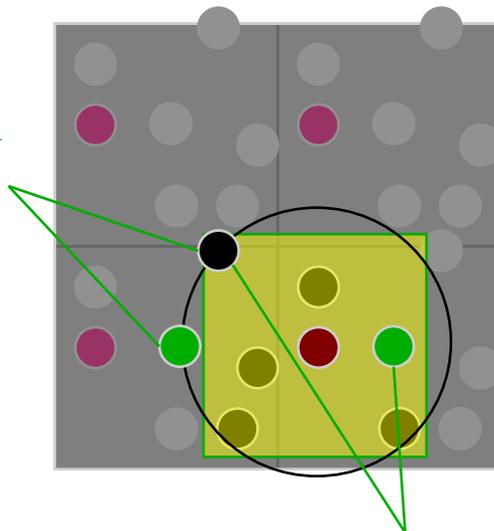


Long-Range Forces

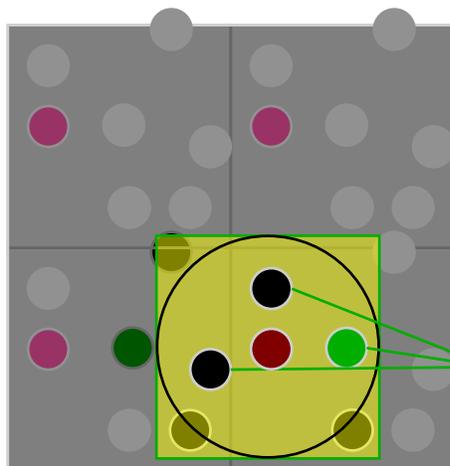
Truncating the potential

When presenting hard-sphere molecular dynamics we introduced the use of periodic boundaries as a means to model a bulk phase without walls. We examined the minimum-image convention, but only in the context of identifying collision pairs. Because the hard-sphere potential is very short-ranged, we were able to avoid the issue of how to treat interactions with atoms that do not lie in the minimum-image shell. But more realistic potentials are of sufficiently long range that interactions with second-nearest and even more distant images cannot be neglected. However, it is not feasible, and fortunately it is not necessary, to sum the interactions an atom has with all these far-away replicas of the atom nearest to it. In fact in many cases it is disadvantageous to include interactions even with some nearest-image atoms. The reason is described in Illustration 1. The yellow square depicts the nearest-image region about the purple atom at its center. Note that the green atom to the right of the central atom is a nearest-image atom, while its (green) image in the box to the left is more distant, and the interaction of the central atom with it

These two are same distance from central atom, yet:
Black atom interacts
Green atom does not



These two are nearest images for central atom



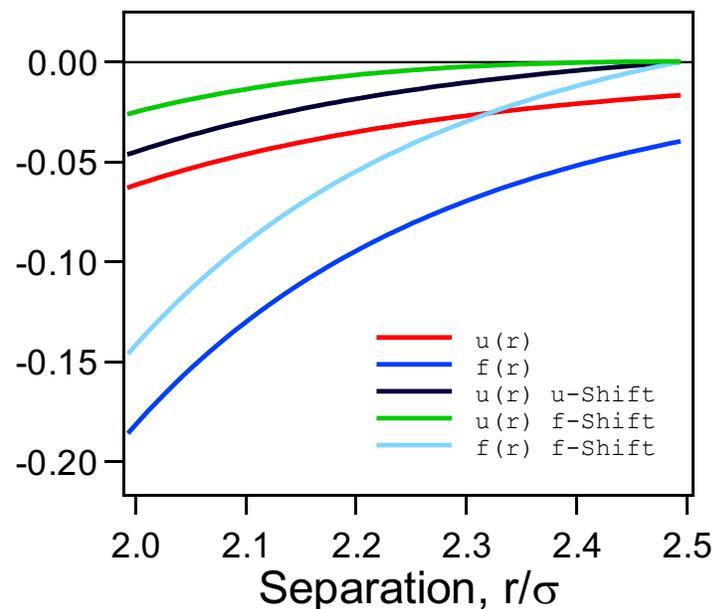
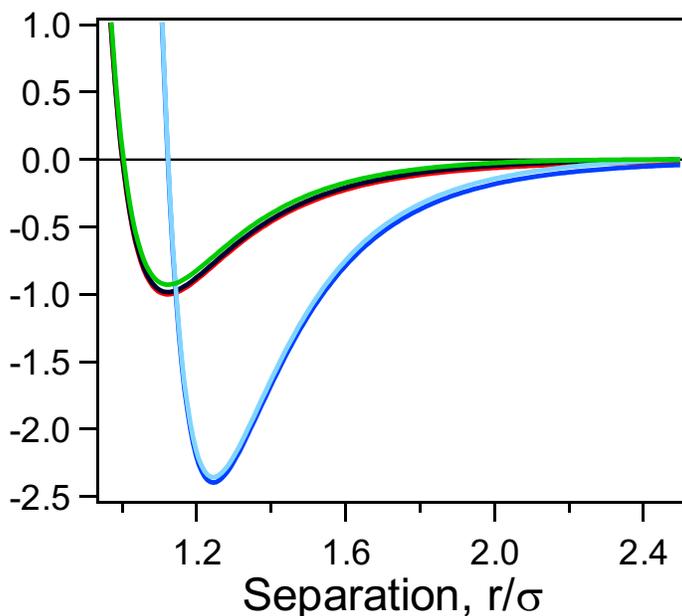
Only interactions considered

is neglected. However, both this second-image atom and the black atom (representing an entirely different atom) in the corner of the nearest-image region are equally distant from the central atom. So if all nearest-image interactions with the central atom were included, there would be an inequity, in that two equally distant atoms would be contributing differently to the total interaction. This situation complicates the application of any adjustment that corrects for all the neglected distant-image interactions. Consequently it is better to neglect all such unequal interactions, even if they are with nearest images. The problem arises only with interactions more distant than half the length of the simulation box, so the interatomic potential should be truncated at this distance, or less.

Truncating the interactions introduces a discontinuity in the potential, which corresponds to an infinite (impulsive) force acting between atoms that cross the discontinuity. Unfortunately, MD algorithms for soft potential are ill equipped to deal with this force correctly, and consequently it gets neglected. The error is manifested by poor energy conservation in the simulation. One possible remedy is to shift the whole potential up by the amount of the discontinuity, thereby bringing the energy to zero at exactly the truncation point. In equation form

$$u_s(r) = \begin{cases} u(r) - u(r_c) & r \leq r_c \\ 0 & r > r_c \end{cases}$$

where r_c is the truncation distance. This remedy helps, but it doesn't go far enough. The potential, while not infinite at the truncation point, is discontinuous. To remedy this, the force can be shifted to zero, now through the application of a term linear in the distance, thus



$$u_{sf}(r) = \begin{cases} u(r) - u(r_c) - \frac{du}{dr}(r - r_c) & r \leq r_c \\ 0 & r > r_c \end{cases}$$

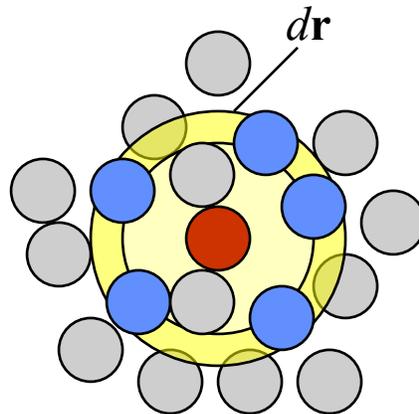
The shifted-force potential represents a larger perturbation on the overall potential. The effect of the shifted and shifted-force alterations are demonstrated on the LJ model in Illustration 2. Note that the shifted/shifted-force modifications of the potential are not normally applied in MC simulations, since energy conservation is not an issue there.

Radial distribution function

One modeling approach takes the potential- and force-shifting as the end of the story, in that the model defines interactions to be zero beyond the cutoff, so there is nothing to be done with them. In quantitative modeling, however, such a radical departure from real-atom behavior is inappropriate, and something must be done to correct for the neglected interactions. Statistical mechanics provides formulas that tell us, for example, how much two model atoms separated by 15 Angstroms will contribute to the internal energy, or to the pressure. The hard part is establishing just how many pairs of molecules will be found at a particular separation. We know that at sufficiently large separations the molecules are uncorrelated, and the number of pairs at each separation is well described by a simple application of probability. This limiting behavior is the key to formulating a correction for the neglected long-range interactions, the so-called *long-range correction*. We assume that the limiting behavior holds for all separations beyond the point where the interactions are truncated. This assumption leads to simple analytic formulas for the long-range correction to almost any thermodynamic quantity of interest.

Before turning to those formulas it is worthwhile to examine the radial distribution function (rdf). The rdf is a key quantity in statistical mechanics because it characterizes how the atom correlations decay with increasing separation. The rdf is defined as follows

$$g(r) = \frac{\rho(r)dr}{\rho^{id} dr}$$



```

//Method in public class MeterRDF extends MeterFunction
//Computes RDF for the current configuration
public double[] currentValue() {
    iterator.reset(); //prepare iterator of atom pairs
    for(int i=0; i<nPoints; i++) {y[i] = 0.0;} //zero histogram
    while(iterator.hasNext()) { //loop over all pairs in phase
        double r = Math.sqrt(iterator.next().r2()); //get pair separation
        if(r < xMax) {
            int index = (int)(r/delr); //determine histogram index
            y[index]+=2; //add once for each atom
        }
    }
    int n = phase.atomCount(); //compute normalization: divide by
    double norm = n*n/phase.volume(); //n, and density*(volume of shell)
    for(int i=0; i<nPoints; i++) {y[i] /= (norm*vShell[i]);}
    return y;
}

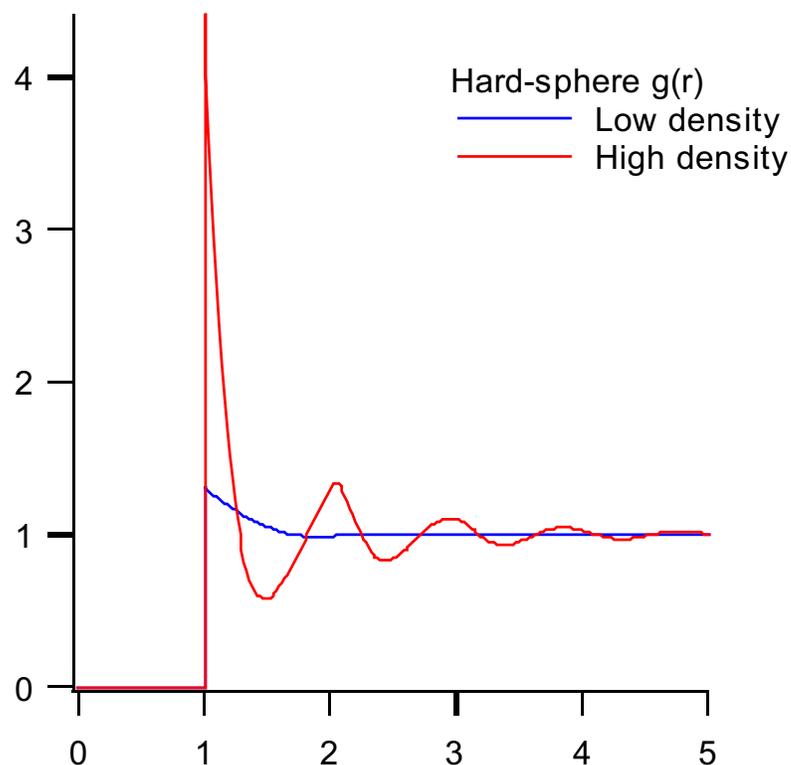
```

The numerator is the number of atoms found in a volume element $d\mathbf{r}$ a distance r from a given atom (see Illustration 3), while the denominator is the same quantity for an ideal gas, a system with no atom correlations at all. The ideal gas term is independent of r and is simply the number density

$$\rho^{id} d\mathbf{r} = \frac{N}{V} d\mathbf{r}$$

Java code for computing the rdf in a pure substance (non a mixture) is presented in Illustration 4. This facility is included in the API as a subclass of MeterFunction. Illustration 5 contains an applet that presents the rdf computed during a simulation.

The radial distribution function for the hard-sphere model at two densities is presented in Illustration 6. The rdf for more realistic models exhibits the same qualitative behavior as that in the illustration, except the behavior near the contact value is not so sharp, reflecting the softer repulsion exhibited by real atoms. Note the prominent structure present at high density, and the lack of almost all correlation at low density. For both



densities, at sufficiently large separations the distribution approaches unity, reflecting the loss of correlation between distant atoms. It is atom pairs in this range of separation and beyond that can be neglected in a molecular simulation. Their influence can be represented instead by formulas such as the following, for the potential energy,

$$U_{lrc} = \frac{N}{2} \rho \int_{r_{cut}}^{\infty} u(r) 4\pi r^2 dr$$

pressure,

$$P_{lrc} = \frac{1}{6} \rho^2 \int_{r_{cut}}^{\infty} r \frac{du}{dr} 4\pi r^2 dr$$

and chemical potential

$$\mu_{lrc} = \rho \int_{r_{cut}}^{\infty} u(r) 4\pi r^2 dr = 2 \frac{U_{lrc}}{N}$$

For example, for the LJ model the energy and pressure corrections are

$$U_{lrc}^{LJ} = \frac{8}{9} \pi N \rho \sigma^3 \varepsilon \left[\left(\frac{\sigma}{r_c} \right)^9 - 3 \left(\frac{\sigma}{r_c} \right)^3 \right]$$

$$P_{lrc}^{LJ} = \frac{32}{9} \pi \rho^2 \sigma^3 \varepsilon \left[\left(\frac{\sigma}{r_c} \right)^9 - \frac{3}{2} \left(\frac{\sigma}{r_c} \right)^3 \right]$$

For a cutoff of about 2.5σ , these corrections make up about 5-10% of the total energy and pressure, depending on density.

Sometimes these corrections can be added to the simulation averages upon completion of the simulation, but at other times it is important to include the corrections during the course of the simulation. For example, the long-range correction to the energy depends on the density, so its contribution to the energy change must be included when deciding acceptance of volume moves in NPT MC simulation. We think it is a good habit to include them at all points in simulation, since doing so has negligible computational cost.

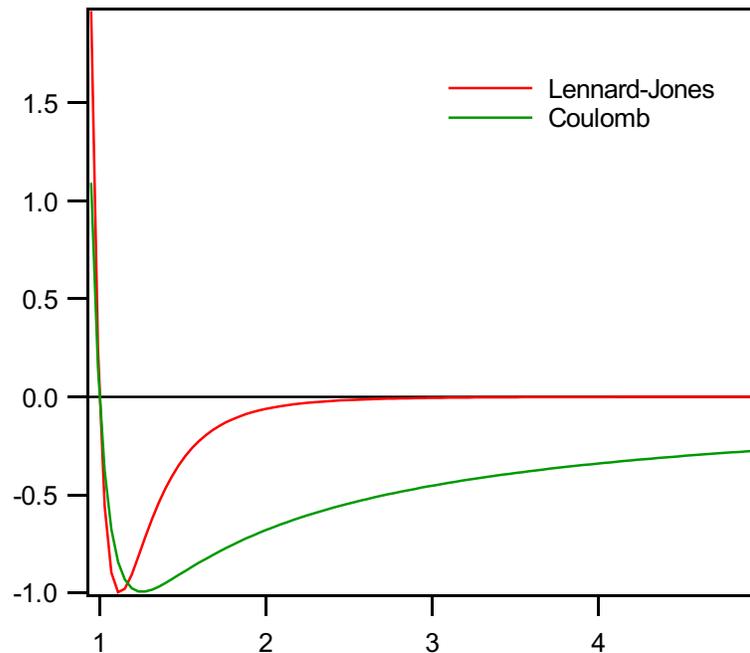
Coulombic interactions

Long-range electrostatic interactions must be treated in a more sophisticated way than that used for van der Waals attraction. The Coulomb potential vanishes as $1/r$, which is a much slower decay than the $1/r^6$ dispersion interaction characterized by the LJ model potential. Representative curves are shown in Illustration 7. Whereas the LJ model can

quite reasonably be truncated at about 2.5 LJ diameters, the Coulomb potential is at 5 LJ diameters nowhere near approaching zero. Moreover, a naïve application of the long-range correction to the energy yields

$$U_{lrc} = \frac{N}{2} \int_{r_c}^{\infty} \frac{1}{r} 4\pi r^2 dr = \infty$$

The influence of the potential never vanishes, and it is only upon adding the positive and negative infinities, arising from the interactions of like and unlike charges, does a finite



energy result.

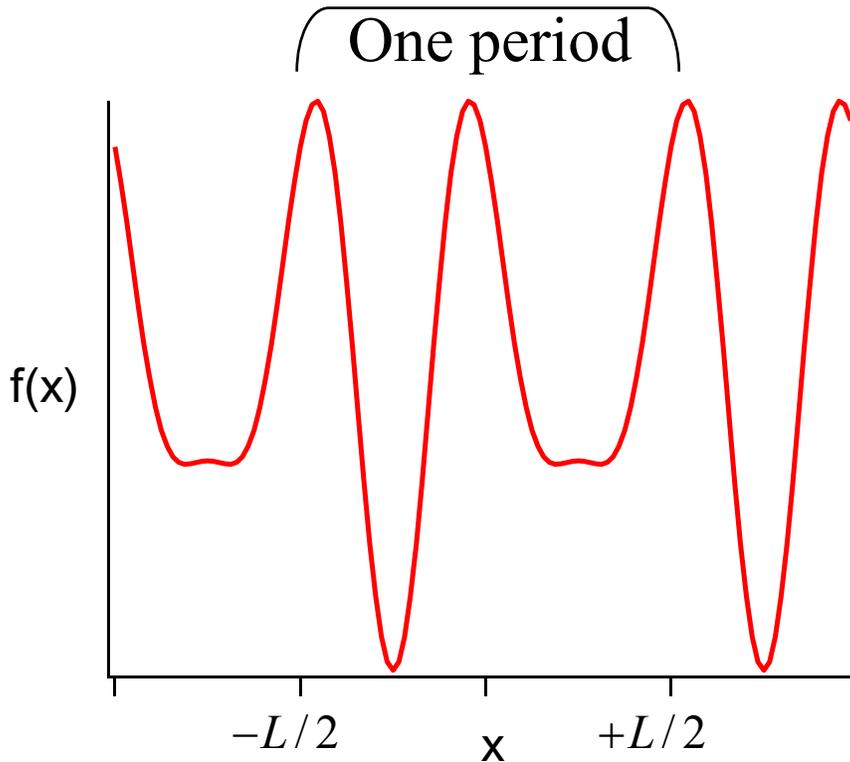
So it should be clear that the treatment of long-range electrostatic interactions must be performed with great care. We will consider two common approaches. One is to perform a full lattice sum, using clever techniques to enhance the convergence; another is to model the surroundings as a dielectric continuum that responds to fluctuations in the simulated system. The former method is known as the Ewald sum, and we will examine it first. Before doing so it is worthwhile to review some of the important elementary methods and concepts related to the problem and its solution.

Fourier series

The periodicity inherent in the lattice sum makes it amenable to application of Fourier techniques. Accordingly, the Ewald method makes use of a Fourier series. Let's review discrete Fourier analysis in simpler context before seeing how it is applied in the Ewald sum.

Consider a periodic function $f(x)$ of period L , such as that exhibited in Illustration 8. A Fourier series provides an equivalent representation of this function through a set of coefficients a_n and b_n

$$f(x) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} a_k \cos kx + b_k \sin kx$$



where the coefficients are

$$a_k = \frac{2}{L} \int_{-L/2}^{+L/2} f(x) \cos(2\pi kx / L) dx$$

$$b_k = \frac{2}{L} \int_{-L/2}^{+L/2} f(x) \sin(2\pi kx / L) dx$$

These relations can be written more compactly using complex numbers, replacing the sine and cosine terms by a complex exponential, using $e^{i\theta} = \cos \theta + i \sin \theta$, where $i = \sqrt{-1}$. Then

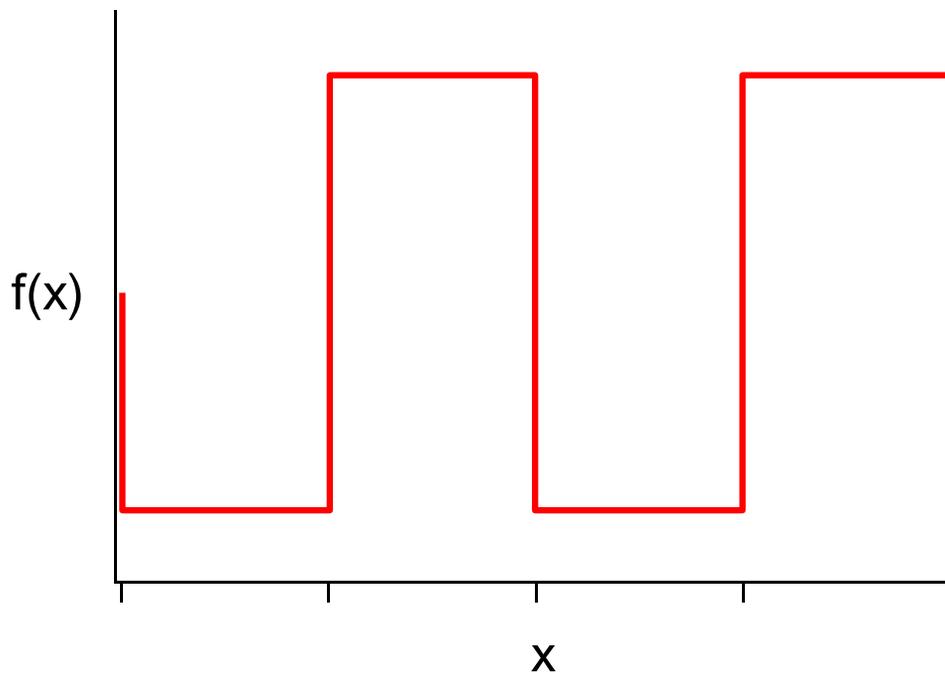
$$f(x) = \sum_{k=-\infty}^{\infty} \hat{f}_k e^{ikx} \quad (1.1)$$

where the coefficients are represented by the real and imaginary parts of \hat{f}_k

$$\hat{f}_k = \begin{cases} \frac{1}{2}(a_k + ib_k) & k \geq 0 \\ \frac{1}{2}(a_{-k} - ib_{-k}) & k \leq 0 \end{cases}$$

which leads to

$$\hat{f}_k = \frac{1}{L} \int_{-L/2}^{+L/2} f(x) e^{-2\pi i k x / L} dx \quad (1.2)$$



As an example, let us consider the square wave shown in Illustration 9. Over the period centered on the origin, the function is

$$f(x) = \begin{cases} +1 & -L/2 < x \leq 0 \\ -1 & 0 < x \leq +L/2 \end{cases}$$

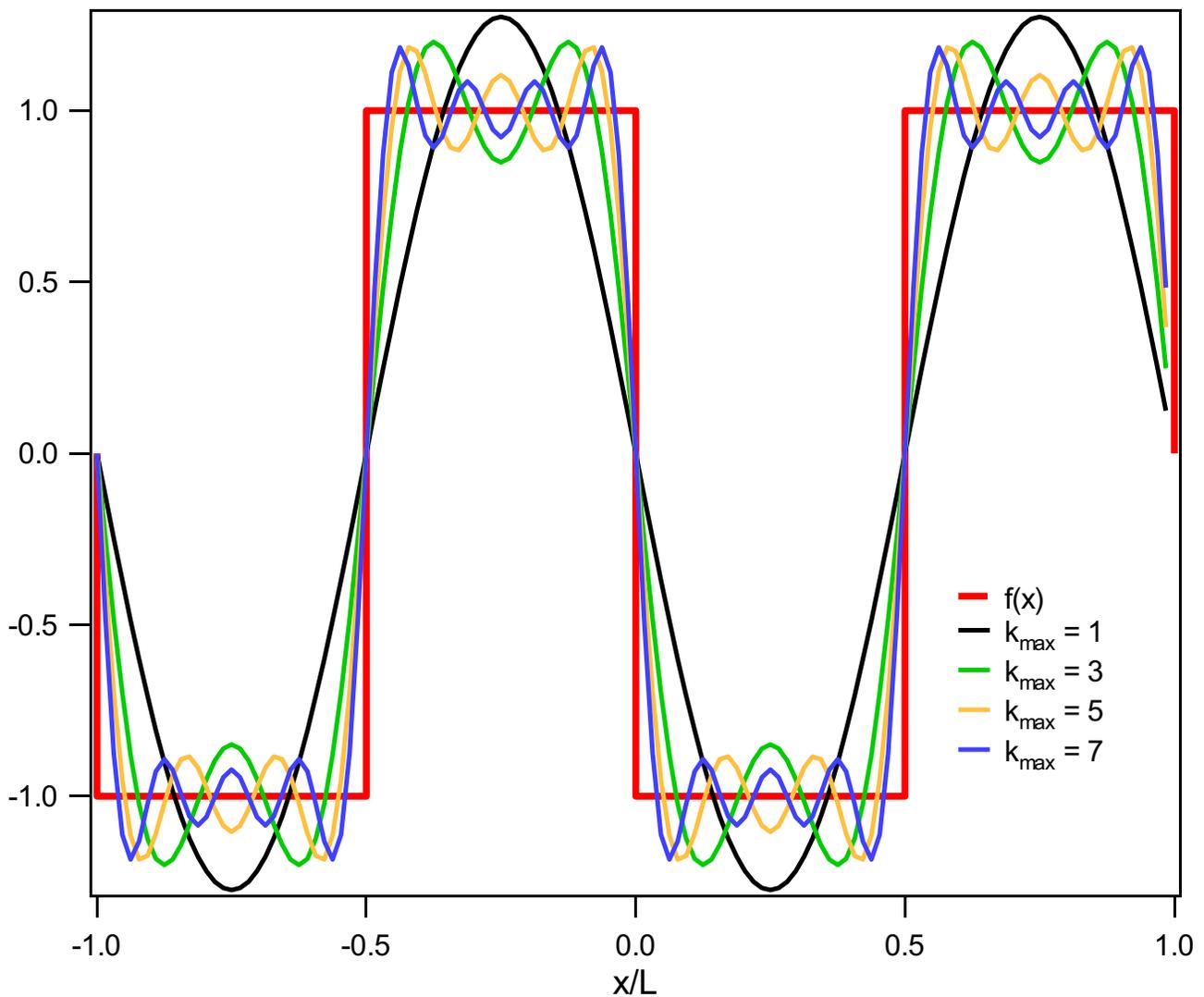
So the transform is

$$\begin{aligned}
\hat{f}_k &= \frac{1}{L} \int_{-L/2}^0 e^{-2\pi i k x / L} dx - \frac{1}{L} \int_0^{+L/2} e^{-2\pi i k x / L} dx \\
&= \frac{1}{L} \int_0^{+L/2} \left(e^{+2\pi i k x / L} - e^{-2\pi i k x / L} \right) dx \\
&= \frac{1}{L} \int_0^{+L/2} 2i \sin(2\pi k x / L) dx \\
&= \frac{i}{k\pi} (1 - \cos n\pi) \\
\hat{f}_k &= \begin{cases} 0 & k \text{ even} \\ \frac{2i}{k\pi} & k \text{ odd} \end{cases}
\end{aligned}$$

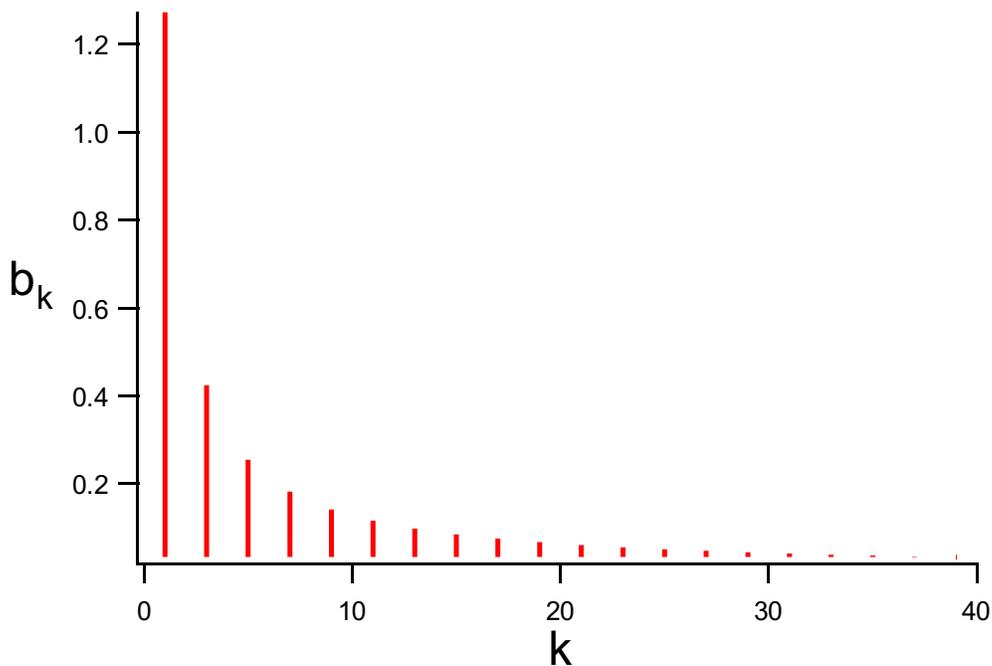
The real part of \hat{f}_k is always zero, indicating that the a_k coefficients are zero; this happens because the function $f(x)$ is odd (*i.e.*, $f(-x) = -f(x)$) and consequently it cannot have any cosine components in its Fourier decomposition. Thus the Fourier series representation of the square wave is simply

$$\begin{aligned}
f(x) &= \sum_{\substack{k>0 \\ k \text{ odd}}} \frac{2}{k\pi} \sin\left(\frac{2\pi k x}{L}\right) + \frac{2}{(-k)\pi} \sin\left(\frac{2\pi(-k)x}{L}\right) \\
&= \sum_{\substack{k>0 \\ k \text{ odd}}} \frac{4}{k\pi} \sin\left(\frac{2\pi k x}{L}\right) \\
&= \frac{4}{\pi} \sin\left(\frac{2\pi x}{L}\right) + \frac{4}{3\pi} \sin\left(\frac{6\pi x}{L}\right) + \frac{4}{5\pi} \sin\left(\frac{10\pi x}{L}\right) + \dots
\end{aligned}$$

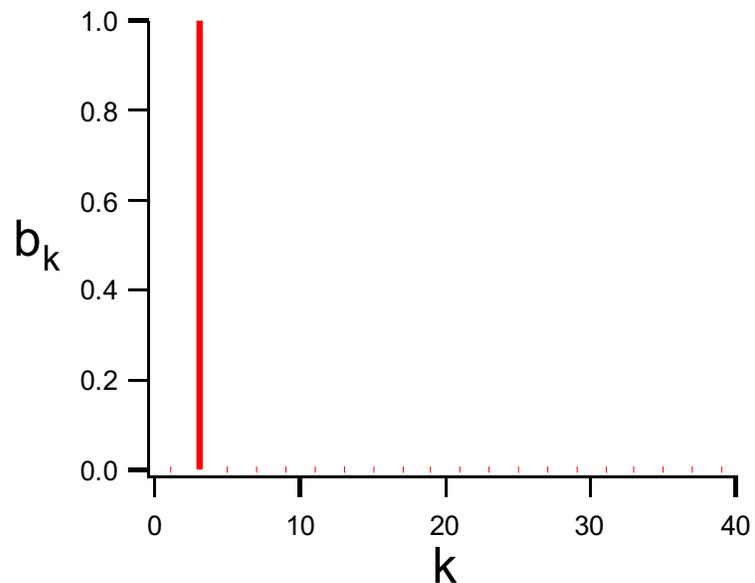
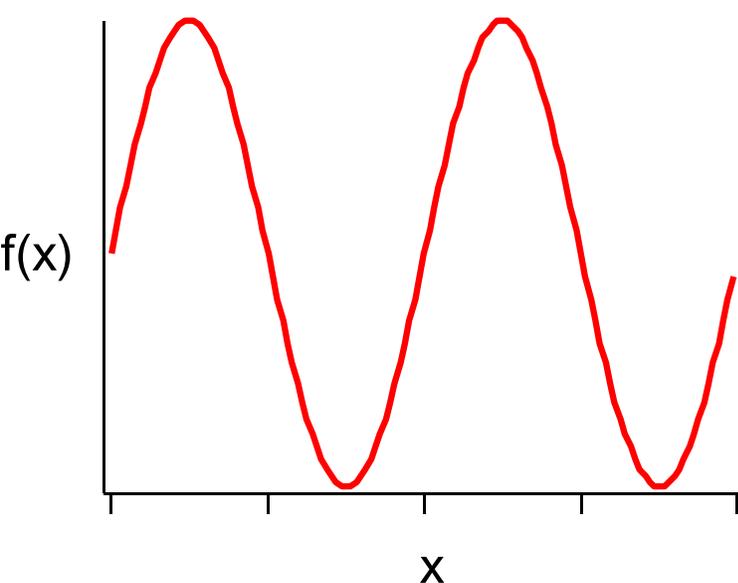
This series, truncated after the first few terms, is presented in Illustration 10.



It is of particular interest to examine the behavior of the transform function \hat{f}_k . Illustration 11 presents the b_k coefficients for the square-wave example. Unlike the original function of x , which persists without decay for arbitrarily large x , this function of k approaches zero over a finite range of k . It does so while retaining complete information about the original function. By working in this alternate representation, we can work with the infinite ranged function of x by doing operations on the finite-ranged function of k . This idea is used by the Ewald method to account for the long-ranging influence of all the periodic images of the simulated atoms.



Let us think now about the rate of convergence of the coefficient function. Note that a simple, smooth sine-wave function $f(x)$ is, quite obviously, represented in Fourier space by a single coefficient. See Illustration 12. This is a particularly simple example of a general feature of the Fourier transformation. A smooth periodic function of infinite range transforms into a very sharp, short ranged function in the Fourier representation. On the other hand, a very sharp, rapidly changing function $f(x)$ transforms into a very long-ranged, slowly decaying Fourier-space function. Some of this behavior was in evidence in the square-wave example, where the sharp features of the square wave require a significant high-frequency (large k) Fourier component (dying off as $1/k$). This general interpretation of the Fourier transform is important to keep in mind. The small- k parts of the transform describe the long-wavelength, low-frequency components of the original function (culminating at $k = 0$, which coefficient gives the simple average of the function $f(x)$). The large- k parts collect all the high-frequency components, which will be small for a smooth function of x .



The Fourier transform is obtained as the limiting behavior of the Fourier series as the period L becomes infinite. In the limit the series representation becomes an integral

$$f(x) = \int_{-\infty}^{+\infty} \hat{f}(k) e^{-2\pi i k x} dk$$

The expression for \hat{f}_k is the same, but for modification to the limits of integration

$$\hat{f}(k) = \int_{-\infty}^{+\infty} f(x) e^{2\pi i k x} dx$$

and we write $\hat{f}(k)$ instead of \hat{f}_k to indicate that the transform variable is continuous in the limit of infinite period.

The Fourier transform exhibits many useful properties. One is in regard to the Fourier transform of the derivative of a function, which is simply related to the transform of the function itself

$$\left[f^{(m)}(x) \right](k) = (-2\pi i k)^m \hat{f}(k)$$

where $f^{(m)}(x)$ is the m^{th} derivative of f with respect to x . We will make use of this relation shortly.

Before returning to the Ewald sum, which motivates this discussion, let us consider one more example. Consider a function $f(x)$ of Gaussian form

$$f(x) = \left(\frac{\alpha}{2\pi} \right)^{1/2} \exp \left[-\alpha \frac{x^2}{2} \right]$$

It turns out that the Fourier transform of this function is also a Gaussian

$$\hat{f}(k) = \exp \left[-\frac{2\pi^2 k^2}{\alpha} \right]$$

Note that the width (variance) of the transform function is the reciprocal of the width of the original function. This is consistent with our earlier observation that a sharply peaked function transforms to a broad, slowly decaying function, and vice versa. As a limiting case, if the original function $f(x)$ is actually a Dirac delta function

$$f(x) = \delta(x - x_0)$$

then it is easy to see that the transform is

$$\hat{f}(k) = e^{2\pi i k x_0}$$

that is, an infinite-ranged sine/cosine function of wavelength x_0 .

Basic electrostatics

A point charge q_1 in an electric field $\mathbf{E}(\mathbf{r})$ experiences a force given by

$$\mathbf{F}(\mathbf{r}) = q_1 \mathbf{E}(\mathbf{r})$$

The electric field is in turn created by the presence of other charges, distributed according to the charge density $\rho(\mathbf{r})$. The fundamental equations of electrostatics dictate how the field is determined from the charge density

$$\begin{aligned} \nabla \cdot \mathbf{E}(\mathbf{r}) &= 4\pi\rho(\mathbf{r}) \\ \nabla \times \mathbf{E}(\mathbf{r}) &= 0 \end{aligned} \tag{1.3}$$

The latter equation can be satisfied automatically if the field is expressed as the gradient of a scalar potential

$$\mathbf{E}(\mathbf{r}) = -\nabla\phi(\mathbf{r})$$

This electrostatic potential ϕ represents a potential energy, in the sense that it is the energy required (in the form of work) to bring a unit charge from infinity to \mathbf{r} . With \mathbf{E} expressed in this manner, the relation between the electrostatic potential and the charge density follows from Eq. (1.3)

$$\nabla^2\phi = -4\pi\rho(\mathbf{r}) \tag{1.4}$$

This is Poisson's equation. If $\rho(\mathbf{r})$ is a simple point charge q_2 , solution of Poisson's equation finds that the potential of a test charge q_1 varies inversely with the distance r from q_2

$$\phi(r) = \frac{q_1 q_2}{r}$$

which is Coulomb's law. We note that this result holds only in a three-dimensional space. In two dimensions, the potential varies as $\ln(r)$, while in 1D it increases linearly with the separation! A very long-ranged interaction indeed.

Ewald sum

The aim of the Ewald method is to perform a sum of the interaction energy of each charge in the simulation cell with all other charges in the cell, and with all periodic images of all charges in the cell. The method proceeds by developing an expression for the electric field $\mathbf{E}(\mathbf{r})$ due to all charges and their images. This field is evaluated at each

point where a charge is located in the simulation cell, thereby obtaining the contribution of that charge to the total potential energy. Because the charge images—and thus the charge density—is periodic, Fourier methods are appropriate for evaluating the electric field due to them. However, the charge density is a very sharply varying function, consisting of delta-function spikes at each charge image, and this prevents the Fourier transform from decaying at large frequencies. To remedy this, the point charges are approximated as Gaussian charge densities (other forms are possible), thereby making for convergent Fourier sums. This step introduces an error that is best removed by performing a rapidly converging sum in real space. Let us now consider the Ewald method in more detail.

The charge density due to all the point charges in the simulation cell, and all their images, is

$$\begin{aligned}\rho(\mathbf{r}) &= \sum_{\substack{\mathbf{l}, \text{image} \\ \text{vectors}}} \sum_{j \text{ in } \mathbf{l}} q_j \delta(\mathbf{r} - \mathbf{r}_j) \\ &= \sum_{\mathbf{n}} \sum_j q_j \delta(\|\mathbf{r} - (\mathbf{r}_j + \mathbf{l})\|)\end{aligned}$$

For cubic periodic boundary conditions, the lattice vectors are

$$\mathbf{l} = \{l_x, l_y, l_z\} \quad l_x, l_y, l_z = 0, \pm L, \pm 2L, \pm 3L, \dots$$

that is, each element of the lattice vector may take on a value equal to any integer multiple of L , where L is the linear dimension of the simulation cell. Thus the first few lattice vectors are $\{0,0,0\}$, $\{0,0,L\}$, $\{0,L,0\}$, $\{L,0,0\}$, $\{0,0,-L\}$, $\{0,-L,0\}$, $\{-L,0,0\}$, $\{0,L,L\}$, ..., $\{0,0,2L\}$, ..., *etc.* The convergence of the sum depends upon the order in which the lattice images are summed. It is necessary to order the terms in a concentric fashion, so that terms with larger $|\mathbf{l}| \equiv l_x^2 + l_y^2 + l_z^2$ are added only after all terms with smaller values of $|\mathbf{l}|$ have been included.

The charge density is a periodic function and, just like the square-wave example examined above, its full behavior can be captured by applying a Fourier transform over just one period, which means in this case over just one simulation cell. At present we are dealing with a three-dimensional function, rather than the simple one-dimensional examples used previously, but the results demonstrated there still apply. The main difference is that our transform variables \mathbf{r} and \mathbf{k} are now vectors instead of the scalars x and k used before. We face a problem now, in that our periodic charge-density is composed of (very sharp) delta functions, and consequently its Fourier transform will be slowly converging. In fact it will be a smooth, infinitely periodic function, and we will have gained no advantage in the trade from real to Fourier space. To permit the useful application of Fourier techniques, we can approximate the sharp charge densities by less-sharp Gaussian functions. We smear the charges. With this modification, the charge density is

$$\rho(\mathbf{r}) = \sum_{\mathbf{l}} \sum_j q_j (\alpha / \pi)^{3/2} \exp \left[-\alpha |\mathbf{r} - (\mathbf{r}_j + \mathbf{l})|^2 \right]$$

Here, α is a parameter that adjusts the degree of smearing of the charges. A small value of α causes the charges to be smeared more, while a large α keeps the charges sharply defined, approaching the original δ -function representation at infinite α . The three-dimensional Fourier representation of this charge density is also a Gaussian

$$\begin{aligned} \hat{\rho}(\mathbf{k}) &= \frac{1}{V} \int d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} \rho(\mathbf{r}) \\ &= \frac{1}{V} \sum_j q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} e^{-k^2/4\alpha} \end{aligned} \quad (1.5)$$

For cubic periodic boundaries, the \mathbf{k} vectors are the same as the \mathbf{l} vectors, except instead of being formed as multiples of L , they are multiples of $2\pi/L$: $\mathbf{k} = \{0,0,0\}$, $\{0,0,+2\pi/L\}$, etc.

The electrostatic potential due to this charge density is obtained via Poisson's equation, Eq. (1.4). First we Fourier transform both sides of the equation. Using the derivative relation reviewed above, the Fourier transform of the Laplacian (a 2nd derivative in 3D) can be written as the square of the transform variable times the Fourier transform of the function; thus

$$|\mathbf{k}|^2 \hat{\phi}(\mathbf{k}) = -4\pi \hat{\rho}(\mathbf{k}) \quad (1.6)$$

This with Eq. (1.5) gives the Fourier-space representation of the electrostatic potential due to the smeared charges. The electrostatic energy of all central-image point charges in this field is given as follows. First we write the energy using the real-space potential

$$U_q = \frac{1}{2} \sum_i q_i \phi(\mathbf{r}_i) \quad (1.7)$$

which we then write as the inverse of its Fourier form (*cf.* Eq. (1.1))

$$U_q = \frac{1}{2} \sum_i q_i \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}_i} \hat{\phi}(\mathbf{k})$$

Now with $\hat{\phi}(\mathbf{k})$ from Eq. (1.6)

$$U_q = \frac{1}{2} \sum_i q_i \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}_i} \frac{-4\pi \hat{\rho}(\mathbf{k})}{|\mathbf{k}|^2}$$

and $\hat{\rho}(\mathbf{k})$ from Eq. (1.5), followed by some rearrangement we have

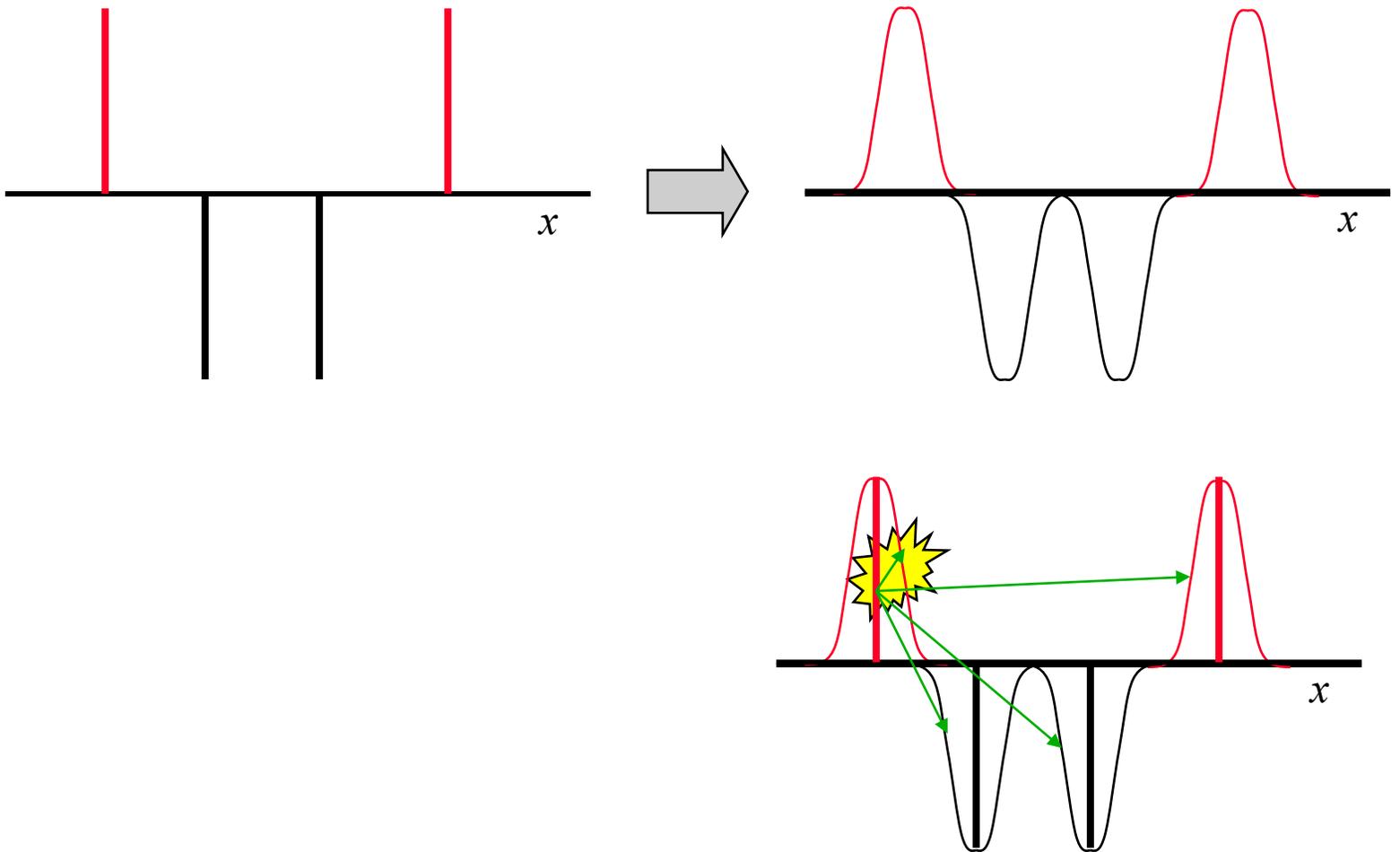
$$\begin{aligned}
 U_q &= \frac{1}{2} \sum_{\mathbf{k}} \frac{4\pi V}{k^2} e^{-k^2/4\alpha} \sum_{i,j} \frac{q_i q_j}{V^2} e^{i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} \\
 &= \frac{1}{2} \sum_{\mathbf{k}} \frac{4\pi V}{k^2} e^{-k^2/4\alpha} |\hat{\rho}_\delta(\mathbf{k})|^2
 \end{aligned}
 \tag{1.8}$$

where we have identified

$$\hat{\rho}_\delta(\mathbf{k}) = \frac{1}{V} \sum_j q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j}$$

as the Fourier transform of the original (unsmeared) charge distribution. We see that the smearing has the effect of attenuating the \mathbf{k} -space sum in Eq. (1.8), and that smaller values of α (more smearing) causes the sum to converge more quickly.

Equation (1.8) gives the electrostatic energy of the central-cell point charges in the electrostatic field generated by smeared charges, both those in the central cell and all image cells. Two corrections are needed. Note that each charge is interacting with smeared versions of all charges in the central cell, including the smeared version of itself. See Illustration 13. We need to subtract this self interaction. We do this by solving for the electric potential due to the smeared charge, and computing the electrostatic energy of a charge in the center of this potential. The potential is given the the solution of Poisson's equation (1.4) for the Gaussian charge density



$$\rho(\mathbf{r}) = q_j (\alpha / \pi)^{3/2} e^{-\alpha |\mathbf{r}|^2}$$

The solution is

$$\phi^G(\mathbf{r}; \mathbf{r}_j) = \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|} \operatorname{erf}(\sqrt{\alpha} |\mathbf{r} - \mathbf{r}_j|)$$

which can be verified by inserting it back into Poisson's equation, and noting that $\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right)$ (in three dimensions; all of this development must be modified if applied in other than a 3D space). The charge at the center, $\mathbf{r} = \mathbf{r}_j$, experiences a field $\phi(0) = 2q_j (\alpha / \pi)^{1/2}$, so its (inappropriate) contribution to the electrostatic energy is $q_j^2 (\alpha / \pi)^{1/2}$. This value is independent of position, and depends only on the magnitude of the charge. Thus to apply a correction to the total electrostatic energy in Eq. (1.8), we subtract this self-interaction term for all charges

$$\begin{aligned} U_{self} &= \frac{1}{2} \sum_j q_j \phi(0) \\ &= \left(\frac{\alpha}{\pi} \right)^{1/2} \sum_j q_j^2 \end{aligned}$$

We must now correct for the use of the smeared charges in determining the electrostatic potential. We can do this by adding the correct potential and subtracting the approximate one. We can do this effectively by staying in real space, because the difference between the potentials decreases rapidly with increasing distance from the charge centers, even though neither potential by itself does. The correction is

$$\begin{aligned} \Delta\phi_j(\mathbf{r}; \mathbf{r}_j) &= \phi_j^\delta(\mathbf{r}; \mathbf{r}_j) - \phi_j^G(\mathbf{r}; \mathbf{r}_j) \\ &= \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|} - \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|} \operatorname{erf}(\sqrt{\alpha} |\mathbf{r} - \mathbf{r}_j|) \\ &= \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|} \operatorname{erfc}(\sqrt{\alpha} |\mathbf{r} - \mathbf{r}_j|) \end{aligned}$$

The full correction is obtained by summing the interactions of each charge in the central cell with the field correction terms due to all charges in the central cell and all image cells.

$$\begin{aligned}\Delta U &= \frac{1}{2} \sum_{\mathbf{l}} \sum_{i \neq j} q_i \Delta \phi_j \left(|\mathbf{r}_{ij} + \mathbf{l}| \right) \\ &= \frac{1}{2} \sum_{\mathbf{l}} \sum_{i \neq j} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{l}|} \operatorname{erfc} \left(\sqrt{\alpha} |\mathbf{r}_{ij} + \mathbf{l}| \right)\end{aligned}$$

where we require $i \neq j$ in the central image only. The rate at which the potential difference vanishes with distance from the each charge center is affected by the parameter α . However, the effect is opposite that on the Fourier sum—the potential difference vanishes most rapidly for large α . Thus the combination of the Fourier sum and the smearing correction requires a compromise in the choice of α . A good rule-of-thumb is to choose $\alpha = 25/L^2$, but testing is recommended in each circumstance to determine an appropriate value.

The total electrostatic energy due to the interaction of the point charges is

$$U_{tot} = U_q(\alpha) + U_{self}(\alpha) + \Delta U(\alpha)$$

As indicated, each term depends on the parameter α , but the sum does not. Collecting the results, we have

$$\begin{aligned}U_{tot} &= \frac{1}{2} \sum_{\mathbf{k} \neq 0} \frac{4\pi V}{k^2} e^{-k^2/4\alpha} |\hat{\rho}_\delta(\mathbf{k})|^2 \\ &\quad + \left(\frac{\alpha}{\pi}\right)^{1/2} \sum_j q_j^2 \\ &\quad + \frac{1}{2} \sum_{\mathbf{l}} \sum_{i \neq j} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{l}|} \operatorname{erfc} \left(\sqrt{\alpha} |\mathbf{r}_{ij} + \mathbf{l}| \right)\end{aligned} \tag{1.9}$$

The sum of these terms is independent of α only if sufficient terms are included in the lattice sums. In practice α is chosen to permit the real-space sum to converge within the central simulation cell, often attempting to limit its range even further, so that it may be truncated at the same point that the van der Waals (*e.g.*, Lennard-Jones) contributions are neglected. Attenuating the real-space part this way increases the number of terms required in the \mathbf{k} -space sum; something of the order of 100-200 vectors are usually needed there to get acceptable convergence.

We have glossed over a subtle but important point in the development. The $\mathbf{k} = 0$ contribution to Eq. (1.8) describes the interaction of each central charge with the infinite-wavelength features of the electrostatic potential. Even though the system is modeled as being periodic forever in every direction, the infinite-wavelength features persist (by definition) over this infinite range. In principle there exists a boundary at some point, and the nature of the surrounding system can influence the infinite-wavelength features of the potential. If the surrounding medium is a perfect conductor, it will respond in a way that exactly offsets any infinite-wavelength features of the electrostatic potential of the

simulated system. This corresponds to a neglect of the $\mathbf{k} = 0$ term in the Ewald sum, as we have done in Eq. (1.9). We give a more thorough explanation for this effect later, explaining how to implement different choices of the boundary.

It is notable that much of the complication of the Ewald method comes from the use of point charges to model the electrostatic interactions. Other choices are possible. It would not be inappropriate to use a model in which the charges are at the outset smeared Gaussians. In this case there would be no need to apply the real-space correction. However, the development would be complicated by the need to evaluate the interactions of the central-cell smeared charges with the electrostatic potential due to the central- and image-cell charges. This requires that Eq. (1.7) be rewritten as a sum of integrals.