Chapter 12

Long-Range Interactions

As computing power continues to increase, we can simulate ever larger systems. For instance, a typical model of a biological system may contain as many as 10^5 particles. In such systems it becomes crucial to avoid computing all pair interactions, as otherwise the computational effort would be proportional to the square of the number of particles. This issue is particularly relevant for long-range interactions (e.g., Coulombic and dipolar potentials) as, for such models, truncation of the potential is never allowed. It then becomes essential to find an efficient technique for computing the long-range part of the intermolecular interactions. In section 3.2.2, we showed that, if we truncate the potential at a distance r_c, the contribution of the tail of the potential $u(r)$ can be estimated (in three dimensions) using

$$U^{\text{tail}} = \frac{N\rho}{2} \int_{r_c}^{\infty} dr\, u(r) 4\pi r^2,$$

where ρ is the average number density. This equation shows that the tail correction to the potential energy diverges, unless the potential energy function $u(r)$ decays faster than r^{-3}. This is why one cannot use truncation plus tail correction for Coulombic and dipolar interactions.

In the literature, one can find numerous examples where the computational cost of evaluating long-range interactions is reduced in a rather drastic way. It is simply assumed that the long-range part of the potential is not important. The problem of the long-range interactions is then "solved" by truncation. This gets rid of the expensive part of the calculation, but it gives rise to serious inaccuracies. A discussion of the artifacts that are introduced by the various truncation schemes is presented in some detail by Steinbach and Brooks [314]. In the present chapter we discuss some of the less draconian (and more reliable) techniques for handling long-range interactions. Such schemes are more expensive than simple truncation, but the advantage is that they do respect the long-range character of the forces.

We discuss three techniques: (1) Ewald summation, (2) fast multipole methods, and (3) particle-mesh-based techniques. Of these, the Ewald summation is (still) the most widely used. As the computational effort for the Ewald summation scales as $\mathcal{O}(N^{3/2})$, this approach becomes prohibitively expensive for large systems. To overcome these restrictions, several alternative algorithms, such as the particle-particle/particle-mesh (PPPM) method of Eastwood and Hockney [315], which scales as $\mathcal{O}(N \log N)$, and the fast multipole method of Greengard and Rokhlin [316], which scales as $\mathcal{O}(N)$, have been proposed. However, these $\mathcal{O}(N)$ algorithms only become more efficient than the Ewald summation for systems containing on the order of 10^5 particles—where it should be noted that the precise location of the break-even point depends on the desired accuracy. For intermediate-size systems ($N \approx 10^3$–10^4) the so-called particle-mesh Ewald summation [317] is an attractive alternative. The complexity of the latter method also scales as $\mathcal{O}(N \log N)$.

12.1 Ewald Sums

In this section we present a simple (and nonrigorous) discussion of the Ewald method [318] for computing long-range contributions to the potential energy in a system with periodic boundary conditions. Readers who are interested in more detail (or more rigor) are referred to a series of articles by De Leeuw *et al.* [319–321] and an introductory paper by Hansen [322]. A discussion of the Ewald sum in the context of solid state physics can be found in [323]. We also refer the reader to the literature for a discussion of an ingenious alternative to the Ewald method that can be used for relatively small fluid systems [324].

12.1.1 Point Charges

Let us first consider a system consisting of positively and negatively charged particles. These particles are assumed to be located in a cube with diameter L (and volume $V = L^3$). We assume periodic boundary conditions. The total number of particles in the fundamental simulation box (the unit cell) is N. We assume that all particles repel each another at sufficiently short distances. In addition we assume that the system as a whole is electrically neutral; that is, $\sum_i q_i = 0$. We wish to compute the Coulomb contribution to the potential energy of this N-particle system,

$$U_{\text{Coul}} = \frac{1}{2} \sum_{i=1}^{N} q_i \phi(r_i), \tag{12.1.1}$$

where $\phi(r_i)$ is the electrostatic potential at the position of ion i:

$$\phi(r_i) = \sum_{j,\mathbf{n}}{}' \frac{q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|}, \tag{12.1.2}$$

where the prime on the summation indicates that the sum is over all periodic images $\mathbf{n}$ and over all particles j, except $j = i$ if $\mathbf{n} = \mathbf{0}$. Note that we assume that particle i interacts with *all* its periodic images, but not, of course, with itself. Here, and in what follows, we use Gaussian units, because it makes the notation more compact. Equation (12.1.2) cannot be used to compute the electrostatic energy in a simulation, because it contains a poorly converging sum (in fact, the sum is only conditionally convergent). To improve the convergence of the expression for the electrostatic potential energy, we rewrite the expression for the charge density. In equation (12.1.2) we have represented the charge density as a sum of δ-functions. The contribution to the electrostatic potential due to these point charges decays as $1/r$. Now consider what happens if we assume that every particle i with charge q_i is surrounded by a diffuse charge distribution of the opposite sign, such that the total charge of this cloud exactly cancels q_i. In that case the electrostatic potential due to particle i is due exclusively to the fraction of q_i that is not screened. At large distances, this fraction rapidly goes to 0. How rapidly depends on the functional form of the screening charge distribution. In what follows, we shall assume a Gaussian distribution for the screening charge cloud.

The contribution to the electrostatic potential at a point r_i due to a set of screened charges can be easily computed by direct summation, because the electrostatic potential due to a screened charge is a rapidly decaying function of r. However, it was not our aim to evaluate the potential due to a set of *screened* charges but due to *point* charges. Hence, we must correct for the fact that we have added a screening charge cloud to every particle. This is shown schematically in Figure 12.1. This compensating charge density varies smoothly in space. We wish to compute the electrostatic energy at the site of ion i. Of course, we should exclude the electrostatic interaction of the ion with itself. We have three contributions to the electrostatic potential: first of all, the one due to the point charge q_i, secondly, the one due to the (Gaussian) *screening* charge cloud with charge $-q_i$, and finally the one due to the *compensating* charge cloud with charge q_i. In order to exclude Coulomb self-interactions, we should not include any of these three contributions to the electrostatic potential at the position of ion i. However, it turns out that it is convenient to retain the contribution due to the *compensating* charge distribution and correct for the resulting spurious interaction afterwards. The reason we retain the compensating charge cloud for ion i is that, if we do so, the compensating charge distribution is not only a smoothly varying function, but it is also periodic. Such a function can be represented by a (rapidly

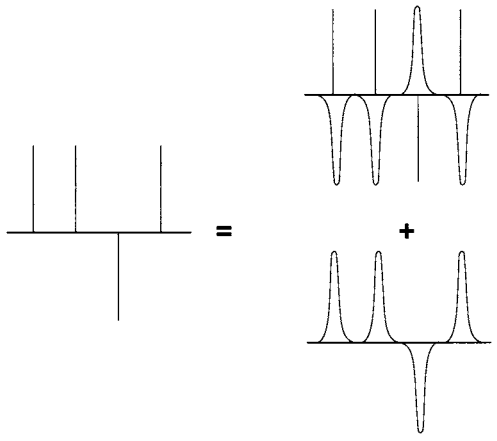

Figure 12.1: A set of point charges may be considered a set of screened charges minus the smoothly varying screening background.

converging) Fourier series, and this will turn out to be essential for the numerical implementation. Of course, in the end we should correct for the inclusion of a spurious "self" interaction between ion i and the compensating charge cloud.

After this brief sketch of the method for evaluating the electrostatic contribution to the potential energy, let us now consider the individual terms. We assume that the compensating charge distribution surrounding an ion i is a Gaussian with width $\sqrt{2/\alpha}$:

$$\rho_{\text{Gauss}}(r) = -q_i(\alpha/\pi)^{\frac{3}{2}} \exp(-\alpha r^2).$$

The choice of α will be determined later by considerations of computational efficiency. We shall first evaluate the contribution to the Coulomb energy due to the continuous background charge, then the spurious "self" term, and finally the real-space contribution due to the screened charges.

Fourier Transformation

This chapter relies heavily on the use of the Fourier transformation. It is therefore instructive to recall some of the basic equations in the context of electrostatics.

The central problem is to compute the energy of a given charge distribution $\rho(\mathbf{r})$. Formally, this corresponds to solving Poisson's equation for the electrostatic potential. Using the Gaussian notation,

$$-\nabla^2\phi(\mathbf{r}) = 4\pi\rho(\mathbf{r}), \tag{12.1.3}$$

where $\phi(\mathbf{r})$ is the electrostatic potential at point $\mathbf{r}$. For a single charge z at the origin the solution of this equation is the Coulomb potential

$$\phi(\mathbf{r}) = \frac{z}{4\pi|\mathbf{r}|}. \tag{12.1.4}$$

For a collection of N point charges we can define a charge density

$$\rho_P(\mathbf{r}) = \sum_{i=1}^{N} q_i \delta(\mathbf{r} - \mathbf{r}_i), \tag{12.1.5}$$

where $\mathbf{r}_i$ and q_i are the position and charge of particle i, respectively. The potential in a point $\mathbf{r}$ follows from a summation of the contributions of the particles

$$\phi(\mathbf{r}) = \sum_{i=1}^{N} \frac{q_i}{4\pi|\mathbf{r} - \mathbf{r}_i|}.$$

A different representation of these equations can be given in Fourier space. Let us consider a periodic system with a cubic box of length L and volume V. Any function $f(\mathbf{r})$ that depends on the coordinates of our system can be represented by a Fourier series:

$$f(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{l}=-\infty}^{\infty} \tilde{f}(k) e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{12.1.6}$$

where $\mathbf{k} = (2\pi/L)\mathbf{l}$ with $\mathbf{l} = (l_x, l_y, l_z)$ are the lattice vectors in Fourier space. The Fourier coefficients $\tilde{f}(k)$ are calculated using

$$\tilde{f}(\mathbf{k}) = \int_V d\mathbf{r}\, f(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}}. \tag{12.1.7}$$

In Fourier space Poisson's equation (12.1.3) has a much simpler form. We can write for the Poisson equation:

$$\begin{aligned} -\nabla^2\phi(\mathbf{r}) &= -\nabla^2 \left(\frac{1}{V} \sum_{\mathbf{k}} \tilde{\phi}(\mathbf{k}) e^{i\mathbf{r}\cdot\mathbf{k}} \right) \\ &= \frac{1}{V} \sum_{\mathbf{k}} k^2 \tilde{\phi}(\mathbf{k}) e^{i\mathbf{r}\cdot\mathbf{k}}. \end{aligned} \tag{12.1.8}$$

For the Fourier transform of the charge density we have

$$\rho(r) = \frac{1}{V} \sum_{\mathbf{k}} \tilde{\rho}(\mathbf{k}) e^{i\mathbf{r}\cdot\mathbf{k}}. \tag{12.1.9}$$

Substitution of equations (12.1.9) and (12.1.8) into equation (12.1.3) yields the Poisson equation in Fourier space

$$k^2\tilde{\phi}(k) = 4\pi\tilde{\rho}(k). \tag{12.1.10}$$

To find the solution of Poisson's equation for a point charge of strength z at the origin, we have to perform the Fourier transform of a delta function:

$$\begin{aligned}\tilde{\rho}(\mathbf{k}) &= \int_V d\mathbf{r}\, z\delta(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}} \\ &= z.\end{aligned}$$

This yields as solution for the Poisson equation

$$\tilde{\phi}(k) = \frac{4\pi z}{k^2}.$$

The solution for a unit charge is often called the Green's function:

$$\tilde{g}(k) = \frac{4\pi}{k^2}. \tag{12.1.11}$$

For a collection of point charges with charge density given by equation (12.1.5), we can write for the Fourier coefficients of the potential

$$\tilde{\phi}(k) = \tilde{g}(k)\tilde{\rho}_P(k)$$

with

$$\begin{aligned}\tilde{\rho}_P(k) &= \int_V d\mathbf{r} \sum_{i=1}^{N} q_i\delta(\mathbf{r}-\mathbf{r}_i)e^{-i\mathbf{k}\cdot\mathbf{r}} \\ &= \sum_{i=1}^{N} q_i e^{-i\mathbf{k}\cdot\mathbf{r}_i}.\end{aligned} \tag{12.1.12}$$

These equations show that in Fourier space the solution of Poisson's equation is simply obtained by multiplying $\tilde{\rho}(k)$ and $\tilde{g}(k)$ for all k vectors.

In the following we will also use another property of the Fourier transform. If we have a function $f_1(x)$, which is the convolution ($\star$) of two other functions $f_2(x)$ and $f_3(x)$,

$$f_1(x) = f_2(x) \star f_3(x) \equiv \int dx'\, f_2(x')f_3(x-x'),$$

then the Fourier coefficients of these functions are related by a simple multiplication:

$$\tilde{f}_1(k) = \tilde{f}_2(k)\tilde{f}_3(k).$$

For example, if we have a charge distribution that does not consist of simple point charges, but a more "smeared out" distribution,

$$\rho(\mathbf{r}) = \sum_i q_i \gamma(\mathbf{r} - \mathbf{r}_i) = \int d\mathbf{r}'\, \gamma(\mathbf{r}')\rho_p(\mathbf{r} - \mathbf{r}'), \tag{12.1.13}$$

where $\gamma(\mathbf{r})$ is the "shape" of the charge distribution of a single charge, then the Poisson equation for this system takes, in Fourier space, the form

$$\tilde{\phi}(k) = \tilde{g}(k)\tilde{\gamma}(k)\tilde{\rho}_P(k).$$

Fourier Part of Ewald Sum

We now apply the properties of the Poisson equation in Fourier form to compute the electrostatic potential at a point r_i due to a charge distribution $\rho_1(r)$ that consists of a periodic sum of Gaussians:

$$\rho_1(r) = \sum_{j=1}^{N} \sum_{\mathbf{n}} q_j (\alpha/\pi)^{\frac{3}{2}} \exp\left[-\alpha |\mathbf{r} - (\mathbf{r}_j + \mathbf{n}L)|^2\right].$$

To compute the electrostatic potential $\phi_1(r)$ due to this charge distribution, we use Poisson's equation:

$$-\nabla^2 \phi_1(r) = 4\pi\rho_1(r),$$

or in Fourier form,

$$k^2 \phi_1(k) = 4\pi\rho_1(k).$$

Fourier transforming the charge density ρ_1 yields

$$\begin{aligned}
\rho_1(\mathbf{k}) &= \int_V d\mathbf{r}\, \exp(-i\mathbf{k}\cdot\mathbf{r})\rho_1(\mathbf{r}) \\
&= \int_V d\mathbf{r}\, \exp(-i\mathbf{k}\cdot\mathbf{r}) \sum_{j=1}^{N} \sum_{\mathbf{n}} q_j (\alpha/\pi)^{\frac{3}{2}} \exp\left[-\alpha |\mathbf{r} - (\mathbf{r}_j + \mathbf{n}L)|^2\right] \\
&= \int_{\text{all space}} d\mathbf{r}\, \exp(-i\mathbf{k}\cdot\mathbf{r}) \sum_{j=1}^{N} q_j (\alpha/\pi)^{\frac{3}{2}} \exp\left[-\alpha |\mathbf{r} - \mathbf{r}_j|^2\right] \\
&= \sum_{j=1}^{N} q_j \exp(-i\mathbf{k}\cdot\mathbf{r}_j) \exp(-k^2/4\alpha).
\end{aligned} \tag{12.1.14}$$

If we now insert this expression in Poisson's equation, we obtain

$$\phi_1(k) = \frac{4\pi}{k^2} \sum_{j=1}^{N} q_j \exp(-i\mathbf{k}\cdot\mathbf{r}_j) \exp(-k^2/4\alpha), \tag{12.1.15}$$

where it should be noted that this expression is defined only for $\mathbf{k} \neq 0$. This is a direct consequence of the conditional convergence of the Ewald sum. For the time being, we shall assume that the term with $\mathbf{k} = 0$ is equal to 0. As we shall see later, this assumption is consistent with a situation where the periodic system is embedded in a medium with infinite dielectric constant.

We now compute the contribution to the potential energy due to ϕ_1, using equation (12.1.1). To this end, we first compute $\phi_1(r)$:

$$\begin{aligned} \phi_1(r) &= \frac{1}{V}\sum_{\mathbf{k}\neq 0}\phi_1(\mathbf{k})\exp(i\mathbf{k}\cdot\mathbf{r}) \qquad (12.1.16)\\ &= \sum_{\mathbf{k}\neq 0}\sum_{j=1}^{N}\frac{4\pi q_j}{k^2}\exp[i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_j)]\exp(-k^2/4\alpha), \end{aligned}$$

and hence,

$$\begin{aligned} \mathcal{U}_1 &\equiv \frac{1}{2}\sum_i q_i\phi_1(r_i)\\ &= \frac{1}{2}\sum_{\mathbf{k}\neq 0}\sum_{i,j=1}^{N}\frac{4\pi q_i q_j}{Vk^2}\exp[i\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_j)]\exp(-k^2/4\alpha)\\ &= \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\frac{4\pi}{k^2}|\rho(\mathbf{k})|^2\exp(-k^2/4\alpha), \qquad (12.1.17) \end{aligned}$$

where we have used the definition

$$\rho(\mathbf{k}) \equiv \sum_{i=1}^{N} q_i \exp(i\mathbf{k}\cdot\mathbf{r}_i). \qquad (12.1.18)$$

Correction for Self-Interaction

The contribution to the potential energy given in equation (12.1.17) includes a term $(1/2)q_i\phi_{self}(r_i)$ due to the interaction between a continuous Gaussian charge cloud of charge q_i and a point charge q_i located at the center of the Gaussian. This term is spurious, and we should correct for it. We therefore must compute the potential energy at the origin of a Gaussian charge cloud. The charge distribution that we have overcounted is

$$\rho_{Gauss}(r) = q_i(\alpha/\pi)^{\frac{3}{2}}\exp(-\alpha r^2).$$

We can compute the electrostatic potential due to this charge distribution using Poisson's equation. Using the spherical symmetry of the Gaussian charge cloud, we can write Poisson's equation as

$$-\frac{1}{r}\frac{\partial^2 r\phi_{Gauss}(r)}{\partial r^2} = 4\pi\rho_{Gauss}(r)$$

or

$$-\frac{\partial^2 r\phi_{\text{Gauss}}(r)}{\partial r^2} = 4\pi r\rho_{\text{Gauss}}(r).$$

Partial integration yields

$$\begin{aligned} -\frac{\partial r\phi_{\text{Gauss}}(r)}{\partial r} &= \int_\infty^r dr\, 4\pi r\rho_{\text{Gauss}}(r) \\ &= -2\pi q_i(\alpha/\pi)^{\frac{3}{2}} \int_r^\infty dr^2\, \exp(-\alpha r^2) \\ &= -2q_i(\alpha/\pi)^{\frac{1}{2}} \exp(-\alpha r^2). \end{aligned} \tag{12.1.19}$$

A second partial integration gives

$$\begin{aligned} r\phi_{\text{Gauss}}(r) &= 2q_i(\alpha/\pi)^{\frac{1}{2}} \int_0^r dr\, \exp(-\alpha r^2) \\ &= q_i \text{erf}\left(\sqrt{\alpha} r\right), \end{aligned} \tag{12.1.20}$$

where, in the last line, we have employed the definition of the error function: $\text{erf}(x) \equiv (2/\sqrt{\pi}) \int_0^x \exp(-u^2)\, du$. Hence,

$$\phi_{\text{Gauss}}(r) = \frac{q_i}{r} \text{erf}\left(\sqrt{\alpha} r\right). \tag{12.1.21}$$

To compute the spurious self term to the potential energy, we must compute $\phi_{\text{Gauss}}(r)$ at $r = 0$. It is easy to verify that

$$\phi_{\text{Gauss}}(r = 0) = 2q_i(\alpha/\pi)^{\frac{1}{2}}.$$

Hence, the spurious contribution to the potential energy is

$$\begin{aligned} \mathcal{U}_{\text{self}} &= \frac{1}{2} \sum_{i=1}^N q_i \phi_{\text{self}}(r_i) \\ &= (\alpha/\pi)^{\frac{1}{2}} \sum_{i=1}^N q_i^2. \end{aligned} \tag{12.1.22}$$

The spurious self-interaction $\mathcal{U}_{\text{self}}$ should be *subtracted* from the sum of the real-space and Fourier contributions to the Coulomb energy. Note that equation (12.1.22) does not depend on the particle positions. Hence, during a simulation, this term is constant, provided that the values of all (partial) charges (and particles) in the system remain fixed.

Real-Space Sum

Finally, we must compute the electrostatic energy due to the point charges screened by oppositely charged Gaussians. Using the results of section 12.1.1, in particular equation (12.1.21), we can immediately write the (short-range) electrostatic potential due to a point charge q_i surrounded by a Gaussian with net charge $-q_i$:

$$\begin{aligned}\phi_{\text{short-range}}(r) &= \frac{q_i}{r} - \frac{q_i}{r}\text{erf}\left(\sqrt{\alpha}r\right) \\ &= \frac{q_i}{r}\text{erfc}\left(\sqrt{\alpha}r\right), \end{aligned} \tag{12.1.23}$$

where the last line defines the complementary error function $\text{erfc}(x) \equiv 1 - \text{erf}(x)$. The total contribution of the screened Coulomb interactions to the potential energy is then given by

$$\mathcal{U}_{\text{short-range}} = \frac{1}{2}\sum_{i\neq j}^{N} q_i q_j \text{erfc}\left(\sqrt{\alpha}r_{ij}\right)/r_{ij}. \tag{12.1.24}$$

The total electrostatic contribution to the potential energy now becomes the sum of equations (12.1.17), (12.1.22), and (12.1.24):

$$\begin{aligned}\mathcal{U}_{\text{Coul}} &= \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\frac{4\pi}{k^2}\left|\rho(\mathbf{k})\right|^2 \exp(-k^2/4\alpha) \\ &\quad - (\alpha/\pi)^{\frac{1}{2}}\sum_{i=1}^{N} q_i^2 \\ &\quad + \frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j \text{erfc}\left(\sqrt{\alpha}r_{ij}\right)}{r_{ij}}. \end{aligned} \tag{12.1.25}$$

12.1.2 Dipolar Particles

It is straightforward to derive the corresponding expressions for the potential energy of a system containing dipolar molecules. The only modification is that we must everywhere replace q_i by $-\boldsymbol{\mu}_i \cdot \boldsymbol{\nabla}_i$. For example, the electro-

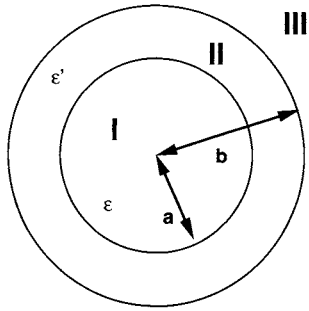

Figure 12.2: Spherical dielectric surrounded by a sphere.

static energy of a dipolar system becomes

$$\begin{aligned} \mathcal{U}_{\text{dipolar}} &= \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\frac{4\pi}{k^2}|\mathbf{M}(\mathbf{k})|^2\exp(-k^2/4\alpha) \\ &\quad -\frac{2\pi}{3}(\alpha/\pi)^{\frac{3}{2}}\sum_{i=1}^{N}\mu_i^2 \\ &\quad +\frac{1}{2}\sum_{i\neq j}^{N}\left[(\boldsymbol{\mu}_i\cdot\boldsymbol{\mu}_j)B(r_{ij})-(\boldsymbol{\mu}_i\cdot\mathbf{r}_{ij})(\boldsymbol{\mu}_j\cdot\mathbf{r}_{ij})C(r_{ij})\right], \end{aligned} \tag{12.1.26}$$

where

$$\begin{aligned} B(r) &\equiv \frac{\operatorname{erfc}\left(\sqrt{\alpha}r\right)}{r^3}+2(\alpha/\pi)^{\frac{1}{2}}\frac{\exp(-\alpha r^2)}{r^2}, \\ C(r) &\equiv 3\frac{\operatorname{erfc}\left(\sqrt{\alpha}r\right)}{r^5}+2(\alpha/\pi)^{\frac{1}{2}}(2\alpha+3/r^2)\frac{\exp(-\alpha r^2)}{r^2}, \end{aligned}$$

and

$$\mathbf{M}(\mathbf{k})\equiv\sum_{i=1}^{N}i\boldsymbol{\mu}_i\cdot\mathbf{k}\exp(i\mathbf{k}\cdot\mathbf{r}_i).$$

Again, this expression applies to a situation where the periodic system is embedded in a material with infinite dielectric constant.

12.1.3 Dielectric Constant

To derive an expression for the dielectric constant of a polar fluid, we consider the system shown in Figure 12.2: a large spherical dielectric with radius a and dielectric constant ϵ (region I) surrounded by a much larger

sphere with radius b and dielectric constant ϵ' (region II). The entire system is placed in vacuum (region III), and an electric field $\mathbf{E}$ is applied. The potential at a given point in this system follows from the solution of the Poisson equation with the appropriate boundary conditions (continuity of the normal component of the displacement $\mathbf{D}$ and tangential component of the electric field $\mathbf{E}$) at the two boundaries between regions I and II, and II and III. In the limit $a \to \infty$, $b \to \infty$, $a/b \to 0$, we can write, for the electric field in region I,

$$\mathbf{E}_I = \frac{9\epsilon'}{(\epsilon'+2)(2\epsilon'+\epsilon)}\mathbf{E}, \tag{12.1.27}$$

which gives, for the polarization $\mathbf{P}$,

$$\mathbf{P} \equiv \frac{\epsilon-1}{4\pi}\mathbf{E}_I = \frac{9\epsilon'(\epsilon-1)}{4\pi(\epsilon'+2)(2\epsilon'+\epsilon)}\mathbf{E}. \tag{12.1.28}$$

In order to make contact with linear response theory, we should compute the polarization of the system as a function of the applied *external* field, i.e., the electric field that would be present in the system in the absence of the particles. Using equation (12.1.27), it is easy to derive that the electrostatic field $\mathbf{E}_I'$ that would be present in region I if it were empty is given by equation (12.1.27) with $\epsilon = 1$:

$$\mathbf{E}_I' = \frac{9\epsilon'}{(\epsilon'+2)(2\epsilon'+1)}\mathbf{E}.$$

The field $\mathbf{E}_I'$ is uniform throughout region I. If we assume that the system is isotropic and that linear response theory is sufficient, we can write for the polarization

$$\begin{aligned}\langle \mathbf{P} \rangle &= \frac{1}{VQ}\int d\mathbf{r}^N \sum_{i=1}^N \boldsymbol{\mu}_i \exp\left[-\beta\left(\mathcal{H}_0 - \sum_{i=1}^N \boldsymbol{\mu}_i \cdot \mathbf{E}_I'\right)\right] \\ &= \frac{\beta}{3V}\left(\langle \mathbf{M}^2\rangle - \langle \mathbf{M}\rangle^2\right)\mathbf{E}_I'. \end{aligned} \tag{12.1.29}$$

Comparison of equations (12.1.29) and (12.1.28) yields

$$\langle \mathbf{P}\rangle = \frac{1}{3}\beta\rho g_k \mu^2 \mathbf{E}_I', \tag{12.1.30}$$

where the g_k is the Kirkwood factor, which is defined as

$$g_k \equiv \frac{1}{N\mu^2}\left(\langle \mathbf{M}^2\rangle - \langle \mathbf{M}\rangle^2\right),$$

where $\mathbf{M}$ is the total dipole moment

$$\mathbf{M} = \sum_{i=1}^N \boldsymbol{\mu}_i.$$

Combining equations (12.1.28) and (12.1.30) gives

$$\frac{(\epsilon - 1)(2\epsilon' + 1)}{(2\epsilon' + \epsilon)} = \frac{4}{3}\pi\beta\rho g_k \mu^2.$$

For a simulation with conducting boundary conditions ($\epsilon' \to \infty$), the expression for the dielectric constant becomes

$$\epsilon = 1 + \frac{4}{3}\pi\rho\beta g_k \mu^2. \tag{12.1.31}$$

This result shows that the fluctuations of the dipole moment depends on the dielectric constant of the *surrounding* medium. This, in turn, implies that, for a polar system, the Hamiltonian itself depends on the dielectric constant ϵ' of the surrounding medium.

12.1.4 Boundary Conditions

It may appear strange that the form for the potential energy of an infinite periodic system of ions or dipoles should depend on the nature of the boundary conditions at infinity. However, for systems of charges or dipoles, this is a very real effect that has a simple physical interpretation. To see this, consider the system shown in Figure 12.2. The fluctuating dipole moment of the unit cell $\mathbf{M}$ gives rise to a surface charge at the boundary of the sphere, which, in turn, is responsible for a homogeneous depolarizing field:

$$\mathbf{E} = -\frac{4\pi\mathbf{P}}{2\epsilon' + 1},$$

where $\mathbf{P} \equiv \mathbf{M}/V$. Now let us consider the reversible work per unit volume that must be performed against this depolarizing field to create the net polarization $\mathbf{P}$. Using

$$dw = -\mathbf{E}d\mathbf{P} = \frac{4\pi}{2\epsilon' + 1}\mathbf{P}d\mathbf{P},$$

we find that the total work needed to polarize a system of volume V equals

$$\mathcal{U}_{\text{pol}} = \frac{2\pi}{2\epsilon' + 1}P^2 V = \frac{2\pi}{2\epsilon' + 1}M^2/V$$

or, using the explicit expression for the total dipole moment of the periodic box,

$$\mathcal{U}_{\text{pol}} = \frac{2\pi}{(2\epsilon' + 1)V}\left|\sum_{i=1}^{N}\mathbf{r}_i q_i\right|^2,$$

in the Coulomb case, and

$$\mathcal{U}_{\text{pol}} = \frac{2\pi}{(2\epsilon' + 1)V} \left| \sum_{i=1}^{N} \boldsymbol{\mu}_i \right|^2 ,$$

in the dipolar case. This contribution to the potential energy corresponds to the $\mathbf{k} = \mathbf{0}$ term that we have neglected thus far. It is permissible to ignore this term if the depolarizing field vanishes. This is the case if our periodic system is embedded in a medium with infinite dielectric constant (a conductor, $\epsilon' \to \infty$), which is what we have assumed throughout.

For simulations of ionic systems, it is essential to use such "conducting" boundary conditions; for polar systems, it is merely advantageous. For a discussion of these subtle points, see [325].

12.1.5 Accuracy and Computational Complexity

In the Ewald summation, the calculation of the energy is performed in two parts: the real-space part (12.1.23) and the part in Fourier space (12.1.17). For a given implementation, we have to choose the parameter α that characterizes the shape of the Gaussian charge distributions, r_c the real-space cutoff distance, and k_c the cutoff in Fourier space. In fact, it is common to write k_c as $2\pi/Ln_c$, where n_c is a positive integer. The total number of Fourier components within this cutoff value is equal to $(4\pi/3)n_c^3$. The values of these parameters depend on the desired accuracy ϵ, that is, the root mean-squared difference between the exact Coulombic energy and the results from the Ewald summation. Expressions for the cutoff errors in the Ewald summation method[1] have been derived in [326, 327]. For the energy, the standard deviation of the real-space cutoff error of the total energy is

$$\delta E_R \approx Q \left(\frac{r_c}{2L^3} \right)^{\frac{1}{2}} \frac{1}{(\alpha r_c)^2} \exp\left(-\alpha^2 r_c^2\right) \tag{12.1.32}$$

and for the Fourier part of the total energy

$$\delta E_F \approx Q \frac{n_c^{1/2}}{\alpha L^2} \frac{1}{(\pi n_c/\alpha L)^2} \exp\left[-(\pi n_c/\alpha L)^2\right], \tag{12.1.33}$$

where

$$Q = \sum_i q_i^2 .$$

Note that for both the real-space part and the Fourier part, the strongest dependence of the estimated error on the parameters α, r_c, and n_c is through

[1] The accuracy is dependent on whether we focus on the energy (for Monte Carlo) or on the forces (for Molecular Dynamics).

a function of the form $\exp(-x^2)/x^2$. We now impose that these two functions have the same value ϵ. The value of x for which $\exp(-x^2)/x^2 = \epsilon$ we denote by s. Hence $\epsilon = \exp(-s^2)/s^2$. Then it follows from Equation (12.1.32) that

$$r_c = \frac{s}{\alpha} \tag{12.1.34}$$

and from equation (12.1.33) we obtain

$$n_c = \frac{sL\alpha}{\pi}. \tag{12.1.35}$$

If we insert these expressions for r_c and n_c back into the expressions (12.1.32) and (12.1.33), we find that both errors have the same functional form:

$$\delta E_R \approx Q\left(\frac{s}{\alpha L^3}\right)^{1/2} \frac{\exp(-s^2)}{s^2}$$

and

$$\delta E_F \approx Q\left(\frac{s}{2\alpha L^3}\right)^{1/2} \frac{\exp(-s^2)}{s^2}.$$

Hence, changing s affects both errors in the same way. We now estimate the computational effort involved in evaluating the Ewald sum. To this end, we write the total computational time as the sum of the total time in real space and the total time in Fourier space

$$\tau = \tau_R N_R + \tau_F N_F, \tag{12.1.36}$$

where τ_R is the time needed to evaluate the real part of the potential of a pair of particles and τ_F is the time needed to evaluate the Fourier part of the potential per particle and per k vector. N_R and N_F denote the number of times these terms need to be evaluated to determine the total energy or the force on the particles. If we assume a uniform distribution of particles, these two numbers follow from the estimates of r_c and n_c:

$$\begin{aligned} N_R &= \frac{4}{3}\pi\frac{s^3N^2}{\alpha^3L^3} \\ N_F &= \frac{4}{3}\pi\frac{s^3\alpha^3L^3N}{\pi^3}. \end{aligned}$$

The value of α follows from minimization of equation (12.1.36)

$$\alpha = \left(\frac{\tau_R\pi^3N}{\tau_FL^6}\right)^{\frac{1}{6}},$$

which yields for the time

$$\tau = \frac{8\sqrt{\tau_r\tau_f}N^{3/2}s^3}{3\sqrt{\pi}} = \mathcal{O}\left(N^{3/2}\right). \tag{12.1.37}$$

Note that, with the above expression for α, the parameters r_c and n_c follow from equations (12.1.34) and (12.1.35) respectively, once we have specified the desired accuracy. To optimize the Ewald summation one has to make an estimate of τ_R/τ_F. This ratio depends on the details of the particular implementation of the Ewald summation and can be obtained from a short simulation.[2]

We conclude this section with a few comments concerning the implementation. First of all, when using equation (12.1.34) to relate r_c to α, one should make sure that $r_c \leq L/2$; otherwise the real part of the energy cannot be restricted to the particles in the box $\mathbf{n} = \mathbf{0}$.

A second practical point is the following: in most simulations, there are short-range interactions between the particles, in addition to the Coulomb interaction. Usually, these short-range interactions also have a cutoff radius. Clearly, it is convenient if the same cutoff radius can be used for the short-range interactions and for the real-space part of the Ewald summation. However, if this is done, the parameters of the Ewald summation need not have their optimum values.

12.2 Fast Multipole Method

An algorithm that is of order $\mathcal{O}(N)$ is the fast multipole method. The multipole method is based on the idea that a group of particles at a large distance can be considered one big cluster, for which it is not necessary to calculate all particle-particle interactions individually. By clustering the system into bigger and bigger groups, the interactions can be approximated. This approach of Appel [329] leads to an order $\mathcal{O}(N)$ algorithm [330]. This algorithm was further refined by Barnes and Hut [331]. In the original algorithm of Appel, the clusters were approximated as a single charge. Greengard and Rokhlin [316] developed an algorithm in which the charge distribution in a cluster is approximated by a multipole expansion. Schmidt and Lee extended this method to systems with periodic boundary conditions [332].

Algorithm

Next, we give a schematic description of the Greengard and Rokhlin algorithm in three dimensions; a more detailed description can be found in [328, 332, 333]. Essential in this algorithm is the use of octal trees and multipole expansions, which are described before we discuss the algorithm.

An octal tree can be constructed in the following way (see Figure 12.3). The original system is defined to be the unique level-zero cell. Level-one

[2] A typical value of this ratio is $\tau_R/\tau_F = 3.6$ [328].

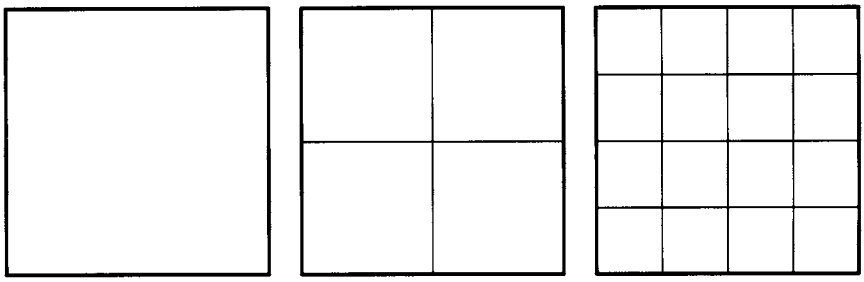

Figure 12.3: Octal tree in two dimensions: the left figure shows the original level-zero system, the middle figure the level-one cells, and the right figure the level-two cells.

cells are obtained by dividing the parent cell into eight (four in two dimensions) children. This subdivision can be continued till a maximum level, denoted level R, has been obtained. We refer to a cell as $C_i^{(l)}$, where l denotes its level and i the index that refers to its position.

In the Greengard and Rokhlin method various multipole expansions are used. The formula for these expansions are given, for example, in [328]; here we give a short description of the four formulas that are used. In the description of the algorithm we refer to the operation by the words in italics.

- *Multipole expansion* of the charges in a cell. Suppose we have a distribution of k charges at location $\mathbf{r}_i = (r_i, \theta_i, \phi_i)$ with charge q_i, then the potential at a point $\mathbf{r}' = (r', \theta', \phi')$ sufficiently far away is given by

$$\Phi(\mathbf{r}') = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{M_n^m}{r^{n+1}} Y_n^m(\theta', \phi') \tag{12.2.1}$$

$$M_n^m = \sum_{i=1}^{k} q_i r_i^n Y_n^{-m}(\theta_i, \phi_i), \tag{12.2.2}$$

  where Y_n^{-m} are the associated Legendre functions.
- *Multipole translation.* To calculate the multipole expansion of the parent cell, for example, one can use the multipole expansion of its children. However, for each of these children, the origin of this expansion (12.2.1) is different for each cell. For example, if multipole expansions have been made for the children, *multipole translations* are used to translate these expansions to a new, common origin, the center of the parent box. This translated multipole can be used only for sufficiently far-away positions.
- *Local expansion.* To calculate the potential energy of all particles in a cell one can determine the potential due to the multipoles in other

cells. This involves a conversion of the multipole expansions into a local expansion that converges in the region of interest. Since, in these multipole expansions, the close-by cells are excluded, within a cell the potential energy does not change much; therefore it is advantageous to make a Taylor expansion of the local field in each cell. This field can then be used to calculate the energy for all atoms in this cell. This expansion can be used only for positions in the same cell.

- *Translation of the local expansion.* If we have calculated the local expansion with respect to a given origin, for example, the local expansion of the parent box, we can translate this local expansion to, for example, the center of its children cells.

The Greengard and Rokhlin algorithm consists of the following steps:

1. An octal tree of level R is constructed; the charge distribution in each of the 8^R cells is described with a multipole expansion (12.2.1) about the center of a box.
2. The multipole expansion of each of the children of a parent is translated to a multipole expansion around the center of the parent box. This procedure is repeated until level zero has been reached.
3. The calculation of the energy is done in two steps. Consider a particle in level-R cell $C_i^{(R)}$. The calculation of the interactions with the particles in the same cell and in the 26 neighboring level-R cells is performed directly:

$$u_{clo} = \sum_{close} \frac{q_i q_j}{r_{ij}}.$$

The interactions with the remainder of the particles are done by calculating the local field expansion in cell $C_i^{(R)}$. To calculate this local field, the following steps are used starting from level two (for levels zero and one, a local expansion does not exist, and therefore they are by definition zero, since there are no cells sufficiently far away).

   (a) For each cell in level r, transform the local expansion of the parent cell to the center of the current cell.
   (b) Add to the local potentials the transformation of the multipole potentials of the same level cells that fulfill the following conditions:
      - The cell must be sufficiently far away for the multipole expansion to be valid; this excludes the neighboring cells.
      - The cells must be sufficiently close that their contribution has not been included in the local field of the parent of the cell;

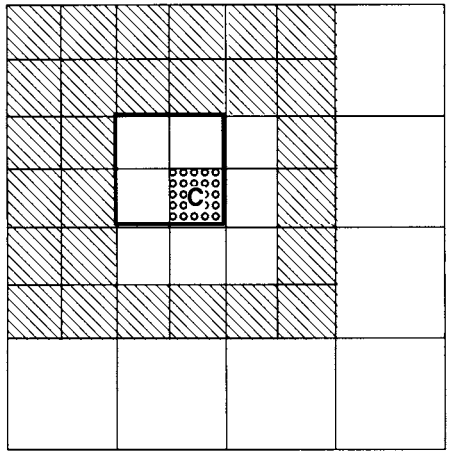

Figure 12.4: The same level cells (shaded) of which the multipole potentials are transferred to the current cell to give the local expansion. The current cell is denoted by C, the neighbors of C do not contribute because they touch C. The white cells of the parent level did not touch the parent cell of C (indicated with a thick line) and therefore did contribute to the local expansion of the parent cell.

this excludes the cells that contribute to local potential initialized in step 3a. In Figure 12.4 these cells are shown. In three dimensions, $8 \times 27 - 27 = 189$ of these cells fulfill this condition.

After level R has been reached the far-way contribution of the charges in cell $C_i^{(R)}$ is calculated from the local expansion $U^{Lo(R)}$:

$$u_{far} = \sum_i U^{Lo}(\mathbf{r}_i).$$

It is important to note that the local expansion $U^{Lo(R)}$ can be used for all particles in the same cell. Therefore in practice the calculation of this term is usually done after step 2 of this algorithm.

The preceding algorithm is for a finite system of charge. To apply this algorithm to a system with periodic boundary conditions, the following modifications have to be made. In a system with periodic boundary conditions the level-zero cell is surrounded by periodic images that have the same multipole expansion around their center. In the algorithm we have started step 3a, assuming that the local field of level zero is 0. For the periodic system, this

has to be replaced by the local expansion of the potential from all images except the nearest 26 neighbors. If we have calculated this term, the remainder of the algorithm remains the same. The summation over all periodic images is done using an Ewald summation technique (see [332] for details). Since this involves the multipole expansion of the parent cell (which does not depend on the number of particles), the overhead in computer time is very small. However, the timings of Schmidt and Lee do show that to arrive at the same accuracy as for the nonperiodic system one needs to use a higher octal tree or more multipole moments.

12.3 Particle Mesh Approaches

The CPU time required for a fully optimized Ewald summation scales with the number of particles as $\mathcal{O}(N^{3/2})$. In many applications we not only have the long-range interactions but short-range interactions as well. For such systems it may be convenient to use the same cutoff radius for the real-space sum in the Ewald summation as for the short-range interactions. For a fixed cutoff, however, the calculation of the Fourier part of the Ewald summation scales as $\mathcal{O}(N^2)$, which makes the Ewald summation inefficient for large systems. Note that it is only the reciprocal-space part of the Ewald sum that suffers from this drawback. Clearly, it would be advantageous to have an approach that handles the Fourier part more efficiently. Several schemes for solving this problem have been proposed. They all exploit the fact that the Poisson equation can be solved much more efficiently if the charges are distributed on a mesh. The efficiency and accuracy of such mesh-based algorithms depend strongly on the way in which the charges are attributed to mesh points. Below, we briefly discuss the basics of the particle-mesh approach. However, a full description of the technical details is beyond the scope of this book.

The earliest particle-mesh scheme for molecular simulations was developed by Hockney and Eastwood [24]. The charges in the systems were interpolated on a grid to arrive at a discretized Poisson equation. For a regular grid this equation can be solved efficiently using the Fast Fourier Transform (FFT) [33]. The computer time associated with the FFT technique scales as $\mathcal{O}(N \log N)$, where N denotes the number of points of discrete Fourier transform. In its simplest implementation, the particle-mesh method is fast, but not very accurate. The technique was subsequently improved by splitting the calculation into a short-range and a long-range contribution. In the spirit of the Ewald method, the short-range part is then calculated directly from the particle-particle interactions while the particle-mesh technique is used for the long-range contribution.

Below, we briefly discuss the particle-mesh methods and their relation to the Ewald-sum approach. We will not attempt to present an exhaustive description of all existing particle-mesh methods. The reason is twofold: first of all, a systematic presentation of this subject—as given, for instance, in a paper by Deserno and Holm [334]—would require a chapter on its own. Secondly, Deserno and Holm [334] have shown that most of the "good" alternative methods, such as the Particle Mesh Ewald (PME) [317] and Smooth Particle Mesh Ewald (SPME) [335], are very similar in spirit and can be seen as variations of the original particle-particle/particle-mesh (PPPM) technique of Hockney and Eastwood [24]. The choice of the method to use, often depends on the application. For example, Monte Carlo simulations require an accurate estimate of the energy, while in Molecular Dynamics simulations we need to compute the forces accurately. Some particle-mesh schemes are better suited to do one, some to do the other.

The idea of the PPPM method is to split the Coulomb potential into two parts by using the following (trivial) identity:

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1-f(r)}{r}. \tag{12.3.1}$$

The idea of using a switching function is similar to the splitting of the Ewald summation into a short-range and a long-range part. Pollock and Glosli [336] found that different choices for $f(r)$ yield comparable results, although the efficiency of the method does depend strongly on a careful choice of this function. Darden *et al.* [317] have shown that, if one uses the same Gaussian screening function as in the Ewald summation, the PPPM technique becomes indeed very similar to the Ewald method.

It is instructive to recall the Fourier-space contribution of the energy:

$$U_1 = \frac{1}{2V} \sum_{\mathbf{k}\neq 0} \frac{4\pi}{k^2} |\tilde{\rho}(\mathbf{k})|^2 \exp(-k^2/4\alpha).$$

Following Deserno and Holm [334], we write the Fourier-space contribution as

$$\begin{aligned} U_1 &= \frac{1}{2} \sum_{i=1}^{N} q_i \left(\frac{1}{V} \sum_{k\neq 0} \tilde{g}(\mathbf{k}) \tilde{\gamma}(\mathbf{k}) \tilde{\rho}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}_i} \right) \\ &= \frac{1}{2} \sum_{i=1}^{N} q_i \phi^k(\mathbf{r}_i), \end{aligned} \tag{12.3.2}$$

where $\phi^k(\mathbf{r}_i)$ can be interpreted as the electrostatic potential due to the second term in equation (12.3.1):

$$\phi^k(\mathbf{r}_i) = \frac{1}{V} \sum_{\mathbf{k}} \tilde{g}(\mathbf{k}) \tilde{\gamma}(\mathbf{k} \neq 0) \tilde{\rho}(\mathbf{k}) \exp(i\mathbf{k}\cdot\mathbf{r}_i).$$

As a product in Fourier space corresponds to a convolution in real space, we see that the potential $\phi^k(\mathbf{r}_i)$ is due to the original charge distribution $\rho(x)$, convoluted by a smearing function $\gamma(r)$. The Ewald summation is recovered if we choose a Gaussian smearing function; in which case $f(r)$ is given by an error function.

In order to evaluate the above expression for the Fourier part of the electrostatic energy using a discrete fast Fourier transform, we have to perform the following steps [334,337]:

1. Charge assignment: Up to this point, the charges in the system are not localized on lattice points. We now need a prescription to assign the charges to the grid points.
2. Solving Poisson's equation: Via a FFT technique the Poisson equation for our discrete charge distribution has to be solved (the Poisson equation on a lattice can also be solved efficiently, using a diffusion algorithm [338].
3. Force assignment (in the case of MD): Once the electrostatic energy has been obtained from the solution of the Poisson equation, the forces have to be calculated and assigned back to the particles in our system.

At every stage, there are several options to choose from. Deserno and Holm have made a careful study of the relative merits of the various options and their combinations [334]. Below we give a brief summary of their observations. For further details the reader is referred to the original article.

To assign the charges of the system to a grid, a charge assignment function, $W(\mathbf{r})$, is introduced. For example, in a one-dimensional system, the fraction of a unit charge at position x assigned to a grid point at position x_p is given by $W(x_p - x)$. Hence, if we have a charge distribution $\rho(x) = \sum_i q_i \delta(x - x_i)$, then the charges at a grid point x_p are given by

$$\rho_M(x_p) = \frac{1}{h} \int_0^L dx\, W(x_p - x)\rho(x), \tag{12.3.3}$$

where L is the box diameter and h is the mesh spacing. The number of mesh points in one dimension, M, is equal to L/h. The factor $1/h$ ensures that ρ_M is a density. Many choices for the function $W(x)$ are possible. Deserno and Holm [334] listed the properties that $W(x)$ should have. $W(x)$ should be an even function and the function should be normalized in such a way that the sum of the fractional charges equals the total charge of the system. Since the computational costs is proportional to the number of particles and the number of mesh points to which the single charge is distributed, a function with a small support decreases the computational cost. In addition, one would like to reduce the errors due to the discretization as much as possible. If a particle moves through the system and passes from one grid point to

another, the function $W(x)$ should not yield abrupt changes in the fractional charges.

A particularly nice way to approach the charge assignment problem was described by Essmann *et al.* [335]. These authors argue that the problem of discretizing the Fourier transform can be viewed as an interpolation problem. Consider a single term in the (off-lattice) Fourier sum $q_i e^{-i\mathbf{k}\cdot\mathbf{r}_i}$. This term cannot be used in a discrete Fourier transform, because $\mathbf{r}$ does not, in general, correspond to a mesh point. However, we can *interpolate* $e^{-i\mathbf{k}\cdot\mathbf{r}_i}$ in terms of values of the complex exponential at mesh points. For convenience, consider a one-dimensional system. Moreover, let us assume that x varies between 0 and L and that there are M equidistant mesh points in this interval. Clearly, the particle coordinate x_i is located between mesh points $[Mx_i/L]$ and $[Mx_i/L]+1$, where $[..]$ denotes the integer part of a real number. Let us denote the real number Mx_i/L by u_i. We can then write an order-$2p$ interpolation of the exponential as

$$e^{-ik_x x_i} \approx \sum_{j=-\infty}^{\infty} W_{2p}(u_i - j) e^{-ik_x Lj/M},$$

where the W_{2p}'s denote the interpolation coefficients. Strictly speaking the sum over j contains only M terms. However, to account for the periodic boundary conditions, we have written it as if $-\infty < j < \infty$. For an interpolation of order $2p$, only the $2p$ mesh point nearest to x_i contributes to the sum. For all other points, the weights W_{2p} vanish. We can now approximate the Fourier transform of the complete charge density as

$$\rho_k \approx \sum_{i=1}^{N} q_i \sum_{j=-\infty}^{\infty} W_{2p}(u_i - j) e^{-ik_x Lj/M}.$$

This can be rewritten as

$$\rho_k \approx \sum_{j} e^{-ik_x Lj/M} \sum_{i=1}^{N} q_i W_{2p}(u_i - j).$$

We can interpret the above expression as a *discrete* Fourier transform of a "meshed" charge density $\rho(j) = \sum_{i=1}^{N} q_i W_{2p}(u_i - j)$. This shows that the coefficients W_{2p} that were introduced to give a good interpolation of $e^{-ik_x x_i}$ end up as the charge-assignment coefficients that attribute off-lattice charges to a set of lattice points.

While the role of the coefficients W is now clear, there are still several choices possible. The most straightforward is to use the conventional Lagrange interpolation method to approximate the exponential (see Darden *et al.* [317] and Petersen [327]). The Lagrange interpolation scheme is useful for

Monte Carlo simulations, but less so for the Molecular Dynamics method. The reason is that although the Lagrangian coefficients are everywhere continuous, their derivative is not. This is problematic when we need to compute the forces acting on charged particles (the solution is that a separate interpolation must be used to compute the forces). To overcome this drawback of the Lagrangian interpolation scheme, Essmann *et al.* suggested the so-called SPME method [335]. The SPME scheme uses exponential Euler splines to interpolate complex exponentials. This approach results in weight functions W_{2p} that are $2p-2$ times continuously differentiable. It should be stressed that we cannot automatically use the *continuum* version of Poisson's equation in all interpolation schemes. In fact equation (12.1.11) is only consistent with the Lagrangian interpolation schemes. To minimize discretization errors, other schemes, such as the SPME method, require other forms of the Green's function $\tilde{g}(k)$ (see ref. [334]).

In section 12.1.5 we discussed how the parameter α in the conventional Ewald sum method can be chosen such that it minimizes the numerical error in the energy (or in the forces). Petersen [327] has derived similar expressions for the PME method. Expressions that apply to the PPPM method [24] and the SPME scheme are discussed by Deserno and Holm [339]. In the case of the force computation, matters are complicated by the fact that, in a particle-mesh scheme, there are several inequivalent ways to compute the electrostatic forces acting on the particles. Some such schemes do not conserve momentum, others do—but at a cost. The choice of what is the "best" method, depends largely on the application [334].

This concludes our discussion of particle-mesh schemes. While we have tried to convey the spirit of these algorithms, we realize that this description is not sufficiently detailed to be of any help in the actual implementation of such an algorithm. We refer readers who are considering implementation of one of the particle-mesh schemes to the articles of Essmann *et al.* [335], Deserno and Holm [334], and, of course, Hockney and Eastwood [24].

Example 14 (Algorithms to Calculate Long-Range Interactions)
A detailed comparison of the various algorithms for determining the long-range interactions has been performed by Esselink [328]. Esselink considered an ensemble of cubic systems with a density $\rho = 1$. Each system consists of N randomly distributed particles. The charges were assigned a random value between -1 and 1 in such a way that the total charge on the system was made 0. The algorithms compared were the naive approach, which is a summation of all pairs of particles (PP), the Appel algorithm (AP), and the Greengard and Rokhlin algorithm (GR); these three methods were applied to a nonperiodic system. The Ewald summation (EW) was tested on a periodic version of the system. All algorithms have been optimized for the accuracy required.

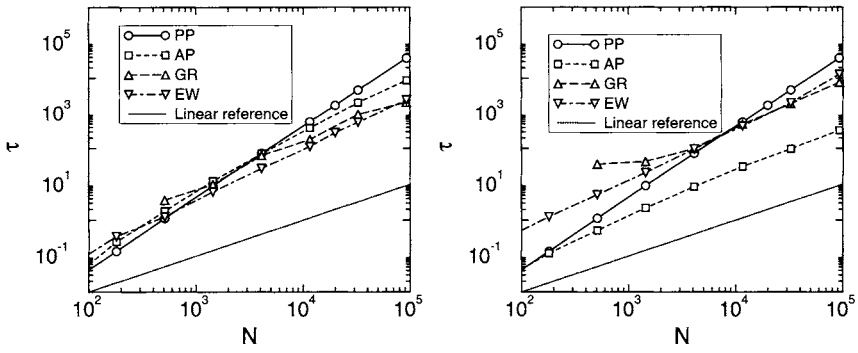

Figure 12.5: Comparison of algorithms for long-range interactions: the left figure shows CPU time τ as a function of the number of particles N for the force and the right figure shows that for the energy. PP denotes the summation over all particle pairs, AP denotes Appel's algorithm, GR the Greengard and Rokhlin algorithm, and EW the Ewald summation. The data are taken from [328].

In Figure 12.5 the efficiencies of the various algorithms for the energy and force calculations are compared.

For nonperiodic systems, for $N > 4000$ both the AP and GR algorithms outperform the naive PP algorithm. The GR algorithm for $N > 300$ is more efficient than the AP algorithm. However, if the energy alone is sufficient (as in the case of Monte Carlo simulations), the AP algorithm is more efficient than the GR algorithm and outperforms the PP algorithm for $N > 200$. Ding *et al.* [333] observed a slightly lower break-even point of the GR and PP algorithms for $N > 300$.

For periodic systems it is not possible to use the PP algorithm. For the force calculation, the EW method is more efficient than the GR algorithm for $N < 100,000$. It is important to note that Esselink used a nonperiodic version for the GR and AP algorithms; for a periodic version this break-even point will shift to a slightly larger number of particles. Ding *et al.* [340] observed a break-even point of the EW and GR methods for only 300 particles, while Schmidt and Lee [332] obtained a break-even point at several thousand particles. The results of Esselink are supported by Petersen [327]. The reason for this large difference is not clear, but may very much depend on different (more efficient) implementations of the Ewald summation.

Esselink did not include the particle-mesh-type methods in his comparison. Petersen has shown that in the range of $N = 10^4$–10^5 the PME technique is superior to the Ewald summation and fast multipole methods. Luty *et al.* [341] and Pollock and Glosli [336] obtained a conclusion result for

the PPPM method. Pollock and Glosli even concluded that for any number of particles they have investigated ($\approx 10^6$) the PPPM method, despite the $\mathcal{O}(N \log N)$ complexity, is more efficient than the fast multipole methods, which have an $\mathcal{O}(N)$ complexity. If the fast multipole method is combined with multiple-time step integration (see section 15.3), a more favorable break-even point is obtained [342]. For a more detailed comparison see ref. [337].

12.4 Ewald Summation in a Slab Geometry

In the previous sections we discussed the treatment of long-range interactions in three-dimensional systems. For some applications one is interested in a system that is finite in one dimension and infinite in the other two dimensions. Examples of such systems are fluids adsorbed in slit-like pores or monolayers of surfactants.

Special techniques are required to compute long-range interactions in such inhomogeneous systems. The most straightforward solution would be to use the same approach as for the three-dimensional Ewald summation, but restrict the reciprocal-space sum to vectors in the x, y directions [343,344]. The energy we wish to calculate is

$$U_{\text{Coul}} = \frac{1}{2} \sum_{i,j=1}^{N} {\sum_{\mathbf{n}}}' \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|},$$

where the summation over $\mathbf{n} = (L_x n_x, L_y n_y, 0)$ indicates that periodicity is only imposed in the x, y directions. As in the ordinary Ewald summation the prime indicates that for cell $(0,0,0)$ the terms $i = j$ should be omitted. We have a two-dimensional periodicity in the x, y directions for which we can use the Fourier representation. The resulting expression for the energy is [345]

$$\begin{aligned} U^{\text{Coul}} = \frac{1}{2} \sum_{i,j=1}^{N} q_i q_j \Bigg[{\sum_{\mathbf{n}}}' \frac{\operatorname{erfc}(\alpha|\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} + \frac{\pi}{L^2} \sum_{h>0} \cos(\mathbf{h} \cdot \mathbf{r}_{ij}) F(h, z_{ij}, \alpha) \\ - g(z_{ij}, \alpha)\Bigg] - \frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^{M} q_i^2, \end{aligned} \tag{12.4.1}$$

where $\mathbf{h} \equiv (2\pi m_x/L_x, 2\pi m_y/L_y, 0)$ denotes a reciprocal lattice vector, z_{ij} is the distance between two particles in the z direction, and α is the screening parameter. The function $F(h, z_{ij}, \alpha)$

$$F(h, z, \alpha) = \frac{\exp(hz)\operatorname{erfc}[\alpha z + h/(2\alpha)] + \exp(-hz)\operatorname{erfc}[-\alpha z + h/(2\alpha)]}{2h} \tag{12.4.2}$$

corrects for the inhomogeneity in the nonperiodic direction. If the system is truly two-dimensional, this term takes a simpler form. The function $g(z, \alpha)$

$$g(z, \alpha) = z \operatorname{erf}(\alpha z) + \exp\left[-(z\alpha)^2\right] / (\alpha\sqrt{\pi}) \tag{12.4.3}$$

is an additional self-term of charge interactions in the central cell that must be subtracted from the reciprocal-space sum. In a neutral system with all particles in the plane $z = 0$ this term disappears. The last term in equation (12.4.1) is the same self-term that appears in the normal Ewald summation (12.1.22). The details of the derivation can be found in refs. [343–347].

From a computational point of view equation (12.4.1) is inconvenient. Unlike the three-dimensional case, the double sum over the particles in the Fourier part of equation (12.4.1) can, in general, not be expressed in terms of the square of a single sum. This makes the calculation much more expensive than its three-dimensional counterpart. Several methods have been developed to increase the efficiency of the evaluation of the Ewald sum for slab geometries. Spohr [348] showed that the calculation can be made more efficient by the use of a look-up table combined with an interpolation scheme and the long-distance limit given by equation (12.4.6).

Hautman and Klein [349] considered the case in which the deviation of the charge distribution from a purely two-dimensional system is small. For such a system one can introduce a Taylor expansion in z, to separate the in-plane contributions x, y in $1/\sqrt{x^2 + y^2 + z^2}$ from the out-of-plane contributions. Using this approach, Hautman and Klein derived an expression in which the Fourier contribution can again be expressed in terms of sums over single particles. However, unless the ratio $z/\sqrt{x^2 + y^2} \ll 1$, the Taylor expansion converges very poorly. Therefore the applicability of this method is limited to systems in which all charges are close to a single plane. An example of such a system would be a self-assembled monolayer in which only the head groups carry a charge [349].

An obvious idea would be to use the three-dimensional Ewald summation by placing a slab of vacuum in between the periodic images (see Figure 12.7). Spohr has shown [348], however, that even with a slab that is four times the distance between the charges one does not obtain the correct limiting behavior (see Example 15). The reason is that a periodically repeated slab behaves like a stack of parallel plate capacitors. If the slab has a net dipole moment, then there will be spurious electric fields between the periodic images of the slab. More importantly, the usual assumption that the system is embedded in a conducting sphere does not correctly account for the depolarizing field that prevails in a system with a (periodic) slab geometry. Yeh and Berkowitz [350] have shown that one can add a correction term to obtain the correct limiting behavior in the limit of an infinitely thin slab. In the limit of an infinitely thin slab in the z direction, the force on a charge

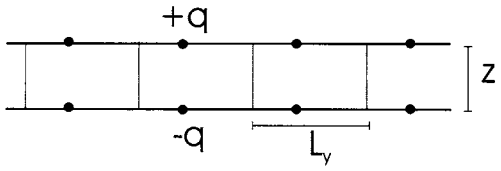

Figure 12.6: A system containing two point charges at positions $(0,0,0)$ and $(0,0,z)$; because of the periodic boundary conditions in the x and y directions, two oppositely charge "sheets" are formed. There are no periodic boundary conditions in the z direction.

q_i due to the depolarizing field is given by [351]

$$F_z = -\frac{4\pi q_i}{V} M_z, \tag{12.4.4}$$

and the total electrostatic energy due to this field is

$$U_c = -\frac{2\pi}{V} M_z^2, \tag{12.4.5}$$

where M_z is the net dipole moment of the simulation cell in the z direction

$$M_z = \sum_{i=1}^{N} q_i z_i.$$

If the slab is not infinitely thin compared to the box dimensions, higher-order correction terms have to be added. However, Yeh and Berkowitz [350] have shown that the lowest-order correction is sufficient if the spacing between the periodically repeated slabs is three to five times larger than the thickness of the slab (see also Crozier *et al.* [352]).

Example 15 (Ewald in Slab)
To illustrate the difficulties that arise when computing long-range forces in a slab geometry, Spohr and co-workers [348,352] considered a simple example of two point charges: a charge $+q$ at $(0,0,z)$ and a charge $-q$ at $(0,0,0)$. The system is finite in the z direction and periodic in the x,y directions with box sizes L_x and L_z (see Figure 12.6). Because of the periodic boundary conditions the system forms two "sheets" of opposite charge.

In the limit $z \to \infty$, the distance between the periodic images of the charge is small compared to the distance between the sheets. We can therefore assume a uniform charge density $q/(L_x L_y)$ on each sheet. In this limit the force acting between the two particles is given by

$$F_z = \frac{2\pi q^2}{L_x L_y}. \tag{12.4.6}$$

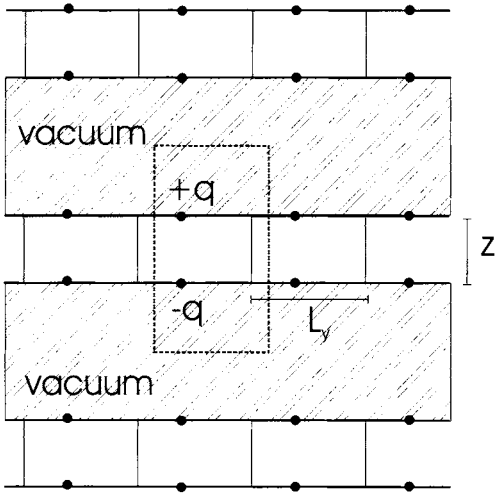

Figure 12.7: The system of Figure 12.6 artificially made periodic in the z direction by adding a slab of vacuum.

It is instructive to compare the various methods to compute the long-range interactions in this geometry. The true forces are given by the two-dimensional Ewald summation (12.4.1) and we can compare the following methods:

- Two-Dimensional Ewald summation, this solution is given by equation (12.4.1), which is the "exact" solution to this problem.
- Bare Coulomb Potential, we simple assume that the periodic images do not exist (or give a zero contribution). The resulting forces follow Coulomb law.
- Truncated and Shifted Coulomb Potential, in this method it is assumed that beyond $r_c = 9$ the potential is zero. To remove the discontinuity at $r = R_c$ the potential is shifted as well (see section 3.2.2).
- Three-Dimensional Ewald Summation, in this approximation a layer of vacuum is added. The total system (vacuum plus slab) is seen as a normal periodic three-dimensional system (see Figure 12.7) for which the three-dimensional Ewald summation (see equation (12.1.25)) is used. To study the effect of the thickness of the slab of vacuum, two systems are considered, one with $L_z = 3L_x$ and a larger one with $L_z = 3L_x$.
- 3-Dimensional Ewald Summation with Correction Term, this method is similar to the previous one; i.e., the normal three-dimensional Ewald summation is used with an additional slab of vacuum, except that now we correct for the spurious dipolar interactions, using equation (12.4.4).

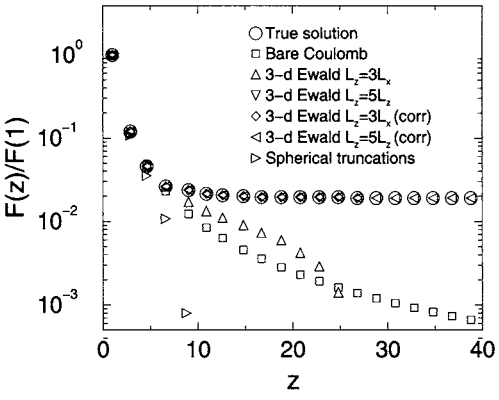

Figure 12.8: Comparison of various methods for approximating the long-range interaction for two charges of the slab geometry shown in Figure 12.6.

In Figure 12.8 we compare the various approximations with the true two-dimensional solution. The bare Coulomb potential and the shifted and truncated Coulomb potential both give a zero force in the limit $z \to \infty$ and therefore do not lead to the correct limiting behavior. Although the three-dimensional Ewald summation gives a better approximation of the correct solution, it still has the incorrect limiting behavior for both a small and a large added slab of vacuum. The corrected three-dimensional Ewald summation, however, does reproduce the correct solution, for both a slab of vacuum of $3L_x$ and that of $5L_x$.