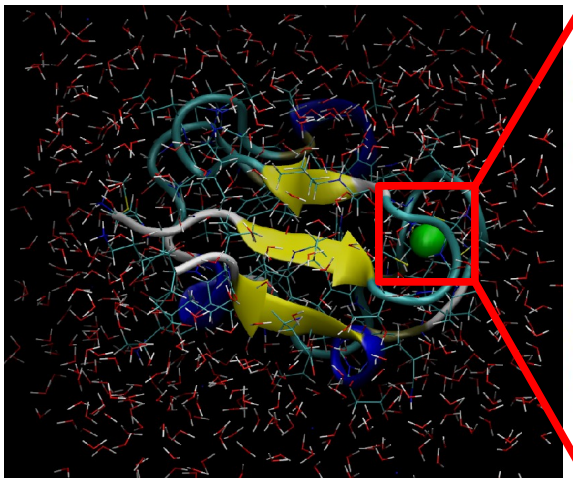
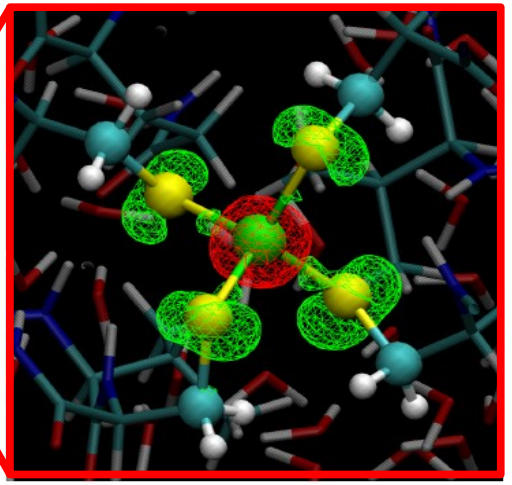
CP2K: summary and new developments



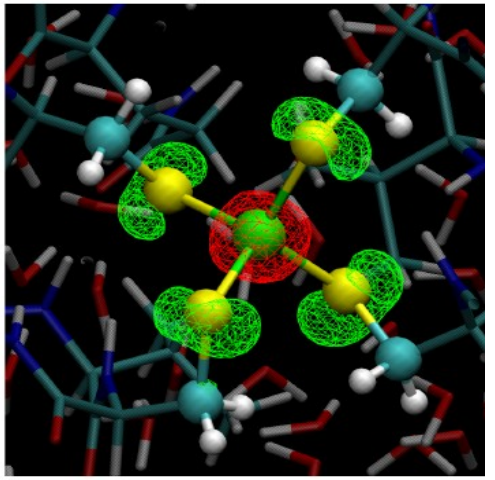
The DFT model: a solvated protein

Standard hybrid



- 48'000 cores
- 1. SCF step: 45 min

ADMM based hybrid



- 1'152 cores
- 1. SCF step: 75 s

Joost VandeVondele
Nanoscale Simulations, ETH Zurich

What is CP2K ?

CP2K is a freely available program to perform atomistic and molecular simulations of solid state, liquid, molecular and biological systems. It provides a general framework for different methods such as e.g. density functional theory (DFT) [...]*

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_s(\vec{r})\right]\phi_i(\vec{r}) = \epsilon_i\phi_i(\vec{r})$$

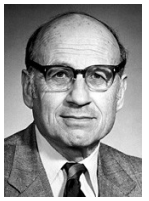
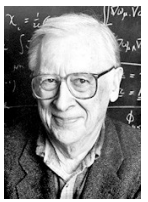
$$V_s(\vec{r}) = V(\vec{r}) + \int \frac{e^2 n_s(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|}\mathrm{d}^3r' + V_{\mathrm{XC}}[n_s(\vec{r})]$$

Nobel Laureates emphasize the importance of atomistic and electronic modeling:

1966:

I would like to emphasize strongly my belief that the era of computing chemists, when hundreds if not thousands of chemists will go to the computing machine instead of the laboratory for increasingly many facets of chemical information, is already at hand.

1998

2013



*) On our web page since the initial version in 2004-10-16

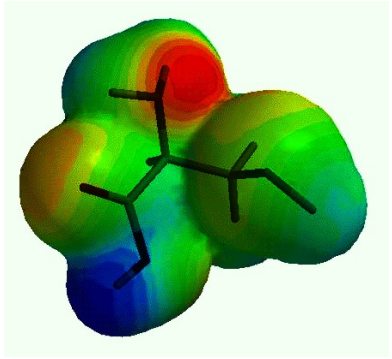
CP2K: the swiss army knife of molecular simulation

- A wide variety of models Hamiltonians
  - classical
  - semi-empirical
  - local and non-local DFT
  - MP2 & RPA
  - Combinations (e.g. QM/MM)
- Various algorithms
  - Molecular dynamics & Monte Carlo
    - NVE, NVT, NPT
  - Free energy and PES tools
  - Ehrenfest MD
- Properties
  - Vibrational
  - NMR, EPR, XAS, TDDFT
- Open source & rapid development
  - 1.000.000 lines of code

CP2K: the swiss army knife of molecular simulation

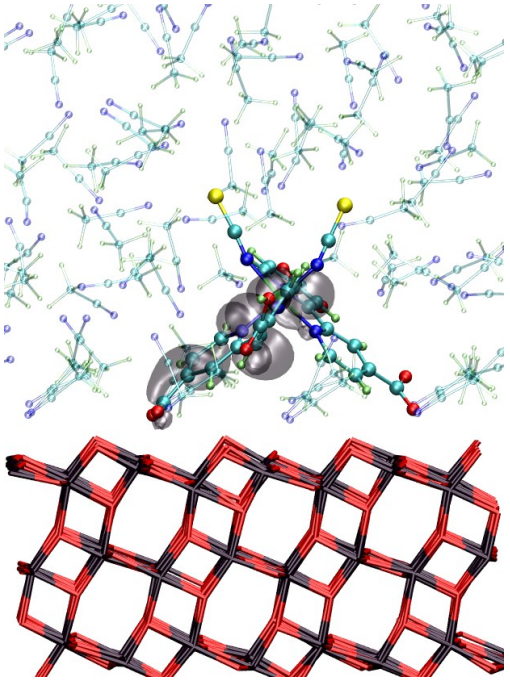
- A wide variety of models Hamiltonians
  - classical
  - semi-empirical
  - local and non-local DFT
  - MP2 & RPA
  - Combinations (e.g. QM/MM)
- Various algorithms
  - Molecular dynamics & Monte Carlo
    - NVE, NVT, NPT
  - Free energy and PES tools
  - Ehrenfest MD
- Properties
  - Vibrational
  - NMR, EPR, XAS, TDDFT
- Open source & rapid development
  - 1.000.000 lines of code

Evolution in Tools

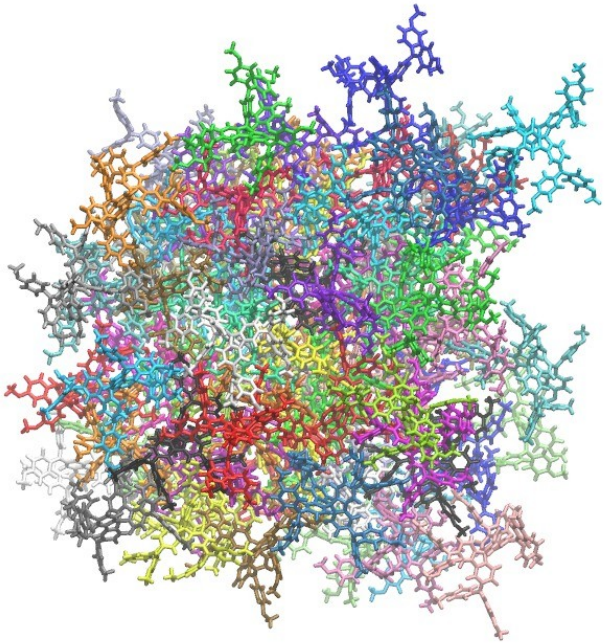


© nobelprize.org (1998)

10 years



5 years



~10 atoms
static

~1000 atoms,
1-100 ps dynamics

~100000 atoms,
or ultra high accuracy

Improved software, hardware and algorithms

Gaussian and plane waves: GPW in CP2K

- Primary basis: Gaussians → Chemistry
  - compact
  - sparse H^{ks} (and P)
  - Many terms analytic
- Auxiliary basis: Plane waves → Physics
  - regular grid for e^- density
  - FFT for Poisson equation
  - No four center integrals needed

The GPW algorithm : compute the GGA Kohn-Sham matrix in O(N) time, PBC are natural.

J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, Comp. Phys. Comm. 167, 103 (2005).
Lippert, G; Hutter, J; Parrinello, M. Mol. Phys., 92 (3), 477-487 (1997).

Orbital transformations (OT)

A cubic, very robust algorithm avoiding the cost of traditional diagonalization

- New variables

$$C(X) = C_0 \cos\left(\sqrt{X^T S X}\right) + X \frac{\sin\left(\sqrt{X^T S X}\right)}{\sqrt{X^T S X}}$$

$$X^T S C_0 = 0 \qquad C(X)^T S C(X) = 1 \;\; \forall X$$

- Direct minimization of $E_{ks}[\{X\}]$
- Linear constraint -> guaranteed convergence!

J. VandeVondele, J. Hutter, *J. Chem. Phys., 2003,* Vol. 118 No. 10, 4365-4369

Enabling high impact simulation research with CP2K

High impact science and engineering funds the code &machines!

Various independent groups have adopted CP2K as their simulation tool of choice:

http://www.cp2k.org/science

Selected high impact publications using&citing CP2K (Dec 2013) :

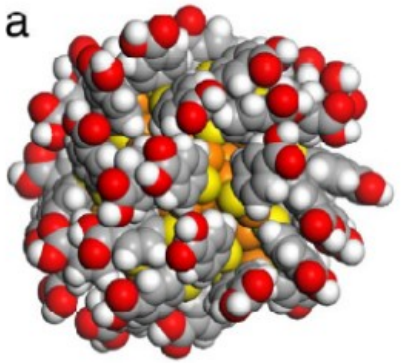
JOURNAL OF PHYSICAL CHEMISTRY LETTERS	19
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	14
PHYSICAL REVIEW LETTERS	12
ANGEWANDTE CHEMIE INTERNATIONAL EDITION	5
ACS NANO	4
NATURE CHEMISTRY	3
NATURE COMMUNICATIONS	3
NANO LETTERS	2
NATURE	2
NATURE MATERIALS	2
NANOSCALE	1
SMALL	1

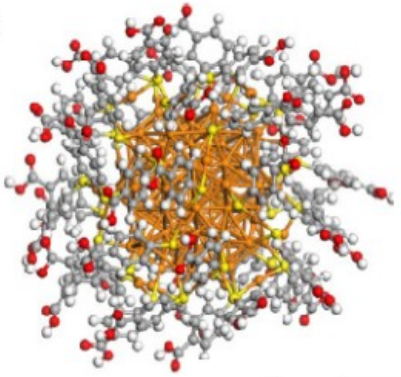
CP2K: science (I)

A unified view of ligand-protected gold clusters as superatom complexes

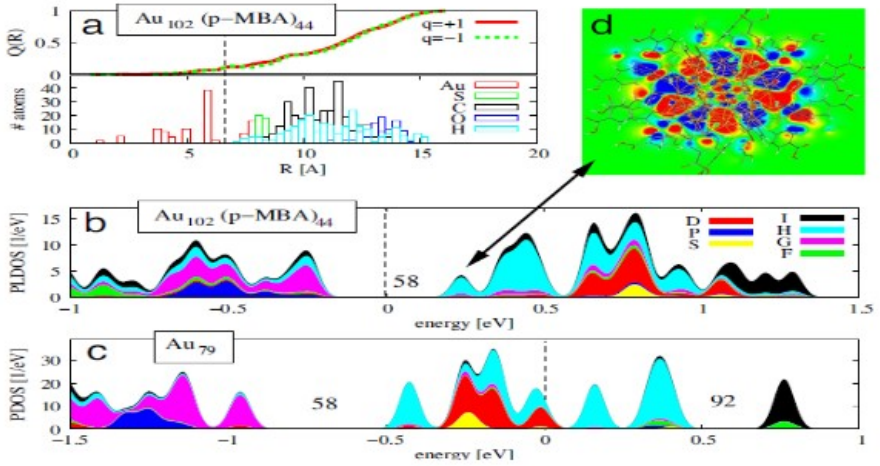
Michael Walter[†], Jaakko Akola[†‡], Olga Lopez-Acevedo[†], Pablo D. Jadzinsky[§¶], Guillermo Calero[§], Christopher J. Ackerson[§‖], Robert L. Whetten[††], Henrik Grönbeck[‡‡], and Hannu Häkkinen[†§§¶¶]

PNAS July 8, 2008 vol. 105 no. 27 9157–9162





System size: 762 Atoms
~3400 electrons



Electronic structure of nanoparticles

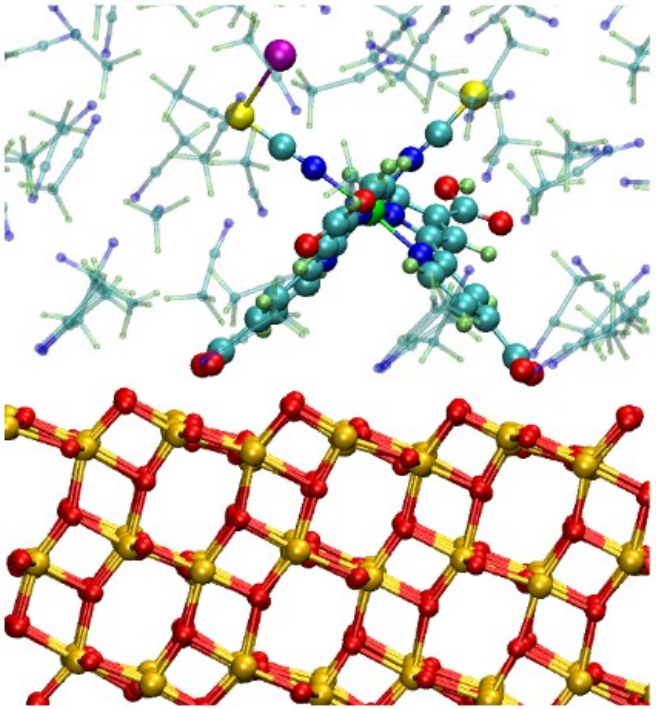
CP2K: science (II)

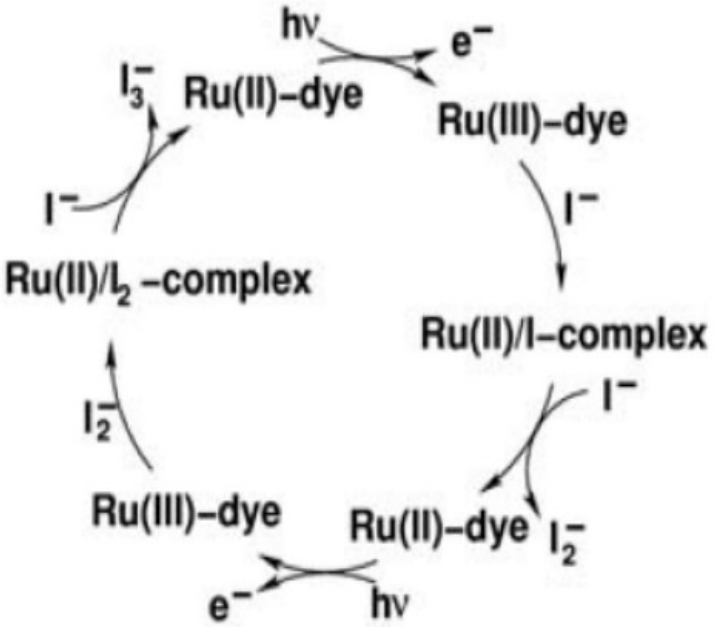
An atomistic picture of the regeneration process in dye sensitized solar cells

Florian Schiffmann[a], Joost VandeVondele[a,1], Jürg Hutter[a], Atsushi Urakawa[b], Ronny Wirz[b], and Alfons Baiker[b]

[a]Institute of Physical Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; and [b]Department of Chemistry and Applied Biosciences, Institute for Chemical and Bioengineering, ETH Zurich, Hönggerberg, HCI, 8093 Zurich, Switzerland

4830–4833 | PNAS | March 16, 2010 | vol. 107 | no. 11





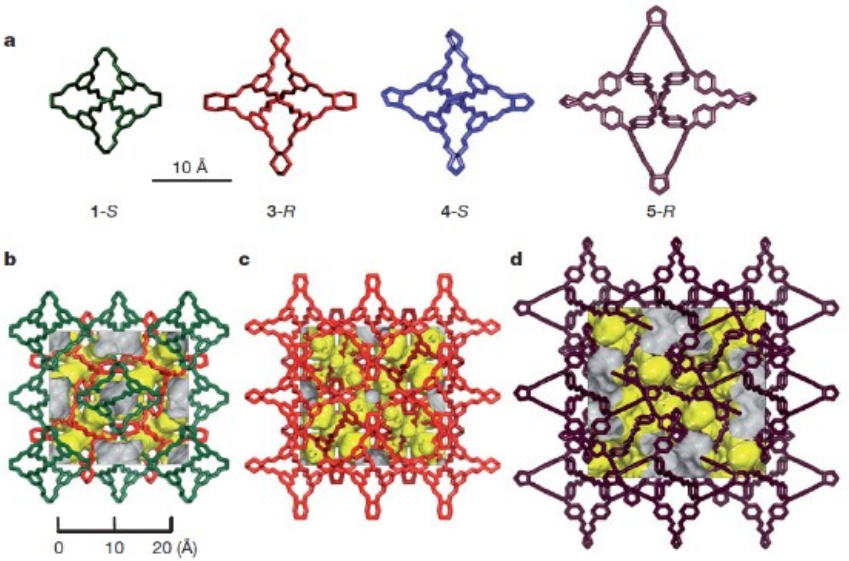
Functionalized solid/liquid interfaces

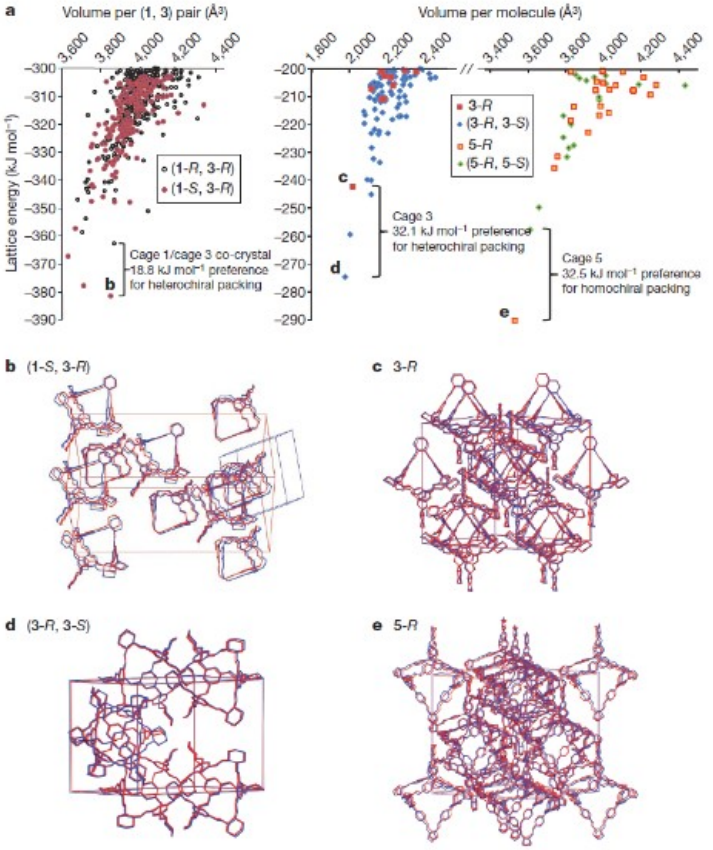
CP2K: science (III)

Modular and predictable assembly of porous organic molecular crystals

James T. A. Jones[1], Tom Hasell[1], Xiaofeng Wu[1], John Bacsa[1], Kim E. Jelfs[1], Marc Schmidtmann[1], Samantha Y. Chong[1], Dave J. Adams[1], Abbie Trewin[1], Florian Schiffman[2], Furio Cora[2], Ben Slater[2], Alexander Steiner[1], Graeme M. Day[3] & Andrew I. Cooper[1]

16 JUNE 2011 | VOL 474 | NATURE | 367





Structure prediction of metal organic frameworks

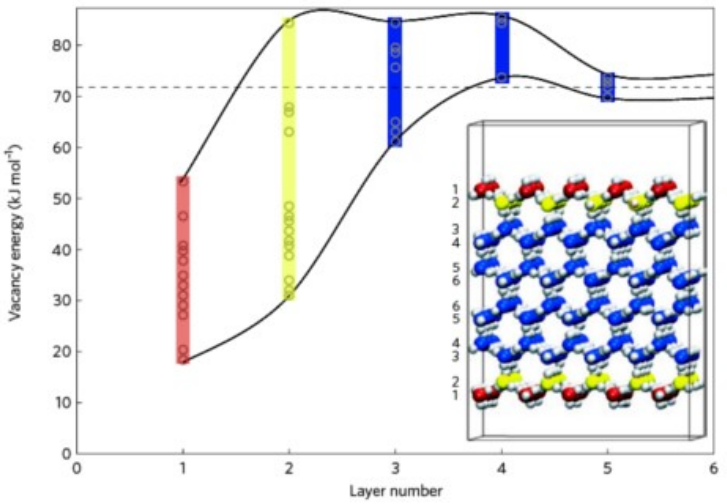
CP2K: science (IV)

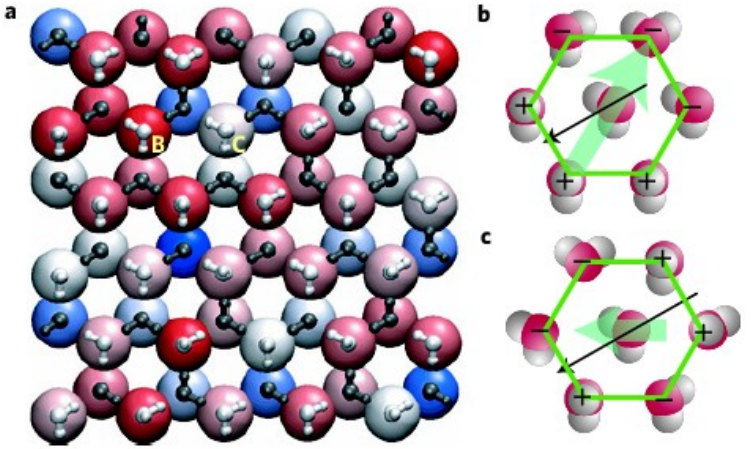
Large variation of vacancy formation energies in the surface of crystalline ice

M. Watkins[1,2,3], D. Pan[4], E. G. Wang[5], A. Michaelides[1,2,3], J. VandeVondele[6] and B. Slater[1,3]*

[1]Department of Chemistry, Christopher Ingold Building, 20 Gordon Street, University College London, London WC1H 0AJ, UK, [2]London Centre for Nanotechnology, University College London, London WC1H 0AJ, UK, [3]TYC@UCL, University College London, London WC1H 0AJ, UK, [4]Institute of Physics, Chinese Academy of Sciences, PO Box 603, Beijing 100190, China, [5]School of Physics, Peking University, Beijing 100871, China, [6]Institute of Physical Chemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. *e-mail: b.slater@ucl.ac.uk.

NATURE MATERIALS | VOL 10 | OCTOBER 2011





Disordered and frustrated materials

Recent developments

Full linear scaling GGA DFT

Efficiently using hybrid functionals

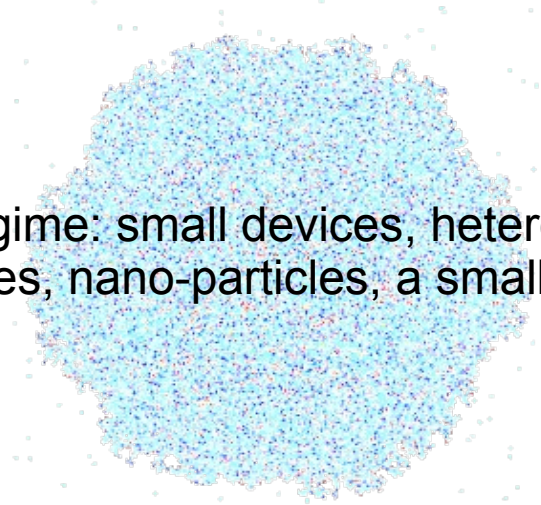
Introducing MP2/RPA/DH correlation

Linear scaling GGA DFT

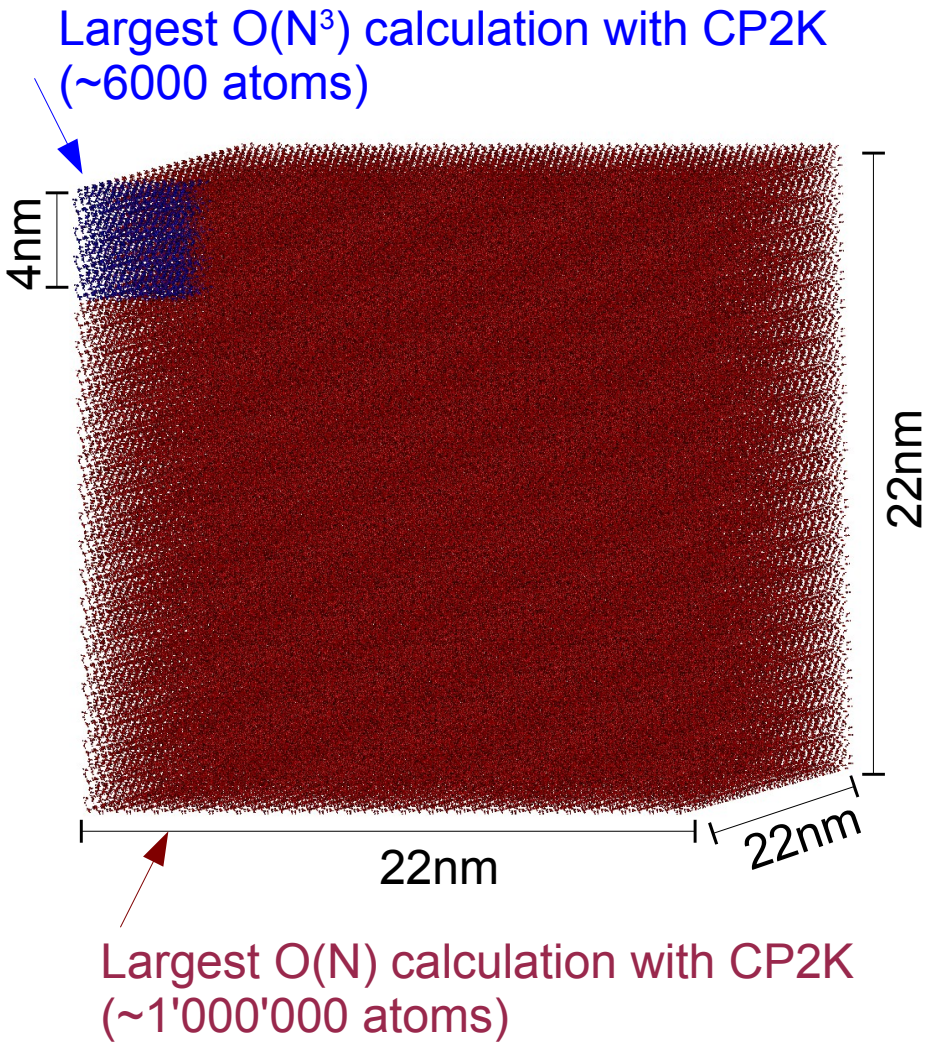
Linear Scaling SCF in CP2K

Traditional approaches to solve the self-consistent field (SCF) equations are $O(N^3)$ limiting system size significantly.

A newly implemented algorithm is O(N), allowing for far larger systems to be studied.



New regime: small devices, heterostructures, interfaces, nano-particles, a small virus.

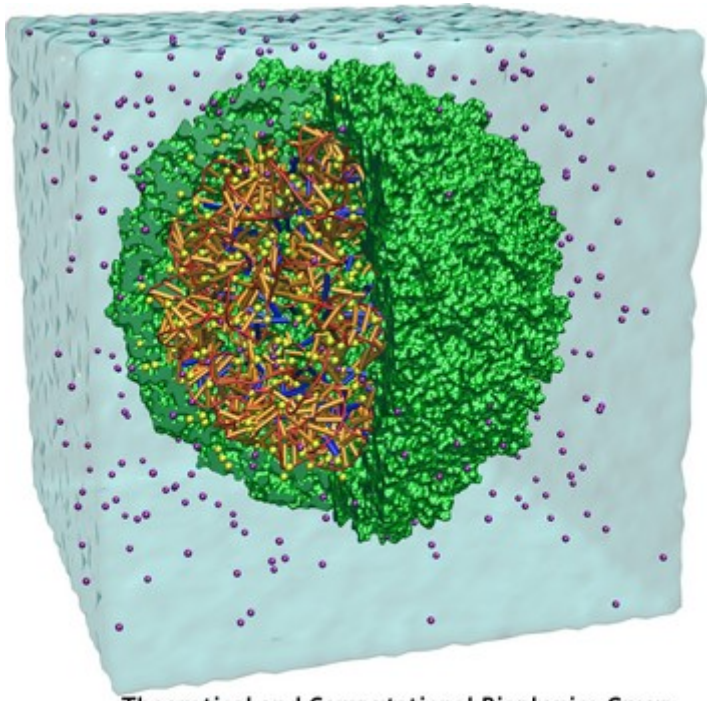


VandeVondele J; Borstnik U; Hutter J; 2012, Linear scaling self-consistent field calculations for millions of atoms in the condensed phase. JCTC 10: 3566 (2012)

Linear Scaling SCF

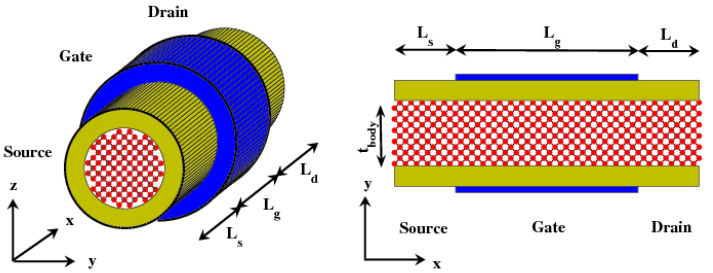
New regime: small devices, heterostructures, interfaces, nano-particles, a small virus.

Solvated STMV: 1M

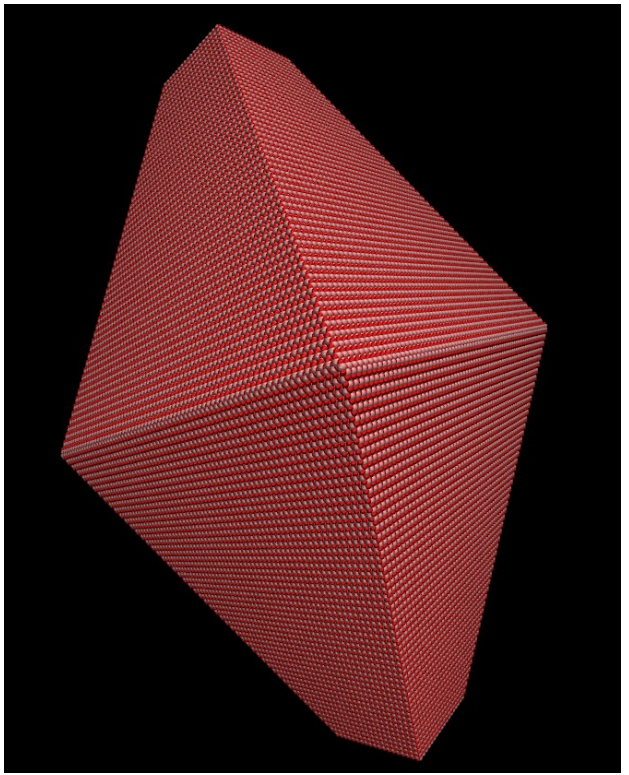


Theoretical and Computational Biophysics Group
Beckman Institute
University of Illinois at Urbana-Champaign

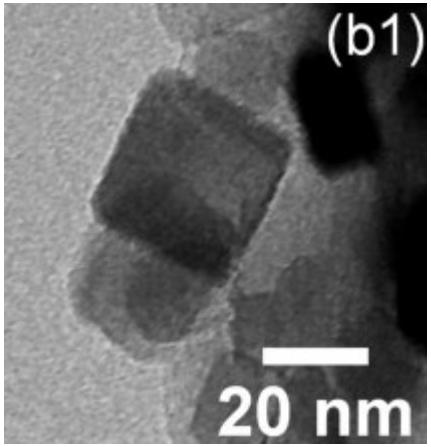
Gate-all-around FET



With Mathieu Luisier



1.5M atoms
Anatase nanocrystal



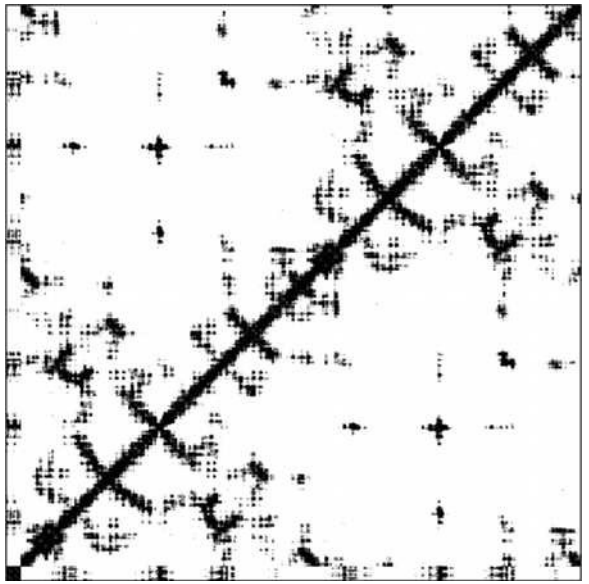
Caplovicova et al.
App. Cat. B, 224, 117

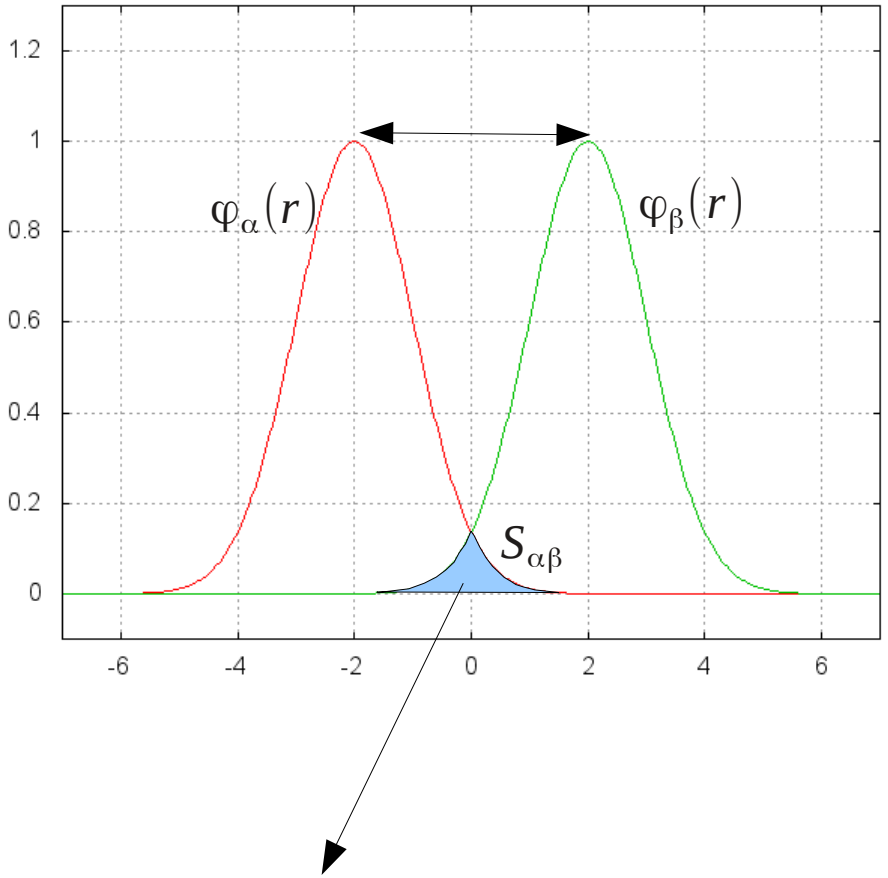
Gaussian basis: The sparsity of H and S

$$S_{\alpha\beta} = \int \varphi_\alpha(r)\varphi_\beta(r)\,dr$$

$$H_{\alpha\beta} = \int \varphi_\alpha(r)v(r)\varphi_\beta(r)\,dr$$

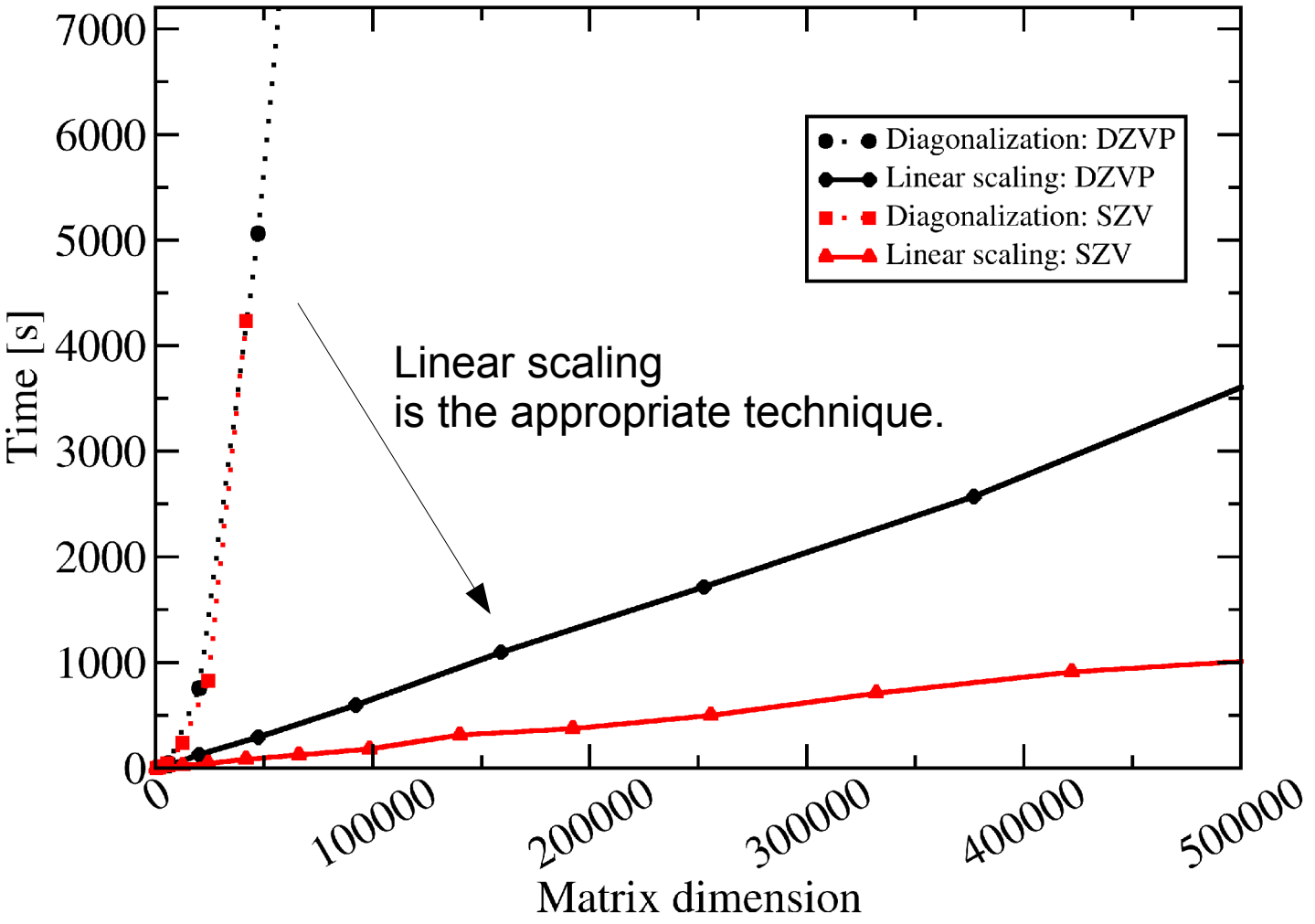
The sparsity pattern of S and H depends on the basis and the spatial location of the atoms, but not on the chemical properties of the system in GGA DFT.





The overlap (integral of the product) rapidly decays with the spatial separation of the basis functions.

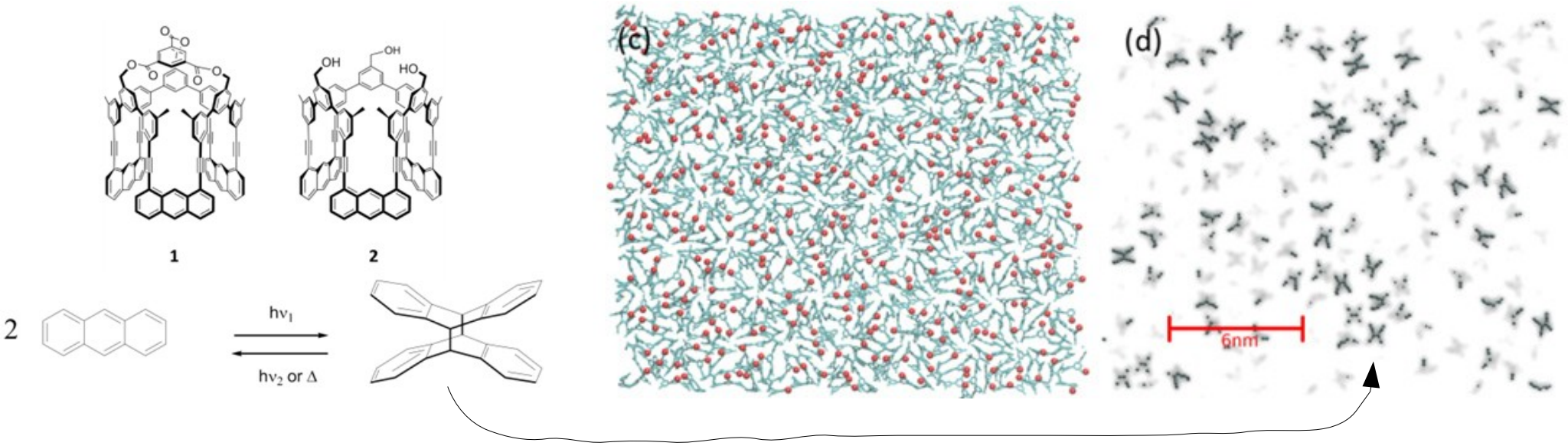
Diagonalization vs. Linear Scaling



Linear scaling based uniquely on sparse matrix matrix multiplication

Bridging from linear scaling SCF to materials properties

2D polymers: synthetically tailored 2D materials beyond graphene

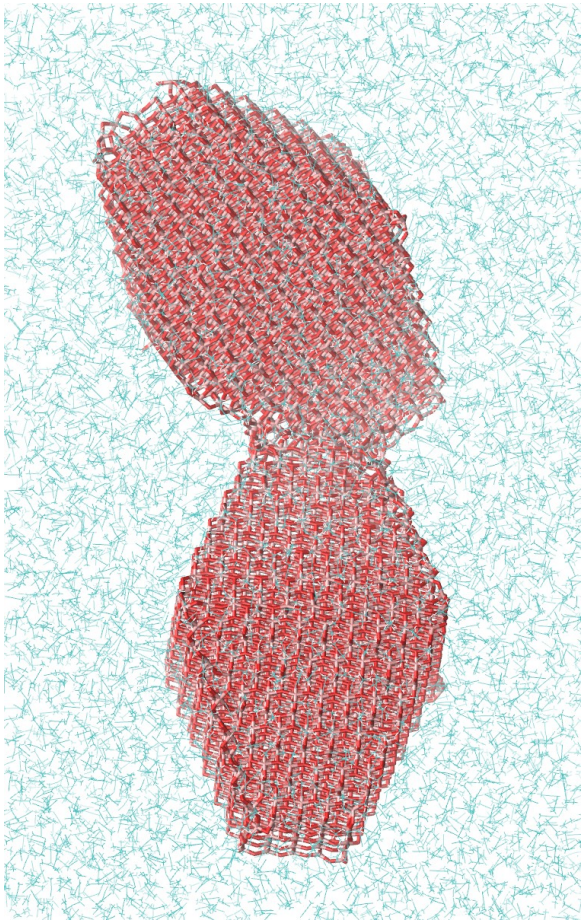


Based on <u>linear scaling MD simulations for 10'000s of atoms</u>, the morphology and properties of the proposed 2D polymer sheets has been investigated

Payam Payamyar, Khaled Kaja, Carlos Ruiz Vargas, Andreas Stemmer, Daniel J. Murray, Carey Johnson, Benjamin T. King, Florian Schiffmann, Joost VandeVondele, Alois Renn, Paola Ceroni, Andri Schütz, Lay-Theng Lee, Zhikun Zheng, Junji Sakamoto, A. Dieter Schlüter, Accepted in ADVANCED MATERIALS (2013).

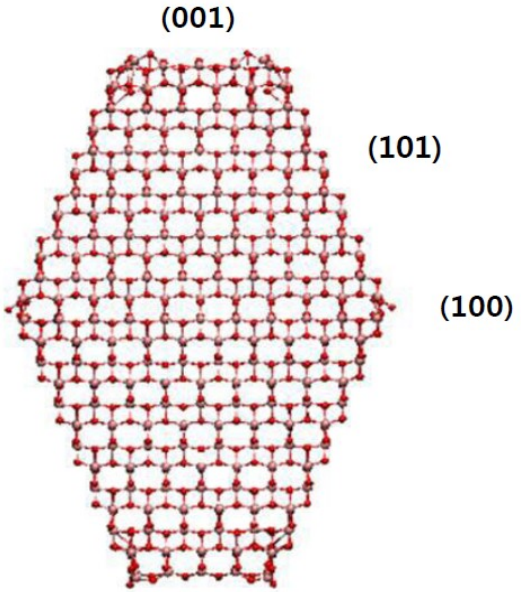
TiO_2 nanocrystals

Model for aggregation

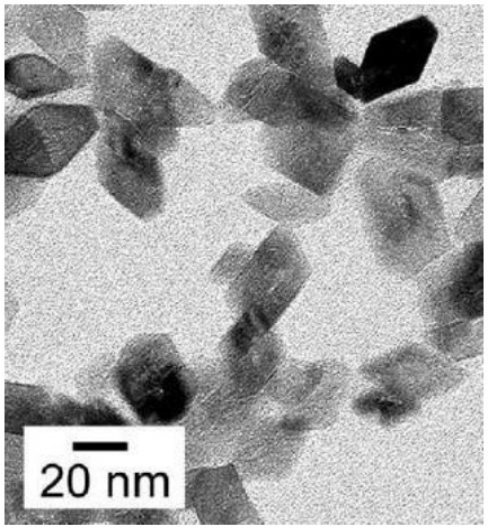


~80'000 atoms with DFT

Models for size effects



Experiment

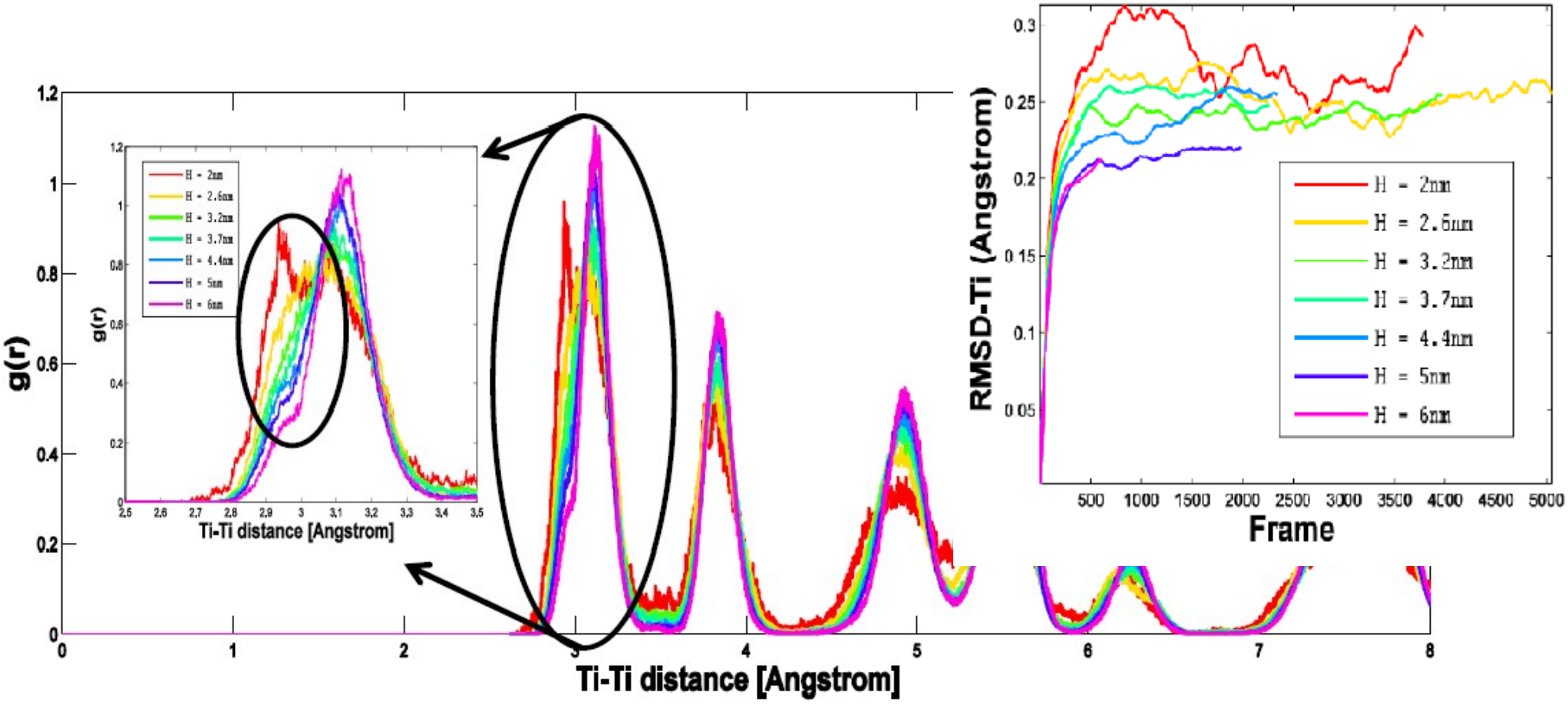


Sizes = 2.0, 2.6, 3.2, 3.7, 4.4, 5.0 and 6.0 nm
(solvent not shown)

Calculations on these models are not possible with traditional DFT approaches

Geometric properties: surface effects

MD based relaxation



Obtaining good structural models is key for large scale simulations!

Sign matrix iterations

The density matrix (P) is function of H

$$P = \frac{1}{2}(I - \mathrm{sign}(S^{-1}H - \mu I))S^{-1}.$$

A simple iterative scheme (Newton-Schultz) gives sign(A):

$$X_{n+1} = \frac{1}{2}X_n(3I - X_n^2).$$

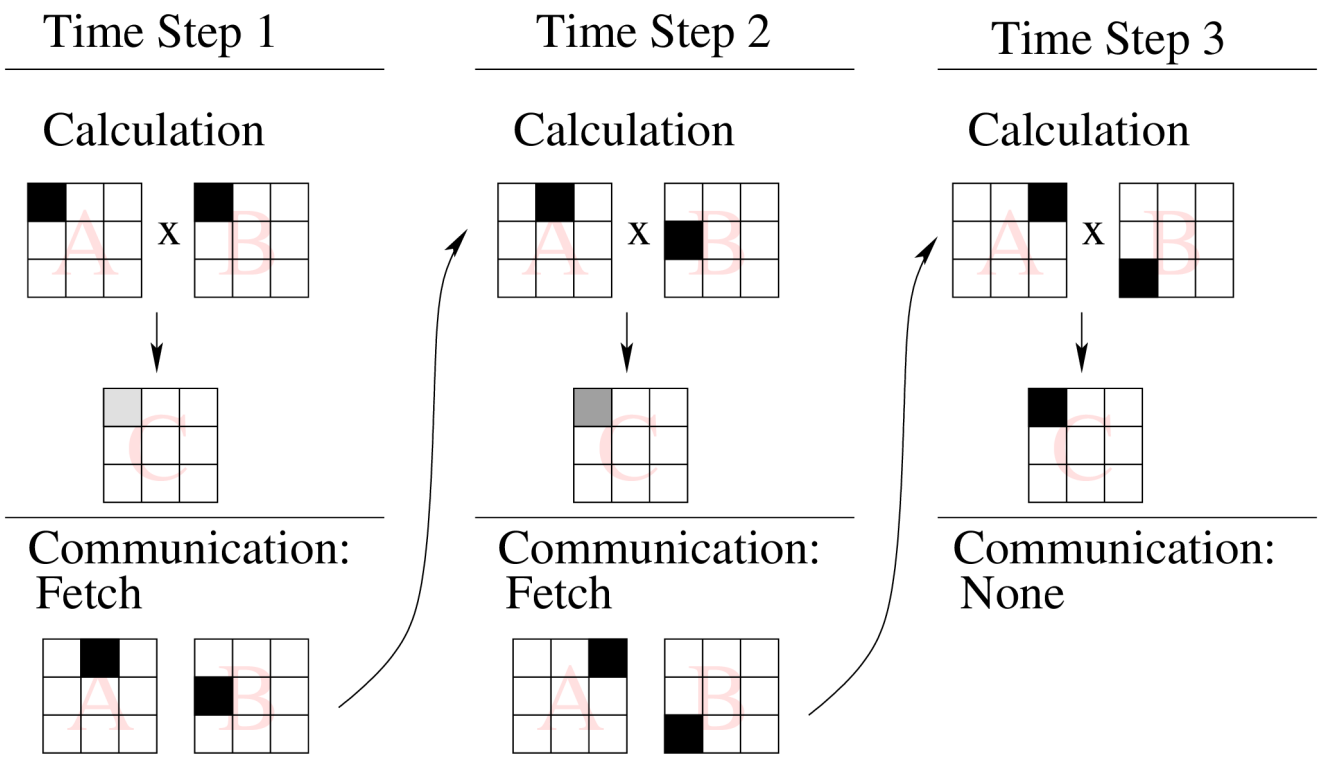
Using only sparse matrix matrix multiplies (not SPMV!)
linear scaling can be obtained

A dedicated sparse matrix multiply library is extremely important
This library has been ported to GPUs

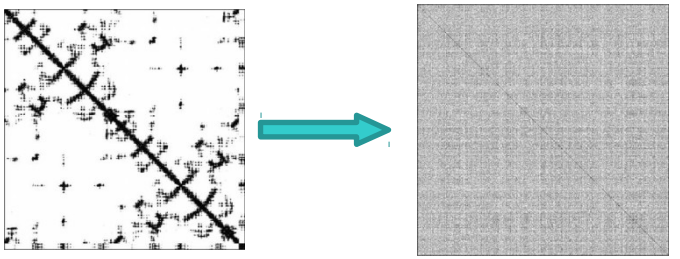
DBCSR: a sparse matrix library

Distributed Blocked Compressed Sparse Row
Distributed Blocked Cannon Sparse Recursive

Optimized for the science case: 10000s of non-zeros per row.
The dense limit as important as the sparse limit.

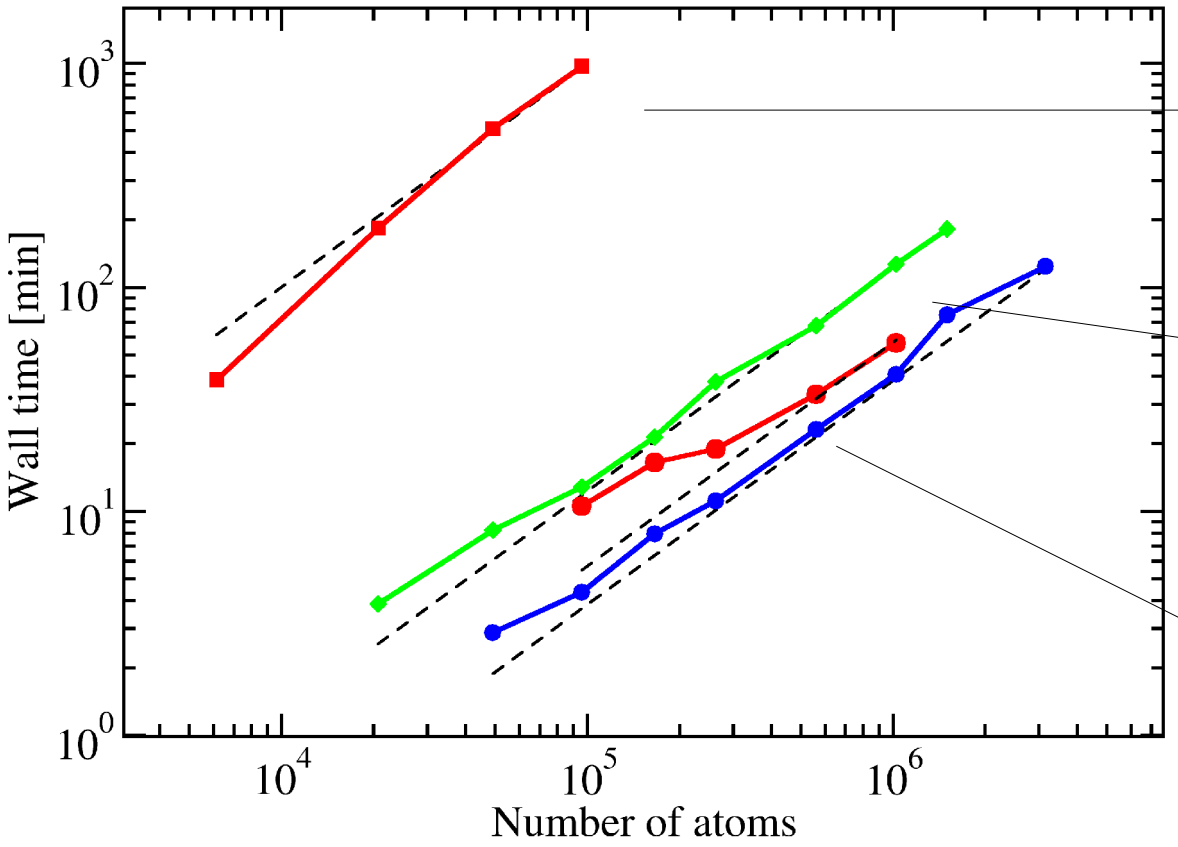


Cannon style communication on a homogenized matrix for strong scaling



Borstnik et al. : submitted

Millions of atoms in the condensed phase



Accurate basis sets, DFT
46656 cores

The electronic structure
$O(10^6)$ atoms in < 2 hours

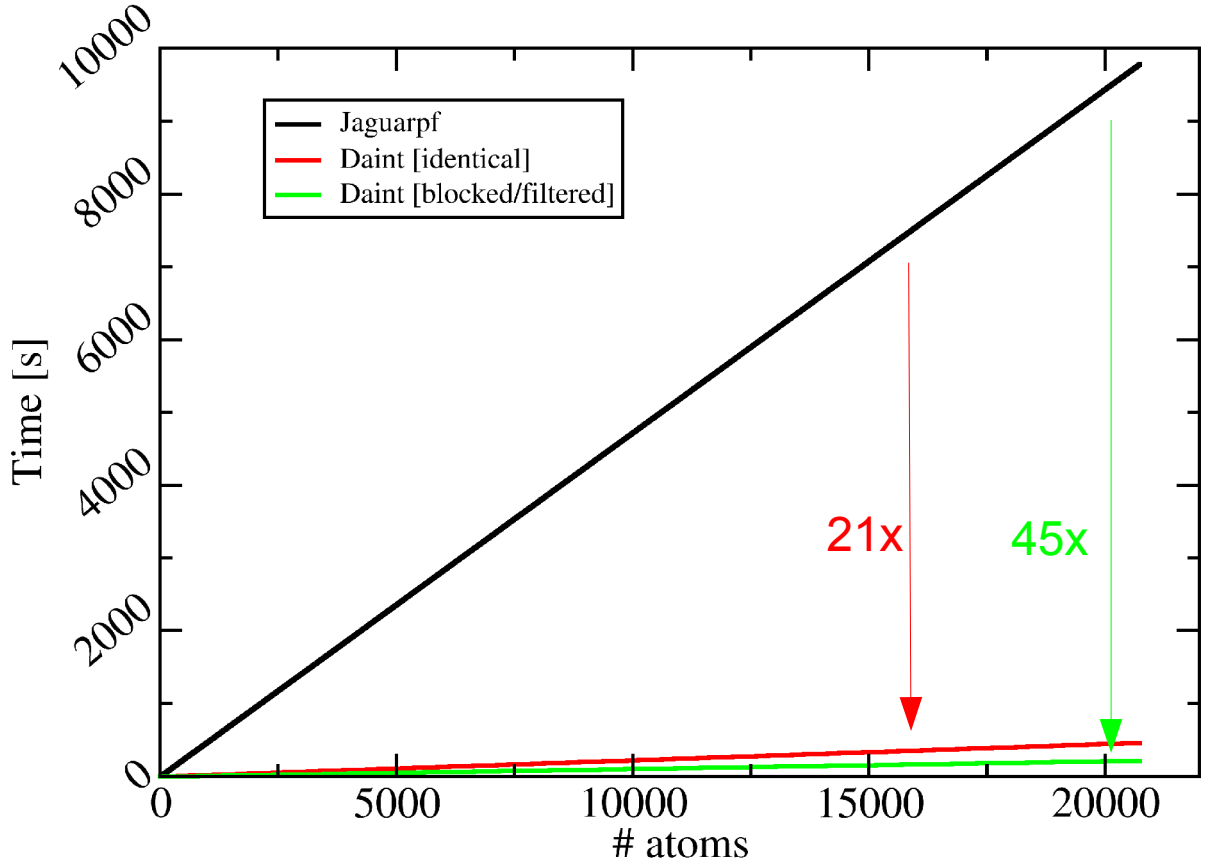
Minimal basis sets:
DFT, NDDO, DFTB
9216 cores

Bulk liquid water. Dashed lines represent ideal linear scaling.

VandeVondele, Borstnik, Hutter, JCTC, DOI: 10.1021/ct200897x

'Historical' comparison

1) Run on Jaguarpf (XT5, 2011-01-01), 3888 nodes (12 cores)
2) Run on Daint (XC30, 2013-11-17), 3844 nodes (8 cores + 1 GPU)



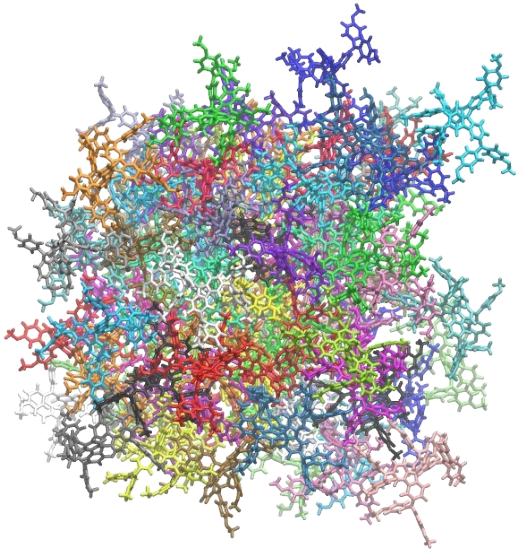
Testcase 'H2O-dft-ls-orig' : 20'000 atoms

Original run published in : VandeVondele, Borstnik, Hutter, JCTC, 2012

Hybrid Daint vs dual SB Daint

Daint: XC30, 8 cores + 1 GPU / per node, ~5200 nodes
Fastest computer in Europe.

Three science benchmarks:
various block sizes, CPU loads, and communication.



Amorph... a hole conducting solar cell material

	Amorph	H2O	TiO2
2 SB	372	275	446
1 SB + 1K20X	272	187	263
Ratio	1.37	1.47	1.70
GPU flop %	92	99	88

Balanced

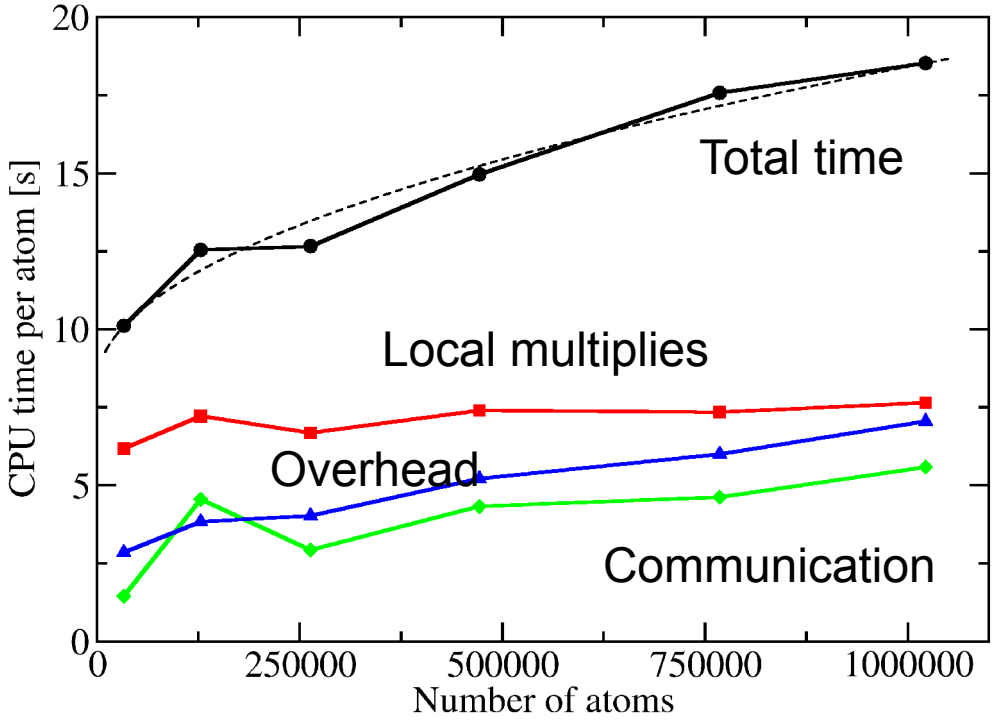
Comm. limited

Small blocks

Canonical benchmarks on 169 nodes
(slightly old code, in particular 2 SB)

Schütt et al. : submitted

Towards O(1) :
constant walltime with proportional resources



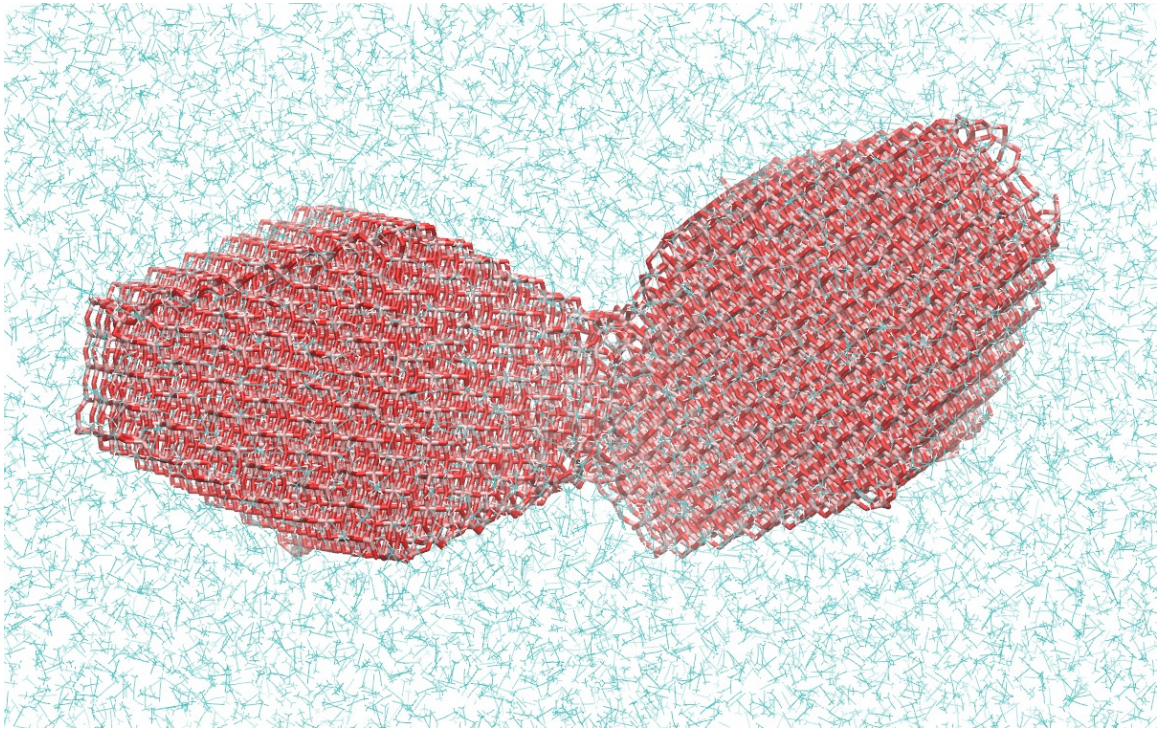
Stringent test:
- Small blocks, large overhead
- Very sparse matrices
- Running with 200 atoms / MPI task

Local multiplies constant (OK!).

Overhead & Communication
- Grows with sqrt(N)
- Needs a replacement for Cannon

Work is underway to replace the Cannon algorithm with something new!
Retain the sqrt(N) max comm, yield constant comm in the limit.

Full system science case



80'000 atoms DFT, high accuracy settings
Aggregated nanoparticles in explicit solution
Relevant for 3rd generation solar cells

Matrix dims ~ 772868 x 772868
Threshold ~1E-6
% non-zero ~ 4%
SCF steps ~ 50
multiplies needed ~ 2000

Dense flops needed =
184661334367982412800O

Actual flops needed =
84992840373629580 2

Sparsity boost = 2172x

GPU flop % = 99.4

Time on 5184 nodes = 6264s

Sustained actual flops = 0.13 PF

Sustained dense flops = 294.7 PF

Open Post Doc position : DBCSR & LS development

A 2yr PDRA position is available in my group (early 2014),
contact Joost.VandeVondele@mat.ethz.ch

Goals:

- Introduce the 'next generation' communication scheme
  Truly reach O(1), gain 10x speedups in applications
- Release DBCSR as a standalone library
  Make this powerful library available for the community
- Establish & enhance linear scaling methods by applications

Requirements:

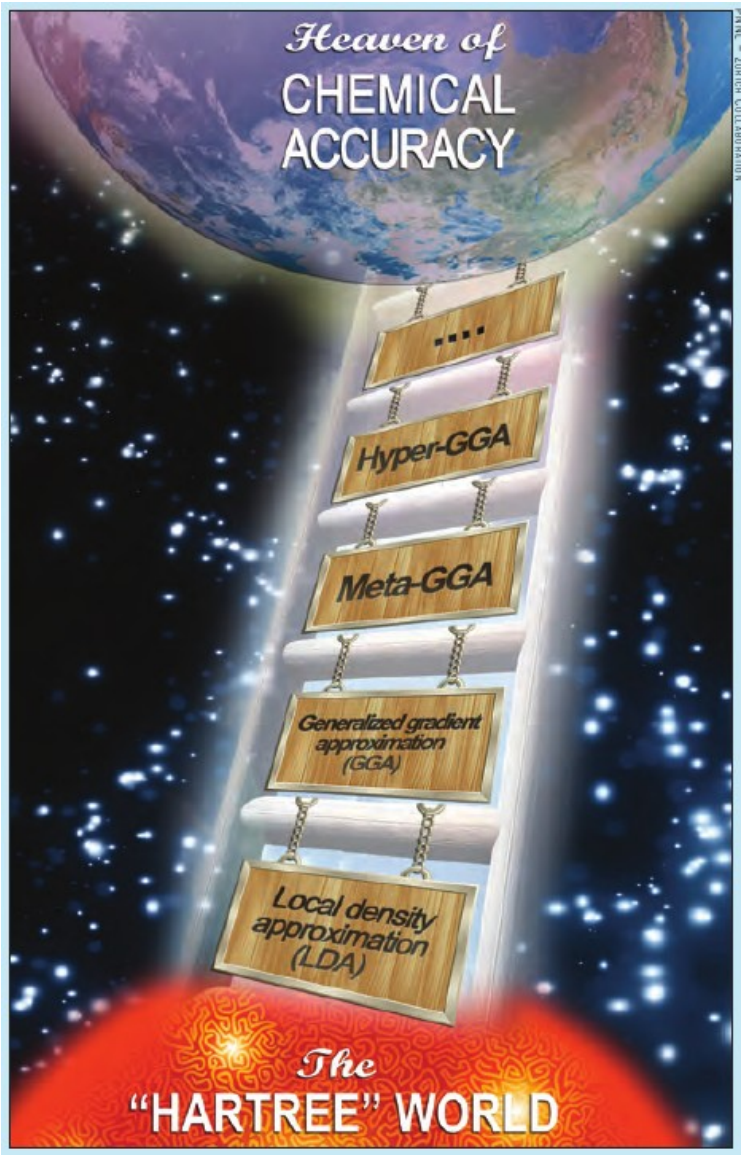
Algorithmic & computationally strong developer with physics/chemistry knowledge, willing to engage with the community and drive the project towards a public release.

Benefits:

Enjoy a Swiss lifestyle at a top institute, embedded in an active group of CP2K developers, access to Europe's largest supercomputer.

Hybrid functionals

Exchange & correlation functionals



$$V_s(\vec{r}) = V(\vec{r}) + \int \frac{e^2 n_s(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \mathrm{d}^3 r' + V_{\mathrm{XC}}[n_s(\vec{r})]$$

Exchange and correlation functionals of improving quality can be constructed by adding new ingredients.

Each rung on the ladder improves accuracy, but also increases complexity

GGA: only relies on the electron density and its gradients (semi-local)

Hyper-GGA: hybrid functionals, includes density and the single particle orbitals directly in a non-local way through Hartree-Fock exchange (HFX)

Double Hybrids: Include MP2/RPA like correlation

Mundy, Kathmann, Rousseau, Schenter, VandeVondele, Hutter, SCIDAC reviews (spring 2010).

Robust Hartree-Fock exchange in the condensed phase

$$E_x^{\mathrm{PBC}} = -\frac{1}{2N_k}\sum_{i,j}\sum_{\mathbf{k},\mathbf{k}'}\int\int \psi_i^{\mathbf{k}}(r_1)\psi_j^{\mathbf{k}'}(r_1)g(|r_1-r_2|)\psi_i^{\mathbf{k}}(r_2)\psi_j^{\mathbf{k}'}(r_2)\,d^3r_1 d^3r_2$$

How to treat this expression, k=k' is only conditionally convergent for g(r)=1/r ?

This 'difficult' point is integrable, but what for a Gamma-point code (k=k'=0) ?

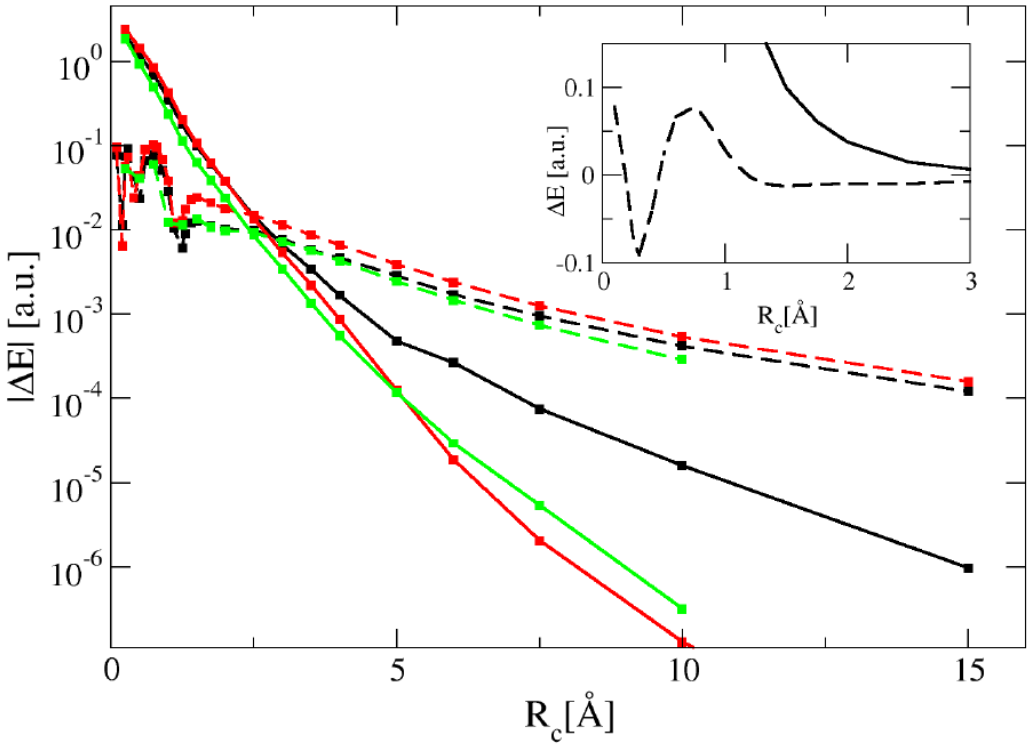
$$g_{TC}(r_{12}) = \begin{cases} \frac{1}{r_{12}}, & r_{12} \leq R_c \\ 0, & r_{12} > R_c \end{cases}$$

Truncated Coulomb:
Needs, Alavi

Avoids spurious self-exchange interactions with images in other cells.
This can be implemented robustly in a simple way.

Guidon, M; Schiffmann, F; Hutter, J; VandeVondele, J. 2008, JCP 128 (21): 214104
Guidon M; Hutter J; VandeVondele J. 2009, JCTC, 5(11): 3010-3021

Exponential Convergence with Truncation Range



A long range (GGA-like) correction, useful for very short truncation, can be derived.

$$\varepsilon_x^{\mathrm{PBE}}(\rho,\nabla\rho) = \varepsilon_x^{\mathrm{LDA}}(\rho)\cdot F_x^{\mathrm{PBE}}(\rho,\nabla\rho) \qquad F_x^{\mathrm{PBE,LRC}}(s) = -\frac{8}{9}\int_{R_c'}^{\infty} y J_x^{\mathrm{PBE}}(s,y)dy$$

$$\text{PBE0-TC-LRC} \quad : \quad E_{\mathrm{xc}}^{\mathrm{PBE0\text{-}TC\text{-}LRC}} = aE_x^{\mathrm{HF,TC}} + aE_x^{\mathrm{PBE,LRC}} + (1-a)E_x^{\mathrm{PBE}} + E_c^{\mathrm{PBE}}$$

$O(N^4) \rightarrow O(N)$

$$E_x^{\mathrm{HF}} = -\frac{1}{2} \sum_{\alpha\beta\gamma\delta} P_{\alpha\beta} P_{\gamma\delta} (\phi_\alpha \phi_\gamma | \phi_\beta \phi_\delta)$$

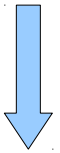
$$(\phi_\alpha \phi_\gamma | \phi_\beta \phi_\delta) = \int d\mathbf{r} d\mathbf{r}' \frac{\phi_\alpha(\mathbf{r}) \phi_\gamma(\mathbf{r}) \phi_\beta(\mathbf{r}') \phi_\delta(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}$$

Based on the fact that for large systems either the integrals
Or the density matrix become zero (to within a threshold eps)

Cauchy-Schwarz screening	$\lvert(ab\lvert cd)\rvert \leq \sqrt{(ab\lvert ab)\lvert(cd\lvert cd)}$	$O(N^2)$
Density matrix screening	$P_{\alpha\beta}$ decays exponentially	O(N)
Operator screening	Operators other than 1/r	O(N)

Auxiliary Density Matrix Methods (ADMM)

For certain density matrices HFX can be computed very efficiently
(e.g. small basis sets or increased sparsity)



Transform an expensive matrix into a cheap one,
use a GGA for estimating the difference

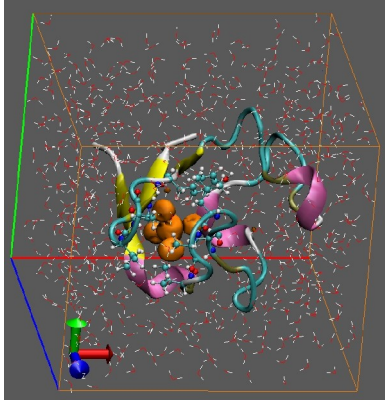
$$\begin{aligned} E_x^{\mathrm{HFX}}[P] &= E_x^{\mathrm{HFX}}[\hat{P}] + \left(E_x^{\mathrm{HFX}}[P] - E_x^{\mathrm{HFX}}[\hat{P}]\right) \\ &\approx E_x^{\mathrm{HFX}}[\hat{P}] + \left(E_x^{\mathrm{DFT}}[P] - E_x^{\mathrm{DFT}}[\hat{P}]\right) \end{aligned}$$

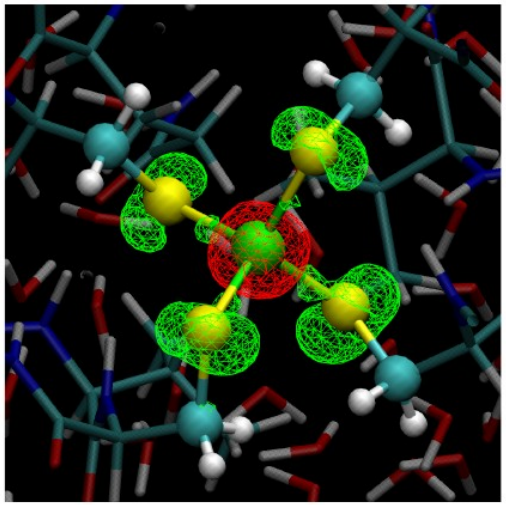
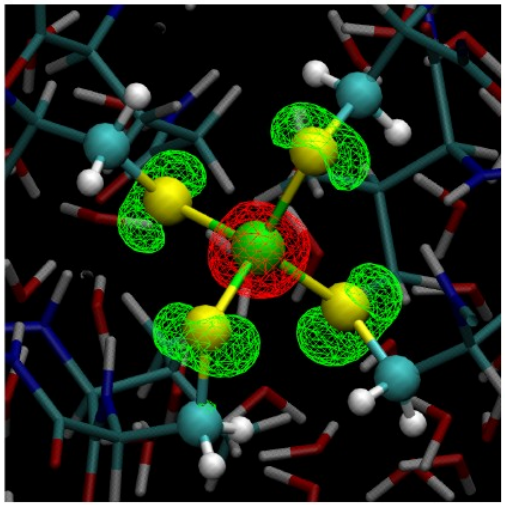
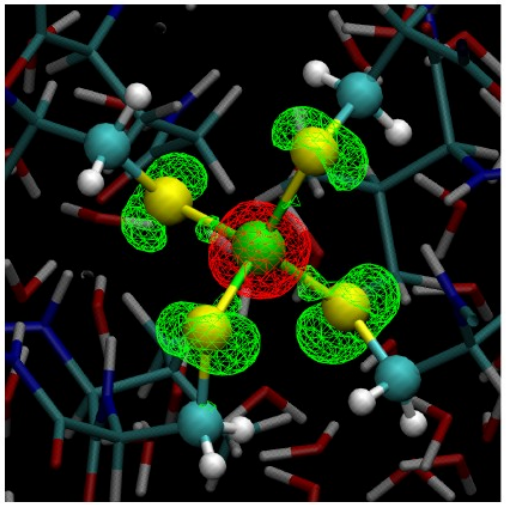
One example: wavefunction fitting, using an auxiliary basis

$$\min_{\tilde{C}} \left[\sum_j \int \left(\psi_j(\mathbf{r}) - \tilde{\psi}_j(\mathbf{r})\right)^2 d\mathbf{r} + \sum_{k,l} \Lambda_{kl} \left(\int \tilde{\psi}_k(\mathbf{r}) \tilde{\psi}_l(\mathbf{r}) d\mathbf{r} - \delta_{kl} \right) \right]$$

Guidon M; Hutter J; VandeVondele J; JCTC 6(8): 2348-2364

ADMM: performance



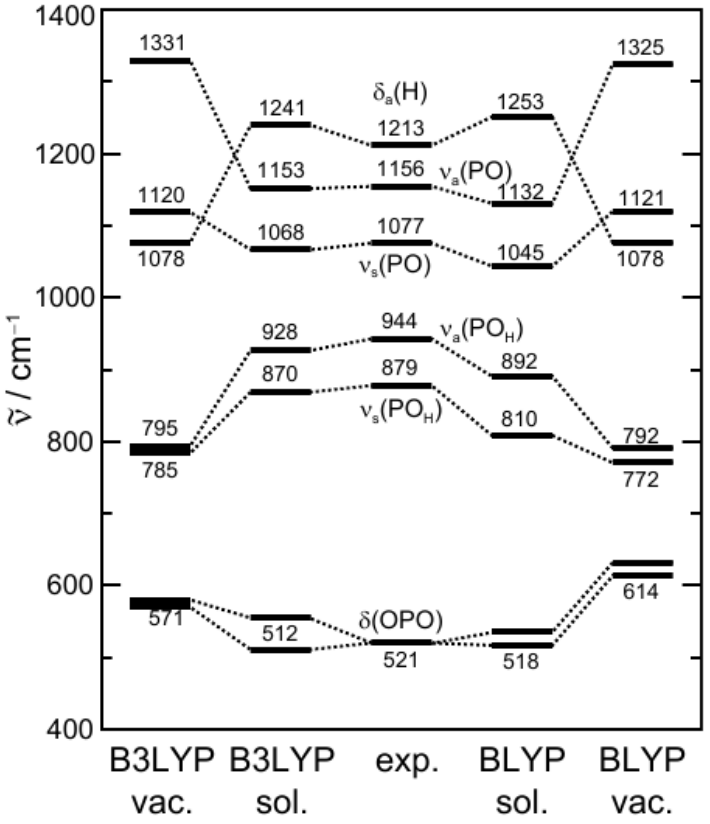
STD DZVP-MOLOPT-SR-GTH	ADMM: MOLOPT/FIT3	ADMM: MOLOPT/EMBED
		
▸ 48'000 cores	▸ 1'152 cores	▸ 1'152 cores
▸ 1. SCF step: 45 min	▸ 1. SCF step: 75 s	▸ 1. SCF step: 80 s
▸ subsequent steps: 25 s	▸ subsequent steps: 25 s	▸ subsequent steps: 25 s
▸ 6.8 TB RAM	▸ 5.2 GB RAM	▸ 5.2 GB RAM

A fully solvated protein computed within minutes using hybrid functionals

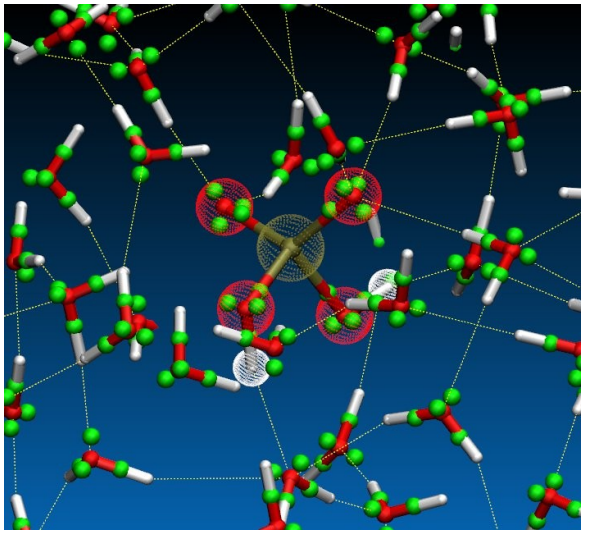
Guidon M; Hutter J; VandeVondele J; JCTC 6(8): 2348-2364

IR spectroscopy from AIMD with hybrid functionals

H_2PO4^-



$H_2PO_4^-$ in explicit water

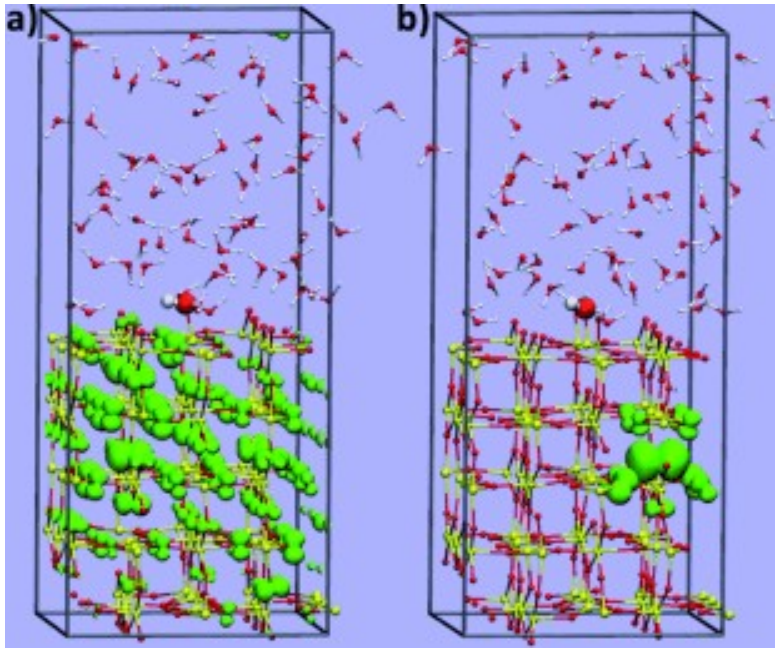


~ 100 ps AIMD / 64 waters

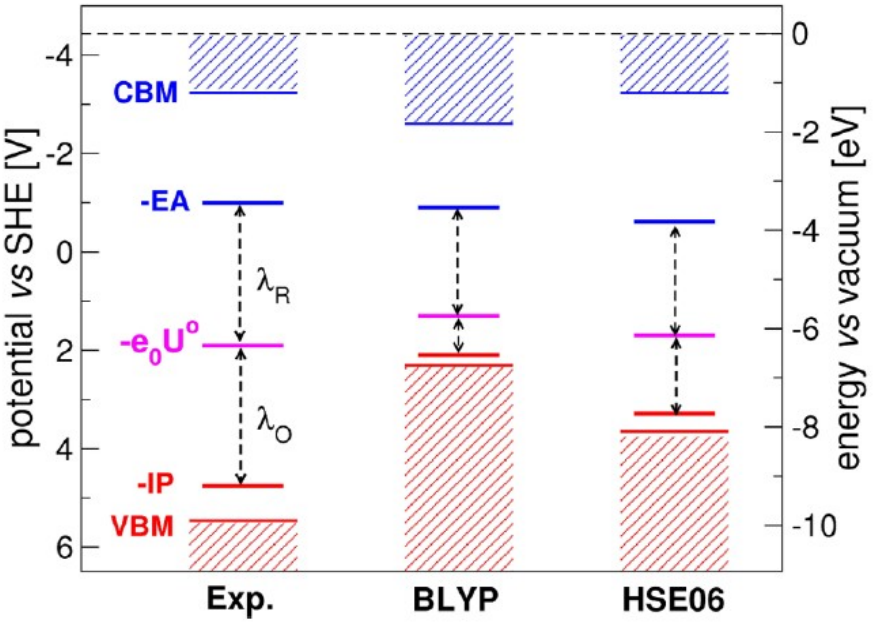
Both Hybrids and GGA capture the main effect of solvation.
Hybrid give 2x more accurate spectra

VandeVondele J; Tröster P; Tavan P; Mathias G; 2012, JPCA 116(10): 2466-2474

Electrochemistry: bulk and interfaces



Hole localization in TiO2.



Redox levels of OH*/OH-

AIMD with hybrid functionals lead to a qualitatively correct description of the interface, and might be essential for electrochemistry.

Adriaanse C; Cheng J; Sulpizi M; Chau V; VandeVondele J; Sprik M; 2012 JPCL 3(23): 3411

Cheng J; Marialore Sulpizi; VandeVondele J; Sprik M; 2012, CHEMCATCHEM 4(5): 636-640

MP2, dRPA, and Double Hybrids

The next rung!

Møller-Plesset Perturbation Theory

The energy:

$$E^{(2)} = -\sum_{ij,ab}^{occ,vir} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j}$$

Two electron integrals over canonical molecular orbitals (MO):

$$(ia|jb) = \int\int \psi_i(\vec{r}_1)\psi_a(\vec{r}_1)\frac{1}{\vec{r}_{12}}\psi_j(\vec{r}_2)\psi_b(\vec{r}_2)d\vec{r}_1 d\vec{r}_2$$

The four index transformation, going from AO to MO

$$(ia|jb) = \sum_{\mu\nu\lambda\sigma} (\mu\nu|\lambda\sigma)C_{\mu i}C_{\nu a}C_{\lambda j}C_{\sigma b}$$

MP2 is relatively expensive O(N^5), not easy to parallelize efficiently, and somewhat tricky in the condensed phase for an AO code.

GPW-MP2

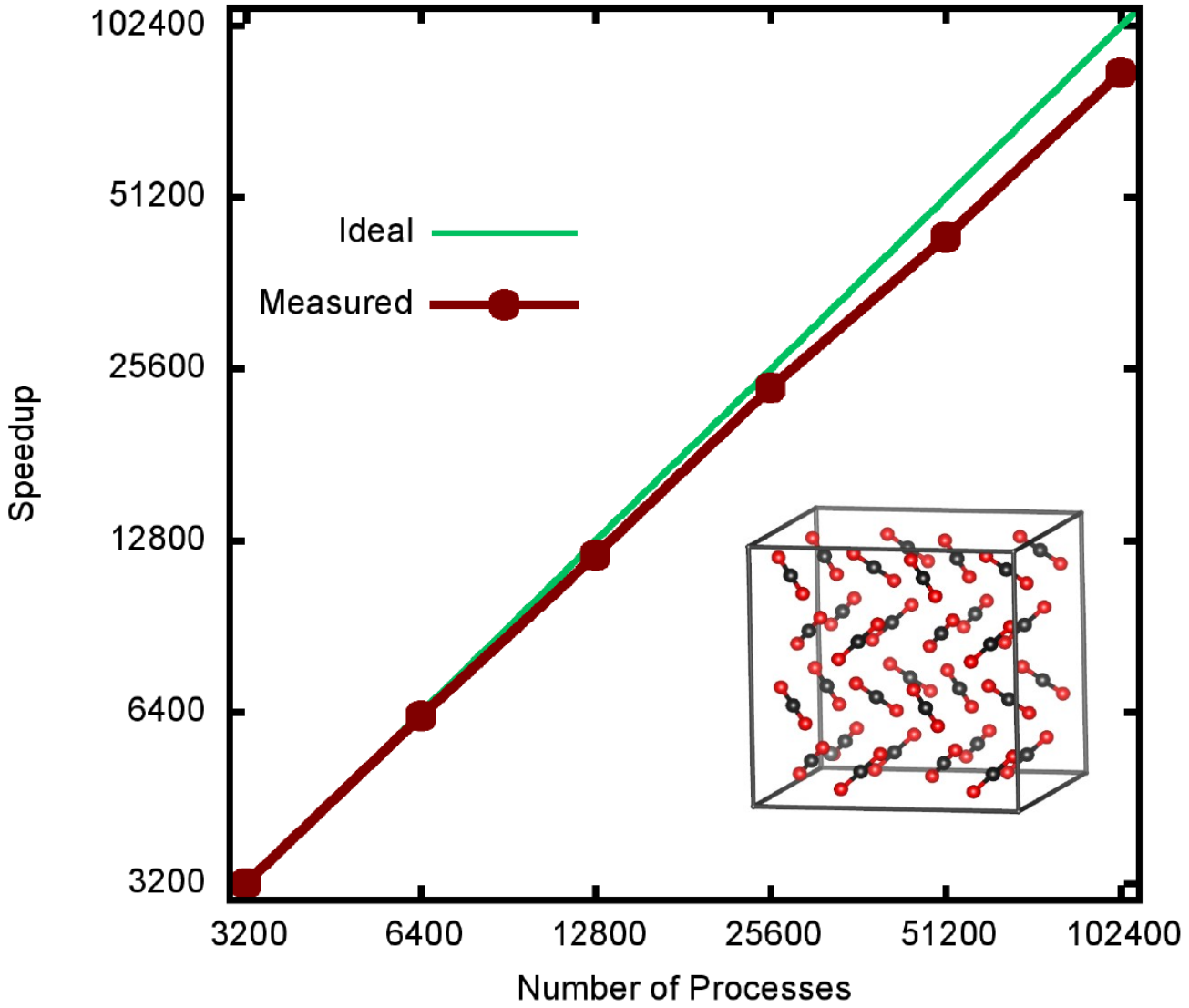
A Gaussian and plane waves approach to MP2

$$
\begin{aligned}
(ia|\lambda\sigma) &= \int\int \psi_i(\vec{r}_1)\psi_a(\vec{r}_1)\frac{1}{\vec{r}_{12}}\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_1 d\vec{r}_2 \\
&= \int\left[\int \frac{\psi_i(\vec{r}_1)\psi_a(\vec{r}_1)}{\vec{r}_{12}}d\vec{r}_1\right]\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_2 \\
&= \int\left[\int \frac{\rho^{ia}(\vec{r}_1)}{\vec{r}_{12}}d\vec{r}_1\right]\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_2 \\
&= \int v^{ia}(\vec{r}_2)\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_2
\end{aligned}
$$

Directly obtain half transformed integrals using the GPW approach.

Del Ben M, Hutter J, VandeVondele J, 2012, JCTC 8(11): 4177–4188

Parallel efficiency

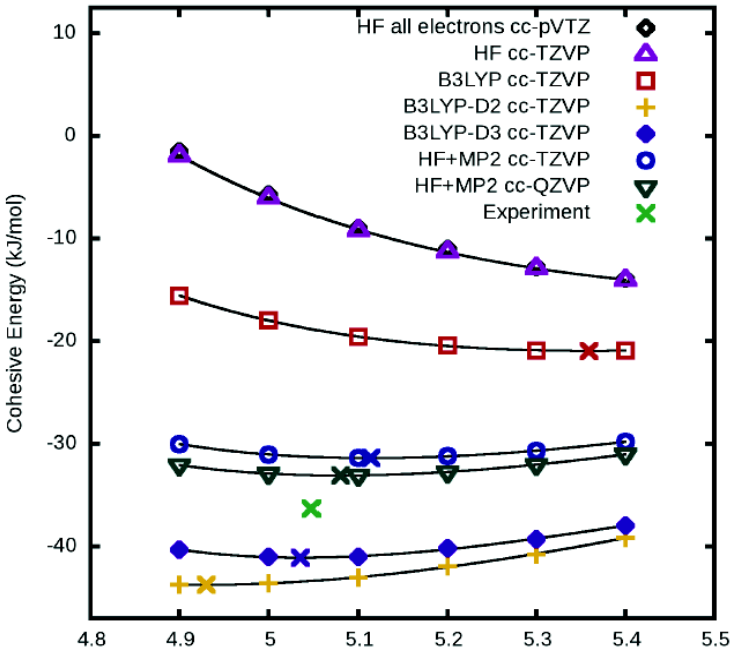


CO_2 crystal (32 molecules)
cc-QZVP basis (5184 BF)
Canonical GPW-MP2 calculation

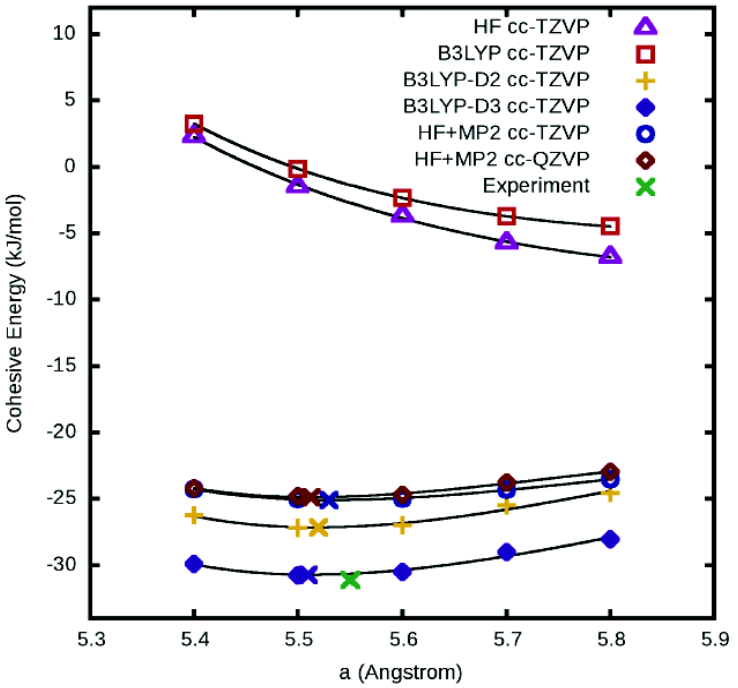
MP2 time: 9min.
parallel efficiency: 80%
On 102400 cores

NH_3 and CO_2 crystals: cohesive energy and lattice constant

NH_3



CO_2



	NH_3		CO_2	
	a	E_{coh}	a	E_{coh}
HF	5.98	-15.7	6.05	-5.00
B3LYP	5.36	-20.9	6.20	-8.20
B3LYP-D2	4.93	-43.7	5.52	-27.2
B3LYP-D3	5.04	-41.1	5.51	-30.7
HF+MP2	5.11	-31.4	5.53	-25.1
HF+MP2 [1]	5.08	-33.1	5.51	-24.9
Exp.	5.048	-36.3	5.55	-31.1

MP2 or DFT+D absolutely needed

RI-GPW

Use an auxiliary Gaussian basis to introduce a resolution of identitiy (RI) approximation:

$$(ia|jb) \approx (ia|jb)_{RI} = \sum_{lm} (ia|l)(l|m)^{-1}(m|jb)$$

- 4 center integrals become 3 center integrals
- Efficiently computable with GPW techniques
- 10 times faster than canonical MP2 (small systems)
- Highly accurate calculations for well selected aux. basis sets
- Allows for O(N^4) implementations (LT-SOS-RI-MP2 & RI-FI-dRPA)

The basis for applications !

Del Ben, Hutter, VandeVondele, 2013, JCTC 9(6): 2654-2671

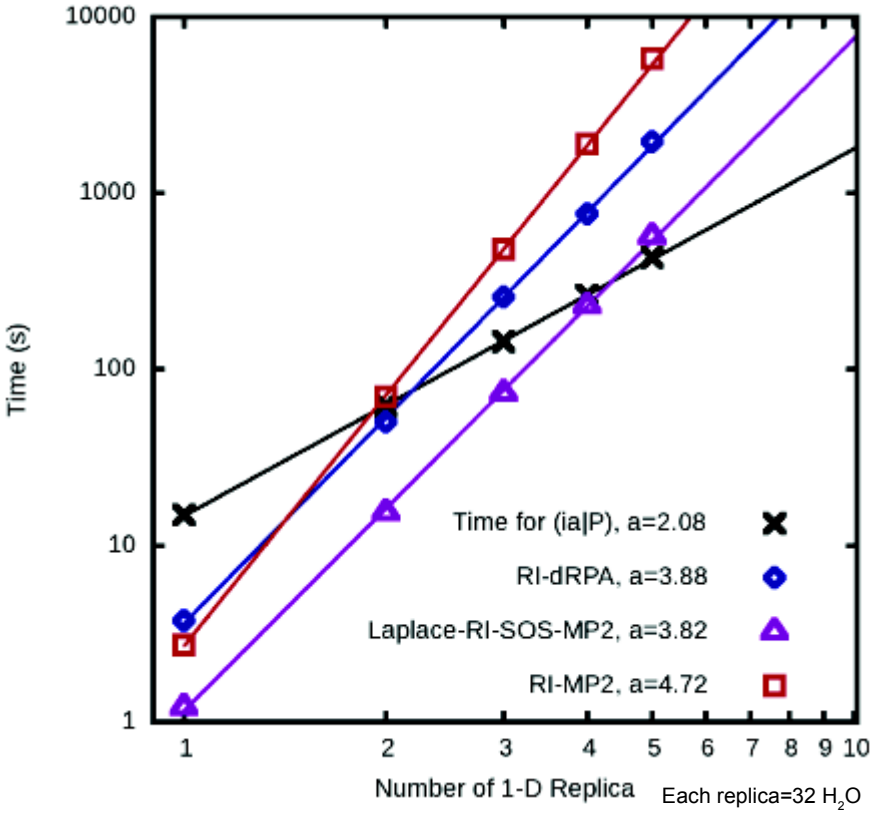
O(N^4) : LT-SOS-MP2 & RI-dRPA

Modern double hybrid functionals:

$$E_{OS}^{(2)} = -\int_0^{\infty} dt \sum_{ia}\sum_{jb} (ia|jb)^2 e^{-t\Delta_{ij}^{ab}}.$$

Efficient RPA:

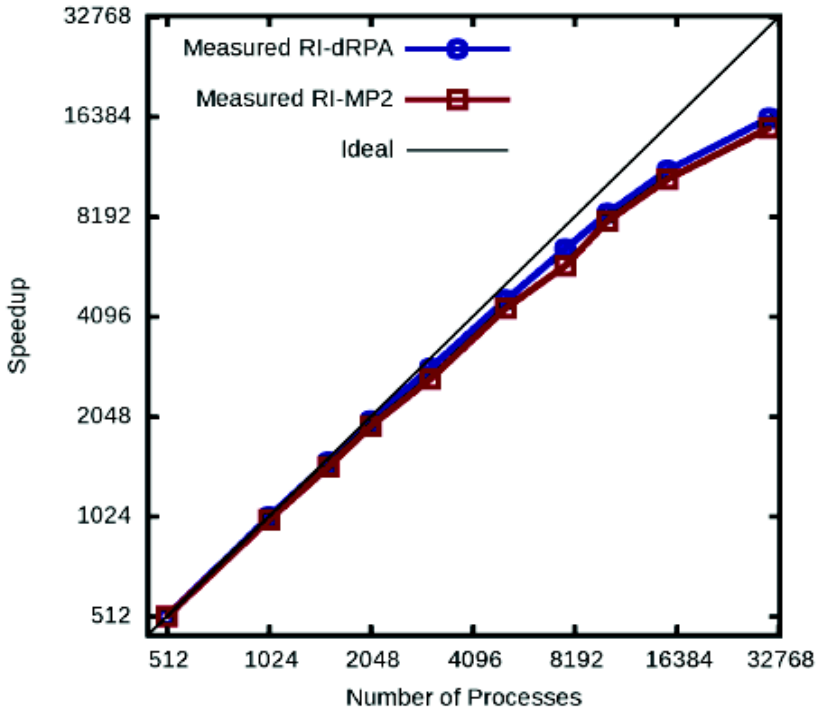
$$E_c^{\text{RI-dRPA}} = \frac{1}{2}\int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \text{Tr}(\ln(\mathbf{1}+\mathbf{Q}(\omega)) - \mathbf{Q}(\omega)).$$



Numerical quadrature with a finite number of points, and introduction of the RI approximation leads to an O(N^4) scheme.

⟶ An O(N^4) scheme allows for 100-750 atoms in < 1h !

RI Performance: scalability & GPUs



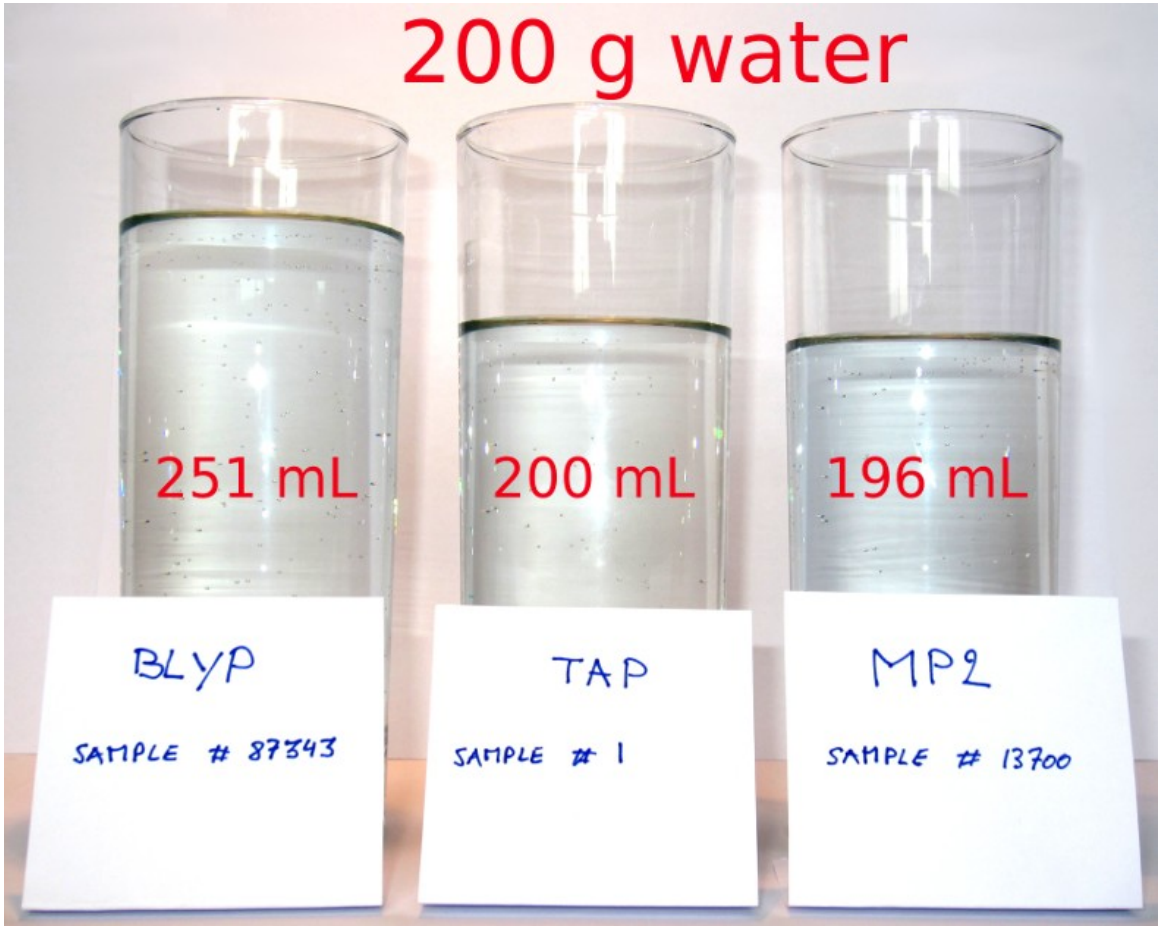
Bulk 64 water, 97s at
30720 cores (50% efficiency, XE6)

	RI-MP2				
	t_{mul}	$t^{\mathrm{H}}_{\mathrm{mul}}$	$\frac{t_{\mathrm{mul}}}{t^{\mathrm{H}}_{\mathrm{mul}}}$	$t^{\mathrm{H}}_{\mathrm{tot}}$	$\frac{t_{\mathrm{tot}}}{t^{\mathrm{H}}_{\mathrm{tot}}}$
U	1.37	0.40	3.4	1.49	1.6
D	1.80	0.61	3.0	2.10	1.6
FA	1.72	0.58	3.0	1.70	1.6
64 H_2O	5.00	1.47	3.4	4.19	1.8
B	7.14	1.93	3.7	5.16	2.0
PD	10.5	2.90	3.6	6.20	2.2
SA	11.2	2.62	4.3	6.14	2.4
CT	13.7	3.78	3.6	7.44	2.3

XK7, speedups without / with GPU

Excellent speedup, good benefit from GPU → ideally suited for XC30 & XC30-ACC

Sampling liquids @ MP2



64 water molecules, cc-pVTZ, 14000 nested MC moves, NPT ensemble

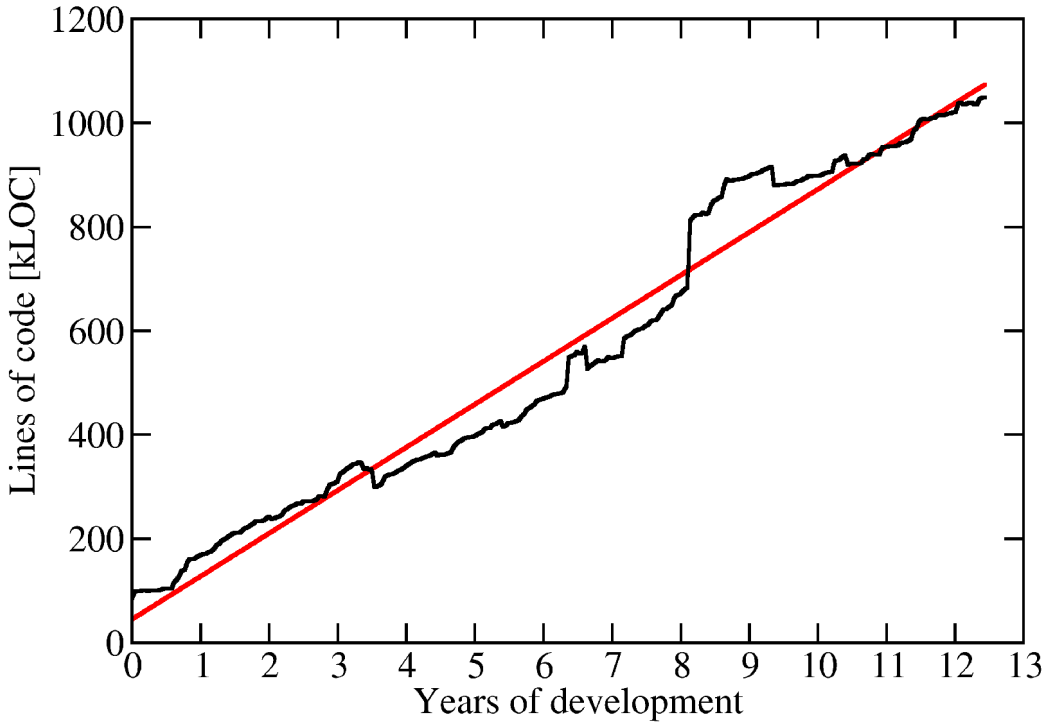
Del Ben M; Schönherr M; Hutter J, VandeVondele J; 2013, JPC L, 4, 3753-3759

Looking to the future...

CP2K: history

Public code repository started in 2001-06-27

Well over 100 years of developer effort
(PhD level, usually piggybacked from science proposals)!



CP2K is an evolving code...
In two years, 436kLOC diff

Significant new functionality requires sustained efforts:

LS MD :
started 2001 → possible 2012
Beyond GGA MD :
started 2006 → MP2 possible 2014

Code base grows at 83000 lines of code per year = 230 LOC/day

CP2K: global team ?

With contributions from: Axel Kohlmeyer Barbara Kirchner Ben Slater Chris Mundy Fawzi Mohamed Florian Schiffmann Gerald Lippert Gloria Tabacchi Greg Schenter Harald Forbert Iain Bethune William Kuo Ken Bagchi Sundaram Balasubramanian Jochen Schmidt Jens Thar Jürg Hutter Matthias Krack Matt Watkins Marcella Iannuzzi Manuel Guidon Matthew McGrath Thomas Chassaing Thomas Heine Thomas Kuehne Teodoro Laino Urban Borstnik Joost VandeVondele Benjamin Levine Luca Bellucci Ari Seitsonen Louis Vanduyfhuys Mathieu Salanne Michele Ceriotti Lukasz Walewski Michael Steinlechner Rodolphe Vuilleumier Sebastiano Caravati Valery Weber Kevin Stratford Toon Verstraelen Marcel Baer Alessandro Laio Stefan Goedecker Luigi Genovese Thierry Deutsch Dieter Gleich Reinout Declerck Kurt Baarman Mauro DelBen Mandes Schönherr Yannik Misteli Fabio Sterpone Gerd Berghold Pietro Ballone Walter Silvestri Pekka Manninen Francois-Xavier Coudert Christiane Pousa Michele Parrinello Michiel Sprik Ilja Siepmann Michel Waroquier … and more

Yet, most of the development is concentrated in just a few groups

Current challenge

CP2K might become a leading free software tool for atomistic simulation with a great impact in science and engineering, but must follow a clear strategy for long term survival & sustainability.

Strategy:
a community project

- Establish an engaged community
  - Broaden the community of developers
    - Need more 'leading groups'
    - Increase the 'bus factor'
    - Increase features and skill-set
  - Broaden the community of users
    - Facilitate access to and use of cp2k
    - Improve Internet presence (interactions & self-help)
    - Provide/support tutorials & workshops
  - Maintain 'computational science' eco-system
    - Interact with vendors & computer centers
    - Improve tool chains (gcc, vmd, …) and distros
  - Facilitate contributing back
    - Wiki, editable manual, …
  - Connect to other community projects
    - Technical (libs)
    - Scientific

Strategy: science driven

- Demonstrate excellence in science applications
  - Dare to tackle important, difficult and new problems
  - Best-of-class quality
    - Critically employ the best methods
    - Know the underlying theory and methods
  - Facilitate reproducing simulations
    - Wiki Tutorials
    - Provide inputs as supplementary materials

Strategy: methodologically leading

- Excellence is method developments
  - Incorporate best methods
  - Work on the next generation methods
  - Build tools, not toys
  - Eat your own dog food, applications not benchmarks
  - Robust, standard conforming, open.
  - Well tested.
  - Modular, simple, efficient
  - Embrace software engineering & refactoring

Strategy: internationally visible

- Visible
  - Quantify and publicly document progress & impact
    - Key for funding
  - Give credit in publications & talks
    - Citations of method papers matter
  - Public relations & lobby work
    - Give keynote talks ;-)
  - Attract major funding (EPSRC, DOE, HP2C, PASC, ...)
  - Large CPU time allocations (PRACE, INCITE, ...)
  - European/International networks
  - Internet presence

Strategy

A community project that is
science driven,
methodologically leading,
and internationally visible!

Conclusions

- CP2K has become established as a successful tool with high impact
- New developments:
  - Full linear scaling is possible
  - Hybrids based on ADMM can be used routinely
  - Next rung : MP2 / DH / RPA are now available
- CP2K's next challenge is long term sustainability

Acknowledgements

ETH
Ole Schütt
Florian Schiffmann
Martin Häufel
Hossein Bani Hashemian
Clelia Spreafico
Florian Thöle
Oliver Gindele
Cha Jinwoong
Samuel Andermatt

UZH
Juerg Hutter
Mauro Del Ben
Urban Borstnik
Manuel Guidon

Cambridge
Michiel Sprik
Marialore Sulpizi
Jun Cheng

World-wide
CP2K Team

UCL&EPCC
Ben Slater
Iain Bethune
Matt Wattkins

EU FP7

HP2C/PASC
INCITE
PRACE

UZH
ETH

CSCS
ORNL

You for your attention!