Large-scale electronic structure methods
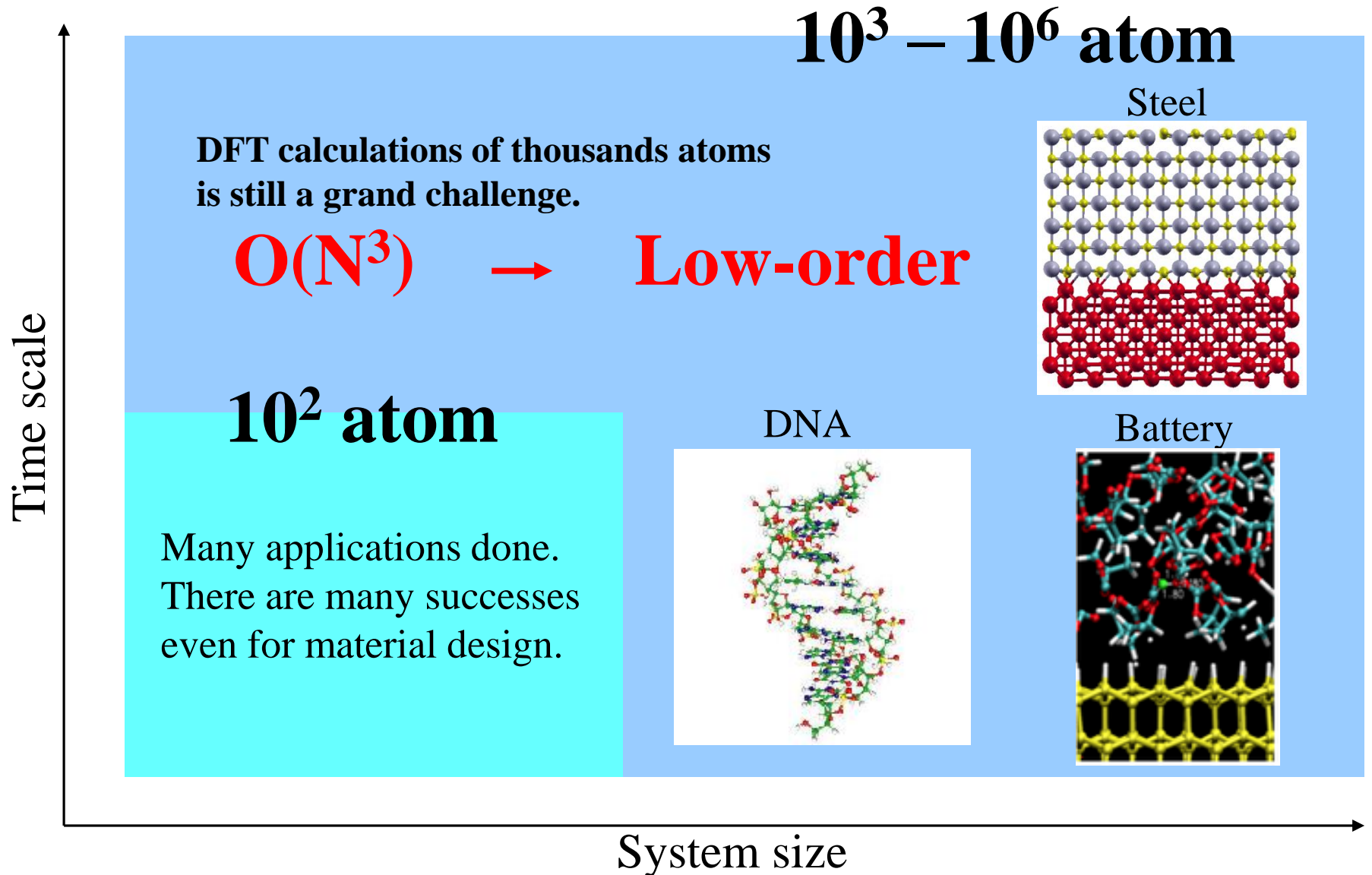
- Introduction
- Lanczos method
- 1D tight-binding model
- O(N) Krylov subspace method
- Applications
- Outlook

Taisuke Ozaki (ISSP, Univ. of Tokyo)

The Summer School on DFT: Theories and Practical Aspects, July 2-6, 2018, ISSP

Towards first-principle studies for industry

$10^3 - 10^6$ atom

DFT calculations of thousands atoms
is still a grand challenge.

$O(N^3)$ $\rightarrow$ Low-order

Steel

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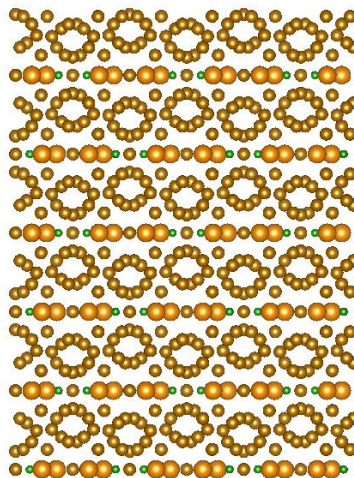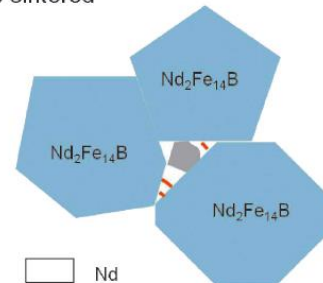
e.g., the coercivity of a permanent magnet of Nd-Fe-B is determined by crystal structure, grain size, and grain boundary.

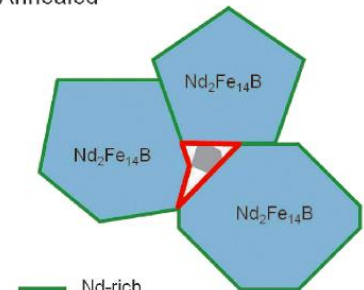http://ev.nissan.co.jp/LEAF/PERFORMANCE/

As-sintered

$Nd_2Fe_{14}B$
$Nd_2Fe_{14}B$
$Nd_2Fe_{14}B$

☐ Nd
▨ NdO_x
— Cu-rich

Annealed

$Nd_2Fe_{14}B$
$Nd_2Fe_{14}B$
$Nd_2Fe_{14}B$

— Nd-rich

Summit in ORNL: 187 Peta flops machine

Summit - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory, United States

Cores: 2,282,544+NVIDIA Tesla V100 GPUs
Rmax: 122,300.0 (TFLOP/sec.)
Pmax: 187,659.3 (TFLOPS/sec.)

https://www.olcf.ornl.gov/olcf-resources/compute-systems/summit/

According to Moore's law…
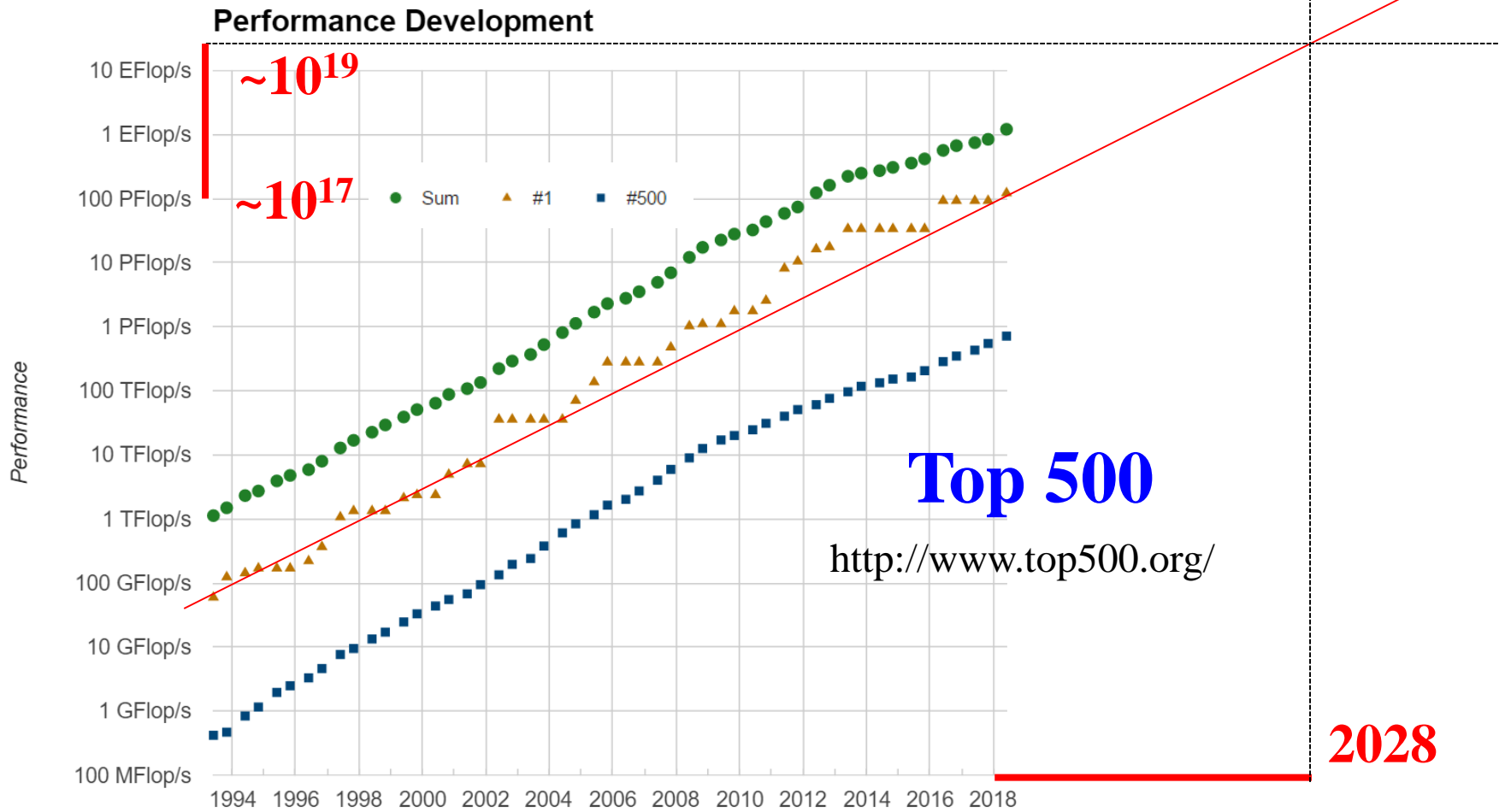
The machine performance may reach to 10 Exa FLOPS around 2028.

How large systems can be treated 10 years later?

The performance increase is about 100 times.

Summit ~100 PFLOPS → 10 years later 10000 PFLOPS

Computational Scaling $O(N^p)$	Computable size
7	1.9
6	2.2
5	2.5
4	3.1
3	4.6 ← DFT
2	10
1	100

The applicability of the $O(N^3)$ DFT method is extended to only 5 times larger systems even if Moore's law continues.

Mathematical structure of KS eq.

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

$O(N^3)$
$$\hat{H}_{\mathrm{KS}}\phi_i = \varepsilon_i\phi_i$$

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

$O(N^2)$
$$\rho(\mathbf{r}) = \sum_i^{\mathrm{occ}} \phi_i^*(\mathbf{r})\phi_i(\mathbf{r})$$

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

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

Red characters indicate the computational order of each calculation.

The largest order appears in the diagonalization, and the whole computational order asymptotically approaches to $O(N^3)$.

Density functional as a functional of *n*

Electron density ρ(r) is calculated by the 1ˢᵗ order reduced density matrix.

$$\rho(r) = \sum_{i,j} n_{ij} \chi_j(r) \chi_i(r)$$

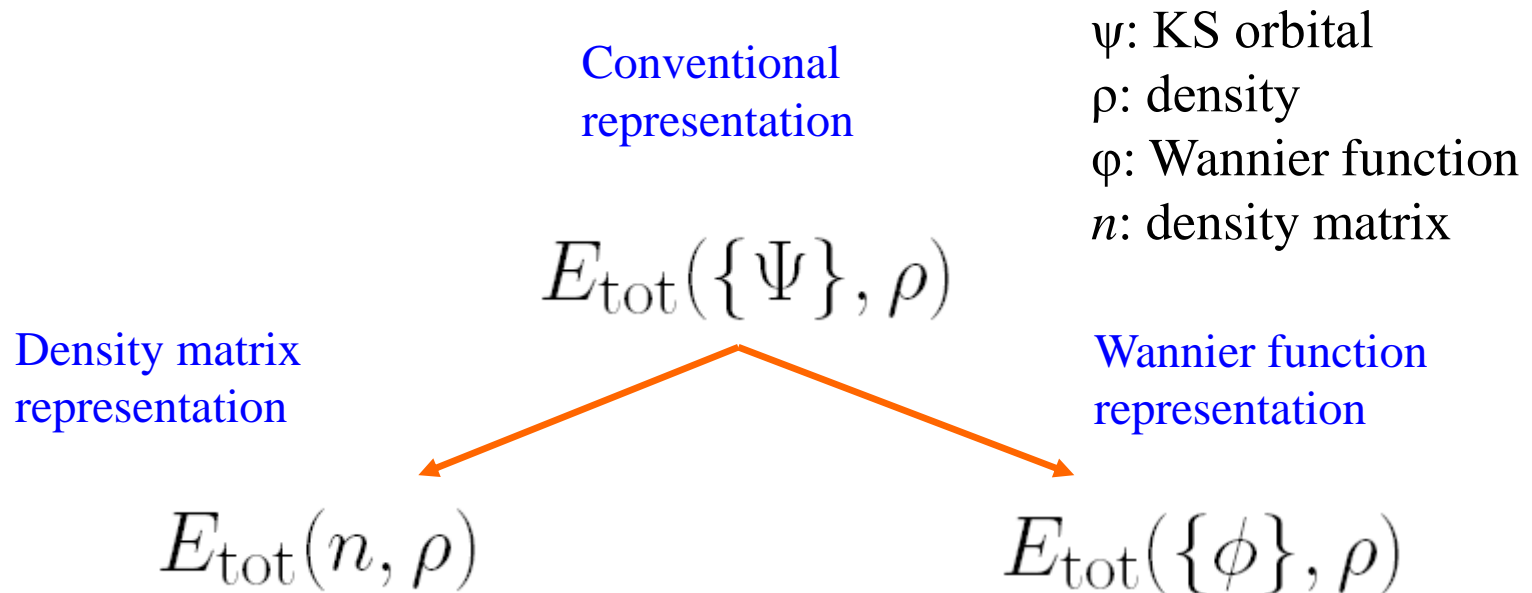Density functional can be rewritten by the first order reduced density matrix: ρ

$$E_{\text{tot}}[\rho, n] = \text{Tr}(n H_{kin}) + \int dr \rho(r) v_{ext}(r)$$

$$+ \frac{1}{2} \iint dr dr' \frac{\rho(r)\rho(r')}{|r-r'|} + E_{\text{xc}}[\rho]$$

If basis functions are localized in real space, the number of elements in the density matrix required to calculate the total energy is O(*N*).

The fact leads to reduction of computational order if only the necessary elements can be calculated.

Two routes towards O(N) DFT

The conventional expression of total energy in DFT is written by electron density and KS orbitals. It is possible to rewrite the energy expression using either density matrix or Wannier functions without introducing approximations.

Conventional
representation

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

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

Density matrix
representation

Wannier function
representation

$$E_{\text{tot}}(n, \rho) \qquad E_{\text{tot}}(\{\phi\}, \rho)$$

It might be possible to reduce the computational order by taking account of locality of density matrix and Wannier functions in real space.

Wannier functions and density matrix

Wannier functions ϕ can be obtained by an unitary transformation of Bloch functions ψ.

$$|\phi_v\rangle = \frac{V}{(2\pi)^3}\int\limits_{BZ} dk \sum_m^{occ} U_{\mu v}|\psi_{\mu k}\rangle \exp(-ik\bullet R)$$

for cases with a gap

Density matrix is obtained through a projection operator of Bloch functions ψ
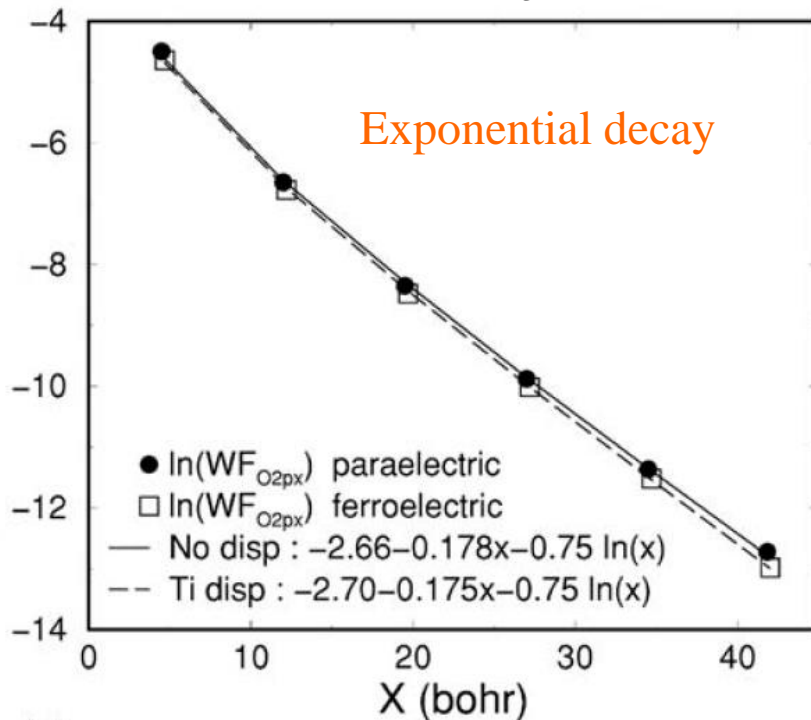
$$n(r,r') = \sum_n n_{ij,R_n}\chi_i(r-\tau_i)\chi_j(r'-\tau_j-R_n)$$
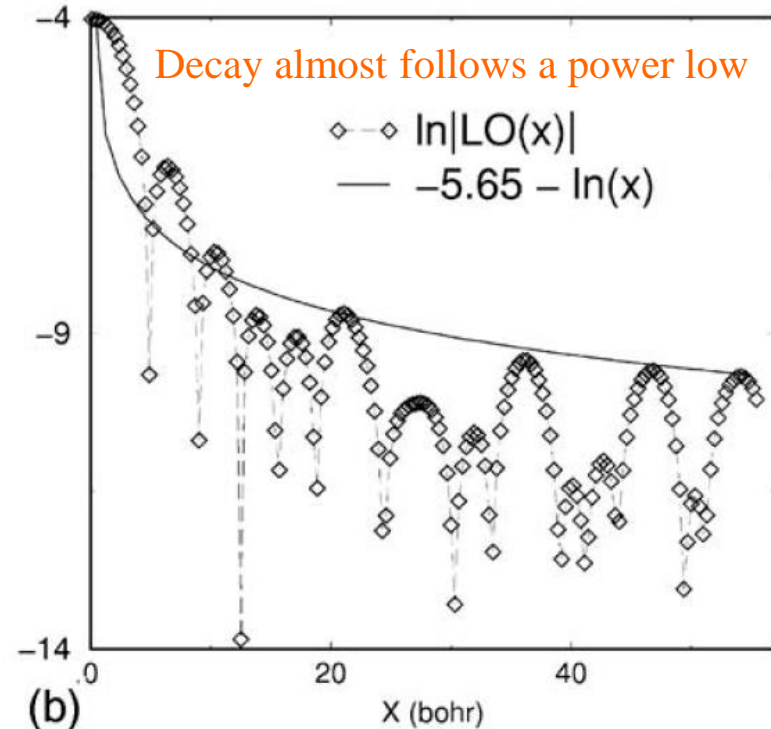
where the matrix representation is given by

$$n_{ij,R_n} = \frac{1}{V_B}\int\limits_B dk \sum_\mu^{occ} \exp(ik\bullet R_n)c_{\mu i,k}c_{\mu j,k}$$

Locality of Wannier functions

O-2px in PbTiO$_3$

An orbital in Aluminum

Exponential decay

Decay almost follows a power low

$\ln(WF_{O2px})$ paraelectric
$\ln(WF_{O2px})$ ferroelectric
No disp : $-2.66-0.178x-0.75\ln(x)$
Ti disp : $-2.70-0.175x-0.75\ln(x)$

$\ln|LO(x)|$
$-5.65-\ln(x)$

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

Wannier functions decay exponentially for semi-conductors and insulators, while for metals they decay algebraically. A mathematical analysis for 1D systems is found in He and Vanderbilt, PRL 86, 5341. A conditional proof for general cases is discussed in Brouder et al., PRL 98, 046402.

Locality of density matrix
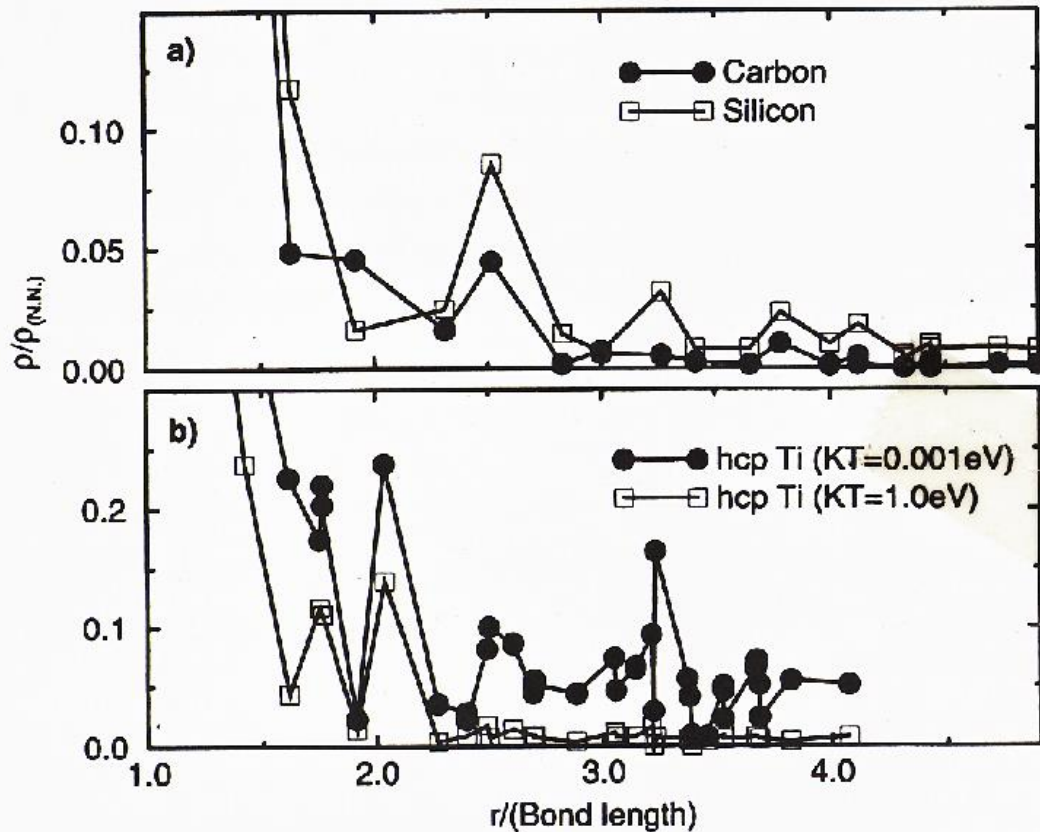
Finite gap systems exponential decay

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

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

At T = 0 K, the density matrix elements decay exponentially for semi-conductors and insulators, while for metals they decay algebraically. For a finite temperature, they decay exponentially even for metals. A mathematical analysis is found in Ismail-Beigi et al, PRL 82, 2127.

Various linear scaling methods

Wannier functions (WF)

Density matrix (DM)

} × {

Variational (V)

Perturbative (P)
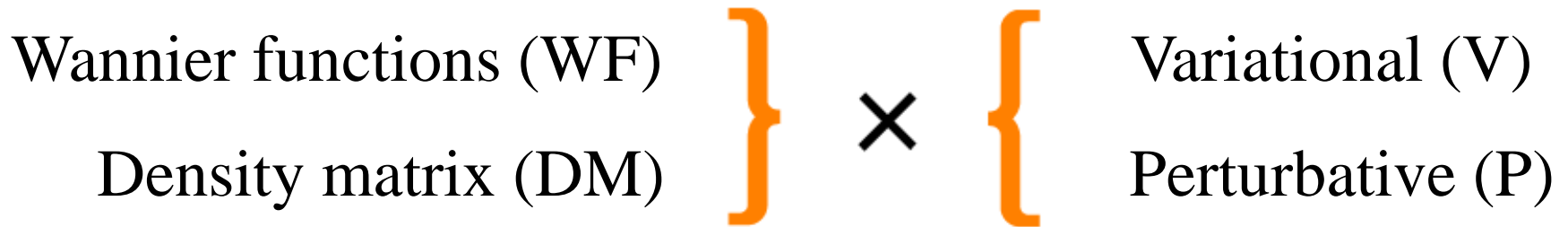
At least four kinds of linear-scaling methods
can be considered as follows:

WF+V

Orbital
minimization
by Galli, Parrinello,
and Ordejon

WF+P

Hoshi
Mostofi

DM+V

Density matrix
by Li and Daw

DM+P

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)

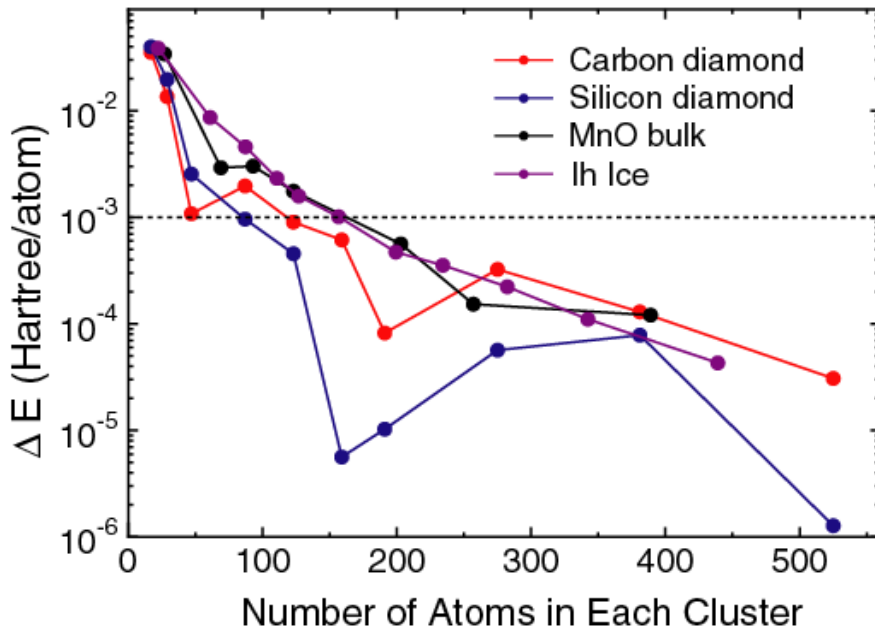# Basic idea behind the O(N) method
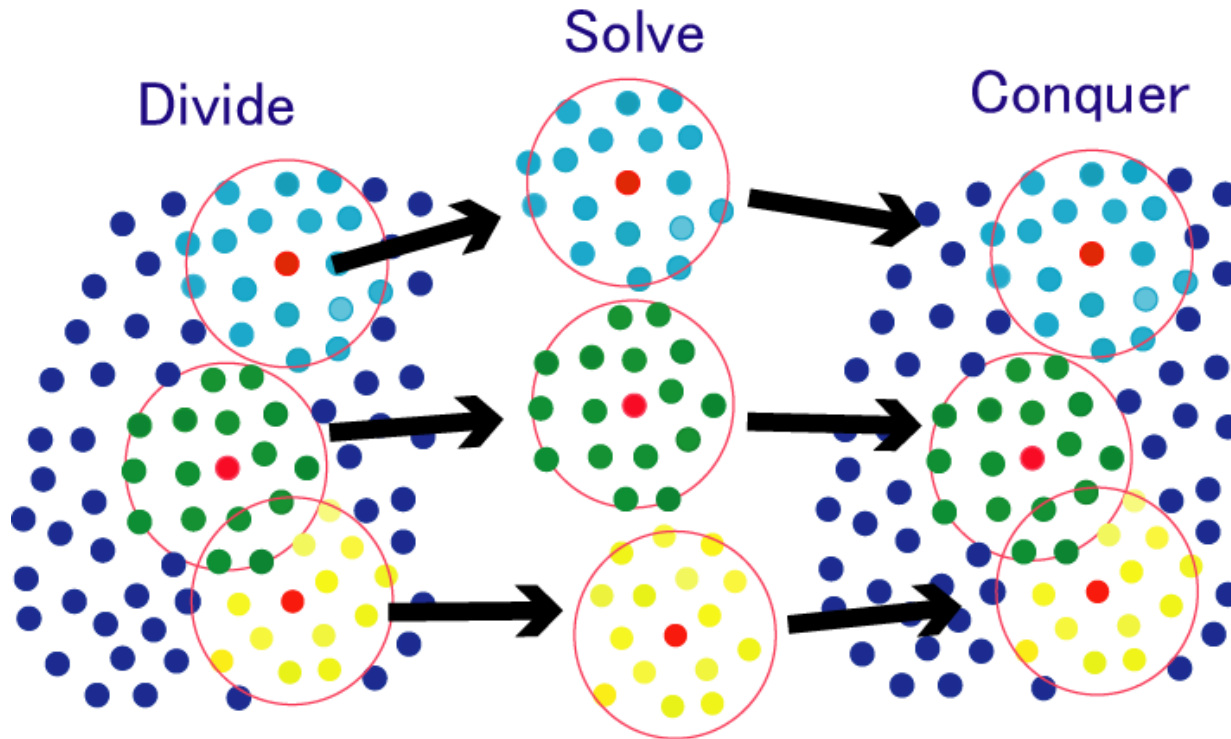
Assumption

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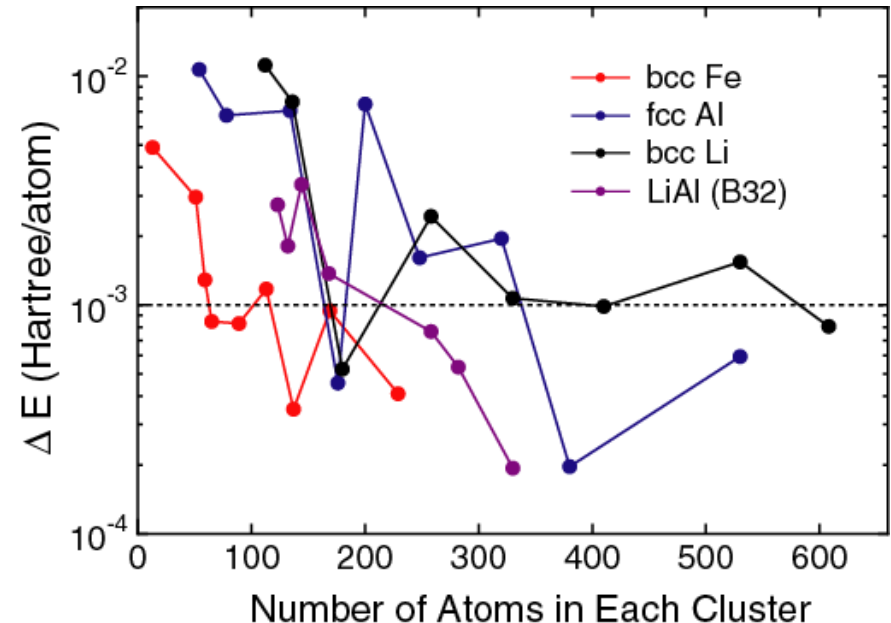
Convergence by the DC method

Just solve the truncated clusters → Divide-Conquer method

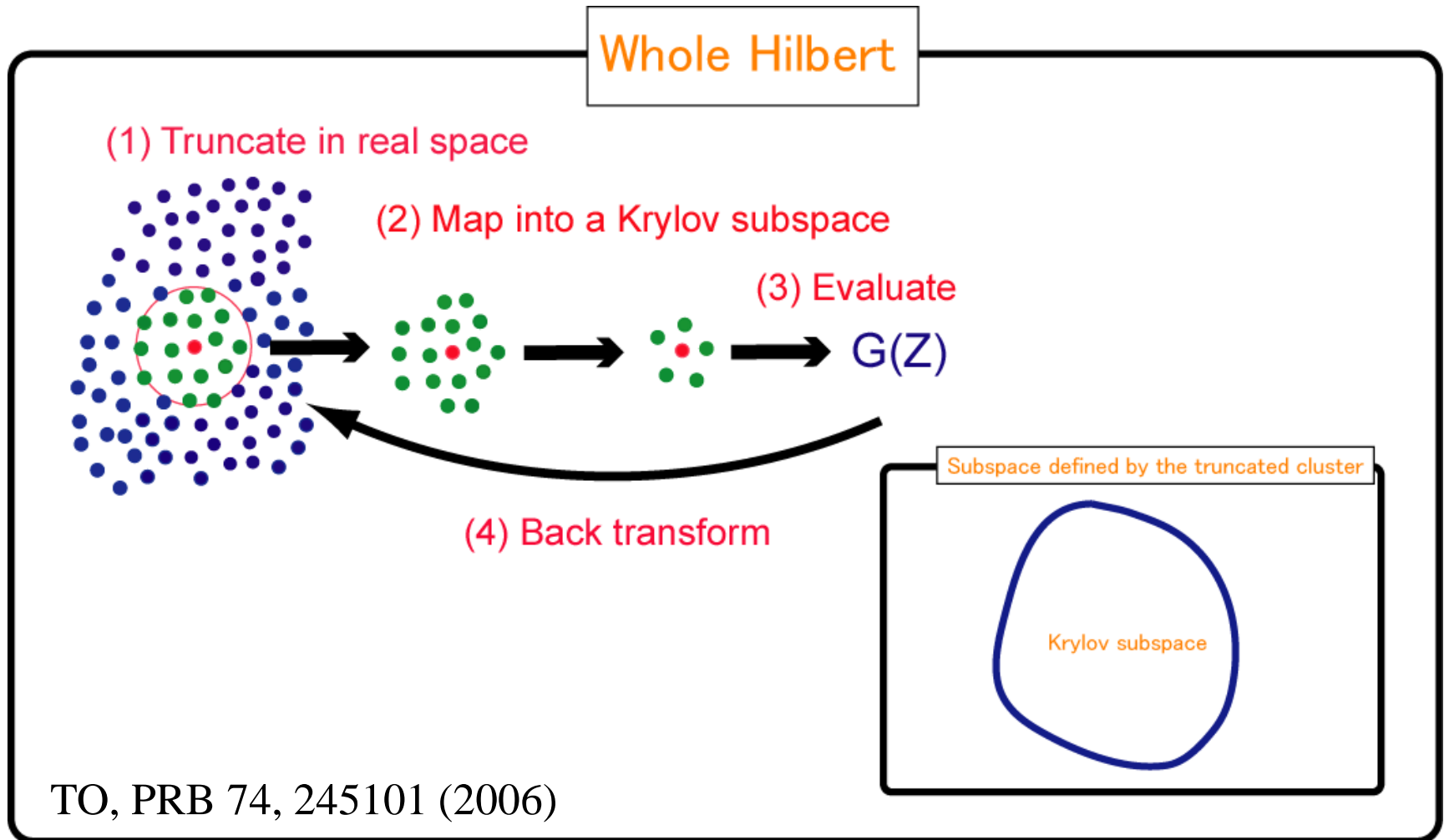W.Yang, PRL 66, 1438 (1991)

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

Whole Hilbert

(1) Truncate in real space

(2) Map into a Krylov subspace

(3) Evaluate

$G(Z)$

(4) Back transform

Subspace defined by the truncated cluster

Krylov subspace

TO, PRB 74, 245101 (2006)

O(N) methods based on Krylov subspace

- ## Based on Lanczos algorithms

  R. Haydock, V. Heine, and M. J. Kelly, J. Phys. C 5, 2845 (1972); R. Haydock, Solid State Phys. 35, 216 (1980).
  T. Ozaki, Phys. Rev. B 59, 16061 (1999); T. Ozaki, M. Aoki, and D. G. Pettifor, ibid. 61, 7972 (2000).

- ## Based on a two-sided block Lanczos algorithm

  T. Ozaki and K. Terakura, Phys. Rev. B 64, 195126 (2001).
  T. Ozaki, Phys. Rev. B 64, 195110 (2001).

- ## Based on an Arnoldi type algorithm

  T. Ozaki, Phys. Rev. B 74, 245101 (2006).

Power method

Can we obtain a convergent result by repeatedly multiplying a random vector by an Hermite matrix H?

The initial vector v_0 can be rewritten by a linear combination.

$$v_1 = Hv_0$$
$$v_2 = Hv_1$$
$$\cdots$$
$$v_n = Hv_{n-1}$$
$$v_\infty \rightarrow \quad ???$$

$$|v_0\rangle = \sum_i a_i |c_i\rangle \qquad H|c_i\rangle = \varepsilon_i |c_i\rangle$$

v_0 is multiplied by H n-th times.

$$
\begin{aligned}
H^n |v_0\rangle &= \left(\sum_k \varepsilon_k |c_k\rangle\langle c_k| \right)^n |v_0\rangle \\
&= \left(\sum_k \varepsilon_k |c_k\rangle\langle c_k| \right)^n \sum_i a_i |c_i\rangle \\
&= \left(\sum_k \varepsilon_k^n |c_k\rangle\langle c_k| \right) \sum_i a_i |c_i\rangle \\
&= \sum_i a_i \varepsilon_i^n |c_i\rangle \\
&\approx a_0 \varepsilon_0^n |c_0\rangle
\end{aligned}
$$

Thus, we see that it converges to the vector corresponding to the largest eigenvalue. Also, it is found that degenerate cases may lead to slow convergence.

ε_0 is the largest eigenvalue in its absolute value.

What is the Krylov subspace?

The Krylov subspace is defined by the following set of vectors:

$$\left(\left| u_0 \right\rangle, \hat{H} \left| u_0 \right\rangle, \hat{H}^2 \left| u_0 \right\rangle, \hat{H}^3 \left| u_0 \right\rangle, \cdots, \hat{H}^q \left| u_0 \right\rangle \right)$$

The Krylov subspace methods try to solve the eigenvalue problem within the subspace, while the power method takes account of only a single vector.

The Lanczos method is one of the most widely used technique based on the Krylov subspace.

Lanczos method

The Lanczos method is an algorithm which generates a Krylov subspace by choosing a vector orthogonal to a subspace generated by the previous step. By repeating the algorithm, one can expand the Krylov subspace step by step.

Idea

Tri-diagonalizaton of a Hermite matrix.

$$H_{\text{TD}} = U^{\dagger} H U.$$

$$H_{\text{TD}} = \begin{pmatrix} \alpha_0 & \beta_1 & & & & \\ \beta_1 & \alpha_1 & \beta_2 & & & \\ & \cdots & \cdots & & & \\ & & \cdots & \cdots & & \\ & & & \beta_{N-1} & \alpha_{N-1} & \beta_N \\ & & & & \beta_N & \alpha_N \end{pmatrix}$$
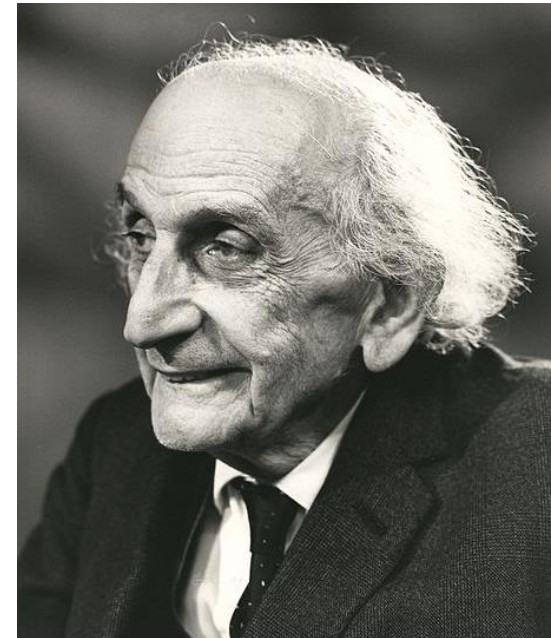
Cornelius Lanczos, 1893-1974

How can we find the unitary matrix?

Quoted from http://guettel.com/lanczos/

Derivation of Lanczos method #1

Writing $H_{TD} = U^{\dagger} H U$ explicitly, ..

$$H\{|u_0>, |u_1>, |u_2>, ..., |u_N>\} = \{|u_0>, |u_1>, |u_2>, ..., |u_N>\} \times$$

$$\begin{pmatrix} \alpha_0 & \beta_1 & & & & \\ \beta_1 & \alpha_1 & \beta_2 & & & \\ & \cdots & & \cdots & & \\ & & & \cdots & & \cdots & \\ & & & \beta_{N-1} & \alpha_{N-1} & \beta_N \\ & & & & \beta_N & \alpha_N \end{pmatrix},$$

We further write column by column.

$$H|u_0> = |u_0>\alpha_0 + |u_1>\beta_1,$$
$$H|u_1> = |u_0>\beta_1 + |u_1>\alpha_1 + |u_2>\beta_2,$$
$$\cdots$$
$$H|u_n> = |u_{n-1}>\beta_n + |u_n>\alpha_n + |u_{n+1}>\beta_{n+1},$$

Then, one has the following three terms recurrence formula:

$$\boxed{|u_{n+1}>\beta_{n+1} = H|u_n> - |u_{n-1}>\beta_n - |u_n>\alpha_n}$$

Derivation of Lanczos method #2

Thus, starring from a given u_0, we can recursively calculate u_n. The process can be summarized as the following algorithm.

$$\text{Set} \qquad < u_0| = (1, 0, 0, ...)$$

$$\text{Compute} \qquad H|u_n >$$

$$\text{Compute} \qquad \alpha_n = < u_n|H|u_n >$$

$$\text{Compute} \qquad |r_n >= H|u_n > - |u_{n-1} > \beta_n - |u_n > \alpha_n$$

$$\text{Compute} \qquad \beta_{n+1} = \sqrt{< r_n|r_n >}$$

$$\text{Compute} \qquad |u_{n+1} >= |r_n > /\beta_{n+1}$$

$$n := n + 1$$

Relation between Lanczos method and Green's function #1

Using the tri-diagonal matrix obtained from the Lanczos transformation, we have an useful expression.

$$G_{\text{TD}}(Z) = (ZI - H_{\text{TD}})^{-1}$$

$$H_{\text{TD}} = \begin{pmatrix} \alpha_0 & \beta_1 & & & & \\ \beta_1 & \alpha_1 & \beta_2 & & & \\ & \cdots & \cdots & & & \\ & & \cdots & \cdots & & \\ & & & \beta_{N-1} & \alpha_{N-1} & \beta_N \\ & & & & \beta_N & \alpha_N \end{pmatrix}$$

Relation between Lanczos method and Green's function #2

The determinant for the tri-diagonal matrix can be expressed by a recurrence formula.

$$\det(ZI - H_{\text{TD}}) = (Z - \alpha_0)A_{11} - \beta_1 A_{12},$$
$$A_{11} = D_1,$$
$$A_{12} = \beta_1 D_2. \quad \longrightarrow \quad D = (Z - \alpha_0)D_1 - \beta_1^2 D_2$$

In general, $\quad D_n = (Z - \alpha_n)D_{n+1} - \beta_{n+1}^2 D_{n+2}$

which is called Laplace expansion.

Using the recurrence formula, one can evaluate the diagonal term of Green's function.

Finally, we have a continued fraction.

$$
\begin{aligned}
G_{00}^L(Z) &= \frac{D_1}{D} \\
&= \frac{D_1}{(Z - \alpha_0)D_1 - \beta_1^2 D_2} \\
&= \frac{1}{Z - \alpha_0 - \dfrac{\beta_1^2 D_2}{D_1}} \\
&= \frac{1}{Z - \alpha_0 - \dfrac{\beta_1^2 D_2}{(Z - \alpha_1)D_2 - \beta_2^2 D_3}}
\end{aligned}
$$

$$
G_{00}^L(Z) = \cfrac{1}{Z - \alpha_0 - \cfrac{\beta_1^2}{Z - \alpha_1 - \cfrac{\beta_2^2}{Z - \alpha_2 - \cfrac{\beta_3^2}{\ddots}}}}.
$$

Green's function and physical quantities

Let's us calculate the imaginary part of Green's function.

$$g(Z) = \frac{1}{Z - E_0}$$

$$\begin{aligned}
\operatorname{Im} g(E + i\varepsilon) &= \frac{1}{2i}\left(\frac{1}{Z - E_0} - \frac{1}{Z^* - E_0}\right) \\
&= \frac{1}{2i}\left(\frac{1}{E - E_0 + i\varepsilon} - \frac{1}{E - E_0 - i\varepsilon}\right)
\end{aligned}$$

$$\operatorname{Im} g(E + i\varepsilon) = \frac{-\varepsilon}{(E - E_0)^2 + \varepsilon}$$

Integrating the imaginary part

$$\begin{aligned}
\int_{-\infty}^{\infty} \operatorname{Im} g(E + i\varepsilon) dE \\
&= \int_{-\infty}^{\infty} \frac{-\varepsilon}{(E - E_0)^2 + \varepsilon} \\
&= -\varepsilon\left[\frac{1}{\varepsilon} \tan^{-1} \frac{E - E_0}{\varepsilon}\right]_{-\infty}^{\infty} \\
&= -\pi
\end{aligned}$$

The following is a plot of the imaginary part.

Thus,

$$\boxed{\lim_{\varepsilon \to 0} -\frac{1}{\pi}\operatorname{Im} g(E + i\varepsilon) = \delta(E - E_0)}$$
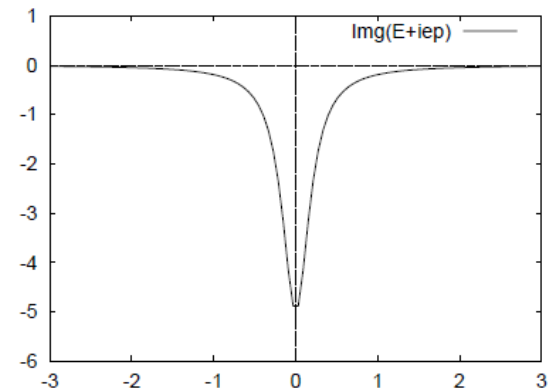
The imaginary part of diagonal part of Green's function is the density of states.

A mathematical analysis on accuracy of O(N) methods

1D tight-binding model #1

By analyzing a 1D-TB model, we discuss accuracy of O(N) methods for gapped and metallic systems.

$$
\begin{array}{c}
\quad\quad \varphi_{-4}\quad \varphi_{-3}\quad \varphi_{-2}\quad \varphi_{-1}\quad \varphi_{0}\quad \varphi_{1}\quad \varphi_{2}\quad \varphi_{3}\quad \varphi_{4}
\end{array}
$$

$$
H =
\begin{array}{c c}
& \begin{array}{c c c c c c c}
\varphi_{0} & \varphi_{1} & \varphi_{-1} & \varphi_{2} & \varphi_{-2} & \varphi_{3} & \varphi_{-3}
\end{array} \\
\begin{array}{c}
\varphi_{0} \\ \varphi_{1} \\ \varphi_{-1} \\ \varphi_{2} \\ \varphi_{-2} \\
\end{array}
&
\left(
\begin{array}{c c c c c c c}
a & b & b & 0 & 0 & 0 & .. \\
b & a & 0 & b & 0 & 0 & .. \\
b & 0 & a & 0 & b & 0 & .. \\
0 & b & 0 & a & 0 & b & .. \\
0 & 0 & b & 0 & a & 0 & b.. \\
.. & .. & .. & .. & .. & .. & ..
\end{array}
\right)
\end{array}
$$

By assuming that the on-site energy is a, and the nearest hopping integral is b, we have the matrix representation above.

1D tight-binding model #2

By applying the Lanczos algorithm to the 1D TB, we transform the model to a semi-infinite model. The following is the procedure.

(1)

$$|u_0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \end{pmatrix}$$

$$|r_1\rangle = H|u_0\rangle - |u_{-1}\rangle\beta_0 - |u_0\rangle\alpha_0,$$

(4)

$$= \begin{pmatrix} 0 \\ b \\ b \\ 0 \\ \cdot \\ \cdot \end{pmatrix}$$

(7)

$$H|u_1\rangle = \begin{pmatrix} 2b \\ a \\ a \\ b \\ b \\ 0 \\ \cdot \\ \cdot \end{pmatrix}$$

(10)

$$\beta_2^2 = \langle r_2|H|r_2\rangle = b^2$$

A similar calculation continues.

(2)

$$H|u_0\rangle = \begin{pmatrix} a \\ b \\ b \\ \cdot \\ \cdot \end{pmatrix}$$

(5)

$$\beta_1^2 = \langle r_1|H|r_1\rangle = 2b^2$$

(8)

$$\alpha_1 = \langle u_1|H|u_1\rangle = a$$

(6)

$$|u_1\rangle = \frac{|r_1\rangle}{\beta_1},$$

(9)

$$|r_2\rangle = H|u_1\rangle - |u_0\rangle\beta_1 - |u_1\rangle\alpha_1,$$

(3)

$$\alpha_0 = \langle u_0|H|u_0\rangle = a$$

$$= \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ \cdot \\ \cdot \end{pmatrix}$$

$$= \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 0 \\ 0 \\ b \\ b \\ 0 \\ \cdot \\ \cdot \end{pmatrix}$$

In summary,

$$\alpha_n = a \quad \text{Arbitrary n}$$

$$\beta_1 = \sqrt{2}b$$

$$\beta_n = b \quad n \neq 1$$

1D tight-binding model #3

Orthogonal bases are generated starting from the initial site, and hopping to the next sites.

$|u_0\rangle$

$|u_1\rangle$

$|u_2\rangle$

In summary

$$\alpha_n = a \quad \text{Arbitrary n}$$

$$\beta_1 = \sqrt{2}b$$
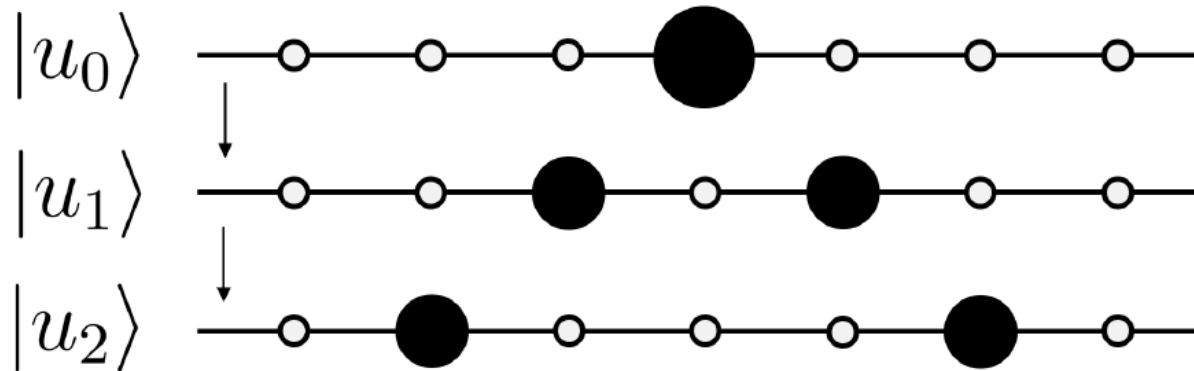
$$\beta_n = b \quad n \neq 1$$

$$|u_0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \end{pmatrix} \quad |u_1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ \cdot \\ \cdot \end{pmatrix} \quad |u_2\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ \cdot \\ \cdot \end{pmatrix}$$

Any system can be transformed to a semi-infinite chain model using the Lanczos algorithm.

1D tight-binding model #4

(1) The diagonal term of Green's function is express by a continued fraction.

The case of K=2 is the DOS of the 1D model.
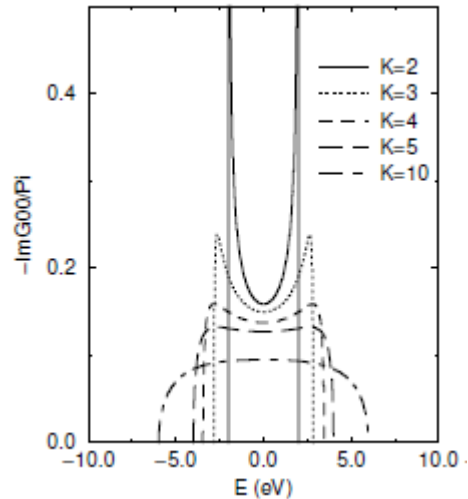
$$G_{00}^{L}(Z) = \cfrac{1}{Z - \alpha_0 - \cfrac{\beta_1^2}{Z - \alpha_1 - \cfrac{\beta_2^2}{Z - \alpha_2 - \cfrac{\beta_3^2}{\ddots}}}}$$

$$= \cfrac{1}{Z - a - \cfrac{2b^2}{Z - a - \cfrac{b^2}{Z - a - \cfrac{b^2}{\ddots}}}}$$

Noting the similarity structure, the last term is obtained.

$$= \cfrac{1}{\sqrt{(Z-a)^2 - 4b^2}}$$

(2) The off-diagonal term of Green's function is express by a recurrence formula.

$$G_{00}^{L}(Z)[ZI - H^L] = I$$

$$G_{0n}^{L} = \frac{1}{\beta_n}\left[G_{0(n-1)}(Z - \alpha_{n-1}) - G_{0(n-2)}\beta_{n-1} - \delta_{1n}\right]$$

1D tight-binding model #5

The off-diagonal term can be expressed by G^L_{00} via the recurrence formula.

$$G^L_{01}(Z) = \frac{\gamma}{\sqrt{2}} G^L_{00}(Z) - \frac{1}{\sqrt{2}b} \qquad G^L_{02}(Z) = \left(\frac{\gamma^2}{\sqrt{2}} - 1 \right) G^L_{00}(Z) - \frac{\gamma}{\sqrt{2}b}$$

$$\gamma = \frac{Z-a}{b}$$

By Taylor-expanding G^L_{00} around $\gamma^{-1} = 0$, one has

$$G^L_{00}(Z) = \frac{1}{\sqrt{2}b} \left(1 + \frac{2}{\gamma^3} + \frac{6}{\gamma^5} + \frac{20}{\gamma^7} + \frac{70}{\gamma^9} + \cdots \right)$$

By inserting the Taylor-expanded G^L_{00} to G^L_{0n}, one obtain the following leading term.

$$G^L_{0n}(Z) \propto \frac{\sqrt{2}}{b\gamma^{n+1}} \qquad \gamma^{-1} < 1 \text{ corresponding to } \left| \frac{b}{Z-a} \right| < 1$$

Under the condition, G^L_{0n} converges to zero as $n \rightarrow \infty$.

1D tight-binding model #6

The density matrix n_{0i} is defined by

$$n_{0n} = \int dE \sum_{\mu} \langle 0 | \psi_{\mu} \rangle \langle \psi_{\mu} | n \rangle \, \delta(E - \varepsilon_{\mu}) f(\varepsilon_{\mu})$$

Rewriting the expression above by Green's function, we have

$$n_{0n} = -\frac{1}{\pi} \mathrm{Im} \int dE \, G_{0n}(E + i0^{+}) f(\varepsilon_{\mu})$$

Using the Cauchy theorem, the integral path can be changed.
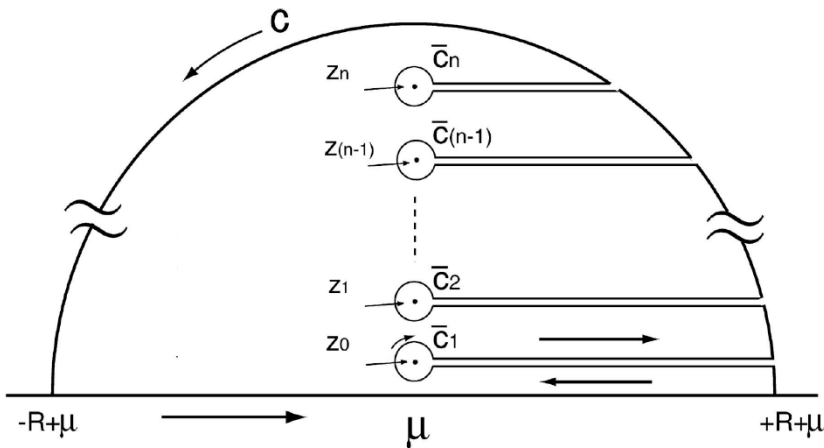
Noting that the Fermi function has the Matsubara poles, we can derive the following formula.

$$n_{0n} = M^{(0)} + \mathrm{Im}\left[-\frac{4i}{\beta} \sum_{p} G\left(\mu + i\frac{z_{p}}{\beta} \right) R_{p} \right]$$

Asymptotic behaviors of G_{0n}

Beyond the circle, the off-diagonal elements of Green's function behave as

$$G_{0n} \propto \frac{1}{y^{n+1}}$$

Metal

Insulator

where y is defined by the Fermi energy μ and a band width of w.

$$y = \frac{z - \mu}{w/4}$$

Matsubara poles in complex plane

Matsubara poles in complex plane

$$r = \frac{w}{4}$$

$$r = \frac{w}{4}$$

$$r = \frac{w}{4}$$
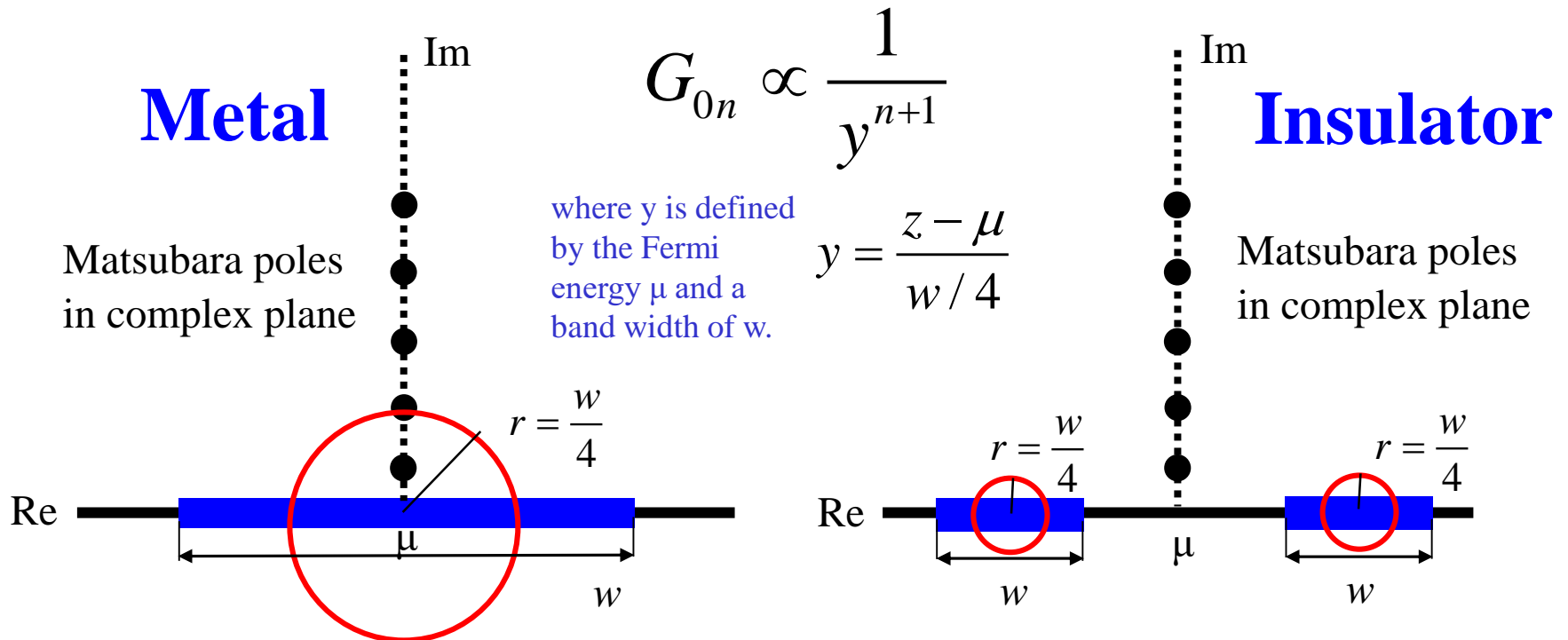
At z=μ, the off-diagonal elements of Green's function behave as

$$G_{0(2m-1)} = (-1)^m \frac{2\sqrt{2}}{w} \qquad G_{0(2m)} = (-1)^m \sqrt{2} G_{00}$$

In the red circle, the Green's function does not localize in real space.
→ leading to long-range correlation.

Extension of O(N) Krylov subspace methods to DFT

- ## Based on Lanczos algorithms

  R. Haydock, V. Heine, and M. J. Kelly, J. Phys. C 5, 2845 (1972); R. Haydock, Solid State Phys. 35, 216 (1980).
  T. Ozaki, Phys. Rev. B 59, 16061 (1999); T. Ozaki, M. Aoki, and D. G. Pettifor, ibid. 61, 7972 (2000).

- ## Based on a two-sided block Lanczos algorithm

  T. Ozaki and K. Terakura, Phys. Rev. B 64, 195126 (2001).
  T. Ozaki, Phys. Rev. B 64, 195110 (2001).

- ## Based on an Arnoldi type algorithm

  T. Ozaki, Phys. Rev. B 74, 245101 (2006).

How can we take account of the overlap matrix S ?

$$Hc_{\mu} = \varepsilon_{\mu} S c_{\mu}$$

O(N) Krylov subspace method

Two step mapping of the whole Hilbert space into subspaces

TO, PRB 74, 245101 (2006)
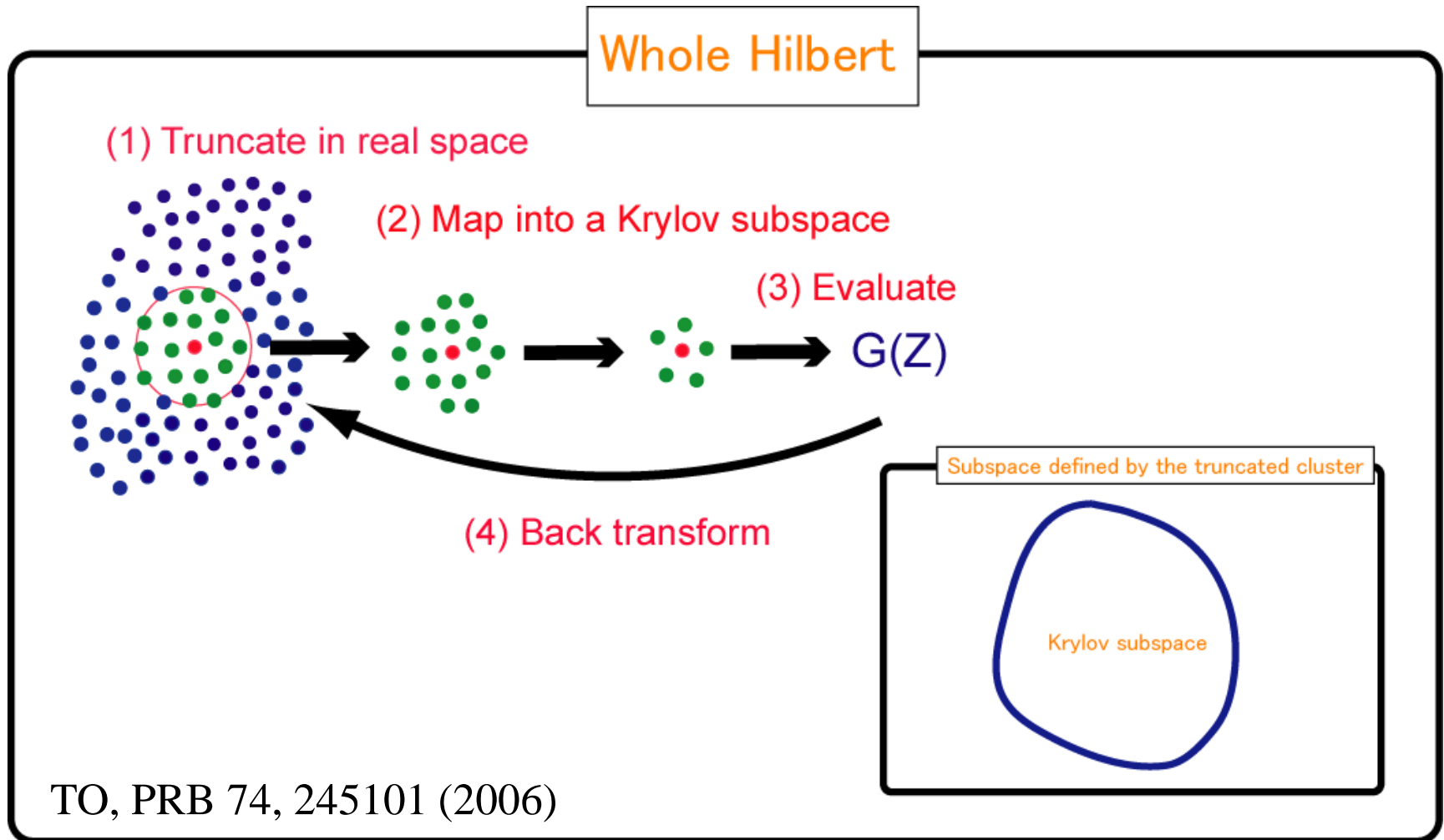
Development of Krylov subspace vectors

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.
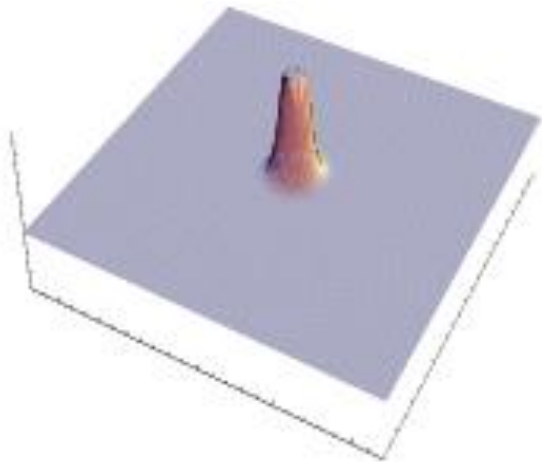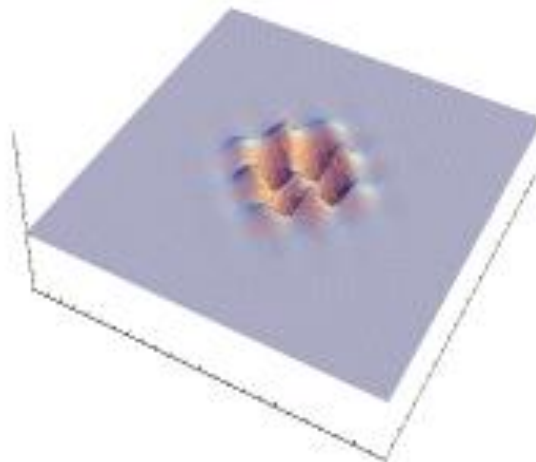
$|K_0>$ $|K_1>$ $|K_5>$
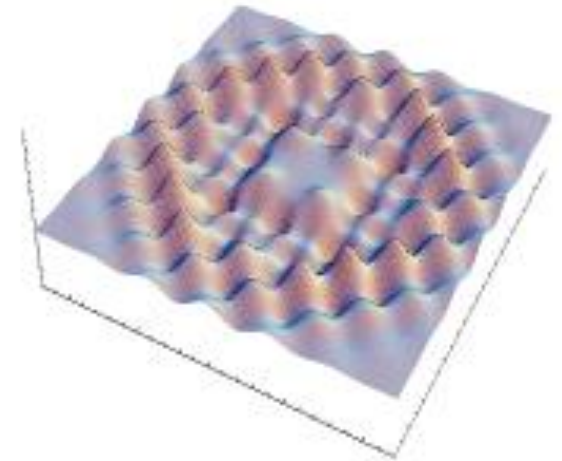
The information on *environment* can be included from near sites step by step, resulting in reduction of the dimension.

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.

$$
\boxed{|R_{n+1}) = S^{-1}H|W_n)}
$$

$$
|W'_{n+1}) = |R_{n+1}) - \sum_{m=0}^{n} |W_m)(W_m|\hat{S}|R_{n+1})
$$

$$
|W_{n+1}) = S - \text{orthonormalized block vector of } |W'_{n+1})
$$

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

$$
\mathbf{U}_K = \mathbf{W}\mathbf{X}\lambda^{-1}
$$

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

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:

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

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

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

updated   **fixed**

Green: core region
Yellow: buffer region
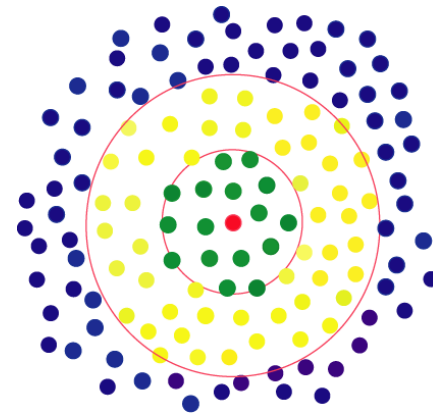
- **The embedded cluster is under the Coulomb interaction from the other parts.**
- **The charge flow from one embedded cluster to the others is allowed.**

Relation between the Krylov subspace and Green's funtion

A Krylov subspace is defined by

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

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

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

Definition of moments

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

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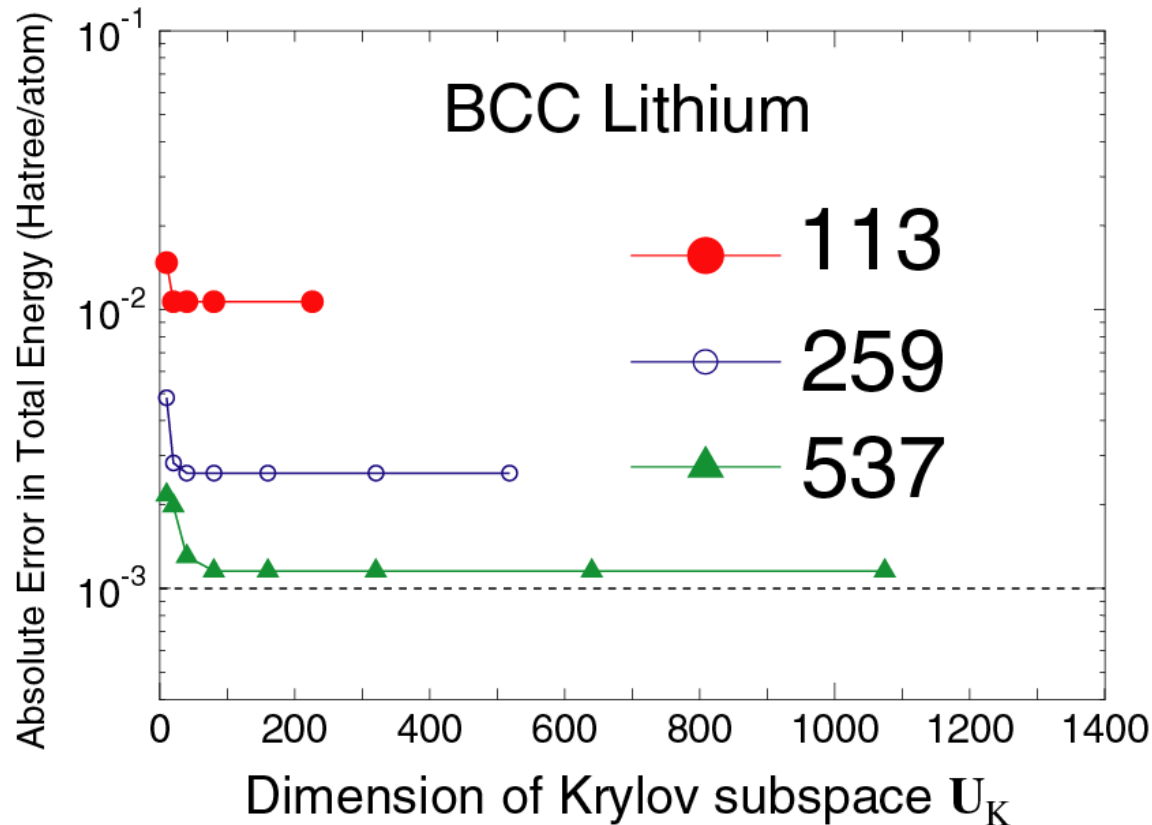
In general, the convergence property is more complicated. See PRB 74, 245101 (2006).

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.
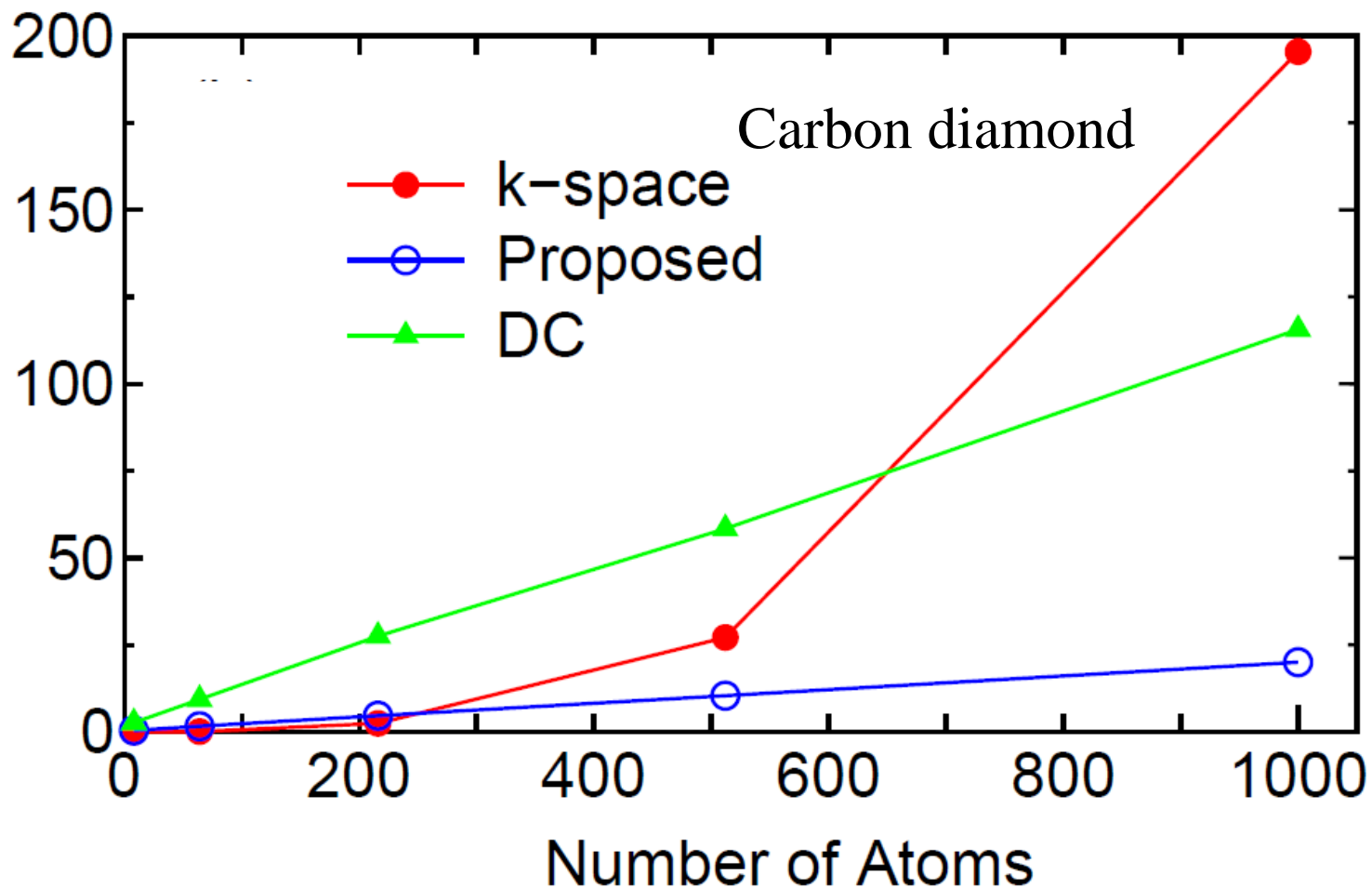
Parallelization

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).
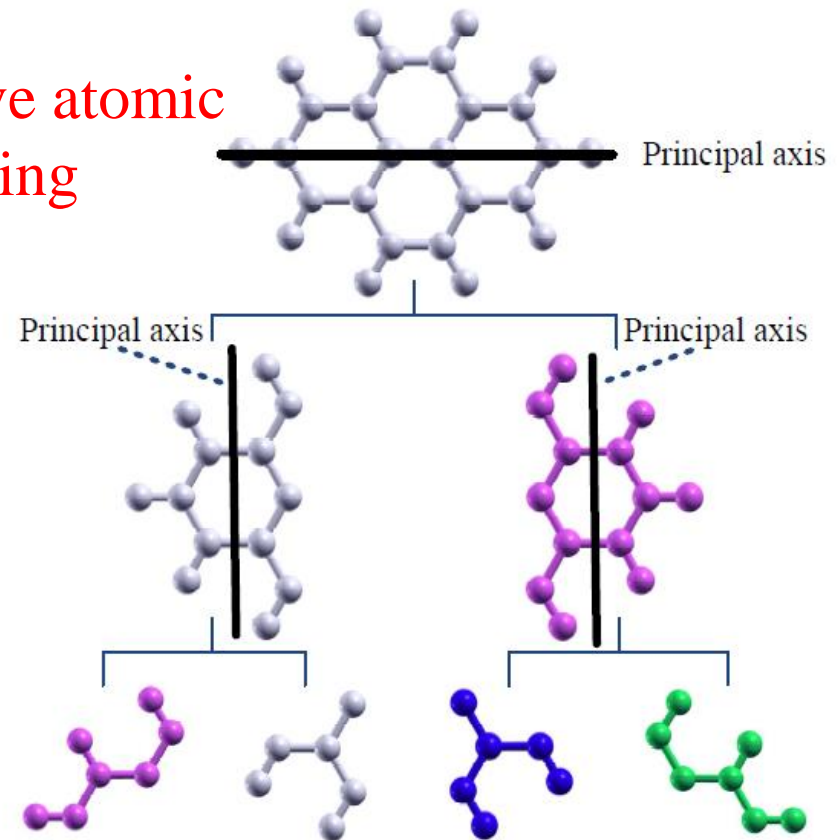
Modified recursive bisection

If the number of MPI processes is 19, then the following binary tree structure is constructed.
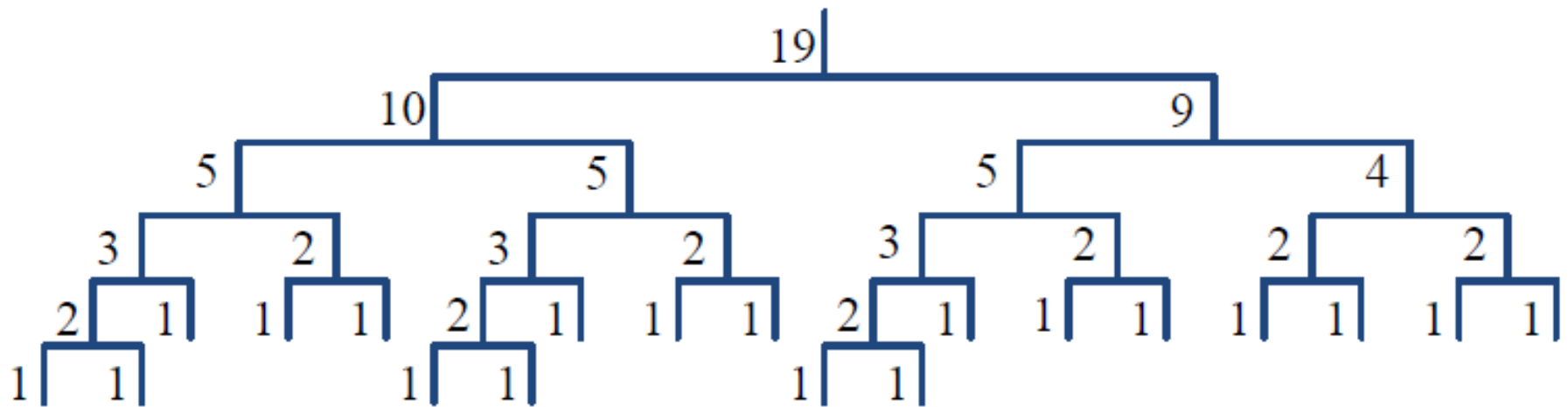
In the conventional recursive bisection, the bisection is made so that a same number can be assigned to each region. However, the modified version bisects with weights as shown above.

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.

3D

1D
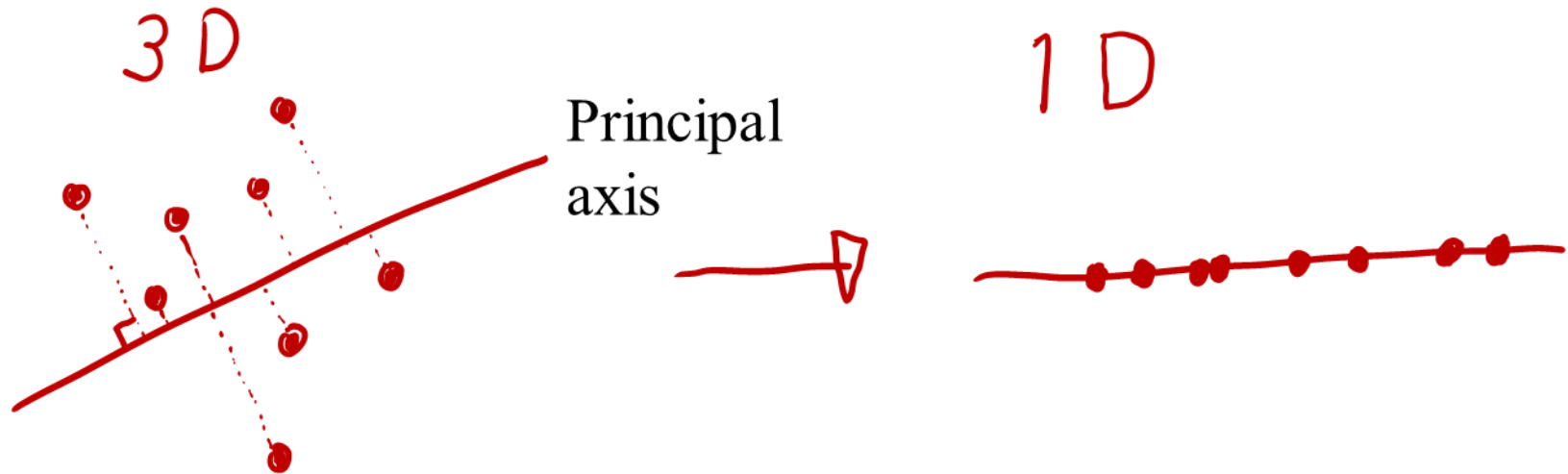
Principal axis

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
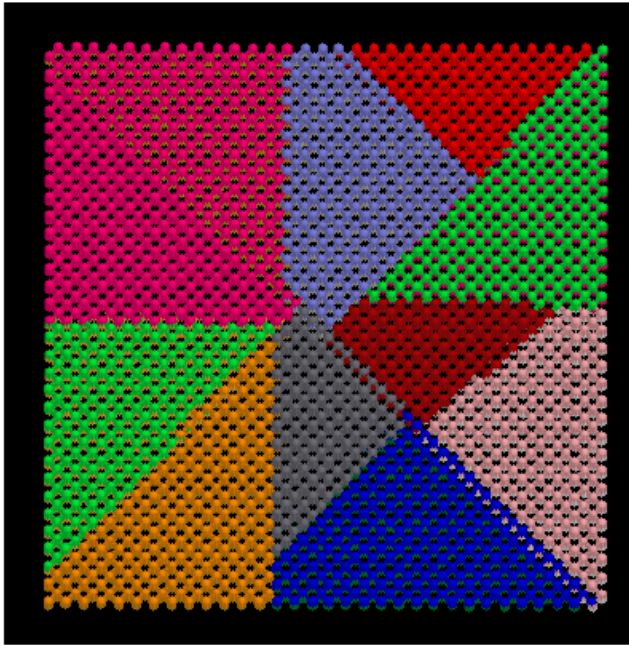
Diamond 16384 atoms, 19 processes
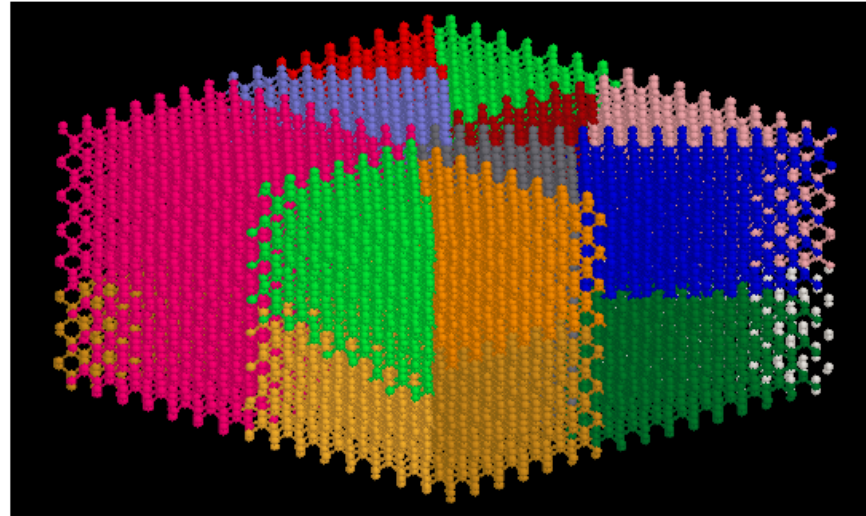
Multiply connected CNT, 16 processes

Parallel efficiency on K

Diamond structure consisting of 131,072 atoms
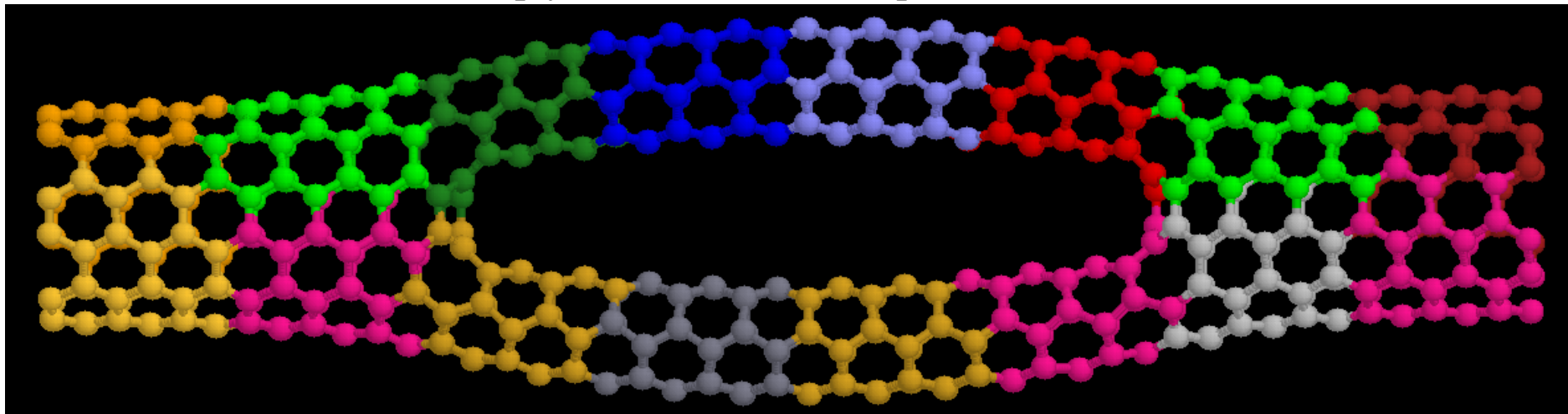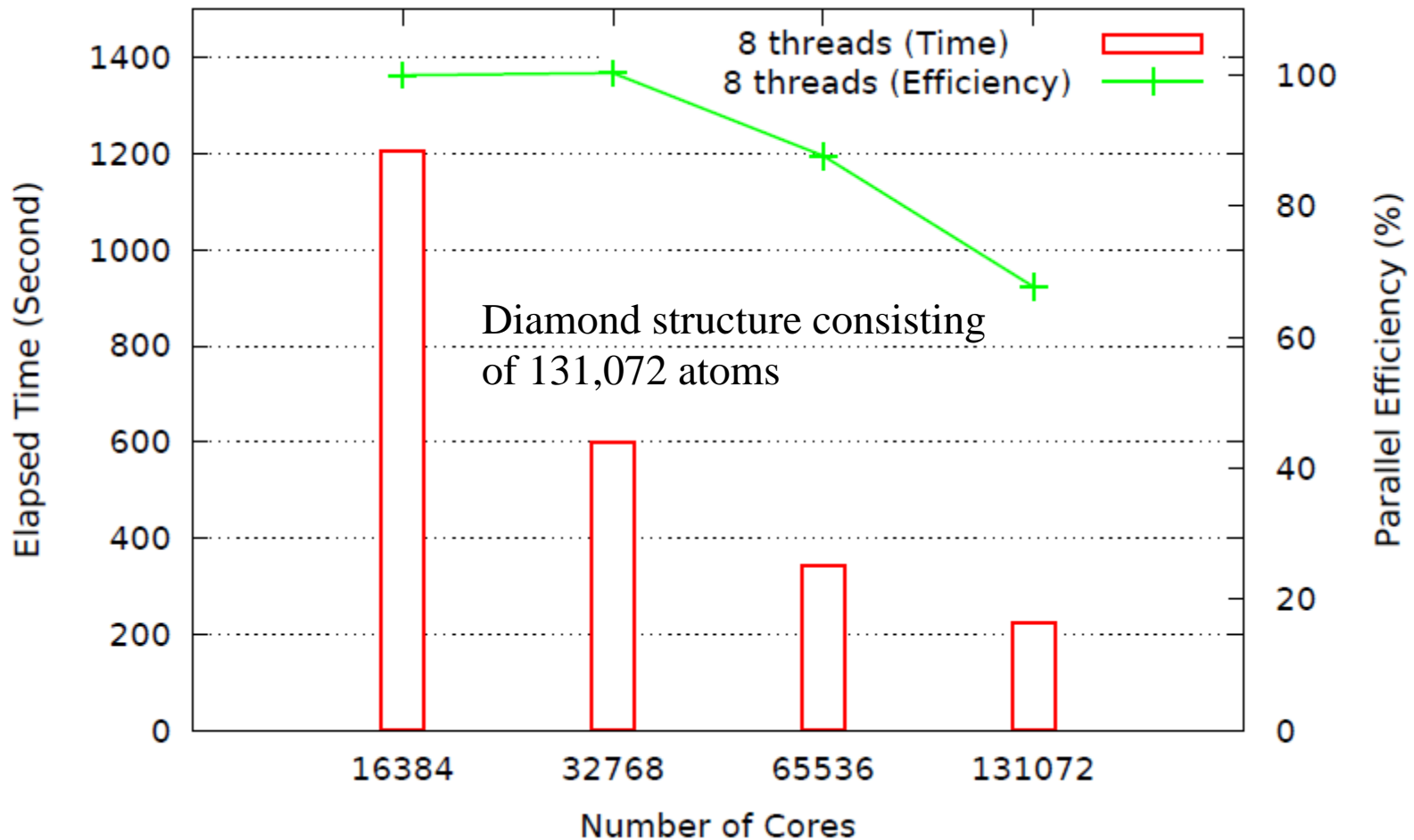
The parallel efficiency is 68 % using 131,072 cores.

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).
   H. Sawada et al., Metals 7, 277 (2017).

2. Desolvation of Li^+

   T. Ohwaki et al., J. Chem. Phys. 136, 134101 (2012).
   T. Ohwaki et al., J. Chem. Phys. 140, 244105 (2014).
   T. Ohwaki et al., Phys. Chem. Chem. Phys. 20, 11586 (2018).

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).
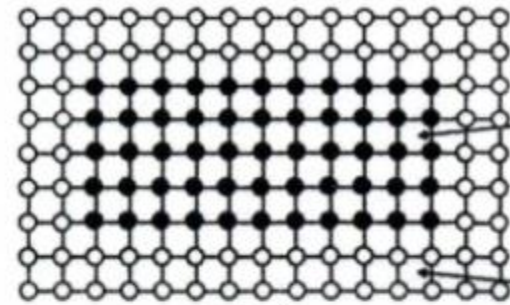
Precipitation in bcc-Fe

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

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

TiC

TiC precipitate steel

3D atom probe image

Ti

10nm

Hardness per precipitate F (N)

Diameter of precipitates R (nm)
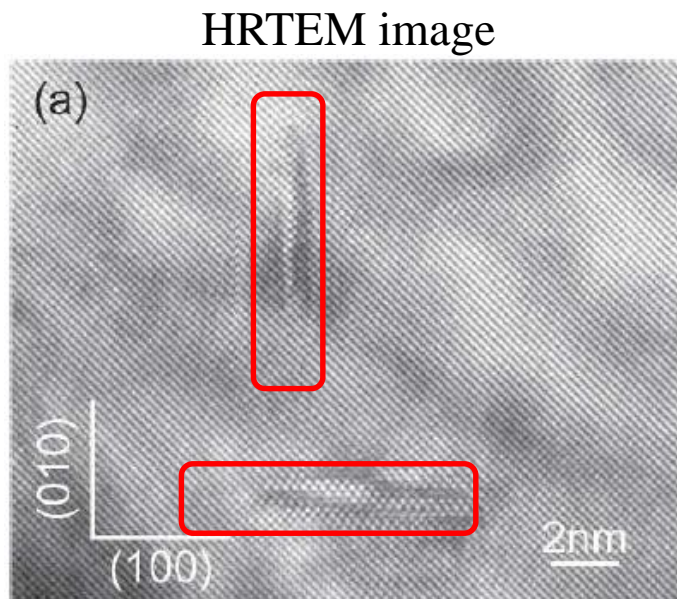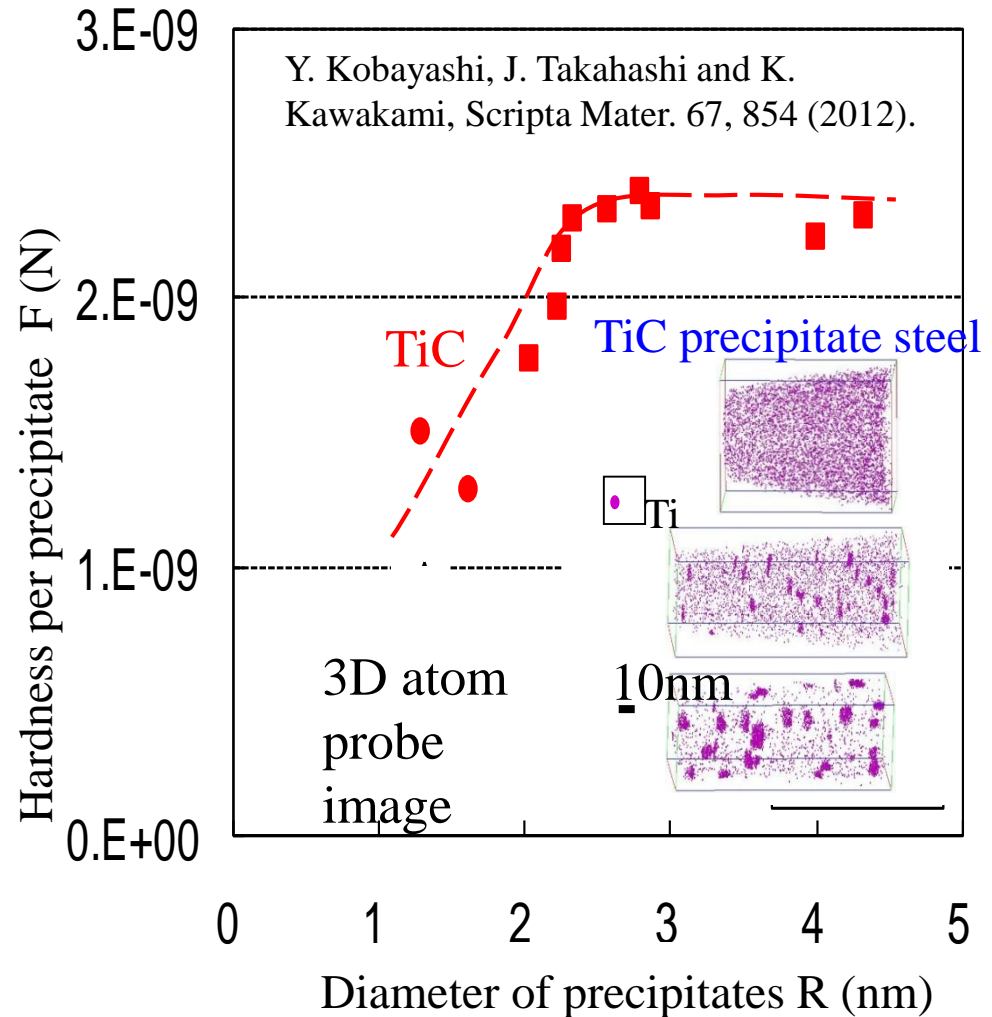
**Precipitating materials:
TiC, VC, NbC**

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
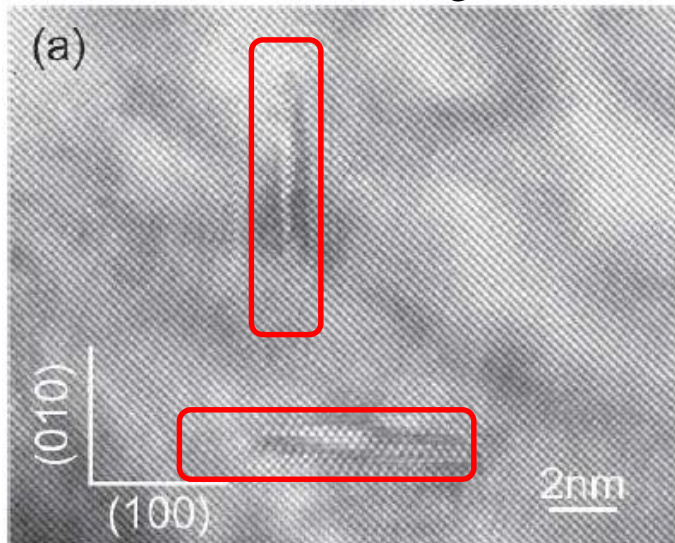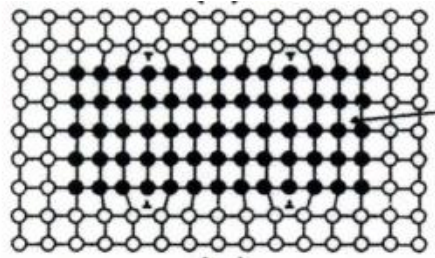

Coherent precipitation
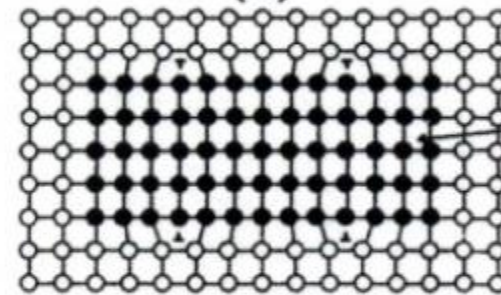
Semicoherent precipitation
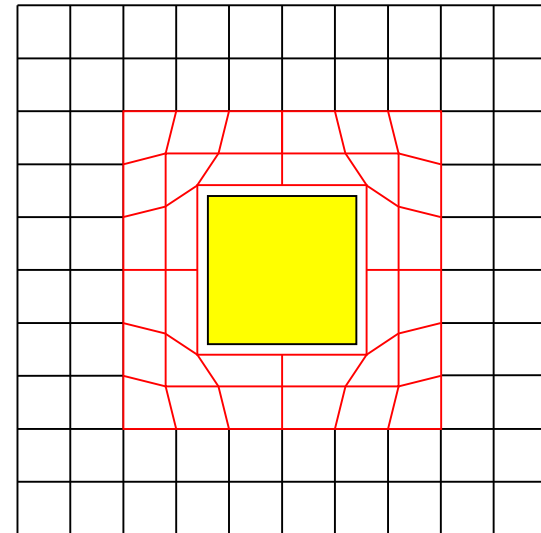
Incoherent precipitation

Precipitating materials: TiC, VC, NbC
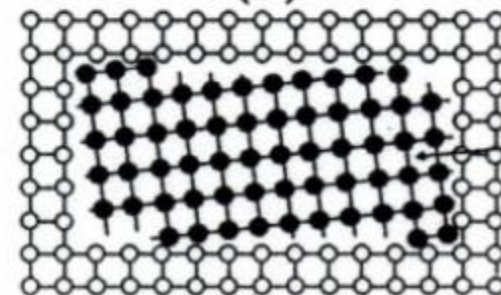
Interface and strain energies

Semi-coherent case

Strain field

Interface energy per area

coherent interface + strain

semi-coherent interface + strain

semi-coherent interface

strain (coherent)

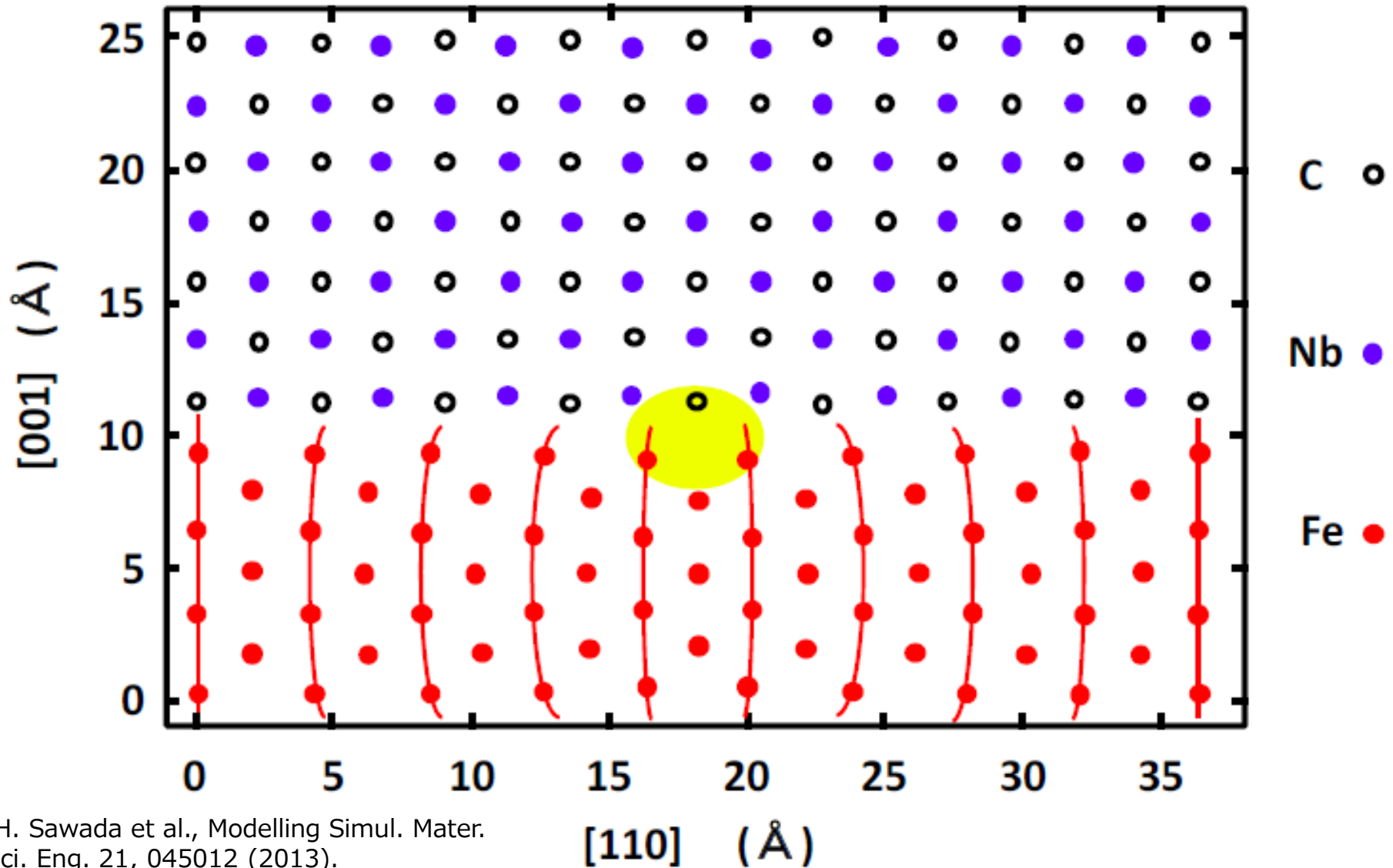coherent interface

strain (semicoherent)

Diameter of precipitate
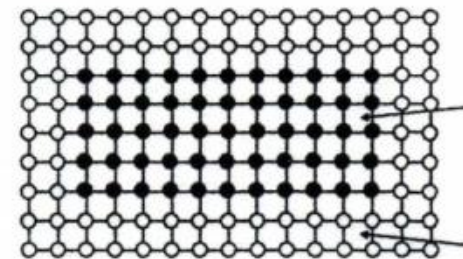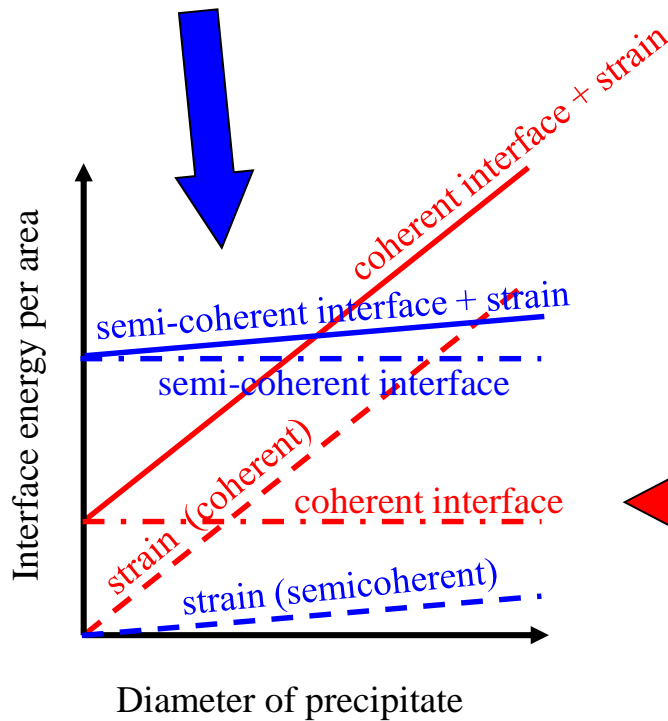
coherent case

Iron

Optimized semi-coherent interface structure

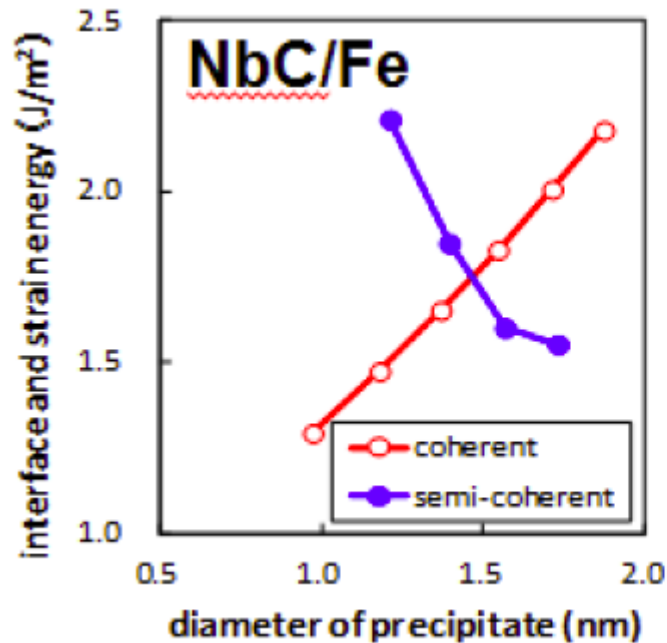

H. Sawada et al., Modelling Simul. Mater.
Sci. Eng. 21, 045012 (2013).

Estimation of strain energy

Model potential method： Finnis-Sinclair

Fe atoms：432,000

Precipitation

Mother phase

8.94Å

Precipitation

Mother phase

Coherent： 10% expansion
Semi-coherent：0% expansion

4% expansion
（Due to dislocation）
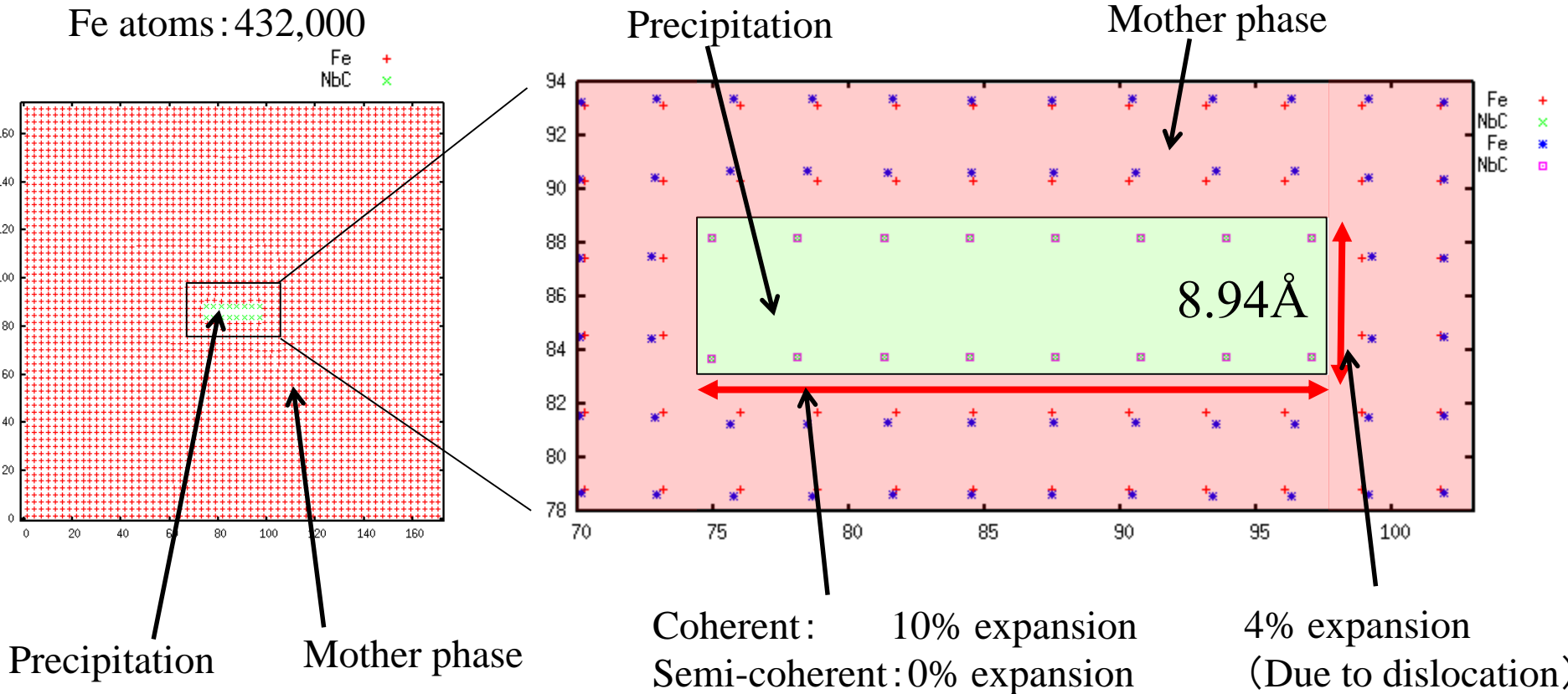
Transition of coherent/semi-coherent interface structure
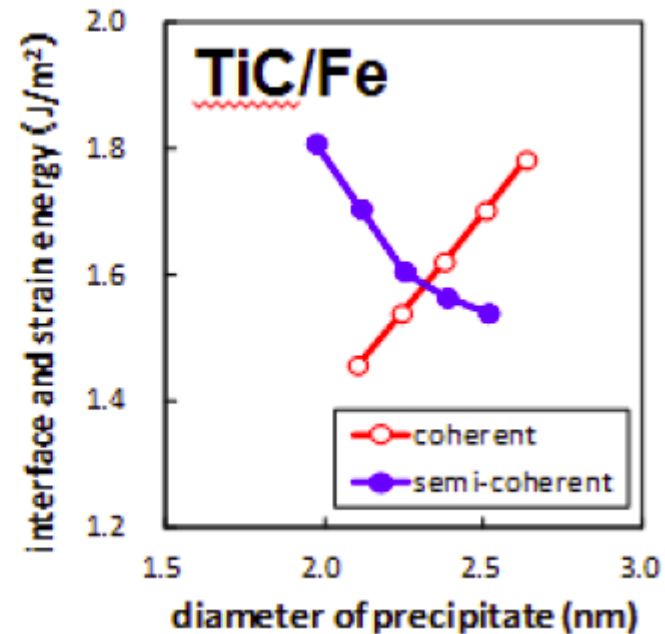
| Calc. | 2.3 nm | |
| Expt. | 2～3 nm | **for TiC/Fe** |

estimated by TEM images and structural properties.

Kobayashi et al., Scripta Materialia 67, 854 (2012).

Outlook

The locality of density matrix and basis function is a key to develop a wide variety of efficient electronic structure methods. In the lecture we have focused theories of O(N) methods, its practical implementations, and discussed applications. By making full use of the locality, in addition to the development of O(N) methods, it may be possible to develop the following methods:

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

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