

# Large Scale Electronic Structure Calculations

Jürg Hutter  
University of Zurich

8. September, 2008 / Speedup08

# CP2K Program System

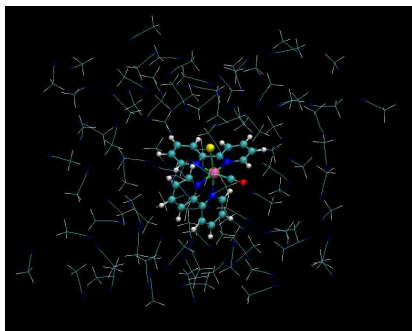
- GNU General Public License
- Community Developers Platform on "Berlios" ([cp2k.berlios.de](http://cp2k.berlios.de))
- Open CVS for download
- Written in Fortran95, 540'000 lines of code
- MPI and OpenMP parallelization
- Quality control: automatic regression and memory leak testing
- Force Methods: 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,  $\mathcal{O}(N)$  in basis set size
- All-electron calculations with projector augmented-waves

## Ru-dye in Acetonitrile Solution

Box:  $(21.43 \text{ \AA})^3$ ,  $\approx 620$  atoms,  $\approx 6000$  bsf,  $\approx 850$  occupied orbitals, 1 min/ MD step / 128 CPUs

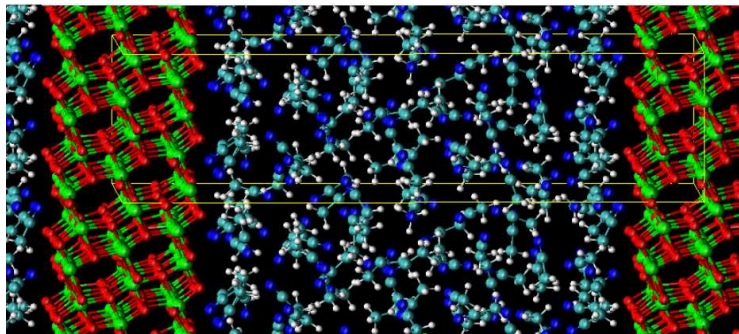


## TiO<sub>2</sub> / Acetonitrile Interface

101 anatase surface, slab calculation,

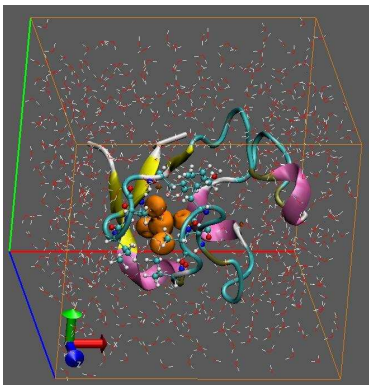
system size:  $22.6 \times 10.0 \times 40 \text{ \AA}$

(TiO<sub>2</sub>)<sub>72</sub> (NCCH<sub>3</sub>)<sub>68</sub>, 624 atoms, 5632 electrons, 10260 bsf



## Rubredoxin in Water Solution

Solvated Rubredoxin (FeS<sub>4</sub>);  $\approx$  2800 atoms;  $\approx$  55000 bsf  
117 s / SCF on 1024 CPUs (XT3), 80% parallel efficiency



# Ab Initio Molecular Dynamics

**Chemistry:** Charged particles (electrons, nuclei)

Coulomb interactions

## Time-dependent Schrödinger equation

$$i\hbar \frac{\partial \Psi(\mathbf{x}; t)}{\partial t} = \mathcal{H} \Psi(\mathbf{x}; t)$$

- $\mathcal{H}$  Hamiltonian operator
- $\Psi(\mathbf{x}; t)$  wavefunction

## Born–Oppenheimer Approximation

Separation of movement of heavy particles (nuclei) and light particles (electrons).

Electrons always assume a stationary state for all nuclear positions.

$$i\hbar \frac{\partial \Psi_{\mathbf{N}}(\mathbf{R}; t)}{\partial t} = \mathcal{H}_{\mathbf{N}} \Psi_{\mathbf{N}}(\mathbf{R}; t)$$

$$\mathcal{H}_{\mathbf{N}} = - \sum_A \frac{1}{2M_A} \delta_A + V(\mathbf{R})$$

$$\mathcal{H}_e \Psi_{\mathbf{N}}(\mathbf{x}; \mathbf{R}) = V(\mathbf{R}) \Psi_{\mathbf{N}}(\mathbf{x}; \mathbf{R})$$

## Newton equation of motion

**Approximation:** Time evolution of nuclei can be described by classical mechanics.

$$M_A \ddot{\mathbf{R}}_A = - \frac{\partial V(\mathbf{R})}{\partial \mathbf{R}_A}$$

$$\mathcal{H}_e \Psi_N(\mathbf{x}; \mathbf{R}) = V(\mathbf{R}) \Psi_N(\mathbf{x}; \mathbf{R})$$

- Newton EOM : Molecular dynamics
- Time-independent Schrödinger equation : Quantum chemistry

# Electronic structure theory

## Kohn–Sham density functional theory

$$V(\mathbf{R}) = \text{Min} (E_{\text{KS}}(\{\Phi\}; \mathbf{R}))$$

$$E_{\text{KS}}(\{\Phi\}; \mathbf{R}) = -\frac{1}{2} \sum_i \int \Phi_i - \frac{1}{2} \delta \Phi_i \, dr + \int V_{\text{ext}}(r) n(r) \, dr \\ + \iint \frac{n(r)n(r')}{|r-r'|} \, dr dr' + E_{\text{xc}}[n]$$

$$n(r) = \sum_i |\Phi_i(r)|^2 \quad \int \Phi_i \Phi_j \, dr = \delta_{ij}$$

- N (number of electrons) coupled, single-particle, non-linear Schrödinger equations
- Constrained minimization problem

## Basis: Atom-centered Gaussians

**Advantage** Localized, analytic integrals, biased, product theorem

**Disadvantage** Non-orthogonal, biased

Other choices used: Exponential functions, plane waves, wavelets, etc.

$$\Phi_i(r) = \sum_{\alpha}^M C_{\alpha i} \varphi_{\alpha}(r)$$

New variational parameters:  $M \times N$  matrix **C**.

## Kohn–Sham Method

Minimize  $E_{\text{KS}}(\mathbf{C})$

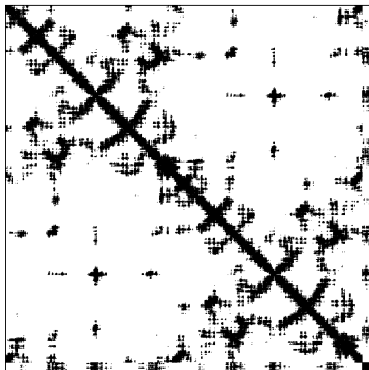
Constraints:  $\mathbf{C}^T \mathbf{S} \mathbf{C} = \mathbf{1}$

$\mathbf{S}$ : Overlap matrix, metric

Optimization using gradient methods: Gradient  $\mathbf{H}_{\text{KS}} \mathbf{C}$   
 $\mathbf{H}_{\text{KS}}$  is the Kohn–Sham matrix

## Kohn–Sham Matrix

- Kohn–Sham matrix is sparse (in localized basis functions) allows for a linear scaling calculation of  $\mathbf{H}_{KS}$



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

## Coulomb energy and potential

$$E_H = \frac{1}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' = \frac{2\pi}{\Omega} \sum_{\mathbf{G}} \frac{n^*(\mathbf{G})n(\mathbf{G})}{\mathbf{G}^2}$$

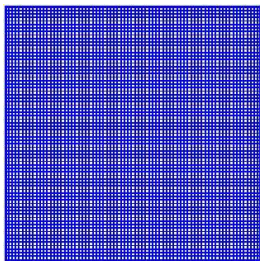
$$V_H(\mathbf{r}) = \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' = \frac{4\pi}{\Omega} \sum_{\mathbf{G}} \frac{n(\mathbf{G})}{\mathbf{G}^2} e^{i\mathbf{G}\cdot\mathbf{r}}$$

Plane wave expansion of charge density

$$n(\mathbf{r}) = \sum_{\mathbf{G}} n(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}$$

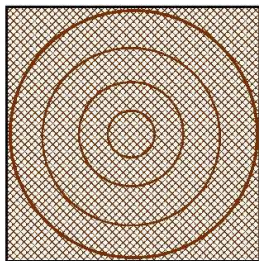
# Real Space Grid

Finite cutoff and computational box define a real space grid  $\{\mathbf{R}\}$



Real Space

FFT  
↔



G-Space

# Coulomb Potential

$$\mathbf{P} = \mathbf{C}\mathbf{C}^T \rightarrow n(\mathbf{R}) \xrightarrow{\text{FFT}} n(\mathbf{G}) \rightarrow V_H(\mathbf{G}) = \frac{n(\mathbf{G})}{G^2} \xrightarrow{\text{FFT}} V_H(\mathbf{R}) \rightarrow \mathbf{V}$$

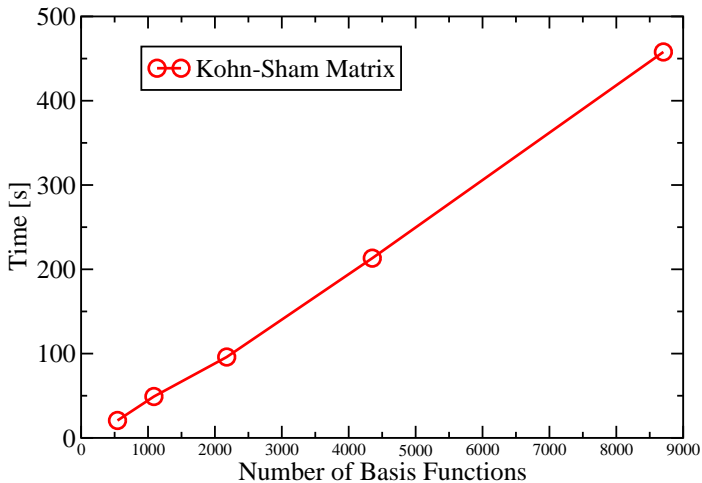
$\underbrace{\hspace{15em}}_{\mathcal{O}(n \log n)}$

$$n(\mathbf{R}) = \sum_{\mu\nu} \mathbf{P}_{\mu\nu} \varphi_\mu(\mathbf{R}) \varphi_\nu(\mathbf{R}) = \sum_{\mu\nu} \mathbf{P}_{\mu\nu} \bar{\varphi}_{\mu\nu}(\mathbf{R})$$

$$\mathbf{V}_{\mu\nu} = \sum_{\mathbf{R}} \mathbf{V}(\mathbf{R}) \varphi_\mu(\mathbf{R}) \varphi_\nu(\mathbf{R}) = \sum_{\mathbf{R}} \mathbf{V}(\mathbf{R}) \bar{\varphi}_{\mu\nu}(\mathbf{R})$$

Efficient screening of sums using  $\bar{\varphi}_{\mu\nu}(\mathbf{R})$ .

## Scaling



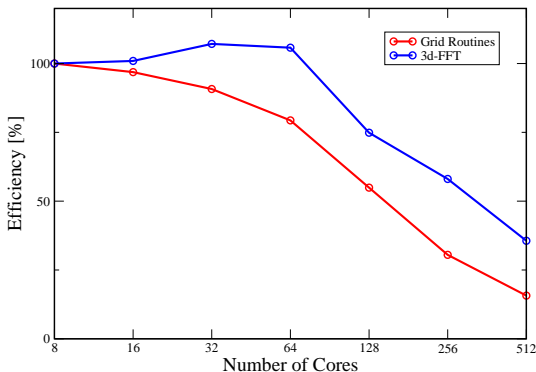
# Parallelization

Conflicting needs for optimal distributions

- Data layout for sparse operator matrices.
- Data layout for real/Fourier space grids.
- Parallel 3d-FFT (typical grid sizes  $200 \times 200 \times 200$ ).
- Typical support of basis function is  $15 \times 15 \times 15$  grid points. Large overlapping regions.

Optimize a complex, empirical work function using Monte Carlo techniques.

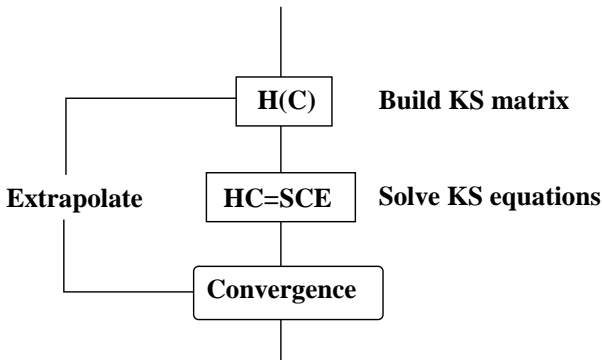
# Parallel efficiency



Test system: 64 water molecules, 192 atoms, 256 orbitals,  
2560 basis functions, 4s/molecular dynamics step on 256 cores  
Cray XT-3, dual core, CSCS Manno

# Optimization

Standard (reference) method : Fix point iteration with diagonalization



# Orbital transformation method

- Direct optimization technique.
- Can use algorithms for unconstrained optimizations.

- 

Memory  $MN$

M Number of basis functions

Scaling  $MN^2$

N Number of occupied orbitals

J. VandeVondele, JH, J. Chem. Phys. **118** 4365 (2003)

J. VandeVondele et al. Comp. Phys. Comm. **167** 103 (2005)

# Orbital transformation method

- Set of reference occupied orbitals:  $\mathbf{C}_0$  ( $M \times N$  Matrix)  
New variables  $\mathbf{X}$  ( $M \times N$  Matrix)

$$\mathbf{C}(\mathbf{X}) = \mathbf{C}_0 \cos \mathbf{U} + \mathbf{X} \mathbf{U}^{-1} \sin \mathbf{U}$$

$$\mathbf{U} = \left( \mathbf{X}^T \mathbf{S} \mathbf{X} \right)^{1/2}$$

- Linear constraint  $\mathbf{X}^T \mathbf{S} \mathbf{C}_0 = 0$
- Analytic gradient for arbitrary  $\mathbf{X}$   
Standard optimization with line search and pre-conditioning

## Avoiding Matrix Functions

If  $\mathbf{C}_0 \approx \mathbf{C}_{\text{opt}} \longrightarrow \mathbf{X} \approx \mathbf{0}$

$$\cos(\mathbf{U}) = \sum_{i=0}^K \frac{(-1)^i}{(2i)!} \mathbf{U}^{2i}, \quad \mathbf{U}^{-1} \sin(\mathbf{U}) = \sum_{i=0}^K \frac{(-1)^i}{(2i+1)!} \mathbf{U}^{2i}$$
$$\mathbf{U} = (\mathbf{X}^T \mathbf{S} \mathbf{X})^{1/2}$$

64 water molecules on 32 CPUs IBM SP4 (seconds)

	DZVP	TZVP	TZV2P	QZV2P	QZV3P
Basis	1472	1856	2560	2944	3684
<b>OT</b>	<b>0.50</b>	<b>0.60</b>	<b>0.77</b>	<b>0.87</b>	<b>1.06</b>
ScaLapack	6.02	8.40	13.80	17.34	24.59

## Simplification of minimization method

Transform constrained minimization problem

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

into an unconstrained minimization of

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

where

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

with  $C = C_0 + \delta C$  and  $C_0^T S C_0 = 1$ .

V. Weber, JH, J. Phys. Comm. **128** 064107 (2008)

## Orbital Refinement Function

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

$$f_2(z) = \frac{1}{2}z(3 - Y),$$

$$f_3(z) = \frac{1}{8}z(15 - 10Y + 3Y^2),$$

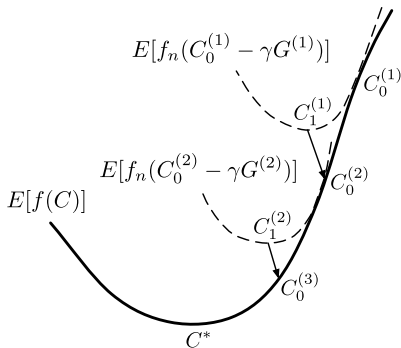
$$f_4(z) = \frac{1}{16}z(35 - 35Y + 21Y^2 - 5Y^3), \text{ and}$$

$$f_5(z) = \frac{1}{128}z(315 - 420Y + 378Y^2 - 180Y^3 + 35Y^4),$$

$$f_n^T(z) S f_n(z) - 1 = \mathcal{O}(\delta z^n).$$

## Optimization

- Nonlinear conjugate gradients with line search
- Includes only matrix additions and multiplications
- Simplifies parallelization and use of sparsity
- Reduces prefactor but is less stable than orbital transformations



## Example: Liquid Water

	$(\text{H}_2\text{O})_{512}$	$(\text{H}_2\text{O})_{1024}$	$(\text{H}_2\text{O})_{2048}$
basis functions	11776	23552	47104
CPUs (XT4)	32	64	64
$n_{\text{iter}} \text{SCF}$	27	25	25
$f_{\text{ks}}$	0.18	0.08	0.04
$f_{\text{other}}$	0.33	0.26	0.20
$f_{\text{mini}}$	0.49	0.66	0.76
$t_{\text{tot}}$ [min]	8.9	23.3	132.7

# Linear scaling optimization

- To reach linear scaling, matrix  $\mathbf{C}$  has to be sparse.
- Solution is invariant to unitary transformations of  $\mathbf{C}$   
We only need a representation of the space spanned by  $\mathbf{C}$ .
- Sparse representations of  $\mathbf{C}$  are called localized orbitals.  
Standard techniques in quantum chemistry based on physical properties.
- Efficient algorithms need parallel sparse matrix-matrix multiply

## L1-Norm Localization

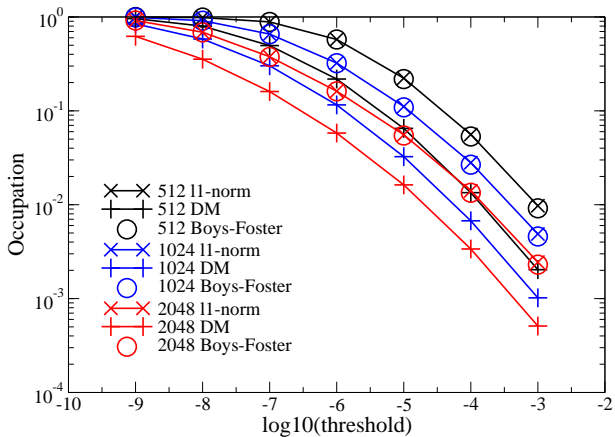
- Smooth sparseness function based on L1 norm

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

- 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

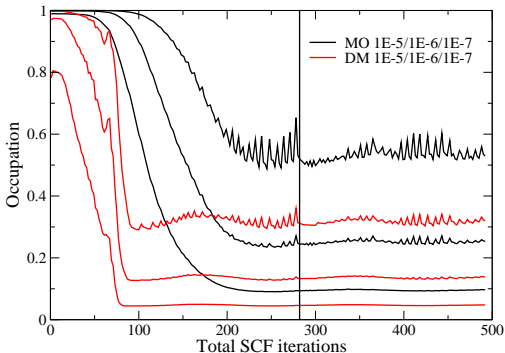
$$U^* = \arg \max_X \phi(U(X)).$$

## Example:Water



# Combine Optimization and Localization

Molecular dynamics, 1024 water molecules



Open question: How does sparseness threshold affect MD stability?

## Multi-level parallelization

- Embarassingly parallel level (MPI)
  - String methods
  - Property calculations
  - Path integral algorithms
- Electronic structure level (MPI)
- Fine-grain parallelisation (OpenMP)
  - Matrix multiply
  - Multiple 1d-FFT
  - Loop level parallelization

# Electronic structure calculations

Applications of CP2K in

- Chemistry
- Biochemistry
- Solid state physics
- Material science
- Nanotechnology

# Acknowledgment

- Joost VandeVondele
- Valery Weber
- Teodoro Laino
- Marcella Iannuzzi
- Florian Schiffmann
- Manuel Guidon

University of Zurich, Swiss National Science Foundation