Low-order scaling methods in density functional theories

- Introduction
- O(N) Krylov subspace method
- Numerically exact low-order scaling method
- O(N) nearly exact exchange functional
- Outlook

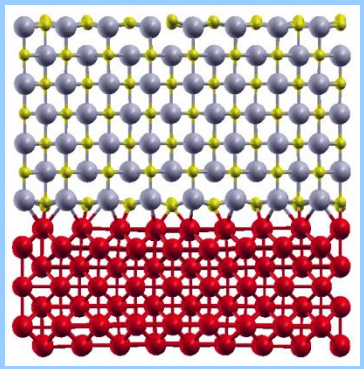
Taisuke Ozaki (ISSP, Univ. of Tokyo)

iDFT @ RIKEN, June 19, 2017

Towards first-principle studies for industry

10^3 – 10^6 atom

Steel



DFT calculations of thousands atoms is still a grand challenge.

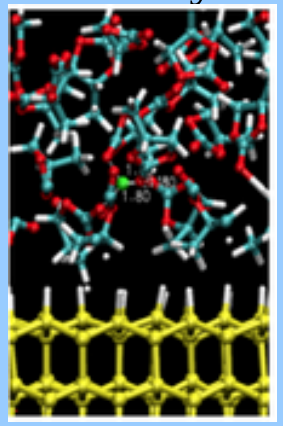
$O(N^3)$ → Low-order

10^2 atom

Many applications done.
There are many successes
even for material design.

DNA

Battery



Time scale

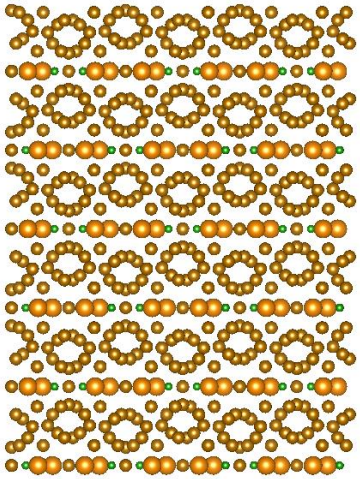
System size

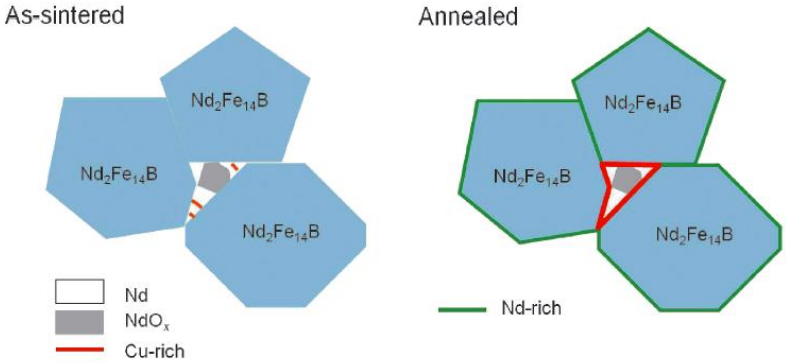
Materials properties

- Materials properties of actual materials are determined by intrinsic properties and secondary properties arising from inhomogeneous structures such as grain size, grain boundary, impurity, and precipitation.
- In use of actual materials, the materials properties can be maximized by carefully designing the crystal structure and higher order of structures .

http://ev.nissan.co.jp/LEAF/PERFORMANCE/

e.g., the coercivity of a permanent magnet of Nd-Fe-B is determined by crystal structure, grain size, and grain boundary.



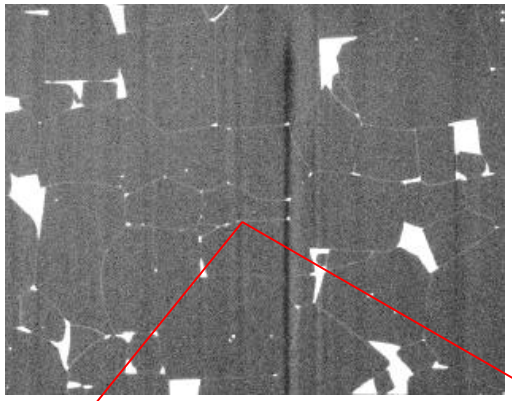


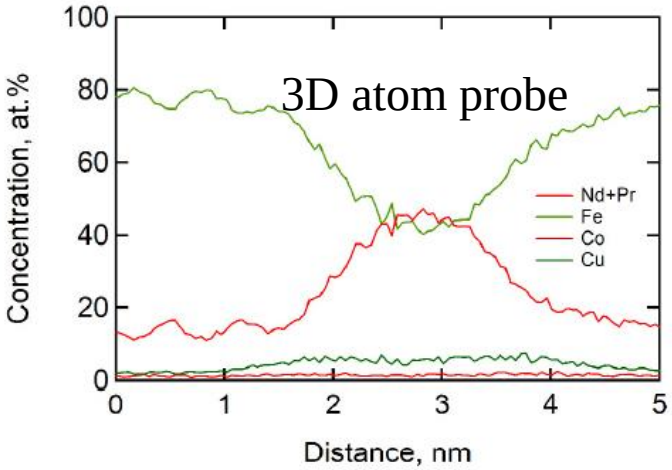
Experimental analysis of inhomogeneous materials

e.g. Grain boundary of a Nd-Fe-B permanent magnet

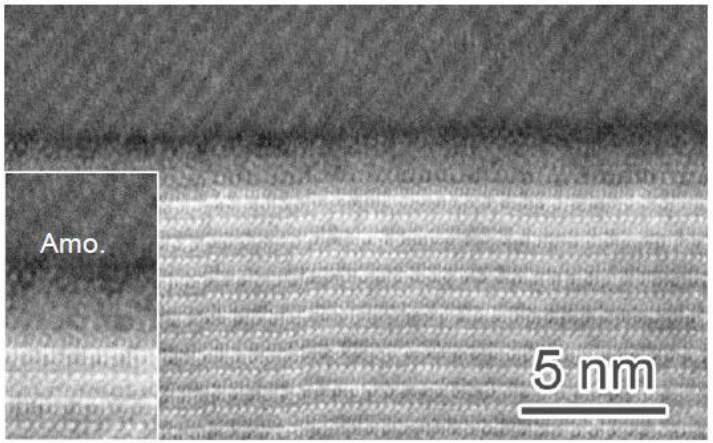
Hono@NIMS

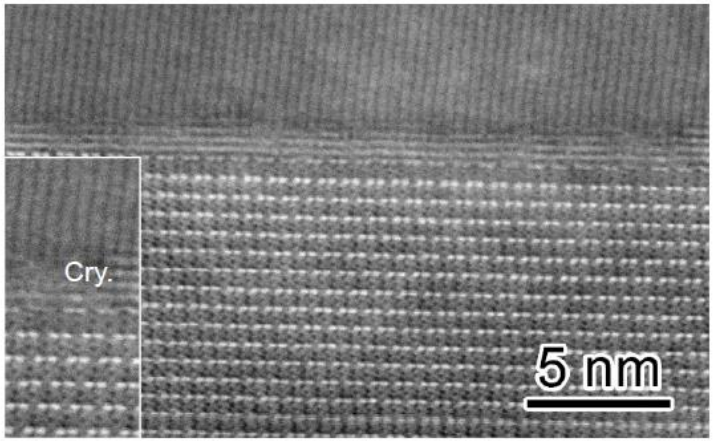
TEM





Around grain boundary





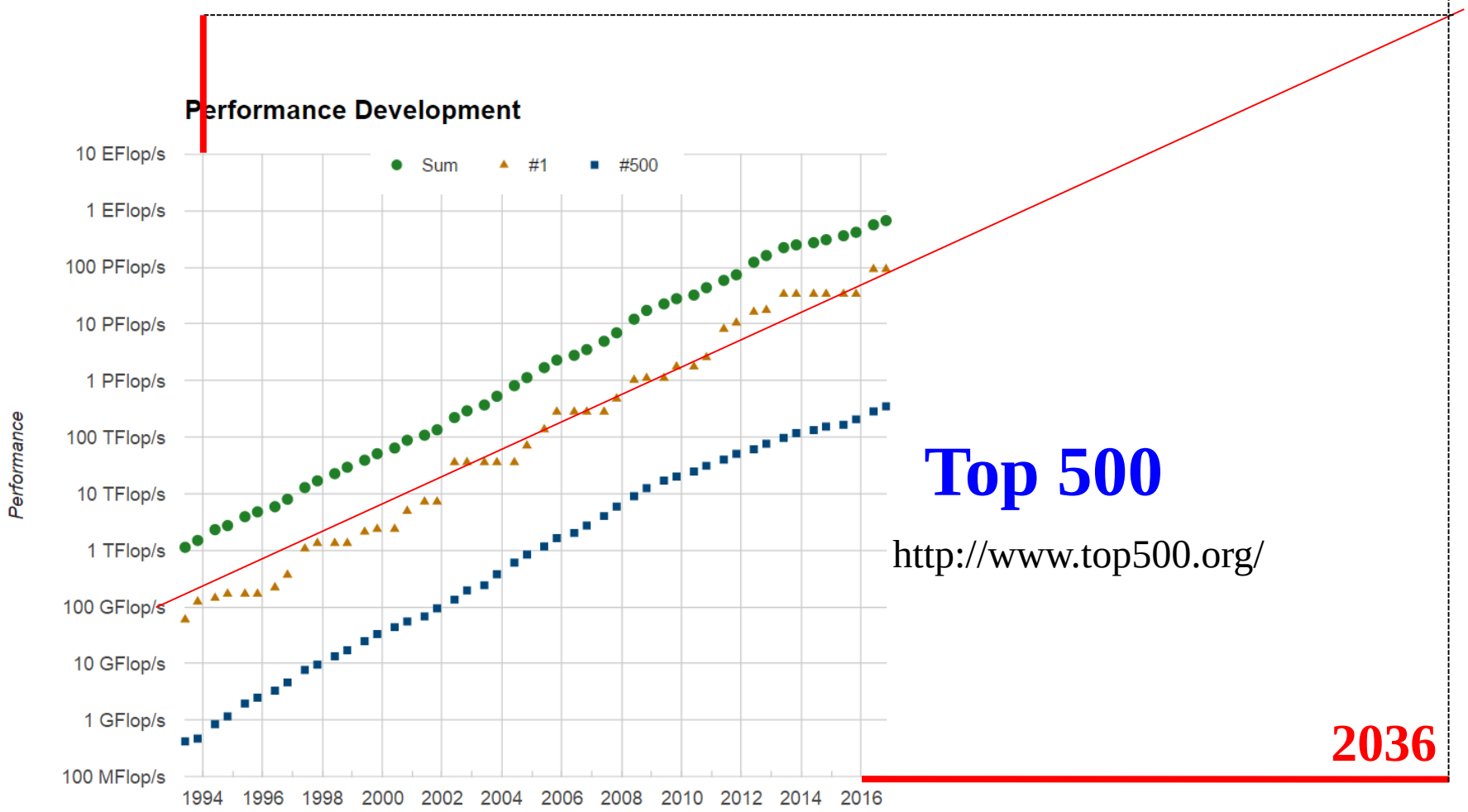
神威·太湖之光：125 Peta flops machine

Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway

NRCPC Cores: 10,649,600
Rmax: 93,014,593.9 (GFLOP/sec.)
Pmax: 125,435,904 (GFLOPS/sec.)

According to Moore's law…

1 Zeta Flops (Zeta = 10^{21})



How large systems can be treated by Zeta machines?

The performance increase is 10,000 times.

Sunway TaihuLight

0.1 Exa FLOPS

→

1 Z Flops machine

1000 Exa FLOPS

Computational Scaling $O(N^p)$	Computable size
7	3.7
6	4.6
5	6.3
4	10
3	22
2	100
1	10000

← DFT

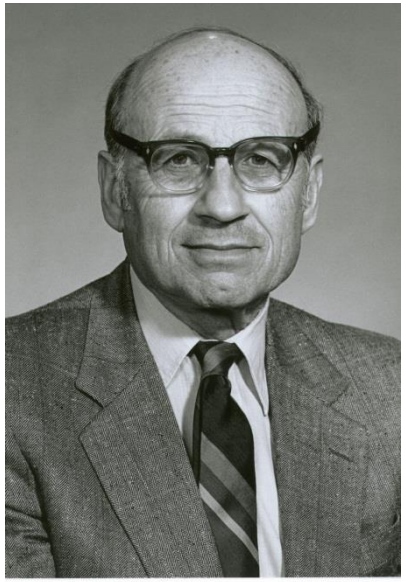
The applicability of the $O(N^3)$ DFT method is extended to only 22 times larger systems.

Linear scaling methods

Density functional theory

The energy of non-degenerate ground state can be expressed by a functional of electron density. (Hohenberg and Kohn, 1964)

$$E[\rho] = \int \rho(\mathbf{r}) v(\mathbf{r}) d + T[\rho] + J[\rho] + E_{\mathrm{xc}}[\rho]$$



W.Kohn (1923-)

The many body problem of the ground state can be reduced to an one-particle problem with an effective potential. (Kohn-Sham, 1965)

$$\hat{H}_{\mathrm{KS}} \phi_i = \varepsilon_i \phi_i$$

$$\hat{H}_{\mathrm{KS}} = -\frac{1}{2}\nabla^2 + v_{\mathrm{eff}}$$

$$v_{\mathrm{eff}} = v_{\mathrm{ext}}(\mathbf{r}) + v_{\mathrm{Hartree}}(\mathbf{r}) + \frac{\delta E_{\mathrm{xc}}}{\delta \rho(\mathbf{r})}$$

Mathematical structure of KS eq.

3D coupled non-linear differential equations have to be solved self-consistently.

$$\hat{H}_{\mathrm{KS}}\phi_i = \varepsilon_i \phi_i \qquad \hat{H}_{\mathrm{KS}} = -\frac{1}{2}\nabla^2 + v_{\mathrm{eff}}$$

O(N^3)

$$\rho(\mathbf{r}) = \sum_i^{\mathrm{occ}} \phi_i^*(\mathbf{r})\phi_i(\mathbf{r})$$

$$\nabla^2 v_{\mathrm{Hartree}}(\mathbf{r}) = -4\pi\rho(\mathbf{r})$$

O(Nlog(N))

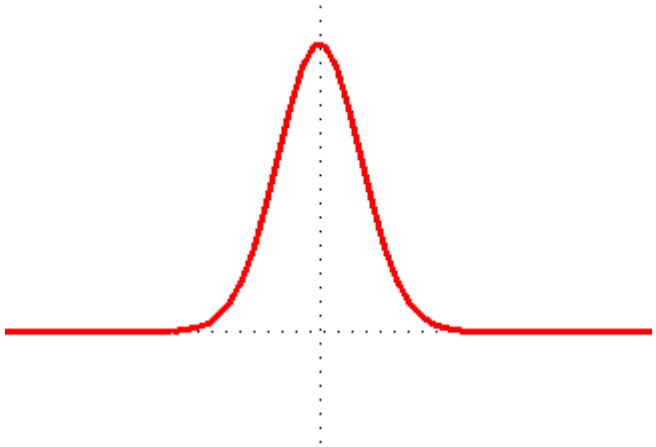
$$v_{\mathrm{eff}} = v_{\mathrm{ext}}(\mathbf{r}) + v_{\mathrm{Hartree}}(\mathbf{r}) + \frac{\delta E_{\mathrm{xc}}}{\delta\rho(\mathbf{r})}$$

Input charge = Output charge → Self-consistent condition

Localized vs. Delocalized

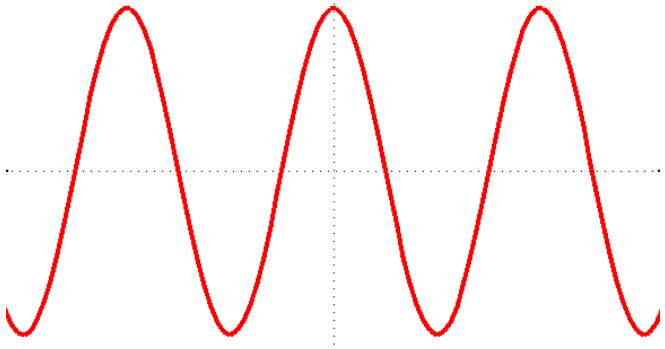
Is it possible to develop efficient DFT methods by making full use of the locality in the quantum physics ?

Localized



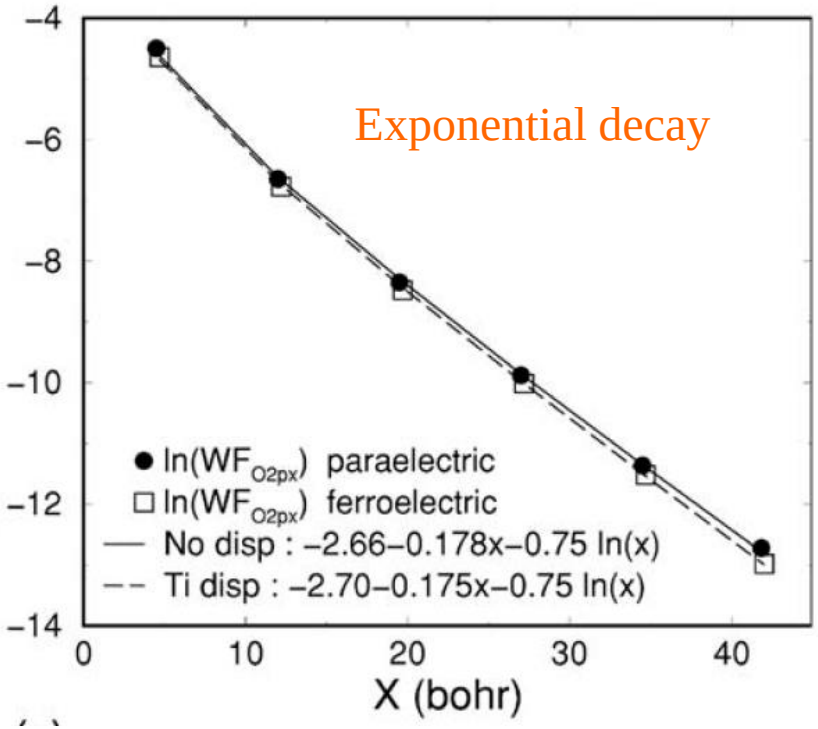
vs.

Delocalized

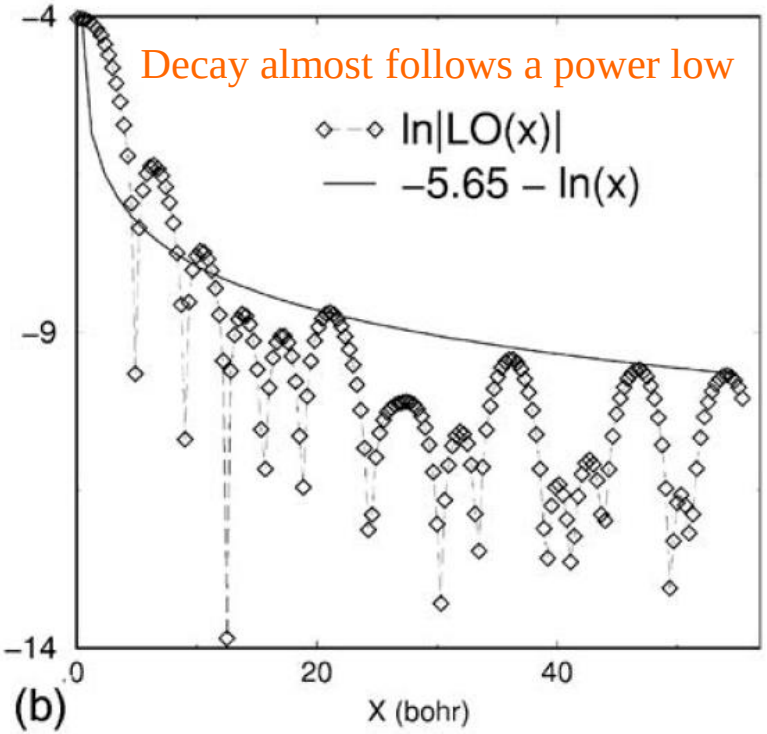


Locality of Wannier functions

O-2px in $PbTiO_3$

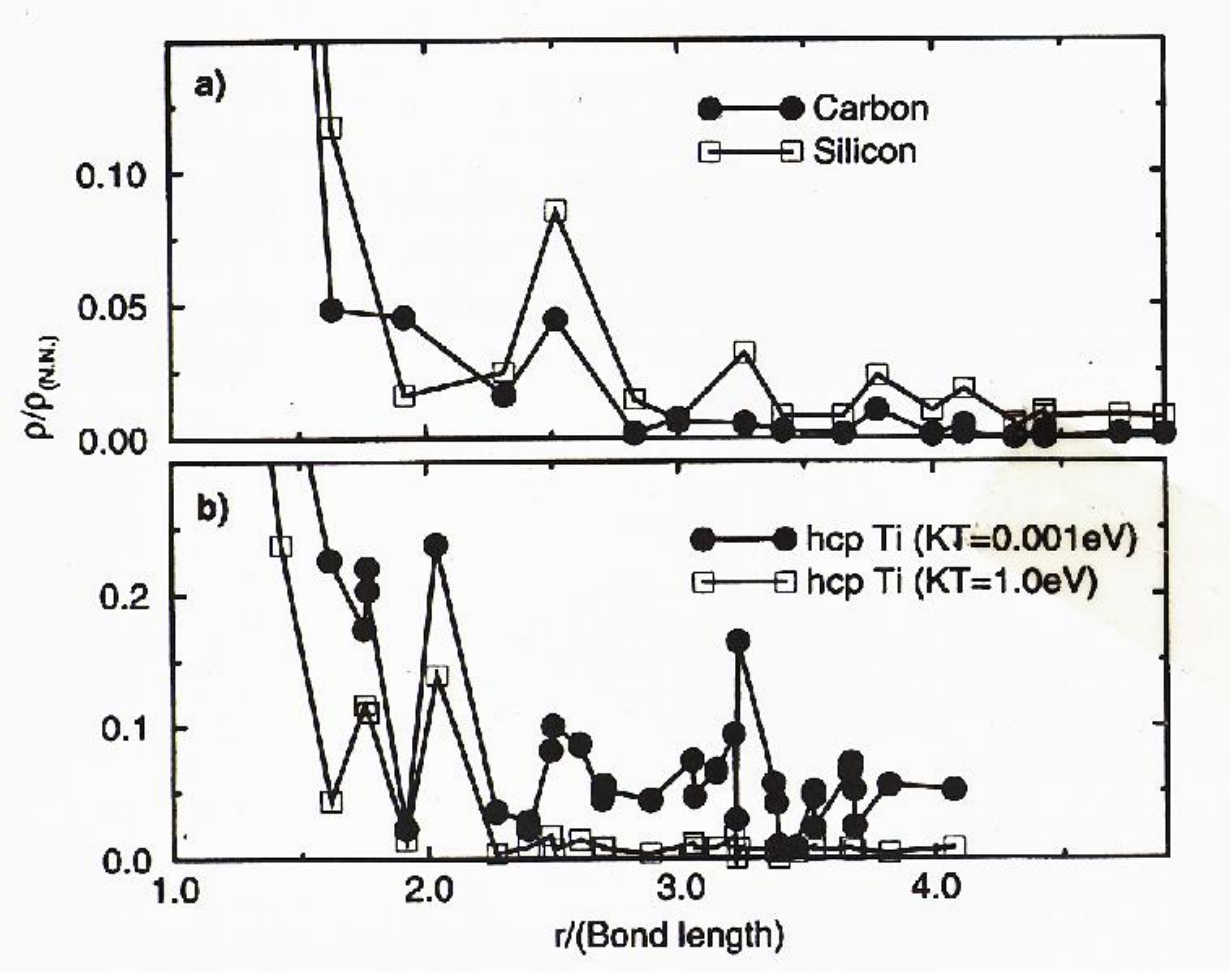


An orbital in Aluminum



J.Battacharjee and U.W.Waghmare, PRB 73, 121102 (2006)

Locality of density matrix



Finite gap systems
exponential decay

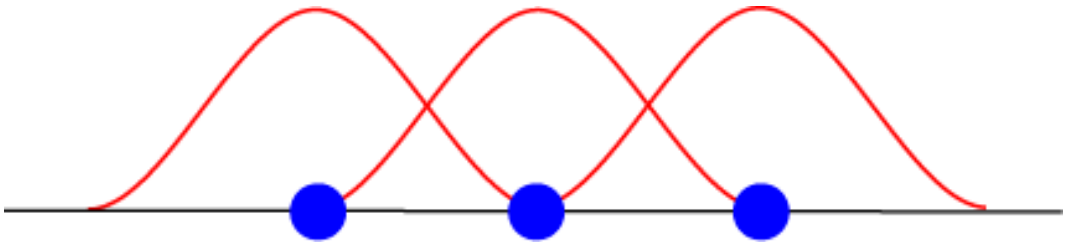
Metals
T=0 power law decay
0<T exponential decay

D.R.Bowler et al.,
Modell.Siml.Mater.Sci.Eng.5, 199 (1997)

Keys to large-scale DFT calculations

The locality of two quantities may lead to development of efficient large-scale electronic structure methods.

Basis function



Density matrix: ρ

Finite gap systems	$\Delta E \neq 0$	$\rho \propto \exp(-\alpha r)$	
Metals	$\Delta E = 0$	$\rho \propto r^{-\alpha p}$	for $T = 0$
		$\rho \propto \exp(-\alpha r)$	for $0 < T$

Ismail-Beigi and Arias, PRL 82, 2127 (1999).
Goededecker, PRB 58, 3501 (1998).

Density functionals as a functional of ρ

Density functionals can be rewritten by the first order reduced density matrix: ρ

$$E_{\mathrm{tot}}[n,\rho] = \mathrm{Tr}(\rho H_{\mathrm{kin}}) + \int d\mathbf{r} n(\mathbf{r}) v_{\mathrm{ext}}(\mathbf{r}) + \int\int d\mathbf{r} d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\mathrm{xc}}[n]$$

where the electron density is given by ρ

$$n(\mathbf{r}) = \sum_{i,j} \rho_{ij} \chi_j(\mathbf{r}) \chi_i(\mathbf{r})$$

Two routes towards O(N) DFT

Conventional representation

$$E_{\mathrm{tot}}(\{\Psi\}, \rho)$$

Density matrix representation

$$E_{\mathrm{tot}}(n, \rho)$$

Wannier function representation

$$E_{\mathrm{tot}}(\{\phi\}, \rho)$$

ψ: KS orbital
ρ: density
φ: Wannier function
n: density matrix

Various linear scaling methods

Wannier functions (WF)
Density matrix (DM)

$\times$

Variational (V)
Perturbative (P)

At least four kinds of linear-scaling methods can be considered as follows:

WF+V	**WF+P**	**DM+V**	**DM+P**
Orbital minimization by Galli, Parrinello, and Ordejon	Hoshi Mostofi	Density matrix by Li and Daw	**Krylov subspace** Divide-conquer Recursion Fermi operator

O(N) DFT codes

OpenMX: (Krylov) Ozaki (U. of Tokyo) et al.

Conquest: (DM) Bowler(London), Gillan(London), Miyazaki (NIMS)

Siesta: (OM) Ordejon et al.(Spain)

ONETEP: (DM) Hayne et al.(Imperial)

FEMTECK: (OM) Tsuchida (AIST)

FreeON: (DM) Challacombe et al.(Minnesota)

O(N) DFT method in OpenMX

1. Variationally optimized local orbitals

   - Reasonably accurate with relatively small # of functions
   - O(N) non-zero matrix elements
   - High compatibility with O(N) methods

2. O(N) Krylov subspace method for diagonalization

   - Numerically very robust
   - Applicable to insulators and metals
   - Suitable for parallel computation

O(N) DFT method in OpenMX

1. Variationally optimized local orbitals

   - Reasonably accurate with relatively small # of functions
   - O(N) non-zero matrix elements
   - High compatibility with O(N) methods

2. O(N) Krylov subspace method for diagonalization

   - Numerically very robust
   - Applicable to insulators and metals
   - Suitable for parallel computation

LCPAO method
(Linear-Combination of Pseudo Atomic Orbital Method)

One-particle KS orbital

$$\psi_{\sigma\mu}^{(\mathbf{k})}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathrm{n}}^{N} \mathrm{e}^{\mathrm{i}\mathbf{R}_{\mathrm{n}}\cdot\mathbf{k}} \sum_{i\alpha} c_{\sigma\mu,i\alpha}^{(\mathbf{k})} \phi_{i\alpha}(\mathbf{r} - \tau_i - \mathbf{R}_{\mathrm{n}}),$$

is expressed by a linear combination of atomic like orbitals in the method.

$$\phi(\mathbf{r}) = Y_l^m(\hat{\mathbf{r}}) R(r)$$

Features:

- It is easy to interpret physical and chemical meanings, since the KS orbitals are expressed by the atomic like basis functions.
- It gives rapid convergent results with respect to basis functions due to physical origin
- The memory and computational effort for calculation of matrix elements are O(N).
- It well matches the idea of linear scaling methods.

Variational optimization of basis functions

One-particle wave functions

$$\psi_\mu(\mathbf{r}) = \sum_{i\alpha} c_{\mu,i\alpha}\phi_{i\alpha}(\mathbf{r}-\mathbf{r}_i)$$

Contracted orbitals

$$\phi_{i\alpha}(\mathbf{r}) = \sum_q a_{i\alpha q}\chi_{i\eta}(\mathbf{r})$$

The variation of E with respect to *c* with fixed *a* gives

$$\partial E_{\mathrm{tot}}/\partial c_{\mu,i\alpha} = 0 \quad \longrightarrow \quad \sum_{j\beta} \langle \phi_{i\alpha}|\hat{H}|\phi_{j\beta}\rangle c_{\mu,j\beta} = \varepsilon_\mu \sum_{j\beta} \langle \phi_{i\alpha}|\phi_{j\beta}\rangle c_{\mu,j\beta}$$

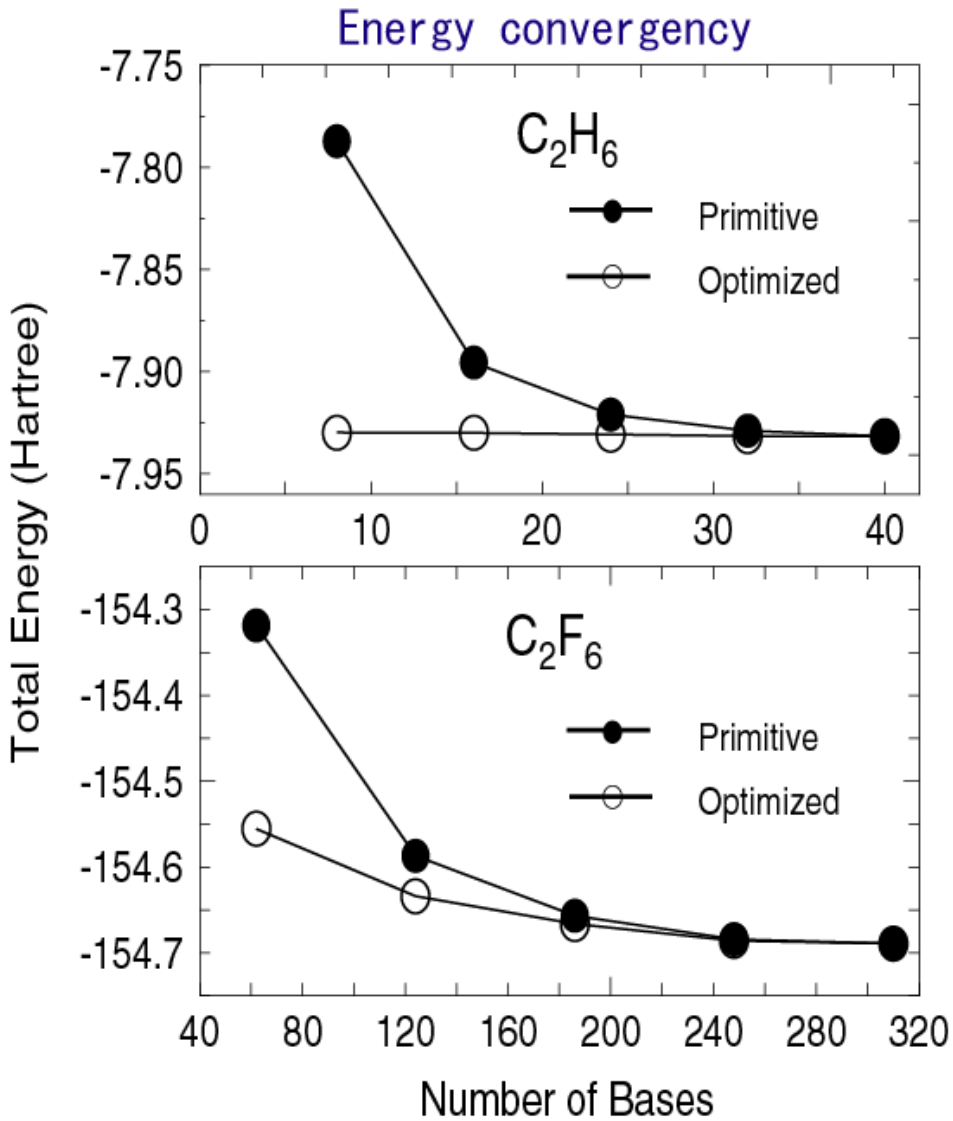
Regarding c as dependent variables on *a* and assuming KS eq. is solved self-consistently with respect to *c*, we have

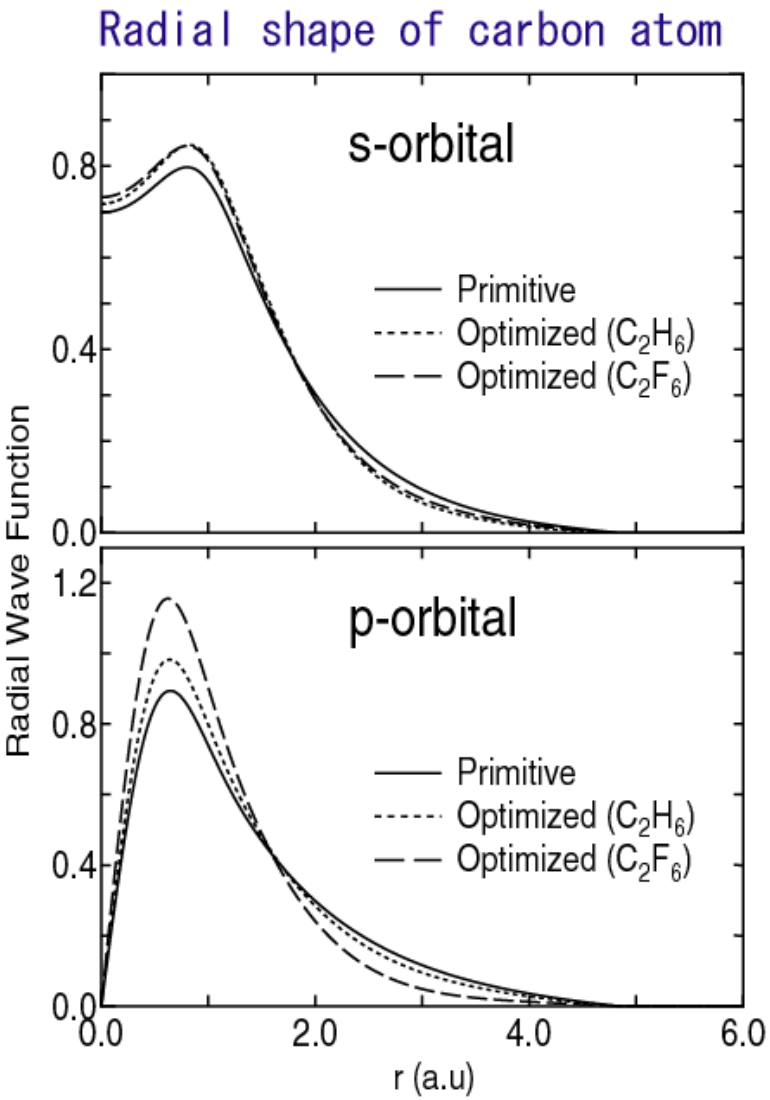
$$\frac{\partial E_{\mathrm{tot}}}{\partial a_{i\alpha q}} = \frac{\delta E_{\mathrm{tot}}}{\delta \rho(\mathbf{r})}\frac{\delta \rho(\mathbf{r})}{\delta a_{i\alpha q}}$$

$$= 2\sum_{j\beta} \left(\Theta_{i\alpha,j\beta}\langle \chi_{i\eta}|\hat{H}|\phi_{j\beta}\rangle - E_{i\alpha,j\beta}\langle \chi_{i\eta}|\phi_{j\beta}\rangle\right)$$

Ozaki, PRB 67, 155108 (2003)

Comparison between primitive and optimized basis functions





Ozaki, PRB 67, 155108 (2003).

Reproducibility in DFT calcs

RESEARCH ARTICLE

Science 351, aad3000 (2016)

DFT METHODS

Reproducibility in density functional theory calculations of solids

Kurt Lejaeghere,[1*] Gustav Bihlmayer,[2] Torbjörn Björkman,[3,4] Peter Blaha,[5] Stefan Blügel,[2] Volker Blum,[6] Damien Caliste,[7,8] Ivano E. Castelli,[9] Stewart J. Clark,[10] Andrea Dal Corso,[11] Stefano de Gironcoli,[11] Thierry Deutsch,[7,8] John Kay Dewhurst,[12] Igor Di Marco,[13] Claudia Draxl,[14,15] Marcin Dułak,[16] Olle Eriksson,[13] José A. Flores-Livas,[12] Kevin F. Garrity,[17] Luigi Genovese,[7,8] Paolo Giannozzi,[18] Matteo Giantomassi,[19] Stefan Goedecker,[20] Xavier Gonze,[19] Oscar Grånäs,[13,21] E. K. U. Gross,[12] Andris Gulans,[14,15] François Gygi,[22] D. R. Hamann,[23,24] Phil J. Hasnip,[25] N. A. W. Holzwarth,[26] Diana Iuşan,[13] Dominik B. Jochym,[27] François Jollet,[28] Daniel Jones,[29] Georg Kresse,[30] Klaus Koepernik,[31,32] Emine Küçükbenli,[9,11] Yaroslav O. Kvashnin,[13] Inka L. M. Locht,[13,33] Sven Lubeck,[14] Martijn Marsman,[30] Nicola Marzari,[9] Ulrike Nitzsche,[31] Lars Nordström,[13] Taisuke Ozaki,[34] Lorenzo Paulatto,[35] Chris J. Pickard,[36] Ward Poelmans,[1,37] Matt I. J. Probert,[25] Keith Refson,[38,39] Manuel Richter,[31,32] Gian-Marco Rignanese,[19] Santanu Saha,[20] Matthias Scheffler,[15,40] Martin Schlipf,[22] Karlheinz Schwarz,[5] Sangeeta Sharma,[12] Francesca Tavazza,[17] Patrik Thunström,[41] Alexandre Tkatchenko,[15,42] Marc Torrent,[28] David Vanderbilt,[23] Michiel J. van Setten,[19] Veronique Van Speybroeck,[1] John M. Wills,[43] Jonathan R. Yates,[29] Guo-Xu Zhang,[44] Stefaan Cottenier[1,45*]

15 codes
69 researchers
71 elemental bulks
GGA-PBE
Scalar relativistic

Comparison of codes by Δ-gauge

	average <Δ>	Elk	exciting	FHI-aims/tier2	FLEUR	FPLO/T+F+s	RSPt	WIEN2k/acc
AE								
Elk	0.6		0.3	0.3	0.6	1.0	0.9	0.3
exciting	0.5	0.3		0.1	0.5	0.9	0.8	0.2
FHI-aims/tier2	0.5	0.3	0.1		0.5	0.8	0.8	0.2
FLEUR	0.6	0.6	0.5	0.5		0.8	0.6	0.4
FPLO/T+F+s	0.9	1.0	0.9	0.9	0.8		0.9	0.9
RSPt	0.8	0.9	0.8	0.8	0.6	0.9		0.8
WIEN2k/acc	0.5	0.3	0.2	0.2	0.4	0.9	0.8	
PAW								
GBRV12/ABINIT	0.9	0.9	0.8	0.8	0.9	1.3	1.1	0.8
GPAW09/ABINIT	1.4	1.3	1.3	1.3	1.3	1.7	1.5	1.3
GPAW09/GPAW	1.6	1.5	1.5	1.5	1.5	1.8	1.7	1.5
JTH02/ABINIT	0.6	0.6	0.6	0.6	0.6	0.9	0.7	0.5
PSlib100/QE	0.9	0.9	0.8	0.8	0.8	1.3	1.1	0.8
VASPGW2015/VASP	0.6	0.4	0.4	0.4	0.6	1.0	0.8	0.3
USPP								
GBRV14/CASTEP	1.1	1.1	1.1	1.0	1.0	1.4	1.3	1.0
GBRV14/QE	1.1	1.0	1.0	0.9	1.0	1.4	1.3	1.0
OTFG9/CASTEP	0.7	0.4	0.5	0.5	0.7	1.0	1.0	0.5
SSSP/QE	0.5	0.4	0.3	0.3	0.5	0.9	0.8	0.3
Vdb2/DACAPO	6.3	6.3	6.3	6.3	6.3	6.4	6.5	6.2
NCPP								
FHI98pp/ABINIT	13.3	13.5	13.4	13.4	13.2	13.0	13.2	13.4
HGH/ABINIT	2.2	2.2	2.2	2.2	2.0	2.3	2.2	2.1
HGH-NLCC/BigDFT	1.1	1.1	1.1	1.1	1.0	1.2	1.1	1.0
MBK2013/OpenMX	2.0	2.1	2.1	2.1	1.9	1.8	1.8	2.0
ONCVPSP(PD0.1)/ABINIT	0.7	0.7	0.7	0.7	0.6	1.0	0.8	0.6
ONCVPSP(SG15)1/QE	1.4	1.4	1.3	1.3	1.3	1.6	1.5	1.3
ONCVPSP(SG15)2/CASTEP	1.4	1.4	1.4	1.4	1.3	1.6	1.5	1.4

The mean Δ-gauge of OpenMX is 2.0meV/atom.

O(N) DFT method in OpenMX

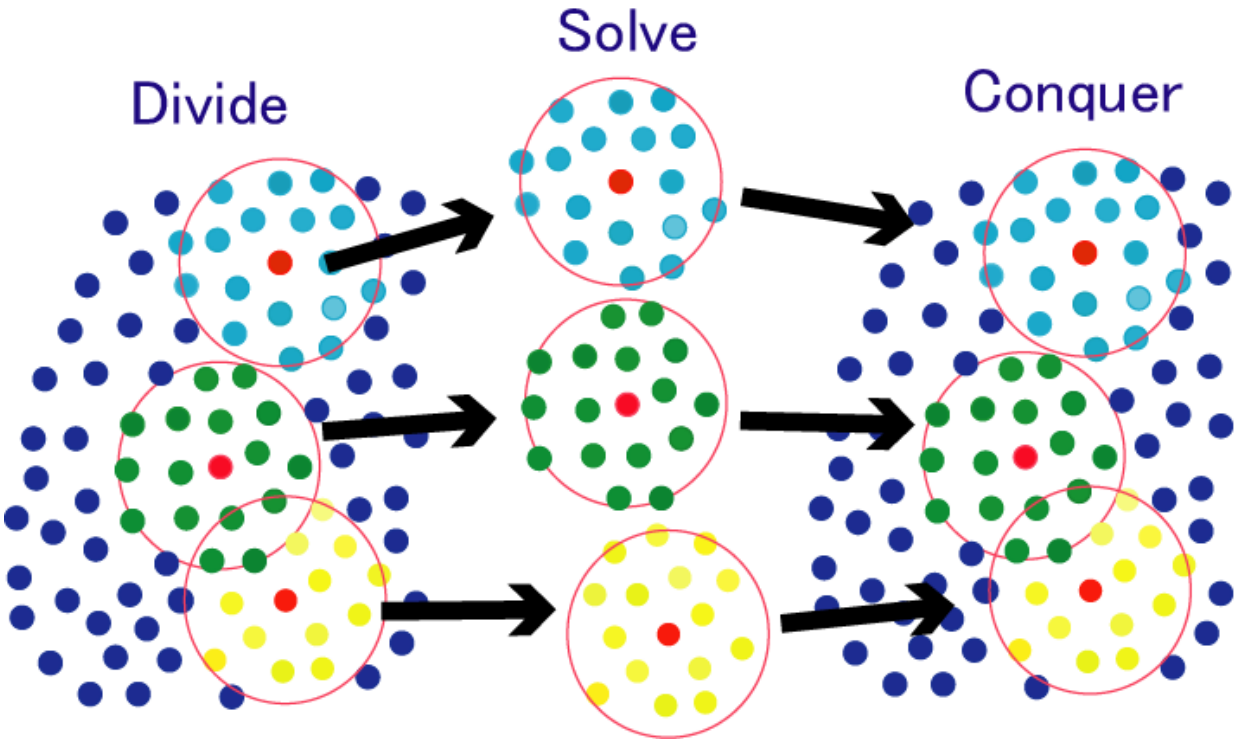
1. Variationally optimized local orbitals

   - Reasonably accurate with relatively small # of functions
   - O(N) non-zero matrix elements
   - High compatibility with O(N) methods

2. O(N) Krylov subspace method for diagonalization

   - Numerically very robust
   - Applicable to insulators and metals
   - Suitable for parallel computation

Basic idea behind the O(N) method



Assumption

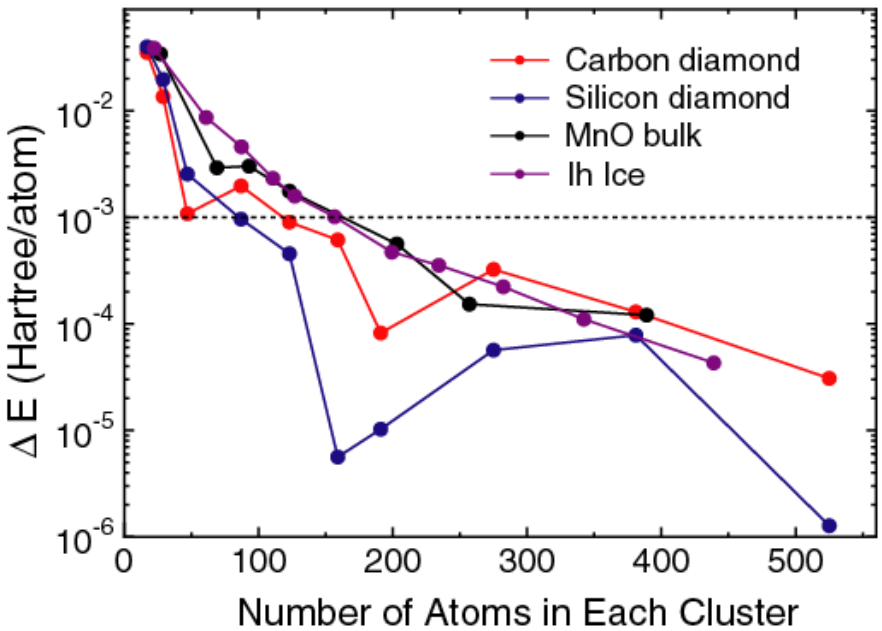
Local electronic structure of each atom is mainly determined by neighboring atomic arrangement producing chemical environment.

Convergence by the DC method

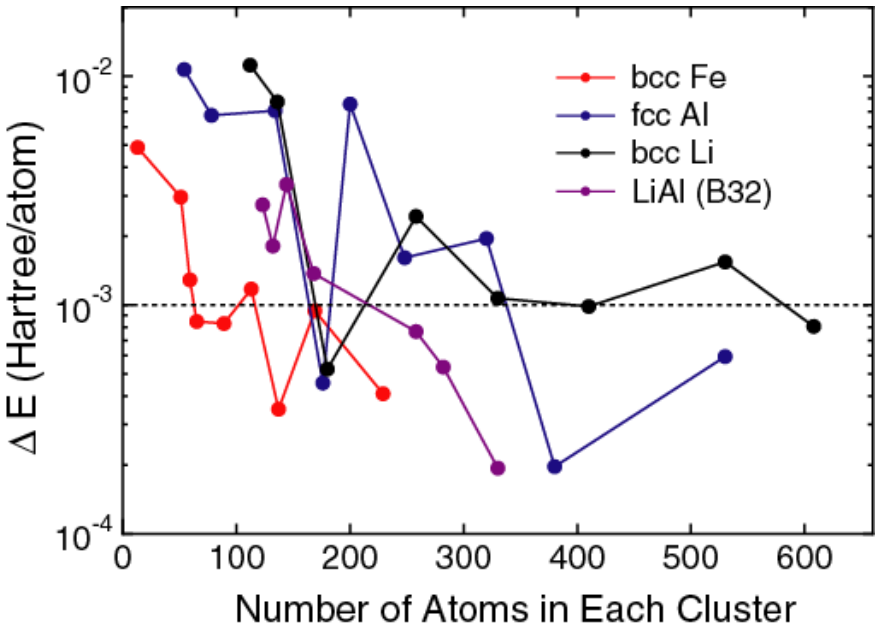
Just solve the truncated clusters → Divide-Conquer method

W.Yang, PRL 66, 1438 (1991)

Insulators, semi-conductors



Metals

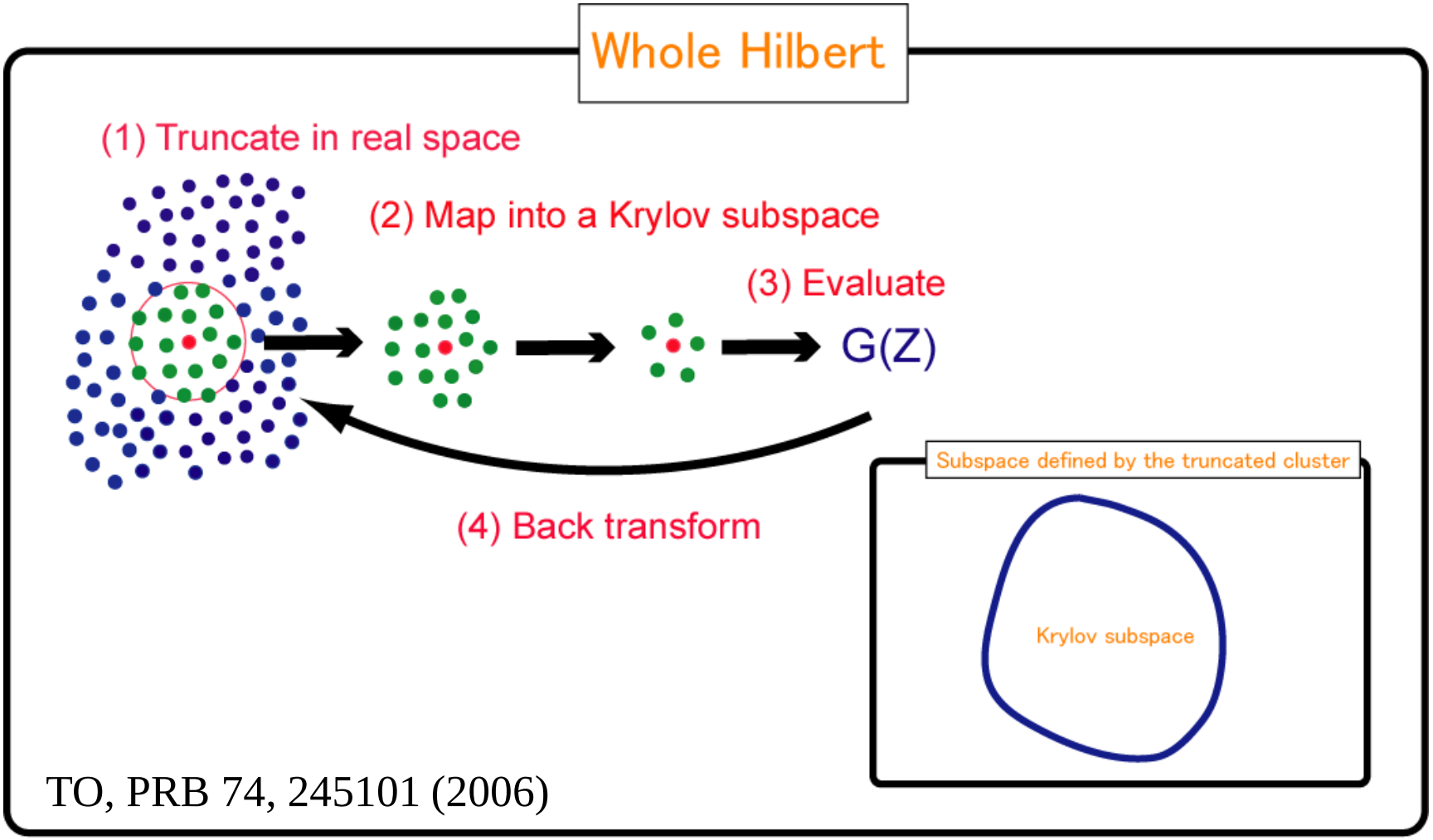


For metals, a large cluster size is required for the convergence.

→ Difficult for direct application of the DC method for metals

O(N) Krylov subspace method

Two step mapping of the whole Hilbert space into subspaces

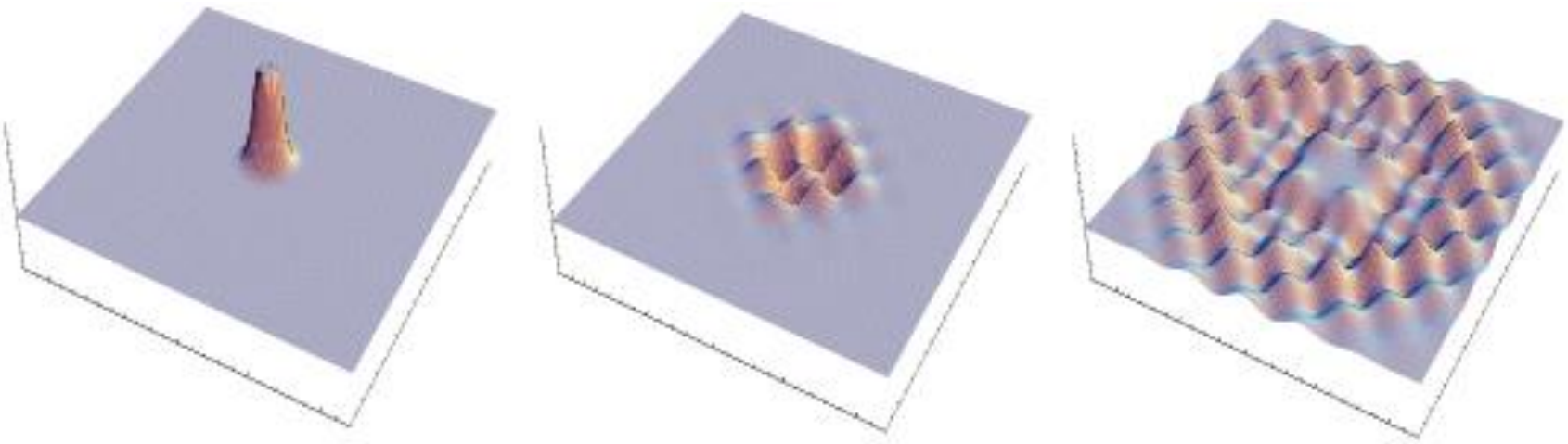


TO, PRB 74, 245101 (2006)

Development of Krylov subspace vectors

The Krylov vector is generated by a multiplication of H by |K>, and the development of the Krylov subspace vectors can be understood as hopping process of electron.

$|K_0\rangle$ **$|K_1\rangle$** **$|K_5\rangle$**



The information on *environment* can be included from near sites step by step, resulting in reduction of the dimension.

Generation of Krylov subspaces

The ingredients of generation of Krylov subspaces is to multiply $|W_n)$ by $S^{-1}H$. The other things are made only for stabilization of the calculation.

$$
\begin{aligned}
|R_{n+1}) &= S^{-1}H|W_n) \\
|W'_{n+1}) &= |R_{n+1}) - \sum_{m=0}^{n} |W_m)(W_m|\hat{S}|R_{n+1}) \\
|W_{n+1}) &= S-\text{orthonormalized block vector of } |W'_{n+1})
\end{aligned}
$$

Furthermore, in order to assure the S-orthonormality of the Krylov subspace vectors, an orthogonal transformation is performed by

$$\mathbf{U}_{\mathrm{K}} = \mathbf{W}\mathbf{X}\lambda^{-1}$$

$$\lambda^2 = \mathbf{X}^\dagger \mathbf{W}^\dagger \hat{S} \mathbf{W} \mathbf{X}$$

For numerical stability, it is crucial to generate the Krylov subspace at the first SCF step.

Embedded cluster problem

Taking the Krylov subspace representation, the cluster eigenvalue problem is transformed to a standard eigenvalue problem as:

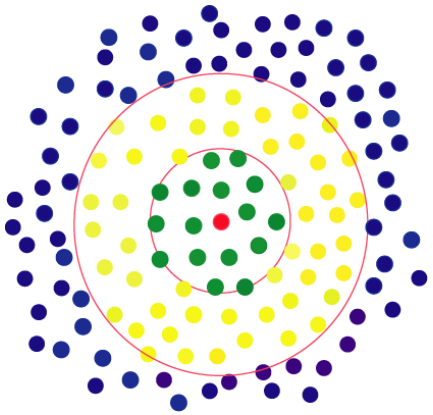
$$H c_\mu = \varepsilon_\mu S c_\mu \longrightarrow H^{\mathrm{K}} b_\mu = \varepsilon b_\mu$$

where H^{K} consists of the short and long range contributions.

$$\begin{aligned} H^{\mathrm{K}} &= U^\dagger H U \\ &= \underbrace{u_c^\dagger H_c u_c}_{\text{updated}} + \underbrace{u_c^\dagger H_{cb}^\dagger u_b + u_b^\dagger H_{bc} u_c + u_b^\dagger H_b u_b}_{\text{fixed}} \\ &= H_s^{\mathrm{K}} + H_l^{\mathrm{K}} \end{aligned}$$

Green: core region

Yellow: buffer region



- **The embedded cluster is under the Coulomb interaction from the other parts.**
- **The charge flow from one embedded cluster to the others is allowed.**

Relation between the Krylov subspace and Green's funtion

A Krylov subspace is defined by

$$\mathbf{U}_{\mathrm{K}} = \left\{|W_0), (S^{-1}H)|W_0), (S^{-1}H)^2|W_0), ..., (S^{-1}H)^q|W_0)\right\}$$

A set of q-th Krylov vectors contains up to information of (2q+1)th moments.

$$\begin{aligned}\underline{\mathrm{H}}^K_{mn} &= (W_0|(A^\dagger)^m H A^n|W_0) \\ &= (W_0|S(S^{-1}H)^{m+n+1}|W_0), \\ &= (W_0|S\mu^{(m+n+1)}S|W_0)\end{aligned}$$

Definition of moments

$$\begin{aligned}\mu^{(p)} &= c\varepsilon^p c^\dagger, \\ &= cc^\dagger Hcc^\dagger Hc \cdots c^\dagger Hcc^\dagger, \\ &= (S^{-1}H)^p S^{-1}\end{aligned}$$

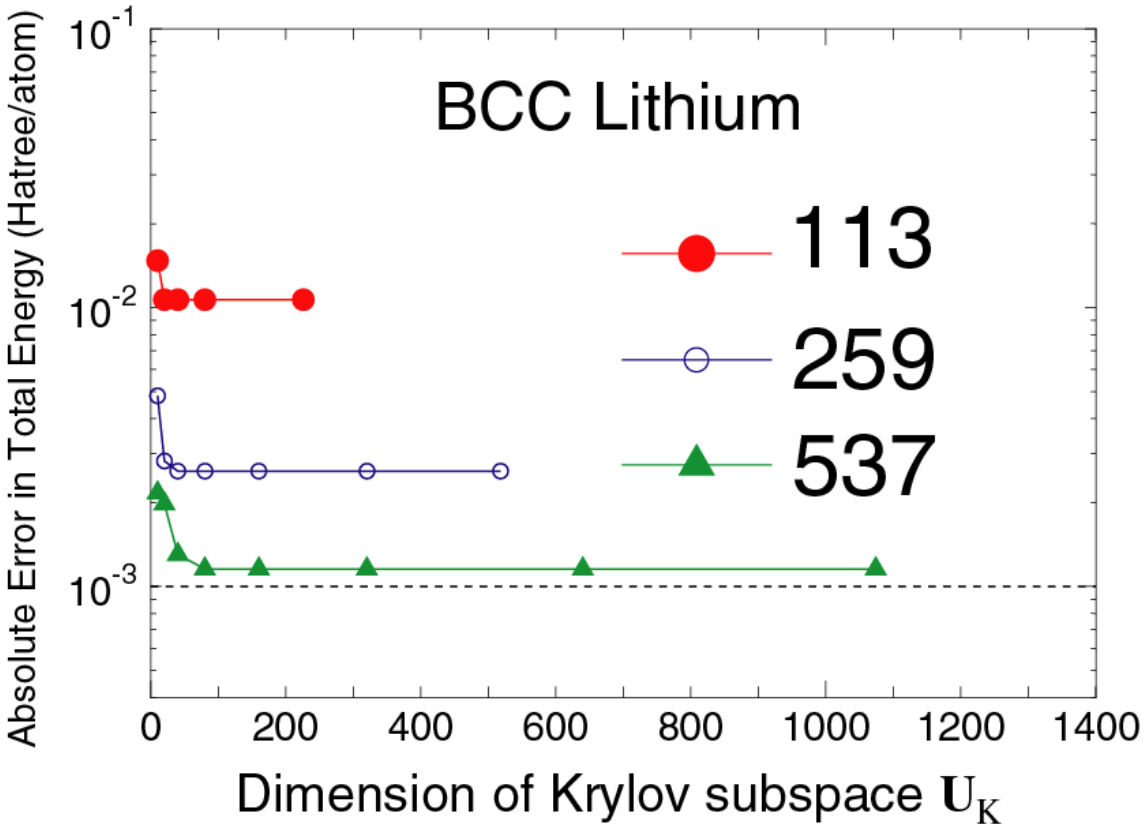
The moment representation of G(Z) gives us the relation.

$$G_{ij}(Z) = \sum_{p=0}^{\infty} \frac{\mu_{ij}^{(p)}}{Z^{p+1}}$$

One-to-one correspondence between the dimension of Krylov subspace and the order of moments can be found from above consideration.

Convergence property

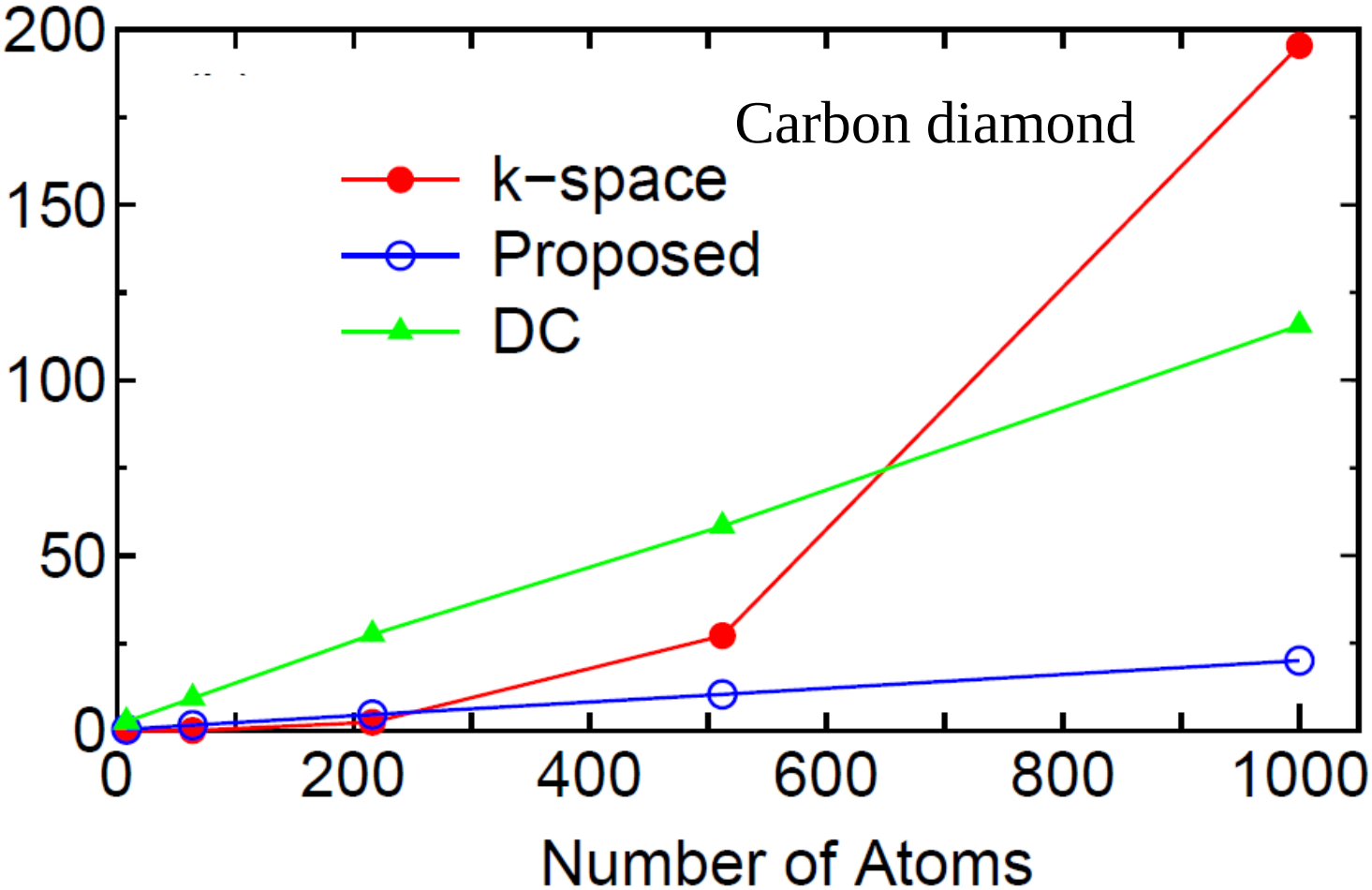
The accuracy and efficiency can be controlled by the size of truncated cluster and dimension of Krylov subspace.



In general, the convergence property is more complicated. See PRB 74, 245101 (2006).

Comparison of computational time

The computational time of calculation for each cluster does not depend on the system size. Thus, the computational time is O(N) in principle.



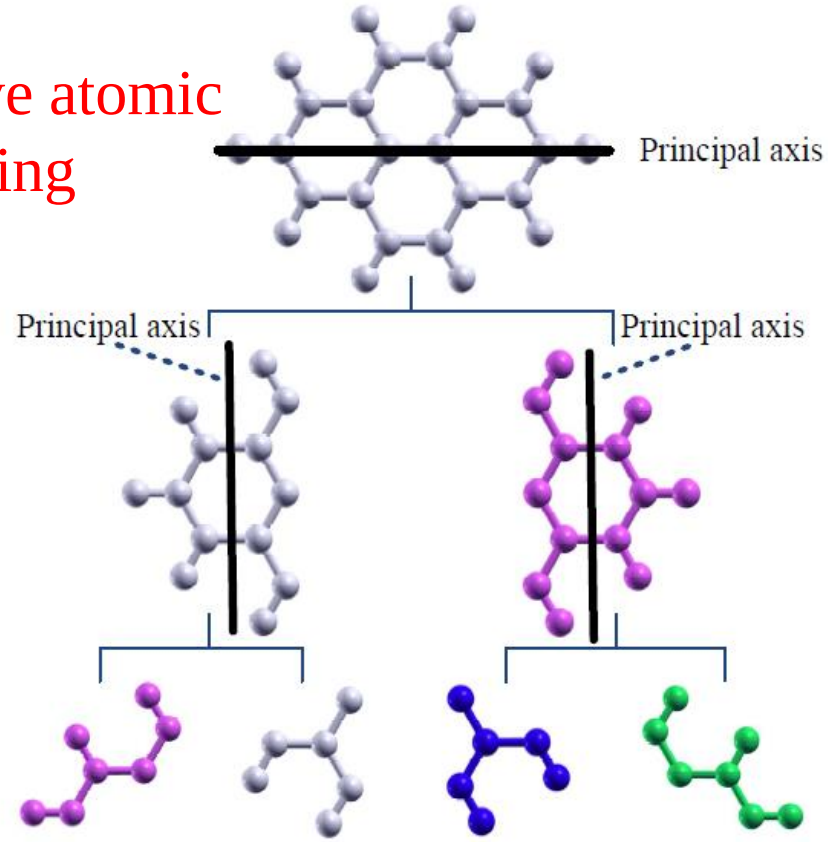
Parallelization

How one can partition atoms to minimize communication and memory usage?

Requirement:

- Locality
- Same computational cost
- Applicable to any systems
- Small computational overhead

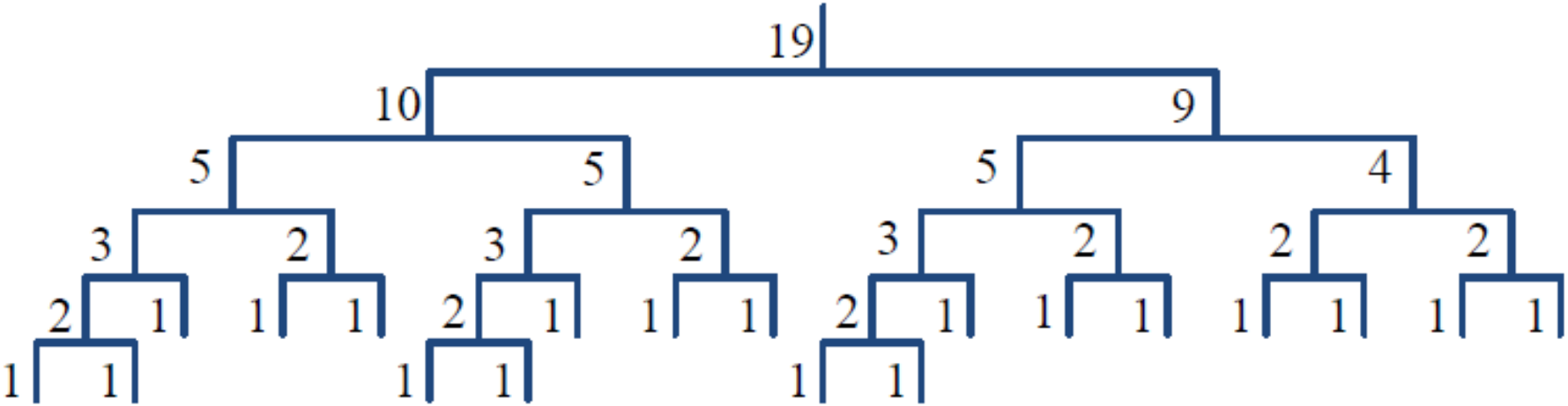
Recursive atomic partitioning



T.V.T. Duy and T. Ozaki, CPC 185, 777 (2014).

Modified recursive bisection

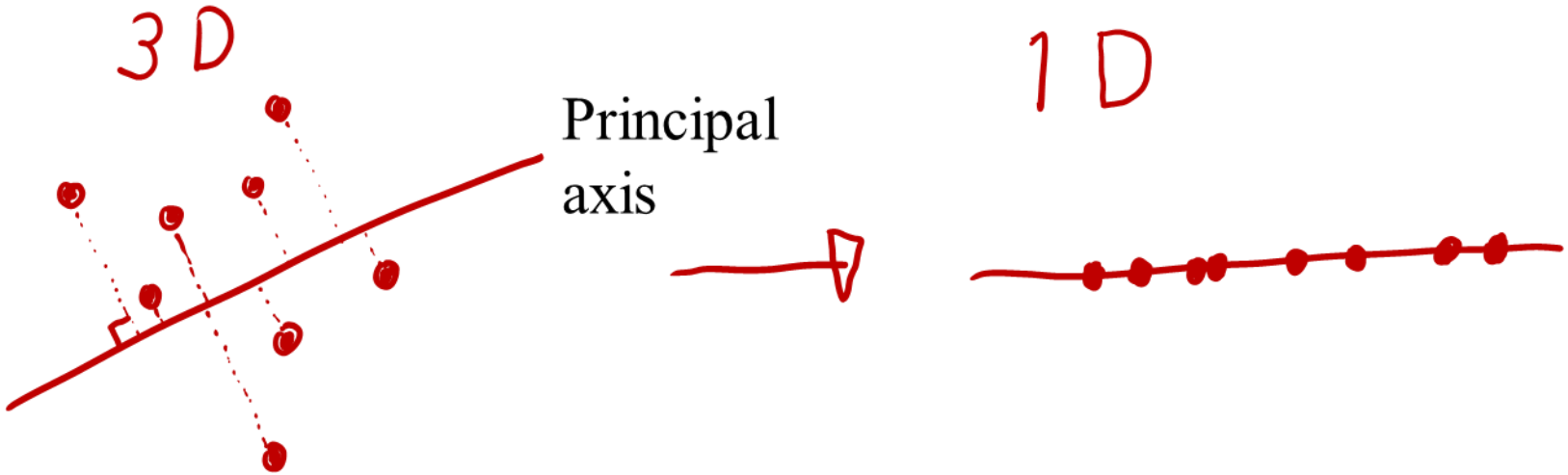
If the number of MPI processes is 19, then the following binary tree structure is constructed.



In the conventional recursive bisection, the bisection is made so that a same number can be assigned to each region. However, the modified version bisects with weights as shown above.

Reordering of atoms by an inertia tensor

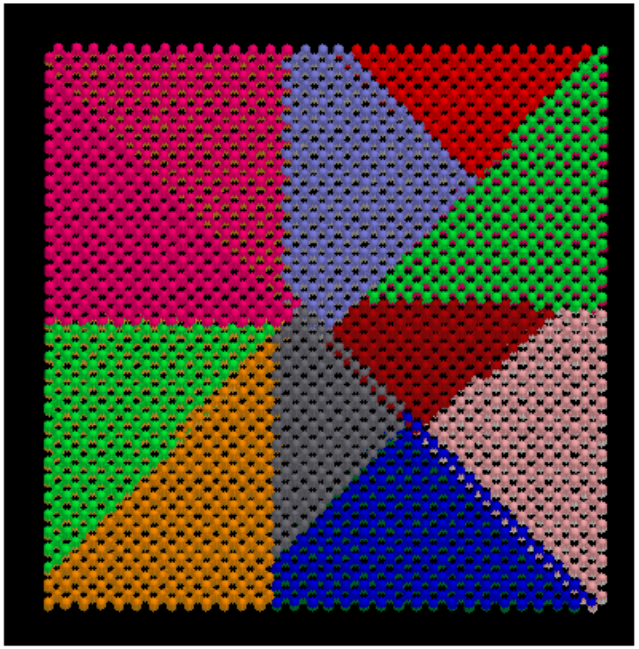
Atoms in an interested region are reordered by projecting them onto a principal axis calculated by an inertia tensor.



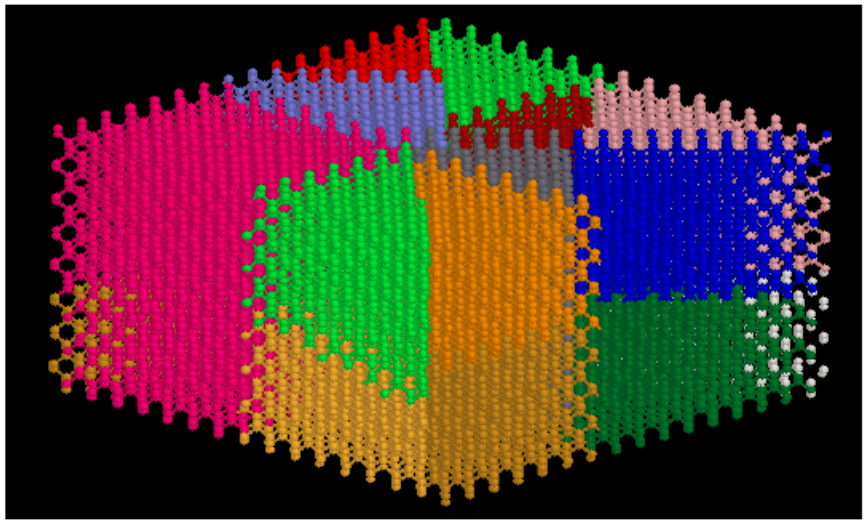
The principal axis is calculated by solving an eigenvalue problem with an inertia tensor:

$$\begin{pmatrix} \sum_i w_i (Y_i^2 + Z_i^2) & -\sum_i w_i X_i Y_i & -\sum_i w_i X_i Z_i \\ -\sum_i w_i Y_i X_i & \sum_i w_i (X_i^2 + Z_i^2) & -\sum_i w_i Y_i Z_i \\ -\sum_i w_i Z_i X_i & -\sum_i w_i Z_i Y_i & -\sum_i w_i (X_i^2 + Y_i \end{pmatrix} \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} = -\lambda \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix}$$

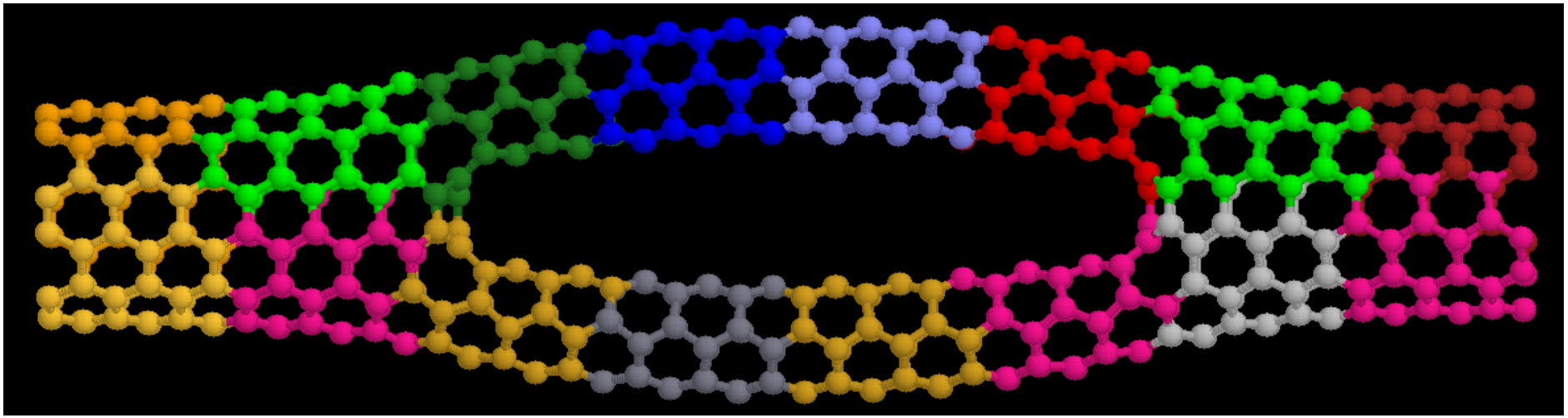
Allocation of atoms to processes



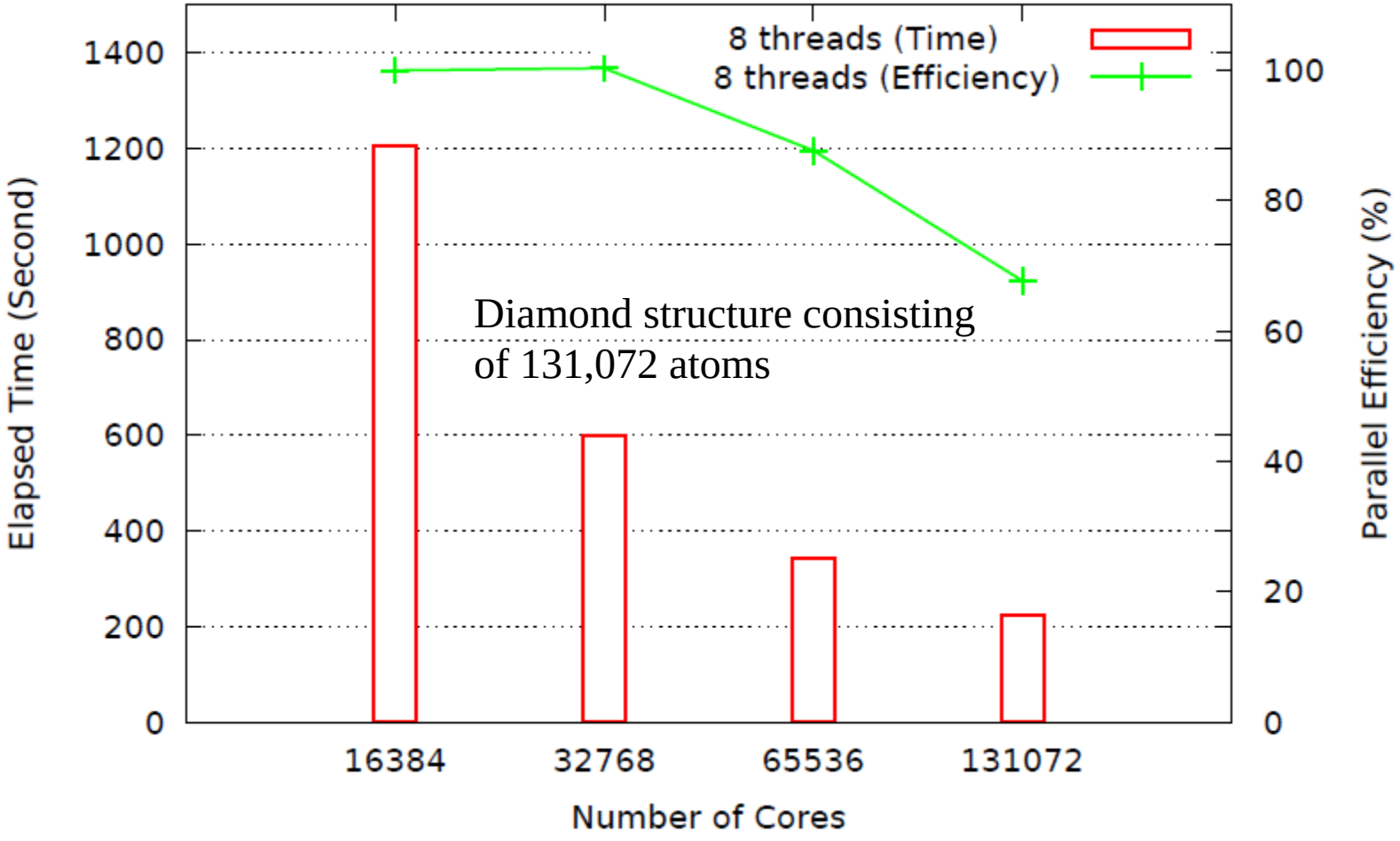
Diamond 16384 atoms, 19 processes



Multiply connected CNT, 16 processes



Parallel efficiency on K



The parallel efficiency is 68 % using 131,072 cores.

Applications of the O(N) method

1. Interface structure between BCC Iron and carbides

   H. Sawada et al., Modelling Simul. Mater. Sci. Eng. 21, 045012 (2013).

2. Desolvation of Li^+

   T. Ohwaki et al., J. Chem. Phys. 136, 134101 (2012).
   T. Ohwaki et al., J. Chem. Phys. 140, 244105 (2014).

3. Electronic transport of graphene nanoribbon

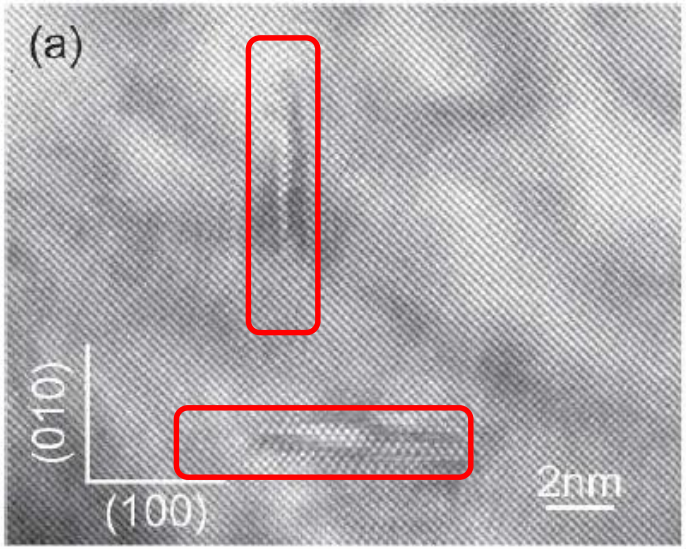
   M. Ohfuchi et al., Appl. Phys. Express 7, 025101 (2014).
   H Jippo, T Ozaki, S Okada, M Ohfuchi, J. Appl. Phys. 120, 154301 (2016).

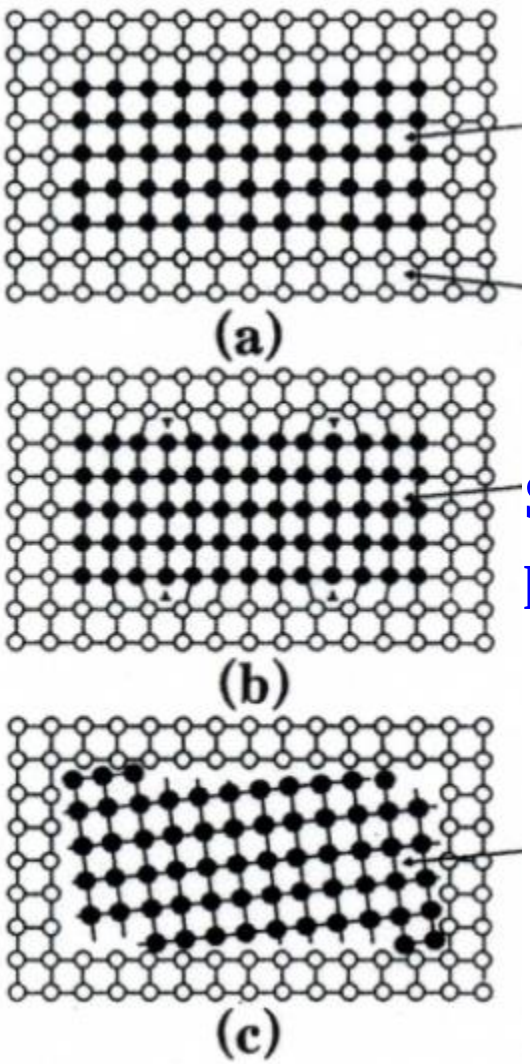
Precipitation in bcc-Fe

In collaboration wit Dr. Sawada (Nippon Steel)

Pure iron is too soft as structural material. Precipitation of carbide can be used to control the hardness of iron.

HRTEM image





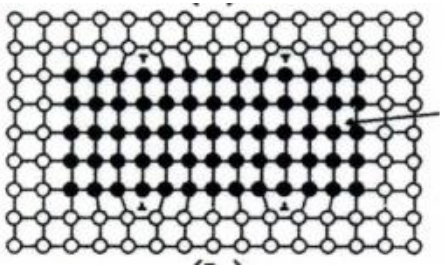
Coherent precipitation

Semicoherent precipitation

Incoherent precipitation

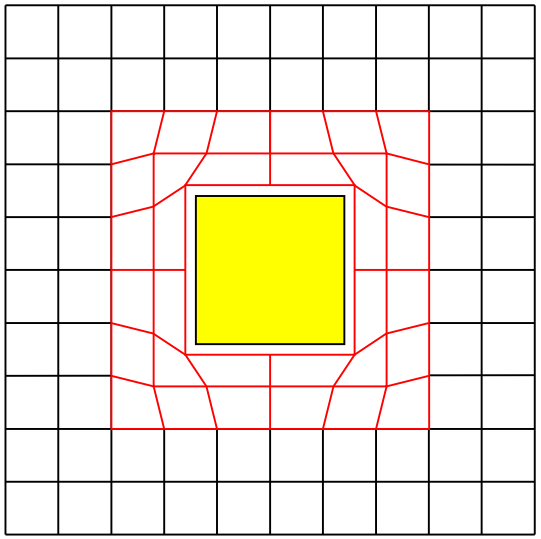
Precipitating materials:
TiC, VC, NbC

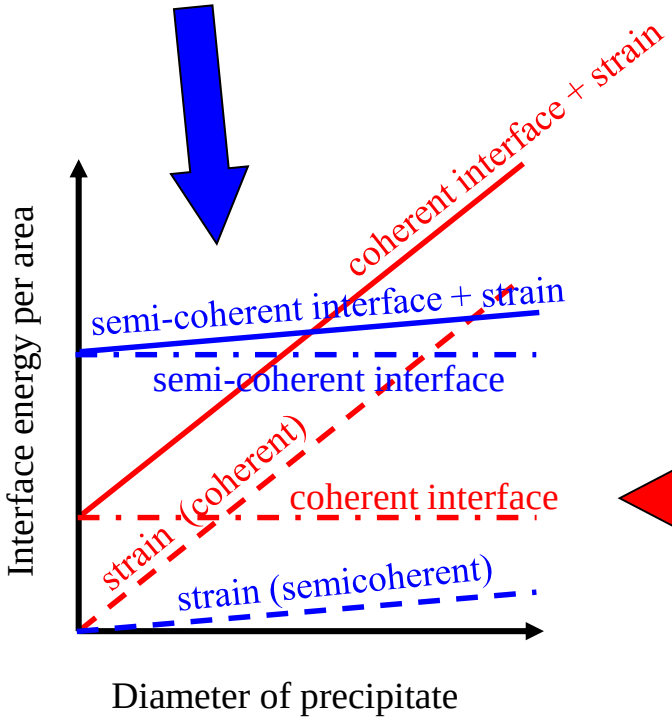
Interface and strain energies

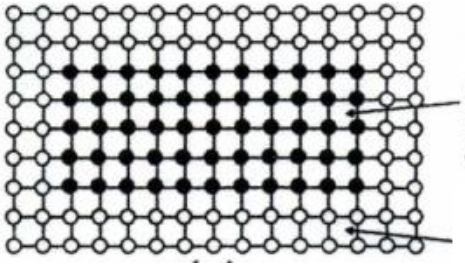


Semi-coherent case

Strain field





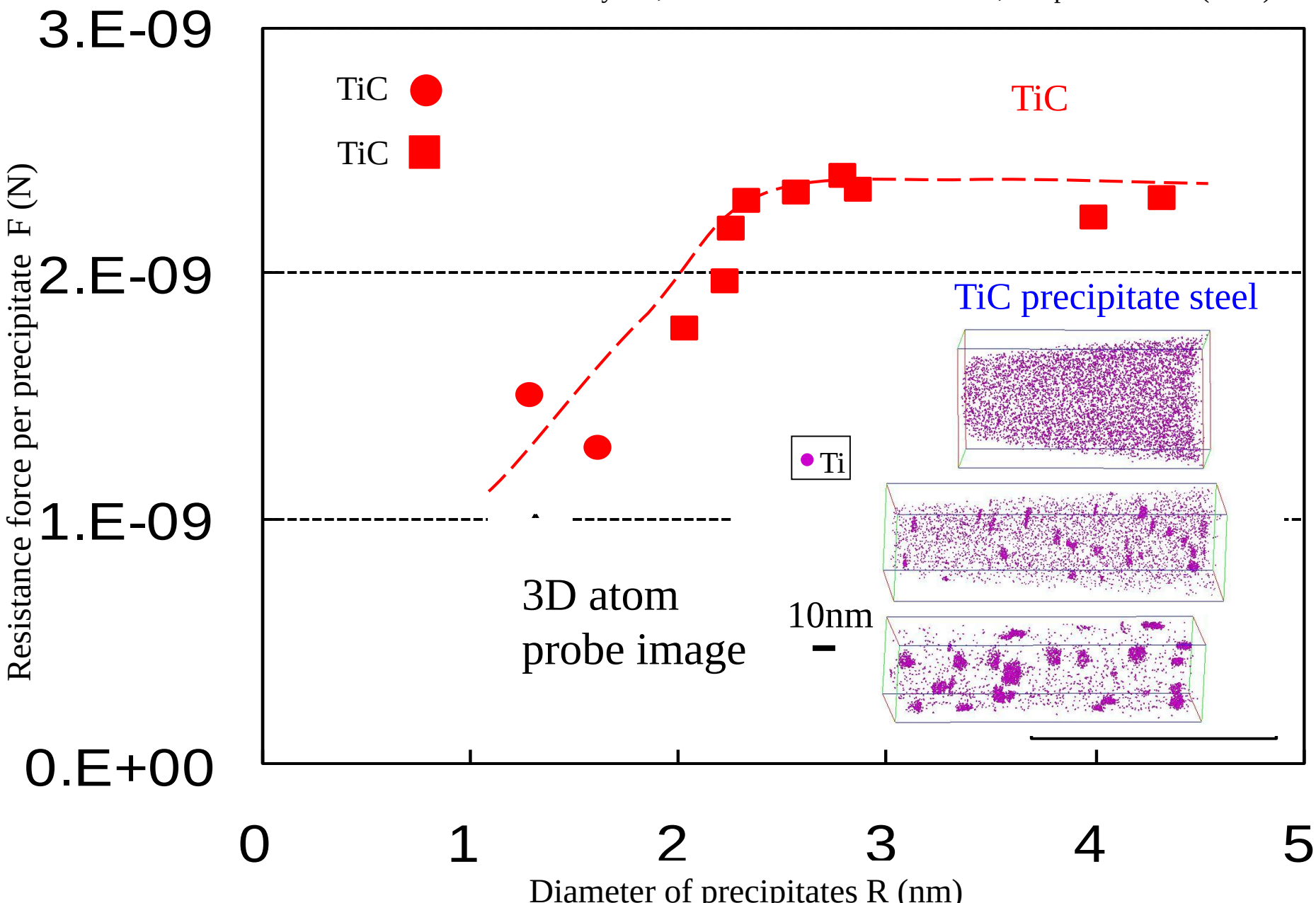


coherent case

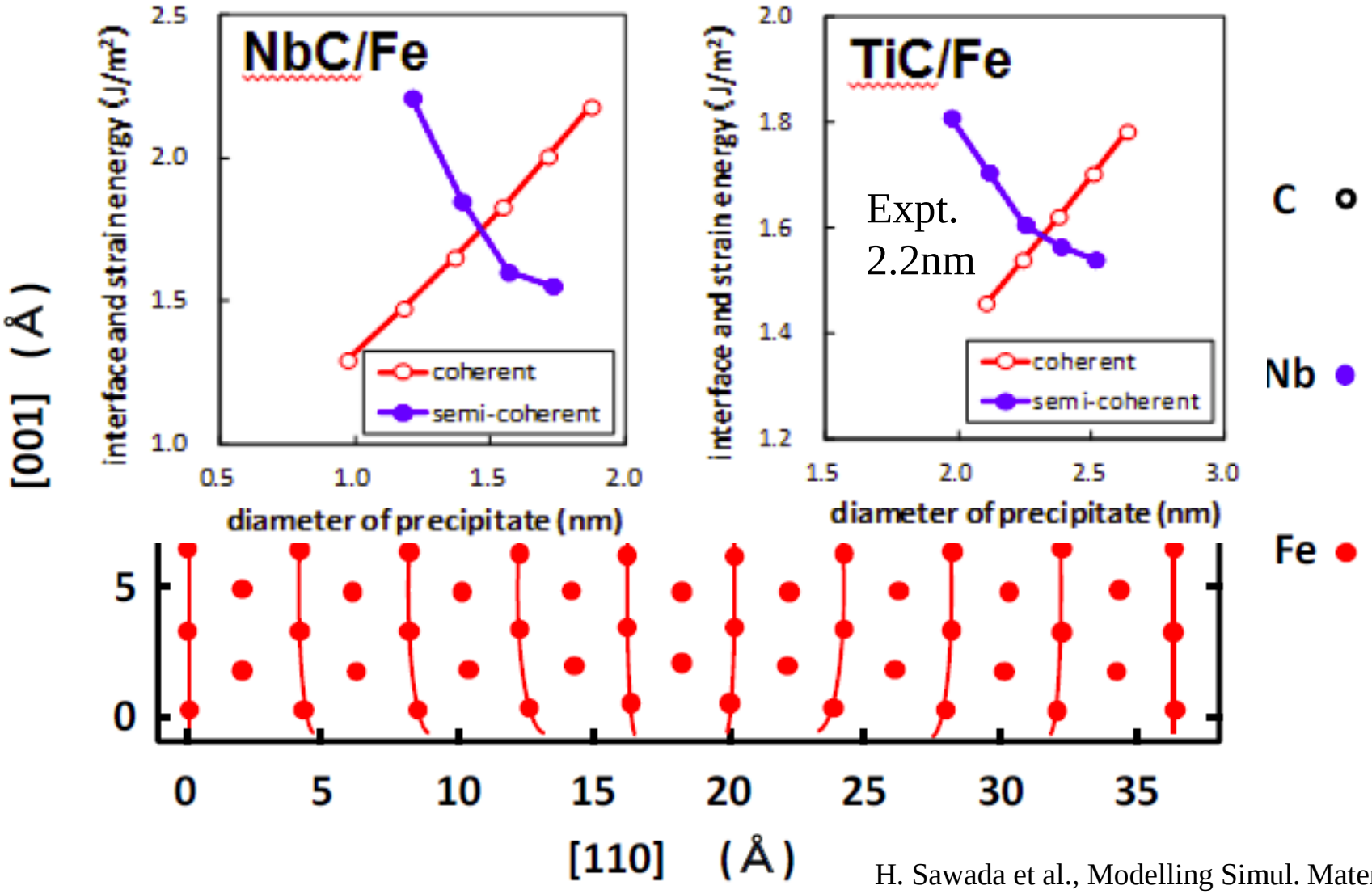
Iron

Resistance force and precipitate diameter

Y. Kobayashi, J. Takahashi and K. Kawakami, Scripta Mater. 67 (2012) 854



Crossover from coherent to semi-coherent



H. Sawada et al., Modelling Simul. Mater. Sci. Eng. 21, 045012 (2013).

Numerically exact low-order scaling method

Main difficulty: 'diagonalization'

$O(N^3)$ method - Numerically exact diagonalization

Householder+QR method

Conjugate gradient (CG) method

Davidson method

Even if basis functions are localized in real space, Gram-Shmidt (GS) type method is needed to satisfy orthonormality among eigenstates, which results in $O(N^3)$ for the computational time.

O(N) method - can be achieved in exchange for accuracy.

O(N) Krylov subspace method,

DC, DM, OM methods, etc..

$O(N^{2\sim})$ method

Is it possible to develop $O(N^{2\sim})$ methods without introducing approximations? → No more GS process.

Possible ways to avoid orthogonalization

1. Green's function method

$$\rho = -\frac{2}{\pi}\mathrm{Im}\int_{-\infty}^{\infty} dE G(E+i0^+) f\left(\frac{E-\mu}{k_{\mathrm{B}}T}\right)$$

2. Density matrix method

LNV, PRB 47, 10891 (1993)

$$\frac{\partial\Omega}{\partial\rho} = 0 \qquad \Omega = \mathrm{Tr}\left[(3\rho^2 - 2\rho^3)(H-\mu I)\right]$$

$$\mathrm{H, S} \longrightarrow \rho$$

3. Iterative method

$$|d_{n+1}\rangle = (H - \varepsilon_n S)^{-1} S |c_n\rangle$$

Numerically exact low-order scaling method

- ✓ Numerically exact
- ✓ Applicable to insulators and metals
- ✓ Suitable for parallel computation
- ✓ Applicable to 1D, 2D, 3D systems
- ✓ Applicable to any local basis functions

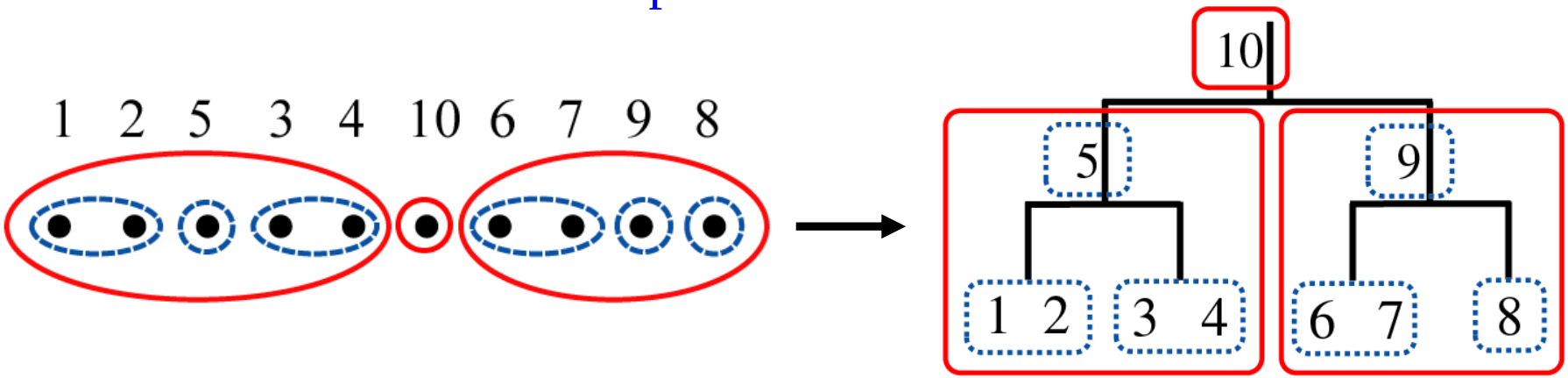
TO, PRB **82**, 075131 (2010)

Numerically exact low-order scaling method

1. Direct evaluation of the selected elements of ρ via a contour integration of the Green's function

$$\rho = M^{(0)} + \mathrm{Im}\left(-\frac{4i}{\beta}\sum_{p=1}^{\infty} G(\alpha_p) R_p\right)$$

2. Nested dissection of sparse matrix



TO, PRB **82**, 075131 (2010)

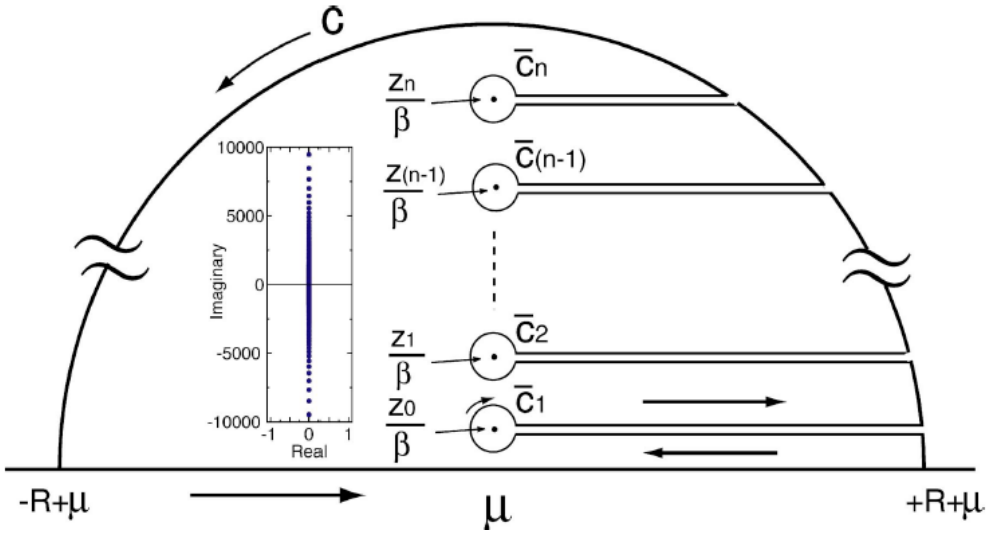
Continued fraction rep. of Fermi function

$$\frac{1}{1+\exp(x)} = \frac{1}{2} - \frac{x}{4}\left(\cfrac{1}{1+\cfrac{\left(\frac{x}{2}\right)^2}{3+\cfrac{\left(\frac{x}{2}\right)^2}{5+\cfrac{\left(\frac{x}{2}\right)^2}{\cfrac{\cdots}{(2M-1)+\ddots}}}}}\right),$$

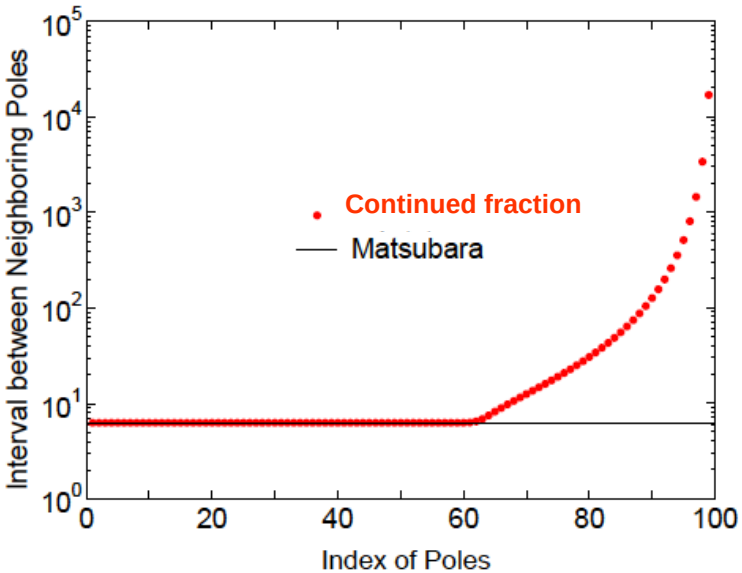
TO, PBR **75**, 035123 (2007)

Contour integration

All the poles are located on the imaginary axis.



The form has a special pole structure, that is, the interval between neighboring poles increases in a faraway region from the real axis, which is very advantageous for the contour integration of Green's function.



Convergence of ρ w.r.t. poles

The calculation of ρ can be expressed by a contour integration:

$$
\begin{aligned}
\rho_{ij} &= \sum_k f\left(\frac{\varepsilon_k - \mu}{k_{\mathrm{B}}T}\right)\langle\chi_i|\phi_k\rangle\langle\phi_k|\chi_j\rangle, \\
&= -\frac{2}{\pi}\mathrm{Im}\int_{-\infty}^{\infty} dE f\left(\frac{E-\mu}{k_{\mathrm{B}}T}\right)G_{ij}(E+i0^+), \\
&= M_{ij}^{(0)} + \mathrm{Im}\left[-\frac{4i}{\beta}\sum_{p=1}^{\infty} G_{ij}(\alpha_p)R_p\right], \\
&= M_{ij}^{(0)} + \mathrm{Im}\left[-\frac{4i}{\beta}\sum_{p=1}^{\infty}\sum_k \frac{\langle\chi_i|\phi_k\rangle\langle\phi_k|\chi_j\rangle}{\alpha_p - \varepsilon_k}R_p\right], \\
&= M_{ij}^{(0)} + \sum_k \mathrm{Im}\left[-\frac{4i}{\beta}\sum_{p=1}^{\infty}\frac{\langle\chi_i|\phi_k\rangle\langle\phi_k|\chi_j\rangle}{\alpha_p - \varepsilon_k}R_p\right],
\end{aligned}
$$

$$M_{ij}^{(0)} = \mathrm{Im}\left[-\frac{1}{\pi}\int_{-\infty}^{\infty} dE G_{ij}(E+i0^+)\right] \simeq iR\, G(iR)$$

$$\alpha_p = \mu_0 + i\frac{z_p}{\beta}$$

Lehmann rep.

The analysis shows that the number of poles for each eigenstate for a sufficient convergence does not depend on the size of system if the spectrum radius does not change. → The scaling property is governed by the calculation of G.

Convergence property of the contour integration

Total energy of aluminum as a function of the number of poles by a recursion method at 600 K.

Nicholson et al., PRB **50**, 14686 (1994).

Poles	Proposed	$\frac{1}{1+\left(1+\frac{x}{n}\right)^n}$	Matsubara
10	−42.933903047211	−33.734015919550	−39.612354360046
20	−47.224346653790	−33.623477214678	−39.849746603905
40	−48.323790725570	−33.346245616679	−40.216055898502
60	−48.324441992259	−33.143128624551	−39.676965494522
80	−48.324441994952	−32.870752577236	−43.523770052176
150	−48.324441994952	−33.837428496424	−41.836938942518
200		−33.418012271726	−42.543354202255
250		−34.003411636691	−43.024756221080
300		−34.003236479262	−43.466729654170
350		−48.324440028792	−43.834528739677
400		−48.324440274509	−44.185100655185
600		−48.324440847749	−45.233651519749
1000		−48.324441306517	−46.331692884149
2000		−48.324441650693	−47.202779497545
5000		−48.324441857239	−47.921384128418
10000		−48.324441926094	−48.122496320516

The energy completely converges using only 80 poles within double precision.

How can Green's funtion be evaluated ?

- The Green's function is the inverse of a sparse matrix *(ZS-H)*.

$$G(Z) = (ZS - H)^{-1}$$

- Selected elements of *G(Z)*, which correspond to non-zero elements of the overlap matrix *S*, are needed to calculate physical properties.

Our idea

1. Nested dissection of *(ZS-H)*
2. LDL^T decomposition for the structured matrix

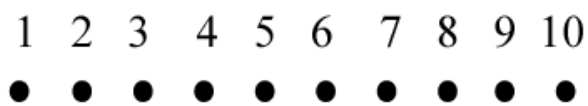
→ a set of recurrence relations

TO, PRB **82**, 075131 (2010)

Nested dissection of a sparse matrix

George, SIAM J. Numer. Anal. 10, 345 (1973).

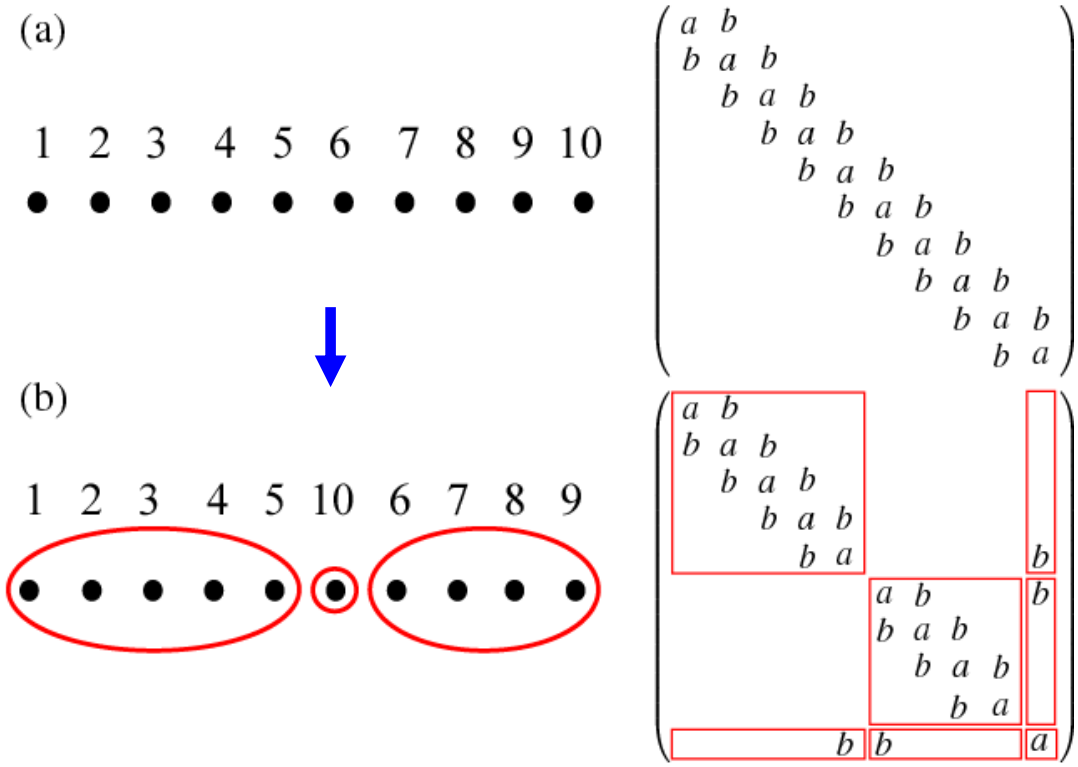
(a)



$$\begin{pmatrix} a & b & & & & & & & & \\ b & a & b & & & & & & & \\ & b & a & b & & & & & & \\ & & b & a & b & & & & & \\ & & & b & a & b & & & & \\ & & & & b & a & b & & & \\ & & & & & b & a & b & & \\ & & & & & & b & a & b & \\ & & & & & & & b & a & b \\ & & & & & & & & b & a \end{pmatrix}$$

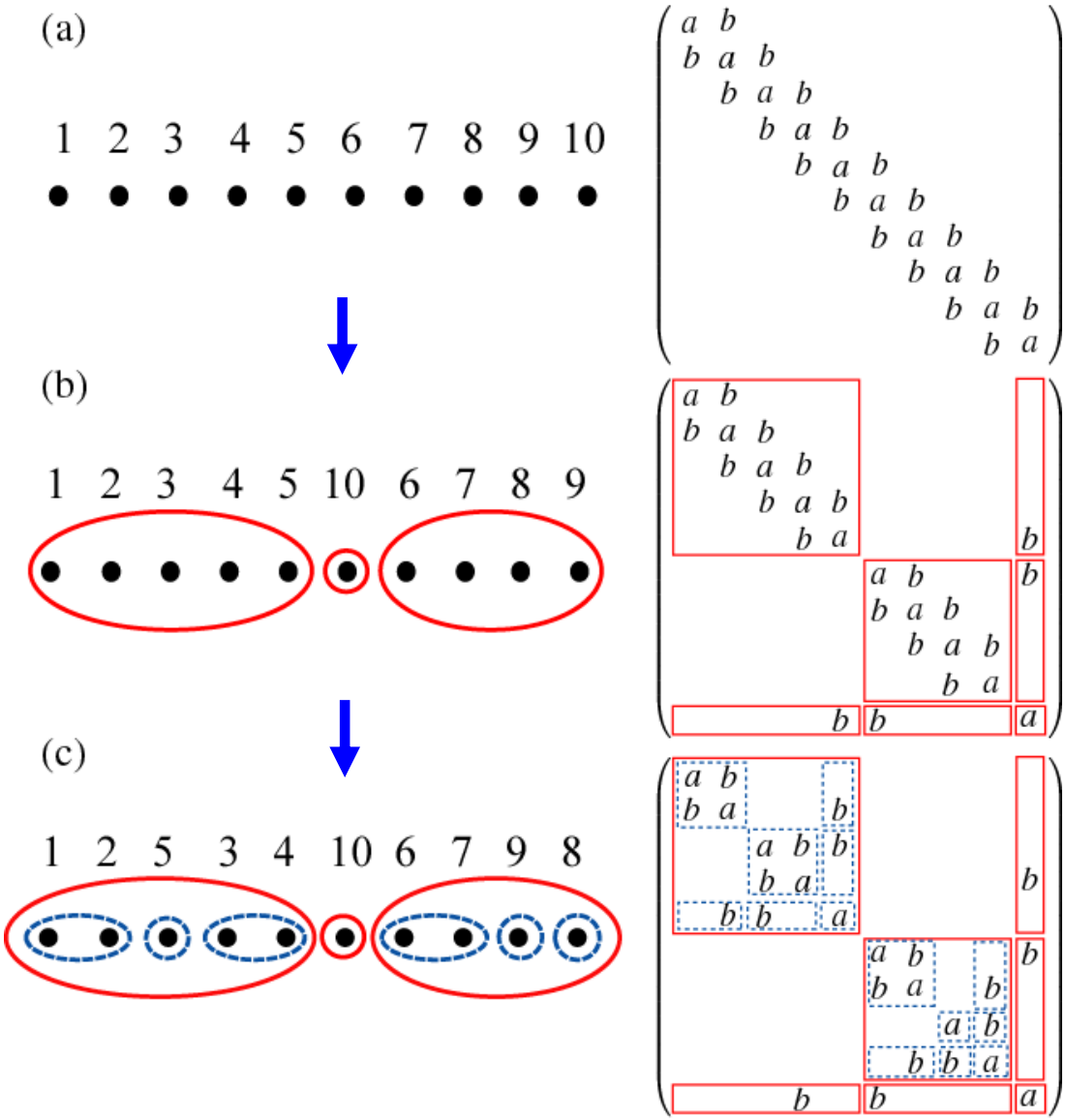
Nested dissection of a sparse matrix

George, SIAM J. Numer. Anal. 10, 345 (1973).



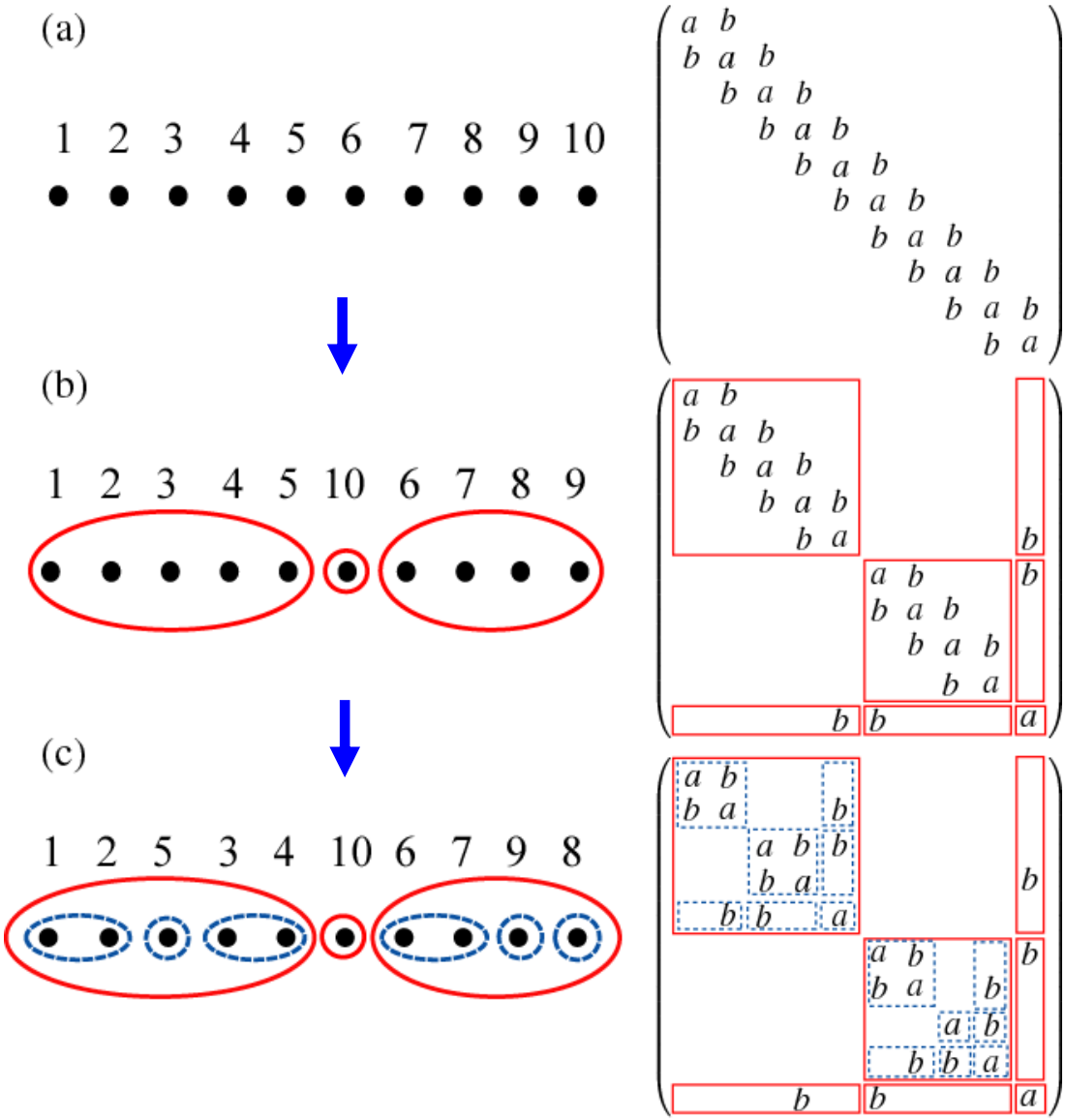
Nested dissection of a sparse matrix

George, SIAM J. Numer. Anal. 10, 345 (1973).

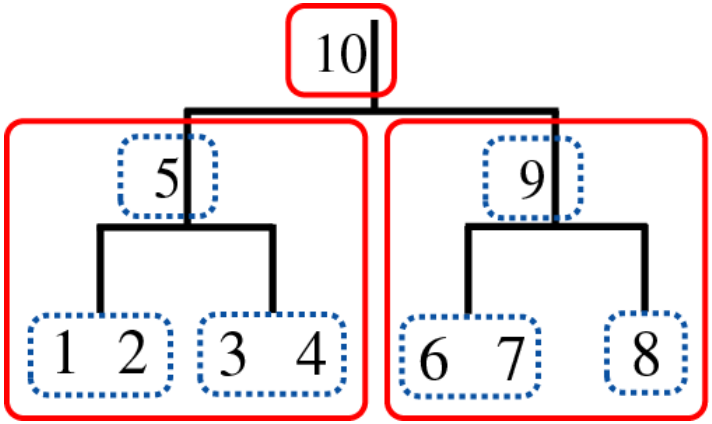


Nested dissection of a sparse matrix

George, SIAM J. Numer. Anal. 10, 345 (1973).



The hierarchical structure of interactions of domains.



Nested dissection of a sparse matrix

The processes (i)-(v) are recursively applied to each domains with computational cost of $O(N(\log_2 N)^2)$ in total.

(i) Ordering:

The basis functions are ordered by coordinates along each direction.

(ii) Screening:

The basis functions with a long tail are assigned as part of the separator.

(iii) Finding of a starting nucleus:

Find a basis function having the smallest number of nonzero overlaps.

(iv) Growth of the nucleus:

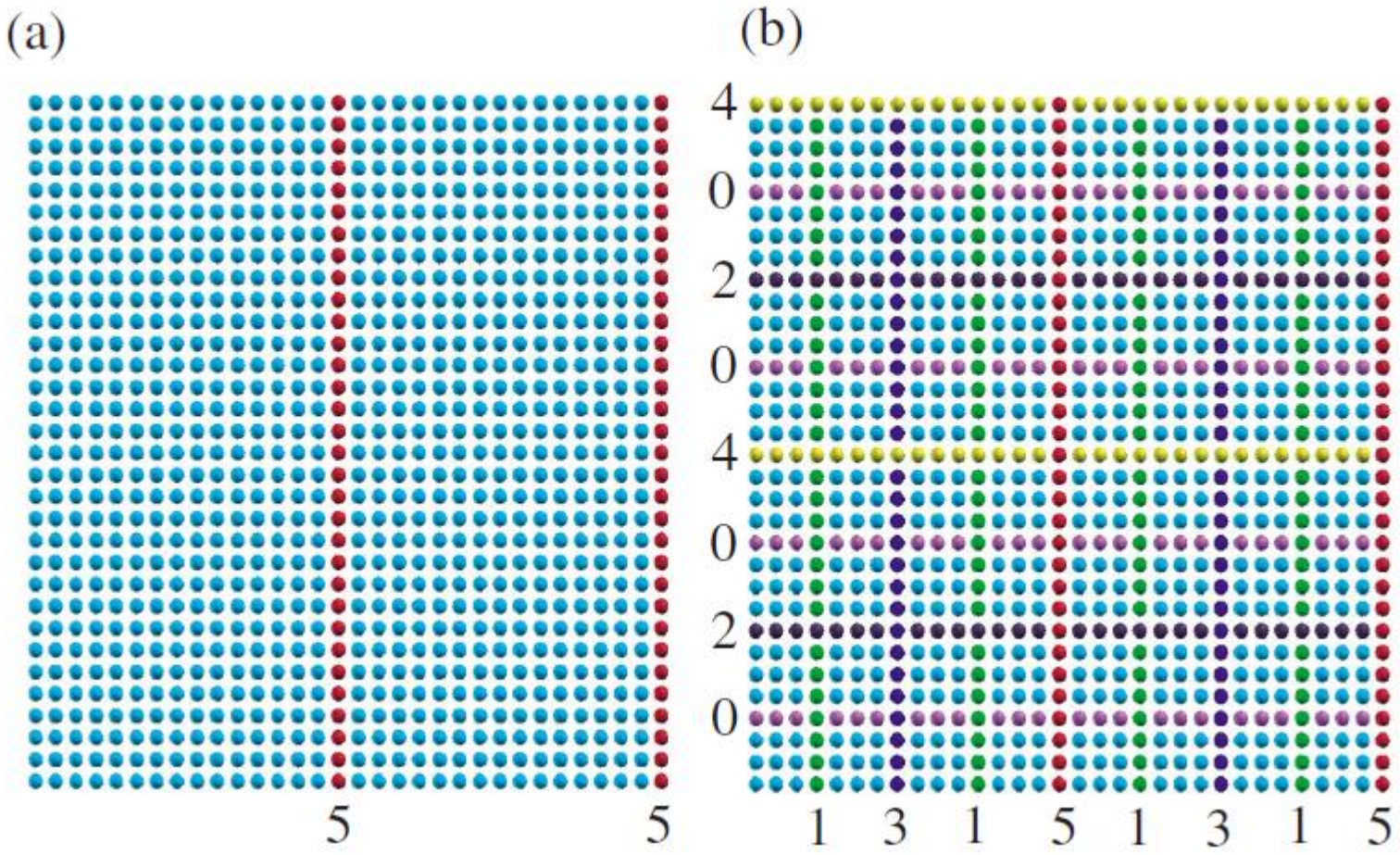
Minimize $|N_0-N_1| + N_s$ by the growth of the nucleus.

N_0: # of bases in domain 0
N_1: # of bases in domain 1
N_s: # of bases in separator

(v) Dissection:

Find a direction with the smallest $|N_0-N_1| + N_s$, make the dissection along the direction.

Square lattice for the nested dissection



Inverse by LDL^T block factorization

A matrix X can be factorized using a Schur complement into a LDL^T form.

$$X = \begin{pmatrix} A & B^T \\ B & C \end{pmatrix} = \begin{pmatrix} I & \\ L & I \end{pmatrix} \begin{pmatrix} A & \\ & S \end{pmatrix} \begin{pmatrix} I & L^T \\ & I \end{pmatrix}$$

$$L = BA^{-1}$$

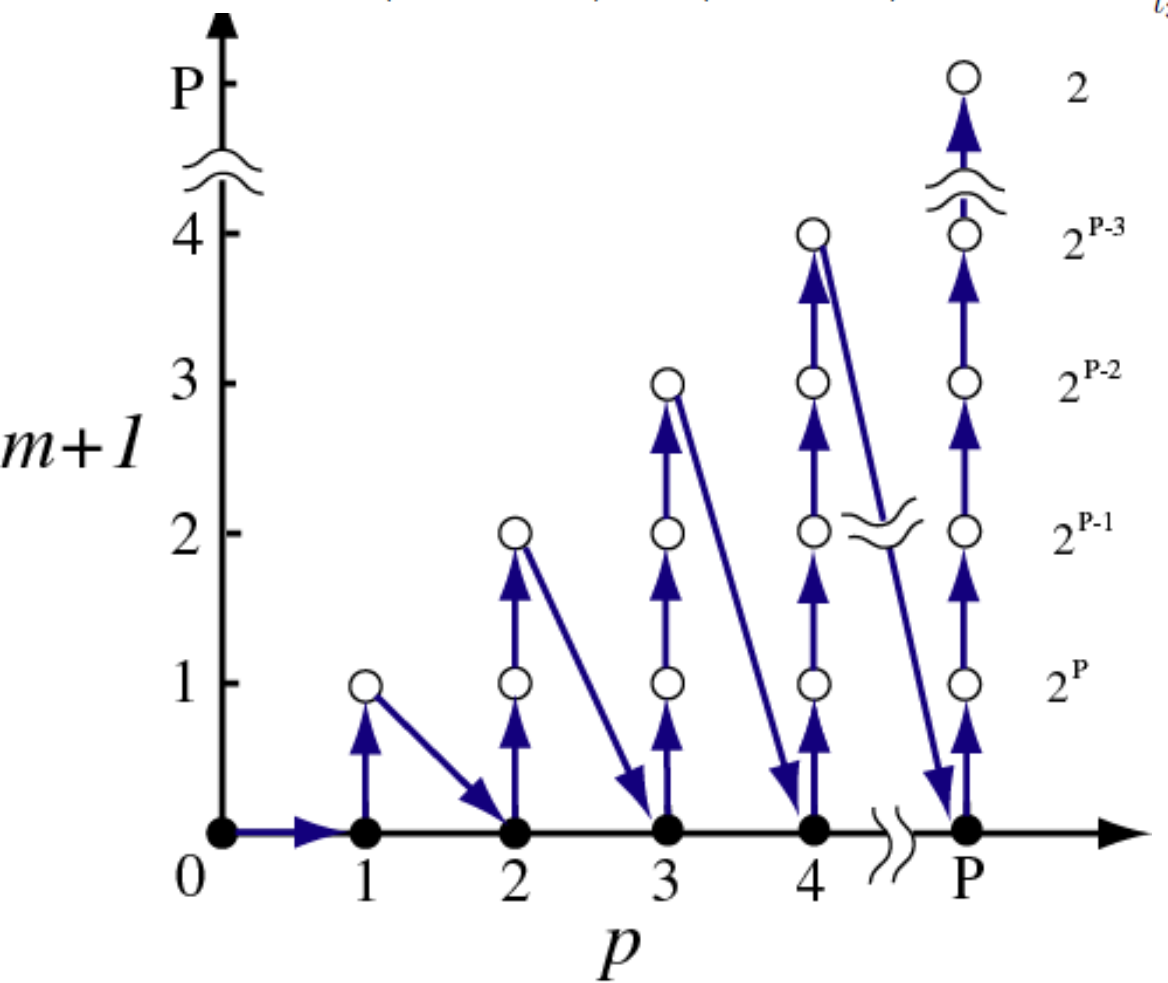
$$S = C - BA^{-1}B^T$$

Then, the inverse of X is given by

$$X^{-1} = \begin{pmatrix} A^{-1} + L^T S^{-1} L & -L^T S^{-1} \\ -S^{-1} L & S^{-1} \end{pmatrix}$$

Analysis of the computational cost

$$V^T_{p,m+1,n} = \begin{pmatrix} V^T_{p,m,2n} \\ V^T_{p,m,2n+1} \\ 0 \end{pmatrix} + \begin{pmatrix} L^T_{m,2n} \\ L^T_{m,2n+1} \\ -I \end{pmatrix} Q^T_{p,m+1,n}$$



$$t_{3D} \propto \sum_{p=1}^{P} \sum_{m=0}^{p-1} \sum_{n=0}^{2^{P-m}-1} \frac{N}{2^{P-m}} N^{(2)}_{p,m,n} N^{(3)}_{p,m,n}$$

$$< \sum_{p=1}^{P} \sum_{m=0}^{p-1} \sum_{n=0}^{2^{P-m}-1} \frac{N}{2^{P-m}} \frac{N^{2/3}}{2^{\frac{2}{3}(P-m-1)}} \frac{N^{2/3}}{2^{\frac{2}{3}(P-p)}}$$

$$= \frac{4N^{7/3}}{2^{2/3}6-9} \left(-1 + 2^{2/3} - \frac{1}{2^{2/3}2^{4P/3}} + \frac{1}{2^{2/3}2^{2P/3}} - \frac{2^{2/3}}{2^{2P/3}} + \frac{1}{2^{4P/3}} \right).$$

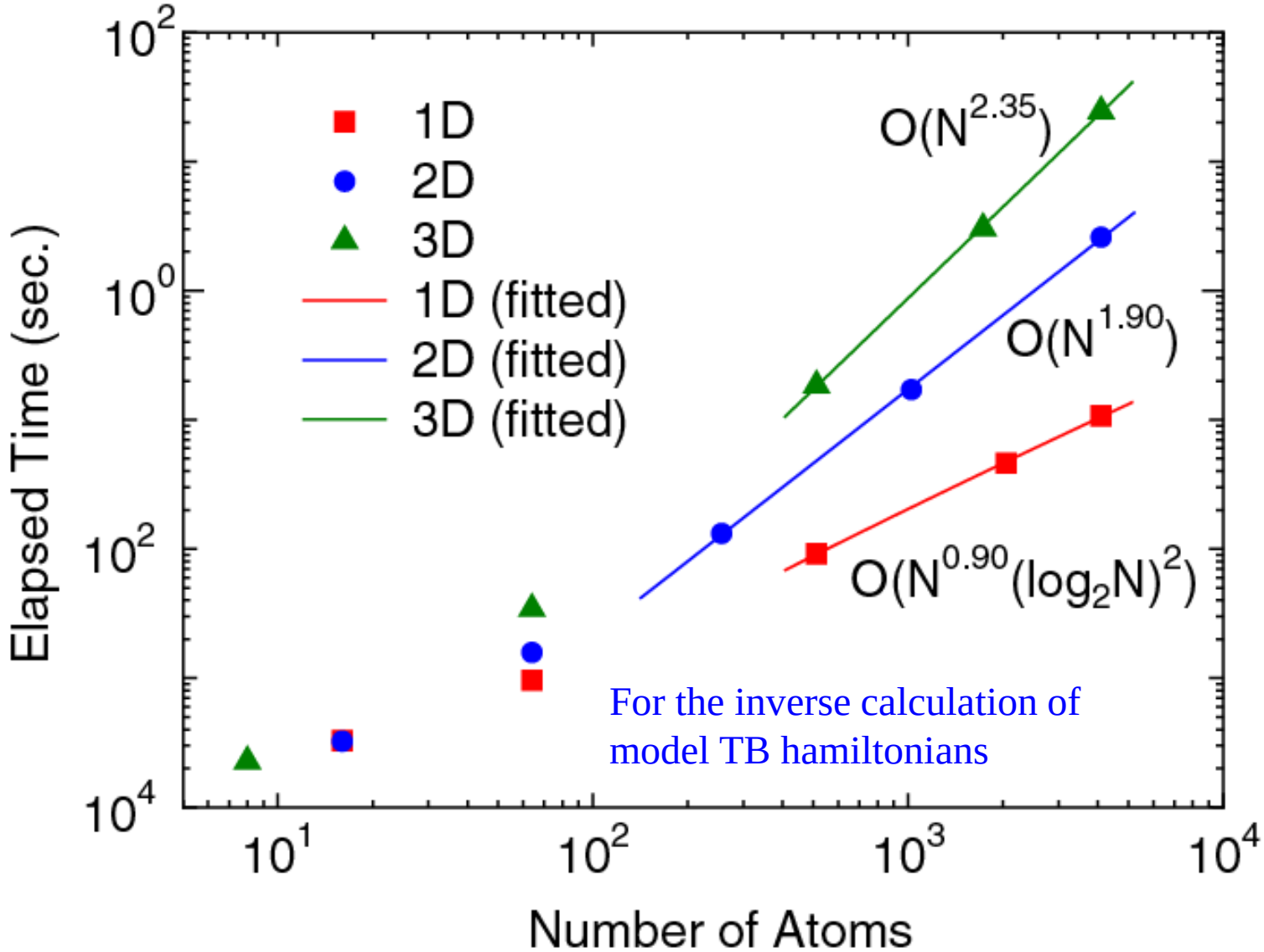
$\rightarrow O(N^{7/3})$

1D $O(N(\log_2 N)^2)$

2D $O(N^2)$

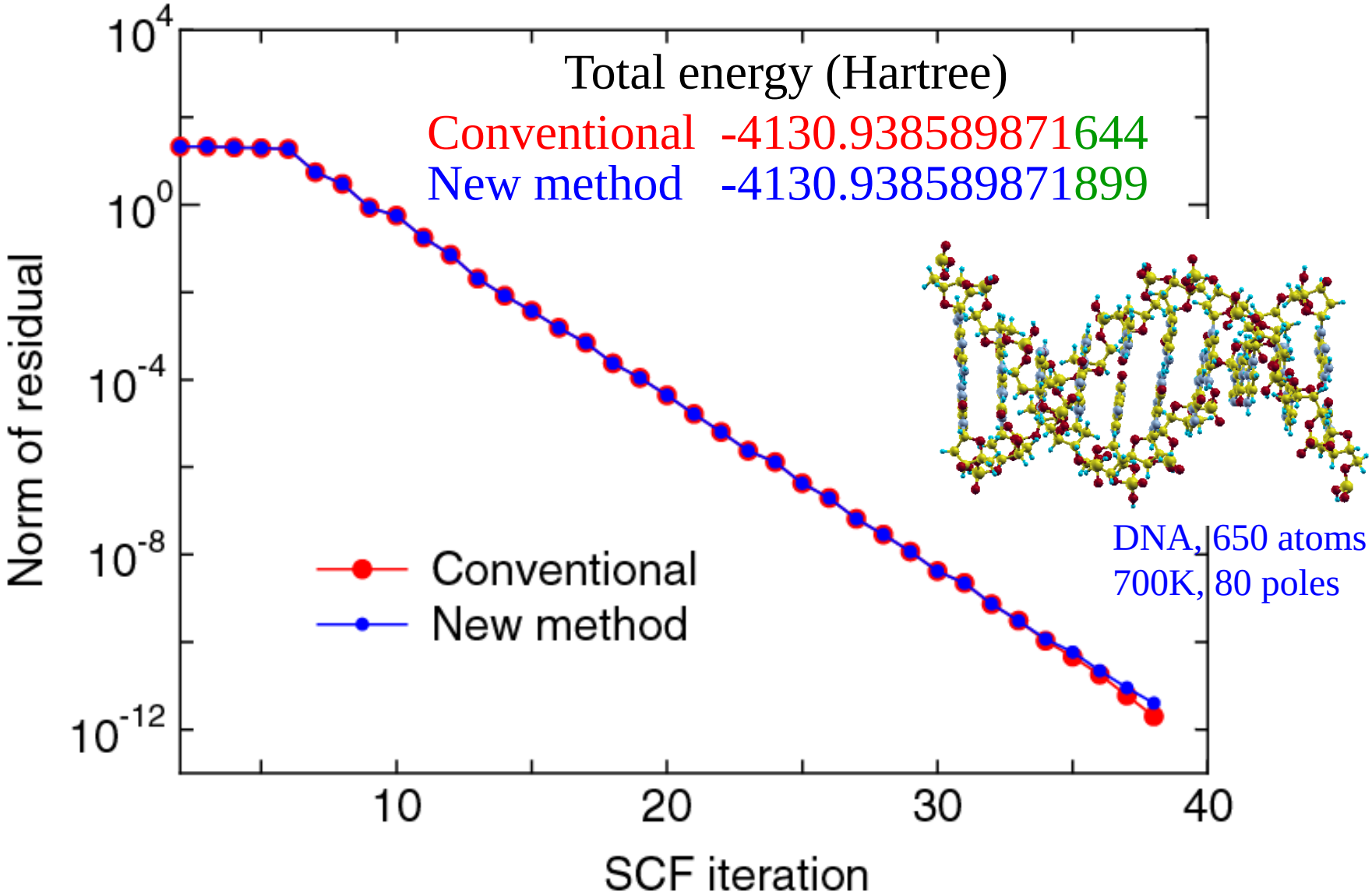
3D $O(N^{7/3})$

Timing result

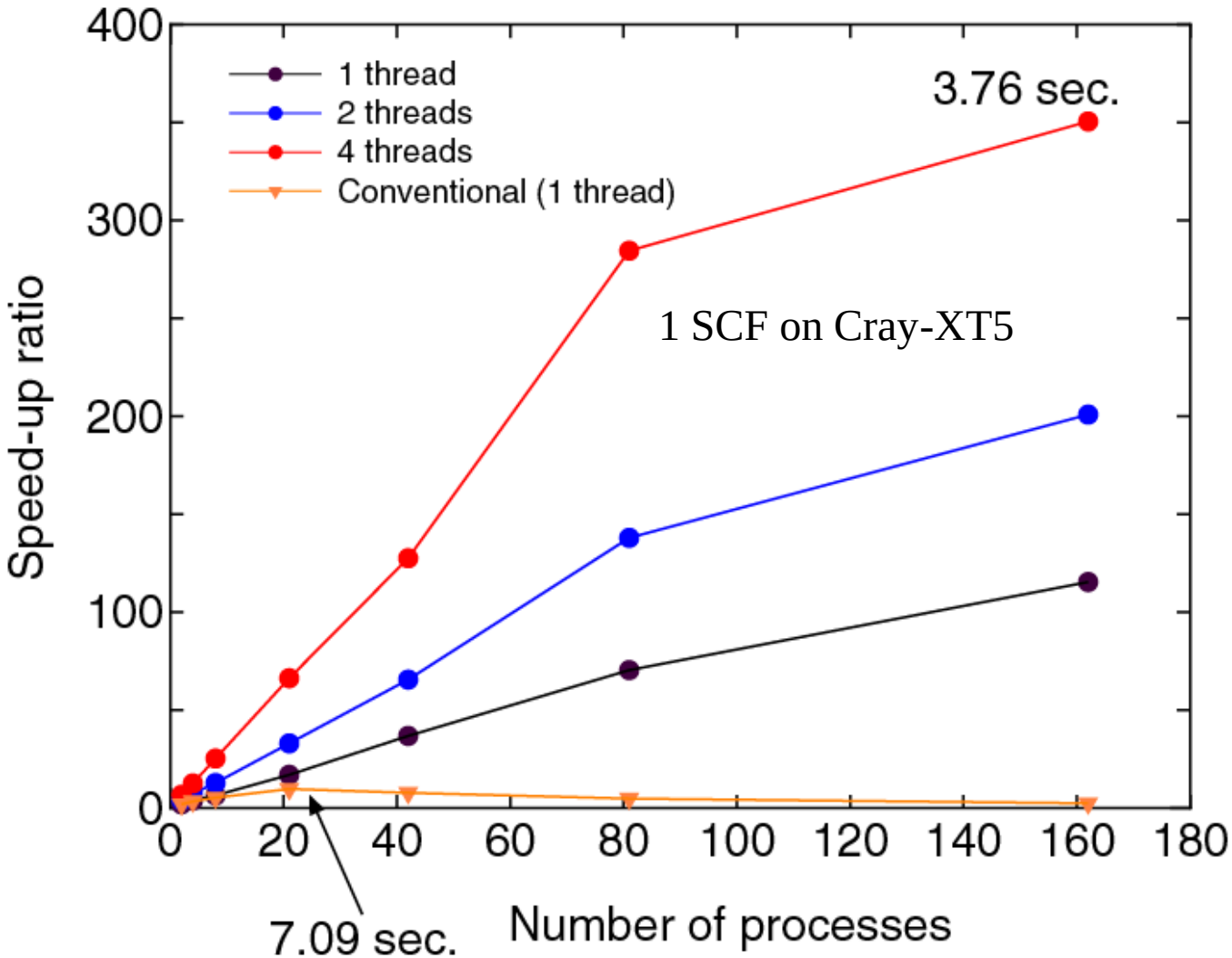


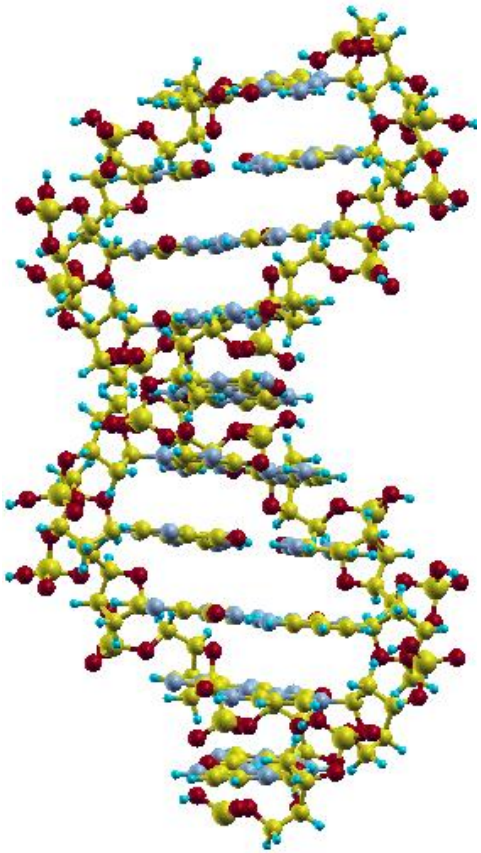
For the inverse calculation of model TB hamiltonians

SCF convergence



Parallel efficiency





O(N) nearly exact exchange functional

Exchange as density *matrix* functional

Hartree-Fock exchange

$$E_{\mathrm{x}} = \frac{1}{2} \int \int \mathbf{r}_1 \mathbf{r}_2 \frac{n(\mathbf{r}_1)\rho_{\mathrm{x}}(\mathbf{r}_1, \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}$$

Exchange hole

$$\rho_{\mathrm{x}}(\mathbf{r}_1, \mathbf{r}_2) = -\frac{|\rho^{(1)}(\mathbf{r}_1, \mathbf{r}_2)|^2}{n(\mathbf{r}_1)}$$

- Exponential decay at finite temperature
- Sum rule: integration of ρ_x over space is -1
- Negativity

O(N) computation of *exact* exchange

Thus, it may be possible to estimate the long range exchange hole by making use of the sum rule and the decaying property of exchange hole. → The long range exchange hole is replaced by a model hole.

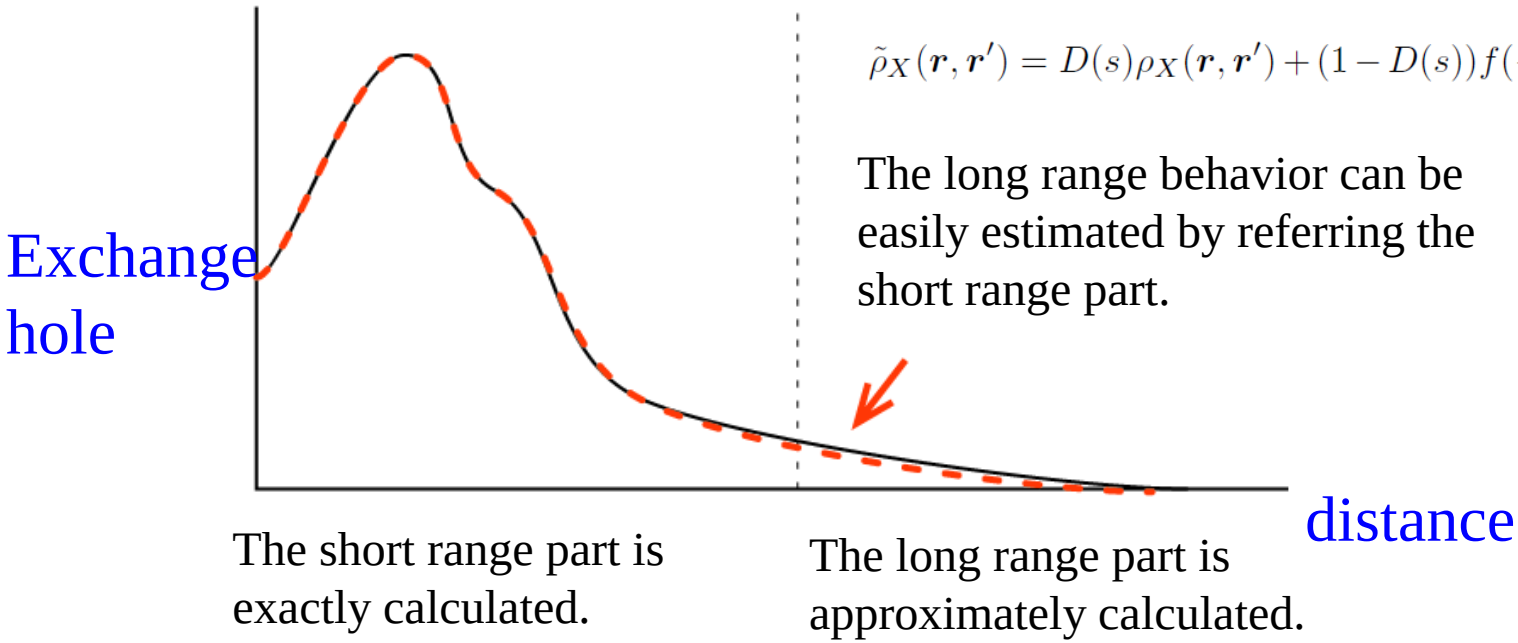
$$\tilde{\rho}_X(\boldsymbol{r},\boldsymbol{r}') = D(s)\rho_X(\boldsymbol{r},\boldsymbol{r}') + (1-D(s))f(\{\alpha(\boldsymbol{r})\};s)$$

Feature

Accuracy:	Comparable to OEP
Self-Int.:	Nearly Self-Int. free
Asymptotic exchange potential:	-1/r
Cost(local basis):	O(N)

M. Toyoda and T. Ozaki, PRA, 83, 032515 (2011).

Range-separation of exchange hole



$$\tilde{\rho}_X(\boldsymbol{r},\boldsymbol{r}') = D(s)\rho_X(\boldsymbol{r},\boldsymbol{r}') + (1-D(s))f(\{\alpha(\boldsymbol{r})\};s)$$

The long range behavior can be easily estimated by referring the short range part.

The model long range hole we used is the hole of hydrogenic atom.

$$f_H(a,b,s) = \frac{a}{16\pi bs}\left[\left(a\mid b-s\mid +1\right)\exp(-a\mid b-s\mid) - \left(a\mid b+s\mid +1\right)\exp(-a\mid b+s\mid)\right]$$

where *a* and *b* are determined by matching the 0th and 1st moments.

0th moment: Sum rule for the short-range part.

1st moment: Exchange potential for the short range part.

Self-consistent field equation

The variational eq. can be analytically derived.

→ The SCF calc. and calculation of forces are possible.

$$\frac{\delta \tilde{E}_X}{\delta \psi_i^*(\boldsymbol{r})} = -\sum_j^{occ.} \psi_j(\boldsymbol{r}) \int^{s<s_{\max}} \tilde{v}(\boldsymbol{r}, s)\psi_i(\boldsymbol{r}')\psi_j^*(\boldsymbol{r}')d^3r'$$

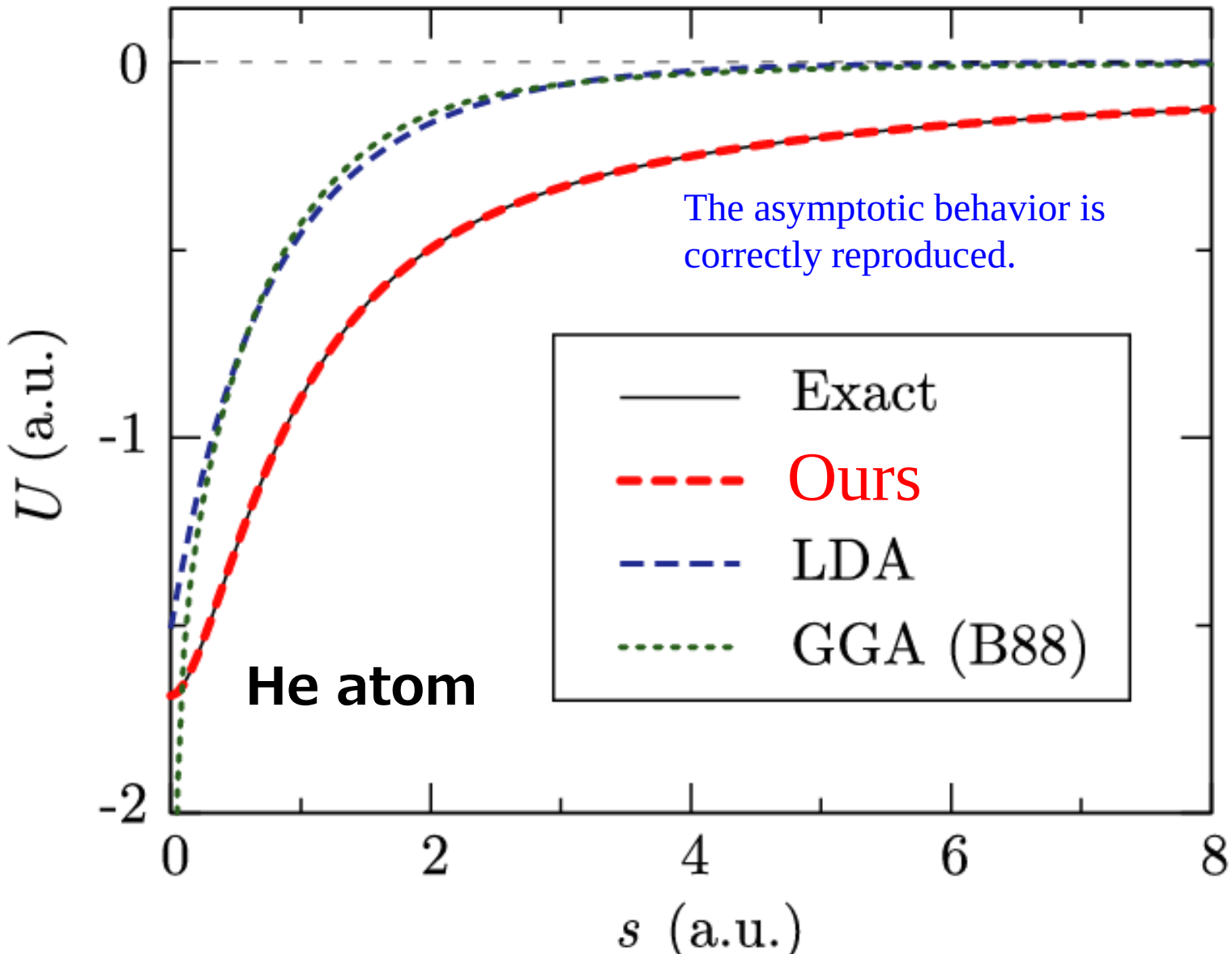
$$-\frac{1}{2}\psi_i(\boldsymbol{r})\left(\tilde{\epsilon}_X^{LR}(\boldsymbol{r}) - \alpha M_0(\boldsymbol{r}) - \beta M_1(\boldsymbol{r})\right)$$

Total energy of rare gases

				Screening length 5.3Å	2.6Å	1.8Å
	Exact	OEP [28]	KLI-OEP [29]	Ours $\mu = 0.1$	$\mu = 0.2$	$\mu = 0.3$
He	−2.8617	−2.8617		−2.8616	−2.8616	−2.8617
Ne	−128.5471	−128.5454	−128.5449	−128.5471	−128.5480	−128.5498
Ar	−526.8175	−526.8122	−526.8105	−526.8177	−526.8193	−526.8236
Kr	−2752.0549	−2752.0430	−2752.0398	−2752.0552	−2752.0584	−2752.0663
Xe	−7232.1384	−7232.1211	−7232.1149	−7232.1388	−7232.1437	−7232.1558
Rn	−21866.7722			−21866.7729	−21866.7797	−21866.7964
AARE[b] (%)		0.0006	0.0009	0.0006	0.0008	0.0007

The accuracy is comparable to OEP.

Exchange potential



OpenMX **Open source package for Material eXplorer**

http://www.openmx-square.org

- Software package for density functional calculations of molecules and bulks
- Norm-conserving pseudopotentials (PPs)
- Variationally optimized numerical atomic basis functions

Basic functionalities

- SCF calc. by LDA, GGA, DFT+U
- Total energy and forces on atoms
- Band dispersion and density of states
- Geometry optimization by BFGS, RF, EF
- Charge analysis by Mullken, Voronoi, ESP
- Molecular dynamics with NEV and NVT ensembles
- Charge doping
- Fermi surface
- Analysis of charge, spin, potentials by cube files
- Database of optimized PPs and basis funcitons

Extensions

- O(N) and low-order scaling diagonalization
- Non-collinear DFT for non-collinear magnetism
- Spin-orbit coupling included self-consistently
- Electronic transport by non-equilibrium Green function
- Electronic polarization by the Berry phase formalism
- Maximally localized Wannier functions
- Effective screening medium method for biased system
- Reaction path search by the NEB method
- Band unfolding method
- STM image by the Tersoff-Hamann method
- etc.

Outlook

The locality of density matrix and basis function is a key to develop a wide variety of efficient electronic structure methods.

We have demonstrated three methods:

- O(N) Krylov subspace method
- Low-order scaling exact method
- O(N) *exact* exchange method

Plenty of developments of new efficient methods might be still possible.