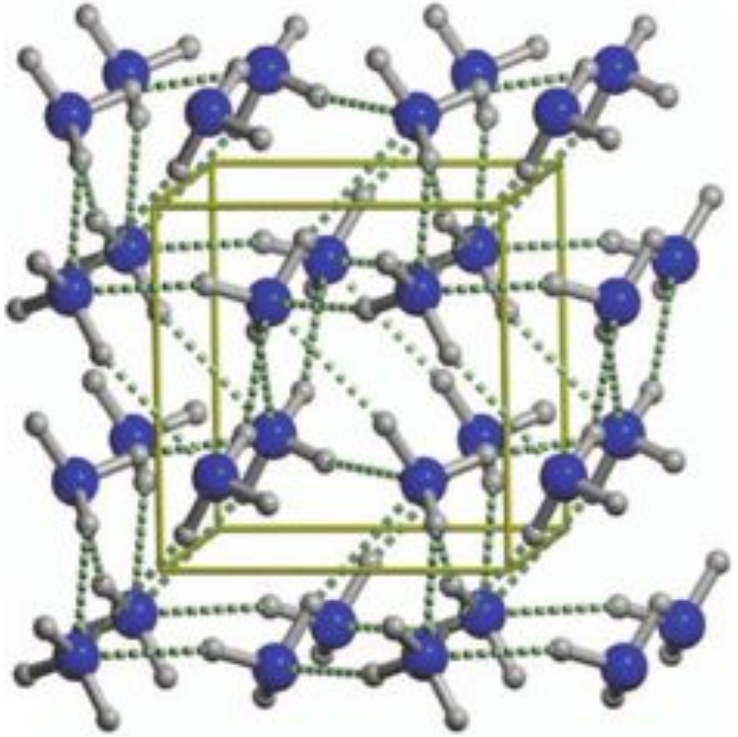
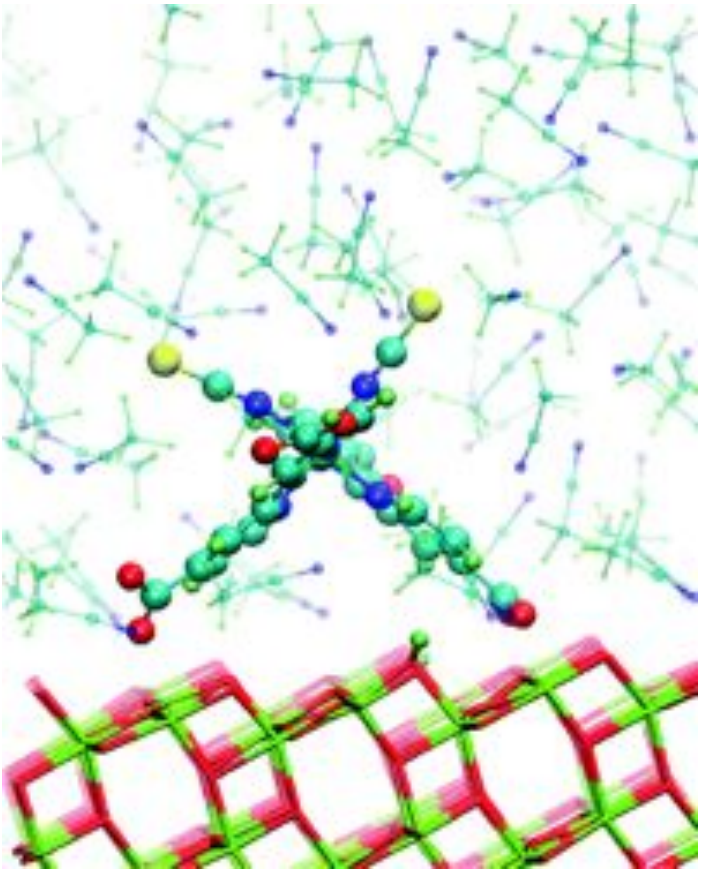
Large scale and hybrid computing with CP2K

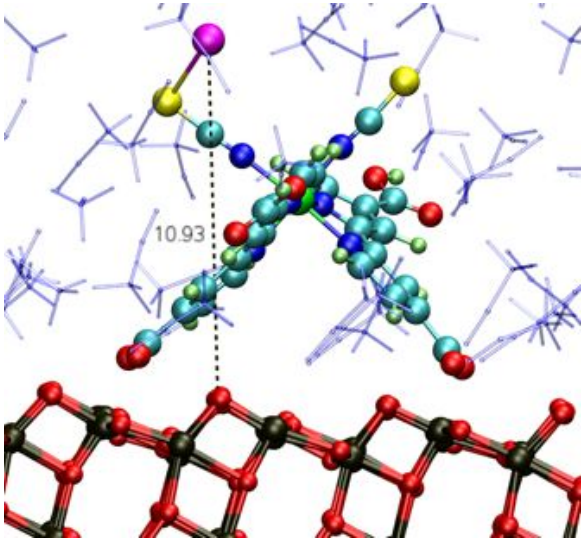


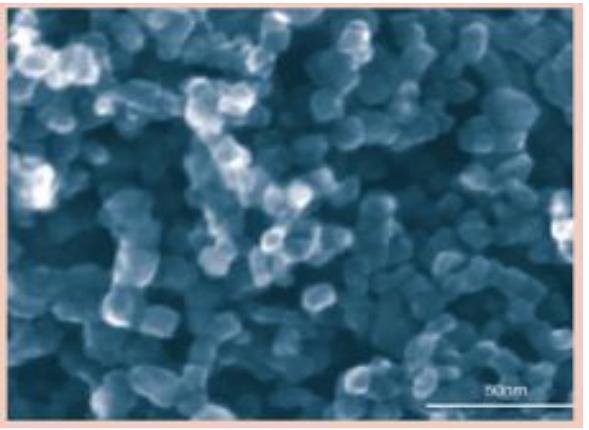


Joost VandeVondele
Nanoscale Simulations, ETH Zurich

Sustainable energy production

A grand challenge for society which will require novel materials engineered with atomistic precision

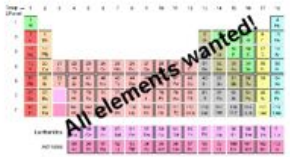


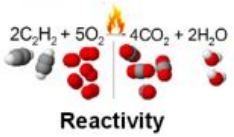


Grätzel, Nature (1991,2001)

Novel simulation tools can contribute towards rational design of complex systems

Modeling complex systems



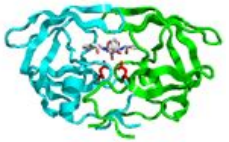


Reactivity

Needs a versatile theory...

Electronic properties

Extreme conditions



Weak interactions

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

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

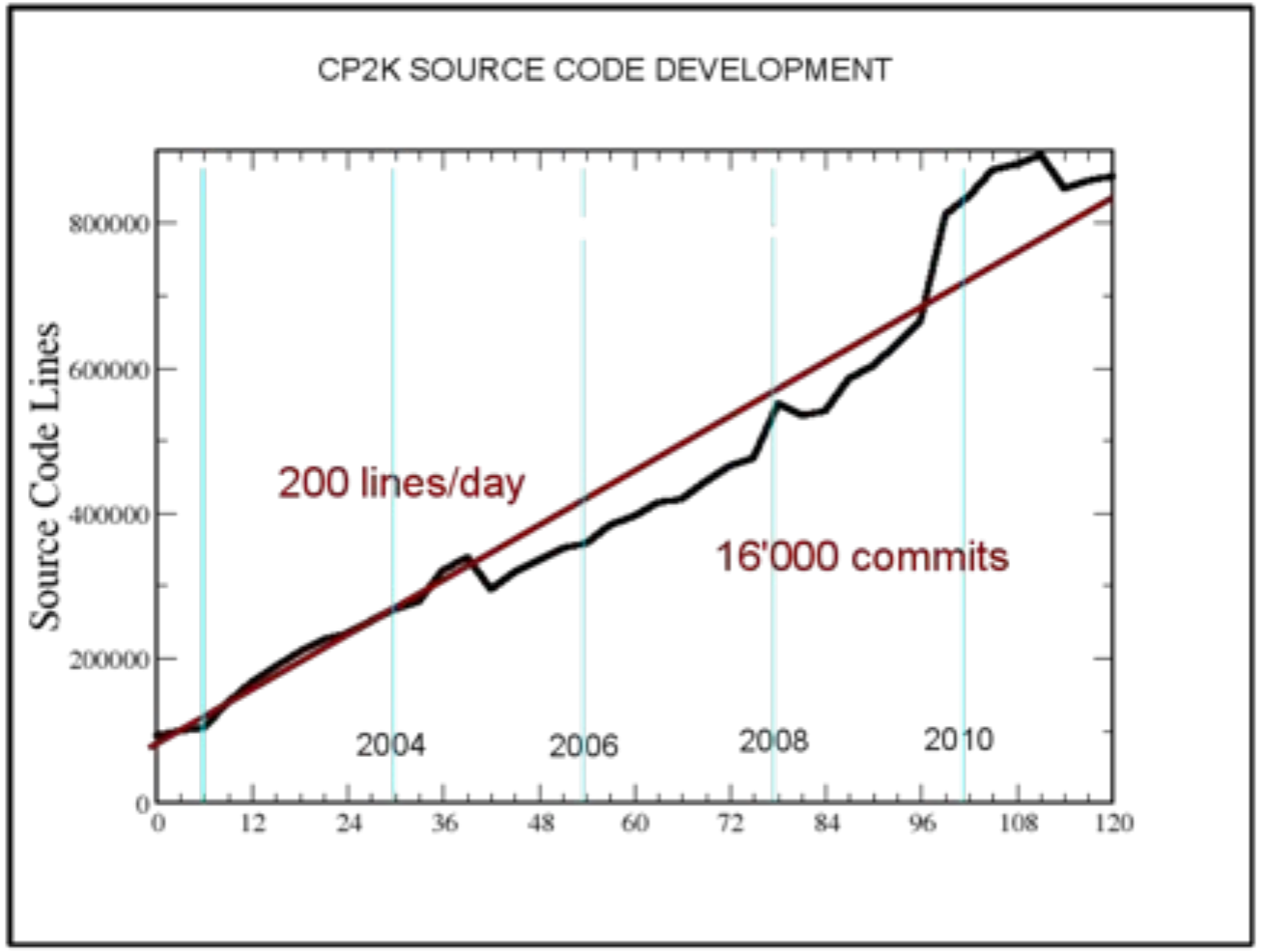
Density functional theory (DFT) & computational methods
(1998, Nobel Prize, Walter Kohn- John A. Pople)

A computationally tractable reference system of non-interacting particles can be used to obtain all properties (in particular the density and energy) of a true electronic system. The required external potential (Vxc) exists but is only known approximately.

CP2K: the swiss army knife of molecular simulation

- A wide variety of models Hamiltonians
  - classical
  - semi-empirical
  - local and non-local DFT
  - 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
  - 900.000 lines of code

What is CP2K: 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

Ten years of development at 200 lines / day = O(1M) SLOC, with a large base of contributors

Computational kernels

- Dense linear algebra:
  - Cholesky
  - Matrix-Matrix multiplication
  - Diagonalization
- Sparse linear algebra:
  - Matrix Matrix
  - Matrix vector
- Regular grids:
  - multigrids and FFT
- Time integration:
  - MD simulation
- Chemical kernels:
  - Grid collocation & integration
  - four center integrals
  - MP2

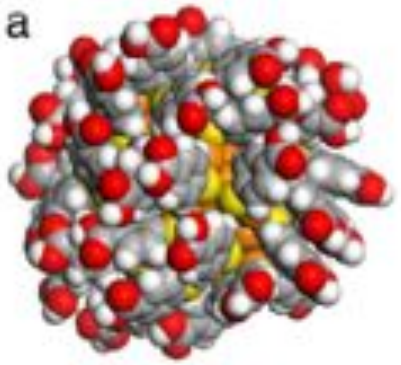
A single kernel rarely dominates
Scaling is O(N)...O(N**5)

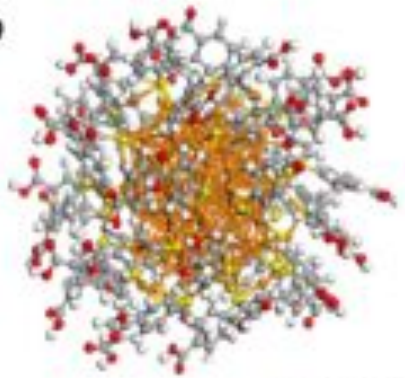
CP2K: science (I)

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

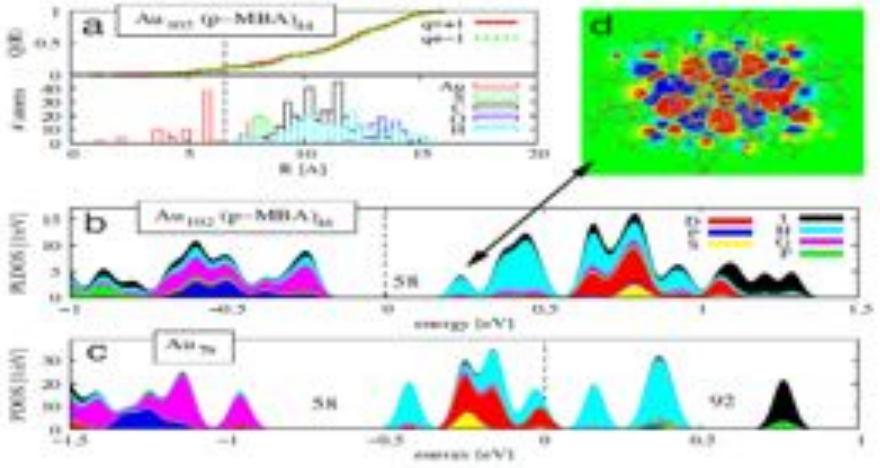
Michael Walter[1], Jaakko Akola[1†], Olga Lopez-Acevedo[1], Pablo D. Jadzinsky[§¶], Guillermo Calero[§], Christopher J. Ackerson[§||], Robert L. Whetten[††], Henrik Grönbeck[††], and Hannu Häkkinen[†‡§¶¶]

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





System size: 762 Atoms
~3400 electrons



Electronic structure of nanoparticles

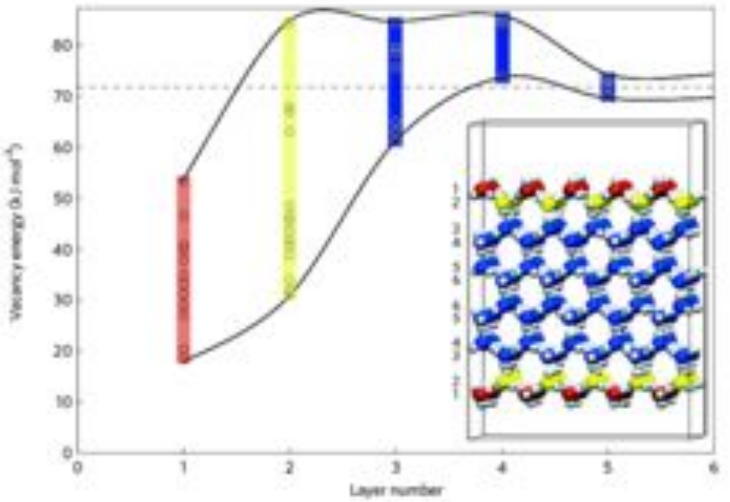
CP2K: science (II)

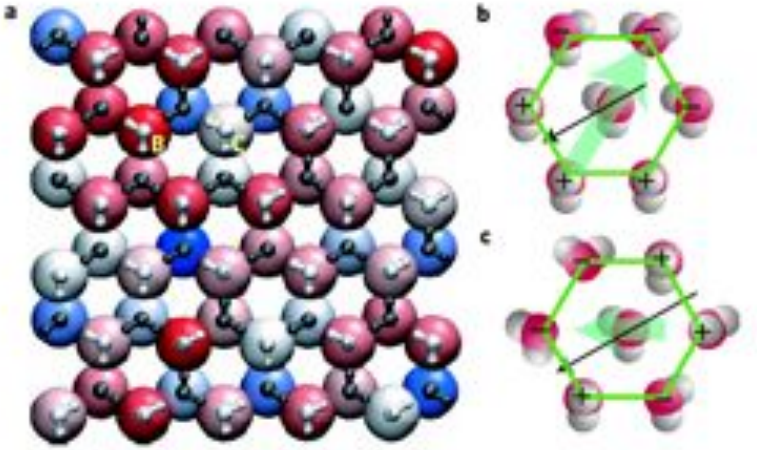
Large variation of vacancy formation energies in the surface of crystalline ice

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





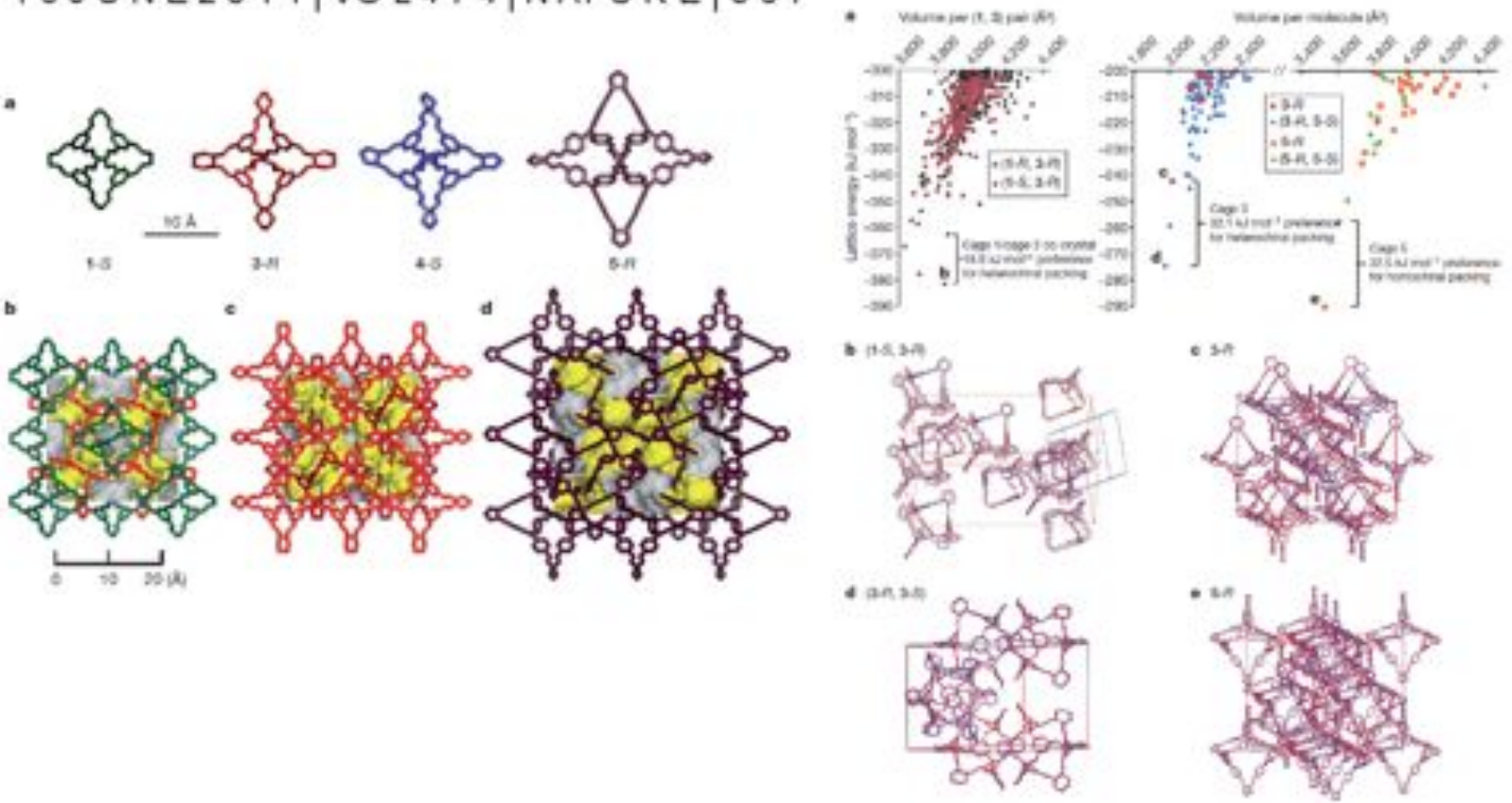
Disordered and frustrated materials

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[2], 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[2] & Andrew I. Cooper[1]

16 JUNE 2011 | VOL 474 | NATURE | 367



Structure prediction of metal organic frameworks

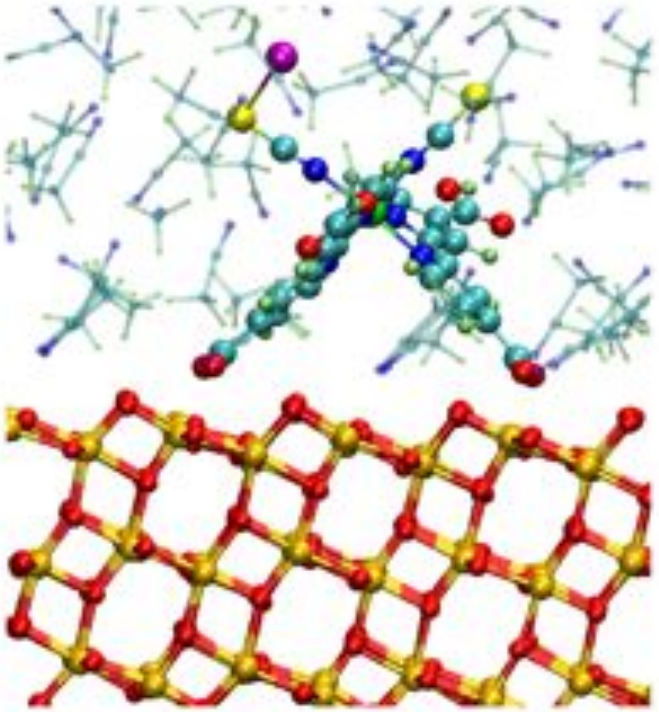
CP2K: science (IV)

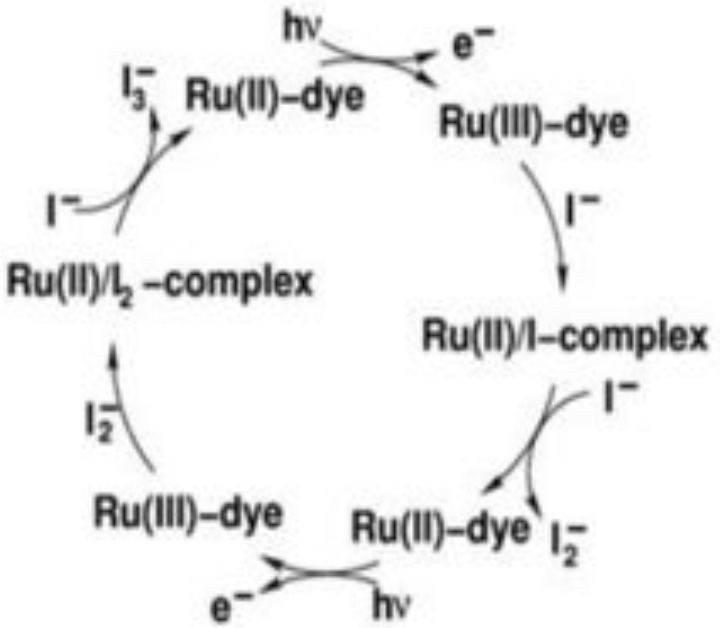
An atomistic picture of the regeneration process in dye sensitized solar cells

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

Modern Compute Resources for atomistic simulation

Perform calculations that

+/- complete faster
- AIMD: at the limit of strong scaling (network)
+ Ensemble simulations: connect to experiment

+ are cheaper

+ are technically better converged (grids, basis, sampling)

+ are based on more accurate theory

+ are based on larger models

More accurate theory:

From 'GGA' to 'HFX' to 'MP2'

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 can be constructed by adding new ingredients.

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

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

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

3) Double hybrids: include some MP2-like terms

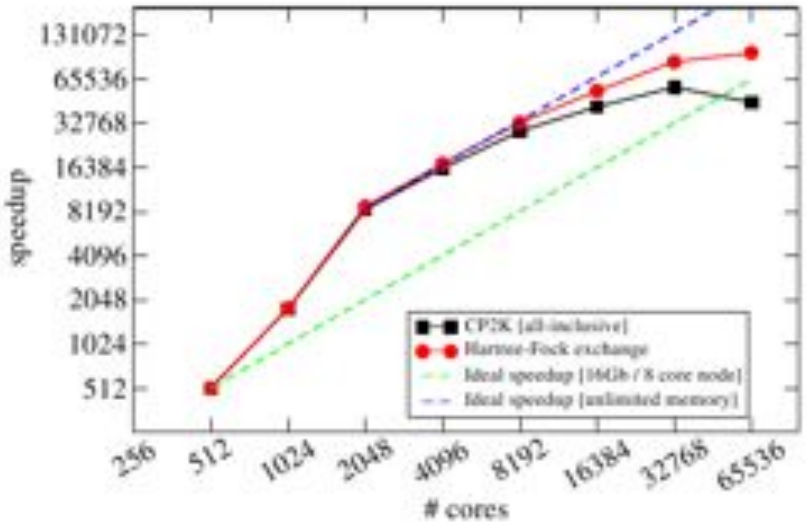
Mundy, Kathmann, Rousseau, Schenter, VandeVondele, Hutter, SCIDAC reviews (spring 2010).

Condensed phase HFX

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

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

216 atoms LiH



Highly accurate calculations at the basis set limit, at gamma are possible... also for systems containing 1000 atoms

- Allows for significantly more accurate calculations of chemical reactions and band alignments.
- Computationally 10-100x more demanding than GGA

Guidon M; Hutter J; VandeVondele J. 2009
JOURNAL OF CHEMICAL THEORY AND COMPUTATION 5(11): 3010-3021

MP2

Closed-shell MP2 energy:

$$E_2 = -\sum_{ij,ab}^{occ,vir} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}$$

Where:

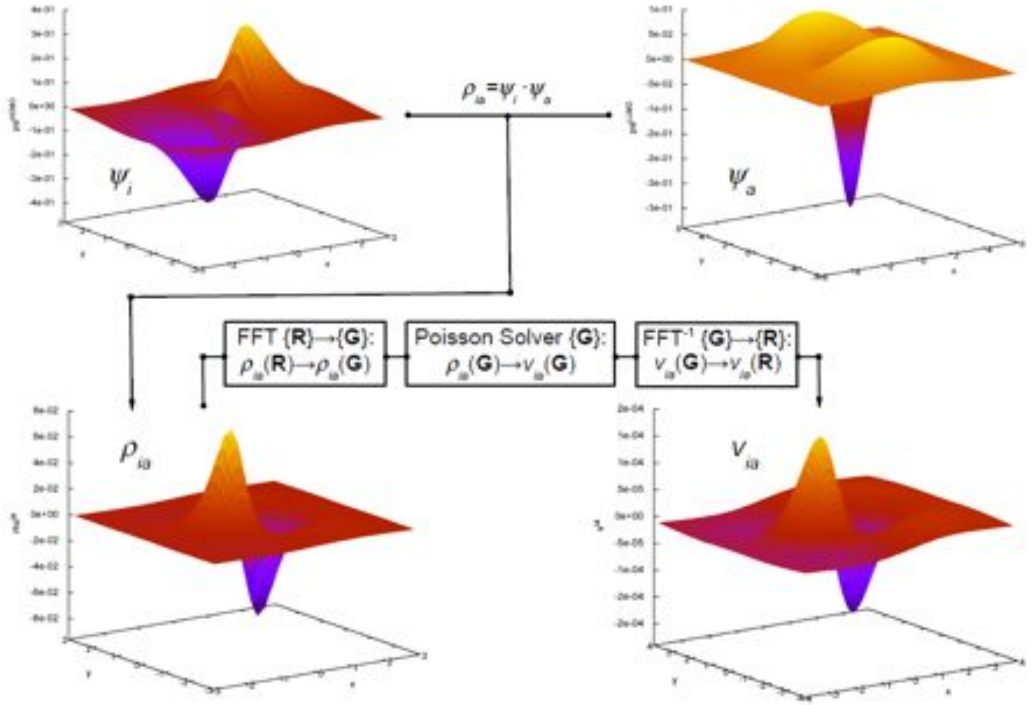
$$(ia|jb) = \int\int \phi_i(\vec{r}_1)\phi_a(\vec{r}_1)\frac{1}{|\vec{r}_1 - \vec{r}_2|}\phi_j(\vec{r}_2)\phi_b(\vec{r}_2)d\vec{r}_1 d\vec{r}_2$$

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

Scales as O(N**5)

MP2 GPW

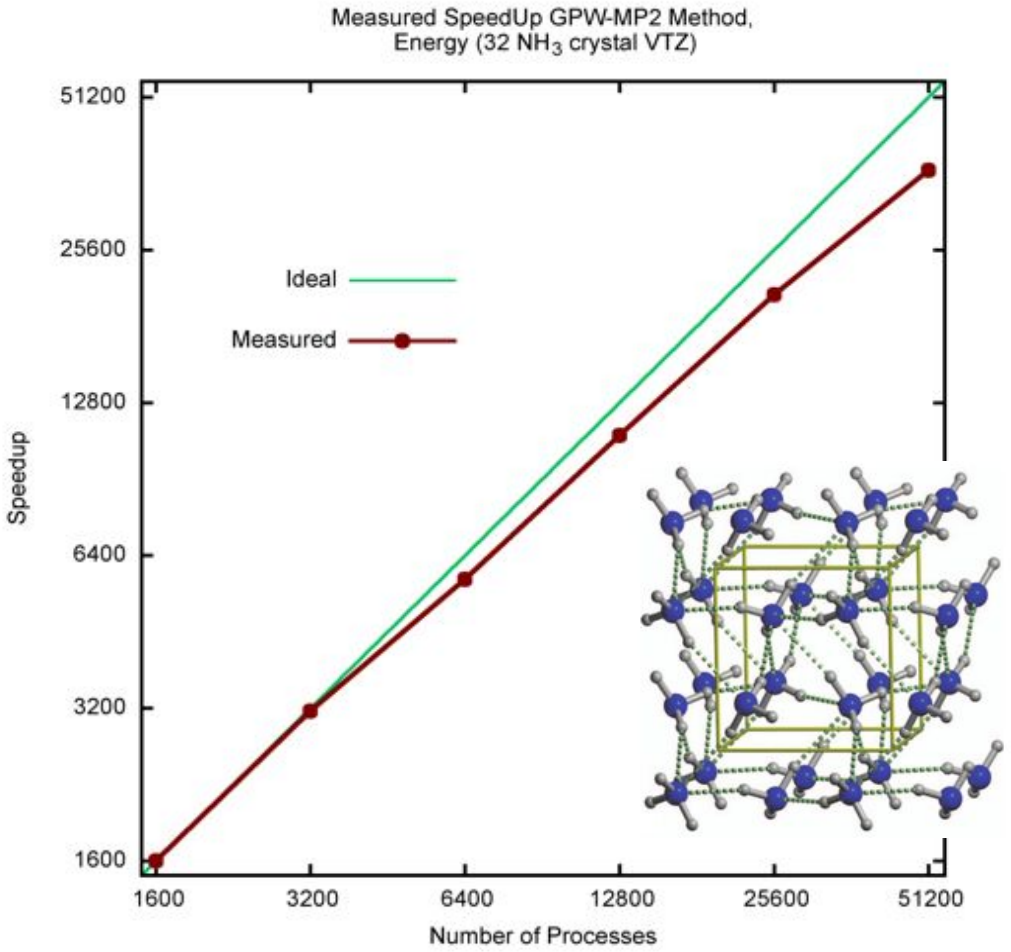
- Directly compute (ia|μν) using the GPW algorithm.
- Transform in the usual way afterwards

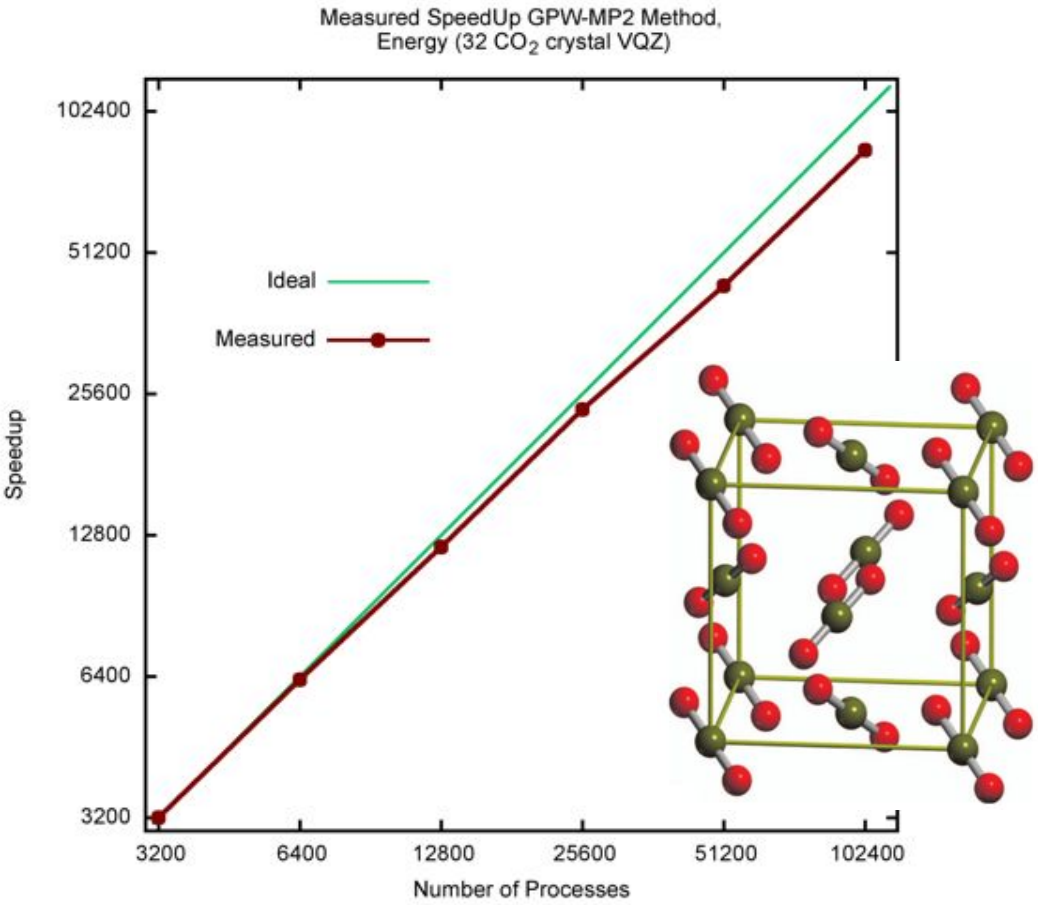


Excellent parallel scalability
'Effective O(N**3)' scaling
Great for the condensed phase

Mauro Del Ben, Hutter, VandeVondele, to be published

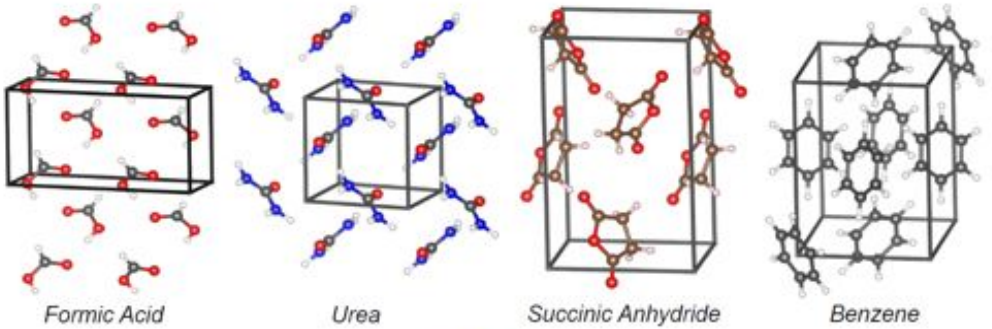
MP2-GPW scalability
CO2 and NH3 crystals



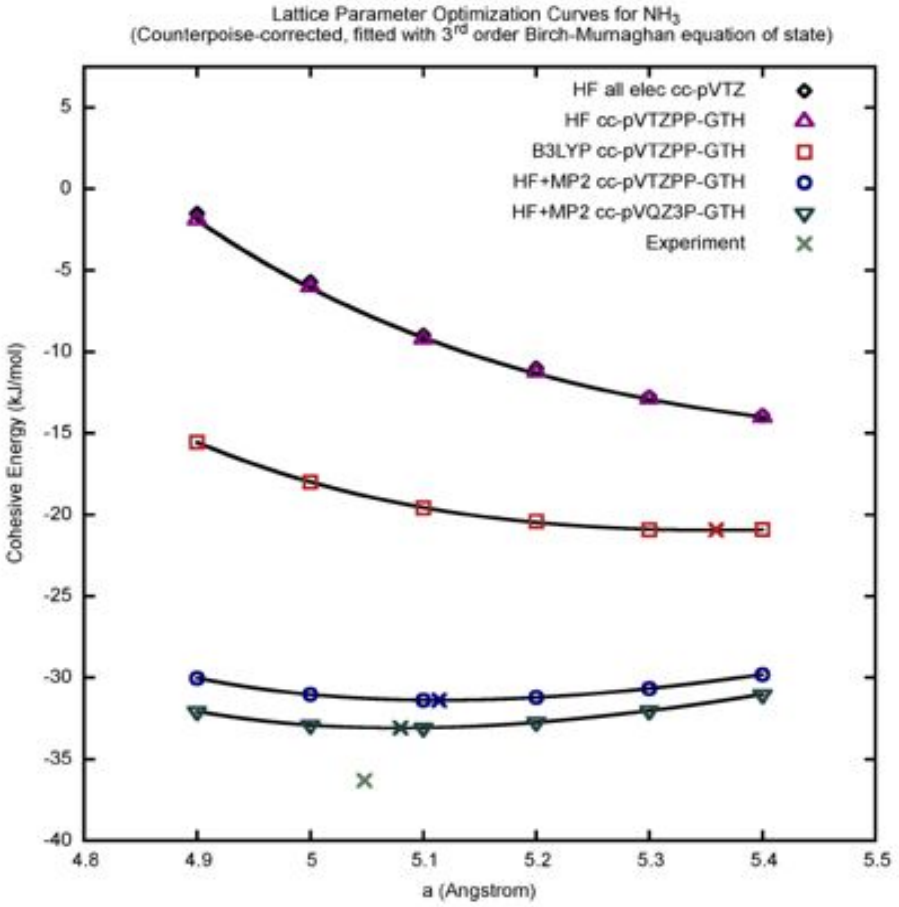


NH3: 74s on 51200 cores, ~cc-pVTZ : 2272BF, 128 occupied orbitals
CO2: 518s on 102400 cores, ~cc-pVQZ : 5184BF, 256 occupied orbitals

Molecular crystals



Cohesive energies (KJ/mol)	NH_3		CO_2	
	HF	HF+MP2 ([a])	HF	HF+MP2 ([a])
cc-VDZP	-8.43	-25.78 (-22.6)	-7.48	-8.92 (-8.9)
cc-VTZPP	-5.98	-30.93 (-29.6)	-5.86	-20.95 (-19.3)
cc-VQZ3P	-5.60	-32.76 (-32.3)	-5.91	-23.28 (-24.4)
Extr.[b]	-5.52	-33.93 (-35.6)	-5.99	-26.09 (-28.4)
Exp.		-36.3		-31.1



MP2 corrects the 6% error in lattice parameter (B3LYP) quite effectively

MP2 summary

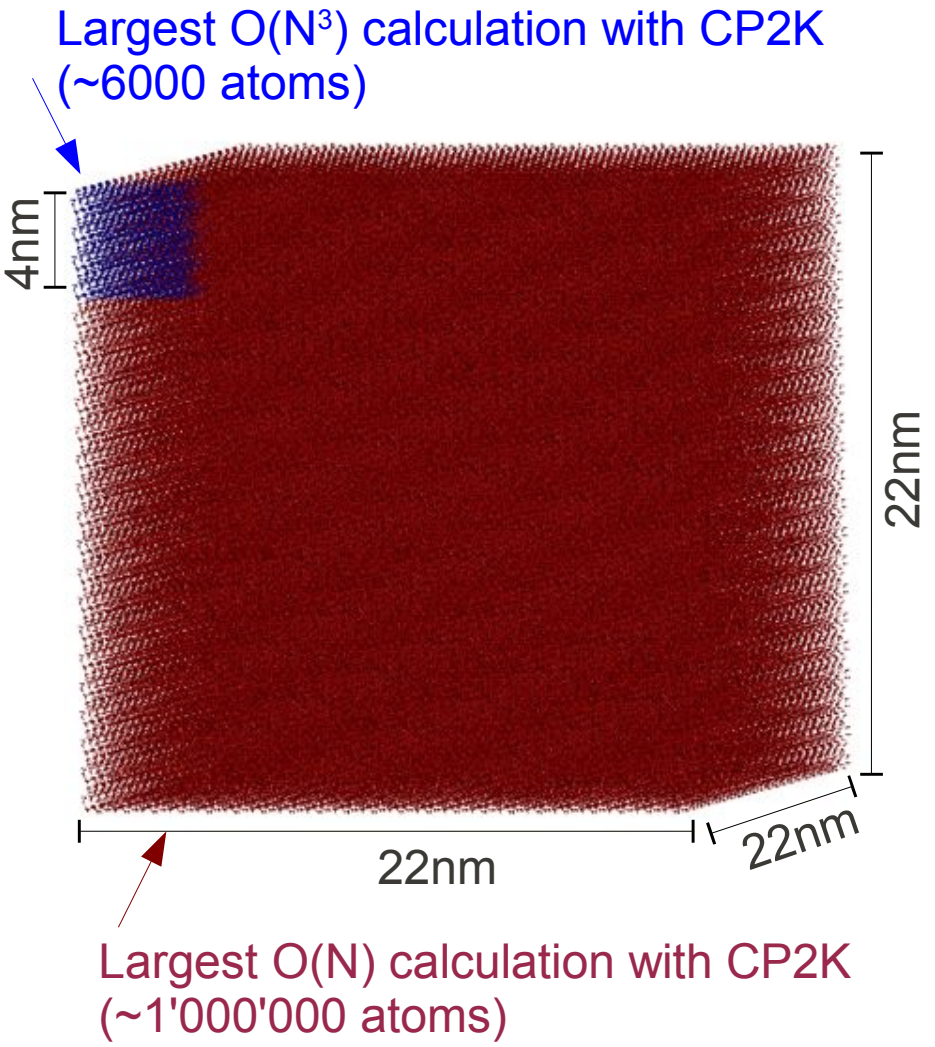
- Excellent scalability
- MP2 GPW: new interesting algorithm
- CPU only algorithm so far, might be an interesting candidate for GPU porting
  - O(N**3) dominating part: complex
  - O(N**5) easier
- 100-10000x more expensive than GGA

Larger model systems

Linear Scaling SCF

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

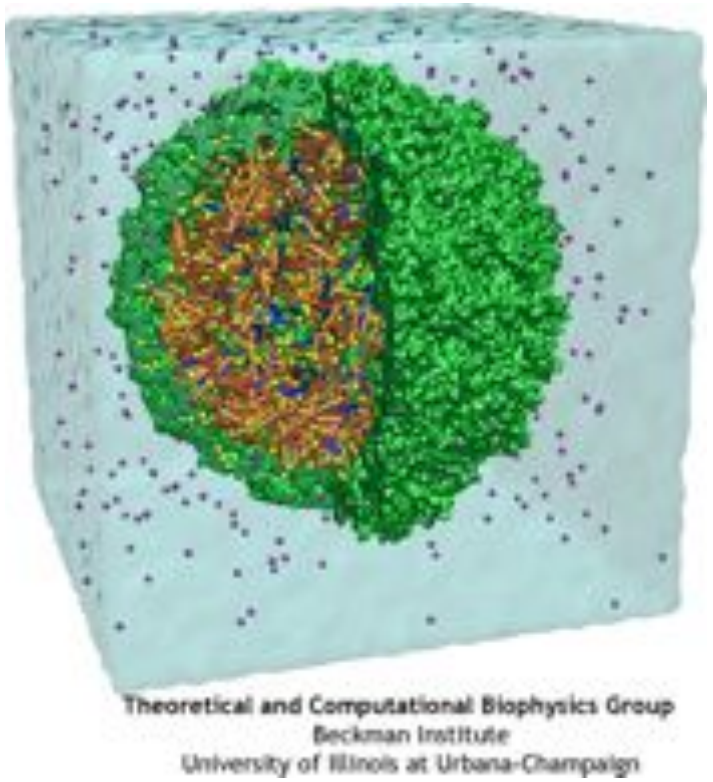
A newly implemented algorithm is O(N), allowing for far larger systems to be studied.



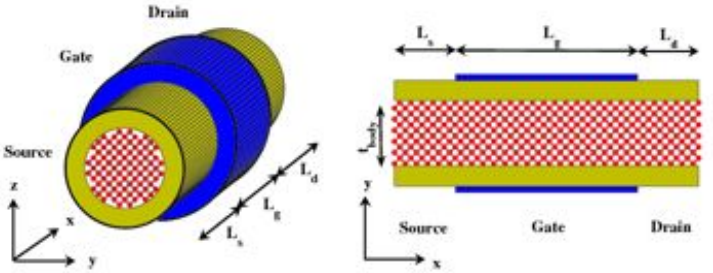
Linear Scaling SCF

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

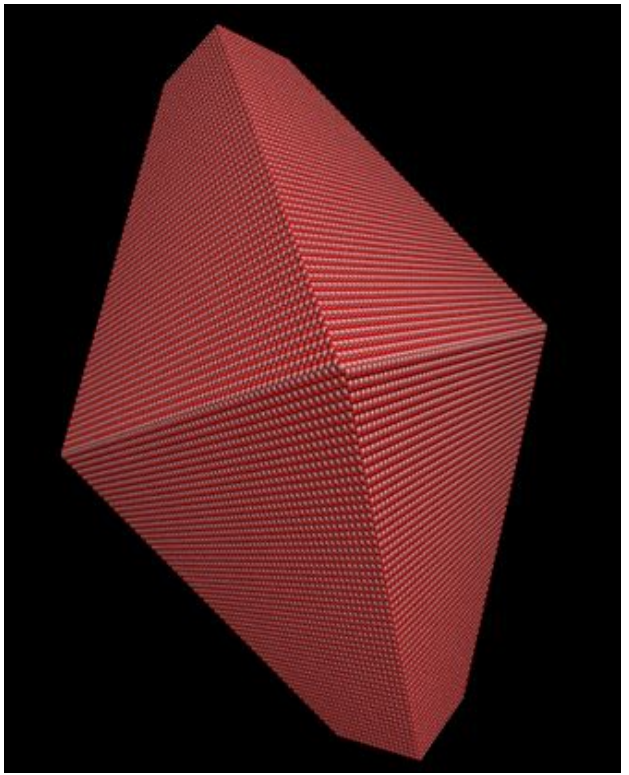
Solvated virus STMV: 1M



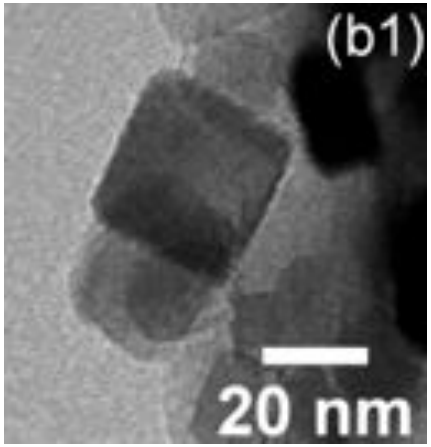
Gate-all-around FET



With Mathieu Luisier



1.5M atoms
Anatase nanocrystal



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

Sign matrix iterations

The density matrix (P) is function of H

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

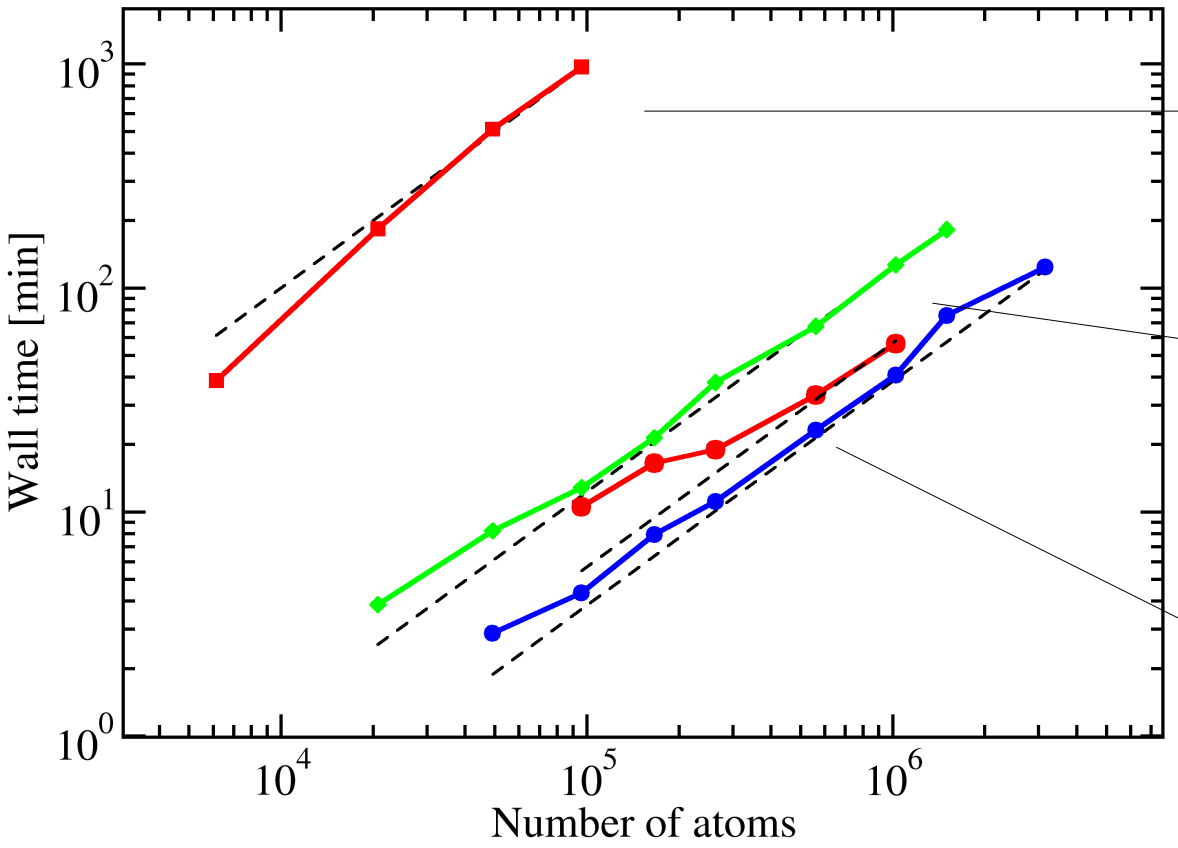
A simple iterative scheme (Newton-Schultz) gives sign(A):

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

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

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

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

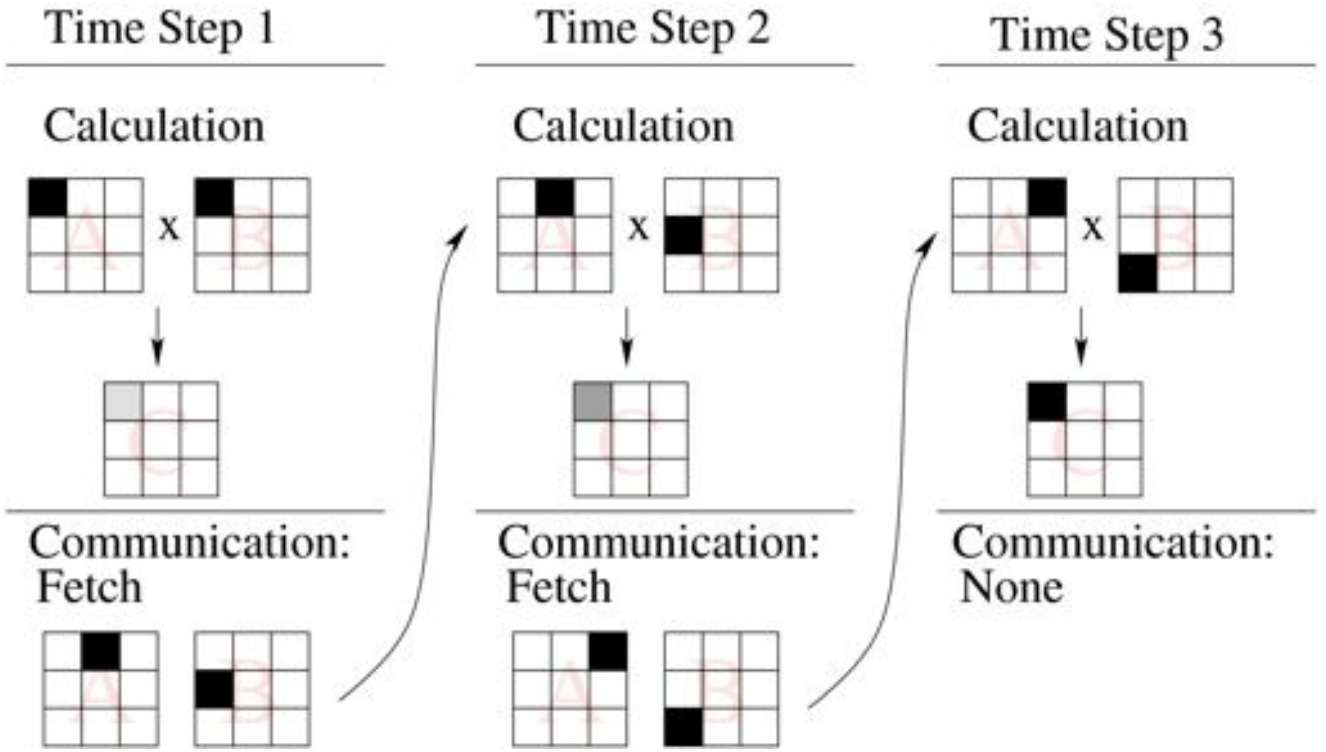
Bulk liquid water. Dashed lines represent ideal linear scaling.

VandeVondele, Borstnik, Hutter, JCTC, http://pubs.acs.org/doi/abs/10.1021/ct200897x

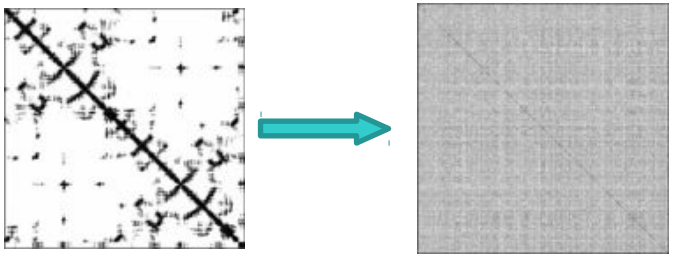
DBCSR: a sparse matrix library

Distributed Blocked Compressed Sparse Row
Distributed Blocked Cannon Sparse Recursive

Target the application: atoms → Blocks (e.g. 5x5, 13x13, 23x23)
Linear scaling → Sparse
Fully dense → Cannon
Large scale → Distributed
High Performance → Recursive



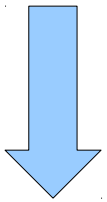
Cannon style communication on a homogenized matrix for strong scaling

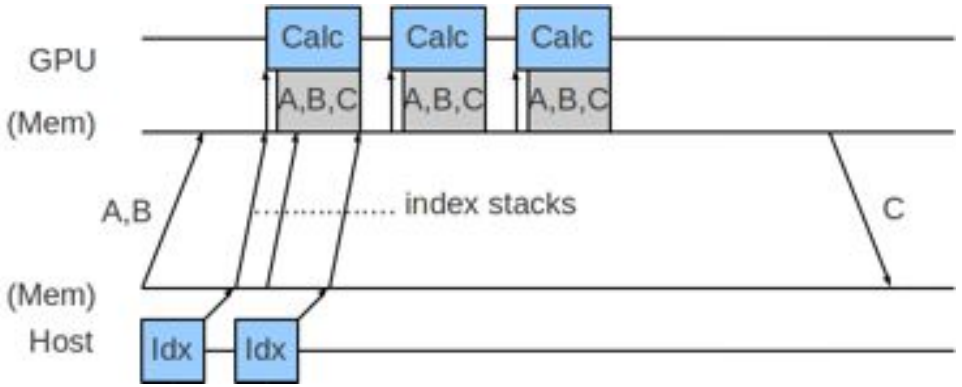


The local multiplication

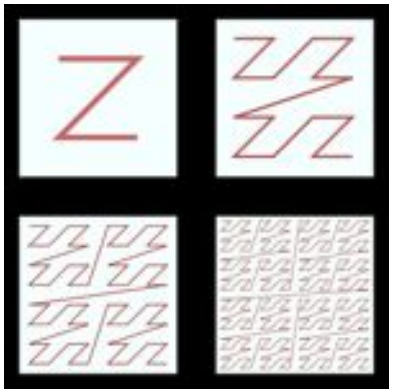
A two stage process:

- Index building: figure out what to calculate
- Computation: Do the calculations





A cache-oblivious recursive approach

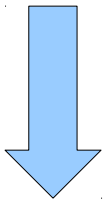


LIBSMM: an autotuned CPU library for small matrix multiplication: 2x faster than optimized blas for applications

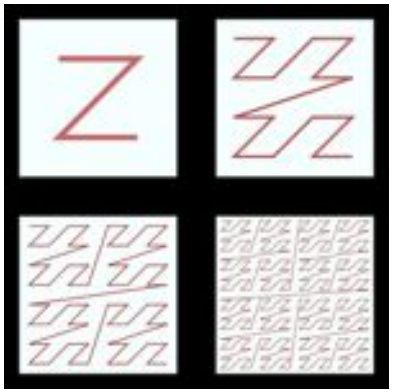
The local multiplication

A two stage process:

- Index building: figure out what to calculate
- Computation: Do the calculations



A cache-oblivious recursive approach



LIBSMM: an autotuned CPU library for small matrix multiplication: 2x faster than optimized blas for applications

GPU Kernel Description

Compute a batch Cij, Aik, Bkj of small matrix block products (5x5 .. 23x23)
Not too different from batched dgemm (Ci=Ci+Ai*Bi) available in cudablas 4.1

CUBLAS Batched dgemm reaches
Only ~50 Gflops for 23x23

Need to:
1) exploit dependencies
(data reuse)
2) Better compute kernel

Sort batch on Cij so we have only 2 data read/writes instead of 4.
Hand-optimize a CUDA kernel for size 23x23 (playing with registers / occupation etc)

The GPU port of DBCSR (I)

Not so easy … ran into many issues with tools
… but received excellent support from Cray/NVIDIA/CP2K team

- Up to 3x performance slowdown in MPI for dynamically linked binaries
  - Fix: use a (not yet released) Cray patched libudreg.so.0 (?)
- ran into CCE miscompilation
  - Fix : compile selected file at -O0
- problems with (virtual) memory management
  - Fix: use hugepages
- A CP2K bug
  - Fix: add a memcpy.
- Up to 10X threading performance difference between compilers
  - Investigating (not crucial for final benchmark)
- Cuda-memcheck crashes/gave false positives
  - Ignore after checking carefully
- Usual tools (valgrind or CP2K internal timing/memory trace) do not work
  - accept

Not a particularly time efficient procedure...
Not for every developer …
Not without excellent support ….

The GPU port of DBCSR (II)

Open Challenges:

•Only one process per node can connect to GPU.
Need a functional OMPed code as well

•Work sharing between GPU and CPU not yet optimal.
Should the CPU also do stacks of multiplications ?

•Sending data panels to GPU not yet overlapping with any work
Could be started as soon as MPI completes. Maybe double buffering

•GPU Kernel optimized for 23x23 matrices (water molecule). We need many sizes.
Needs an auto-tuning and auto-generating framework for the GPU
A strategy to deal with smaller blocks.

GPU multiplication benchmark

Benchmark matrix-matrix multiplication with random [0..1] matrix, quasi-realistic fill-in pattern

Hybrid MPI+OpenMP+CUDA Fortran/C code

NREP=6: Matrix dimension N=159'000;
50% occupation

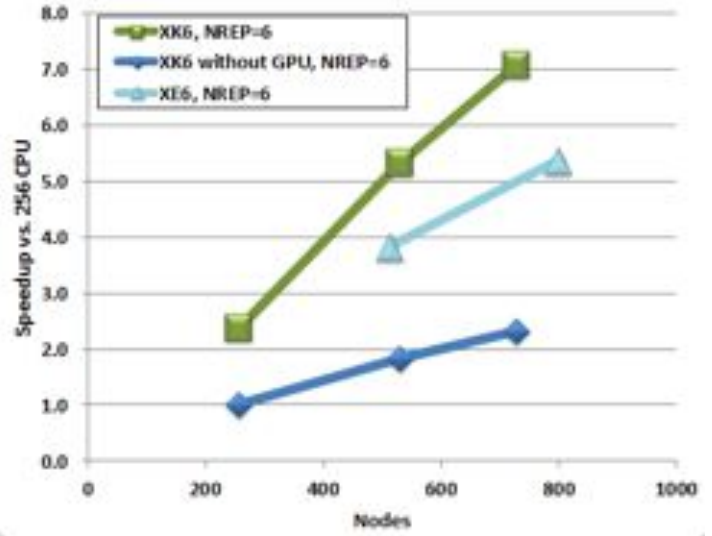
<370Mb of data per node on 256 nodes....

CPU (1xIL, XK6): 38 GFLOPS
CPU (2xIL, XE6): 77 GFLOPS (2.3-2.5 GFLOPs/core)
GPU : 114 GFLOPS (110 – 125 GFLOPs)

Current Kernel performance 170 Gflops.

Estimated Kepler performance: X.Xx speedup on kernel

Performance difference: time in MPI



(data Peter Messmer)

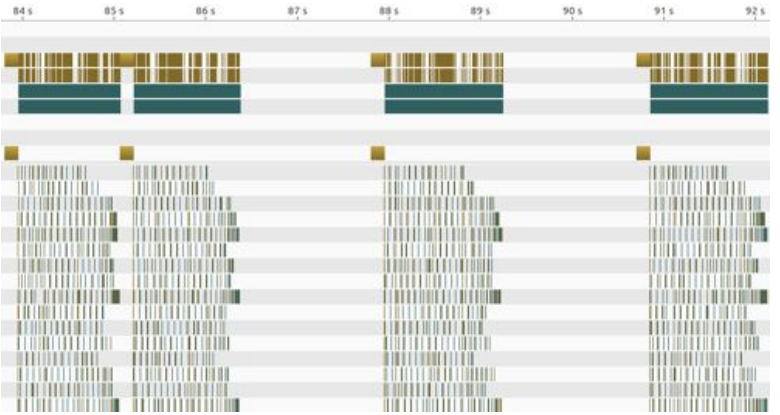
GPU application benchmark

- >400 multiplications for 1 run.
- Additional thresholding in multiplications (less flops for same data)
- This week's results.... subject to change

20736 atoms (6912 water molecules), matrix dim 159000, on 576 nodes XK6, ~60 matrix multiplications / iter.

XK6 without GPU : 1965s per iteration
XK6 with GPU : 924s per iteration

Speedup 2.12x



MPI performance (bandwidth) appears to be the bottleneck (e.g. 50% slowdown without custom rank reordering) :

- Still need to figure out MPI performance (incl. effectiveness of overlap).
- Is the dynamic linking still an issue ?
- Any interference between GPU+CPU ?
- One Communication thread per node enough ?

Summary Linear Scaling

- O(N) scaling reached
- CUDA kernel can deliver good performance
- MPI performance is crucial
- Applications will need
  - CUDA; more general auto-tuned library.
  - CP2K: improved LS algorithms.

Acknowledgements

Zürich
Juerg Hutter
Urban Borstnik
Christian Pousa
Mauro Del Ben
Valery Weber
Manuel Guidon

World-wide
CP2K Team

PNL&Minnesota
Chris Mundy

CSCS&EPCC&
CRAY&NVIDIA
Neil Stringfellow
Iain Bethune
Peter Messmer
Roberto Ansaloni
Thomas Schulthess
John Levesque

Flops&More
HP2C
CSCS
ORNL
UZH
ETH
SNF
INCITE
EU-FP
DEISA
PRACE

You for your attention!