Linear-Scaling KKR Green-Function Method for Large-Scale Systems

Rudolf Zeller and Alexander Thiess

Institute of Advanced Simulation, Institut für Festkörperforschung
and German Research School for Simulation Sciences
Forschungszentrum Jülich

Collaboration: Matthias Bolten (Wuppertal), Irad Yavneh (Haifa)

Support: Stefan Blügel, Peter Dederichs, Heiner Müller-Krumbhaar (Jülich)

Outline

Problem: standard scaling in DFT is $O(N^3)$

Our method: KKR Green function method

Accuracy of our method

Reduction of scaling to $O(N^2)$ and $O(N)$

Parallelization strategies

Experience in applications and **outlook** to improvements

Standard density functional equations

wavefunctions

$$[-\nabla^2_{\underline{r}} + V(\underline{r}) - E_i]\Psi_i(\underline{r}) = 0$$

density

$$n(\underline{r}) = \sum_{E_i \leq E_F} |\Psi_i(\underline{r})|^2$$

potential $V = V_H + V_{XC}$

$$\nabla^2_{\underline{r}} V_H(\underline{r}) = -4\pi n(\underline{r})$$

Standard density functional equations

wavefunctions

$$[-\nabla^2_{\underline{r}} + V(\underline{r}) - E_i]\Psi_i(\underline{r}) = 0$$

density

$$n(\underline{r}) = \sum_{E_i \leq E_F} |\Psi_i(\underline{r})|^2$$

potential $V = V_H + V_{XC}$

$$\nabla^2_{\underline{r}} V_H(\underline{r}) = -4\pi n(\underline{r})$$

$O(N^3)$ **bottleneck for large systems**

Key observations to achieve reduced scaling

$$[-\nabla^2_{\underline{r}} + V(\underline{r}) - E_i]\Psi_i(\underline{r}) = 0$$

1. observation: $V(\underline{r})$ **is local,** $\nabla_{\underline{r}}$ **is short ranged**

simple discretization

$\Rightarrow$ **sparse matrices**

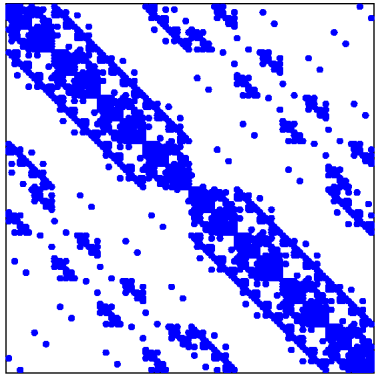
or localized basis sets

sparse matrices with $O(N)$ **matrix elements instead of** $O(N^2)$

How to exploit sparse matrices?

direct solution methods

	1D	2D	3D
full matrices:	$O(N^3)$	$O(N^3)$	$O(N^3)$
band matrices:	$O(N)$	$O(N^2)$	$O(N^{7/3})$
optimal algorithms	$O(N)$	$O(N^{3/2})$	$O(N^2)$



but implementation and parallelization difficult

iterative solution methods

$\Rightarrow$ $O(N)$ work to obtain one wavefunction

$O(N)$ wavefunctions are required $\Rightarrow$ $O(N^2)$ scaling

Difficulty exists with wavefunctions

wavefunction orthogonalization $\Rightarrow O(N^3)$ scaling

2. observation density matrix or Green function can be calculated directly

$$\rho(\underline{r}, \underline{r}', T) = -\frac{1}{\pi}\mathrm{Im}\int_{-\infty}^{\infty} f(E - E_F, T)G(\underline{r}, \underline{r}', E)\mathrm{d}E$$

$$\Rightarrow O(N^2) \text{ scaling}$$

Difficulty exists with wavefunctions

wavefunction orthogonalization $\Rightarrow O(N^3)$ **scaling**

2. observation density matrix or Green function can be calculated directly

$$\rho(\underline{r}, \underline{r}', T) = -\frac{1}{\pi}\mathrm{Im}\int_{-\infty}^{\infty} f(E - E_F, T)G(\underline{r}, \underline{r}', E)\mathrm{d}E$$

$$\Rightarrow O(N^2) \text{ scaling}$$

How to obtain linear scaling?

3. observation: (nearsightedness) density matrix decays exponentially in semiconductors and insulators

spatial truncation $\Rightarrow$ **only fixed number of density matrix elements are needed around each atom**

$$\Rightarrow O(N) \text{ scaling}$$

Are large computing facilities needed?

4. observation: computational work increases as $n_z n_{it} n_r N$ instead of N^3

- $n_z N$ is number of non-zero matrix elements instead of N^2
- n_{it} is number of iterations
- n_r is number of atoms in the truncation region

assume realistic values as $n_z = 50$, $n_{it} = 50$ and $n_r = 200$

$$\Rightarrow n_z n_{it} n_r N < N^3 \text{ for } N > 700$$

- supercomputing (with efficient parallelization) is required
- accuracy depends on n_z, n_{it} and n_r

Basic features of the KKR-GF method

all electron method (no pseudopotentials)

applicable to metallic, semiconducting and insulating systems

useful for disordered systems (KKR-CPA)

available in relativistic version (Dirac equation)

KKR: Korringa (1947), Kohn-Rostoker (1954)
and earlier Lord Rayleigh (1892)

Green-function version
in Jülich developed originally for impurity calculations
with Peter Dederichs and many other coworkers

Comparison wavefunction and Green function methods

wavefunction method

$$n(\underline{r}) = \sum_{E_i \leq E_F} |\Psi_i(\underline{r})|^2$$

with basis-sets methods

$\Rightarrow$ linear eigenvalue problem

$$[-\nabla^2_{\underline{r}} + V(\underline{r}) - E_i]\Psi_i(\underline{r}) = 0$$

Green function method

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

with multiple-scattering theory

$\Rightarrow$ system of linear equations

$$\left[-\nabla^2_{\underline{r}} + V(\underline{r}) - E\right] G(\underline{r}, \underline{r}', E) = -\delta(\underline{r} - \underline{r}')$$

$\Downarrow$

or alternatively by a Dyson equation

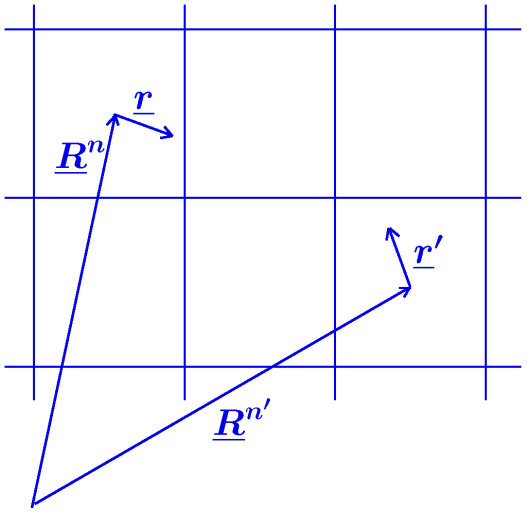
$$G(\underline{r}, \underline{r}', E) = G^r(\underline{r}, \underline{r}', E) + \int G^r(\underline{r}, \underline{r}'', E)\,[V(\underline{r}'') - V^r(\underline{r}'')]\,G(\underline{r}'', \underline{r}', E)\mathrm{d}\underline{r}''$$

with known Green function G^r of a suitably chosen reference system

e.g. for $V^r(\underline{r}) \equiv 0 \quad \Rightarrow \quad G^0(\underline{r}, \underline{r}', E) = -\dfrac{\exp(\mathrm{i}\sqrt{E}|\underline{r} - \underline{r}'|)}{4\pi|\underline{r} - \underline{r}'|}$

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

Accuracy in comparison with FLAPW results

lattice constant [a.u.]

	Al	Fe	Ni	Cu	Rh	Pd	Ag
FPKKR	7.52	5.20	6.46	6.63	7.09	7.24	7.53
FLAPW	7.51	5.18	6.46	6.63	7.09	7.25	7.54

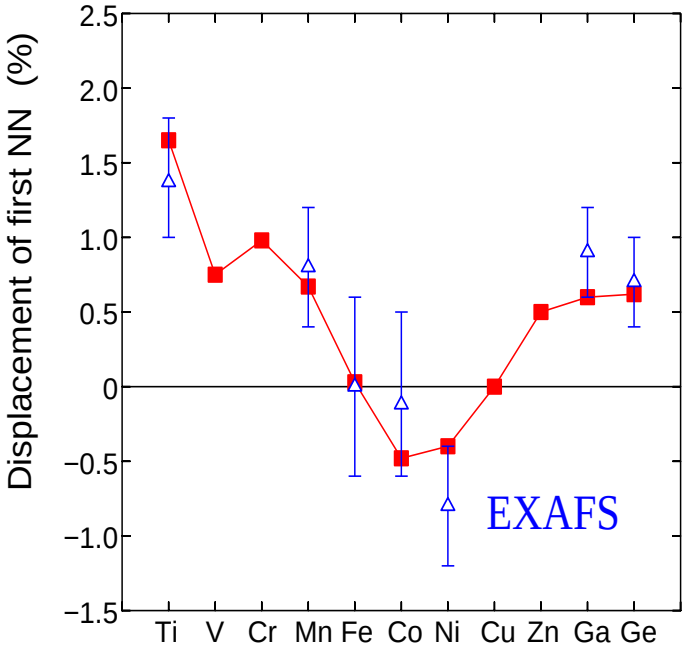
bulk modulus [Mbar]

	Al	Fe	Ni	Cu	Rh	Pd	Ag
FPKKR	0.82	2.43	2.54	1.88	3.18	2.28	1.39
FLAPW	0.85	2.57	2.56	1.90	3.12	2.29	1.41

from Asato et al., PRB 60, 5202 (1999)

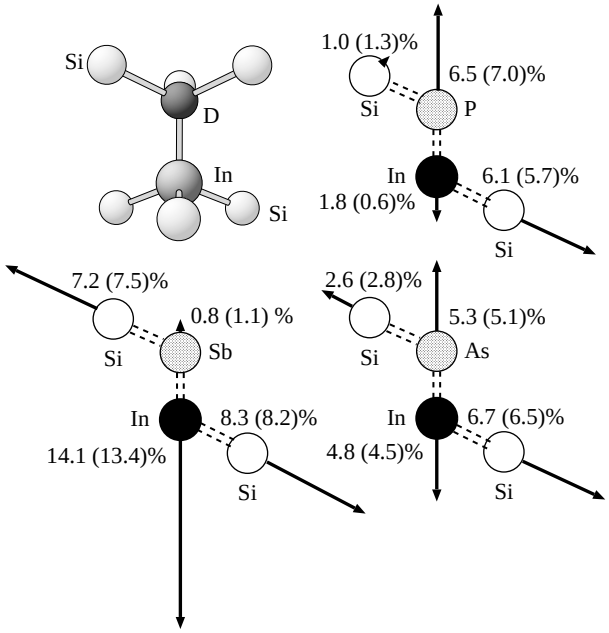
Forces and relaxations

comparison with experiment



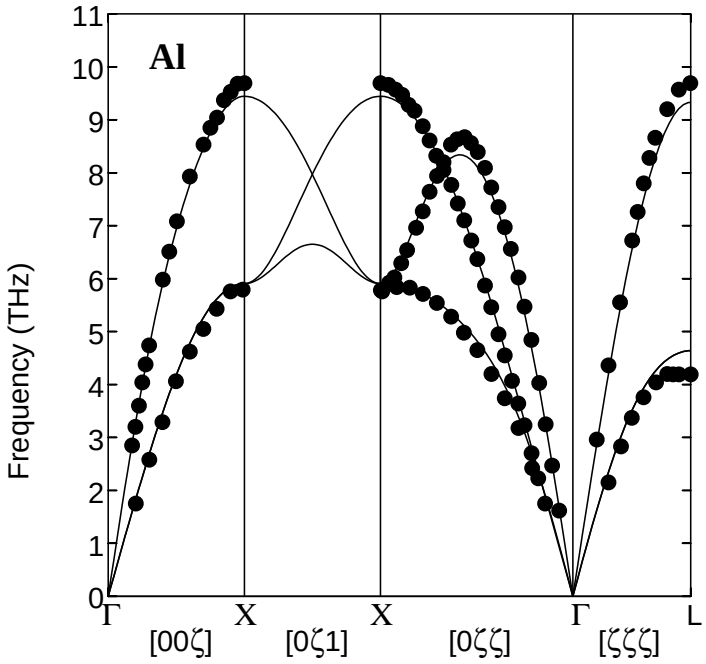
displacement of nearest Cu neighbour atoms around impurities

comparison with ab-initio plane-wave results



indium-donor complexes in Si

Phonons (example aluminium)



- displace central Al atom by 0.5 %
- calculate forces on six shells of neighbours
- Fourier transform force constant matrix
- determine eigenvalues of dynamical matrix

How to achieve sparsity in the KKR matrix equation?

originally screening transformation of structure constants

- TB-LMTO method
  Andersen and Jepson (1984)
- screened (or TB) KKR method
  Andersen et al. (1992)
  energy-dependent screening parameters $\Rightarrow$ are difficult to obtain
- introduction of least-square fits
  Szunyogh et al. (1994)
- hard sphere solid and unitary spherical waves
  Andersen et al. (1994)
- concept of a repulsive reference system
  Zeller et al. (1995)

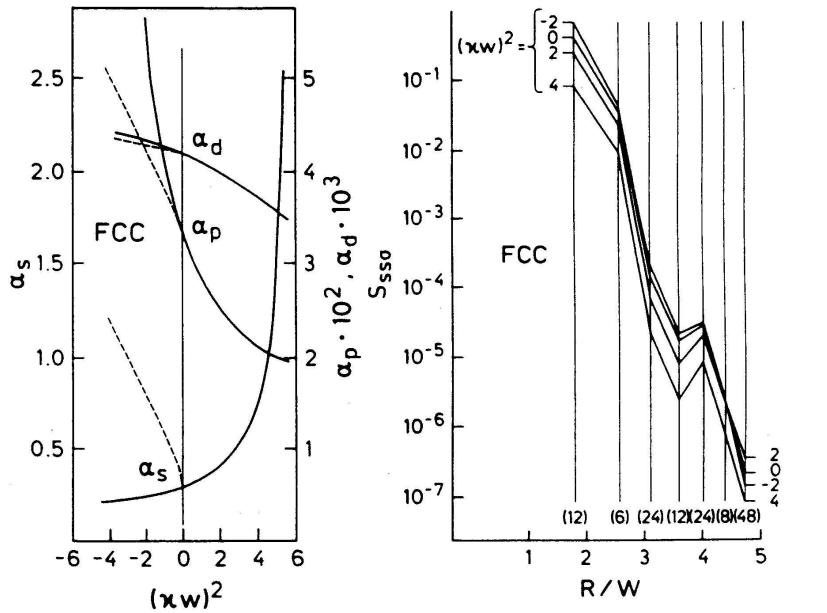


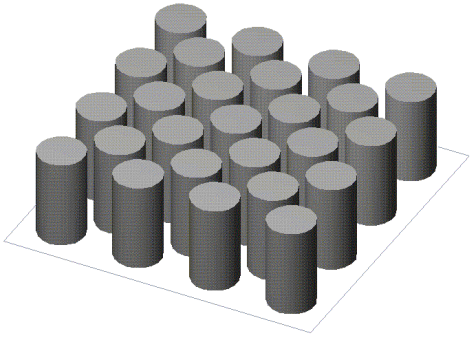
Fig. 3. *Left:* Screening parameters for the most localized structure matrix found as function of the energy κ^2 times the WS-radius squared w^2. *Right:* ss–matrix element of the most localized structure matrix in the fcc structure as a function of the energy and the interatomic distance (logarithmic scale).

What is a suitable repulsive reference system?

we choose an infinite array of repulsive potentials

$\Rightarrow$ a finite energy E_0 exists such that below E_0

- reference system has no eigenstates
- reference Green function decays exponentially



with $n_z = 50$ total energy accuracy better than

1 meV is achieved

How to solve KKR matrix equation by iteration

$$G^{nn'}(E) = G^{r,nn'}(E) + \sum_{n''} G^{r,nn''}(E)\Delta t^{n''}(E)G^{n''n'}(E)$$

$O(N^2)$ computational effort, $O(N)$ storage

easy parallelization over the atoms and L components

simple iterations $\boxed{G^{(i+1)} = G^r + G^r \Delta t G^{(i)}}$ do not work

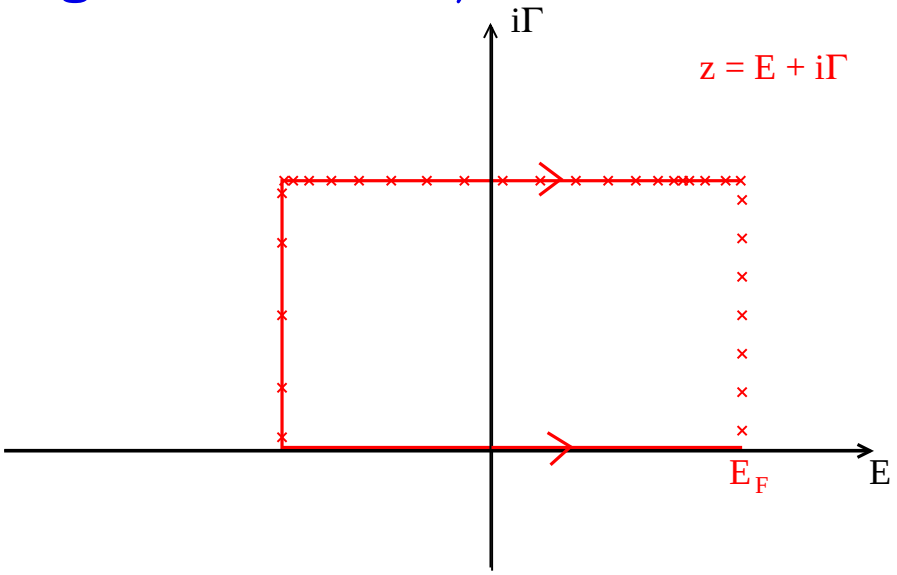
---

QMR method and complex energy integration with $T \neq 0$

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

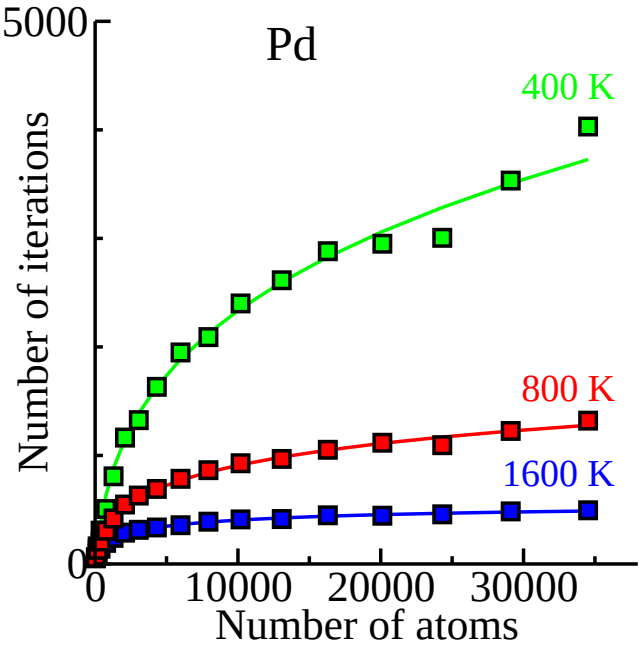
$f(E,T)$ = Fermi-Dirac function

usually 30 to 40 energy points are enough



How many iterations are needed?

as function of T and size



curves fitted to $n_{it}^{\infty} - \alpha e^{-\gamma N^{1/3}}$

crossover : $\beta n_{it} n_z N^2 = n_E N^3$

$$\Rightarrow N = \beta n_{it} n_z / n_E$$

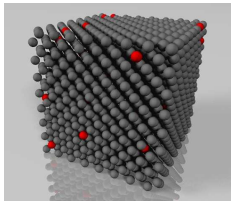
$n_{it} \approx 400$, $n_z \approx 50$ and $n_E \approx 40$

$$\Rightarrow N \approx 500\beta$$

β depends on flop rate, parallel efficiency

KKRnano

- our newly developed code (presently implemented in supercell mode)
- why nano?

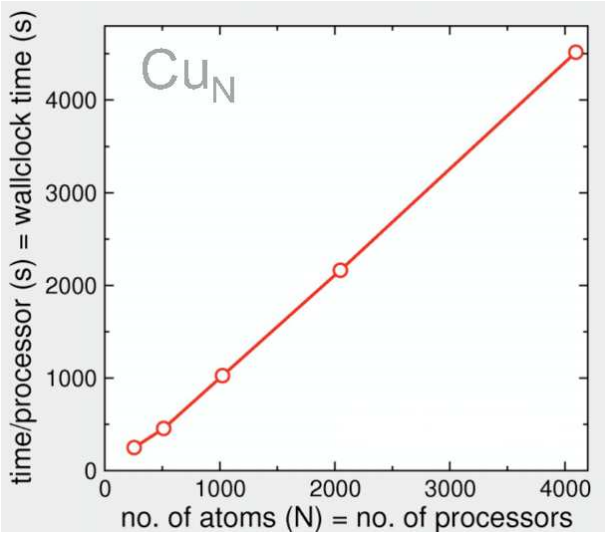


nanosystems contain many atoms
(2000 in a cube of 3 nm length)

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

scaling behaviour



present applications

MgO: 1-10 % N, C $\Rightarrow J_{ij} \Rightarrow T_c$ (Poster P517)

GaN:Gd with 1-5 % interstitial N, O (LDA+U) $\Rightarrow T_c$

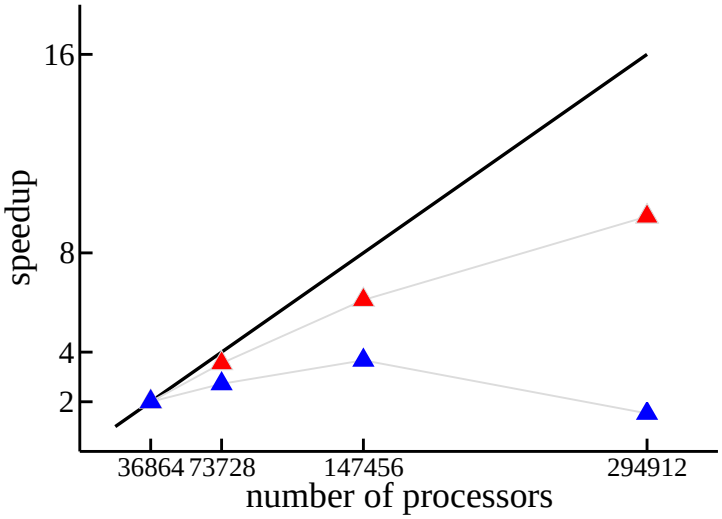
Si + shallow donors

How to use more processors than atoms?

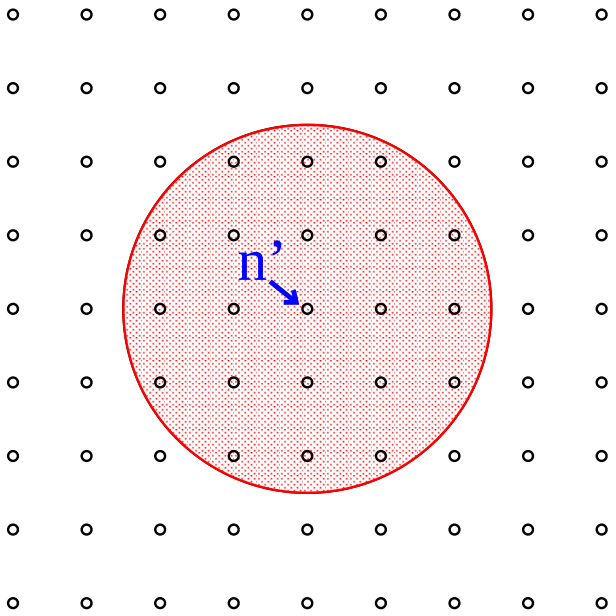
KKRnano uses four levels of parallelization MPI groups and communicators and point-to-point and collective messages

- 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 (implemented until now only in matrix equation)

test system: NiPd3071



How to achieve linear scaling

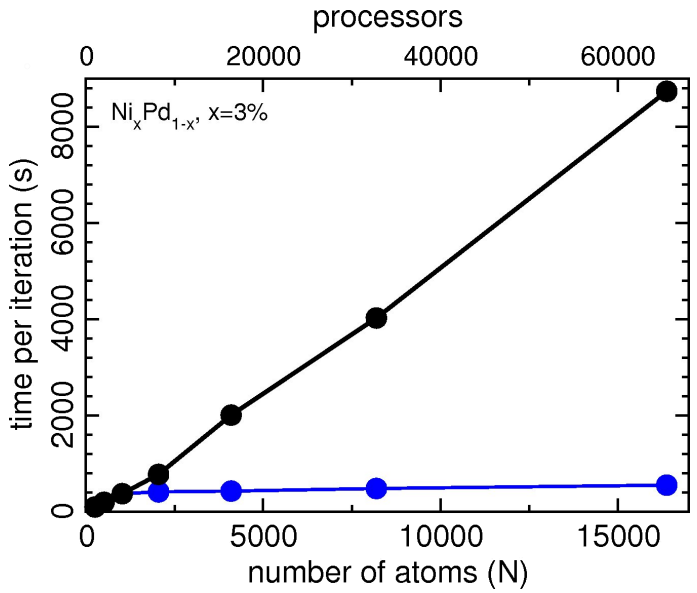


- **truncate:** $G^{nn'}_{LL'} = 0$ for $|\underline{R}^n - \underline{R}^{n'}| > r_{cut}$

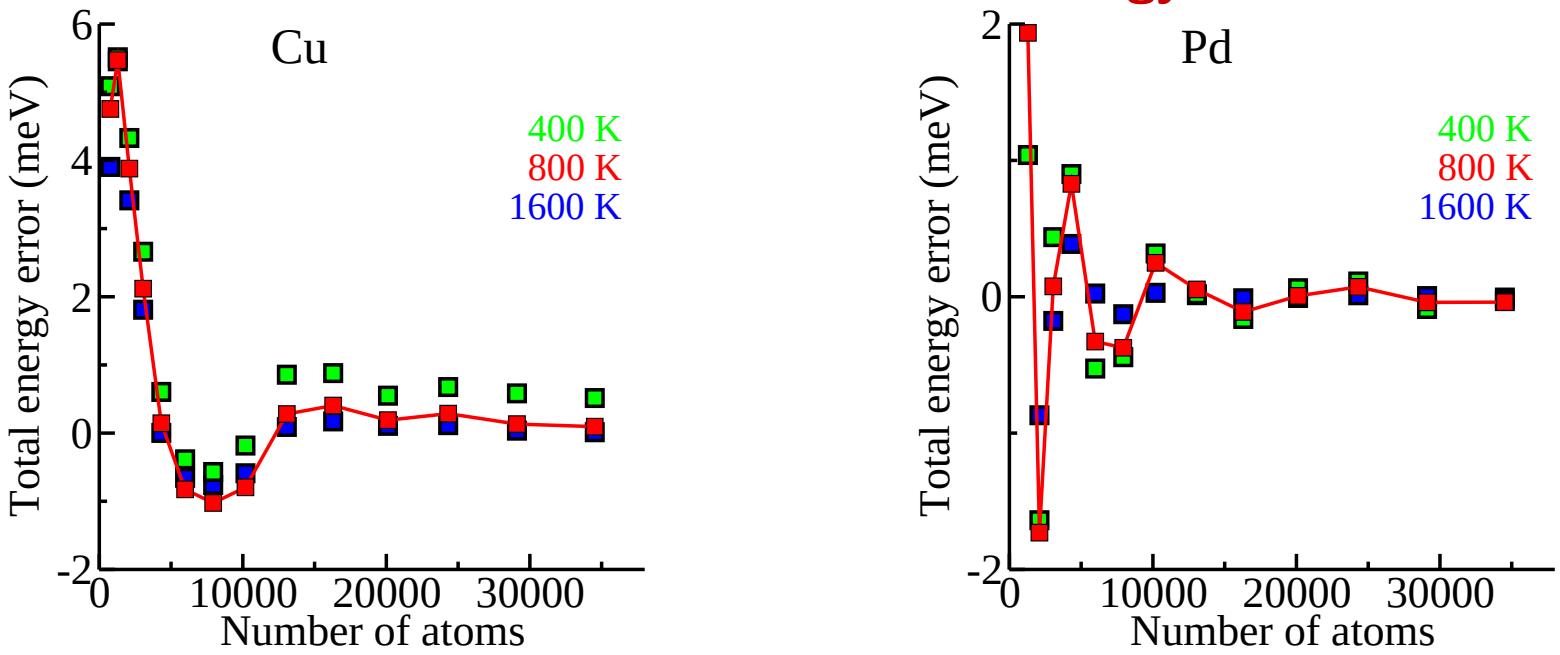
  G decays like the density matrix
- truncation leads to $O(1)$ storage
- truncation leads to $O(N)$ computing time

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

use $X_C^{(i+1)} = A_{CC} X_C^{(i)}$ and $X_R^{(i+1)} = 0$



Truncation error for total energy



---

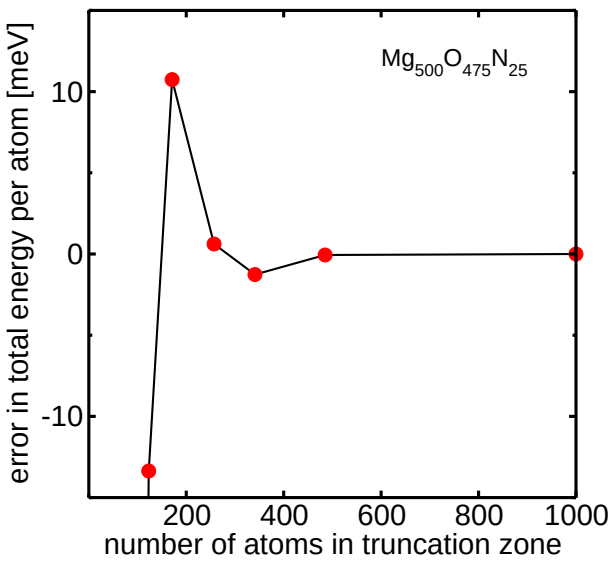
computational details:

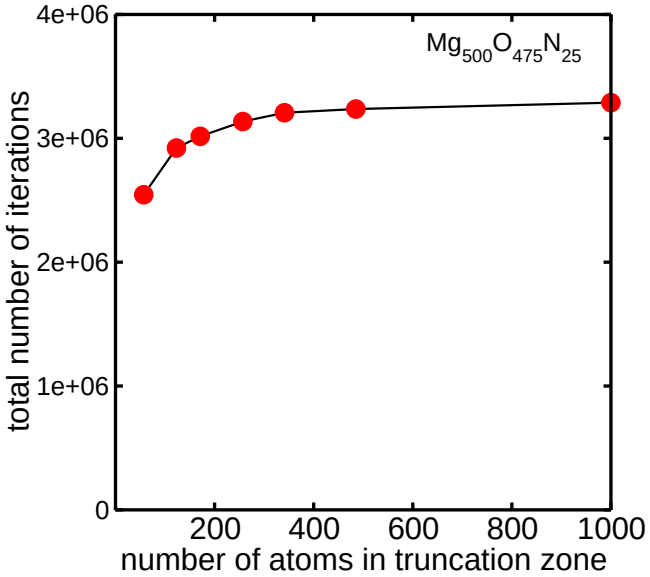
supercell with 131072 identical atoms in fcc geometry

energy without truncation error was obtained for a small cubic cell with 5984 $\underline{k}$ points

this corresponds to one special point $\underline{k} = \frac{2\pi}{a}(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ in the large supercell

KKRnano for band gap materials





remark: **use of Lloyd's formula is implemented**
to correct for the density normalization error
caused by l_{max} cutoff

How to improve the efficiency?

we use four strategies

1. use higher temperature without losing accuracy
2. use preconditioning
3. use good initialization
4. use postprocessing by Newton step

How can higher "temperature" be used?

increasing T reduces number of iterations but decreases accuracy

observation T corresponds to smoothing $f(E,T)$

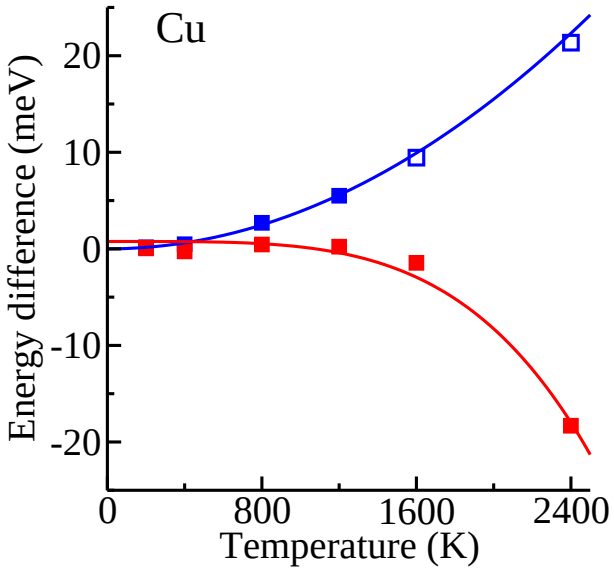
use more efficient smoothing $f(E,T) \Rightarrow 4/3 f(E,T) - 1/3 f(E,2T)$

$$E_{tot}(T) = E_{tot}(0) + \alpha_2 T^2 + \alpha_4 T^4 ... \Rightarrow$$

$$\frac{4}{3}E_{tot}(T) - \frac{1}{3}E_{tot}(2T) = E_{tot}(0) - 4\alpha_4 T^4 ...$$

only modified quadrature rules

no problems with negative occupancies



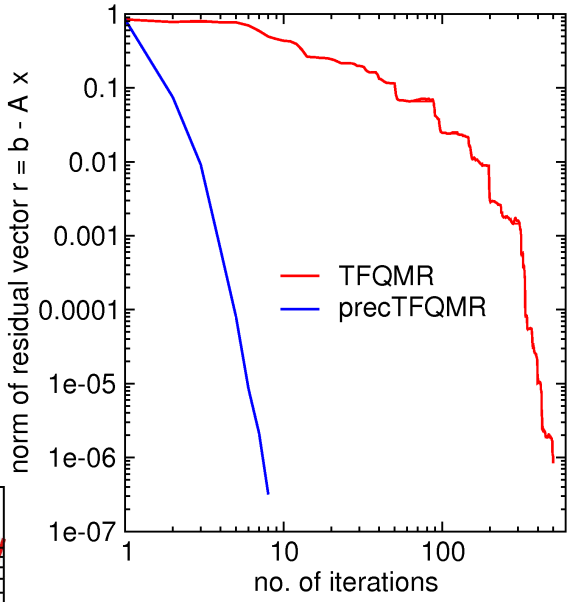
Is preconditioning useful?

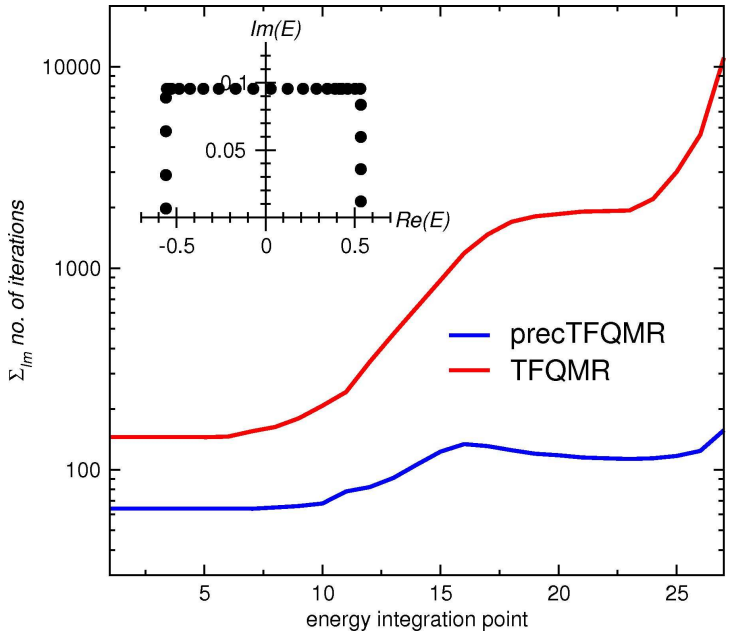
instead of $(1 - gt)G = g$ solve $P(1 - gt)G = Pg$

if $P(1 - gt) \approx 1$ fewer iterations are required

1. multigrid methods
2. incomplete LU
3. circular preconditioners

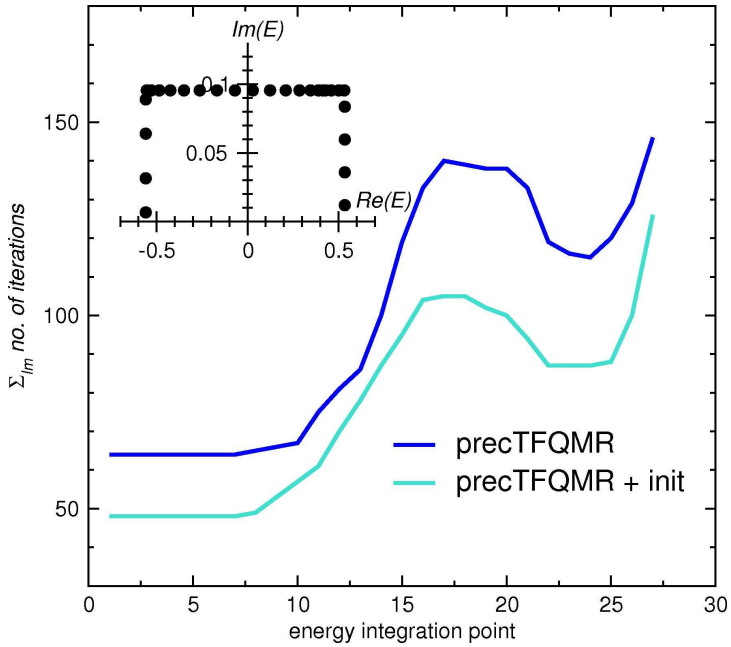
test system NiPd255 with randomly displaced atoms (about 2 % displacement)
convergence quality 10^{-7}
energy error $< \mu$eV





better than random initialization?

1. extrapolation along the energy contour
2. use results of previous self-consistency iteration



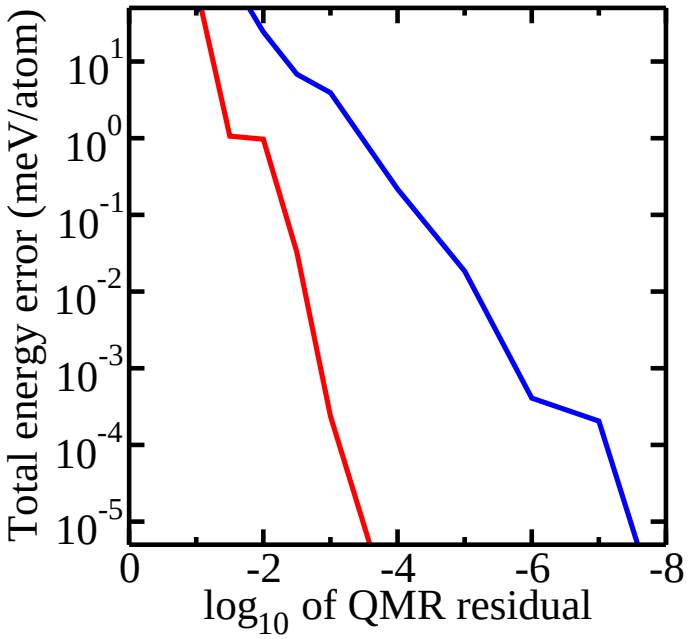
Are Newton steps possible?

Newton step for matrix inversion $X^{(i+1)}_{Newton} = X^{(i)}(2 - AX^{(i)})$

Advantage: number of correct digits doubled

Disadvantage: multiplication with dense matrix $X^{(i)}$

seems to scale as $O(N^3)$ and destroys parallel efficiency



Use particular features of KKR-GF method:

- density calculation requires only $O(N)$ on-site blocks of $G \approx X^{(i+1)}_{Newton}$
- using symmetry $G^{nn'}_{LL'} = G^{n'n}_{L'L}$ avoids data transfer between processors

Summary

1. Key concepts for linear-scaling DFT calculations:
   - sparse Hamiltonians (matrices)
   - iterative solutions
   - nearsightedness of electronic matter
2. Implementation in Korringa-Kohn-Rostocker Green function method by:
   - repulsive reference system
   - complex energy integration with $T \neq 0$ and QMR method
   - local truncation of the Green function
3. Accuracy of the KKR-GF method in general and for $O(N)$ calculations
4. Jülich computer code KKRnano is useful in $O(N^2)$ mode for thousands of atoms and is efficiently parallelized
5. For many thousand atoms $O(N)$ calculations are possible also in metals
6. Large reduction of computational work can be expected by:
   smoothing, preprocessing, preconditioning and postprocessing