

Efficient Evaluation of Forces

Jim Phillips

Theoretical Biophysics Group

Beckman Institute, UIUC

Components of the force field

$$\begin{aligned}
 U(\vec{R}) = & \underbrace{\sum_{\text{bonds}} k_i^{\text{bond}} (r_i - r_0)^2}_{U_{\text{bond}}} + \underbrace{\sum_{\text{angles}} k_i^{\text{angle}} (\theta_i - \theta_0)^2}_{U_{\text{angle}}} + \\
 & \underbrace{\sum_{\text{dihedrals}} k_i^{\text{dihe}} [1 + \cos(n_i \phi_i + \delta_i)]}_{U_{\text{dihedral}}} + \\
 & \underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]}_{U_{\text{nonbond}}} + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}
 \end{aligned}$$

Nonbonded cutoffs

- Lennard-Jones additive but decays as $1/r^6$.
- Electrostatic decays as $1/r$, tends to cancel.
- Naive N^2 algorithm expensive for large N .
- Neglect interactions beyond r_{cut} .
- Reduces interaction work to order $N r_{\text{cut}}^3$.
- May be based on neutral groups of atoms.

Switching functions

- Force discontinuities cause heating artifacts.
- (Most errors cause energy gain since $T > 0$.)
- Applied function brings force and energy smoothly to zero at r_{cut} .

Pairlists

- Pairlist contains all atom pairs within r_{cut} .
- Determining pairlist contents is order N^2 .
- By expanding pairlist cutoff, it is calculated less often, but contains extra atoms.
- Ideally pairlist is recalculated only when an atom has moved more than $r_{\text{pairlist}} - r_{\text{cut}}$.

Pairlist update frequencies

- Biomolecular simulations have certain timescales as liquid phase systems.
- Local oscillations on 10 fs timescales.
- Slow diffusion on 1 ps timescales.
- Collective motions don't affect pairlist.
- Typically use 20 fs for a 1.5 Å tolerance.

Cells and neighbor lists

- Distance testing for a pairlist is order N^2 .
- Decompose space into a 3D lattice of cells, assigning atoms to cells is order N .
- List of pairs of cells within r_{cut} is constant.
- Some waste at corners of cells, but using a larger number of cells is also inefficient, as is combining pairlists and cells.

Cache based microprocessors

- Accessing memory randomly is expensive.
- Data in a small region of memory can be accessed quickly until other data is needed.
- Cell-based designs perform all calculations involving a pair of cells at once and may therefore be more efficient than pairlists.
- Memory access is only getting “slower”.

Long range electrostatics

- Cancellation of long range electrostatic terms does not occur in all systems.
- Lipids bilayers have a low dielectric.
- Nucleic acids have many charged groups.
- Long range electrostatics may be needed for qualitative or quantitative accuracy.
- If it can be fast, then no need for cutoffs.

Fast multipole algorithm

- Early method, order N (or at least $N \log N$).
- Multilevel approximation of distant groups of charges as a multipole expansion.
- Fields due to distant multipoles approximated as a Taylor expansion.
- Also used in gravitational simulations.
- Can be adapted to periodic systems.

Limitations of FMA

- Complex algorithm, hard to code well.
- Scales well, but slow for small systems.
- Cell-based multipoles result in discontinuous forces and energies.
- More accuracy is needed to conserve energy than is required by the physics or biology.

Particle-particle particle-mesh

- Surround each point charge by smooth, localized clouds of opposite charge.
- Particle-particle interactions cutoff exactly.
- Long-range interactions are smooth and can be approximated as a regular grid.
- Convolve electrostatic kernel with grid using (fast) Fourier transforms.

Particle mesh Ewald

- Similar in operation to PPPM, but based on Ewald formulation for periodic systems.
- Smoothly separate interactions into real and reciprocal space terms via Gaussians.
- Evaluate reciprocal sum using FFT.
- Can be formulated to conserve momentum or to conserve energy, but not both.

Fast Fourier transforms

- Discrete Fourier transform is order N^2 .
- If $N = 2^m$, can factor common operations to create an algorithm of order $N \log N$.
- Similar factorizations for $N = 2^m 3^n 5^p$ etc.
- Therefore each grid dimension should have only small factors (2, 3, 5, 7) to be efficient.
- PME grid spacing should be about 1 Å.

Interpolation tables

- Ewald adds $\text{erfc}()$ to short range calculation.
- Interpolation table can subsume $\text{erfc}()$, as well as $\text{sqrt}()$, switching functions, etc.
- Cubic interpolation of $U(r^2)$ based on energies and forces at endpoints.
- Using $a + b(x-x_0) + c(x-x_0)^2 + d(x-x_0)^3$ and exactly representable x_0 avoid cancellation.

Accuracy and artifacts

- Accuracy is the degree to which the physical interactions embodied in the force field and parameters are well represented.
- Accuracy of the force field is limited!
- Artifacts are nonphysical behaviors in the simulation which are introduced by systematic errors in force calculation, etc.