> **Today's lecture:** considerations for setting up and running simulations of bulk, iso-
> tropic phases (e.g., liquids and gases)

Periodic boundaries

Cubic boxes

In simulations of bulk phases, the presence of explicit interfaces or "walls" in a simulation box
would have a profound effect on the resulting properties of the system, since these interactions
are very significant compared to the small size of the system. Instead, for bulk phases we typically
implement **periodic boundary conditions** that realize an infinite number of copies of the simula-
tion box.
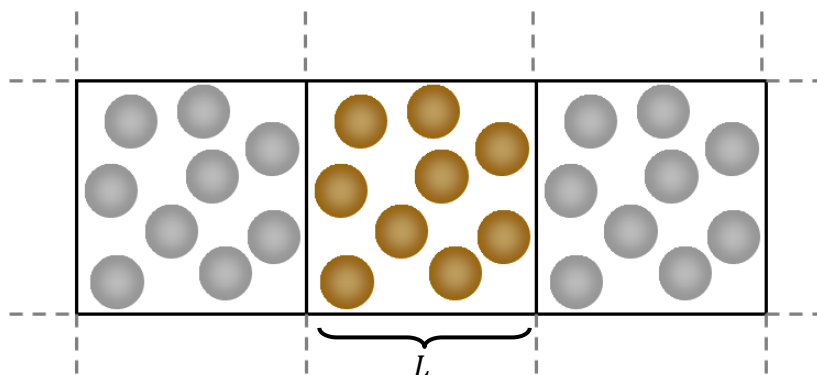
Particles have copies of themselves inside every periodic repetition of the simulation box, called
image particles. For any given particle, we can find its position in the central box using the near-
est integer operation:

$$\mathbf{r}^0 = \mathbf{r} - L\,\mathrm{nint}(\mathbf{r}/L)$$

where L is the length of the imaginary cubic volume in which the particles sit. It may also be a
vector for rectangular cuboids (i.e., boxes with different side lengths) provided that L is replaced
by the vector $\mathbf{L}$. This equation implies a separate operation for each of the x, y, z coordinates.
After its application, each coordinate will be in the range $-L/2$ to $L/2$.

Oftentimes we are interested in the vector and scalar distances between two particles. We ex-
amine the **minimum image distance**, the smallest distance between any two images of the par-
ticles. This distance can be found using an expression similar to that above:

 last modified 11/30/2022

$$\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$$

$$\mathbf{r}_{ij}^0 = \mathbf{r}_{ij} - L \, \text{nint}(\mathbf{r}_{ij}/L)$$

$$r_{ij}^0 = \left|\mathbf{r}_{ij}^0\right|$$

Note that we must first find the vector minimum image distance in order to compute its scalar norm. That is, we must first minimum image each of the x, y, z coordinates separately.

Other geometries

While a cubic box is the most frequently used geometry for periodic boundary conditions, other simulation cell geometries are possible:

- rectangular cuboid

- parallelpiped

- hexagonal prism

- truncated octahedron

- rhombic dodecahedron

All of these geometries will regularly tile space and thus can serve to replicate the infinite number of periodic images. The expression for the minimum image distance changes for each one. The choice of one of these alternate geometries can be motivated by

- the desire to apply a particular symmetry. For example, a hexagonal simulation box may better accommodate hexagonal ordering in a structured fluid or crystal

- to save on the number of simulation atoms required. The octahedron resembles a cube with its corners cut off and thus requires less atoms. For systems in which a macromole-cule is solvated (e.g., a protein), the octahedron reduces the number of solvent molecules required (e.g., water).

Potential truncation

For **nonbonded pairwise interactions** we typically truncate the interactions between pairs of particles that are separated by large distances. However, we never apply such treatments to bonded interactions.

Cutting the potential

Many nonbonded pairwise potentials decay rapidly with separation distance. This enables us to ignore interactions between pairs of particles that are separated by large distances. We typically implement a cutoff where the energy is approximately only a percent or two of the minimum energy.

For example, consider the Lennard-Jones (LJ) potential:

$$u(r) = 4\epsilon\left[\left(\frac{r}{\sigma}\right)^{-12} - \left(\frac{r}{\sigma}\right)^{-6}\right]$$

A typical **cutoff distance** is $r_c = 2.5\sigma$, beyond which the potential is chosen to be zero:

$$u(r) = \begin{cases} 4\epsilon\left[\left(\frac{r}{\sigma}\right)^{-12} - \left(\frac{r}{\sigma}\right)^{-6}\right] & r \leq r_c \\ 0 & r > r_c \end{cases}$$

This is termed a **simple cut** of the potential. Notice that the energy at $r = 2.5\sigma$ is $\approx -0.016\epsilon$.

For efficiency reasons, we often avoid the computation of the square root for atom pairs beyond the cutoff using the following generic pairwise interaction loop:

```
loop over atom i = 1 to N − 1
      loop over atom j = i+1 to N
            compute r²ᵢⱼ
            if r²ᵢⱼ < r²_c then:

                  compute the interaction energy; if needed, compute rᵢⱼ = √(r²ᵢⱼ)
```

Considerations with periodic boundary conditions

With periodic boundary conditions, we want to avoid having to account for multiple periodic images of the same particles in the pairwise interaction loop. We can avoid dealing with multiple images if the cutoff distance is less than half of the simulation box width:
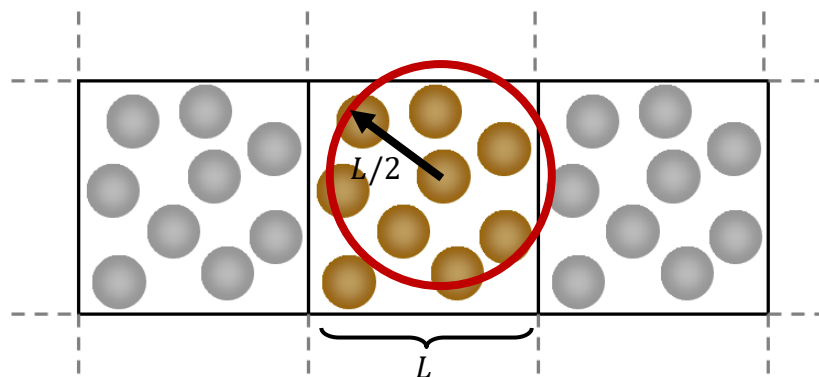
Notice that, for pairwise distances larger than $L/2$, a particle begins to interact with *multiple* images of the neighboring particles, rather than just the minimum image distance one. Therefore, simulations are generally set up so that all cutoff distances obey

$$r_c < \frac{L}{2}$$

Typically, a cutoff is chosen first and the minimum size of the simulation box is then computed. Given a desired bulk density, this will then also determine a minimum number of particles.

Cutting and shifting the potential

The simple truncation of the potential described above leads to a discontinuity in the pairwise potential energy function, across which the forces are undefined. A better approach is to **shift** the potential for $r < r_c$ so that the energy continuously approaches a zero value at r_c. To do this, we subtract from the pairwise potential the value of the energy evaluated at the cutoff. For the LJ system,

$$u(r) = \begin{cases} 4\epsilon \left[\left(\frac{r}{\sigma}\right)^{-12} - \left(\frac{r}{\sigma}\right)^{-6} \right] - 4\epsilon \left[\left(\frac{r_c}{\sigma}\right)^{-12} - \left(\frac{r_c}{\sigma}\right)^{-6} \right] & r \leq r_c \\ 0 & r > r_c \end{cases}$$

The **cut and shift** approach to potential truncation is perhaps the most popular treatment and results in better energy conservation in molecular dynamics than simple cutting alone.

Smooth truncation

The cut and shift approach still results in a potential that has discontinuities in the first derivative (the forces) and in higher order derivatives. For better stability, in particular during energy minimization, some authors employ a **switching function** that smoothly and continuously tapers the pair potential to zero between two cutoff distances.

A typical form that results in continuous first and second derivatives is:

$$u_{\text{smoothed}}(r) = \begin{cases} u(r) & r < r_{c,1} \\ S(t)u(r) & r_{c,1} \leq r < r_{c,2} \\ 0 & r_{c,2} \leq r \end{cases}$$

where the smoothing function $S(t)$ is given by:

$$S(t) = 1 - 10t^3 + 15t^4 - 6t^5 \quad \text{with} \quad t \equiv \frac{r - r_{c,1}}{r_{c,2} - r_{c,1}}$$

Corrections for cut interactions

Truncating pair interactions systematically removes a contribution to the net potential energy and pressure. For moderate cutoffs, such as $r_c = 2.5\,\sigma$ for the LJ system, this contribution can constitute a nontrivial fraction of the total.

For interactions that are cut but not shifted, one can approximately add the interactions beyond r_c back into the total energy by assuming $g(r > r_c) \approx 1$ and using the integral expressions for pairwise interactions:

$$U = U_{\text{pairs}} + U_{\text{tail}}$$

$$= \left[\sum_{i<j, r_{ij}<r_c} u(r_{ij}) \right] + \left[\frac{2\pi N^2}{V} \int_{r_c}^{\infty} r^2 u(r) dr \right]$$

Here, U_{pairs} is the energy for the explicit summation over atom pairs separated by $r_{ij} < r_c$. The second term on the RHS is the **tail correction**, and can be evaluated analytically. A similar approach can be applied to the pressure,

$$P = P_{\text{pairs}} + P_{\text{tail}}$$

$$= \left[\frac{Nk_BT}{V} - \frac{1}{3V} \sum_{i<j, r_{ij}<r_c} r_{ij} \frac{\partial}{\partial r_{ij}} u(r_{ij}) \right] + \left[-\frac{2\pi N^2}{3V^2} \int_{r_c}^{\infty} r^3 \frac{du(r)}{dr} dr \right]$$

For the Lennard-Jones system, we find that

$$U_{\text{tail}} = \frac{8\pi N^2}{3V} \epsilon \sigma^3 \left[\frac{1}{3} \left(\frac{r_c}{\sigma} \right)^{-9} - \left(\frac{r_c}{\sigma} \right)^{-3} \right]$$

$$P_{\text{tail}} = \frac{16\pi N^2}{3V^2} \epsilon \sigma^3 \left[\frac{2}{3} \left(\frac{r_c}{\sigma} \right)^{-9} - \left(\frac{r_c}{\sigma} \right)^{-3} \right]$$

If multiple atom types are present and all have the same interaction cutoff distance,

$$U_{\text{tail}} = \frac{2\pi}{V} \int_{r_c}^{\infty} r^2 \left[\sum_{X=1}^{M} \sum_{Y=1}^{M} N_X N_Y u_{XY}(r) \right] dr$$

$$P_{\text{tail}} = -\frac{2\pi}{3V} \int_{r_c}^{\infty} r^3 \left[\sum_{X=1}^{M} \sum_{Y=1}^{M} N_X N_Y \frac{du_{XY}(r)}{dr} \right] dr$$

It can also be shown that cut potentials (not shifted) also incur an **impulse** correction term in the pressure due the discontinuous change in energy at r_c, but this is rarely used in simulations.

Corrections for cut and shifted interactions

For potentials that are cut and shifted, the expression for the correction to the energy also includes the averaged ($g(r) \approx 1$) interactions below the cutoff:

$$U_{\text{tail}} = \frac{2\pi N^2}{V} \int_0^{r_c} r^2 u(r_c) dr + \frac{2\pi N^2}{V} \int_{r_c}^{\infty} r^2 u(r) dr$$

$$= \frac{2\pi N^2}{3V} r_c^3 u(r_c) + \frac{2\pi N^2}{V} \int_{r_c}^{\infty} r^2 u(r) dr$$

For multiple atom types with the same r_c,

$$U_{\text{tail}} = \frac{2\pi}{3V} r_c^3 \left[\sum_{X=1}^{M} \sum_{Y=1}^{M} N_X N_Y u_{XY}(r_c) \right] + \frac{2\pi}{V} \int_{r_c}^{\infty} r^2 \left[\sum_{X=1}^{M} \sum_{Y=1}^{M} N_X N_Y u_{XY}(r) \right] dr$$

The pressure tail correction, however, is the same as in the cut case.

Considerations for potential truncation corrections

Generally speaking, these corrections should *not* be applied if:

- The system is not a bulk phase, or involves interactions with interfaces.

- The system is anisotropic.

- The pairwise interactions themselves are anisotropic (e.g., rotation of an atom affects the pairwise energy… this would be the case if the atom had a dipole interaction on it).

Generally speaking, unless quantitative agreement is desired for bulk averages, one can perform a simulation without using the tail corrections and obtain the same qualitative (and sometimes even quantitative) results. By performing runs for various values of r_c, it is then possible to assess the influence of neglecting the cutoff interactions. It is not uncommon to neglect the tail corrections, but one should always be careful to report the actual tail treatment in any presentation of simulation results.

Treating long-ranged interactions

Pair interactions that do not decay fast enough with distance can be problematic for simulations. Let a generic pair interaction have the form

$$u(r) = ar^{-n}$$

Here, a is a constant (positive or negative) and n is an exponent that is greater than one. Consider the tail correction:

$$U_{\text{tail}} = \frac{2\pi N^2}{V} \int_{r_c}^{\infty} r^2 a r^{-n} dr$$

$$= \frac{2\pi N^2 a}{V} \int_{r_c}^{\infty} r^{2-n} dr$$

This integral **diverges** for $n \leq 3$. Interactions of this sort are termed **long-ranged**, and they require special treatment. For $n > 3$, the interactions are **short-ranged** and can be handled with a tail correction, or simply truncated without correction.

Physically, what happens with long-ranged interactions is the following. The number of atoms interacting with a central atom grows as r^2. On the other hand, the energy decays as r^{-n}. If the energy does not decay fast enough, the net contribution of energies from atoms farther away will outweigh that from those nearby; the total energy therefore becomes infinite in a bulk system.

Coulomb interactions

A common long-ranged interaction is the Coulomb law that applies to atoms with partial charges:

$$u(r_{ij}) = \frac{q_i q_j}{4\pi \epsilon_0 r_{ij}}$$

We could imagine that the tail integral for a central particle i might look like:

$$U_{\text{tail},i} = \frac{2\pi N^2}{V} \int_{r_c}^{\infty} r^2 \frac{q_i \langle q \rangle_r}{4\pi \epsilon_0 r} dr$$

$$= \frac{N^2 q_i}{2\epsilon_0 V} \int_{r_c}^{\infty} r \langle q \rangle_r dr$$

Here $\langle q \rangle_r$ is the average charge of particles in a spherical shell a distance r away from the central particle. For a bulk neutral system (zero net charge), we expect that $\langle q \rangle_r \to 0$ as $r \to \infty$. Therefore, it is possible that this integral could converge.

It turns out that Coulombic interactions, while long-ranged, are in fact **convergent** for **neutral systems**. However, these interactions require special treatment because it is not possible to compute a tail correction directly; the quantity $\langle q \rangle_r$ is not known as a function of r. Instead, special techniques are used to perform the sum of the interactions for an infinite number of replicas of the simulation cell.

For periodic systems with a net charge, the total energy depends on the number of periodic replicas. In the limit of an infinite number of replicas, these systems have an infinite energy and are not thermodynamic (i.e., the energy does not scale extensively with system size).

Note that these considerations **do not apply to the screened Coulomb potential**, which is short-ranged due to the damping exponential involving the Debye length. Keep in mind that the screened Coulomb potential is a highly approximate description of electrostatic interactions.
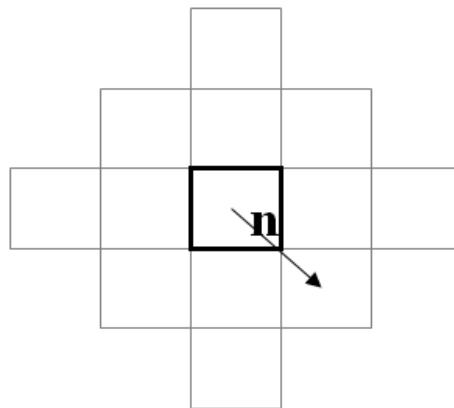
The Ewald summation

General approach

A formal way to treat Coulombic interactions was devised by Ewald in 1921. At a fundamental level, the goal of this approach is to compute the electrostatic energies summed over the central simulation box and all periodic image replicas of it:

$$U = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}^0} + \sum_{\mathbf{n}}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\epsilon_0 |\mathbf{r}_{ij}^0 + \mathbf{n}|}$$

The first term denotes the interactions within the central box, and the second denotes those between atoms in the central box and images and between atoms in various periodic replicas. Here, the superscript 0 indicates the minimum image distance. Moreover, the vector $\mathbf{n}$ denotes all lattice vectors pointing from the central box to a periodic image:

$$\mathbf{n} = \left(\pm n_x L, \pm n_y L, \pm n_z L\right) \quad n_x, n_y, n_z \text{ positive integers such that } n_x^2 + n_y^2 + n_z^2 \geq 1$$


As written above, the summation over an infinite number of periodic replicas converges very slowly. Moreover, it is **conditionally convergent**, meaning that care must be taken in the order in which the terms are added.

Ewald's solution to this problem was to turn the sum above into a mathematically equivalent version that converges much faster. Note that a point charge i can be expressed as a delta function in the charge density:

$$\rho_i(\mathbf{r}) = q_i \delta(\mathbf{r} - \mathbf{r}_i)$$

The Ewald solution is to subtract from the point charges equivalent smeared charges whose density is Gaussian in space. These charges must then be identically added to maintain the original system:
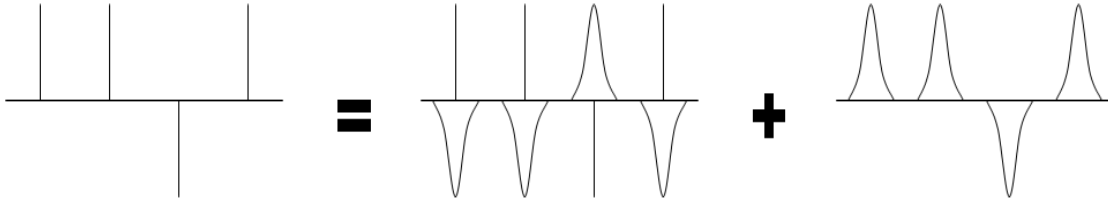
The two terms that result from this approach (on the RHS of the diagram above) are termed the **real-space** and **reciprocal-space** components of the electrostatic energy. They are so-named because the latter is solved in Fourier space.

Real-space component

For the real-space component, we subtract the smeared from the discrete charges,

$$\rho_i(\mathbf{r}) = q_i \left[\delta(\mathbf{r} - \mathbf{r}_i) - \alpha^3 \pi^{\frac{3}{2}} e^{-\alpha^2 |\mathbf{r} - \mathbf{r}_i|^2} \right]$$

Here, α is often called the **decay parameter** and controls the width of the Gaussian distribution. The limit $\alpha \to \infty$ returns an infinitely peaked Gaussian. The factors in front of the exponential on the RHS ensure normalization across space of the smeared distribution.

By solving Poisson's equation for the charge density, one obtains the following electrostatic energy due to this charge density:

$$U_{\text{real}} = \sum_{i<j} \frac{q_i q_j}{4\pi\epsilon_0} \text{erfc}(\alpha r_{ij})$$

Here, erfc is the complementary error function. This function decays very fast with respect to its argument, such that the pair interaction represented here is no longer long-ranged. As a result, a simple truncation of this interaction at $r_c = L$ usually has no effect on the summation.

Reciprocal-space component

To the real-space term, we must add the reciprocal energy, given by the sum of the smeared-out charges in the central box and all image particles in the infinite replication of the central box. One can find this energy by summing the Gaussian charges in a Fourier series. A detailed derivation is available in Frenkel and Smit. The result is:

$$U_{\text{recip}} = \frac{1}{2L^3\epsilon_0}\sum_{\mathbf{k}\neq 0}\frac{1}{|\mathbf{k}|^2}|\rho(\mathbf{k})|^2\exp\left(-\frac{|\mathbf{k}|^2}{4\alpha}\right) - \frac{\alpha^{\frac{1}{2}}\pi^{\frac{1}{2}}}{4\pi\epsilon_0}\sum_{i=1}^{N}q_i^2$$

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

Here, the summation proceeds over all **reciprocal-space vectors**:

$$\mathbf{k} = \left(\pm\frac{2\pi n_x}{L}, \pm\frac{2\pi n_y}{L}, \pm\frac{2\pi n_z}{L}\right) \quad n_x, n_y, n_z \text{ positive integers such that } n_x^2 + n_y^2 + n_z^2 \geq 1$$

Typically the summation is not taken to infinity but to a finite number of vectors. A common criterion is to include all vectors such that

$$|\mathbf{k}| < k_{\text{max}}$$

Notice that the reciprocal summation involves a reciprocal space density $\rho(\mathbf{k})$ that is a complex number. The computation of this quantity is usually performed using complex data types in Fortran or other programming languages.

Polarization energy

There is one final interaction that it often included in computations with electrostatics in addition to the two terms above. Neutral periodic systems involving partial charges can still have a net nonzero dipole, and this dipole can interact with whatever boundary exists around the system. This **polarization energy** is given by:

$$U_{\text{pol}} = \frac{1}{(2\epsilon_{\text{ext}} + 1)8\pi\epsilon_0 L^3}\left|\sum_{i=1}^{N}q_i\mathbf{r}_i\right|^2$$

Here, ϵ_{ext} is the relative permittivity of the medium at the boundaries. For good conductors like metals, $\epsilon_{\text{ext}} \to 0$ and the polarization energy vanishes.

Parameters of the Ewald method

There are two parameters of the Ewald approach that affect the convergence behavior of the sum above:

- α – controls the rate of decay of the Gaussian smeared charges, and of the real-space interactions $\sim\text{erfc}(\alpha r)$. Faster decay rates increase the accuracy of the real-space calculation but increase the number of reciprocal space vectors $\mathbf{k}$ required for equivalent accuracy in the reciprocal summation.

- k_max - controls the number of reciprocal space vectors used, proportional roughly to k_max^3. More vectors increases accuracy, but requires greater memory storage and computation time.

A typical choice for these parameters that reflects a compromise between the two is:

$$\alpha = \frac{5.6}{L} \qquad k_\text{max} = 5 \times \frac{2\pi}{L}$$

A longer discussion of the choice of values for these can be found in Frenkel and Smit.

Performance of the Ewald method

Ewald summations add substantial computational expense to the evaluation of the energy and forces, and typically dominate the simulation time. Nominally, the Ewald method scales as N^2 for the real-space term and as N for the reciprocal-space term. However, by choosing

$$\alpha \propto \frac{N^{1/6}}{L}$$

the algorithm can be made to scale as $N^{3/2}$ [Frenkel and Smit].

Typically the reciprocal-space summation is the more computationally intensive. Fast Fourier transforms (FFTs) can speed this part of the method substantially, and result in better cost scalings of $N \log N$ for the algorithm.

For very large systems, typically of 1000s or greater numbers of atoms, the Ewald method becomes extremely expensive. In these cases, alternative **mesh-based** approaches can be used in which the partial charges on the atoms are distributed to grid points on a finely-discretized cubic lattice in space. These methods take advantage of the fact that the reciprocal-space summation can be solved much faster using FFTs on lattices. Some methods are:

- **Particle-Mesh Ewald (PME)**

- **Smooth Particle-Mesh Ewald (SPME)**

- **Particle-Particle/Particle-Mesh Ewald (PPME)**

These algorithms typically scale as $N \log N$ in computational expense. The choice of which to use often depends on whether or not the forces are needed to high accuracy (e.g., for MD simulations).

Slab geometries

Systems that are periodic in only two dimensions are said to have **slab geometry**. Examples include fluids confined between walls and thin films. For these systems, the Ewald summation must be modified because the original equations assume infinite replication in all spherical dimensions. Several approaches have been developed for these situations. The reader is referred to Frenkel and Smit.