Duke
UNIVERSITY
MECHANICAL ENGINEERING
& MATERIALS SCIENCE

Practical Implementations: The Nuts and Bolts of DFT

Part II

Björn Lange

Outline

- *Part I: The Self Consistency Cycle (SCF)*
  - *Density mixing*
  - *Metallic systems: Fermi level broadening*
  - *Preconditioning*
- *Part II: Structure optimization*
  - *Local structure optimization*
  - *„Forces“ in density functional theory*
  - *Vibrations in the harmonic approximation*
- *Part III: Beyond conventional CPUs*
  - *Calculating with Graphical Processing Units (GPUs)*

Recap: Electron Density

- *All ground state properties are related to the electron distribution [1]:*

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n) \Longleftrightarrow n(\mathbf{r})$$

- *Map electron density on effective one-particle orbitals with energy εᵢ [2]:*

$$n(\mathbf{r}) \Longrightarrow \sum_i^{\infty} f_i \left|\phi_i(\mathbf{r})\right|^2 \qquad H[n]\phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r})$$

- *„Occupation numbers“*

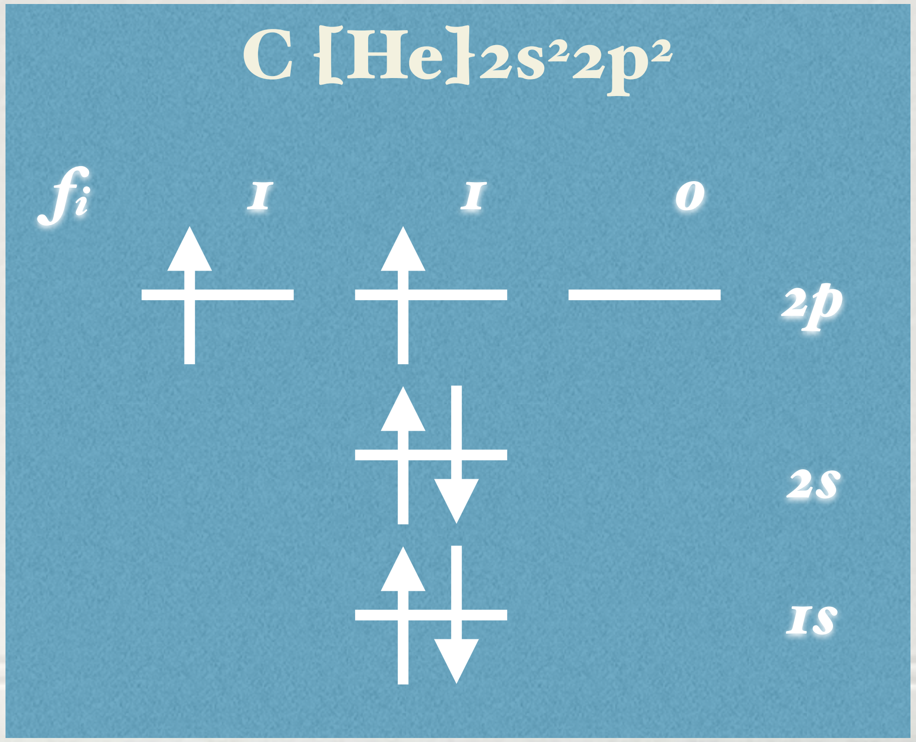
$$f_i = \begin{cases} 0 \ldots 2 & \text{not spin-polarized} \\ 0 \ldots 1 & \text{spin-polarized} \end{cases}$$

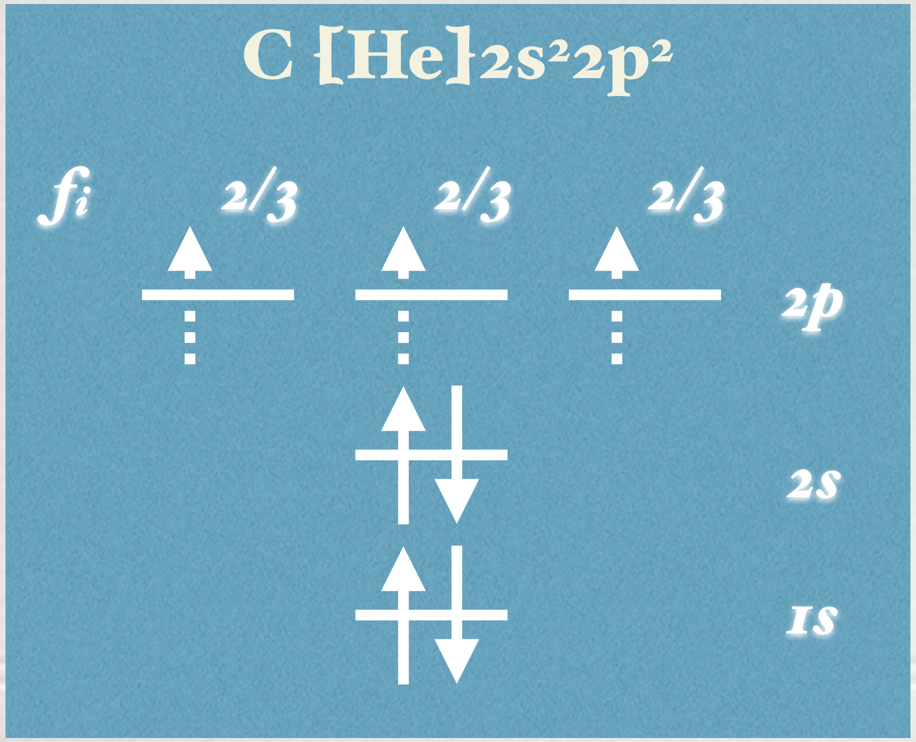
[1] P. Hohenberg, W. Kohn, Phys Rev. (1964), B864

[2]: W. Kohn, L.J. Sham, Phys. Rev. (1965), A1133

Fractional Occupation Numbers

- *Justified as a statistical average* $$n(\mathbf{r}) \Longrightarrow \sum_{i}^{\infty} f_i \left|\phi_i(\mathbf{r})\right|^2$$
- *Practical necessity: Brillouin zone integrals for metallic systems (later in this talk)*
- *Formal necessity: Physics of metals needs statistical mechanics*
- *Can lead to unexpected occupations for pure state systems*


C [He]$2s^22p^2$


C [He]$2s^22p^2$

Recap: Kohn-Sham Equations

$$H[n]\phi_i(\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r})$$

$$\left(\boxed{-\frac{1}{2}\nabla^2} + \boxed{\int_\Omega \frac{n(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' + V_{\mathrm{xc}}(\mathbf{r})} + \boxed{V_{\mathrm{ext}}(\mathbf{r})}\right)\phi_i(\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r})$$

kinetic contribution

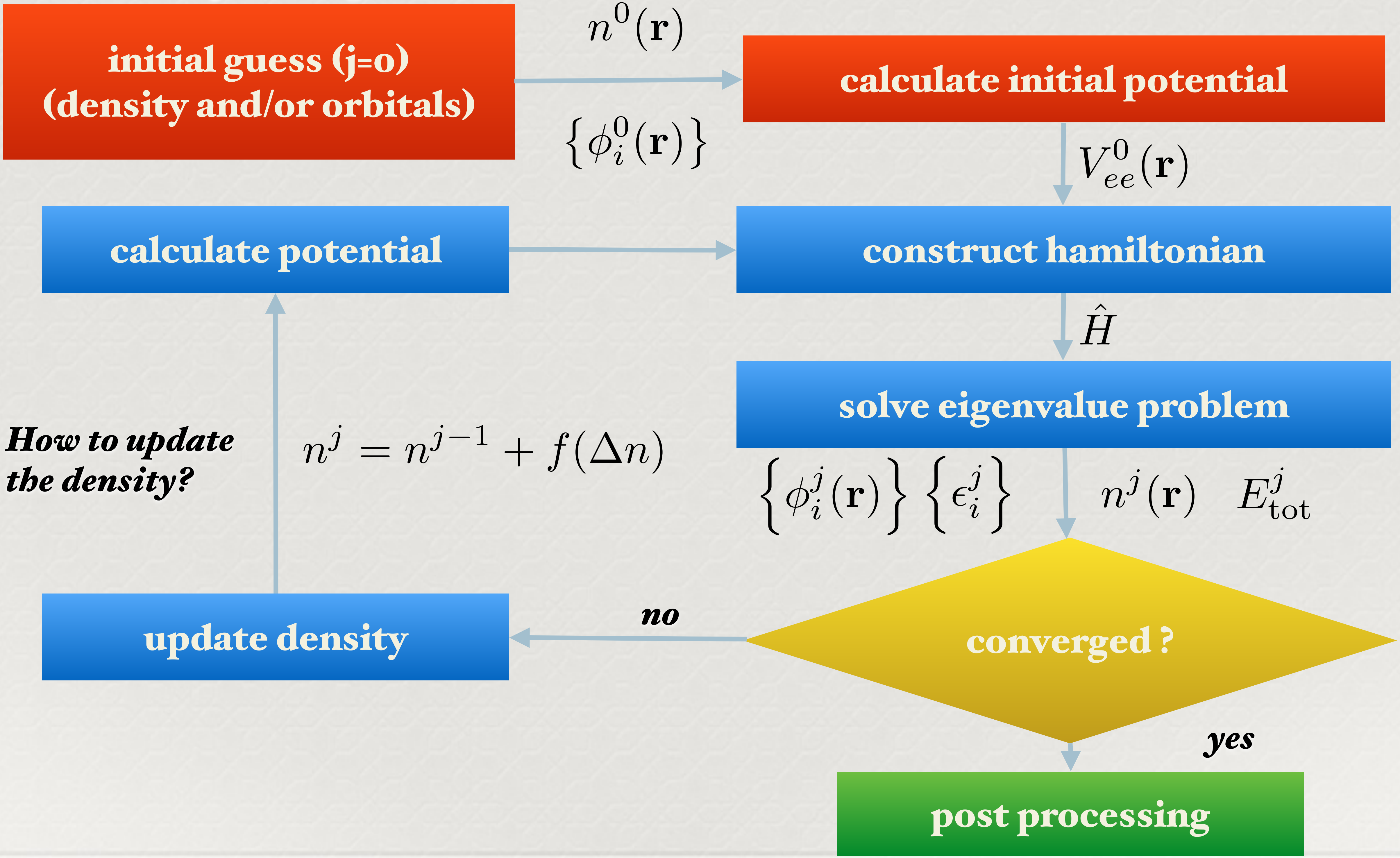
electron-electron interaction

electron-ion interaction
external fields
…

- *We want to determine all solutions Φ_i such that*
  - *the total energy E[n] is minimized*
  - *H[n] is consistent with $\{\Phi_i\}$*

Self Consistent Field Method (S.C.F.)

What kind of initial guess?



initial guess (j=0)
(density and/or orbitals)

$n^0(\mathbf{r})$

$\{\phi_i^0(\mathbf{r})\}$

calculate initial potential

$V_{ee}^0(\mathbf{r})$

calculate potential

construct hamiltonian

$\hat{H}$

solve eigenvalue problem

$\left\{\phi_i^j(\mathbf{r})\right\} \left\{\epsilon_i^j\right\}$

$n^j(\mathbf{r}) \quad E_{\mathrm{tot}}^j$

converged ?

no

update density

How to update the density?

$n^j = n^{j-1} + f(\Delta n)$

yes

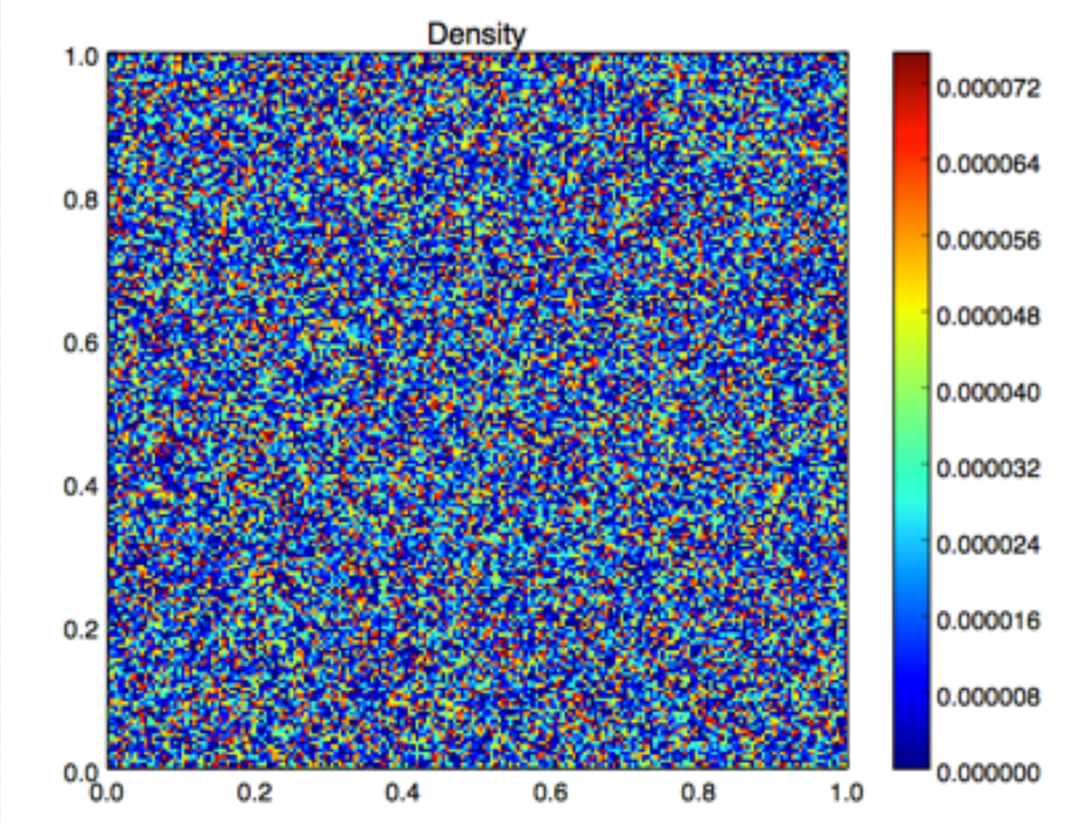
post processing

The Initial Guess

- *Random Numbers ?*
  - *mainly in plane-wave based methods*
  - *easy to implement*
  - *no chemical motivation (unbiased)*
  - *no chemical motivation (slow)*
  - *Completely impractical for an all electron code (density is dominated by core electrons close to the nuclei)*

$$n^0(\mathbf{r}_i) = RND([0\ldots 1])$$

$$\int_\Omega n^0(\mathbf{r})d\mathbf{r} = N_e$$

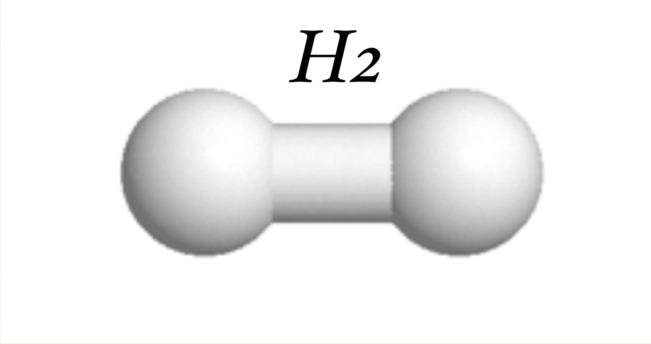


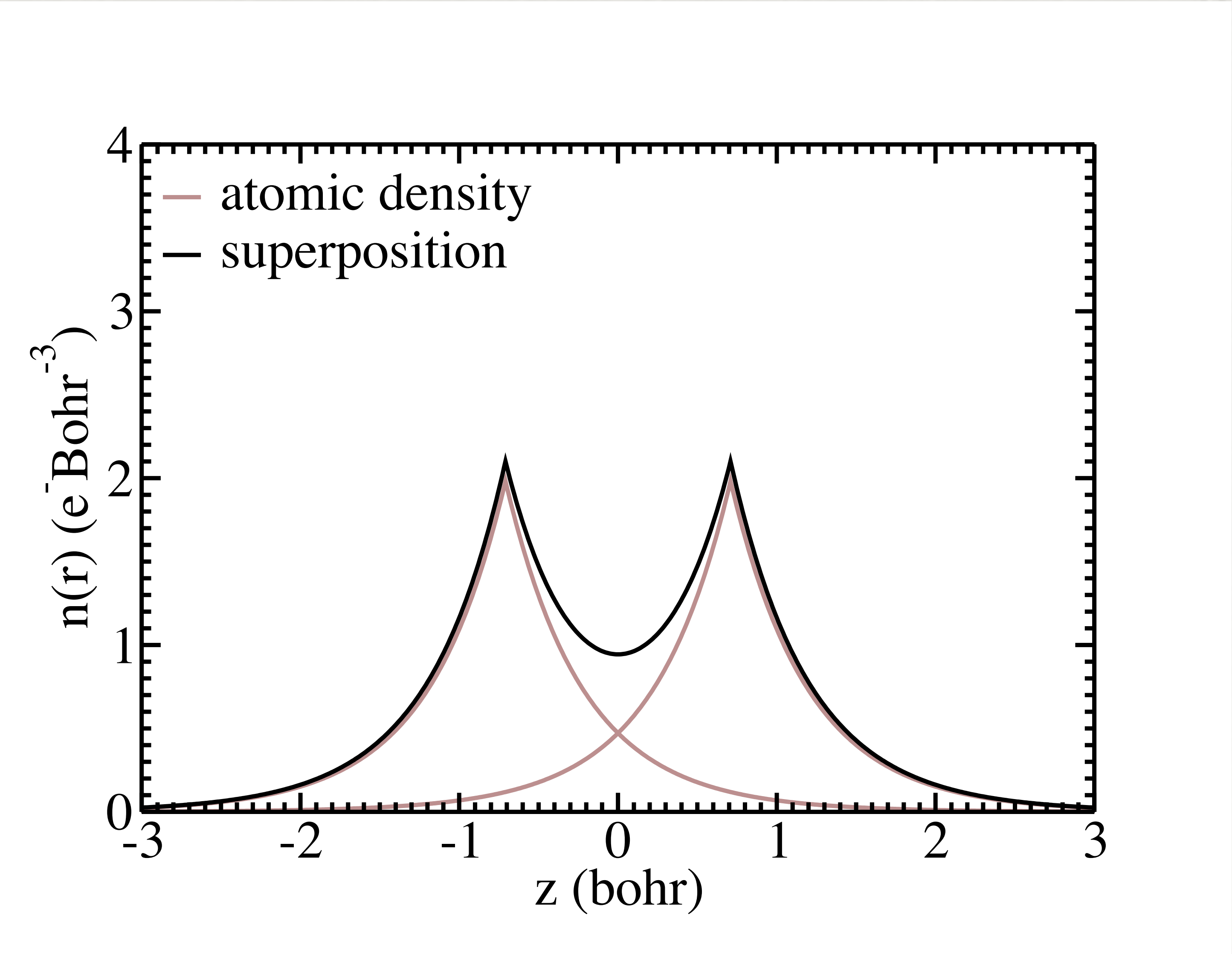
The Initial Guess

- *Superposition of atomic densities (used in FHI-aims):*

$$n^{\mathrm{at}}(\mathbf{r}) = \sum_i f_i^{\mathrm{occ}} \left|\chi_i(\mathbf{r})\right|^2 \qquad n^0(\mathbf{r}) = \sum_I^{N_{\mathrm{at}}} n_I^{\mathrm{at}}(\mathbf{r} - \mathbf{R}_I)$$

- *straightforward to implement*
- *Can overemphasise symmetry*





- *Extended Hückel Theory [1] (in Quantum Chemistry codes)*
- *not implemented in FHI-aims*
  - *Linear Combination of Atomic Orbitals*

$$\phi_i^0(\mathbf{r}) = \sum_j c_{ij}\chi_j(\mathbf{r} - \mathbf{R}_j)$$

  - *Hamiltonian [2,3]:*

$$H_{ii}\text{: parameterized atomic valence ionization energies}$$

$$H_{ij} = KS_{ij}\frac{H_{ii} + H_{jj}}{2}$$

  - *Solve set of linear equations* $\hat{H}\mathbf{c}_i = \epsilon_i \hat{S}\mathbf{c}_i$
  - *Improved flexibility, specific orbitals can be populated*

$$n^0(\mathbf{r}) = \sum_i f_i^{\mathrm{occ}}|\phi_i^0(\mathbf{r})|^2$$

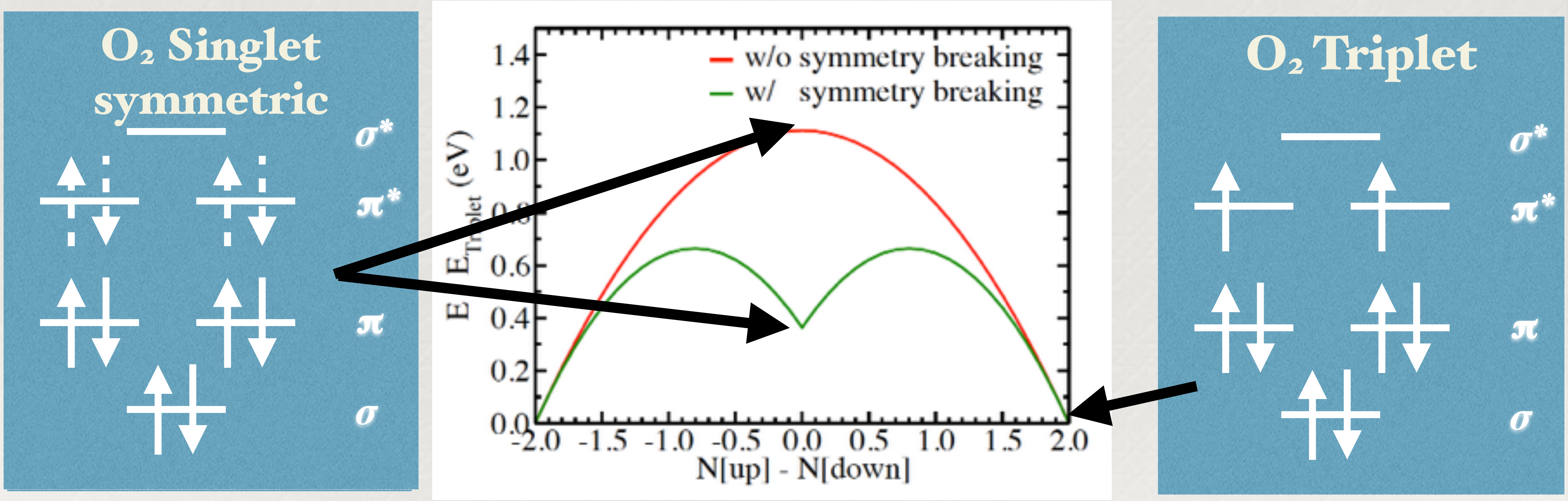
[1] *R. Hoffmann, J Chem. Phys* (**1963**), *1397*

[2] *R. S. Mulliken, J. Chem. Phys.* (**1946**) *497*

[3] *M. Wolfsberg and L. Helmholtz, J. Chem. Phys* (**1952**), *837*

Importance of Initial Guess

- *Different minima might exist, leading to qualitative different properties*
- *e.g. O_2: singlet vs. triplet*
- *symmetric solutions can be higher in energy*



Density Update

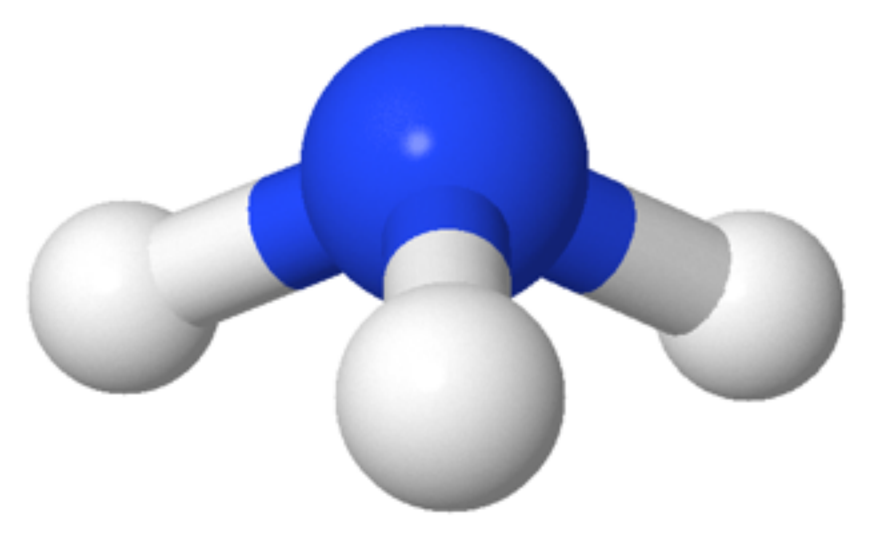
- *Evaluate new KS density from current KS orbitals*

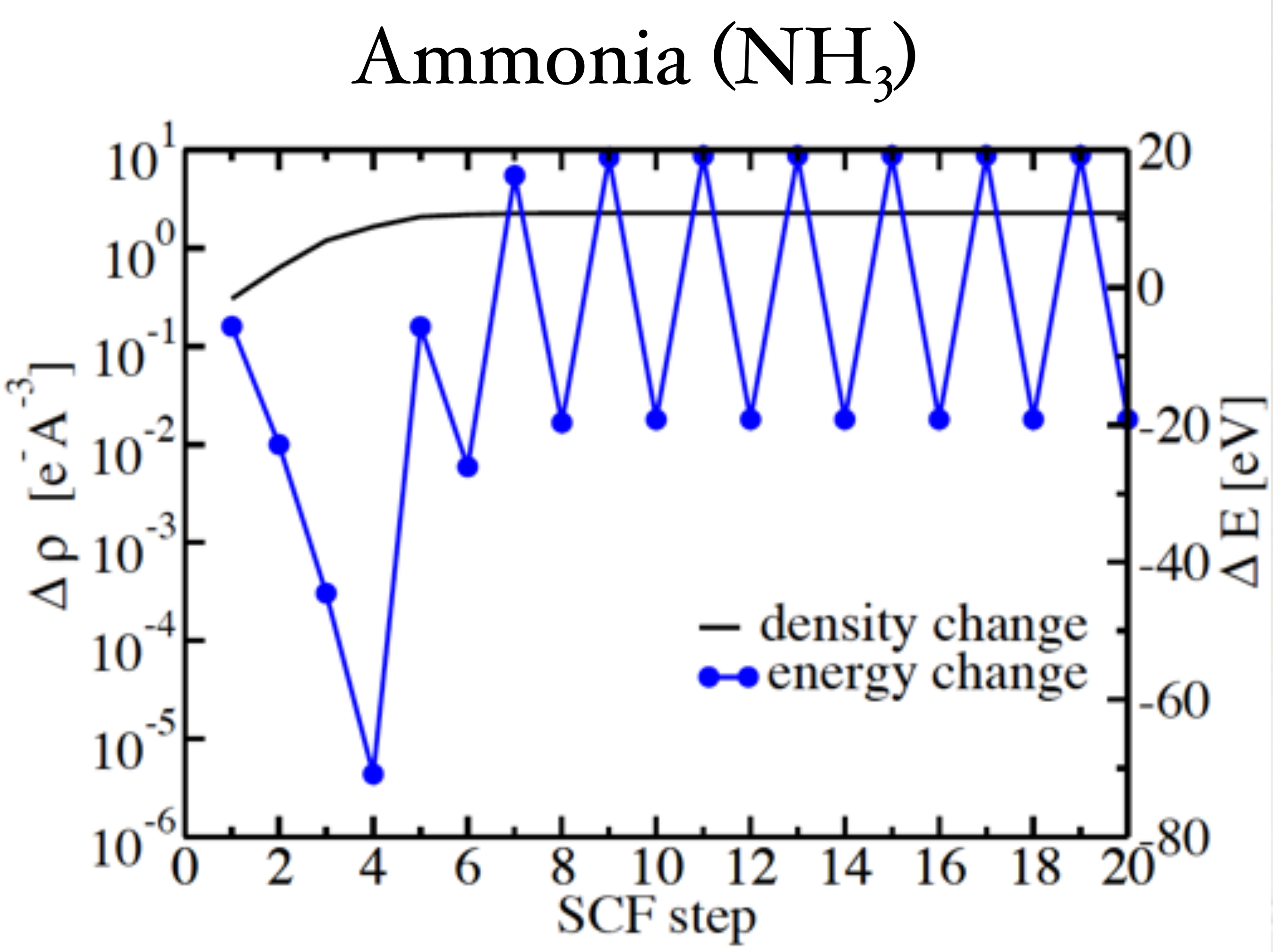
$$n^j_{\mathrm{KS}}(\mathbf{r}) = \sum_i f_i |\phi^j_i(\mathbf{r})|^2$$

- *Naive way: Replace old density by new density*

$$n^j(\mathbf{r}) = n^j_{\mathrm{KS}}(\mathbf{r})$$

- *no convergence*
- *oscillating total energy*




Ammonia (NH$_3$)

Simple Linear Mixing

- *Define difference density*

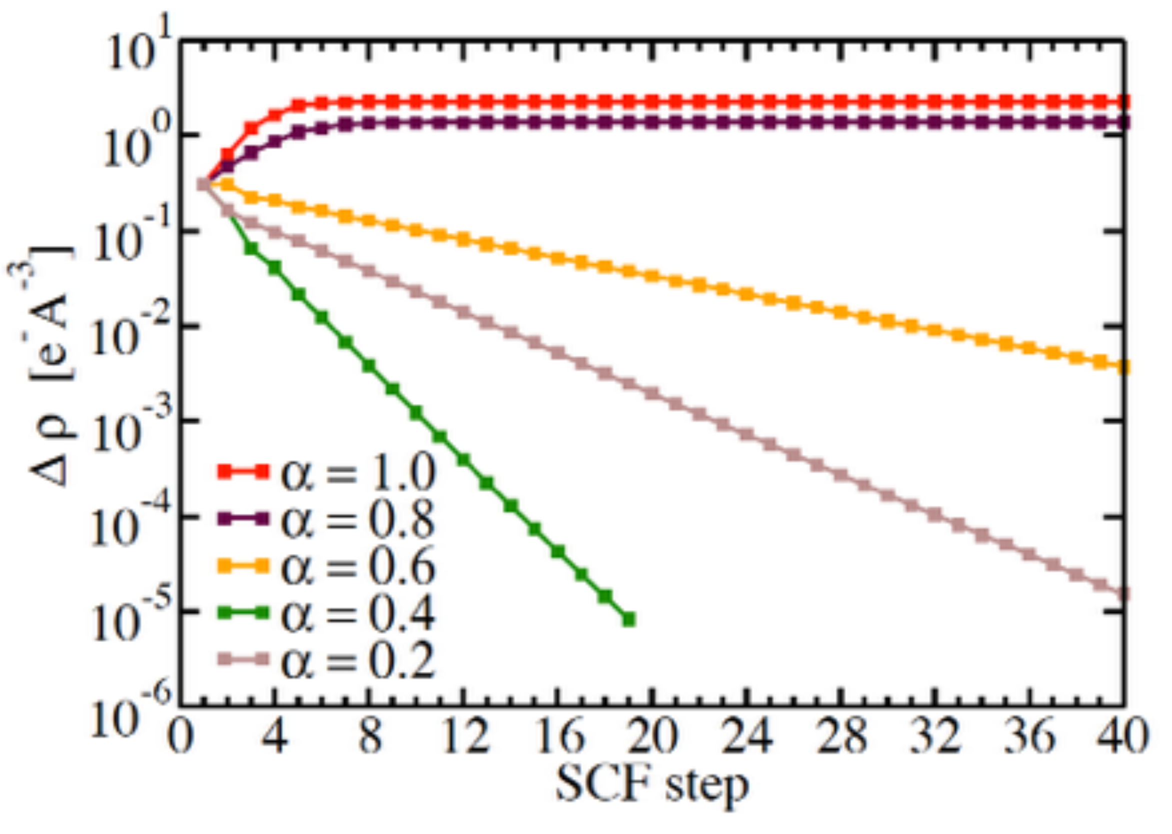
$$\Delta n(\mathbf{r}) = n^{j}_{\mathrm{KS}}(\mathbf{r}) - n^{j-1}(\mathbf{r})$$

- *Take only a fraction of this (underrelaxation method):*

$$n^{j}(\mathbf{r}) = n^{j-1}(\mathbf{r}) + \alpha\Delta n(\mathbf{r})$$

  - *too high α:*
    unstable SCF behaviour
  - *too small α:*
    slow convergence
  - *optimal α is*
    system dependent

Ammonia (NH_3)



Pulay mixing (Direct inversion of iterative subspace)

- *the residual density: change between KS density and input density*

$$R^j(\mathbf{r}) = n^j(\mathbf{r}) - n^j_{\mathrm{KS}}(\mathbf{r})$$

- *Try to predict R^{j+1} from N previous residues*

$$R^{j+1}(\mathbf{r}) = \sum_{i=0}^{N-1} \beta_i R^{j-i}(\mathbf{r})$$

  - *find β_i by minimizing*

$$F = ||R^{j+1}(\mathbf{r})||^2 - \lambda \left(1 - \sum_{i=0}^{N} \beta_i \right)$$

[1] *P. Pulay, Chem. Phys. Lett. 73 , 393 (1980)*

Pulay mixing (Direct inversion of iterative subspace)

- *Construct „optimal density“:*

$$n^{\mathrm{opt}}(\mathbf{r}) = n^{j}_{\mathrm{KS}} + R^{j}(\mathbf{r})$$

- *be conservative*

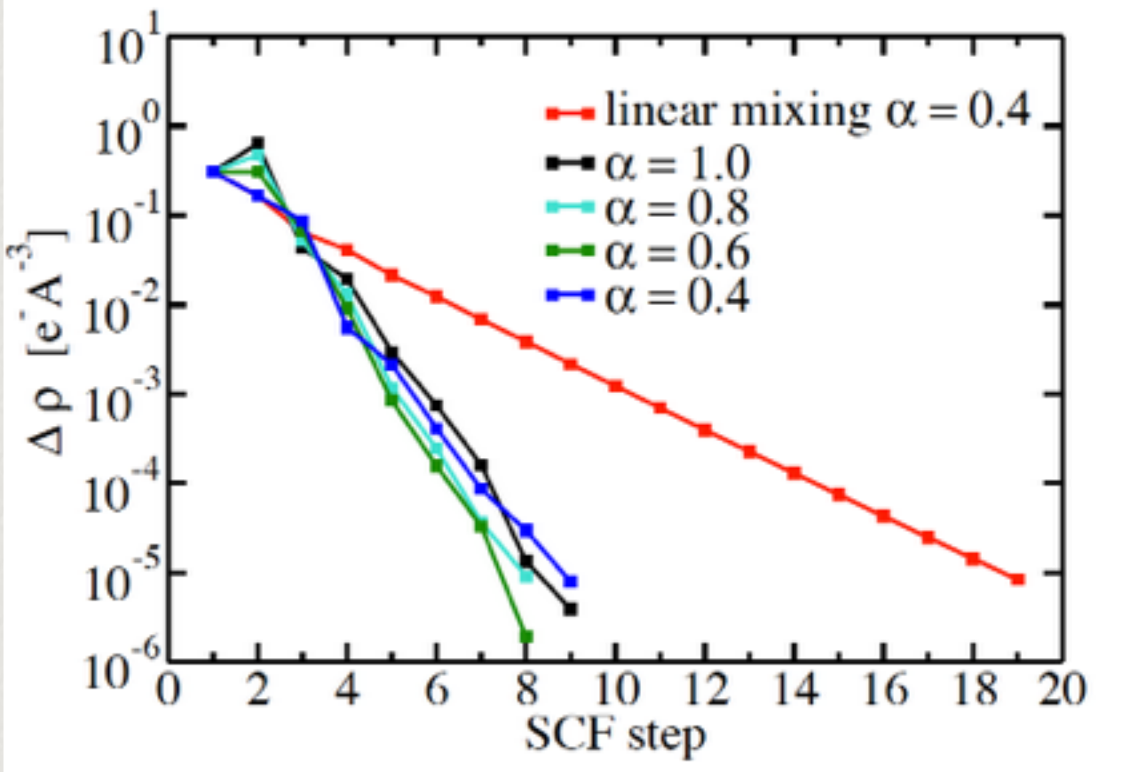
$$n^{j}(\mathbf{r}) = (1-\alpha)n^{j-1}(\mathbf{r}) + \alpha n^{\mathrm{opt}}(\mathbf{r})$$

- *increase of convergence speed*
- *can be nearly independent of α*
- *for insulating systems*

  α = *0.2 - 0.6*
- *for metallic systems*

  α *can be 0.05*

Ammonia (NH_3)



Metallic systems

- *Metallic systems are more demanding in convergence*
- *Band crossing at the Fermi level*
- *Bad convergence of band structure energy in reciprocal space:*

$$E_{\mathrm{Band}} = \sum_n \frac{1}{\Omega} \int_\Omega \epsilon_{n\mathbf{k}} \Theta(\epsilon_{n\mathbf{k}} - \epsilon_{\mathrm{Fermi}}) d\mathbf{k}$$

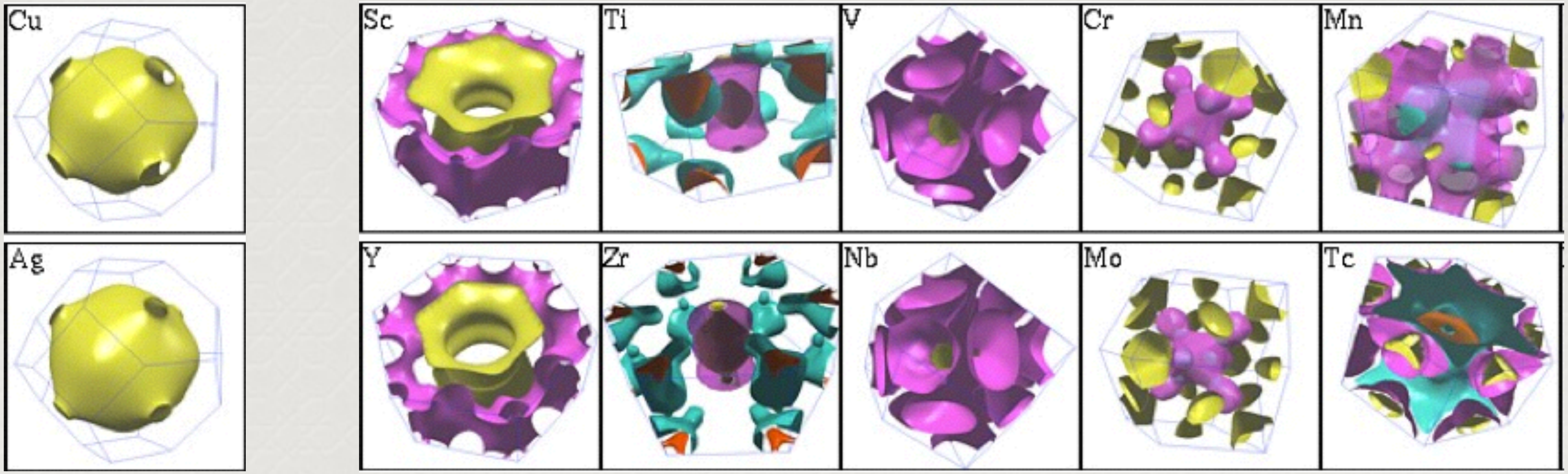
- *We operate with a finite k-mesh*

$$\frac{1}{\Omega} \int_\Omega d\mathbf{k} \rightarrow \sum_{N_k} w_{\mathbf{k}}$$

- *for metals : T=0K occupancies would jump* ***discontinuously*** *from 1 to 0 at the Fermi level*

Metallic systems

- *Fermi surfaces = collection of k-points with energy* ε_{Fermi}
- *related to Fermi level crossing of bands*
  (structure-rich Fermi surface[1])



- *„level switching“ of partially filled orbitals during SCF possible*

[1] *http://www.phys.ufl.edu/fermisurface/periodic_table.html*

Fermi level broadening

- *Stabilize integration by replacing Θ with a smoother function*

$$f_{\text{Fermi-Dirac}}\left(\frac{\epsilon-\epsilon_{\text{Fermi}}}{\sigma}\right)=\frac{1}{\exp\left(\frac{\epsilon-\epsilon_{\text{Fermi}}}{\sigma}\right)+1}$$ *[1]*

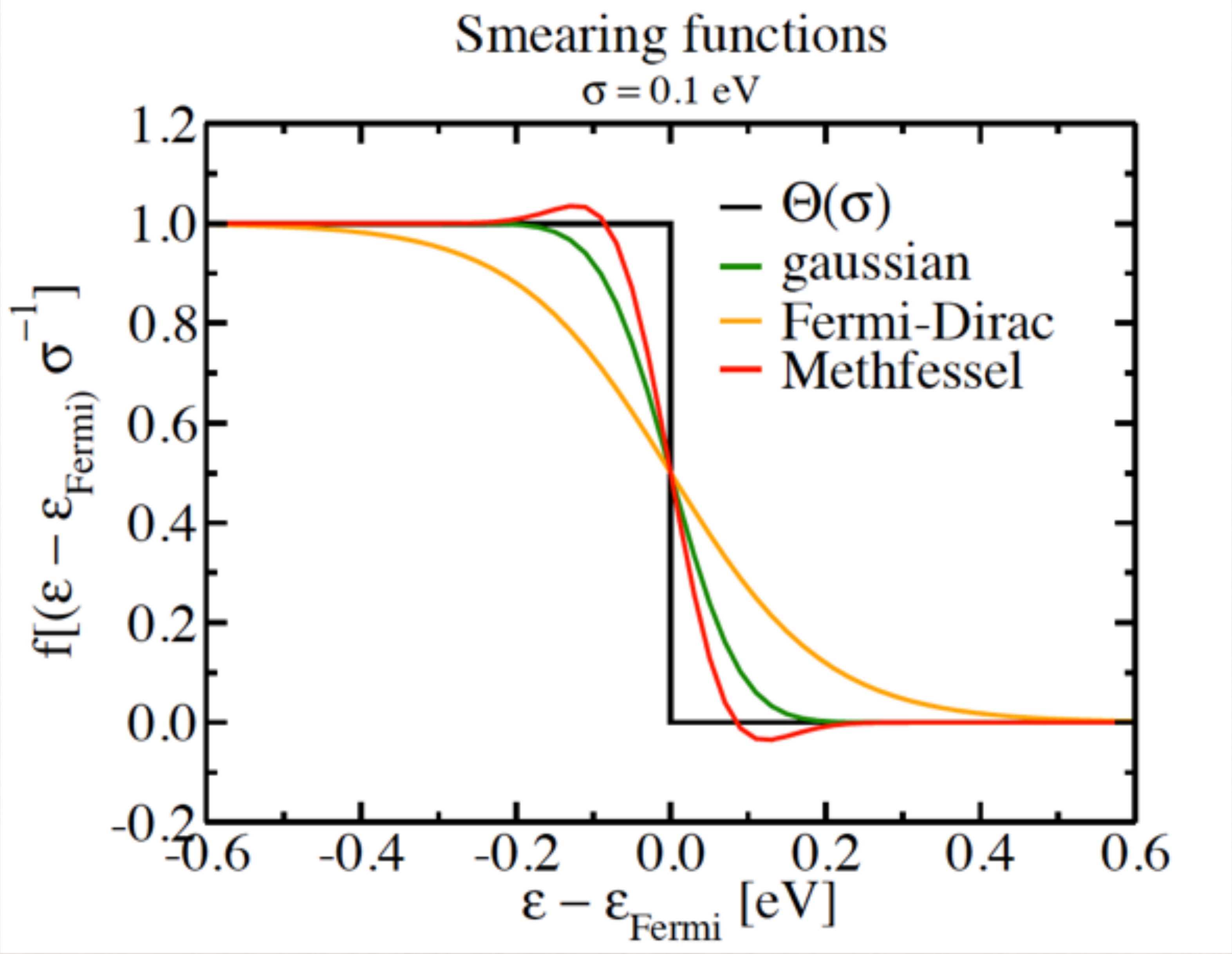
$$f_{\text{Gaussian}}\left(\frac{\epsilon-\epsilon_{\text{Fermi}}}{\sigma}\right)=\frac{1}{2}erfc\left(\frac{\epsilon-\epsilon_{\text{Fermi}}}{\sigma}\right)$$ *[2]*

$$f_{\text{Methfessel-Paxton}}\left(\frac{\epsilon-\epsilon_{\text{Fermi}}}{\sigma}\right)=\frac{1}{2}erfc\left(\frac{\epsilon-\epsilon_{\text{Fermi}}}{\sigma}\right)+\sum_{i=1}^{n}A_iH_{2i-1}\left(\frac{\epsilon-\epsilon_{\text{Fermi}}}{\sigma}\right)\exp\left(-\left(\frac{\epsilon-\epsilon_{\text{Fermi}}}{\sigma}\right)^2\right)$$ *[3]*

[1] Mermin, N. D. (1965), Phys.Rev., 137(5A), A1441–A1443
[2] Fu, C. L., & Ho, K. M. (1983), Physical Review B, 28, 5480–5486.
[3] Methfessel, M., & Paxton, A. T. (1989) Physical Review B, 40(6), 3616–3621

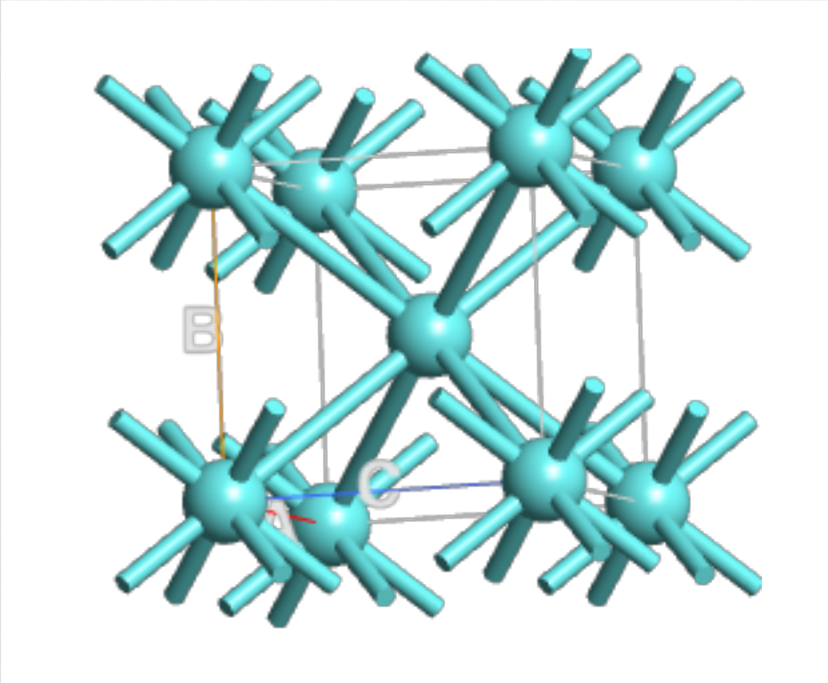
Fermi Level Broadening

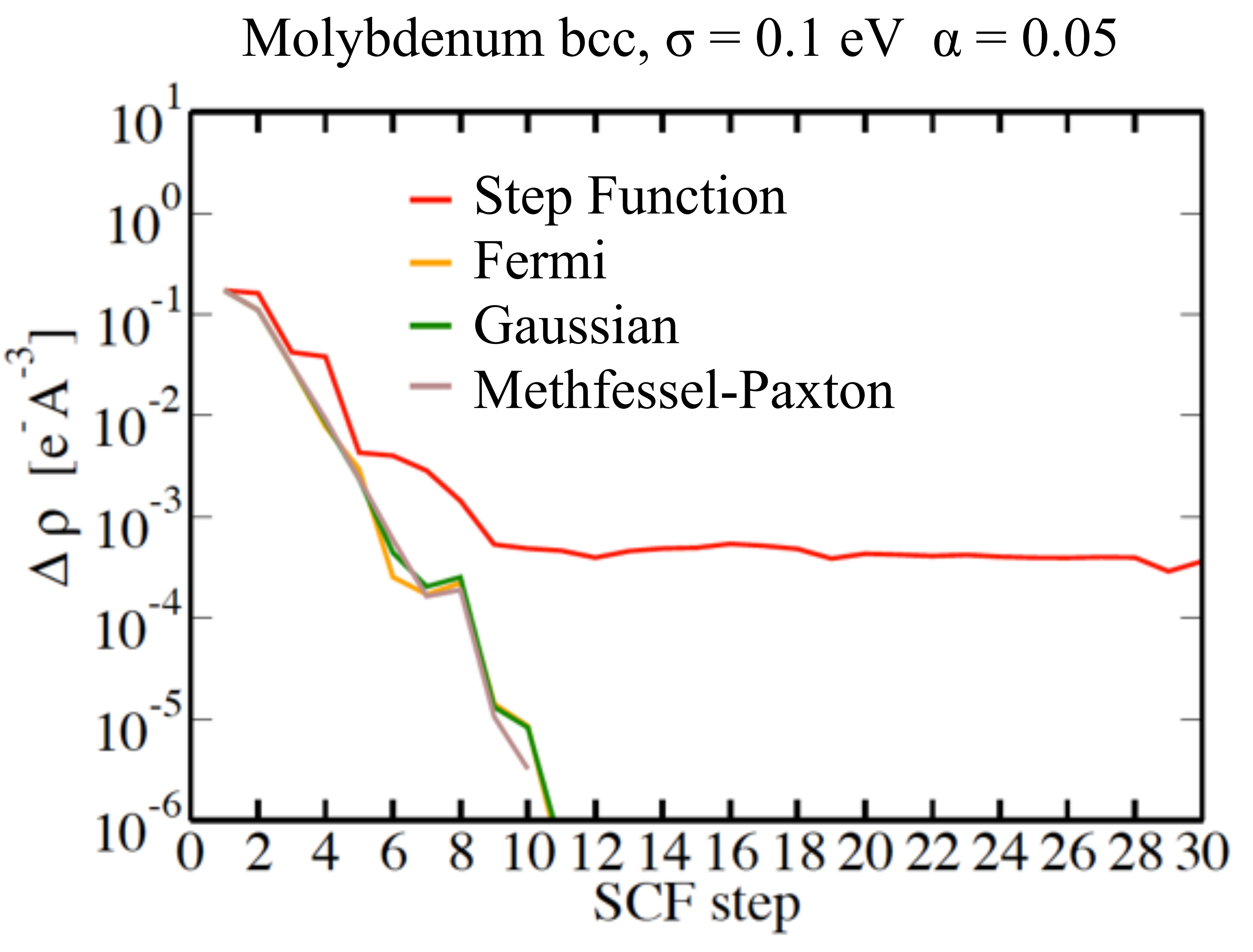
- *electronic smearing avoids discontinuity at the Fermi level*
- *can be related to electronic temperature in Fermi-Dirac method*



Convergence

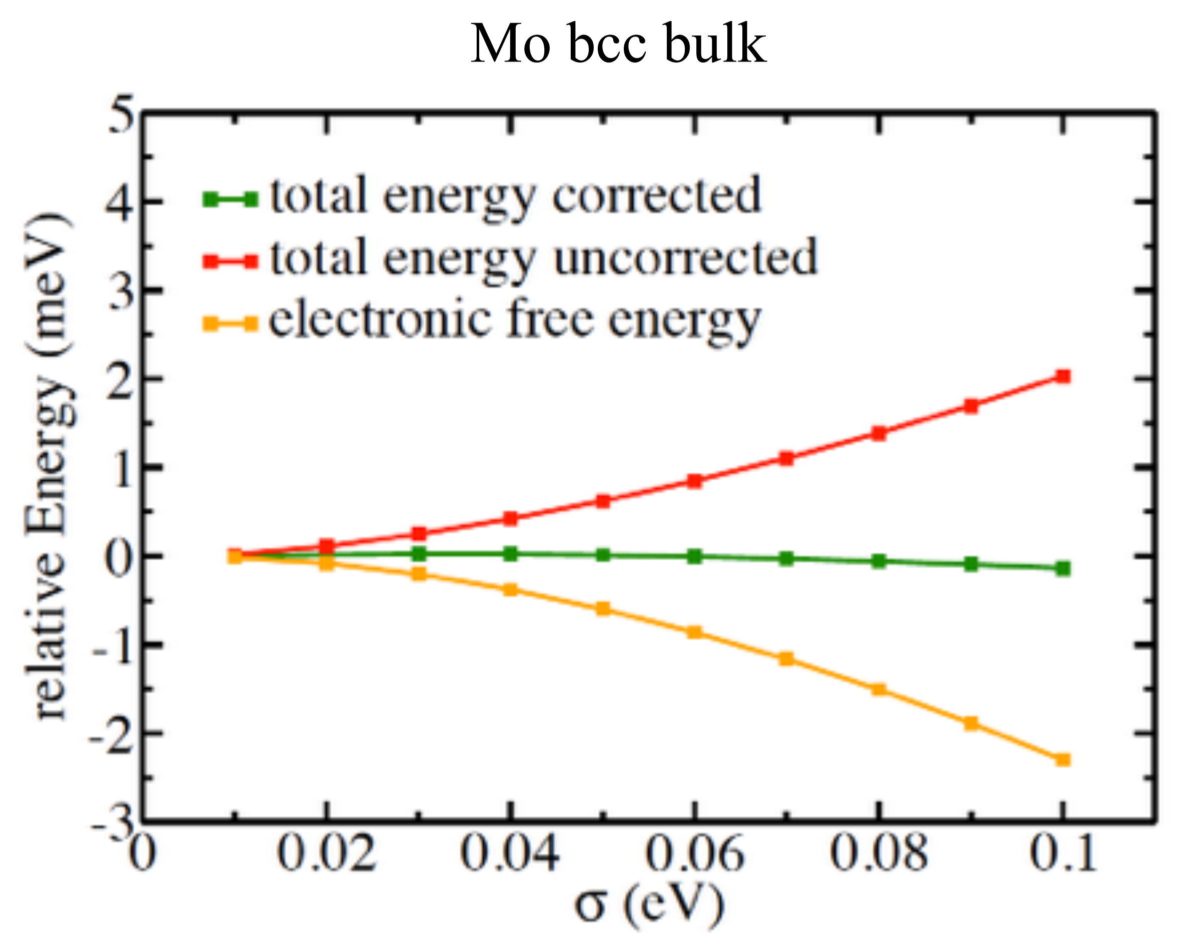
- *Step function approach fails to converge*
- *Electronic broadening crucial to speed up the convergence*





Free Energy

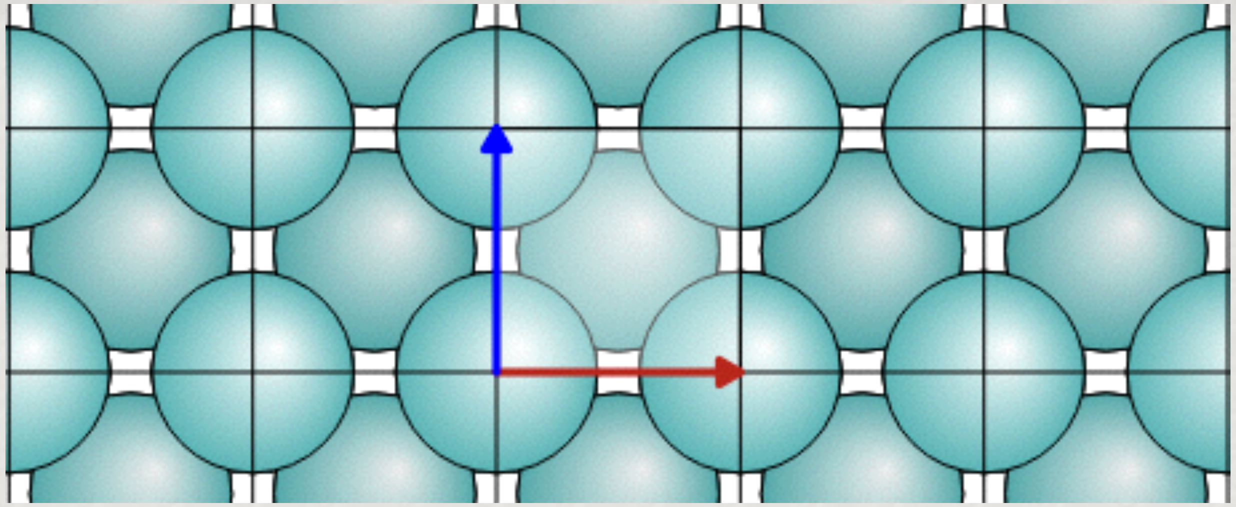
- *Total energy depends on variations in* f_i
- *Optimize free energy instead* $F = E - \sigma S(\sigma)$
- *Backextrapolation possible* $\sigma \longrightarrow 0$
- *Extrapolation Error << 1 meV*
- *Applicable for metals,*
  but not, in general, for atoms
  or small molecules
  with discrete electronic levels



Kresse, G., & Furthmüller, J. (1996). Computational Materials Science, 6(1), 15–50
M. Weinert and J.W. Davenport. Phys. Rev. B. 45 (1992) 13709.
R.M. Wentzcovitch, J.L. Martins and P.B. Allen, Phys. Rev B 45 (1992) 11372

Surfaces, Interfaces, Defects, Thin Films

- *e.g. Mo(100) surface*
  - *difficult electronic structure*



- *Long range density fluctuations in Surfaces, Interfaces, Defects, and Thin Films can lead to* ***charge sloshing*** *[1]*
- *Let mixing depend on the density* $n^j(\mathbf{r}) = n^{j-1}(\mathbf{r}) + \hat{G}\Delta n(\mathbf{r})$
- *Idea: Treat density in Fourier space and damp long range density fluctuations*

$$\Delta n(\mathbf{G}) = \frac{1}{\Omega}\int_{\Omega} \Delta n(\mathbf{r}) e^{-\mathbf{G}\cdot\mathbf{r}} d\mathbf{r}$$

[1] Kerker, G. P. (1981). Physical Review B, 23(6), 3082–3084

Kerker Preconditioning

- *Based on Thomas Fermi Theory in reciprocal space[1,2]:*

$$\hat{G}(\mathbf{G}) = \beta \frac{|\mathbf{G}|^2}{|\mathbf{G}|^2 + \lambda^2}$$

β: mixing parameter
λ: screening constant

- *For G >> λ*

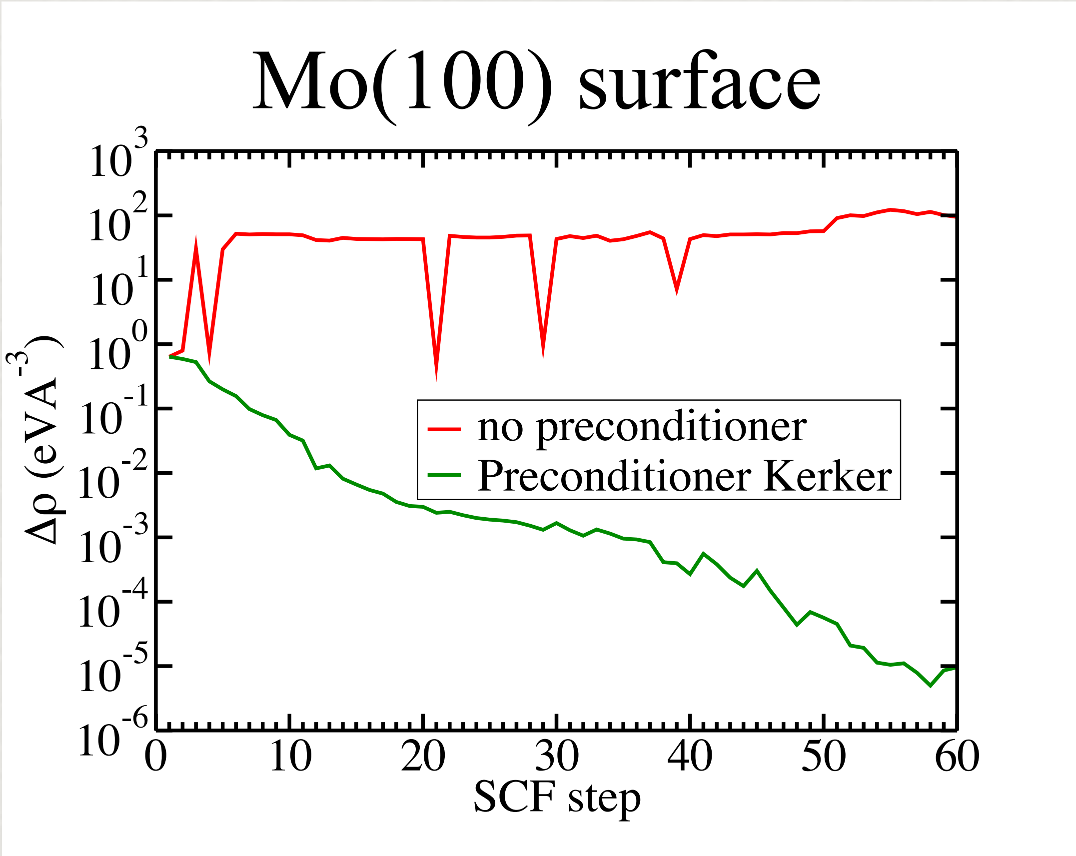
  (short-range density fluctuations)

$$\hat{G}(\mathbf{G}) \approx \beta \quad \text{normal mixing}$$

- *For G << λ*

  (long-range density fluctuations)

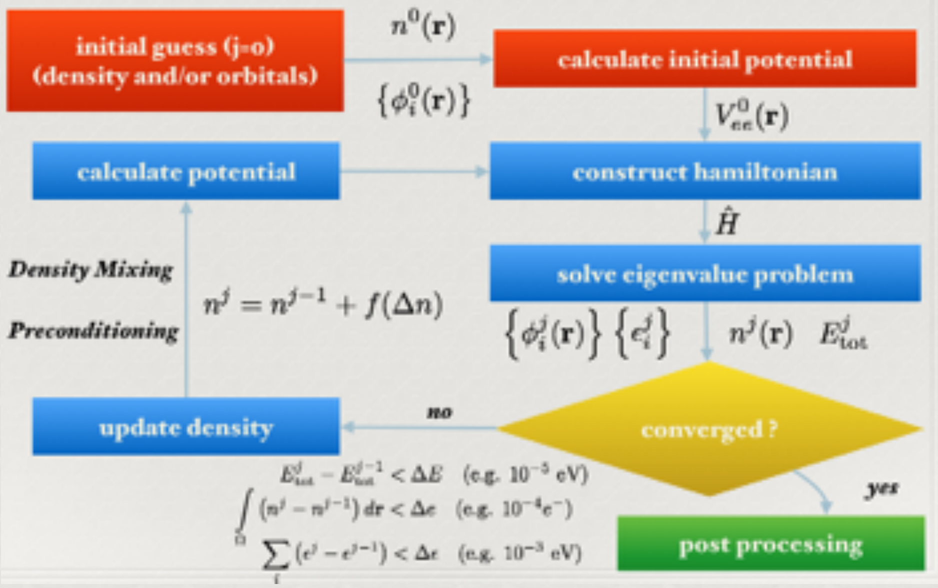
$$\hat{G}(\mathbf{G}) \approx \beta \frac{|G|^2}{\lambda^2} \quad \text{effective damping}$$



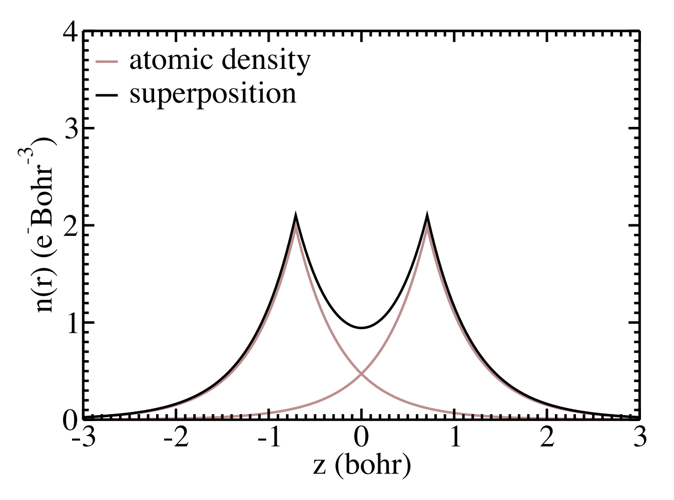
[1] M. Manninen, R. Nieminen, and P. Hautojavri, Phys. Rev. B 12, 4012 (1975)
[2] Kerker, G. P. (1981). Physical Review B, 23(6), 3082–3084

Summary SCF

- *the SCF cycle*



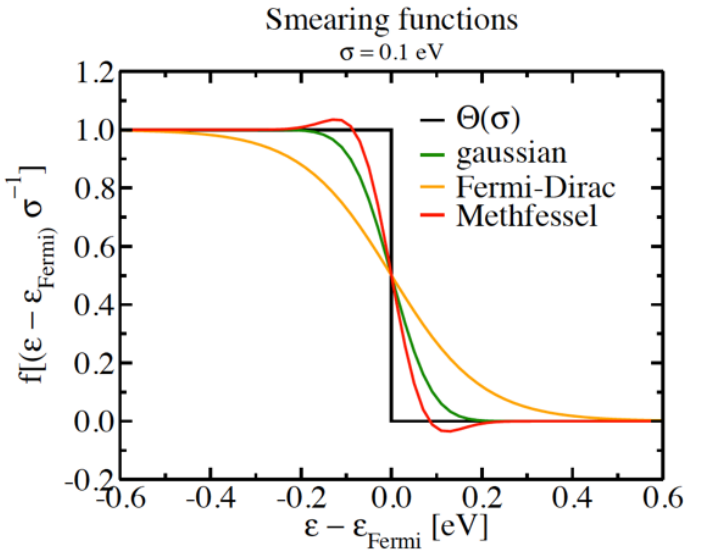
- *initial guess*



- *approaches of density mixing*

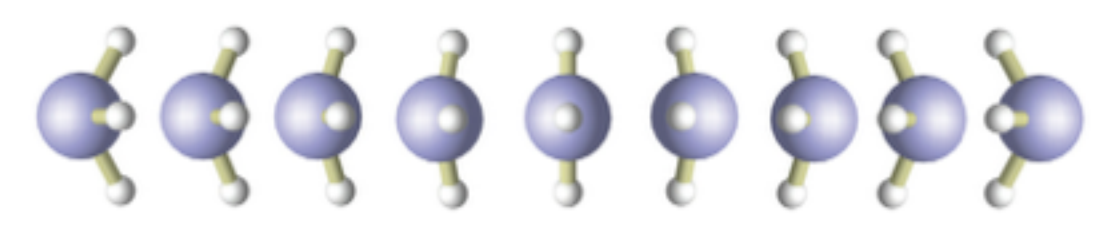
$$n^j(\mathbf{r}) = (1-\alpha)n^{j-1}(\mathbf{r}) + \alpha n^{\text{opt}}(\mathbf{r})$$

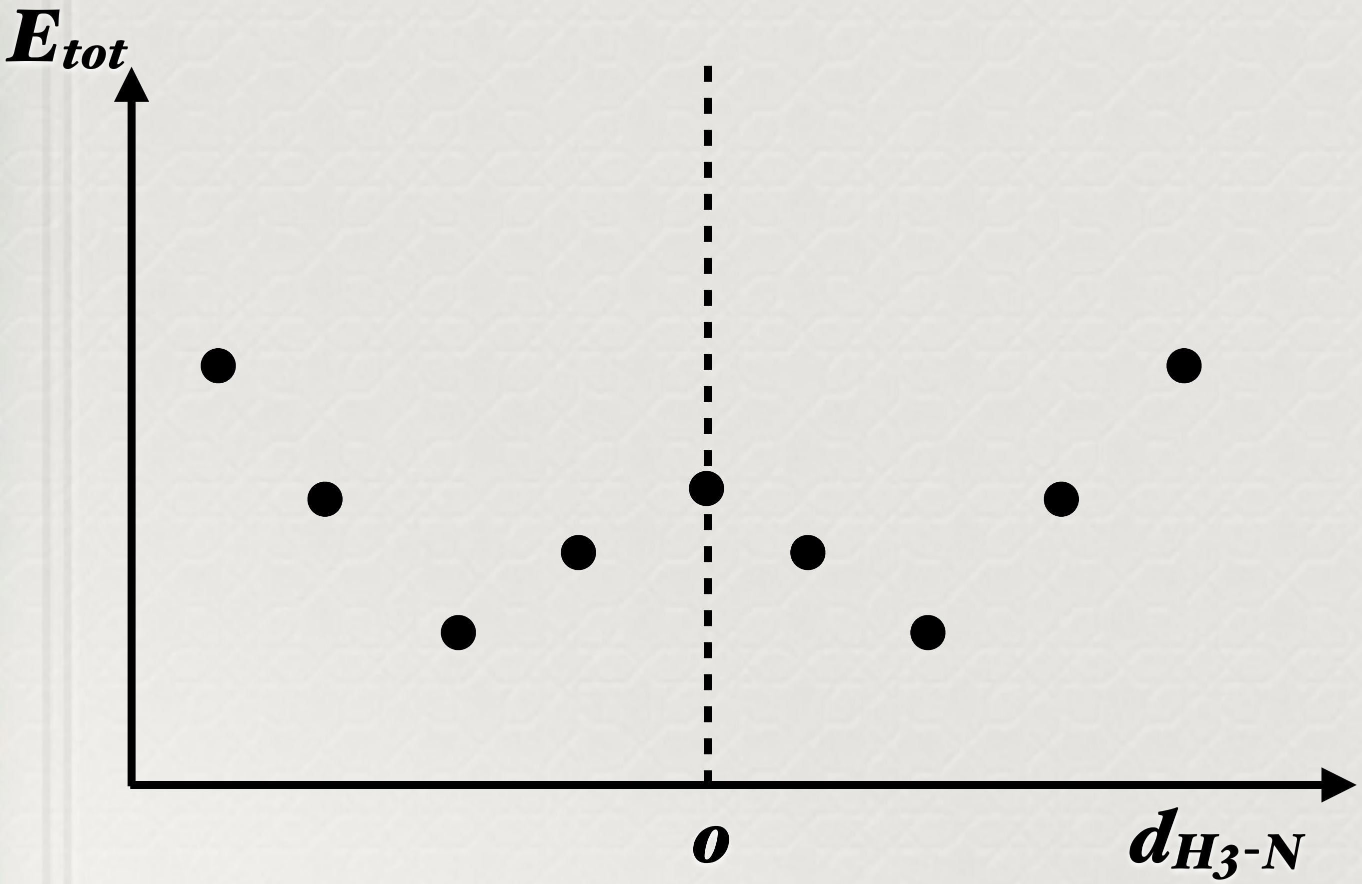
- *Fermi level broadening*

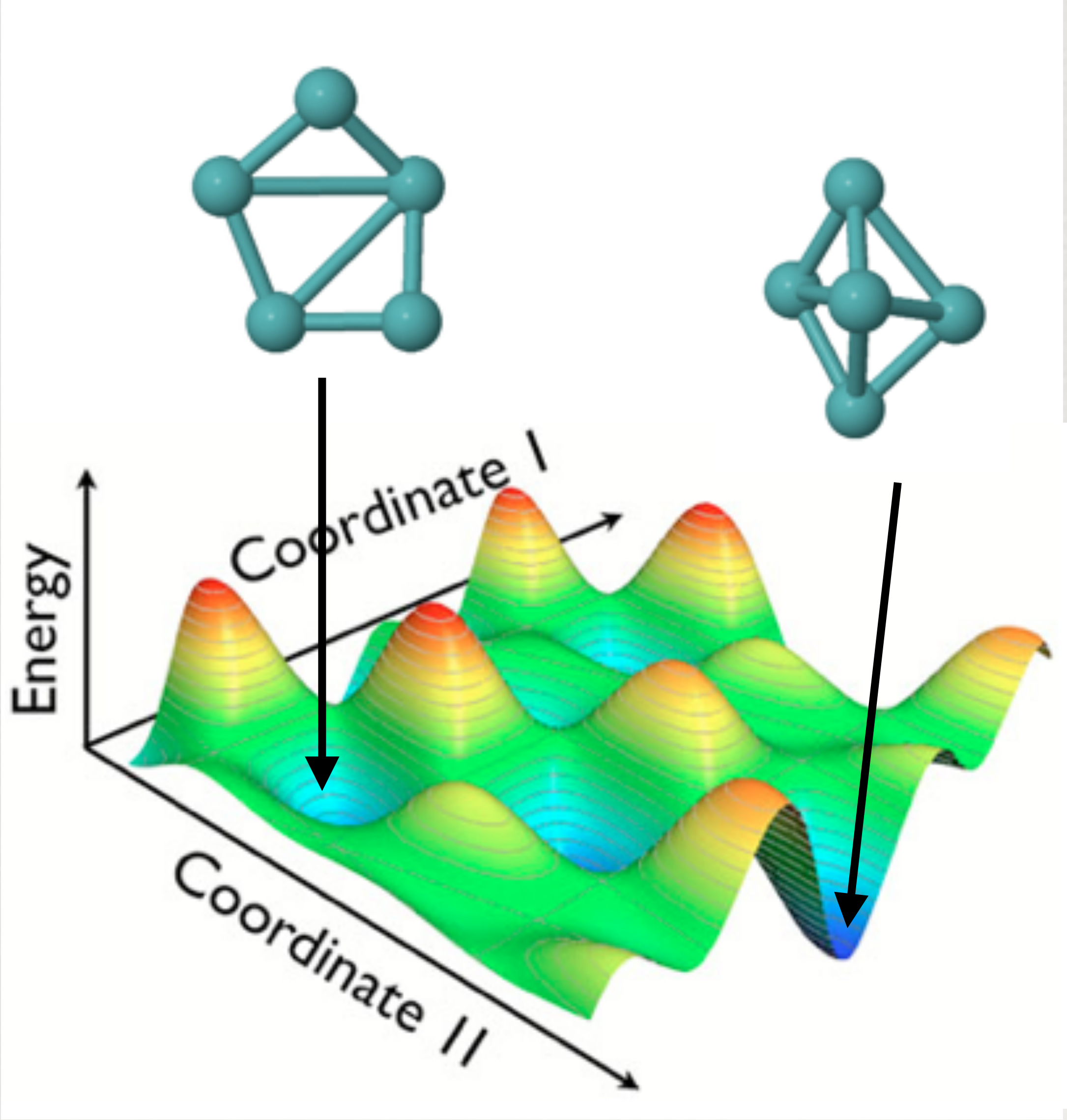


- *preconditioner* $\hat{G}(\mathbf{G}) = \beta \dfrac{|\mathbf{G}|^2}{|\mathbf{G}|^2 + \lambda^2}$

Geometry optimization





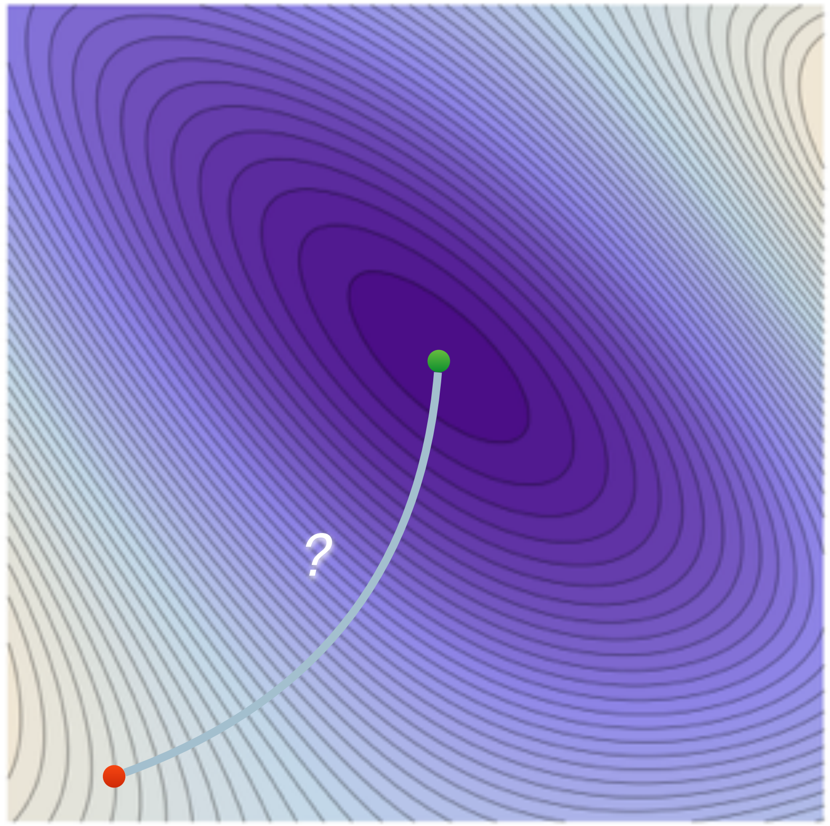


Picture taken from
Frontiers in Bioscience 13, 4495-4516, May 1, 2008

Finding Stable Structures

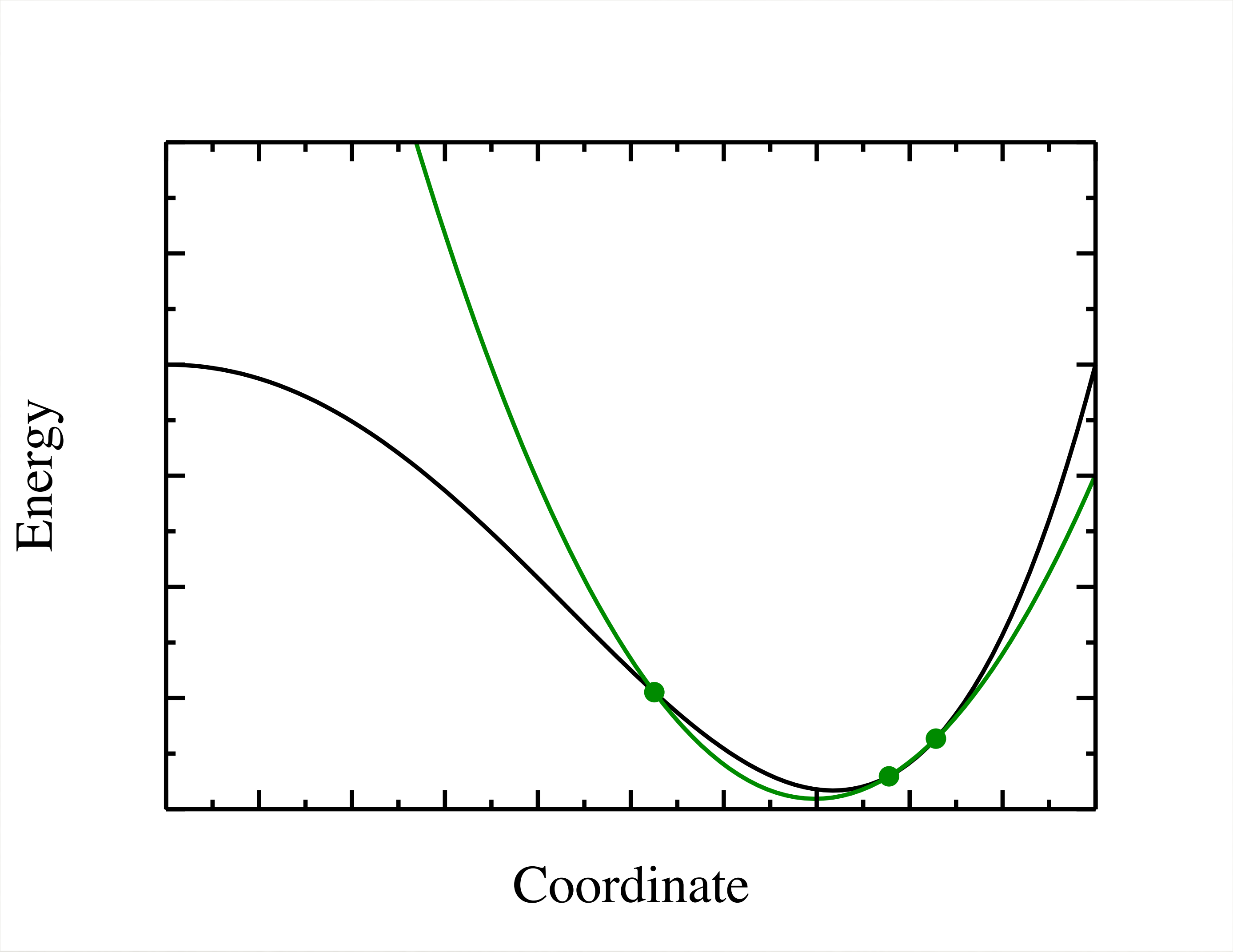
- *„The space of possible structures is endless, as well as the number of methods to find them“ :*
    - *Stochastical or Monte Carlo Methods (Talk: Peter Kratzer)*
    - *Molecular Dynamics (Talk: Luca Ghiringhelli, MauroMaggioni)*
    - *Genetic Algorithms*
    - *Diffusion Methods*
    - *Experimental structure determination*

- *However, once you are near a minimum, you have to find it...*



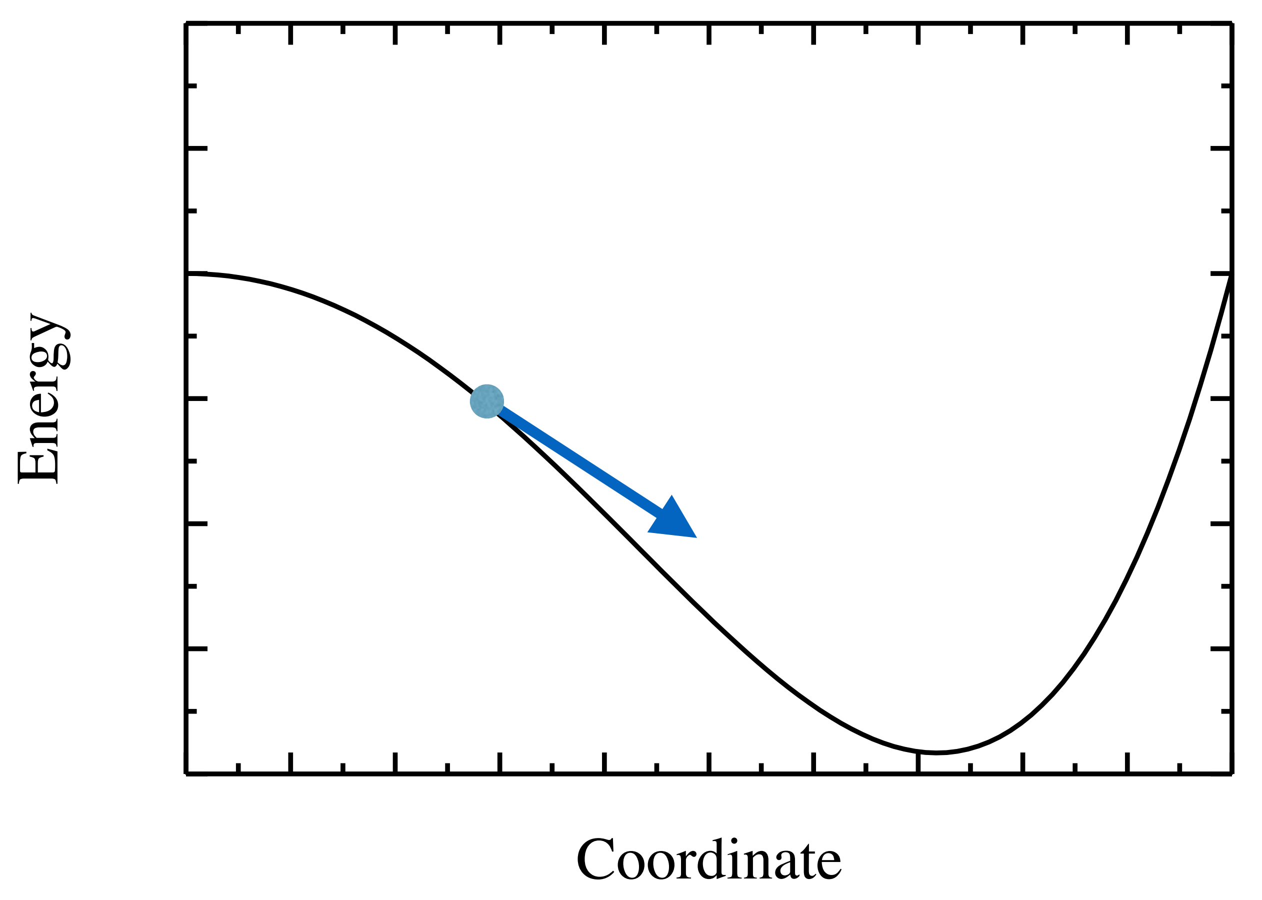
Local Structure Optimization

- *gradient free methods (a „total energy only“ approach)*
- *evaluate a few points*
- *construct new point,*
  e.g. assume harmonicity
- *repeat*
  until self-consistency



Local Structure Optimization

- *gradient based method („Forces“)* $\mathbf{F}_i = -\frac{\delta E_{\text{tot}}}{\delta \mathbf{R}_i}$
- *evaluate Force(s) on the atoms analytically*
- *use standard scheme to optimize the structure*



- *steepest descent (SD)*
  just for illustration
- *conjugate gradient (CG)*
  better than SD
- *(Quasi-) Newton*
  approaches (BFGS)
  state-of-the-art approach

- *How to evaluate the forces?*

Forces (light version)

- *Energy*

$$E = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle}$$

- *Forces*

$$\mathbf{F}_i = -\frac{dE}{d\mathbf{R}_i}$$

- *factor rule:*

$$\frac{dE}{d\mathbf{R}_i} = \frac{d}{d\mathbf{R}_i}\frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle}$$

$$= \frac{\frac{d}{d\mathbf{R}_i}\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} - \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle}\frac{\frac{d}{d\mathbf{R}_i}\langle\Psi|\Psi\rangle}{\langle\Psi|\Psi\rangle}$$

$$= \frac{1}{\langle\Psi|\Psi\rangle}\left(\langle\Psi|\frac{d\hat{H}}{d\mathbf{R}_i}|\Psi\rangle + \langle\frac{d\Psi}{d\mathbf{R}_i}|\hat{H} - E|\Psi\rangle + \langle\Psi|\hat{H} - E|\frac{d\Psi}{d\mathbf{R}_i}\rangle\right)$$

Hellman-Feynman Forces

- *Analyze first term*

$$\mathbf{F}_i = -\frac{1}{\langle\Psi|\Psi\rangle}\left(\boxed{\langle\Psi|\frac{d\hat{H}}{d\mathbf{R}_i}|\Psi\rangle} + \langle\frac{d\Psi}{d\mathbf{R}_i}|\hat{H} - E|\Psi\rangle + \langle\Psi|\hat{H} - E|\frac{d\Psi}{d\mathbf{R}_i}\rangle\right)$$

  - *affects only electron-nucleon and nucleon-nucleon interaction[1]*

$$\mathbf{F}_i^{\text{Hellman-Feynman}} = \sum_j \frac{\delta}{\delta\mathbf{R}_i}\frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} - \int_\Omega n(\mathbf{r})\frac{\delta}{\delta\mathbf{R}_i}\frac{Z_i}{|\mathbf{R}_i - \mathbf{r}|}d\mathbf{r}$$

[1] Feynman, R. P. (1939). "Forces in Molecules". Phys. Rev. 56 (4): 340

Pulay Forces

- *What about the second term ?*

$$\mathbf{F}_i = -\frac{1}{\langle\Psi|\Psi\rangle}\left(\langle\Psi|\frac{d\hat{H}}{d\mathbf{R}_i}|\Psi\rangle + \boxed{\langle\frac{d\Psi}{d\mathbf{R}_i}|\hat{H} - E|\Psi\rangle + \langle\Psi|\hat{H} - E|\frac{d\Psi}{d\mathbf{R}_i}\rangle}\right)$$

- *applying basis expansion results in*
  - *Pulay Forces = artificial force due to moving basis:*

$$\mathbf{F}_i^{\text{Pulay}} = -2\sum_{j,j',n} f_n^{\text{occ}} c_{jn}^{*}\langle\frac{\delta\chi_j}{\delta\mathbf{R_i}}|\hat{H} - \epsilon_n|\chi_{j'}\rangle c_{j'n}$$

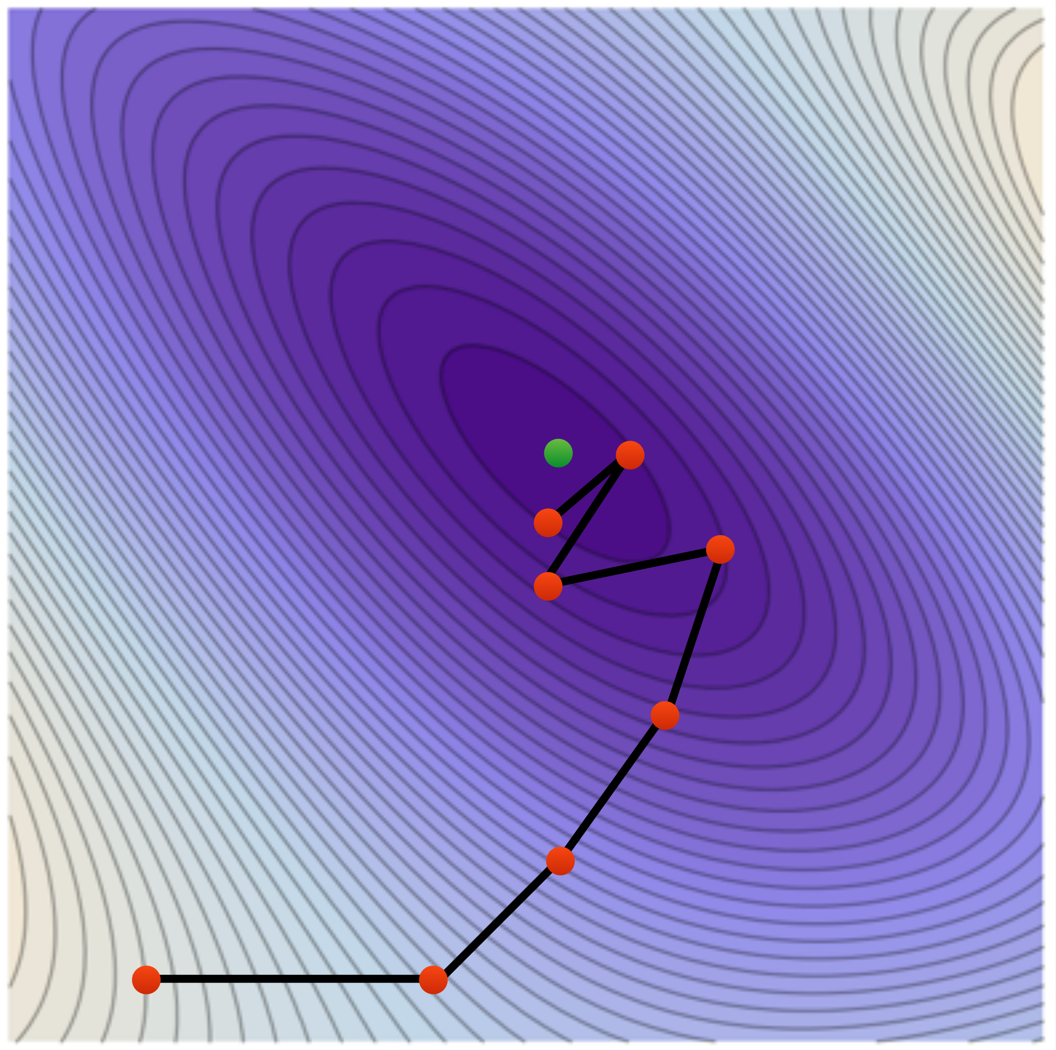
  - *vanishes if basis set is stationary, e.g. plane waves (Talk: Anderson Janotti)*

Steepest Descent

- *Naive approach: Follow gradient in negative direction*

$$\mathbf{R}^n = \mathbf{R}^{n-1} - \alpha \mathbf{F}[\mathbf{R}^{n-1}]$$

  - *Step length as parameter*
  - *Convergence depends on the step length*
  - *Slow, oscillates around path*



- *Improvements:*
  - *Line minimization for optimal step length*
  - *Conjugate Gradient method*

(Quasi-) Newton Method

- *Approximate energy surface as quadratic form*

$$E(\mathbf{R}+\Delta\mathbf{R}) \approx E(\mathbf{R}) + \mathbf{F}^T(\mathbf{R})\Delta\mathbf{R} + \frac{1}{2}(\Delta\mathbf{R})^T\mathbf{H}(\Delta\mathbf{R})$$

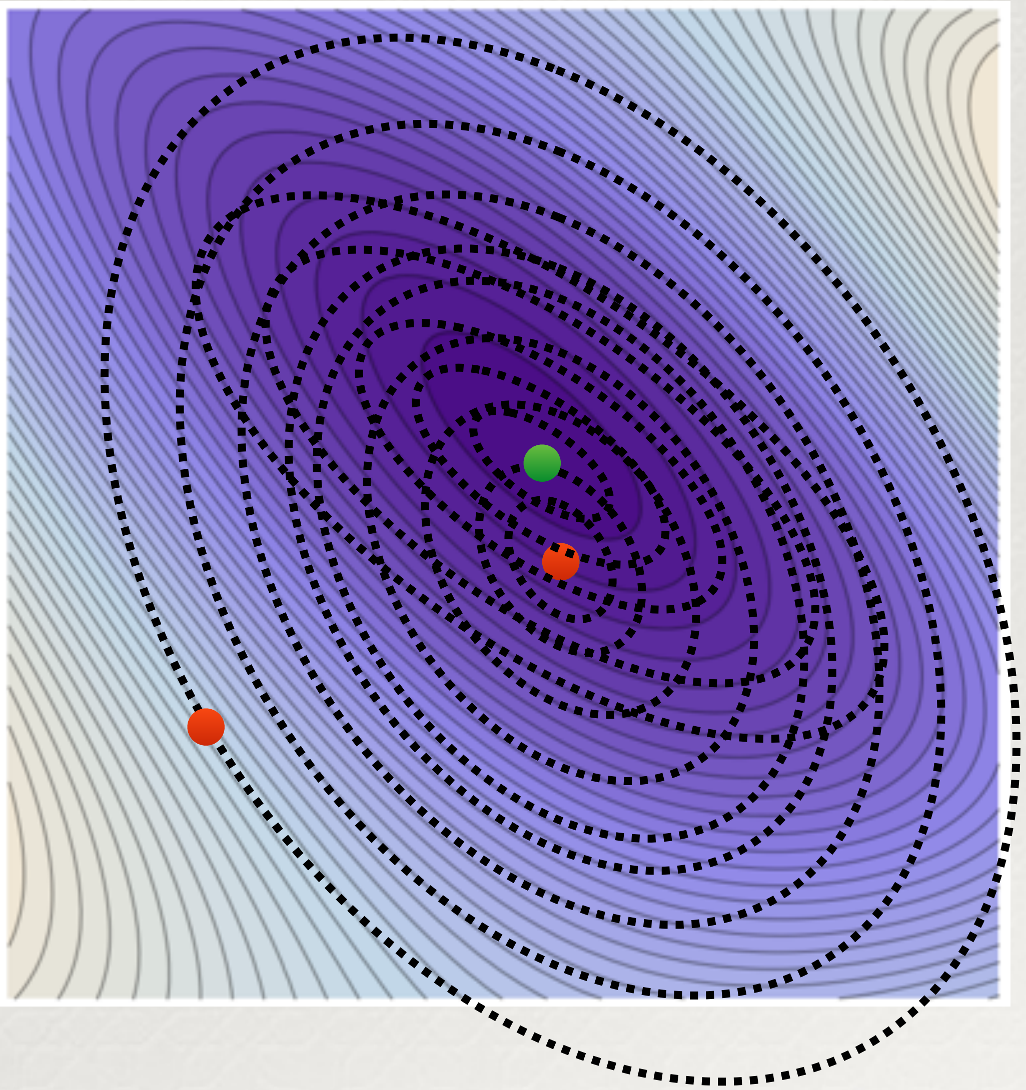
- *with Hesse matrix*

$$H_{ij} = \frac{dE}{d\mathbf{R}_i d\mathbf{R}_j}$$

- *Minimum:*

$$\Delta\mathbf{R} = \mathbf{H}^{-1}\mathbf{F}(\mathbf{R})$$

- *But H is unknown / expensive:*
- *cheaper construction of Hesse matrix needed*



Hessian construction „on the fly“

- *Make a guess for the Hessematrix*
  - *scaled unity* $\tilde{\mathbf{H}}^0 = \beta \mathbf{1}$
  - *force field model based on internal coordinates [1]* *(e.g. Lindh initialization)*

- *update H during optimization*
  - *Broyden-Fletcher-Goldfarb-Shanno update (BFGS)[2-5]*

$$\tilde{\mathbf{H}}^n = \tilde{\mathbf{H}}^{n-1} + \frac{\Delta\mathbf{F}\Delta\mathbf{F}^T}{\Delta\mathbf{F}^T\Delta\mathbf{R}} - \frac{(\tilde{\mathbf{H}}^{n-1}\Delta\mathbf{R})(\tilde{\mathbf{H}}^{n-1}\Delta\mathbf{R})^T}{\Delta\mathbf{R}^T\tilde{\mathbf{H}}^{n-1}\Delta\mathbf{R}}$$

  - *H stays positive definite*

[1] Lindh, Bernhardsson, Karlström, Malmqvist, Chem. Phys. Lett. 241, 423 (1995)
[2] Broyden, C. G. (1970), Journal of the Institute of Mathematics and Its Applications 6: 76–90
[3] Fletcher, R. (1970), Computer Journal 13 (3): 317–322
[4] Goldfarb, D. (1970), Mathematics of Computation 24 (109): 23–26
[5] Shanno, David F. (1970), Math. Comput. 24 (111): 647–656

Stability and Improvements

- *Hesse Matrix is an approximation*
- *Step control: Allow only finite step sizes to avoid overshooting*

$$\Delta \mathbf{R} = \alpha H^{-1} \mathbf{F}(\mathbf{R})$$

- *Trust radius method:*
    - *Evaluate quality of quadratic model*

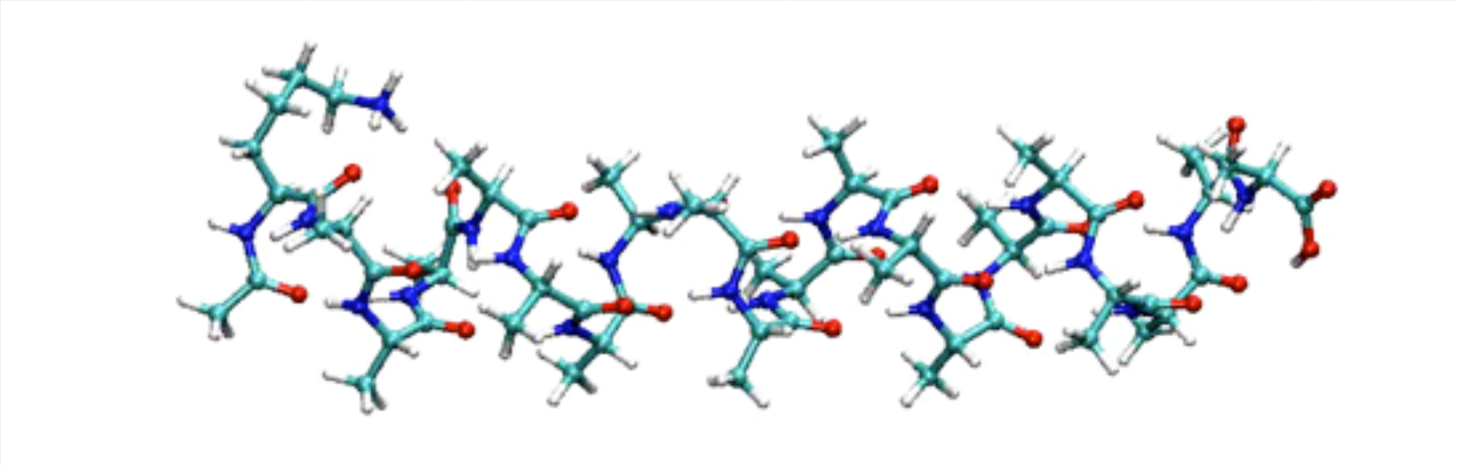
$$q = \frac{E_{\mathrm{true}}}{E_{\mathrm{model}}}$$

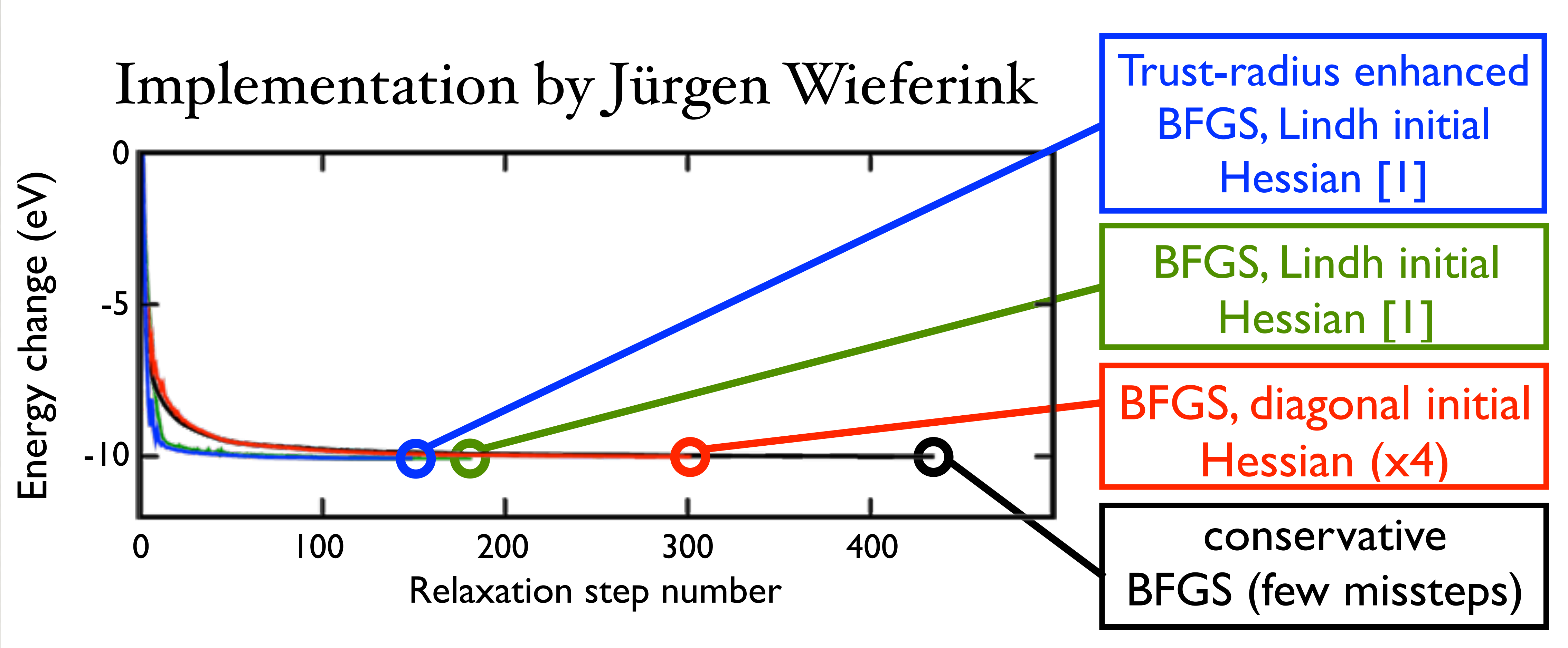
    - *Adjust maximal step size based on q*
- *Introduce internal coordinates like rotations rather the cartesian steps*

Performance

α-helical peptide
Ac-LysH$^+$-Ala$_{19}$, 220 atoms

initial guess:
MD snapshot



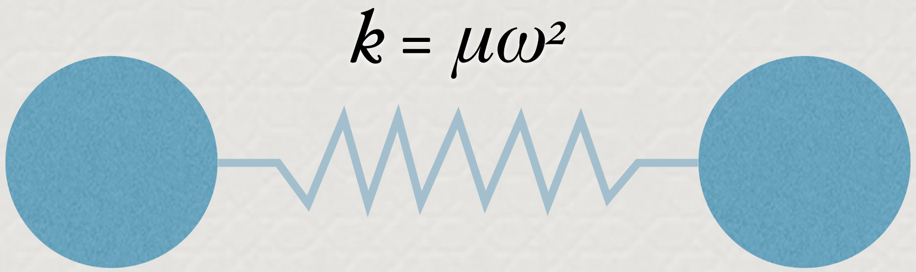


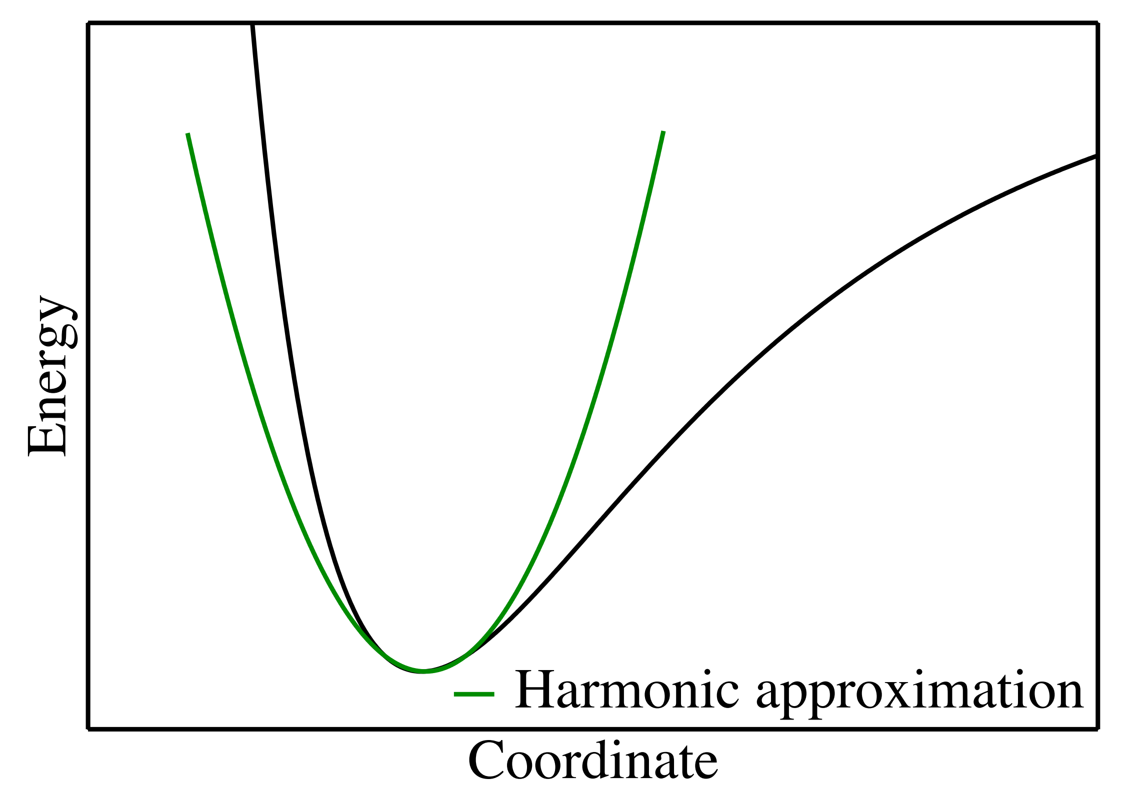
Implementation by Jürgen Wieferink

[1] Lindh, Bernhardsson, Karlström, Malmqvist, Chem. Phys. Lett. 241, 423 (1995)

Vibrations

- *Based on harmonic approximation (Harmonic oscillator)*
- *Perform Taylor expansion of energy up to second order*





$$E(r) = E_0 - F \cdot r + \frac{1}{2}\omega^2 r^2 + \mathcal{O}(r^3)$$

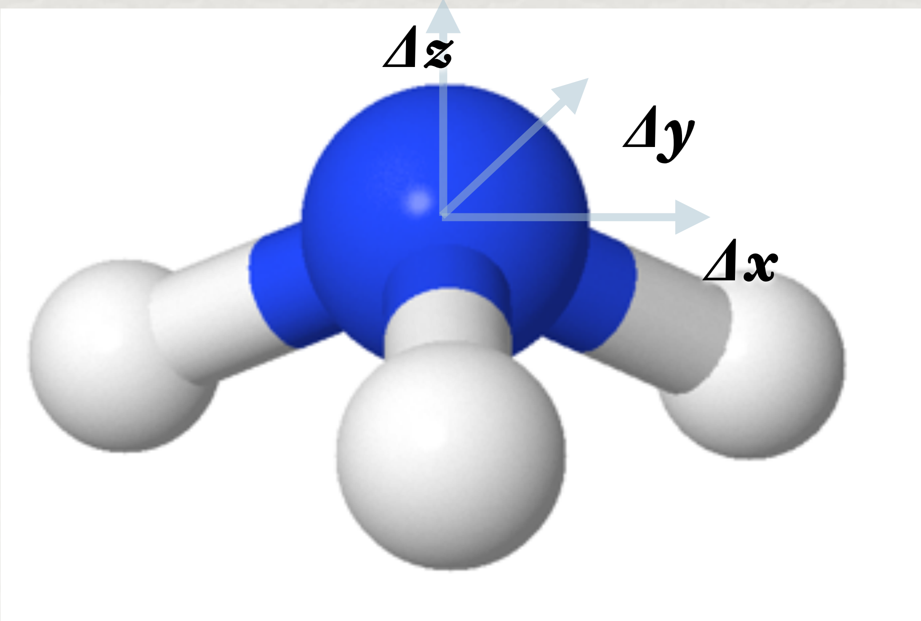
- *Forces vanish for optimized geometries:*

$$E(r) \approx E_0 + \frac{1}{2}\omega^2 r^2$$

Vibrations

- *Assume harmonic approximation in 3N space:*

$$E(\mathbf{R}) \approx E_0 + \frac{1}{2}\Delta\mathbf{R}^T\mathbf{H}\Delta\mathbf{R}$$



- *Hesse matrix H needed*

- *Numerical evaluation of Hessian via finite differences (BFGS Hessian not accurate enough)*

$$H_{\alpha\beta} = \frac{d^2E}{dR_\alpha dR_\beta} \approx -\frac{\mathbf{F}_\alpha(\mathbf{R}+h\mathbf{e}_\beta) - \mathbf{F}_\alpha(\mathbf{R}-h\mathbf{e}_\beta)}{2h}$$

Vibrations

- *Exponential Ansatz to solve Newtons equations*

$$\mathbf{M}\frac{d^2}{dt^2}\Delta\mathbf{R} = -\mathbf{H}\Delta\mathbf{R} \qquad \Delta\mathbf{R} = \mathbf{u}e^{-\mathrm{i}\omega t}$$

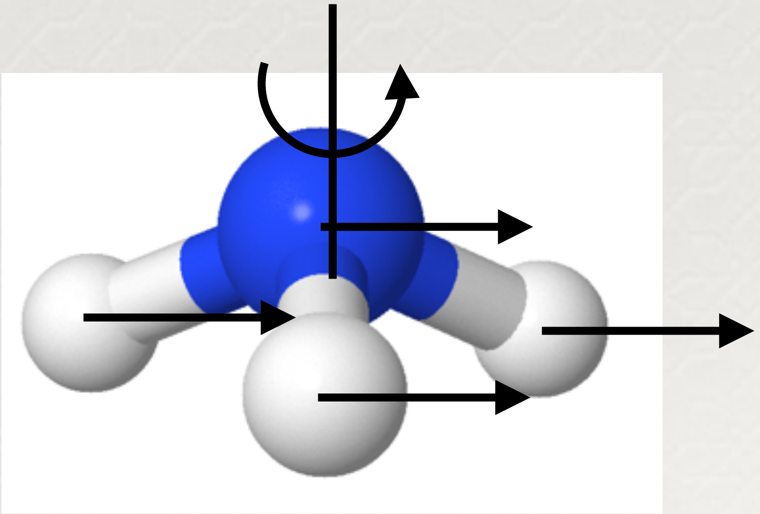
- *Generalized eigenvalue problem:*

$$\mathbf{H}\mathbf{u}_n = \omega_n^2\mathbf{M}\mathbf{u}_n$$

- *Solution gives eigenmodes and eigenfrequencies*

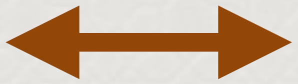
$$\mathbf{R} = \mathbf{R}_0 + \sum_n^{3N-6} c_n\mathbf{u}_n\cos\left(\omega_n t + \varphi_n\right)$$

- *3N - 6:*
  - *3 modes describe translations*
  - *3 modes describe rotations*

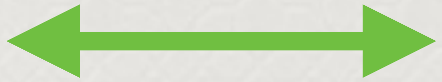


Interpreting Vibrations

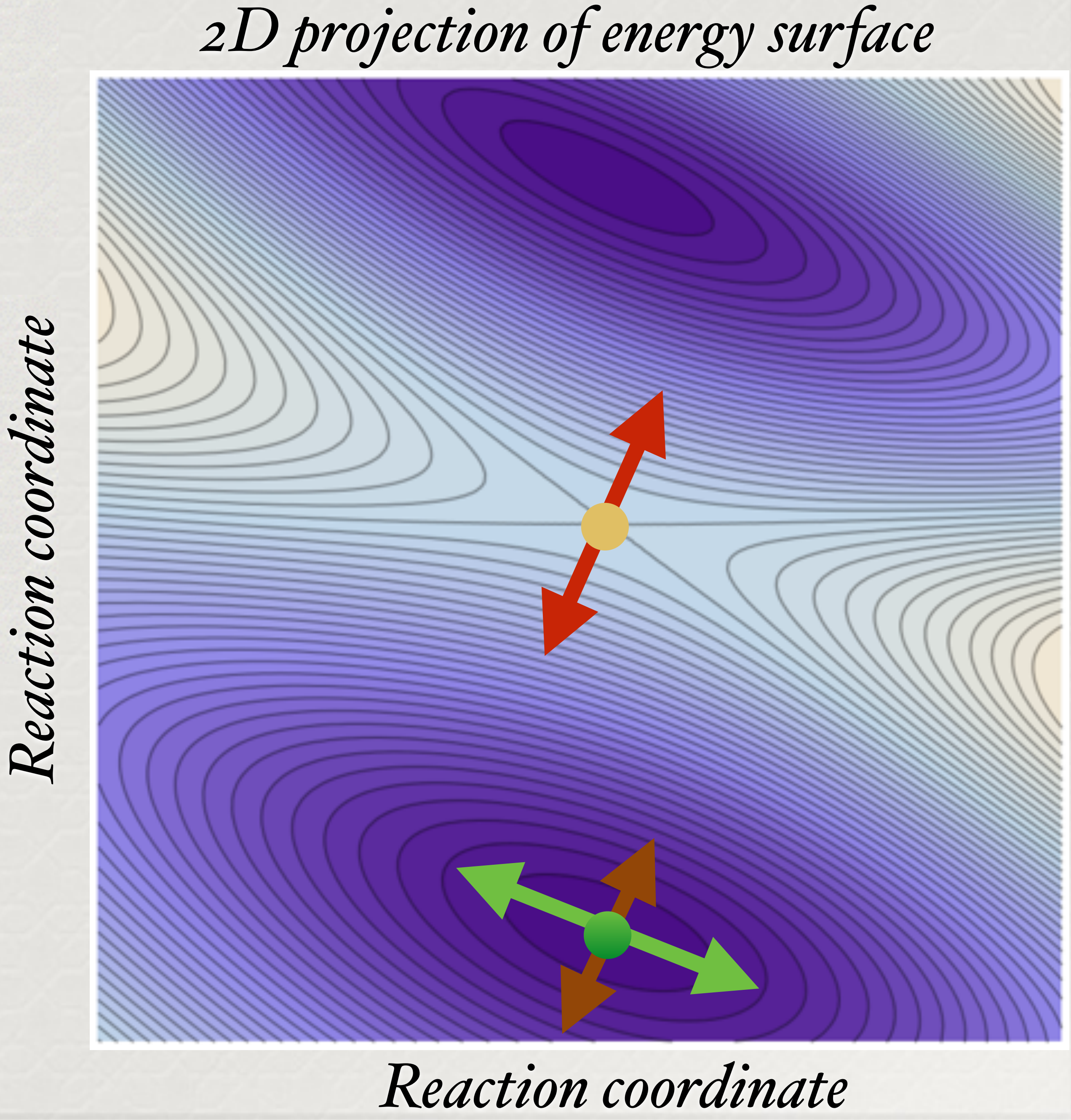
- *hard modes*
  high ω
  e.g. stretching



- *soft modes*
  low ω
  e.g wagging



- *imaginary modes*
  indicates sattle point,
  e.g. flat NH3



Beyond classical nuclei

- *Free energy for T> 0 with quantum vibrational effects [1]*

$$F(T) = E_0 + \sum_i \frac{\hbar\omega_i}{2} + k_b T \sum_i \ln\left(1 - e^{-\hbar\omega_i/k_b T}\right)$$

- *Zero-point vibrational energy correction*

$$F(0) = E_0 + \sum_i \frac{\hbar\omega_i}{2}$$

- *Connect to experimental data*
    - *Infrared spectroscopy*
    - *Raman spectroscopy*
- *More details (Tutorial: Christian Carbogno)*

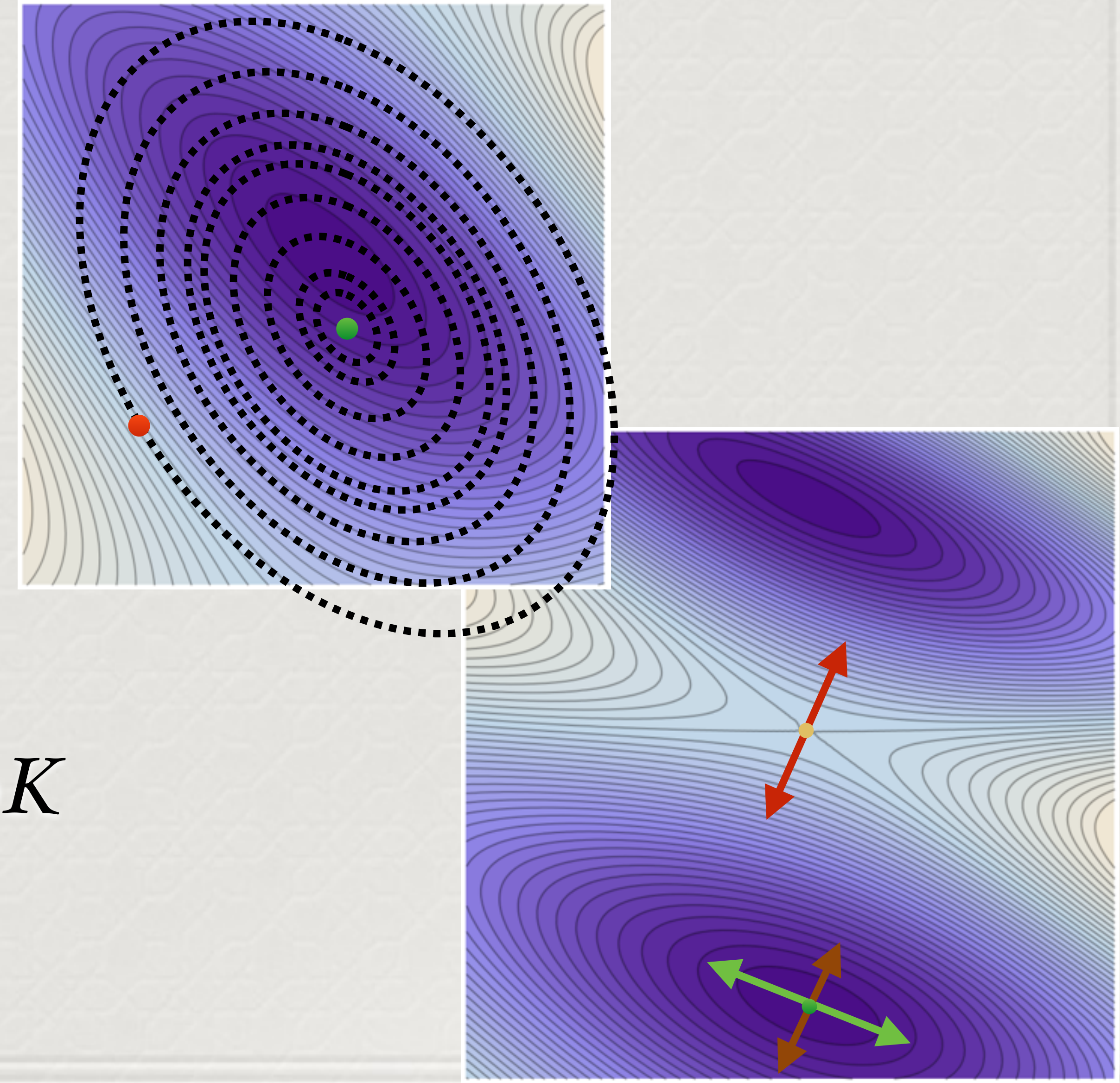
[1] F. Jensen, Introduction to Computational Chemistry

Summary Geometry Optimization

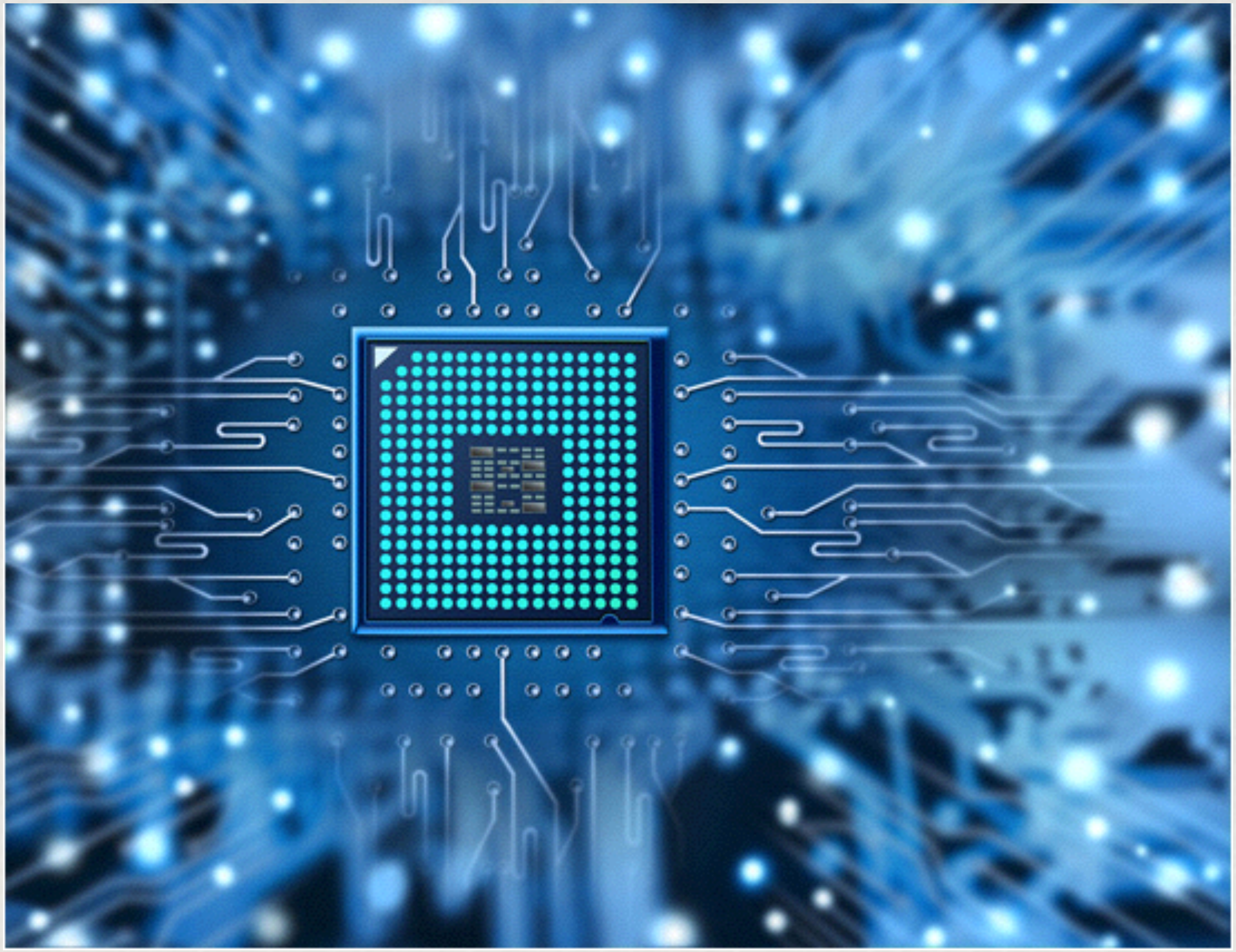
- *Forces in electronic structure theory*
  - *Hellman-Feynman and Pulay Forces*

$$\mathbf{F}_i = -\frac{\delta E_{\text{tot}}}{\delta \mathbf{R}_i}$$

- *Minimization techniques*
  - *Steepest descent*
  - *Quasi-Newton (BFGS)*
- *Atomic Vibrations*
  - *structure stability*
  - *zero-point energy*
  - *one first step to go beyond T=0K*



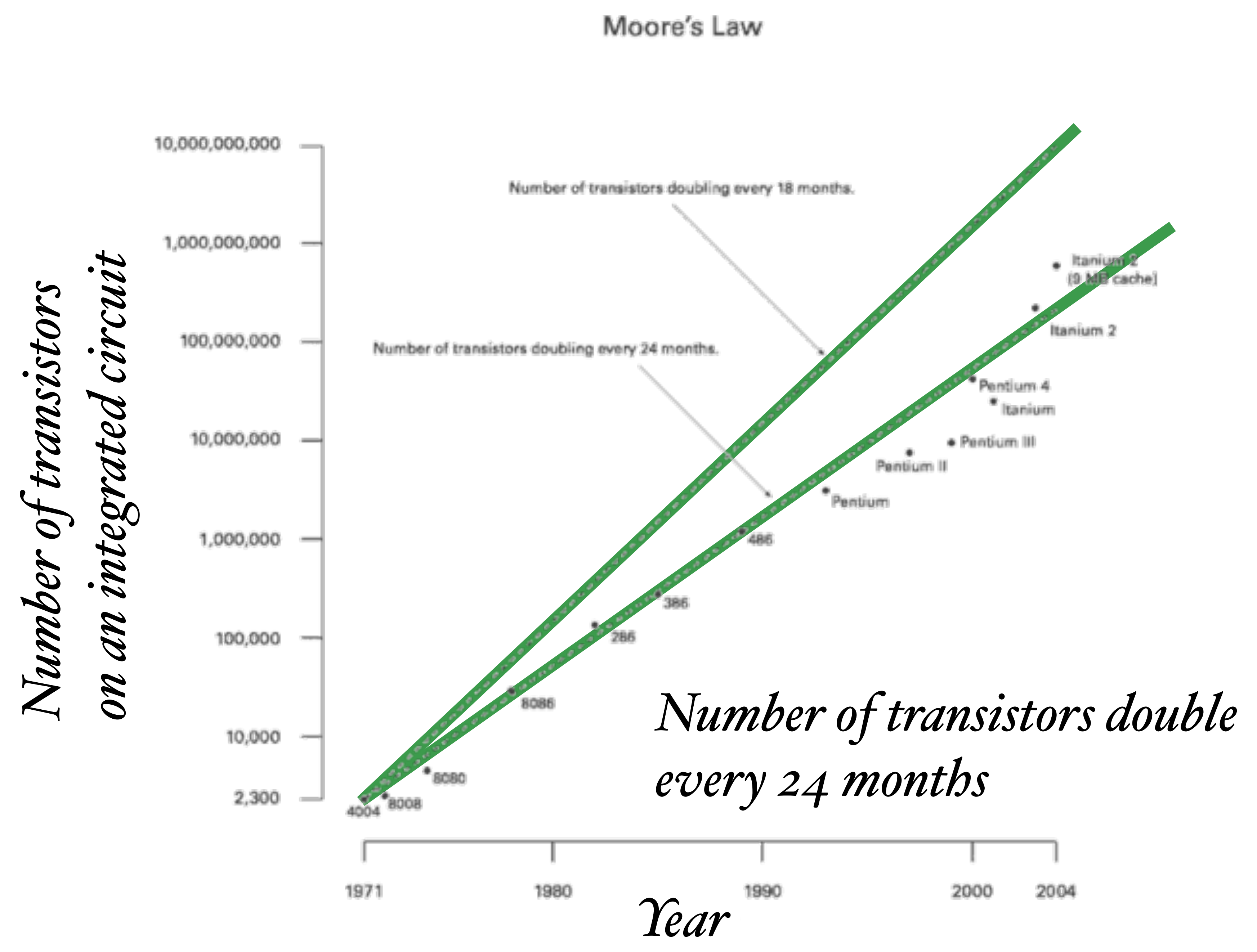
Beyond conventional CPUs



http://www.futuretimeline.net/21stcentury/2019.htm#.U8RgGagmAZs

Every 24 months ...

- *Moore's observation*



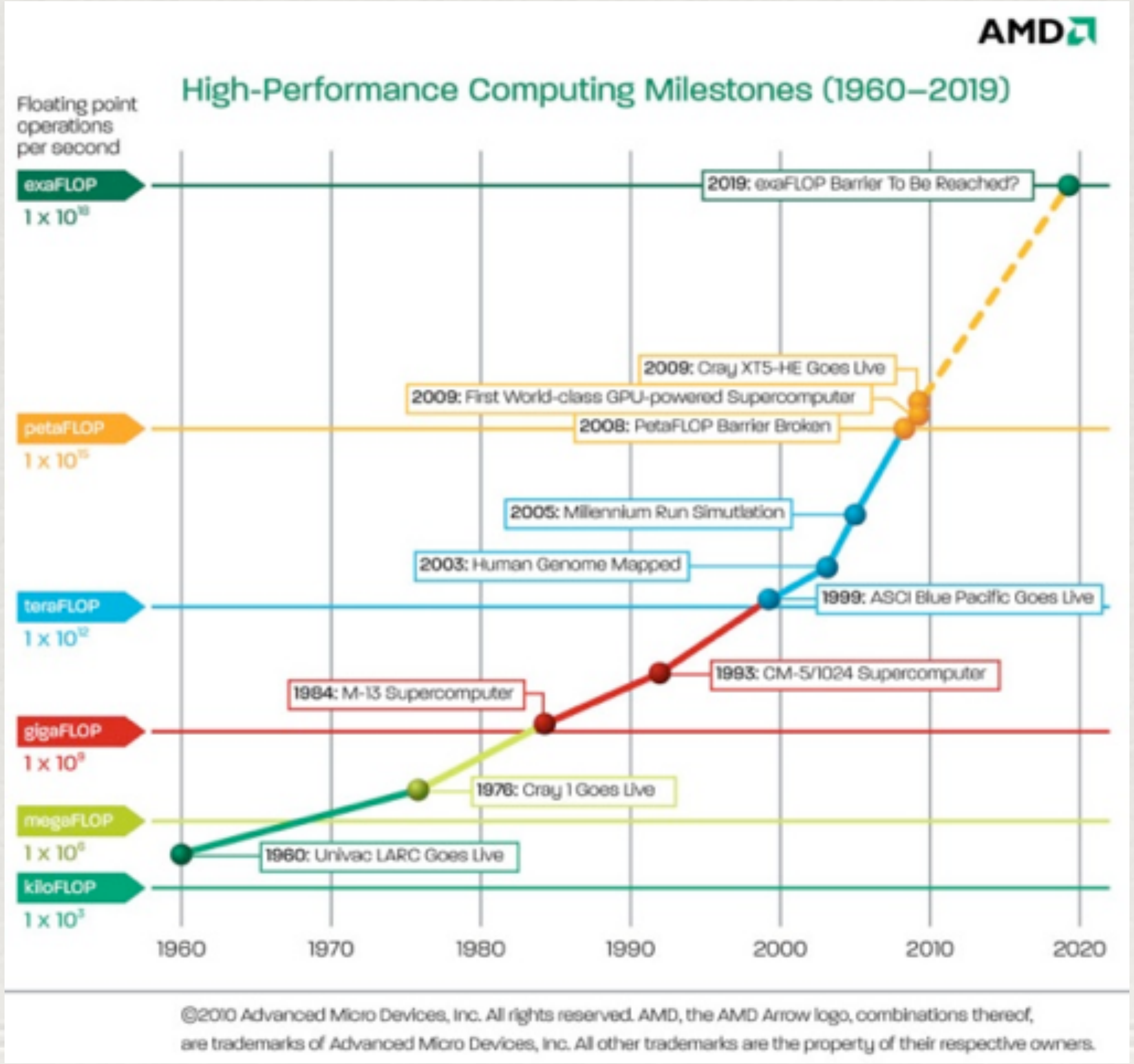
Number of transistors on an integrated circuit

Number of transistors double every 24 months

Year

Towards the ExaFLOP scale

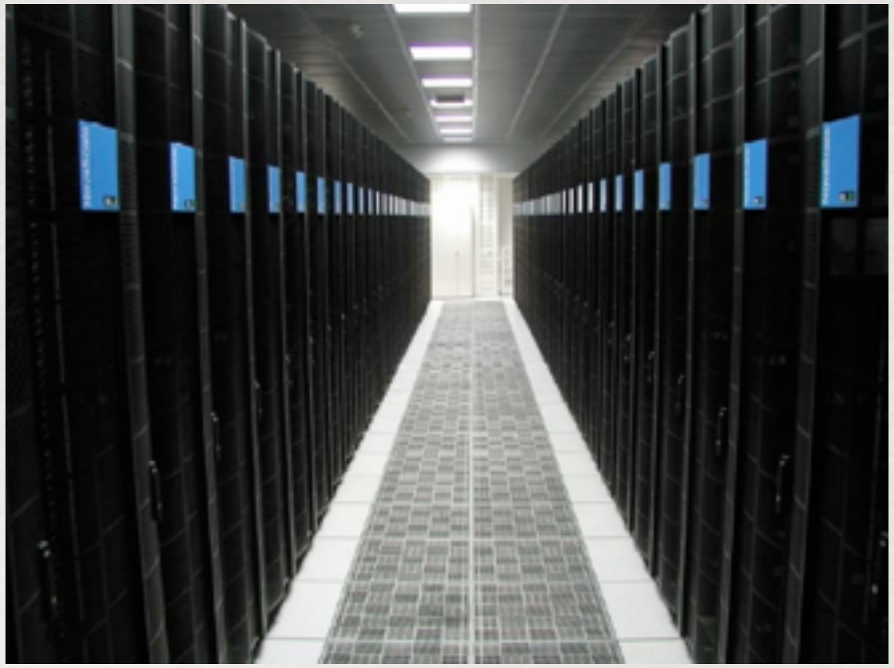
- *Moore's observation for **FL**oating-point **OP**erations per second*
- *Prediction: Overcome ExaFLOP barrier in 2018*



High-Performance Computing Milestones (1960–2019)

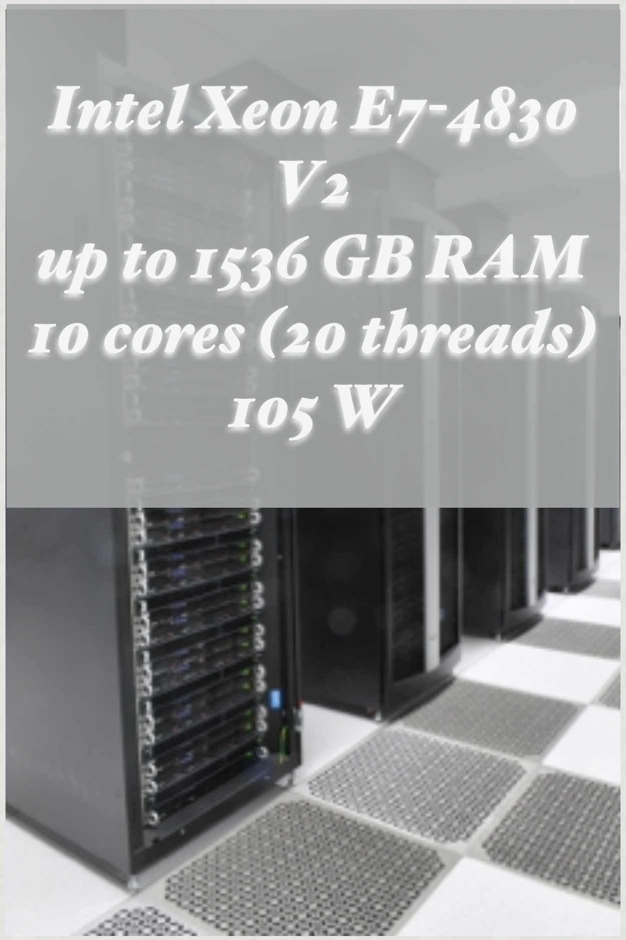
Are we ready for the ExaScale ?

- *Current ExaScale model predicts a machine with 1 billion cores*
- *If 1 billion cores would be typical CPUs then the predicted energy consumption would be more than 1 GW*

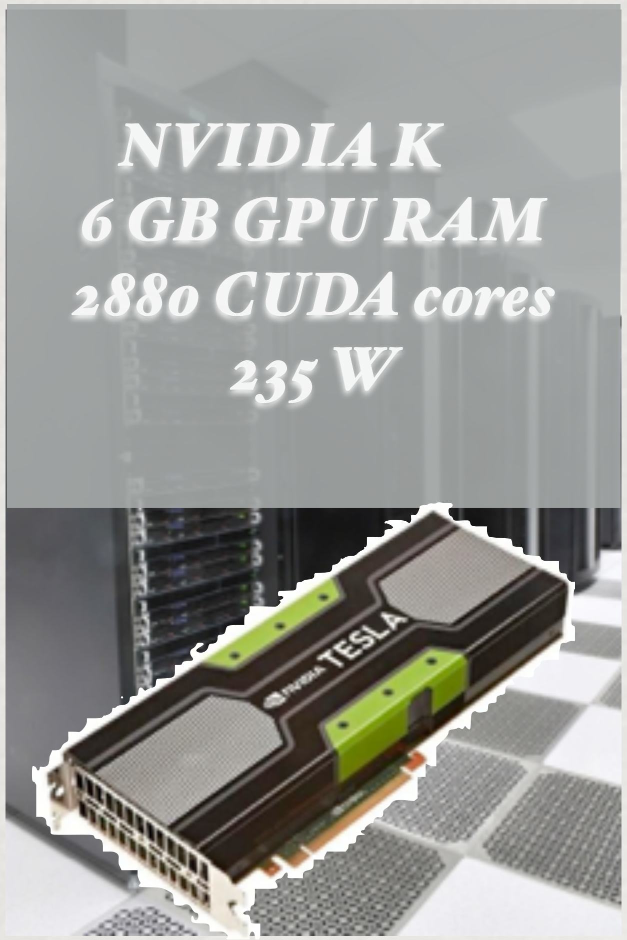


Modern Compute Architectures

- *New architectures: GPU computing*
- *Power consumption:*
  10.5 Watt / CPU core
  0.08 Watt / GPU core
  0.77 % Watt / core
- *GPUs are a possibility towards exaFLOP scaling*
- *… and are available*



Intel Xeon E7-4830 V2
up to 1536 GB RAM
10 cores (20 threads)
105 W



NVIDIA K
6 GB GPU RAM
2880 CUDA cores
235 W

The TOP 10 Supercomputers

43rd List: The TOP10

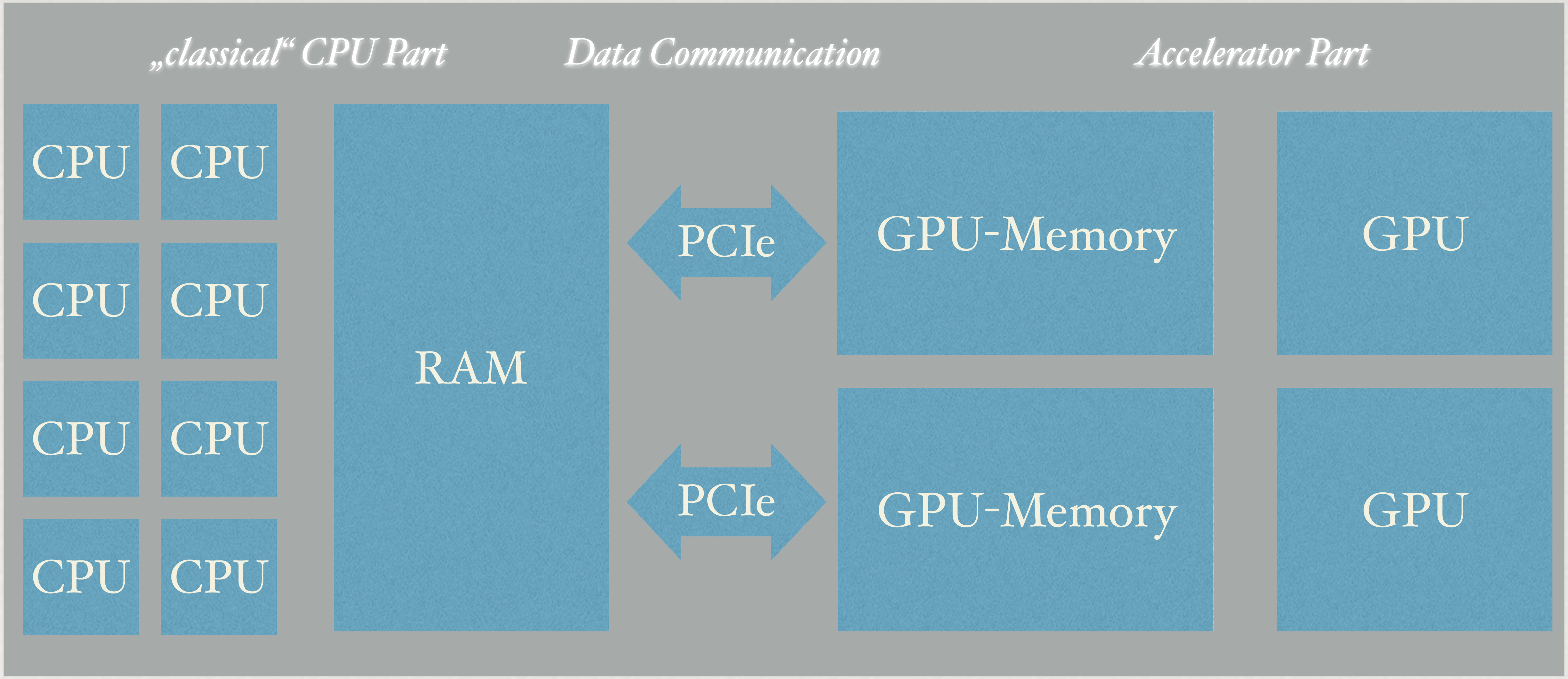
#	Site	Manufacturer	Computer	Country	Cores	Rmax [Pflops]	Power [MW]
1	National University of Defense Technology	NUDT	Tianhe-2 NUDT TH-IVB-FEP, Xeon 12C 2.2GHz, IntelXeon Phi	China	3,120,000	33.9	17.8
2	Oak Ridge National Laboratory	Cray	Titan Cray XK7, Opteron 16C 2.2GHz, Gemini, NVIDIA K20x	USA	560,640	17.6	8.21
3	Lawrence Livermore National Laboratory	IBM	Sequoia BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	1,572,864	17.2	7.89
4	RIKEN Advanced Institute for Computational Science	Fujitsu	K Computer SPARC64 VIIIfx 2.0GHz, Tofu Interconnect	Japan	795,024	10.5	12.7
5	Argonne National Laboratory	IBM	Mira BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	786,432	8.59	3.95
6	Swiss National Supercomputing Centre (CSCS)	Cray	Piz Daint Cray XC30, Xeon E5 8C 2.6GHz, Aries, NVIDIA K20x	Switzerland	115,984	6.27	2.33
7	Texas Advanced Computing Center UT	Dell	Stampede PowerEdge C8220, Xeon E5 8C 2.7GHz, Intel Xeon Phi	USA	462,462	5.17	4.51
8	Forschungszentrum Juelich (FZJ)	IBM	JuQUEEN BlueGene/Q, Power BQC 16C 1.6GHz, Custom	Germany	458,752	5.01	2.30
9	Lawrence Livermore National Laboratory	IBM	Vulcan BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	393,216	4.29	1.97
10	Government	Cray	Cray XC30, Xeon E5 12C 2.7GHz, Aries	USA	225,984	3.14	

1) Tianhe-2 (Heavenrider-2) China
32,000 Intel Xeon E5 12 Cores
48,000 Intel Xeon Phi
33.86 PFLOPS

2) Titan USA
18,688 AMD Opteron 6274 16 Cores
18,688 Nvidia Tesla K20X GPUs
17.59 PFLOPS (27 theoretical)

6) Piz Daint Switzerland
5272 Intel Xeon E5 8Core
5272 Nvidia Tesla K20X GPUs
7.78 PFLOPS

Modern Computer Architectures



„classical" CPU Part

Data Communication

Accelerator Part

CPU CPU
CPU CPU
CPU CPU
CPU CPU

RAM

PCIe

PCIe

GPU-Memory

GPU-Memory

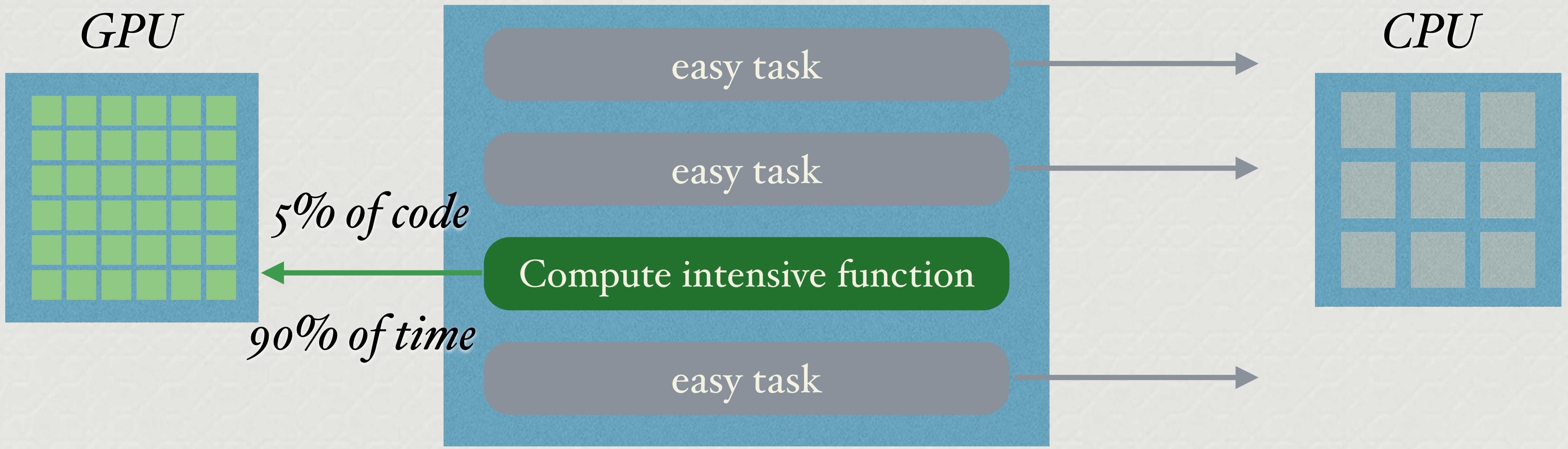
GPU

GPU

- *distinct memory for CPUs and GPUs*
- *data has to be copied using (slow) PCIe bus*
- *GPU memory (4-12 GB / GPU) < CPU RAM (2-4 GB / CPU)*

Program structure

- *typically suggested way:*



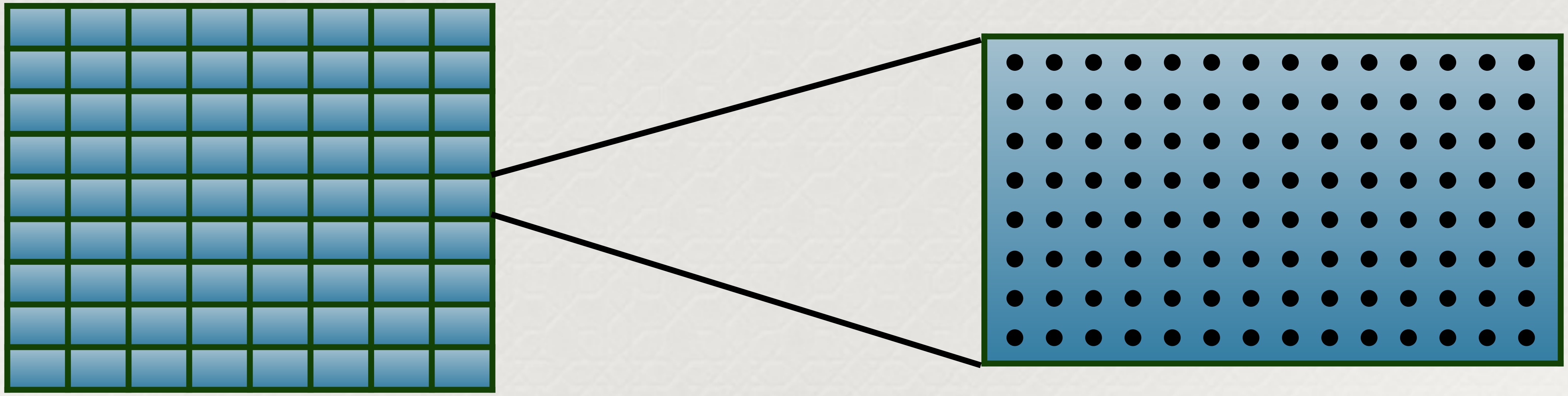
- *What about the the CPU cores, while the GPU is working*
  - *CPUs should be working as well*
  - *One GPU „beating" 10 - 20 CPU cores is a much harder task*

Example: Construct Hamiltonian matrix

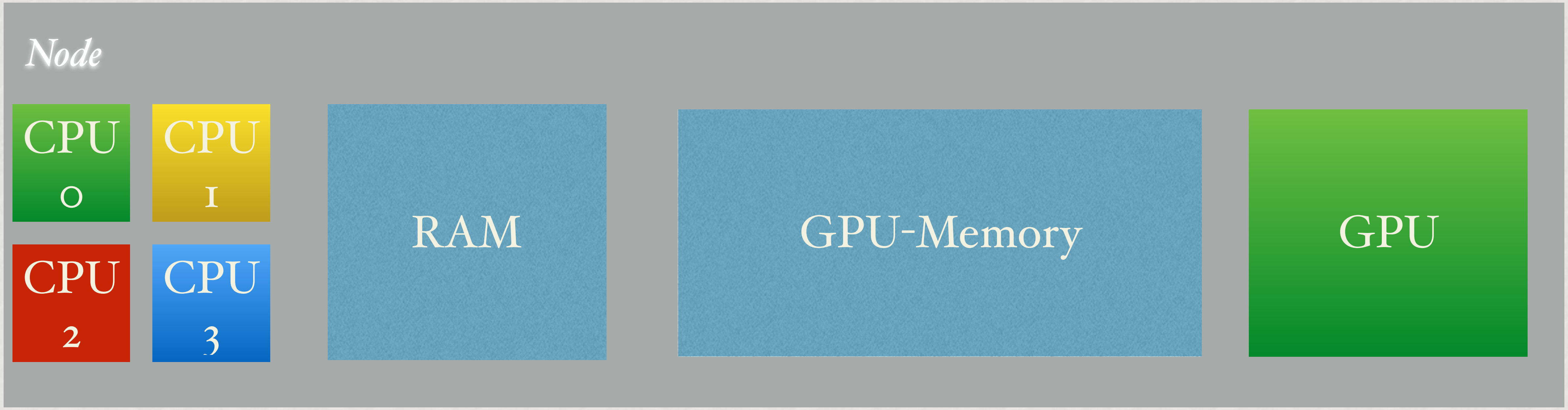
- *The Hamilton matrix construction is a compute intense operation*

$$H_{ij} = \int\limits_{\Omega} \chi_i^*(\mathbf{r})\hat{H}\chi_j(\mathbf{r})d\mathbf{r}$$

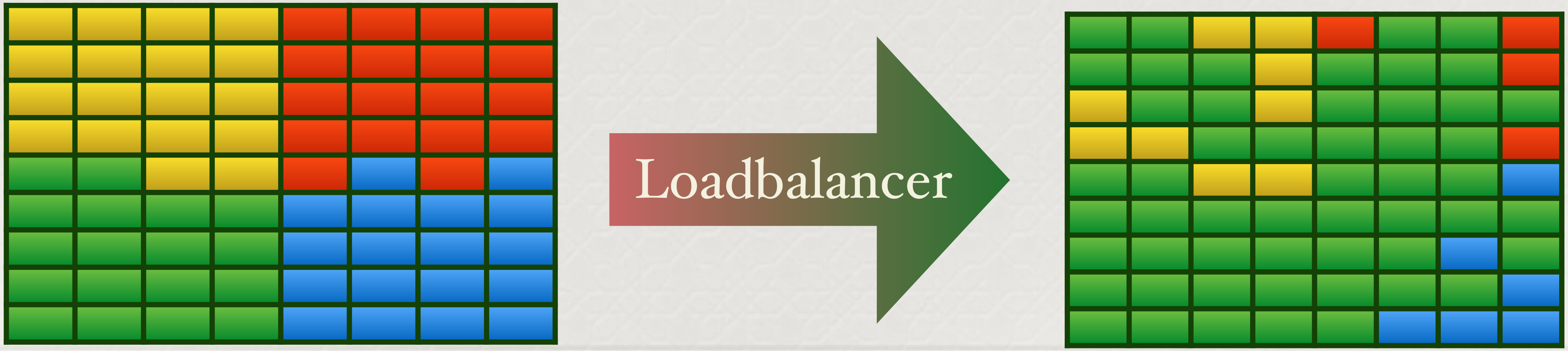
- *Operation is performed point-wise in real space*
- *Operation is organised in batches = collection of points in real space*



Parallelization + Loadbalancing



e.g. space integration



Load Balancing Hamiltonian

- *Ongoing work: Example C (diamond) - Gamma point only 8000 basis functions, 512 batches a 70 points*

- *Speedup:*

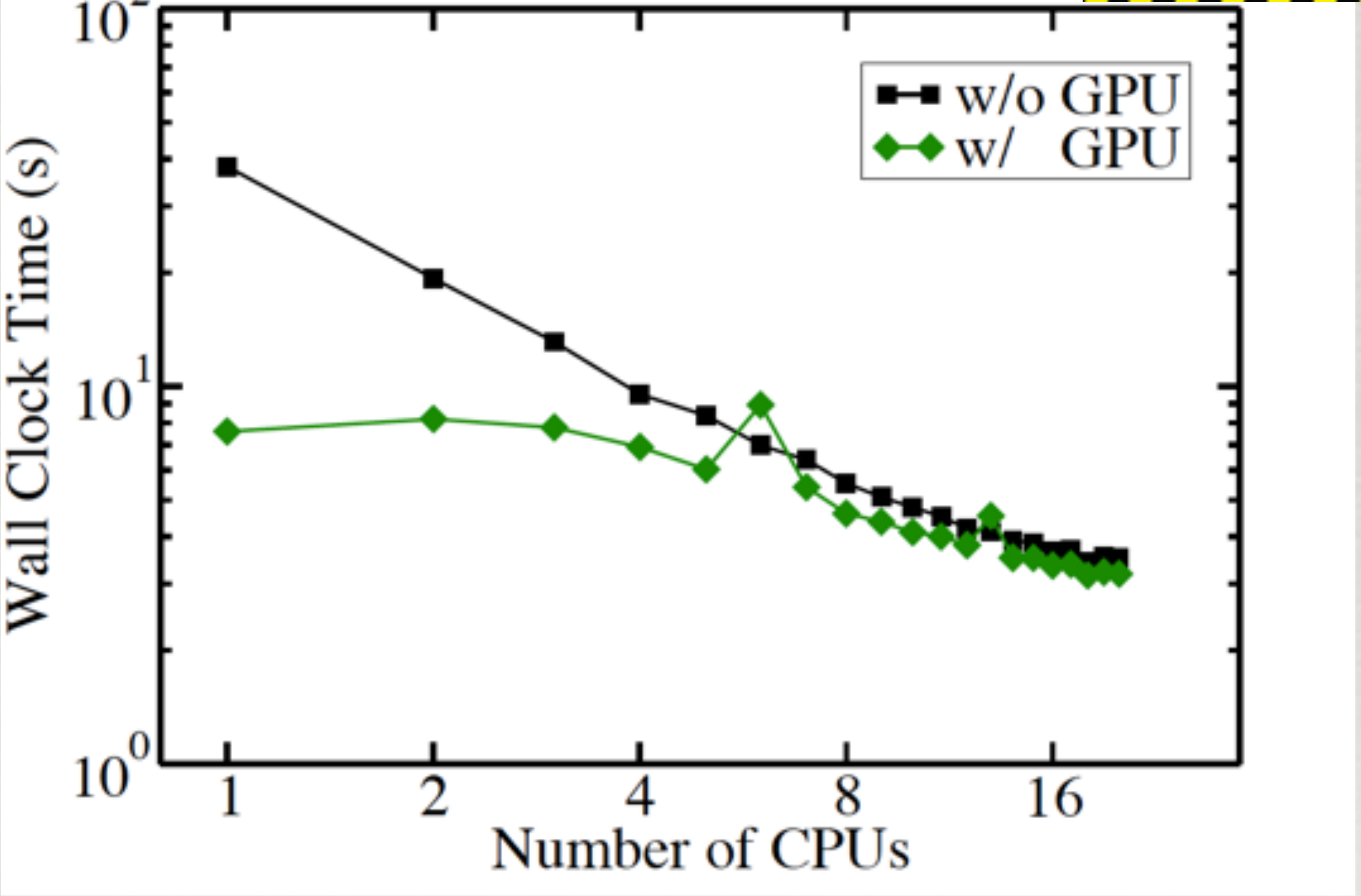
$$S = \frac{t_{\mathrm{CPU}}}{t_{\mathrm{GPU}}}$$

- *Speedup on entire node*

$$S_{\mathrm{node}} = \frac{S}{N_{\mathrm{CPU}}} + \frac{N_{\mathrm{CPU}} - 1}{N_{\mathrm{CPU}}}$$

Load Balancing

Integrate Hamiltonian



- *We have to aim for a Speedup of 21 to achieve a node speedup of 2.*

Summary

- *CPUs cannot be the way for ExaFLOP scale (Energy consumption)*

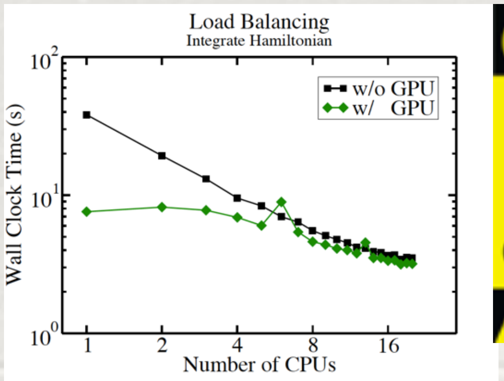
- *Accelerator cards such as GPUs are a promising architecture*

43rd List: The TOP10

#	Site	Manufacturer	Computer	Country	Cores	Rmax [Pflops]	Power [MW]
1	National University of Defense Technology	NUDT	Tianhe-2 NUDT TH-IVB-FEP, Xeon 12C 2.2GHz, IntelXeon Phi	China	3,120,000	33.9	17.8
2	Oak Ridge National Laboratory	Cray	Titan Cray XK7, Opteron 16C 2.2GHz, Gemini, NVIDIA K20x	USA	560,640	17.6	8.21
3	Lawrence Livermore National Laboratory	IBM	Sequoia BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	1,572,864	17.2	7.89
4	RIKEN Advanced Institute for Computational Science	Fujitsu	K Computer SPARC64 VIIIfx 2.0GHz, Tofu Interconnect	Japan	795,024	10.5	12.7
5	Argonne National Laboratory	IBM	Mira BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	786,432	8.59	3.95
6	Swiss National Supercomputing Centre (CSCS)	Cray	Piz Daint Cray XC30, Xeon E5 8C 2.6GHz, Aries, NVIDIA K20x	Switzer-land	115,984	6.27	2.33
7	Texas Advanced Computing Center/UT	Dell	Stampede PowerEdge C8220, Xeon E5 8C 2.7GHz, Intel Xeon Phi	USA	462,462	5.17	4.51
8	Forschungszentrum Juelich (FZJ)	IBM	JuQUEEN BlueGene/Q, Power BQC 16C 1.6GHz, Custom	Germany	458,752	5.01	2.30
9	Lawrence Livermore National Laboratory	IBM	Vulcan BlueGene/Q, Power BQC 16C 1.6GHz, Custom	USA	393,216	4.29	1.97
10	Government	Cray	Cray XC30, Xeon E5 12C 2.7GHz, Aries	USA	225,984	3.14	

- *Accelerator cards are available on compute facilities*

- *GPUs and CPUs have to work together (Load balancing)*



- *Ongoing work*

Thank you !

- *Part I: The Self Consistency Cycle (SCF)*
  - *Density mixing*
  - *Metallic systems: Fermi level broadening*
  - *Preconditioning*
- *Part II: Structure optimization*
  - *Local structure optimization*
  - *„Forces“ in density functional theory*
  - *Vibrations in the harmonic approximation*
- *Part III: Beyond conventional CPUs*
  - *Calculating with Graphical Processing Units (GPUs)*