Linear scaling QM Methods

Gérald MONARD

Equipe de Chimie et Biochimie Théoriques
UMR 7565 CNRS - Université Henri Poincaré
Faculté des Sciences - B.P. 239
54506 Vandœuvre-les-Nancy Cedex - FRANCE

http://www.monard.info/
Outline

- Fundamentals of Linear Scaling methods
  - QM Bottlenecks
  - General ideas and solutions
- Focus on the Divide & Conquer method
  - D&C scheme at the semiempirical level
  - Energy Decomposition; Charge Transfer & Polarization
  - SemiEmpirical Born-Oppenheimer Molecular Dynamics
- Other linear scaling methods
  - Mozyne
  - CG-DMS
- Comments on the Parallelization of QM/MM and Linear Scaling methods
How to solve the QM scaling problem?

Theoretical CPU scaling order for different QM methods

<table>
<thead>
<tr>
<th>QM method</th>
<th>Scaling</th>
<th>bottleneck</th>
</tr>
</thead>
<tbody>
<tr>
<td>semiempirical</td>
<td>$O(N^3)$</td>
<td>diagonalization</td>
</tr>
<tr>
<td>DFT</td>
<td>$O(N^3)$</td>
<td>electronic integrals</td>
</tr>
<tr>
<td>ab initio</td>
<td>$O(N^4)$</td>
<td>electronic integrals</td>
</tr>
<tr>
<td>MP2</td>
<td>$O(N^5)$</td>
<td></td>
</tr>
<tr>
<td>Full CI</td>
<td>$O(\exp^N)$</td>
<td></td>
</tr>
</tbody>
</table>

- change the methods: use approximate quantum methods
  - semiempirical QM methods
  - molecular mechanics (MM) force fields
  - combined QM/MM methods

- change the algorithms
  - Linear scaling algorithms
**Linear Scaling in Quantum Chemistry**

**Linear Scaling objectives**

- A QM calculation with a $O(N)$ CPU scaling
- Standard algorithms must be changed
- New algorithms interesting beyond the crossover point

![Graph showing linear, cubic, and quartic scaling](image)
Changing the algorithms towards linear scaling

**CPU bottleneck in Hartree-Fock methods**

- bielectronic integrals
  
  \[ O(N^4) \]  
  
  HF: \[ O(N^4) \]  
  
  DFT: \[ O(N^3) \]

- orthogonal transformation  
  
  \[ O(N^3) \]

- fock matrix diagonalization  
  
  \[ O(N^3) \]

-reaching linear scaling for HF (or DFT) methods involves multiple targets and changes of algorithms

**CPU bottleneck in semiempirical methods**

- fock matrix diagonalization  
  
  \[ O(N^3) \]

-reaching linear scaling is/should be easier with semiempirical methods
Exploiting Electronic Density Locality

\[ \rho(r, r') \sim \exp\left(-\sqrt{E_{\text{gap}}} |r - r'| \right) \]

- \( E_{\text{gap}} \) is the energy difference between HOMO and LUMO
- Thus, in large atomic systems, \( \mathbf{P} \) is sparse
- Except in metallic or conjugated systems
- But true for proteins, small solute in solutions, etc
Some Linear Scaling Algorithms

An example: Revisiting the Hartree-Fock SCF algorithm

- Compute mono- and bielectronic integrals  
  Fast Multiple Method
- Build the Fock matrix $\mathbf{F}$  
  sparse matrix
- Orthogonal transformation  
  Cholesky decomposition
- Diagonalization of $\mathbf{F}$ replaced by Conjugate-Gradient Density Matrix Search

Among the published linear scaling methods

- Divide & Conquer  
  (SemiEmp., DFT, HF, etc.)
- Localized Molecular Orbitals  
  (SemiEmp. only)
- Density Matrix Minimization  
  (SemiEmp., DFT, HF, etc.)
The Divide and Conquer Method

If $\chi_\mu$ and $\chi_\nu$ are well separated in space ($> 9-11$ Å): $P_{\mu \nu} = 0$

Division of the system into $N_{\text{sub}} \mathcal{R}^\alpha$ subsystems: $P = \sum_{\alpha=1}^{N_{\text{sub}}} P_\alpha$

Full System       Small Systems

$F \rightarrow \{F_\alpha\}$ (for all systems)

$P \leftarrow \{C^i_\alpha\}$ (subsystem M.O.)

Theoretical speed-up $\sim N_{\text{sub}} (N^\alpha / N)^3$

Diagram of amino acid structure with buffer regions and core.
The Divide and Conquer Method

- Fock matrix of subsystem $\mathcal{R}^\alpha$:

$$F^\alpha_{\mu\nu} = \begin{cases} F_{\mu\nu} & \text{if } \chi_\mu \in \mathcal{R}^\alpha \text{ and } \chi_\nu \in \mathcal{R}^\alpha \\ 0 & \text{otherwise} \end{cases}$$

- For all subsystems, Roothan equations are solved:

$$F^\alpha C^\alpha = C^\alpha \varepsilon^\alpha$$

- Density submatrices are built:

$$P^\alpha_{\mu\nu} = \sum_{i}^{\text{MOs}} n_i^\alpha (c_{\mu i}^\alpha)^* c_{\nu i}^\alpha \quad \text{with} \quad n_i^\alpha = \frac{2}{1 + \exp[(\varepsilon_i^\alpha - \varepsilon_F)/kT]}$$

- Fermi energy $\varepsilon_F$ is obtained by solving the equation:

$$n_{\text{elec}} = \sum_{\mu=1}^{N} P_{\mu\mu} = \sum_{\mu=1}^{N} \sum_{\alpha=1}^{N_{\text{sub}}} \sum_{i}^{\text{MOs}} D_{\mu\mu}^\alpha \sum_{i} n_i^\alpha |c_{\mu i}^\alpha|^2 = \sum_{\alpha=1}^{N_{\text{sub}}} \sum_{i} n_i^\alpha B_i^\alpha$$

- Total density matrix $P$ is built from:

$$P_{\mu\nu} = \frac{\sum_{\alpha=1}^{N_{\text{sub}}} D_{\mu\nu}^\alpha P_{\mu\nu}^\alpha}{\sum_{\alpha=1}^{N_{\text{sub}}} D_{\mu\nu}^\alpha} \quad \text{with} \quad D_{\mu\nu}^\alpha = \begin{cases} 1 & \text{if } (\mu, \nu) \in (\text{core}^\alpha, \text{buffer} 1^\alpha) \\ 0 & \text{unless} \end{cases}$$
Semiempirical SCF algorithm

1. Compute mono- and bielectronic integrals
2. Build core hamiltonian (invariant) $H_c$
3. Guess an initial density matrix
4. Build the Fock matrix $F$
5. For each $\mathcal{R}^\alpha$ subsystems at iteration $(k)$:
   - Assemble $F^\alpha$ from $F^{(k)}$
   - Solve $F^\alpha \ C^\alpha = C^\alpha \ e^\alpha$
   - Store $B_i^\alpha$
   - Build $P^\alpha$ from $C^\alpha$, $e^\alpha$, and $e_{F}^{(k-1)}$
   - Add $P^\alpha$ to $P^{(k+1)}$

6. determine $e_{F}^{(k)}$ from all $B_i^\alpha$ and $N_{\text{elec}}$
   Back to 4. unless convergence
Semiempirical Divide & Conquer: timings

- (H₂O)ₙ water cluster (n from 1 to 500)
- 1 energy calculations
- DivCon99 (Merz et al. + modifications from Nancy Group)
- Standard algorithm or Divide & Conquer (ncore = 1, buffer1 = 4Å, buffer2 = 2Å)
- Intel(R) Xeon(R) CPU E5620 2.40GHz (8 cores) 32Gb RAM
Semiempirical Divide & Conquer implementations

- Linear-scaling semiempirical quantum calculations for macromolecules

- Semiempirical molecular orbital calculations with linear system size scaling

- Fast, accurate semiempirical molecular orbital calculations for macromolecules

- Critical assessment of the performance of the semiempirical divide and conquer method for single point calculations and geometry optimizations of large chemical systems
Other Divide & Conquer implementations


Scaling factor in Divide & Conquer


Figure 4. Average computational time to diagonalize the Fock matrix in each SCF cycle using traditional HF and DC-HF for a series of extended polyglycines at the HF/6-31G* level.

Figure 6. Similar to Figure 4 but for the polyaniline systems in an α-helical structure α-(Ala)₉.

1700 basis functions ~ 20 alanine amino acids (200 atoms)
Divide & Conquer Interaction Energy Decomposition

➢ The theoretical framework of the Divide & Conquer method allows for a natural decomposition of the interaction energy into electrostatic, polarization, and charge transfer components:

\[ E_{int} = E_{es} + E_{pol} + E_{CT} \]

➢ The total energy \( E[\varepsilon_F, P_r, r, b] \) of a multimolecular system is obtained from

\[ E[\varepsilon_F, P_r, r, b] = \frac{1}{2} \sum_{\mu} \sum_{\nu} P_{\mu\nu} [H_{\mu\nu}^c + F_{\mu\nu}] + E_{core} \]

➢ The total energy is a function of:

- ✔ the Fermi energy \( \varepsilon_F \)
- ✔ the intermolecular distance \( r \)
- ✔ the density matrix \( P_r \) at this intermolecular distance
- ✔ and the buffer region \( b \)
The interaction energy of a multimolecular system can then be expressed as

\[ E_{int} = E[\varepsilon_f, P_r, r, b] - E[\varepsilon_F, P_\infty, \infty, b] \]

The number of electrons per subsystem can be constrained by using multiple Fermi energies \( \varepsilon_F^\alpha \)

Electron flow between subsystems can be inhibited by applying zero-buffering (= every subsystem consists of one core)

For an infinitely separated system, the buffer region will always be empty between interacting molecules, therefore:

\[ E[\varepsilon_F, P_\infty, \infty, b] = E[\varepsilon_F^\alpha, P_\infty, \infty, 0] \]
Divide & Conquer Interaction Energy Decomposition

The electrostatic contribution to the interaction energy $E_{es}$ can be obtained from:

$$E_{es} = E[\varepsilon^\alpha_F, P_\infty, r, 0] - E[\varepsilon_F, P_\infty, \infty, 0]$$

This corresponds to the energy obtained by bringing the infinitely separated system to equilibrium distance, without a change in the charge distribution.

By allowing intramolecular charge rearrangement, without charge flow, the contribution of polarization is obtained:

$$E_{pol} = E[\varepsilon^\alpha_F, P, r, 0] - E[\varepsilon^\alpha_F, P_\infty, r, 0]$$

By allowing intermolecular charge flow, the contribution of charge transfer is:

$$E_{CT} = E[\varepsilon_f, P_r, r, b] - E[\varepsilon^\alpha_F, P, r, 0]$$
**Divide & Conquer Interaction Energy Decomposition**

64 water system (*JPCA* 1999, 103, 3321–3329)

- classical molecular dynamics with 64 TIP3P water molecules
- NPT: $T = 330$ K, $P = 1$ bar, $\Delta t = 1.5$ fs, $t = 330$ps
- 3 snapshots at 210, 270 and 330 ps
- single point PM3 energy calculations within D & C framework
- Interaction energy decomposition for all water

**TABLE 4: Interaction Energy Decomposition of 64 Waters**

<table>
<thead>
<tr>
<th>time (ps)</th>
<th>opt$^a$</th>
<th>$E_{\text{int}}$</th>
<th>$E_{\text{es}}$</th>
<th>$E_{\text{pol}}$</th>
<th>$E_{\text{CT}}$</th>
<th>%</th>
<th>%</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>210</td>
<td>no</td>
<td>$-225.75$</td>
<td>485.70</td>
<td>$-17.75$</td>
<td>$-693.70$</td>
<td>40.6</td>
<td>1.5</td>
<td>57.9</td>
</tr>
<tr>
<td>270</td>
<td>no</td>
<td>$-180.54$</td>
<td>607.89</td>
<td>$-15.64$</td>
<td>$-772.79$</td>
<td>43.5</td>
<td>1.1</td>
<td>55.4</td>
</tr>
<tr>
<td>330</td>
<td>no</td>
<td>$-221.38$</td>
<td>476.55</td>
<td>$-17.07$</td>
<td>$-680.86$</td>
<td>40.6</td>
<td>1.4</td>
<td>58.0</td>
</tr>
<tr>
<td>210</td>
<td>yes</td>
<td>$-385.73$</td>
<td>622.62</td>
<td>$-27.52$</td>
<td>$-980.83$</td>
<td>38.2</td>
<td>1.7</td>
<td>60.1</td>
</tr>
<tr>
<td>270</td>
<td>yes</td>
<td>$-380.00$</td>
<td>622.56</td>
<td>$-26.26$</td>
<td>$-976.30$</td>
<td>38.3</td>
<td>1.6</td>
<td>60.1</td>
</tr>
<tr>
<td>330</td>
<td>yes</td>
<td>$-389.60$</td>
<td>592.54</td>
<td>$-26.95$</td>
<td>$-955.19$</td>
<td>37.6</td>
<td>1.7</td>
<td>60.7</td>
</tr>
</tbody>
</table>

$^a$ Yes if a PM3 geometry optimization was performed.  
$^b$ In kcal/mol.
Divide & Conquer Interaction Energy Decomposition

64 water system (JPCA 1999, 103, 3321–3329)

Divide and Conquer Interaction Energy Decomposition

\[ E_{\text{int}} = E_{\text{cs}} + E_{\text{pol}} + E_{\text{CT}} \]

**Figure 4.** Effect of electrostatics, polarization, and charge transfer on the oxygen and hydrogen CM1 charge distribution of the PM3 optimized, 64 water system: (A) Oxygen charge distribution, \( E_{\text{int}} = E_{\text{cs}} \), \( E_{\text{pol}} = E_{\text{CT}} = 0 \); (B) hydrogen charge distribution, \( E_{\text{int}} = E_{\text{cs}}, E_{\text{pol}} = E_{\text{CT}} = 0 \); (C) oxygen charge distribution, \( E_{\text{int}} = E_{\text{cs}} + E_{\text{pol}}, E_{\text{CT}} = 0 \); (D) hydrogen charge distribution, \( E_{\text{int}} = E_{\text{cs}} + E_{\text{pol}}, E_{\text{CT}} = 0 \); (E) oxygen charge distribution, \( E_{\text{int}} = E_{\text{cs}} + E_{\text{pol}} + E_{\text{CT}} \); (F) hydrogen charge distribution, \( E_{\text{int}} = E_{\text{cs}} + E_{\text{pol}} + E_{\text{CT}} \).
Divide & Conquer Interaction Energy Decomposition

64 water system (JPCA 1999, 103, 3321–3329)

Figure 4. Effect of electrostatics, polarization, and charge transfer on the oxygen and hydrogen CM1 charge distribution of the PM3 optimized, 64 water system: (A) Oxygen charge distribution, $E_{\text{int}} = E_{\text{es}}, E_{\text{pol}} = E_{\text{CT}} = 0$; (B) hydrogen charge distribution, $E_{\text{int}} = E_{\text{es}}, E_{\text{pol}} = E_{\text{CT}} = 0$; (C) oxygen charge distribution, $E_{\text{int}} = E_{\text{es}} + E_{\text{pol}}, E_{\text{CT}} = 0$; (D) hydrogen charge distribution, $E_{\text{int}} = E_{\text{es}} + E_{\text{pol}}, E_{\text{CT}} = 0$; (E) oxygen charge distribution, $E_{\text{int}} = E_{\text{es}} + E_{\text{pol}} + E_{\text{CT}}$; (F) hydrogen charge distribution, $E_{\text{int}} = E_{\text{es}} + E_{\text{pol}} + E_{\text{CT}}$. 
Divide & Conquer Interaction Energy Decomposition

64 water system (JPCA 1999, 103, 3321–3329)

Figure 4. Effect of electrostatics, polarization, and charge transfer on the oxygen and hydrogen CM1 charge distribution of the PM3 optimized, 64 water system: (A) Oxygen charge distribution, $E_{\text{int}} = E_{\text{es}}$, $E_{\text{pol}} = E_{\text{CT}} = 0$; (B) hydrogen charge distribution, $E_{\text{int}} = E_{\text{es}}$, $E_{\text{pol}} = E_{\text{CT}} = 0$; (C) oxygen charge distribution, $E_{\text{int}} = E_{\text{es}} + E_{\text{pol}}$, $E_{\text{CT}} = 0$; (D) hydrogen charge distribution, $E_{\text{int}} = E_{\text{es}} + E_{\text{pol}}$, $E_{\text{CT}} = 0$; (E) oxygen charge distribution, $E_{\text{int}} = E_{\text{es}} + E_{\text{pol}} + E_{\text{CT}}$; (F) hydrogen charge distribution, $E_{\text{int}} = E_{\text{es}} + E_{\text{pol}} + E_{\text{CT}}$. 
Divide & Conquer Interaction Energy Decomposition

CspA + water (JACS 1999, 121, 9182–9190)

- CspA: cold shock protein A (69 residues, charged -1e)
- Molecular Dynamics 500ps
- MM force field: AMBER + TIP3P
- single point PM3 energy calculations within D & C framework (keep only water molecules < 9Å ~ 4500 atoms)

Table 2. Interaction Energy Decomposition for Solvated CspA

<table>
<thead>
<tr>
<th>time</th>
<th>method</th>
<th>$E_{\text{int}}$</th>
<th>$E_{\text{es}}$</th>
<th>$E_{\text{pol}}$</th>
<th>$E_{\text{CT}}$</th>
<th>%$E_{\text{es}}$</th>
<th>%$E_{\text{pol}}$</th>
<th>%$E_{\text{CT}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>PM3</td>
<td>$-1501.495$</td>
<td>$678.918$</td>
<td>$-256.474$</td>
<td>$-1923.939$</td>
<td>23.7</td>
<td>9.0</td>
<td>67.3</td>
</tr>
<tr>
<td></td>
<td>AM1</td>
<td>$-1394.793$</td>
<td>$-33.369$</td>
<td>$-235.601$</td>
<td>$-1125.823$</td>
<td>2.4</td>
<td>16.9</td>
<td>80.7</td>
</tr>
<tr>
<td>200</td>
<td>PM3</td>
<td>$-1489.493$</td>
<td>$698.240$</td>
<td>$-248.075$</td>
<td>$-1939.658$</td>
<td>24.2</td>
<td>8.6</td>
<td>67.2</td>
</tr>
<tr>
<td></td>
<td>AM1</td>
<td>$-1354.087$</td>
<td>$8.731$</td>
<td>$-224.368$</td>
<td>$-1138.450$</td>
<td>0.6</td>
<td>16.4</td>
<td>83.0</td>
</tr>
<tr>
<td>300</td>
<td>PM3</td>
<td>$-1494.170$</td>
<td>$743.241$</td>
<td>$-234.367$</td>
<td>$-2003.044$</td>
<td>24.9</td>
<td>7.9</td>
<td>67.2</td>
</tr>
<tr>
<td></td>
<td>AM1</td>
<td>$-1333.471$</td>
<td>$50.513$</td>
<td>$-214.238$</td>
<td>$-1169.746$</td>
<td>3.5</td>
<td>14.9</td>
<td>81.6</td>
</tr>
<tr>
<td>400</td>
<td>PM3</td>
<td>$-1550.705$</td>
<td>$869.492$</td>
<td>$-279.010$</td>
<td>$-2141.186$</td>
<td>26.4</td>
<td>8.5</td>
<td>65.1</td>
</tr>
<tr>
<td></td>
<td>AM1</td>
<td>$-1414.340$</td>
<td>$81.045$</td>
<td>$-249.260$</td>
<td>$-1246.124$</td>
<td>5.1</td>
<td>15.8</td>
<td>79.1</td>
</tr>
<tr>
<td>500</td>
<td>PM3</td>
<td>$-1376.857$</td>
<td>$929.004$</td>
<td>$-235.015$</td>
<td>$-2070.846$</td>
<td>28.7</td>
<td>7.3</td>
<td>64.0</td>
</tr>
<tr>
<td></td>
<td>AM1</td>
<td>$-1227.472$</td>
<td>$209.451$</td>
<td>$-217.185$</td>
<td>$-1219.738$</td>
<td>12.7</td>
<td>13.2</td>
<td>74.1</td>
</tr>
<tr>
<td>average</td>
<td>PM3</td>
<td>$-1482.544$</td>
<td>$783.779$</td>
<td>$-250.588$</td>
<td>$-2015.735$</td>
<td>25.6</td>
<td>8.2</td>
<td>66.2</td>
</tr>
<tr>
<td></td>
<td>AM1</td>
<td>$-1344.833$</td>
<td>$63.274$</td>
<td>$-228.130$</td>
<td>$-1179.976$</td>
<td>4.9</td>
<td>15.4</td>
<td>79.7</td>
</tr>
</tbody>
</table>

*a In ps. *b In kcal/mol. *c $%E_x = 100|E_x|/(|E_{\text{es}}| + |E_{\text{pol}}| + |E_{\text{CT}}|)$, $x = \text{es, pol, CT.}$
Figure 1. The effect of polarization on the charge distribution of CspA. Shown is the difference in the calculated CM2 charges when polarization is added to electrostatics. Charges are listed by residue. The bars indicate the observed range of charges from the five snapshots and the bold bars indicate the charged residues. (a) AM1. (b) PM3.
Figure 2. The effect of charge transfer on the charge distribution of CspA. Shown is the difference in the calculated CM2 charges when charge transfer is added to polarization + electrostatics. Charges are listed by residue. The bars indicate the observed range of charges from the five snapshots and the bold bars indicate the charged residues. (a) AM1. (b) PM3.
**Figure 3.** The effect of solvation on the charge distribution of CspA. Shown is the difference in the calculated CM2 charges between the solvated and vacuum protein. Charges are listed by residue. The bars indicate the observed range of charges from the five snapshots and the bold bars indicate the charged residues. (a) AM1. (b) PM3.
Divide & Conquer Interaction Energy Decomposition

Docking of urate in Urate Oxidase (Theochem, 2009, 898, 31–41)

Active site (12Å box) ~ 700 atoms, full QM
Divide & Conquer Interaction Energy Decomposition

Docking of urate in Urate Oxidase (Theochem, 2009, 898, 31–41)

\[ Pol_{res} = \frac{1}{N_{res}} \sum_{a \in res} |q_{a}^{pol} - q_{a}^{vac}| \]

\[ CT_{res} = \frac{1}{N_{res}} \sum_{a \in res} (q_{a}^{E:S} - q_{a}^{pol}) \]
**What is SEBOMD?**

*SemiEmpirical Born-Oppenheimer Molecular Dynamics*

**Born-Oppenheimer** Only the positions of the nuclei are defined.

At each step of a molecular dynamics, molecular orbitals, energy, and gradient are obtained after a **fully converged SCF** calculation.

Ӏ Full quantum description of a molecular system.

**Semiempirical** NDDO approximation is performed (MNDO, AM1, PM3, . . .).

This enables fast and repetitive computations and allows for “long” MD trajectories (∼ 100ps) of “large” systems (several hundred of atoms) on commodity computers (single PC, clusters).
A SEBOMD program developed in Nancy

The software

The coupling of two programs:

- a classical MD program (AMBER\textsuperscript{1})
- + a semiempirical program (DivCon\textsuperscript{2})
- MD routines are modified: instead of computing energy and forces from a classical force field, the semiempirical program is filled with the system coordinates and outputs the energy and the gradient.

Why DivCon?

1. It can compute molecular energies in a linear scaling way (Divide & Conquer methodology)
2. It is parallelized (D&C calculations are spread among nodes)

\[ \text{access to “large” systems} \]

\textsuperscript{1}D. A. Case \textit{et al.}, AMBER9, Univ. of California (2006)
\textsuperscript{2}S. L. Dixon \textit{et al.}, DivCon 99, Penn State Univ. (1999)
Simulation protocol:

- box of 64 or 216 water molecules (H$_2$O or D$_2$O)
- cubic box ($a = 12.417\text{Å}$ or $18.626\text{Å}$)
- $dt = 1\text{fs}$
- NVT, 20 ps with strong temperature coupling (Andersen thermostat)
  
  \[
  0 \rightarrow 5\text{ps} = 10\text{K}, \quad 5 \rightarrow 10\text{ps} = 100\text{K}, \quad 10 \rightarrow 20\text{ps} = 300\text{K}
  \]
- sampling: 80 ps NVT at 300K
- comparison with experimental values:
  Radial Distribution Functions, auto-diffusion constant, heat of vaporization
- Various semiempirical methods
**Liquid Water: MNDO**

![Radial Distribution Functions (RDF)](image)

- 64 water molecules
- cubic box ($a = 12.417 \text{Å}$)
- 100 ps, NVT, 300K, dt=1 fs
- MNDO, SCF standard (full diagonalization)
LIQUID WATER: AM1

Radial Distribution Functions (RDF)

0 0.5 1 1.5 2 2.5 3
0 1 2 3 4 5 6 7 8
g(OO) r (Angstroms)

Radial Distribution Functions (RDF)

0 0.5 1 1.5 2 2.5 3
0 1 2 3 4 5 6 7 8
g(OH) r (Angstroms)

Radial Distribution Functions (RDF)

0 0.5 1 1.5 2 2.5 3
0 1 2 3 4 5 6 7 8
g(HH) r (Angstroms)

✓ 64 water molecules
✓ cubic box (a = 12.417 Å)
✓ 100ps, NVT, 300K, dt=1fs
✓ AM1, SCF standard (full diagonalization)

**Liquid Water: PM3**

![Radial Distribution Functions (RDF)](image1)

![Radial Distribution Functions (RDF)](image2)

- 64 water molecules
- cubic box \((a = 12.417\text{Å})\)
- 100ps, NVT, 300K, \(dt=1\text{fs}\)
- PM3, SCF standard
  (full diagonalization)
PM3-PIF (Parametrized Interaction Function)\(^6\)

- **Original PM3:**
  \[
  V_{AB} = Z'_A Z'_B (s_A s_A | s_B s_B) [1 + e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}}] + g_{AB}
  \]

- **PM3-PIF (Intermolecular interactions only):**
  \[
  g_{AB} = \frac{Z'_A Z'_B}{R_{AB}} \left( \sum_i K_{A,i} e^{-L_{A,i} (R_{AB} - M_{A,i})^2} + \sum_i K_{B,i} e^{-L_{B,i} (R_{AB} - M_{B,i})^2} \right)
  \]

- parameterized from ab initio potential energy surface (700 water dimer configurations, MP2/aug-cc-pVQZ)

LIQUID WATER: PM3-PIF

Radial Distribution Functions (RDF)

- 64 / 216 water molecules
- cubic box (12.417Å/ 18.626Å)
- 100ps, NVT, 300K, dt=1fs/2fs
- PM3-PIF, SCF standard or Divide & Conquer
- Hydrogen / Deuterium
A comparison between PM3-PIF, SPCE and TIP3P models

Radial Distribution Functions (RDF)

- Experimental (exp.)
- Minimum Image
- Ewald CM1
- TIP3P
- SPCE

$r$ (Angstroms) vs. $g(00)$
## Heat of Vaporization / Autodiffusion Constant

<table>
<thead>
<tr>
<th>Method</th>
<th>Heat of vaporization (kcal/mol)</th>
<th>Autodiffusion constant ($10^{-9} m^2 s^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>exp</td>
<td>10.5</td>
<td>2.3</td>
</tr>
<tr>
<td>TIP3P</td>
<td>10.27</td>
<td>5.0</td>
</tr>
<tr>
<td>SPCE</td>
<td>11.84</td>
<td>2.2</td>
</tr>
<tr>
<td>AM1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>7.72</td>
<td></td>
</tr>
<tr>
<td>AM1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>8.99</td>
<td></td>
</tr>
<tr>
<td>PM3&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4.11</td>
<td></td>
</tr>
<tr>
<td>PM3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.38</td>
<td></td>
</tr>
<tr>
<td>PM3-PIF&lt;sup&gt;a&lt;/sup&gt;</td>
<td>8.64</td>
<td></td>
</tr>
<tr>
<td>PM3-PIF&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10.12</td>
<td>3.2</td>
</tr>
<tr>
<td>PM3-PIF&lt;sup&gt;c&lt;/sup&gt;</td>
<td>8.71</td>
<td>4.7</td>
</tr>
<tr>
<td>PM3-PIF&lt;sup&gt;d&lt;/sup&gt;</td>
<td>9.98</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>: 64 water molecules, standard SCF, Minimum Image  
<sup>b</sup>: 64 water molecules, standard SCF, Ewald Summation, CM1 atomic charges  
<sup>c</sup>: 64 water molecules, standard SCF, Ewald Summation, Mulliken atomic charges  
<sup>d</sup>: 216 water molecules, D&C, Ewald Summation, CM1 atomic charges
## Molecular Geometry / CM1 Dipole moment

<table>
<thead>
<tr>
<th>Method</th>
<th>gas phase</th>
<th></th>
<th>liquid phase</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>d (Å)</td>
<td>α (°)</td>
<td>µ (D)</td>
<td>d (Å)</td>
</tr>
<tr>
<td>exp</td>
<td>0.957</td>
<td>104.5</td>
<td>1.85</td>
<td>0.970</td>
</tr>
<tr>
<td>TIP3P</td>
<td>0.957</td>
<td>104.5</td>
<td>2.35</td>
<td>0.957</td>
</tr>
<tr>
<td>SPCE</td>
<td>1.000</td>
<td>109.5</td>
<td>2.35</td>
<td>1.000</td>
</tr>
<tr>
<td>AM1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.961</td>
<td>103.5</td>
<td>2.02</td>
<td>0.967</td>
</tr>
<tr>
<td>AM1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.961</td>
<td>103.5</td>
<td>2.02</td>
<td>0.970</td>
</tr>
<tr>
<td>PM3&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.951</td>
<td>107.7</td>
<td>1.90</td>
<td>0.962</td>
</tr>
<tr>
<td>PM3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.951</td>
<td>107.7</td>
<td>1.90</td>
<td>0.966</td>
</tr>
<tr>
<td>PM3-PIF&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.951</td>
<td>107.7</td>
<td>1.90</td>
<td>0.966</td>
</tr>
<tr>
<td>PM3-PIF&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.951</td>
<td>107.7</td>
<td>1.90</td>
<td>0.972</td>
</tr>
<tr>
<td>PM3-PIF&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.951</td>
<td>107.7</td>
<td>1.90</td>
<td>0.966</td>
</tr>
<tr>
<td>PM3-PIF&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.951</td>
<td>107.7</td>
<td>1.90</td>
<td>0.971</td>
</tr>
</tbody>
</table>

<sup>a</sup>: 64 water molecules, standard SCF, Minimum Image  
<sup>b</sup>: 64 water molecules, standard SCF, Ewald Summation, CM1 atomic charges  
<sup>c</sup>: 64 water molecules, standard SCF, Ewald Summation, Mulliken atomic charges  
<sup>d</sup>: 216 water molecules, D&C, Ewald Summation, CM1 atomic charges
Polarization / Charge transfer (PM3-PIF)

Distribution of the molecular net charge

- **Mulliken**
- **CM1**
- **CM2**

Distribution of the molecular dipole moment

- **Mulliken**
- **CM1**
- **CM2**

Fluctuation of the net molecular charge of one water molecule (SEBOMD: 64 H₂O, NVT, 300K, dt=1fs, standard SCF, Ewald-CM1)
BEYOND LIQUID WATER …

- PM3-PIF provides a good model for liquid water
- What about small organic solute in water?
  - on-going tests of:
    - formamide,
    - ethanol,
    - dimethyl ether,
    - N-methylacetamide,
    - glycine (zwitterionic or “neutral”),
    - etc.

INTERMOLECULAR INTERACTIONS

- Water-Water interactions PM3-PIF
- Solute-Water interactions PM3 or PM3-PIF
DIMETHYL ETHER IN WATER (SOLUTE-WATER = PM3)

Radial Distribution Functions (RDF)

exp. (liquid water)
O(Water)-O(Water)
O(DME)-O(Water)

Radial Distribution Functions (RDF)

exp. (liquid water)
O(Water)-H(Water)
O(DME)-H(Water)
H(DME)-O(Water)

Radial Distribution Functions (RDF)

exp. (liquid water)
H(Water)-H(Water)
H(DME)-H(Water)

✓ 1 DME + 61 H$_2$O
✓ cubic box (12.417Å)
✓ 100ps, NVT, 300K, dt=1fs/2fs
✓ PM3-PIF, SCF standard
✓ DME ↔ Water = PM3
H-H intermolecular interactions in PM3

CH₄ ··· CH₄

PM3-MAIS
MP2
PM3

E_{int} kcal/mol
R (H-H) Å

PM3-PIF extensions to hydrated systems:

<table>
<thead>
<tr>
<th>Solute</th>
<th>H</th>
<th>C</th>
<th>N</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solvent</td>
<td>H</td>
<td>C</td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 DME + 61 H₂O
✓ cubic box (12.417Å)
✓ 100ps, NVT, 300K, dt=1fs/2fs
✓ PM3-PIF, SCF standard
✓ DME ↔ Water = PIF
Infra-Red spectra of N-methylacetamide in Water: a SEBOMD study

N-methyl acetamide (NMA) is a simple model for the amide linkage of peptides

Two conformers

IR signature of a peptide bond

Amide I: major contribution C-O stretch.

Amide II: major contribution HNC bend; C-N stretch.

Amide III: major contribution C-N stretch; HCN bend.
Infra-Red spectra of N-methylacetamide in Water: a SEBOMD study

Computational details (JCTC, 2011, 7, 1840–1849)

- 1 NMA + 64 water molecules in a cubix box ($\rho = 0.996$ g/ml)
- 500 ps+1 ns AMBER+SPC/E MD (NVT 300 K, $\Delta t = 1$ fs)
- 100+300 ps PM3-PIF SEBOMD (NVT 300 K, $\Delta t = 1$ fs)
- Ewald sum to treat long-range electrostatics (CM1 charges)
- IR spectra calculated from the Fourier Transform of the electric dipole time correlation function:

$$I(\omega) \sim Q(\omega) \int_{-\infty}^{+\infty} dt \exp^{-i\omega t} \langle \mu(0).\mu(t) \rangle$$

- NMA molecular dipole moment is computed from Mulliken, CM1, or CM2 atomic charges
  (with the NMA center of masses as the cartesian space origin)
Infra-Red spectra of N-methylacetamide in Water: a SEBOMD study

AMBER classical MD

Isolated NMA

NMA in water
Infra-Red spectra of N-methylacetamide in Water: a SEBOMD study

PM3-PIF SEBOMD

Isolated NMA

NMA in water

Mulliken

CM1

CM2

cis NMA
trans NMA

ω (cm\(^{-1}\))

Mulliken

CM1

CM2

cis NMA
trans NMA

ω (cm\(^{-1}\))
Infra-Red spectra of N-methylacetamide in Water: a SEBOMD study

Comparison of the IR pics

Table 3. Collection of Results for the Infrared Frequencies Which Are Characteristic of the Peptide Bond of NMA in the Gas Phase and in Water: Amide I (AI), Amide II (AII), Amide III (AIII), and N–H Stretch ($\delta_{\text{NH}}$)$^a$

<table>
<thead>
<tr>
<th>mode</th>
<th>Amber03</th>
<th>SEBOMD</th>
<th>experiment$^{18,78}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GP</td>
<td>sol.</td>
<td>$\Delta$</td>
</tr>
<tr>
<td>AI</td>
<td>1680$^b$$^{1800}^c$</td>
<td>1680$^b$$^{1800}^c$</td>
<td>0</td>
</tr>
<tr>
<td>AII</td>
<td>1780$^b$$^{1800}^c$</td>
<td>1780$^b$$^{1800}^c$</td>
<td>0</td>
</tr>
<tr>
<td>AIII</td>
<td>1250$^b$$^{1320}^c$</td>
<td>1250$^b$$^{1330}^c$</td>
<td>5</td>
</tr>
<tr>
<td>$\delta_{\text{NH}}$</td>
<td>3380$^b$$^{3370}^c$</td>
<td>3380$^b$$^{3370}^c$</td>
<td>0</td>
</tr>
</tbody>
</table>

$^a$ All data in the gas phase (GP), in water (sol.), and the average shift of the condensed phase results with respect to the gas phase ($\Delta$) are reported in cm$^{-1}$. $^b$ cis-NMA. $^c$ trans-NMA. $^d$ In CCl$_4$ solution.

What we learn from SEBOMD studies

- SemiEmpirical methods can tackle very large molecular systems in static or dynamical simulations
- SemiEmpirical methods are fast enough to run molecular dynamics simulations (SEBOMD) on commodity computers or small clusters
- Results are mainly dependent on the parameters of the “semiempirical force field” which naturally includes intermolecular polarization and charge transfer
- Now, there is a need for an improvement of the semiempirical parameters that would correctly take into account intermolecular interactions

“SemiAbInitio” parameters: the lack of empirical data to represent intermolecular interactions imposes the fitting of semiempirical parameters from ab initio data
Localized Molecular Orbitals
Semiempirical foundations (Stewart, 1996)\(^1\)

- \(P\) is sparse, thus MOs have a localized extension i.e. MOs are developed on a limited set of atoms
- For SCF convergence, \(F\) does not need to be fully diagonalized;
- It is only necessary to block diagonalize \(F\) between occupied and non-occupied MOs

Algorithm

- Starting MO set: a set of strictly localized MOs (LMOs)
- For all non-orthogonal (occupied MO, non-occupied MO) tuples that share at least one atom, combine them to form two new orthogonal MOs
- Repeat until \(F\) is block diagonal

---

\(^1\)Stewart, J. J. P. *Int. J. Quant. Chem.* 1996, 58, 133–146
Localized Molecular Orbitals

Starting set of LMOs

- They are built from a hybrid orbital basis set
- To each bonding LMO corresponds an orthogonal antibonding LMO
- They are usually built from a Lewis structure

LMO Orthogonalization

- Performed through the annihilation of $F$ elements connecting occupied and unoccupied MOs sharing at least one atom in common (pseudo-diagonalization)
- Semiempirical fact: if an occupied LMO does not extend to any atom of an unoccupied LMO, thus they are orthogonal ($S = I$)
Localized Molecular Orbitals

MO energy:

$$\varepsilon_{ij} = \langle \psi_i | F | \psi_j \rangle = \sum_{\mu} \sum_{\nu} c_{\mu i} F_{\mu \nu} c_{\nu j}$$

with:

$$\psi_i = \sum_{\mu} c_{\mu i} \varphi_{\mu} \quad \text{LCAO}$$

Pseudo-diagonalization step:

$$\begin{align*}
\psi'_i &= \alpha \psi_i + \beta \psi_j \\
\psi'_j &= -\beta \psi_i + \alpha \psi_j
\end{align*}$$

$$\alpha = \sqrt{\frac{1}{2} \left( 1 + \frac{D}{\sqrt{4\varepsilon_{ij}^2 + D^2}} \right)} \quad \text{with} \quad D = \varepsilon_{jj} - \varepsilon_{ii}$$

$$\beta = \phi \sqrt{1 - \alpha^2} \quad \text{with} \quad \phi = \begin{cases} 1 & \text{if } \varepsilon_{ij} < 0 \\ -1 & \text{unless} \end{cases}$$

Procedure stops when all occupied LMOs are two-by-two orthogonal to all unoccupied LMOs. Then, $\mathbf{P}$ can be build.
Localized Molecular Orbitals

Pros and cons

Pros

- Avoid diagonalization
- Fast computations (small prefactor)
- Can handle very large molecular systems

Cons

- Tricks not published
- No extension to ab initio or DFT theories
- Parallelization?
- Linear Scaling?

available as MOZYME in MOPAC
Localized Molecular Orbitals

Is LMO a linear scaling methods?

<table>
<thead>
<tr>
<th>Number of Residues</th>
<th>Number of Atoms</th>
<th>$\Delta \Delta H_f^a$ (kcal/mol)</th>
<th>Time for one SCF (s)</th>
<th>Ratio of times MOPAC / MOZYME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MOZYME - MOPAC</td>
<td>MOPAC 93</td>
<td>MOZYME</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>0.00</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>-0.01</td>
<td>3.5</td>
<td>8.2</td>
</tr>
<tr>
<td>4</td>
<td>62</td>
<td>+0.02</td>
<td>27.8</td>
<td>24.3</td>
</tr>
<tr>
<td>8</td>
<td>114</td>
<td>-0.07</td>
<td>189.6</td>
<td>118.8</td>
</tr>
<tr>
<td>16</td>
<td>255</td>
<td>+0.32</td>
<td>2662.8</td>
<td>541.7</td>
</tr>
<tr>
<td>32</td>
<td>481</td>
<td></td>
<td>(17871.2)</td>
<td>1244.7</td>
</tr>
<tr>
<td>64</td>
<td>992</td>
<td></td>
<td>(156766.4)</td>
<td>5489.0</td>
</tr>
<tr>
<td>128</td>
<td>1995</td>
<td></td>
<td>(1275107.2)</td>
<td>17592.2</td>
</tr>
<tr>
<td>256</td>
<td>3924</td>
<td></td>
<td>(9702977.0)</td>
<td>63321.2</td>
</tr>
<tr>
<td>265</td>
<td>4046</td>
<td></td>
<td>(10636424.0)</td>
<td>65266.0</td>
</tr>
</tbody>
</table>

$^a$Difference in heats of formation calculated using MOPAC and MOZYME. Numbers in parentheses are estimated based on the assumption that the time required for a MOPAC run increases as $N^3$.

Localized Molecular Orbitals

Is LMO a linear scaling methods?

Localized Molecular Orbitals

Some LMOs applications

- Calculation of the geometry of a small protein using semiempirical methods

- A practical method for modeling solid using semiempirical methods

- Application of the PM6 method to modeling proteins

- Application of the PM6 method to modeling the solid state
Localized Molecular Orbitals

Another method using LMOs in a linear scaling fashion: LocalSCF

- LocalSCF method for semiempirical quantum-chemical calculation of ultralarge biomolecule

- Validation of linear scaling semiempirical LocalSCF method

- Charge Transfer Effects in the GroEL-GroES Chaperonin Tetramer in Solution
Density Matrix Minimization

Mc Weeny, 1962²

- Eigensolution of the SCF equations can be avoided by direct minimization of $E_{elec}$ subject to the constraint of normalization:

$$N_e = \text{Tr}\{P\}$$

and idempotency

$$P = PP$$

- Mc Weeny purification to obtain idempotency:

$$\tilde{P} = 3PP - 2PPP$$

---

²McWeeny, R. Rev. Mod. Phys. 1960, 32, 335–369
Density Matrix Minimization

Li, Nunes, Vanderbilt, 1993

- \( \mathbf{P} \) is the quantity minimizing:

\[
\Omega(\mathbf{P}) = \text{Tr}\{(3\mathbf{PP} - 2\mathbf{PPP})\mathbf{F}\} + \mu(\text{Tr}\{\mathbf{P}\} - N_{el})
\]

(in the orthogonal basis set)

- \( \mu \): Lagrange constraint to enforce the correct number of electrons

- To achieve \( O(N) \) complexity: sparse matrix algebra + gradient-only minimization

---


Density Matrix Minimization

Conjugate-Gradient Density Matrix Search (CG-DMS) \(^5\)

1. initial search direction (steepest descent step):

\[
G_0 = H_0 = -\nabla \Omega(P_0)
\]

2. conjugate gradient search:

\[
P_{i+1} = P_i + \lambda_i H_i \quad \text{the step}
\]
\[
G_i = -\nabla_{P_i} \Omega(P_0) \quad \text{the gradient}
\]
\[
H_i = G_{i+1} + \gamma_i H_i \quad \text{the search direction}
\]
\[
\gamma_i = \frac{(G_{i+1} - G_i) \cdot G_{i+1}}{G_i \cdot G_i}
\]

Density Matrix Minimization

Conjugate-Gradient Density Matrix Search (CG-DMS)\textsuperscript{6}

\[ \Omega(P) = \text{Tr}\{(3PP - 2PPP)F\} + \mu(\text{Tr}\{P\} - N_{\text{el}}) \]
\[ \nabla \Omega(P) = 3(PF + FP) - 2(PPF + PFP + FPP) - \mu I \]

\( \mu \) is chosen such that the gradient of the functional \( \Omega(P) \) is traceless:

\[ \mu = \text{Tr}\{3(PF + FP) - 2(PPF + PFP + FPP)\} / N_{\text{basis}} \]

Then, any finite step preserves the number of electrons\textsuperscript{7}.


Density Matrix Minimization


FIG. 4. Cumulative CPU timings for CG-DMS and diagonalization for polyglycine chains using 3-21 G/LDA. (Last four points in the diagonalization curve are extrapolated using $N^3$ scaling.)

<table>
<thead>
<tr>
<th>Basis set</th>
<th>Diagonalization</th>
<th>CG-DMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SCF cycles</td>
<td>CPU time (h)</td>
</tr>
<tr>
<td>STO-3G</td>
<td>11</td>
<td>0.05</td>
</tr>
<tr>
<td>3-21G</td>
<td>10</td>
<td>0.4</td>
</tr>
<tr>
<td>6-31G</td>
<td>11</td>
<td>0.5</td>
</tr>
<tr>
<td>6-31G**</td>
<td>10</td>
<td>4.9</td>
</tr>
</tbody>
</table>
Basis set affects computational cost significantly. For a given molecule, if we change the basis from 3-21G to 6-31G**, the number of basis functions increases substantially, and so does the CPU time. However, for a given number of basis functions, a 6-31G** calculation on a given system is more expensive than a 3-21G calculation with identical number of basis functions (and consequently on a larger similar molecule) because polarized basis sets normally yield much less sparse matrices than unpolarized basis sets.
Density Matrix Minimization


1D is different than 3D!

FIG. 4. Total CPU time for single point energy calculations on polyglycine chains demonstrating the crossover point for various neglect thresholds. Each calculation took 10 SCF iterations with 4 CG cycles per SCF iteration. Two purification transformations were applied after each conjugate gradient cycle. Except for the curve that explicitly says so, no soft diagonal cutoffs were used.

FIG. 5. Total CPU time for single point energy calculations on water clusters demonstrating the crossover point for the $5 \times 10^{-3}$ a.u. neglect threshold with no diagonal cutoff. Each calculation took 8 SCF iterations with 4 CG cycles per SCF iteration. Two purification transformations were applied after each conjugate gradient cycle.
Density Matrix Minimization

Pros and cons

Pros

- Scales linearly if sparse matrix algebra is well implemented
- Easy to implement
- Parallelism is possible (matrix multiplication only)

Cons

- Not enough to ensure linear scaling for ab initio methods (other limiting factors like integral computations, etc)
- Threshold in the matrix multiplication (affects the prefactor)
- Efficient for small basis sets
Conclusions on Linear Scaling Methods

As of today

- Standard algorithms are being modified to allow $O(N)$
- $O(N)$ publications: semiempirical, DFT, HF, MPx, CCSD(T), etc
- The smaller basis set, the smaller prefactor
- Computer codes: fine Chemical answer : ?

Some Available codes

- semiempirical: MOPAC(Mozyme), LocalSCF, DivCon(AMBER)
- DFT: Siesta, ADF
- ab initio: Gaussian, MondoSCF (Freeon), MolPro
Linear Scaling QM methods

Some selected reviews

- Linear scaling electronic structure methods

- Order-N methodologies and their applications

- Linear-Scaling Methods in Quantum Chemistry

- \( O(N) \) methods in electronic structure calculations

- Linear-scaling self-consistent field methods for large molecules
QM/MM methods, Linear Scaling methods and Parallelism

Bottlenecks for parallelization

- bielectronic integrals
- matrix multiplication
- matrix diagonalization

Linear scaling methods employ scheme that can be better parallelized, thus reducing the prefactor.

GPUs in quantum chemistry (ab initio and semiempirical)


What about linear scaling + GPU?