Special Topic 3

Computation of Force Fields for Molecular Simulation

Robert Krasny (Michigan)

outline

- Ewald summation
- treecodes
- Fast Multipole Method

warm-up quiz

"Everyone knows that the real reason universities have students is ....."

warm-up quiz

"Everyone knows that the real reason universities have students is to educate the professors."

-John Wheeler

(as quoted in *Physics Today*, April 2009)

1. Introduction

Coulomb potential : $\phi(x) = \dfrac{1}{4\pi|x|}$

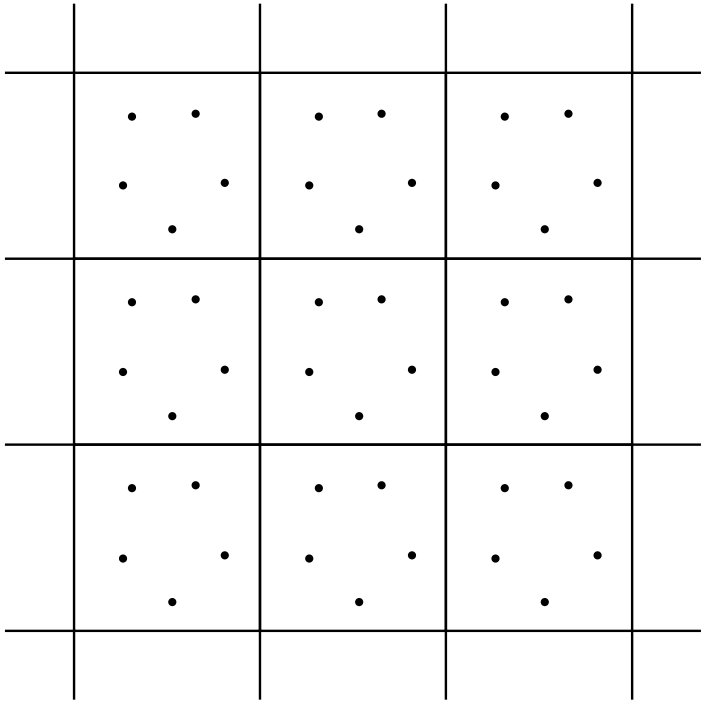
two properties : singularity at $x = 0$, slow decay as $x \to \infty$

consider a system of charged particles : x_i , q_i , $i = 1 : N$

potential energy : $V = \dfrac{1}{2} \displaystyle\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{q_i q_j}{4\pi|x_i - x_j|}$: $O(N^2)$

force field : $F_i = -\, q_i \displaystyle\sum_{\substack{j=1 \\ j \neq i}}^{N} q_j \frac{x_i - x_j}{4\pi|x_i - x_j|^3}$: $O(N^2)$

2. periodic boundary conditions



potential energy : $V = \frac{1}{2} \sum_{n} {\sum_{i,j=1}^{N}}' \frac{q_i \, q_j}{4\pi |x_i - x_j + nL|}$

cell index : $n = (n_1, n_2, n_3)$, $n_i = 0, \pm 1, \ldots$

$'$ means omit $i = j$ for $n = (0, 0, 0)$

3. issues with convergence

$$V = \frac{1}{2}\sum_{n}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{q_i q_j}{4\pi|x_i - x_j|}$$

assume $\sum_{i=1}^{N} q_i = 0$: charge neutrality

$$\Rightarrow \begin{cases} \text{(1) the series for } V \text{ is conditionally convergent} \\ \text{(2) the result depends on the order of summation} \\ \text{(3) regardless of the order, the series converges slowly} \end{cases}$$

example

$$1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} + \tfrac{1}{7} - \tfrac{1}{8} + \cdots = \ln 2$$

$$1 + \tfrac{1}{3} - \tfrac{1}{2} + \tfrac{1}{5} + \tfrac{1}{7} - \tfrac{1}{4} + \tfrac{1}{9} + \tfrac{1}{11} - \cdots = \tfrac{3}{2}\ln 2 \quad : \text{ reordered}$$

$$\tfrac{2}{1\cdot 3} + \tfrac{2}{3\cdot 3^3} + \tfrac{2}{5\cdot 3^5} + \tfrac{2}{7\cdot 3^7} + \tfrac{2}{9\cdot 3^9} + \cdots\cdots\cdots = \ln 2 : \quad \text{better}$$

4. Ewald summation

$$V = \frac{1}{2} \sum_{n} {\sum_{i,j=1}^{N}}' \frac{q_i \, q_j}{4\pi |x_i - x_j + nL|}$$

$$= V^r + V^k + V^s \text{ : one choice}$$

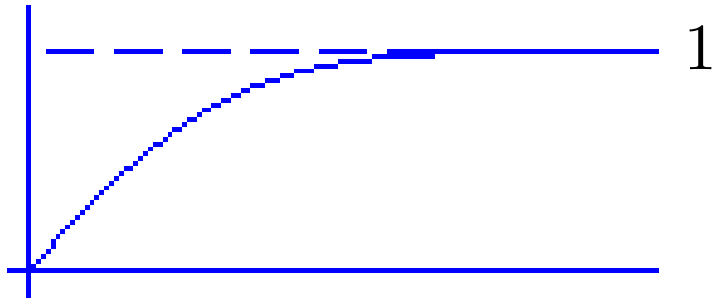
$$V^r = \frac{1}{2} \sum_{n} {\sum_{i,j=1}^{N}}' q_i \, q_j \frac{\mathrm{erfc}(\alpha |x_i - x_j + nL|)}{4\pi |x_i - x_j + nL|} \text{ : real-space term}$$

$$V^k = \frac{1}{2} \sum_{i,j=1}^{N} q_i \, q_j \sum_{k \neq 0} \frac{e^{-\frac{\pi^2 |k|^2}{L^2 \alpha^2} + 2\pi i k \cdot (x_i - x_j)/L}}{4\pi^2 L \, |k|^2} \text{ : reciprocal-space term}$$

$$V^s = \frac{-\alpha}{4\pi^{3/2}} \sum_{i=1}^{N} q_i^2 \text{ : self-energy term}$$

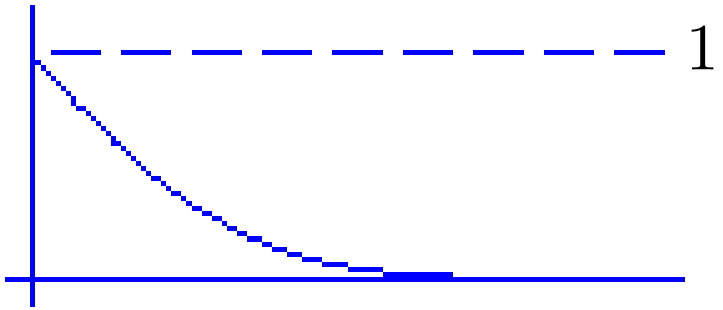
5. some ingredients

- $\mathrm{erf}(r) = \dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^r e^{-s^2} ds$:



- $\mathrm{erfc}(r) = \dfrac{2}{\sqrt{\pi}} \displaystyle\int_r^\infty e^{-s^2} ds$:

$$= 1 - \mathrm{erf}(r)$$



$\mathrm{erfc}(0/1/2) = 1.0000/0.1573/0.0047$: rapidly decaying

- $f_\alpha(x) = \dfrac{\alpha^3}{\pi^{3/2}} e^{-\alpha^2 |x|^2}$: Gaussian charge distribution

$$\int_{\mathbb{R}^3} f_\alpha(x) dx = 1 \ , \ \lim_{\alpha \to \infty} f_\alpha(x) = \delta(x) \ , \ -\Delta \frac{\mathrm{erf}(\alpha|x|)}{4\pi|x|} = f_\alpha(x)$$

6. operation count for Ewald summation

V^r, V^k are evaluated using cutoffs : $|x_i - x_j + nL| < r_c$, $|k| \leq k_c$

large α $\Rightarrow$ $\begin{cases} V^r \text{ converges rapidly} & : \ O(N) \\ V^k \text{ converges slowly} & : \ O(N^2) \end{cases}$

small α $\Rightarrow$ $\begin{cases} V^r \text{ converges slowly} & : \ O(N^2) \\ V^k \text{ converges rapidly} & : \ O(N) \end{cases}$

- Choosing α, r_c, k_c properly, the cost can be reduced to $O(N^{3/2})$ (Fincham, 1993).

- For large α, the cost of computing V^k can be reduced to $O(N \log N)$ using the FFT-based particle-mesh Ewald method (Darden-York-Pedersen, 1993). PME is the state of the art.

- For small α, the cost of computing V^r can be reduced to $O(N \log N)$ using a treecode (Duan-Krasny, 2000).

7. insight into Ewald's method (1/3)

$\sum_n \frac{1}{4\pi|x+nL|}$: lattice sum, diverges, but ok as a distribution

$$-\Delta\frac{1}{4\pi|x|} = \delta(x) \;\Rightarrow\; -\Delta\sum_n \frac{1}{4\pi|x+nL|} = \sum_n \delta(x+nL)$$

define $\rho(x) = \sum_n \delta(x+nL) - \frac{1}{L^3}$, so $\int_{\text{cell}} \rho(x)dx = 0$

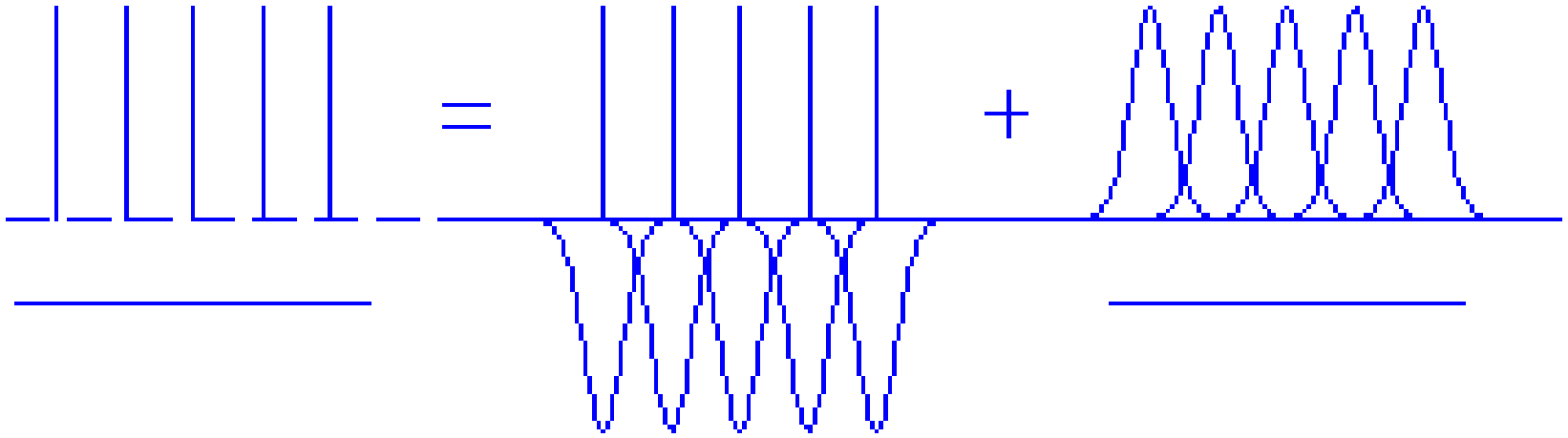
define $\psi(x)$: $-\Delta\psi(x) = \rho(x)$, PBC

Then $\psi(x)$ is well-defined (up to an additive constant) and pointwise values can be efficiently computed. How?

7. insight into Ewald's method (2/3)

$$\rho(x) = \sum_n \delta(x+nL) - \frac{1}{L^3}$$

$$= \sum_n \left(\delta(x+nL) - f_\alpha(x+nL)\right) + \sum_n f_\alpha(x+nL) - \frac{1}{L^3}$$



$\rho(x) = \rho_1(x) + \rho_2(x)$: both terms are charge neutral

$-\Delta\psi_1(x) = \rho_1(x)$, $-\Delta\psi_2(x) = \rho_2(x)$, PBC

then $\psi(x) = \psi_1(x) + \psi_2(x)$

7. insight into Ewald's method (3/3)

- $\psi_1(x) = \sum_n \frac{\mathrm{erfc}(\alpha|x+nL|)}{4\pi|x+nL|} - \frac{1}{4L^3\alpha^2}$

- $\psi_2(x) = \sum_{k\neq 0} \frac{e^{-\frac{\pi^2|k|^2}{L^2\alpha^2} + 2\pi i k\cdot x/L}}{4\pi^2 L\,|k|^2}$: Fourier series

These series are rapidly converging and hence $\psi(x) = \psi_1(x) + \psi_2(x)$ can be efficiently computed.

proof :

7. insight into Ewald's method (3/3)

- $\psi_1(x) = \sum_n \frac{\mathrm{erfc}(\alpha|x+nL|)}{4\pi|x+nL|} - \frac{1}{4L^3\alpha^2}$

- $\psi_2(x) = \sum_{k\neq 0} \frac{e^{-\frac{\pi^2|k|^2}{L^2\alpha^2} + 2\pi i k\cdot x/L}}{4\pi^2 L\,|k|^2}$: Fourier series

These series are rapidly converging and hence $\psi(x) = \psi_1(x) + \psi_2(x)$ can be efficiently evaluted.

proof : homework

7. insight into Ewald's method (3/3)

- $\psi_1(x) = \sum_n \frac{\mathrm{erfc}(\alpha|x+nL|)}{4\pi|x+nL|} - \frac{1}{4L^3\alpha^2}$

- $\psi_2(x) = \sum_{k\neq 0} \frac{e^{-\frac{\pi^2|k|^2}{L^2\alpha^2} + 2\pi i k\cdot x/L}}{4\pi^2 L\,|k|^2}$: Fourier series

These series are rapidly converging and hence $\psi(x) = \psi_1(x) + \psi_2(x)$ can be efficiently evaluted.

proof : homework

now that we understand all about Ewald summation ...

THE JOURNAL OF CHEMICAL PHYSICS **124**, 234104 (2006)

Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics

Christopher J. Fennell and J. Daniel Gezelter[a)]
Department of Chemistry Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556

(Received 24 March 2006; accepted 27 April 2006; published online 19 June 2006)

We investigate pairwise electrostatic interaction methods and show that there are viable computationally efficient ($\mathcal{O}(N)$) alternatives to the Ewald summation for typical modern molecular simulations. These methods are extended from the damped and cutoff-neutralized Coulombic sum originally proposed by Wolf *et al.* [J. Chem. Phys. **110**, 8255 (1999)]. One of these, the damped shifted force method, shows a remarkable ability to reproduce the energetic and dynamic characteristics exhibited by simulations employing lattice summation techniques. Comparisons were performed with this and other pairwise methods against the smooth particle-mesh Ewald summation to see how well they reproduce the energetics and dynamics of a variety of molecular simulations. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2206581]

8. free-space boundary conditions

$$V = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N}\frac{q_i q_j}{|x_i - x_j|} \quad : \quad \text{potential energy}$$

- direct summation

  particle-particle , $O(N^2)$

- treecode

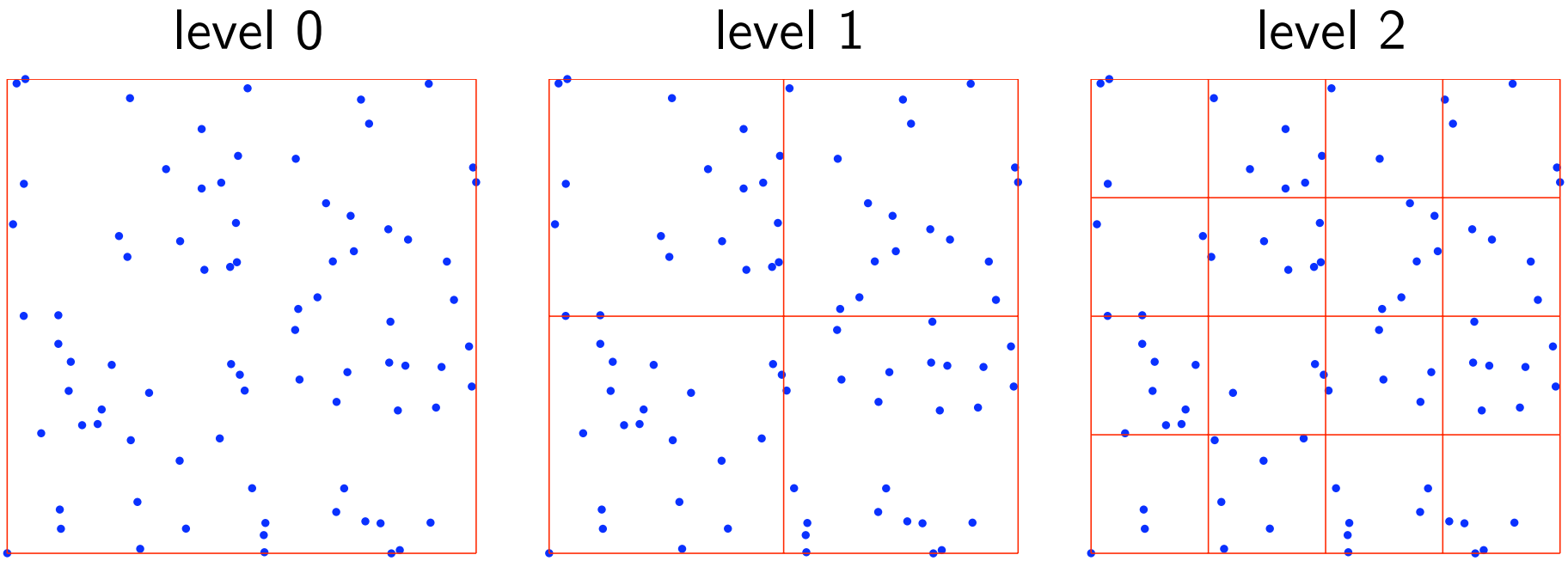
  Barnes-Hut (1986) , particle-cluster , $O(N \log N)$

- Fast Multipole Method

  Greengard-Rokhlin (1987) , cluster-cluster , $O(N)$ claimed

DISCLOSURE : I work on treecodes. -RK

9. treecode

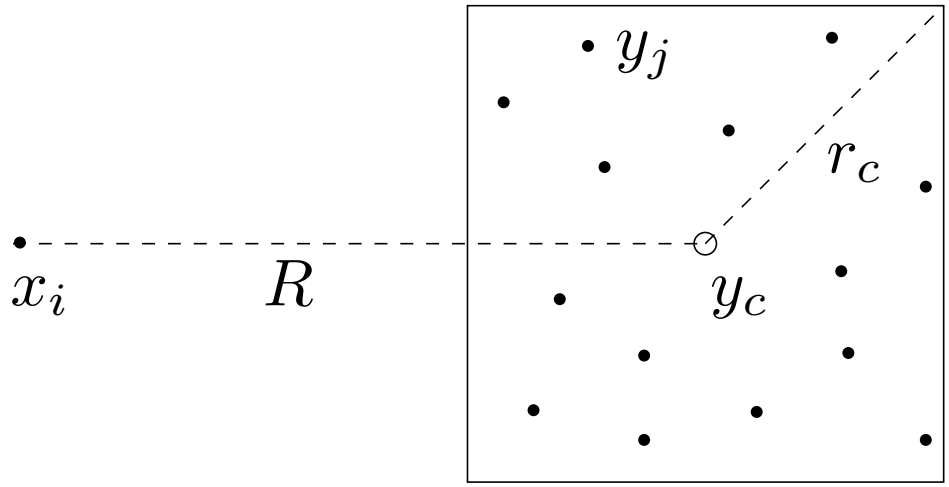


$$V = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{q_i q_j}{|x_i - x_j|} = \frac{1}{2}\sum_{i=1}^{N} q_i \sum_{c}\sum_{y_j\in c}\frac{q_j}{|x_i - y_j|}$$

BH used a monopole approximation for the particle-cluster interaction.

$$\sum_{y_j\in c}\frac{q_j}{|x_i - y_j|} \approx \frac{Q_c}{|x_i - y_c|}\ ,\ Q_c = \sum_{y_j\in c} q_j$$

10. multipole approximation for particle-cluster interaction



R : particle-cluster distance

y_c : cluster center

r_c : cluster radius

spherical coordinates

$$\sum_{y_j \in c} \frac{q_j}{|x_i - y_j|} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{M_l^m}{r_i^{l+1}} Y_l^m(\theta_i, \phi_i) \;:\; \text{multipole expansion}$$

$$M_l^m = \sum_{y_j \in c} q_j \, r_j^l \, Y_l^{-m}(\theta_j, \phi_j) \;:\; \text{cluster moments}$$

Cartesian coordinates

see posters of Boateng , Krasny for $\phi(x) = \dfrac{\mathrm{erfc}(\alpha|x|)}{|x|}$, $\dfrac{e^{-\kappa|x|}}{|x|}$

11. outline of treecode

• program main

construct tree

do $i = 1, N$

    compute-potential$(\mathbf{x}_i, root)$

end program

• subroutine **compute-potential**$(\mathbf{x}, c)$

if particle x and cluster c are well-separated

    compute and store moments of c (unless already available)

    compute particle-cluster interaction by multipole approximation

else

    if c is a leaf

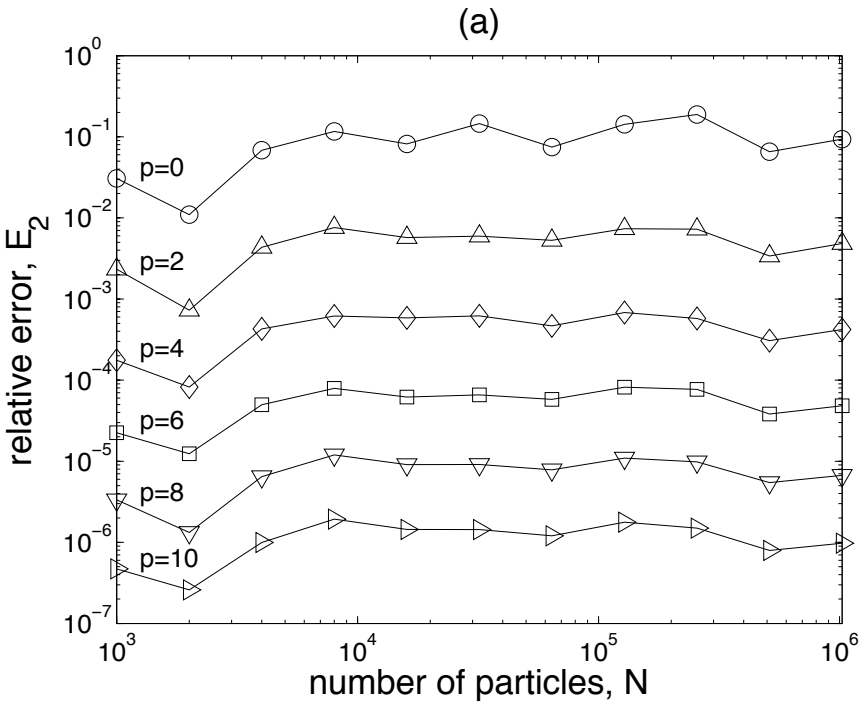
        compute particle-cluster interaction by direct summation

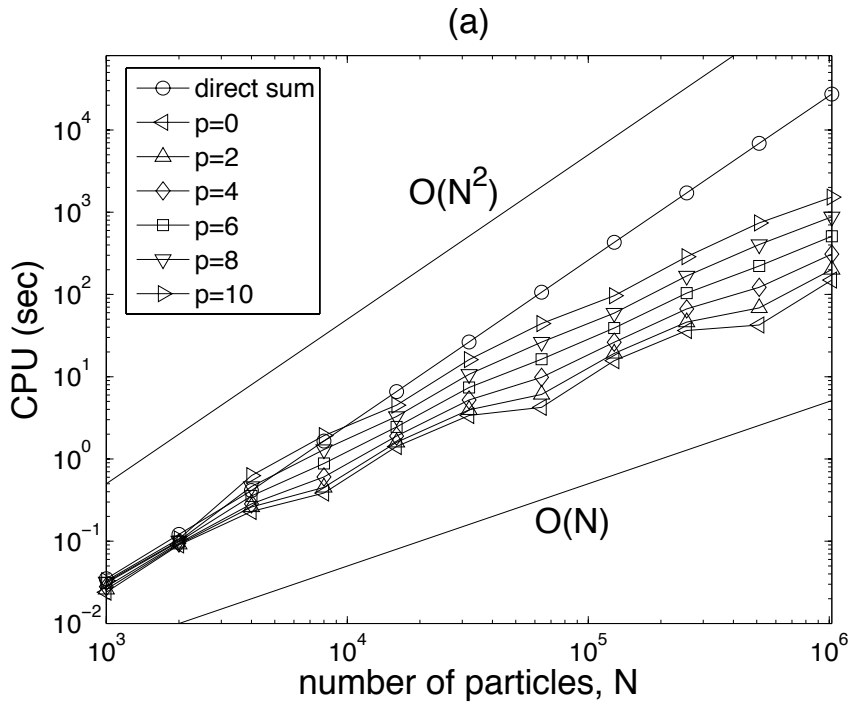
    else

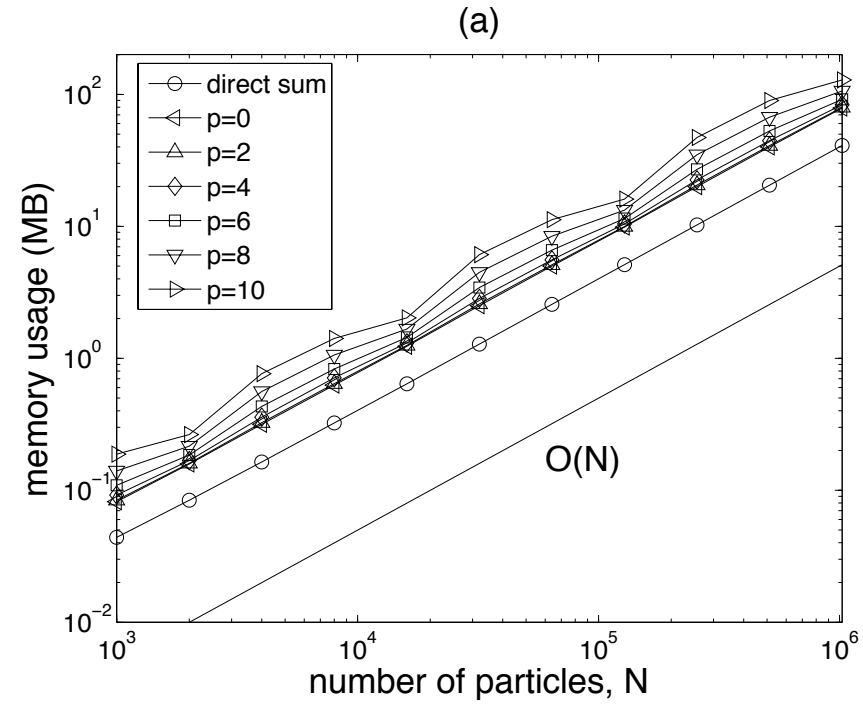
        call **compute-potential**$(\mathbf{x}, c\%\text{child})$ for each child of c

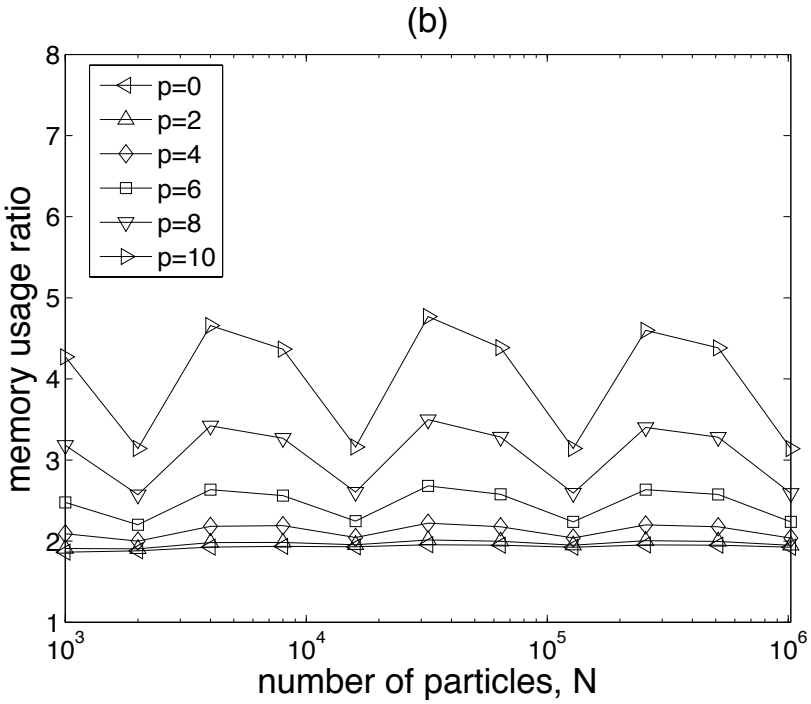
end subroutine

12. treecode performance for random particles in a cube



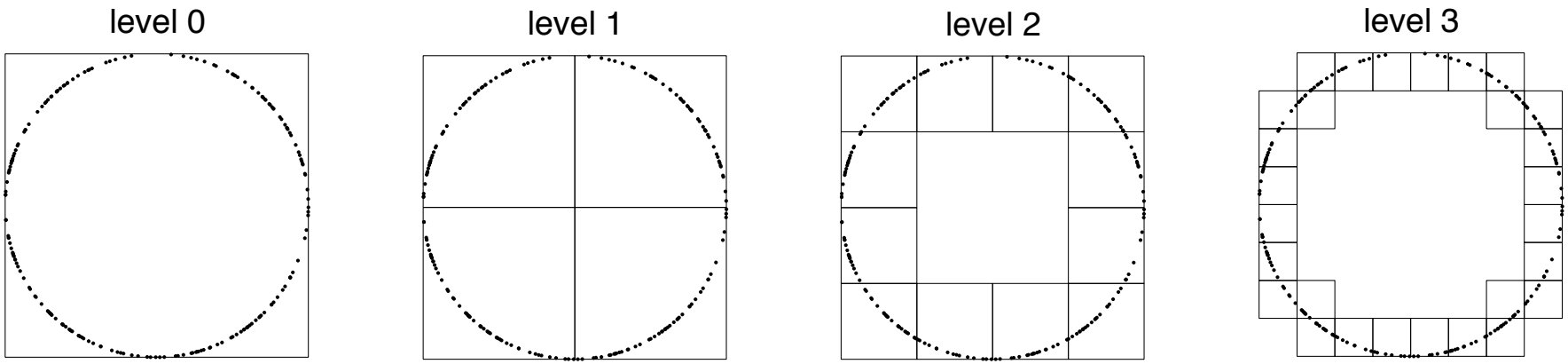




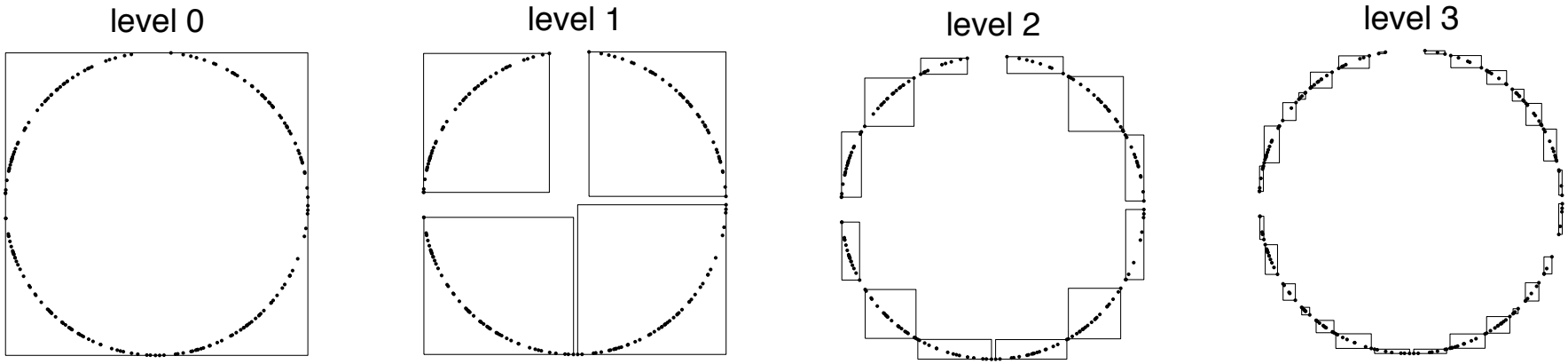


13. two types of clusters

- uniform cubes



- adapted rectangular boxes



14. fast multipole method : Greengard-Rokhlin (1987)

level 0

n	n	n	i
n	b	n	i
n	n	n	i
i	i	i	i

level 1

i	i	i	i	i	i		
i	i	i	i	i	i		
i	i	n	n	n	i		
i	i	n	b	n	i		
i	i	n	n	n	i		
i	i	i	i	i	i		

b : box , n : neighbor , i : interaction list

tools

- shift the center of a multipole expansion (M2M)
- convert a multipole expansion into a local expansion (M2L)
- shift the center of a local expansion (L2L)

15. outline of FMM algorithm

upward pass

- form multipole expansion of each box at finest level of tree

- form multipole expansion of each parent by shifting multipole expansions of children to center of parent using M2M

downward pass

- starting at coarsest level of tree, convert multipole expansion of each box b into local expansion about center of each box i in b's interaction list using M2L and add to local expansion in box i

- starting at coarsest level of tree, shift local expansion of parent to center of each child using L2L and add to local expansion of child

- at finest level of tree, evaluate local expansion at each particle and add to the direct sum with neighbor particles

16. variations of FMM (partial list)

- Poisson integral formula : Anderson (1992)
- plane-wave expansions : Greengard-Rokhlin (1997)
- kernel-independent FMM : Ying-Biros-Zorin (2004)

17. other fast summation methods

- particle-particle/particle-mesh : Hockney-Eastwood (1981)
- multilevel summation : Brandt-Lubrecht (1990)
  Skeel-Tezcan-Hardy (2002)

17. conclusions

- I discussed Ewald summation, treecodes, and the FMM.
- There are several options available for reducing the cost of computing long-range electrostatic interactions.
- There is a need for :
  - more comparison of different methods
  - open source versions
  - better understanding the range of application of PBC
  - automatic tuning of parameters
  - parallel/GPU implementations