The Block-P$^3$M Algorithm for Fast Integral Equation Solvers

Abstract — The Block-P$^3$M solver calculates action-at-a-distance convolution sums that arise in electromagnetics and, unlike the plane-wave fast multipole method (FMM), it does not suffer from low-frequency breakdown. The solution uses $O(N \log N)$ operations and $O(N)$ storage for surface as well as volume distributions of sources. It differs from single-mesh P$^3$M [1, 2] primarily in that sparse meshes are used and the mesh convolution sum is evaluated using Fast Fourier Transforms on a nested hierarchy of blocks. Block-P$^3$M gives optimal accuracy, and is applicable from statics to high frequency electromagnetics.

I. INTRODUCTION

This article describes the Block-P$^3$M generalisation of the P$^3$M algorithm. The work was motivated by the needs of large Method of Moments (MoM) calculations, but is also applicable to low frequency electromagnetic calculations that use integral formulations. Block-P$^3$M has advantages over multi-level fast multipole (MLFMA) and multigrid algorithms [3, 4, 5].

Block-P$^3$M evaluates the convolution sum

$$\phi_i = \sum_{j=1}^{N} G(x_i; x_j) \sigma_j \quad \forall i \in [1, N]$$

in $O(N \log N)$ operations using $O(N)$ storage. The distribution of sources $\{\sigma_j\}$ may be on surfaces or throughout volumes and the displacement-invariant Green’s function $G(x_i; x_j)$ may take any functional form, for example, the Helmholtz kernel (for scattering) or the Laplace kernel (for potential fields).

Block-P$^3$M is used in solving action-at-a-distance integral formulations of the equations of electromagnetism. Fields described by the differential equation

$$L \phi = \sigma$$

where $L$ is the differential operator, $\phi$ is the field and $\sigma$ is the source are equivalently given by the integral equation

$$\phi(x) = \int G(x; x') \sigma(x') dx'$$

where $x$ is the field point, $x'$ is the source point and the integral is over all volumes containing sources. The Green’s function satisfies the differential equation

$$L G = \delta(x - x')$$

In applications where Block-P$^3$M is applicable, the boundary conditions are such that $G$ takes either a displacement invariant form $G = G(r)$ where $r = |x - x'|$, or may be written as the sum of displacement invariant functions.

The simplest instance of action-at-a-distance formulation is in electrostatics, where Eq. (2) is Poisson’s equation, Eq. (3) is Coulomb’s Law and the Green’s function becomes the Laplace kernel $G = 1/4 \pi r \epsilon_0 |r|$. In magnetostatics, the differential Ampere’s Law can be transformed to the integral Biot-Savart Law. Maxwell’s equations have action-at-a-distance equivalents of the Liénard-Weichert equations in the time domain and the Stratton-Chu equations for time harmonic signals [6].

Historically, computational complexity has favoured the differential equations as the starting point for numerical computations. These are approximated using volume-filling finite difference or finite element nets within a computational volume on which boundary conditions are applied. The resulting sparse matrix equations are then solved using either direct or iterative solvers. For example, given a volume distribution of $N$ charges, rapid elliptic solvers can find potentials in $O(N \log N)$ operations [2, Chap. 6], whereas a direct Coulomb sum takes $O(N^2)$. Fast solvers such as P$^3$M [1, 2] can reduce the complexity of the Coulomb sum to better than $O(N \log N)$, making the action-at-a-distance approach faster than differential equation approaches for large problems. The Block-P$^3$M method applied to the Helmholtz kernels that arise in MoM again gains complexity gains over volume mesh methods. Fast solvers plus the absence of body conforming volume meshes makes integral equation methods the preferred choice for complex geometries, especially if there are moving conductors.

A. Background

All fast solver methods for evaluating the so-called ‘N-body’ problem (Eq. (1)) rely on

1. decomposing the Green’s function into the sum of short-range and long-range parts;
2. explicitly evaluating the convolution sums for the short-range parts;
3. approximately evaluating the long-range parts on meshes.

P$^3$M variants use continuous finite element or spline approximations of the Green’s function, whereas the fast multipole methods use discontinuous point expansions about cell centres.

The P$^3$M method[1, 2] uses additive splitting of the Green’s function into a short-range part and a part that can be accurately approximated on a mesh. It works for any displacement-invariant kernel, provided that the mesh spacing is smaller than the scale-length of the long-range Green’s function. The Laplace Green’s function is scale-free, and this has been successfully exploited by Couchman in developing the Adaptive-P$^3$M algorithm [7, 8]. He showed that extra speedup is obtained by recursively applying P$^3$M over reducing scale lengths in astrophysical problems where there are large density contrasts. The scale-free nature of the Laplace kernel is also key to multigrid methods (c.f. Section III.).

Rokhlin [9] introduced the FMM for astrophysical $N$-body calculations. The FMM relies on analytic expansions that allow the long-range Green’s function to be extrapolated from lattice points using spherical wave expansions [10]. Recursive application of the extrapolation on increasing scale-lengths led to the tree-based MLFMA [4, 5]. Recently, Dehnen has shown that replacing the spherical-wave expansion by a Cartesian Taylor expansion offers further speedup gains [11].
The FMM was extended to electromagnetic scattering [12, 13] by replacing the spherical-wave expansion by a plane-wave expansion. Fast multipole methods have been the subject of intensive research over the past decade. The multi-level fast multiple algorithm further reduces the complexity of the FMM.

Another fast method for scattering is the Adaptive Integral Method (AIM) [14, 15]. This uses a multipole expansion on a lattice, and like P³M uses Fast Fourier Transforms (FFTs) to evaluate the long-range part of the interaction. The operations count scaling for AIM (and for single-mesh P³M) degrades to \( O(N^{3/2} \log N) \) for surface distributions of sources.

B. P³M and MoM

In MoM, the Helmholtz kernel \( G \propto \exp(-ik_0|x|)/|x| \) has scale length \( \lambda = 2\pi/k_0 \) and this limits to low frequency the use of fast methods that rely on the Green’s function being scale-free.

The quadrature points for the integrals of MoM play the same role as ‘particles’ in particle applications of P³M. For example, the scattered field in the electric field integral equation (EFIE) \( \mathbf{E}^s = -\nabla \phi - i\omega \mathbf{A} \) is evaluated using the integral equations for \( \phi \) and \( \mathbf{A} \). Currents are approximated by finite elements on the elements and integrals are evaluated using quadrature, allowing the approximations for \( \phi \) and \( \mathbf{A} \) to be written in the form of Eq. (1)

\[
\phi(x) = \frac{1}{\mu} \sum_j G(x; x_j)Q_j; \quad \mathbf{A}(x) = \mu \sum_j G(x; x_j)\mathbf{I}_j
\]

where \( Q_j \) and \( \mathbf{I}_j \) are the charge and current associated with quadrature point \( j \).

The building blocks for a fast P³M-MoM solver are MoM integrations routines to evaluate sources and fields at quadrature points, a Krylov solver [16] and a Block-P³M solver. These allow fast computation of the voltage excitation vector \( \mathbf{V} \) and the product \( \mathbf{Z}I \) in the iterative solution for the current vector \( I = \mathbf{Z}^{-1}\mathbf{V} \).

Block-P³M for MoM uses a Helmholtz kernel with the singularity removed, and is used in the following steps for the fast computation of the \( \mathbf{Z}I \) product:

- map MoM currents to the quadrature points;
- use Block-P³M to find fields (c.f. Eq. (1));
- add in singular integral corrections;
- map fields at quadrature points to the MoM basis functions;
- sum field contributions to obtain the \( \mathbf{Z}I \) product vector.

II. SINGLE MESH P³M

The original P³M method used a mesh Poisson solver with short-range corrections [17]. Eastwood [1, 2] generalised and optimised it by introducing the FFT solver, Green’s function formulation, spline function assignment/interpolation and least-squares optimisation. P³M using point collocation is also known as the PME (Particle-Mesh Ewald) method [18].

The Particle-Particle/Particle-Mesh (P³M) method splits the potential and Green’s function into a short-ranged (P-P) part and a smoothly varying (P-M) part

\[
\phi_i = \phi^\text{sr}_i + \phi^\text{mesh}_i; \quad G = G^\text{sr} + G^\text{mesh}
\]

\( \phi^\text{sr}_i \) is evaluated using direct pairwise sums over the set \( S_i \) of near neighbours of \( i \)

\[
\phi^\text{sr}_i = \sum_{j \in S_i} G^\text{sr}(x_i - x_j)\sigma_j; \quad i \in [1, N]
\]

If the spectral content of \( G^\text{mesh} \) is negligible beyond wavenumber \( k_\infty \), then by the sampling theorem [19] it can be accurately represented by a sample of nodal values at an interval \( \Delta \leq \pi/k_\infty \). This allows \( \phi^\text{mesh} \) to be accurately computed using the FFT on a mesh. \( G^\text{mesh} \) is approximated at general points using finite elements:

\[
G^\text{mesh}(x_i - x_j) \approx W_p(x_i)G^\text{fc}(x_p - x_j)W_{p'}(x_j)
\]

Nodal amplitudes \( G^\text{fc} \) are precomputed using a Galerkin (or point collocation) method and their FFTs \( G^k \) are stored. The use of the FFT on a regular lattice favours the use of spline functions for \( W_p \) rather than the more common finite element choices of Lagrange or Hermite polynomials. The P-M calculation reduces to

\[
\text{assign:} \quad \sigma_{p'} = \sum_j W_p(x_j)\sigma_j \quad (9)
\]

\[
\text{solve:} \quad \phi_p = \sum_{p' = 0}^{P-1} G_{p-p'}^\text{fc} \sigma_{p'} \quad (10)
\]

\[
\text{interpolate:} \quad \phi^\text{mesh}_i = \sum_p W_p(x_i)\phi_p \quad (11)
\]

where \( p \) and \( p' \) label mesh nodes and \( i \) and \( j \) are quadrature points in the MoM model. Figure 1 illustrates the assign and interpolate steps for quadratic splines in 2D. In 3D, the currents from the quadrature points are assigned to the nearest 27 mesh points, and fields are interpolated from the nearest 27 mesh points. Cubic splines use 64 points, and order \( p \) splines use the nearest \( (p + 1)^3 \) points. The correction for the smoothing effect of the spline functions is incorporated in the FFT solve.

The mesh convolution sum (Eq. (10)) is evaluated using FFTs; the FFT of source values \( \sigma_p \) gives harmonics \( \sigma^k \). Multiplication by Green’s function harmonics \( G^k \) gives potential harmonics \( \phi^k = G^k\sigma^k \), and the inverse FFT converts these to values \( \phi_p \). If periodic boundary conditions are used, Eq. (10) can be transformed using a length \( P \) FFT and short range contributions from periodic images need adding to the fields. If isolated boundary conditions are required, then zero padding sources and FFT lengths greater than \( 2P - 1 \) [20] are needed to diagonalise Eq. (10). The field \( \phi_i \) is given by combining the contributions of the P-P and P-M calculations.

P³M in two and three dimensions is given by interpreting \( x \) and \( p \) as 2D or 3D vectors and indices, and \( W_p \) as products of 1D polynomial functions [2]. For spatially uniform distributions of sources, the operations count scales as \( O(N\sqrt{\log N}) \), but this increases to \( O(N^{3/2}\log N) \) for surface distributions in a 3D volume. The Block-P³M algorithm described below recovers the more favourable scaling for non-uniform and surface distributions that are encountered in MoM and other integral equation applications.

![Figure 1: The assign-solve-interpolate steps of the mesh calculation.](image-url)
P³M avoids low-frequency breakdown by using interpolation for approximating the Green’s function, and this should also lead to better accuracy at a given order. Figure 2 gives a low-order 1D illustration. The finite element approximation with $\mathcal{O}(p)$ spline functions $W_p$ gives continuity of value and derivatives to $\mathcal{O}(p-1)$, whereas FMM uses independent power expansions about each cell centre. In Figure 2 the exact $\cos(r)/r$ Green’s function is approximated using 2 degrees of freedom in each 0.1 interval using linear splines (as in P³M) or point expansions about cell centres (as in FMM). The accuracy of the latter degrades rapidly with decreasing $r$.

![Figure 2: A 1D illustration of the FMM and P³M representations of the Green’s function. See text for details.](image)

### III. Recursive Interpolation

If the scale length of the Green’s function is much larger than the size of the mesh, then the assign and interpolate steps can be repeated to reduce the size of the solve mesh.

Recursive interpolation repeats the decomposition of the convolution as described above in Eqs. (6)-(11). Eq. (10) has the same form as the original problem, Eq. (1), but now for sources on the mesh. It can be decomposed in the same way as the original problem but on a coarser mesh. For the Laplace kernel, this decomposition is limited to cases where the mesh size is less than $\lambda$. The next section describes how to overcome this limitation.

The recursion can be terminated at any level, leaving the solve step at the coarsest mesh level to be completed as for the single-mesh P³M algorithm.

### IV. Block-P³M

Recursive interpolation decomposes the solve step by repeating the short-range/long-range split of the Green’s function and the assign/solve steps on increasingly coarse spatial meshes. A key feature of Block-P³M is to decompose the solve step by repeatedly grouping the mesh into blocks but now applying the FFT solve only to the non-empty subset of blocks.

Block decomposition splits 1D into 2D convolution sums (i.e., 3D to 6D in $\mathbb{R}^3$) and introduces inter-block propagators $H$. Given Eq. (10), where $P = QR$, we can set $p = q + qr$ and rewrite it as a 2D convolution

$$
\phi_{q,r} = \sum_{r'=0}^{R-1} \sum_{q'=0}^{Q-1} G_{q'-r',q+r'} \sigma_{q',r'}
$$

(12)

The convolution over index $q$ is not periodic. However, zero padding [20] with $1 + w/Q$ zeroes (to give $\rho$) and setting $H_{q,r} = F_q G_{q,r}$ gives a periodic convolution of period $(2 + w)Q$ that yields $M = (2 + w)Q$ values of $\psi_{q,r}$ where $\phi_{q',r} \equiv \psi_{q,r}$ for $q \in [0, Q - 1]$.

$F_q$ is a filtering function which is 1 for $|q| < Q$ and goes to zero over the interval $Q \leq |q| < (1 + w)Q$. If the padding width $w$ is zero, then $F$ is a rectangular window of width $2Q$, but there are advantages in taking $w \approx 1$ and using smooth transitions such as cosine bells to reduce spectral infilling.

The periodic convolution (period $M = (2 + w)Q$) and its Fourier transform over index $q$ are

$$
\psi_{q,r} = \sum_{r'=0}^{R-1} \sum_{q'=0}^{Q-1} H_{q'-r',q+r'} \rho_{q',r'}
$$

(13)

$$
\psi^k_r = \sum_{r'=0}^{R-1} H^k_{r'-r',r'} \rho^k_r
$$

(14)

where $k$ is the harmonic index corresponding to $q$, and $q$ indices are modulo $M$. Any block $r$ where $\rho_{q,r} = 0$ for all $q$ is ignored.

Inspection of Eq. (14) shows that the original problem (Eq. (10)) has been recovered, but now for different harmonics $k$ on the large blocks with indices $r$. These harmonics are split into three groups, depending on the range of $H^k_{r'}$. If $|H^k_{r'}|$ are below the $k$-space truncation amplitude they are ignored, if $H^k_{r'}$ are short ranged (i.e., non-zero only for small $x$) so that direct evaluation is faster than FFT evaluation, Eq. (14) is evaluated by direct summation, otherwise a further level of decomposition is used, setting $R = ST$ and repeating the above sequence.

At the final level, all remaining harmonics are short ranged. The computation of the potential is completed by repeatedly computing inverse FFTs (IFFTs) at each level and combining the results with the short-range contributions until the first level is reached.

#### A. One-level algorithm

The P-P part of the one-level Block-P³M algorithm is a direct application of Eqs. (13)-(14). Figure 3 summarises the algorithm. If $R = 1$, it reduces to single-mesh P³M described above in Sec. II, but in general covers the situation with $R > 1$ illustrated in Figure 4.

In Figure 4, the surface distribution of sources is represented by the oval curve. The computational region containing the curve is divided into coarse blocks $r$ and these in turn are subdivided into fine blocks $q$. Data are stored only for non-empty (shaded) blocks. Starting with source $\sigma_j$, the direct P-P contribution to $\phi_j$ is given by Eq. (7), and mesh sources $\sigma_{qr}$ are found by assignment to the fine blocks in each coarse block. Each non-empty coarse block is zero-padded and transformed to give $\rho^k_r$. Direct evaluation of Eq. (14) followed by inverse Fourier transformation gives $\psi^k_r$, and this is interpolated to test point positions to give $\phi^\text{mesh}_j$. If Eq. (14) is evaluated by padding and FFT, then it would become a two-level scheme.

A naive estimate of the operation count for summing Eq. (14) is $O(N_{p,r}^2 N_k)$, where $N_p$ is the number of non-empty coarse blocks and $N_k$ is the number of $k$ values. However, the nature of the inter-block propagators reduces this to close to $O(N_p)$.

Figure 5(a) shows an example of contours of constant real part of $H_{qr}$ that arise from the filtered section of the Helmholtz Green’s function; the larger box is the range over which it is non-zero, and the smaller is the FFT periodic box. Figure 5(b)
shows the 1% of peak level contour of amplitude of the transformed propagator $H^k_r$, shaded by $\Re(H^k_r)$. The spectral content is strongly peaked about $|k| = k_0$ along the vector direction between the source and test blocks, where $k_0$ is the signal wavenumber. For larger $r$, $H_{qr}$ becomes closer to a plane wave pulse, and so its transform becomes even more localised. For $|k|$ away from $k_0$, $H^k_r$ are only non-negligible for near-neighbour blocks. These properties substantially reduce the operations count for evaluating the sum Eq. (14). If blocks are larger in terms of wavelengths, the peaking of the propagator transform is even more pronounced.

The benefit of using a filter width $w > 0$ is that spectral infilling from the sharp cutoff of $H$ is avoided, and so more inter-block propagator harmonics can be discarded. For example, using a cosine bell truncation with $w = 1$ for blocks of $32^3$ cells with cell width of $\lambda/8$ and a threshold of $10^{-4}$ peak amplitude for discarding harmonics retains less than 3% of harmonics for blocks beyond second neighbour and gives an rms error $\approx 10^{-5}$, a threshold of $10^{-5}$ retains $< 0.5\%$ and gives an error $\approx 10^{-3}$.

**B. Multi-level algorithm**

In multilevel Block-P$^3$M the assign and interpolate steps may be single level (c.f. Sec. II) or multilevel (c.f. Sec. III). The solve step goes through $L$ levels of blocks computing harmonics from the smallest blocks at level 1 to the largest at level $L$, and then from largest to smallest computing field values. The one-level case described in the previous section generalises to:

For level 1 to $L$:

1. For all harmonics at the current level where $G$ is short ranged, directly evaluate the sum, Eq. (10);
2. Terminate if all harmonics have been processed, otherwise, group sub-blocks into blocks of side $Q$, yielding the convolution sum, Eq. (12);
3. Zero-pad and FFT the finer mesh index $q$ for non-empty $r$ blocks giving for each harmonic $k$ Eq. (14);

**Figure 3:** The one-level Block-P$^3$M algorithm.

**Figure 4:** One-level Block-P$^3$M blocks.

**Figure 5:** (a) the real part of the level 1 propagator for $r = (2, 2, 0)$ and (b) the 1% peak amplitude contour of the transform of the propagator for coarse blocks of side one wavelength.

**Figure 6:** Flow diagram for three-level Block-P$^3$M.
4. Set $r \rightarrow p$, $R \rightarrow P$, $\psi \rightarrow \phi$, $H \rightarrow G$, $\rho \rightarrow \sigma$ in Eq. (14) to recover Eq. (10) and go to the next level.

For level $L$ to 1:

1. IFFT harmonics to get field values at the lower level;

2. discard field values for padding index values;

3. combine with directly-summed terms at this level.

Figures 6 and 7 illustrate a three-level scheme for the mesh field calculation with single assign/interpolate levels. Figure 6 gives the flow diagram and Figure 7 illustrates the steps for a real MoM model [21]. Sources at quadrature points (Figure 7(a)) on the MoM model are assigned to the fine blocks (Figure 7(b)) to give $\sigma_{pqr}$. These are grouped in the medium-level blocks (larger blocks in Figure 7(b)), and the data are Fourier transformed to give $\sigma_{kqr}^{i}$ on the medium blocks. Zero-padding is used to avoid aliasing. The result is a mesh of values for each retained $k$ on the finer mesh in Figure 7(c).

For wavenumbers $k$ where $H_{qr}^{k}$ is short-ranged, the convolution is summed directly, and where it is long-ranged, the data are Fourier transformed to give $\sigma_{q}^{k}$ on the coarse mesh. The result is a mesh of values for each retained $kl$ on the mesh in Figure 7(d). The same procedure as is used for lower levels is followed to get $\sigma_{klm}$ on the bounding box.

The direct sum on the bounding box gives $\phi_{klm}^{r}$. This is inverse transformed and combined with the directly calculated $\phi_{kl}^{r}$. The process is repeated to get $\phi_{qr}^{k}$ and then $\phi_{pqr}$ and finally the result is interpolated to the test point positions.

Cubic blocks are used to allow storage saving for the precomputed filtered Green’s functions by exploiting the 48-fold symmetry:

$$H(i) \equiv H(\text{Perm}\{\pm i_x, \pm i_y, \pm i_z\})$$

where indices $i = (i_x, i_y, i_z)$ are triplets of $pqr$, $kqr$, $klr$ or $klm$ depending on the level.

V. FINAL REMARKS

The Block-P$^3$M algorithm overcomes the problem of the reduced performance of the original P$^3$M algorithm for non-uniform source distributions that are typically found in surface integral calculations. It gives $O(N \log N)$ complexity and $O(N)$ storage for non-uniform distributions of sources and for any displacement-invariant Green’s function.

The complexity for surface distributions of sources and test points is achieved through avoiding empty blocks at all levels and through discarding harmonics where the amplitudes of the propagator harmonics are below the chosen threshold level. Direct summation of convolutions is used where the range of the propagator is small, and Fourier transforms are used where the FFT is computationally faster. Storage reduction is achieved by storing only blocks that are non-empty. When blocks do contain non-empty sub-blocks, then sparse storage methods are used to store only those sub-blocks. Computations and storage needed...
for the precomputed inter-block propagators are reduced by exploiting 48-fold symmetry on cubic lattices and by storing only those harmonics of the propagators with amplitudes above the cutoff threshold.

The algorithm is built on FFTs on uniform lattices and works equally well for Laplace and Helmholtz kernels. Unlike FMM, it does not rely on truncated divergent series, and so does not suffer breakdown when point separations become small compared to the wavelength.

If a Galerkin approximation is used in Block-PMP then it gives optimal least-squares accuracy. The trade-off between computational cost and accuracy remains optimal even when \( k \)-space truncation is used. How to make the best choice for combining the recursive interpolation, recursive block decomposition and \( k \)-space truncation is currently being investigated. However, Block-PMP is expected to provide a competitively fast method that exploits optimised FFT routines and is amenable to parallelisation.

Block-PMP is also of value for volume-filling source distributions. The block decomposition offers a method for breaking a global mesh in large computations on distributed-memory parallel computers into a set of spatially-localised blocks on which the FFT can be used without communication between processors. The lower levels of the calculation are embarrassingly parallel, and higher levels involve few harmonics so there is a much reduced interprocessor communication when compared to applying the FFT to the global mesh.

The Block-PMP algorithm is particularly attractive for electromagnetic scattering calculations using the method of moments. It has the same complexity and storage scaling as the MLFMA, but superior properties: MLFMA has a point-matched discontinuous approximation to \( G \) that is recursively extrapolated using a mixture of rectangular and polar meshes and suffers from low-frequency breakdown. In contrast, Block-PMP has a least-squares fit, continuous (to the order of the splines used) approximation to \( G \) that is recursively interpolated on rectangular meshes, and does not suffer low-frequency breakdown.

VI. References


Authors Name and Affiliation

James Eastwood, Culham Electromagnetics Ltd, Culham Science Centre, Abingdon, Oxfordshire, OX14 3DB, UK

Telephone: +44 1865 408236
email: James.Eastwood@CulhamEM.co.uk