Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 216 (2006) 300–325
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

A wideband fast multipole method for the Helmholtz equation in three dimensions

Hongwei Cheng [a], William Y. Crutchfield [a], Zydrunas Gimbutas [a], Leslie F. Greengard [c], J. Frank Ethridge [a], Jingfang Huang [d], Vladimir Rokhlin [b], Norman Yarvin [a,*], Junsheng Zhao [a]

[a] *Plain Sight Systems, 1020 Sherman Avenue, Hamden, CT 06514, United States*
[b] *Department of Computer Science, Yale University, New Haven, CT 06520, United States*
[c] *Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, United States*
[d] *Department of Mathematics, University of North Carolina, Chapel Hill, CB 3250 Phillips Hall, Chapel Hill, NC 27599, United States*

Received 12 August 2005; accepted 2 December 2005
Available online 18 January 2006

Abstract

We describe a wideband version of the Fast Multipole Method for the Helmholtz equation in three dimensions. It unifies previously existing versions of the FMM for high and low frequencies into an algorithm which is accurate and efficient for any frequency, having a CPU time of $O(N)$ if low-frequency computations dominate, or $O(N \log N)$ if high-frequency computations dominate. The performance of the algorithm is illustrated with numerical examples.

MSC: 65R99; 78A45

Keywords: Helmholtz equation; Fast multipole method; Scattering problems

1. Introduction

One of the standard approaches to the solution of scattering problems is to convert them into integral equations, which are then discretized using appropriate quadrature formulae. This usually leads to large-scale systems of linear algebraic equations, which are in turn solved via appropriately chosen iterative schemes (such as GMRES). Most iterative schemes for the solution of linear systems of this type require the application of the matrix of the system to a sequence of recursively generated vectors. Applying a dense matrix to a vector is an order N^2 procedure, where N is the dimensionality of the matrix, which in this case is equal to the number of nodes in the discretization of the domain of the integral equation (or to a small multiple of that number of

* Corresponding author.
E-mail address: norman.yarvin@snet.net (N. Yarvin).

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2005.12.001

nodes). As a result, the whole process is at least of order N^2, which is prohibitive for many large-scale problems.

During the last 30 years or so, a number of algorithms have been constructed for the rapid application to arbitrary vectors of the matrices resulting from the discretization of integral equations of scattering theory. Historically, the first group of algorithms for this purpose were the so-called k-space methods, which take advantage of the fact that the free-space Green's function for the Helmholtz equation is translation invariant, and use Fast Fourier Transforms (see, for example, [6]). This approach can be quite efficient when applied to volume integrals, but usually requires order $N^{3/2} \cdot \log N$ operations when used for the solution of boundary integral equations in three dimensions. Its performance deteriorates in environments where some parts of the domain of the integral equation must be discretized at a higher resolution than other parts.

Another class of techniques is known as Fast Multipole Methods (FMMs). Algorithms in this class construct a hierarchical subdivision of the domain of the integral equation, in which it is recursively divided into smaller and smaller regions; for each region in the hierarchy, they use a "far-field expansion" (in the original FMM for the Laplace equation, a multipole expansion; for low-frequency scattering problems, a partial-wave expansion) to represent the potential of that region, on regions distant from it. Since these expansions are used only on distant regions, arbitrary specified precision can be achieved (though the cost of the calculation grows somewhat when the required accuracy is increased). The CPU time requirements of the FMMs are proportional to N in the low-frequency regime, and to $N \cdot \log N$ for high-frequency problems involving boundary integrals; the latter estimate becomes $O(N)$ for volume integral equations. Fast Multipole Methods can easily be designed to be "adaptive", that is, to subdivide the domain of the integral equation more finely in regions where the discretization contains more nodes; the performance of the FMM is only weakly affected (and often improved) in such environments.

While the algorithms known by the term "FMM" share the same basic computational structure, the expansions they use are based on either of two very different principles. For the Laplace equation, or for the Helmholtz equation in the low-frequency regime, the principle used is that large submatrices of the matrix to be applied are of low rank (to high but finite precision); this permits a wide variety of expansions to be used successfully in these environments, as in, for instance, "fast" methods based on wavelets and similar structures (see, for example, [3,2,8]).

In the high-frequency regime, the ranks of submatrices tend to be proportional to their sizes, and rank considerations cannot be used to construct asymptotically fast algorithms. In this environment, existing fast schemes are based on a somewhat more subtle mathematical apparatus. Specifically, it turns out that the diagonal forms for the translation operators for the Helmholtz equation (see Section 4) are available analytically, and that on every level of subdivision *all translation operators are diagonalized by the same unitary operator* (see [27]). This observation leads to algorithms of order $N \cdot \log N$ for the application of discretized operators of scattering theory to arbitrary vectors, but not to algorithms of order N.

Each of these two types of expansions fails in some way outside its preferred regime: attempts to use rank-based approaches in the high-frequency regime result in algorithms whose CPU time requirements are proportional to N^2; attempts to use diagonal forms of translation operators at low frequencies result in numerically unstable schemes. In each case, the difficulty is fundamental, and cannot be removed by simple expedients such as scaling, etc. Thus, there exist problems (e.g. scattering from an object many wavelengths in size which has significant subwavelength structure) for which neither of the two approaches performs well.

Fortunately, it is possible to construct "hybrid" schemes, conducting all calculations via partial-wave expansions on the subwavelength levels of subdivision, and transitioning to the diagonal form once the groups are of such size that the diagonal forms are stable. In this paper, we present such a procedure. More specifically, we describe an algorithm for the rapid evaluation of expressions of the form

$$\Phi_k = \sum_{j=1,\ j\neq k}^{N} s_j \cdot h_0(\omega \cdot \|x_k - x_j\|), \tag{1}$$

or minor variants thereof, in which $\Phi_1, \ldots, \Phi_N \in \mathbb{C}$ are to be calculated, given $x_1, \ldots, x_N \in \mathbb{R}^3$, $s_1, \ldots, s_N \in \mathbb{C}$ and $\omega \in \mathbb{C}$, and where h_0 denotes the spherical Hankel function of order zero ($h_0(z) = e^{iz}/iz$). In our description, we use the customary terminology of electromagnetics, referring to the formula

$$\Phi(x) = h_0(\omega \cdot \|x - x_0\|) \tag{2}$$

as the potential at the point x of a unit source of wavenumber ω located at the point x_0. The potentials satisfy the law of superposition, so that, for instance, we refer to (1) as the potential at the points $x_1, \ldots, x_N$ which is generated by sources $s_1, \ldots, s_N$ located at those same respective points, all the sources being of the same wavenumber ω.

The algorithm presented here is a variant of the Fast Multipole Method (FMM); it is accurate and efficient for any value of ω whose real and imaginary parts are greater than zero, and we thus term it "wideband". The basic computational structure of the wideband FMM is the same as that of previous variants of the FMM in three dimensions, and can be summarized as follows (for more detail, see Section 6). The algorithm finds the smallest cubical box which encloses all the points $\{x_j\}$, and then constructs a hierarchical subdivision of that box, in which it is divided into eight boxes of equal size (also cubes), each of which is likewise subdivided, the subdivision process continuing recursively until the lowest-level boxes have O(1) points in them. The eight boxes into which a box is divided are termed its "children", and it their "parent". For each box, a far-field expansion is produced, which represents the potential due to the sources on that box, at distances more than its own length away from it. The far-field expansion for each childless (i.e. lowest-level) box is calculated from the sources on that box; the far field expansion for each parent box is calculated from the far-field expansions of its children. These far-field expansions are not evaluated directly (except in rare instances, having to do with the adaptive aspect of the algorithm), but rather translated into "local expansions", which represent the potential inside a box due to sources distant from that box. Local expansions on parent boxes are evaluated by translating them into local expansions on their children; then the local expansion on each childless box is evaluated on the points $\{x_j\}$ in that box.

Fig. 1 illustrates the various expansions and methods for the translation and conversion of expansions used by the algorithm of this paper. At low frequencies, the principal expansions used are partial-wave expansions

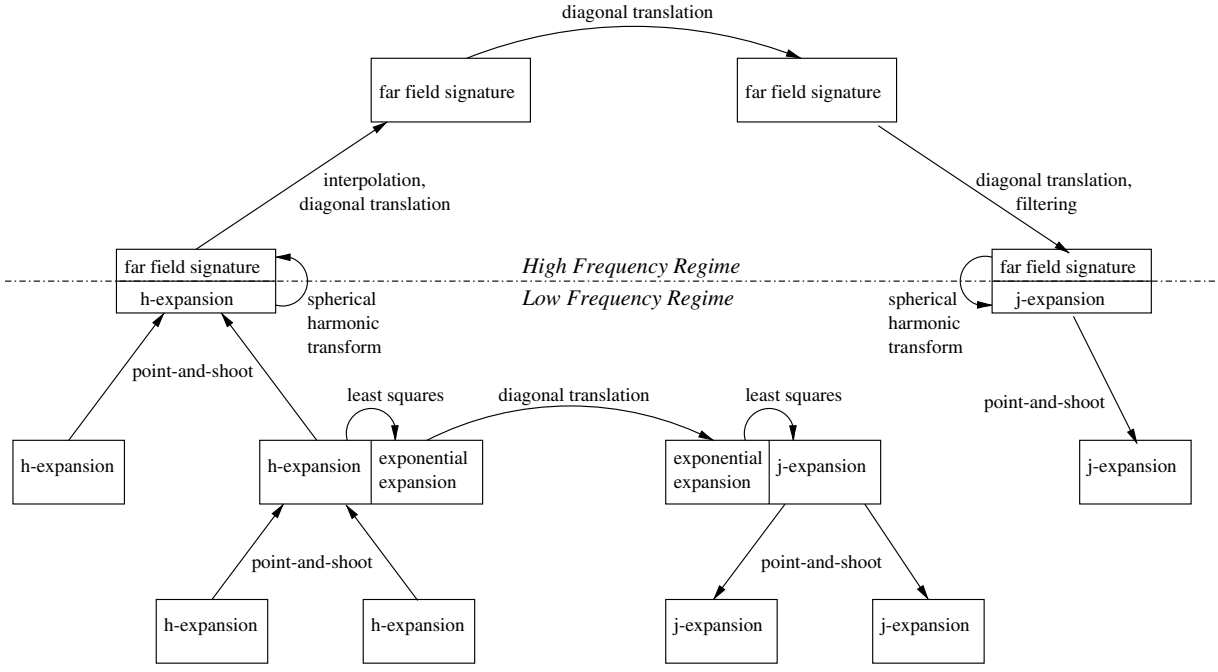

Fig. 1. A small but representative fraction of the boxes used in a typical computation of the wideband FMM. Boxes are placed in the figure according to their positions in the hierarchy, with parents on top of children. Inside each box is the type of expansion used for that box; on the left half of the figure, only far-field expansions are listed, while on the right half, only local expansions are listed (although in the full computation, all boxes have both types of expansions). Where a box is divided into two, that indicates that two different types of expansions are used for that box. An arrow indicates a transformation performed between two expansions, and the text alongside it indicates the method used for that transformation. Exponential expansions are not fully represented in this figure: in the full computation, each box in the low-frequency regime has them; and not just one but six exponential expansions are used for each box.

(Section 3), far-field expansions of this type being referred to as "h-expansions" and local expansions as "j-expansions". Translations of these expansions up and down in the hierarchy (for h-expansions, from child to parent; for j-expansions, from parent to child) are done using the "point-and-shoot" method (Section 3.3), in which the expansion is first rotated so that its z-axis points in the direction in which it is to be translated, then translated, then rotated back. If the order of the expansion is denoted by p (the number of terms being p^2), then point-and-shoot is an $O(p^3)$ operation, whereas performing the translation via a single matrix multiplication would be $O(p^4)$. Translations from h-expansions to j-expansions are performed using exponential expansions (Section 5). This does not affect the asymptotic CPU time of the algorithm, but improves the constant substantially: conversion of far-field expansions into local expansions consumes most of the CPU time of the low-frequency portion of the algorithm (since each far-field expansion might be translated to as many as $6^3 - 3^3 = 189$ local expansions, that being the maximum possible number of boxes on that level which are not adjacent to the source box, but whose parents are adjacent to its parents); since translation operators for exponential expansions are diagonal each of them can be applied in $O(p^2)$ time.

It has been empirically determined that the CPU time requirements of the scheme are minimized when high-frequency techniques are used wherever possible, which is whenever the box size exceeds a certain threshold (which depends on the desired accuracy of the computation, but is on the order of a wavelength). The high-frequency techniques are reviewed in Section 4; there, the far-field expansion for a box (its "far-field signature") consists of the potential due to sources on that box, sampled on the surface of a sphere centered on the box, appropriately scaled, and taken in the limit as the radius of the sphere approaches infinity. The analytical machinery treats far-field signatures as functions on the sphere; translation is performed by multiplying far-field signatures by other functions on the sphere. Numerically, far-field signatures must also be resampled when they are translated; for downward translation, this amounts to filtering, and for upward translation, interpolation. Both interpolation and filtering are performed via a version of the algorithm introduced in [22] and summarized in Section 2.4.

The history of "fast" techniques of the type used in this paper goes back about 20 years; an excellent review can be found in [25]. An FMM for the Laplace equation in two dimensions was published in [18]. It has been understood for a long time that trivial modifications convert the scheme of [18] into a viable FMM for the low-frequency Helmholtz equation; for algorithms of this type see [25] and references therein. In [21], the use of "intermediate" exponential expansions to accelerate the 2D Laplace FMM was proposed, and in [19], a three-dimensional version of that accelerated FMM was introduced. In [20], exponential expansions were worked out for the Helmholtz equation in three dimensions.

A high-frequency FMM for the Helmholtz equation in two dimensions was published in [26]; diagonal forms for translation operators in three dimensions are described in [27], and a single-level FMM in three dimensions is described in [9]. In [16], the complete theory of such diagonal forms is developed, and full-scale FMMs for the Helmholtz and Maxwell's equations are constructed in [28,33,12,13,15].

In [5], an algorithm of a somewhat different type for the compression of matrices resulting from the discretization of integral equations of scattering theory is presented, both in the Helmholtz and Maxwell environments; the scheme is based on the observation that the free space Green's functions for these equations are convolutions, and uses FFTs. An interesting development of this approach (under certain conditions, leading to remarkably efficient schemes) can be found in [32].

A detailed exposition of both (low-frequency and high-frequency) FMMs can be found in [7]. An outline of the wide-band FMM can also be found in [7]; a version of it based entirely on exponential expansions of the type described in Section 5 was published in [14]. A scheme somewhat related to the one described here can be found in [23].

The structure of this paper is as follows. Section 2 contains the mathematical facts to be used in the paper that are either well-known or trivially follow from well-known facts. In Section 3, we discuss various properties of partial-wave expansions that are relevant to the algorithm to be constructed, and introduce the concept of translation operators for partial-wave expansions. Diagonal forms of translation operators in the high-frequency regime are studied in Section 4, and Section 5 is devoted to the diagonal forms of translation operators in the low-frequency regime (obtained via the so-called exponential expansions). The wideband FMM is described in Section 6, and the performance of the FMM is illustrated in Section 7 with a number of numerical examples. Finally, Section 8 contains very brief conclusions.

2. Mathematical preliminaries

2.1. Symmetry

In describing various expansions and conversions between them, this paper makes use of a symmetry between far-field and local expansions, as follows. If we view (1) as a matrix–vector multiplication, and denote by A the matrix involved, then clearly A is symmetric; that is to say, the potential at any point x_1 due to a unit source at another point x_2 is equal to the potential at x_2 due to a unit source at x_1. Suppose X_1 and X_2 are two subsets of the points $\{x_k\}$; denote by A_{12} the submatrix of A mapping sources at points in X_2 to potentials at points in X_1, and by A_{21} the submatrix of A mapping sources at points in X_1 to potentials at points in X_2. Since A is symmetric, $A_{12} = A_{21}^{\mathrm{T}}$. If X_1 consists of the points on a box, and X_2 consists of the points for which that box's far-field expansion is valid, then for any type of far-field expansion for that box, there is a method of creating two matrices: a matrix C which creates the coefficients of the expansion, and a matrix E which evaluates the expansion, such that $\|A_{21} - EC\| < \varepsilon$, where ε is the accuracy of the expansion. Since A is symmetric, this also means that $\|A_{12} - C^{\mathrm{T}}E^{\mathrm{T}}\| < \varepsilon$; that is to say, for any type of far-field expansion, there is a corresponding type of local expansion, such that the creation matrix for the far-field expansion on any given box is the transpose of the evaluation matrix for the local expansion on that box, and the evaluation matrix for the far field expansion is the transpose of the creation matrix for the local expansion. It does not necessarily follow that the types of far-field and local expansion used in any particular version of the FMM have this symmetry; but in almost all of the versions we are aware of, and in all the expansions used in this paper, this is in fact the case, if minor changes such as rescalings and sign changes of the coefficients of the expansions are neglected.

This symmetry also extends to conversions between two far-field expansions, as compared to conversions between the two corresponding local expansions: in that case we have a far-field creation matrix C, a far-field evaluation matrix E for a different expansion (perhaps of a different type, or perhaps of the same type but for the parent box), and a conversion matrix B, such that $\|A_{21} - EBC\| < \varepsilon$. Then $\|A_{12} - C^{\mathrm{T}}B^{\mathrm{T}}E^{\mathrm{T}}\| < \varepsilon$; that is to say, if we have a conversion matrix between two far-field expansions, then the conversion matrix between the corresponding local expansions is its transpose.

Issues of numerical stability are likewise common to both types of expansion: if applying a sequence of matrices to a vector, one after another, is a numerically stable process, then so is applying, in the reverse order, the sequence of transposes of those matrices.

The situation is slightly complicated by the presence of dipoles or of other more complicated sources: if those are used, the resulting analog of (1) is no longer symmetric. However, the algorithm still uses the same types of expansions in that situation – that is, although the creation matrices for far-field expansions change, the evaluation matrices do not, nor do creation and evaluation matrices for local expansions. Thus translation matrices and conversion matrices between different types of expansions remain the same (and thus retain the above-described symmetry). Furthermore, if the desired result at each point were not the potential but an appropriate directional derivative of the potential, then the symmetry would be fully restored. Thus, even with this complication, the symmetry of far-field and local expansions permits us, in many places in this paper, after treating far-field expansions in full, to merely state that the local expansions are the corresponding ones, rather than repeating from a slightly different perspective their identical mathematics.

2.2. Spherical harmonics

We use the usual spherical coordinates, in which the mapping from spherical coordinates (r, θ, ϕ) to Cartesian coordinates (x, y, z) is given by the formulae:

$$x = r\sin\theta\cos\phi, \tag{3}$$

$$y = r\sin\theta\sin\phi, \tag{4}$$

$$z = r\cos\theta. \tag{5}$$

We denote by S^2 the two-dimensional sphere, each point $s \in S^2$ having coordinates $(\theta(s), \phi(s))$; for brevity, a point in spherical coordinates may be written (r, s), meaning $(r, \theta(s), \phi(s))$.

A function $f : S^2 \to \mathbb{C}$ is referred to as a spherical harmonic of degree n if the function

$$r^n \cdot f(\theta, \phi) \tag{6}$$

satisfies the Laplace equation in $\mathbb{R}^3$. As is well known (see, for example, [11]), for any integer $n \geqslant 0$, there exist exactly $2n + 1$ linearly independent spherical harmonics of degree n, an orthonormal basis for which consists of the functions

$$Y_n^m(\theta, \phi) = \overline{P}_n^{|m|}(\cos\theta)\, e^{im\phi} \tag{7}$$

for integer $m \in [-n, n]$, where $\overline{P}_n^m$ denotes the normalized associated Legendre function of degree n and order m, defined by the formula

$$\overline{P}_n^m(z) = \sqrt{\frac{2n+1}{4\pi}} \sqrt{\frac{(n-|m|)!}{(n+|m|)!}} \cdot P_n^m(z), \tag{8}$$

where P_n^m denotes the associated Legendre function of degree n and order m, which is given by Rodrigues' formula

$$P_n^m(x) = (-1)^m (1-x^2)^{(m/2)} \frac{d^m}{dx^m} P_n(x), \tag{9}$$

and where P_n denotes the Legendre polynomial of degree n. Any two spherical harmonics of different degrees are orthogonal, so the functions Y_n^m are all orthogonal to each other.

2.3. Integration of spherical harmonics

The grid of points used in the wideband FMM for sampling of functions on the sphere is the one introduced in [27]. It consists of the points (θ_k, ϕ_j), for $k = 1, \ldots, n$ and $j = 1, \ldots, 2n$, where the points $\{\phi_j\}$ are equispaced nodes on the circle (that is, on $[0, 2\pi]$), and the points $\{\theta_k\}$ are given by the formula

$$\theta_k = \arccos t_k, \tag{10}$$

where $t_1, \ldots, t_n \in [-1, 1]$ are the nodes of the n-point Gaussian quadrature on $[-1, 1]$. As is well known (see, for instance, [29]), the quadrature consisting of n equispaced nodes on the circle, with all quadrature weights equal to $2\pi/n$ (the trapezoidal rule) integrates exactly all functions of the form $e^{im\phi}$ with integer m such that $-n < m < n$, and the n-point Gaussian quadrature integrates polynomials of degree $2n - 1$ exactly: denoting the weights of that quadrature by $w_1, \ldots, w_n$, we have

$$\sum_{j=1}^{n} w_j t_j^m = \int_{-1}^{1} t^m \, dt \tag{11}$$

for any integer $m \in [0, 2n - 1]$.

For integration on the sphere, the weight at each grid point (θ_k, ϕ_j) is taken to be $w_k \cdot (2\pi/n)$. This quadrature integrates exactly any spherical harmonic of degree less than $2n$: when applied to a function Y_n^m, if $m \neq 0$, the integral on the sphere is zero, as is the result of applying the quadrature, while if $m = 0$, the integral of Y_n^m on the sphere is equal to

$$2\pi \int_0^{\pi} \overline{P}_n^{|m|}(\cos\theta) \sin\theta \, d\theta. \tag{12}$$

The substitution $t = \cos\theta$ converts (12) into the integral of a polynomial (due to (9), from which can be seen that the product of any two Legendre functions of the same order is a polynomial); it is thus integrated exactly by the Gaussian quadrature in t.

Since the product of a spherical harmonic of degree n_1 and a spherical harmonic of degree n_2 is a sum of spherical harmonics of degree $n_1 + n_2$ or less, the above quadrature integrates exactly products of spherical harmonics, provided the sum of their degrees is less than $2n$.

2.4. Filtering and interpolation of spherical harmonics

Translations using the high-frequency diagonal form (Section 4) involve filtering and interpolation of functions on the sphere of the form

$$f(\theta,\phi)=\sum_{n=0}^{p}\sum_{m=-n}^{n}\alpha_n^m Y_n^m(\theta,\phi). \tag{13}$$

In that context, the ideal filtering scheme is one in which the filtered function values, sampled on a coarser grid, are given by the same formula (13), with the same coefficients $\{\alpha_n^m\}$, except that coefficients with $n>p'$ are set to zero, for some $p'<p$. Likewise, the ideal interpolation scheme is one in which the interpolated function values are given by the same formula (13), with the same coefficients $\{\alpha_n^m\}$, only sampled on a finer grid. Either of these schemes can be accomplished by computing the coefficients $\{\alpha_n^m\}$, then using them to evaluate (13) on the desired grid of points.

The computation of the coefficients $\{\alpha_n^m\}$ from the values of f on S^2 is referred to as a spherical harmonic transform. Since the functions $\{Y_n^m\}$ are orthonormal on S^2, with the complex conjugate of $Y_n^m(s)$ being $Y_n^{-m}(s)$, it can be performed via the formula

$$\alpha_n^m=\int_{S^2} f(s)Y_n^{-m}(s)\,\mathrm{d}s. \tag{14}$$

As shown in Section 2.3, the quadrature defined in that section is exact for such integrals, provided that the number of nodes in the grid is sufficient. Using that quadrature with sufficiently many nodes, the number of nodes in the θ direction being denoted by q, and the number of nodes in the ϕ direction being $2q$, (14) becomes

$$\alpha_n^m=\frac{2\pi}{q}\sum_{k=1}^{q} w_k \sum_{j=1}^{2q} f(\theta_k,\phi_j)Y_n^{-m}(\theta_k,\phi_j). \tag{15}$$

This computation can be split into in two steps: first the calculation of the sums

$$\beta_k^m=\frac{2\pi}{q}\sum_{j=1}^{2q} f(\theta_k,\phi_j)\mathrm{e}^{-\mathrm{i}m\phi_j} \tag{16}$$

for all $k=1,\ldots,q$ and $m=-p,\ldots,p$, where p denotes the maximum order n for which the coefficients $\{\alpha_n^m\}$ are to be computed, then the calculation of the sums

$$\alpha_n^m=\sum_{k=1}^{p} w_k \beta_k^m \overline{P}_n^m(\cos\theta_k) \tag{17}$$

for all $n=0,\ldots,p$ and $m=-n,\ldots,n$. Since the nodes $\{\phi_j\}$ are equispaced on the circle, the first of these steps can be performed by q invocations of the Fast Fourier Transform (FFT), each taking $\mathrm{O}(q\log q)$ CPU time.

If filtering or interpolation of the above-described type is desired, the reverse of the above process must then be performed – that is, the evaluation of (13) on a (possibly different) grid of points $(\tilde{\theta}_k,\tilde{\phi}_j)$, of size $\tilde{q}\times 2\tilde{q}$. This process can likewise be accelerated by splitting it into two steps: first the calculation of the sums

$$\tilde{\beta}_k^m=\sum_{n=|m|}^{p}\alpha_n^m\overline{P}_n^m(\cos\tilde{\theta}_k) \tag{18}$$

for all $k=1,\ldots,\tilde{q}$ and $m=-k,\ldots,k$, then the calculation of the sums

$$f_k^j=\sum_{m=-n}^{n}\tilde{\beta}_k^m\,\mathrm{e}^{\mathrm{i}m\phi_j} \tag{19}$$

for all $k=1,\ldots,\tilde{q}$ and $j=1,\ldots,2\tilde{q}$. The second of these steps can be performed by $\tilde{q}$ invocations of the FFT, provided that the nodes $\{\tilde{\phi}_j\}$ are equispaced on the circle.

In [22], it was observed that the second and third stages of this interpolation or filtering process (that is, the computations (17) and (18)), combined, can be evaluated efficiently using the one-dimensional FMM, and that this, together with the use of the FFT for the other two stages, constitutes an efficient algorithm for filtering or interpolation on the sphere. In outline, that observation is as follows: the combination of (17) and (18), consists, for each $m = -p, \ldots, p$, of multiplication by a matrix whose (i,j)'th entry is given by the formula

$$w_i \cdot \sum_{n=|m|}^{p} \overline{P}_n^m(\cos\theta_i)\overline{P}_n^m(\cos\theta_j); \tag{20}$$

substituting into that formula the Christoffel–Darboux formula for the associated Legendre functions, which is

$$(\tilde{\mu} - \mu) \sum_{n=|m|}^{p} \overline{P}_n^m(\mu)\overline{P}_n^m(\tilde{\mu}) = \varepsilon_{p+1}^m \left(\overline{P}_{p+1}^m(\tilde{\mu})\overline{P}_p^m(\mu) - \overline{P}_p^m(\tilde{\mu})\overline{P}_{p+1}^m(\mu) \right), \tag{21}$$

where

$$\varepsilon_n^m = \sqrt{(n^2 - m^2)/(4n^2 - 1)}, \tag{22}$$

shows that the matrix (20) can be applied to a vector using two invocations of the one-dimensional FMM, each of which computes sums of the form

$$f_i = \sum_{j=1}^{N} \frac{q_j}{x_i - y_j}. \tag{23}$$

The wideband FMM uses this algorithm both for filtering and interpolation on the sphere; the one-dimensional FMM used in it is a version of the algorithm described in [31].

2.5. Spherical Bessel and Hankel functions

In accordance with standard practice [1], we denote by j_n the spherical Bessel function of the first kind of order n, and by h_n the spherical Hankel function of order n. Functions of both types are 'elementary' functions; in particular,

$$j_0(z) = \frac{\sin z}{z}, \tag{24}$$

$$h_0(z) = \frac{e^{iz}}{iz}. \tag{25}$$

For large n, the asymptotic behavior of $j_n(z)$ and $h_n(z)$ is given by the formulae (see [1, 9.3.1–9.3.3])

$$j_n(z) \sim \frac{z^n \cdot e^{n+\frac{1}{2}}}{2 \cdot (2n+1)^{n+1}}, \tag{26}$$

$$h_n(z) \sim \frac{\sqrt{2} \cdot (2n+1)^n}{z^{n+1} \cdot e^{n+\frac{1}{2}}}. \tag{27}$$

For large z with $\mathrm{Im}(z) \geqslant 0$, the asymptotic behavior of $j_n(z)$ and $h_n(z)$ is given by the formulae (see [1, 9.2.5, 9.2.7, 10.1.1])

$$\lim_{z\to\infty} z \cdot j_m(z) = \cos(z - (m+1)\pi/2) + O\left(\frac{e^{\mathrm{Im}(z)}}{|z|^2}\right), \tag{28}$$

$$\lim_{z\to\infty} z \cdot h_m(z) = e^{i(z-(m+1)\pi/2)} + O\left(\frac{e^{-\mathrm{Im}(z)}}{|z|^2}\right). \tag{29}$$

3. Partial-wave expansions

Classical partial-wave expansions are the primary type of expansion used by the low-frequency Helmholtz FMM. They are given by the following theorem (which can be found, for instance, in [24]), which expresses as a series the potential from sources inside a disk in $\mathbb{R}^3$ on points outside the disk:

Theorem 1. *Suppose that the function $\Phi : \mathbb{R}^3 \to \mathbb{C}$ is given by the formula*

$$\Phi(x) = \sum_{k=1}^{N} q_k h_0(\|x - x_k\|), \tag{30}$$

and that $\|x_k\| \leqslant a$ for each $k = 1, \ldots, N$. Then, denoting the spherical coordinates of the point x by (r, θ, ϕ), and the spherical coordinates of each point x_k by (r_k, θ_k, ϕ_k),

$$\Phi(x) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_n^m Y_n^m(\theta, \phi) h_n(\omega r) \tag{31}$$

for any point x such that $r > a$, where

$$\alpha_n^m = \sum_{k=1}^{N} q_k Y_n^{-m}(\theta_k, \phi_k) j_n(\omega r_k). \tag{32}$$

Expansions of the form (31) will be referred to as h-expansions. Since the potential at any point x due to a unit source at another point y is the same as the potential at y due to a unit source at x, the above expansion can also be used with the positions of sources and measurement points reversed:

Corollary 2. *Suppose the conditions of Theorem 1 hold, except that $x_k > a$ for $k = 1, \ldots, N$. Then*

$$\Phi(x) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \beta_n^m Y_n^{-m}(\theta, \phi) j_n(\omega r) \tag{33}$$

for any point x such that $r < a$, where

$$\beta_n^m = \sum_{k=1}^{N} q_k Y_n^m(\theta_k, \phi_k) h_n(\omega r_k). \tag{34}$$

Expansions of the form (33) will be referred to as j-expansions. In the FMM, whenever an h- or a j-expansion is used for some box in the hierarchy of boxes, the center of that box is taken to be the origin of the system of spherical coordinates used in (31), (32) or (33), (34).

Since j-expansions are related to h-expansions in the manner described in Section 2.1 – that is, the creation and evaluation operators for each type of expansion are the transposes of the evaluation and creation operators, respectively, for the other type of expansion – the following error analysis for h-expansions applies equally well to j-expansions.

3.1. Truncation error

Each h-expansion used in the 3D FMM represents the potential from sources inside a cubical box, at locations outside that box and its immediate neighbors. Denoting the length of the box by $2R$, and taking the center of the box as the origin of the system of spherical coordinates, any point (r_k, θ_k, ϕ_k) inside the box satisfies the inequality

$$r_k \leqslant \sqrt{3}R, \tag{35}$$

and any point (r, θ, ϕ) at which the h-expansion for that box must be accurate satisfies the inequality

$$r \geqslant 3R. \tag{36}$$

These constraints permit the h-expansion to be truncated to a manageable number of terms. Examination of (31), (32) shows that the appropriate truncation point depends on the possible magnitudes of the products

$$Y_n^m(\theta,\phi)\cdot h_n(\omega r)\cdot Y_n^{-m}(\theta_k,\phi_k)\cdot j_n(\omega r_k) \tag{37}$$

if, for some n and m, the magnitude of (37) is bounded by some number ε, and the accuracy budget can handle an error of size ε, then the term α_n^m can be omitted from the calculation. Since the functions $\{Y_n^m\}$ are orthonormal and smooth, and thus have maximum values roughly equal to one, they may be dropped from (37) with only a modest loss of accuracy, leaving it as:

$$h_n(\omega r)\cdot j_n(\omega r_k). \tag{38}$$

Since this is not a function of m (and for reasons of general convenience), the truncation used in the FMM is purely in n – that is to say, the terms retained in the series (31) are those for which $n=0,\ldots,p$, and $m=-n,\ldots,n$, where p depends on the desired accuracy and on the size R of the box. That dependence is a matter of the numerical behavior of the spherical Bessel functions h_n and j_n.

For real arguments, that behavior can be summarized as follows. For $x>n$, both $h_n(x)$ and $j_n(x)$ are oscillatory; as $x\to\infty$, the period of their oscillation tends toward a constant, and the magnitude of their oscillation decays as $1/x$. For $x<n$, $h_n(x)$ is monotonically increasing, starting from large negative values; in the limit as $x\to 0$, it is proportional to x^{-n-1}. For $x<n$, $j_n(x)$ is also monotonically increasing, but starting from small positive values; in the limit as $x\to 0$, it is proportional to x^n. Thus the non-negative values of n can be partitioned into three regions, in each of which the behavior of the product (38), under the constraints (35), (36), is of a different character:

- *Region I*: $0\leqslant n\leqslant\sqrt{3}\omega R$; here, in (38), j_n might be evaluated in its oscillatory regime, and h_n always is; since both factors of the product can be on the order of one, coefficients α_n^m for all these values of n must generally be included in the computation to achieve even minimal accuracy. For a box W wavelengths on each side (the wavelength being equal to $2\pi/\omega$), there are approximately $5W$ integer values of n in this region.
- *Region II*: $\sqrt{3}\omega R<n\leqslant 3\omega R$; here, in (38), j_n is evaluated on an argument which is less than n, and h_n is evaluated in its oscillatory regime. Thus the value of (38) is bounded by $j_n(\sqrt{3}\omega R)$, which decreases with increasing n. If the box is sufficiently large and/or the desired precision sufficiently small, this decrease permits the truncation point p to be placed in this region. For a box W wavelengths in size, there are approximately $4W$ integer values of n in this region.
- *Region III*: $3\omega R<n$; here, in (38), both j_n and h_n are evaluated on arguments which are less than n; thus j_n continues its decrease, and h_n starts to increase in magnitude. If $3\omega R\ll n$, then the approximations (26), (27) apply, and thus the maximum possible value of (38) under the constraints (36), (35) can be approximated by

$$\left(\frac{\sqrt{3}}{3}\right)^n\frac{1}{3R\cdot(2n+1)}. \tag{39}$$

  The accuracy of this asymptotic bound is not satisfactory for numerical use outside the asymptotic regime; accordingly, in the wideband FMM, we bound the significance of the coefficients α_n^m by evaluating the product $h_n(3\omega R)\cdot j_n(\sqrt{3}\omega R)$ numerically. Experimentally, this product decreases with increasing n throughout this region, not just in the asymptotic regime.

The calculation and use of coefficients α_n^m, for n in Region III, clearly is more delicate than in Regions I or II, since in Region III the coefficients α_n^m can have small values yet large significance, being multiplied by a large number ($h_n(x)$, for $x<n$) when used in the series (31). The most delicate case is that of the coefficients at the truncation point (those for which $n=p$), which at worst may be multiplied by $h_p(3\omega R)$ when the expansion is evaluated.

This difference in the scaling of coefficients renders the high-frequency diagonal form (described below in Section 4) unusable for boxes below a certain size (which depends on the desired accuracy, but is on the order of a wavelength); the details of the problem are described further in that section. However, the difference in scaling does not cause problems if formulae (31), (32) are used directly, except that if the boxes are very small,

it introduces a different type of numerical problem, namely overflow or underflow of the exponents of floating-point numbers. This can be avoided by rescaling expansions: replacing $h_n(x)$ with $x^n h_n(x)$, and $j_n(x)$ with $j_n(x)/x^n$. This rescaling must be carried through the entire analysis; but as that is both straightforward and tedious, we omit it in our descriptions in this paper.

The above count of terms in Regions I and II, combined with (39), yields the result that the order p of partial-wave expansion needed to achieve a given accuracy ε is $O(W + \log \varepsilon)$, where W is the size of the box in wavelengths. (The number of terms in a partial-wave expansion of order p is $(p+1)^2$).

In the above, only the case of real wavenumber ω has been considered. In the case of a complex wavenumber whose imaginary part is positive (yielding waves that decrease exponentially away from sources), it suffices to apply the above truncation criterion to the real part of the wavenumber. For wavenumbers with large imaginary parts, this is, however, not an optimal truncation; further savings are possible. If the wavenumber is purely imaginary, the method of [17] should be used, as it is optimized for that case.

3.2. Rotation

By "rotating" an h-expansion, we refer to the following task: given an expansion of the form (31) relative to one system of spherical coordinates, convert it into an equivalent expansion of the same form relative to another system of spherical coordinates which shares the same origin. Denoting the coordinates in the first system by (r, θ_1, ϕ_1) and in the second by (r, θ_2, ϕ_2) (since they share the same origin, the radius is the same in both systems), denoting the coefficients of the expansion in the first system by $\{\alpha_n^m\}$, and denoting the order of that expansion by p, the task is to find coefficients $\{\beta_n^m\}$ such that

$$\sum_{n=0}^{p} \sum_{m=-n}^{n} \alpha_n^m Y_n^m(\theta_1, \phi_1) h_n(\omega r) = \sum_{n=0}^{p} \sum_{m=-n}^{n} \beta_n^m Y_n^m(\theta_2, \phi_2) h_n(\omega r). \tag{40}$$

Since this is to be true for any r, it must be the case that

$$\sum_{m=-n}^{n} \alpha_n^m Y_n^m(\theta_1, \phi_1) = \sum_{m=-n}^{n} \beta_n^m Y_n^m(\theta_2, \phi_2), \tag{41}$$

for each $n = 0, \ldots, p$. Each of these $p+1$ subproblems amounts to converting a spherical harmonic of degree n from one orthonormal basis to another, which is an exact operation (thus justifying our above demand for equality between the two expansions). The wideband FMM uses only a small number of angles of rotation, and uses partial-wave expansions only in the low-frequency regime, where they have limited numbers of terms; thus rotation matrices to accomplish this task can easily be precomputed and stored. Various methods of computing rotation matrices are given in [4]; methods sufficient for an FMM are summarized in [17]. However, here, we need only use the brute-force numerical procedure of evaluating each function Y_n^m, rotated as desired, at suitable grid points (such as those defined in Section 2.3), then performing a spherical harmonic transform (see Section 2.4) to convert those values at grid points to spherical harmonic expansion coefficients: this yields the m'th column of the n'th rotation matrix. Using precomputed rotation matrices, a rotation takes $O(p^3)$ CPU time.

Rotation matrices for j-expansions are identical to those for h-expansions, as is evident either from symmetry considerations (as per Section 2.1) or by noting that the function j_n may be substituted for h_n in the above derivation without changing its result.

3.3. Translation

One of the basic building blocks of the FMM is a procedure for the translation of a far-field expansion for a child box into a far-field expansion for its parent. For partial-wave expansions, due to the ease of rotations (see the preceding section), we need only examine translations along the z-axis: any other translation can be performed by rotating the expansion so that its z-axis points along the direction of the desired translation (that is, from the center of the child box to the center of the parent box), translating it along the z-axis, then rotating it back—a procedure referred to as "point-and-shoot". Since each of these three steps can be

performed in $O(p^3)$ time (as shown above for rotations, and below for translation in z), this is more economical than performing the whole translation via the application of a single dense matrix, which would take $O(p^4)$ time.

Formulae for the translation of h-expansions along the z-axis can be found in Chapter 5 of [7], and in [17] (which gives formulae for the case of a purely imaginary wavenumber; to use them in the present context, the wavenumber must be divided by i). They are given in the following theorem:

Theorem 3. *Suppose a point c_1 lies on the positive z-axis at a distance ρ from the origin, and suppose that the potential Φ, at all points whose distance from c_1 is greater than r_1, is given by an expansion of the form (31) centered at c_1, whose coefficients are $\{\alpha_n^m\}$. Then at all points whose distance from the origin is greater than $r_1 + \rho$, Φ is given by an expansion of the same form centered at the origin, whose coefficients $\{\beta_n^m\}$ are given by the formula*

$$\beta_n^m = \sum_{n'=m}^{\infty} C_{n,n'}^m \alpha_{n'}^m, \tag{42}$$

where

$$C_{n,n'}^m = \sum_{k=m}^{\min n,n'} \left(\frac{1}{2}\right)(-1)^{n'+k}(2n'+1)\cdot \frac{(n'-m)!(n+m)!(2k)!\mathrm{i}^{-(n'+n)}(-\omega\rho)^{-k} j_n(\omega\rho)}{(k+m)!(k-m)!(n'-k)!(n-k)!k!}. \tag{43}$$

In the wideband FMM, matrices containing the coefficients $\{C_{n,n'}^m\}$ are precomputed; since only one such set of coefficients is needed at each level of the hierarchy, and since they are only computed in the low-frequency regime, the time for this precomputation is negligible. Working with these precomputed matrices, the CPU time taken to perform a translation via (42) is $O(p^3)$.

Due to the symmetry discussed in Section 2.1, translation of a j-expansion from parent to child along the z-axis is the transpose of the above translation operation; thus a similar point-and-shoot procedure is used for it, using the transpose of each of the precomputed matrices used in the above procedure, and applying them in reverse order.

As a remark, the fact that translations along the z-axis do not mix together coefficients with different values of m (and thus that they are $O(p^3)$) can be seen without delving into the full proof of Theorem 3. Such translations involve two h-expansions, each relative to a different system of spherical coordinates, but with the coordinate ϕ of any given point being the same in both systems. In each expansion, the coefficients whose superscript is m are multiplied by $\mathrm{e}^{\mathrm{i}m\phi}$, forming the mth Fourier component in ϕ of the potential. For the two expansions to yield the same potential, it is necessary and sufficient that their respective Fourier components be equal, which implies that the coefficients of each expansion with any given m must depend only on the coefficients of the other expansion for that m.

4. High-frequency diagonal form

In [27], forms of far-field and local expansion for the Helmholtz equation were introduced which have the property that translation operators (far field to far field, far field to local, and local to local) are diagonal. They are defined in terms of partial-wave expansions, as follows. The far-field expansion for a given box, which is referred to as its "far-field signature", consists of the function $f : S^2 \to \mathbb{C}$ given by the formula

$$f(s) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (-\mathrm{i})^{n+1} \alpha_n^m Y_n^m(s), \tag{44}$$

where $\{\alpha_n^m\}$ are the coefficients of the h-expansion for that box. The local expansion for a given box (also referred to as a far-field signature) consists of the function $g : S^2 \to \mathbb{C}$ such that

$$g(s) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (-\mathrm{i})^{n+1} \beta_n^m Y_n^m(s), \tag{45}$$

where $\{\beta_n^m\}$ are the coefficients of the j-expansion for that box.

Since h- and j-expansions are related to each other in the manner described in Section 2.1 (that is, the creation operator for the former is the transpose of the evaluation operator for the latter), far-field signatures inherit, through (44) and (45), that same relationship.

From (44), the definition (31) of h-expansions, and the asymptotic behavior of the function h_n for large arguments (28), it follows that the far-field signature f of any given box has the property that

$$f(s) = \lim_{r\to\infty} \Phi(r,s) \cdot r \cdot \mathrm{e}^{-\mathrm{i}\omega r}, \tag{46}$$

where $\Phi : \mathbb{R}^3 \to \mathbb{C}$ is the potential due to the h-expansion on that box (and thus, in the FMM, due to sources on that box), and where the system of spherical coordinates (r,s) has as its origin the center of that box.

Likewise, one may regard the function g for a given box, as given by (45), as a source distribution on a sphere centered on that box, which possesses the property that as the radius of that sphere is taken to infinity, the source distribution being appropriately scaled and adjusted in phase, the potential generated by it approaches the potential on the box.

4.1. Discretization

It is evident from (44), (45) that conversion of a partial-wave expansion to a far-field signature, or vice versa, amounts to a spherical harmonic transform, forward or inverse (plus appropriate scaling by powers of i). Thus the grid points defined in Section 2.3 are suitable for sampling far-field signatures; as shown in Section 2.4, if a $p \times 2p$ grid of that type is used, it allows for exact transformations from functions sampled on the grid to coefficients $\{\alpha_n^m\}$ of a partial-wave expansion of degree $p-1$, and vice versa. Numerically, since the basis functions $\{Y_n^m\}$ are orthonormal, these conversions are very well conditioned. However, in the case that the coefficients of the partial-wave expansion are scaled differently from each other, as they are for boxes less than roughly a wavelength in size, these conversions degrade or destroy the accuracy of the expansion, depending on how severe the difference in scaling is. As shown in Section 3.1, the worst case ill-conditioning is $h_p(3\omega R)$, where p is the number of terms in the h-expansion, and R is half the length of the box. Accordingly, the wideband FMM proceeds by calculating this value for each box size, and not using the high-frequency diagonal form for any boxes for which it exceeds the allowable loss of precision (as given by the ratio between the accuracy desired and the accuracy provided by the form of floating-point arithmetic in use). Since boxes at a given level in the hierarchy are all of the same size, this means that a cutoff line is drawn, above which the high-frequency form is used, and below which low-frequency techniques are used; the more precision is required, the higher in the hierarchy the line is. (See Table 1 for experimentally determined cutoff lines for the accuracies 10^{-3}, 10^{-6}, and 10^{-9}.)

4.2. Far field to far field and local to local translations

The following theorem is proven in [27]:

Theorem 4. *Suppose $f_1 : S^2 \to \mathbb{C}$ and $f_2 : S^2 \to \mathbb{C}$ are two far-field signatures for the same potential, centered at points c_1 and c_2, respectively. Then*

$$f_2(s) = f_1(s) \cdot \mathrm{e}^{\mathrm{i}\omega\cdot(c_2-c_1,E(s))} \tag{47}$$

for all $s \in S^2$, where $E(s)$ denotes the unit vector in the direction s.

Table 1
Transition points to diagonal forms for translation operators

Requested accuracy	Box size (wavelengths)
10^{-3}	0.25
10^{-6}	3.50
10^{-9}	12.0

Theorem 4 provides a way to translate far-field signatures upwards in the hierarchy of boxes. Numerically, both the function f_1 (the child's expansion) and the function f_2 (the child's expansion recentered on the parent box) are discretized on grids of the type defined in Section 2.3, with the grid used for f_2 having roughly twice as many points in each direction. Upwards translation thus consists of interpolating f_1 to the finer grid, then performing the operation (47) at each grid point. Due to the relation (44), the type of interpolation defined in Section 2.4 is ideal for this purpose.

Theorem 4 also provides a way to generate far-field signatures for the potential generated by a set of point sources, without going through partial-wave expansions: each point source has a far-field signature, centered on itself, which for simple sources (of the type (2)) is a constant; such far-field signatures, one for each point in the box, can each be translated to the center of the box via (47), then added together. (Formulae for this, for both simple sources and dipoles, can be found in [27].) This can be used to generate far-field signatures for boxes whose size is above the high-frequency cutoff but which contain too few source points to be worth subdividing (although as implemented, the wideband FMM simply subdivides those boxes anyway).

For the reasons described in Section 2.1, each local-to-local translation operator is the transpose of the far-field-to-far-field translation operator for the same geometry. Considered as operators on functions on the sphere, both are diagonal, and thus are identical to each other. Numerically, for translation from parent to child, the translated expansion is then filtered via the procedure described in Section 2.4, so as to reduce the number of grid points at which it is discretized to that used at the child's level, discarding those parts of the far-field signature which have insignificant effects on the child box.

4.3. Far field to local translation

The following theorem provides a means for translating a far-field expansion of the high-frequency diagonal form to a local expansion in the same form.

Theorem 5. *Suppose $f_1 : S^2 \to \mathbb{C}$ is a far-field signature centered at a point c_1 and valid outside a ball D_1 of center c_1 and radius R_1; denote the potential it yields by $\Phi : \mathbb{R}^3 \setminus D_1 \to \mathbb{C}$. Suppose D_3 is another ball, which has center c_3 and radius R_3, such that $D_3 \cap D_1 = \emptyset$. Let the functions $\mu_k : S^2 \to \mathbb{C}$, for positive integer k, be defined by the formula*

$$\mu_k(s) = \sum_{m=0}^{k} \mathrm{i}^m \cdot (2m+1) \cdot P_m(\cos(\theta(c_3 - c_1, s))) \cdot h_m(\omega \|c_3 - c_1\|), \tag{48}$$

where $\theta(c_3 - c_1, s)$ denotes the angle between the vector $c_3 - c_1$ and the direction s. Let $g_k : S^2 \to \mathbb{C}$ be given by the formula

$$g_k(s) = f_1(s) \cdot \mu_k(s) \tag{49}$$

and let the functions $g_{k,n} : S^2 \to \mathbb{C}$ be defined by the formula

$$g_k(s) = \sum_{n=0}^{\infty} g_{k,n}(s) \tag{50}$$

together with the restriction that $g_{k,n}(s)$ must be a spherical harmonic of degree n. Let $\Phi_k : D_3 \to \mathbb{C}$ be given by the formula

$$\Phi_k(r, s) = \sum_{n=0}^{\infty} g_{k,n}(s) j_n(\omega r), \tag{51}$$

where r and s are coordinates in the system of spherical coordinates whose origin is c_3. Then, for any $x \in D_3$,

$$\lim_{k \to \infty} \Phi_k(x) = \Phi(x). \tag{52}$$

Furthermore,

$$\max_{D_3} |\Phi_k(x) - \Phi(x)| = \mathrm{O}\left(\left(\frac{R_1 + R_3}{\|c_3 - c_1\|} \right)^k \cdot \|f_1\| \right). \tag{53}$$

The above theorem is proven in [27]. The proof is based on the same formula (the addition theorem for spherical Bessel functions) on which Theorem 1 is based; accordingly, the truncation error, as given by (53), behaves similarly to the truncation error of partial-wave expansions (described above in Section 3.1): the significance of the kth term in the series is indicated by the magnitude of the product

$$h_k(\omega r) \cdot j_k(\omega \rho), \tag{54}$$

with

$$r = \|c_3 - c_1\|, \tag{55}$$

$$\rho < R_1 + R_3. \tag{56}$$

In the context of the FMM, far-field-to-local translation is done between boxes on the same level of the hierarchy which are separated by at least one box of the same size; thus the bounds (55), (56) become

$$r \geqslant 4R, \tag{57}$$

$$\rho < 2\sqrt{3}R, \tag{58}$$

where R is half the length of each of the boxes. Comparison of (58) to the corresponding bound (35) for partial-wave expansions shows that for sufficiently high frequencies, the appropriate truncation points to achieve a given accuracy, in the context of the FMM, are related as follows: when translating a far-field signature of degree p between boxes on the same level, the appropriate truncation point in (48) is $2p$. At lower frequencies, it is somewhat less than that; in the wideband FMM, it is determined numerically by evaluation of the worst-case value of (54), which is obviously

$$h_k(4\omega R) \cdot j_k(2\sqrt{3}\omega R). \tag{59}$$

With the above truncation point, (49) is a product of two sums of spherical harmonics, one of maximum degree p, the other of maximum degree $2p$. The resulting product is thus a sum of spherical harmonics of maximum degree $3p$. That product is then to be filtered, eliminating all but components of degree p or less; this can be done if the grid on which the product is sampled is of the type defined in Section 2.3, with at least $2p$ points in θ and $4p$ points in ϕ. Thus, numerically, the process of translating a far-field expansion to a local expansion consists of resampling the far-field expansion on a twice-finer grid, multiplying by the function μ_{2p} at each grid point, then filtering the result back onto the original grid, which yields the local expansion.

Three remarks on the numerical implementation of diagonal forms of translation operators seem in order.

Remark 6. The evaluation of the Far-Field-to-Local translation operators is where numerical instability manifests itself if an attempt is made to use high-frequency expansions for boxes that are too small. Indeed, the series (48) obviously does not converge as $k \to \infty$; in fact, $h_m(\omega\|c_3 - c_1\|)$ starts growing in magnitude, once m exceeds $\omega\|c_3 - c_1\|$. For sufficiently large $\omega\|c_3 - c_1\|$ (or, equivalently, for sufficiently large R), this problem does not occur, since the series is truncated before $h_m(\omega\|c_3 - c_1\|)$ becomes too large. As a practical matter, given a certain precision of calculations and a certain desired accuracy, one can determine the minimum size of the box for which the diagonal form of the translation operator provided by Theorem 5. When the calculations are performed in 64-bit floating-point arithmetic, Table 1 lists some of the cut-off points. We refer the reader to [25] for an excellent discussion of this class of issues.

Remark 7. The filtering and interpolation operations on the sphere are quite expensive compared to multiplication by a diagonal translation matrix, and at first glance appear to be the dominant element of the algorithm (so far as the CPU time requirements are concerned). Fortunately, each of them only needs to be performed once for each box: in the entire wideband FMM, there is only one function being interpolated for each box (the far-field expansion), and it is interpolated to the same grid each time. Likewise, as regards filtering, a number of functions, in the above descriptions, get filtered then added together to produce the local expansion for each box; but, since filtering is a linear operation, they can be added together before filtering, and the filter applied once.

Remark 8. The final remark concerns the calculation of the function μ_{2p} on grid points on the sphere. This calculation would be overly expensive if done naively: direct evaluation of (48) at each of $O(p^2)$ points would require $O(p^3)$ time. At the highest levels in the hierarchy, p is, in many important cases, on the order of $\sqrt{N}$ (where N is the total number of input points, as in (1)); thus a single $O(p^3)$ operation would result in an $O(N^{3/2})$ algorithm. But the only place that s enters into the calculation of $\mu_{2p}(s)$ is in the taking of the angle between s and the vector $c_3 - c_1$; thus μ_{2p} can be tabulated as a function of that angle, then interpolated as necessary for points on the sphere; if a local interpolation method is used for this, the tabulation takes $O(p^2)$ time ($O(p)$ points, with $O(p)$ time per point), as does the interpolation (to $O(p^2)$ points, with $O(1)$ time per point).

5. Exponential expansions

For translation from far-field expansions to local expansions in the low-frequency regime (that is, from h-expansions to j-expansions), the wideband FMM uses the scheme introduced in [20]. In that scheme, the translation is performed by means of "exponential expansions" (also known as "plane wave expansions"), which are based on the following formula for the Green's function of the Helmholtz equation, which is valid for $z > 0$:

$$\frac{e^{i\omega r}}{r} = \frac{1}{2\pi}\int_0^\infty e^{-\sqrt{\lambda^2-\omega^2}\cdot z}\int_0^{2\pi} e^{i\lambda(x\cos\alpha + y\sin\alpha)}\,d\alpha\frac{\lambda}{\sqrt{\lambda^2-\omega^2}}\,d\lambda. \tag{60}$$

In [20], the outer integral in (60) was divided into two parts, a "propagating" part and an "evanescent" part, on each of which a different quadrature was used. In the wideband FMM, we instead use a single quadrature for the whole of the outer integral. To determine the appropriate quadrature, we first evaluate the inner integral analytically, which transforms the right-hand side of (60) into

$$\frac{1}{2\pi}\int_0^\infty e^{-\sqrt{\lambda^2-\omega^2}\cdot z} J_0(\lambda\sqrt{x^2+y^2})\frac{\lambda}{\sqrt{\lambda^2-\omega^2}}\,d\lambda. \tag{61}$$

An appropriate quadrature is thus one which integrates (61) accurately for any x, y, and z within the ranges used by the algorithm. Those ranges are as follows:

$$L \leqslant z \leqslant 4L, \tag{62}$$

$$-4L \leqslant x, y \leqslant 4L, \tag{63}$$

where L is the length (on each side) of the box to which the exponential expansion belongs.

The appropriate quadrature obviously depends on two variables: the wavenumber ω and the size L of the box. However, it depends chiefly on the product ωL (the real part of which is proportional to the size of the box in wavelengths); it is easy to show that if this is held constant while ω and L are varied, then the number of quadrature nodes required to achieve the same relative accuracy remains the same, while the weights and nodes are rescaled by a common factor.

The wideband FMM uses quadratures custom-tailored to this exact problem using a variant of the algorithm presented in [10]. These quadratures have a constant term; that is, (60) is approximated via the formula

$$\frac{e^{i\omega r}}{r} \approx a + \frac{1}{2\pi}\sum_{k=1}^{s} w_k\, e^{-\sqrt{\lambda_k^2-\omega^2}\cdot z}\int_0^{2\pi} e^{i\lambda_k(x\cos\alpha + y\sin\alpha)}\,d\alpha\frac{\lambda_k}{\sqrt{\lambda_k^2-\omega^2}}, \tag{64}$$

in which the accuracy of approximation is selected to be on the same order as the accuracy desired of the wideband FMM as a whole. The quadrature-generation algorithm takes as input that desired accuracy, and yields the number s of weights w_k and nodes λ_k, as well as the weights and nodes themselves and the constant a, to satisfy (64) to that accuracy. These quadratures are not computed at runtime, but are precomputed for each of several distinct ranges of the product ωL, which together encompass all the values of that product which are possible in the low-frequency regime.

The inner integral of (60) is obviously an integral of a smooth, periodic function; thus the trapezoidal rule is appropriate for it. The amount of oscillation of the function to be integrated obviously depends on the value

of the coefficient λ_k, as does thus the number of nodes required for the trapezoidal rule to achieve the desired accuracy; we denote that number by M_k, and denote the total number of quadrature nodes needed by S_{exp}; clearly, $S_{\text{exp}} = 1 + \sum_{k=1}^{s} M_k$. A somewhat involved analysis, which we omit, shows that $s \approx p$, where p is the order of multipole expansion required to achieve the same accuracy, and also that $S_{\text{exp}} \approx p^2$.

The ranges (62), (63) arise from the way exponential expansions are used, which is as follows. In the FMM, the "interaction list" of a box is defined to be the list of boxes on the same level of the hierarchy which are not adjacent to it (adjacency being defined as having any common boundary point, even just a corner), but whose parents are adjacent to its parent. The number of such boxes is clearly at most $6^3 - 3^3 = 189$—fewer for boxes near the edge of the problem domain, or near larger boxes which were not subdivided to their level. For each box F in the interaction list of a box B, the FMM applies a far-field-to-local translation operator to convert the far-field expansion on B into a local expansion on F.

To use exponential expansions for those translations, the interaction list of a box is partitioned into six lists, one for each face of the box:

- the $+z$-list: boxes separated by at least one box in the $+z$-direction,
- the $-z$-list: boxes separated by at least one box in the $-z$-direction,
- the $+y$-list: boxes separated by at least one box in the $+y$-direction, and not contained in the $+z$ or $-z$ lists,
- the $-y$-list: boxes separated by at least one box in the $-y$-direction, and not contained in the $+z$ or $-z$ lists,
- the $+x$-list: boxes separated by at least one box in the $+x$-direction, and not contained in the $+z$, $-z$, $+y$, or $-y$ lists,
- the $-x$-list: boxes separated by at least one box in the $-x$-direction, and not contained in the $+z$, $-z$, $+y$, or $-y$ lists.

A $+z$-list is depicted in Fig. 2. An exponential expansion as described above is referred to as facing in the $+z$-direction (that being the direction in which the exponentials in it decay). Exponential expansions facing in the other five directions are defined similarly, the only difference being in the names of the coordinate axes. For each of the six lists, an exponential expansion facing in that direction is used for translations to boxes in that list. It is evident that if a point $(x_0, y_0, z_0) \in \mathbb{R}^3$ is in a box B of length L, and a point $(x, y, z) \in \mathbb{R}^3$ is in a box of B's $+z$-list, then $L \leqslant z - z_0 \leqslant 4L$, $-4L \leqslant x - x_0 \leqslant 4L$, and $-4L \leqslant y - y_0 \leqslant 4L$; thus those are the ranges for which quadratures are constructed, as described above.

Substituting those quadratures into (60), the potential at a point (x, y, z) due to a source at a point (x_0, y_0, z_0) can be approximated via the formula

$$\frac{e^{i\omega\|(x,y,z)-(x_0,y_0,z_0)\|}}{\|(x,y,z)-(x_0,y_0,z_0)\|} \approx a + \sum_{k=1}^{s} w_k e^{-\sqrt{\lambda_k^2-\omega^2}\cdot(z-z_0)} \cdot \frac{\lambda_k}{\sqrt{\lambda_k^2-\omega^2}} \cdot \frac{1}{M_k} \sum_{j=1}^{M_k} e^{i\lambda_k((x-x_0)\cos\alpha_{kj}+(y-y_0)\sin\alpha_{kj})}, \tag{65}$$

provided that $L \leqslant z - z_0 \leqslant 4L$, $-4L \leqslant x - x_0 \leqslant 4L$, and $-4L \leqslant y - y_0 \leqslant 4L$. Each of the terms summed up on the right-hand side of (65) is evidently a plane wave; that is to say, it is of the form

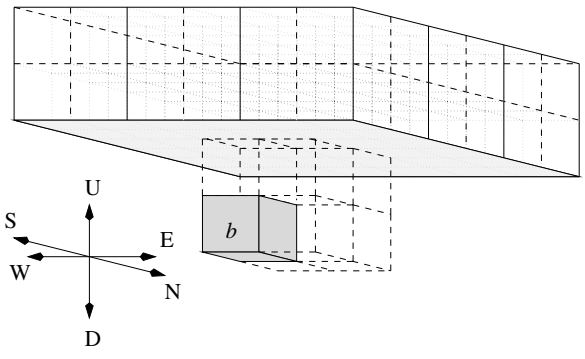

Fig. 2. The $+z$-list of the box b.

$$W_j e^{(C_j, u-u_0)}, \tag{66}$$

where $u = (x, y, z)$ and $u_0 = (x_0, y_0, z_0)$, with $W_j \in \mathbb{R}$ and $C_j \in \mathbb{C}^3$. Using this notation, if sources at points $u_1, \ldots, u_n$, with respective source strengths $s_1, \ldots, s_n$, are located in a box B, then the resulting potential at any point u located in any of the boxes of B's $+z$-list can be written as

$$\Phi(u) = \sum_{k=1}^{n} s_k e^{i\omega\|u-u_k\|} \approx \sum_{k=1}^{n} s_k \sum_{j=1}^{J} W_j e^{(C_j, u-u_k)}. \tag{67}$$

This, in turn, can be broken up into three parts: first, the creation of an exponential expansion located at a point v_1, via the formula

$$\alpha_j = W_j \sum_{k=1}^{n} s_k e^{(C_j, v_1-u_k)}, \tag{68}$$

for each $j = 1, \ldots, J$, then (optionally) its translation to another point v_2, via the formula

$$\beta_j = \alpha_j e^{(C_j, v_2-v_1)}, \tag{69}$$

for each $j = 1, \ldots, J$, then its evaluation at the point u:

$$\Phi(u) \approx \sum_{j=1}^{J} \beta_j e^{(C_j, u-v_2)}. \tag{70}$$

The translation (69) can obviously be performed any number of times, with no loss of accuracy; the accuracy depends only on the relative locations of each pair of source and measurement points, which must satisfy conditions (62), (63). In the wideband FMM, the location v_1 of the exponential expansion for a box B is chosen to be the center of the box, and the location v_2 from which the exponential expansion is evaluated is chosen to be the center of the box of B's $+z$-list that u resides in. With these choices, the creation operator (68) and the evaluation operator (70) are transposes of each other (thus possessing symmetry of the type described in Section 2.1), except that the former operator contains the weights $\{W_j\}$; having each operator contain the square roots of those weights would render the symmetry complete.

The advantage of using exponential expansions is, as previously mentioned, that their translation operator (69) is diagonal, and thus can be applied in $O(p^2)$ time. Since translation is the most frequently applied operation, being applied up to 189 times per box, this yields significant savings. The ability of exponential expansions to be translated more than once, with no loss of accuracy nor change of region of validity, allows for further savings: examining the $+z$-lists of the eight children of a parent box, and referring to the $+z$-direction as being "upwards", it is evident that the $+z$-lists of the top four children are the same, as are the $+z$-lists of the bottom four children; and the former is a subset of the latter. Thus the $+z$ translations of those eight children can be accomplished by translating the exponential expansions of the bottom four children to a common point, adding them together, evaluating the resulting expansion on the boxes which are only in the bottom four children's $+z$-list, then translating the exponential expansions of the top four children to the same common point and adding them in, then evaluating the resulting expansion on the boxes which are in all eight children's $+z$-lists. This optimization reduces the maximum number of translations per box from 189 to roughly 40.

5.1. Conversion between exponential and partial-wave expansions

In the wideband FMM, exponential expansions for each box are created not via (68) but rather from the h-expansion for that box; likewise, exponential expansions are not evaluated on a box via (70), but rather are converted into a j-expansion on that box. Since exponential expansions and partial-wave expansions each possess the symmetry described in Section 2.1, these two types of conversions (h-expansion to exponential expansion and exponential expansion to j-expansion) are transposes of each other; thus we describe only the latter type of conversion.

The representation of a plane wave as a partial-wave expansion belongs to a well-studied part of classical mathematical physics; the formula used in the wideband FMM for the conversion is easily derived from the following formula from [24, vol. 2, 11.3.46]:

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{n=0}^{\infty}(2n+1)i^n \sum_{m=0}^{n} \epsilon_m \frac{(n-m)!}{(n+m)!} \cos(m(\phi - v)) \cdot P_n^m(\cos u) P_n^m(\cos\theta) j_n(kr), \tag{71}$$

where the vector $\mathbf{r}$ is of length r and has the spherical angles θ, ϕ, where the vector $\mathbf{k}$ is of length k and has the spherical angles u, v, and where ϵ_m is the Neumann factor ($\epsilon_m = 1$ when $m = 0$, $\epsilon_m = 2$ when $m > 0$). Using the formula $\cos z = (e^{iz} + e^{-iz})/2$, (71) becomes

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{n=0}^{\infty}(2n+1)i^n \sum_{m=-n}^{n} \frac{(n-|m|)!}{(n+|m|)!} e^{im(\phi - v)} \cdot P_n^{|m|}(\cos u) P_n^{|m|}(\cos\theta) j_n(kr). \tag{72}$$

To convert (72) to the form used in this paper, we make the substitutions $k = \omega$, $v = \alpha_{kj}$, and $\cos u = (i\sqrt{\lambda_k^2 - \omega^2})/\omega$, the last of which yields the formula $\sin u = \lambda_k/\omega$. The Cartesian components of the vector $\mathbf{k}$ which result from these substitutions are:

$$k_x = k \sin u \cos v = \lambda_k \cos \alpha_{kj}, \tag{73}$$

$$k_y = k \sin u \sin v = \lambda_k \sin \alpha_{kj}, \tag{74}$$

$$k_z = k \cos u = i\sqrt{\lambda_k^2 - \omega^2}. \tag{75}$$

Making the above substitutions into (76) yields the formula

$$e^{-\sqrt{\lambda_k^2 - \omega^2}\cdot z} \cdot e^{i\lambda_k(x\cos\alpha_{kj} + y\sin\alpha_{kj})} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \beta_n^m \, Y_n^m(\theta, \phi) j_n(\omega r), \tag{76}$$

in which (x, y, z) are the Cartesian equivalents of the spherical coordinates (r, θ, ϕ), and where

$$\beta_n^m = 4\pi i^n \overline{P}_n^{|m|}\left(\frac{i\sqrt{\lambda_k^2 - \omega^2}}{\omega}\right) e^{im\alpha_{kj}}. \tag{77}$$

The left-hand side of (76) is a single term of an exponential expansion (facing in the $+z$-direction), corresponding to the quadrature nodes λ_k and α_{kj}; the right-hand side is the equivalent j-expansion, with (77) being the formula for the coefficients of that expansion.

Denoting the coefficients of the exponential expansion by $\{E_{kj}\}$, the formula for converting the entire exponential expansion into a j-expansion is thus

$$\beta_n^m = \sum_{k=1}^{s} 4\pi i^n \overline{P}_n^{|m|}\left(\frac{i\sqrt{\lambda_k^2 - \omega^2}}{\omega}\right) \sum_{j=1}^{M_k} E_{kj} e^{im\alpha_{kj}}. \tag{78}$$

Since the nodes $\alpha_{k1}, \ldots, \alpha kM_k$ are equispaced, the inner sums of (78) can be performed using the FFT; doing so reduces the CPU time required for the whole conversion to $O(p^3)$.

The exponential expansion to which the above conversion formula applies faces in the $+z$-direction; thus, using formula (78) on an exponential expansion facing in another direction results in a j-expansion relative to a system of coordinates whose z-axis points in that direction; the latter must then be rotated (as described in Section 3.2) to the standard orientation.

6. Algorithm

The wideband FMM, as implemented, is adaptive; that is, the hierarchical subdivision of the problem domain is deeper in places where there are more source points: the rule used is that a box is subdivided if it encloses more than a certain number d of source points. (The number d is chosen so as to roughly minimize

the CPU time.) However, in the high-frequency regime, no adaptivity is used: any box above the high-frequency cutoff line which contains any points whatsoever is subdivided. This removes the necessity to consider the cases of nearby boxes of different size which interact, one or more of those boxes being above the high-frequency cutoff line; instead, all such cases are entirely in the low-frequency regime. (Since the normal practice is to discretize objects using more than two points per wavelength, this is not a major limitation of the code.) After the problem domain is subdivided, the main computation is performed. It consists of three stages: first, creation of far-field expansions for each box; second, translation of far-field expansions to local expansions; third, evaluation of the local expansions.

The first stage is a bottom-up pass through the hierarchy of boxes. For each childless box, an h-expansion is created via (32). Then, for each parent box below the high-frequency cutoff line, an h-expansion is created from the h-expansions of its children, using the "point-and-shoot" method described in Section 3.3. At the high-frequency cutoff line, each h-expansion is converted into a far-field signature via a spherical harmonic transform (Section 2.4). Above the cutoff line, far-field signatures for parents are created from the far-field signatures of their children as described in Section 4.2. At the end of this stage, a far-field expansion (either an h-expansion or a far-field signature) has been computed for all the boxes in the hierarchy.

In the second stage, for each box below the high-frequency cutoff line, six exponential expansions are created from the h-expansion for that box, as described in Section 5.1. These are then translated to each box in its interaction list, as described in Section 5. The resulting six exponential expansions (of the local-expansion variety) on each box are converted into j-expansions on that box, as described in Section 5.1. This accounts for all interactions between boxes of the same size in the low-frequency regime. Interactions between boxes of different sizes only occur when the larger of those two boxes has not been subdivided, and thus has O(1) points on it; accordingly, such interactions are handled by evaluating the smaller box's h-expansion directly on the larger box, or (for interactions going the other direction) by creating a j-expansion on the smaller box directly from the source points on the larger box. (In all cases, interactions are handled at as high a level as possible; thus interactions between boxes of different size only occur when the parent of the smaller box is closer than its own length to the larger box, and thus its expansions are invalid on the larger box.) For boxes above the high-frequency cutoff line, the far-field signature for each box is translated to each of the boxes in its interaction list via the procedure described in Section 4.3.

The third stage is a top-down pass through the hierarchy of boxes. Starting from the top level, the local expansion on each parent box is translated into a local expansion on each of its children, and added to the existing local expansion (from the second stage) on that child. Above the high-frequency cutoff line, the translation process is performed as described in Section 4.2; to cross the line, a spherical harmonic transform is used; below the line, the translation process is performed as described in Section 3.3. After this, the j-expansion on each childless box is evaluated to yield the potential at the points in that box.

The above three stages handle all parts of (1) except for the interactions between points $\{x_j\}$ which are in childless boxes which are adjacent to each other, and between points $\{x_j\}$ which are in the same childless box; those interactions are evaluated directly.

7. Numerical results

The wideband FMM has been applied to several test cases. The first of these is an aircraft-shaped object (Fig. 3), 50 wavelengths in size. The surface of the aircraft was divided into 706,300 triangles, on each of which a single node was placed; the size of the smallest triangle was 1.06×10^{-6} wavelengths, and the size of the largest was 2.86×10^{-1} wavelengths. The wideband FMM was run at each of three different levels of accuracy, and the error calculated by comparing the results to those produced by direct application of (1). Table 2 shows the results, as well as the time (in seconds) taken by the direct method; the "error" columns contain the relative error in the L^2 sense (the L^2 norm of the error, divided by the L^2 norm of the correct result), and the first column contains the time taken by the direct evaluation of (1). For comparison, the results of applying the FMM for the Laplace potential to the same geometry are shown in Table 3; the Laplace FMM used was the one described in [8], which is at a level of technology similar to that of the wideband FMM of this paper: it is adaptive, and it uses exponential expansions for diagonal translations.

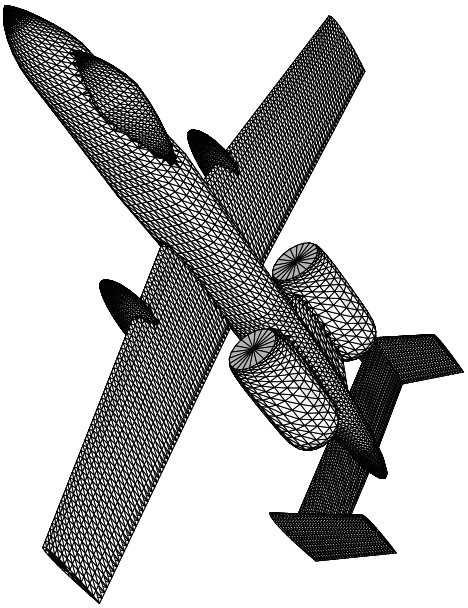

Fig. 3. Surface triangulation of an aircraft-shaped object 50 wavelengths in size. The size of the smallest triangle is 1.06×10^{-6} wavelengths, and the size of the largest is 2.86×10^{-1} wavelengths.

Table 2
Example 1: aircraft-shaped object

Time (direct)	Requested accuracy	Error in potential	Error in gradient	Time (s)	Memory (MB)
337,329	10^{-3}	0.43E − 3	0.56E − 3	485	300
337,329	10^{-6}	0.48E − 6	0.50E − 6	1291	790
337,329	10^{-9}	0.11E − 9	0.95E − 10	2947	1143

Table 3
Aircraft-shaped object – Laplace potential

Time (direct)	Requested accuracy	Error in potential	Error in gradient	Time (s)	Memory (MB)
60,590	10^{-3}	0.27E − 3	0.37E − 4	48.3	211
60,590	10^{-6}	0.19E − 6	0.43E − 7	119	292
60,590	10^{-9}	0.85E − 10	0.61E − 11	2437	376

Another example to which the wideband FMM was applied is a horse (Fig. 4), also 50 wavelengths in size. The surface of the horse was divided into 872,694 triangles, on each of which a single node was placed; the size of the smallest triangle was 9.34×10^{-3} wavelengths, and the size of the largest one was 3.27×10^{-1} wavelengths. The results are depicted in Table 4, whose columns have the same meanings as the corresponding columns in Table 2. Again, the FMM for the Laplace potential was applied to the same geometry; the results are shown in Table 5.

The FMM of this paper was also applied to points on the surface of a sphere 50 wavelengths in size. The surface of the sphere was divided into 619,520 triangles (the smallest being 4.91×10^{-2} wavelengths in size, and the largest 6.27×10^{-2} wavelengths), on each of which a single node was placed. The results are tabulated in Table 6; the results for the Laplace potential in the same geometry are shown in Table 7.

Finally, we applied the FMM to points on the surface of a cube 50 wavelengths in size. The surface of the cube was divided into 619,520 triangles (each being 9.12×10^{-2} wavelengths in size), on each of which a single node was placed. The results can be found in Table 8; the results for the Laplace potential in the same geometry are shown in Table 9.

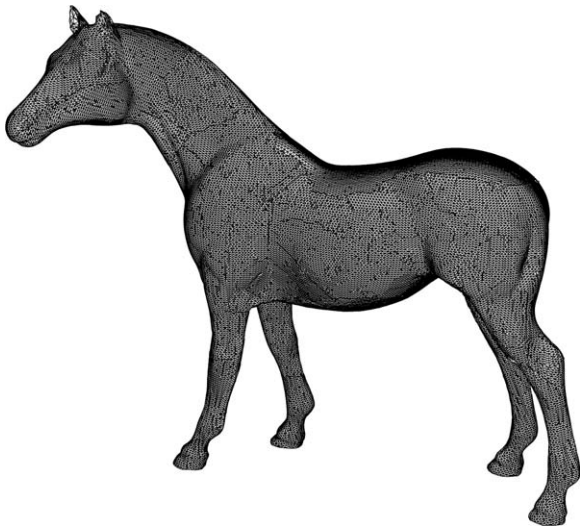

Fig. 4. Surface triangulation of a horse 50 wavelengths in size. The size of the smallest triangle is 9.34×10^{-3} wavelengths, and the size of the largest is 3.27×10^{-1} wavelengths.

Table 4
Example 2: Horse

Time (direct)	Requested accuracy	Error (potential)	Error (gradient)	Time (s)	Memory (MB)
646143	10^{-3}	0.65E − 3	0.31E − 3	672	549
646143	10^{-6}	0.66E − 6	0.92E − 7	1832	1111
646143	10^{-9}	0.33E − 9	0.33E − 11	3515	2027

Table 5
Horse – Laplace potential

Time (direct)	Requested accuracy	Error (potential)	Error (gradient)	Time (s)	Memory (MB)
107833	10^{-3}	0.91E − 3	0.57E − 3	63.7	328
107833	10^{-6}	0.46E − 6	0.31E − 6	139.7	322
107833	10^{-9}	0.25E − 9	0.10E − 9	298	584

Table 6
Sphere

Time (direct)	Requested accuracy	Error (potential)	Error (gradient)	Time (s)	Memory (MB)
324381	10^{-3}	0.27E − 3	0.19E − 3	521	416
324381	10^{-6}	0.15E − 6	0.42E − 7	1358	914
324381	10^{-9}	0.91E − 10	0.24E − 10	2873	1474

Table 7
Sphere – Laplace potential

Time (direct)	Requested accuracy	Error (potential)	Error (gradient)	Time (s)	Memory (MB)
52936	10^{-3}	0.79E − 3	0.90E − 3	45	245
52936	10^{-6}	0.33E − 6	0.45E − 6	97.7	244
52936	10^{-9}	0.19E − 9	0.12E − 9	223	402

Table 8
Cube

Time (direct)	Requested accuracy	Error (potential)	Error (gradient)	Time (s)	Memory (MB)
376950	10^{-3}	0.97E − 3	0.74E − 3	393	364
376950	10^{-6}	0.73E − 6	0.26E − 7	1022	1295
376950	10^{-9}	0.23E − 9	0.17E − 10	2077	1001

Table 9
Cube – Laplace potential

Time (direct)	Requested accuracy	Error (potential)	Error (gradient)	Time (s)	Memory (MB)
56433	10^{-3}	0.94E − 3	0.60E − 3	52	201
56433	10^{-6}	0.41E − 6	0.34E − 6	132	272
56433	10^{-9}	0.28E − 9	0.17E − 9	231	362

The orders of the various expansions used for boxes of various sizes, and for three different levels of requested accuracy ε, are listed in the Tables 10, 11. Table 10 lists the orders of partial-wave expansion used, in the columns labeled "LF", and the orders of truncation of far-field-to-local translation operators (the number k in (48)), in the columns labeled "HF". In each case the number of terms in the expansion is on the order of the square of the order: a partial-wave expansion of order p has $(p+1)^2$ terms; and, for translation, a far-field signature is sampled on a grid of size roughly $2p \times 4p$ (the latter number being generally increased slightly so as to enable FFTs of that size to be performed efficiently), while the tabulated orders of truncation are, at high frequencies, roughly equal to $2p$. When a number is in parentheses (which only occurs in the "LF" column), it indicates that partial-wave expansions of that order are not used in the wideband FMM; instead, the high-frequency version is used.

The numbers of terms used in exponential expansions, for boxes of various sizes, are listed in Table 11. (Each number is slightly larger than strictly necessary, since each of the numbers M_k which are summed to yield each entry in the table is adjusted upwards slightly to be a product of small primes, so that the FFT can be performed efficiently.)

The following observations can be made from the results of the numerical experiments described above, and from the more extensive experimentation we have performed.

1. The observed CPU times are compatible with the $n \cdot \log(n)$ estimate. One apparent exception is the CPU time for the 9-digit calculation for the aircraft-shaped object in Table 3. In fact, the FMM algorithm for the Laplace equation (we used the version from [8]) ran out of physical memory, and we see the CPU time

Table 10
Orders of partial-wave expansions

Box size (wavelengths)	Expansion order					
	$\varepsilon = 10^{-3}$		$\varepsilon = 10^{-6}$		$\varepsilon = 10^{-9}$	
	LF	HF	LF	HF	LF	HF
⩽0.1	10		22		28	
0.25	11	7	23		38	
1	(15)	17	26		41	
2	(22)	32	31		46	
3.5	(31)	50	39	56	52	
5	(40)	68	(48)	77	59	
10	(70)	126	(80)	137	91	
12.5	(85)	155	(95)	166	107	176
15		183		195		206
20		240		253		265
30		352		367		380
40		463		480		494
50		575		593		608

Table 11
Numbers of terms in exponential expansions

Box size (wavelengths)	Number of exponential terms		
	$\varepsilon = 10^{-3}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-9}$
$<10^{-4}$	415	1251	2576
10^{-4}	415	1251	2576
10^{-3}	415	1251	2575
10^{-2}	427	1271	2570
10^{-1}	426	1337	2561
0.5	722	1742	3149
1	898	2134	3646
2	1991	3407	5261
3	3540	5261	7156
4	5487	7720	9767
5	7619	10522	12662
6	10665	12785	16191
7	13860	16363	19499
8	18511	20399	24433
9	23445	24963	28774
10	28032	30071	33815

penalty associated with the use of the virtual memory (the so-called paging). Using a computer with more memory or a more carefully written memory allocation for the Laplace FMM would eliminate the anomaly.

2. For larger boxes, the number of terms in the partial wave expansions (and, consequently, the cost of applying the translation operators) is almost independent of the requested precision (see Table 10); this is a well-known aspect of the behavior of such expansions. The number of terms in the exponential expansions is also remarkably insensitive to the accuracy requirements (see Table 11).
3. In terms of CPU time requirements of the algorithm, it is advantageous to switch to the high-frequency (diagonal) form as soon (for cubes as small) as the accuracy considerations permit. When calculations are conducted in double precision (64 bit) arithmetic, and the answer is desired with three digits, the transition can be made very early (for boxes only 1/4 of a wavelength in size). When 9-digit accuracy is required, the transition is pushed to boxes that are as large as 12 wavelengths. This aspect of the algorithm is the principal reason for the relatively high cost of the scheme when higher accuracy is required.
4. The scheme tends to be about 10 times slower than the scheme for the Laplace equation in the same geometry. A factor of about 3 can be accounted for by the fact that the Laplace code uses predominantly real arithmetic, while the Helmholtz code is mostly complex. The remaining difference is related to the structure of the Helmholtz algorithm, principally at the boundary between the low and the high-frequency regimes.
5. As expected, both the time requirements and the accuracy of the algorithm are fairly insensitive to the nature of the charge distribution. Virtually arbitrary accuracies can be obtained (within the limitations of one's computational environment).
6. The scheme of this paper permits scattering problems involving hundreds of thousands of unknowns on the boundaries of the scatterers to be handled on modern desk-top computers, though the resulting CPU (and more importantly, wall-clock) times can be inconveniently long. Using more powerful (but still very accessible) systems, one can easily solve problems involving many millions of unknowns.

The reader might have observed that in the numerical examples in this paper, the size of scattering objects is limited by about fifty wavelengths. The reason for this limitation is that at about this point, the desktop computer used by the authors (Pentium 4 with 1.5 gigabytes of memory, using double precision arithmetic) tends to run out of memory. With other computer systems, considerably larger-scale scattering problems have been solved (see, for example, [30]).

8. Conclusions

We have presented a wideband version of the Fast Multipole Method for the Helmholtz equation in three dimensions. Although the method has considerable internal complexity, it does not expose that complexity to the user by breaking down in any regime. It has asymptotic CPU time $O(N \log N)$, and, as demonstrated by numerical examples, even at high accuracies delivers very substantial speed increases over the direct method—more than two orders of magnitude—at problem sizes which fit on an ordinary personal computer at the time of this writing.

Acknowledgements

The authors were supported in part by DARPA/AFOSR under the contracts F49620-03-C-0052 and F49620-03-C-0041, and by DARPA under contract HR0011-05-P-0001.

References

[1] M. Abramowitz, I. Stegun (Eds.), Handbook of Mathematical Functions, National Bureau of Standards, 1964.
[2] B. Alpert, G. Beylkin, R. Coifman, V. Rokhlin, Wavelet-like bases for the fast solution of second kind integral equations, SIAM Journal of Scientific and Statistical Computing 14 (1) (1993) 159–184.
[3] G. Beylkin, R.R. Coifman, V. Rokhlin, Fast wavelet transforms and numerical algorithms I, Comm. Pure Appl. Math. XLIV (1991) 141–183.
[4] L.C. Biedenharn, J.D. Louck, Angular Momentum in Quantum Physics: Theory and Application, Addison-Wesley, Reading, 1981.
[5] E. Bleszinsky, M. Bleszinsky, T. Jaroszewicz, AIM: Adaptive integral method for solving large-scale electromagnetic scattering and radiation problems, Radio Science 31 (1996) 1225–1251.
[6] N.N. Bojarski, K-space formulation of the electromagnetic scattering problems, Air Force Avionic Lab. Technical Report AFAL-TR-71-75, 1971.
[7] W.C. Chew, J.M. Jin, E. Michielssen, J. Song (Eds.), Fast and Efficient Algorithms in Computational Electromagnetics, Artech House, Boston, 2001.
[8] H. Cheng, L. Greengard, V. Rokhlin, A fast adaptive multipole algorithm in three dimensions, J. Comput. Phys. 155 (1999) 468–498.
[9] R. Coifman, V. Rokhlin, S. Wandzura, The fast multipole method for the wave equation: a pedestrian prescription, IEEE Antenn. Propagat. Mag. 35 (3) (1993).
[10] H. Cheng, V. Rokhlin, N. Yarvin, Non-linear optimization, quadrature, and interpolation, SIAM J. Optim. 9 (4) (1999) 901–923.
[11] R. Courant, D. Hilbert, Methods of Mathematical Physics, Interscience Publishers, New York, 1953.
[12] E. Darve, The fast multipole method I: error analysis and asymptotic complexity, SIAM J. Numer. Anal. 38 (1) (2000) 98–128.
[13] E. Darve, P. Have, The fast multipole method numerical implementation, J. Comput. Phys. 160 (2000) 195–240.
[14] E. Darve, A fast multipole method for Maxwell's equations stable at all frequencies, Phil. Trans. Roy. Soc. Lond. A 362 (2004) 1–27.
[15] B. Dembart, M. Epton, B.H.L. Fong, L. Greengard, S. Jiang, J.J. Ottusch, V. Rokhlin, J.L. Visher, S.M. Wandzura, Advanced electromagnetic modeling, Final Technical Report, HRL REF K0006, Computational Physics Department, HRL Laboratories, December 2003.
[16] M.A. Epton, B. Dembart, Multipole translation theory for three-dimensional Laplace and Helmholtz equations, SIAM J. Sci. Comput. 16 (4) (1995) 865–897.
[17] L. Greengard, J. Huang, A new version of the fast multipole method for screened coulomb interactions in three dimensions, J. Comput. Phys. 180 (2002) 642–658.
[18] L. Greengard, V. Rokhlin, A fast algorithm for particle simulations, J. Comput. Phys. 73 (1987) 325–348.
[19] L. Greengard, V. Rokhlin, A new version of the fast multipole method for the Laplace equation in three dimensions, Acta Numerica (1997) 229–269.
[20] L. Greengard, J. Huang, V. Rokhlin, S. Wandzura, Accelerating fast multipole methods for the Helmholtz equation at low frequencies, IEEE Comput. Sci. Eng. 5 (3) (1998) 32–38.
[21] T. Hrycak, V. Rokhlin, An improved fast multipole algorithm for potential fields, SIAM J. Scient. Comput. 19 (6) (1998) 1804–1826.
[22] R. Jakob-Chien, B. Alpert, A fast spherical filter with uniform resolution, J. Comput. Phys. 136 (2) (1997) 580–584.
[23] L.J. Jiang, W.C. Chew, A mixed-form fast multipole algorithm, Research Report: CCEM No. 03–05, Center for Computational Electromagnetics and Electromagnetics Laboratory, University of Illinois, 2005; IEEE T. Antenn. Propag. 53 (12) (2005) 4145–4156.
[24] P.M. Morse, H. Feshbach, Methods of Theoretical Physics, McGraw-Hill, New York, 1953.
[25] N. Nishimura, Fast multipole accelerated boundary integral equation methods, Appl. Mech. Rev. 55 (4) (2002).
[26] V. Rokhlin, Rapid solution of integral equations of scattering theory in two dimensions, J. Comput. Phys. 86 (1990) 414–439.
[27] V. Rokhlin, Diagonal Forms of Translation Operators for the Helmholtz Equation in Three Dimensions, Applied and Computational Harmonic Analysis, vol. 1, Academic Press, San Diego, 1993, pp. 82–93.

[28] J.M. Song, W.C. Chew, Multilevel fast multipole algorithm for solving combined field integral equations of electromagnetic scattering, Microwave Opt. Technol. Lett. 10 (1) (1995).
[29] J. Stoer, R. Bulirsch, Introduction to Numerical Analysis, second ed., Springer, Berlin, 1993.
[30] S. Velamparambil, W.C. Chew, J.M. Song, 10 million unknowns, is it that large, IEEE Antenn. Propagat. Mag. 45 (2) (2003) 43–58.
[31] N. Yarvin, V. Rokhlin, An improved fast multipole algorithm for potential fields on one-dimensional structures, SIAM J. Numer. Anal. 36 (2) (1999) 629–666.
[32] L. Ying, G. Biros, D. Zorin, A kernel-independent adaptive fast multipole algorithm in two and three dimensions, J. Comput. Phys. 196 (2004) 591–626.
[33] E. Yip, B. Dembart, in: 13th Annual Review of Progress in Applied Computational Electromagnetics at the Naval Postgraduate School, Monterey, CA, March 17, 1997.