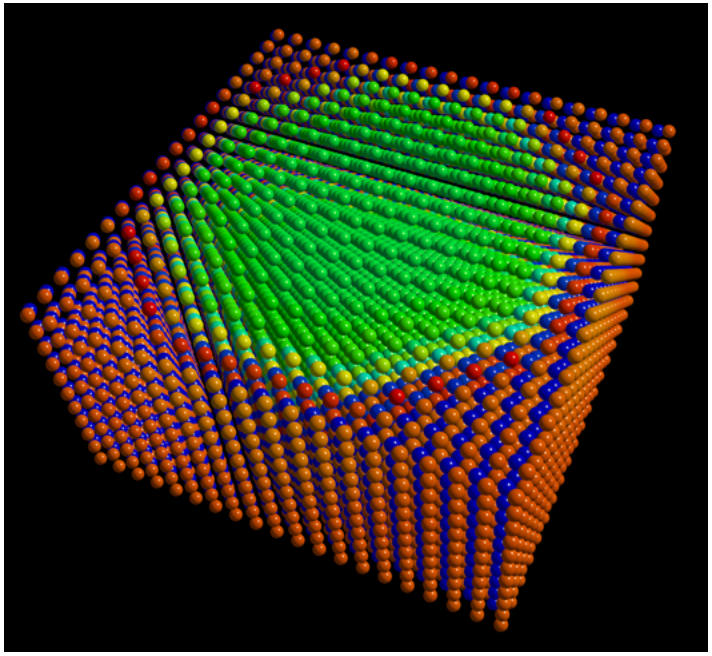
KKR and massively parallel computing

Rudolf Zeller

Institute for Advanced Simulation
Forschungszentrum Jülich, D-52425 Jülich, Germany

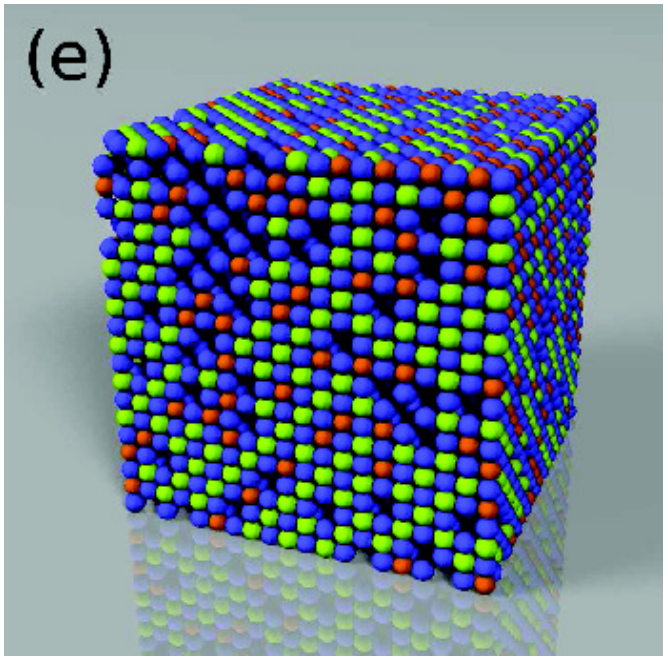
KKR for large systems



16000 atoms

Fe nanoparticle in FeAl matrix
LSMS method (Oak Ridge)

from http://www.psc.edu/science/2006/nano.html



8000 atoms

disordered $GeSb_2Te_4$ alloy
KKRnano (Jülich)

from A. Thiess, thesis (2011)

Computational bottleneck

work in standard methods scales as $O(N^3)$
for N occupied orbitals (or electrons, atoms, ...)
because of eigenvalue problem and orthogonalization

today petascale computing

Titan (Oak Ridge)

Blue Gene/Q (Jülich)

future? exascale computing

increase of computing power by a factor 1000
increase of number of atoms by a factor 10
increase of length scale by a factor $\sqrt[3]{10} \approx 2$

better scaling needed $\Rightarrow$ $O(N^2)$ or $O(N)$ methods

Supercomputing complexity

- old well tested codes are not well suited for modern supercomputing
- development of new codes for supercomputing takes many years
- this dilemma concerns codes as well as the underlying ideas

advantage of KKR

- main work consists in solving linear equations

JUQUEEN: 28 racks (458,752 cores)

Rack: 32 nodeboards (16,384 cores)

Nodeboard: 32 compute nodes

Node: 16 cores with 16 GB memory per node

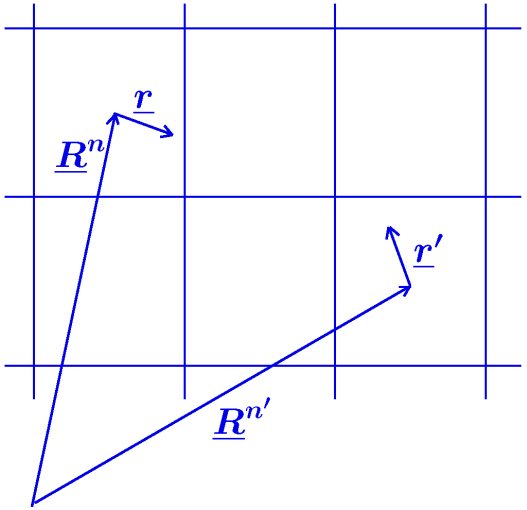
Core: 16-way SMP processor

Maximal parallelisation: 1835008 MPI tasks

Power: on average 1.9 MW in 2012
Peak performance: 5.872 Petaflops
Linpack performance: 5.009 Petaflops

Structure of the KKR Green function equations

$$G(\underline{r}+\underline{R}^n,\underline{r}'+\underline{R}^{n'}) = \delta_{nn'}G_s^n(\underline{r},\underline{r}') + \sum_{LL'} R_L^n(\underline{r})G_{LL'}^{nn'}R_{L'}^{n'}(\underline{r}')$$



- divide space into cells n
- solve single-cell problems

$$G_s^n(\underline{r},\underline{r}') = G^0(\underline{r},\underline{r}') + \int_n G^0(\underline{r},\underline{r}'')V(\underline{r}')G_s^n(\underline{r}'',\underline{r}')\mathrm{d}\underline{r}''$$

$$R_L^n(\underline{r}) = J_L(\underline{r}) + \int_n G^0(\underline{r},\underline{r}')V(\underline{r}'')R_L^n(\underline{r}')\mathrm{d}\underline{r}'$$

- use matrix equation

$$G_{LL'}^{nn'} = G_{LL'}^{r,nn'} + \sum_{n''L''L'''} G_{LL''}^{r,nn''}\Delta t_{L''L'''}^{n''}G_{L'''L'}^{n''n'}$$

$$\Delta t_{LL'}^n = \int_n R_L^{r,n}(\underline{r})\Delta V(\underline{r})R_{L'}^n(\underline{r})\mathrm{d}\underline{r}$$

a single cutoff parameter $l_{\max}$ determines accuracy and matrix size

single-cell problems can be solved in parallel with $O(N)$ work

matrix equation is independent of the radial resolution used

Linear scaling

physics and chemistry both tell us that properties of materials are local

Locality principle in wave mechanics

W. KOHN AND A. YANIV

Proc. Natl. Acad. Sci. USA
Vol. 75, No. 11, pp. 5270–5272, November 1978
Physics

Nearsightedness of electronic matter

E. Prodan[†‡§] and W. Kohn[†]

PNAS | **August 16, 2005** | vol. 102 | no. 33 | **11635–11638**

nearsightedness principle Kohn PRL 1996
in systems without long range electric fields (and for fixed chemical potential) the density change at a point in space is negligibly affected, if the electronic potential is changed sufficiently far away from this point

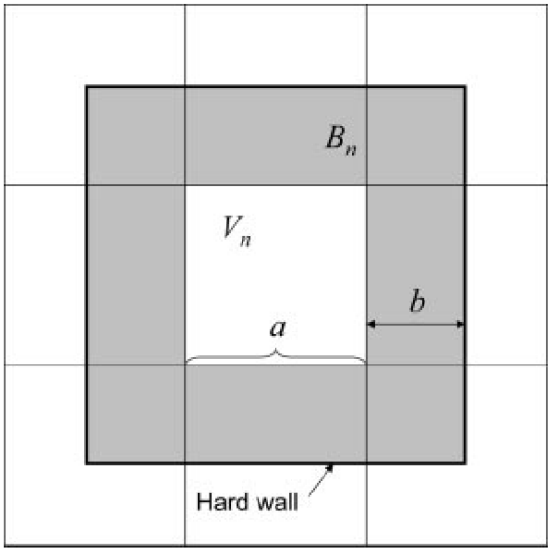
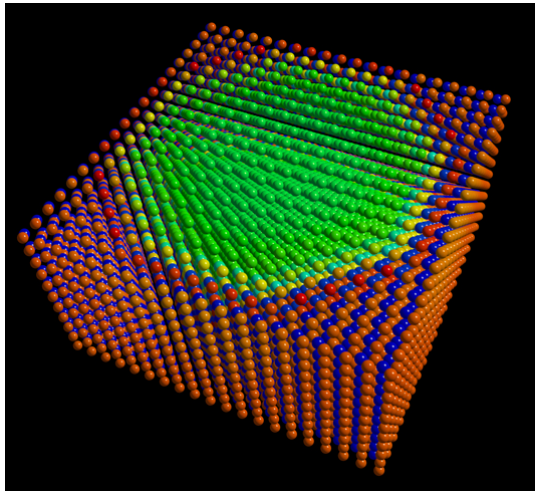


Fig. 3. The system is divided into smaller volumes V_n (nine in this example), with buffer zones B_n (gray).

Divide and conquer techniques

- original DC method: Yang PRL 1991
- charge patching method: Wang PRL 2002
- LSMS method: Wang et al. PRL 1995
- LSGF method: Abrikosov et al. PRL 1996
- and others

an example:



16000 atoms

charge distribution of
Fe nanoparticle in FeAl matrix
LSMS method

Stocks, Wang et al.

from http://www.psc.edu/science/2006/nano.html

General basis for linear scaling

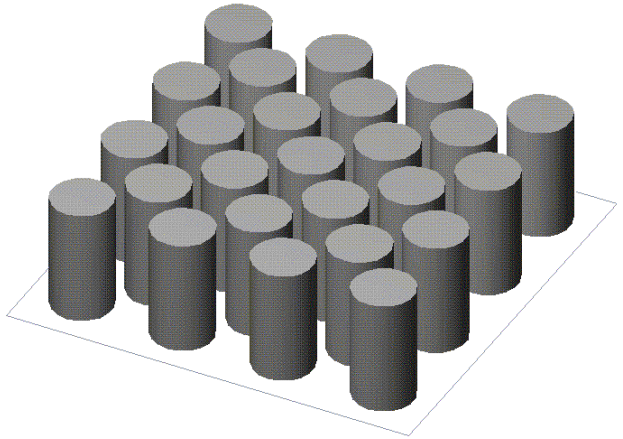
- reduce computing effort by tolerating small loss of precision
- exploit locality of the potential and quasi-locality of $\nabla^2_{\underline{r}}$
  $\Rightarrow$ sparse matrix computations
- apply iterative solution for sparse matrix equations
- use nearsightedness by neglecting potential changes for away
- use supercomputing with massive parallelization

implementation in KKR?

Concept of a repulsive reference system

infinite array of repulsive potentials $\Rightarrow$ a finite energy E_0 exists such that

- **reference system has no eigenstates below E_0**
- **relevant energies E in DFT satisfy $E < E_0$**
- **reference Green function decays exponentially for $E < E_0$**
- **neglect of exponentially small elements $\Rightarrow$ sparse matrices**

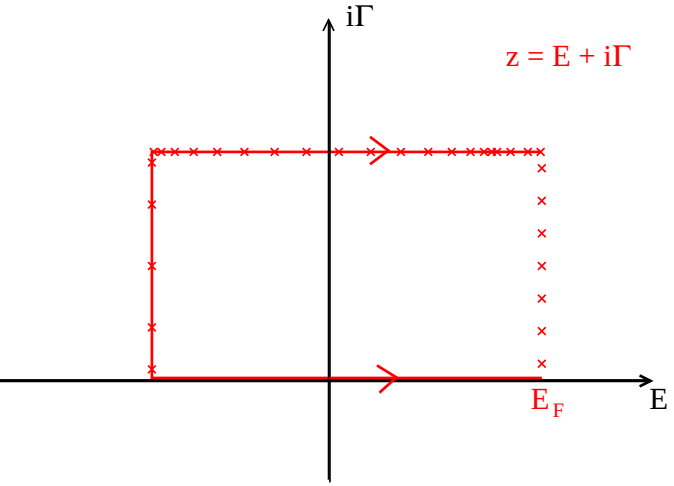


- **real space calculation of structure constants**
- **clusters of about 50 atoms are sufficient**
- **decay is property of the reference system**

  Zeller et al. PRB 1995

Iterative solution

- the Green function matrices are complex and non-Hermitian
- the Green function as function of energy has singularities for real E due to atomic core states and valence and conduction band states



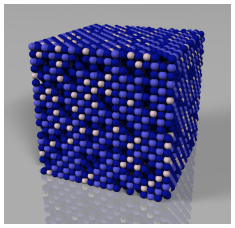
complex energy integration Zeller et al. SSC 1982

$$n(\underline{r}) = -\frac{1}{\pi}\mathrm{Im}\int_{-\infty}^{\infty} f(E,T)G(\underline{r},\underline{r},E)\mathrm{d}E$$

- finite temperature DFT Mermin PR 1965, Wildberger et al. PRB 1995
- straightforward iterations diverge $\boxed{G^{(i+1)} = G^r + G^r \Delta t G^{(i)}}$
  quasi-minimal-residual (QMR) method works
- highly parallelizable, no loss of precision $\Rightarrow O(N^2)$ method

KKRnano

- KKRnano is a new code (presently implemented in supercell mode)
- why nano?



nanosystems contain many atoms
(8000 in a cube of 6 nm length)

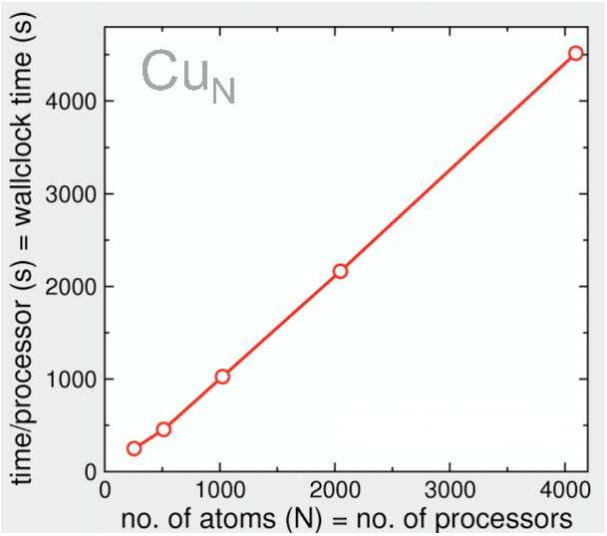
- work with A. Thiess, E. Rabel, M. Bolten, P. H. Dederichs, S. Blügel

accuracy (in meV)

	Cu		Pd	
$\|\|r\|\|$	ΔE_{tot}	N_{it}	ΔE_{tot}	N_{it}
10^{-3}	5.3740	403	2.3790	234
10^{-4}	0.3456	528	0.4179	315
10^{-5}	0.0055	670	0.0167	397
10^{-6}	0.0003	814	0.0015	463

any desired accuracy can be achieved
computing time $O(N_{\mathrm{it}} N_{\mathrm{cl}} N^2)$
efficient parallelization is possible

scaling behavior

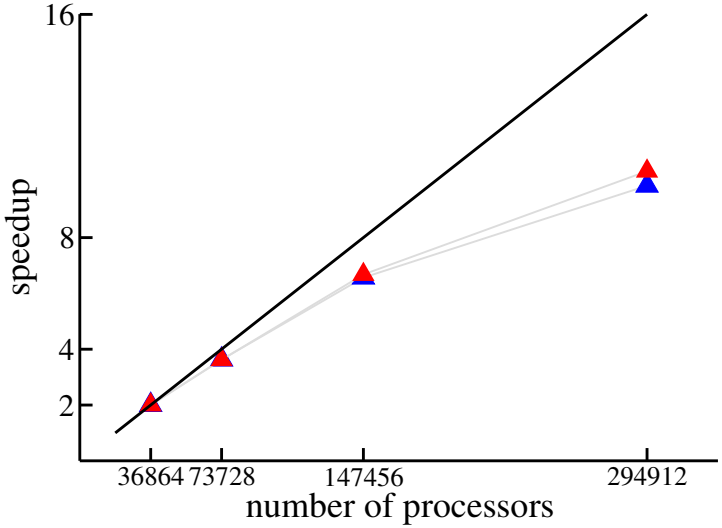


Parallelization strategy

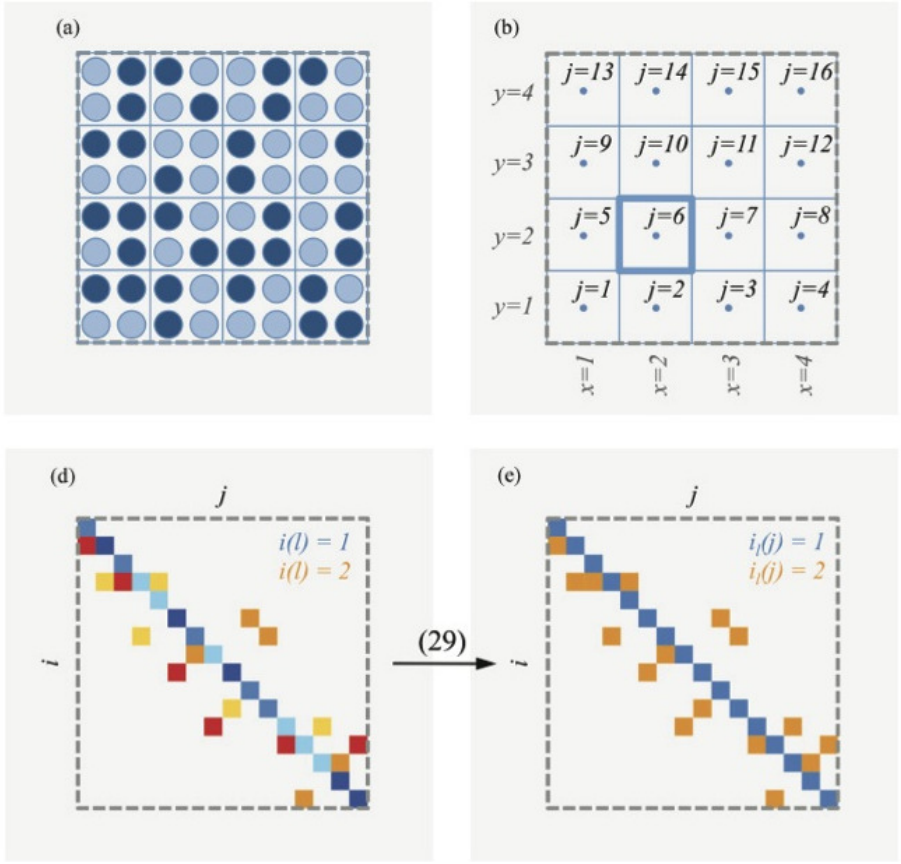
KKRnano uses four levels of parallelization with MPI groups and communicators and point-to-point and collective messages

- parallelization over atoms (is efficient)
- parallelization over two spin directions (is trivial and efficient)
- parallelization over energy points (2 or 3 panels dynamically load balanced)
- parallelization over L components (until now only in matrix equation)
- optionally OpenMP threads instead of L parallelization

test system: NiPd3071



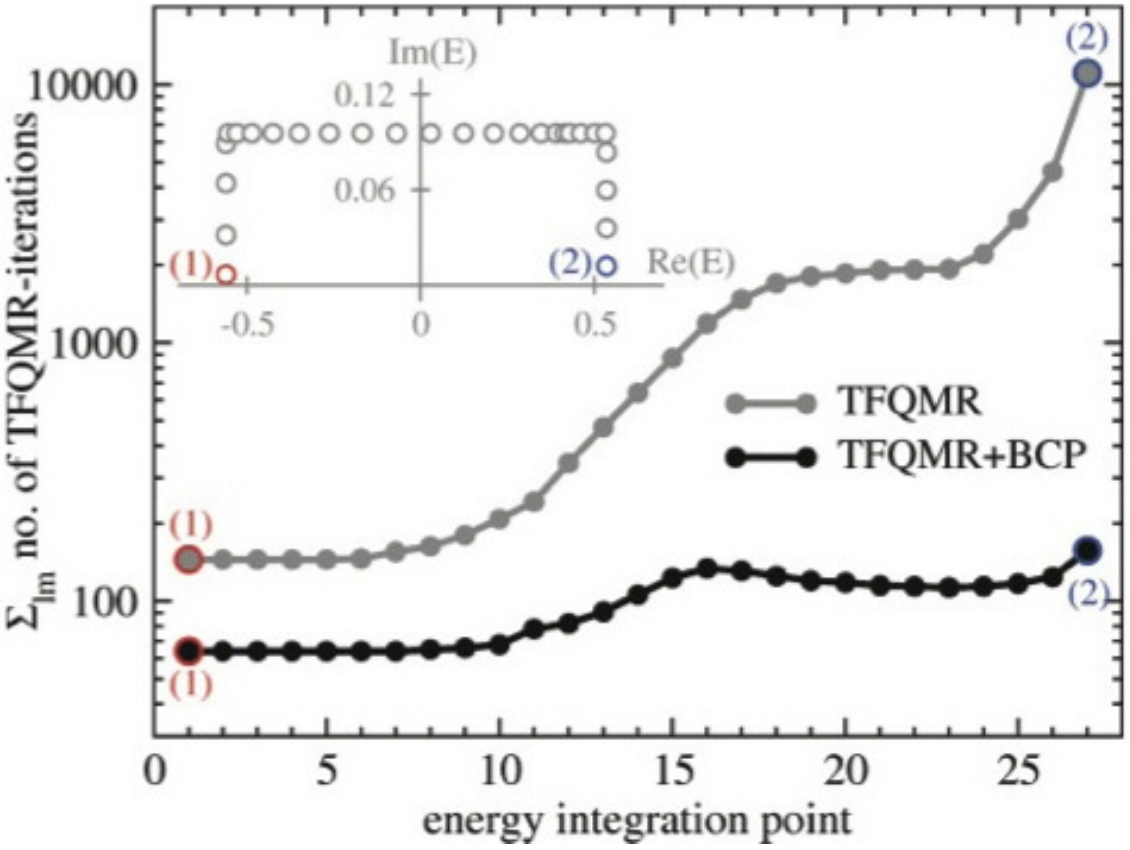
Preconditioning



1. replace $AX = B$
   by $(AP^{-1})(PX) = B$
2. split unit cell into subcells
3. divide matrix A into subblocks
4. average over equivalent subblocks
5. construct P from the averages
6. invert P in reciprocal space

multi-level block-circulant preconditioning

Convergence at all energy points



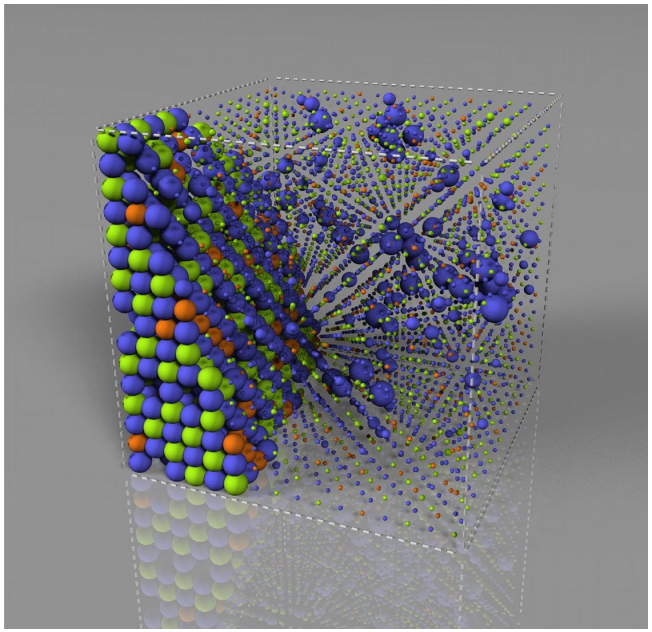
Bolten et al. Lin. Alg. Appl. 2012

Phase change material: $GeSb_2Te_4$

ARTICLES
PUBLISHED ONLINE: 14 OCTOBER 2012 | DOI: 10.1038/NMAT3456
nature materials

Role of vacancies in metal–insulator transitions of crystalline phase-change materials

W. Zhang[1], A. Thiess[2,3], P. Zalden[4], R. Zeller[2], P. H. Dederichs[2], J-Y. Raty[5], M. Wuttig[4,6]*, S. Blügel[2,6] and R. Mazzarello[1,6]*



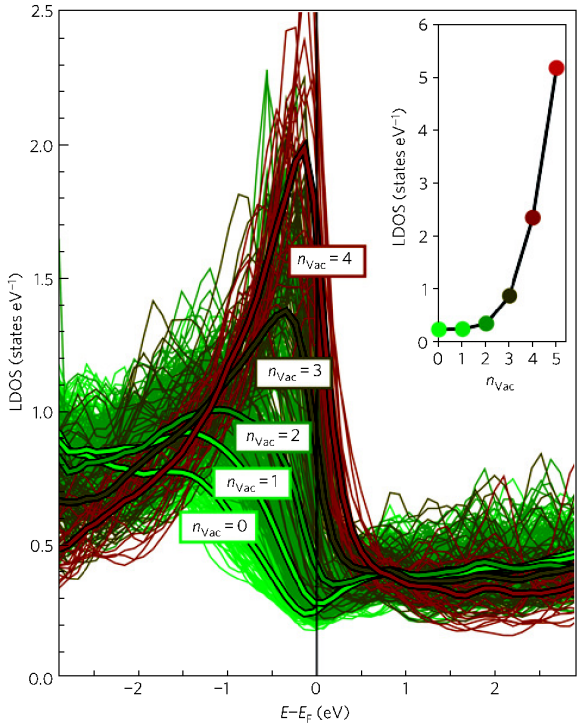
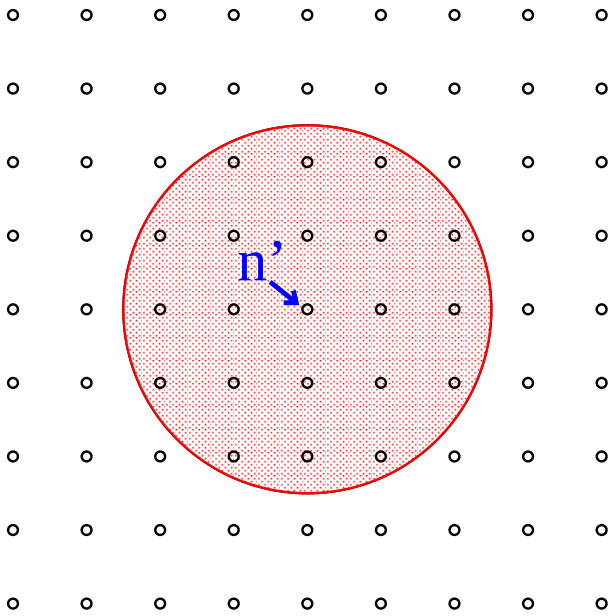


Figure 1 | Local density of *p* states (LDOS) on the 500 Te sites in a $Ge_{125}Sb_{250}Te_{500}$ supercell. Different colours are used to distinguish between Te atoms with different number of nearest-neighbour vacancies, n_{Vac}. For each of these groups the average LDOS is shown as a thick line i the corresponding colour. An increasing number of nearest-neighbour vacancies leads to a pronounced increase in the LDOS near E_F. This is further corroborated in the inset, which shows the average LDOS on Te atoms at E_F as a function of n_{Vac}, calculated from the larger $Ge_{512}Sb_{1024}Te_{2048}$ supercell.

Linear scaling mode



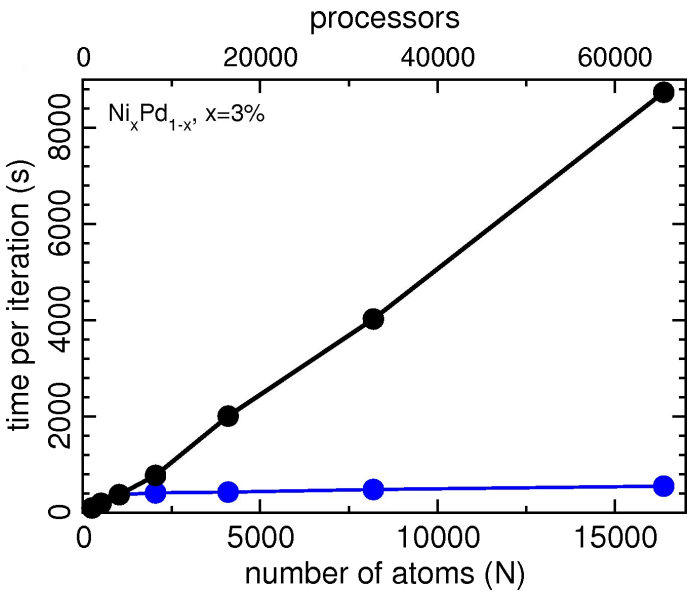
nearsightedness principle

- **truncate:** $G^{nn'}_{lml'm'} = 0$ for $|\underline{R}^n - \underline{R}^{n'}| > r_{cut}$
- **truncation leads to** $O(1)$ **memory/processor**
- **truncation leads to** $O(N)$ **computing time**

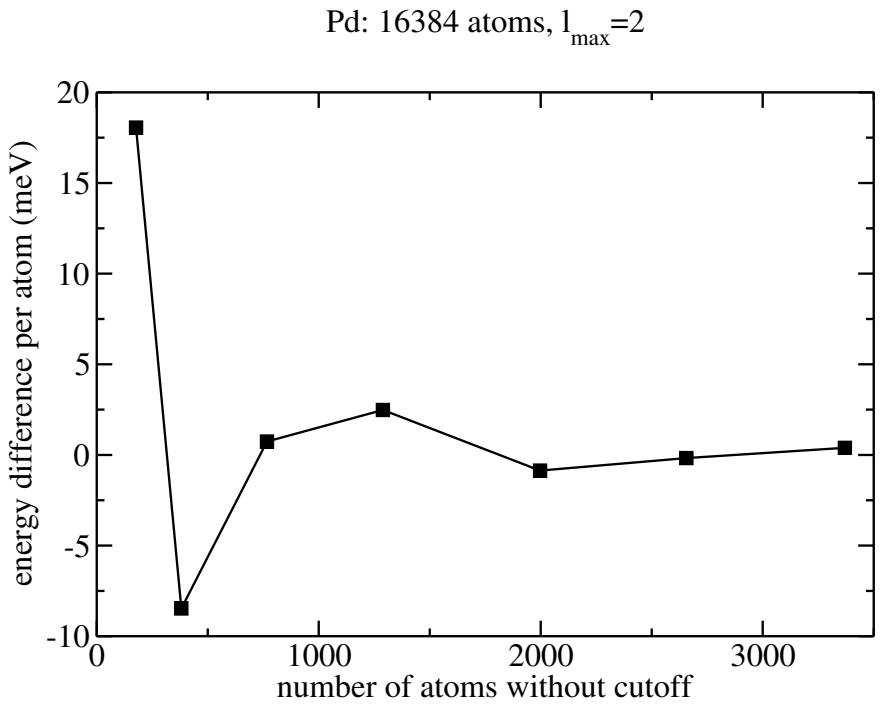
$$\begin{pmatrix} A_{CC} & A_{CR} \\ A_{RC} & A_{RR} \end{pmatrix} \begin{pmatrix} G^{(i)}_{CC} \\ 0 \end{pmatrix} = \begin{pmatrix} A_{CC} G^{(i)}_{CC} \\ A_{RC} G^{(i)}_{CC} \end{pmatrix}$$

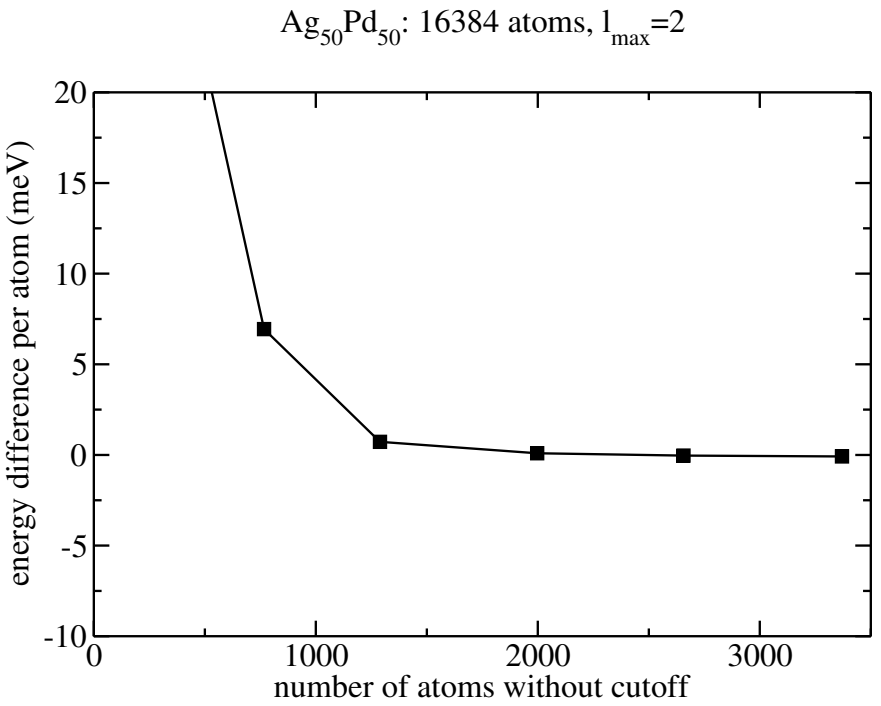
use $G^{(i+1)}_{CC} = A_{CC} G^{(i)}_{CC}$ **and**
replace $G^{(i+1)}_{RC} = A_{RC} G^{(i)}_{CC}$ **by 0**
C **denotes inner space and** R **outer space**

wall clock time $O(N^2)$ **vs.** $O(N)$



Truncation error for total energies

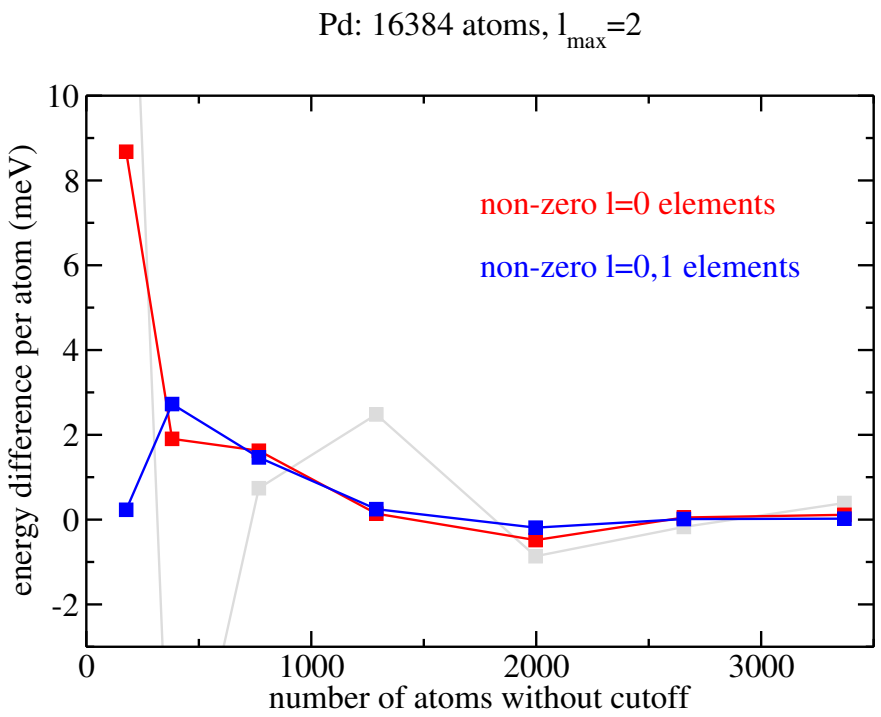


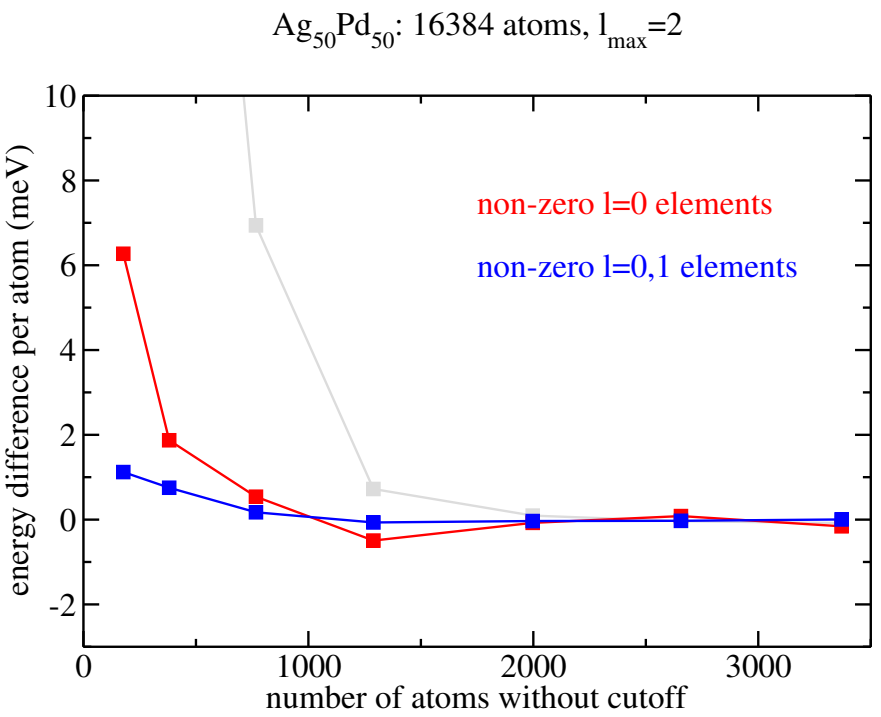


if total energy errors of several meV/atom are tolerated,

truncation regions with 1000 to 2000 atoms seem to be large enough

Truncation error for total energies





attainable total energy precisions

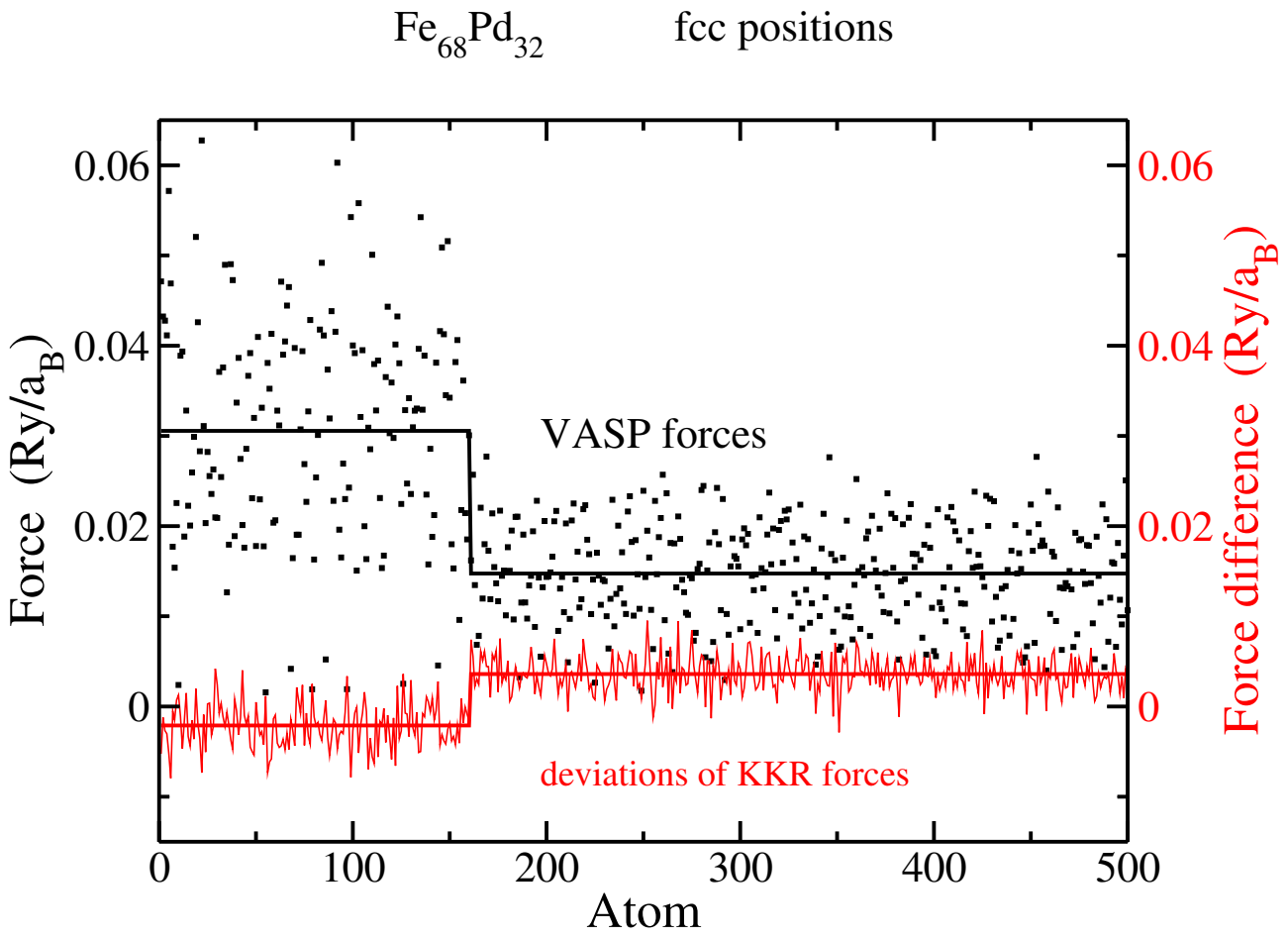
about 0.1 meV/atom with truncation regions of a few thousand atoms

about 1 meV/atom with truncation regions of a few hundred atoms

important: s channels are described by matrix blocks of size 1×1 instead of size $(l_{\max}+1)^2 \times (l_{\max}+1)^2 \Rightarrow$ reduces number of flops by a factor of $(l_{\max}+1)^6$

Accuracy ?

Forces: KKR compared to VASP



disordered alloy simulated by supercell with 500 atoms

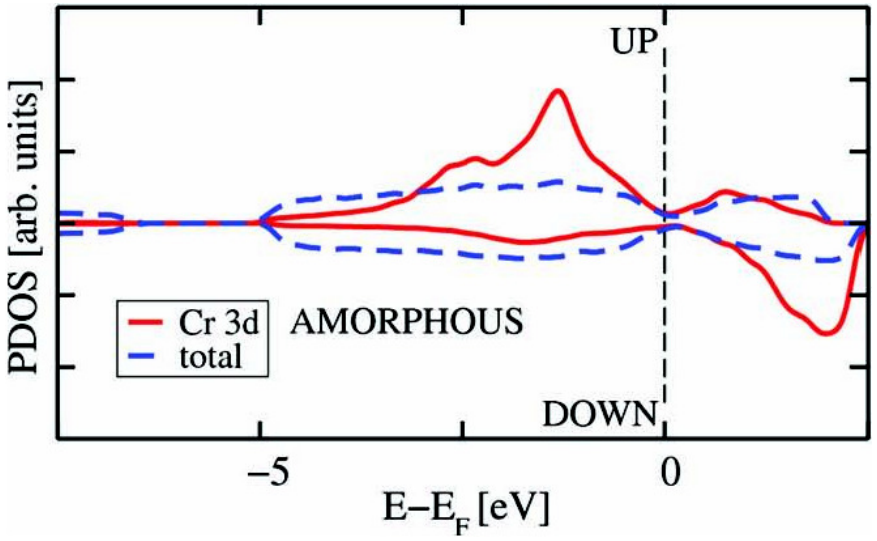
Amorphous system: $Cr_{15}Ge_{15}Sb_{41}Te_{120}$

Materials Views
www.MaterialsViews.com

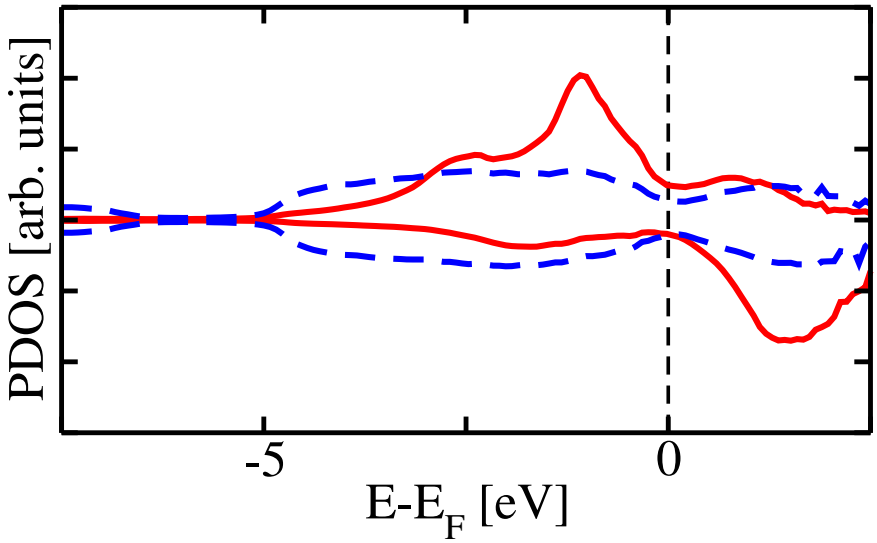
ADVANCED MATERIALS
www.advmat.de

Magnetic Properties of Crystalline and Amorphous Phase-Change Materials Doped with 3*d* Impurities

*Wei Zhang, Ider Ronneberger, Yan Li, and Riccardo Mazzarello**



PWSCF (Quantum Espresso)

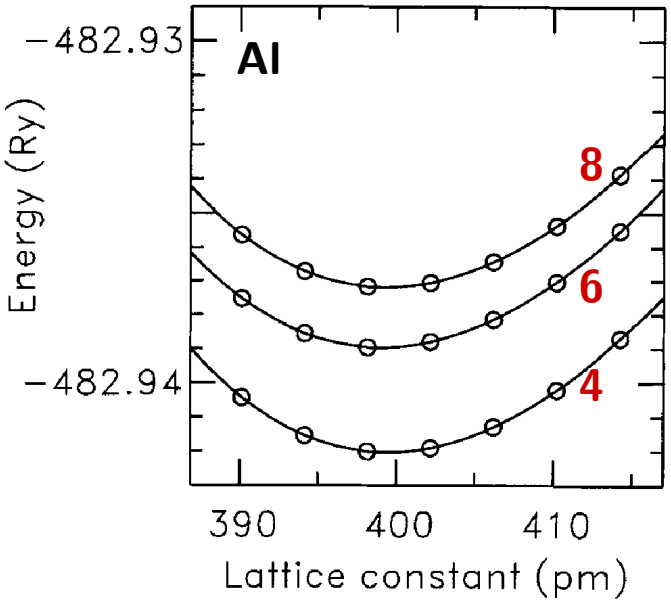


KKRnano

216 atomic positions were determined by PWSCF

125 empty cells were added for KKRnano (at sites determined by E. Rabel)

KKR total energy convergence with $l_{\max}$



Zeller et al. Phil. Mag. 1998

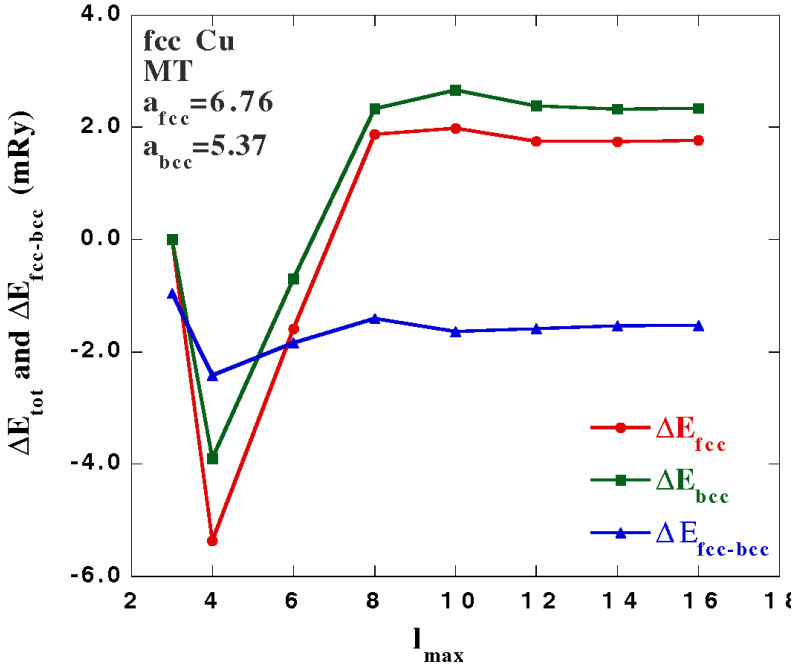


Figure 3. Convergence properties of ΔE_{tot} for fcc (circles) and bcc (squares)

Moghadam et al. JPCM 2002

Mathematical basis for the KKR method

Instead of $G^0(\underline{r},\underline{r}';\epsilon) = -\mathrm{i}\sqrt{\epsilon}\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l} j_l(r_<\sqrt{\epsilon})h_l(r_>\sqrt{\epsilon})Y_{lm}(\hat{\underline{r}})Y_{lm}(\hat{\underline{r}}')$

use $G^0(\underline{r},\underline{r}';\epsilon) = -\mathrm{i}\sqrt{\epsilon}\sum_{l=0}^{l_{\max}}\sum_{m=-l}^{m=l} j_l(r_<\sqrt{\epsilon})h_l(r_>\sqrt{\epsilon})Y_{lm}(\hat{\underline{r}})Y_{lm}(\hat{\underline{r}}')$

and solve the integral equations for this approximation. This can be done exactly.

Zeller JPCM 2013

---

important results:

- $G(\underline{r}+\underline{R}^n,\underline{r}'+\underline{R}^{n'};\epsilon) = \sum_{LL'}^{l_{\max}} \mathcal{G}^{nn'}_{LL'}(r,r';\epsilon)Y_L(\hat{\underline{r}})Y_{L'}(\hat{\underline{r}}')$ **is exact**
- **rate of convergence of** $\mathrm{Im}G$ **with** $l_{\max}$ **is exponential**

  consequence of optical theorem $G - G^+ = (1+GV)(G^0 - G^{0+})(1+VG^+)$ **and** $j_l(x) \approx x^l/(2l+1)!!$

Total energy functional

$$E_{\mathrm{tot}}[n(\underline{r})] = T_{\mathrm{s}}[n(\underline{r})] + U[n(\underline{r})] + E_{\mathrm{en}}[n(\underline{r})] + E_{\mathrm{nn}} + E_{\mathrm{xc}}[n(\underline{r})]$$

$E_{\mathrm{xc}}[n(\underline{r})]$ **must be approximated, for** $T_{\mathrm{s}}[n(\underline{r})]$ **exact result is known:**

$$T_{\mathrm{s}}[n(\underline{r})] = 2\sum_i \int \mathrm{d}\underline{r}\varphi_i^\star(\underline{r})(-\nabla_{\underline{r}}^2)\varphi_i(\underline{r})$$

but exact solution of $\left[-\nabla_{\underline{r}}^2 + V(\underline{r})\right]\varphi_i(\underline{r}) = \epsilon_i\varphi_i(\underline{r})$ **is necessary**

Principal challenge is the finite number of potential matrix elements

in plane wave methods: $V(\underline{G},\underline{G}') = \int \mathrm{d}\underline{r}\ \mathrm{e}^{-\mathrm{i}\underline{G}\underline{r}}\ V(\underline{r})\ \mathrm{e}^{\mathrm{i}\underline{G}'\underline{r}}$

in the KKR method: $V^n_{lm,l'm'}(r) = \int_n \mathrm{d}\hat{\underline{r}}\ Y_{lm}(\hat{\underline{r}})\ V(\underline{r})\ Y_{l'm'}(\hat{\underline{r}})$

...

Additional complication in the KKR method

non-linear dependence on ϵ prevents invariance for constant potential shifts

Projection potentials and total energy convergence in the KKR method 15

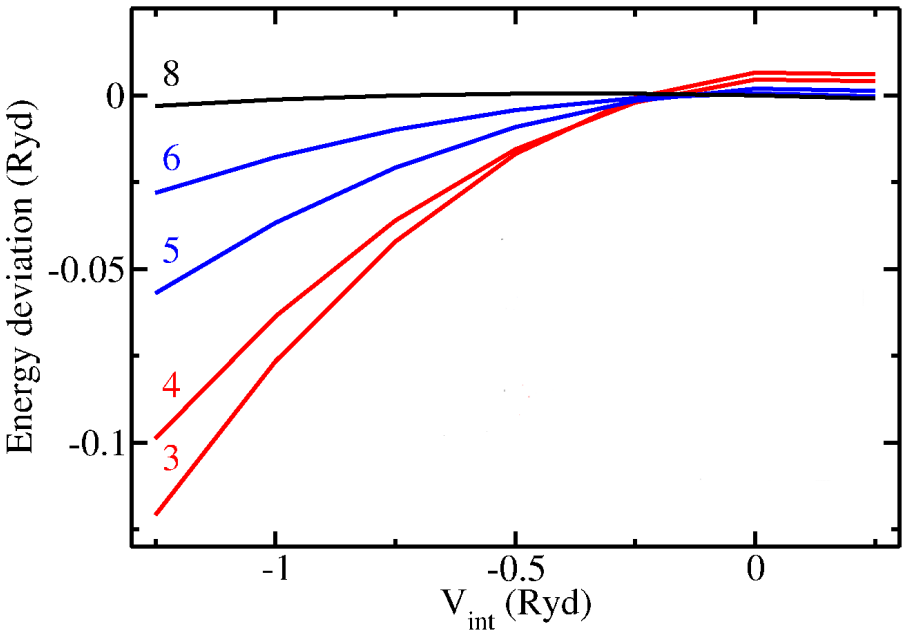
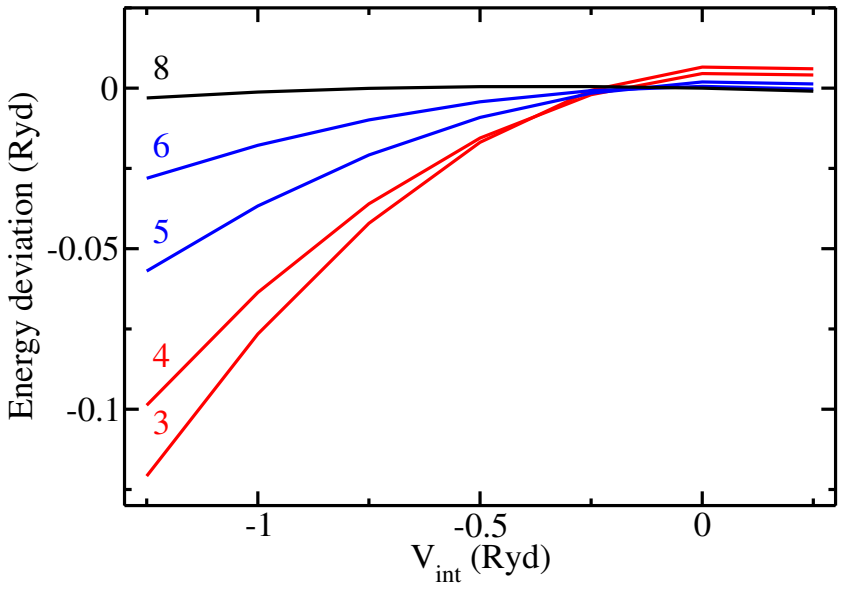
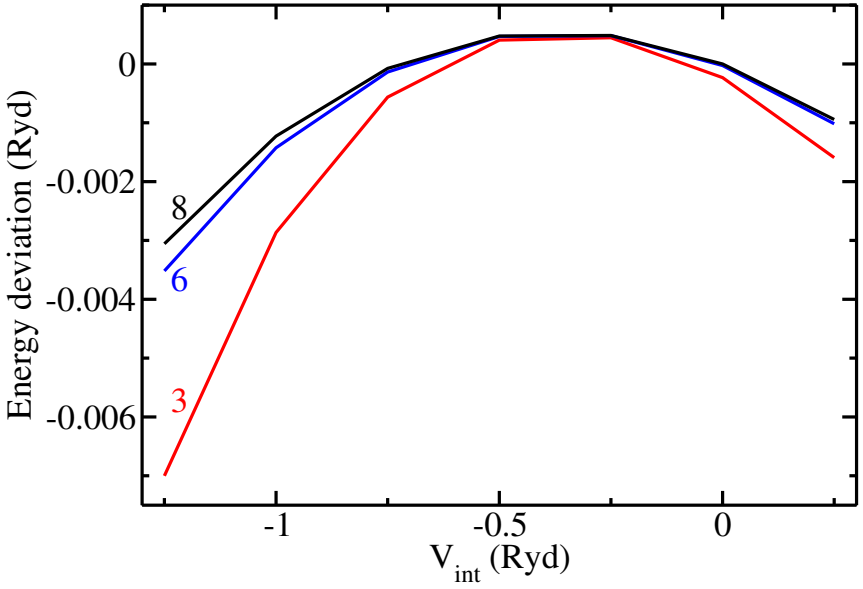


Figure 2. Total energy for Al as function of $V_{\rm int}$, the prescribed value for the average of the interstitial potential. The numbers at the curves indicate the value used for $l_{\rm max}$.

- **error arises from setting $V_{lml'm'} = 0$ for $l, l' > l_{\rm max}$**
- **this approximation is harmless during the selfconsistency iterations and for calculating $E_{\rm dc}$**
- **this approximation crucially affects single-particle energies $E_{\rm sp}$ $\Rightarrow$ use high $l_{\rm max}$ only for $E_{\rm sp}$**

Correction for single-particle energies





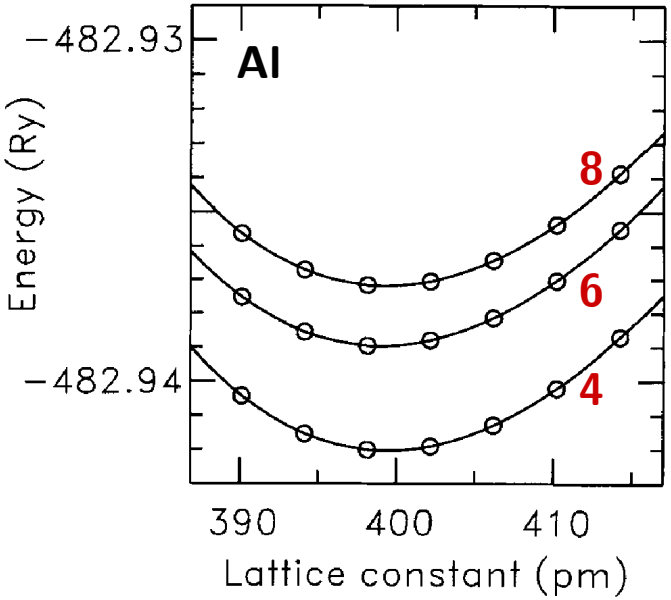
$l_{\max} = 8$ correction applied only to single-particle energies

KKR matrix equation solved with $l_{\max} < 8$

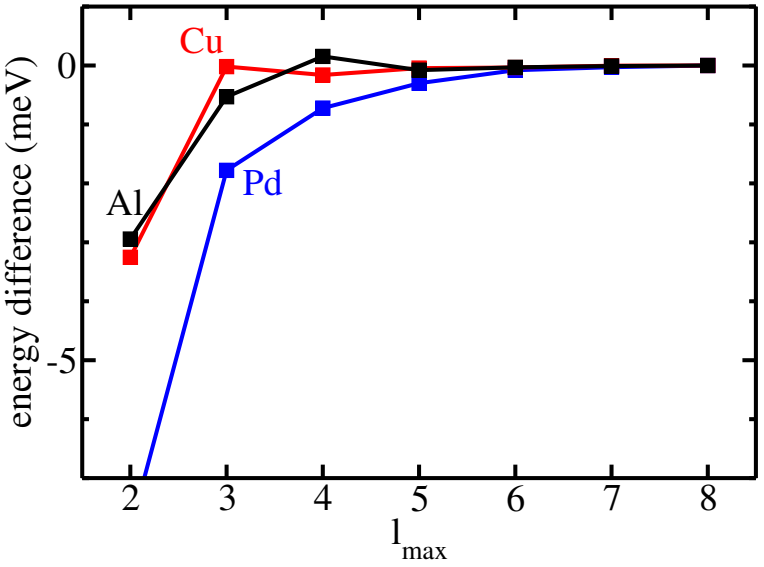
single-particle energies calculated with Lloyd's formula for $l_{\max} = 8$

correction necessary only at the end of the self-consistency steps

KKR total energy convergence with $l_{\max}$



total energy calculated
with correction applied to $E_{\rm sp}$



Conclusions

- precise DFT calculations for large system are possible
  (needed for advanced material science problems)
- our approach KKRnano uses
  - repulsive reference system $\Rightarrow$ sparse matrices
  - iterative solutions with the QMR method
  - efficient parallelization on modern supercomputers
- KKRnano requires
  - $O(N^2)$ computing time and $O(N)$ memory
    if no compromise on accuracy is made
  - $O(N)$ computing time and $O(1)$ memory
    if total energy errors of meV are tolerated ($>\approx 2000$ atoms needed)
- largest systems up-to-date
  - 65536 atoms with $l_{\max} = 3$ (shape memory alloy Ni_2MnGa)
  - 262144 atoms with $l_{\max} = 2$ (disordered AgPd alloy)
  - work in progress: half a million atoms

Conclusions

- the KKR method is accurate and efficient for solving the KS equation

- number of potential matrix elements determines the total energy accuracy
  - higher number required for E_{sp}
  - smaller number sufficient for $n(\underline{r})$ and E_{dc}
- improvements planned for KKRnano
  - efficient Ewald method for the electrostatic potential in large systems
  - removal of near field errors in the electrostatic potential
  - accurate calculation of irregular single-site solutions near the origin
- open question

  can a total energy functional be formulated that is stationary
  when potential matrix elements are neglected