Gamma(\mathbf{y}) \\ &- \frac{ik_e}{4\pi}\int_\Gamma \exp\left(\frac{-ik_e\mathbf{x}\cdot\mathbf{y}}{|\mathbf{x}|}\right)\mathbf{v}(\mathbf{y})\cdot\frac{\mathbf{x}}{|\mathbf{x}|}d\Gamma(\mathbf{y}).\end{aligned} \tag{7.6}$$

7.2.1 Amplitude Scattering Matrix

The amplitude scattering matrix $\mathbf{S}$ describes the transformation of the $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ components of the incident plane wave into the $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ components of the scattered spherical wave in the far-field. It depends on the directions of incident and scattering waves as well as the size, morphology, composition, and orientation of the scattering object with respect to the coordinate system. It also depends on the choice of origin of the coordinate system inside the scattering object. The elements of the amplitude matrix can be evaluated as follows

$$S_{11} = \hat{\boldsymbol{\theta}}^{sca}\cdot\mathbf{F}^{s,f}(\hat{\mathbf{n}}^{sca},\hat{\mathbf{n}}^{inc},\hat{\boldsymbol{\theta}}^{inc}), \qquad S_{12} = \hat{\boldsymbol{\theta}}^{sca}\cdot\mathbf{F}^{s,f}(\hat{\mathbf{n}}^{sca},\hat{\mathbf{n}}^{inc},\hat{\boldsymbol{\phi}}^{inc}), \tag{7.7}$$

$$S_{21} = \hat{\boldsymbol{\phi}}^{sca}\cdot\mathbf{F}^{s,f}(\hat{\mathbf{n}}^{sca},\hat{\mathbf{n}}^{inc},\hat{\boldsymbol{\theta}}^{inc}), \qquad S_{22} = \hat{\boldsymbol{\phi}}^{sca}\cdot\mathbf{F}^{s,f}(\hat{\mathbf{n}}^{sca},\hat{\mathbf{n}}^{inc},\hat{\boldsymbol{\phi}}^{inc}). \tag{7.8}$$

By $\mathbf{F}^{s,f}(\hat{\mathbf{n}}^{sca},\hat{\mathbf{n}}^{inc},\hat{\boldsymbol{\theta}}^{inc})$ we mean the scattered far-field created by the $\hat{\boldsymbol{\theta}}^{inc}$ component of the incident plane wave, i.e. $E_\theta^{inc}\hat{\boldsymbol{\theta}}^{inc}\exp(ik_e\hat{\mathbf{n}}^{inc}\cdot\mathbf{r})$ and evaluated in the scattering direction $\hat{\mathbf{n}}^{sca}$. Analogous definitions can be made for $\mathbf{F}^{s,f}(\hat{\mathbf{n}}^{sca},\hat{\mathbf{n}}^{inc},\hat{\boldsymbol{\phi}}^{inc})$.

7.2.2 Phase matrix

The elements of the phase matrix, $\mathbf{Z}$, in terms of the amplitude matrix are

$$Z_{11} = \frac{1}{2}\left(|S_{11}|^2+|S_{12}|^2+|S_{21}|^2+|S_{22}|^2\right), \tag{7.9}$$

$$Z_{12} = \frac{1}{2}\left(|S_{11}|^2-|S_{12}|^2+|S_{21}|^2-|S_{22}|^2\right), \tag{7.10}$$

$$Z_{13} = -\Re\left(S_{11}S_{12}^*+S_{22}S_{21}^*\right), \tag{7.11}$$

$$Z_{14} = -\Im\left(S_{11}S_{12}^{*} - S_{22}S_{21}^{*}\right), \tag{7.12}$$

$$Z_{21} = \frac{1}{2}\left(|S_{11}|^2 + |S_{12}|^2 - |S_{21}|^2 - |S_{22}|^2\right), \tag{7.13}$$

$$Z_{22} = \frac{1}{2}\left(|S_{11}|^2 - |S_{12}|^2 - |S_{21}|^2 + |S_{22}|^2\right), \tag{7.14}$$

$$Z_{23} = -\Re\left(S_{11}S_{12}^{*} - S_{22}S_{21}^{*}\right), \tag{7.15}$$

$$Z_{24} = -\Im\left(S_{11}S_{12}^{*} + S_{22}S_{21}^{*}\right), \tag{7.16}$$

$$Z_{31} = -\Re\left(S_{11}S_{21}^{*} + S_{22}S_{12}^{*}\right), \tag{7.17}$$

$$Z_{32} = -\Re\left(S_{11}S_{21}^{*} - S_{22}S_{12}^{*}\right), \tag{7.18}$$

$$Z_{33} = \Re\left(S_{11}S_{22}^{*} + S_{12}S_{21}^{*}\right), \tag{7.19}$$

$$Z_{34} = \Im\left(S_{11}S_{22}^{*} + S_{21}S_{12}^{*}\right), \tag{7.20}$$

$$Z_{41} = -\Im\left(S_{21}S_{11}^{*} + S_{22}S_{12}^{*}\right), \tag{7.21}$$

$$Z_{42} = -\Im\left(S_{21}S_{11}^{*} - S_{22}S_{12}^{*}\right), \tag{7.22}$$

$$Z_{43} = \Im\left(S_{22}S_{11}^{*} - S_{12}S_{21}^{*}\right), \tag{7.23}$$

$$Z_{44} = \Re\left(S_{22}S_{11}^{*} - S_{12}S_{21}^{*}\right), \tag{7.24}$$

where $\Re$ and $\Im$ denote the real and imaginary parts respectively.

7.2.3 Single Scattering Properties (SSPs)

Knowledge of the scattered far field allows us to compute the optical properties of the scatterer. For the purposes of the database these are:

- the *extinction cross section*, C_{ext}: it refers to the area that when multiplied by the incident monochromatic energy flux, gives the total monochromatic power removed from the incident wave by the effect of scattering and absorption,

- the *scattering cross section*, C_{sca}: as above, but the product gives the total monochromatic power removed from the incident wave as a result of scattering of the incident radiation in all directions,

- the *back-scattering cross section*, C_{bsca}: similar to the scattering cross section but only for the scattering of the incident wave back in the di-

rection of the incident wave,

- the *asymmetry parameter*, $g \in [-1, 1]$: it is defined as the average cosine of the scattering angle (i.e. the angle between the incidence and the scattering directions). It is positive if the particle scatters more light toward the forward direction, negative if more light is scattered toward the back-scattering direction, and vanishes if the scattering is symmetric with respect to the plane perpendicular to the incident direction,

- the *scattering albedo*, $\varpi_0 \in [0, 1]$: it is interpreted as the probability that a photon interacting with the particle will be scattered rather than absorbed. In the case of non-absorbing particles it has a value of 1.

The extinction, scattering and back-scattering cross sections can be computed as follows

$$C_{ext} = \frac{1}{2}\left(C_{ext,\theta} + C_{ext,\phi}\right), \tag{7.25}$$

$$C_{sca} = \frac{1}{2}\left(C_{sca,\theta} + C_{sca,\phi}\right), \tag{7.26}$$

$$C_{bsca} = 4\pi Z_{11}(\theta^{sca} = \pi), \tag{7.27}$$

where the subscripts θ and ϕ correspond to the two components of the incident wave (as described in (7.2)). The individual components can be computed as follows

$$C_{ext,x} = \frac{4\pi}{k_e |E_x^{inc} \hat{\boldsymbol{x}}^{inc}|^2} \Im\left(\mathbf{F}^{s,f}(\hat{\mathbf{n}}^{sca}, \hat{\mathbf{n}}^{inc}, \hat{\boldsymbol{x}}^{inc}) \cdot \left(E_x^{inc} \hat{\boldsymbol{x}}^{inc}\right)^*\right), \tag{7.28}$$

$$C_{sca,x} = \frac{1}{|E_x^{inc} \hat{\boldsymbol{x}}^{inc}|^2} \int_{4\pi} |\mathbf{F}^{s,f}(\hat{\mathbf{r}}, \hat{\mathbf{n}}^{inc}, \hat{\boldsymbol{x}}^{inc})|^2 d\hat{\mathbf{r}}, \tag{7.29}$$

where $\hat{\boldsymbol{x}}^{inc}$ corresponds to one of $\hat{\boldsymbol{\theta}}^{inc}$ or $\hat{\boldsymbol{\phi}}^{inc}$. The scattering albedo is defined as the ratio of the scattering and extinction cross sections

$$\varpi_0 = \frac{C_{sca}}{C_{ext}}. \tag{7.30}$$

The asymmetry parameter $g = \langle \cos\Theta \rangle$ is defined as the average cosine of the scattering angle $\Theta = \arccos(\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}^{inc})$, i.e. the angle between the incidence and the scattering directions

$$g = \langle \cos\Theta \rangle = \frac{1}{2}\left(g_\theta + g_\phi\right), \tag{7.31}$$

where the individual components can be computed by

$$g_x = \frac{1}{C_{sca,x}} \int_{4\pi} |\mathbf{F}^{s,f}(\hat{\mathbf{r}}, \hat{\mathbf{n}}^{inc}, \hat{\boldsymbol{x}}^{inc})|^2 \hat{\mathbf{r}} \cdot \hat{\mathbf{n}}^{inc} d\hat{\mathbf{r}}. \tag{7.32}$$

We note that there are alternative ways of computing C_{sca} and g through the phase matrix element Z_{11} and the relative difference between the two definitions is often used to check the accuracy of the implementation

$$C_{sca} = \int_0^{2\pi} \int_0^{\pi} Z_{11}(\theta) \sin\theta d\theta d\phi, \tag{7.33}$$

$$g = \int_0^{2\pi} \int_0^{\pi} Z_{11}(\theta) \sin\theta \cos\theta d\theta d\phi. \tag{7.34}$$

7.2.3.1 Integrals over the sphere: Lebedev quadrature

For the evaluation of integrals over the sphere (e.g. for C_{sca} and g) we use Lebedev quadrature [111, 112, 113, 114, 115, 116]. This is to take advantage of the discrete grid points (θ_i, ϕ_i) being available as a pair and hence reducing the number of loops to one.

We briefly introduce the method but refer to the original sources [111, 112, 113, 114, 115, 116] for a more detailed discussion. A surface integral of a function over the unit sphere is evaluated by

$$I[f] = \int_{4\pi} f(\hat{\mathbf{r}}) d\hat{\mathbf{r}} = \int_0^{\pi} sin(\theta) d\theta \int_0^{2\pi} f(\theta, \phi) d\phi. \tag{7.35}$$

Let S be a unit sphere in $\mathbb{R}^3$ and let Ω be the octahedron inscribed in it, with

vertices on the coordinate axes as follows

$$\mathbf{a}_i^1 = (0,0,\pm 1),(0,\pm 1,0),(\pm 1,0,0). \tag{7.36}$$

The Lebedev grid points are constructed to lie on the surface of S and to be invariant under the octahedral rotation group with inversion [111]. For any point on the sphere, there are either five, seven, eleven, twenty-three, or forty-seven equivalent points with respect to the octahedral group given by $\mathbf{a}_i^1$ (defined above) and $\mathbf{a}_i^3$, $\mathbf{a}_i^2$, $\mathbf{b}_i^k$, $\mathbf{c}_i^k$, $\mathbf{d}_i^k$ respectively. These are defined as [112]

- $\mathbf{a}_i^2 = \frac{\sqrt{2}}{2}(\pm 1,\pm 1,0)$, $\frac{\sqrt{2}}{2}(\pm 1,0,\pm 1)$ and $\frac{\sqrt{2}}{2}(0,\pm 1,\pm 1)$, representing the mid-points of the edges of the octahedron Ω,
- $\mathbf{a}_i^3 = \frac{\sqrt{3}}{3}(\pm 1,\pm 1,\pm 1)$, representing the centres of the faces of Ω,
- $\mathbf{b}_i^k : (\pm\ell_k,\pm\ell_k,\pm m_k)$, $(\pm\ell_k \pm m_k,\pm\ell_k,)$, and $(\pm m_k,\pm\ell_k,\pm\ell_k)$, with the constraint $2\ell_k^2 + m_k^2 = 1$,
- $\mathbf{c}_i^k = (\pm p_k,\pm q_k,0)$, $(\pm p_k,0,\pm q_k)$, $(0,\pm p_k,\pm q_k)$, $(\pm q_k,\pm p_k,0)$, $(\pm q_k,0,\pm p_k)$, $(0,\pm q_k,\pm p_k)$, with the constraint $p_k^2 + q_k^2 = 1$, and
- $\mathbf{d}_i^k = (\pm r_k,\pm u_k,\pm w_k)$, $(\pm r_k,\pm w_k,\pm u_k)$, $(\pm u_k,\pm r_k,\pm w_k)$, $(\pm u_k,\pm w_k,\pm r_k)$, $(\pm w_k,\pm u_k,\pm r_k)$, $(\pm w_k,\pm r_k,\pm u_k)$ with the constraint $r_k^2 + w_k^2 + u_k^2 = 1$.

For the Lebedev scheme all points equivalent under the rotational and inversion group share the same weights; $A_1,\ A_2,\ A_3,\ B_k,\ C_k,\ D_k$ for $\mathbf{a}_i^1$, $\mathbf{a}_i^2\,\mathbf{a}_i^3$, $\mathbf{b}_i^k\,\mathbf{c}_i^k$, $\mathbf{d}_i^k$ respectively, and the general form of the Lebedev quadrature is given by

$$\begin{aligned}\tilde{I}_N[f] =& A_1 \sum_{i=1}^{6} f(\mathbf{a}_i^1) + A_2 \sum_{i=1}^{12} f(\mathbf{a}_i^2) + A_3 \sum_{i=1}^{8} f(\mathbf{a}_i^3) \\ &+ \sum_{k=1}^{N_1} B_k \sum_{i=1}^{24} f(\mathbf{b}_i^k) + \sum_{k=1}^{N_2} C_k \sum_{i=1}^{24} f(\mathbf{c}_i^k) + \sum_{k=1}^{N_3} D_k \sum_{i=1}^{48} f(\mathbf{d}_i^k),\end{aligned} \tag{7.37}$$

with the total number of grid points given by

$$N = 26 + 24(N_1 + N_2) + 48N_3. \tag{7.38}$$

To determine the weights one has to integrate the invariant spherical harmonics up to a given order and solve the resulting non-linear equations.

Construction of the Lebedev grid points and weights can be expensive as one has to solve large systems of non-linear equations but we take advantage of pre-computed numerical values for the nodes and weights [2] [117]. For the evaluation of C_{sca} and g we use Lebedev quadrature of order 59 resulting in 1202 grid points.

Compared to other methods, such as using products of one-dimensional Gaussian quadratures approximating the integral by

$$\tilde{I}[f] = \sum_{i=1}^{N}\sum_{j=1}^{M} \omega_i v_j f(\theta_i, \phi_j), \tag{7.39}$$

for appropriate weight functions ω_i and v_j, Lebedev quadrature is faster to evaluate as fewer grid points are used and the two loops reduce down to one.

7.2.4 Random orientation

The aggregates of Figures 7.4–7.6 are assumed to be randomly oriented and therefore in this section we discuss how the SSPs and phase matrix of randomly oriented scatterers can be evaluated. Averaging some quantity f over random particle orientations is represented by [118]

$$\langle f \rangle = \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} f(\alpha, \beta, \gamma) \sin\beta \; d\alpha \; d\beta \; d\gamma. \tag{7.40}$$

where α, β, γ are the three Euler angles used to describe the orientation of a particle (see Figure 7.8). The traditional approach to evaluate the SSPs and phase matrix of randomly oriented particles is to fix the direction of the incident wave, and then rotate the scatterer using three Euler angles (as described for example in [119]). The Euler angles and coordinate systems before and after the transformations can be seen in Figure 7.8. The rotation matrix

[2] available on `bitbucket.org/CasperBeentjes/quadratures-on-unit-sphere`

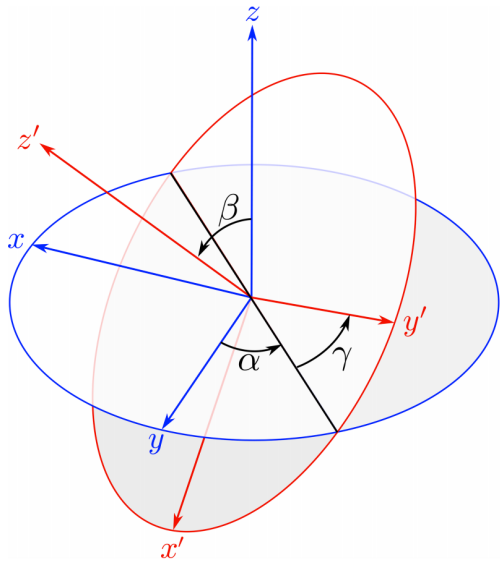


Figure 7.8: Euler angles of rotation α, β and γ. The blue coordinate system, xyz, represents the original coordinate system, while the red one, $x'y'z'$, represents the coordinate system obtained after the three rotations have been applied. Taken from [9].

representing the three rotations is given by

$$\begin{aligned} R_{\alpha\beta\gamma} &= R_z(\gamma)R_y(\beta)R_z(\alpha) \\ &= \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{7.41}$$

The SSPs and phase matrix are then evaluated for some scattering angles $(\theta^{sca}, \phi^{sca})$. Once many orientations have been considered one can use (7.40) to compute the orientationally averaged quantity.

A naive approach to this would be to re-assemble the operator and preconditioner for each orientation considered, however this would be very expensive using BEM. Instead, two reference frames are often used when considering multiple orientations; a 'laboratory' reference frame and a particle one [8]. The laboratory frame is often chosen so that it corresponds to the geometry of the scattering situation, i.e., the Cartesian coordinate system. The particle reference frame is used to describe the orientation of the scatterer with respect

to the laboratory frame. In order to solve the scattering problem with respect to the laboratory reference frame, one must first solve in the particle reference frame and then through suitable rotation matrices transition to the laboratory frame. Full details of this process and explicit formulas for the necessary rotation matrices are given in Chapter 2.4 of [8]. Fixing the direction of the incident wave and rotating the scatterer by $R_{\alpha\beta\gamma}$ is equivalent to fixing the orientation of the scatterer and instead rotating the direction of the incident wave by $R^{-1}_{\alpha\beta\gamma}$; and this methodology has also been considered in [9]. To avoid transitioning through particle and laboratory frames we instead follow this approach. In this case, we assemble the operator matrix and preconditioner and re-use for all incident wave solutions. We note that for each incident wave, two GMRES solves need to be performed; one for each polarisation vector as described in (7.1)–(7.2). However, different GMRES solves can be distributed over different CPUs to reduce the total computation time.

Assuming a fixed incident wave direction with fixed $\theta^{inc} = \phi^{inc} = 0$ in (7.1) (which is what is usually assumed in the traditional way of simulating random orientation), gives

$$\hat{\mathbf{n}}^{inc} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \hat{\boldsymbol{\theta}}^{inc} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \hat{\boldsymbol{\phi}}^{inc} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}. \tag{7.42}$$

These can then be rotated by $R^{-1}_{\alpha,\beta,\gamma}$ given by

$$R^{-1}_{\alpha,\beta,\gamma} = R_z^{-1}(\alpha) R_y^{-1}(\beta) R_z^{-1}(\gamma), \tag{7.43}$$

giving the rotated incident directions and polarisation vectors

$$\hat{\mathbf{n}}^{inc} = \begin{bmatrix} \cos\alpha \sin\beta \\ \sin\alpha \sin\beta \\ \cos\beta \end{bmatrix}, \quad \hat{\boldsymbol{\theta}}^{inc} = \begin{bmatrix} \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma \\ \sin\alpha\cos\beta\cos\gamma + \cos\alpha\sin\gamma \\ -\sin\beta\cos\gamma \end{bmatrix}, \tag{7.44}$$

$$\hat{\boldsymbol{\phi}}^{inc} = \begin{bmatrix} -\cos\alpha\cos\beta\sin\gamma - \sin\alpha\cos\gamma \\ -\sin\alpha\cos\beta\sin\gamma + \cos\alpha\cos\gamma \\ \sin\beta\sin\gamma \end{bmatrix}, \tag{7.45}$$

that can be varied for different α, β, γ. Comparing with the definitions of (7.1), we can see that both $\hat{\mathbf{n}}^{inc}$ are equivalent with $\alpha = \phi^{inc}$ and $\beta = \theta^{inc}$, i.e. the direction of the incident wave does not depend on the angle γ and can be defined using a pair of grid points $(\theta^{inc}, \phi^{inc})$. It is only the polarisation vectors $\hat{\boldsymbol{\theta}}^{inc}$ and $\hat{\boldsymbol{\phi}}^{inc}$ that need to be altered by an additional angle.

This allows us to make use of Lebedev quadrature to obtain the grid points $(\theta_i^{inc}, \phi_i^{inc})$ that define the direction of the incident wave, reducing the two integrals over θ^{inc} and ϕ^{inc} (or α and β as defined in (7.40)) to just one. The third integral over γ can then be evaluated by a simple Gaussian integration. For the rest of this thesis we use the terms '**number of incident waves**' to refer to the number of pairs $(\theta_i^{inc}, \phi_i^{inc})$ obtained by the Lebedev scheme, and '**number of polarisation rotations/vectors**' to refer to the number of points γ_i obtained by the Gaussian quadrature scheme.

7.2.4.1 Testing our interpretation of random orientation

To test the accuracy of our implementation of random orientation as well as to find the number of incident waves and polarisation rotations that are needed to approximate random orientation, we compare with results obtained from a T-matrix method [22]. We consider scattering by hexagonal columns of different size parameters as detailed in Table 7.2.

In Figure 7.9 we present the relative errors of SSPs computed for the problems of Table 7.2, for a number of incidents waves and for different mesh resolutions. We find that the prescribed mesh size of 10 elements per wavelength suggested in [5] is not a useful scheme for all the hexagonal columns considered, as some are very small and require a refined mesh size to achieve a sufficient discretisation of the surfaces. In particular, this is the case for $X = 0.05$ and $X = 0.1$, as the maximum mesh size occurring from a rule of 10

X	Dimensions	Frequency	Refractive Index
0.05	$L = D = 96\mu m$	50 GHz	$1.7746 + 0.00064\mathrm{i}$
0.1	$L = D = 191\mu m$		
1	$L = D = 1.91mm$		
5	$L = D = 9.55mm$		
10	$L = D = 1.04mm$	664 GHz	$1.7746 + 0.00940\mathrm{i}$

Table 7.2: Information on the geometry, frequency and refractive index for the test cases. The top four cases have the same frequency and refractive index. The height of the hexagonal column is denoted by L, with the diameter of the hexagonal face by D. We refer back to Figure 6.1 for a schematic representation of the hexagonal columns.

elements per wavelength is larger than the length L of the columns, resulting in the minimum amount of elements per surface produced by Gmsh. A mesh size of 100 elements per wavelength was sufficient to reduce the relative errors to a few percent. We also observe that 14 incident waves in combination with the refined mesh are sufficient to provide accurate SSPs at a relative error of 1% or below, for size parameters $X = 0.05$ and 0.1. For size parameter $X = 1$, 14 waves are again sufficient, however one can get away with the usual 10 elements per wavelength. A discretisation of 20 elements per wavelength increases accuracy but this is at the expense of additional computation time and memory consumption.

For the hexagonal column of size parameter $X = 5$ we see that a mesh size of 10 elements per wavelength produces relative errors between 1-2%. A mesh size of 20 elements per wavelength reduces those relative errors below 1%. There is no obvious change in assuming 110 or 194 incident waves, except for a small increase in the relative error of $\langle C_{bsca} \rangle$ when 20 elements per wavelength are used. For the hexagonal column of size parameter $X = 10$, a mesh size of 10 elements per wavelength results in relative errors between 1-4%. These fall below 1% when a mesh of 20 elements per wavelength is used, except for $\langle g \rangle$ at 2% and $\langle C_{bsca} \rangle$ at 3%. Our numerical experiments show that the accuracy of the SSPs is not affected by varying γ; these can be computed accurately just by varying $(\theta^{inc}, \phi^{inc})$.

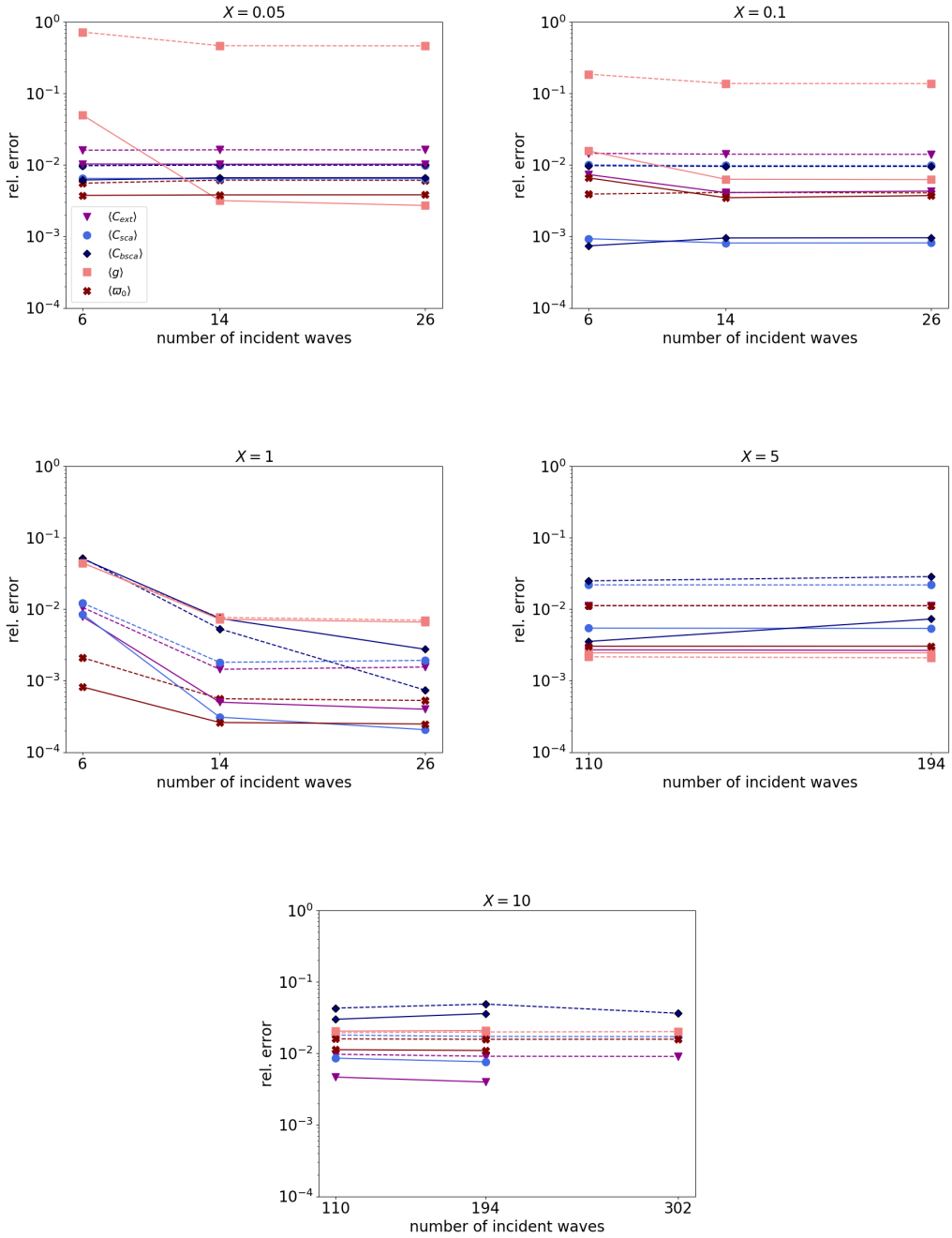
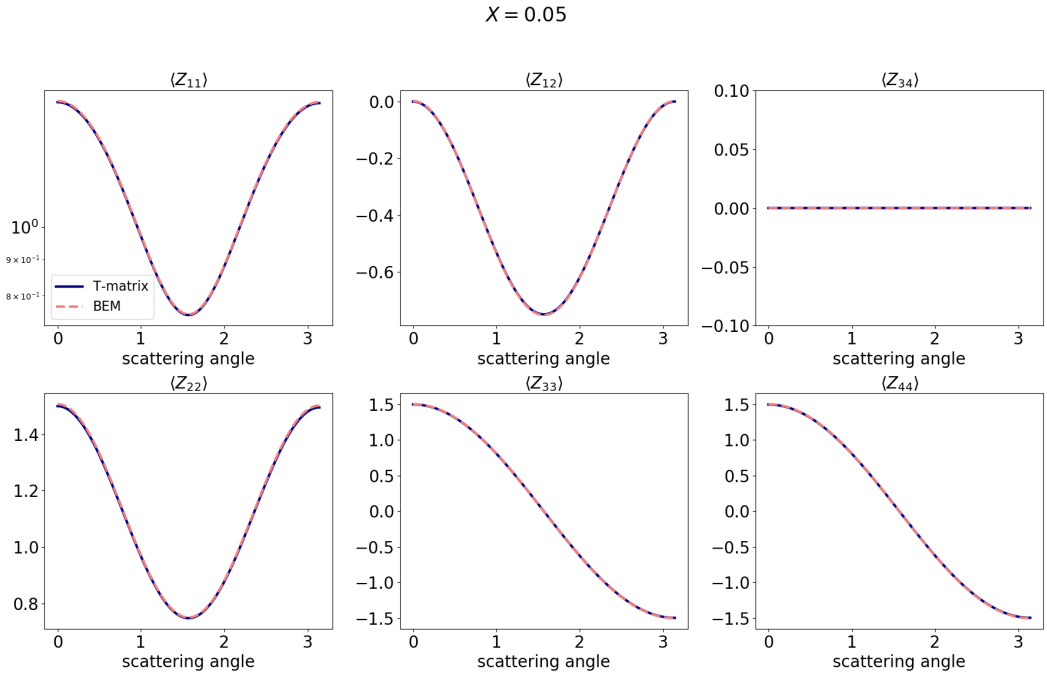
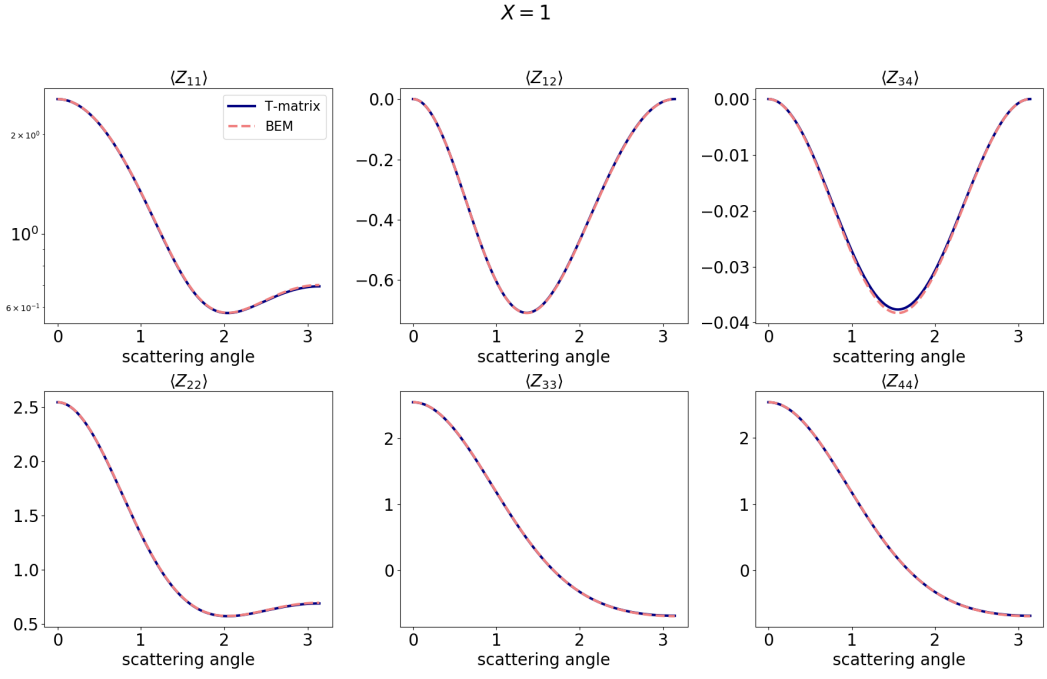


Figure 7.9: Relative errors of the SSPs as a function of increasing number of incident waves for hexagonal columns of size parameter $X = 0.05$ (top left), $X = 0.1$ (top right), $X = 1$ (centre left), $X = 5$ (centre right) and $X = 10$ (bottom). A mesh size of 10 elements per wavelength was prescribed for the dashed lines resulting in 54, 54, 270, 4452 and 9612 dofs respectively. The solid lines represent refined meshes with 100 elements per wavelength for $X = 0.05$ and $X = 0.1$, and 20 elements per wavelength for $X = 1$, 5 and 10 resulting in 90, 270, 918, 17478 and 21378 dofs respectively.

In Figures 7.10–7.11 we compare the phase matrix elements computed using our boundary element method against the T-matrix method of [22]. We note that the only elements of the phase matrix that remain non-zero when considering random orientation are $\langle Z_{11}\rangle$, $\langle Z_{22}\rangle$, $\langle Z_{33}\rangle$, $\langle Z_{44}\rangle$, $\langle Z_{12}\rangle$ $(=\langle Z_{21}\rangle)$ and $\langle Z_{34}\rangle$ $(=-\langle Z_{43}\rangle)$ and hence we only present comparisons for those. For the small size parameters $X=0.05$ and $X=1$ considered in Figure 7.10 we see a complete agreement between the methods with just 14 incident waves, in agreement with earlier findings on the accuracy of SSPs. We note though that a different number of polarisation rotations had to be considered to achieve this accuracy; 5 for $X=0.05$ and 10 for $X=1$. For the larger size parameters $X=5$ and $X=10$ of Figure 7.11, 194 incident waves were considered for both cases, but 10 polarisation rotations for $X=5$ and 15 for $X=10$. A good agreement is observed between the BEM and T-matrix methods although not a complete match for some parts of $\langle Z_{11}\rangle$ and $\langle Z_{12}\rangle$. This could be remedied by using a finer mesh or by assuming a larger number of incident waves/polarisation rotations but this level of accuracy was deemed acceptable at the testing stage, as our implementation is tested again in the following section for the aggregate model of Figures 7.4–7.6.

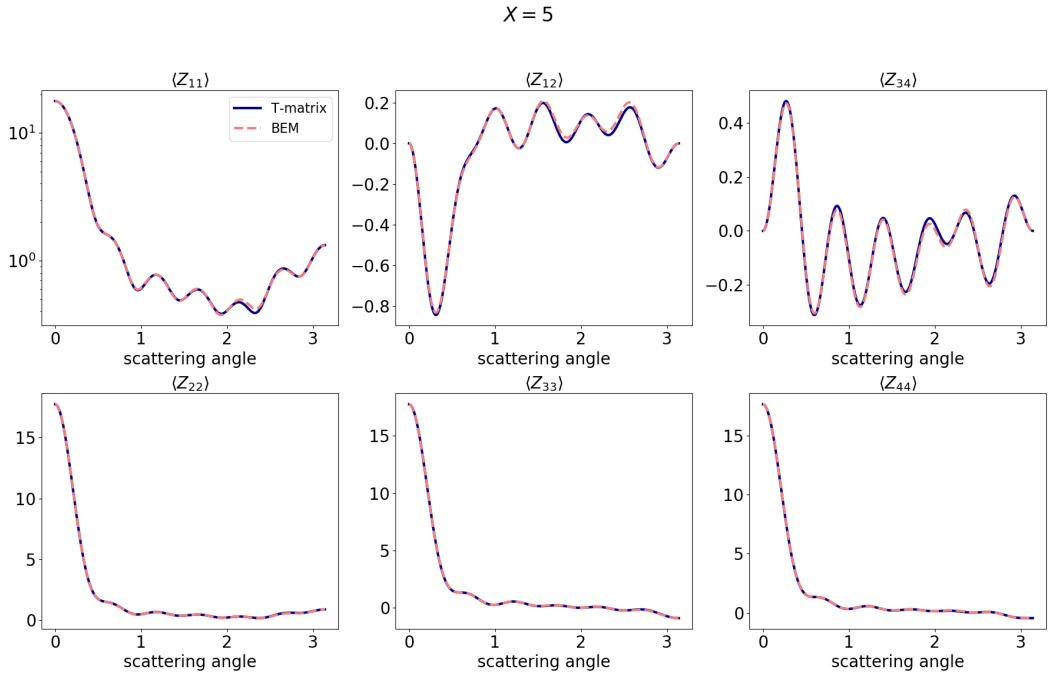


(a)

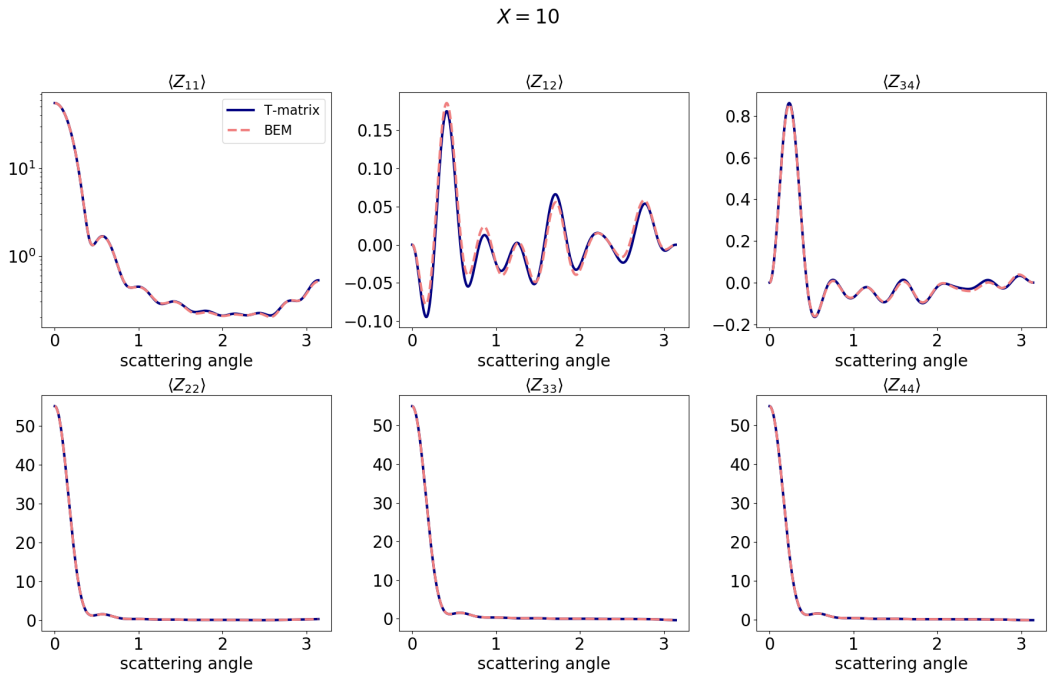


(b)

Figure 7.10: Phase matrix elements for hexagonal columns of size parameter (a) $X = 0.05$ and (b) $X = 1$. For (a) 14 incident waves and 5 polarisation rotations have been considered, while for (b) 14 incident waves and 10 polarisation rotations. We note that the remaining phase matrix elements are 0 when random orientation is assumed.



(a)



(b)

Figure 7.11: As in Figure 7.10 but for (a) $X = 5$ and (b) $X = 10$. In (a) 194 incident waves and 10 polarisations are considered and in (b) 194 incident waves and 15 polarisation rotations.

7.3 Accuracy

The question now arises as to how many incident waves and polarisation directions are needed and what level of mesh refinement is required to obtain the SSPs and phase matrix elements of the aggregate model in random orientation. In particular, we need to test if our findings from the comparisons between T-matrix methods for hexagonal columns carry over to more complex shapes such as the budding rosettes and rosette aggregates of our model. For the purposes of the database, accuracy of a few percent is desired and ideally around 1%. This is to minimise costs from memory consumption, and total computation time, since the final simulations were carried out on AWS machines. Since no other (numerical or analytical) solutions exist for the aggregate model to compare with, we compare results between different mesh discretisations, number of incident waves and polarisation rotations.

7.3.1 Mesh size

With regards to mesh refinement, we find that the usual rule of 10 elements per wavelength [5] is not always a useful scheme as each aggregate consists of multiple smaller monomers, for which a sufficient discretisation needs to be achieved for each surface. Taking a look at the aggregate model in Figures 7.4–7.6, we see our aggregates fall into 4 main categories, on which we also base our discretisation rules:

- $D_{max} = 10 - 492\mu m$: These are all re-scaled versions of $D_{max} = 492\mu m$ implying that the surfaces of each monomer shrink as D_{max} decreases. A minimum number of elements is imposed on such aggregates. After testing, these are: roughly 800 for 50 GHz, 4000 for 183 GHz, 8000 for 243 GHz and 13 000 for 664 GHz.

- $D_{max} = 547 - 875\mu m,\ 1111 - 1190\mu m,\ 1510 - 1630\mu m$: These are budding rosettes with very short branches. Each aggregate consists of identical monomers that are aligned in a different way to form the aggregate (see for example aggregates of $D_{max} = 547\mu m$ and $621\mu m$ in Figure 7.4;

both consist of 6 identical monomers aggregated together in a different manner as to produce the different maximum dimensions). A constant number of elements per wavelength is be used. After testing, these are: 100 elements per wavelength for 50 GHz, 80 for 183, 70 for 243 and 40 for 664 GHz.

- $D_{max} = 958 - 1045\mu m$, $1221 - 1461\mu m$, $1700 - 1759\mu m$, $2258\mu m$, $4115\mu m$, $4539\mu m$ and $5791\mu m$: These are budding rosettes with longer branches compared to the above category. Again, the individual monomers are identical but aggregated together in a different manner as for the above category (see for example aggregates of $D_{max} = 958\mu m$ and $1008\mu m$ in Figure 7.4). A constant number per wavelength is again prescribed but slightly relaxed compared to the above category, owing to the larger surfaces of the individual monomers. These are: 100 elements for 50 GHz, 50 for 183, 40 for 243 and 20 for 664 GHz.

- Remaining D_{max}: These are rosette aggregates, with each individual monomer being longer than the previous two categories. A relaxed mesh rule can be applied to those. We use 50 elements per wavelength at 50 GHz, 20 at 183, 20 at 243 and 10 or 20 at 664 GHz (we explain below).

Comparison between different mesh resolutions is performed for some test cases at all frequencies. In Tables 7.3–7.6 we present some of those test cases at 664 GHz and 190 K, each table representing one of the above four categories.

We note that at Table 7.6, a relative difference of 1-2% is observed for some of the scattering cross-sections. For aggregates $D_{max} > 4000\mu m$, at 664 GHz, the choice was made to run the simulations with a mesh size of 10 elements per wavelength despite the relative difference being more than 1%. This is because as D_{max} increases, so does the size parameter X and as seen in the following subsections, the number of incident waves and polarisation rotations need to increase with increasing X. Since the relative difference between the two mesh resolutions was below 5% it was deemed acceptable to proceed with

D_{max}	$60\mu m$	$200\mu m$	$350\mu m$	$450\mu m$	$492\mu m$
$\langle C_{ext} \rangle$	0.02%	0.38%	0.05%	0.31%	0.37%
$\langle C_{sca} \rangle$	0.06%	0.08%	0.10%	0.11%	0.12%
$\langle C_{bsca} \rangle$	0.05%	0.06%	0.03%	0.13%	0.27%
$\langle g \rangle$	0.17%	0.05%	0.04%	0.01%	0.03%
$\langle \varpi_0 \rangle$	0.10%	0.32%	0.07%	0.18%	0.46%
# of dofs N	12828	12843	12825	12852	12624
	20496	**20382**	**20454**	**20400**	**20172**

Table 7.3: Relative difference between two different discretisation schemes for the smaller aggregates sizes $D_{max} \leq 492\mu m$ at 664 GHz, 190 K, with 14 incident waves. Bold font indicates the discretisation chosen for the database.

D_{max}	$547\mu m$	$697\mu m$	$1111\mu m$	$1190\mu m$	$1510\mu m$
$\langle C_{ext} \rangle$	0.60%	0.42%	0.22%	0.18%	0.25%
$\langle C_{sca} \rangle$	0.16%	0.10%	0.14%	0.18%	0.10%
$\langle C_{bsca} \rangle$	0.52%	0.66%	0.12%	0.09%	0.19%
$\langle g \rangle$	0.06%	0.06%	0.03%	0.09%	0.09%
$\langle \varpi_0 \rangle$	0.43%	0.33%	0.34%	0.36%	0.35%
# of dofs N	9603	14409	28749	38391	43359
	18438	**23790**	**47556**	**63057**	**71736**

Table 7.4: Relative difference between two different discretisation schemes (30 and 40 elements per wavelength) for some of the short budding rosettes at 664 GHz, 190 K, with 194 incident waves. We note that relative differences between 20 and 30 elements per wavelength were all above 1% (and as high as 6% in some cases). Bold font indicates the discretisation chosen for the database.

D_{max}	$958\mu m$	$1221\mu m$	$1700\mu m$	$2258\mu m$
$\langle C_{ext} \rangle$	0.05%	0.09%	0.11%	0.11%
$\langle C_{sca} \rangle$	0.25%	0.36%	0.37%	0.39%
$\langle C_{bsca} \rangle$	0.18%	0.40%	0.58%	0.16%
$\langle g \rangle$	0.04%	0.04%	0.05%	0.05%
$\langle \varpi_0 \rangle$	0.20%	0.26%	0.26%	0.29%
# of dofs N	3096	4380	8721	13989
	8196	**12279**	**24372**	**39498**

Table 7.5: Relative difference between two different discretisation schemes (10 and 20 elements per wavelength) for some of the longer budding rosettes at 664 GHz, 190 K, with 194 incident waves. Bold font indicates the discretisation chosen for the database.

D_{max}	$2121\mu m$	$3004\mu m$	$5508\mu m$	$6456\mu m$	$8582\mu m$
$\langle C_{ext} \rangle$	0.59%	0.75%	0.77%	0.78%	0.81%
$\langle C_{sca} \rangle$	1.22%	1.29%	1.46%	1.50%	1.55%
$\langle C_{bsca} \rangle$	0.70%	0.46%	0.65%	1.14%	1.68%
$\langle g \rangle$	0.11%	0.28%	0.17%	0.19%	0.19%
$\langle \varpi_0 \rangle$	0.63%	0.54%	0.69%	0.72%	0.74%
# of dofs N	6606	10560	**44127**	**59358**	**65631**
	24774	**39189**	165951	220599	244347
# of waves	194	230	230	230	230

Table 7.6: Relative difference between two different discretisation schemes (10 and 20 elements per wavelength) for some of the aggregate rosettes at 664 GHz, 190 K. Red underlined values indicate that the relative difference is above 1% which is the desired accuracy. Bold font indicates the discretisation chosen for the database.

the mesh size of 10 elements per wavelength to reduce computational cost. The probability of occurrence of ice crystals with such large dimensions is also low, so a greater inaccuracy in the computation of SSPs is not going to have as big an impact as the intermediate ones.

In addition, the SSPs and phase matrix elements of aggregates of size parameter $X = 0.2$ or smaller were computed using the Rayleigh approximation [120] for equivalent mass ice spheres. The asymmetry parameters $\langle g \rangle$ were computed from the T-matrix method of [121] otherwise the asymmetry parameters under Rayleigh approximation would be zero. Contributions of such small size parameter ice crystals are very small so the choice was made to reduce computation time and cost and focus on the remaining size parameters that have a larger contribution in numerical weather and climate models. The aggregate sizes for which the Rayleigh approximation was used are

- 50 GHz: $D_{max} = 10 - 450\mu m$,
- 183 GHz: $D_{max} = 10 - 80\mu m$,
- 243 GHz: $D_{max} = 10 - 60\mu m$, and
- 664 GHz: $D_{max} = 10 - 20\mu m$.

7.3.2 Number of incident waves

With regards to the number of incident waves required to achieve accuracy of a few percent of the SSPs, we find that these are dependent on the size parameter. For our aggregate model we use

- $X < 1$: 14 incident waves (Lebedev scheme of order 5),
- $1 \leq X < 3$: 50 incident waves (Lebedev scheme of order 11),
- $3 \leq X < 5$: 110 incident waves (Lebedev scheme of order 17),
- $5 \leq X < 8$: 194 incident waves (Lebedev scheme of order 23),
- $8 \leq X < 10$: 230 incident waves (Lebedev scheme of order 25),
- $X \geq 10$: 302 incident waves (Lebedev scheme of order 29).

Numerical tests for some of the aggregates at 664 GHz can be seen in Tables 7.7–7.8. We note that it is likely that as X increases further than 10, the number of incident waves will also have to increase. However, for the purposes of the database we limit the number to 302 to limit computation costs. We note that all SSPs except for $\langle C_{bsca} \rangle$ are approximated well at about 50 incident waves and it is the back-scattering cross-section that requires the additional incident waves. For future simulations where the back-scattering cross-section is not of interest, computational costs can be reduced by reducing the number of incident waves. We also note, that consistent with the comparisons of T-matrix results for hexagonal columns, no polarisation rotations are required for the accurate computation of SSPs.

D_{max}	$60\mu m$	$200\mu m$	$350\mu m$		$492\mu m$	
$\langle C_{ext} \rangle$	0.001%	0.009%	0.012%	0.000%	0.000%	0.000%
$\langle C_{sca} \rangle$	0.035%	0.009%	0.016%	0.000%	0.000%	0.000%
$\langle C_{bsca} \rangle$	0.039%	0.323%	9.448%	0.063%	2.442%	0.008%
$\langle g \rangle$	0.048%	0.006%	0.274%	0.000%	0.004%	0.000%
$\langle \varpi_0 \rangle$	0.068%	0.001%	0.007%	0.000%	0.000%	0.000%
# of waves	**14**	**14**	14	**50**	50	**110**
	50	50	50	110	110	194
X	0.41	1.31	2.43		3.42	
# of dofs N	20496	20382	20454		20172	

Table 7.7: Relative difference between different number of incident waves for the smaller aggregates sizes $D_{max} \leq 492\mu m$ at 664 GHz, 190 K. Red underlined values indicate that the relative difference is above 1% which is the desired accuracy. A relative difference of 0.000% indicates that the relative difference was below 0.001%. Bold font indicates the number of waves chosen for the database.

D_{max}	$547\mu m$	$697\mu m$	$805\mu m$	$1111\mu m$	$1190\mu m$	$1461\mu m$
$\langle C_{ext} \rangle$	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
$\langle C_{sca} \rangle$	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
$\langle C_{bsca} \rangle$	0.001%	0.249%	1.263%	0.198%	1.217%	2.618%
$\langle g \rangle$	0.000%	0.000%	0.001%	0.000%	0.09%	0.001%
$\langle \varpi_0 \rangle$	0.000%	0.000%	0.000%	0.000%	0.36%	0.000%
# of waves	**110**	**110**	110	**194**	194	230
	194	194	**194**	230	**230**	**302**
X	3.80	4.84	5.59	7.73	8.28	10.2
# of dofs N	18438	23790	24033	47556	63057	16920

Table 7.8: Relative difference between different number of incident waves for some of the short budding rosettes at 664 GHz, 190 K. Red underlined values indicate that the relative difference is above 1% which is the desired accuracy. A relative difference of 0.000% indicates that the relative difference was below 0.001%. Bold font indicates the number of waves chosen for the database.

7.3.3 Number of polarisation rotations

To evaluate the phase matrix elements of randomly oriented aggregates, in addition to the number of incident waves, we also need to consider the number of the polarisation rotations. After testing we find that those again depend on the size parameter as follows

- aggregates $X < 1$: 10 rotations,
- aggregates $1 \leq X < 24$: 15 rotations,
- aggregates $X \geq 24$: 20 rotations.

A selection of numerical experiments validating the above choice is shown in Table 7.9. We present the relative difference between the scattering cross-section, $\langle C_{sca} \rangle$, and the asymmetry parameter, $\langle g \rangle$, calculated through the far field (from equations (7.29) and (7.32)) and calculated through $\langle Z_{11} \rangle$ (from equations (7.33) and (7.34)) for some of the model aggregates at 664 GHz and 190 K. We note that a higher relative difference is shown for $D_{max} = 150\mu m$ but looking closer at the actual values $\langle g \rangle = 0.1149$ and $\langle g \rangle = 0.1111$ they are still in agreement in up to 2 decimal places.

Again, it is likely that more polarisation rotations will be needed as X grows but in combination with the number of incident waves the computational cost becomes increasingly high. For the largest aggregate sizes, 302×20 incident waves will have to be considered. These correspond to $2 \times 302 \times 20$ GMRES solves (recall the decomposition of the incident wave in (7.2)) that need to be performed for each aggregate of $X \geq 24$, where the number of degrees of freedom also increases with D_{max}. For the purposes of the database we therefore limit the number of polarisation rotations to 20, but the reader should bear in mind that this might imply a reduced accuracy.

D_{max}	$150\mu m$	$697\mu m$	$1422\mu m$	$2121\mu m$	$3443\mu m$	$4346\mu m$
$\langle C_{sca} \rangle$	0.06%	0.06%	0.11%	0.24%	0.29%	0.23%
$\langle g \rangle$	3.29%	0.93%	0.29%	0.09%	0.17%	0.22%
# of polarisations	10	15	15	15	15	20
# of waves	50	110	194	230	230	302
X	1.04	4.85	9.89	14.75	23.94	30.22

Table 7.9: Relative difference between C_{sca} and g calculated through the far field (from equations (7.29) and (7.32)) and calculated through $\langle Z_{11} \rangle$ (from equations (7.33) and (7.34)) for some of the model aggregates at 664 GHz, 190 K. Red underlined values indicate that the relative difference is above 1% which is the desired accuracy.

7.3.4 Comparing with other databases

In the database by Guosheng Liu [73], the discrete dipole approximation (DDA) method of [47] was used to calculate the SSPs and phase matrix elements of ice columns and plates, rosettes and snowflakes of different maximum dimension and for different frequencies of incoming radiation. Two different causes were considered for the accuracy of the computed results: the interdipole spacing (similar to the mesh size of our method) and the number of orientations considered in simulating random orientation. An interdipole spacing scheme based on the wavelength (similar to a fixed number of elements per wavelength in our case) was considered. Comparison between the scheme used for the database and a scheme with half the spacing showed less than 2% relative errors for $\langle C_{sca} \rangle$ and $\langle g \rangle$ for all frequencies and particle sizes. The relative difference for $\langle C_{bsca} \rangle$ was higher but less than 5%. A total of $16 \times 17 \times 16$ ($= 4352$) orientations were considered for the simulations of the database. To test whether this number of orientations was 'random' enough tests were also performed with $32 \times 33 \times 32$ ($= 33792$) orientations. The relative difference for $\langle C_{sca} \rangle$ and $\langle g \rangle$ was less than 1%, however the relative difference for $\langle C_{bsca} \rangle$ increased with frequency and particle size. While for lower frequencies the relative difference was less than 5% this increased to more than 20% for higher frequencies and larger particles.

These observations are consistent with ours, i.e. the number of incident waves required to accurately compute $\langle C_{bsca} \rangle$ increases with X, while the remaining scattering properties are accurately computed with a smaller number of them, about 50 incident waves in our case. We note that we managed to significantly reduce the number of orientations considered for smaller size parameters by taking advantage of the Lebedev scheme and the fact that no polarisation rotations (hence no second integral) are required for the computation of SSPs. Even with the inclusion of the polarisation rotations (needed for the accurate computation of the phase matrix) we required only 14×10 $(= 140)$, 50×15 $(= 750)$, 110×15 $(= 1650)$, 194×15 $(= 2910)$ or 230×15 $(= 3450)$ orientations for size parameters up to 10, compared with 4352 orientations considered in [73]. The number of orientations is greater for X larger than 10 with 302×15 $(= 4530)$ and 302×20 $(= 6040)$ orientations, compared to [73], but that was only the case for 60 out of the 260 size parameters considered.

7.4 Comments on the implementation

For the completion of the database, the accelerating methods of Chapters 4 and 5 were used. Given the extremely large number of GMRES solves that had to be performed, our focus was on minimising solution time. Memory consumption and assembly time were not a concern except for the larger D_{max} at 664 GHz.

For the simulations at 50 GHz, a bi-parametric implementation has shown an increased solution time (we refer back to Figure 6.8), and hence we only used a reduced preconditioner with $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{D}}^i$ and a non bi-parametric implementation. For the simulations at 183 GHz, testing beyond the results of Figure 6.8 showed an increased solution time when $\chi_{\mathbf{P}} = 0$, so we used the reduced bi-parametric preconditioner $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{D}}^i$ and $(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (0.1, \infty, (1,1,1,1))$. For the simulations at 243 GHz, a bi-parametric implementation with $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{D}}^i$ and $(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q}_{\mathbf{P}}) = (0.1, \infty, (1,1,1,1))$ was used as the cutoff parameter $\chi_{\mathbf{P}}$ was not implemented at that time, so the default $\chi_{\mathbf{P}} = \infty$ was assumed. For aggregates

of $D_{max} < 4000\mu m$ at 664 GHz, the same implementation was used as for 243 GHz (again the cutoff parameter was not implemented at that stage). For the larger aggregates at 664 GHz, we expect memory consumption to become increasingly high, in which case we aim to use a bi-parametric implementation of $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{S}}^i$ with $(\nu_{\mathbf{P}}, \chi_{\mathbf{P}}, \mathbf{q_P}) = (0.1, 0, (1,1,1,1))$.

Amazon Web Services (AWS) EC2 instances were used to perform our simulations with up to 96 CPUs available at a time and 384 GB of RAM. Smaller size parameter problems (such as the aggregates at 50 and 183 GHz) were ran on a desktop machine with 40 CPUs and 188 GB RAM. Given the large availability of CPUs, the Multiprocessing library in Python was used to distribute the different incident waves to different CPUs and therefore run the multiple GMRES solves in parallel aiming to reduce the total computation time for each aggregate.

7.5 Some results from our simulations

We finish this chapter by comparing the SSPs between our rosette aggregate model and other ice crystal models, such as those considered in the Ding *et al.* database [10]. In particular, we consider the 10 and 5 plate aggregates, the 8-column aggregate and the hollow bullet rosette, images of which are shown in Figure 7.12. We show comparisons for frequencies 50 and 243 GHz and temperature 230 K. These are shown in Figure 7.13 and 7.14 respectively.

The comparisons show that for aggregates $D_{max} \leq 492$ (where all aggregate shapes are the same but re-scaled to different dimensions), the extinction and scattering cross-sections behave very similar to the 8-column aggregate, but for $D_{max} > 492$ these come closer to the behaviour of the hollow bullet rosette for both frequencies. The scattering albedo, $\langle \varpi_0 \rangle$, is smaller than that of the 8-column aggregate at 50 GHz and $D_{max} \leq 492$, but larger for 243 GHz, although we note that the Ding *et al.* database [10] is insufficiently discretised for D_{max} between $100\mu m$ and $500\mu m$. For $D_{max} > 492$ the albedo is between the 8-column aggregate and hollow bullet rosette but as D_{max} increases further

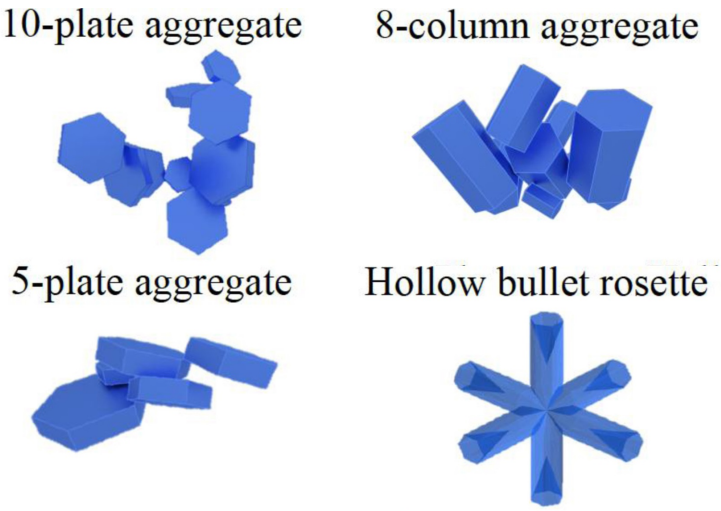


Figure 7.12: Ice crystal models from the Ding *et al.* database [10]. Image from [10].

the behaviour becomes closer to that of the hollow bullet rosette.

We note that our aggregate model consists of different ice crystal configurations, explaining why the some of the SSPs vary from aggregate to aggregate, compared with the SSPs from [10]. This change in shape was necessary to constrain the Monte Carlo simulations to follow the Cotton *et al.* [72] mass - dimension relationship to be consistent with the Unified Model [110] followed at the Met Office. This variation in shape, mass and area is more obvious in the scattering albedo $\langle \varpi_0 \rangle$ and the asymmetry parameter $\langle g \rangle$.

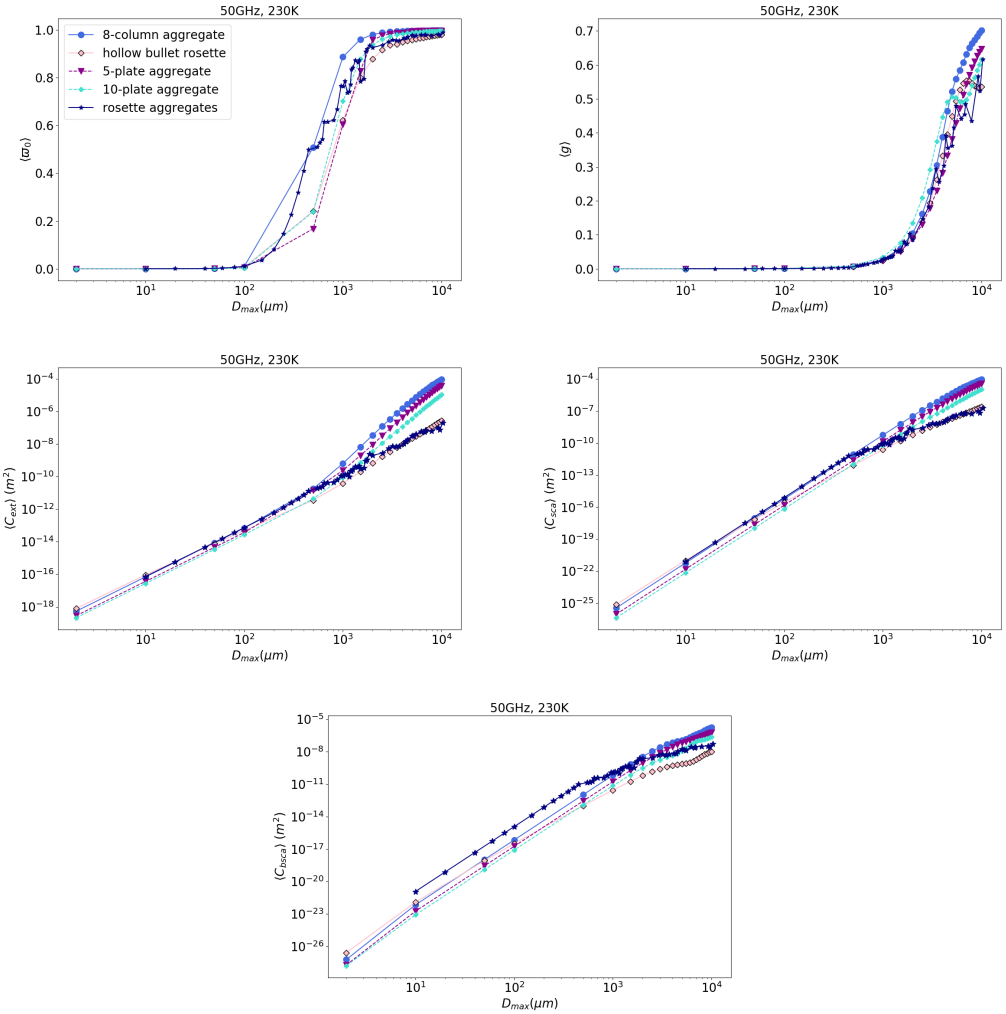


Figure 7.13: Comparison of the SSPs for the aggregate rosette model compared to SSPs of some of the ice crystals considered in the Ding *et al.* database [10], for 50 GHz and 230 K.

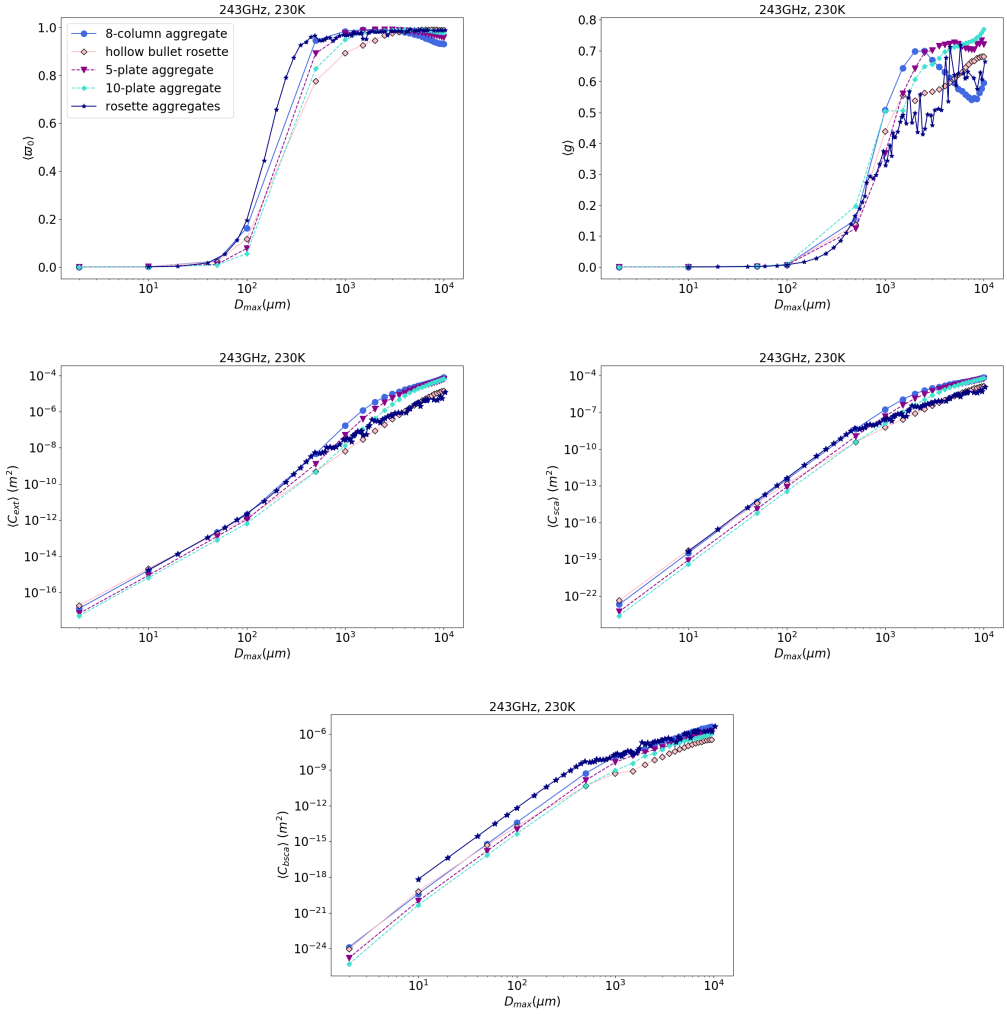


Figure 7.14: Same as in Figure 7.13 but for 243 GHz.

Chapter 8

Concluding Remarks and Future Research Avenues

In this thesis, we have presented two complementary approaches in accelerating Calderón preconditioning for electromagnetic scattering by multiple absorbing dielectric objects. We recall from Chapter 1, that we have been looking for some operator $\boldsymbol{\mathcal{P}}$ applied to the PMCHWT formulation so that $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$ has better properties than $\boldsymbol{\mathcal{A}}$ and its discrete version is easier to solve numerically. The overall computational cost of solving the preconditioned discrete system is affected by both

(i) the choice of preconditioning operator $\boldsymbol{\mathcal{P}}$, and

(ii) the choice of discretisation for the operator product $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$.

With regards to (i), the choice of $\boldsymbol{\mathcal{P}}$ should lead to a smaller number of iterations compared to $\boldsymbol{\mathcal{A}}$. At the same time, the operator product $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$ incurs additional matrix-vector operations, so the ideal preconditioner should consist of as few operators as possible. Regarding (ii), the discretisation of $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$ should be as cheap (memory and time-wise) as possible, while still producing a sufficiently accurate numerical solution. To achieve a stable discretisation of the operator product $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}$, one has to use both a primal and a dual mesh, the latter defined on a barycentric refinement of the primal mesh leading to a 6-fold increase in the number of elements required. To capture the oscilla-

tory solution of the electromagnetic waves, the mesh needs to be refined with respect to frequency. A dense assembly and storage of the matrix would therefore scale as $\mathcal{O}(N^2)$, where N are the degrees of freedom, or $\mathcal{O}(k^4)$, making the simulation of high-frequency problems very expensive.

Two complementary approaches have been considered in this thesis in order to minimise memory cost and computation time: a modification of the preconditioning operator, and a bi-parametric implementation. The former aimed to minimise the number of operators used in the preconditioner. This was in order to reduce the additional matrix-vector products performed, together with memory costs, while still maintaining a sufficient preconditioning effect. The latter involved using two distinct sets of parameters for the assembly of the operator product; one for the operator and one for the preconditioner. The operator was assembled with a more expensive set of parameters to obtain an accurate solution. The preconditioner, which was discretised using the expensive dual basis functions, was assembled with a cheaper set of parameters to minimise assembly and solution time as well as memory cost. The two approaches have been used in a series of model problems and later combined together for realistic complex ice crystal configurations.

The different reduced Calderón preconditioners have been defined and numerically tested for model problems in Chapter 4. We have demonstrated that depending on the problem of interest one can achieve a reduction of memory cost by 50-75% and a 60-80% reduction in total computational time. Some of the reduced preconditioners presented have shown some erratic behaviour, by showing a sudden increase of GMRES iterations (and hence increased matvecs and GMRES time) for specific exterior wavenumbers. Further work to understand the theoretical basis of this behaviour would be desirable as it could provide ideas on different preconditioning techniques that avoid this sudden increase. A theoretical analysis of the reduced preconditioners presented in Chapter 4 could also provide an insight into whether the observed performance is consistent with other libraries and not just for our specific implementation.

This would also indicate that our methods perform better in general, and not just for our particular implementation and software library.

Different bi-parametric implementations have been discussed and tested in Chapter 5. This consisted of using two different sets of parameters; an expensive one for the accurate assembly of the operators and a cheaper one for the preconditioner. This included a larger value for the ACA parameters, and smaller values for the quadrature orders and far-field cutoff parameter. Our numerical experiments demonstrated that a 60-90% reduction in memory consumption and a 50-60% reduction in total computational time can be achieved with such a bi-parametric implementation. A bi-parametric implementation of the reduced preconditioners of Chapter 4 reduced costs even further by 95% and 80% respectively. Again, some of the reduced preconditioners in combination with a bi-parametric implementation have shown an erratic behaviour (increased GMRES iterations, matvecs and GMRES time for specific exterior wavenumbers). We note that a careful analytical investigation in this case is difficult due to the methods combining properties of the operators at the continuous level together with properties of the numerical implementation at the discrete level. Although we have presented our ideas using results from the $\mathcal{H}$-matrix implementation within Bempp, we note that the fundamental ideas of ignoring far-field interactions can be extended to FMM type preconditioning. In fact, any compression method that separates interactions into 'near' and 'far' field can use the ideas presented in this thesis.

In Chapter 6, we have combined the accelerating techniques of Chapters 4 and 5, and compared their performance for realistic complex ice crystal configurations. Our numerical experiments have demonstrated a 99% reduction in memory cost and at least an 80% reduction in total computational time, for the highest frequency (664 GHz) considered. These findings are significant as the otherwise prohibitive cost imposed by the use of the dual basis functions is now alleviated, allowing us to consider large scale simulations that were otherwise too expensive to do.

In Chapter 7, we have discussed the scattering database that is being generated using our accelerating methods at the Met Office to improve numerical weather prediction. The microphysical model had been discussed along with the relevant definitions required to compute the scattering properties of the assumed model. Our implementation of random orientation has also been tested and compared with other methods and scattering databases. We found that one can decrease the number of incident waves and polarisation vectors needed to simulate random orientation by using a combination of Gaussian and Lebedev quadrature to reduce the three integrals to two (and reduce the number of corresponding loops from three to two). We also found that depending on the size parameter, one can reduce the number of incident waves and polarisation vectors needed to simulate random orientation. Early results from the scattering database have also been presented.

The ideas presented in this thesis lay the foundations for a number of future applications. A natural extension of the work presented here is the extension of the PMCHWT formulation (which we presented in Chapter 2 for the multi-particle case $M > 1$), to a more generalised set up used to describe scattering by multiple objects with inclusions. This could be used to simulate scattering by ice crystals with air bubbles (such as in [122, 123, 124]) or other trapped particles such as soot impurities [122] or mineral aerosol [123], or scattering by ice crystals falling through the atmosphere which are covered by a liquid surface. Some initial work on the generalised PMCHWT formulation in this case has been included in Appendix A and can be used for future work. From a scientific computing point of view, it would be interesting to investigate the performance of the bi-parametric reduced preconditioners showcased in this thesis, to determine whether the observations carry over to more complex examples. From an application's perspective, it would be interesting to investigate how many incident waves and polarisation vectors are needed to simulate random orientation of ice crystals with inclusions, and whether the shape and position of the inclusions affect the scattering properties and phase

matrix.

Finally, our implementation of random orientation presented in Chapter 7, could be adjusted to simulate azimuthally random orientation. This is the case where particles have a fixed orientation in the polar direction (fixed θ^{inc}) but are randomly oriented in the azimuth direction. While the Lebedev scheme can no longer be used, as only one of the two angles needs to be varied, the implementation still requires only two integrals, this time implemented by two Gaussian quadrature products. Early experiments have demonstrated that our implementation could be easily adjusted to simulate this but further tests will need to be performed to identify the number of waves and polarisation vectors needed for large size parameters. A natural extension would then be the combination of the two ideas; for example simulating hexagonal plates falling through the atmosphere which are covered by a liquid surface and are usually azimuthally randomly oriented.

Appendix A

Multi-particle scattering with inclusions

This Appendix describes a generalisation of the scattering problem and PM-CHWT formulation that was presented in Chapter 2, for scattering by multiple particles that might have inclusions, such as air bubbles or other trapped particles, or aggregates that are surrounded by liquid.

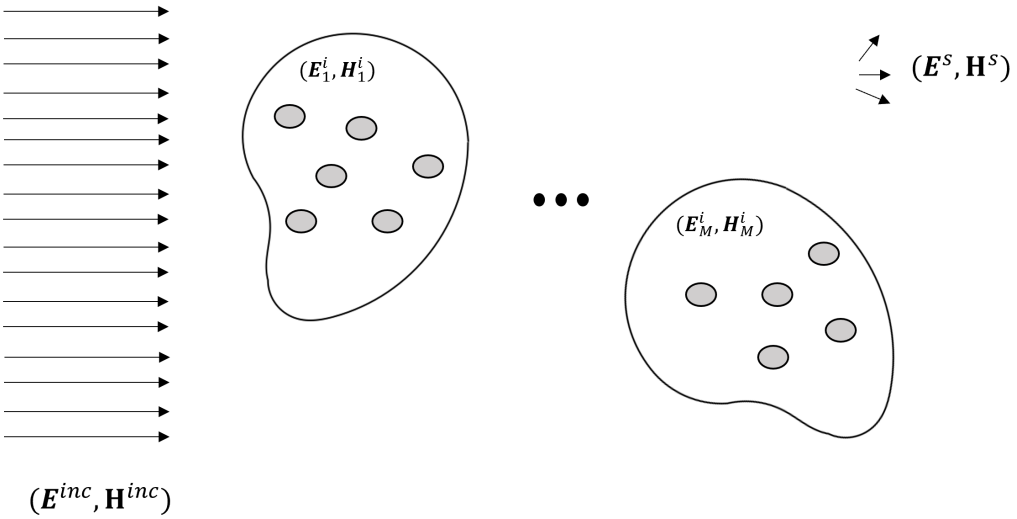


Figure A.1: Electromagnetic scattering by multiple scatterers of arbitrary shape each with an arbitrary number of inclusions.

A.1 The scattering problem

We consider M disjoint arbitrary 3D isotropic homogeneous dielectric scatterers occupying bounded domains $\Omega_m^i \subset \mathbb{R}^3$, $m = 1, \ldots, M$, with boundaries $\Gamma_m = \partial\Omega_m^i$, in a homogeneous exterior medium $\Omega^e = \mathbb{R}^3 \backslash \overline{\cup_{m=1}^M \Omega_m^i}$, as in Figure 1.1. Each scatterer Ω_m^i might contain N_m, $m = 1, \ldots, M$, smaller 3D isotropic homogeneous dielectric scatterers, occupying domains $\Omega_{m_n}^i \subset \Omega_m^i, n = 1, \ldots, N_m$, with boundaries $\Gamma_{m_n} = \partial\Omega_{m_n}^i$.

The electric and magnetic fields in the exterior domain Ω^e and interior domains $\Omega_m^i \backslash \overline{\cup_{n=1}^{N_m} \Omega_{m_n}^i}$, $m = 1, \ldots, M$, and $\Omega_{m_n}^i$, $n = 1, \ldots, N_m$, $m = 1, \ldots, M$, will be denoted $(\mathbf{E}^e, \mathbf{H}^e)$, $(\mathbf{E}_m^i, \mathbf{H}_m^i)$ and $(\mathbf{E}_{m_n}^i, \mathbf{H}_{m_n}^i)$ respectively. They are assumed to satisfy the time-harmonic Maxwell equations

$$\nabla \times \mathbf{E}^e = i\omega\mu_e \mathbf{H}^e, \qquad \text{in } \Omega^e, \tag{A.1}$$

$$\nabla \times \mathbf{H}^e = -i\omega\epsilon_e \mathbf{E}^e, \qquad \text{in } \Omega^e, \tag{A.2}$$

$$\nabla \times \mathbf{E}_m^i = i\omega\mu_m \mathbf{H}_m^i, \qquad \text{in } \Omega_m^i \backslash \overline{\cup_{n=1}^{N_m} \Omega_{m_n}^i},\ m = 1, \ldots, M, \tag{A.3}$$

$$\nabla \times \mathbf{H}_m^i = -i\omega\epsilon_m \mathbf{E}_m^i, \qquad \text{in } \Omega_m^i \backslash \overline{\cup_{n=1}^{N_m} \Omega_{m_n}^i},\ m = 1, \ldots, M, \tag{A.4}$$

and

$$\nabla \times \mathbf{E}_{m_n}^i = i\omega\mu_{m_n} \mathbf{H}_{m_n}^i, \qquad \text{in } \Omega_{m_n}^i,\ n = 1, \ldots, N_m,\ m = 1, \ldots, M, \tag{A.5}$$

$$\nabla \times \mathbf{H}_{m_n}^i = -i\omega\epsilon_{m_n} \mathbf{E}_{m_n}^i, \qquad \text{in } \Omega_{m_n}^i,\ n = 1, \ldots, N_m,\ m = 1, \ldots, M, \tag{A.6}$$

together with the transmission boundary conditions

$$\mathbf{E}_m^i(\mathbf{x}) \times \mathbf{n}_m = \mathbf{E}^e(\mathbf{x}) \times \mathbf{n}_m, \quad \mathbf{x} \in \Gamma_m,\ m = 1, \ldots, M, \tag{A.7}$$

$$\mathbf{H}_m^i(\mathbf{x}) \times \mathbf{n}_m = \mathbf{H}^e(\mathbf{x}) \times \mathbf{n}_m, \quad \mathbf{x} \in \Gamma_m,\ m = 1, \ldots, M, \tag{A.8}$$

and

$$\mathbf{E}^i_{m_n}(\mathbf{x}) \times \mathbf{n}_{m_n} = \mathbf{E}^i_m(\mathbf{x}) \times \mathbf{n}_{m_n}, \quad \mathbf{x} \in \Gamma_{m_n},\ n = 1, \ldots, N_m, m = 1, \ldots, M, \tag{A.9}$$

$$\mathbf{H}^i_{m_n}(\mathbf{x}) \times \mathbf{n}_{m_n} = \mathbf{H}^i_m(\mathbf{x}) \times \mathbf{n}_{m_n}, \quad \mathbf{x} \in \Gamma_{m_n},\ n = 1, \ldots, N_m, m = 1, \ldots, M. \tag{A.10}$$

Here we assume a time-dependence of the form $e^{-i\omega t}$, with angular frequency $\omega > 0$. The parameters ϵ_e, ϵ_m, ϵ_{m_n} and μ_e, μ_m, μ_{m_n} represent respectively the electric permittivity and the magnetic permeability of the domains, and $\mathbf{n}_m$ and $\mathbf{n}_{m_n}$ are the unit normal vectors on Γ_m and Γ_{m_n} pointing into Ω^e and Ω^i_m respectively.

In the scattering problem, an incident field $(\mathbf{E}^{inc}, \mathbf{H}^{inc})$ (for instance, a plane wave) gives rise to internal fields $(\mathbf{E}^i_m, \mathbf{H}^i_m)$ in Ω^i_m and $(\mathbf{E}^i_{m_n}, \mathbf{H}^i_{m_n})$ in $\Omega^i_{m_n}$, and a scattered field $(\mathbf{E}^s, \mathbf{H}^s)$ in the exterior domain Ω^e. The latter is assumed to satisfy the Silver-Müller radiation condition

$$\sqrt{\mu_e}\mathbf{H}^s \times \frac{\mathbf{x}}{|\mathbf{x}|} - \sqrt{\epsilon_e}\mathbf{E}^s = \mathcal{O}\left(\frac{1}{|\mathbf{x}|}\right), \quad \text{as } |\mathbf{x}| \to \infty, \tag{A.11}$$

and the total exterior field is then the sum of incident and scattered fields

$$\mathbf{E}^e = \mathbf{E}^{inc} + \mathbf{E}^s, \qquad \text{in } \Omega^e, \tag{A.12}$$

$$\mathbf{H}^e = \mathbf{H}^{inc} + \mathbf{H}^s, \qquad \text{in } \Omega^e. \tag{A.13}$$

It is sufficient to solve for either the electric or magnetic fields and then recover the remaining fields by (A.1)-(A.2), (A.3)-(A.4) and (A.5)-(A.6). In what follows, we will solve for the electric fields $\mathbf{E}^e$, $\mathbf{E}^i_m$, $\mathbf{E}^i_{m_n}$ which satisfy

$$\nabla \times (\nabla \times \mathbf{E}^e) - k_e^2 \mathbf{E}^e = 0, \qquad \text{in } \Omega^e, \tag{A.14}$$

$$\nabla \times (\nabla \times \mathbf{E}^i_m) - k_m^2 \mathbf{E}^i_m = 0, \qquad \text{in } \Omega^i_m,\ m = 1, \ldots, M, \tag{A.15}$$

$$\nabla \times (\nabla \times \mathbf{E}^i_{m_n}) - k^2_{m_n} \mathbf{E}^i_{m_n} = 0, \qquad \text{in } \Omega^i_{m_n},\ n = 1, \ldots, N_m, m = 1, \ldots, M, \tag{A.16}$$

where $k_e = \omega\sqrt{\mu_e \epsilon_e}$, $k_m = \omega\sqrt{\mu_m \epsilon_m}$ and $k_{m_n} = \omega\sqrt{\mu_{m_n} \epsilon_{m_n}}$ are the wavenumbers in the respective domains.

A.2 Boundary Integral Formulations

Using the definitions of the Dirichlet and Neumann traces, and equations (A.1), (A.3), (A.5), we can express the transmission boundary conditions (A.7)-(A.10) in terms of traces of the respective electric fields

$$\mathbf{u}^i_m = \mathbf{u}^s_m + \mathbf{u}^{inc}_m, \quad \mathbf{x} \in \Gamma_m,\ m = 1, \ldots, M, \tag{A.17}$$

$$\mathbf{u}^i_{m_n} = \mathbf{u}^e_{m_n}, \qquad \mathbf{x} \in \Gamma_{m_n},\ n = 1, \ldots, N_m, m = 1, \ldots, M, \tag{A.18}$$

where [1]

$$\mathbf{u}^i_m = \begin{bmatrix} \gamma^-_{D,m} \mathbf{E}^i_m \\ \dfrac{k_m}{\mu_m} \gamma^-_{N,m} \mathbf{E}^i_m \end{bmatrix}, \quad \mathbf{u}^s_m = \begin{bmatrix} \gamma^+_{D,m} \mathbf{E}^s \\ \dfrac{k_e}{\mu_e} \gamma^+_{N,m} \mathbf{E}^s \end{bmatrix}, \quad \mathbf{u}^{inc}_m = \begin{bmatrix} \gamma^+_{D,m} \mathbf{E}^{inc} \\ \dfrac{k_e}{\mu_e} \gamma^+_{N,m} \mathbf{E}^{inc} \end{bmatrix}, \tag{A.19}$$

and

$$\mathbf{u}^i_{m_n} = \begin{bmatrix} \gamma^-_{D,m_n} \mathbf{E}^i_{m_n} \\ \dfrac{k_{m_n}}{\mu_{m_n}} \gamma^-_{N,m_n} \mathbf{E}^i_{m_n} \end{bmatrix}, \quad \mathbf{u}^e_{m_n} = \begin{bmatrix} \gamma^+_{D,m_n} \mathbf{E}^i_m \\ \dfrac{k_m}{\mu_m} \gamma^+_{N,m_n} \mathbf{E}^i_m \end{bmatrix}. \tag{A.20}$$

The Stratton-Chu representation formulae can be written as follows: The exterior and interior fields $\mathbf{E}^s$, $\mathbf{E}^i_m$, $m = 1, \ldots, M$ and $\mathbf{E}^i_{m_n}$, $n = 1, \ldots, N_m$,

[1] the subscript defines the scatterer on which we take the relevant trace, and the superscript whether it's from the interior i, or exterior e. The trace $\mathbf{u}^e_{m_n}$ denotes the trace on $\Omega^i_{m_n}$ from its exterior Ω^i_m and not the global exterior Ω^e.

$m = 1, \ldots, M$ can be represented as [2]

$$-\sum_{m}^{M} \mathcal{H}_m^e \left(\gamma_{D,m}^{+} \mathbf{E}^s\right) - \sum_{m}^{M} \mathcal{E}_m^e \left(\gamma_{N,m}^{+} \mathbf{E}^s\right) = \begin{cases} \mathbf{E}^s(\mathbf{x}), & \mathbf{x} \in \Omega^e, \\ 0, & \mathbf{x} \notin \overline{\Omega^e}. \end{cases} \tag{A.21}$$

$$\begin{aligned} &\mathcal{H}_m^i \left(\gamma_{D,m}^{-} \mathbf{E}_m^i\right) + \mathcal{E}_m^i \left(\gamma_{N,m}^{-} \mathbf{E}_m^i\right) - \\ &\sum_{n}^{N_m} \mathcal{H}_{m_n}^e \left(\gamma_{D,m_n}^{+} \mathbf{E}_m^i\right) - \sum_{n}^{N_m} \mathcal{E}_{m_n}^e \left(\gamma_{N,m_n}^{+} \mathbf{E}_m^i\right) = \begin{cases} \mathbf{E}_m^i(\mathbf{x}), & \mathbf{x} \in \Omega_m^i, \\ 0, & \mathbf{x} \notin \overline{\Omega_m^i}, \end{cases} \end{aligned} \tag{A.22}$$

and

$$\mathcal{H}_{m_n}^i \left(\gamma_{D,m_n}^{-} \mathbf{E}_{m_n}^i\right) + \mathcal{E}_{m_n}^i \left(\gamma_{N,m_n}^{-} \mathbf{E}_{m_n}^i\right) = \begin{cases} \mathbf{E}_{m_n}^i(\mathbf{x}), & \mathbf{x} \in \Omega_{m_n}^i, \\ 0, & \mathbf{x} \notin \overline{\Omega_{m_n}^i}. \end{cases} \tag{A.23}$$

Taking appropriate traces of (A.21) and (A.22) on the boundary Γ_m gives

$$\left(\frac{1}{2}\boldsymbol{\mathcal{I}}_m - \boldsymbol{\mathcal{A}}_m^e\right) \mathbf{u}_m^s - \sum_{\ell \neq m}^{M} \boldsymbol{\mathcal{A}}_{m\ell} \mathbf{u}_\ell^s = \mathbf{u}_m^s, \quad m = 1, \ldots, M, \tag{A.24}$$

$$\left(\frac{1}{2}\boldsymbol{\mathcal{I}}_m + \boldsymbol{\mathcal{A}}_m^i\right) \mathbf{u}_m^i - \sum_{n}^{N_m} \boldsymbol{\mathcal{A}}_{mm_n} \mathbf{u}_{m_n}^{i,m} = \mathbf{u}_m^i, \ m = 1, \ldots, M, \tag{A.25}$$

and appropriate traces of (A.22) and (A.23) on the boundary Γ_{m_n}

$$\left(\frac{1}{2}\boldsymbol{\mathcal{I}}_{m_n} - \boldsymbol{\mathcal{A}}_{m_n}^e\right) \mathbf{u}_{m_n}^e - \sum_{s \neq n}^{N_m} \boldsymbol{\mathcal{A}}_{m_n m_s} \mathbf{u}_{m_s}^e - \boldsymbol{\mathcal{A}}_{m_n m} \mathbf{u}_m^i = \mathbf{u}_{m_n}^e, \tag{A.26}$$

$$\left(\frac{1}{2}\boldsymbol{\mathcal{I}}_{m_n} + \boldsymbol{\mathcal{A}}_{m_n}^i\right) \mathbf{u}_{m_n}^i = \mathbf{u}_{m_n}^i, \tag{A.27}$$

for $n = 1, \ldots, N_m, \ m = 1, \ldots, M$.

Subtracting (A.24) from (A.25), (A.26) from (A.27) and using the boundary conditions (A.17) and (A.18) gives the PMCHWT formulation in the con-

[2] Equation (A.22), is a combination of the interior and exterior Stratton-Chu formulae, as Ω^i is the interior of the large scatterer but acts as the exterior of the smaller scatterers $\Omega_{m_n}^i$, $n = 1, \cdots, N_m$.

text of multi-particle scattering where scatterers might have an arbitrary number of inclusions

$$\left(\boldsymbol{\mathcal{A}}_m^i + \boldsymbol{\mathcal{A}}_m^e\right)\mathbf{u}_m^s + \sum_{\ell\neq m}^{M} \boldsymbol{\mathcal{A}}_{m\ell}\mathbf{u}_\ell^s - \sum_{n}^{N_m} \boldsymbol{\mathcal{A}}_{mm_n}\mathbf{u}_{m_n}^i = \left(\frac{1}{2}\boldsymbol{\mathcal{I}}_m - \boldsymbol{\mathcal{A}}_m^i\right)\mathbf{u}_m^{inc}, \tag{A.28}$$

$$\left(\boldsymbol{\mathcal{A}}_{m_n}^i + \boldsymbol{\mathcal{A}}_{m_n}^e\right)\mathbf{u}_{m_n}^i + \sum_{s\neq n}^{N_m} \boldsymbol{\mathcal{A}}_{m_n m_s}\mathbf{u}_{m_s}^i - \boldsymbol{\mathcal{A}}_{m_n m}\mathbf{u}_m^s = \boldsymbol{\mathcal{A}}_{m_n m}\mathbf{u}_m^{inc}, \tag{A.29}$$

for each $n = 1,\ldots,N_m,\ m = 1,\ldots,M$. These $M + \sum_m N_m$ systems can be combined to form a block structured system

$$\tilde{\boldsymbol{\mathcal{A}}}\mathbf{u} = \mathbf{b}, \tag{A.30}$$

with the block-diagonal entries given by

$$\tilde{\boldsymbol{\mathcal{A}}}_{m,m} = \begin{bmatrix} \boldsymbol{\mathcal{A}}_m^e + \boldsymbol{\mathcal{A}}_m^i & -\boldsymbol{\mathcal{A}}_{m,m_1} & \cdots & \cdots & -\boldsymbol{\mathcal{A}}_{m,m_{N_m}} \\ -\boldsymbol{\mathcal{A}}_{m_1,m} & \boldsymbol{\mathcal{A}}_{m_1}^e + \boldsymbol{\mathcal{A}}_{m_1}^i & \boldsymbol{\mathcal{A}}_{m_1,m_2} & \cdots & \boldsymbol{\mathcal{A}}_{m_1,m_{N_m}} \\ \vdots & \boldsymbol{\mathcal{A}}_{m_2,m_1} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \boldsymbol{\mathcal{A}}_{m_{N_m-1},m_{N_m}} \\ -\boldsymbol{\mathcal{A}}_{m_{N_m},1} & \boldsymbol{\mathcal{A}}_{m_{N_m},m_1} & \cdots & \boldsymbol{\mathcal{A}}_{m_{N_m},m_{N_m-1}} & \boldsymbol{\mathcal{A}}_{m_{m_1}}^e + \boldsymbol{\mathcal{A}}_{m_{N_m}}^i \end{bmatrix} \tag{A.31}$$

and the off-diagonals by

$$\tilde{\boldsymbol{\mathcal{A}}}_{m,\ell} = \begin{bmatrix} \boldsymbol{\mathcal{A}}_{m,\ell} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 \end{bmatrix}. \tag{A.32}$$

We recall the definitions of

$$\boldsymbol{\mathcal{A}}_m^i = \begin{bmatrix} \mathcal{C}_m^i & \frac{\mu_m}{k_m}\mathcal{S}_m^i \\ -\frac{k_m}{\mu_m}\mathcal{S}_m^i & \mathcal{C}_m^i \end{bmatrix}, \quad \boldsymbol{\mathcal{A}}_m^e = \begin{bmatrix} \mathcal{C}_m^e & \frac{\mu_e}{k_e}\mathcal{S}_m^e \\ -\frac{k_e}{\mu_e}\mathcal{S}_m^e & \mathcal{C}_m^e \end{bmatrix},$$
$$\boldsymbol{\mathcal{A}}_{m\ell} = \begin{bmatrix} \mathcal{C}_{m\ell}^e & \frac{\mu_e}{k_e}\mathcal{S}_{m\ell}^e \\ -\frac{k_e}{\mu_e}\mathcal{S}_{m\ell}^e & \mathcal{C}_{m\ell}^e \end{bmatrix}. \tag{A.33}$$

These can be adapted to $\boldsymbol{\mathcal{A}}_{m_n}^i$, $\boldsymbol{\mathcal{A}}_{m_n}^e$ and $\boldsymbol{\mathcal{A}}_{m_n,m_s}$, remembering that the exterior of $\Omega_{m_n}^i$ is Ω_m^i and therefore any exterior parameters in the above should be replaced by those for Ω_m^i. The vector of unknowns $\mathbf{u}$ is a combination of scattered fields $\mathbf{u}_m^s$ created by Ω_m^i and interior fields $\mathbf{u}_{m_n}^i$ for $\Omega_{m_n}^i$

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_M \end{bmatrix}, \tag{A.34}$$

where

$$\mathbf{u}_m = \begin{bmatrix} \mathbf{u}_m^s \\ \mathbf{u}_{m_1}^i \\ \vdots \\ \mathbf{u}_{m_{N_m}}^i \end{bmatrix}. \tag{A.35}$$

Bibliography

[1] Wojciech Śmigaj, Timo Betcke, Simon Arridge, Joel Phillips, and Martin Schweiger. Solving boundary integral problems with BEM++. *ACM Transactions on Mathematical Software (TOMS)*, 41(2):6, 2015.

[2] Antigoni Kleanthous, Timo Betcke, David P Hewett, Matthew W Scroggs, and Anthony J Baran. Calderón preconditioning of PMCHWT boundary integral equations for scattering by multiple absorbing dielectric particles. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 224:383–395, 2019.

[3] Antigoni Kleanthous, Timo Betcke, David P Hewett, Paul Escapil-Inchauspé, Carlos Jerez-Hanckes, and Anthony J Baran. Accelerated Calderón preconditioning for Maxwell transmission problems. *arXiv preprint arXiv:2008.04772*, 2020.

[4] Anthony J Baran. A review of the light scattering properties of cirrus. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 110(14):1239–1260, 2009.

[5] Samuel P Groth, Anthony J Baran, Timo Betcke, Stephan Havemann, and Wojciech Śmigaj. The boundary element method for light scattering by ice crystals and its implementation in BEM++. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 167:40–52, 2015.

[6] Ping Yang and KN Liou. Single-scattering properties of complex ice crystals in terrestrial atmosphere. *Beitrage zur Physik der Atmosphare-Contributions to Atmospheric Physics*, 71(2):223–248, 1998.

[7] RP Lawson, S Woods, E Jensen, E Erfani, C Gurganus, M Gallagher, P Connolly, J Whiteway, AJ Baran, P May, et al. A review of ice particle shapes in cirrus formed in situ and in anvils. *Journal of Geophysical Research: Atmospheres*, 124(17-18):10049–10090, 2019.

[8] Michael I Mishchenko, Larry D Travis, and Andrew A Lacis. *Scattering, absorption, and emission of light by small particles*. Cambridge university press, 2002.

[9] Manfred Brath, Robin Ekelund, Patrick Eriksson, Oliver Lemke, and Stefan A Buehler. Microwave and submillimeter wave scattering of oriented ice particles. *Atmospheric Measurement Techniques*, 13(5):2309–2333, 2020.

[10] Jiachen Ding, Lei Bi, Ping Yang, George W Kattawar, Fuzhong Weng, Quanhua Liu, and Thomas Greenwald. Single-scattering properties of ice particles in the microwave regime: Temperature effect on the ice refractive index with implications in remote sensing. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 190:26–37, 2017.

[11] Stephen G Warren and Richard E Brandt. Optical constants of ice from the ultraviolet to the microwave: A revised compilation. *Journal of Geophysical Research: Atmospheres*, 113(D14), 2008.

[12] Christian Mätzler. Microwave dielectric properties of ice. *Thermal microwave radiation: applications for remote sensing*, 52:455–462, 2006.

[13] Wolfgang Hackbusch. *Hierarchical matrices: algorithms and analysis*, volume 49. Springer, 2015.

[14] Mario Bebendorf. *Hierarchical matrices*. Springer, 2008.

[15] Steffen Börm. *Efficient numerical methods for non-local operators: $\mathcal{H}^2$-matrix compression, algorithms and analysis*, volume 14. European Mathematical Society, 2010.

[16] Saad Omar and Dan Jiao. $\mathcal{O}(N)$ iterative and $\mathcal{O}(N \log N)$ direct volume integral equation solvers for large-scale electrodynamic analysis. In *2014 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, pages 593–596. IEEE, 2014.

[17] Ben Dembart and Elizabeth Yip. The accuracy of fast multipole methods for Maxwell's equations. *IEEE Computational Science and Engineering*, 5(3):48–56, 1998.

[18] Eric Darve. The fast multipole method: numerical implementation. *Journal of Computational Physics*, 160(1):195–240, 2000.

[19] Nail A Gumerov and Ramani Duraiswami. Fast multipole method for the biharmonic equation in three dimensions. *Journal of Computational Physics*, 215(1):363–383, 2006.

[20] Olaf Steinbach. *Numerical approximation methods for elliptic boundary value problems: finite and boundary elements*. Springer Science & Business Media, 2007.

[21] Michael I Mishchenko, Larry D Travis, and Daniel W Mackowski. T-matrix computations of light scattering by nonspherical particles: A review. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 55(5):535–575, 1996.

[22] S. Havemann and A.J. Baran. Extension of T-matrix to scattering of electromagnetic plane waves by non-axisymmetric dielectric particles: application to hexagonal ice cylinders. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 70(2):139–158, 2001.

[23] Michael Kahnert. The T-matrix code Tsym for homogeneous dielectric particles with finite symmetries. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 123:62–78, 2013.

[24] Lei Bi, Ping Yang, George W Kattawar, and Michael I Mishchenko. Efficient implementation of the invariant imbedding T-matrix method and the separation of variables method applied to large nonspherical inhomogeneous particles. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 116:169–183, 2013.

[25] Andrew J Poggio and Edmund K Miller. *Integral equation solutions of three-dimensional scattering problems*. MB Assoc., 1970.

[26] Te-Kao Wu and Leonard L Tsai. Scattering from arbitrarily-shaped lossy dielectric bodies of revolution. *Radio Science*, 12(5):709–718, 1977.

[27] Joseph R Mautz and Roger F Harrington. Electromagnetic scattering from a homogeneous body of revolution. Technical report, Syracuse Univ. NY Dept. of Electrical and Computer Engineering, 1977.

[28] Roger F Harrington. Boundary integral formulations for homogeneous material bodies. *Journal of electromagnetic waves and applications*, 3(1):1–15, 1989.

[29] Carl Müller. *Foundations of the mathematical theory of electromagnetic waves*, volume 155. Springer Science & Business Media, 2013.

[30] Pasi Yla-Oijala and Matti Taskinen. Well-conditioned Muller formulation for electromagnetic scattering by dielectric objects. *IEEE Transactions on Antennas and Propagation*, 53(10):3316–3323, 2005.

[31] Anthony J Baran and Samuel P Groth. The application of the boundary element method in BEM++ to small extreme Chebyshev ice particles and the remote detection of the ice crystal number concentration

of small atmospheric ice particles. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 198:68–80, 2017.

[32] Angelika Bunse-Gerstner and Ignacio Gutiérrez-Cañas. A preconditioned GMRES for complex dense linear systems from electromagnetic wave scattering problems. *Linear algebra and its applications*, 416(1):135–147, 2006.

[33] Tahir Malas and Levent Gürel. Incomplete LU preconditioning with the multilevel fast multipole algorithm for electromagnetic scattering. *SIAM Journal on Scientific Computing*, 29(4):1476–1494, 2007.

[34] Bruno Carpentieri, Iain S Duff, and Luc Giraud. Experiments with sparse preconditioning of dense problems from electromagnetic applications. *CERFACS, Toulouse, France, Tech. Rep. TR/PA/00/04*, page 9, 2000.

[35] Ralf Hiptmair. Operator preconditioning. *Computers and mathematics with Applications*, 52(5):699–706, 2006.

[36] Kristof Cools, Francesco P Andriulli, and Eric Michielssen. A Calderón multiplicative preconditioner for the PMCHWT integral equation. *IEEE transactions on Antennas and Propagation*, 59(12):4579–4587, 2011.

[37] Su Yan, Jian-Ming Jin, and Zaiping Nie. A comparative study of Calderón preconditioners for PMCHWT equations. *IEEE transactions on Antennas and Propagation*, 58(7):2375–2383, 2010.

[38] K Niino and N Nishimura. Calderón preconditioning approaches for PMCHWT formulations for Maxwell's equations. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 25(5-6):558–572, 2012.

[39] Pasi Ylä-Oijala and Sami P Kiminki. Challenges in developing efficient Calderón preconditioners for resonating or high material contrast pen-

etrable objects. *Journal of Computational and Applied Mathematics*, 289:296–305, 2015.

[40] Paul Escapil-Inchauspé and Carlos Jerez-Hanckes. Fast Calderón Preconditioning for the Electric Field Integral Equation. *IEEE Transactions on Antennas and Propagation*, 2019.

[41] Ignacia Fierro and Carlos Jerez-Hanckes. Fast Calderón preconditioning for Helmholtz boundary integral equations. *Journal of Computational Physics*, page 109355, 2020.

[42] Anthony J Baran. From the single-scattering properties of ice crystals to climate prediction: A way forward. *Atmospheric Research*, 112:45–69, 2012.

[43] Stuart Fox. An Evaluation of Radiative Transfer Simulations of Cloudy Scenes from a Numerical Weather Prediction Model at Sub-Millimetre Frequencies Using Airborne Observations. *Remote Sensing*, 12(17):2758, 2020.

[44] Kuo-Nan Liou and Ping Yang. *Light scattering by ice crystals: fundamentals and applications*. Cambridge University Press, 2016.

[45] Michael I Mishchenko. *Electromagnetic scattering by particles and particle groups: an introduction*. Cambridge University Press, 2014.

[46] Bruce T Draine and Piotr J Flatau. Discrete-dipole approximation for scattering calculations. *JOSA A*, 11(4):1491–1499, 1994.

[47] Bruce T Draine and Piotr J Flatau. DDSCAT: The discrete dipole approximation for scattering and absorption of light by irregular particles. *Astrophysics Source Code Library*, pages ascl–0008, 2000.

[48] Maxim A Yurkin and Alfons G Hoekstra. The discrete dipole approximation: an overview and recent developments. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 106(1-3):558–589, 2007.

[49] Ping Yang and KN Liou. Finite-difference time domain method for light scattering by small ice crystals in three-dimensional space. *JOSA A*, 13(10):2072–2085, 1996.

[50] Ping Yang and KN Liou. Finite difference time domain method for light scattering by nonspherical and inhomogeneous particles. In *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications*, volume 1, page 174. Academic, 2000.

[51] Maxim A Yurkin, Alfons G Hoekstra, R Scott Brock, and Jun Q Lu. Systematic comparison of the discrete dipole approximation and the finite difference time domain method for large dielectric scatterers. *Optics Express*, 15(26):17902–17911, 2007.

[52] Chao Liu, Lei Bi, R Lee Panetta, Ping Yang, and Maxim A Yurkin. Comparison between the pseudo-spectral time domain method and the discrete dipole approximation for light scattering simulations. *Optics Express*, 20(15):16763–16776, 2012.

[53] Anthony J Baran, Ping Yang, and Stephan Havemann. Calculation of the single-scattering properties of randomly oriented hexagonal ice columns: a comparison of the T-matrix and the finite-difference time-domain methods. *Applied optics*, 40(24):4376–4386, 2001.

[54] Lei Bi and Ping Yang. Accurate simulation of the optical properties of atmospheric ice crystals with the invariant imbedding T-matrix method. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 138:17–35, 2014.

[55] Karri Muinonen. Scattering of light by crystals: a modified Kirchhoff approximation. *Applied optics*, 28(15):3044–3050, 1989.

[56] K Muinonen, T Nousiainen, P Fast, K Lumme, and JI Peltoniemi. Light scattering by Gaussian random particles: ray optics approximation.

Journal of Quantitative Spectroscopy and Radiative Transfer, 55(5):577–601, 1996.

[57] Andreas Macke, Johannes Mueller, and Ehrhard Raschke. Single scattering properties of atmospheric ice crystals. *Journal of the Atmospheric Sciences*, 53(19):2813–2825, 1996.

[58] Ping Yang and KN Liou. Geometric-optics–integral-equation method for light scattering by nonspherical ice crystals. *Applied optics*, 35(33):6568–6584, 1996.

[59] Michael I Mishchenko and Andreas Macke. Incorporation of physical optics effects and computation of the Legendre expansion for ray-tracing phase functions involving δ-function transmission. *Journal of Geophysical Research: Atmospheres*, 103(D2):1799–1805, 1998.

[60] Kuo-Nan Liou. *An introduction to atmospheric radiation*. Elsevier, 2002.

[61] Anatoli G Borovoi and Igor A Grishin. Scattering matrices for large ice crystal particles. *JOSA A*, 20(11):2071–2080, 2003.

[62] E Hesse. Modelling diffraction during ray tracing using the concept of energy flow lines. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 109(8):1374–1383, 2008.

[63] Lei Bi, Ping Yang, George W Kattawar, Bryan A Baum, Yong X Hu, David M Winker, R Scott Brock, and Jun Q Lu. Simulation of the color ratio associated with the backscattering of radiation by ice particles at the wavelengths of 0.532 and 1.064 μm. *Journal of Geophysical Research: Atmospheres*, 114(D4), 2009.

[64] Lei Bi, Ping Yang, George W Kattawar, Yongxiang Hu, and Bryan A Baum. Scattering and absorption of light by ice particles: solution by a new physical-geometric optics hybrid method. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 112(9):1492–1508, 2011.

[65] Evelyn Hesse, Andreas Macke, Stephan Havemann, AJ Baran, Zbigniew Ulanowski, and Paul H Kaye. Modelling diffraction by facetted particles. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 113(5):342–347, 2012.

[66] Lei Bi, Ping Yang, Chao Liu, Bingqi Yi, Bryan A Baum, Bastiaan Van Diedenhoven, and Hironobu Iwabuchi. Assessment of the accuracy of the conventional ray-tracing technique: Implications in remote sensing and radiative transfer involving ice clouds. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 146:158–174, 2014.

[67] Yuzo Mano. Exact solution of electromagnetic scattering by a three-dimensional hexagonal ice column obtained with the boundary-element method. *Applied optics*, 39(30):5541–5546, 2000.

[68] Takashi Y Nakajima, Teruyuki Nakajima, Kyu Yoshimori, Sumit K Mishra, and Sachchida N Tripathi. Development of a light scattering solver applicable to particles of arbitrary shape on the basis of the surface-integral equations method of Müller type. I. Methodology, accuracy of calculation, and electromagnetic current on the particle surface. *Applied optics*, 48(19):3526–3536, 2009.

[69] Mei Ping Yu, Yi Ping Han, Zhi Wei Cui, and An Tao Chen. Electromagnetic scattering by multiple dielectric particles under the illumination of unpolarized high-order Bessel vortex beam. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 195:107–113, 2017.

[70] Paul A Martin. *Multiple scattering: interaction of time-harmonic waves with N obstacles*. Number 107. Cambridge University Press, 2006.

[71] CD Westbrook, RC Ball, PR Field, and AJ Heymsfield. Theory of growth by differential sedimentation, with application to snowflake formation. *Physical Review E*, 70(2):021403, 2004.

[72] RJ Cotton, PR Field, Z Ulanowski, Paul H Kaye, Edwin Hirst, RS Greenaway, I Crawford, J Crosier, and James Dorsey. The effective density of small ice particles obtained from in situ aircraft observations of mid-latitude cirrus. *Quarterly Journal of the Royal Meteorological Society*, 139(676):1923–1934, 2013.

[73] Guosheng Liu. A database of microwave single-scattering properties for nonspherical ice particles. *Bulletin of the American Meteorological Society*, 89(10):1563–1570, 2008.

[74] Annalisa Buffa and Ralf Hiptmair. Galerkin boundary element methods for electromagnetic scattering. In *Topics in Computational Wave Propagation*, pages 83–124. Springer, 2003.

[75] Matthew W Scroggs, Timo Betcke, Erik Burman, Wojciech Śmigaj, and Elwin van't Wout. Software frameworks for integral equations in electromagnetic scattering based on Calderón identities. *Computers & Mathematics with Applications*, 74(11):2897–2914, 2017.

[76] Simon N Chandler-Wilde, David P Hewett, and Andrea Moiola. Sobolev Spaces on Non-Lipschitz Subsets of $\mathbb{R}^n$ with Application to Boundary Integral Equations on Fractal Screens. *Integral Equations and Operator Theory*, 87(2):179–224, 2017.

[77] Annalisa Buffa, Martin Costabel, and Dongwoo Sheen. On traces for $\mathbf{H}(\mathbf{curl}, \Omega)$ in Lipschitz domains. *Journal of Mathematical Analysis and Applications*, 276(2):845–867, 2002.

[78] Annalisa Buffa and Patrick Ciarlet Jr. On traces for functional spaces related to Maxwell's equations Part II: Hodge decompositions on the boundary of Lipschitz polyhedra and applications. *Mathematical Methods in the Applied Sciences*, 24(1):31–48, 2001.

[79] Jean-Claude Nédélec. *Acoustic and electromagnetic equations: integral representations for harmonic problems*. Springer Science & Business Media, 2001.

[80] Harry Contopanagos, Benjamin Dembart, Michael Epton, John J Ottusch, Vladimir Rokhlin, John L Visher, and Stephen M Wandzura. Well-conditioned boundary integral equations for three-dimensional electromagnetic scattering. *IEEE Transactions on Antennas and Propagation*, 50(12):1824–1830, 2002.

[81] Xavier Claeys, Ralf Hiptmair, and Carlos Jerez-Hanckes. Multi-trace boundary integral equations. *Direct and inverse problems in wave propagation and applications*, 14:51–100, 2012.

[82] Timo Betcke, Matthew W Scroggs, and Wojciech Śmigaj. Product algebras for Galerkin discretisations of boundary integral operators and their applications. *ACM Transactions on Mathematical Software (TOMS)*, 46(1):1–22, 2020.

[83] Stefan A Sauter and Christoph Schwab. Boundary element methods. In *Boundary Element Methods*, pages 183–287. Springer, 2010.

[84] Timo Betcke, Elwin van't Wout, and Pierre Gélat. Computationally efficient boundary element methods for high-frequency Helmholtz problems in unbounded domains. In *Modern Solvers for Helmholtz Problems*, pages 215–243. Springer, 2017.

[85] Pierre-Arnaud Raviart and Jean-Marie Thomas. A mixed finite element method for 2-nd order elliptic problems. In *Mathematical aspects of finite element methods*, pages 292–315. Springer, 1977.

[86] Sadasiva Rao, D Wilton, and Allen Glisson. Electromagnetic scattering by surfaces of arbitrary shape. *IEEE Transactions on antennas and propagation*, 30(3):409–418, 1982.

[87] Francesco P Andriulli, Kristof Cools, Hakan Bagci, Femke Olyslager, Annalisa Buffa, Snorre Christiansen, and Eric Michielssen. A multiplicative Calderón preconditioner for the electric field integral equation. *IEEE Transactions on Antennas and Propagation*, 56(8):2398–2412, 2008.

[88] Annalisa Buffa and Snorre Christiansen. A dual finite element complex on the barycentric refinement. *Mathematics of Computation*, 76(260):1743–1769, 2007.

[89] Mario Bebendorf. Approximation of boundary element matrices. *Numerische Mathematik*, 86(4):565–589, 2000.

[90] Stefan Kurz, Oliver Rain, and Sergej Rjasanow. The adaptive cross-approximation technique for the 3D boundary-element method. *IEEE transactions on Magnetics*, 38(2):421–424, 2002.

[91] Mario Bebendorf and Sergej Rjasanow. Adaptive low-rank approximation of collocation matrices. *Computing*, 70(1):1–24, 2003.

[92] Youcef Saad and Martin H Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on scientific and statistical computing*, 7(3):856–869, 1986.

[93] Lloyd N Trefethen and David Bau III. *Numerical linear algebra*, volume 50. Siam, 1997.

[94] Oscar Bruno, Tim Elling, Randy Paffenroth, and Catalin Turc. Electromagnetic integral equations requiring small numbers of Krylov-subspace iterations. *Journal of Computational Physics*, 228(17):6169–6183, 2009.

[95] Yassine Boubendir and Catalin Turc. Well-conditioned boundary integral equation formulations for the solution of high-frequency electromagnetic scattering problems. *Computers & Mathematics with Applications*, 67(10):1772–1805, 2014.

[96] Alan Ayala, Xavier Claeys, Paul Escapil-Inchauspé, and Carlos Jerez-Hanckes. Local multiple traces formulation for electromagnetics: Stability and preconditioning for smooth geometries. *arXiv preprint arXiv:2003.08330*, 2020.

[97] Xavier Antoine and Marion Darbas. An introduction to operator preconditioning for the fast iterative integral equation solution of time-harmonic scattering problems. *Multiscale Science and Engineering*, pages 1–35, 2021.

[98] George C Hsiao and Ralph E Kleinman. Mathematical foundations for error estimation in numerical solutions of integral equations in electromagnetics. *IEEE transactions on Antennas and Propagation*, 45(3):316–328, 1997.

[99] Felipe Vico, Leslie Greengard, and Zydrunas Gimbutas. Boundary integral equation analysis on the sphere. *Numerische Mathematik*, 128(3):463–487, 2014.

[100] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python. 2001.

[101] Xavier Claeys and Ralf Hiptmair. Electromagnetic scattering at composite objects: a novel multi-trace boundary integral formulation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(6):1421–1445, 2012.

[102] Christophe Geuzaine and Jean-François Remacle. Gmsh: A 3-D finite element mesh generator with built-in pre-and post-processing facilities. *International journal for numerical methods in engineering*, 79(11):1309–1331, 2009.

[103] Graeme L Stephens, Deborah G Vane, Ronald J Boain, Gerald G Mace, Kenneth Sassen, Zhien Wang, Anthony J Illingworth, Ewan J O'connor, William B Rossow, Stephen L Durden, et al. The CloudSat mission and

the A-Train: A new dimension of space-based observations of clouds and precipitation. *Bulletin of the American Meteorological Society*, 83(12):1771–1790, 2002.

[104] Graeme L Stephens and Deborah G Vane. Cloud remote sensing from space in the era of the A-Train. *Journal of Applied Remote Sensing*, 1(1):013507, 2007.

[105] J Schulz, P Albert, H-D Behr, D Caprion, H Deneke, S Dewitte, B Dürr, P Fuchs, A Gratzki, P Hechler, et al. Operational climate monitoring from space: the EUMETSAT Satellite Application Facility on Climate Monitoring (CM-SAF). *Atmospheric Chemistry and Physics*, 9(5):1687–1709, 2009.

[106] Andreas Macke, Michael I Mishchenko, and Brian Cairns. The influence of inclusions on light scattering by large ice particles. *Journal of Geophysical Research: Atmospheres*, 101(D18):23311–23316, 1996.

[107] Ping Yang, George W Kattawar, Gang Hong, Patrick Minnis, and Yongxiang Hu. Uncertainties associated with the surface texture of ice particles in satellite-based retrieval of cirrus clouds—Part I: Single-scattering properties of ice crystals with surface roughness. *IEEE transactions on geoscience and remote sensing*, 46(7):1940–1947, 2008.

[108] Ping Yang, Gang Hong, George W Kattawar, Patrick Minnis, and Yongxiang Hu. Uncertainties associated with the surface texture of ice particles in satellite-based retrieval of cirrus clouds: Part II—Effect of particle surface roughness on retrieved cloud optical thickness and effective particle size. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7):1948–1957, 2008.

[109] Junshik Um and Greg M McFarquhar. Single-scattering properties of aggregates of bullet rosettes in cirrus. *Journal of applied meteorology and climatology*, 46(6):757–775, 2007.

[110] Terry Davies, Mike JP Cullen, Andrew J Malcolm, MH Mawson, Andrew Staniforth, AA White, and Nigel Wood. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(608):1759–1782, 2005.

[111] Vyacheslav Ivanovich Lebedev. Values of the nodes and weights of ninth to seventeenth order gauss-markov quadrature formulae invariant under the octahedron group with inversion. *USSR Computational Mathematics and Mathematical Physics*, 15(1):44–51, 1975.

[112] Vyacheslav Ivanovich Lebedev. Quadratures on a sphere. *USSR Computational Mathematics and Mathematical Physics*, 16(2):10–24, 1976.

[113] Vyacheslav Ivanovich Lebedev. Spherical quadrature formulas exact to orders 25–29. *Siberian Mathematical Journal*, 18(1):99–107, 1977.

[114] Vyacheslav Ivanovich Lebedev and AL Skorokhodov. Quadrature formulas of orders 41, 47 and 53 for the sphere. In *Russian Academy of Sciences-Doklady Mathematics*, volume 45, pages 587–592, 1992.

[115] Vyacheslav Ivanovich Lebedev. A quadrature formula for the sphere of 59th algebraic order of accuracy. *Russian Academy of Sciences-Doklady Mathematics-AMS Translation*, 50(2):283–286, 1995.

[116] Vyacheslav Ivanovich Lebedev and DN Laikov. A quadrature formula for the sphere of the 131st algebraic order of accuracy. In *Doklady Mathematics*, volume 59, pages 477–481. Pleiades Publishing, Ltd.(,), 1999.

[117] Casper HL Beentjes. Quadrature on a spherical surface. *Working note available on the website* `https://cbeentjes.github.io`, 2015.

[118] Michael I Mishchenko and Maxim A Yurkin. On the concept of random orientation in far-field electromagnetic scattering by nonspherical particles. *Optics letters*, 42(3):494–497, 2017.

[119] Michael I Mishchenko. Calculation of the amplitude matrix for a nonspherical particle in a fixed orientation. *Applied optics*, 39(6):1026–1031, 2000.

[120] Lord Rayleigh. V. On the incidence of aerial and electric waves upon small obstacles in the form of ellipsoids or elliptic cylinders, and on the passage of electric waves through a circular aperture in a conducting screen. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 44(266):28–52, 1897.

[121] Michael I Mishchenko and Larry D Travis. Capabilities and limitations of a current FORTRAN implementation of the T-matrix method for randomly oriented, rotationally symmetric scatterers. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 60(3):309–324, 1998.

[122] Ping Yang, KN Liou, Michael I Mishchenko, and Bo-Cai Gao. Efficient finite-difference time-domain scheme for light scattering by dielectric particles: application to aerosols. *Applied Optics*, 39(21):3727–3737, 2000.

[123] Laurent C.-Labonnote, Gérard Brogniez, Jean-Claude Buriez, Marie Doutriaux-Boucher, Jean-François Gayet, and Andreas Macke. Polarized light scattering by inhomogeneous hexagonal monocrystals: Validation with ADEOS-POLDER measurements. *Journal of Geophysical Research: Atmospheres*, 106(D11):12139–12153, 2001.

[124] Yu Xie, Ping Yang, George W Kattawar, Patrick Minnis, and Yong X Hu. Effect of the inhomogeneity of ice crystals on retrieving ice cloud optical thickness and effective particle size. *Journal of Geophysical Research: Atmospheres*, 114(D11), 2009.