CP2K: Past, Present, Future

Jürg Hutter
Department of Chemistry, University of Zurich
Outline

• Past
  • History of CP2K
  • Development of features

• Present
  • Quickstep DFT code
  • Post-HF methods (RPA, MP2)
  • Libraries

• Future
  • Algorithms for KS-DFT
  • Post-HF methods
  • k-points
CP2K source repository goes online on berlios.de
Now on sourceforge.net
13 years of open development
Origins

- **Quickstep** DFT Code, Max-Planck-Institute, Stuttgart
  
  Gerald Lippert, Matthias Krack, JH

- **Fist** MD Code, UPenn, Philadelphia

  Chris Mundy, S. Balasubramanian, Ken Bagchi
CP2K SOURCE CODE DEVELOPMENT

Hartree-Fock

200 lines/day

16'000 commits
Recent Versions

90 lines/day and 2.6 commits/day

<table>
<thead>
<tr>
<th>Date</th>
<th>Version</th>
<th>Source lines</th>
<th>Commit</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.2011</td>
<td>2.2</td>
<td>891’928</td>
<td>11883</td>
</tr>
<tr>
<td>09.2012</td>
<td>2.3</td>
<td>919’583</td>
<td>12358</td>
</tr>
<tr>
<td>06.2013</td>
<td>2.4</td>
<td>968’038</td>
<td>12977</td>
</tr>
<tr>
<td>02.2014</td>
<td>2.5</td>
<td>947’762</td>
<td>13637</td>
</tr>
<tr>
<td>12.2014</td>
<td>2.6</td>
<td>990’282</td>
<td>14881</td>
</tr>
</tbody>
</table>
## CP2K: Application Fields

<table>
<thead>
<tr>
<th></th>
<th>Chemistry</th>
<th>Materials Science</th>
<th>Physics</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP2K</td>
<td>70%</td>
<td>31%</td>
<td>41%</td>
</tr>
<tr>
<td>Quantum-Espresso</td>
<td>35%</td>
<td>57%</td>
<td>69%</td>
</tr>
<tr>
<td>VASP</td>
<td>36%</td>
<td>50%</td>
<td>65%</td>
</tr>
</tbody>
</table>

¹scopus.com
CP2K Main Modules

- Kohn–Sham DFT
  GPW, GAPW, {R,LR}-TDDFT, EPR, NMR, NQR, XAS, IR

- Atomic DFT code
  Optimize pseudopotentials, basis sets

- Semi–empirical and Tight-binding
  MNDO, AM1, PM6, DFTB

- Classical Potentials
  CHARMM/AMBER-ff, EAM, COS, polarizable FF
CP2K Main Modules

- **Molecular Dynamics**
  NVE, NPT, GLE thermostats, Nose-Hoover thermostats, Ehrenfest dynamics

- **Meta-Dynamics, String Methods**
  multiple walkers, NEB

- **Monte Carlo**
  NVT, NPT, GEMC, TMC
and few other things CP2K can do

- QM/MM
  DFT, semi-empirical, ..., fully periodic, ...

- Multiple force evaluation
  Define your personal energy function

- General non-bonded pair interaction
  Any analytic potential possible

- Powell optimizer
  Optimize (almost) any variable defined in your input file

- Farming
  Run as many inputs as you want in a single job
CP2K: Features

2002  OT optimizer
2004  GAPW, TD-DFT
2006  QM/MM, XAS
2008  Ehrenfest dynamics, Hartree-Fock
2010  ADMM, Metals, NMR
2012  Linear scaling KS
2014  RI-MP2, RI-RPA, libxc, nl-vdW
Quickstep DFT Code

- Water benchmarks and scaling
- Metals
- MP2/RPA
- Internal/external libraries
Water Benchmarks

[Graph showing the relationship between the number of cores and the time per MD step for different water benchmarks: H2O-32, H2O-64, H2O-128, H2O-256, H2O-512, H2O-1024, and H2O-2048. Two sets of data are compared: XT3 Stage 0 (2005) and XC30 ARCHER (2013).]
System Size Scaling

- 512 water: $67\%$ LinAlg
- 256 water: $58\%$ LinAlg
- 128 water: $43\%$ LinAlg
- 64 water: $25\%$ LinAlg
- 32 water: $14\%$ LinAlg

Time MD Step [s] vs Number of Cores

- 64 cores: $X 2.3$
- 128 cores: $X 2.9$
- 256 cores: $X 3.4$
- 512 cores: $X 4.2$
Large Systems

Dominated by linear algebra

Key Operation : Sparse Matrix Multiplication

=DBCSR Library

Distributed Block Compressed Sparse Row Format
Water on $h$-BN

833 water molecules on $h$-BN layer on metal surface. QM/MM using optimized force field (QM: water, MM: $h$-BN and metal)
\[ \Delta \cos \Theta \approx 0.04 \quad \rightarrow \quad \text{smaller contact angle for NM} \]

D. Golze et al, PCCP in press
$h$-BN/Rh(111): Structure

Corrugated Monolayer Model
3.2 nm periodic structure


R. Lakowski et al. PRL **98** 106802 (2007)
Methods

- **Gaussian and Plane Waves (GPW)**

- **Basis sets: DZVP MOLOPT type**
  500 Ry PW cutoff (electron density)

- **Dual-space pseudopotentials: Rh [17/9 e], Ru [16/8 e]**

- **Revised PBE + D2/3 vdW correction**
  Y. Zhang, W. Yang, PRL 80, 890 (1998)
  S. Grimme et al. JCP 132, 154104 (2010).

- **Fermi-Dirac smearing with T=300 K**
Computational Models

Slab models, 3d periodic, 20 Å empty space

- \textit{h-BN/Rh}(111) (small)
  4 layer Rh + 1 BN: 914 atoms, 19370 BSF, 11144 el

- \textit{h-BN/Rh}(111) (large)
  7 layer Rh + 2 BN: 1684 atoms, 34996 BSF, 19840 el

- Multiple cells (2x2)
  4 layer Rh/Cu + 1 BN/gr: 3656 atoms, 77480 BSF, 44576 el
DFT Optimized Structure

25 on 23 reconstruction
Electronic corrugation in rotated $h$-BN on Cu(111)

**Fig. 1** Registry of the N and B atoms in the corresponding sublattices relative to the metal substrate. The top, hcp and fcc registries are shown as the red, green and blue components of the point colors, respectively. Mixed colors indicate intermediate and bridging positions. The black line indicates the (11) diagonal of the monolayer surface from the top-right N atom, used for later evaluation (cf. Fig. 5). The frame shows the unit cell of the metal substrate for reference. Axis ticks at intervals of 10 Å.

**Fig. 2** Simulated STM using Tersoff–Hamann approximation at a bias of +1.0 V. Color bar is in units of Å relative to the topmost Cu layer, the $x$ and $y$ axis ticks at intervals of 10 Å. Black lines indicate the unit cell.

**Fig. 3** Lateral map of the electrostatic potential relative to the Fermi energy (eV) at constant height, approx. 3.4 Å above the hBN layer. The area is the same as the unit cell drawn in Fig. 2.
Figure 2. Speedup (a) and efficiency (b) with respect to 64 nodes for the calculation of the RI-MP2 energy gradients and stress of 64 bulk water molecules (cc-TZVP basis). Calculation performed on a CRAY-XC30 machine, each node consists of 8 processes.
Solid NH₃ and CO₂

Figure 5. Location of the minima for NH₃ (a) and CO₂ (b), computed at the RI-MP2 level of theory with different basis sets obtained with different approaches. The lattice parameter optimization curves have been fitted with a third order Birch-Murnaghan equation, the crosses represent the location of the minimum point for each curve. CP means that the cohesive energy have been counterpoise corrected.
## Timings

<table>
<thead>
<tr>
<th>o</th>
<th>n</th>
<th>$N_a$</th>
<th>$t_{\text{tot}}$</th>
<th>$t_D$</th>
<th>$\frac{t_D}{t_E}$</th>
<th>$t_{\text{GPU}}$</th>
<th>$\frac{t_D}{t_{\text{GPU}}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH$_3$</td>
<td>128</td>
<td>2272</td>
<td>5312</td>
<td>3.15</td>
<td>1.53</td>
<td>4.20</td>
<td>1.47</td>
</tr>
<tr>
<td>U</td>
<td>192</td>
<td>2752</td>
<td>6784</td>
<td>5.97</td>
<td>3.58</td>
<td>4.59</td>
<td>2.89</td>
</tr>
<tr>
<td>FA</td>
<td>216</td>
<td>2760</td>
<td>6912</td>
<td>5.83</td>
<td>3.87</td>
<td>4.28</td>
<td>2.95</td>
</tr>
<tr>
<td>D</td>
<td>192</td>
<td>2992</td>
<td>7520</td>
<td>12.84</td>
<td>5.27</td>
<td>5.15</td>
<td>4.26</td>
</tr>
<tr>
<td>CO$_2$</td>
<td>256</td>
<td>2784</td>
<td>7296</td>
<td>7.94</td>
<td>4.99</td>
<td>4.15</td>
<td>3.50</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>256</td>
<td>3648</td>
<td>8704</td>
<td>10.17</td>
<td>9.34</td>
<td>4.00</td>
<td>5.85</td>
</tr>
<tr>
<td>B</td>
<td>240</td>
<td>4128</td>
<td>10176</td>
<td>23.01</td>
<td>13.77</td>
<td>4.45</td>
<td>8.81</td>
</tr>
<tr>
<td>PD</td>
<td>312</td>
<td>3936</td>
<td>10208</td>
<td>28.96</td>
<td>17.48</td>
<td>4.13</td>
<td>9.80</td>
</tr>
<tr>
<td>SA</td>
<td>304</td>
<td>4144</td>
<td>10432</td>
<td>27.00</td>
<td>19.29</td>
<td>4.26</td>
<td>10.94</td>
</tr>
<tr>
<td>CT</td>
<td>336</td>
<td>4152</td>
<td>10560</td>
<td>29.71</td>
<td>22.30</td>
<td>4.16</td>
<td>11.97</td>
</tr>
</tbody>
</table>
Libraries

- DBCSR: Sparse matrix multiplication
- libxc: XC functionals
- libint: Two-electron repulsion integrals
- Eigensolvers
DBCSR

- Distributed Block CSR format
- Key library
- Backend for multi-core, GPU, MIC
- OT: large systems
- Linear scaling code
Software Architecture
Figure 9: Performance comparison of the multi-threaded DBCSR library based on 23x23 matrix blocks, and was not using the MPI capabilities. The benchmark was run on a dual Sandy Bridge (E5-2620, 2.0GHz, 6 cores) machine, equipped with one NVIDIA Tesla K20 card.
Limitations

Communication scales with $\sqrt{(P)}$
libxc

- XC functionals library from

- Interface to CP2K (V2.03)

- Problem: needs code intervention when library changes

- Problem: new Fortran interface (V2.2.0)
libint

- ERI over Gaussian functions from https://sourceforge.net/projects/libint/
- Interface to CP2K (V1.1.4)
- Needed for HFX and ADMM
- New version 2.0 not compatible https://github.com/evaleev/libint
Eigen solver

- Essential for calculations of metals
- Interface in CP2K to ScaLapack
- ELPA: Improved performance and MPI scalability
  http://elpa.rzg.mpg.de/
- Raffaele Solca (ETHZ): new solver using OpenACC
  Multi-thread and GPU acceleration
  not yet fully available
Future: Current CP2K Developments

- GPW/GAPW Quickstep code
- Post-HF code
- k-points
Quickstep Algorithms

- LRI (local density fitting)

\[
\rho(r) = \sum_{AB} \rho_{AB} \approx \sum_{AB} \left[ \sum_u f^{AB}_u \chi^A_u(r) + \sum_v f^{AB}_v \chi^B_v(r) \right]
\]

- Two step collocation of density

\[
\rho(r) \rightarrow \sum_A \rho_A(r) \rightarrow \rho(R)
\]

- Much faster collocation, better parallelization
target are small to medium systems (100-500 atoms)

- Can we control accuracy?
ADMM

- Additional projection methods
  ADMMS, ADMMP, ADMMQ

- Additional functionals for error correction
  OPTX
Post-HF

- UMP2 gradients/stress
- RPA gradients/stress
- Extensions to dRPA
- Quasi-particle energies from RPA/MP2 (G0W0)
k-points

\[ F^k C^k = S^k C^k E^k \]

where the operator matrices \((F^k, S^k)\) are calculated in real space and then Fourier transformed.

\[ A^k = \sum_g A^0 g e^{i k \cdot g} \]

the density matrix is calculated in Fourier space from \(C^k\) and then transformed to real space

\[ P^0 g = \frac{1}{\Omega} \int P^k e^{i k \cdot g} dk \]
Status of implementation

- Input and setup
- Data structures in real space (DBCSR)
  Generalization of matrices to include \( g \) index
- Data structures in \( k \) space (full matrices)
- General complex eigensolver in subgroups (MPI groups)
- Parallelization over \( k \)-points
- Fourier transforms and redistribution between processor groups
What is missing

- Calculation of operator matrices in DFT
  Generalize integral routines to matrices $A^{0g}$
  Generalize integrate potential the same way
- General collocate density routine to handle $P^{0g}$
- Symmetrization of density
- Forces and stress for k-points
- HFX and ADMM
- Band structure calculation routine
Acknowledgement

CP2K Community