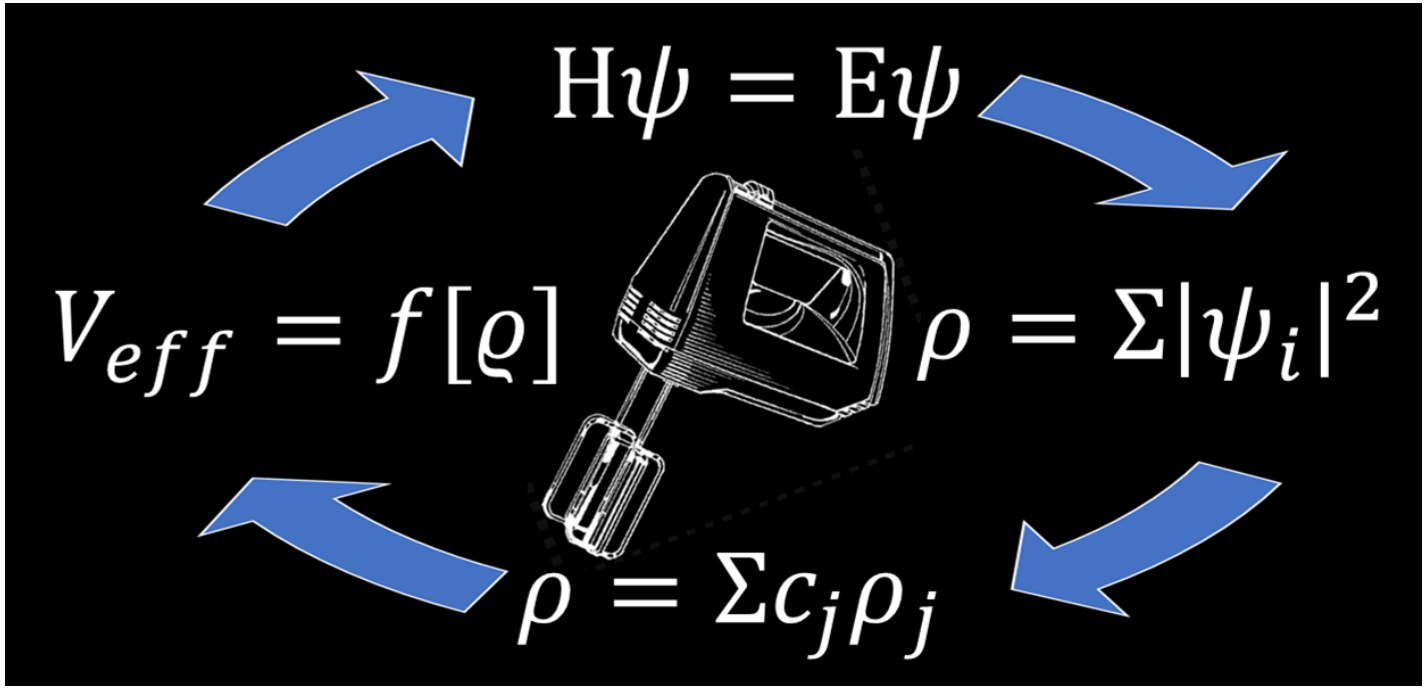
Mixing for Dummies

L. D. Marks
Department of Mat Sci & Eng
Northwestern University



Acknowledgements

Peter Blaha
TU Wien

Russel Luke
U. Göttingen

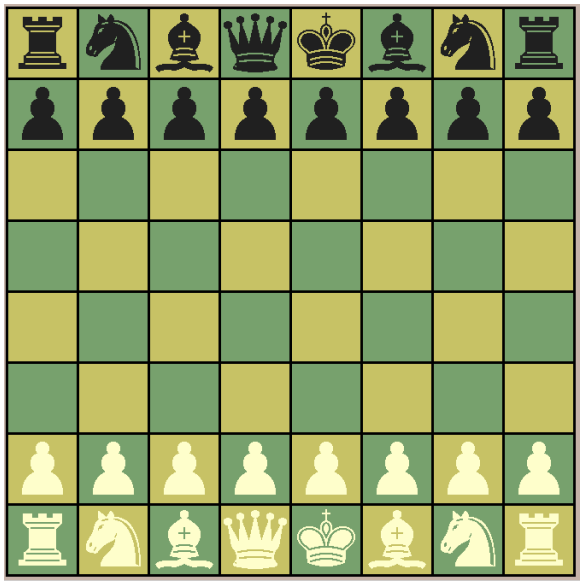
National Science Foundation
WHERE DISCOVERIES BEGIN

Overview

- Role of mixing and force minimization
- Nuts and bolts in Wien2k
- Underlying math
  - Simple form
  - What is really going on
  - Predictive mixing
- What to do when you mess up

Overview

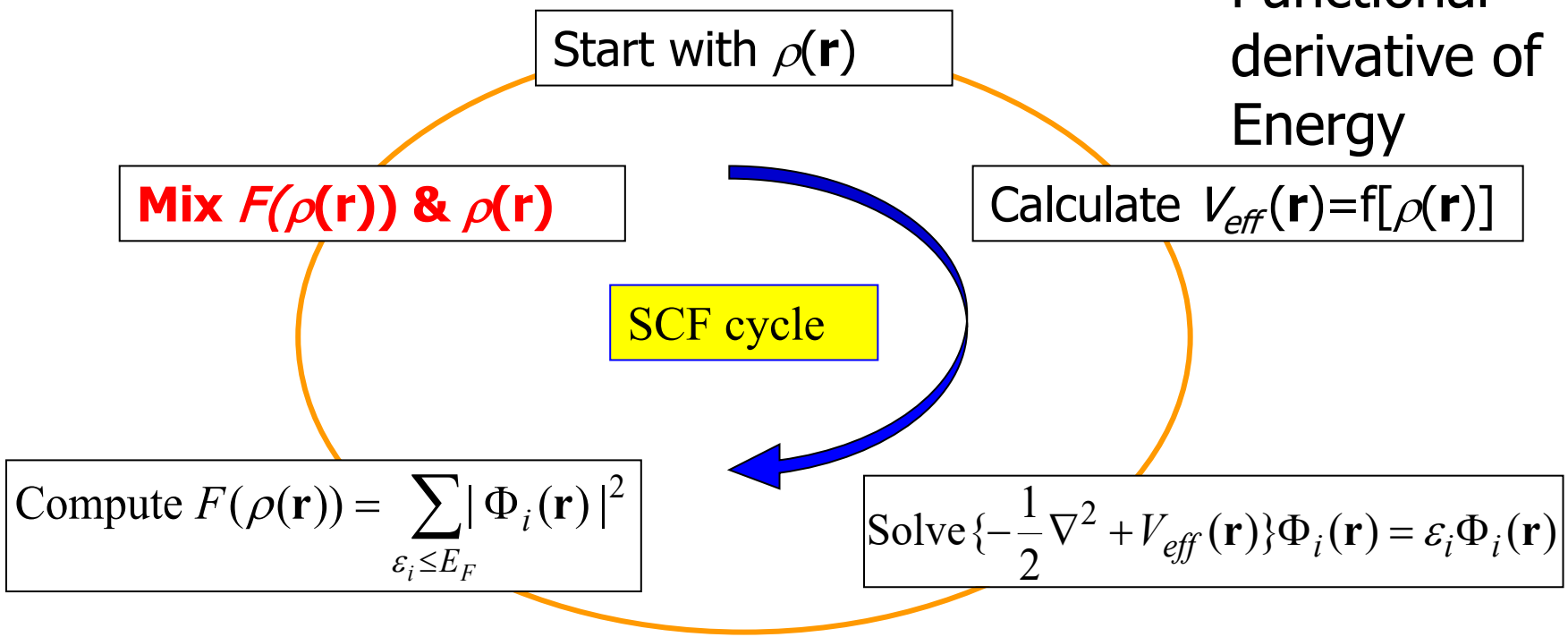
- Role of mixing and force minimization
- Nuts and bolts in Wien2k
- Underlying math
  - Simple form
  - What is really going on
  - Predictive mixing
- What to do when you mess up



What is DFT?

- Density Functional Theory
  - Solve the quantum mechanics for tens to thousands of atoms
  - Fundamental to many areas of current chemistry, materials science and physics
  - One of the major killer of electrons on supercomputers

Fixed Point cycle



Solves a variational energy by a fixed-point method, implicit minimization

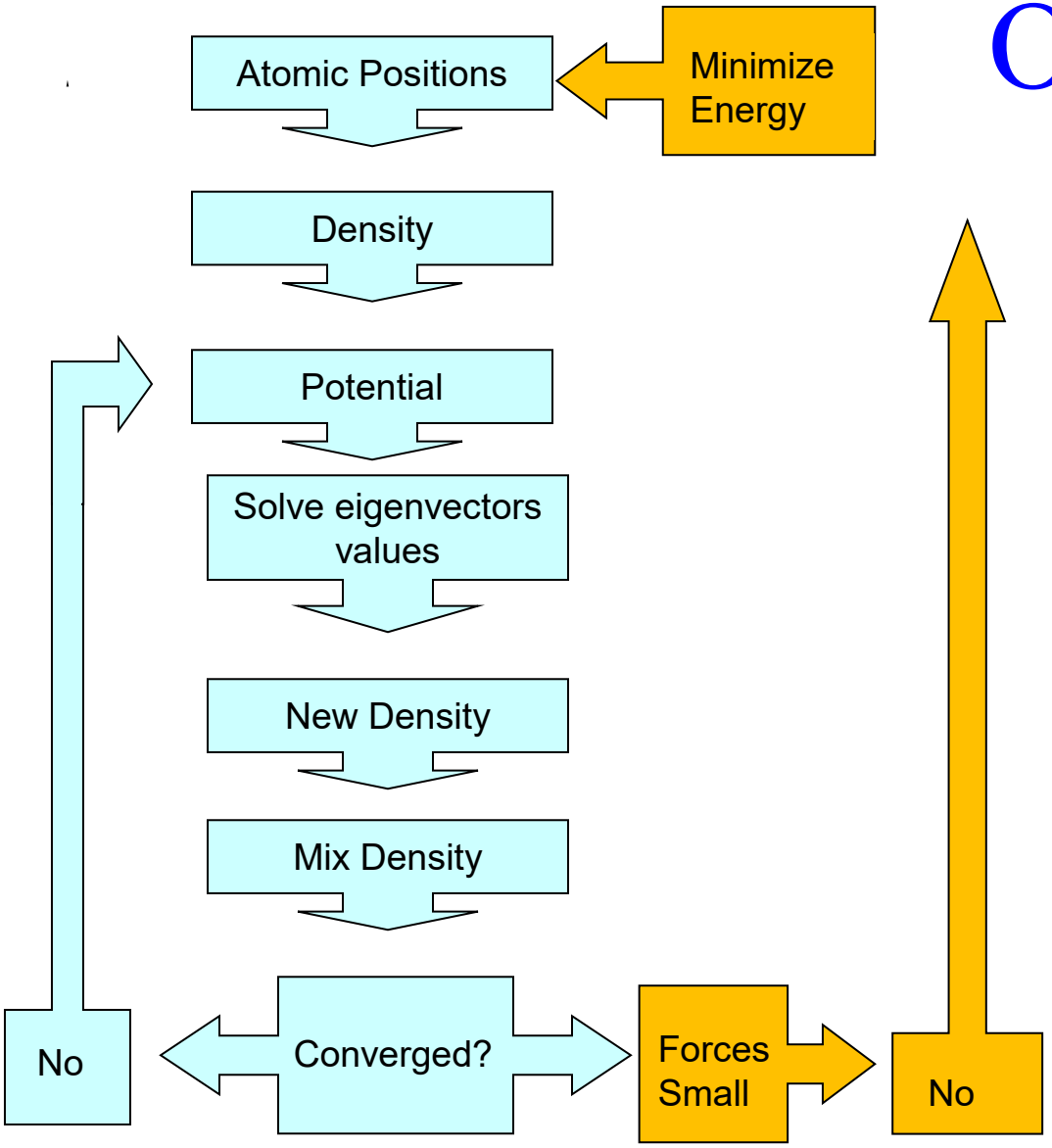
General Structure

- Density contains 10^3 to 10^6 components
- No gradients plausible
- Fortunately the eigenvectors/values are much, much smaller (10-1000)
- Often converges in 20-40 iterations
- Iff well posed, the stationary point is a variational minimum of the energy, so has good properties

What has to be computed

1. The atomic positions need to be minimized
2. A self-consistent density has to be found

- Conventional is to do nested loops, 2 inside 1
- An alternative is to solve the two at the same time (MSR1a)

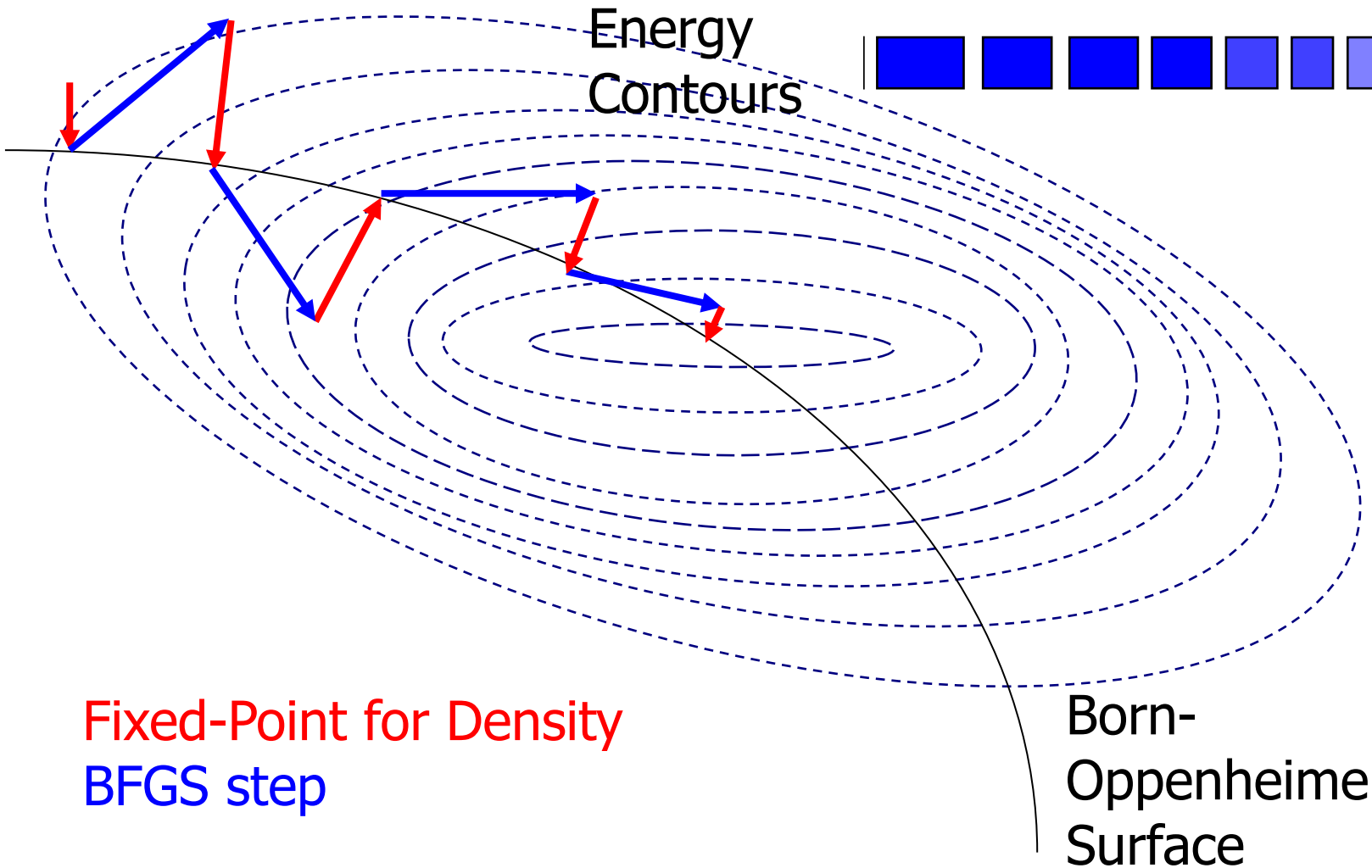


Old (PORT)

- Inner loop to obtain fixed-point for given atom positions
- Outer loop to optimize atomic positions

J. Chem. Theory Comput, DOI:10.1021/ct4001685

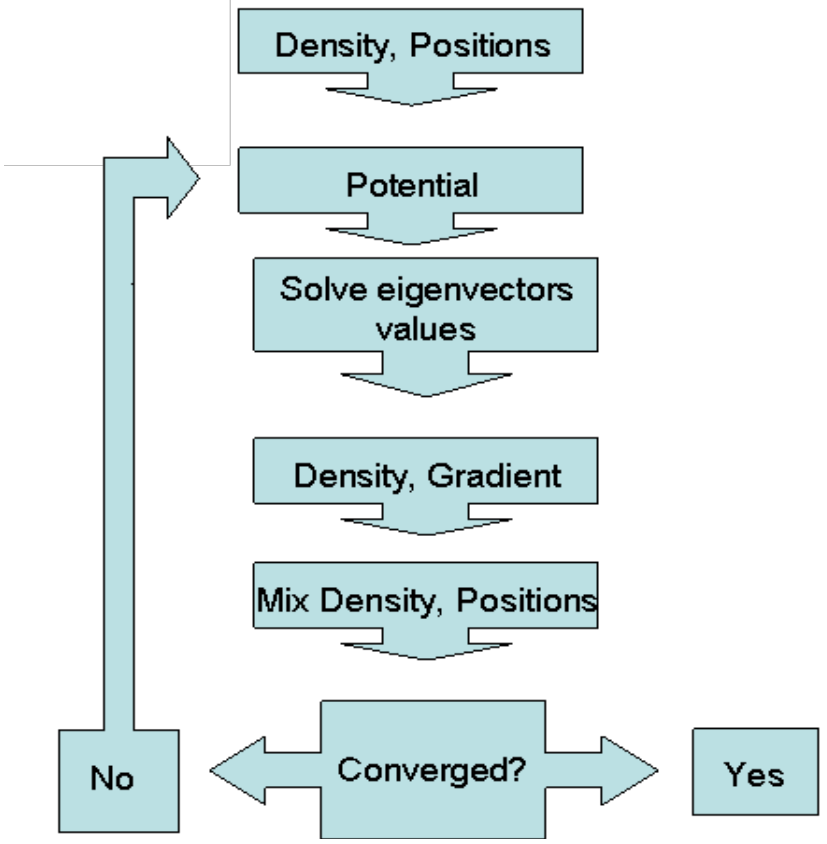
Double-Loop

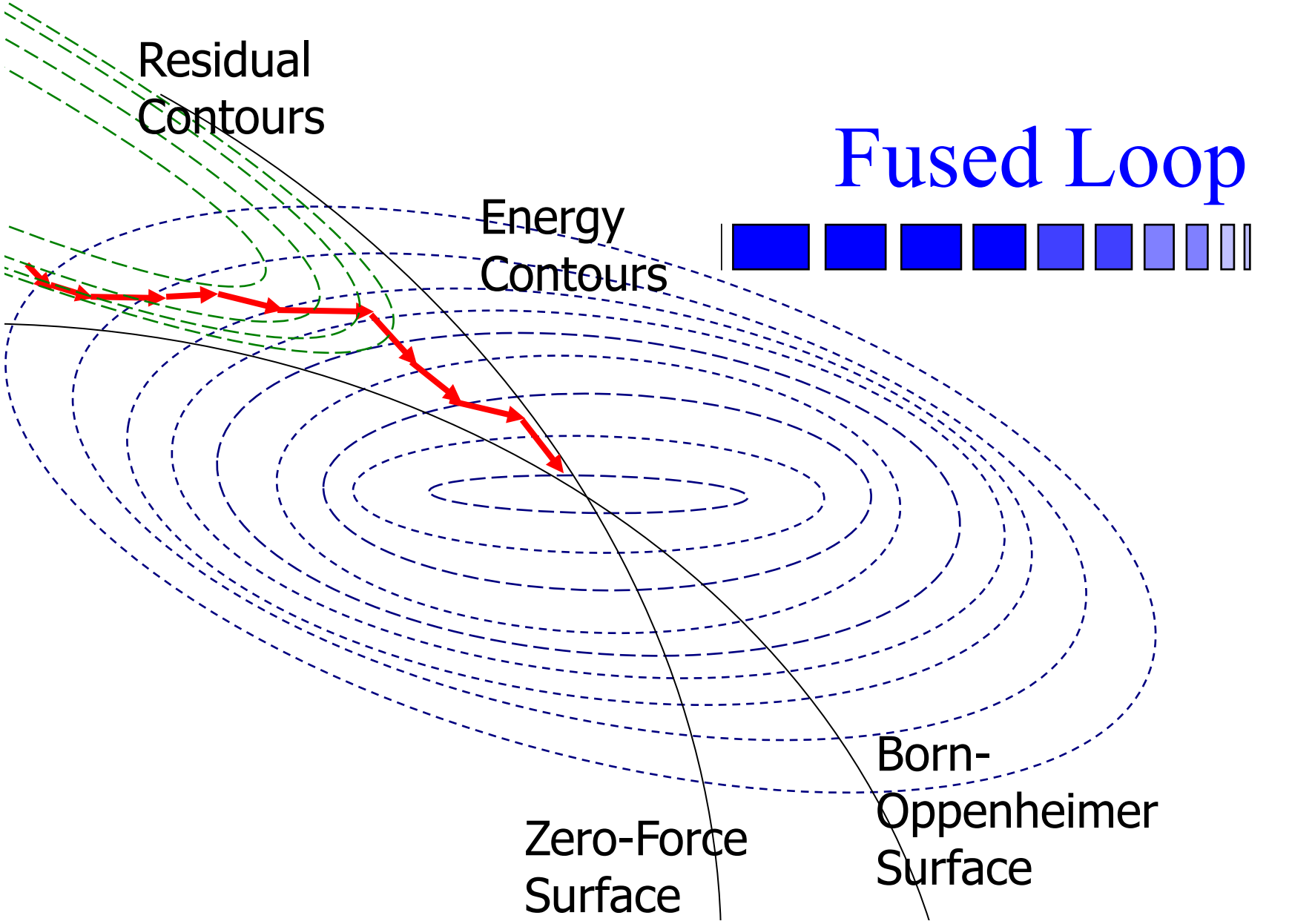


J. Chem. Theory Comput, DOI: 10.1021/ct4001685

MSR1a Fused Loop

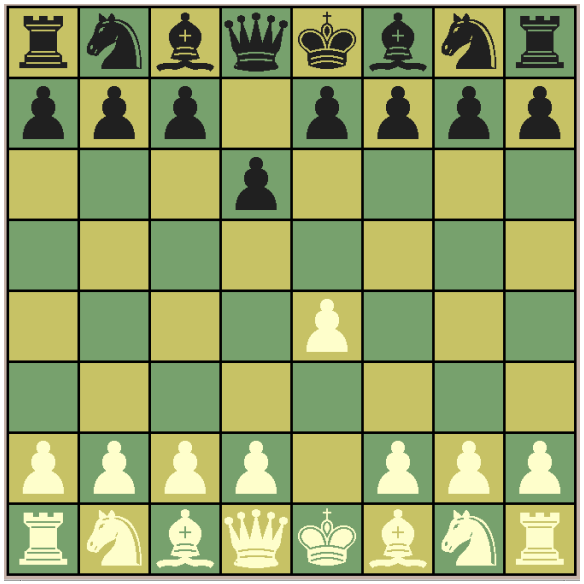
- Treat the density and atomic positions (as well as hybrid potentials etc as needed) *all* at the same time.
- No restrictions to "special" cases, general algorithm has to work for insulators, metals, semiconductors, surfaces, defects, hybrids….
- Few to no user adjustable parameters




Residual
Contours
Fused Loop
Energy
Contours
Born-
Oppenheimer
Surface
Zero-Force
Surface

Overview

- Role of mixing and force minimization
- Nuts and bolts in Wien2k
- Underlying math
  - Simple form
  - What is really going on
  - Predictive mixing
- What to do when you mess up



Calculating a system

- Choose the functional (LDA, PBE, WC)
- Find the DFT equilibrium volume for the bulk, with similar RMT, RKMAX (may have to redo later)
- Create the structure – supercell or other (e.g. Cryscon from Shape Software)
- Look at it – you probably did something wrong!

Converge density

- Done with the “runXX_lapw” scripts
- In 99% of cases no need to change mixing parameters
- First iteration might be problematic, could use “echo .001 > .pratt”
- Most issues are due to bad models (e.g. cold fusion)

Converge positions

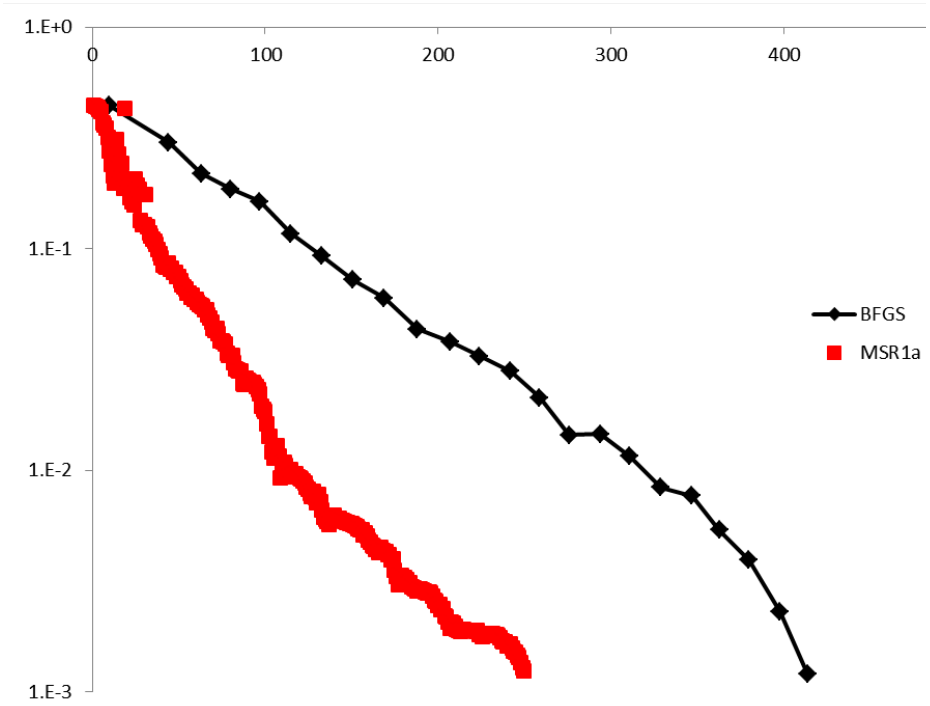
- Done with runXX_lapw and “-min” option
- Again, in 99% of cases no need to change anything
- Sometimes it is better to slightly converge the density first
- Can converge better than density alone, but takes longer
- Be careful about iterative modes

Alternative for positions

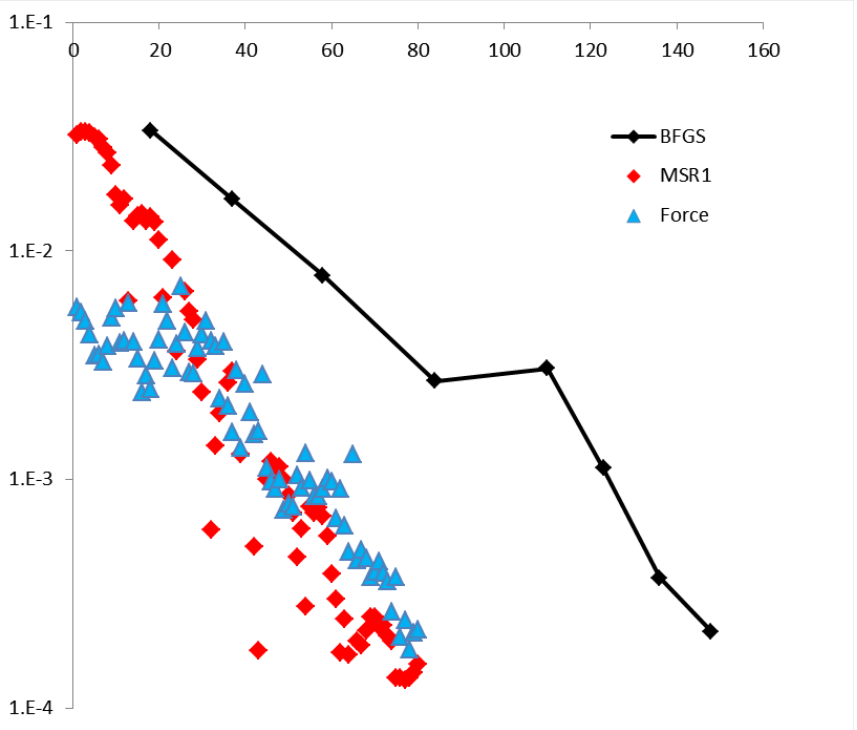
- Slower double loop (PORT), using min_lapw
- Uses gradients, and builds an approximation to the Hessian (2nd derivative matrix)
- If you are doing many almost identical calculations it can be faster due to the Hessian from prior calculations

Comparison

52 atoms, MgO (111)+H_2O

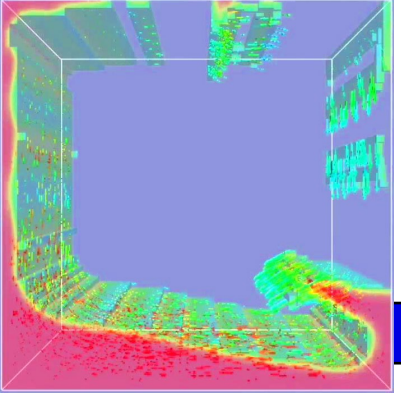


108 atoms AlFe



J. Ciston, A. Subramanian, L.D. Marks, PRB, 79 (2009) 085421.

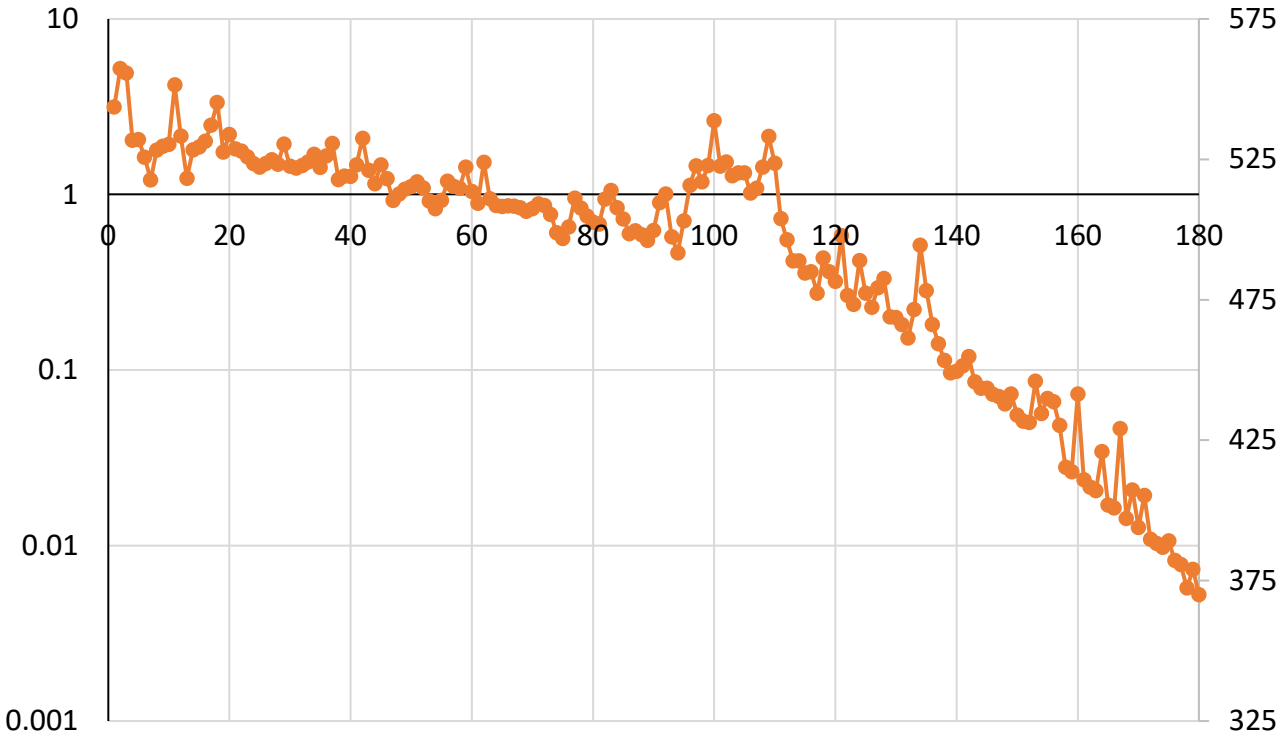
Lyudmila V. Dobysheva (2011)



Sloshing

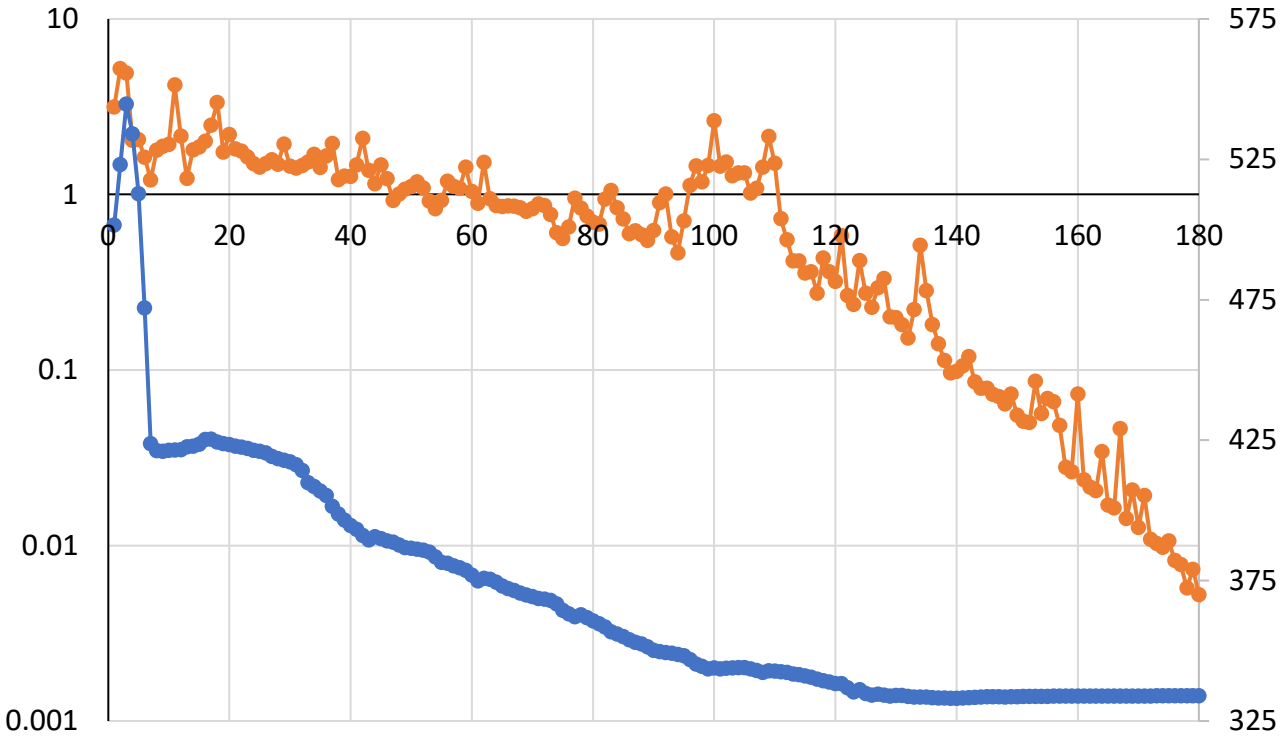
- Oscillations of the charge density
- Density moving from one end of a surface to another
- Appears as oscillations
- Due to too large steps (non-linear terms)
- Or….

Not really oscillations



Density apparently oscillating

Not really oscillations



Density apparently oscillating

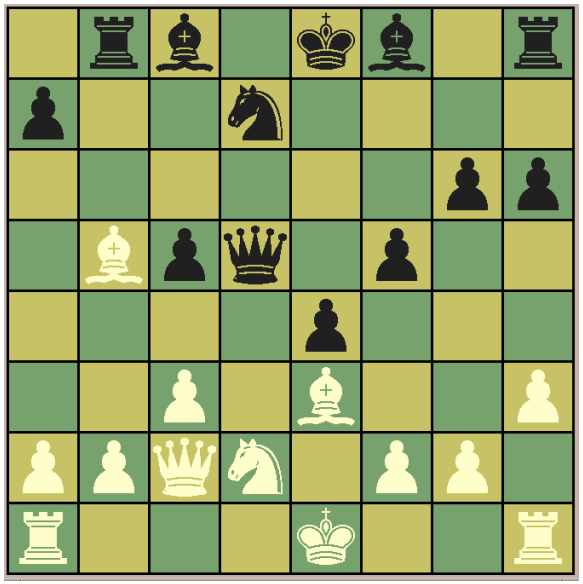
Total PW density in sphere converging

Beware of Fake News

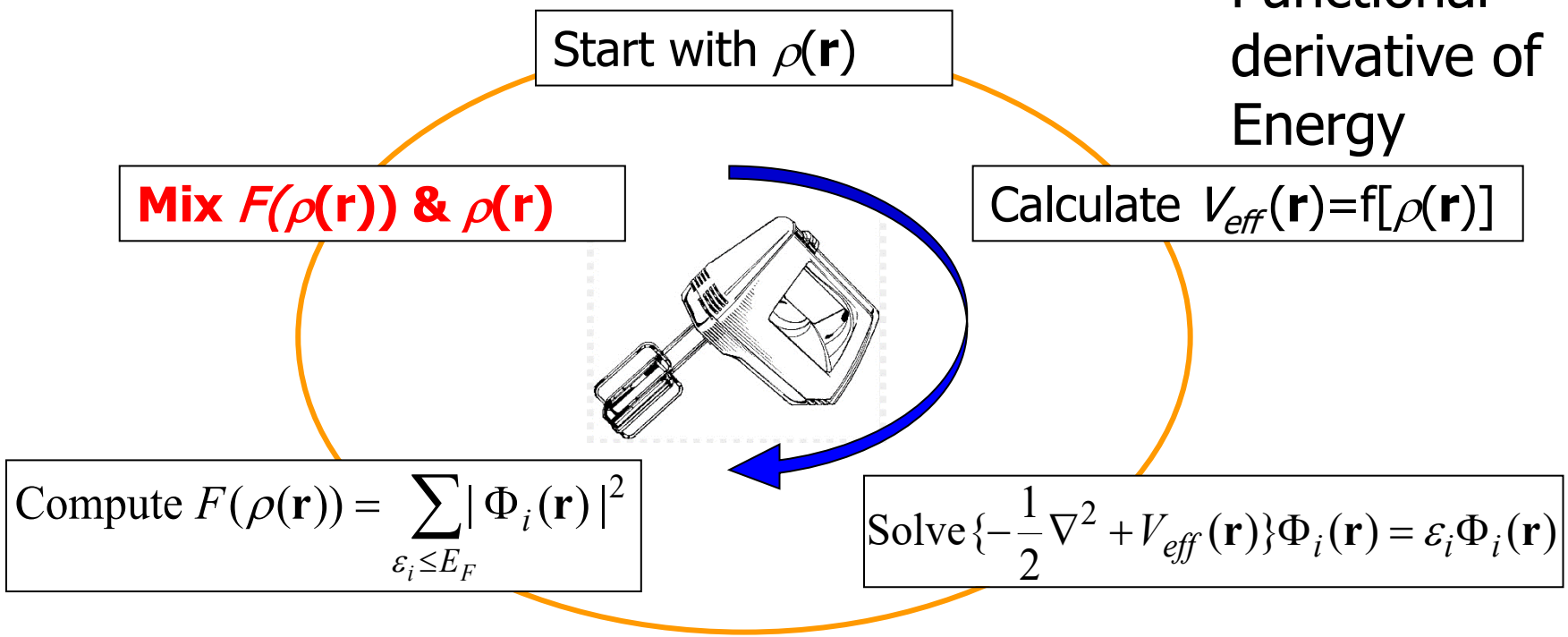
- "Reduce the mixing factor" wisdom is for other codes – don't
- Fixing atoms probably wont make the calculation converge faster
- Bigger problems are worse
  - Density convergence scales as the number of bands and their width
  - Positions scale as the number of electron-phonon bands (dielectric eigenvalues)

Overview

- Role of mixing and force minimization
- Nuts and bolts in Wien2k
- Underlying math
  - Simple form
  - What is really going on
  - Predictive mixing
- What to do when you mess up



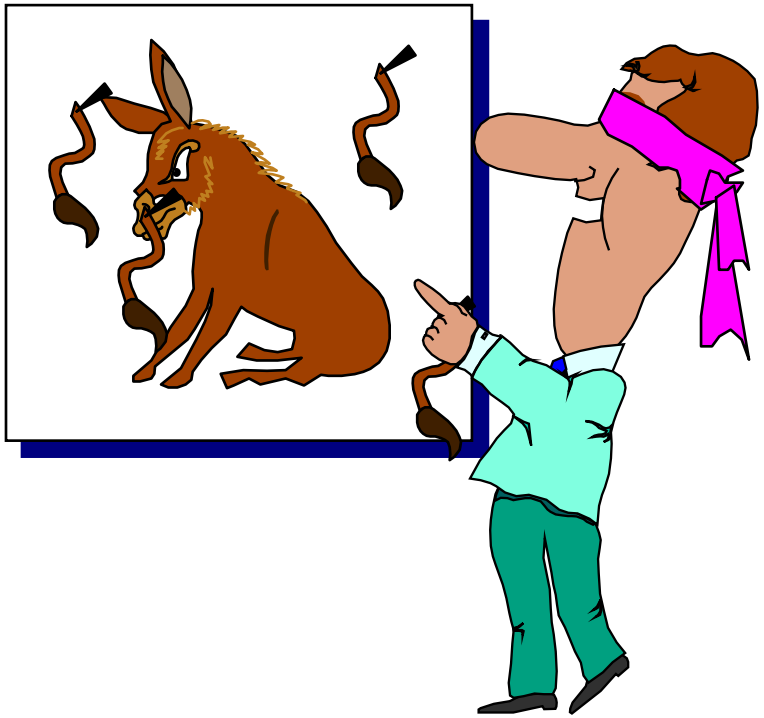
Fixed Point cycle



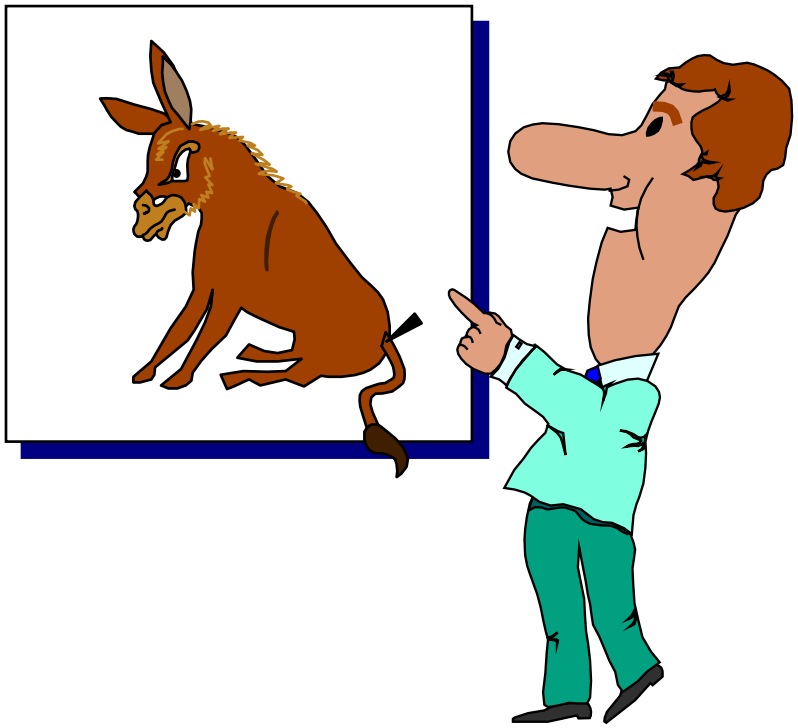
Solves a variational energy by a fixed-point method, implicit minimization

Underlying principle

Current density



Use prior densities



Information available

- History of prior densities, and the density at the end of the SCF iteration
- History of atomic positions, and the pseudo-forces
- Use these to build an approximate mathematical model

Historical perspective

- New point $F(\rho_k)$
- Define a residue $R_k = F(\rho_k) - \rho_k$
- Simplest is linear

$$\rho_{k+1} = \rho_k + \alpha R_k$$

- Slow, and what is α ? (user choice)
- Pratt method often works if the user chooses right

General form

- $s_i = \rho_i - \rho_k \; ; \; y_i = R_i - R_k$
- $R = \; R_k + B_k \, s_{k+1} + \cancel{s_{k+1} \, C_k \, s_{k+1}} + \cdots$
- Form a matrix for s & y values
- The key question is how to construct the solution, for which there are more than a few issues.
- Require that the reproduce prior steps

Broyden Fixed-Point Methods

- Broyden's "Good Method"

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) s_k^T}{s_k^T s_k}$$

- Broyden's "Bad Method" ($H=B^{-1}$)

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k) y_k^T}{y_k^T y_k}$$

C.G. Broyden, A Class of Methods for Solving Nonlinear Simultaneous Equations, Mathematics of Computation, 19 (1965) 577-593.

J. Chem. Theory Comput, DOI: 10.1021/ct4001685

Why Good & Bad?

- Originally the “Good” method worked for Broyden (PRB…) and the “Bad” method did not
- For many years the “Bad” method was never used…..except in mixing as DFT developers probably never read the literature
- Subtle difference

  Good Broyden: finds optimal density with current model (greedy algorithm)

  Bad Broyden: finds most conservative density to minimize residue (least greedy algorithm)

DFT history

- Initially people tried the “standard” Broyden method (1965) – it did not work
- Sequential Broyden, fair (Srivastava, 1984)
- Better, multisecant (Johnson 1988)

$$\rho_{k+1} = \rho_k + \Sigma\, \alpha_j R_j$$

- The coefficients from L2 (LS) solution.
- *Some LS are more equal than others. Different regularization & conditioning*

General form

- Unpredicted step, greed α_k

  $$u_k = R_k - B_k p_k$$

- Predicted step

  $$p_k = B_k^{-1} R_k = H_k R_k$$

- General form

  $$\rho_{k+1} = \rho_k + \alpha_k u_k + \beta_k p_k$$

General form

- Unpredicted step, greed α_k

  $u_k = R_k - B_k p_k$

- Predicted step

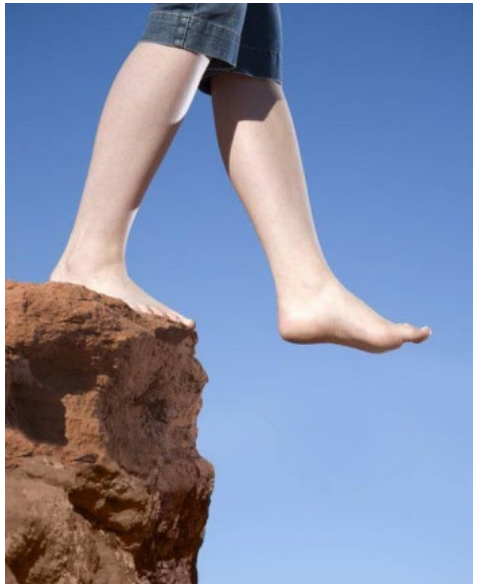
  $p_k = B_k^{-1} R_k = H_k R_k$

- General form

  $\rho_{k+1} = \rho_k + \alpha_k u_k + \beta_k p_k$

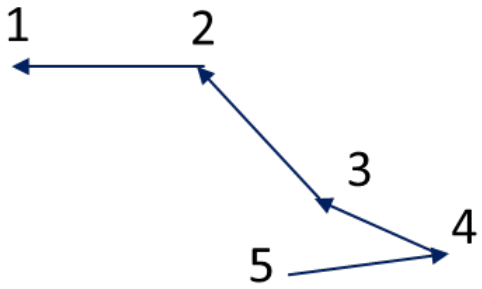
Issue 1: Unpredicted step

- No information available
- Must be controlled
- If too large, simplex gradients are unreliable
- Implicit trust, i.e. increase if improving “enough”
- It is hard to know what enough is

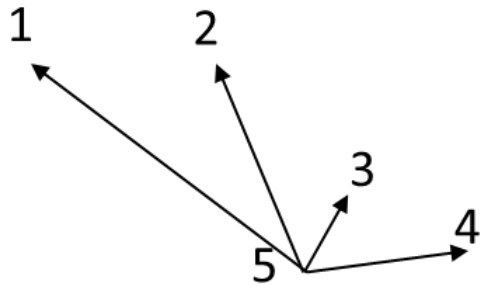


Issue 2: Scaling

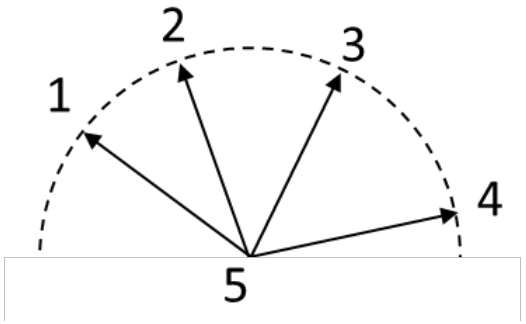
- How to treat the previous steps?
  - Sequential (Broyden) or multisecant (DIIS, Simplex)?
  - As is, with different magnitudes?
  - All my tests support a simplex gradient



Broyden



DIIS



Simplex

General form

- Unpredicted step, greed α_k

$$u_k = R_k - B_k p_k$$

- Predicted step

$$p_k = B_k^{-1} R_k = H_k R_k$$

- General form

$$\rho_{k+1} = \rho_k + \alpha_k u_k + \beta_k p_k$$

An issue for fixed-point problems

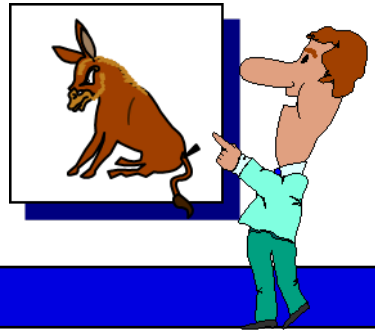
- Most early methods are somewhat causal, appropriate for a minimization where the energy is decreasing
- This does not need to be the case for a fixed-point problem as the gradient is not available, instead a residual
- There is nothing a-priori to say that point 2 should be better than point 1
- Hence….

Multisecant Approach

- Consider a number of values:
- $S = (s_0, s_1, \ldots . s_n)$; $Y = (y_0, y_1, \ldots . y_n)$
- Expand to a simultaneous solution:
- $BS = Y$; or $HY = S$
- Minimum-Norm Solution (MSEC)

$$H_{k+1} = H_k + (S_k - H_k Y_k)(Y_k^T Y_k)^{-1} Y_k^T$$

Take $H_k = I$



What form?

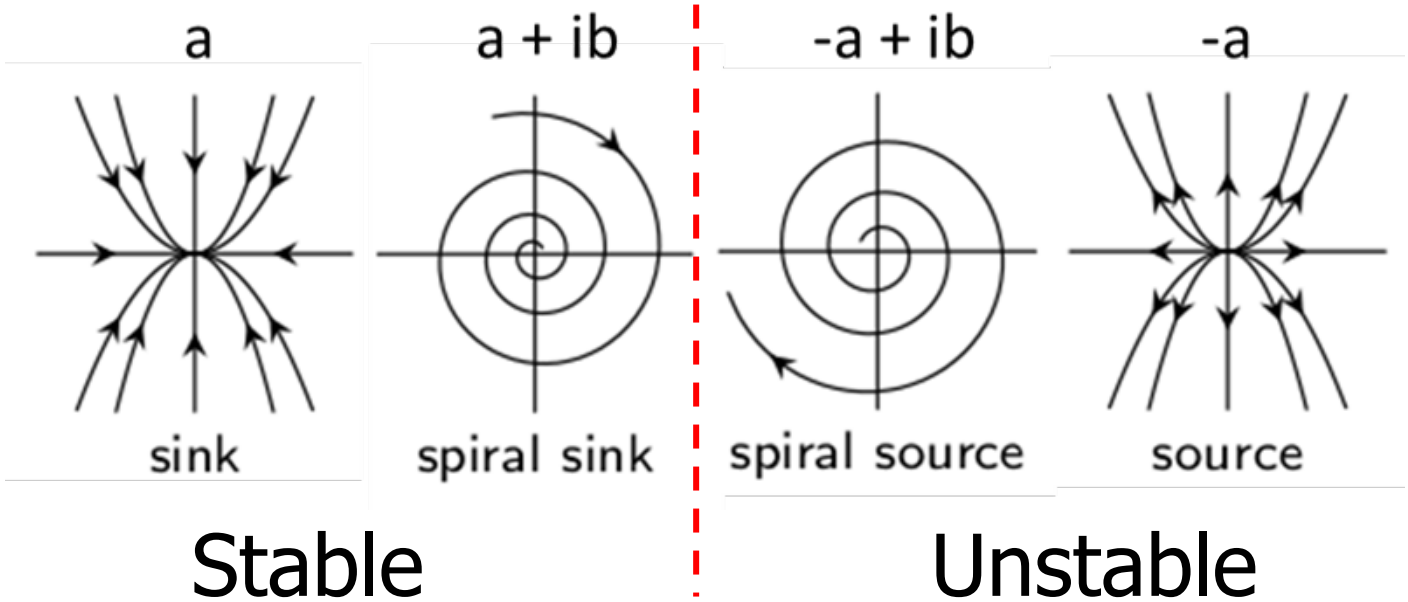
- $H_k = S_k \left(T_k^T Y_k\right)^{-1} T_k^T$
- $T_k = (1-\lambda)S_k + \lambda Y_k$
- $\lambda = 0$ is conventional “Good” Broyden, most greedy, may diverge, soft case
- $\lambda = 1 \cong$ DIIS, Anderson, “Bad” Broyden, least greedy, may stagnate (different scalings…), stiff case
- 8-16 memories (not critical)

What is a greedy algorithm?

- A greedy algorithm takes decisions on the basis of information at hand without worrying about the consequences. In many cases “greed is good”, but not always.
- Example: make 41c with 25c, 10c, 4c coins
- Optimum solution: 25+4x4
- Greedy solution: start with 41c, use largest reduction
  - 25c Remainder 16
  - 10c Remainder 6
  - 4c Remainder 2

Ansatz: inspired by optimization

- $T_k = \lambda S_k + Y_k$
- Search up for largest λ where $T_k^T Y_k$ does not have source eigenvalues

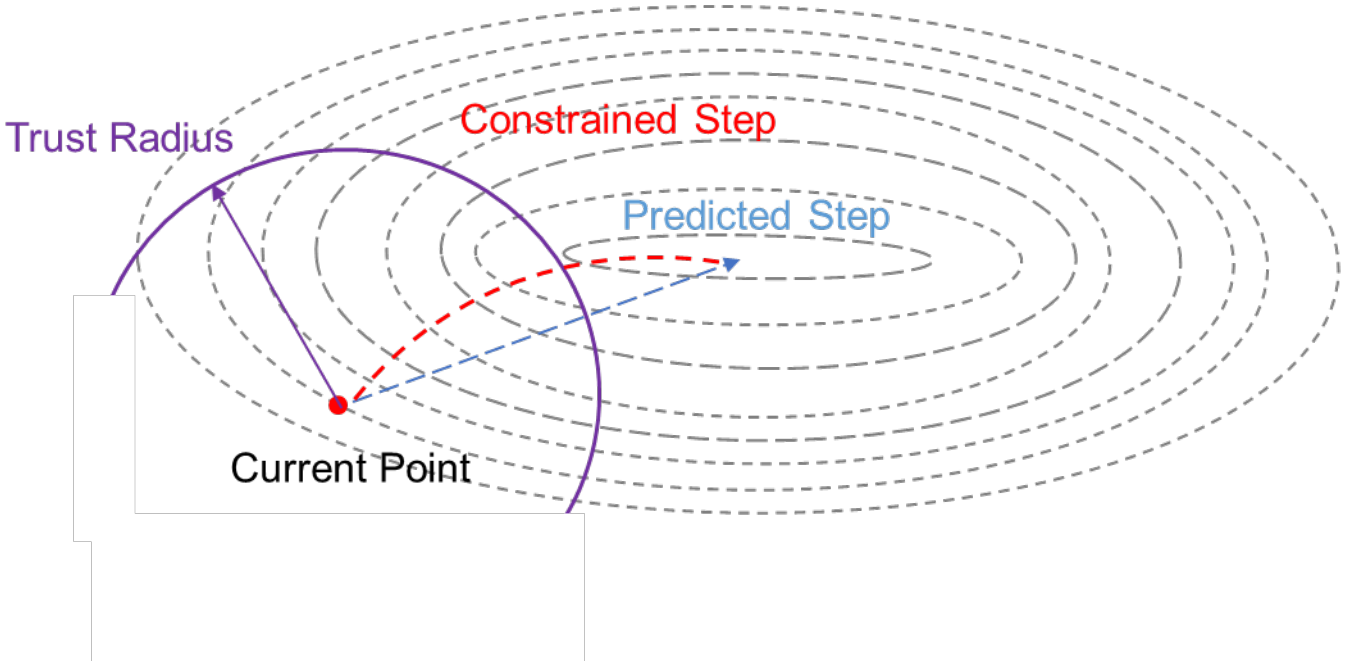


Issue 4: Predicted

- $R = R_k + B_k\, s_{k+1} + s_{k+1}\, C_k\, s_{k+1} + \cdots$
- What about the higher-order term?
  - The linear model (predicted step is only valid for small enough steps)
  - What is small enough – needs a trust region
  - Surprisingly I don't think any code except Wien2k has trust regions for DFT

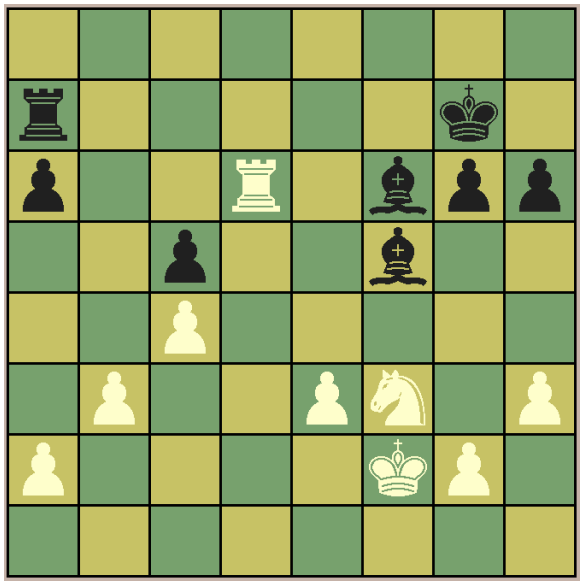
Trust Region

- Total magnitude of step limited by a “Trust Region”; solve quadratic form to minimize the energy with this as a constraint



Overview

- Role of mixing and force minimization
- Nuts and bolts in Wien2k
- Underlying math
  - Simple form
  - What is really going on
  - Predictive mixing
- What to do when you mess up



Math issues

- There are cases with multiple fixed points
- There are cases where there is **no** downhill route from certain densities to the solution
- *At the solution* properties are good – but unpredictable otherwise
- Radius of convergence etc. changes with problem, details unclear
- Global minimum with atoms as well

Numerical issues

- Algorithms involve numerical integrations/differentiations – stability and conditioning errors
- Some codes are more equal than others, e.g. break symmetry boundary conditions
- Variables often effectively single-precision
- Not a well researched topic (most codes have been written by physicists/chemists)

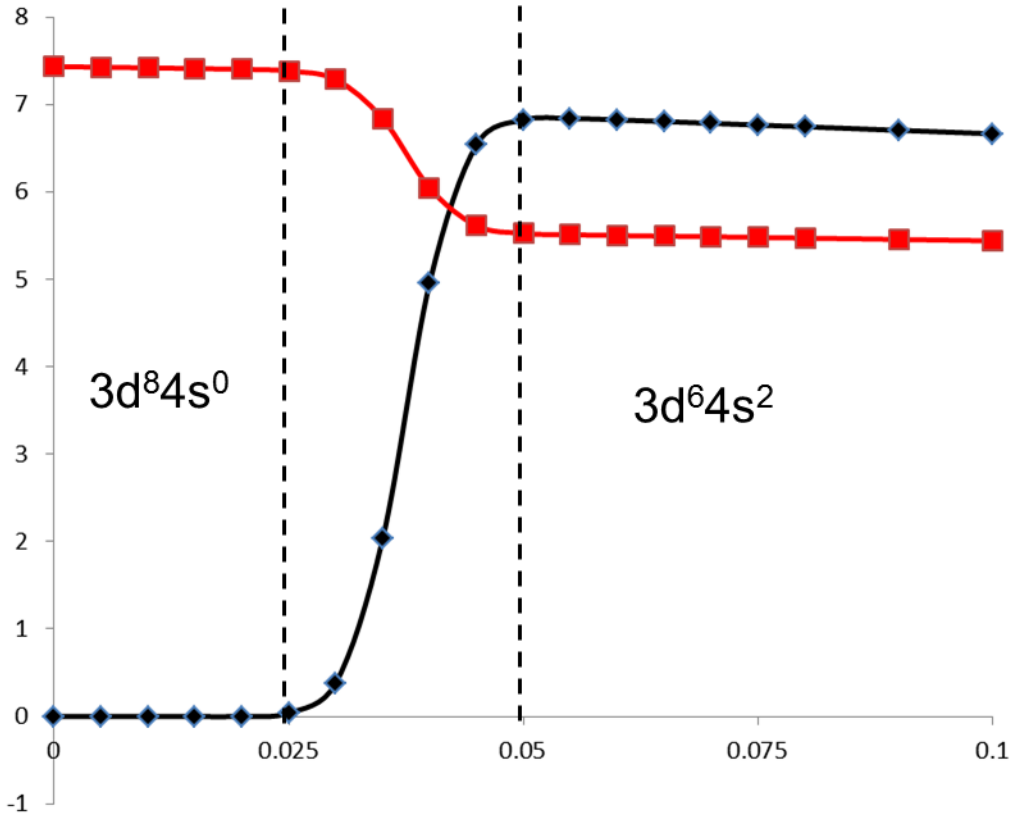
Transparency issues

- All approaches require an inverse – how is rarely mentioned (and regularization)
- Codes can have rubber bands to hold them together (Wien2k did 15 years ago)
- They are not always written to be transparent
- Can have undisclosed variables (sometime not deliberate)

Physics issues

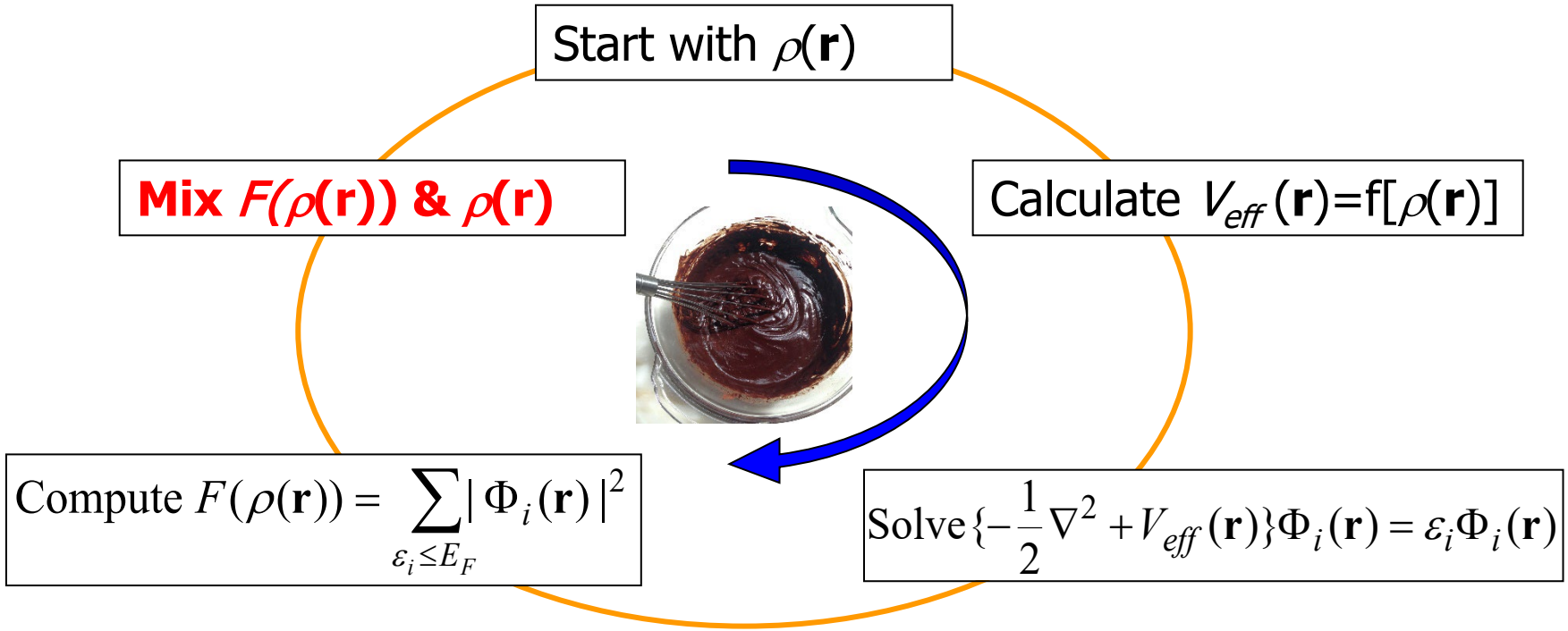
- The fundamental character can change discontinuously, to illustrate
  - Problem starts as steam
  - Then condenses into water
  - Ends as ice
  - There might be a storm/lightning in the mix
- *This cannot be ignored in the math or algorithm.*

Phase Transitions: Can change discontinuously



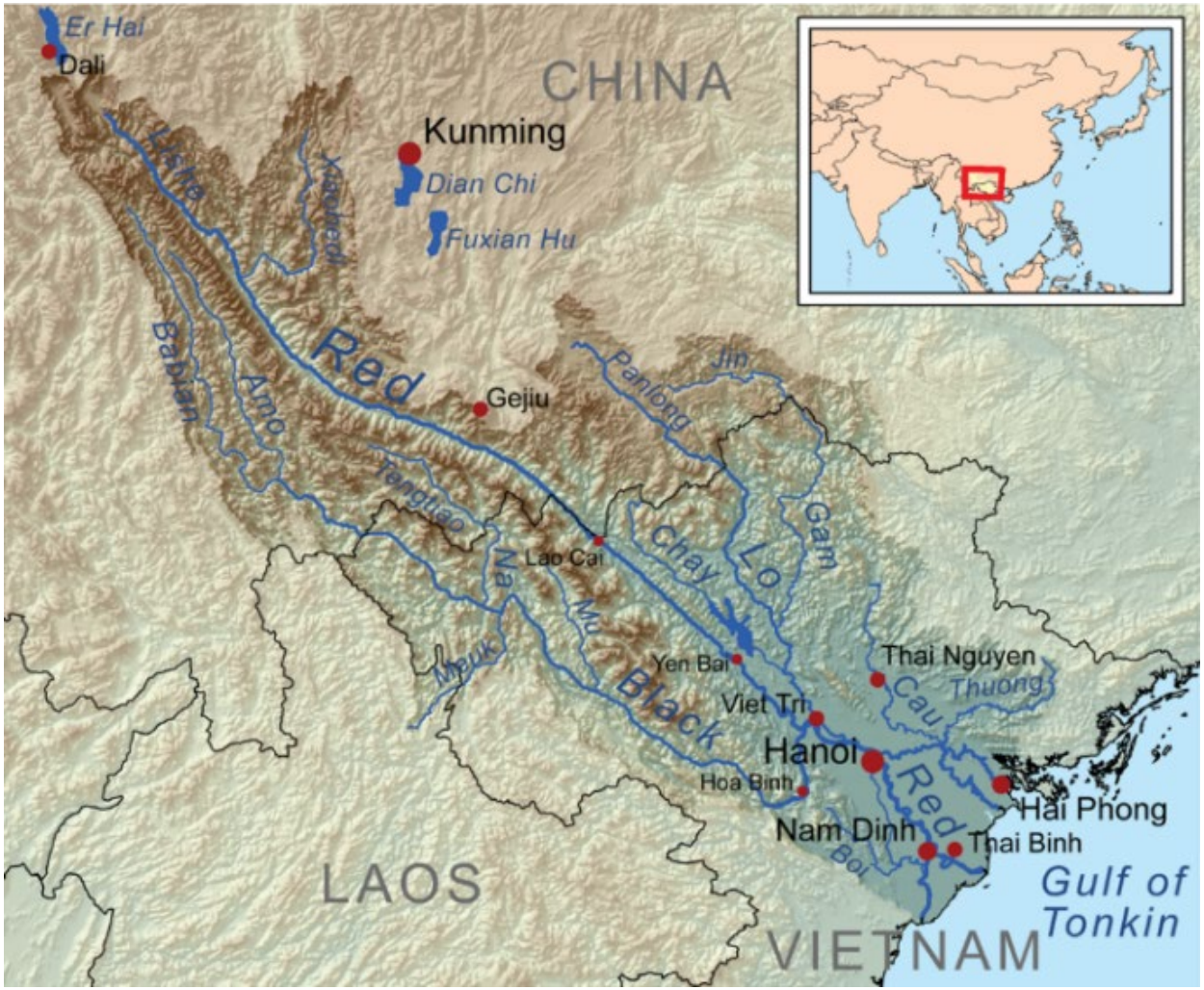
Electronic configuration of F(ρ) in the second step as a function of the size of the first Pratt step for an Fe atom, with the 4s occupancy within the muffin-tins in black (x10) and the 3d in red

Analogy: water running downhill



Solves a variational energy by a fixed-point method, implicit minimization

Flowing downhill can be simple



- Red River

Not so simple

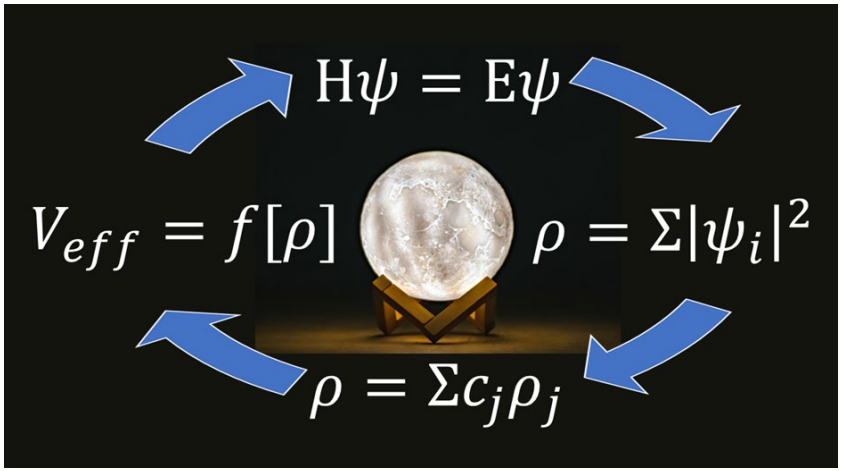
- Mississippi
- Many wiggles, but no hard walls

Problem Changes

- Starts in the mountains
- Lake Powell
- Grand Canyon
- Down to sea…not!

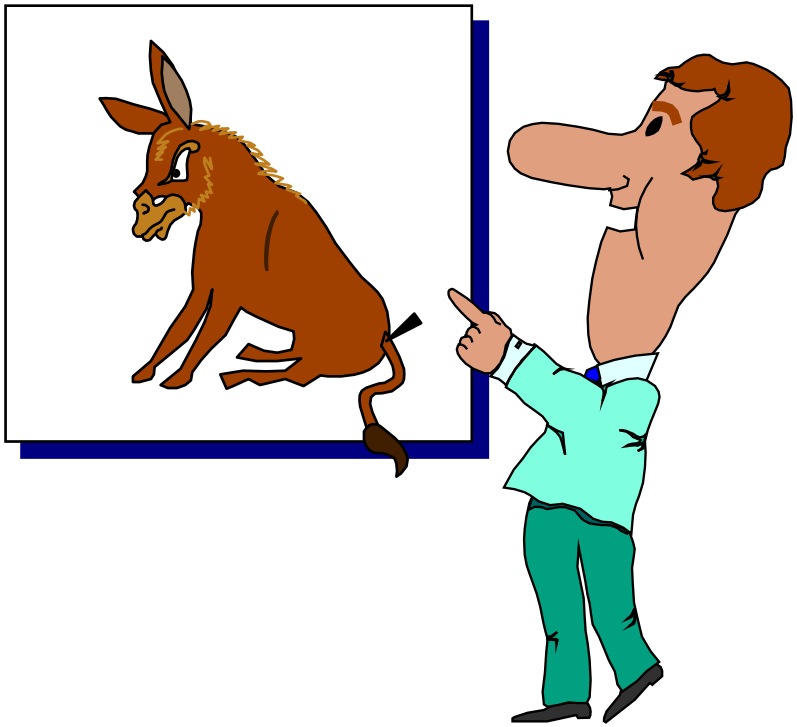
Overview

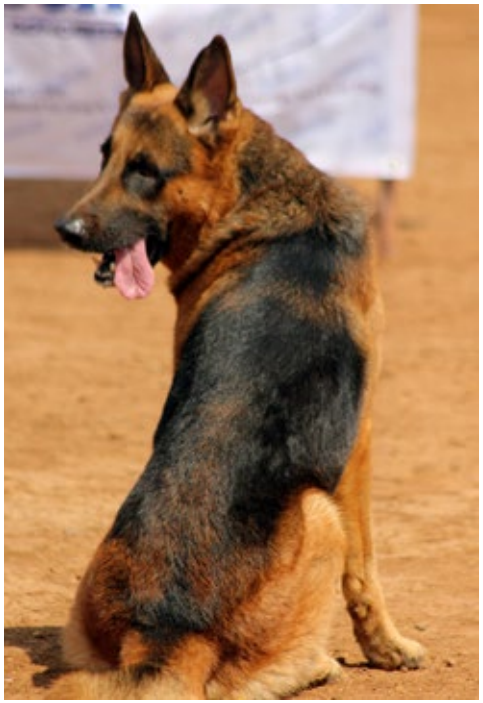
- Role of mixing and force minimization
- Nuts and bolts in Wien2k
- Underlying math
  - Simple form
  - What is really going on
  - Predictive mixing
- What to do when you mess up



Predictive mixing (2021)

Use best values for the last step





DOI: 10.1021/acs.jctc.1c00630

Reminder: general form

- Predicted step & residue

$$p_k = B_k^{-1} R_k = H_k R_k$$

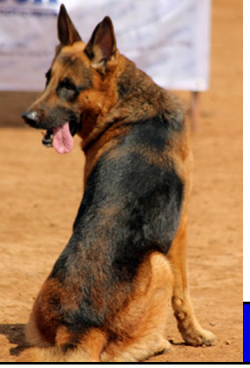
$$R_k^P = B_k p_k$$

- Unpredicted step, residue greed α_k

$$u_k = R_k - B_k p_k = R_k^U$$

- General form

$$\rho_{k+1} = \rho_k + \alpha_k u_k + \beta_k p_k$$



What we *should* have use

- Predicted step, find β' that minimizes

  $||R_{k-1}^{P} - \beta'\ B_k p_{k-1}||$ then $\beta_k = (\beta_{k-1} + \beta')/2$

- Unpredicted step find α' that minimizes

  $||R_{k-1}^{U} - \alpha'\ B_k u_{k-1}||$ then $\alpha_k = (\alpha_{k-1} + \alpha')/2$

- For trust radius, use

  $||R_{k-1} - t'\ B_k s_k||$ then $t_k = (t_{k-1} + t')/2$

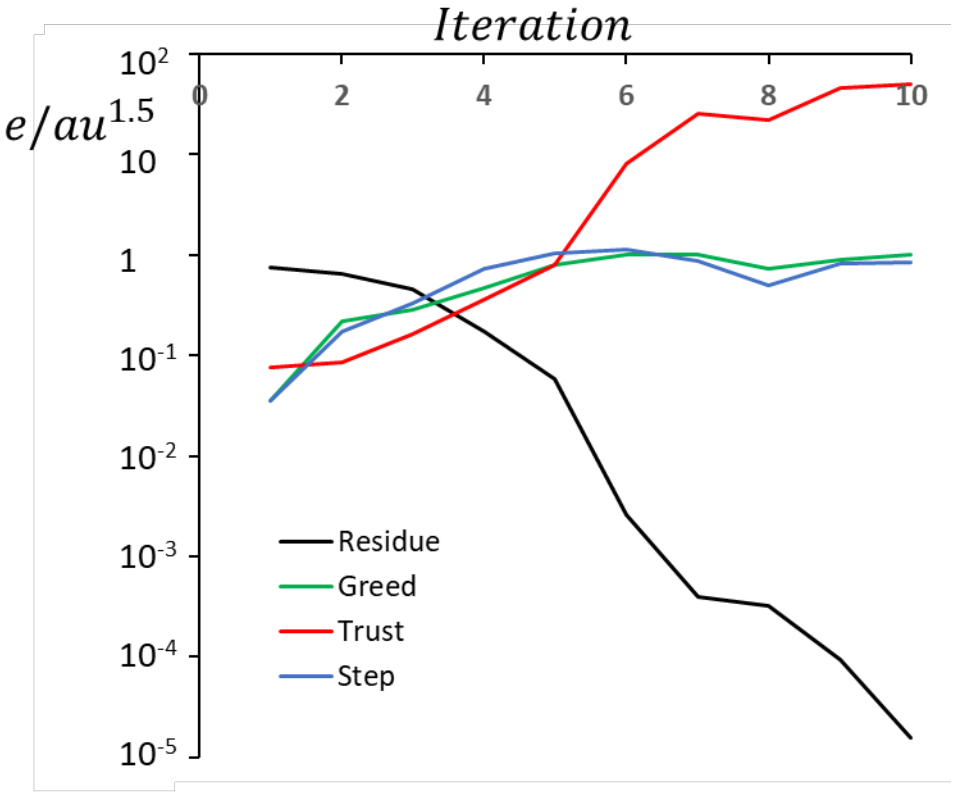
- General form, bounded by size

$$\rho_{k+1} = \rho_k + \alpha_k u_k + \beta_k p_k$$

Standard additions

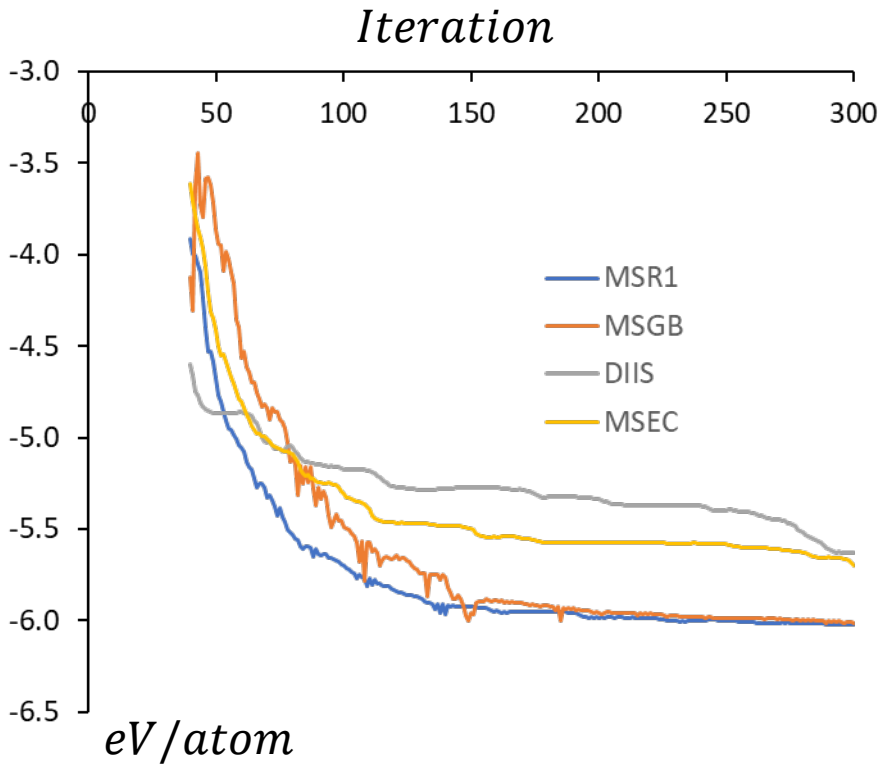
- For a really bad step, recalculate along direction and discard new point (don't contaminate simplex gradients)
- If a little bad, add to matrices & recalculate
- SVD inversion with regularization
- Start conservative
- No user parameters except a couple that they think work but don't (they will fiddle)

Does it work?



- Simple case, bulk MgO
- Greed/Trust start low, then rapidly increase

More complex: atoms+density

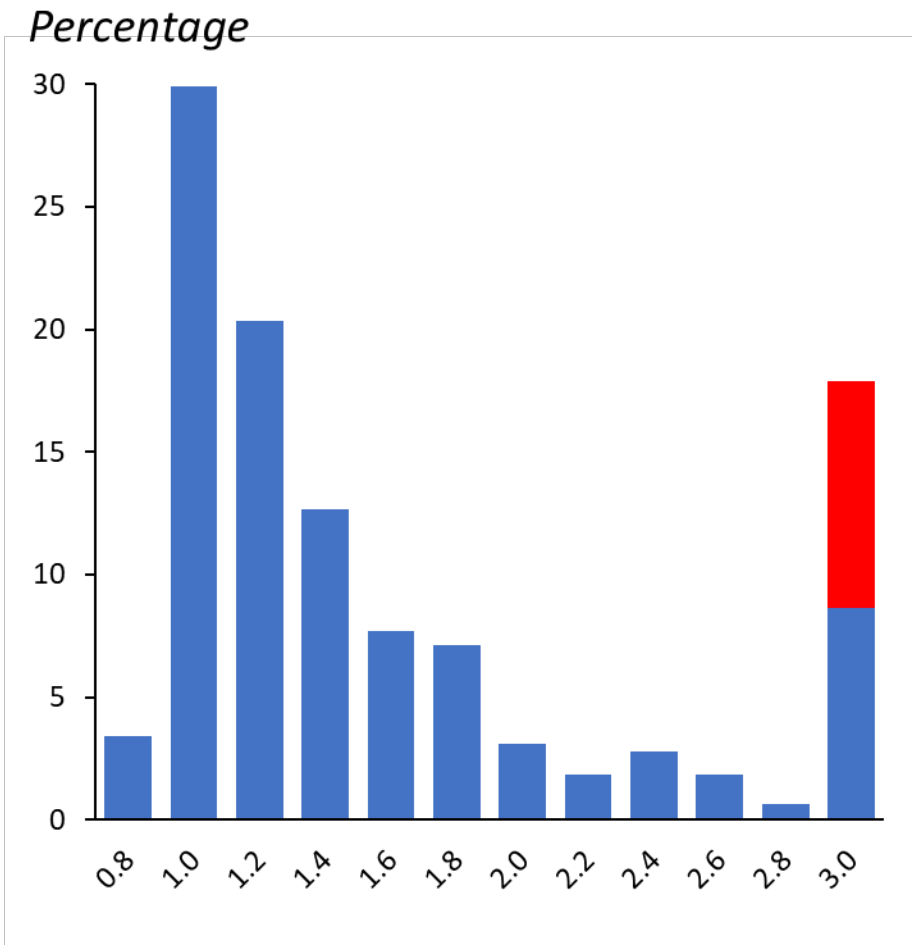


- MSR1: Predictive
- MSEC: λ=0
- MSGB: λ=1
- DIIS: λ=0, different scaling

DIIS/MSEC don't converge
MSGB is noisy (too greedy)

Statistics

Ration of iterations to converge of predictive approach versus “best” (by search) of greed for 36 cases (blue), and 4 which never converged in red

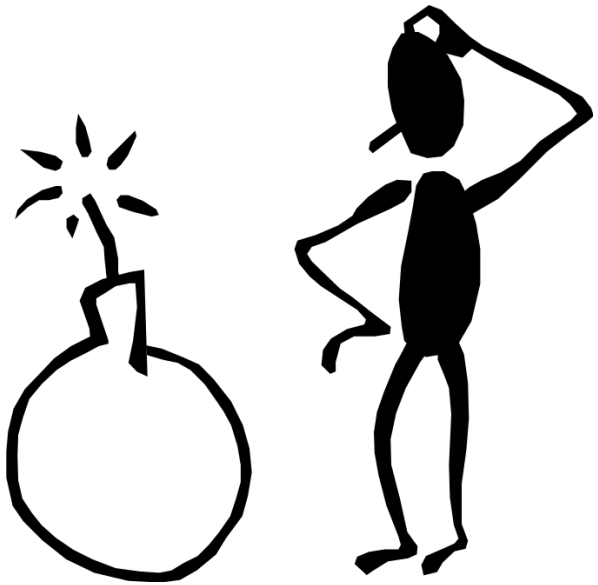


Current status

- A detailed search can often find parameters as good – at the cost of doing 10 times or more as many calculations
- It (and earlier versions) have survived the acid test – novice users
- Follows the math
- No rubber bands

Overview

- Role of mixing and force minimization
- Nuts and bolts in Wien2k
- Underlying math
  - Simple form
  - What is really going on
  - Predictive mixing
- What to do when you mess up

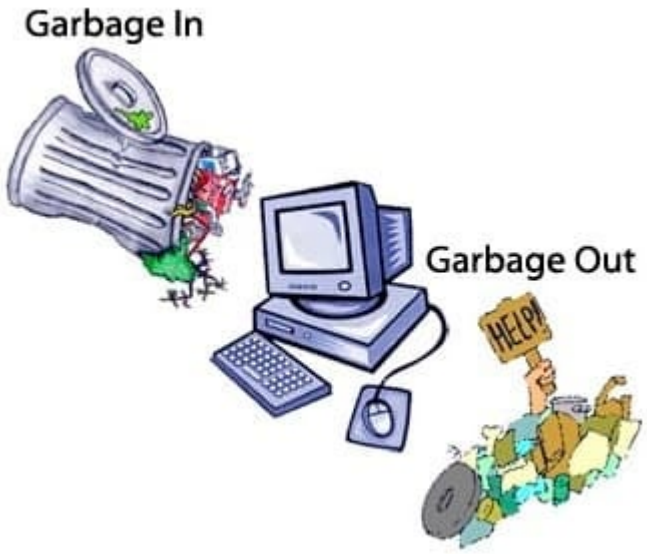


If something goes wrong…

- In most cases this is because:
  - The problem is poorly posed
  - Too few k-points
  - The structure is very far from equilibrium
- So
  - Increase k-points
  - Maybe increase T (TEMPS), sometimes does not work
  - Think

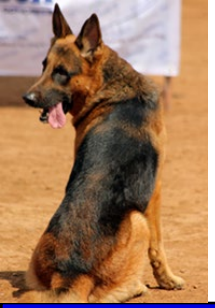
Some solutions are more equal than others

- How do we know the model is correct?
  - All models will refine to an improved fit, but this does not prove they are right
  - It is rarely plausible to test with DFT calculations all possible models



Be careful

- All software has bugs – by definition
- Many computers are “broke”
- Many calculations are at the limit of what can be done
- Few computer codes will tell you “don’t do that you stupid #!?”
- Never assume theory (or experiment) is correct
- Think about what you are doing as computer experiments
- Think science, not just typing at a terminal



Questions ?

It is through science that we prove, but through intuition that we discover.
Jules H. Poincaré

Marks, L.D. and D.R. Luke, *Robust mixing for ab initio quantum mechanical calculations.* Physical Review B, 2008. **78**(7): p. 075114-12 http://doi.org/10.1103/PhysRevB.78.075114.

Marks, L.D., *Fixed-Point Optimization of Atoms and Density in DFT.* J Chem Theory Comput, 2013. **9**(6): p. 2786-800 http://doi.org/10.1021/ct4001685.

Marks, L.D., *Predictive Mixing for Density Functional Theory (and Other Fixed-Point Problems).* J Chem Theory Comput, 2021. **17**(9): p. 5715-5732 http://doi.org/10.1021/acs.jctc.1c00630.