Tools and Methods for Multiscale Biomolecular Simulations

Celeste Sagui

Department of Physics, NC State University, Raleigh, NC

Outline

1. Introduction (motivation and people)
2. Science (multiscale methods with quantum Monte Carlo, density functional, molecular mechanics, and continuum methods; interfaces; electrostatics; biology)
3. Education (CHiPS)
4. Summary
Multiscale Methods for Biomolecular Simulations

MM
<400,000 atoms
~10 ns

QM
~100 atoms
~400 electrons
~0–10 ps

MM/cont interface

QM/MM interface
~1000 atoms

continuum
• scientific aims are to produce a set of scalable and portable computational tools for multiscale calculations incorporating Quantum Monte Carlo (QMC), density functional theory (DFT), Molecular Mechanics (MM) and Continuum Methods along with suitable interfaces between them

• DFT, MM, and continuum solvent parts of the code will be based on real-space grids with multigrid acceleration and convergence

• electrostatics will treated in a highly efficient and accurate manner, which facilitates the development of a robust interface between them

• code will be used to solve paradigmatic biomolecular problems

• code will be distributed to public under the Open Source GPL license
we represent a partnership of 7 researchers located at 3 universities and the National Institute of Environmental and Health Sciences (NIEHS), all located within North Carolina’s Research Triangle

NC State Members: Jerry Bernholc, Lubos Mitas, Christopher Roland, and Celeste Sagui (PI) (Physics)

UNC Member: Lee Pedersen (Chemistry and NIEHS)

Duke Member: John Board (Computer Science and Electrical Engineering)

NIEHS Member: Thomas Darden (Structural Biology Lab)
Prior Accomplishments of Team Impacting ITR

• development of diffusion and variational QMC codes for molecules and condensed systems (L. Mitas)

• development of real-space DFT code with multigrid extension, along with newer O(N)-like electronic structure methods (J. Bernholc)

• development of the Particle Mesh Ewald (PME) method for electrostatics currently used in biomolecular codes such as AMBER and CHARMM (T. Darden and L. Pedersen)

• efficient scheme for treating dipolar contributions to electrostatics now in SANDER module of AMBER (C. Sagui, J. Board, T. Darden)

• development of new O(N) multigrid method for parallel implementation of electrostatics (C. Sagui, T. Darden)

• parallel implementation of Fast Multipole Method (FMM) (J. Board)

• suitable expertise in continuum-scale and statistical mechanics simulations (C. Roland, C. Sagui)
**Motivation:** accurate and efficient electrostatics is absolutely essential because:

(i) “partial charges” assigned to every atom in simulation

(ii) bottleneck in accuracy in classical force fields

(iii) crucial for QM/MM interface

**Challenges:**

(i) Eliminate artifacts associated with classical point charges

(ii) efficient simulation of these very costly long-range interactions
PME APPROACH TO FIXED AND INDUCED DIPOLAR INTERACTIONS


CPU-time [s]

Number of atoms

225MHz R10000 1 proc
Efficient Methods for Accurate Electrostatics

- Implement higher-order multipoles and polarizabilities using both PME and multigrid methods.
- Fit ab initio charge densities (or wavefunctions) of different species to atom- or bond-centered functions (Slaters and Wannier functions) to determine multipoles and improve electrostatic potentials.
- Introduce polarizabilities into biomolecular simulations.
- Address the issues of long-range multipolar energy and penetration effects.
### Results for 4096 water molecules

<table>
<thead>
<tr>
<th></th>
<th>$\beta$ (Ang$^{-1}$)</th>
<th>$R_c$ (Ang)</th>
<th>Spline Order</th>
<th>$H_x$ (Ang)</th>
<th>Direct (sec)</th>
<th>Reciprocal (sec)</th>
<th>Overall (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>charges</td>
<td>0.55</td>
<td>5.1</td>
<td>6</td>
<td>0.77</td>
<td>0.99</td>
<td>0.72</td>
<td>2.02</td>
</tr>
<tr>
<td>dipoles</td>
<td>0.6</td>
<td>4.7</td>
<td>6</td>
<td>0.68</td>
<td>1.33</td>
<td>1.02</td>
<td>2.73</td>
</tr>
<tr>
<td>quadrupoles</td>
<td>0.7</td>
<td>4.25</td>
<td>7</td>
<td>0.55</td>
<td>2.32</td>
<td>2.04</td>
<td>4.87</td>
</tr>
<tr>
<td>octupoles</td>
<td>0.85</td>
<td>3.75</td>
<td>10</td>
<td>0.51</td>
<td>3.79</td>
<td>3.38</td>
<td>8.13</td>
</tr>
<tr>
<td>hexadecapoles</td>
<td>1.00</td>
<td>3.55</td>
<td>12</td>
<td>0.41</td>
<td>5.70</td>
<td>5.30</td>
<td>13.07</td>
</tr>
<tr>
<td>hexadecapoles</td>
<td>1.00</td>
<td>3.55</td>
<td>12</td>
<td>0.38</td>
<td>7.90</td>
<td>7.70</td>
<td>17.60**</td>
</tr>
</tbody>
</table>

Relative RMS force error: $5 \times 10^{-4}$

Hexadecapoles with $R_c = 8$ Angstroms : cost $t = 92.5$ sec ($rms \sim 0.05$)

** RMS = $4 \times 10^{-5}$
Advantages of using real-space methods…

1. Real-space methods provide for a natural way of achieving $O(N)$ scaling in both quantum and classical simulations – FFTs entail substantial communication cost in parallel computing

2. Real-space methods are more flexible in terms of boundary conditions -- i.e., periodic, non-periodic, fixed, and mixed boundary conditions are possible

3. Grid-based methods can attain high-accuracy while preserving stability with additional gains to be achieved with non-uniform grids

→ many of our real-space algorithms aim to take advantage of the multigrid method
MULTIGRID ACCELERATION TECHNIQUE

STANDARD ITERATIVE TECHNIQUE WITH FIXED BOUNDARY CONDITIONS

SOLUTION ERROR FOR 1D POISSON'S EQUATION
LINEAR SCALING WITH NUMBER OF ATOMS

![Graph showing linear scaling with number of atoms. The x-axis represents the number of atoms, and the y-axis represents CPU time in seconds. Different runs (A, B, C, D, E, F) are indicated with distinct markers.](image-url)
LINEAR SCALING WITH NUMBER OF PROCESSORS
EXCELLENT ENERGY CONSERVATION
Multigrid method for quantum simulations

- Density functional equations solved directly on the grid
- Multigrid techniques remove instabilities by working on one length scale at a time
- Convergence acceleration and automatic preconditioning on all length scales
- Non-periodic boundary conditions are as easy as periodic
- Compact “Mehrstellen” discretization
- Allows for efficient massively parallel implementation


Speedup on Cray T3E with number of processors

Runs also on IBM SP, Origin 2000 and Linux clusters
Grid-optimized orbitals for nearly $O(N)$ DFT

• Large-scale electronic structure calculations scale as $O(N^3)$

• Current algorithmic developments aim at $O(N)$ scaling, see Goedecker, Rev. Mod. Phys. (1999) for a review.


• We keep: (i) accurate ab initio results; (ii) multigrid preconditioning; (iii) unoccupied orbitals to increase the convergence rate and handle metals.
Ab initio $O(N)$-like DFT calculations

• Expansion of the DFT total energy in localized, variationally optimized orbitals – very few orbitals needed, e.g., 3-4 orbitals per carbon atom
• Same computational cost as in tight-binding models for computing quantum conductances
• All operations performed on real-space grid with multigrid acceleration – fast convergence rate
• Main parts scale linearly with the number of atoms
• Unoccupied orbitals are essential (small $O(N^3)$ part)
• Fully parallel on Cray T3E, tested on > 1000 atoms
  • New code being developed for IBM SP, Beowulf
• Forces, geometry optimization

  Shape of an optimized orbital: valence bond function
  Slice through a plane tangent to carbon nanotube, $R_c = 6$ a.u.
Timings

- Implementation in C on CRAY T3E.
- Based on BLAS, Lapack, BLACS, PBLAS, and ScaLapack libraries
- Timing for 1 SC step (T3E, DEC alpha 450 MHz Processors, 256 MB RAM) $R_C = 6.2a.u.$, 3 orbitals/atom, $h = 0.34a.u.$: (grid 56 x 56 x 96 for 140 atoms)

<table>
<thead>
<tr>
<th></th>
<th>140</th>
<th>280</th>
<th>560</th>
<th>1120</th>
</tr>
</thead>
<tbody>
<tr>
<td># atoms</td>
<td>140</td>
<td>280</td>
<td>560</td>
<td>1120</td>
</tr>
<tr>
<td># orbitals</td>
<td>420</td>
<td>840</td>
<td>1680</td>
<td>3360</td>
</tr>
<tr>
<td># PEs</td>
<td>32</td>
<td>64</td>
<td>128</td>
<td>256</td>
</tr>
<tr>
<td># storage func.</td>
<td>237</td>
<td>252</td>
<td>255</td>
<td>255</td>
</tr>
<tr>
<td>CPU time/PE [s]</td>
<td>69</td>
<td>82</td>
<td>104</td>
<td>173</td>
</tr>
<tr>
<td>Subdiag. [s]</td>
<td>1.4</td>
<td>2.6</td>
<td>9</td>
<td>30</td>
</tr>
</tbody>
</table>

90 Mflops/PE for test on 128 PEs
Some Biological Applications:

I. Enzyme Reaction Mechanisms: map out using QM/MM reaction path energy/geometry profiles for

(a) Transfer of mononucleotide to growing strand of DNA
(b) the His/Asp/Ser active site of activated coagulation proteins
(c) cleavage of GTP to GDP by p21-H vas/GHP complex

a) Sawaya Biochem 1997  
b) Topf TCAccts 2001  
c) Cavalii JACS 2002
II. PROTEIN/PROTEIN COMPLEXES:
(Blood Coagulation Cascade)

- Find solvent-ion equilibrated structures using Multigrid MD
- Prepare complexes for QM/MM study

a) Factor Xa/Prothrombin (Pedersen Lab 2002)
b) Tenase (Pedersen Lab 2002)

a) Prothrombinase
   (factor Xa/prothrombin)
b) Tenase (factor X/factor viia/tissue factor)
III. ION DISTRIBUTION ABOUT DNA

- Find solvent-ion uni/divalent distributions around B-A-Z forms of DNA using multigrid MD

Hamelberg JACS 2002
Educational Aims of ITR

- ITR aims to provide a rich set of educational opportunities for students, faculty and research partners aimed at promoting the field of high performance biomolecular simulations within the NC triangle region.

- Specific thrusts aimed at facilitating this goal include:
  
  (a) Curriculum Development for Center for High Performance Supercomputing (CHiPS)
  
  (b) Interdisciplinary Short Courses
  
  (c) Simulation and Code Dissemination Workshops
  
  (d) Research Experience for Undergraduates (REU)
  
  (e) Technical Courses at the North Carolina Supercomputing Center (NCSC)
Summary

• scientific aims are to produce a set of scalable and portable computational tools for multiscale biomolecular calculations incorporating QMC, DFT, MM and continuum methods, along with suitable interfaces between different regions

• development of new and efficient methods for the treatment of the electrostatic forces is integral part of program

• code will be distributed to public under the Open Source GPL license