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

25. June 2001

CP2K source repository goes online on berlios.de

Now on sourceforge.net

13 years of open development

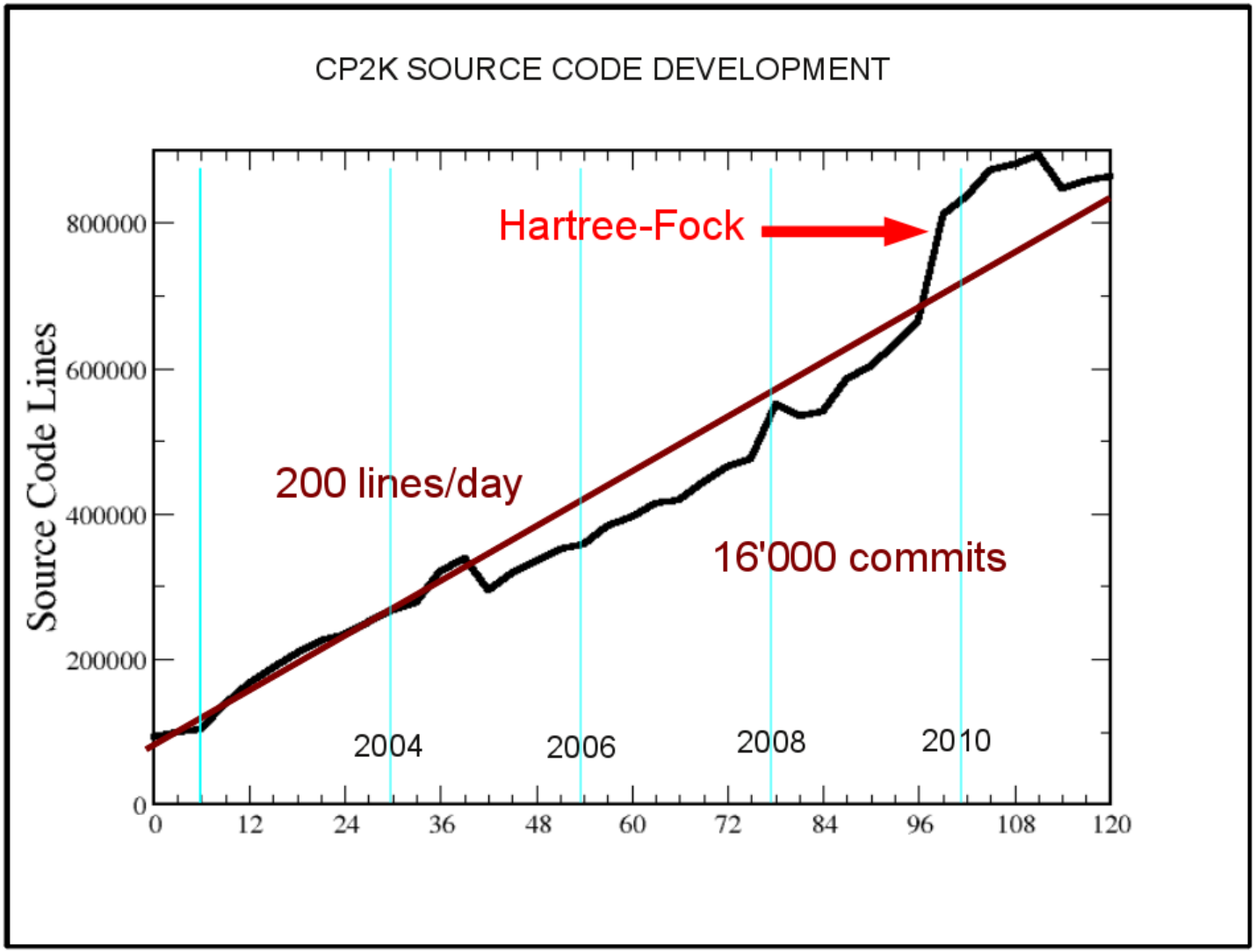
Origins

- Quickstep DFT Code, Max-Planck-Institute, Stuttgart

  Gerald Lippert, Matthias Krack, JH

- Fist MD Code, UPenn, Philadelphia

  Chris Mundy, S. Balasubramanian, Ken Bagchi

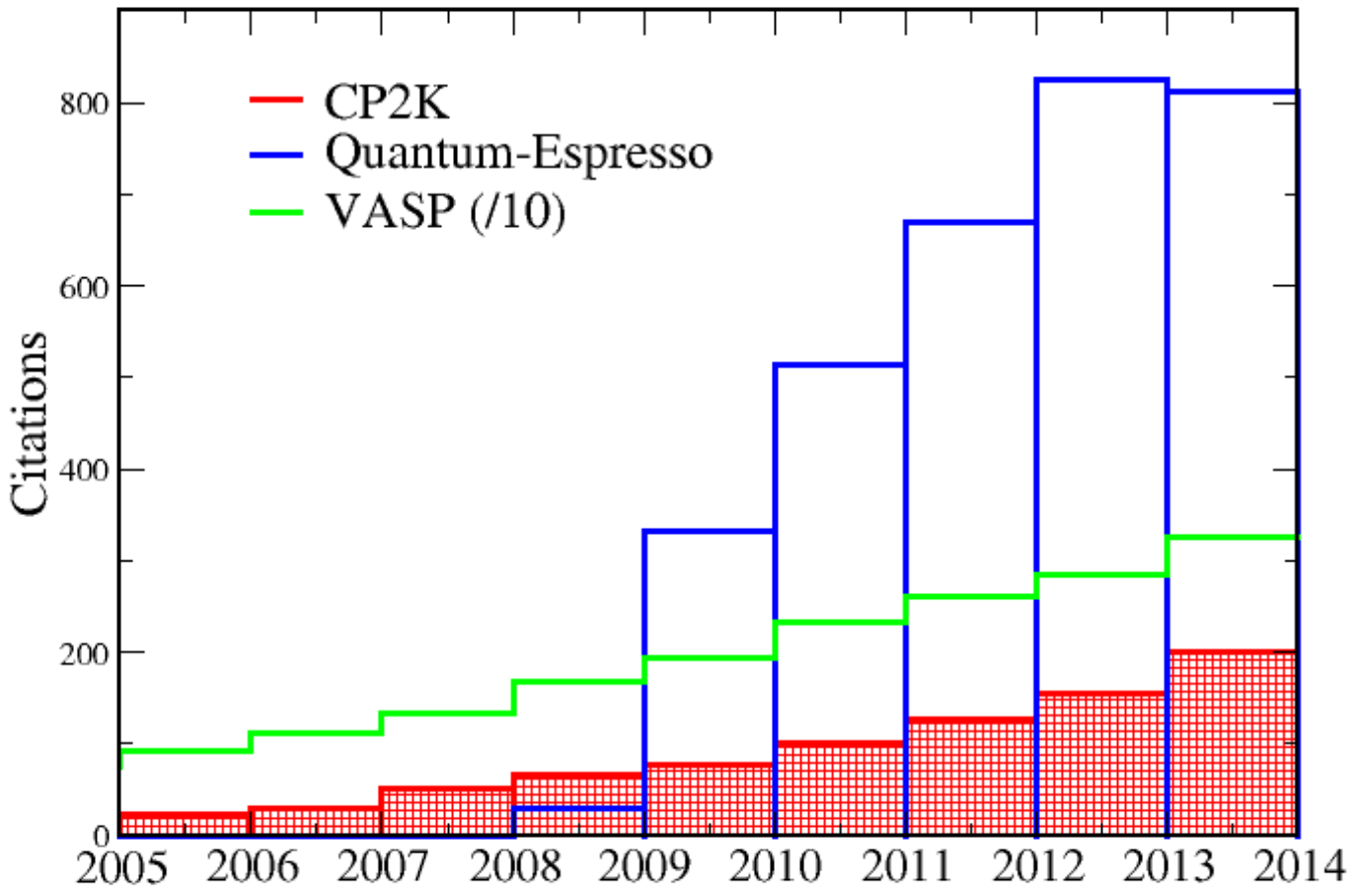

CP2K SOURCE CODE DEVELOPMENT
Source Code Lines
800000
600000
400000
200000
0
0
12
24
36
48
60
72
84
96
108
120
Hartree-Fock
200 lines/day
16'000 commits
2004
2006
2008
2010

Recent Versions

90 lines/day and 2.6 commits/day

Date	Version	Source lines	Commit
10.2011	2.2	891’928	11883
09.2012	2.3	919’583	12358
06.2013	2.4	968’038	12977
02.2014	2.5	947’762	13637
12.2014	2.6	990’282	14881

CP2K: Impact on Science



CP2K: Application Fields[1]

	Chemistry	Materials Science	Physics
CP2K	70%	31%	41%
Quantum-Espresso	35%	57%	69%
VASP	36%	50%	65%

[1] scopus.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

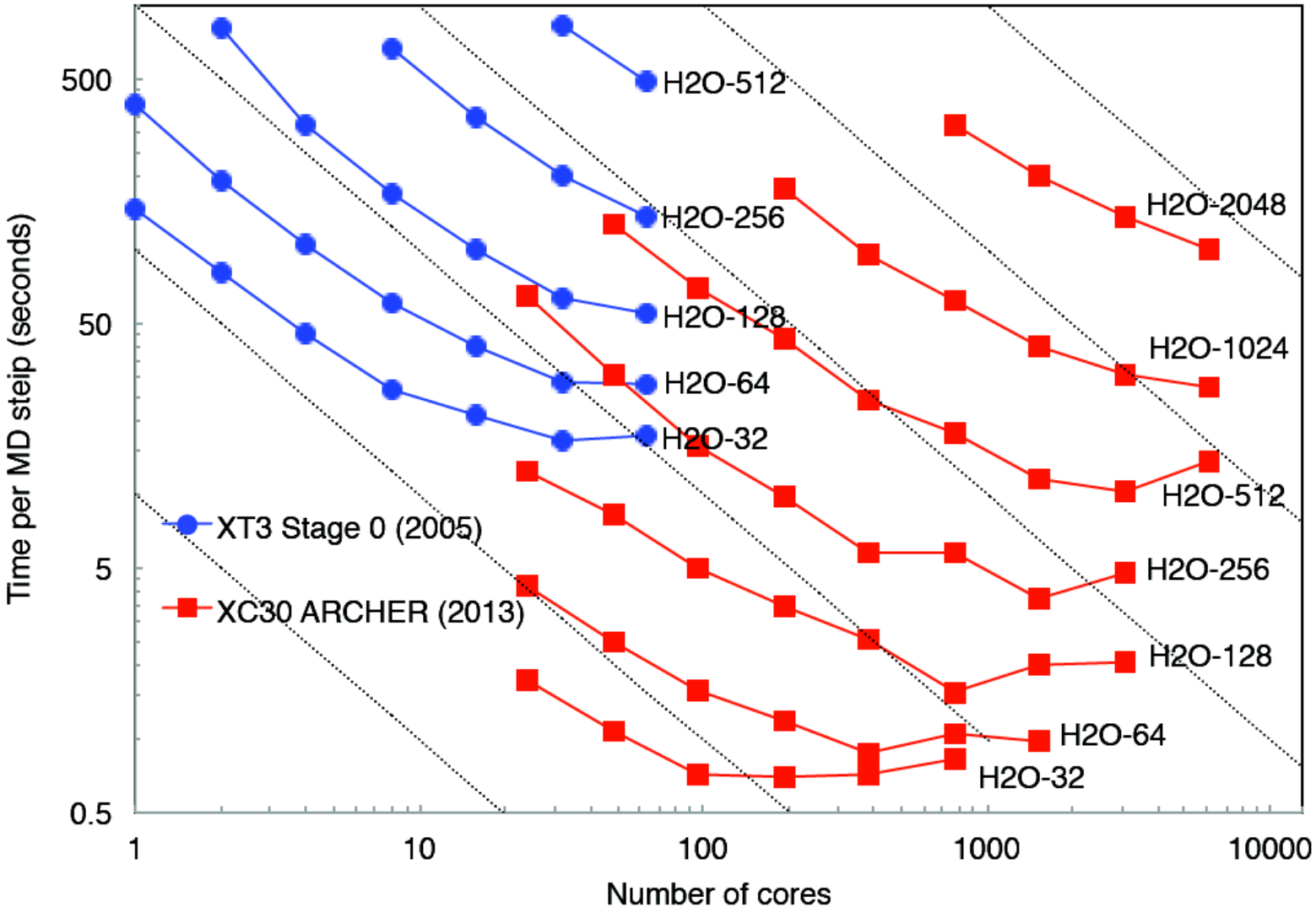
2012 Linear scaling KS

2014 RI-MP2, RI-RPA, libxc, nl-vdW

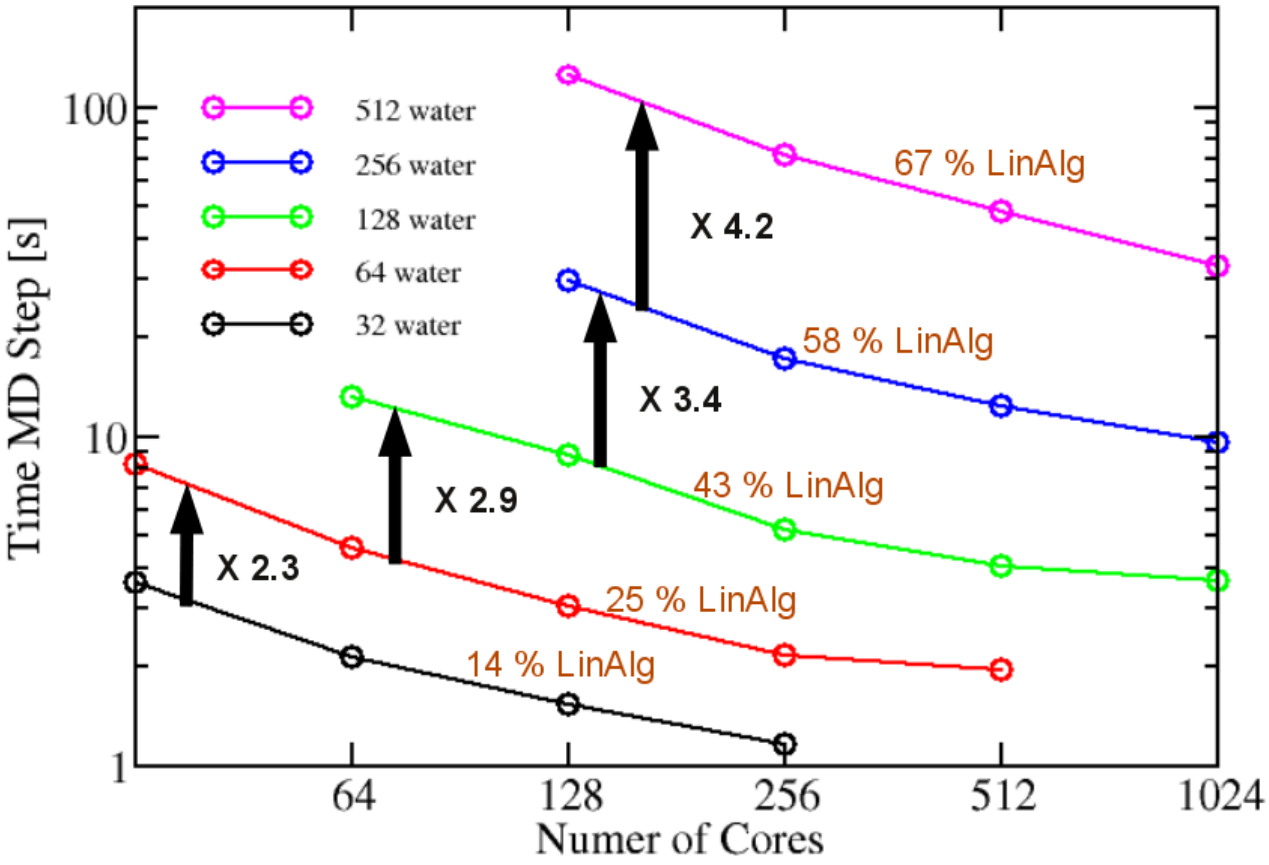
Quickstep DFT Code

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

Water Benchmarks



System Size Scaling



Large Systems

Dominated by linear algebra

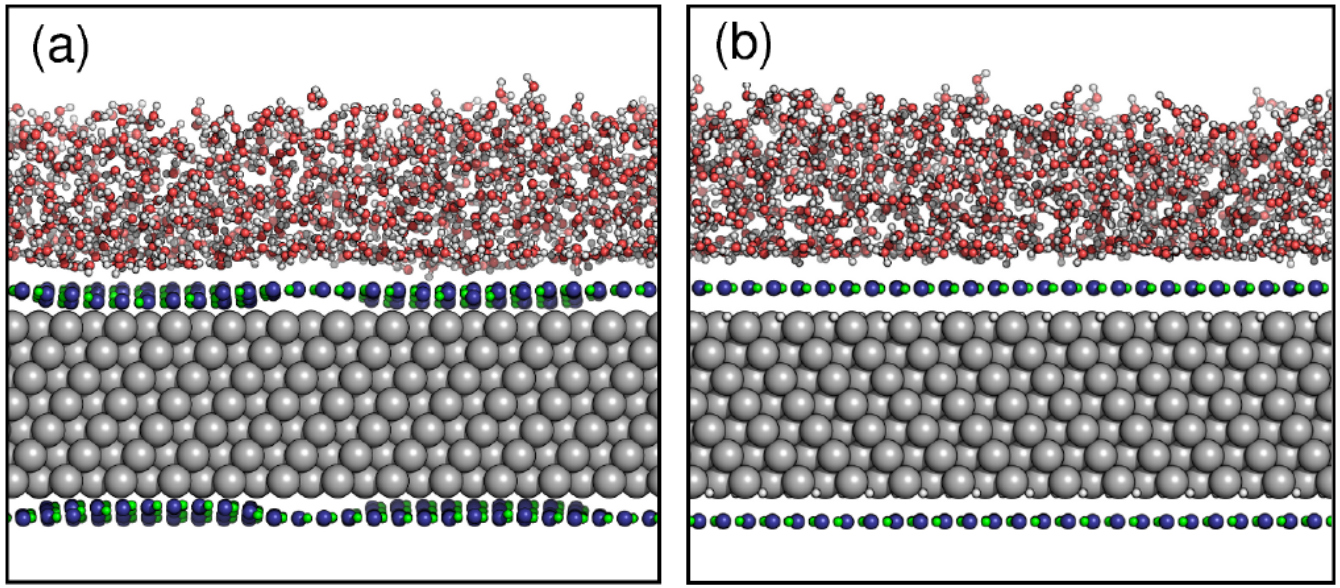
Key Operation : Sparse Matrix Multiplication

$\Downarrow$

DBCSR Library

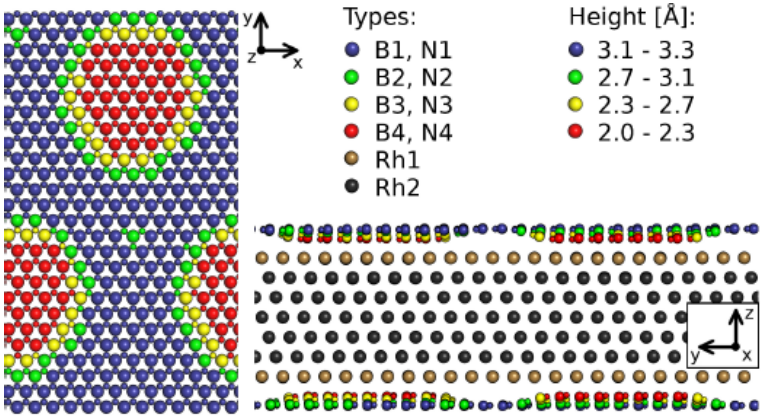
Distributed Block Compressed Sparse Row Format

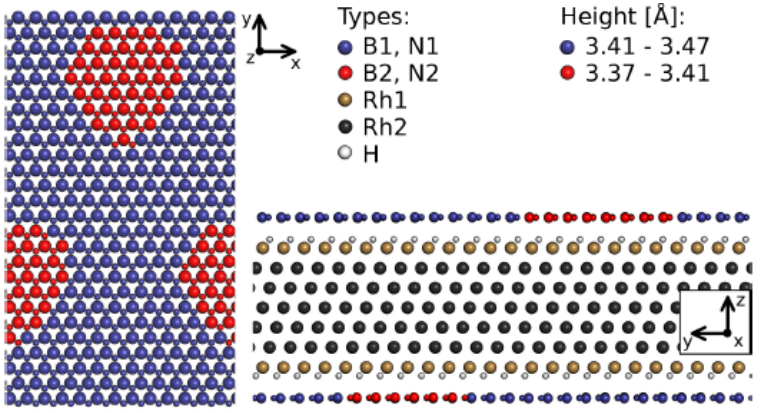
Water on *h*-BN

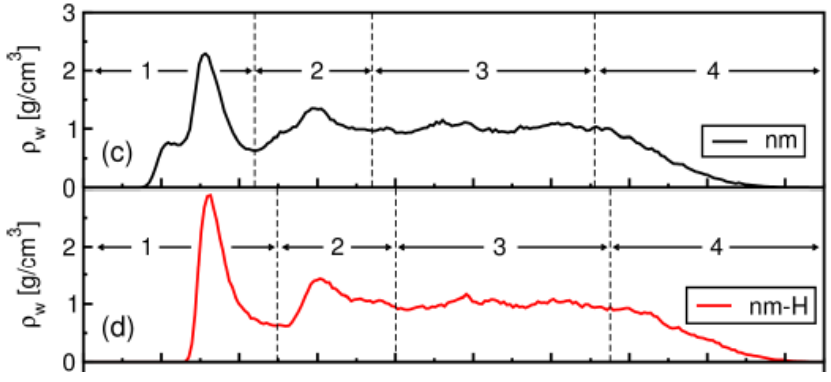


833 water molecules on *h*-BN layer on metal surface. QM/MM using optimized force field (QM: water, MM: h-BN and metal)

Wetting Angle







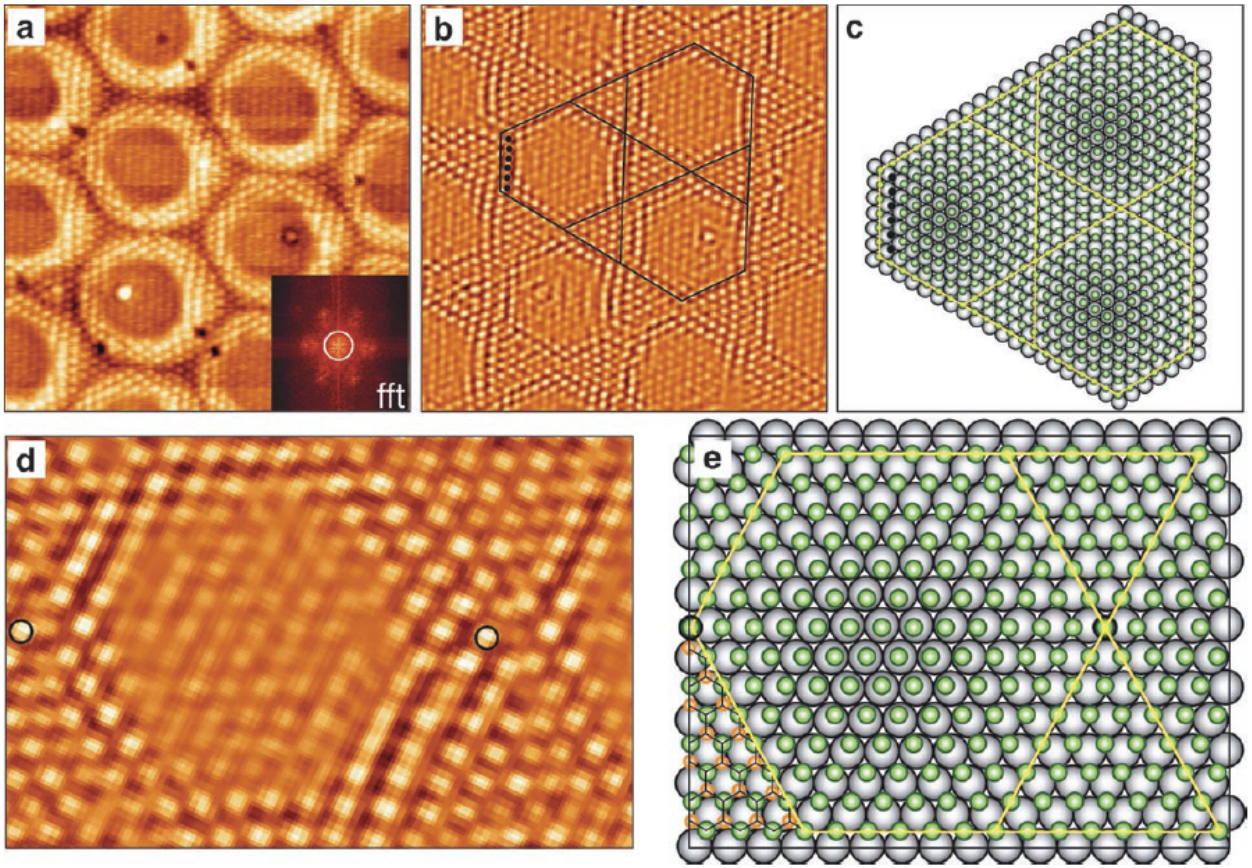
$\Delta \cos \Theta \approx 0.04 \quad \rightarrow \quad$ smaller contact angle for NM

D. Golze et al, PCCP in press

h-BN/Rh(111): Structure

Corrugated Monolayer Model

3.2 nm periodic structure



S. Berner et al. Angew. Chem. **46** 5115 (2007)

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

Methods

- Gaussian and Plane Waves (GPW)

  G. Lippert, JH, and M. Parrinello; Mol. Phys. 92, 477-488 (1997)

- Basis sets: DZVP MOLOPT type

  J. VandeVondele, JH, Journal of Chemical Physics 127, 114105 (2007)

  500 Ry PW cutoff (electron density)

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

  C. Hartwigsen, S. Goedecker and JH; Phys. Rev. B 58, 3641-3662 (1998)

- 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

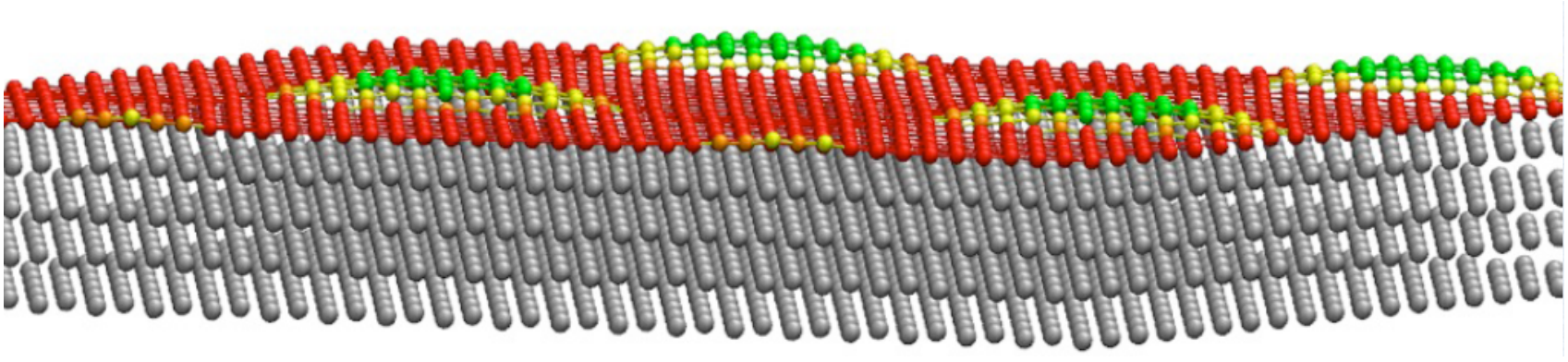
Computational Models

Slab models, 3d periodic, 20 Å empty space

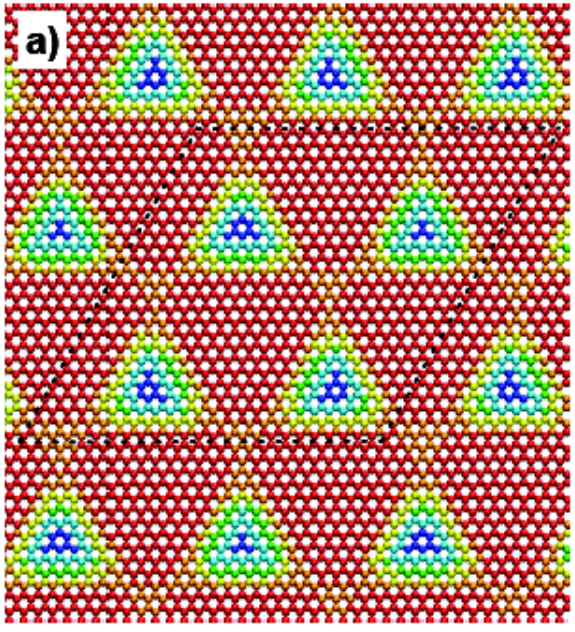
- *h*-BN/Rh(111) (small)
  4 layer Rh + 1 BN: 914 atoms, 19370 BSF, 11144 el
- *h*-BN/Rh(111) (large)
  7 layer Rh + 2 BN: 1684 atoms, 34996 BSF, 19840 el
- Multiple cells (2x2)
  4 layer Rh/Cu + 1 BN/gr: 3656 atoms, 77480 BSF, 44576 el

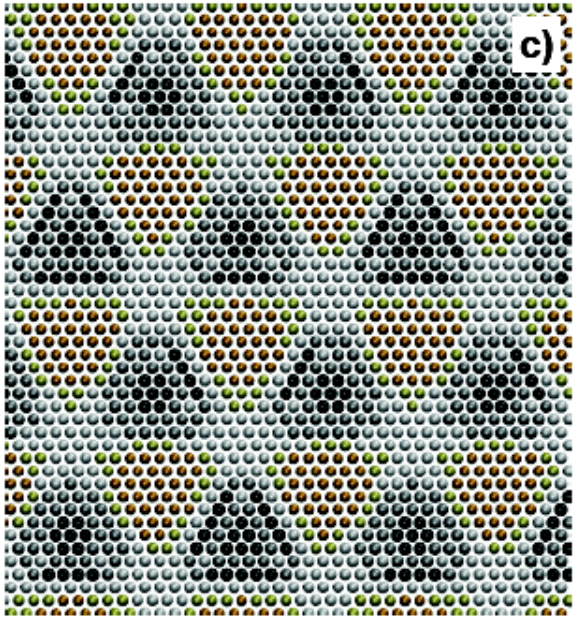
DFT Optimized Structure

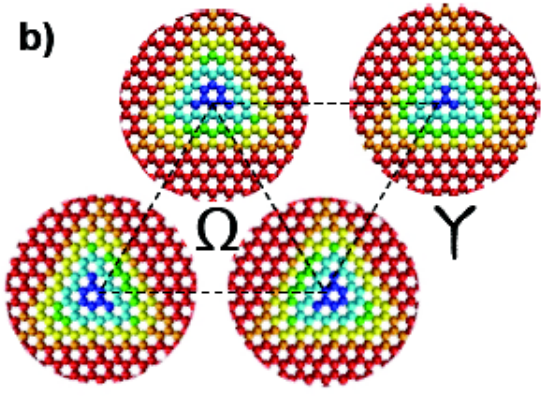
25 on 23 reconstruction

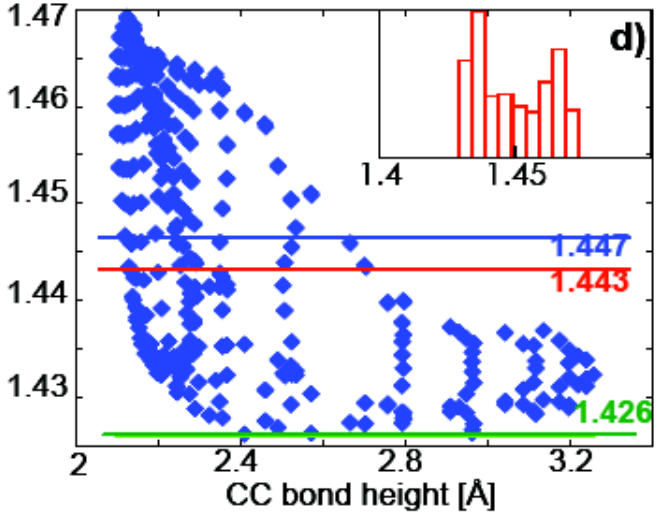


DFT Structure Analysis









Electronic corrugation in rotatet *h*-BN on Cu(111)

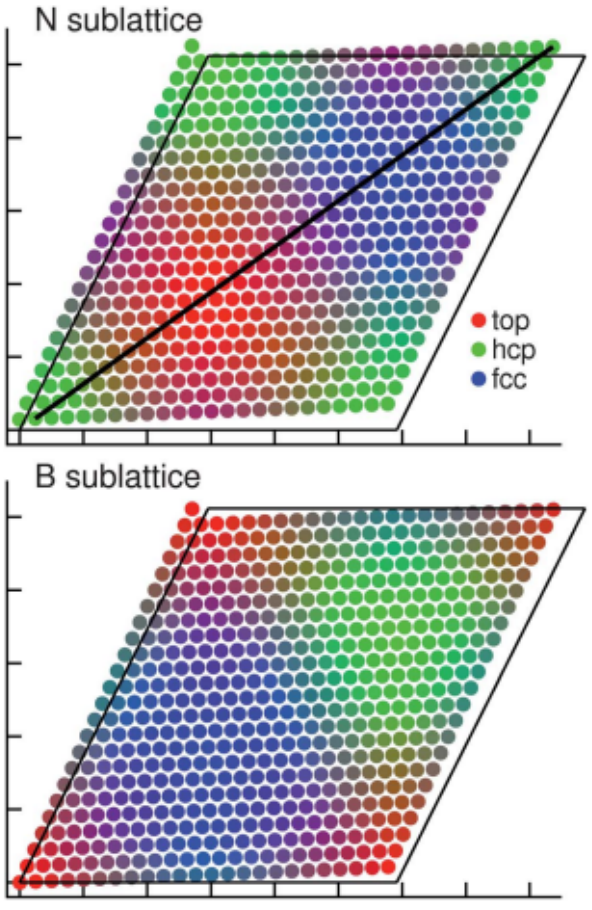


Fig. 1 Registry of the N and B atoms in the corresponding sublattices relative to the metal substrate. The top, hcp and fcc registries are shown as the red, green and blue components of the point colors, respectively. Mixed colors indicate intermediate and bridging positions. The black line indicates 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.

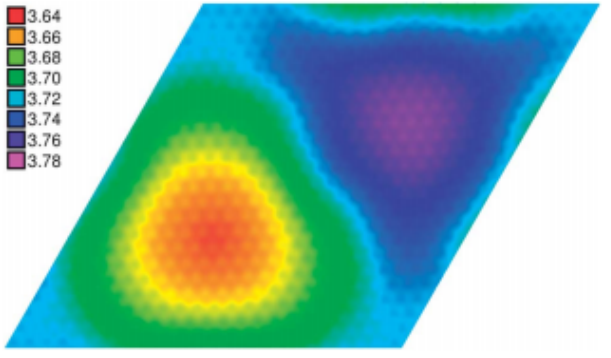
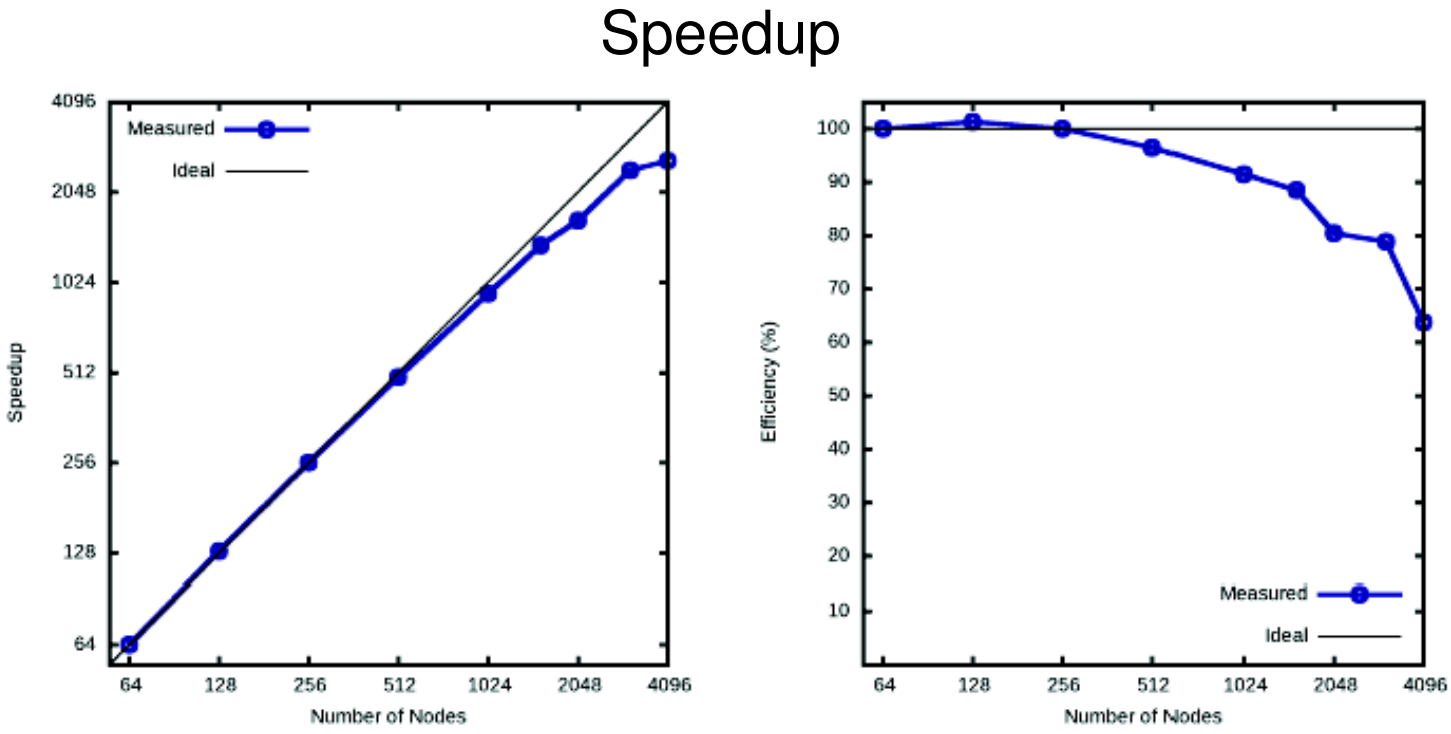


Fig. 3 Lateral map of the electrostatic potential relative to the Fermi energy (eV) at constant height, approx. 3.4 Å above the *h*-BN layer. The area is the same as the unit cell drawn in Fig. 2.

MP2: Forces and Stress



(a) (b)

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

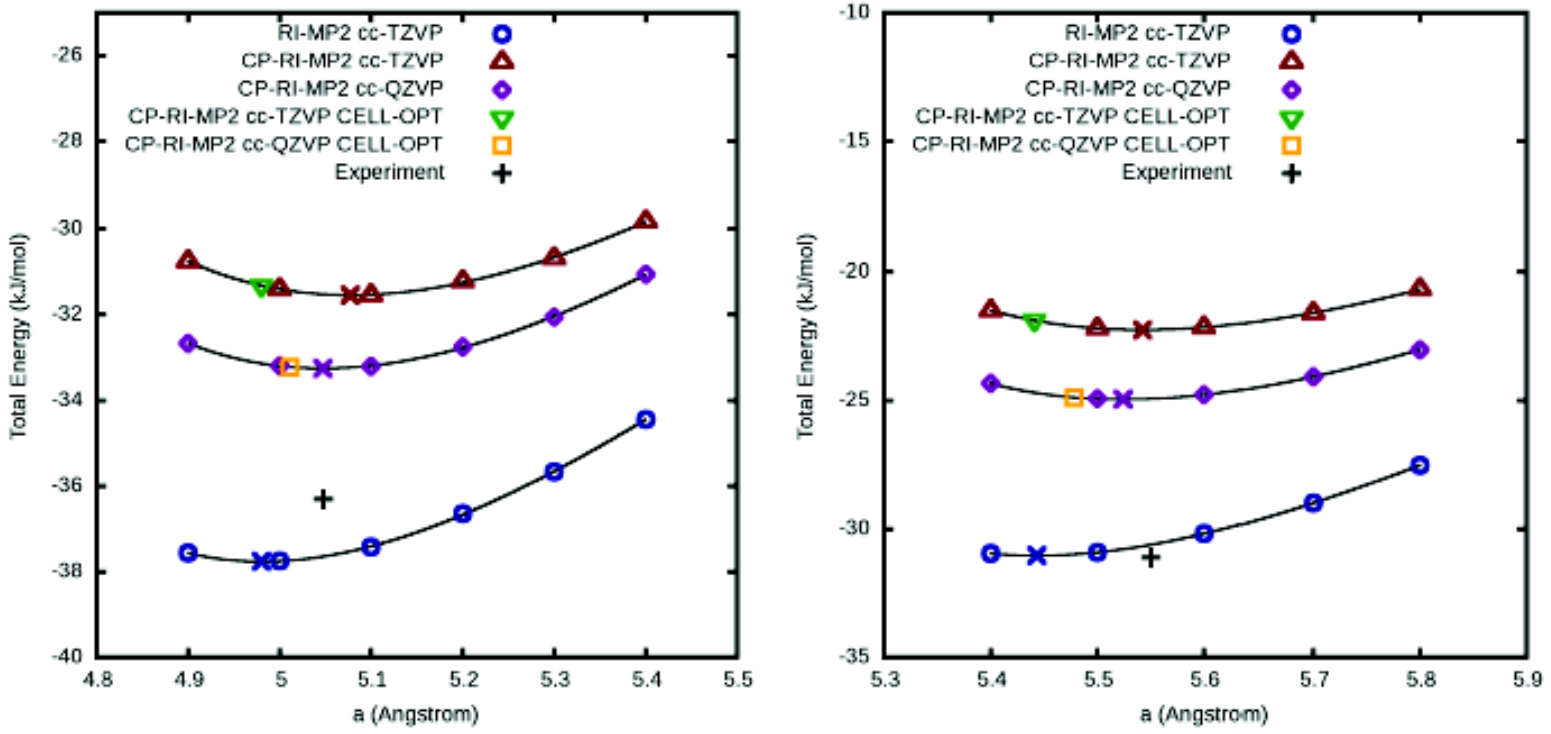


Figure 5. Location of the minima for NH_3 (a) and CO_2 (b), computed at the RI-MP2 level of theory with different basis sets obtained with different approaches. The lattice parameter optimization curves have been fitted with a third order Birch-Murnagham equation, the crosses represent the location of the minimum point for each curve. CP means that the cohesive energy have been counterpoise corrected.

Timings

	o	n	N_a	t_{tot}	t_D	$\frac{t_D}{t_E}$	t_D^{GPU}	$\frac{t_D}{t_D^{\text{GPU}}}$
NH_3	128	2272	5312	3.15	1.53	4.20	1.47	1.04
U	192	2752	6784	5.97	3.58	4.59	2.89	1.24
FA	216	2760	6912	5.83	3.87	4.28	2.95	1.31
D	192	2992	7520	12.84	5.27	5.15	4.26	1.24
CO_2	256	2784	7296	7.94	4.99	4.15	3.50	1.43
H_2O	256	3648	8704	10.17	9.34	4.00	5.85	1.60
B	240	4128	10176	23.01	13.77	4.45	8.81	1.56
PD	312	3936	10208	28.96	17.48	4.13	9.80	1.78
SA	304	4144	10432	27.00	19.29	4.26	10.94	1.76
CT	336	4152	10560	29.71	22.30	4.16	11.97	1.86

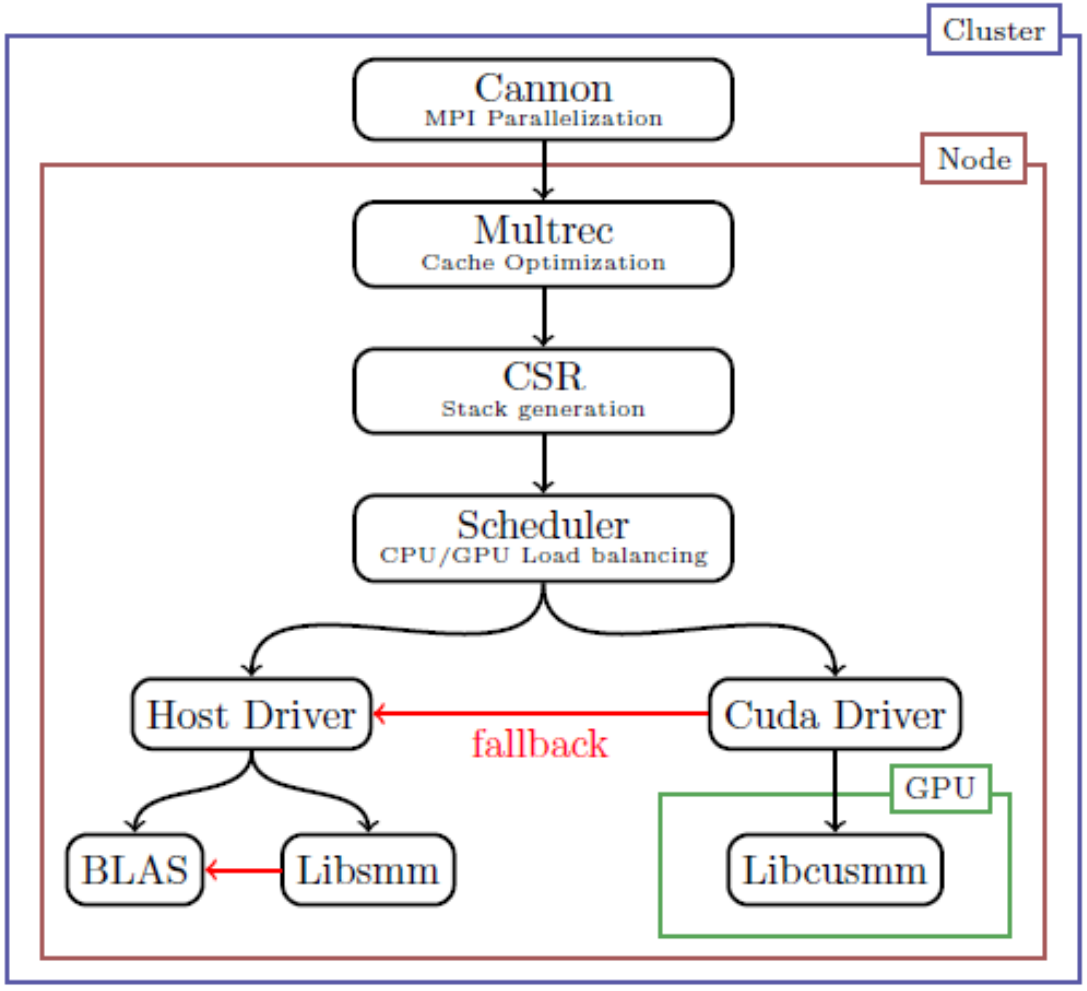
Libraries

- 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



Performance

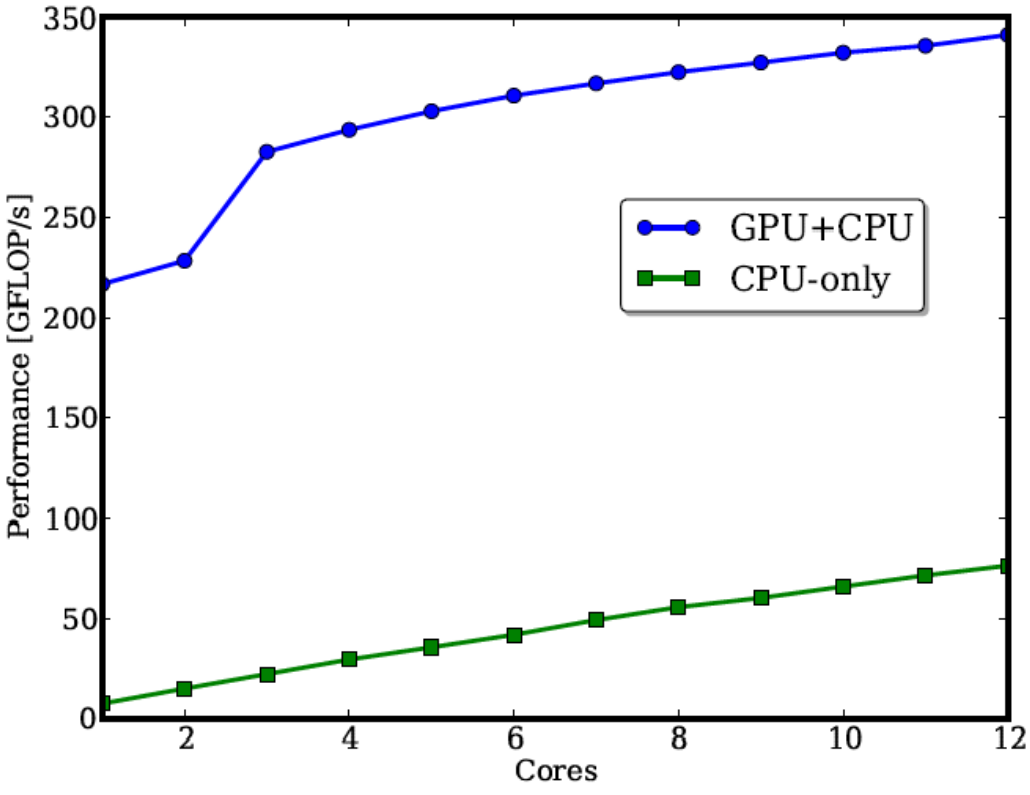
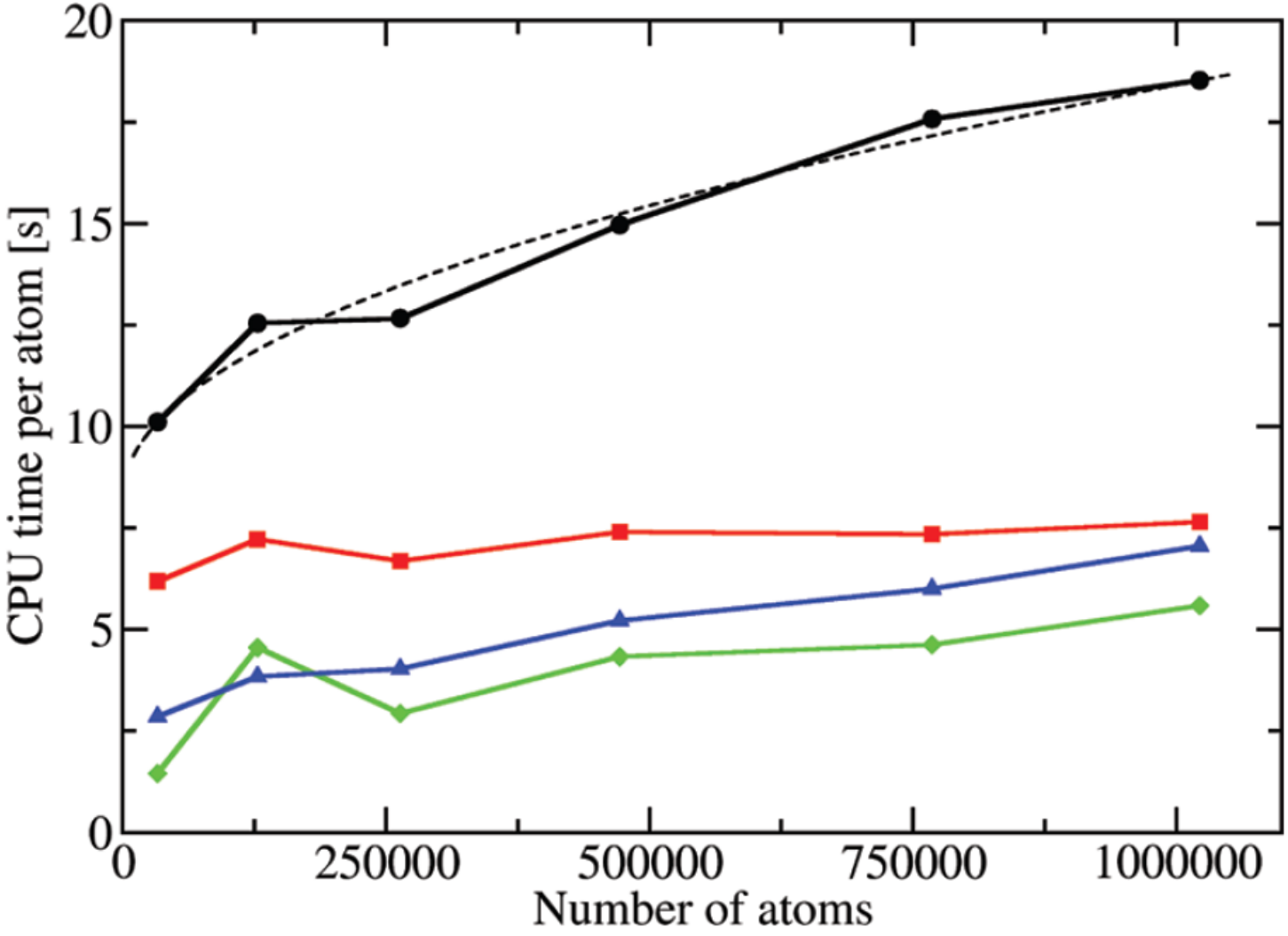


Figure 9: Performance comparison of the multi-threaded DBCSR library based on 23x23 matrix blocks, and was not using the MPI capabilities. The benchmark was run on a dual Sandy Bridge (E5-2620, 2.0GHz, 6 cores) machine, equipped with one NVIDIA Tesla K20 card.

Limitations



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

libxc

- XC functionals library from http://www.tddft.org/programs/octopus/wiki/index.php/Libxc
- 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

Eigensolver

- 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)

$$\rho(r) = \sum_{AB} \rho_{AB} \approx \sum_{AB} \left[\sum_{u} f_u^{AB} \chi_u^A(r) + \sum_{v} f_v^{AB} \chi_v^B(r) \right]$$

- 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

$$\mathbf{F^k C^k} = \mathbf{S^k C^k E^k}$$

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

$$\mathbf{A^k} = \sum_{\mathbf{g}} \mathbf{A^{0g}} e^{i\mathbf{k}\cdot\mathbf{g}}$$

the density matrix is calculated in Fourier space from $\mathbf{C^k}$ and then transformed to real space

$$\mathbf{P^{0g}} = \frac{1}{\Omega} \int \mathbf{P^k} e^{i\mathbf{k}\cdot\mathbf{g}} d\mathbf{k}$$

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 $\mathbf{A}^{0g}$
  Generalize integrate potential the same way
- General collocate density routine to handle $\mathbf{P}^{0g}$
- Symmetrization of density
- Forces and stress for k-points
- HFX and ADMM
- Band structure calculation routine

Acknowledgement

CP2K Community