Optimization: The Difference Between Theory and Practice ...

Juan Meza

Senior Scientist and Department Head

High Performance Computing Research

Lawrence Berkeley National Laboratory

SIAM Annual Meeting Denver, CO July 6-10, 2009

In theory there is no difference between theory and practice.

In practice there is.

Yogi Berra

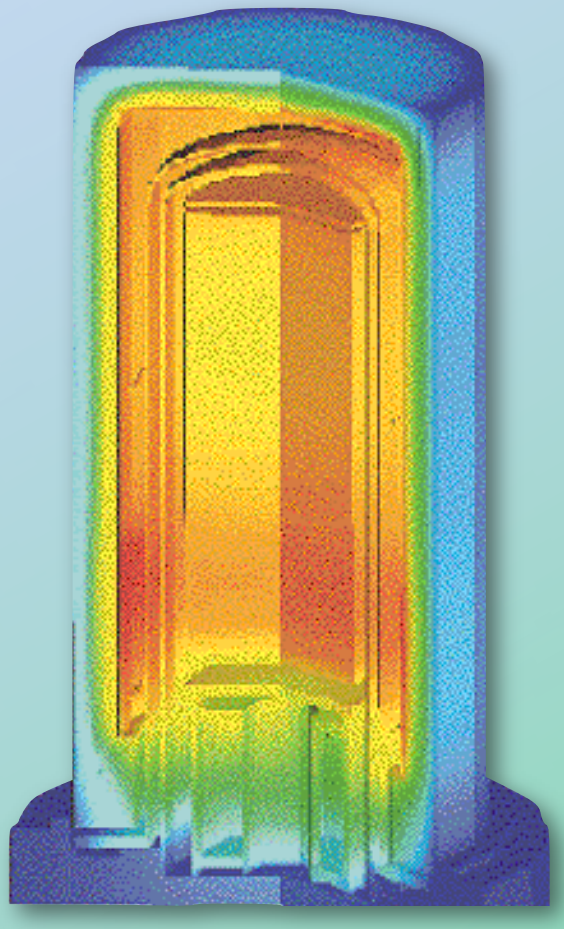
Optimization Status

- Optimization has become increasingly widespread as computational science has expanded.
- Variety of forms/applications including inverse problems, parameter estimation, energy minimization, etc.
- Wide range of problems with diverse characteristics, including nonlinear, nonconvex, mixed integer, bi-level, stochastic, global, etc.
- Most problems exceedingly difficult to analyze theoretically, yet solutions must be obtained.

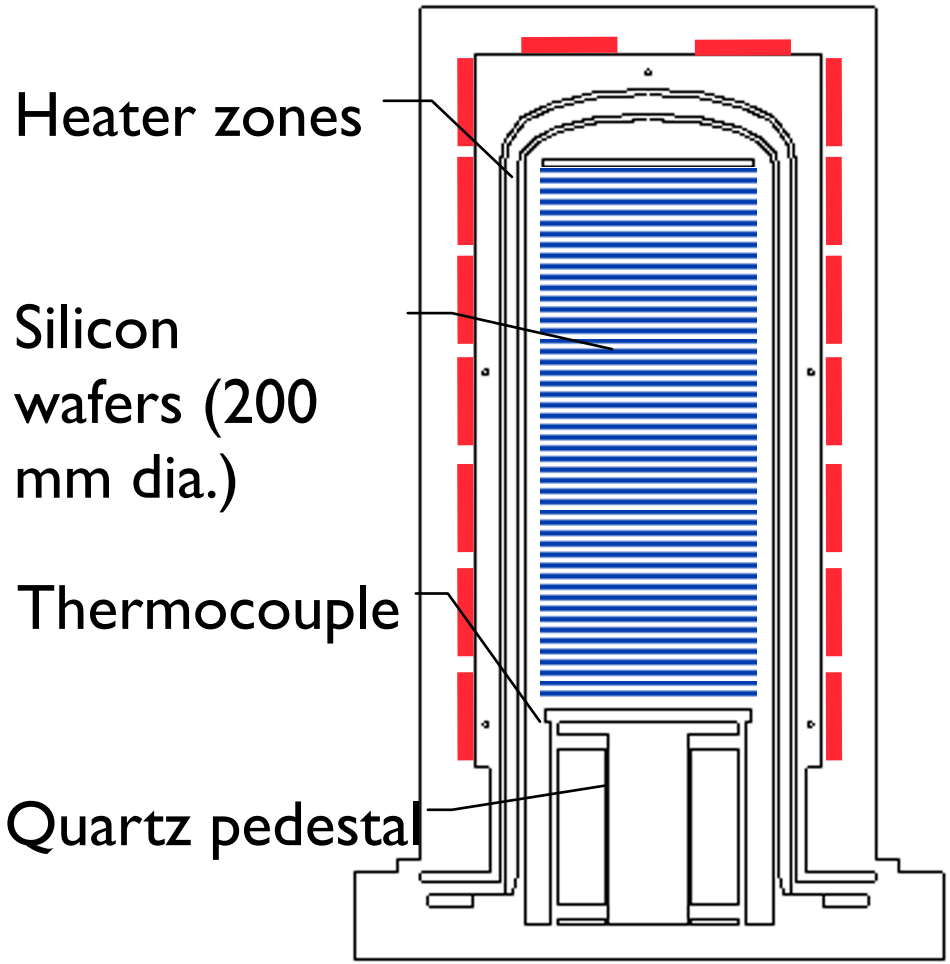
Challenges in simulation-based optimization

- Objective function is smooth
  - Often true, but simulations can create noisy behavior
- Twice continuously differentiable
  - Difficult to prove
- Constraints
  - Users can sometimes over-specify or incorrectly guess constraints
  - Require strict feasibility
- Computationally expensive objective functions
  - Dominant cost is evaluation of objective function

Chemical Vapor Deposition Control



The design of a small-batch fast-ramp LPCVD furnace



- Temperature uniformity across the wafer stack is critical
- Independently controlled heater zones regulate temperature
- Wafers are radiatively heated
- Design parameters:
  - Number of heater zones
  - Size/position of heater zones
  - Wafer pitch
  - Pedestal configuration
  - Insulation thickness
  - Baseplate cooling

Standard least squares problem formulation

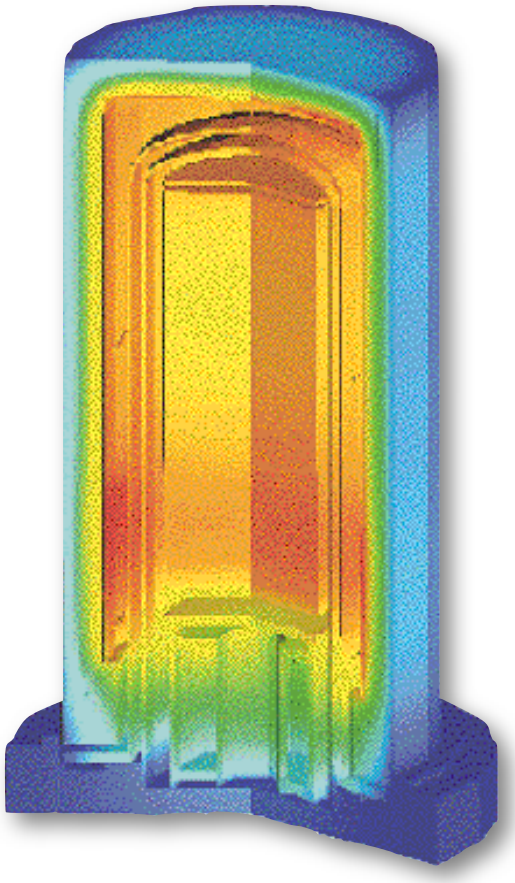
$$\min f(p) = \sum_i (T_i(p;x) - T_{target})^2$$

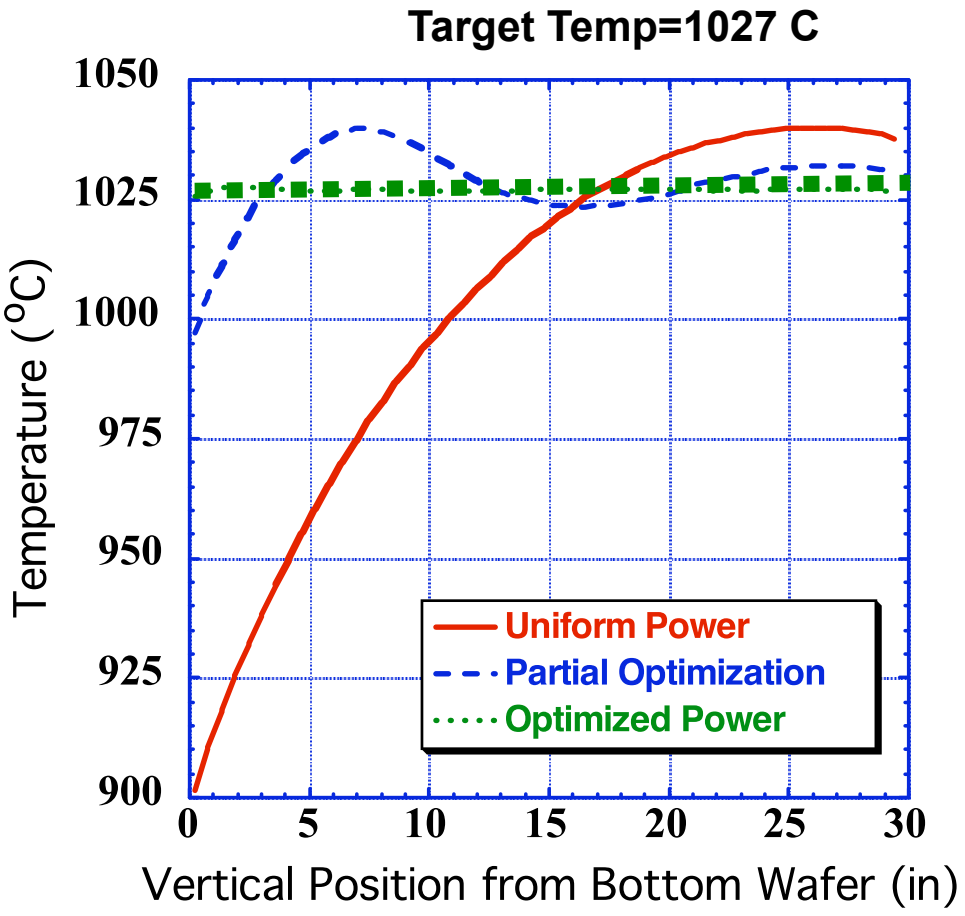
- Finding temperatures involves solving a heat transfer problem with radiation
- Adjusting tolerances in the PDE solution trades off noise with CPU time
- Larger tolerances lead to
  - Less accurate PDE solutions
  - Less time per function evaluation

In Theory You Have Derivatives ...

- But in practice you don't
- Lack of derivatives reduces the choice of optimization methods one can readily use.
  - Need to consider alternative methods
- Accurately computing derivatives much harder than one might think at first glance

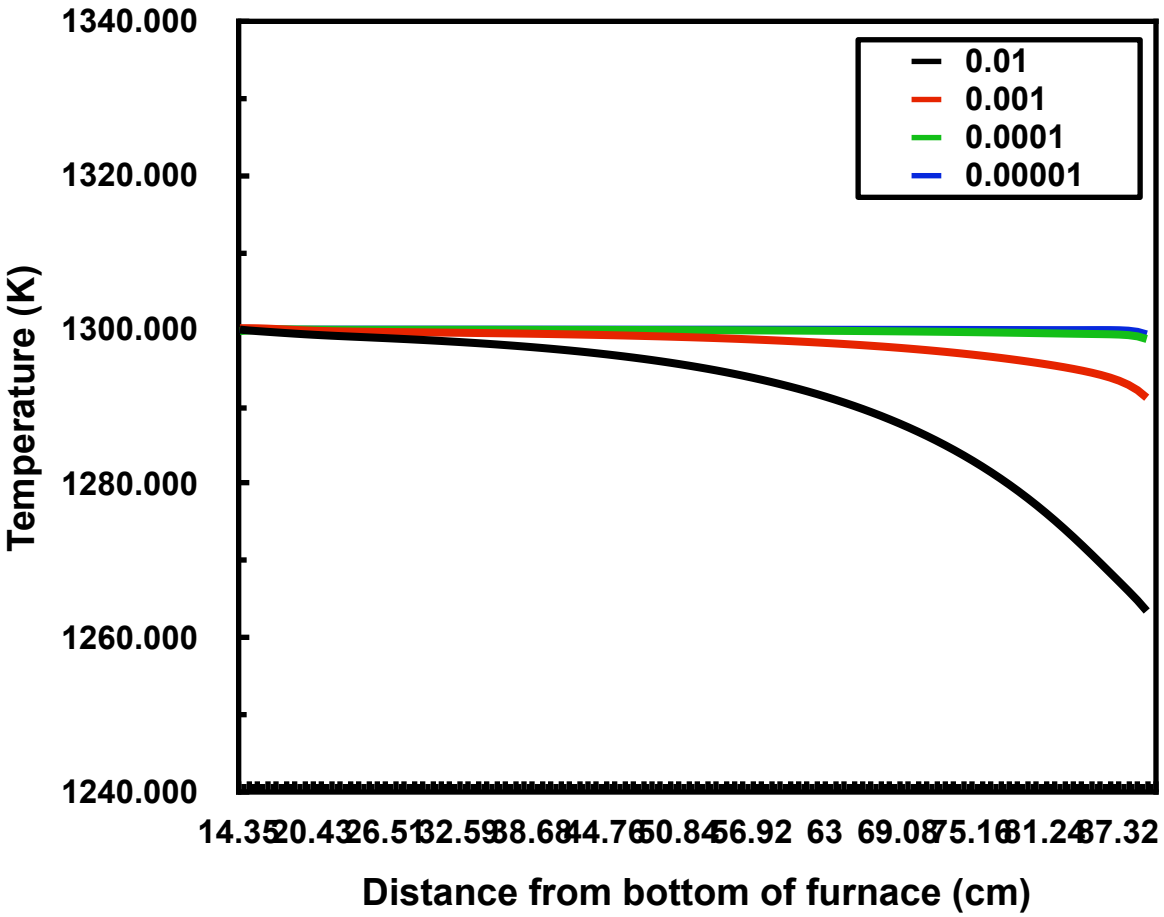
Optimized power distribution enhances wafer temperature uniformity for steady-state





Simulation of Equipment Design Optimization in Microelectronics Manufacturing, **J.C. Meza, C.H. Tong, C.D. Moen, Proc. 30th Annual Simulation Symposium, Atlanta, GA, April 7-9, 1997.**

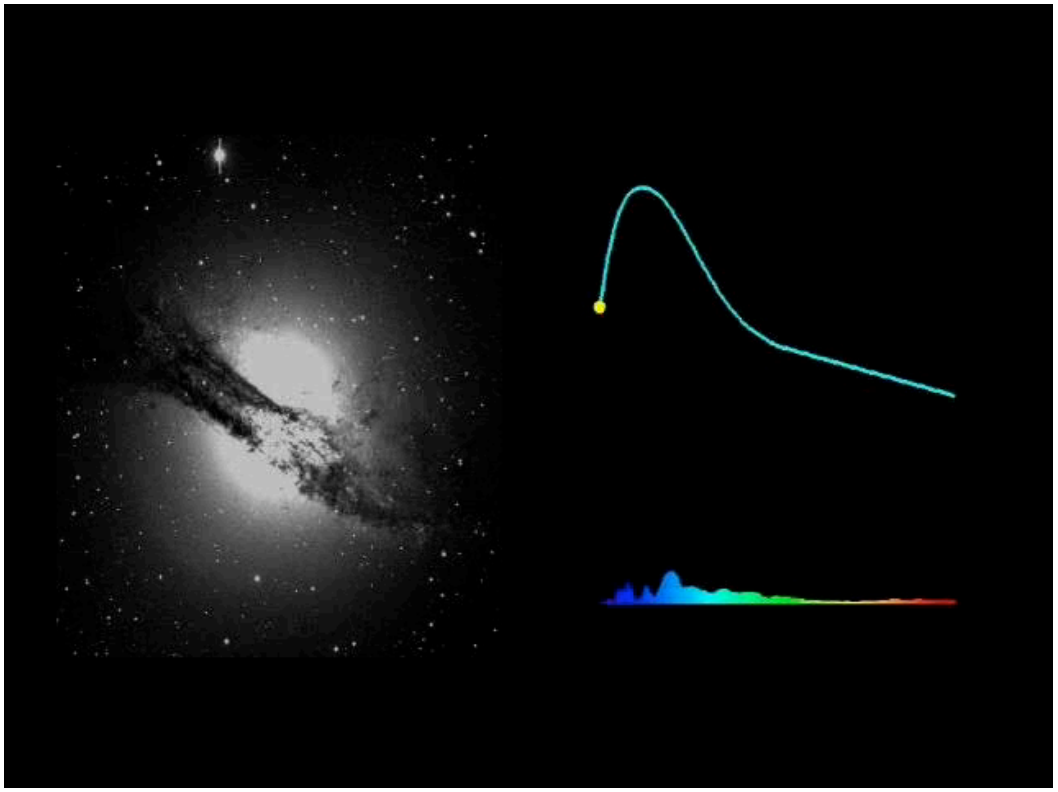
Wafer Temperatures for Optimal Powers Obtained by TRPDS



- Automatic Differentiation
- Pattern Search Method
- Parallelization combined within a trust-region framework

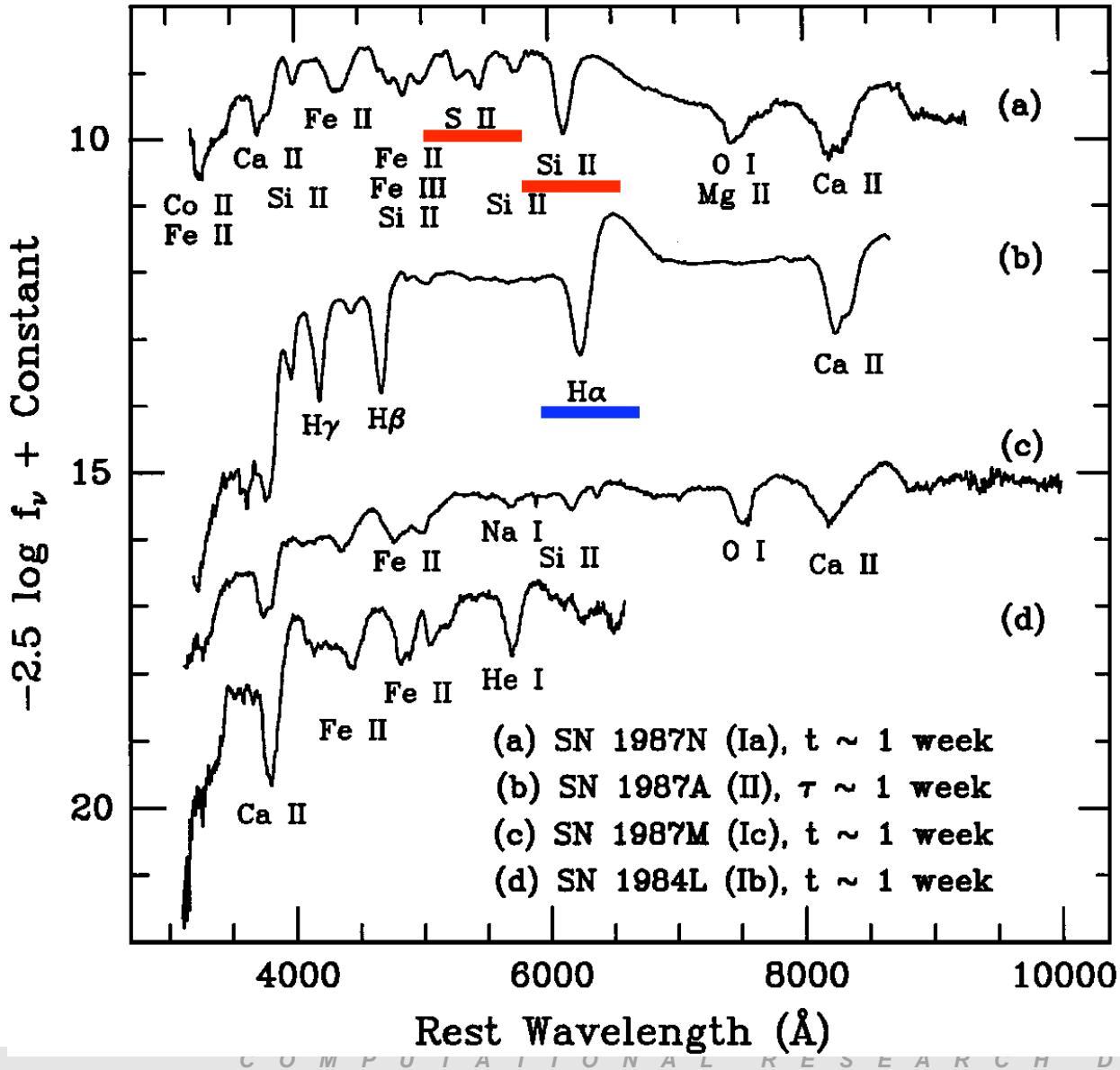
P.D. Hough and J.C. Meza, A Class of Trust Region Methods for Parallel Optimization. SIAM Journal of Optimization, Vol. 13, No. 1, pp. 64-282, 2002.

Supernova Spectra



- Supernova (SN) cosmology is the most powerful and best proven technique to date for probing the dark energy
- Most SN reach maximum light a few weeks after outburst
- Light curves evolve on day to week time scales
- Fade away over months to years.

Spectra



Composition
Density
Temperature
Energy deposited
Kinetic energy
Environment

Spectra indicative of rapid ejecta outflow up to 0.1c.

Introduction to Spectrum Synthesis

- Spectrum synthesis is the computation of a theoretical spectrum of a model atmosphere based upon the known laws of physics.
- The goals of spectrum synthesis for supernovae are three-fold:
  - Calculate the spectra of detailed explosion models, compare with observations and falsify or validate the models
  - Calculate various "toy models" in order to predict or explain a set of observations
  - Parameterize a model atmosphere and calculate the subsequent spectra, compare these directly to observation

Courtesy of Peter Nugent, LBNL

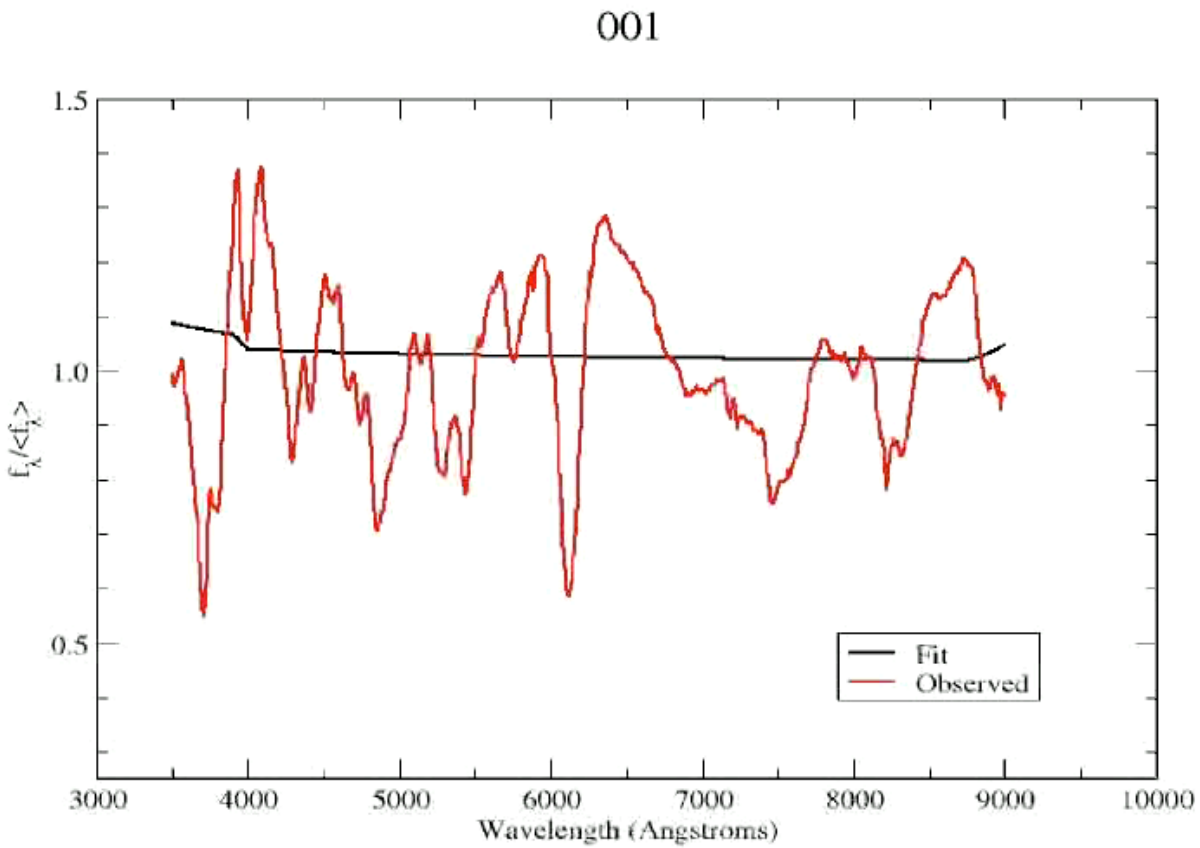
Problem is a least squares problem

- A simulation code produces synthetic spectra based on inputs that describe the elements that are present in the model atmosphere
- Goal is to find the correct set of elements that best fit the synthetic spectra to the observed data
- Must have an solution within 24 hours
- No analytic derivatives available

In Theory The Functions are Inexpensive ...

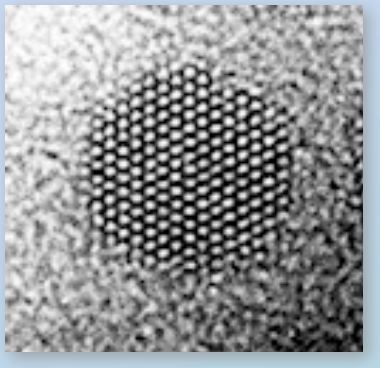
- But in practice the objective functions can dominate the total cost of the solution
- May not matter how expensive the optimization algorithm is as long as you can minimize the total number of function calls
- Sometimes convergence is dictated by how big your computer budget is and/or time constraints

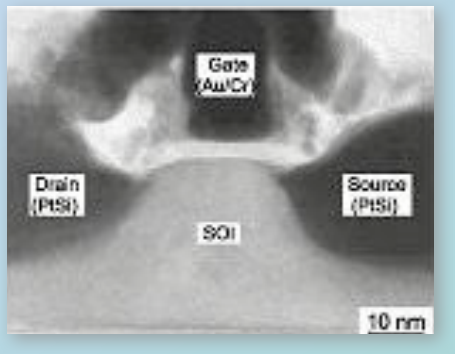
Real-Time Assessment of Data for SN Factory

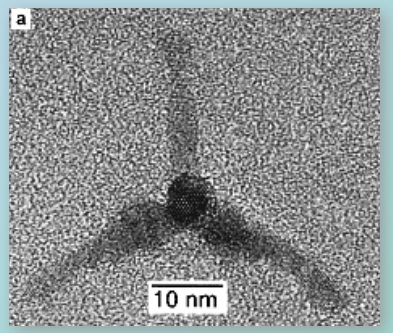


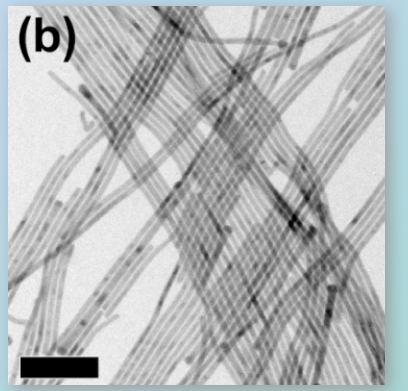
- Simulation Inputs:
  - Temperature
  - Velocity
  - Density
  - Optical depth
  - Min and Max velocity
  - T_{eff}
- Parallel Pattern Search
- Solution found within 1 hour

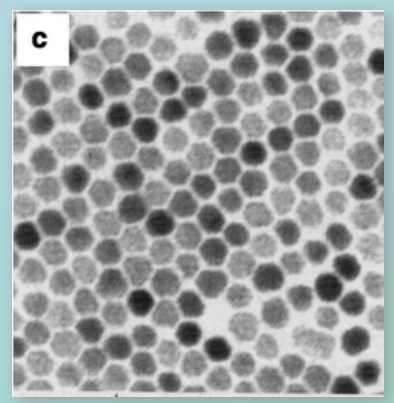
Computational Nanoscience

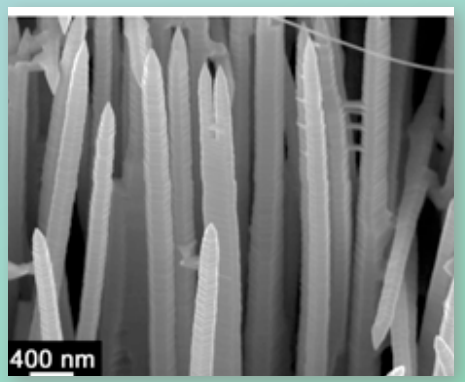




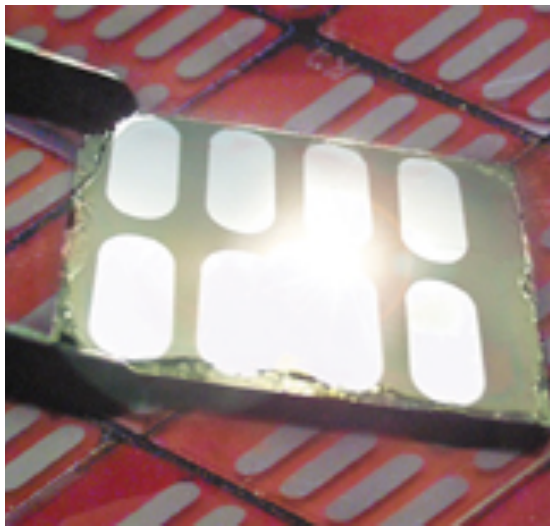


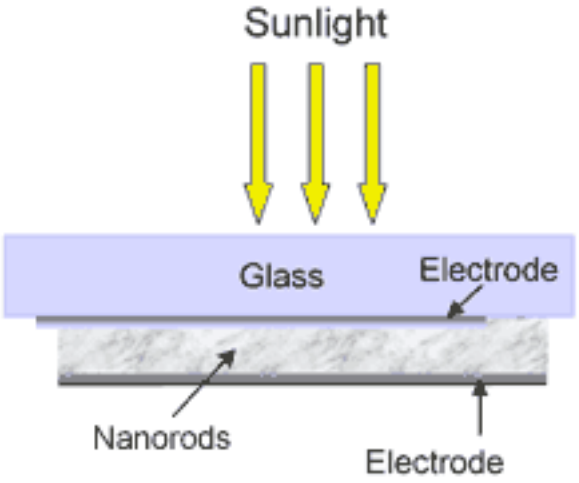






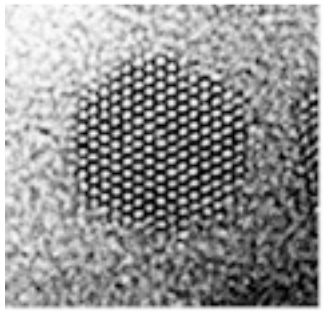
Photovoltaic Solar Cells





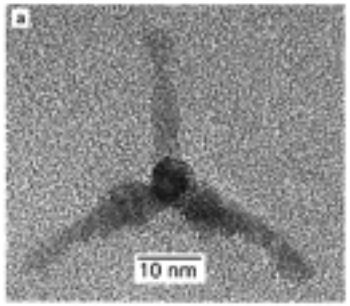
- Solar cells based on inorganic nanorods and semiconducting polymers
- Nanorods can be made of CdSe, a semiconducting material
- Nanorods act like wires, absorbing light and generating hole-electron pairs
- Biggest challenge is cost, ~30 cents/kWh

Why do we need simulations



CdSe QD

1000-5000

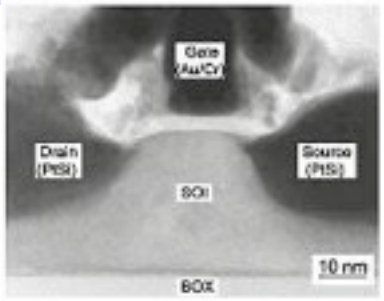


Tetrapod

10,000

Solar Cell

50,000



CMOS

1,000,000 atoms

*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 transfer 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

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 and the Kohn-Sham equations

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

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

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

$$\left[-\frac{1}{2}\nabla^2 + V_{ext}(r) + \int \frac{\rho}{|r-r'|} + V_{xc}(\rho)\right]\psi_i = \epsilon_i\psi_i$$

Kohn-Sham Equations

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

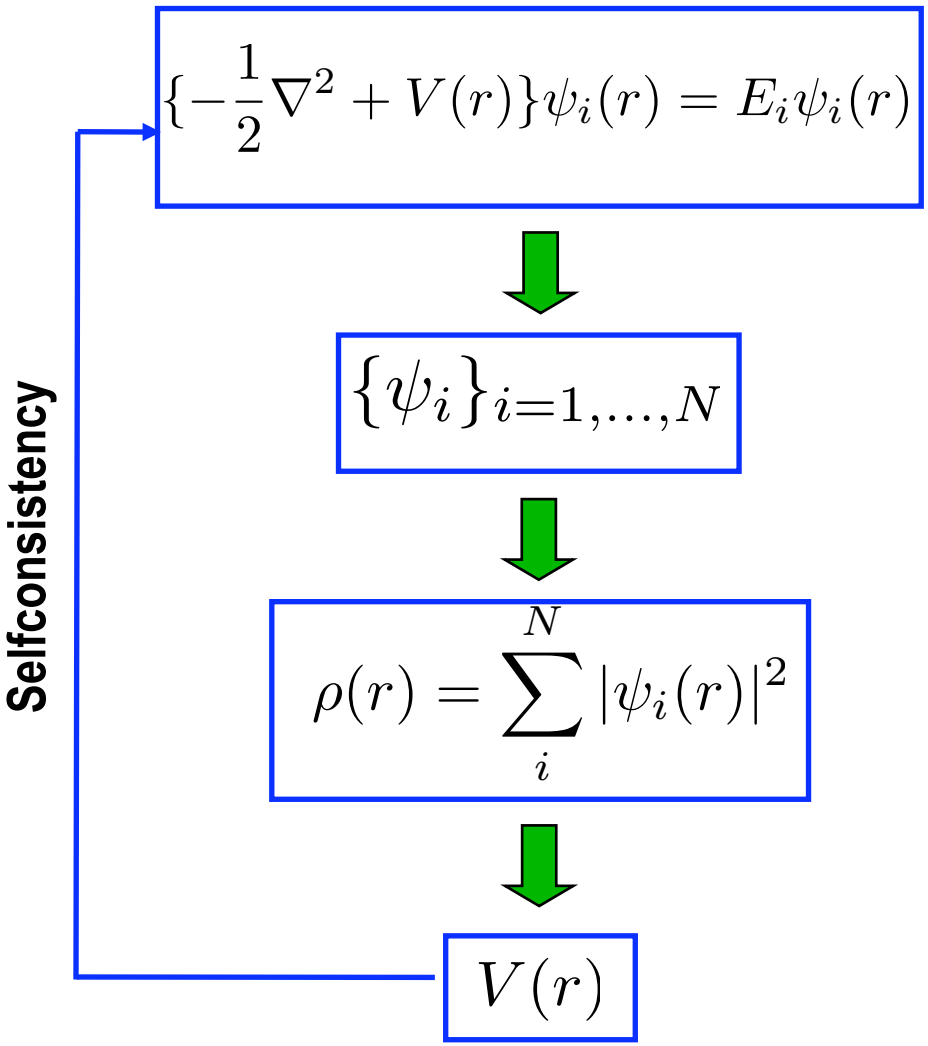
$$
\begin{aligned}
H\psi_i &= \epsilon_i \psi_i, \quad i = 1, 2, ..., n_e \\
H &= \left[-\frac{1}{2}\nabla^2 + V_{ion}(r) + \int \frac{\rho}{|r - r'|} + V_{xc}(\rho) \right]
\end{aligned}
$$

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

Solving the Kohn-Sham equations

- Self-Consistent Field (SCF) iteration
  - view as a (large) linear eigenvalue problem
  - usually used with other acceleration techniques to improve (ensure) convergence
  - convergence theory sparse
- Direct Constrained Minimization
  - minimize the total energy directly
  - also requires globalization techniques
  - convergence theory still difficult, but can use existing frameworks

Basic SCF Iteration



- Most of the work is in solving the sequence of linear eigenvalue problems
- Orthogonality constraint for the wavefunctions must be enforced explicitly
- May converge slowly and sometimes doesn't converge at all

In theory ...

- Derivatives are available!
- Orthogonalization constraints imply that we need to find not just one solution but an invariant subspace
- Working with large-scale problems again reduces the number of options available.

Improving SCF

- Construct better surrogate – cannot afford to use local quadratic approximations (Hessian too expensive)
- Charge mixing to improve convergence (heuristic)
- Use trust region to restrict the update to stay within a neighborhood of the gradient matching point
- TRSCF – Thogersen, Olsen, Yeager & Jorgensen (2004)
- DCM – Yang, Meza, Wang (2007)

Electronic Structure Theory

Mathematically, all problems arising in electronic structure theory are nonlinear minimization problems, with possible lacks of compactness at infinity ... and are subject to a constraint.

C. Le Bris, Proc. of the International Congress of Mathematicians (2006)

Direct Constrained Minimization

- Trick is to solve a sequence of smaller nonlinear eigenvalue problem
- Minimize energy in a particular subspace
- Construct minimization problem so that constraints are satisfied automatically
- Add trust region to ensure robustness
- Resulting method is more robust and faster (especially for larger systems)

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)

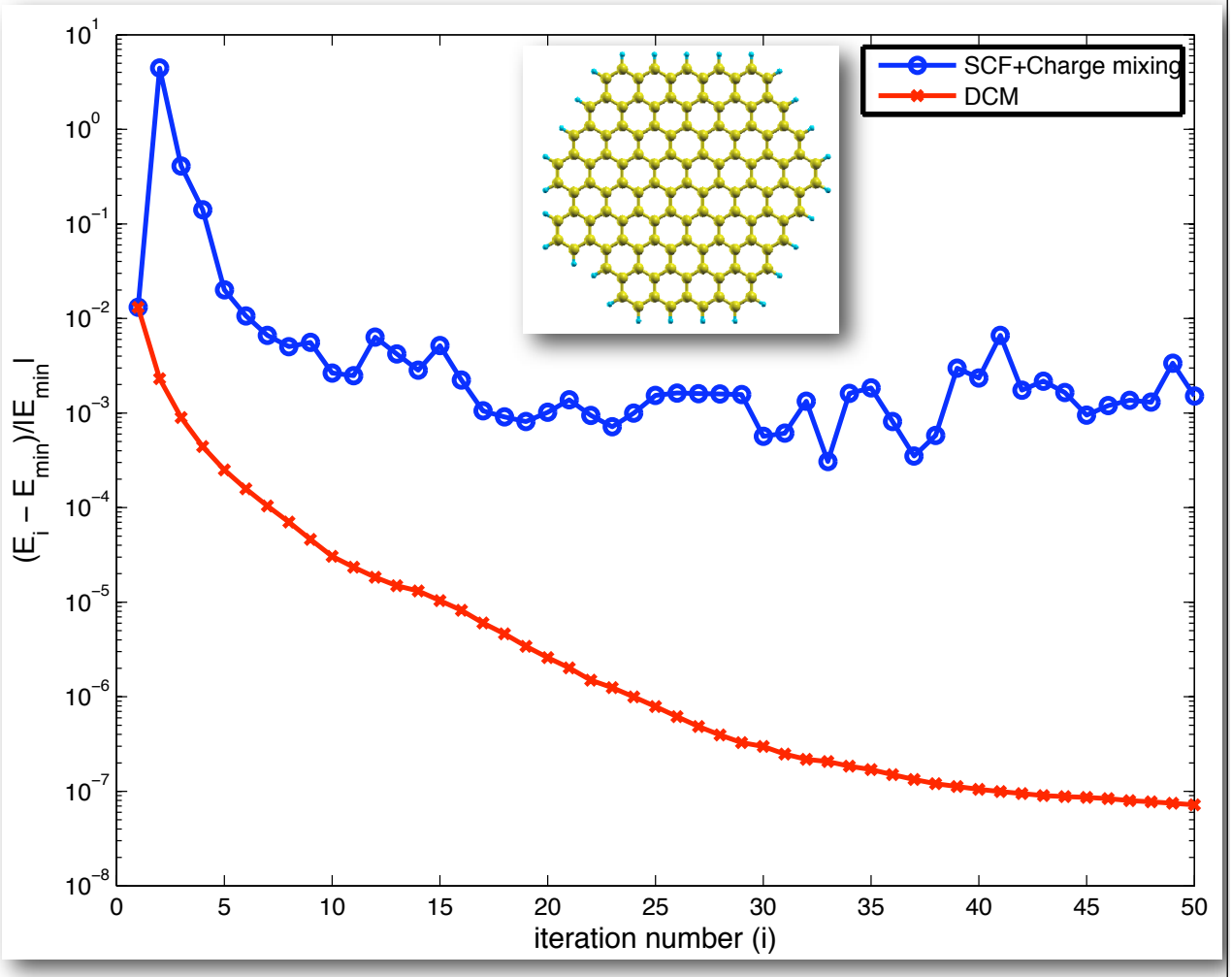
Direct Constrained Minimization

- 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
    - $x_{k+1}^T x_{k+1} = 1$;
    - $E(x_{k+1}) < E(x_k)$;

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)

Example: Graphene

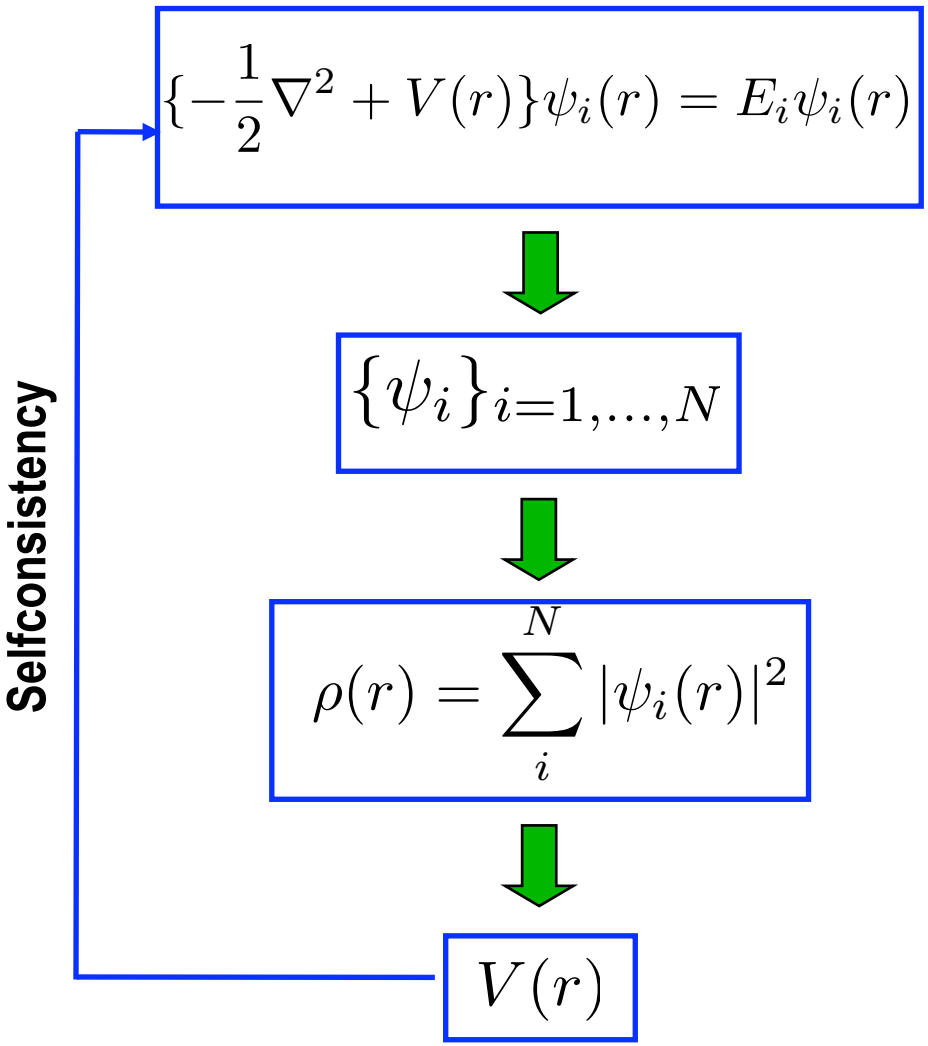
- sampling grid:
  - 114 x 114 x 15
- 10 PCG iterations / SCF outer iteration
- 5 inner SCF iteration / DCM outer iteration
- supercell:
  - 40 x 40 x 5



C. Yang, J.C. Meza, B. Lee, and L.W. Wang, "KSSOLV - A MATLAB Toolbox for Solving the Kohn-Sham Equation," ACM Transactions on Mathematical Software, Vol. 36, No. 2, 2009

Next Steps

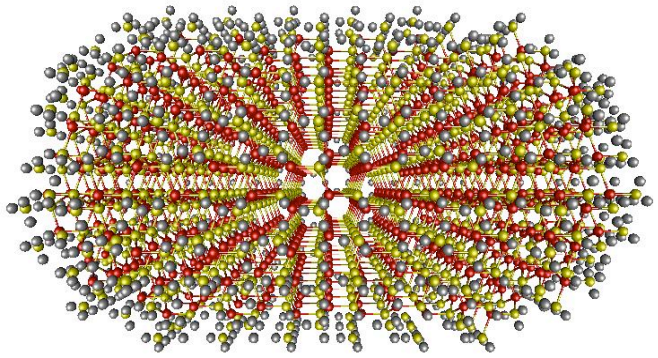
Basic SCF Iteration



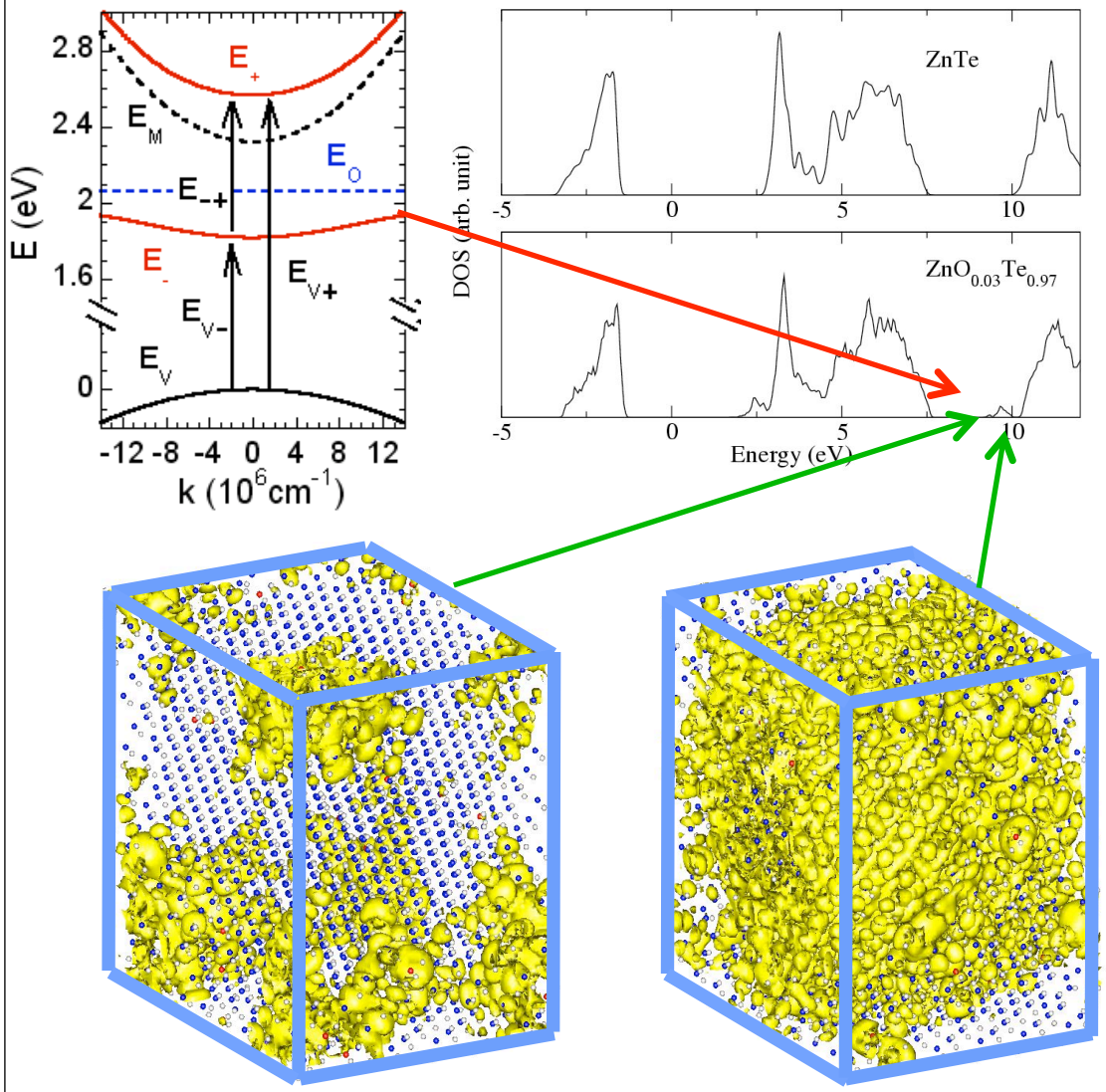
- Overall Complexity $O(N^3)$
- Major computational work (for plane wave codes):
  - 3D FFT
  - Orthogonalization
  - Nonlocal potential
- Parallel efficiencies can be quite high

Nanoscience Calculations and Linearly Scaling Algorithms

- Linear Scaling 3D Fragment (LS3DF)
- Density Function Theory (DFT) calculation numerically equivalent to direct DFT
- Scales with *O(N)* in the number of atoms rather than $O(N^3)$.
- Up to 400X faster than direct DFT
- Achieved 442 TFlops/sec on Cray XT5 with 147K processors and 225 TFlops/s on a Blue Gene/P with 165K processors



Can one use an intermediate state to improve solar cell efficiency?



- Single band material theoretical PV efficiency is 30%
- With an intermediate state, the PV efficiency could be 60%
- One proposed material ZnTe:O
  - Is there really a gap?
- LS3DF calculation for 3500 atom 3% O alloy [one hour on 17,000 processors]
- ***Yes, there is a gap, and O induced states are very localized.***

Highest O induced state *ZnTe bottom of cond. band state*

INCITE project, NERSC, NCCS

Summary - The Difference Between Theory and Practice ...

- Theory gives us the framework for analyzing our problems and guides our solutions
- Practice gives us the experience of real-world problems and helps us improve algorithms
- New techniques (as well as old techniques) are always arising that should be investigated
- We are in a unique position to contribute solutions to emerging problems in energy, climate, and many others.

Acknowledgements

- Chemical Vapor Deposition
  - Patricia Hough, Chris Moen, Todd Plantenga, Paul Spence, Charles Tong
- Supernova
  - Peter Nugent, Rollin Thomas
- Nanoscience
  - Chao Yang
  - Zhengji Zhao, David Bailey, Byounhak-Lee, Hongzhang Shan, Erich Strohmaier, Lin-Wang Wang
- DOE Office of Advanced Scientific Computing Research

Thank you!