Journal of Computational Physics: X 5 (2020) 100047

Contents lists available at ScienceDirect

Journal of Computational Physics: X

www.elsevier.com/locate/jcpx

A fast solver for the narrow capture and narrow escape problems in the sphere

Jason Kaye [a,*], Leslie Greengard [a,b]

[a] Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, United States
[b] Flatiron Institute, Simons Foundation, New York, NY 10010, United States

ARTICLE INFO

Article history:
Received 24 July 2019
Received in revised form 4 December 2019
Accepted 5 December 2019
Available online 12 December 2019

Keywords:
Narrow capture
Narrow escape
Mean first passage time
Mixed boundary value problems
Integral equations
Fast multipole method

ABSTRACT

We present an efficient method to solve the narrow capture and narrow escape problems for the sphere. The *narrow capture* problem models the equilibrium behavior of a Brownian particle in the exterior of a sphere whose surface is reflective, except for a collection of small absorbing patches. The *narrow escape* problem is the dual problem: it models the behavior of a Brownian particle confined to the interior of a sphere whose surface is reflective, except for a collection of small patches through which it can escape.

Mathematically, these give rise to mixed Dirichlet/Neumann boundary value problems of the Poisson equation. They are numerically challenging for two main reasons: (1) the solutions are non-smooth at Dirichlet-Neumann interfaces, and (2) they involve adaptive mesh refinement and the solution of large, ill-conditioned linear systems when the number of small patches is large.

By using the Neumann Green's functions for the sphere, we recast each boundary value problem as a system of first-kind integral equations on the collection of patches. A block-diagonal preconditioner together with a multiple scattering formalism leads to a well-conditioned system of second-kind integral equations and a very efficient approach to discretization. This system is solved iteratively using GMRES. We develop a hierarchical, fast multipole method-like algorithm to accelerate each matrix-vector product. Our method is insensitive to the patch size, and the total cost scales with the number N of patches as $\mathcal{O}(N \log N)$, after a precomputation whose cost depends only on the patch size and not on the number or arrangement of patches. We demonstrate the method with several numerical examples, and are able to achieve highly accurate solutions with 100 000 patches in one hour on a 60-core workstation. For that case, adaptive discretization of each patch would lead to a dense linear system with about 360 million degrees of freedom. Our preconditioned system uses only 13.6 million "compressed" degrees of freedom and a few dozen GMRES iterations.

© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

We consider the numerical solution of two related problems which arise in the study of Brownian diffusion by a particle in the exterior or interior of a porous sphere. We denote the open unit ball centered at the origin in $\mathbb{R}^3$ by Ω, and assume

* Corresponding author.
E-mail addresses: jkaye@cims.nyu.edu (J. Kaye), greengard@cims.nyu.edu (L. Greengard).

https://doi.org/10.1016/j.jcpx.2019.100047
2590-0552/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

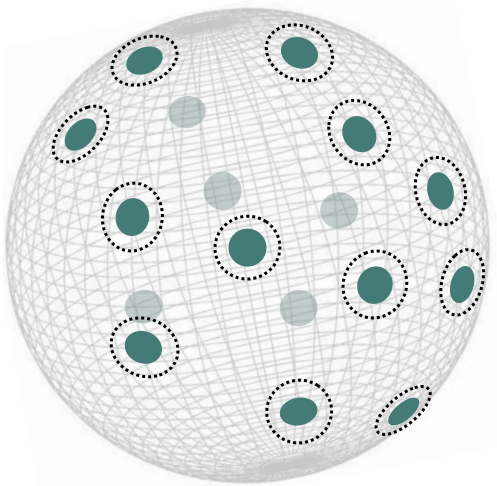

Fig. 1. A sphere partially covered by disk-shaped patches. We assume each patch is of radius ε. We also assume that distinct patches are separated by a distance of at least ε. In the figure, this means that the regions bounded by the dashed lines do not overlap.

that the sphere $\partial\Omega$ is partially covered by N small patches of radius ε, measured in arclength (Fig. 1). For the sake of simplicity, we assume that the patches are disk-shaped and comment briefly on more general shapes in the conclusion.

The union of the patches is referred to as the *absorbing boundary* and denoted by Γ_A. The remainder of the boundary, $\Gamma_R = \partial\Omega \backslash \Gamma_A$, is referred to as the *reflecting boundary*. The first problem, called the narrow capture problem, is to calculate the concentration $\bar{u}(x)$, at equilibrium, of Brownian particles at $x \in \mathbb{R}^3 \backslash \overline{\Omega}$ with a given fixed concentration far from the origin, assuming that particles are absorbed (removed) at Γ_A. The second problem, called the narrow escape problem, is to calculate the mean first passage time (MFPT) in Ω, namely the expected time $\bar{v}(x)$ for a Brownian particle released at $x \in \Omega$ to first reach Γ_A. In both settings, particles are reflected from Γ_R. In this paper, we sometimes refer to the narrow capture problem as the exterior problem, and the narrow escape problem as the interior problem.

These problems have received quite a lot of attention in the mathematics and biophysics communities since the seminal work of Berg and Purcell [1]. We do not seek to review the biophysical background here, but note that the absorbing patches serve as a simplified model for either surface receptors (the capture mechanism) or pores (the escape mechanism) in an otherwise impermeable membrane. We refer the reader to [1–7] for more detailed discussions of applications and a selection of work on related biophysical models.

Standard arguments from stochastic analysis show that $\bar{u}$ and $\bar{v}$ satisfy the Laplace and Poisson equations, respectively, with mixed Dirichlet-Neumann boundary conditions [8,9]. More precisely, for the capture problem, if the far-field particle concentration is set to be 1, then $\bar{u}$ satisfies the exterior Laplace equation:

$$\begin{cases} \Delta \bar{u} = 0 & x \in \mathbb{R}^3 \backslash \overline{\Omega} \\ \bar{u} = 0 & x \in \Gamma_A \\ \frac{\partial \bar{u}}{\partial n} = 0 & x \in \Gamma_R \\ \bar{u}(x) \to 1 & |x| \to \infty. \end{cases} \tag{1}$$

A scalar quantity of interest is the total equilibrium flux J of particles through Γ_A:

$$J = \int_{\Gamma_A} \frac{\partial \bar{u}}{\partial n} \, dS. \tag{2}$$

This flux J is related to the so-called capacitance via the formula $C = \frac{J}{4\pi}$ [10]. For the escape problem, the MFPT $\bar{v}$ satisfies the interior Poisson equation:

$$\begin{cases} \Delta \bar{v} = -1 & x \in \Omega \\ \bar{v} = 0 & x \in \Gamma_A \\ \frac{\partial \bar{v}}{\partial n} = 0 & x \in \Gamma_R. \end{cases} \tag{3}$$

Here, the quantity of interest is the average MFPT μ – that is the average, over all possible initial particle positions, of the expected time to escape from Ω through Γ_A:

$$\mu = \frac{1}{|\Omega|} \int_{\Omega} \bar{v} \, dV. \tag{4}$$

In the above, $\frac{\partial}{\partial n}$ refers to the derivative in the outward normal direction; n points towards the interior of Ω for the exterior problem, and towards the exterior of Ω for the interior problem. In order to understand how the distribution of absorbing patches on the surface affects $\bar{u}(x)$, $\bar{v}(x)$ and the associated quantities J and μ, a variety of asymptotic and numerical methods have been developed (see [1,11,12,10,13,4,5] and the references therein).

Remark 1. The total flux J is computed directly from the Neumann data on Γ_A, as seen from (2). Likewise, the average MFPT μ can be computed directly from the Dirichlet data $\bar{v}$ on Γ_R. For this, we use Green's second identity,

$$\int_\Omega (\psi \Delta\varphi - \varphi \Delta\psi)\, dV = \int_{\partial\Omega} \left(\psi \frac{\partial \varphi}{\partial n} - \varphi \frac{\partial \psi}{\partial n} \right) dS$$

with $\psi(x) \equiv \bar{v}(x)$ and $\varphi(x) \equiv \frac{|x|^2}{6}$. Using that $\Delta \frac{|x|^2}{6} = 1$, $\int_\Omega \frac{|x|^2}{6} dV(x) = \frac{2\pi}{15}$, and that for $|x| = 1$, $n \equiv x$ and $\frac{\partial}{\partial n} \frac{|x|^2}{6} = \frac{1}{3}$, we obtain

$$\int_\Omega \bar{v}\, dV = \frac{1}{3} \int_{\partial\Omega} \bar{v}\, dS - \frac{1}{6} \int_{\partial\Omega} \frac{\partial \bar{v}}{\partial n} dS - \frac{2\pi}{15}.$$

Applying the divergence theorem to the second term, dividing by $|\Omega|$, and using that $|\Omega| = \frac{4\pi}{3}$, $|\partial\Omega| = 4\pi$ gives an alternative expression for μ:

$$\mu = \frac{1}{|\partial\Omega|} \int_{\partial\Omega} \bar{v}\, dS + \frac{1}{15} \equiv \frac{1}{|\partial\Omega|} \int_{\Gamma_R} \bar{v}\, dS + \frac{1}{15}. \tag{5}$$

Thus the average MFPT over Ω may be obtained from the average MFPT on $\partial\Omega$.

Given an arrangement of patches, we present here a fast, high-order accurate numerical scheme for the evaluation of $\bar{u}$, J, $\bar{v}$, and μ, of particular use when N is large and ε is small. Such computations are numerically challenging, partly because solutions of elliptic boundary value problems of mixed type are singular near Dirichlet-Neumann interfaces [14,15]. Direct discretization, using either PDE-based methods or integral equation methods, would require many degrees of freedom to resolve the singularities in $\bar{u}$ and $\bar{v}$. Further, the resulting linear systems would be large and ill-conditioned, especially in cases involving large numbers of small patches.

The formulation presented here is *well-conditioned*, is nearly identical for the capture and escape problems, and suffers no loss in accuracy or increase in computational cost as ε is decreased. To make large-scale problems practical, we have developed a fast algorithm, so that the cost per GMRES iteration [16] is of the order $\mathcal{O}(N \log N)$, rather than $\mathcal{O}(N^2)$. Our method involves the following ingredients:

- We make use of the Neumann Green's functions for the interior and exterior of the sphere to recast (1) and (3) as first-kind integral equations for a density σ on Γ_A.
- Given a patch radius ε, we precompute the solution operator for the corresponding one-patch integral equation, assuming smooth Dirichlet data which is expanded in a rapidly converging series of Zernike polynomials. We analytically incorporate a square root singularity in the induced density at the Dirichlet/Neumann interface.
- To solve the many-patch integral equation, we use the solution operator for the one-patch integral equation as a block-diagonal "right preconditioner". This yields a second-kind Fredholm system of equations which, upon discretization, is well-conditioned and has a small number of degrees of freedom per patch.
- We solve the resulting linear system by iteration, using GMRES, and accelerate each matrix-vector product by means of a fast algorithm modeled after the fast multipole method (FMM). The fast algorithm uses the *interpolative decomposition* [17] to derive a compressed representation of the outgoing field induced by the density on a patch, a hierarchical organization of patches into groups at different length scales, and a spectral representation of the smooth incoming field due to densities on distant patches.

Though most of the past work on the narrow capture and narrow escape problems is based on asymptotics, we wish to highlight the numerical work of Bernoff and Lindsay, who also proposed an integral equation method for the narrow capture problem for the sphere and the plane based on the Neumann Green's function [10]. Our approach to discretization shares several characteristics with theirs: both methods incorporate a square root singularity into the density on each patch analytically, and both use a representation in terms of Zernike polynomials for smooth Dirichlet data on each patch. The solver itself is quite different, incorporating the ingredients listed above as well as adaptive refinement toward the Dirichlet/Neumann interfaces.

The paper is organized as follows. In Section 2, we introduce the analytical framework for our method, reformulate the boundary value problems as first-kind integral equations using single layer potentials, and explain how to calculate the scalar quantities J and μ directly as functionals of the layer potential densities. In Section 3, we show how to transform

the first-kind integral equations into Fredholm equations of the second-kind, using the solution operator for the one-patch integral equation as a preconditioner. In Sections 4, 5, and 6 we describe our discretization approach for the full system of equations, and in Section 7 we introduce the technical tools involved in our fast algorithm. In Section 8 we describe the full method, including our fast algorithm to accelerate the application of the system matrix. In Section 9, we provide a detailed description of the solver for the one-patch integral equation. We demonstrate the performance of the method with numerical experiments in Section 10.

2. Analytical setup

Our approach to solving the exterior and interior problems (1) and (3) uses a representation of each solution as an integral involving the corresponding Neumann Green's function. This representation leads to an integral equation, and the scalar quantity of interest – J or μ – can be calculated directly from its solution.

2.1. Neumann Green's functions for the sphere

Let us first consider the exterior Neumann problem:

$$\begin{cases} \Delta u = 0 & x \in \mathbb{R}^n \backslash \overline{\Omega} \\ \frac{\partial u}{\partial n} = g & x \in \partial\Omega \\ u(x) \to 0 & |x| \to \infty. \end{cases} \tag{6}$$

Here Ω is a bounded domain, and g a given continuous function on $\partial\Omega$. This problem has a unique solution, and if Ω is the unit ball in $\mathbb{R}^3$, it may be obtained using the *exterior Neumann Green's function* $G_E(x, x')$, which is known analytically [18,19]. G_E is symmetric, and satisfies

$$\begin{cases} -\Delta G_E(x, x') = 4\pi\delta(x - x') & x, x' \in \mathbb{R}^3 \backslash \Omega \\ \frac{\partial}{\partial n_{x'}} G_E(x, x') = 0 & x \in \mathbb{R}^3 \backslash \Omega,\ x' \in \partial\Omega,\ x \neq x', \end{cases} \tag{7}$$

with $G_E(x, x') = \mathcal{O}\left(|x|^{-1}\right)$ as $|x| \to \infty$ for fixed $x' \in \mathbb{R}^3 \backslash \Omega$. It can be shown, using Green's second identity, that

$$u(x) = \frac{1}{4\pi} \int_{\partial\Omega} G_E(x, x') g(x')\, dS(x') \tag{8}$$

solves the exterior Neumann problem (6). When $x' \in \partial\Omega$, G_E is given explicitly by

$$G_E(x, x') = \frac{2}{|x - x'|} + \log\left(\frac{|x| - x \cdot x'}{1 - x \cdot x' + |x - x'|}\right). \tag{9}$$

If, in addition, $x \in \partial\Omega$, then

$$G_E(x, x') = \frac{2}{|x - x'|} - \log\left(\frac{2}{|x - x'|}\right) - \log\left(1 + \frac{1}{2}|x - x'|\right). \tag{10}$$

The interior Neumann problem is given by

$$\begin{cases} \Delta v = 0 & x \in \Omega \\ \frac{\partial v}{\partial n} = g & x \in \partial\Omega, \end{cases} \tag{11}$$

where Ω is a bounded domain and g is a continuous function defined on the boundary, with the additional constraint that g must satisfy the consistency condition

$$\int_{\partial\Omega} g\, dS = 0.$$

This problem has a solution which is unique up to an additive constant. The consistency condition precludes the existence of an interior Green's function with zero Neumann data. Rather, for Ω the unit ball in $\mathbb{R}^3$, we have an *interior Neumann Green's function* $G_I(x, x')$, also known analytically [18,19]. It is again symmetric and satisfies

$$\begin{cases} -\Delta G_I(x, x') = 4\pi\delta(x - x') & x, x' \in \Omega \\ \frac{\partial}{\partial n_{x'}} G_I(x, x') = -1 & x \in \overline{\Omega},\ x' \in \partial\Omega,\ x \neq x'. \end{cases} \tag{12}$$

As before,

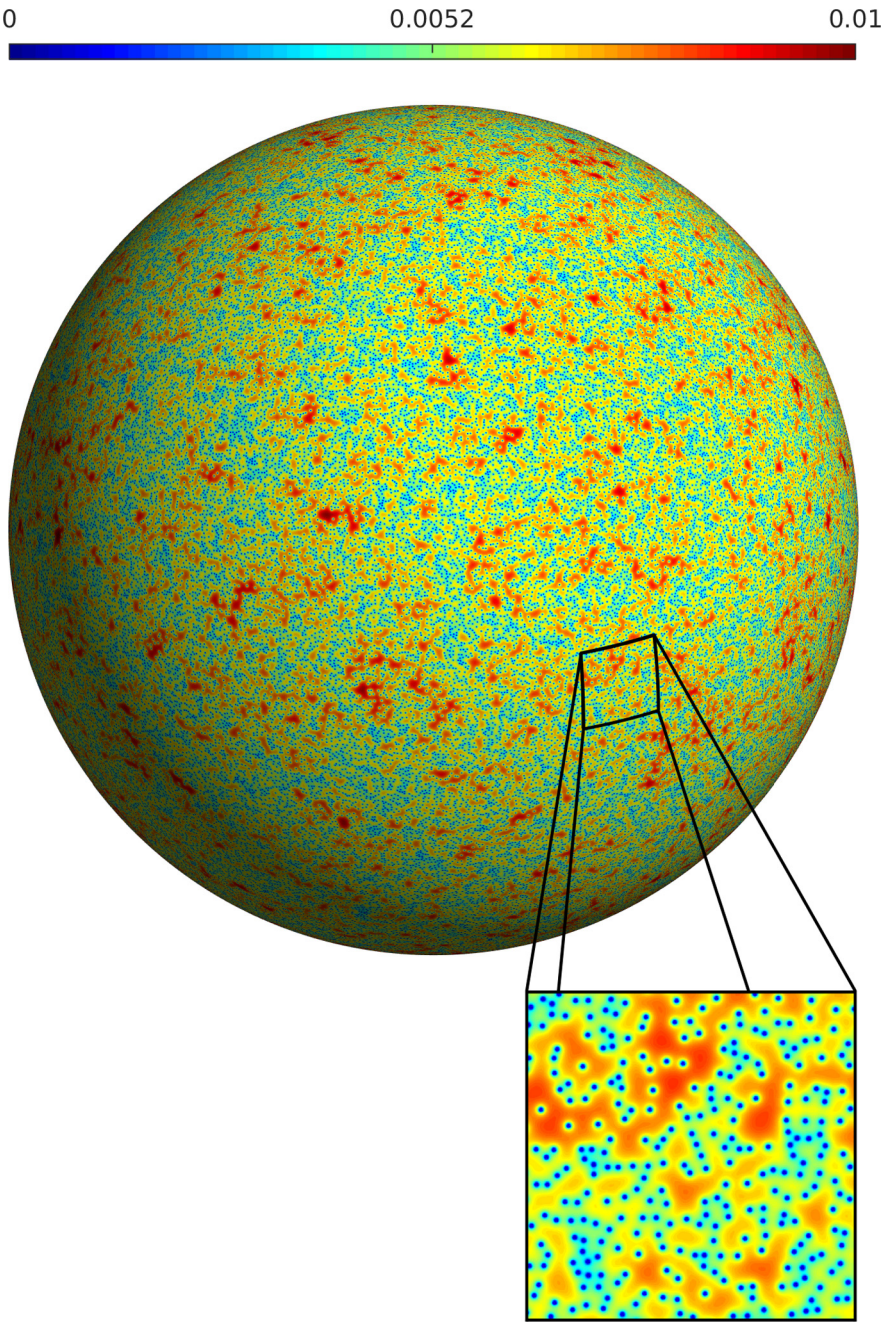

Fig. 2. MFPT $\bar{v}$ plotted just inside the unit sphere for an example with $N = 100\,000$ random well-separated patches of radius $\varepsilon \approx 0.00141$. The integral equation associated with this problem was solved in 63 minutes on a 60-core workstation, to an L^2 residual error of approximately 2.2×10^{-8}. Further details are given in Section 10.2.

$$v(x) = \frac{1}{4\pi} \int_{\partial\Omega} G_I(x, x') g(x')\, dS(x') \tag{13}$$

solves the interior Neumann problem (11). When $x' \in \partial\Omega$, G_I is given by

$$G_I(x, x') = \frac{2}{|x - x'|} + \log\left(\frac{2}{1 - x \cdot x' + |x - x'|}\right). \tag{14}$$

If, in addition, $x \in \partial\Omega$, this reduces to

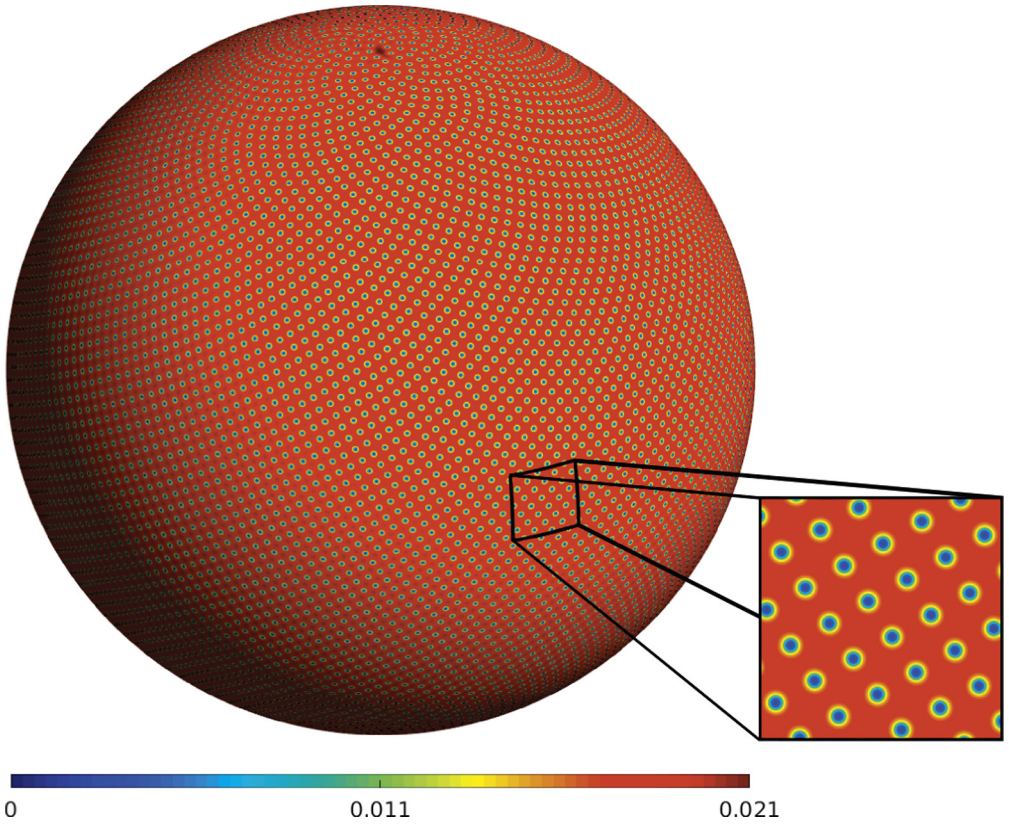

Fig. 3. MFPT $\bar{v}$ plotted just inside the unit sphere for an example with $N = 10\,000$ uniformly distributed patches of radius $\varepsilon \approx 0.00447$. The integral equation associated with this problem was solved in 114 seconds on a 60-core workstation, and in 15 minutes on a four-core, eight-thread laptop, to an L^2 residual error of approximately 6.4×10^{-8}. Further details are given in Section 10.2.

$$G_I(x,x') = \frac{2}{|x-x'|} + \log\left(\frac{2}{|x-x'|}\right) - \log\left(1+\frac{1}{2}|x-x'|\right). \tag{15}$$

This is the same as (10) except for the sign of the second term. In other words, the restrictions of the interior and exterior Green's functions to the boundary $\partial\Omega$ are nearly identical.

The following lemma, which we will require in the next section, follows from the second property in (12) and the symmetry of G_I.

Lemma 1. *Let Γ be an open subset of $\partial\Omega$ and let σ be continuous on Γ. Then for $x \in \partial\Omega\backslash\bar{\Gamma}$,*

$$\frac{\partial}{\partial n_x}\int_\Gamma G_I(x,x')\sigma(x')\,dS(x') = -\int_\Gamma \sigma(x')\,dS(x').$$

2.2. The narrow capture problem

We turn now to the narrow capture problem, which is the simpler of the two. We first modify the BVP (1) by defining $u = 1 - \bar{u}$, so that solutions decay as $|x| \to \infty$. The function u satisfies the modified equations

$$\begin{cases} \Delta u = 0 & x \in \mathbb{R}^3\backslash\Omega \\ u = 1 & x \in \Gamma_A \\ \frac{\partial u}{\partial n} = 0 & x \in \Gamma_R \\ u(x) \to 0 & |x| \to \infty. \end{cases} \tag{16}$$

Let us denote the unknown Neumann data on Γ_A by $\sigma(x')$. Then (8) implies that for $x \in \mathbb{R}^3 \backslash \overline{\Omega}$, we have

$$u(x) = \frac{1}{4\pi} \int_{\Gamma_A} G_E(x, x') \frac{\partial u}{\partial n}(x')\, dS(x') \equiv \int_{\Gamma_A} G_E(x, x')\sigma(x')\, dS(x'). \tag{17}$$

By analogy with classical potential theory, we refer to this as a *single layer potential* representation with density σ supported on Γ_A. Since the dominant singularity of the kernel G_E is that of the free-space Green's function for the Laplace equation, this single layer potential is continuous up to $\partial\Omega$. Taking the limit as $x \to \Gamma_A$ and using the second condition in (16), we obtain the first-kind integral equation

$$\int_{\Gamma_A} G_E(x, x')\sigma(x')\, dS(x') = f(x), \quad x \in \Gamma_A, \tag{18}$$

where $f(x) \equiv 1$, with the weakly singular kernel G_E. Assuming that we can solve (18) for σ, it follows that $u(x)$, given by (17), is the solution to (16), and that $\bar{u} = 1 - u$ solves (1). Furthermore, since $\sigma \equiv \frac{\partial u}{\partial n} \equiv -\frac{\partial \bar{u}}{\partial n}$ on Γ_A, the total flux J from (2) will be given by

$$J = -I_\sigma$$

where we have introduced the shorthand

$$I_\sigma := \int_{\Gamma_A} \sigma\, dS. \tag{19}$$

We will not prove the existence of a solution to (18), but sketch a possible approach. If we replace the kernel G_E in (18) with its first term $\frac{2}{|x-x'|}$, which is the free-space Green's function for the Laplace equation (up to a constant scaling factor), we obtain the first-kind integral equation for the Dirichlet problem on an open surface, which we can denote in operator form by

$$\mathcal{S}_0 \sigma = f.$$

This is a well-studied problem, which has a unique solution in the Sobolev space $H^{-\frac{1}{2}}(\Gamma_A)$ given data in $H^{\frac{1}{2}}(\Gamma_A)$ [20]. Writing the full single layer potential operator in the form $\mathcal{S}_0 + K$, where K is a compact pseudodifferential operator of order -2, we may rewrite (18) in the form of a Fredholm integral equation of the second kind:

$$(I + \mathcal{S}_0^{-1} K)\sigma = \mathcal{S}_0^{-1} f. \tag{20}$$

Thus, to prove existence and uniqueness for the single patch equation, one can apply the Fredholm alternative to (20). That is, one need only show that the homogenous version of the single patch equation has no nontrivial solutions. This is straightforward to prove when ε is sufficiently small, since the norm of K goes to zero as ε goes to zero and the corresponding Neumann series converges. We conjecture that the result holds for any ε.

2.3. The narrow escape problem

The analytical formulation of the narrow escape problem is somewhat more complicated than that of the narrow capture problem, largely because of the non-uniqueness of the interior Neumann problem, but it leads to a similar integral equation. We first recast the Poisson problem (3) as a Laplace problem with inhomogeneous boundary conditions. Assume that v satisfies

$$\begin{cases} \Delta v = 0 & x \in \Omega \\ v = 1 & x \in \Gamma_A \\ \frac{\partial v}{\partial n} = D & x \in \Gamma_R, \end{cases} \tag{21}$$

for some non-zero constant D. Then $\bar{v}$ given by

$$\bar{v} = \frac{v - 1}{3D} + \frac{1 - |x|^2}{6} \tag{22}$$

solves (3). We will therefore seek a method to produce a solution of (21) for some $D \neq 0$.

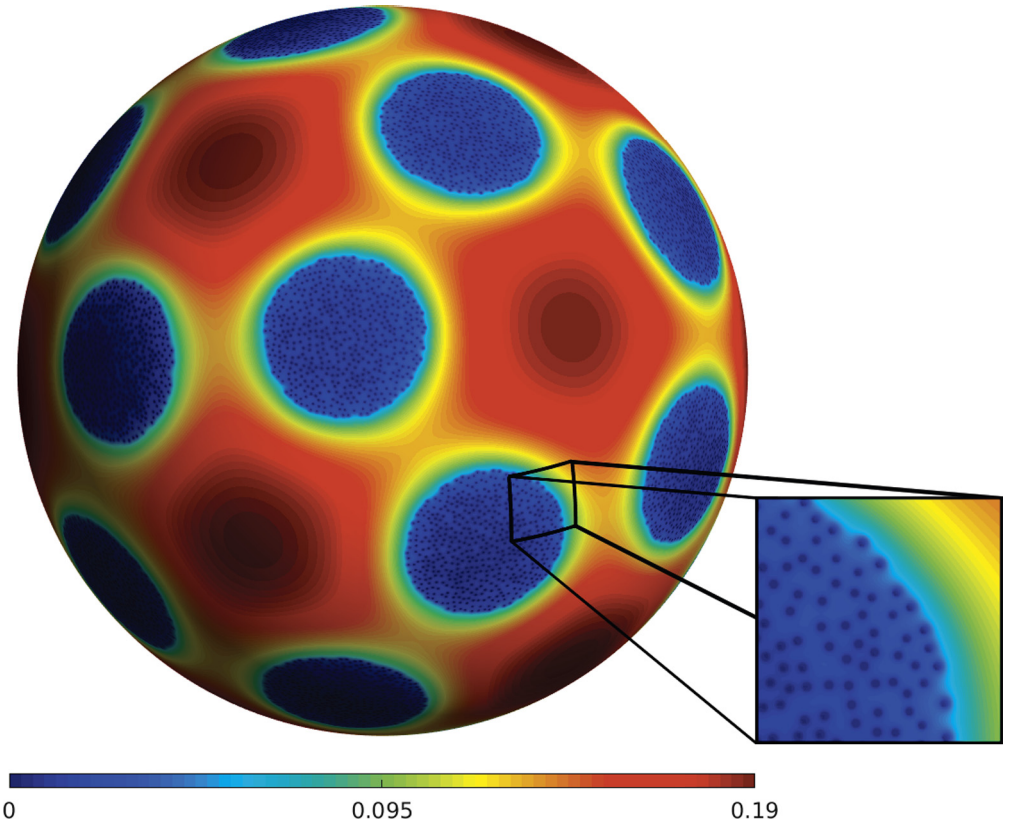

Fig. 4. MFPT $\bar{v}$ plotted just inside the unit sphere for an example with $N = 10\,000$ random, clustered patches of radius $\varepsilon \approx 0.0035$. The integral equation associated with this problem was solved in 269 seconds on a 60-core workstation, and in 35 minutes on a four-core, eight-thread laptop, to an L^2 residual error of approximately 6.5×10^{-8}. Further details are given in Section 10.2.

Lemma 2. *Let*

$$v(x) = \int_{\Gamma_A} G_I(x, x')\sigma(x')\, dS(x'), \tag{23}$$

where σ satisfies the first-kind integral equation

$$\int_{\Gamma_A} G_I(x, x')\sigma(x')\, dS(x') = 1 \tag{24}$$

for $x \in \Gamma_A$. Then v solves (21) with $D = -I_\sigma$, for I_σ defined as in (19), and $I_\sigma \neq 0$.

Proof. The function $v(x)$ is harmonic in Ω, and by Lemma 1, it satisfies the third condition of (21) with $D \equiv -I_\sigma$, as long as $I_\sigma \neq 0$. Taking x to Γ_A and using the continuity of the single layer potential up to Γ_A, we find that v will satisfy the second condition of (21) as long as σ satisfies (24).

It remains only to show that if σ satisfies (24), then $I_\sigma \neq 0$. If not, then v given by (23) satisfies (21) with $D = 0$, as does the constant function 1. It follows from Green's identity that solutions to (21) with the same value of D are unique, so we must have $v \equiv 1$. The formula (14) for G_I shows that if $|x'| = 1$, then $G_I(0, x') = 2$, so if $v \equiv 1$ we have

$$1 = v(0) = 2 \int_{\Gamma_A} \sigma(x')\, dS(x') = 2I_\sigma,$$

a contradiction. □

The question of the existence of a solution to (24) is analogous to that for (18), which was discussed in Section 2.2.

To calculate the average MFPT μ directly from σ, we plug (22) into (5) to obtain

$$\mu = \frac{1}{3D|\partial\Omega|}\int_{\partial\Omega} v\,dS - \frac{1}{3D} + \frac{1}{15}. \tag{25}$$

To calculate $\frac{1}{|\partial\Omega|}\int_{\partial\Omega} v\,dS$, we use the representation (23):

$$\begin{aligned}\frac{1}{|\partial\Omega|}\int_{\partial\Omega} v\,dS &= \frac{1}{|\partial\Omega|}\int_{\partial\Omega}\int_{\Gamma_A} G_I(x,x')\sigma(x')\,dS(x')\,dS(x)\\ &= \int_{\Gamma_A}\sigma(x')\left(\frac{1}{|\partial\Omega|}\int_{\partial\Omega} G_I(x,x')\,dS(x)\right)dS(x').\end{aligned}$$

A calculation using the explicit form (15) of G_I gives

$$\frac{1}{|\partial\Omega|}\int_{\partial\Omega} G_I(x,x')\,dS(x) = 2$$

for any $x' \in \partial\Omega$. We therefore have

$$\frac{1}{|\partial\Omega|}\int_{\partial\Omega} v\,dS = 2I_\sigma.$$

Plugging this into (25) and replacing D by $-I_\sigma$ gives

$$\mu = \frac{1}{3I_\sigma} - \frac{3}{5}. \tag{26}$$

3. A multiple scattering formalism

We have shown that the solutions of the two boundary value problems of interest, as well the associated scalars J and μ, may be obtained by solving (18) and (24), respectively, on the collection of absorbing patches. These integral equations differ only by the sign of one term in their respective kernels, as seen in Section 2.1. Since our treatment of the two cases is the same, we drop the subscripts on G_E and G_I, and discuss the solution of

$$\int_{\Gamma_A} G(x,x')\sigma(x')\,dS(x') = 1 \quad x \in \Gamma_A,$$

where σ is an unknown density on Γ_A. Letting $\Gamma_A = \cup_{i=1}^N \Gamma_i$, where Γ_i is the ith patch, and letting σ_i be the restriction of σ to Γ_i, we write this equation in the form

$$\sum_{j=1}^N \int_{\Gamma_j} G(x,x')\sigma_j(x')\,dS(x') = 1 \quad x \in \Gamma_i,\ i = 1,\ldots,N. \tag{27}$$

For the sake of simplicity, we assume that each patch has the same radius ε. We also assume that the patches are *well-separated*, in the sense that the distance between the centers of any two patches in arc length along the surface of the sphere is at least 3ε. That is, any two patches are separated by a distance greater than or equal to their own radius. For $x \in \Gamma_i$, we define $\mathcal{S}_{ij}$ by

$$(\mathcal{S}_{ij}\sigma_j)(x) := \int_{\Gamma_j} G(x,x')\sigma_j(x')\,dS(x').$$

More specifically, we define each such operator in a coordinate system fixed about the center of Γ_j. Since all the patches have the same radius, the operators $\mathcal{S}_{ii}$ are therefore identical, and we denote $\mathcal{S}_{ii}$ by $\mathcal{S}$. Thus we may rewrite the many-patch integral equation (27) in the form

$$\mathcal{S}\sigma_i + \sum_{j\neq i}^N \mathcal{S}_{ij}\sigma_j = 1 \quad i = 1,\ldots,N. \tag{28}$$

The aim of this section is to reformulate (28) as a Fredholm system of the second kind in an efficient basis.

Definition 1. Let f be a smooth function on some patch Γ_i. The *one-patch integral equation with data* f is defined by

$$\mathcal{S}\sigma_i = f, \tag{29}$$

where σ_i is an unknown density on Γ_i.

Remark 2. Writing (28) in the form

$$\mathcal{S}\sigma_i = 1 - \sum_{j\neq i}^{N} \mathcal{S}_{ij}\sigma_j,$$

and observing that $\mathcal{S}_{ij}\sigma_j$ is a smooth function for Γ_j well-separated from Γ_i, we see that each σ_i satisfies a one-patch integral equation with smooth data. Conversely, if $\sigma_1, \ldots, \sigma_N$ satisfy (28), then each $\mathcal{S}\sigma_i$ is smooth on Γ_i.

It is convenient to make use of an orthonormal basis $\{q_1, q_2, \ldots\}$ of smooth functions on each patch, so that for smooth f on Γ_i we have

$$f(x) = \sum_{n=1}^{\infty} \hat{f}_n q_n(x), \tag{30}$$

in the usual L^2 sense, with

$$\hat{f}_n = \int_{\Gamma_i} f(x) q_n(x)\, dx.$$

We postpone until Section 4 a discussion of our particular choice of the basis $\{q_n\}$, which will be constructed using Zernike polynomials. We will denote the vector of the first K coefficients by $\hat{f}^K$:

$$\hat{f}^K = (\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_K)^T.$$

Definition 2. Let f be a smooth function on some patch Γ_i defined by (30), with $\hat{f}$, $\hat{f}^K$ computed as above. The *projection* operators $\mathcal{P}$ and $\mathcal{P}^K$ are defined by

$$(\mathcal{P}[f])_n = \hat{f}_n,$$

with $\mathcal{P}^K$ defined in the same manner for $n \leq K$. The *synthesis* operators $\mathcal{Q}$ and $\mathcal{Q}^K$ are defined by

$$\mathcal{Q}[\hat{f}](x) = \sum_{n=1}^{\infty} \hat{f}_n q_n(x), \quad \mathcal{Q}^K[\hat{f}_K](x) = \sum_{n=1}^{K} f_n q_n(x).$$

$\mathcal{P}$ and $\mathcal{P}^K$ are left inverses of $\mathcal{Q}$ and $\mathcal{Q}^K$, respectively.

Finally, we define b_n to be the solution of the one-patch integral equation with data given by the basis element q_n:

$$b_n = \mathcal{S}^{-1} q_n. \tag{31}$$

Thus, if a smooth function f on Γ_i is expanded as $f = \sum_{n=1}^{\infty} \hat{f}_n q_n$, then the solution of the one-patch integral equation with data f is given by $\mathcal{S}^{-1} f = \sum_{n=1}^{\infty} \hat{f}_n b_n$. This motivates the following definition.

Definition 3. We denote the solution operator of the one-patch integral equation in the basis $\{q_n\}$ by

$$\mathcal{B} = \mathcal{S}^{-1}\mathcal{Q}.$$

For $\hat{f} = \{\hat{f}_1, \hat{f}_2, \ldots\}$ and $f(x) = \sum_{n=1}^{\infty} \hat{f}_n q_n(x)$, $\mathcal{B}$ satisfies

$$\mathcal{B}[\hat{f}](x) = \sum_{n=1}^{\infty} \hat{f}_n b_n(x).$$

We denote the solution operator of the one-patch integral equation in the truncated basis $\{q_n\}_{n=1}^{K}$ by

$$\mathcal{B}^K = \mathcal{S}^{-1}\mathcal{Q}^K.$$

For $\hat{f} = (\hat{f}_1, \hat{f}_2, \dots \hat{f}_K)$ and $f(x) = \sum_{n=1}^{K} \hat{f}_n q_n(x)$, $\mathcal{B}_K$ satisfies

$$\mathcal{B}^K[\hat{f}](x) = \sum_{n=1}^{K} \hat{f}_n b_n(x).$$

Note that the construction of $\mathcal{B}$ requires solving the one-patch integral equations with data $q_1, q_2, \dots$ to obtain $b_1, b_2, \dots$, and that the construction of $\mathcal{B}^K$ requires solving the first K of these equations. For a fixed patch radius ε, these solutions are universal and do not depend on the number or arrangement of patches in the full problem.

Given $\mathcal{B}$, we are now able to rewrite the integral equation (28) as a well-conditioned Fredholm system of the second kind in the basis $\{q_n\}$. On Γ_i, we *define* a function f_i by

$$f_i = \mathcal{S}\sigma_i.$$

Substituting into (28), we have

$$f_i + \sum_{j \neq i}^{N} \mathcal{S}_{ij}\mathcal{S}^{-1} f_j = 1 \quad i = 1, \dots, N.$$

To transform to the basis $\{q_n\}$, we write f_i in the form $f_i = \mathcal{Q}\hat{f}_i$ and multiply on the left by $\mathcal{P}$ to obtain

$$\hat{f}_i + \mathcal{P} \sum_{j \neq i}^{N} \mathcal{S}_{ij}\mathcal{B}\hat{f}_j = \mathcal{P}\,1 \quad i = 1, \dots, N. \tag{32}$$

Since the patches Γ_i and Γ_j are well-separated, $\mathcal{P}\mathcal{S}_{ij}\mathcal{B}$ is a compact operator for $i \neq j$, so that (32) is a Fredholm system of the second kind. The corresponding truncated system takes the form

$$\hat{f}_i^K + \mathcal{P}^K \sum_{j \neq i}^{N} \mathcal{S}_{ij}\mathcal{B}^K \hat{f}_j^K = \mathcal{P}^K\,1 \quad i = 1, \dots, N, \tag{33}$$

where we have used the approximation $f_i \approx \mathcal{Q}^K \hat{f}_i^K$.

Remark 3. We refer to the approach described above as a *multiple scattering* formalism by analogy with the problem of wave scattering from multiple particles in a homogeneous medium. In the language of scattering theory, one would say that for the ith patch, the boundary data is the known data ($\mathcal{S}\sigma_i = 1$), perturbed by the potential "scattered" from all other patches, namely $\sum_{j \neq i}^{N} \mathcal{S}_{ij}\sigma_j$. Solving the system (28) corresponds to determining how the collection of uncoupled single patch solutions $\mathcal{S}\sigma_i = 1$ needs to be perturbed to account for the "multiple scattering" effects. The approach developed above, where $f_i = \mathcal{S}\sigma_i$ are the unknowns, has many advantages over solving (28) directly, even with $\mathcal{S}^{-1}$ as a left preconditioner. By working in the spectral basis, we avoid the need to discretize σ_i on each patch, the number of degrees of freedom per patch is significantly reduced, and the linear system is a well-conditioned Fredholm equation of the second kind.

The multiple scattering framework is used in a variety of contexts. In fracture mechanics, for example, local preconditioners like $\mathcal{S}^{-1}$ play an important role in calculations involving many disjoint cracks [21,22].

Remark 4. The original unknowns σ_i may be recovered from the solution of (32) or (33) using the formula

$$\sigma_i = \mathcal{B}\hat{f}_i \approx \mathcal{B}^K \hat{f}_i^K. \tag{34}$$

Thus, we may think of the unknowns $\hat{f}_i$ as a representation of the unknown density σ_i in the basis $\{b_n\}$.

We turn now to the construction of an orthonormal basis $\{q_n\}$ for smooth functions on a patch, the construction of the singular solutions $b_n = \mathcal{S}^{-1}q_n$, and the efficient solution of the discretized multiple scattering system (33).

4. A basis for smooth functions on a patch

It is well-known that the Zernike polynomials are a spectrally accurate, orthogonal basis for smooth functions on the disk. For a thorough discussion of these functions, we refer the reader to [23]. Here, we simply summarize their relevant properties.

The Zernike polynomials on the unit disk $0 \leq r \leq 1$, $0 \leq \theta < 2\pi$ are given by

$$\begin{cases} Z_n^m(r,\theta) & = R_n^m(r)\cos(m\theta) \\ Z_n^{-m}(r,\theta) & = R_n^m(r)\sin(m\theta), \end{cases}$$

with $0 \le m < \infty$, $m \le n < \infty$, and

$$R_n^m(r) = (-1)^{(n-m)/2} r^m P^{m,0}_{(n-m)/2}(1-2r^2),$$

where $P_n^{\alpha,\beta}(x)$ is a Jacobi polynomial on $[-1,1]$. The Jacobi polynomials are orthogonal on $[-1,1]$ with respect to the weight function $(1-x)^\alpha(1+x)^\beta$. Thus, for fixed m, the functions $R_n^m(r)$ are orthogonal on $[0,1]$ with respect to the weight function r. This gives the orthogonality relation

$$\int_0^{2\pi}\int_0^1 Z_{n_1}^{m_1}(r,\theta) Z_{n_2}^{m_2}(r,\theta) r\, dr\, d\theta = \frac{(1+\delta_{m_1,0})\pi}{2n_1+2}\delta_{n_1,n_2}\delta_{m_1,m_2}. \tag{35}$$

The natural truncation of this basis is to fix a cutoff mode M in both the radial and angular variables, and to let $0 \le m \le n \le M$. This yields $K = (M+1)(M+2)/2$ basis functions. To use this basis on a generic patch Γ_i, we define a polar coordinate system (r,θ) about the patch center, for which r is the distance in arc length along the sphere from the center, and θ is the polar angle. We rescale the radial variable from $[0,1]$ to $[0,\varepsilon]$, transforming the Zernike polynomials to functions on Γ_i. Finally, the basis functions $q_1,\ldots,q_K$ discussed in Section 3 can be defined as the scaled Zernike polynomials up to mode M.

From the orthogonality relation (35), the projection operators $\mathcal{P}$ and $\mathcal{P}^K$ are obtained as normalized inner products against Zernike polynomials in polar coordinates. This *Zernike transform* can be implemented numerically using a tensor product quadrature with a Gauss-Legendre rule in the radial variable and a trapezoidal rule in the angular variable. The number of grid points required to obtain the exact Zernike coefficients of a function in the space spanned by $q_1,\ldots,q_K$ is $\mathcal{O}(K)$; we denote this number by K^*. We refer to these points as the *Zernike sampling nodes* $x_1^z,\ldots,x_{K^*}^z$ (see [23] for further details).

Remark 5. Rewriting (33) in the form

$$\hat{f}_i^K = \mathcal{P}^K\left(1 - \sum_{j\neq i}\mathcal{S}_{ij}\mathcal{B}^K\hat{f}_j^K\right), \tag{36}$$

we see that the truncation error compared with (32) depends on how well the smooth function

$$1 - \sum_{j\neq i}\mathcal{S}_{ij}\mathcal{B}^K\hat{f}_j^K$$

is represented in the space spanned by $q_1,\ldots,q_K$. In the one-patch case, the summation term vanishes, and $K=1$ is sufficient. For multiple patches, the choice of K depends largely on how well-separated the patches are. Since the Zernike basis is spectrally accurate, M grows only logarithmically with the desired precision. In practice, *a posteriori* estimates are easily obtained for any fixed configuration by inspection of the decay of the Zernike coefficients $\hat{f}_i^K$ in the computed solution.

5. Informal description of the one-patch solver

While the details of our solver for the one-patch integral equation

$$\mathcal{S}\sigma_i = f$$

are deferred to Section 9, we outline the general approach here. First, we note that in the absence of curvature (i.e. a flat disk on a half-space) and with the associated terms of the Green's function removed, the solution σ_i is known to have a square root singularity at the disk edge [10,14,15,20,24]. In our case, we will explicitly include this square root singularity in the representation of σ_i, but also allow for weaker singularities – which we have observed and will demonstrate in Section 9.3 – by using a discretization that is adaptively refined toward the edge $\partial\Gamma_i$.

Assume then that we have discretized the patch Γ_i using a suitable polar mesh with n_f *fine grid points*, denoted by $x_{i,1}^f,\ldots,x_{i,n_f}^f$. The fine grid points for different patches are identical relative to the coordinate systems of their own patches. We denote the corresponding samples of the right-hand side f and σ_i by

$$\vec{f} = (f(x_{i,1}^f),\ldots,f(x_{i,n_f}^f))^T,$$
$$\vec{\sigma}_i = ((\vec{\sigma}_i)_1,\ldots,(\vec{\sigma}_i)_{n_f})^T \approx (\sigma_i(x_{i,1}^f),\ldots,\sigma_i(x_{i,n_f}^f))^T.$$

We assume that $\mathcal{S}$ is discretized to high-order accuracy by a matrix S with

$$\mathcal{S}[\sigma_i](x^f_{i,k}) \approx \sum_{l=1}^{n_f} S(k,l)(\vec{\sigma}_i)_l, \tag{37}$$

so that the discretized system takes the form

$$S\vec{\sigma}_i = \vec{f}. \tag{38}$$

We will also require a set of quadrature weights, denoted by $w^f_1, \ldots, w^f_{n_f}$ and identical for each patch, that permit the accurate integration over Γ_i of the product of an arbitrary smooth function with the discretized density $\vec{\sigma}_i$, taking into account the fact that σ_i has an edge singularity. That is, we assume that

$$\int_{\Gamma_i} g(x)\sigma_i(x)\, dS(x) \approx \sum_{l=1}^{n_f} g(x^f_l)(\vec{\sigma}_i)_l w^f_l \tag{39}$$

for any smooth g, with high-order accuracy. In the next section, we will use this quadrature to discretize the operators $\mathcal{S}_{ij}$.

The solutions of the K one-patch integral equations (31) may be obtained in a precomputation, after which we have access to the functions $b_1, \ldots, b_K$ sampled on the fine grid. We assemble these functions into an $n_f \times K$ matrix B with

$$B(n,m) = b_m(x^f_n).$$

B is then the discretization of the operator $\mathcal{B}^K$, mapping the first K Zernike coefficients of a smooth function to the solution of the corresponding one-patch integral equation sampled on the fine grid. If we denote by Q the discretization of the synthesis operator $\mathcal{Q}^K$ as an $n_f \times K$ matrix,

$$Q(i,j) = q_j(x^f_i),$$

then we have, as in Definition 3,

$$SB = Q.$$

In short, the precomputation amounts to solving this matrix system for B.

6. Discretization of the multiple scattering system

We return now to the multiple scattering system (33). The unknowns on Γ_i are defined in the truncated Zernike basis as $\hat{f}^K_i$. We will need as intermediate variables the fine grid samples of $\sigma_i(x)$. From Remark 4, we define the sampling vector $\vec{\sigma}_i$ by

$$\vec{\sigma}_i = B\hat{f}^K_i \approx \mathcal{B}^K \hat{f}^K_i.$$

In order to discretize the integral operators $\mathcal{S}_{ij}$ for $i \neq j$, we note that $G(x,x')$ is smooth for $x \in \Gamma_i$, $x' \in \Gamma_j$, and use the quadrature (39). This yields

$$\int_{\Gamma_j} G(x,x')\sigma_j(x')\, dS(x') \approx \sum_{l=1}^{n_f} G(x, x^f_{j,l})(\vec{\sigma}_j)_l w^f_l. \tag{40}$$

Setting $x = x^z_{i,k}$ to be the kth Zernike sampling node on Γ_i, we define the matrix S_{ij} by

$$S_{ij}(k,l) = G(x^z_{i,k}, x^f_{j,l}) w^f_l.$$

Thus, S_{ij} maps a density sampled on the fine grid on Γ_j to the smooth field it induces at the Zernike sampling nodes on Γ_i. Lastly, we discretize the truncated Zernike transform $\mathcal{P}^K$ as a $K \times K^*$ matrix P using the trapezoidal-Legendre scheme described in Section 4.

Definition 4. The *discrete Zernike transform* P is defined to be the mapping of a smooth function sampled on the K^* Zernike sampling nodes to its K Zernike coefficients.

We can now write the multiple scattering system (33) in a fully discrete form,

$$\hat{f}^K_i + P\sum_{j\neq i} S_{ij} B \hat{f}^K_j = P\vec{1} \quad i = 1, \ldots, N, \tag{41}$$

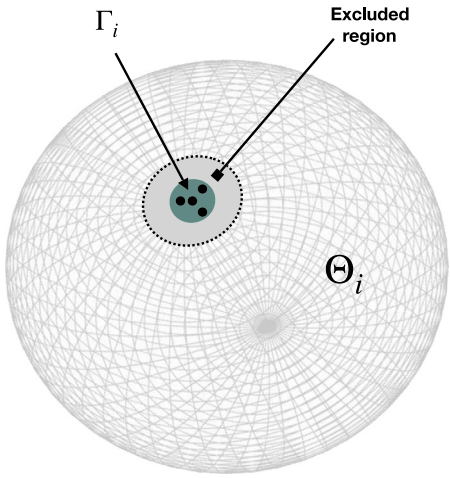

Fig. 5. For a patch Γ_i, the far field region Θ_i is defined as the complement on the surface of the sphere of a disk of radius 2ε, measured in arclength, about the center of Γ_i. The black dots in the figure represent the subset of the fine grid points used to efficiently represent the outgoing field induced by the density σ_i.

where $\vec{1}$ is the vector of length K^* with all entries equal to 1 – the discretization of the function $f = 1$. Since $P \in \mathbb{R}^{K\times K^*}$, $S_{ij} \in \mathbb{R}^{K^*\times n_f}$, and $B \in \mathbb{R}^{n_f\times K}$, this is a linear system of dimensions $KN \times KN$, with $K << n_f$ degrees of freedom per patch. As a discretization of a Fredholm system of the second kind, it is amenable to rapid solution using an iterative method such as GMRES [16].

We now describe how to calculate the constants J and μ from the solution of (41). We saw in Sections 2.2 and 2.3 that these can be computed directly from $I_\sigma = \sum_{i=1}^{N} \int_{\Gamma_i} \sigma_i \, dS$. Using the fine grid quadrature (39), we have

$$I_\sigma = \sum_{i=1}^{N} \int_{\Gamma_i} \sigma_i \, dS \approx \sum_{i=1}^{N} \sum_{k=1}^{n_f} (B \hat{f}_i^K)_k w_k^f = (w_1^f, \ldots, w_{n_f}^f) B \sum_{i=1}^{N} \hat{f}_i^K. \tag{42}$$

Since we may precompute the row vector $I := (w_1^f, \ldots, w_{n_f}^f)B$ of length K, the cost to compute I_σ is $\mathcal{O}(NK)$.

When the system (41) is solved iteratively, each matrix-vector product is dominated by the computation of the "multiple scattering events"

$$P \sum_{j\neq i} S_{ij} B \hat{f}_j^K \tag{43}$$

for $i = 1, \ldots, N$. That is, for each patch Γ_i, we must compute the Zernike coefficients of the field induced on that patch by the densities on all other patches. Note that if we were to calculate the above sums by simple matrix-vector products, the cost would be $\mathcal{O}(n_f K N^2)$. We turn now to the description of a scheme that permits the computation of these sums using $\mathcal{O}(KN \log N)$ operations, with a constant which depends only on the desired precision, but not on n_f.

7. Efficient representation of outgoing and incoming fields

Our fast algorithm relies on what is variously referred to as a compressed, skeletonized, or sparsified representation of the far field induced by a source density σ_i on a single patch Γ_i (Fig. 5). We define the far field region Θ_i for a patch Γ_i to be the set of points whose distance from the center of Γ_i (measured in arc length along the surface of the sphere) is greater than 2ε. In light of our restriction on the minimum patch separation distance, this ensures that the far field region of a particular patch contains every other patch.

We start from (40), which was used to define the matrix S_{ij}. We will show that there is a subset of p fine grid points with $p << n_f$ and modified source strengths $\vec{\rho}_i = (\rho_{i,1}, \rho_{i,2}, \ldots, \rho_{i,p})^T$ so that

$$\int_{\Gamma_i} G(x, x') \sigma_i(x') \, dS(x') \approx \sum_{l=1}^{n_f} G(x, x_{i,l}^f)(\vec{\sigma}_i)_l w_l^f \approx \sum_{m=1}^{p} G(x, x_{i,\pi(m)}^f) \rho_{i,m}, \tag{44}$$

for any $x \in \Theta_i$. Moreover, there is a stable algorithm for obtaining this *compressed* or *skeletonized outgoing representation*. Here, $\pi(m)$ is an indexing function which maps $\{1, \ldots, p\} \to \{1, \ldots, n_f\}$, and identifies which of the original fine grid points are used in the representation. The number p represents the numerical rank, to a specified precision, of the n_f functions $\{G(x, x_{i,l}^f)\}$ on Θ_i.

Remark 6. The existence of such low-rank factorizations is discussed in detail in [25–27]. For the purposes of computation, we will use the interpolative decomposition (ID) [17,25,28], described briefly below. The ID and related compression schemes are essential and widely used in hierarchical, fast algorithms for applying and inverting dense matrices (see for example [29–38] and the references therein).

7.1. The interpolative decomposition

We consider a generic patch Γ_i and, for simplicity, drop the patch index i on all quantities. We first discretize Θ on a training grid $x_1^t, \ldots, x_{n_t}^t$ of n_t points chosen to be sufficiently fine to accurately represent smooth functions on Θ. We can then obtain a matrix A of size $n_t \times n_f$, with entries $A_{jl} = G(x_j^t, x_l^f)$, so that the lth column of A is a discretization of the function $G(x, x_l^f)$ on the training grid. Given a user-specified tolerance ϵ, the ID takes as input a matrix A, and returns the factorization $\widetilde{A}\Pi$ with

$$\|A - \widetilde{A}\Pi\|_2 = O(\epsilon), \tag{45}$$

where $\widetilde{A}$ is $n_t \times p$ and Π is $p \times n_f$. The parameter p is the numerical rank of A determined by the ID as part of the factorization. The columns of $\widetilde{A}$ are a p-column subset of the original matrix A, chosen so that the column space of $\widetilde{A}$ approximates that of A. The matrix Π contains the coefficients needed to approximately reconstruct the columns of A from those of $\widetilde{A}$. If we define the indexing function π so that the mth column of $\widetilde{A}$ is the $\pi(m)$th column of A, then the approximation (45) implies that

$$G(x_j^t, x_l^f) \approx \sum_{m=1}^{p} G(x_j^t, x_{\pi(m)}^f)\Pi_{ml}$$

for $l = 1, \ldots, n_f$. Since the columns of A represent the functions $\{G(x, x_l^f)\}$ on a fine training grid, the expression above holds not just for $x \in \{x_j^t\}$, but more generally for $x \in \Theta$. That is,

$$G(x, x_l^f) \approx \sum_{m=1}^{p} G(x, x_{\pi(m)}^f)\Pi_{ml}.$$

Summing both sides of this expression against $(\vec{\sigma})_l w_l^f$ and rearranging yields

$$\sum_{l=1}^{n_f} G(x, x_l^f)(\vec{\sigma})_l w_l^f \approx \sum_{l=1}^{n_f}\sum_{m=1}^{p} G(x, x_{\pi(m)}^f)\Pi_{ml}(\vec{\sigma})_l w_l^f = \sum_{m=1}^{p} G(x, x_{\pi(m)}^f)(\Pi W \vec{\sigma})_m$$

where W is a diagonal $n_f \times n_f$ matrix with $W_{ll} = w_l^f$. Since $\vec{\sigma} = B\hat{f}^K$, we let $T := \Pi W B$ to obtain the representation (44) with

$$\vec{\rho} = T\hat{f}^K. \tag{46}$$

T is a generic $p \times K$ matrix which may be formed and stored once Π, W, and B are available. We emphasize that each of these matrices is identical for all patches of a given radius ε and may therefore be precomputed. Π is obtained from a single interpolative decomposition, W is a simply a matrix of quadrature weights, and B is computed by solving a sequence of one-patch integral equations as explained in Section 5.

Using this compression scheme alone, it is straightforward to reduce the cost of computing the sums (43) from $\mathcal{O}(Kn_fN^2)$ to $\mathcal{O}(KpN^2)$. The tools introduced in the remainder of this section will allow us to reduce the cost further to $\mathcal{O}(KpN\log N)$.

7.2. Quadtree on the sphere

We now describe a data structure which will enable us to organize groups of patches in a hierarchical fashion. We first inscribe the sphere in a cube (see Fig. 6). We then project each patch center onto the surface of the cube via the ray from the origin through the patch center (indicated by the arrows in the figure). This defines a set of points on the surface of the cube. We then build a quadtree on each face of the cube, subdividing boxes until there is only one point per box, and pruning empty boxes in the process. The union of these six quadtrees is an FMM-like full tree data structure, which provides a subdivision of the sphere itself into a hierarchy of levels. The patches assigned to a particular box in the full tree will be said to form a *patch group*. Each patch is a member of one patch group at each level of the full tree. At the leaf level, each group consists of a single patch.

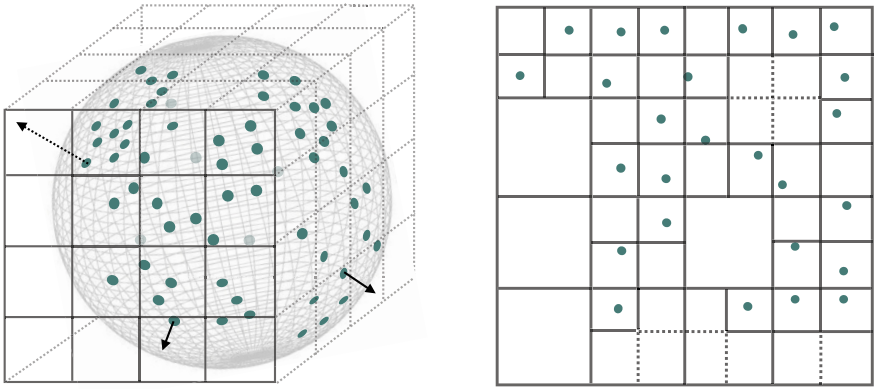

Fig. 6. The sphere is inscribed in a cube and each patch center is projected to a face of the cube by a ray emanating from the sphere center (left). An adaptive quad tree is then built on each face until, at the finest level, there is one patch in every non-empty leaf node in the quad tree (right).

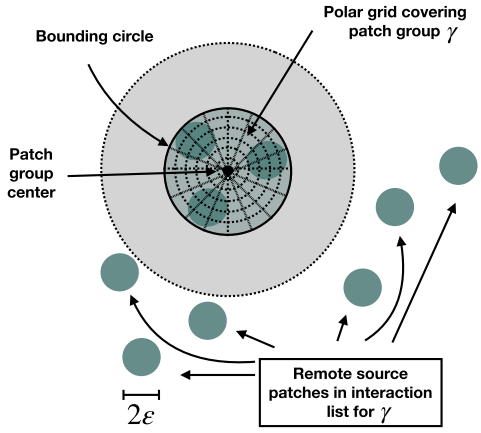

Fig. 7. For a group of m patches, the field due to well-separated source patches may be captured with high order accuracy on a polar grid which covers all m patches.

We define parent, child, and neighbor boxes in the full tree in the same way as in an ordinary quadtree. The only modification to the definition of a neighbor box is that it wraps across cube edges and corners. Thus, a box adjacent to an edge has eight neighbors (like an interior box) unless it is a corner box, in which case it has seven neighbors. Well-separatedness and the interaction list for boxes or their corresponding patch groups are defined as in the usual FMM. Two boxes at a given level are well-separated if they are not neighbors, and the interaction list for a particular box is comprised of the well-separated children of its parent's neighbors. We will sometimes refer to a patch Γ_i as being in the interaction list of some patch group γ, by which we mean that Γ_i is contained in a group which is in the interaction list of γ.

7.3. The representation of incoming fields on patch groups

Since the incoming field due to remote source patches in the interaction list of a patch group γ is smooth, it can be efficiently represented on a spectral polar grid (see Fig. 7). This requires the construction of a *bounding circle* on the surface of the sphere, enclosing all of the patches in γ, which circumscribes the grid. Incoming field values can then be obtained at arbitrary points inside the bounding circle by interpolation. We refer to the grid samples of the incoming field as an *incoming representation.*

The bounding circle is straightforward to construct using a "smallest circle algorithm" for a collection of points in the plane, suitably adapted to the sphere (see [39–41] and the references therein for discussion of the smallest circle problem).

Given a bounding circle for a patch group, we can build a local polar coordinate system (r, θ), for which $r = 0$ corresponds to the center of the patch group, and $r = R$ corresponds to the bounding circle. We must select an *incoming grid* in these coordinates which can represent a smooth incoming field in a high order manner with as few grid points as possible. For this, we will use a parity-restricted Chebyshev-Fourier basis, formed by taking products of scaled Chebyshev polynomials in the radial variable $r \in [-R, R]$ with trigonometric functions in the angular variable $\theta \in [0, 2\pi)$. The coefficients of an expansion in these basis functions corresponding to Chebyshev and Fourier modes of different parity can be shown to be zero, hence the name of the basis. This is an efficient and spectrally accurate basis with a simple associated grid

[23]. Namely, the coefficients of the expansion may be computed from function samples on a polar grid comprised of the scaled Chebyshev nodes in $r \in [0, R]$ and equispaced nodes in $\theta \in [0, 2\pi)$. The desired field may then be evaluated at any point inside a patch group's bounding circle by evaluating the resulting Chebyshev-Fourier expansion. It is straightforward to verify that the number of grid points and coefficients required to obtain an accuracy ϵ is $\mathcal{O}(\log^2(1/\epsilon))$.

8. Solution of the multiple scattering system

We now describe our method to solve the discretized many-patch system (41), including the fast algorithm for accelerating the computation of the multiple scattering interactions (43) within a GMRES iteration.

Step 1: Precomputation (for each choice of ε)

Given the patch radius ε, select the Zernike truncation parameter K and form the matrix Q.

(a) Solve the system $SB = Q$ described in Section 9.

(b) Construct the matrix T defined in Section 7.1 by building and composing the matrices Π, W, and B. Π need not be stored after T is formed.

(c) Construct the vector $I = (w_1^f, \ldots, w_{n_f}^f)B$, used to obtain the quantities J and μ in (42). At this point we no longer need to store B, only the $p \times K$ matrix T and the $1 \times K$ vector I. The storage associated with the outputs of the precomputation phase is therefore negligible.

Step 2: Construction of hierarchical data structure

Let N denote the number of patches on the surface of the sphere, assumed to satisfy the minimum patch separation condition introduced in Section 3.

(a) Form the quadtree on the sphere described in Section 7.2. The data structure should associate each patch with its group at every level, and identify the interaction list of every patch group.

(b) For each patch group, construct the incoming grid described in Section 7.3. For each patch, construct the Zernike sampling grid described in Section 4.

Step 3: Iteration

We use GMRES to solve the system (41). At each iteration, we must apply the system matrix; that is, we must compute

$$\hat{f}_i^K + P \sum_{j \neq i} S_{ij} B \hat{f}_j^K \tag{47}$$

for $i = 1, \ldots, N$, where here $(\hat{f}_1^K, \ldots, \hat{f}_N^K)^T \in \mathbb{R}^{KN}$ is the input vector at a given iteration. The following algorithm computes this expression in $\mathcal{O}(N \log N)$ operations.

1. Compute and store the outgoing coefficients $\vec{\rho}_i = T \hat{f}_i^K$ for each patch, $i = 1, \ldots, N$.
   Cost: Approximately pKN.
2. Loop through every patch group in every level. For each patch group γ, loop through all patches in its interaction list. For each such patch Γ_i, evaluate the field induced by the density on Γ_i on the incoming grid of γ, using the outgoing representation (44). Add together all such field values to obtain the total incoming field on the incoming grid.
   Cost: If q is an upper bound on the number of points in each incoming grid, the cost of evaluating a single outgoing representation on an incoming grid is at most qp. At each level, the outgoing representation corresponding to each patch must be evaluated on at most 27 incoming grids, since the interaction list of each patch's group at that level contains at most 27 other groups. There are approximately $\log_4 N$ levels. Therefore, the cost of this step is approximately $27qpN \log_4 N$.
3. At the leaf level of the tree, each patch group γ contains a single patch, say Γ_i. Though we have already evaluated the outgoing representation for Γ_i on the incoming grids of all (single-patch) groups in the interaction list of γ, we now do so also for the neighbors of γ, which are also single-patch groups but are not contained in the interaction list of γ. We add these contributions to the field values already stored on the incoming grids of these neighbor patches.
   Cost: Since each leaf-level single-patch group has at most 8 neighbors, the cost of this step is approximately $8qpN$.
   Note: For each patch Γ_i, the incoming field due to every other patch has now been stored in the incoming grid of exactly one patch-group of which Γ_i is a member. Indeed, every other patch is either a neighbor of Γ_i at the leaf level, or it is contained in exactly one of the interaction lists of the patch groups containing Γ_i.
4. Loop through each patch group. For every patch Γ_i in a group γ, evaluate the interpolant of the incoming field stored on the incoming grid of γ at the Zernike sampling nodes on Γ_i.
   Cost: There are $\mathcal{O}(K)$ Zernike sampling nodes, so the cost of each interpolation is approximately q^2 to form the interpolant and Kq to evaluate it. Each patch is a member of a single group at each level, so we must carry out approximately $N \log_4 N$ such interpolations. The total cost is therefore approximately $(q^2 + Kq)N \log_4 N$. (For large q, this step could be accelerated with fast transform methods but q is generally too small for this to provide any significant benefit.)
   At this point, we have computed the field due to all other patches on the Zernike sampling grid on each patch. That is, we have computed the sums $\sum_{j \neq i} S_{ij} B \hat{\sigma}_j$ for $i = 1, \ldots, N$.

5. Apply the matrix P to the values stored on the Zernike sampling grid on each patch and add $\hat{f}_i^K$ to the result to obtain (47).
   Cost: *Approximately* $K^2 N$.

The total cost of each iteration is therefore $\mathcal{O}(N \log N)$, with asymptotic constants which involve the parameters K, q, and p associated with the resolution of smooth functions on spectral grids. The singular character of the problem is dealt with entirely during the precomputation phase.

8.1. Optimizations and parallelization

While the algorithm described above has the desired computational complexity, there are several practical considerations that are worth discussing to optimize its performance.

Selection of incoming grid parameters: Rather than making a uniform choice of the radial and azimuthal truncation parameters for the incoming grid, we can compute these adaptively as follows. For each patch group γ, we determine the distance from its bounding circle to the nearest patch in its interaction list. We then adaptively construct an incoming grid which accurately interpolates a collection of point sources $G(x, x')$ at points x' this distance away. This adaptive interpolation is carried out by increasing the incoming grid truncation parameters until the last few Legendre-Fourier coefficients of the interpolant fall below some specified tolerance.

Additional compression of the outgoing representation: Instead of using the same outgoing coefficients $\vec{\rho}_i$ for each level of the quadtree, we can associate with each patch a different outgoing representation for each level. Recall that the far field regions Θ_i were constructed identically for each patch Γ_i to be as large as possible, consistent with the minimum patch separation. This way, one could build a single generic matrix T taking a density on a patch to its outgoing representation. T was built by compressing the outgoing field due to a generic patch Γ against a grid on a generic far field region Θ. Instead, we can build one such matrix for each level of the quadtree by constructing a generic far field region for each level. Each such far field region is an annulus or disk on the surface of the sphere. For each level, it is taken to be just large enough so that for any $i = 1, \ldots, N$, in the coordinate system of Γ_i, it covers the bounding circle of every group γ containing Γ_i in its interaction list at that level. Using the interpolative decomposition, we can then recompress the outgoing representation for a generic patch against training grids on each of the approximately $\log_4 N$ new far field regions. We obtain one matrix T per level, each of which has fewer rows and therefore yields fewer outgoing coefficients than the original.

Parallelization: Each step of the algorithm to compute (47) may be straightforwardly parallelized. Steps (1) and (5) are parallelized over all patches; steps (2) and (4) are parallelized over all patch groups at all levels; step (3) is parallelized over all patch groups at the leaf level.

9. The one-patch integral equation

In this section, we describe in detail a solver for the integral equation (29), as well as the construction of the far-field quadrature nodes $x^f_{i,1}, \ldots, x^f_{i,n_f}$ and weights $w^f_1, \ldots, w^f_{n_f}$ discussed in Section 5.

We assume that a patch Γ has radius ε and make use of cylindrical coordinates (r, θ, z). If we take the center of the patch to be the north pole of the sphere, then $r = 0$ corresponds to the z-axis, $r = 0$ and $z = \pm 1$ to the north and south poles, respectively, and $\theta = 0$ to the x-axis. Following the approach of [42,43], we use the rotational symmetry of Γ to reduce the integral equation over the patch to a sequence of one-dimensional integral equations, each corresponding to a Fourier mode in the variable θ. More precisely, we denote by C the arc which generates Γ via rotation about the z-axis: $C(t) \equiv (r(t), z(t)) = (\sin(t), \cos(t))$ for $t \in [0, \varepsilon]$. In this parametrization, t is simply the arclength along the sphere.

Let $x = (r, \theta, z)$ and $x' = (r', \theta', z')$. Since G_E and G_I are functions of $|x - x'|$ and

$$|x - x'| = \sqrt{r^2 + r'^2 + (z - z')^2 - 2rr' \cos(\theta - \theta')},$$

we can write the dependence of the Green's function in cylindrical coordinates as $G(x - x') = G(r, r', z - z', \theta - \theta')$. In these coordinates, the one-patch integral equation (29) takes the form

$$\int_0^{\varepsilon} \int_0^{2\pi} G(r(t), r'(t'), z(t) - z'(t'), \theta - \theta') \sigma(r'(t'), z'(t'), \theta') r'(t') \, dt' \, d\theta' = f(r(t), z(t), \theta).$$

Representing σ as a Fourier series in θ,

$$\sigma(r(t), z(t), \theta) = \sum_{n=-\infty}^{\infty} \sigma_n(t) e^{in\theta},$$

and taking the Fourier transform of both sides of this equation, upon rearrangement, gives the following integral equation for the Fourier modes:

$$2\pi \int_0^{\varepsilon} G_n(t,t')\sigma_n(t')\sin(t')\,dt' = f_n(t). \tag{48}$$

Here $G_n(t,t')$, $\sigma_n(t)$, and $f_n(t)$ are the Fourier transforms of $G(r(t),r'(t'),z(t)-z'(t'),\theta)$, $\sigma(r(t),z(t),\theta)$ and $f(r(t),z(t),\theta)$ with respect to θ. Thus, after solving the one-dimensional modal equations (48), we can recover $\sigma(r(t),z(t),\theta)$ from its Fourier series. Note that the Fourier series is spectrally convergent because $\sigma(r(t),z(t),\theta)$ is smooth as a function of θ, even though it is singular as a function of t at the edge $t=\varepsilon$.

9.1. Evaluation of the modal kernels

Let

$$G_n^{(1)}(t,t') = \frac{1}{\pi}\int_0^{\pi} \frac{2}{|x-x'|}\cos(n\tilde{\theta})\,d\tilde{\theta}$$

$$G_n^{(2)}(t,t') = \frac{1}{\pi}\int_0^{\pi} \log\left(\frac{2}{|x-x'|}\right)\cos(n\tilde{\theta})\,d\tilde{\theta}$$

$$G_n^{(3)}(t,t') = \frac{1}{\pi}\int_0^{\pi} \log\left(1+\frac{1}{2}|x-x'|\right)\cos(n\tilde{\theta})\,d\tilde{\theta}.$$

Then, using the formulae (10) and (15), it is straightforward to show that $G_n = G_n^{(1)} + G_n^{(2)} - G_n^{(3)}$ for $G_E(x,x')$ and $G_n = G_n^{(1)} - G_n^{(2)} - G_n^{(3)}$ for $G_I(x,x')$. We can write $|x-x'|$ in terms of t, t' and $\tilde{\theta}=\theta-\theta'$ as

$$|x-x'| = \sqrt{2\left(1-\cos(t)\cos(t')-\sin(t)\sin(t')\cos(\tilde{\theta})\right)}.$$

The integrands are not smooth at $t=t'$, $\tilde{\theta}=0$, so we must use specialized methods to evaluate each kernel.

$G_n^{(1)}(t,t')$ is simply the cosine transform of the Coulomb kernel and arises in boundary integral equations for electrostatics on axisymmetric surfaces. In [43], an efficient evaluation algorithm is described which involves writing the modal kernel in terms of Legendre functions of half-integer order and using their associated three-term recurrence. We refer the reader to this paper for further details.

The kernel $G_n^{(2)}(t,t')$ is weakly singular and may be evaluated by adaptive Gaussian quadrature. However, the following formula, discovered by a combination of analytical manipulation and symbolic calculation with Mathematica, has been numerically verified to machine precision on a dense grid of values of $t,t'\in[0,\varepsilon]$ and up to $n=1000$:

$$\frac{1}{\pi}\int_0^{\pi} \log\left(\frac{2}{|x-x'|}\right)\cos(n\tilde{\theta})\,d\tilde{\theta} = \begin{cases} -\log\left(\cos(t_1/2)\sin(t_2/2)\right) & n=0 \\ \frac{1}{2n}\left(\tan(t_1/2)\cot(t_2/2)\right)^n & n>0 \\ t_1=\min(t,t'),\ t_2=\max(t,t'). \end{cases}$$

The integrand in the expression for $G_n^{(3)}(t,t')$ is even more weakly singular, so $G_n^{(3)}(t,t')$ may be evaluated relatively quickly by adaptive Gaussian quadrature.

9.2. Discretization of the modal integral equations

Since (48) is a singular integral equation, care must be taken to discretize it accurately. The dominant singularity of the kernel $G_n(t,t')$ at $t=t'$ is the logarithmic singularity of $G_n^{(1)}(t,t')$. An analogous classical problem is therefore the first-kind integral equation arising from the solution of the Dirichlet problem on an open arc in two dimensions by a single layer potential. Stable and accurate numerical schemes for this problem can be found, for example, in [44–46]. As described in [46], when the domain is the interval $[-1,1]$, the solution of

$$\int_{-1}^{1} \log|t-s|\,\sigma(s)\,ds = f(t) \tag{49}$$

can be computed with spectral accuracy in the form $\sigma(t) = g(t)/\sqrt{(1+t)(1-t)}$, where g is a smooth function whose Chebyshev coefficients depend in a simple manner on those of f. For an open arc, the corresponding integral equation can

be preconditioned using the solution of (49). This procedure results in a Fredholm equation of the second kind for which the density may be represented as a Chebyshev expansion and computed stably with high order accuracy.

In the present context, the inclusion of the additional weakly singular kernels $G_n^{(2)}$ and $G_n^{(3)}$ cause the singularity of $\sigma_n(t)$ to be more complex, but our numerical evidence suggests that there is still a dominant square root singularity at $t = \varepsilon$. To be more precise, if we represent σ_n by

$$\sigma_n(t) = g_n(t)/\sqrt{\varepsilon - t} \tag{50}$$

near $t = \varepsilon$, we can investigate the effectiveness of representing g_n in a basis of orthogonal polynomials. While the exact behavior of $g_n(t)$ is not understood analytically, the numerical results presented in Section 9.3 suggest that it is only mildly non-smooth. We note that there is no singularity at the endpoint $t = 0$, since this point corresponds to the patch center, at which there is no physical singularity.

To resolve the endpoint singularity of σ_n, we discretize it on a set of panels $[a_0, a_1], [a_1, a_2], \ldots, [a_{m-1}, a_m]$ on $[0, \varepsilon]$ which are dyadically refined towards $t = \varepsilon$:

$$a_0 = 0,\ a_1 = \frac{\varepsilon}{2},\ a_2 = \frac{3\varepsilon}{4}, \ldots,\ a_{m-1} = \frac{(2^{m-1} - 1)\varepsilon}{2^{m-1}},\ a_m = \varepsilon.$$

On each panel, except the last, σ_n is represented as a Legendre series of fixed order k. Since σ_n is smooth on each such panel and separated from its singularity by a distance equal to the panel length, it can be shown that this representation has an error of size $\mathcal{O}(e^{-k}\log_2(1/\varepsilon))$. This argument is widely used in handling endpoint and corner singularities in the context of boundary integral equations [47–52].

On the last panel, we analytically incorporate a square root singularity into our representation of σ_n as above, and expand $g_n(t) = \sigma_n(t)\sqrt{\varepsilon - t}$ as a series of Jacobi polynomials with $\alpha = -\frac{1}{2}$ and $\beta = 0$. If the singularity of σ_n at $t = \varepsilon$ were exactly of square root type, this would yield a spectrally accurate representation of σ_n. Instead, as we will show in Section 9.3, we obtain a representation which is finite order but resolves the solution quite well even for modest truncation parameters.

Thus we have rewritten (48) as

$$f_n(t) = 2\pi \sum_{j=1}^{m-1} \int_{a_{j-1}}^{a_j} G_n(t, t')\sigma_n(t')\sin(t')\,dt' + 2\pi \int_{a_{m-1}}^{\varepsilon} \frac{G_n(t, t')}{\sqrt{\varepsilon - t'}} \left(\sigma_n(t')\sqrt{\varepsilon - t'}\right) \sin(t')\,dt'$$

and discretized σ_n by Legendre polynomials for the first $m - 1$ panels and by Jacobi polynomials for the last. Sampling the resulting equations at the corresponding quadrature nodes – Gauss-Legendre for the first $m - 1$ panels and Gauss-Jacobi for the last – yields a collocation method for σ_n, in which σ_n is determined by its piecewise polynomial basis coefficients. For each collocation node t_i, we compute the system matrix entries by adaptively integrating $G_n(t_i, t')$ in t' against the piecewise polynomial basis functions. We compute the values $f_n(t_i)$ by discretizing the Fourier transform of $f(r(t_i), z(t_i), \theta)$ in θ by the trapezoidal rule, which is spectrally accurate for smooth, periodic functions. We solve the resulting set of linear systems – one for each Fourier mode – by LU factorization and back substitution. The factorizations may be reused, since we must solve a one-patch integral equation for many different right hand sides.

We can now define the fine grid points and the smooth quadrature weights introduced in Section 5. The points $x_{i,1}^f, \ldots, x_{i,n_f}^f$ are the tensor products of the collocation nodes in the radial direction with equispaced points – the trapezoidal rule quadrature nodes – in the azimuthal direction. $w_1^f, \ldots, w_{n_f}^f$ are the corresponding quadrature weights – products of the panel-wise Gauss weights with the trapezoidal rule weight.

9.3. Numerical investigation of the singularity of σ_n

In this section, we contrast two strategies for representing σ_n in (50). In the first, we use $m = 1$ panels, and represent g_n in a basis of Jacobi polynomials, which takes into account the square root singularity in σ_n. This approach would yield spectral accuracy with respect to g_n if σ_n only contained a square root singularity. The second strategy is the one described above; we use $m > 1$ panels with a Jacobi polynomial basis of fixed degree only in the last panel. These experiments give us some insight into the nature of the true singularity in σ_n, and justify our discretization choice.

In both cases, we solve the interior one-patch integral equation by the method described above for a basis of Zernike polynomials with truncation parameter $M = 15$. The results do not change significantly if we solve the exterior equation instead. We do this for several different choices of ε. The Fourier series truncation is fixed sufficiently large to resolve the highest azimuthal Zernike mode. For each solution, we measure the residual error in L^2, normalized by the patch size:

$$\|\mathcal{S}\sigma - f\|_{L^2(\Gamma)}/|\Gamma|. \tag{51}$$

Here $|\Gamma|$ is the surface area of the patch, and f is a Zernike polynomial. This measures the extent to which the computed solution of the one-patch BVP satisfies the Dirichlet boundary condition. This solution automatically satisfies the Neumann

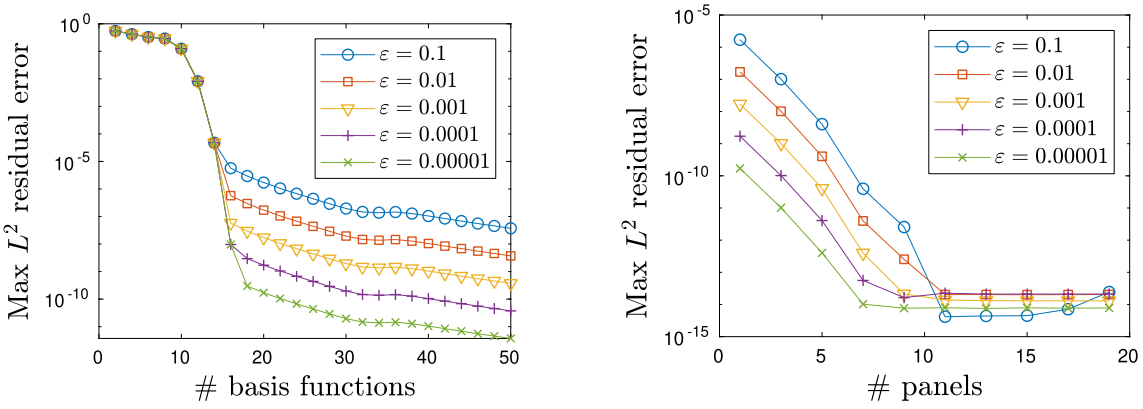

Fig. 8. Left panel: g_n is represented by a basis of Jacobi polynomials on a single panel. We plot the maximum residual error (51) vs. the number of Jacobi basis functions. Right panel: g_n is represented in a Legendre basis on every panel except the last, where a Jacobi basis is used. We plot the maximum residual error vs. the number of panels.

boundary condition and the PDE, because of its representation as a single layer potential with the Neumann Green's function, so a small L^2 residual error corresponds to a solution which nearly satisfies the boundary value problem. This error is computed by quadrature on a Legendre-Fourier grid which does not overlap with the grid on which the integral equation is solved, so it is *not* the same as the residual of the solution to the discrete linear system.

Using the first strategy ($m = 1$), we measure the error (51) for each Zernike polynomial, as the number of Jacobi basis functions is increased. The error is defined to be the maximum taken over all Zernike polynomials. The results are presented in the left panel of Fig. 8. We observe an initial regime of rapid convergence, followed by much slower convergence. Indeed, 15 basis functions are required to resolve the highest Zernike modes we have used as data. Afterward, the slow regime of convergence suggests that σ_n has a dominant square root singularity and a subdominant term which is nonsmooth, but much smaller. We also notice that performance improves as ε is decreased, which is not surprising since as $\varepsilon \to 0$, we approach the flat case in which σ_n has a pure square root singularity.

The second strategy is explored in the right panel of Fig. 8. Here, we fix 20 basis functions per panel – sufficient to begin with a good error constant, according to the first experiment. We then increase the number m of panels. Although we can already obtain quite good accuracy using the first strategy, the second allows us to reach near-machine precision. The improvement is particularly dramatic for larger choices of ε.

10. Numerical experiments

An important parameter in studying narrow escape and narrow capture problems is the *patch area fraction* $f_{N,\varepsilon}$. Since the surface area of a single patch of radius ε is given by

$$A_\varepsilon = 4\pi \sin^2(\varepsilon/2),$$

we have

$$f_{N,\varepsilon} = N \sin^2(\varepsilon/2). \tag{52}$$

Assuming ε is sufficiently small, we may write

$$f_{N,\varepsilon} \approx \varepsilon^2 N/4. \tag{53}$$

Given N, we will use (53) to compute the patch radius ε for a given patch area fraction.

10.1. Convergence with respect to the Zernike basis

We first investigate the convergence of the solution with respect to the Zernike truncation parameter M, which determines the largest radial and azimuthal Zernike modes used to represent the smooth incoming field on each patch. We fix the patch area fraction at $f_{N,\varepsilon} = 0.05$ and carry out experiments with $N = 10$, 100, and 1000 patches. ε is computed from (53). The patch locations are drawn from a uniform random distribution on the sphere, with a minimal patch separation of 2ε enforced. In each case, we solve the one-patch problems with the truncation parameter M set to $1, 3, 5, \ldots, 15$. The one-patch solutions are obtained, guided by the results in Fig. 8, using 13 panels with 20 basis functions per panel, and the number of Fourier modes set equal to the number of azimuthal modes in the Zernike basis. The ID and GMRES tolerances are set to 10^{-15}, and the incoming grid tolerance is set to 10^{-12}.

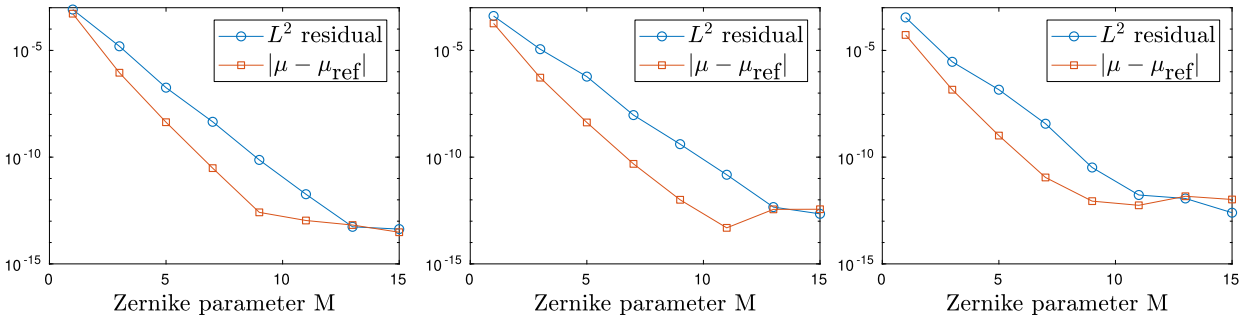

Fig. 9. L^2 residual error and self-consistent convergence error of the average MFPT μ for random patches with $f_{N,\varepsilon} = 0.05$. Left panel: $N = 10$, $\varepsilon \approx 0.141$. Middle panel: $N = 100$, $\varepsilon \approx 0.0447$. Right panel: $N = 1000$, $\varepsilon \approx 0.0141$.

We measure error in two ways. The first, as in (51), is to examine the relative L^2 residual of the multiple scattering system (28) (the discrepancy of the computed boundary values with the Dirichlet data) on a random patch Γ_i:

$$\frac{1}{|\Gamma_i|}\left\| \left(\mathcal{S}\sigma_i + \sum_{j\neq i}^{N} \mathcal{S}_{ij}\sigma_j \right) - 1 \right\|_{L^2(\Gamma_i)} . \tag{54}$$

The second is to examine the difference between the computed average mean first passage time (MFPT) μ and a reference value, denoted by μ_{ref}. We obtain μ_{ref} by carrying out a more refined simulation, with $M = 17$ on each patch, while also increasing the number of panels and basis functions used to solve the one-patch problem to 19 and 30, respectively, and doubling the numbers of both radial and azimuthal modes used in the incoming grids of all patch groups. This is a self-consistent convergence test for μ.

The results are presented in Fig. 9. In all cases, we observe the expected spectral convergence with respect to M, and can reach errors of approximately 10^{-12} or less. We also find that the residual error appears to provide a good upper bound on the error of μ until convergence is reached.

10.2. Large scale simulations

We next study the performance of our solver as N is increased and ε is decreased. The error is measured by computing the L^2 residual (54) on a random patch. The parameters for the one-patch solver are set as in the previous section with $M = 15$, but we fix the ID tolerance at 10^{-11}, the GMRES tolerance at 10^{-10}, and the incoming grid truncation tolerance at 10^{-8}. This selection of parameters yields errors in range $10^{-7} - 10^{-10}$ for all of our experiments. Our calculations are performed on either a laptop with a 4-core Intel i7-3630QM 2.40 GHz processor or a workstation with four Intel Xeon E7-4880 2.50 GHz processors. each of which has 15 cores. The algorithm has been implemented in Fortran, and in both cases, the hierarchical fast algorithm is parallelized over all available cores using OpenMP.

We consider randomly located patches, uniformly located patches and patches that are highly clustered. For each experiment we report N, ε, the computed value of the average MFPT μ, truncated at 8 significant digits, the L^2 residual error on a random patch, the total number of GMRES iterations, the total solve time, and the time per GMRES iteration. We also compute the *parallel scaling factor* – namely, the ratio of the time to compute the matrix-vector product (47) using a single core to the time required using all cores on the 60-core workstation.

10.2.1. Example 1: random patches with area fraction $f_{N,\varepsilon} = 0.05$

Fixing the patch area fraction at $f_{N,\varepsilon} = 0.05$, we let ε be given by (53) for $N = 10, 100, 1000, 10\,000, 100\,000$, with patches randomly distributed on the sphere with a minimum patch separation of 2ε. The corresponding results are given in Table 1. In the left panel of Fig. 10, we plot the time per GMRES iteration as a function of N using the 4-core laptop and the 60-core workstation, as well as a reference curve with $\mathcal{O}(N \log N)$ scaling. In Fig. 11, we also plot the computed MFPT $\bar{v}$ just inside the unit sphere – on a sphere of radius $1 - \varepsilon/5$ – for $N = 10, 100, 1000, 10\,000$. The case $N = 100\,000$ case was plotted earlier, in Fig. 2.

Note that the number of GMRES iterations increases with N, as one would expect from the increased complexity of the problem, but slowly. The computation with $N = 100\,000$ required just over an hour to complete using the 60-core workstation. The computation with $N = 10\,000$ required just over 45 minutes to solve on the 4-core laptop, and the computation with $N = 1000$ required approximately one minute. (The case $N = 100\,000$ was not attempted on the laptop because of memory requirements.) Note from the data in Table 1 that we achieve approximately 85% parallel efficiency at $N = 1000$ and an efficiency near 90% for the largest calculation. Note also from Fig. 10 that the complexity of the fast algorithm is consistent with the expected $O(N \log N)$ scaling.

Table 1
Narrow escape problem with random patches at patch area fraction $f_{N,\varepsilon} = 0.05$.

N	10	100	1000	10 000	100 000
ε	≈ 0.14	≈ 0.045	≈ 0.014	≈ 0.0045	≈ 0.0014
Average MFPT μ	0.64277353	0.24999828	0.12308716	0.084405945	0.072275200
L^2 residual error	3.6×10^{-9}	1.6×10^{-9}	5.3×10^{-9}	4.8×10^{-8}	2.2×10^{-8}
# GMRES iterations	7	12	17	25	35
Total iteration time (s) (60 cores)	0.11	0.54	8.9	215	3793
Time per iteration (s) (60 cores)	0.02	0.05	0.5	8.6	108
Total iteration time (s) (laptop)	0.10	2.63	68.9	1731	
Time per iteration (s) (laptop)	0.01	0.22	4.1	69	
Parallel scaling factor (60 cores)	2.1	25.7	51.4	52.3	53.5

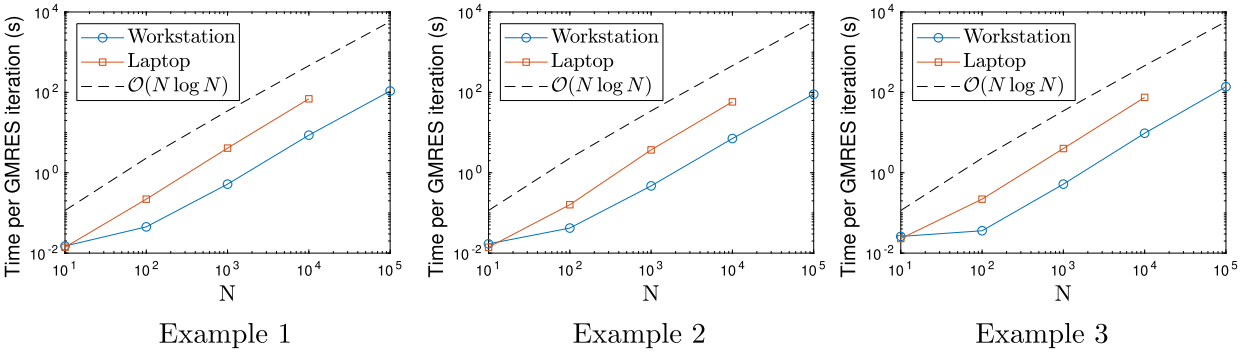

Fig. 10. Time per GMRES iteration for the 4-core laptop and 60-core workstation. A reference curve with $\mathcal{O}(N \log N)$ scaling is also plotted.

Table 2
Narrow escape problem with uniform patches at patch area fraction $f_{N,\varepsilon} = 0.05$.

N	10	100	1000	10 000	100 000
ε	≈ 0.14	≈ 0.045	≈ 0.014	≈ 0.0045	≈ 0.0014
Average MFPT μ	0.62771752	0.23201408	0.11813387	0.082870386	0.071784189
L^2 residual error	3.0×10^{-9}	1.5×10^{-9}	3.2×10^{-8}	6.4×10^{-8}	8.4×10^{-8}
# GMRES iterations	6	9	11	16	20
Total iteration time (s) (60 cores)	0.10	0.38	5.1	114	1803
Time per iteration (s) (60 cores)	0.02	0.04	0.47	7.1	90
Total iteration time (s) (laptop)	0.087	1.45	40.7	926	
Time per iteration (s) (laptop)	0.014	0.16	3.7	58	
Parallel scaling factor (60 cores)	5.0	29.9	53.7	54.0	54.8

10.2.2. Example 2: uniform patches with area fraction $f_{N,\varepsilon} = 0.05$

Using the same patch area fraction as in the previous example, we let N take the same values, but place the patch centers at the Fibonacci spiral points, which are approximately uniform on the sphere [10]. Results are shown in Table 2 and the middle panel of Fig. 10. The computed MFPT $\bar{v}$ on the sphere of radius $1 - \varepsilon/5$ was plotted in Fig. 3 for the case $N = 10\,000$. The MFPT is plotted for the $N = 100$ and $N = 1000$ cases in Fig. 11.

10.2.3. Example 3: clustered patches

In our final example, we configure the patches to form a collection of 20 clusters. Each cluster is contained within a disk on the surface of the sphere centered at the vertices of a dodecahedron inscribed in the sphere, and the radii of the disks are chosen so that all 20 disks cover one quarter of the area of the sphere. Patch centers are placed randomly on the sphere, and a proposed center is accepted if it falls within one of the disks, while enforcing a minimum patch separation distance of 2ε. We choose ε empirically to be as large as possible so that our random placement process yields the desired number N of patches in a reasonable amount of time. For sufficiently large N, this results in a much denser packing of patches within each cluster than we had in our previous examples.

The results of our simulations are provided in Table 3 and the right panel of Fig. 10. The MFPT is plotted on a sphere of radius $1 - \varepsilon/5$ in Fig. 4 for the $N = 10\,000$ case and in Fig. 11 for the $N = 100$ and $N = 1000$ cases. The denser packing of patches leads to a greater number of GMRES iterations than in the previous examples and longer computation times, but the difference is mild. The case with $N = 100\,000$ required just over an hour and a half to solve on our 60-core workstation. The simulation with $N = 10\,000$ required 75 minutes on a laptop, and the simulation with $N = 1000$ required about one minute.

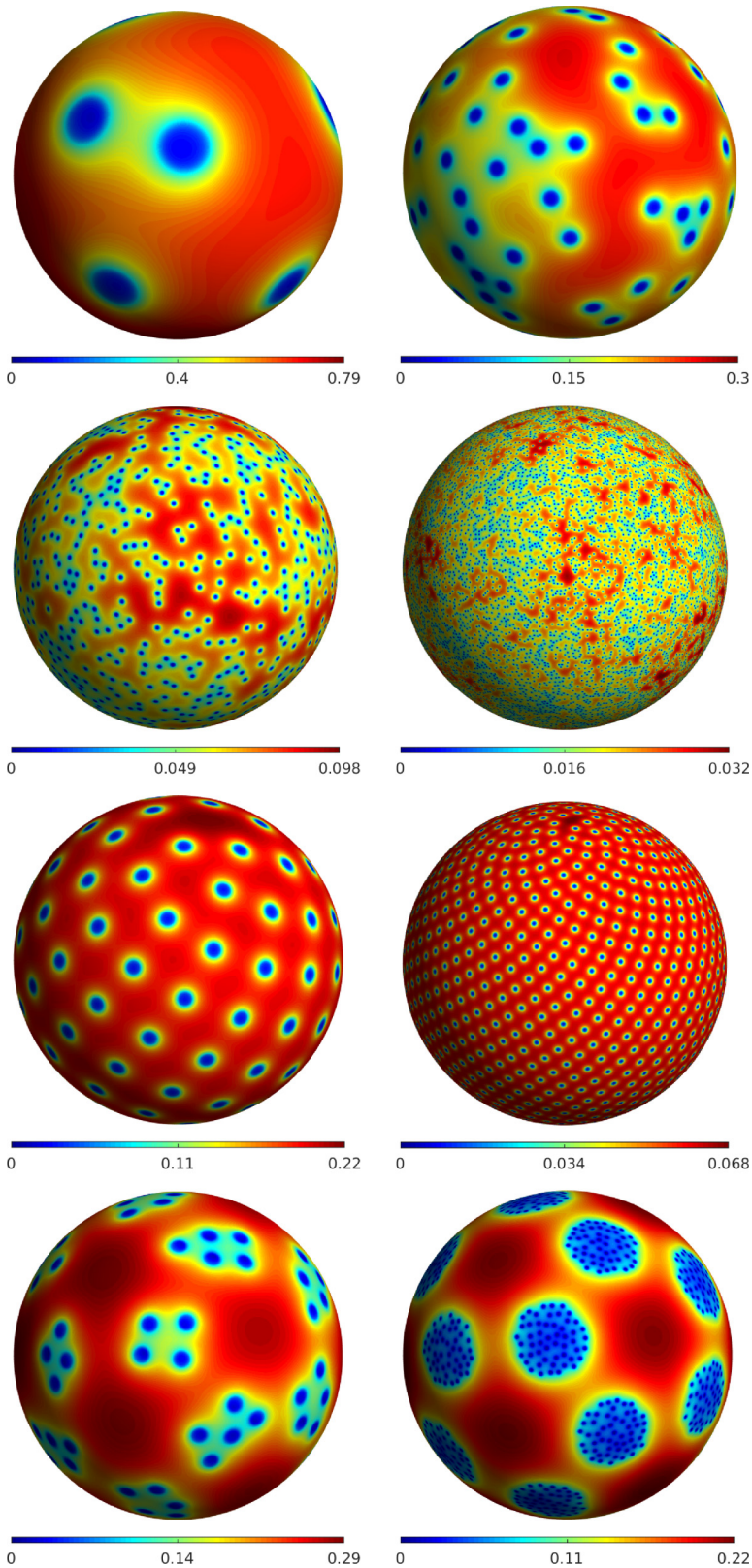

Fig. 11. Plots of the MFPT $\bar{v}$ on a sphere of radius $1-\varepsilon/5$ for the experiments described in Section 10.2. The first two rows correspond to Example 1 with $N = 10, 100, 1000, 10\,000$. The third row corresponds to Example 2 with $N = 100, 1000$. The final row corresponds to Example 3 with $N = 100, 1000$.

Table 3
Narrow escape problem with clustered patches.

N	10	100	1000	10 000	100 000
ε	0.25	0.047	0.012	0.0035	0.001
Average MFPT μ	0.29687267	0.25519357	0.20318506	0.17622000	0.16531162
L^2 residual error	4.9×10^{-10}	3.9×10^{-9}	1.2×10^{-8}	6.5×10^{-8}	1.2×10^{-7}
# GMRES iterations	8	12	19	28	42
Total iteration time (s) (60 cores)	0.21	0.43	9.9	269	5795
Time per iteration (s) (60 cores)	0.03	0.04	0.52	9.6	138
Total iteration time (s) (laptop)	0.18	2.7	76.4	2112	
Time per iteration (s) (laptop)	0.02	0.22	4.0	75	
Parallel scaling factor (60 cores)	2.9	43.9	49.3	51.4	55.5

Remark 7. We carried out the simulations above for the corresponding exterior problem as well (the narrow capture problem). As expected (since the integral equations are nearly identical), the timings and errors are similar and are therefore omitted.

11. Conclusions

We have developed a fast solver for the narrow capture and narrow escape problems on the sphere with arbitrarily-distributed well-separated disk-shaped patches. We solve the corresponding mixed boundary value problems by an integral equation scheme derived using the Neumann Green's functions for the sphere. Our numerical method combines a high order accurate solver for the one-patch problem, a multiple scattering formalism, and a hierarchical fast algorithm. We have demonstrated the scheme on examples with N as large as 100 000, significantly larger than previously accessible. The ability to carry out such large-scale simulations will permit a systematic study of the asymptotic approaches described, for example, in [11] and [12].

Possible extensions of our method include the consideration of narrow escape and narrow capture problems when the patches are asymmetric and have multiple shapes. Assuming some separation between patches, the multiple scattering formalism still applies, but the single patch integral equation will not be solvable by separation of variables and the compressed representation of outgoing fields will need to be computed for each distinct patch type. Neither of these extra steps, however, affects the asymptotic $\mathcal{O}(N \log N)$ scaling of the fast algorithm. Exterior problems involving multiple spheres with different arrangements of patches could also be simulated by a simple modification of our multiple scattering approach.

A more challenging problem is to extend our method to non-spherical geometries. For this, one would either have to discretize the entire domain surface, rather than just the absorbing patches, or construct the Neumann Green's function for such a domain numerically. In the latter case, aspects of our multiple scattering approach would carry over. We are currently investigating these issues and will report on our progress at a later date.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Michael Ward for suggesting this problem and for several valuable insights. We would also like to thank Mike O'Neil for many useful conversations. J.K. was supported in part by the Research Training Group in Modeling and Simulation funded by the National Science Foundation via grant RTG/DMS-1646339.

References

[1] H.C. Berg, E.M. Purcell, Physics of chemoreception, Biophys. J. 20 (1977) 193–219.
[2] F.W. Wiegel, Diffusion and the physics of chemoreception, Phys. Rep. 95 (5) (1983) 283–319.
[3] P.C. Bressloff, J.M. Newby, Stochastic models of intracellular transport, Rev. Mod. Phys. 85 (1) (2013) 135.
[4] D. Holcman, Z. Schuss, Stochastic Narrow Escape in Molecular and Cellular Biology: Analysis and Applications, Springer, New York, 2015.
[5] D. Holcman, Z. Schuss, The narrow escape problem, SIAM Rev. 56 (2) (2014) 213–257.
[6] K. Kaizu, W. De Ronde, J. Paijmans, K. Takahashi, F. Tostevin, P.R. Ten Wolde, The Berg-Purcell limit revisited, Biophys. J. 106 (4) (2014) 976–985.
[7] S. Isaacson, D. McQueen, C.S. Peskin, The influence of volume exclusion by chromatin on the time required to find specific DNA binding sites by diffusion, Proc. Natl. Acad. Sci. USA 108 (9) (2011) 3815–3820.
[8] S. Redner, A Guide to First-Passage Processes, Cambridge University Press, 2001.
[9] G. Pavliotis, Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations, Springer, 2014.
[10] A.J. Bernoff, A.E. Lindsay, Numerical approximation of diffusive capture rates by planar and spherical surfaces with absorbing pores, SIAM J. Appl. Math. 78 (1) (2018) 266–290.
[11] A.F. Cheviakov, M.J. Ward, R. Straube, An asymptotic analysis of the mean first passage time for narrow escape problems: part II: the sphere, Multiscale Model. Simul. 8 (3) (2010) 836–870.

[12] A.E. Lindsay, A.J. Bernoff, M.J. Ward, First passage statistics for the capture of a Brownian particle by a structured spherical target with multiple surface traps, Multiscale Model. Simul. 15 (1) (2017) 74–109.
[13] A. Singer, Z. Schuss, D. Holcman, R.S. Eisenberg, Narrow escape, part I, J. Stat. Phys. 122 (3) (2006) 437–463.
[14] I.N. Sneddon, Mixed Boundary Value Problems in Potential Theory, North-Holland, Amsterdam, 1966.
[15] V.I. Fabrikant, Applications of Potential Theory in Mechanics: A Selection of New Results, Kluwer Academic Publishers, Dordrecht, Boston, 1989.
[16] Y. Saad, M.H. Schultz, GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems, SIAM J. Sci. Stat. Comput. 7 (3) (1986) 856–869.
[17] E. Liberty, F. Woolfe, P.-G. Martinsson, V. Rokhlin, M. Tygert, Randomized algorithms for the low-rank approximation of matrices, Proc. Natl. Acad. Sci. USA 104 (51) (2007) 20167–20172.
[18] N. Koshlyakov, M.M. Smirnov, E.B. Gliner, Differential Equations of Mathematical Physics, North-Holland Publishing Company, 1964.
[19] O. Kellogg, Foundations of Potential Theory, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Dover Publications, 1953.
[20] E.P. Stephan, Boundary integral equation for screen problems in R^3, Integral Equ. Oper. Theory 10 (1987) 236–257.
[21] J. Helsing, G. Peters, Integral equation methods and numerical solutions of crack and inclusion problems in planar elastostatics, Abstr. Appl. Anal. (2013) 938167, 20.
[22] Q.T. Trinh, S. Mouhoubi, C. Chazallon, M. Bonnet, Solving multizone and multicrack elastostatic problems: a fast multipole symmetric Galerkin boundary element method approach, Eng. Anal. Bound. Elem. 50 (2015) 486–495.
[23] J.P. Boyd, F. Yu, Comparing seven spectral methods for interpolation and for solving the Poisson equation in a disk, J. Comput. Phys. 230 (Feb. 2011) 1408–1438.
[24] M. Costabel, M. Dauge, R. Duduchava, Asymptotics without logarithmic terms for crack problems, Commun. Partial Differ. Equ. 28 (2003) 869–926.
[25] H. Cheng, Z. Gimbutas, P.-G. Martinsson, V. Rokhlin, On the compression of low rank matrices, SIAM J. Sci. Comput. 26 (4) (2005) 1389–1404.
[26] S. Goreinov, E. Tyrtyshnikov, N. Zamarashkin, A theory of pseudoskeleton approximations, Linear Algebra Appl. 261 (1997) 1–21.
[27] M. Gu, S.C. Eisenstat, Efficient algorithms for computing a strong rank-revealing QR factorization, SIAM J. Sci. Comput. 17 (4) (1996) 848–869.
[28] P.-G. Martinsson, V. Rokhlin, M. Tygert, A randomized algorithm for the decomposition of matrices, Appl. Comput. Harmon. Anal. 30 (1) (2011) 47–68.
[29] S. Ambikasaran, E.F. Darve, An $O(N \log N)$ fast direct solver for partial hierarchically semi-separable matrices, J. Sci. Comput. 57 (3) (2013) 477–501.
[30] S. Börm, L. Grasedyck, W. Hackbusch, Hierarchical matrices, Lecture Notes 21 (2003).
[31] E. Corona, P.-G. Martinsson, D. Zorin, An $O(N)$ direct solver for integral equations on the plane, Appl. Comput. Harmon. Anal. (2014).
[32] W. Fong, E. Darve, The black-box fast multipole method, J. Comput. Phys. 228 (23) (2009) 8712–8725.
[33] A. Gillman, P.M. Young, P.-G. Martinsson, A direct solver with $O(N)$ complexity for integral equations on one-dimensional domains, Front. Math. China 7 (2) (2012) 217–247.
[34] Z. Gimbutas, V. Rokhlin, A generalized fast multipole method for nonoscillatory kernels, SIAM J. Sci. Comput. 24 (3) (2002) 796–817.
[35] K.L. Ho, L. Greengard, A fast direct solver for structured linear systems by recursive skeletonization, SIAM J. Sci. Comput. 34 (2011) A2507–A2532, 10.
[36] P.-G. Martinsson, V. Rokhlin, A fast direct solver for boundary integral equations in two dimensions, J. Comput. Phys. 205 (1) (2005) 1–23.
[37] V. Minden, K.L. Ho, A. Damle, L. Ying, A recursive skeletonization factorization based on strong admissibility, Multiscale Model. Simul. 15 (2016) 768–796.
[38] L. Ying, G. Biros, D. Zorin, A kernel-independent adaptive fast multipole algorithm in two and three dimensions, J. Comput. Phys. 196 (2) (2004) 591–626.
[39] S. Skyum, A simple algorithm for computing the smallest enclosing circle, Inf. Process. Lett. 37 (3) (1991) 121–125.
[40] E. Welzl, Smallest enclosing disks (balls and ellipsoids), in: Results and New Trends in Computer Science, Springer-Verlag, 1991, pp. 359–370.
[41] S. Xu, R.M. Freund, J. Sun, Solution methodologies for the smallest enclosing circle problem, Comput. Optim. Appl. 25 (1–3) (2003) 283–292.
[42] P. Young, S. Hao, P.-G. Martinsson, A high-order Nyström discretization scheme for boundary integral equations defined on rotationally symmetric surfaces, J. Comput. Phys. 231 (June 2012) 4142–4159.
[43] J. Helsing, A. Karlsson, An explicit kernel-split panel-based Nyström scheme for integral equations on axially symmetric surfaces, J. Comput. Phys. 272 (Supplement C) (2014) 686–703.
[44] Y. Yan, I.H. Sloan, On integral equations of the first kind with logarithmic kernels, J. Integral Equ. Appl. 1 (4) (1988) 549–580.
[45] K.E. Atkinson, I.H. Sloan, The numerical solution of first-kind logarithmic-kernel integral equations on smooth open arcs, Math. Comput. 56 (193) (1991) 119–139.
[46] S. Jiang, V. Rokhlin, Second kind integral equations for the classical potential theory on open surfaces II, J. Comput. Phys. 195 (1) (2004) 1–16.
[47] J. Bremer, On the Nyström discretization of integral equations on planar curves with corners, Appl. Comput. Harmon. Anal. 32 (1) (2012) 45–64.
[48] J. Bremer, V. Rokhlin, Efficient discretization of Laplace boundary integral equations on polygonal domains, J. Comput. Phys. 229 (7) (2010) 2507–2525.
[49] J. Helsing, Solving integral equations on piecewise smooth boundaries using the RCIP method: a tutorial, Abstr. Appl. Anal. (2013) 938167, 20.
[50] J. Helsing, R. Ojala, Corner singularities for elliptic problems: integral equations, graded meshes, quadrature, and compressed inverse preconditioning, J. Comput. Phys. 227 (20) (2008) 8820–8840.
[51] K. Serkh, V. Rokhlin, On the solution of elliptic partial differential equations on regions with corners, J. Comput. Phys. 305 (2016) 150–171.
[52] L. Trefethen, Numerical computation of the Schwarz-Christoffel transformation, SIAM J. Sci. Stat. Comput. 1 (1980) 82–102.