Large Scale Electronic Structure Calculations with CP2K

Jürg Hutter
University of Zurich

6. October, 2007 / CSCS User day
Overview

CP2K: Ab initio Molecular Dynamics Simulations
   Introduction
   Examples

Towards Linear Scaling
   Conjugate Gradient Optimizer
   L1-Norm Localization

HF/Exact Exchange
   Implementation in CP2K
CP2K Program System

- GNU General Public License
- Community Developers Platform on "Berlios"(cp2k.berlios.de)
- Open CVS for download
- Written in Fortran95, 470’000 lines of code
- MPI and OpenMP parallelization
- Quality control: automatic regression and memory leak testing
- Force Methods: Kohn–Sham DFT, Classical Force Fields, QM/MM
- Sampling Methods: Molecular Dynamics, Meta-Dynamics, Monte Carlo
CP2K/Quickstep

- Gaussian basis sets
- Plane waves auxiliary basis for Coulomb integrals
- Sparse matrices, efficient screening linear scaling KS matrix computation
- Fast/robust direct wavefunction optimizer (OT)
  Memory scaling: $MN$
  CPU time scaling: $MN^2$
  $M$ number of basis functions, $N$: number of states
- For given number of electrons, $O(N)$ in basis set size
- All-electron calculations with projector augmented-waves
Computational Tasks

- Calculation of matrices $H, S$
- Calculation of forces
  - Linear scaling through screening
  - Fast Poisson solver with FFT
  - Problems with HF/exact exchange

- Minimization of energy functional
- Calculation of orbitals $C$
  - Orbital transformation method: fast and robust
  - Problems with efficient preconditioners (memory, time)
  - Problems with very large systems (cubic scaling)
Ru-dye in Acetonitrile Solution
Box: \((21.43 \text{ Å})^3\), \(\approx 620\) atoms, \(\approx 6000\) bsf, \(\approx 850\) occupied orbitals, 1 min/ MD step / 128 CPUs
TiO₂ / Acetonitrile Interface
101 anatase surface, slab calculation, system size: 22.6 × 10.0 × 40 Å
(TiO₂)₇₂ (NCCH₃)₆₈, 624 atoms, 5632 electrons, 10260 bsf
Rubredoxin in Water Solution
Solvated Rubredoxin (FeS4); \( \approx 2800 \) atoms; \( \approx 55000 \) bsf
117 s / SCF on 1024 CPUs (XT3), 80% parallel efficiency
Simplification of minimization method
Transform constrained minimization problem

$$C^* = \arg \min_C \left\{ E[C] \mid C^T SC = 1 \right\},$$

into an unconstrained minimization of

$$C^* = \arg \min_C \left\{ E[f(C)] \right\},$$

where

$$E[f(C)] = E[f_n(C)] + O(E[\delta C^n]).$$

with $C = C_0 + \delta C$ and $C_0^T SC_0 = 1$. 
Orbital Refinement Function

\[ z = z_0 + \delta z \text{ and } z_0^T S z_0 = 1 \]

\[ f_2(z) = \frac{1}{2} z (3 - Y), \]
\[ f_3(z) = \frac{1}{8} z \left( 15 - 10 Y + 3 Y^2 \right), \]
\[ f_4(z) = \frac{1}{16} z \left( 35 - 35 Y + 21 Y^2 - 5 Y^3 \right), \text{ and} \]
\[ f_5(z) = \frac{1}{128} z \left( 315 - 420 Y + 378 Y^2 - 180 Y^3 + 35 Y^4 \right), \]

\[ f_n^T(z) S f_n(z) - 1 = O(\delta z^n). \]
Optimization

- Nonlinear conjugate gradients with line search
- Includes only matrix additions and multiplications
- Simplifies parallelization and use of sparsity
Example: Liquid Water

<table>
<thead>
<tr>
<th></th>
<th>$(\text{H}<em>2\text{O})</em>{512}$</th>
<th>$(\text{H}<em>2\text{O})</em>{1024}$</th>
<th>$(\text{H}<em>2\text{O})</em>{2048}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>basis functions</td>
<td>11776</td>
<td>23552</td>
<td>47104</td>
</tr>
<tr>
<td>CPUs (XT4)</td>
<td>32</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>$n_{\text{iter SCF}}$</td>
<td>27</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>$f_{\text{ks}}$</td>
<td>0.18</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>$f_{\text{mini}}$</td>
<td>0.49</td>
<td>0.66</td>
<td>0.76</td>
</tr>
<tr>
<td>$t_{\text{tot}}$ [min]</td>
<td>8.9</td>
<td>23.3</td>
<td>132.7</td>
</tr>
</tbody>
</table>
Example: DNA Decamer

- **System**
  - DNA decamer: \((\text{C}_{196}\text{H}_{230}\text{N}_{74}\text{O}_{116}\text{P}_{18})\)
  - Water box: \(45.4 \times 46.1 \times 61.2 \text{ Å}^3\) with 4003 water molecules, 18 potassium cations
  - Total number of atoms: **12661**

- **Basis functions**: **103333**

- **Optimization**
  - Cray XT3 with 1024 processors
  - 159 SCF iterations (no preconditioner - memory!)
  - 148 s + 255 s (energy + gradient) / SCF
Sparse Matrices

- Operator matrices are rather sparse

HIV-1 Protease-DMP323 complex in solution (3200 atoms)

- Orbital matrices are invariant under unitary transformation
  Chemical localization: Boys, Edmiston-Rudenberg, etc.
  Mathematical localization
L1-Norm

- Smooth sparseness function based on L1 norm

\[ \phi(y) = \begin{cases} 
|y| - \varepsilon_1/2 & |y| \geq \varepsilon_1 \\
|y|^2/2\varepsilon_1 & |y| < \varepsilon_1,
\end{cases} \]

- Parameterization of orbitals: \( C(U) = CU \),
  
  \( U \) is a \( M \times M \) unitary matrix
  
  \( U(X) = \exp X \) where \( X \) is an antisymmetric matrix.

- Optimization of

\[ U^* = \arg \max_X \phi(U(X)). \]
Example: Water
Combine Optimization and Localization
Molecular dynamics, 1024 water molecules

Open question: How does sparseness threshold affect MD stability?
Hartree-Fock/Exact Exchange

- Efficency of Quickstep is based on algorithms for local operators
  Only local density functionals (LDA, GGA)
- Functionals with Hartree-Fock exchange (non-local) work often better

\[
E_{\text{HFX}} = \sum_{\alpha\gamma} P_{\alpha\gamma} \sum_{\beta\delta} P_{\beta\delta} \int \int \frac{\phi_{\alpha}(r)\phi_{\beta}(r)\phi_{\gamma}(r')\phi_{\delta}(r')}{|r - r'|} \, drdr'
\]

- Implementation through two-electron integrals (analytic)
  Standard in all quantum chemistry programs
Parallelization

- Distribute integrals \((\alpha\beta \mid \gamma\delta)\)
- Store integrals (incore algorithm)
- Full use of permutation symmetry \(\alpha \leftrightarrow \beta, (\alpha\beta) \leftrightarrow (\gamma\delta)\)
- Replicate matrices \(H, P\)

<table>
<thead>
<tr>
<th>液态水，64个水分子</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Basis</td>
<td>6-31G**</td>
<td>DZVP</td>
</tr>
<tr>
<td>CPUs</td>
<td>PBE</td>
<td>HSE</td>
</tr>
<tr>
<td>4</td>
<td>124</td>
<td>2276</td>
</tr>
<tr>
<td>8</td>
<td>83</td>
<td>689</td>
</tr>
<tr>
<td>16</td>
<td>59</td>
<td>388</td>
</tr>
<tr>
<td>32</td>
<td>46</td>
<td>222</td>
</tr>
<tr>
<td>64</td>
<td>43</td>
<td>157</td>
</tr>
<tr>
<td>128</td>
<td>126</td>
<td>533</td>
</tr>
<tr>
<td></td>
<td>PBE</td>
<td>HSE</td>
</tr>
<tr>
<td>4</td>
<td>105</td>
<td>N/A</td>
</tr>
<tr>
<td>8</td>
<td>71</td>
<td>&gt;20000</td>
</tr>
<tr>
<td>16</td>
<td>46</td>
<td>11070</td>
</tr>
<tr>
<td>32</td>
<td>33</td>
<td>1881</td>
</tr>
<tr>
<td>64</td>
<td>30</td>
<td>924</td>
</tr>
<tr>
<td>128</td>
<td></td>
<td>533</td>
</tr>
</tbody>
</table>
Example: Liquid Water

64 waters periodic boundary conditions (zeroth shell),
HSE functional (omega = 0.15),
Max Memory per CPU 600 MB

Basis DZVP; Total Number of integrals to be stored: 2449434176

<table>
<thead>
<tr>
<th>CPUS</th>
<th>1.step[s]</th>
<th>2.step[s]</th>
<th>mem/cpu[MB]</th>
<th>forces[s]</th>
<th>energy[s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6935</td>
<td>6200</td>
<td>600</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td>3600</td>
<td>3000</td>
<td>600</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>8</td>
<td>2018</td>
<td>1450</td>
<td>600</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>16</td>
<td>1246</td>
<td>625</td>
<td>600</td>
<td>1707</td>
<td>9300</td>
</tr>
<tr>
<td>32</td>
<td>522</td>
<td>40</td>
<td>585</td>
<td>811</td>
<td>935</td>
</tr>
<tr>
<td>64</td>
<td>272</td>
<td>18</td>
<td>294</td>
<td>411</td>
<td>480</td>
</tr>
<tr>
<td>128</td>
<td>154</td>
<td>11</td>
<td>147</td>
<td>221</td>
<td>279</td>
</tr>
</tbody>
</table>
Current Developments and Problems

- Better memory usage through compressed integral storage (compression factor 4 - 6)
- Optimized load balancing using Monte Carlo techniques
  Minimize number of atoms touched by each CPU
  Partial distribution of matrices
- Fortran C-binding (currently only available in g95 compiler)
- Alternative methods using grids
Summary

- Ab initio MD simulations with up to 1000 atoms possible with CP2K
- New optimization and localization techniques for larger systems
- HF exchange (hybrid functionals): efficient implementation but still a challenge
Acknowledgment

- Joost VandeVondele
- Valery Weber
- Teodoro Laino
- Fanny Masson
- Florian Schiffmann
- Manuel Guidon