A Direct Constrained Optimization Method for the Kohn-Sham Equations

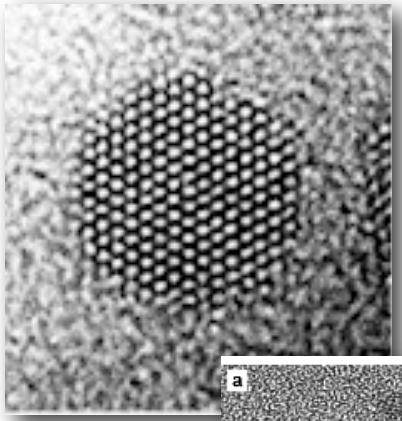
Juan Meza

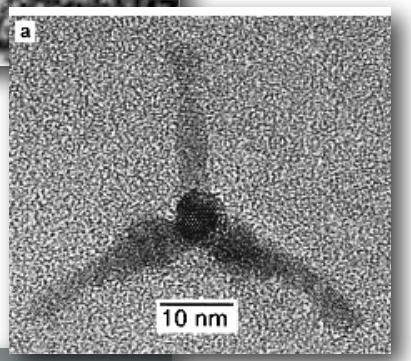
(joint with Chao Yang, Lin-Wang Wang)

Department Head and Senior Scientist
High Performance Computing Research
Lawrence Berkeley National Laboratory

North Carolina State November 13, 2008

What do all of these have in common?





On using mathematics for chemistry

Every attempt to employ mathematical methods in the study of chemical questions must be considered ***profoundly irrational*** *and contrary to the spirit of chemistry. If mathematical analysis should ever hold a prominent place in chemistry – an aberration which is happily* ***almost impossible*** *– it would occasion a rapid and widespread degeneration of that science.*

Auguste Comte, 1830

100 years later – the problem is solved!

...in the Schrödinger equation we very nearly have the mathematical foundation for the solution of the whole problem of atomic and molecular structure ...

almost

…the problem of the many bodies contained in the atom and the molecule cannot be completely solved without a great further development in mathematical technique.

G.N. Lewis, J. Chem. Phys. 1, 17 (1933)

Many-body Schrödinger equation

$$H\Psi_i(r_1, r_2, ..., r_N) = E_i\Psi(r_1, r_2, ..., r_N)$$

$$H = -\frac{h}{2m}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i=1}^{N}v(r_i) + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|r_i - r_j|}$$

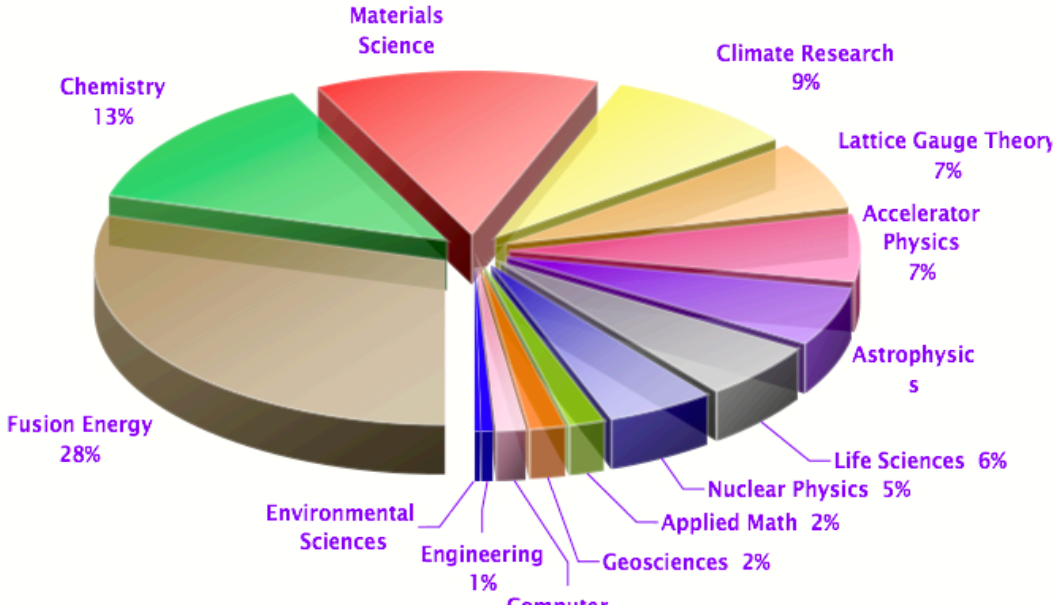
- Ψ_i contains all the information needed to study a system
- $|\Psi_i|^2$ probability density of finding electrons at a certain state
- E_i quantized energy
- Computational work goes as 10^{3N}, where N is the number of electrons

Density Functional Theory

- The unknown is simple – the electron density, ρ
- Hohenberg-Kohn Theory
  - There is a unique mapping between the ground state energy and density
  - Exact form of the functional is unknown
- Independent particle model
  - Electrons move independently in an average effective potential field
  - Add correction for correlation
- Good compromise between accuracy and feasibility

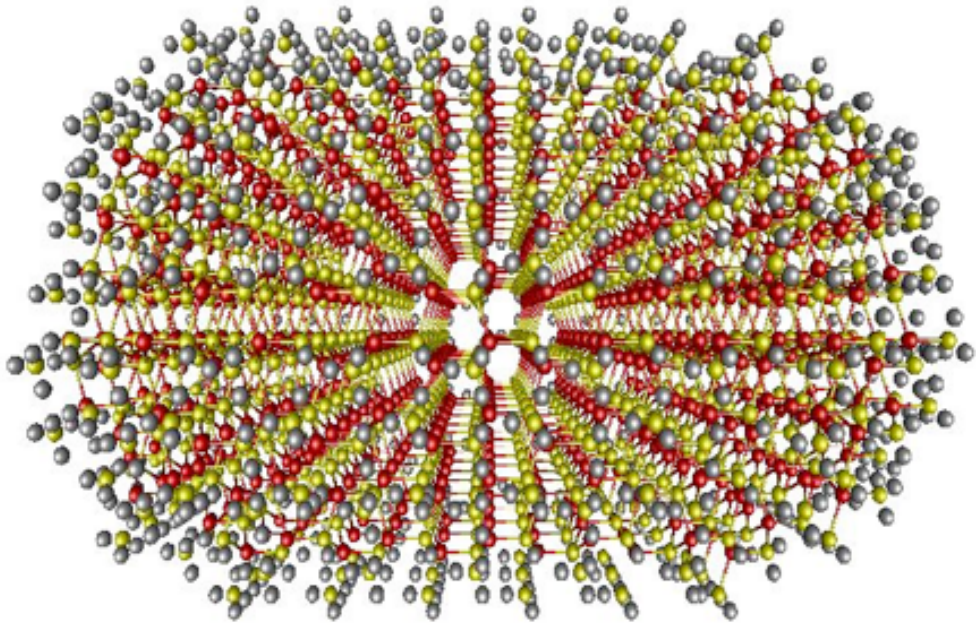
DFT codes play a major role in computational science

Franklin (NERSC-5): Cray XT4



- 9,660 nodes; 38,640 2.3 GHz cores
- 356 Tflops/s peak
- DFT methods account for 75% of the materials sciences simulations at NERSC, totaling over 5 million hours of computer time in 2006

Fast forward to today: we can now simulate realistic nanosystems



The calculated dipole moment of a 2633 atom CdSe quantum rod, $Cd_{961}Se_{724}H_{948}$. Using 2560 processors at NERSC the calculation took about 30 hours.

Wang, Zhao, Meza, Phys. Rev. B, 77, 165113 (2008)

Advances in ***density functional theory coupled with multinode computational clusters*** now enable accurate simulation of the behavior of multi-thousand atom complexes that mediate the electronic and ionic transfers of solar energy conversion. These new and emerging nanoscience capabilities bring a ***fundamental understanding of the atomic and molecular processes of solar energy utilization within reach.***

Basic Research Needs for Solar Energy Utilization,

Report of the BES Workshop on Solar Energy Utilization,April 18-21, 2005

Brief Review of Fundamental Equations

Kohn-Sham Formulation

- Use *N* noninteracting electrons as a reference
- Replace many-particle wavefunctions, Ψ_i, with single-particle wavefunctions, ψ_i
- Write Kohn-Sham total energy as:

$$E_{total}[\{\psi_i\}] = \frac{1}{2}\sum_{i=1}^{n_e}\int_\Omega |\nabla\psi_i|^2 + \int_\Omega V_{ion}\rho$$

$$+\frac{1}{2}\int_\Omega \frac{\rho(r)\rho(r')}{|r-r'|}drdr' + E_{xc}(\rho),$$

$$\rho(r) = \sum_{i=1}^{n_e}|\psi_i(r)|^2,\ \int_\Omega \psi_i\psi_j = \delta_{i,j},\ n_e$$

- Exchange-correlation term, E_{xc}, contains quantum mechanical contributions, plus part of K.E. not covered by first term when using single-particle wavefunctions

Kohn-Sham Equations

- Goal is to find the ground state energy by minimizing total energy, E_{total}
- Leads to:

$$
\begin{aligned}
H\psi_i &= \epsilon_i \psi_i, \quad i = 1, 2, ..., n_e \\
H &= \left[-\frac{1}{2}\nabla^2 + V_{ion}(r) + \int \frac{\rho}{|r - r'|} + V_{xc}(\rho) \right]
\end{aligned}
$$

Discretization Options

- Finite difference $\psi'(r_j) \approx [\psi(r_j + h) - \psi(r_j - h)]/h$
- Finite elements

$$\psi(r) \approx \sum_j^n \alpha_j \phi_j(r), \quad \phi_j(r) \text{ functions with local support}$$

- Local orbital method (good for molecules)
  - Choose $\phi_j(r)$ as Gaussian or other "nice" functions
- Planewave expansion
  - Choose $\phi_j(r)$ as $e^{ig_j \cdot r}$

Finite Dimensional Problem

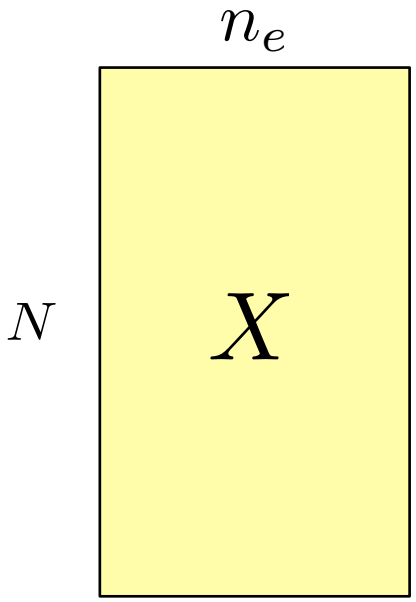
We want to find the ground state energy

$$\min E[\{\psi_i\}] = \frac{1}{2}\sum_{i=1}^{n_e}\int_\Omega |\nabla\psi_i|^2 + \int_\Omega V_{ion}\rho + \frac{1}{2}\int_\Omega \frac{\rho(r)\rho(r')}{|r-r'|}drdr' + E_{xc}(\rho)$$

After discretization we have:

$$\min_{s.t. X^*X = I_{n_e}} E_{KS}(X) \equiv E_{kinetic}(X) + E_{ion}(X) + E_{Hartree}(X) + E_{xc}(X),$$

where



$$\begin{aligned} E_{kinetic} &= \frac{1}{2}\mathrm{trace}(X^*LX) \\ E_{ionic} &= \mathrm{trace}(XV_{ion}X^*) \\ E_{Hartree} &= \frac{1}{2}\rho(X)^T L^\dagger \rho(X) \\ E_{xc} &= \rho(X)^T(\mu_{xc}[\rho(X)]) \\ \rho(X) &= \mathrm{diag}(XX^*) \end{aligned}$$

KKT Conditions

- KKT conditions

$$\begin{aligned} \nabla_X \mathcal{L}(X, \Lambda) &= 0, \\ X^* X &= I_{n_e}. \end{aligned}$$

- Leads to ***Discretized Kohn-Sham equations***

$$\begin{aligned} H(X)X &= X\Lambda, \\ X^* X &= I_{n_e}, \\ H(X) &= \frac{1}{2}L + V_{ion} + \mathrm{Diag}\,(L^\dagger \rho(X)) + \mathrm{Diag}\, g_{xc}(\rho(X)) \end{aligned}$$

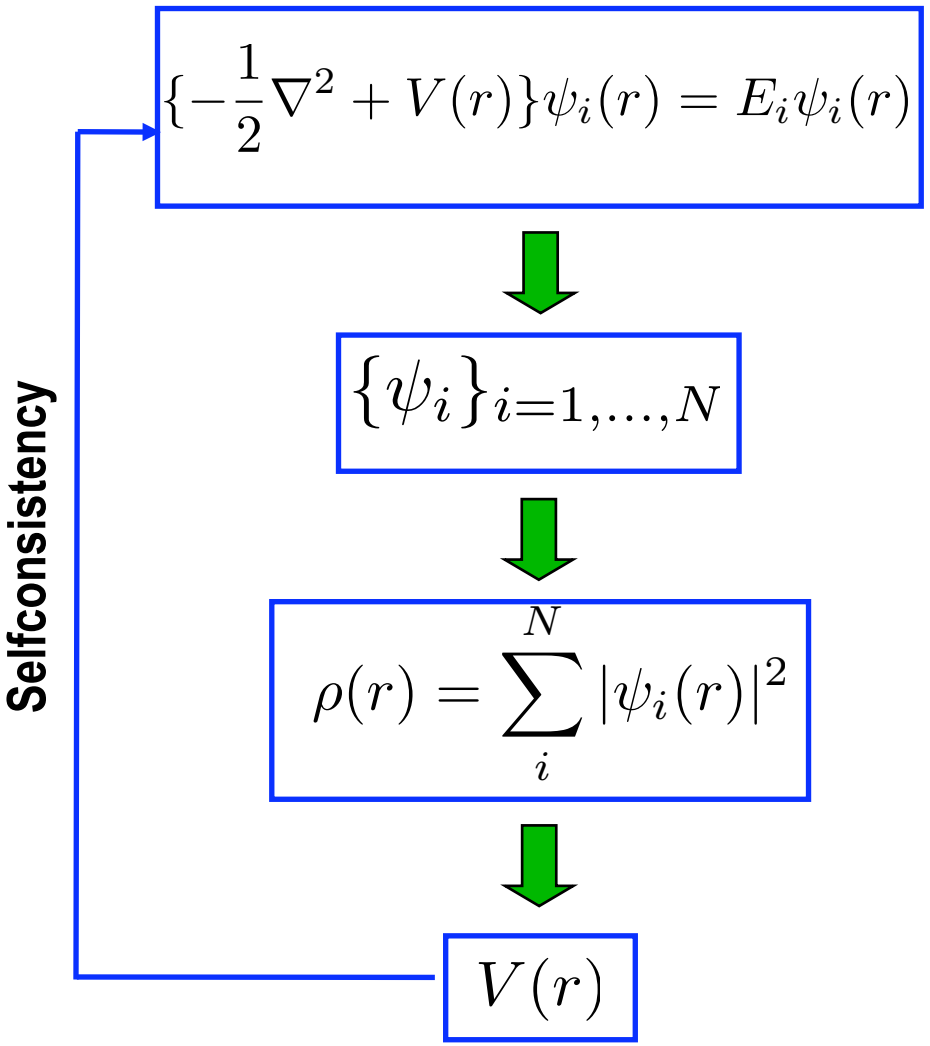
Approaches for solving the Kohn-Sham Equations

Solving the Kohn-Sham equations

- Self-Consistent Field (SCF) iteration
  - view as a linear eigenvalue problem
  - need to precondition
  - usually used with other acceleration techniques to improve convergence
  - no good convergence theory
- Direct Constrained Minimization
  - minimize the total energy directly
  - pose as a constrained optimization problem
  - also requires globalization techniques
- Invariance property

$$\begin{aligned} E(XQ) &= E(X) \\ H(XQ) &= H(X) \end{aligned} \quad \text{for any} \quad Q^*Q = I_{n_e}$$

Basic SCF Iteration



- Overall Complexity $O(N^3)$
- Major computational work (for plane wave codes):
  - 3D FFT
  - Orthogonalization
  - Nonlocal potential
- Parallel efficiencies can be quite high
- May converge slowly and sometimes doesn't converge at all
- Energy need not decrease monotonically

When can we expect SCF to work?

- SCF seeks to minimize a sequence of surrogate models
- Gradients match at $x^{(i)}$, i.e. $\nabla E(x^{(i)}) = \nabla E_{sur}(x^{(i)})$
- Consider simple 2D example:

$$E(x) = \frac{1}{2} x^T L x + \frac{\alpha}{4} \rho(x)^T L^{-1} \rho(x)$$

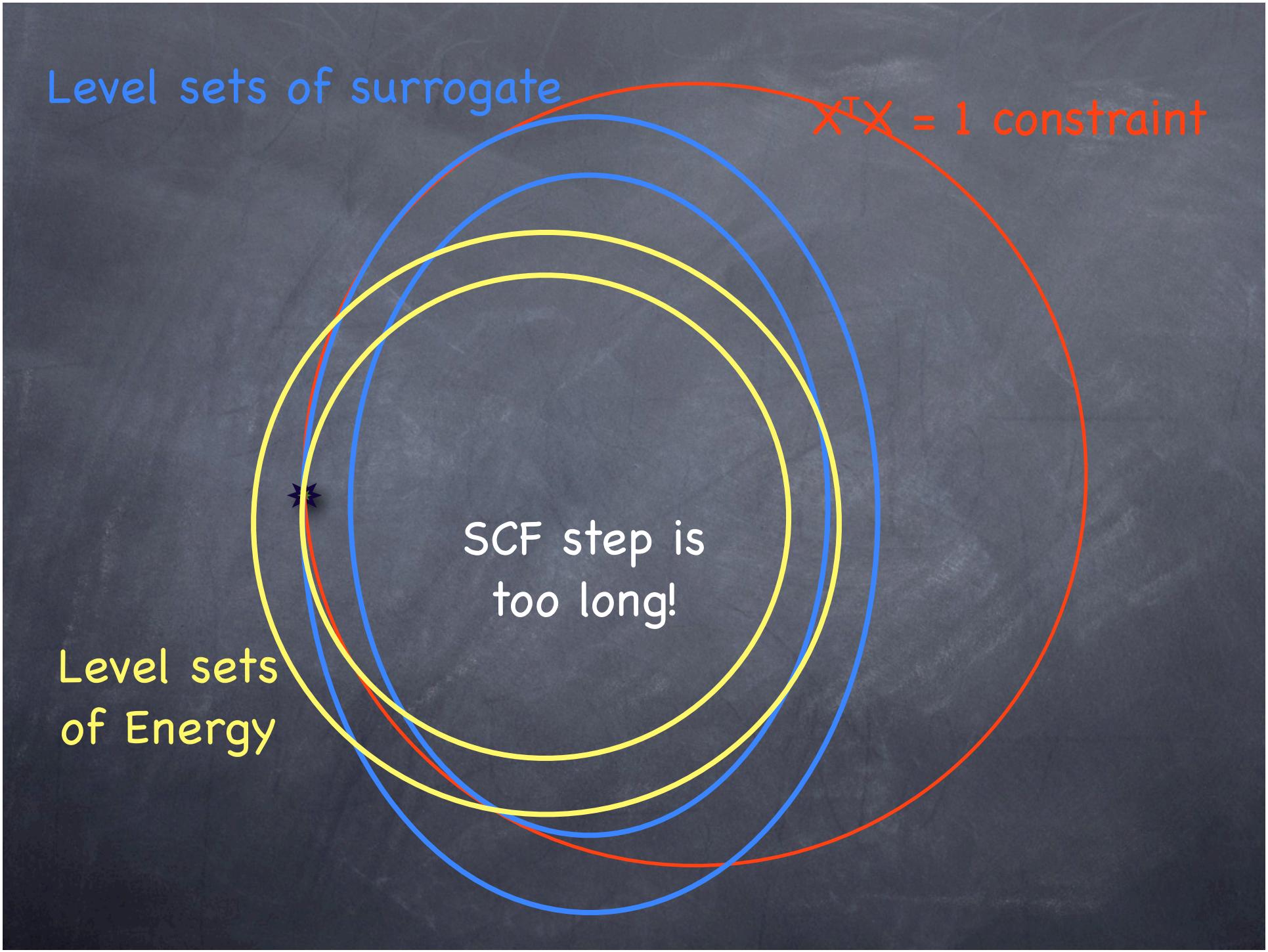
$$L = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \rho(x) = \begin{pmatrix} x_1^2 \\ x_2^2 \end{pmatrix}$$

$$\begin{aligned} &\min E(x) \\ &\text{s.t.} \quad x_1^2 + x_2^2 = 1 \end{aligned}$$

$$\left[L + \alpha \mathrm{Diag}(L^{-1} \rho(x)) \right] x = \lambda_1 x$$

Convergence of SCF on toy problem

α	$\|res\|$	iter
2	$1.47 \cdot 10^{-11}$	14
3	$4.01 \cdot 10^{-11}$	22
4	$3.78 \cdot 10^{-11}$	35
5	$7.60 \cdot 10^{-11}$	58
6	$1.07 \cdot 10^{-10}$	126
7	$1.41 \cdot 10^{-1}$	200
8	$1.03 \cdot 10^{0}$	200


Level sets of surrogate
$X^TX = 1$ constraint
SCF step is too long!
Level sets of Energy

Improving SCF

- Construct better surrogate – cannot afford to use local quadratic approximations (Hessian too expensive)
- **Charge mixing** to improve convergence; related to Broyden methods
- Use **trust region** to restrict the update to stay within a neighborhood of the gradient matching point
  - Level-Shifting (Saunders & Hillier 1973)
  - Cances & LeBris 2000
  - TRSCF – Thogersen, Olsen, Yeager & Jorgensen 2004; Francisco, Martinez, Martinez 2006; Yang, Meza, Wang (2007)

Charge Mixing Schemes

- Simple mixing

$$\rho^{(i+1)} \leftarrow \tau\rho_{in}^{(i)} + (1-\tau)\rho_{out}^{(i)}, \quad 0 < \tau < 1.$$

- Pulay mixing (Direct Inversion of Iterative Subspace)

$$\rho^{(i+1)} = \sum_{j=1}^{i} \alpha_j \rho^{(j)}, \quad \sum_{j=1}^{i} \alpha_j = 1$$

- Broyden mixing

$$\rho^{(i+1)} = \rho^{(i)} + \tau C_{i+1} r_i$$

- Anderson mixing

$$\rho^{(i+1)} = \rho^{(i)} + \tau r_i + (S_i - \tau Y_i) Y_i^{\dagger} r_i$$

Trust Region Subproblem

- Solve

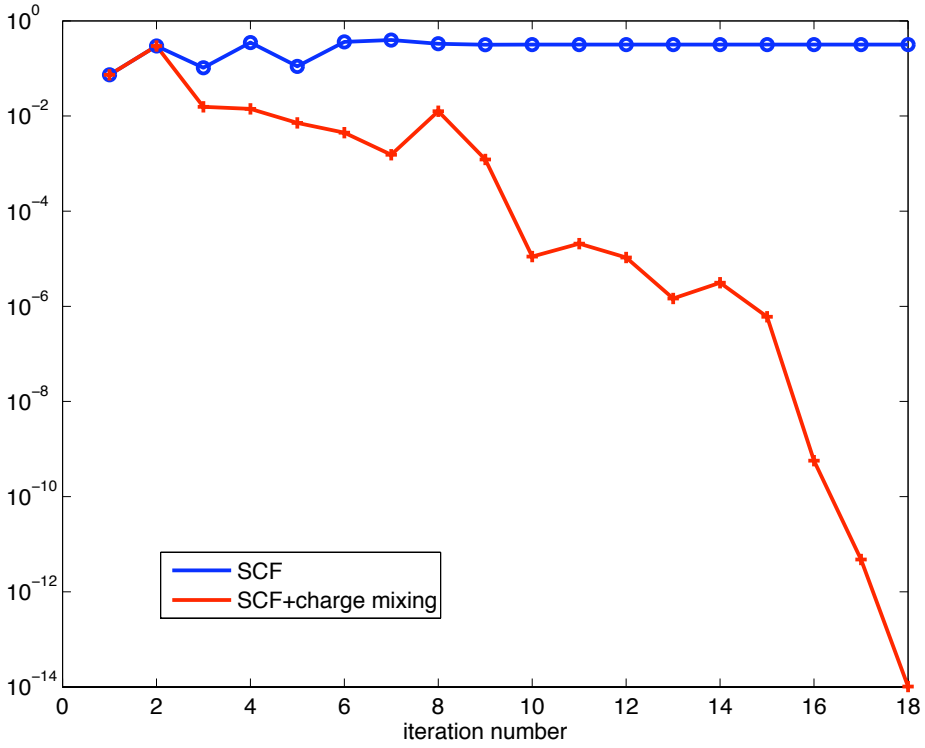
$$\begin{aligned} \min \quad & E_{sur}(x) \\ \text{s.t.} \quad & \\ & x^T x = 1, \\ & \|xx^T - x^{(i)}(x^{(i)})^T\|_F^2 \leq \Delta \qquad \text{trust region constraint} \end{aligned}$$

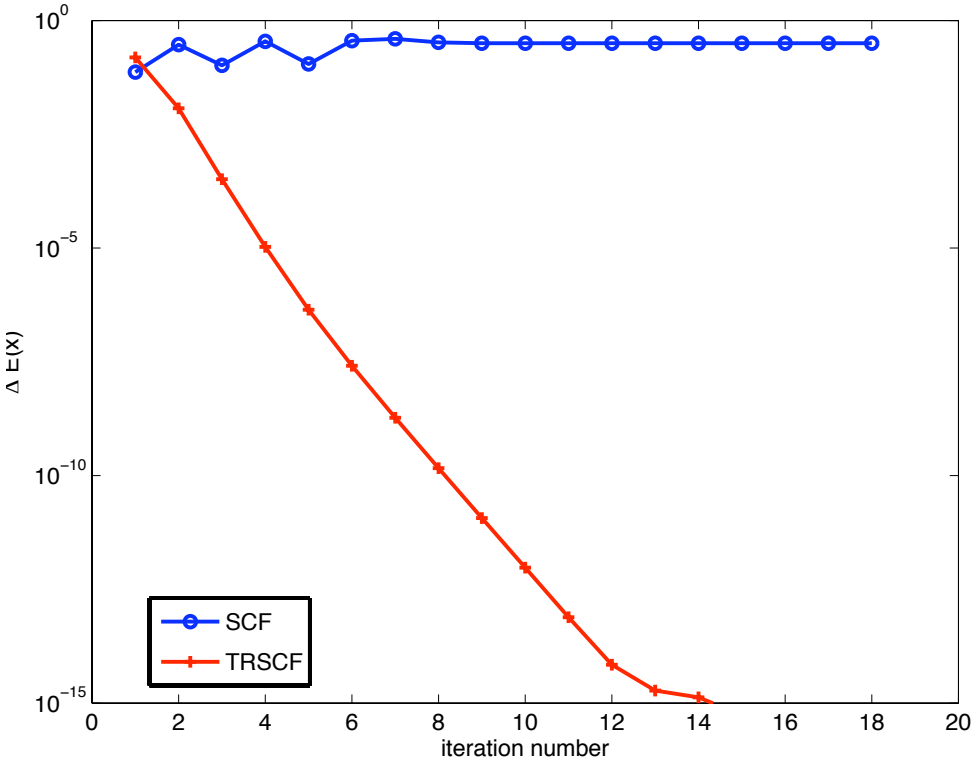
- Equivalent to solving

$$\begin{aligned} \left[H(x^{(i)}) - \sigma x^{(i)}(x^{(i)})^T\right] x &= \lambda x \\ x^T x &= 1 \end{aligned}$$

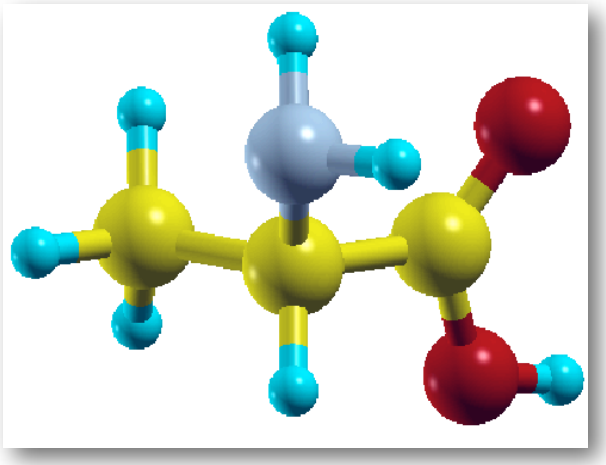
- σ is a penalty parameter (Lagrange multiplier for TR)

Comparison of TRSCF vs. mixing

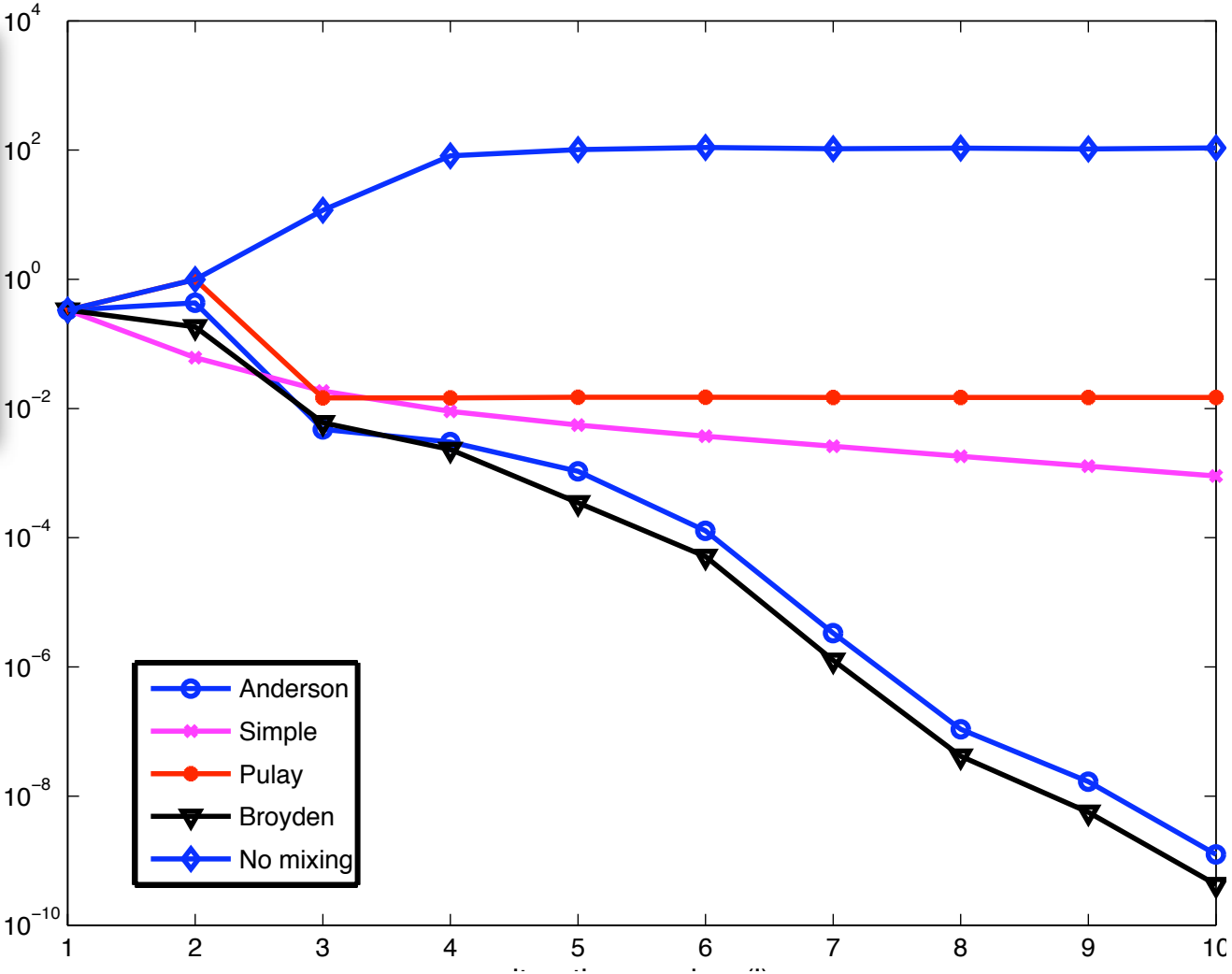




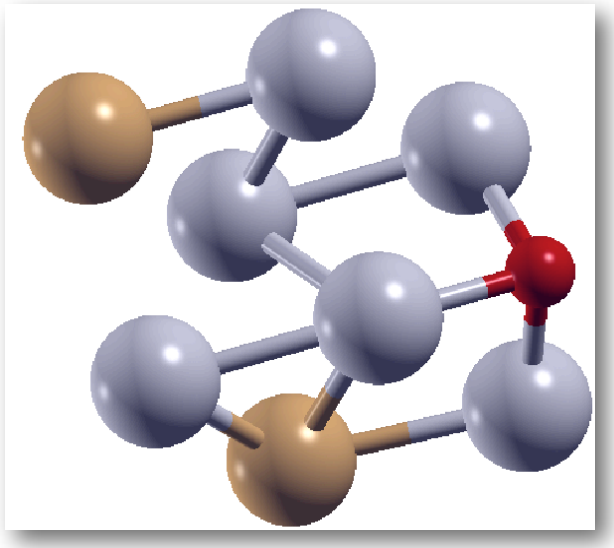
Comparison of other charge mixing schemes



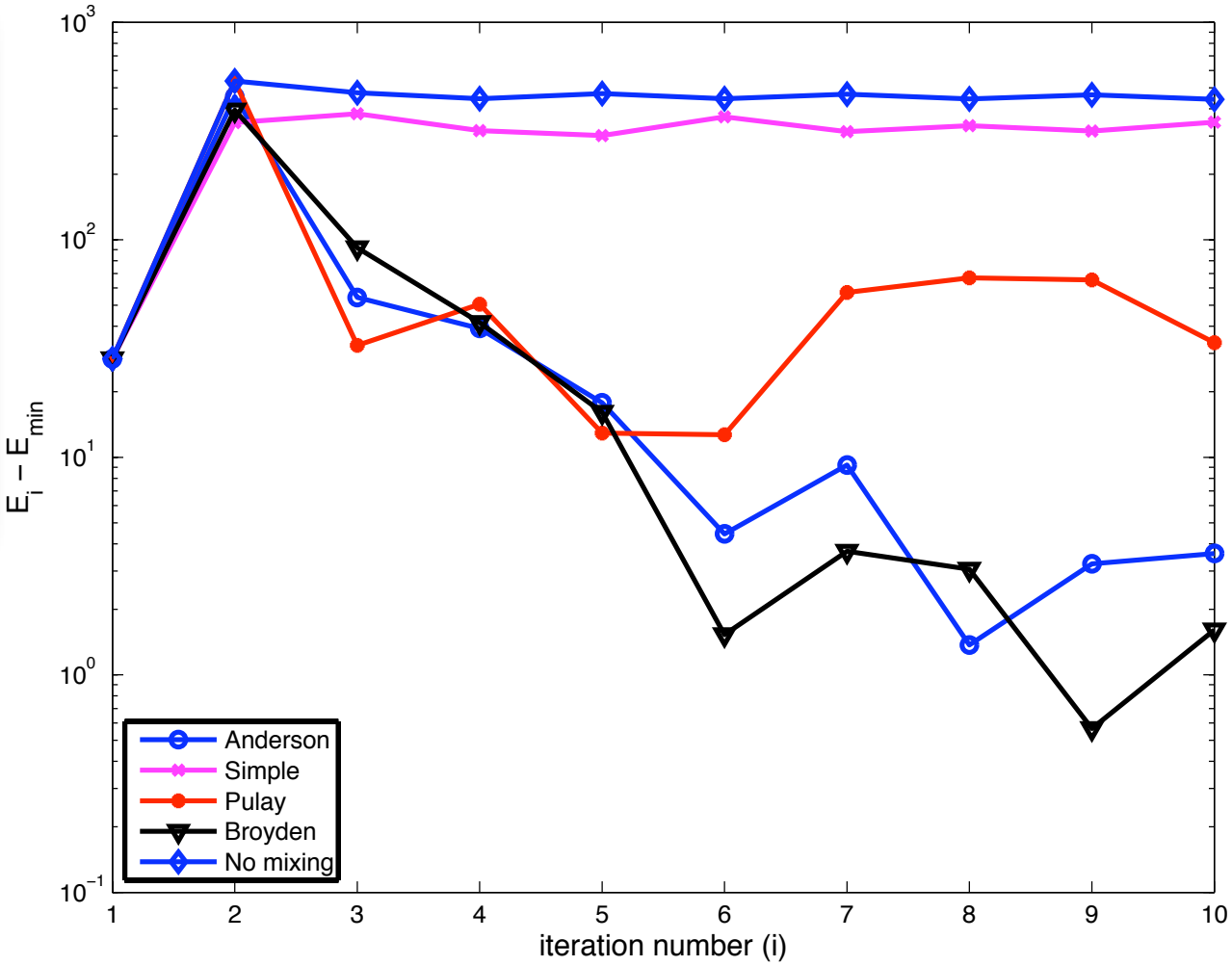
Alanine



Charge mixing can fail



Pt_6Ni_2O



Direct Constrained Minimization

- Assume $x^{(i)}$ is the current approximation
- Idea: minimize the energy in a certain (smaller) subspace
- Update $x^{(i+1)} = \alpha x^{(i)} + \beta p^{(i-1)} + \gamma r^{(i)}$;
  - $p^{(i-1)}$ previous search direction;
  - $r^{(i)} = H^{(i)} x^{(i)} - \theta^{(i)} x^{(i)}$;
  - choose α, β and γ so that
    - $x_{k+1}^T x_{k+1} = 1$;
    - $E(x_{k+1}) < E(x_k)$;

Remark 1: A nonlinear CG-like algorithm

Remark 2: Extension of LOBPCG (Knyazev) to nonlinear EV

Subspace minimization

- Let $V = (x^{(i)}, p^{(i-1)}, r^{(i)})$; $x^{(i+1)} = Vy$, for some y;
- Solve

$$\min_{y^T V^T V y = 1} E(Vy)$$

- Equivalent to solving

$$\begin{aligned} G(y)y &= \lambda By \\ y^T By &= 1 \end{aligned}$$

where $B = V^T V$ and $G(y) = V^T[L + \alpha \mathrm{Diag}(L^{-1}\rho(Vy))]V$

DCM Algorithm

- Input: Initial guess
- Output: X such that E_{KS} is minimized

1. $P^{(0)} = [], \mathrm{i} = 0;$
2. while (not converged)
   (a) $\Theta^{(i)} = X^{(i)^*} H^{(i)} X^{(i)};$
   (b) $R^{(i)} = H^{(i)} X^{(i)} - X^{(i)} \Theta^{(i)};$
   (c) Set $Y = (X^{(i)}, P^{(i-1)}, K^{-1} R^{(i)});$
   (d) Solve $\min_{G^* Y^* Y G = I_k} E_{tot}(YG);$
   (e) $X^{(i+1)} = YG(1 : n_e, :);\ P^{(i+1)} = YG(n_e + 1 : 3n_e, :);$
   (f) $i \leftarrow i + 1;$

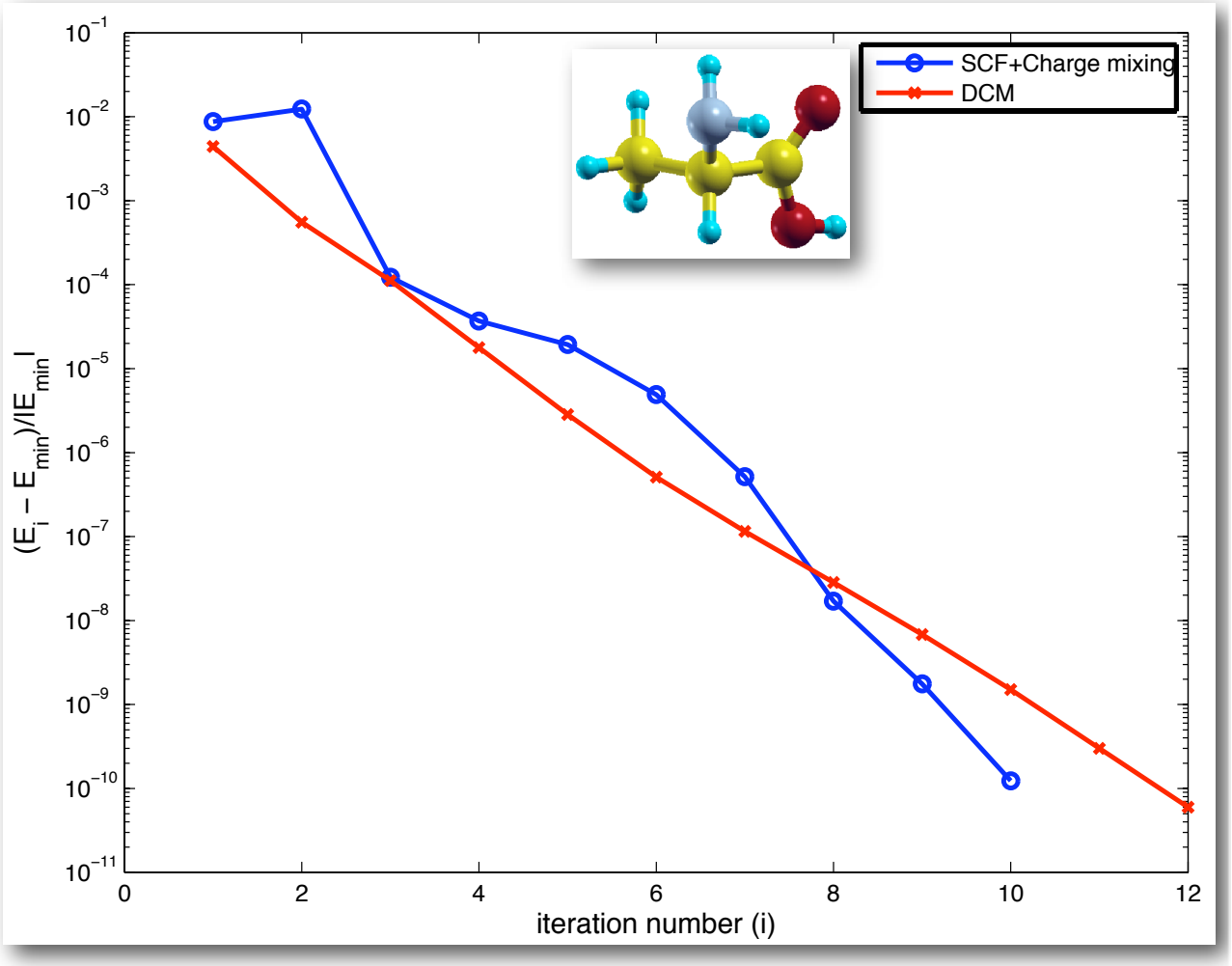
C. Yang, J. Meza, L. Wang, A Constrained Optimization Algorithm for Total Energy Minimization in Electronic Structure Calculation, J. Comp. Phy., 217 709-721 (2006)

Test problems

- KSSOLV Matlab code for solving the Kohn-Sham equations
  - Open source package
  - Handles SCF, DCM, Trust Region
  - Various mixing strategies
- Example problems: alanine and graphene
- Tests run on desktop computer

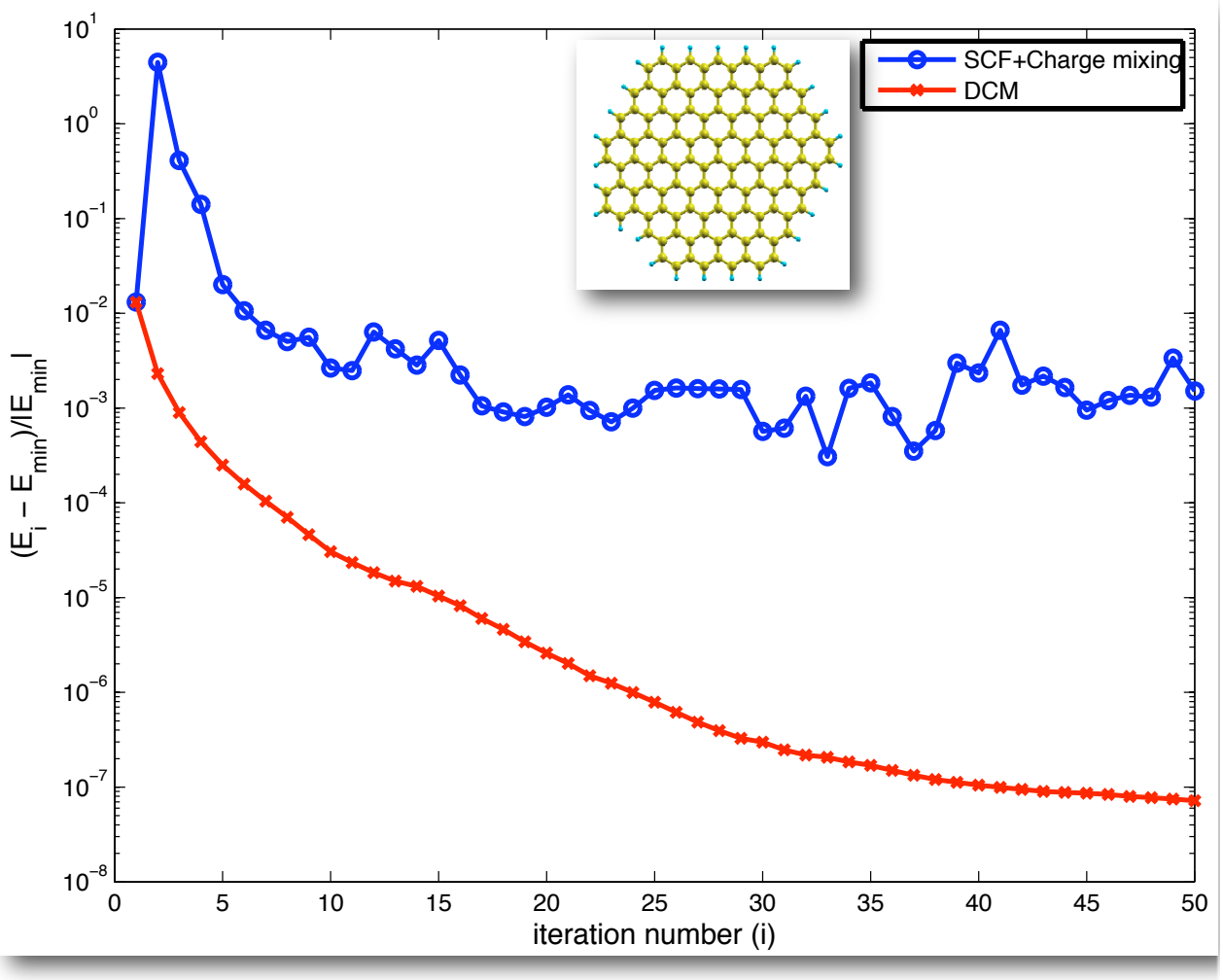
Example 1: Alanine

- sampling grid:
  - 96 x 48 x 96 (ecut=25 Ryd)
- 10 PCG iterations / SCF outer iteration
- 3 inner SCF iteration / DCM outer iteration
- supercell:
  - 20 x 15 x 20



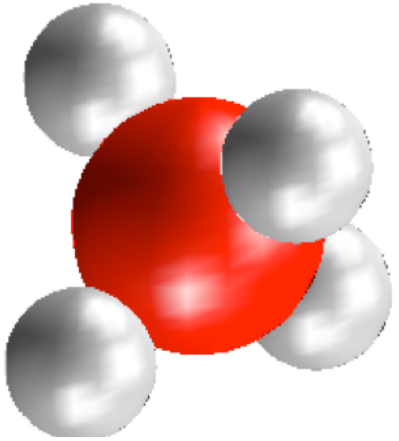
Example 2: Graphene

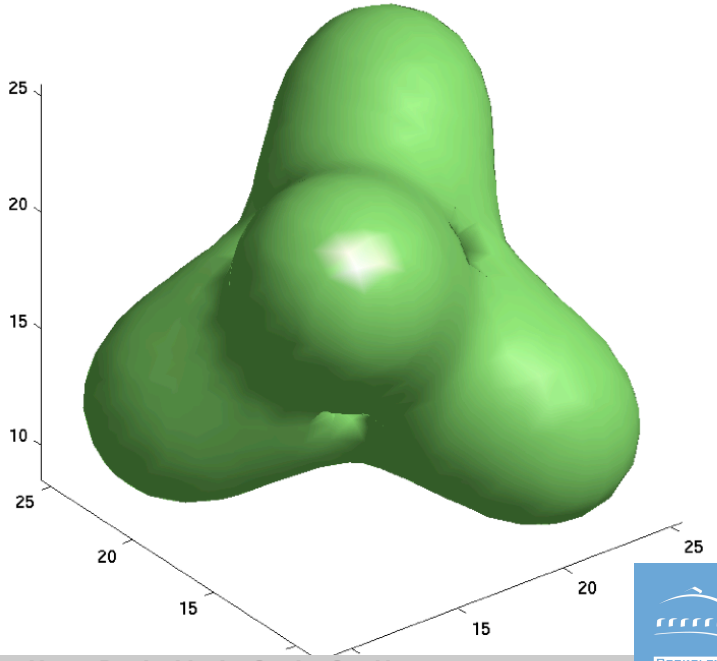
- sampling grid:
  - 114 x 114 x 15
- 10 PCG iterations / SCF outer iteration
- 5 inner SCF iteration / DCM outer iteration
- supercell:
  - 40 x 40 x 5



Example: SiH_4

```
a1 = Atom('Si');
a2 = Atom('H');
alist = [a1 a2 a2 a2 a2];
xyzlist= [
 0.0   0.0   0.0
 1.61  1.61  1.61
 ... ];
mol = Molecule();
mol = set(mol,'Blattice',BL);
mol = set(mol,'atomlist',alist);
mol = set(mol,'xyzlist' ,xyzlist);
mol = set(mol,'ecut', 25);
mol = set(mol,'name','SiH4');
...
[Etot, X, vtot, rho] = dcm(mol);
isosurface(rho);
```





Comparison of DCM vs. SCF using KSSOLV

system	SCF time	DCM time	SCF error	DCM error
C_2H_6	26	25	9.4 e-6	3.5 e-6
CO_2	26	23	3.1 e-3	1.1 e-4
H_2O	16	16	5.7 e-5	2.2 e-5
$HNCO$	34	32	7.4 e-3	6.8 e-5
Quantum dot	18	16	5.0 e-3	3.7 e-1
Si_2H_4	25	23	1.8 e-3	2.7 e-4
silicon bulk	15	15	3.0 e-4	9.6 e-6
SiH_4	20	19	9.7 e-6	4.9 e-7
Pt_2Ni_6O	415	281	3.7 e0	4.9 e-2
pentacene	887	493	5.2 e-1	2.5 e-2

Summary

- Despite dire warnings, mathematical techniques actually help in chemistry
- New approach for solving the Kohn-Sham equations using a direct optimization method improves convergence
- Trust region modification increases robustness of both SCF and DCM
- New computational software tools for modeling and simulation of nanosystems

Where do we go from here

- Investigate new algorithms to speed up convergence
- Develop more accurate methods
- Expand applicability of methods to new systems, (metals)
- Develop linear scaling versions of DCM

First Nanoscientists?

The New York Times

February 21, 2005

The First Nanotechnologists

Ancient stained-glass makers knew that by putting varying, tiny amounts of gold and silver in the glass, they could produce the red and yellow found in stained-glass windows. Similarly, today's scientists and engineers have found that it takes only small amounts of a nanoparticle, precisely placed, to change a material's physical properties.

Gold particles in glass

Size*: 25 nm
Shape: sphere
Color reflected:

100 nanometers = 0.0001 millimeter

Silver particles in glass

Size*: 100 nm
Shape: sphere
Color reflected:

Had medieval artists been able to control the size and shape of the nanoparticles, they would have been able to use the two metals to produce other colors. Examples:

Size*: 50 nm
Shape: sphere
Color reflected:

Size*: 40 nm
Shape: sphere
Color reflected:

Size*: 100 nm
Shape: sphere
Color reflected:

Size*: 100 nm
Shape: prism
Color reflected:

Source: Dr. Chad A. Mirkin, Institute of Nanotechnology, Northwestern University

*Approximate