Particle mesh multipole method: An efficient solver for gravitational/electrostatic forces based on multipole method and fast convolution over a uniform mesh

Keigo Nitadori

Co-Design Team, Exascale Computing Project,
RIKEN Advanced Institute for Computational Science,
7-1-26, Minatojima-minami-machi, Kobe, Japan

Abstract

We propose an efficient algorithm for the evaluation of the potential and its gradient of gravitational/electrostatic N-body systems, which we call particle mesh multipole method (PMMM or PM^3). PMMM can be understood both as an extension of the particle mesh (PM) method and as an optimization of the fast multipole method (FMM). In the former viewpoint, the scalar density and potential held by a grid point are extended to multipole moments and local expansions in $(p+1)^2$ real numbers, where p is the order of expansion. In the latter viewpoint, a hierarchical octree structure which brings its $O(N)$ nature, is replaced with a uniform mesh structure, and we exploit the convolution theorem with fast Fourier transform (FFT) to speed up the calculations. Hence, independent $(p+1)^2$ FFTs with the size equal to the number of grid points are performed.

The fundamental idea is common to PPPM/MPE by Shimada et al. (1993) and FFTM by Ong et al. (2003). PMMM differs from them in supporting both the open and periodic boundary conditions, and employing an irreducible form where both the multipole moments and local expansions are expressed in $(p+1)^2$ real numbers and the transformation matrices in $(2p+1)^2$ real numbers.

The computational complexity is the larger of $O(p^2N)$ and $O(N\log(N/p^2))$, and the memory demand is $O(N)$ when the number of grid points is $\propto N/p^2$.

Keywords: particle mesh method, fast multipole method, fast Fourier transform, Ewald summation, molecular dynamics

Email address: keigo@riken.jp (Keigo Nitadori)

1. Introduction

Particle mesh (PM) methods [including the particle–particle particl–mesh (PPPM or P^3M) method (Hockney and Eastwood, 1988), and the particle mesh Ewald (PME) method (Darden et al., 1993)], and the fast multipole method (FMM) by Greengard and Rokhlin (1987, 1988b) have been developed to speed up the calculation of long range forces in particle simulations. An efficient unification of these two methods was accomplished by Shimada, Kaneko and Takada (1993) as a multipole expansion version of PPPM (PPPM/MPE) to improve the accuracy in periodic boundary conditions. Later, they demonstrated the performance advantage over PPPM/MPE to the FMM in low-order cases (Shimada et al., 1994). Their multipole formulation was based on direct Cartesian gradients where the number of multipole terms scales as p^3 for the order of expansion p. A version with an irreducible form based on spherical harmonics where the number of terms is $(p + 1)^2$ was introduced by Ong et al. (2003) for open boundary conditions, as fast Fourier transform on multipoles (FFTM).

These hybrid schemes have several practical advantages over both the PPPM and FMM. Compared to PPPM, fairly high accuracy can be easily achieved. Since a grid point holds the information of charge distribution and the potential field as multipole moments and local expansions, a particle needs to interact only with the nearest grid point. In the PM methods, interaction with the nearest $(p + 1)^3$ grid points is necessary. Moreover, relatively coarse mesh can be used and multiple FFTs with small size can be performed independently instead of one big FFT. This is suitable for cache based and distributed parallel computers of today. For the particle–particle interaction part, a simple Newtonian force can be used. A cutoff function, which requires an expensive table look-up or mathematical functions, is not necessary. The advantage over FMM is the arithmetic operation cost. In the case of the periodic boundary, the hybrid scheme requires only one multipole-to-local (M2L) transformation per cell (a factor 8 overhead exists, however, for open boundary systems). In the octree-based FMM, each cell requires 189 transformations when the minimum separation between cells is set to one cell size (27 cells cutoff), and it increases to 875 when the minimum separation is set to two (125 cells cutoff) for better accuracy. In the hybrid scheme, the number of M2L transformations remains constant irrespective of the cell separation criterion, and this relaxes the motivation to use very high order of expansion p. Thus, the M2L transformation is usually not the bottleneck. Another practical advantage is that any

products of three integers can be used for the number of cells, while it is usually limited to powers of 8 in FMM.

Despite the above benefits, the hybrid approach of PM and FMM does not seem to be widely used in the particle simulation field, though similar approaches are used in other fields (Sezai et al., 2007; Hesford and Waag, 2010). One reason might be that the formulation is complicated, and another is that algorithms in previous studies are either sub-optimal (not irreducible) or limited in applicability (e.g. only for open boundary). In this paper, we present an optimal and general scheme based on this approach, which we named particle mesh multipole method (PMMM) [1]. In PMMM, a simple solid harmonics notation (Wang and LeSar, 1996) is employed for the multipole transformations, and both multipole moments and local expansions are expressed in $(p+1)^2$ real numbers. A matrix for a $\mathbb{R}^{(p+1)^2} \to \mathbb{R}^{(p+1)^2}$ transformation has effectively $(2p+1)^2$ real numbers, not $(p+1)^4$. In this way, the computational cost and memory requirements are minimized. As far as we know, ours is the first implementation that supports a periodic boundary condition and an irreducible form generalized to higher orders.

Throughout this paper, N is referred to as the total number of particles, K the total number of cells or grid points, and p the order of multipole expansions. We may assume $K \sim N/p^2$ for the optimum value, however, we leave N and K as separate parameters for convenience.

This paper is organized as follows. In section 2, we present the algorithm in detail. In section 3, we discuss the computational complexity of PMMM for a given set of parameters, and give a guide to chose the parameters. A test of numerical accuracy is carried out in section 4. Finally in section 5, we discuss a possible parallelization and hierarchical version of PMMM, and applications to classical simulation of biomolecular systems and cosmological N-body simulations. The appendix supplies some materials useful for implementation.

2. Construction

2.1. Algorithm in detail

Consider N particles distributed in K uniform cells. For each cell, the set of particles inside is known. This condition is achieved in an $O(N)$ procedure. The multipole moments of each cell in $(p+1)^2$ real numbers are evaluated with

[1] A scheme named pseudoparticle multipole method (PPMM or P^2M^2) exists as well (Makino, 1999).

(A.15). Let us now express the multipole moments of cell $\boldsymbol{i} = (i_x, i_y, i_z)$ as a vector $\boldsymbol{M_i} \in \mathbb{R}^{(p+1)^2}$ where $\boldsymbol{i} = (i_x, i_y, i_z) \in \mathbb{Z}^3$ are three dimensional indices of the cell. The local expansions of each cell $\boldsymbol{L_i} \in \mathbb{R}^{(p+1)^2}$ is available in

$$\boldsymbol{L_i} = \sum_{\boldsymbol{j}} \mathsf{G}_{\boldsymbol{i}-\boldsymbol{j}} \boldsymbol{M_j}. \tag{1}$$

Here, $\mathsf{G}_{\boldsymbol{i}-\boldsymbol{j}}$ is a square matrix of size $(p+1)^2$, which depends on the displacement vector $\boldsymbol{i} - \boldsymbol{j}$ of cells $\boldsymbol{i}$ and $\boldsymbol{j}$, and also referred to as a Green's function. In the summation, $\boldsymbol{j}$ iterates over all K cells, and the indices $\boldsymbol{i} - \boldsymbol{j}$ are cyclic. An explicit form of $\mathsf{G}_{\boldsymbol{i}-\boldsymbol{j}}$ is given by (A.17), (A.27), and (B.5), and it effectively consists of $(2p+1)^2$ real numbers, not $(p+1)^4$ numbers.

The calculation cost of the summation can be reduced from $O(K^2)$ to $O(K \log K)$ using a convolution theorem with FFT. The following gives the final procedure for periodic systems:

$$\begin{aligned} \{\widetilde{\mathsf{G}}_{\boldsymbol{k}}\} &:= \mathcal{F}_{\boldsymbol{i}\to\boldsymbol{k}}\{\mathsf{G}_{\boldsymbol{i}}\}, \\ \{\widetilde{\boldsymbol{M}}_{\boldsymbol{k}}\} &:= \mathcal{F}_{\boldsymbol{i}\to\boldsymbol{k}}\{\boldsymbol{M_i}\}, \\ \widetilde{\boldsymbol{L}}_{\boldsymbol{k}} &:= \widetilde{\mathsf{G}}_{\boldsymbol{k}} \widetilde{\boldsymbol{M}}_{\boldsymbol{k}}, \\ \{\boldsymbol{L_i}\} &:= \mathcal{F}^{-1}_{\boldsymbol{k}\to\boldsymbol{i}}\{\widetilde{\boldsymbol{L}}_{\boldsymbol{k}}\}. \end{aligned} \tag{2}$$

Here, $\{\,\}$ denotes the set of all K points, $\mathcal{F}$ and $\mathcal{F}^{-1}$ are forward and backward discrete Fourier transforms, tilde is the value in wave space, and there are wave-number indices $\boldsymbol{k} \in \mathbb{Z}^3$. The first line requires $(2p+1)^2$ Fourier transforms of size K which can be performed and saved at the beginning of the simulation. For the second and the fourth equation, we perform $(p+1)^2$ independent FFTs for each. And in the third, we perform element-by-element M2L transformations. Strictly speaking, the effective number of points in complex numbers after the FFT of K real numbers is $K/2$. And in the M2L transformations in the wave space, all the real operations in (A.27) are turned into complex operations which are expected to be four times more expensive. In total, the cost is equivalent to about $2K$ transformations in real numbers. For three-dimensional open boundary systems, this convolution procedure need to be performed on $8K$ points.

After the local expansions of each cell are obtained, the potential of a particle is available in (A.18), and its gradient in (A.19) and (A.21).

A cutoff for the nearest 27 or 125 cells is expressed in a mask in Green's function and contributions from the masked cells are evaluated in direct particle–particle interactions. An example layout of Green's function is shown in Fig. C.1

for a two-dimensional open boundary system. In a periodic system, the value of Green's function for the closest interactions is not exactly zero, and has the contributions from mirror images.

2.2. Extension of the PM method

At the limit of the spatial order $p = 0$, PMMM agrees with the PM method in the nearest grid point (NGP) mode, where equation (1) reduces to a scalar equation

$$\Phi_{\boldsymbol{i}} = \sum_j G_{\boldsymbol{i}-j}\rho_j, \tag{3}$$

with a discreet scalar potential and density field $\Phi_{\boldsymbol{i}}$ and ρ_j, and Green's function $G_{\boldsymbol{i}-j}$. From this baseline, PM and PMMM increase the spatial order in different ways. The PM method increases the order with a diffusive interaction between a particle and its nearest $(p+1)^3$ grid points. For $p = 0, 1, 2$, they are called NGP, CIC (cloud in cell), and TSC (triangular shaped cloud) mode (Hockney and Eastwood, 1988), and higher order generalization is given in B-spline functions. In PMMM, a particle interacts with the nearest grid point while a grid point holds multiple information in $(p+1)^2$ terms. From the PM method, scalar density is extended to multipole moments, scalar potential is extended to local expansions, and the scalar Green's function is extended to matrix form.

The particle–particle interaction takes different forms. In the PM series, it is sometimes omitted to give a mesh softening (pure PM), or it has a cutoff function so as to make the total force Newtonian (PPPM and PME). In PMMM, the short range cutoff of the mesh part is expressed by the nearest cells as a mask of the Green's function, and the particle-particle interaction takes a pure Newtonian form.

2.3. Optimization from FMM

At the level of the smallest cells, PMMM agrees with FMM. The relative positions and charges (masses) of particles are assigned to the cell center as multipole moments, and the potentials are assigned back to the particles from local expansions of the cell. Differences exist in the process to compute the local expansions from the multipole moments of all the other cells. For the number of cells K, a simple summation takes $K(K-1)$ multipole-to-local (M2L) transformations. FMM exploits a hierarchical octree structure for reduction to $O(K)$. Instead, PMMM employs a uniform mesh structure and the transformations are

accelerated by the fast convolution theorem using FFT. It only requires K transformations, associated with an extra cost of FFT which is $O(K \log K)$. However, a factor 8 overhead exists for an isolated system.

The $O(K \log K)$ scaling does not immediately mean it is slower than that of $O(K)$ in practical cases. The $O(K)$ method tends to have a relatively large factor, about 189 to 875, depending on the cell separation criterion. In the $O(K \log K)$ scheme, the number of transformations remains constant, K or $8K$ irrespective of the criterion. In both schemes, it is common that the contributions from the nearest cells are evaluated through direct particle-particle interactions.

3. Computational and memory complexity

3.1. Order estimation

1. Each particle interacts with $(p+1)^2$ coefficients of the nearest grid point. This part is $O(p^2 N)$.
2. The forward and backward FFTs of size K are performed for $(p+1)^2$ independent terms, and is $O(p^2 K \log K)$.
3. The M2L transformations $\mathbb{C}^{(p+1)^2} \to \mathbb{C}^{(p+1)^2}$ are performed on the number of K reciprocal grids, and is $O(p^4 K)$.
4. Each particle interacts with $O(N/K)$ particles in nearby cells, with a multiplying factor 27 or 125. In total, this part is $O(N^2/K)$.
5. The memory demand is $O(N + p^2 K)$ including the transformation matrices.

If we set the parameter $K \propto N/p^2$, the total computational cost becomes either $O(p^2 N)$ or $O(N \log(N/p^2))$, and the memory demand $O(N)$. In this article, we follow the original $O(p^4)$ transformation method by (Greengard and Rokhlin, 1988b), however, a possible reduction to $O(p^2 \log p)$ of this part is discussed in §5.5.

3.2. Choice of parameters

We put the minimum cell separation as c (≥ 1), and try to find the optimum value of the parameters including K and p, for given error tolerance. Let us write the cost of the short range particle–particle interactions and the long range particle-mesh interactions as

$$C_{\mathrm{PP}}(2c+1)^3 N^2/K, \quad \text{and} \quad C_{\mathrm{PM}} K (p+1)^4.$$

Here, we assume that the $O(p^4K)$ translation part costs more than the $O(p^2K \log K)$ FFT part. The factor 8 overhead for the open boundary case can be included to the coefficient C_{PM}. The balancing point of these two is given by

$$K = \sqrt{C_{\mathrm{PP}}/C_{\mathrm{PM}}}(2c+1)^{3/2}N/(p+1)^2,$$

with the resulting total cost

$$2\sqrt{C_{\mathrm{PP}}C_{\mathrm{PM}}}N \cdot (2c+1)^{3/2}(p+1)^2.$$

The error in the worst case is estimated by

$$\varepsilon = \left(\frac{\sqrt{3}/2}{(c+1) - \sqrt{3}/2}\right)^{p+1}. \tag{4}$$

This gives scalings 0.76^p, 0.41^p for $c = 1,2$ (Greengard, 1988) and 0.28^p, 0.21^p for $c = 3,4$. The order of expansion p and the cost scaling $(2c+1)^{3/2}(p+1)^2$ for the given tolerance ε are plotted in Fig. 1. A large offset in the efficiency exists between $c = 1$ and $c = 2$, and from $c = 2$, they all behave similarly. Thus, $c = 2$ (125 cells cutoff) seems satisfactory in most cases, though larger cutoff can be considered when $p \geq 10$ is needed for $c = 2$ (i.e. $\varepsilon < 10^{-4}$).

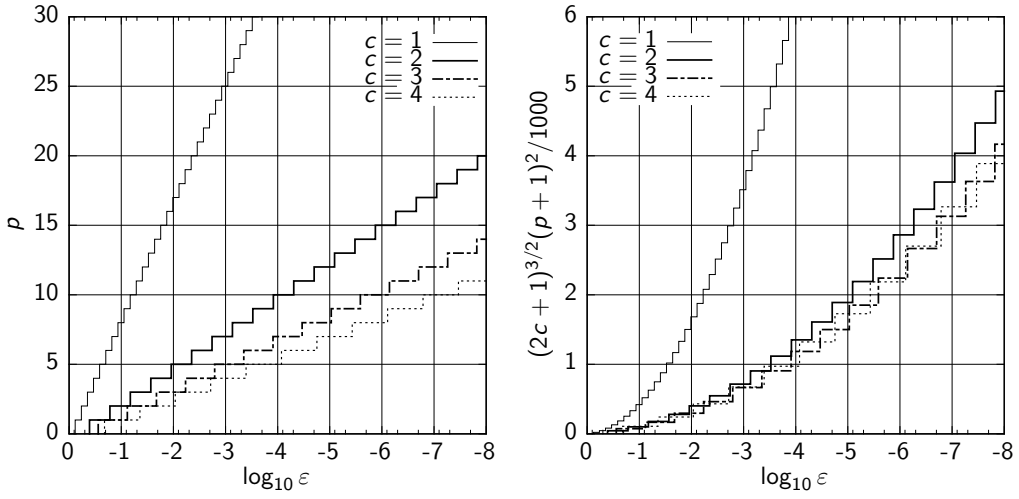

Figure 1: The order of expansion p (left) and the cost scaling $(2c+1)^{3/2}(p+1)^2$ (right) for given error tolerances ε, for $c = 1,2,3,4$.

4. Numerical test

Even the error behavior of PMMM can be expected — it should be the same level or slightly better than that of FMM because the result is mathematically equivalent to the summation of K^2 transformations of all cells — we carried out minimum numerical tests for the verification of the scheme. The potential and its gradient obtained from an open boundary PMMM were compared with those from the $O(N^2)$ direct summation, and those of a periodic bondary PMMM were compared with the Ewald summation. The test code was implemented in C++ with double precision arithmetics, and we used the FFTW 3.3.3 [2] library for the three-dimensional real-to-complex and complex-to-real Fourier transforms.

For the test condition, we randomly distributed $N = 16,384$ particles in a unit box, and the box was split into $K = 8^3$ cells, while the FFT and M2L translations in the wave space were performed on 16^3 cells for the open boundary condition. The charges of the particles were also randomly distributed in the range $[0, 1/N)$, but were later subtracted by the average value to make the total charge of the system zero. We tested both the 27 cells cutoff (one cell for the minimum separation) and the 125 cells cutoff (two cells separation) cases.

Fig. 2 shows cumulative distributions of absolute and relative error in potential for the expansion order $1 \le p \le 7$, in open boundary calculations. To make it a dimension free comparison, the distribution of the absolute value of the potential itself is plotted as a reference. Fig. 3 is essentially the same, but the norm of the error in the gradient of the potential is plotted. All of the plots represent well what we can expect from the multipole theory. The error decreases by a ratio as p increases, and the larger cutoff improves the convergence, though the scaling looks even better than the estimation in (4).

Errors in a periodic boundary system with the same particle distribution above are plotted in Fig. 4. The potential and its gradient were evaluated as described in Appendix B and compared with those from the Ewald summation. The error behavior is common to that of an open boundary case.

5. Discussion

5.1. Parallelization

Parallelization of PMMM for distributed memory machines is a straightforward task and the communication pattern is totally regular. As an extreme case,

[2] `http://www.fftw.org`

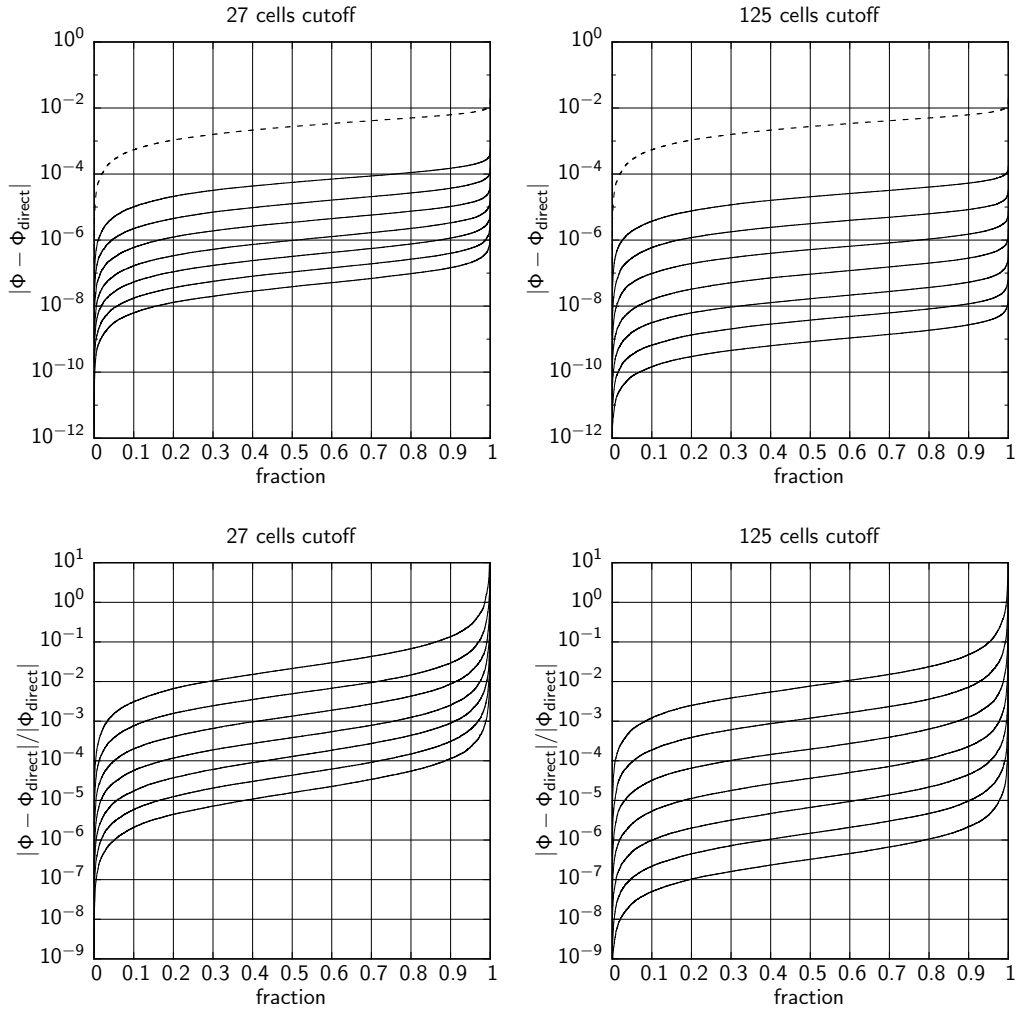

Figure 2: Cumulative plots of the error in potential, absolute (top) and relative (bottom) error for 27 cells cutoff (left) and 125 cells cutoff (right), for the order of expansion $1 \leq p \leq 7$ (upper curve to lower curve). The cumulative distribution of potential itself is plotted in the dashed curve.

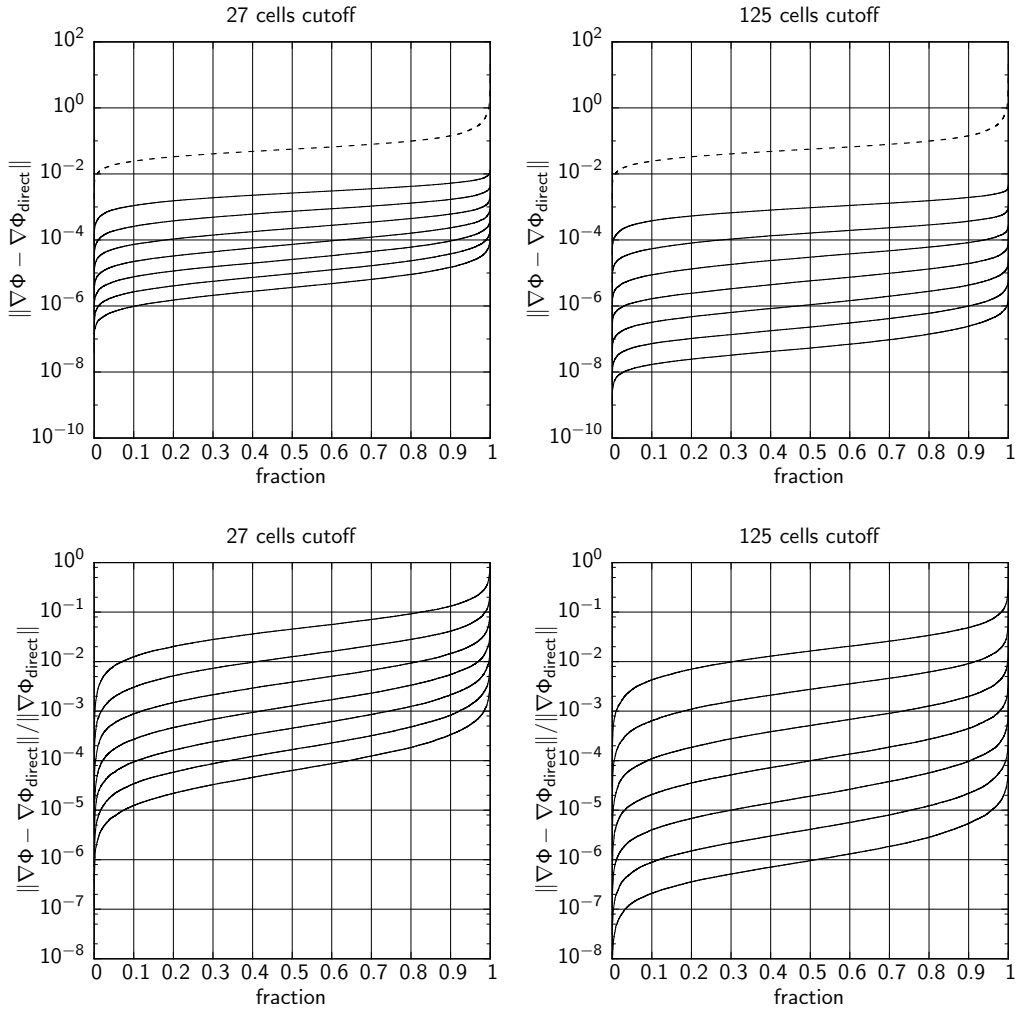

Figure 3: Same as Fig. 2, but for the gradient of potential $\nabla\Phi$.

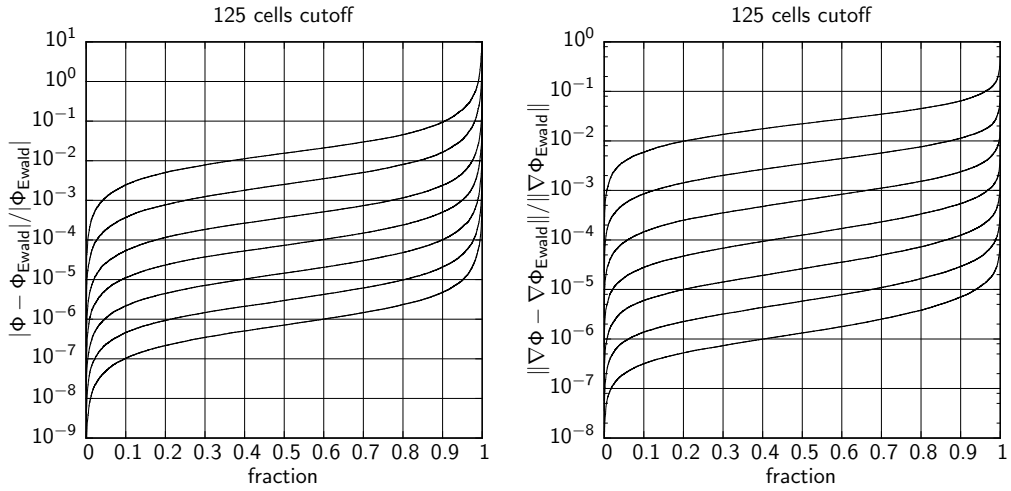

Figure 4: Relative error in potential (left) and its gradient (right) in a periodic boundary system, for $1 \le p \le 7$.

we consider using the number of processors equal to the number of cells K and assume that each processor is responsible for one cell and the particles contained in it. The interactions between particles and cell, particle-to-multipole (P2M) and local-to-particle (L2P) are totally local and parallel. The cell–cell interactions, M2L translations in the wave space after the FFT are also local and parallel, with a distributed Green's function (transformation matrices). Half the number of processors can join this part if we employ real-to-complex and complex-to-real transforms. The particle–particle interactions of nearby cells need to be solved by communications with neighbor processors, and this part is simple. Thus, almost all of the efforts will be focused on the efficient communication for the forward and backward FFT part for the $(p+1)^2$ independent terms.

One choice is to use $(p+1)^2$ processors for the FFT part where each FFT processor gathers/scatters the elements from/to all other processors. This can be easily written with nonblocking collective operations supported as a new feature in the message passing interface (MPI) version 3.0 (Message Passing Interface Forum, 2012). The communication size that one cell exports/imports during a step is kept minimum and usually smaller than that of an octree based FMM; sending and receiving the $(p+1)^2$ terms for the forward and backward FFTs. However, heavy communication traffic can concentrate on the processor that performs FFT.

In more practical cases, we may use the number of processors less than the

number of cells K, and use more than one processors for each FFT. Still the discussion above is unchanged that all we have to take care of is the parallel FFT part for the distributed coefficients.

An implementation for a graphics processing unit (GPU) would be easy by assigning each cell for each thread, and $(p+1)^2$ thread-blocks for the multiple FFT tasks.

We remark that a parallel version of FFTM already exists (Ong et al., 2004).

5.2. Hierarchical treatment

It might be a common argument from the computational science side, that an FFT and uniform grid based scheme wastes the *local* nature of the system. In principle, distant interactions can be performed in large scale cells. Moreover, the contribution from the higher order moment is more local — equation (A.17) expresses that the contribution from the λ-th order multipole moment to the ℓ-th order local expansion decays as $1/r^{\ell+\lambda+1}$.

Here, we discuss the case that we use a global coarse mesh and local fine mesh instead of a uniform mesh, with $K = K_{\text{coarse}}K_{\text{fine}}$. First, PMMM is directly applied on the coarse mesh resulting in the cost of K_{coarse} transformations in total, and the contribution from/to the nearest 27 or 125 coarse cells (including the self cell) are left over. Then a coarse cell is refined to K_{fine} (e.g. 4^3) fine cells. From a set of the multipole moments of K_{fine} cells, local expansions of $27K_{\text{fine}}$ or $125K_{\text{fine}}$ cells can be evaluated in a convolution form. Fig. C.2 illustrates a possible layout for the two-dimensional case. Including the buffer zone for the aliasing effect, one fine cell costs 64 or 216 transformations [3] irrespective of the refinement size K_{fine}. These numbers are slightly better than those of the original FMM, 189 or 875, where the extra cost comes from the asymmetry of the octree structure. However, they are much more expensive than the uniform mesh case. This hierarchical treatment relaxes the demand for the global network bandwidth in exchange for increased local computations.

For the refinement factor K_{fine}, any cubic number can be used, like 4^3, 8^3 or even 6^3, etc. Thus, the hierarchy can be quite shallow compared with the octree structure, and two levels are usually enough. As a bonus of the hierarchical treatment, an $O(N)$ scaling is recovered from the $O(N \log N)$, where the cost associated with the fine cells dominates.

[3] $64 = (3+1)^3$, $216 = (5+1)^3$, $189 = 6^3 - 3^3$, $875 = 10^3 - 5^3$.

5.3. Application to molecular dynamics (MD)

In the simulations of bimolecular systems, the most significant bottleneck is, especially for distributed memory parallel computers, in solving the global electrostatic force from charged atoms. Here, we discuss the use of PMMM for such simulations with typical parameters. We consider a simulation box with a size $(100\text{Å})^3$, and number density ~ 0.1 atom/Å^3, i.e. $N = 10^5$. We take the cell size as 5Å and 125 cutoff neighbors, to make the cutoff length 10Å. This is a typical value of the cutoff length for the Lennard-Jones potential. In the PME method (Darden et al., 1993) and its variants, smooth particle mesh Ewald (SPME, Essmann et al., 1995) and Gaussian split Ewald (GSE, Shan et al., 2005) being the most popular solvers in this field, a typical grid spacing in the above case is 1Å. Hence, the comparison is between single FFT of size 100^3 or multiple $(p+1)^2$ FFTs of size 20^3. At $p = 7$, the total number of coefficients on the grid points becomes about half of PME, keeping the accuracy at a reasonable level. This helps for reducing the FFT overhead. Moreover, lower precision arithmetics can be used for the higher order coefficients; e.g. assigning $3 \cdot (p + 1 - \ell)$ bits for the ℓ-th order coefficient results into $3 \cdot (p+1)(p+2)(2p+3)/6$ bits per cell and it is only 612 bits when $p = 7$.

In addition, the number of coefficients that a particle interacts with is not a negligible factor for the efficiency. In the SPME method, a particle usually interacts with 216 grid points for $p = 5$, while 64 terms of the nearest grid point are needed in PMMM with $p = 7$. Furthermore, PMMM has little difficulty in the charge assignment part in a multi-thread environment because a particle only interacts with its nearest grid point.

In future work, we will address the question of the sufficient value of p in practical simulations, and compare the accuracy and efficiency of PMMM with PME series. Finally, we emphasize the benefit of the relatively simple form of the particle–mesh interaction without diffusion and particle–particle interaction without a cutoff function.

5.4. Application to cosmological N-body simulations

PMMM has the same dilemma as PPPM when applied directly to cosmological N-body simulations. At the late stage of the simulation after structure formation, a high density contrast starts to require either increased particle–particle interactions or finer mesh size. Thus, a relatively young TreePM method (Bagla, 2002) which replaces the PP part of PPPM with an adaptive octree algorithm (Barnes and Hut, 1986) has become the most popular solver in this field, due to its simplicity and efficiency. Exactly the same approach can be applied for PMMM,

where the global force field is solved by PMMM with the periodic boundary condition. For the contribution from the 27 or 125 cutoff cells, either multipole-to-particle (M2P) based Barnes & Hut type tree method or M2L based FMM on adaptive octree structure can be applied.

The practical benefit of such combinations is the simplicity of both the code structure and accuracy control. The adaptive octree structure can be started from each cell of PMMM when it is needed, and constructed locally. The accuracy of the mesh part is simply controlled with one parameter p. Since the PP part has no cutoff function, it is easier to introduce higher order treecode than the center-of-mass approximation ($p = 1$), which most of TreePM implementations employ (Springel, 2005; Ishiyama et al., 2009).

5.5. Another use of FFT convolution

Historically, the use of fast convolution with FFT was suggested at an early time by Greengard and Rokhlin (1988a) to reduce the complexity of each M2L translation from $O(p^4)$ to $O(p^2 \log p)$. Later, a full study was carried out by Elliott and Board (1996) including the treatment of numerical instabilities. A two-dimensional convolution form of an M2L transformation is visualized in Fig. C.3. This approach should not be confused with the other where the FFT convolution is performed on uniform grids as in Shimada et al. (1993), Ong et al. (2003), and this work, although the approaches may be categorized by the keywords 'FFT based FMM'. Ultimately, these two methods can be combined through a five-dimensional FFT resulting in the total cost $O(p^2 K \log(p^2 K))$. Still it is not clear whether an application exists that requires an extremely high order of expansion.

Acknowledgement

The author thanks Dr. Rio Yokota of King Abdullah University of Science and Technology for his useful comments on FMM. The author also thanks Dr. Yousuke Ohno of RIKEN QBiC for his comments on biomolecular simulations.

A minimum implementation of PMMM can be obtained at
`https://github.com/nitadori/PMMM`.

Appendix A. Implementation note for FMM

The original formulation of the three-dimensional FMM for $1/r$ potential based on spherical harmonics (Greengard and Rokhlin, 1988b) seems to be a bit complicated and not straightforward for implementers. In this Appendix, we provide

all the equations needed for the implementation. Strict proofs are mostly omitted, see Epton and Dembart (1995) and van Gelderen (1998) for full descriptions.

Appendix A.1. Solid harmonics

A simple formulation of the multipole transformations is established with two solid harmonics functions (Wang and LeSar, 1996). We first prepare definitions of 'regular' and 'singular' solid harmonics base functions in spherical coordinates,

$$R_\ell^m(r,\theta,\phi) = \frac{r^\ell}{(\ell+m)!}P_\ell^m(\cos\theta)e^{im\phi}, \tag{A.1}$$

$$S_\ell^m(r,\theta,\phi) = (-1)^{\ell+m}\frac{(\ell-m)!}{r^{\ell+1}}P_\ell^m(\cos\theta)e^{im\phi}, \tag{A.2}$$

for $0 \le m \le \ell$. Extensions for negative m is given by the conjugate relation $R_\ell^{-m} = (-1)^m[R_\ell^m]^*$ and $S_\ell^{-m} = (-1)^m[S_\ell^m]^*$. A sign factor $(-1)^{\ell+m}$ is included in the definition of S_ℓ^m for simplicity of later formulation. Here, P_ℓ^m is the associated Legendre polynomial which follows the definition,

$$\begin{aligned} P_\ell^m(\cos\theta) &= (-\sin\theta)^m\frac{d^m}{(d\cos\theta)^m}P_\ell(\cos\theta) \\ &= \frac{(-\sin\theta)^m}{2^\ell\ell!}\frac{d^{\ell+m}}{(d\cos\theta)^{\ell+m}}\left[(\cos\theta)^2-1\right]^\ell. \end{aligned} \tag{A.3}$$

Note that another definition exists for the factor $(-1)^m$.

These functions represent a general solution of the Laplace equation in spherical coordinates $\nabla^2\Phi(r,\theta,\phi) = 0$, where

$$\Phi(r,\theta,\phi) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\left[M_\ell^m S_\ell^{-m}(r,\theta,\phi) + L_\ell^m R_\ell^m(r,\theta,\phi)\right], \tag{A.4}$$

for the outer (former term) and inner (latter term) solutions. The coefficients M_ℓ^m and L_ℓ^m are referred to as *multipole moments* and *local expansions*, respectively.

Some low order Cartesian expressions of these base functions are listed in table A.1. It turns out that both functions have simple Cartesian forms and hereafter we employ the notation $R_\ell^m(\boldsymbol{r}) = R_\ell^m(r,\theta,\phi) = R_\ell^m(x,y,z)$.

Appendix A.2. Ladder operators

This section is for the preparation of the derivations of transformation relations, and could be skipped when the interest is only in the results.

Table A.1: R_ℓ^m and S_ℓ^m in low orders.

(l,m)	R_ℓ^m	S_ℓ^m
$(0,\ 0)$	1	$1/r$
$(1,\ 0)$	z	$-z/r^3$
$(1,\pm1)$	$-(\pm x+iy)/2$	$-(\pm x+iy)/r^3$
$(2,\ 0)$	$(3z^2-r^2)/4$	$(3z^2-r^2)/r^5$
$(2,\pm1)$	$-z(\pm x+iy)/2$	$3z(\pm x+iy)/r^5$
$(2,\pm2)$	$(\pm x+iy)^2/8$	$3(\pm x+iy)^2/r^5$

We define the following three *ladder operators*

$$\partial_+ = \frac{\partial}{\partial x} + i\frac{\partial}{\partial y}, \quad \partial_z = \frac{\partial}{\partial z}, \quad \partial_- = -\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}, \tag{A.5}$$

where the base functions satisfy

$$\begin{pmatrix}\partial_+\\ \partial_z\\ \partial_-\end{pmatrix} R_\ell^m(\boldsymbol{r}) = \begin{pmatrix}R_{\ell-1}^{m+1}(\boldsymbol{r})\\ R_{\ell-1}^{m}(\boldsymbol{r})\\ R_{\ell-1}^{m-1}(\boldsymbol{r})\end{pmatrix}, \quad \begin{pmatrix}\partial_+\\ \partial_z\\ \partial_-\end{pmatrix} S_\ell^m(\boldsymbol{r}) = \begin{pmatrix}S_{\ell+1}^{m+1}(\boldsymbol{r})\\ S_{\ell+1}^{m}(\boldsymbol{r})\\ S_{\ell+1}^{m-1}(\boldsymbol{r})\end{pmatrix}. \tag{A.6}$$

When the indices of R_ℓ^m fall outside the range from $0 \le |m| \le \ell$, they merely become zero. The Laplace equation is confirmed with $\partial_+\partial_-[\,] = {\partial_z}^2[\,]$, where $[\,]$ is either R_ℓ^m or S_ℓ^m which is one of the solutions [4] .

From these operators, we compose

$$\mathcal{D}_\ell^{\pm|m|} = (\partial_\pm)^{|m|}(\partial_z)^{\ell-|m|}, \tag{A.7}$$

to make $\mathcal{D}_\ell^m R_\lambda^\mu = R_{\lambda-\ell}^{\mu+m}$ and $\mathcal{D}_\ell^m S_\lambda^\mu = S_{\lambda+\ell}^{\mu+m}$. By remembering $R_0^0 = 1$ and $S_0^0 = 1/r$, we have $\mathcal{D}_\ell^{-m} R_\ell^m = 1$ and $\mathcal{D}_\ell^m(1/r) = S_\ell^m$.

Let us examine the relation

$$\left[\mathcal{D}_\ell^{-m} R_\lambda^\mu(\boldsymbol{r})\right]_{\boldsymbol{r}=\boldsymbol{0}} = \delta_{\ell\lambda}\delta_{m\mu}, \tag{A.8}$$

with δ_{ij} the Kronecker delta. When $\ell \ne \lambda$ or $m \ne \mu$, $\mathcal{D}_\ell^{-m} R_\lambda^\mu(\boldsymbol{r})$ becomes just 0 or a homogeneous polynomial of Cartesian coordinates, which vanishes with the

[4] This explains that, if we define $\partial_- = \frac{\partial}{\partial x} - i\frac{\partial}{\partial y}$ and $Y_\ell^{-m} = [Y_\ell^m]^*$ (without the factor $(-1)^m$), the transformation formulae tend to have unwieldy sign factors.

substitution $\boldsymbol{r} = \boldsymbol{0}$. Then we consider a vacuum potential field around $\boldsymbol{r}_L$ which is expressed in local expansions

$$\Phi(\boldsymbol{r}) = \sum_{\lambda=0}^{\infty} \sum_{\mu=-\lambda}^{\lambda} L_\lambda^\mu R_\lambda^\mu(\boldsymbol{r} - \boldsymbol{r}_L). \tag{A.9}$$

By applying (A.8), we can extract the coefficient L_ℓ^m as

$$L_\ell^m = \left[\mathcal{D}_\ell^{-m}\Phi(\boldsymbol{r})\right]_{\boldsymbol{r}=\boldsymbol{r}_L}. \tag{A.10}$$

This enables us to obtain a set of local expansion coefficients at an arbitrary vacuum point of a given potential field $\Phi(\boldsymbol{r})$.

Now, a Taylor expansion of a function $\Phi(\boldsymbol{r})$ that satisfies the Laplace equation can be written as

$$\Phi(\boldsymbol{r} + \Delta\boldsymbol{r}) = \sum_{\lambda=0}^{\infty} \sum_{\mu=-\lambda}^{\lambda} \left[\mathcal{D}_\lambda^{-\mu}\Phi(\boldsymbol{r})\right] R_\lambda^\mu(\Delta\boldsymbol{r}). \tag{A.11}$$

The expansion is valid as long as the sphere of expansion is vacuum. As the base function $R_\ell^m(\boldsymbol{r})$ or $S_\ell^m(\boldsymbol{r})$ is a solution of the Laplace equation, both of them can be expanded in the same way, yielding addition theorems of solid harmonics,

$$R_\ell^m(\boldsymbol{r} + \Delta\boldsymbol{r}) = \sum_{\lambda=0}^{\infty} \sum_{\mu=-\lambda}^{\lambda} R_{\ell-\lambda}^{m-\mu}(\boldsymbol{r}) R_\lambda^\mu(\Delta\boldsymbol{r}), \tag{A.12}$$

$$S_\ell^{-m}(\boldsymbol{r} + \Delta\boldsymbol{r}) = \sum_{\lambda=0}^{\infty} \sum_{\mu=-\lambda}^{\lambda} S_{\ell+\lambda}^{-(m+\mu)}(\boldsymbol{r}) R_\lambda^\mu(\Delta\boldsymbol{r}). \tag{A.13}$$

Appendix A.3. Transformations

In this section, we list the transformations needed in FMM. Each translation has a shortened name where the prefix 'P' reads particle or potential, 'M' is for multipole moment, 'L' for local expansion, and '2' for 'to'. We assume that multipole moments and local expansions have finite order p, which means ℓ takes the range $0 \le \ell \le p$.

As a special case of (A.13), we have an expansion of $1/r$ potential by

$$\begin{aligned} \frac{1}{\|\boldsymbol{r}_S - \boldsymbol{r}_R\|} &= S_0^0(\boldsymbol{r}_S - \boldsymbol{r}_R) \\ &= \sum_{\lambda=0}^{\infty} \sum_{\mu=-\lambda}^{\lambda} S_\lambda^{-\mu}(\boldsymbol{r}_S) R_\lambda^\mu(-\boldsymbol{r}_R) \quad \text{for} \quad \|\boldsymbol{r}_S\| > \|\boldsymbol{r}_R\|. \end{aligned} \tag{A.14}$$

Thus, if we take the following definition for the multipole moments

$$\text{P2M:} \quad M_\ell^m = \sum_i q_i \cdot R_\ell^m(\boldsymbol{r}_M - \boldsymbol{r}_i), \tag{A.15}$$

the potential field outside the multipole sphere $\Phi(\boldsymbol{r}) = \sum_i (q_i / \|\boldsymbol{r} - \boldsymbol{r}_i\|)$ is given by,

$$\text{M2P:} \quad \Phi(\boldsymbol{r}) = \sum_{\lambda=0}^{p} \sum_{\mu=-\lambda}^{\lambda} M_\lambda^\mu S_\lambda^{-\mu}(\boldsymbol{r} - \boldsymbol{r}_M). \tag{A.16}$$

Here, $\boldsymbol{r}_M$ is the center of expansion, $\boldsymbol{r}_i$ and q_i the position and charge of particle i. The coefficients of local expansions are available from the potential field and equation (A.10), which leads directly to a multipole-to-local transformation

$$\text{M2L:} \quad L_\ell^m = \sum_{\lambda=0}^{p} \sum_{\mu=-\lambda}^{\lambda} M_\lambda^\mu S_{\ell+\lambda}^{-(m+\mu)}(\boldsymbol{r}_L - \boldsymbol{r}_M). \tag{A.17}$$

From a given set of local expansion coefficients at $\boldsymbol{r}_L$, the potential field expands to

$$\text{L2P:} \quad \Phi(\boldsymbol{r}) = \sum_{\lambda=0}^{p} \sum_{\mu=-\lambda}^{\lambda} L_\lambda^\mu R_\lambda^\mu(\boldsymbol{r} - \boldsymbol{r}_L), \tag{A.18}$$

and again (A.10) makes a new expansion at $\boldsymbol{r}_{L'}$ (local-to-local transformation) as

$$\text{L2L:} \quad L'^m_\ell = \sum_{\lambda=\ell}^{p} \sum_{\mu=-\lambda}^{\lambda} L_\lambda^\mu R_{\lambda-\ell}^{\mu-m}(\boldsymbol{r}_{L'} - \boldsymbol{r}_L). \tag{A.19}$$

Finally, we derive a multipole-to-multipole transformation for a new center $\boldsymbol{r}_{M'}$. From the addition theorem (A.12) with $\boldsymbol{r} = \boldsymbol{r}_M - \boldsymbol{r}_i$ and $\Delta\boldsymbol{r} = \boldsymbol{r}_{M'} - \boldsymbol{r}_M$,

$$R_\ell^m(\boldsymbol{r}_{M'} - \boldsymbol{r}_i) = \sum_{\lambda=0}^{\infty} \sum_{\mu=-\lambda}^{\lambda} R_{\ell-\lambda}^{m-\mu}(\boldsymbol{r}_M - \boldsymbol{r}_i) R_\lambda^\mu(\boldsymbol{r}_{M'} - \boldsymbol{r}_M),$$

and the definition of multipole moments (A.15), the transformation is given by

$$\text{M2M:} \quad M'^m_\ell = \sum_{\lambda=0}^{\ell} \sum_{\mu=-\lambda}^{\lambda} M_{\ell-\lambda}^{m-\mu} R_\lambda^\mu(\boldsymbol{r}_{M'} - \boldsymbol{r}_M). \tag{A.20}$$

In practice, μ iterates from $\max(-\lambda, m - (\ell - \lambda))$ to $\min(\lambda, m + (\ell - \lambda))$.

Appendix A.4. Potential gradient

In N-body simulations, the gradient of the potential is more important than the potential itself. We can exploit that the local expansions have the information of Cartesian gradient of the potential in their first order coefficients. The expansion is given by

$$\begin{aligned}\Phi(\boldsymbol{r}_L + d\boldsymbol{r}) &= L_0^0 - \frac{1}{2}\left(L_1^1 - L_1^{-1}\right)dx - \frac{i}{2}\left(L_1^1 + L_1^{-1}\right)dy + L_1^0\,dz + \dots \\ &= L_0^0 + \left(-\Re L_1^1\right)dx + \left(\Im L_1^1\right)dy + L_1^0\,dz + \dots\end{aligned} \tag{A.21}$$

Here, $\Re$ and $\Im$ are operators to extract the real and imaginary part. Thus, we can use the L2L or M2L implementation instead of L2P or M2P by setting the destination order as one, when we need the gradient at the particle position.

Appendix A.5. Computing solid harmonics

Table A.1 motivates us to have a full Cartesian derivation of the two solid harmonics. With help of the scaled version of the associated Legendre polynomials which we define as $\tilde{P}_\ell^m(r,z) = r^\ell \cdot (r\sin\theta)^{-m} \cdot P_\ell^m(\cos\theta)$, we have $\tilde{P}_\ell^m(r,z)\cdot(x+iy)^m = r^\ell \cdot P_\ell^m(\cos\theta)\cdot e^{im\phi}$. For $m \geq 0$, this obeys a recursion,

$$\tilde{P}_\ell^m(r,z) = \begin{cases} (-1)^m(2m-1)!! & (\ell = m) \\ (2\ell-1)z\tilde{P}_{\ell-1}^m(r,z) & (\ell = m+1) \\ \dfrac{2\ell-1}{\ell-m}z\tilde{P}_{\ell-1}^m(r,z) - \dfrac{\ell+m-1}{\ell-m}r^2\tilde{P}_{\ell-2}^m(r,z) & (\ell \geq m+2) \end{cases}. \tag{A.22}$$

Now we can compute the solid harmonics by

$$\begin{aligned} R_\ell^m(x,y,z) &= \frac{1}{(\ell+m)!}Q_\ell^m(x,y,z), \\ S_\ell^{-m}(x,y,z) &= \frac{(\ell-m)!}{r^{2\ell+1}}Q_\ell^m(-x,y,-z), \\ \text{with}\quad Q_\ell^m(x,y,z) &= \tilde{P}_\ell^m(r,z)\cdot(x+iy)^m. \end{aligned} \tag{A.23}$$

Here, we use octant symmetries,

$$Q_\ell^m(-x,y,-z) = (-1)^\ell Q_\ell^m(x,-y,z) = (-1)^{\ell+m}Q_\ell^{-m}(x,y,z). \tag{A.24}$$

Appendix A.6. Storage format

All R_ℓ^m, S_ℓ^m, M_ℓ^m, and L_ℓ^m satisfy the conjugate relation $R_\ell^{-m} = (-1)^m [R_\ell^m]^*$, etc. Thus, they effectively have $(p+1)^2$ real numbers when the order of expansion is p. We save the elements with non-negative m to one-dimensional arrays X_i with an index $i = \ell(\ell+1)+m$, for ranges $0 \le \ell \le p$, $-\ell \le m \le \ell$, and thus $0 \le i < (p+1)^2$, as

$$X_{\ell(\ell+1)+m} = \begin{cases} \Re R_\ell^{|m|} & (m \ge 0) \\ \Im R_\ell^{|m|} & (m < 0) \end{cases}. \tag{A.25}$$

Note that S_ℓ^{-m} for the M2L translation in an order p calculation requires a size $(2p+1)^2$.

Appendix A.7. Real Matrix form

Equation (A.17) can be regarded as a linear transformation $\mathbb{R}^{(p+1)^2} \to \mathbb{R}^{(p+1)^2}$. First we consider a transformation in complex numbers,

$$\begin{aligned} L_\ell^m &= \sum_{\lambda=0}^{p} \sum_{\mu=-\lambda}^{\lambda} G_{\ell,\lambda}^{m,\mu} M_\lambda^\mu \\ &= \sum_{\lambda=0}^{p} \left[G_{\ell,\lambda}^{m,0} M_\lambda^0 + \sum_{\mu=1}^{\lambda} \left(G_{\ell,\lambda}^{m,\mu} M_\lambda^\mu + G_{\ell,\lambda}^{m,-\mu} M_\lambda^{-\mu} \right) \right], \end{aligned} \tag{A.26}$$

with a matrix element $G_{\ell,\lambda}^{m,\mu} \in \mathbb{C}$ that depends on the relative position of two cells. From the conjugate relation $\Re M_\ell^{-m} = (-1)^m \Re M_\ell^m$ and $\Im M_\ell^{-m} = (-1)^{m+1} \Im M_\ell^m$, a transformation in real numbers is expressed by

$$\begin{aligned} \Re L_\ell^m &= \sum_{\lambda=0}^{p} \left[A_{\ell,\lambda}^{m,0} M_\lambda^0 + \sum_{\mu=1}^{\lambda} \left\{ \left(A_{\ell,\lambda}^{m,\mu} + C_{\ell,\lambda}^{m,\mu} \right) \Re M_\lambda^\mu - \left(B_{\ell,\lambda}^{m,\mu} - D_{\ell,\lambda}^{m,\mu} \right) \Im M_\lambda^\mu \right\} \right], \\ \Im L_\ell^m &= \sum_{\lambda=0}^{p} \left[B_{\ell,\lambda}^{m,0} M_\lambda^0 + \sum_{\mu=1}^{\lambda} \left\{ \left(B_{\ell,\lambda}^{m,\mu} + D_{\ell,\lambda}^{m,\mu} \right) \Re M_\lambda^\mu + \left(A_{\ell,\lambda}^{m,\mu} - C_{\ell,\lambda}^{m,\mu} \right) \Im M_\lambda^\mu \right\} \right], \end{aligned}$$

with

$$\begin{aligned} A_{\ell,\lambda}^{m,\mu} &= \Re G_{\ell,\lambda}^{m,\mu}, & B_{\ell,\lambda}^{m,\mu} &= \Im G_{\ell,\lambda}^{m,\mu}, \\ C_{\ell,\lambda}^{m,\mu} &= (-1)^\mu \Re G_{\ell,\lambda}^{m,-\mu}, & D_{\ell,\lambda}^{m,\mu} &= (-1)^\mu \Im G_{\ell,\lambda}^{m,-\mu}. \end{aligned} \tag{A.27}$$

For $0 \le m \le \ell \le p$, this gives a linear transformation $\mathbb{R}^{(p+1)^2} \to \mathbb{R}^{(p+1)^2}$.

Appendix B. Green's function for periodic boundary conditions

In a periodic system, the transformation matrix might be given by an infinite summation of mirror images, as in

$$G^{m,\mu}_{\ell,\lambda}(\boldsymbol{r}_L - \boldsymbol{r}_M) = \sum_{\boldsymbol{n}\in\mathbb{Z}^3} S^{-(m+\mu)}_{\ell+\lambda}\Big(\boldsymbol{r}_L - (\boldsymbol{r}_M + \boldsymbol{r}_{\boldsymbol{n}})\Big), \tag{B.1}$$

where $\boldsymbol{r}_{\boldsymbol{n}} = (n_x b_x, n_y b_y, n_z b_z)$ is a displacement of a mirror image for the root box dimension (b_x, b_y, b_z), and the summation covers all the three integers $\boldsymbol{n} = (n_x, n_y, n_z) \in \mathbb{Z}^3$. The contributions from the nearest 27 or 125 cells need to be subtracted which we do not write explicitly under the summation symbol.

Appendix B.1. Infinite summation of periodic FMM

For the infinite summation of the singular solid harmonics, we can apply a rapid convergence method for the periodic FMM (Figueirido et al., 1997; Amisaki, 2000). The basic idea of the method is based on a splitting with usual and incomplete gamma functions

$$\gamma(a,x) + \Gamma(a,x) = \Gamma(a),$$

which have definitions for each term as

$$\int_0^x t^{a-1}e^{-t}dt + \int_x^\infty t^{a-1}e^{-t}dt = \int_0^\infty t^{a-1}e^{-t}dt. \tag{B.2}$$

By substituting $a = \ell + \frac{1}{2}$ and $x = (\alpha r)^2$, we have a splitting of a power function

$$\frac{1}{r^{\ell+1}} = \frac{\Gamma\left(\ell+\frac{1}{2},(\alpha r)^2\right)}{\Gamma\left(\ell+\frac{1}{2}\right)}\frac{1}{r^{\ell+1}} + \frac{\gamma\left(\ell+\frac{1}{2},(\alpha r)^2\right)}{\Gamma\left(\ell+\frac{1}{2}\right)}\frac{1}{r^{\ell+1}}. \tag{B.3}$$

Then, the latter term with $\gamma()$ is transformed into a summation in the reciprocal space. The final form is

$$\begin{aligned}\sum_{\boldsymbol{n}\in\mathbb{Z}^3\setminus\boldsymbol{0}} S^m_\ell(\boldsymbol{r}_{\boldsymbol{n}}) = &\sum_{\boldsymbol{n}\in\mathbb{Z}^3\setminus\boldsymbol{0}} \frac{\Gamma\left(\ell+\frac{1}{2},(\alpha r_{\boldsymbol{n}})^2\right)}{\Gamma\left(\ell+\frac{1}{2}\right)} S^m_\ell(\boldsymbol{r}_{\boldsymbol{n}}) \\ &+ \sum_{\boldsymbol{n}\in\mathbb{Z}^3\setminus\boldsymbol{0}} \frac{(i\pi)^\ell}{\sqrt{\pi}} \frac{\exp\left(-(\pi k_{\boldsymbol{n}}/\alpha)^2\right)}{\Gamma\left(\ell+\frac{1}{2}\right)} \frac{k_{\boldsymbol{n}}^{2\ell-1}}{V} S^m_\ell(\boldsymbol{k}_{\boldsymbol{n}}).\end{aligned} \tag{B.4}$$

Here, $\boldsymbol{k_n} = (n_x/b_x, n_y/b_y, n_z/b_z)$ is a reciprocal space vector [5], $V = b_x b_y b_z$ the volume of the root box, α a splitting parameter, $r_{\boldsymbol{n}} = \|\boldsymbol{r_n}\|$, and $k_{\boldsymbol{n}} = \|\boldsymbol{k_n}\|$. Both terms decay quickly for increasing $\|\boldsymbol{n}\|$. Example parameters for a double precision calculation on a unit box are $\alpha = 1.5$ and $\|\boldsymbol{n}\| \leq 4$ for both summations.

Appendix B.2. Green's function for PMMM

With an offset vector $\boldsymbol{r}$, (B.4) is modified slightly to

$$\begin{aligned}\sum_{\boldsymbol{n}\in\mathbb{Z}^3} S_\ell^m(\boldsymbol{r}+\boldsymbol{r_n}) = \\ \sum_{\boldsymbol{n}\in\mathbb{Z}^3} \frac{\Gamma\left(\ell+\frac{1}{2}, \left(\alpha\|\boldsymbol{r}+\boldsymbol{r_n}\|\right)^2\right)}{\Gamma\left(\ell+\frac{1}{2}\right)} S_\ell^m(\boldsymbol{r}+\boldsymbol{r_n}) + \\ \sum_{\boldsymbol{n}\in\mathbb{Z}^3\setminus\boldsymbol{0}} \frac{(-i\pi)^\ell}{\sqrt{\pi}} \frac{\exp\left(-(\pi k_{\boldsymbol{n}}/\alpha)^2\right)}{\Gamma\left(\ell+\frac{1}{2}\right)} \frac{k_{\boldsymbol{n}}^{2\ell-1}}{V} S_\ell^m(\boldsymbol{k_n}) \cdot \exp(2\pi i \boldsymbol{k_n}\cdot\boldsymbol{r}).\end{aligned} \tag{B.5}$$

Equality is valid only for $\ell > 2$ and the left-hand side diverges otherwise.

When $\ell = 0$, the right-hand side of (B.5) agrees with the well known Ewald form

$$\begin{aligned}\sum_{\boldsymbol{n}\in\mathbb{Z}^3} \frac{\operatorname{erfc}\left(\alpha\|\boldsymbol{r}+\boldsymbol{r_n}\|\right)}{\|\boldsymbol{r}+\boldsymbol{r_n}\|} + \\ \sum_{\boldsymbol{n}\in\mathbb{Z}^3\setminus\boldsymbol{0}} \frac{\exp\left(-(\pi k_{\boldsymbol{n}}/\alpha)^2\right)}{\pi V k_{\boldsymbol{n}}^2} \cdot \exp(2\pi i \boldsymbol{k_n}\cdot\boldsymbol{r}),\end{aligned} \tag{B.6}$$

and $\ell = 1$ gives its gradient. Thus, the right-hand side of (B.5) could be expected to give an identical result to the Ewald method. However, it turns out that we need several corrections in the potential which are due to the conditional convergence at $\ell = 2$. After several attempts, the following correction terms are added to the

[5] We only consider a rectangular box.

potential:

$$\begin{aligned}\Phi_i^{(\mathrm{corr})} =& \frac{2\pi}{3V}\sum_{j=1}^{N} q_j \left\| \boldsymbol{r}_i^{(\mathrm{rel})} - \boldsymbol{r}_j^{(\mathrm{rel})} \right\|^2 \\ =& -\frac{4\pi}{3V}\left[\sum_{j=1}^{N} q_j \boldsymbol{r}_j^{(\mathrm{rel})}\right] \cdot \boldsymbol{r}_i^{(\mathrm{rel})} \\ &+ \frac{2\pi}{3V}\left[\sum_{j=1}^{N} q_j \left\| \boldsymbol{r}_j^{(\mathrm{rel})} \right\|^2\right] + \frac{2\pi}{3V}\left[\sum_{j=1}^{N} q_j\right] \left\| \boldsymbol{r}_i^{(\mathrm{rel})} \right\|^2 .\end{aligned} \tag{B.7}$$

Here, $\boldsymbol{r}_i^{(\mathrm{rel})}$ is the coordinate of particle i relative to the center of cell in which it resides. Since the result agrees with the Ewald method when all the particles reside at the centers of cells, the relative positions are involved in the correction.

An intuitive interpretation of this potential correction is as follows. The density field which corresponds to the solution of the Ewald method is

$$\rho(\boldsymbol{r}) = \sum_{j=1}^{N} q_j \left[\delta\Big(\boldsymbol{r} - (\boldsymbol{r}_j + \boldsymbol{r}_{\boldsymbol{n}})\Big) - \frac{1}{V}\right]. \tag{B.8}$$

The last term inside the square bracket is due to the omitted wave number $\mathbf{0}$ in the Fourier space. Now let us examine the potential around $\boldsymbol{r}_j$ due to the uniform counter charge field $\rho(\boldsymbol{r}) = -q_j/V$. A total charge in a solid sphere $\{\boldsymbol{r} \mid \|\boldsymbol{r} - \boldsymbol{r}_j\| \le \|\boldsymbol{r}_i - \boldsymbol{r}_j\|\}$ is $-\frac{4\pi}{3V} q_j \|\boldsymbol{r}_i - \boldsymbol{r}_j\|^3$, and it makes a radial electric field $-\nabla_i \Phi(\boldsymbol{r}_i) = -\frac{4\pi}{3V} q_j (\boldsymbol{r}_i - \boldsymbol{r}_j)$ and hence a potential $\Phi(\boldsymbol{r}_i) = \frac{2\pi}{3V} q_j \|\boldsymbol{r}_i - \boldsymbol{r}_j\|^2$ which explains (B.7). Even in a charge neutral system where $\sum_{i=1}^{N} q_i = 0$, the first term of (B.7) still remains [6]. Thus, a computation based on a vacuum boundary, i.e. $1/r$ potential, which does not include the contribution from the uniform counter charge field $\rho(\boldsymbol{r}) = -q_j/V$, requires the correction (B.7) to obtain an identical potential to the Ewald method or the solution of the Poisson equation for (B.8).

See De Leeuw, Perram and Smith (1980) for a mathematical insight of the correction term.

[6] Consider a two-body charge neutral system with $+q$ at $\boldsymbol{r}_1$ and $-q$ at $\boldsymbol{r}_2$. The Ewald method gives a well defined energy that depends on the relative position of the two, $U = -q^2 G(\boldsymbol{r}_1 - \boldsymbol{r}_2) = -q^2 G(\boldsymbol{r}_2 - \boldsymbol{r}_1)$ with a Green's function $G(\boldsymbol{r})$, of which the Laplacian is not zero and $\nabla^2 G(\boldsymbol{r}) = -4\pi(\delta(\boldsymbol{r}) - 1/V)$. Thus, each particle feels the uniform counter charge field of the other, even in a charge neutral case of the Ewald method.

Appendix B.3. Correction procedure

The first term of (B.7) can be reflected as a correction to the first order expansion L_1^m from the summation of the first order moment M_1^m, as in

$$\begin{aligned}\left[L_1^0\right]_{\boldsymbol{i}} &:= \left[L_1^0\right]_{\boldsymbol{i}} + \frac{4\pi}{3V}\sum_{\boldsymbol{j}}\left[M_1^0\right]_{\boldsymbol{j}},\\ \left[L_1^1\right]_{\boldsymbol{i}} &:= \left[L_1^1\right]_{\boldsymbol{i}} + \frac{8\pi}{3V}\sum_{\boldsymbol{j}}\left[M_1^1\right]_{\boldsymbol{j}}^*,\end{aligned} \tag{B.9}$$

where $\boldsymbol{i}$ or $\boldsymbol{j}$ is an index of a cell, and the summation on $\boldsymbol{j}$ takes over all the cells.

The second term is a correction to the potential:

$$\left[L_0^0\right]_{\boldsymbol{i}} := \left[L_0^0\right]_{\boldsymbol{i}} - \frac{2\alpha}{\sqrt{\pi}}\left[M_0^0\right]_{\boldsymbol{i}} + \frac{2\pi}{3V}\sum_{\boldsymbol{j}}\left[\sum_k q_k\left\|\boldsymbol{r}_k^{(\mathrm{rel})}\right\|^2\right]_{\boldsymbol{j}}. \tag{B.10}$$

The self energy term is corrected as well. The summation $\sum_k q_k\|\boldsymbol{r}_k^{(\mathrm{rel})}\|^2$ of each cell has the same dimension as the five quadrupole moments M_2^m $(-2 \le m \le 2)$, however, is independent from either of the five.

In a charged system where $\sum_{j=1}^N q_j \neq 0$, the potential of each particle needs an explicit correction together with a self term:

$$\begin{aligned}\Phi_i &:= \Phi_i + \frac{2\pi}{3V}\left[\sum_{j=1}^N q_j\right]\left(\left\|\boldsymbol{r}_i^{(\mathrm{rel})}\right\|^2 - \frac{3}{2\alpha^2}\right),\\ \nabla\Phi_i &:= \nabla\Phi_i + \frac{4\pi}{3V}\left[\sum_{j=1}^N q_j\right]\boldsymbol{r}_i^{(\mathrm{rel})}.\end{aligned} \tag{B.11}$$

The potential of Ewald summation for the reference value is defined in

$$\begin{aligned}\Phi_i = &\sum_{\boldsymbol{n}\in\mathbb{Z}^3}\sum_{j=1}^N \frac{\mathrm{erfc}\left(\alpha\|\boldsymbol{r}_{ij}+\boldsymbol{r}_{\boldsymbol{n}}\|\right)}{\|\boldsymbol{r}_{ij}+\boldsymbol{r}_{\boldsymbol{n}}\|}\\ &+ \sum_{\boldsymbol{n}\in\mathbb{Z}^3\setminus\boldsymbol{0}} \frac{\exp\left(-(\pi k_{\boldsymbol{n}}/\alpha)^2\right)}{\pi V k_{\boldsymbol{n}}^2}\sum_{j=1}^N \exp(2\pi i\boldsymbol{k}_{\boldsymbol{n}}\cdot\boldsymbol{r}_{ij})\\ &+ \frac{2\alpha}{\sqrt{\pi}}q_i - \frac{\pi}{V\alpha^2}\sum_{j=1}^N q_j,\end{aligned} \tag{B.12}$$

and the total energy $U = \frac{1}{2}\sum_{i=1}^{N} q_i\Phi_i$. In the first line of (B.12), the interaction for $j = i$ is suppressed when $\boldsymbol{n} = \boldsymbol{0}$. See also a manual [7] of ProtoMol framework (Matthey et al., 2004) for the comments on each term.

Appendix C. Array layout of Green's function

To avoid the aliasing effect, we need eight times more volume to perform the convolution operation in a three-dimensional open boundary system. Figure C.1 shows an example layout of the Green's function in the case of two-dimensional system with 4^2 cells.

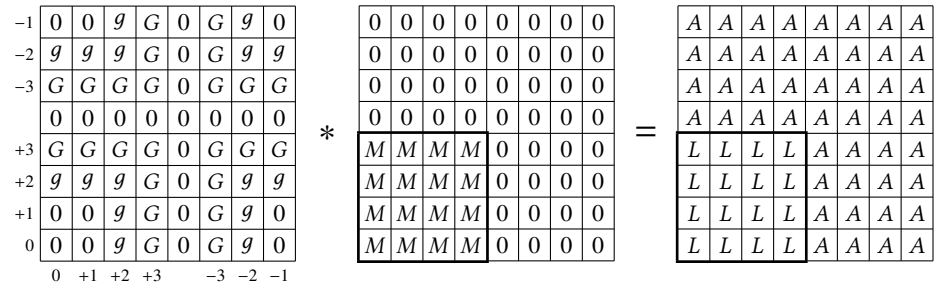

Figure C.1: A convolution for an open boundary system with 4^2 cells, computing local expansions (right) from a Green's function (left) and multipole moments (middle), with G an element of the Green's function, M multipole moments, L local expansions, A an aliasing element. The asterisk symbol ($*$) denotes a convolution operation. The interactions reach up to ±3 cells for each direction, and the nearest nine interactions are masked with 0, whereas g takes either G or 0 depending on the cutoff distance.

Here, a convolution operation between two-dimensional arrays is defined by,

$$h = f * g \Leftrightarrow h(i_0, j_0) = \sum_{i,j} f(i_0 - i, j_0 - j) \times g(i, j) \tag{C.1}$$

with periodic indices, and $f * g = g * f$.

Figure C.2 illustrates a hierarchical treatment of PMMM in a two-dimensional system. Contribution of a coarse cell to 9 nearby cells were masked out. Then, the coarse cell is refined to 4^2 fine cells, and contributions of 4^2 multipole moments to 12^2 local expansions are evaluated by a fast convolution method. Again, contributions of 9 nearby fine cells are masked out. In this case, 16^2 FFT and M2L

[7] `http://protomol.sourceforge.net/ewald.pdf` [term (7) needs to be doubled)]

transformations are needed including the margin region. Thus, one fine cell costs the same as 16 transformations. This cost does not depend on the size of refined cells.

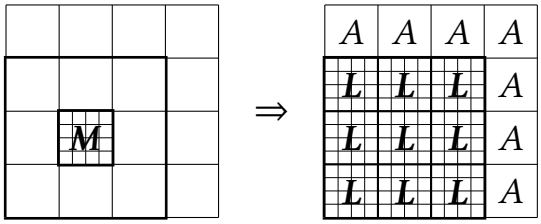

Figure C.2: Computing local expansions of 12^2 cells ($\boldsymbol{L}$) from multipole moments of 4^2 cells ($\boldsymbol{M}$) , with aliasing cells (A). Including the margin region, it requires 16^2 transformations. The Green's function looks quite similar to Fig. C.1, and the interaction reaches up to ± 7 cells for each direction.

The transformation $L_\ell^m = \sum_{\lambda,\mu} S_{\ell+\lambda}^{-(m+\mu)} M_\lambda^\mu$ itself has also a convolution form. A possible convolution form is illustrated in Fig. C.3. This reduces the computational complexity of one transformation from $O(p^4)$ to $O(p^2 \log p)$ (Greengard and Rokhlin, 1988a; Elliott and Board, 1996).

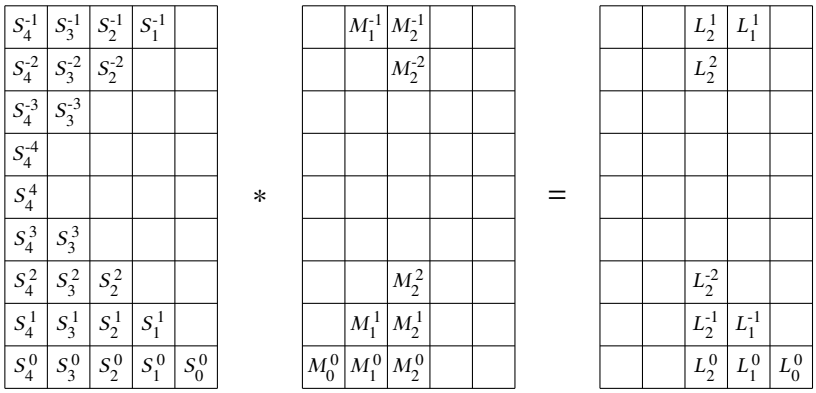

Figure C.3: An M2L transformation in a convolution form with an order $p = 2$.

Appendix D. Use of complex numbers

A complication arises when both the potential theory in spherical harmonics and Fourier transformation are formulated in complex numbers but what matters is charge and potential in real numbers. Several optimizations in this topic have

been discussed in the text, however, one simple choice is to perform two transformations simultaneously in full complex operations. From the linearity, we can exploit

$$L_\ell^m + iL'^m_\ell = \sum_{\lambda,\mu} G_{\ell,\lambda}^{m,\mu} (M_\lambda^\mu + iM'^\mu_\lambda). \tag{D.1}$$

First, we compose two sets of coefficients as

$$C_\ell^m = A_\ell^m + iB_\ell^m,$$

where $A_\ell^{-m} = (-1)^m [A_\ell^m]^*$ and $B_\ell^{-m} = (-1)^m [B_\ell^m]^*$. The composite C_ℓ^m effectively has $2(p+1)^2$ words in real numbers. Then, from

$$\begin{aligned} C_\ell^m &= (\Re A_\ell^m - \Im B_\ell^m) + i(\Re B_\ell^m + \Im A_\ell^m), \\ (-1)^m C_\ell^{-m} &= (\Re A_\ell^m + \Im B_\ell^m) + i(\Re B_\ell^m - \Im A_\ell^m), \end{aligned} \tag{D.2}$$

the splitting is given by

$$\begin{aligned} A_\ell^m &= \frac{1}{2} \Re \left[C_\ell^m + (-1)^m C_\ell^{-m} \right] + \frac{i}{2} \Im \left[C_\ell^m - (-1)^m C_\ell^{-m} \right], \\ B_\ell^m &= \frac{1}{2} \Im \left[C_\ell^m + (-1)^m C_\ell^{-m} \right] - \frac{i}{2} \Re \left[C_\ell^m - (-1)^m C_\ell^{-m} \right]. \end{aligned} \tag{D.3}$$

The equations above can be applied for composing two sets of multipole moments and splitting the composite of two local expansions.

References

Amisaki, T., 2000. Precise and efficient Ewald summation for periodic fast multipole method. J. Comput. Chem. 21, 1075–1087.

Bagla, J.S., 2002. TreePM: a code for cosmological N-body simulations. Journal of Astrophysics and Astronomy 23, 185–196.

Barnes, J.E., Hut, P., 1986. A hierarchical $O(N \log N)$ force-calculation algorithm. Nature 324, 446–449.

Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald — an $N \log N$ method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092.

De Leeuw, S.W., Perram, J.W., Smith, E.R., 1980. Simulation of electrostatic systems in periodic boundary conditions. I. lattice sums and dielectric constants. Proc. R. Soc. Lond. A 373, 27–56.

Elliott, W.D., Board, Jr., J.A., 1996. Fast Fourier transform accelerated fast multipole algorithm. SIAM J. Sci. Comput. 17, 398–415.

Epton, M.A., Dembart, B., 1995. Multipole translation theory for the three-dimensional Laplace and Helmholtz equations. SIAM J. Sci. Comput. 16, 865–897.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L., 1995. A smooth particle mesh Ewald method. J. Chem. Phys. 103, 8577.

Figueirido, F., Levy, R.M., Zholl, R., Berne, B.J., 1997. Erratum: "Large scale simulation of macromolecules in solution: Combining the periodic fast multipole method with multiple time step integrators". J. Chem. Phys. 107, 7002–7002.

van Gelderen, M., 1998. The shift operators and translations of spherical harmonics. DEOS Progress Letter 98, 57–67.

Greengard, L., 1988. The rapid evaluation of potential fields in particle systems. ACM distinguished dissertations, MIT Press, Cambridge, MA, USA.

Greengard, L., Rokhlin, V., 1987. A fast algorithm for particle simulations. J. Comput. Phys. 73, 325–348.

Greengard, L., Rokhlin, V., 1988a. On the efficient implementation of the fast multipole algorithm. Technical Report TR-602. Yale University. New Haven, CT, USA.

Greengard, L., Rokhlin, V., 1988b. The rapid evaluation of potential fields in three dimensions, in: Anderson, C.R., Greengard, C. (Eds.), Vortex Methods: Proceedings of the U.C.L.A. workshop held in Los Angeles, May 20–22, 1987, Springer-Verlag, Berlin, Germany / Heidelberg, Germany / London, UK / etc.. pp. 121–141.

Hesford, A.J., Waag, R.C., 2010. The fast multipole method and Fourier convolution for the solution of acoustic scattering on regular volumetric grids. J. Comput. Phys. 229, 8199–8210.

Hockney, R.W., Eastwood, J.W., 1988. Computer Simulation Using Particles. Adam Hilger Ltd., Bristol, UK.

Ishiyama, T., Fukushige, T., Makino, J., 2009. GreeM: Massively parallel TreePM code for large cosmological N-body simulations. Publications of the Astronomical Society of Japan 61, 1319–1330.

Makino, J., 1999. Yet another fast multipole method without multipoles—pseudoparticle multipole method. J. Comput. Phys. 151, 910–920.

Matthey, T., Cickovski, T., Hampton, S., Ko, A., Ma, Q., Nyerges, M., Raeder, T., Slabach, T., Izaguirre, J.A., 2004. Protomol, an object-oriented framework for prototyping novel algorithms for molecular dynamics. ACM Trans. Math. Softw. 30, 237–265.

Message Passing Interface Forum, 2012. MPI: A Message-Passing Interface Standard Version 3.0. Chapter author for Collective Communication, Process Topologies, and One Sided Communications.

Ong, E.T., Lee, H.P., Lim, K.M., 2004. A parallel fast fourier transform on multipoles (FFTM) algorithm for electrostatics analysis of three-dimensional structures. IEEE Trans. Computer-Aided Design 23, 1063–1072.

Ong, E.T., Lim, K.M., Lee, K.H., Lee, H.P., 2003. A fast algorithm for three-dimensional potential fields calculation: fast Fourier transform on multipoles. J. Comput. Phys. 192, 244–261.

Sezai, T., Hisada, Y., Zhai, H., Chen, Q., Sawaya, K., 2007. Improvement of calculation speed and memory of the MoM by the CG-FMM-FFT method. Technical report of IEICE SPS2007, 7–14.

Shan, Y., Klepeis, J.L., Eastwood, M.P., Dror, R.O., Shaw, D.E., 2005. Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. J. Chem. Phys. 122, 054101+.

Shimada, J., Kaneko, H., Takada, T., 1993. Efficient calculations of Coulombic interactions in biomolecular simulations with periodic boundary conditions. J. Comput. Chem. 14, 867–878.

Shimada, J., Kaneko, H., Takada, T., 1994. Performance of fast multipole methods for calculating electrostatic interactions in biomacromolecular simulations. J. Comput. Chem. 15, 28–43.

Springel, V., 2005. The cosmological simulation code GADGET-2. Monthly Notices of the Royal Astronomical Society 364, 1105–1134.

Wang, H.Y., LeSar, R., 1996. An efficient fast multipole algorithm based on an expansion in the solid harmonics. J. Chem. Phys. 104, 4173–4179.