Large-scale electronic structure methods in OpenMX

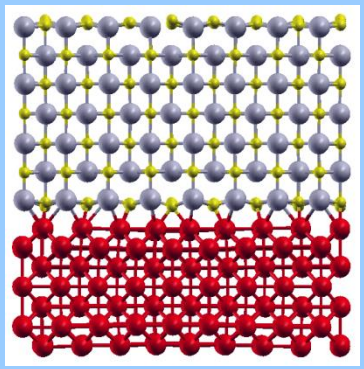
- Introduction
- O(N) Krylov subspace method
- Applications
- Summary

Taisuke Ozaki (ISSP, Univ. of Tokyo)
Nov. 23rd, OpenMX hands-on workshop in KAIST

Towards first-principle studies for industry

10^3 – 10^6 atom

Steel



DFT calculations of thousands atoms is still a grand challenge.

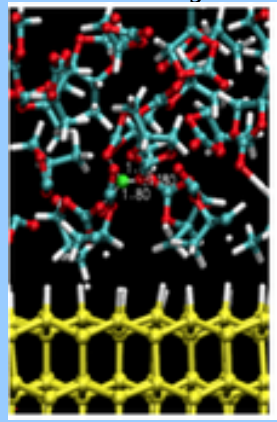
$O(N^3)$ → Low-order

10^2 atom

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

DNA

Battery



Time scale

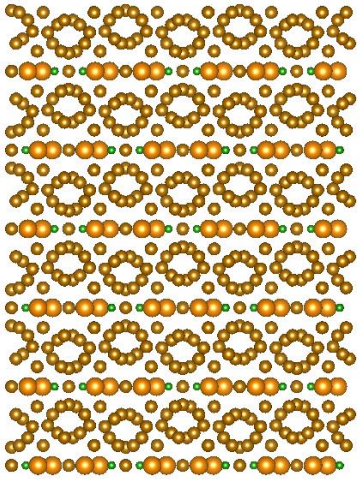
System size

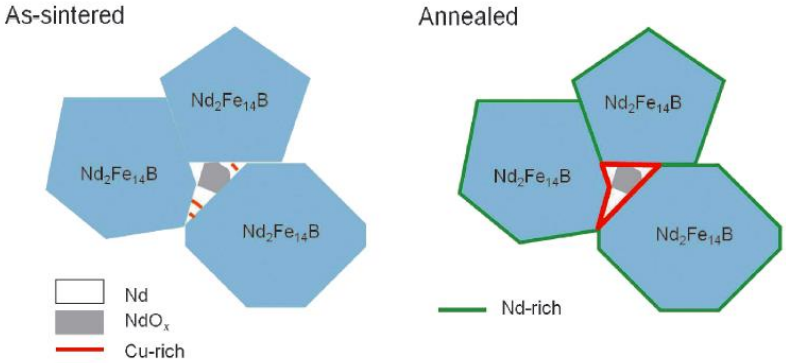
Materials properties

- 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.



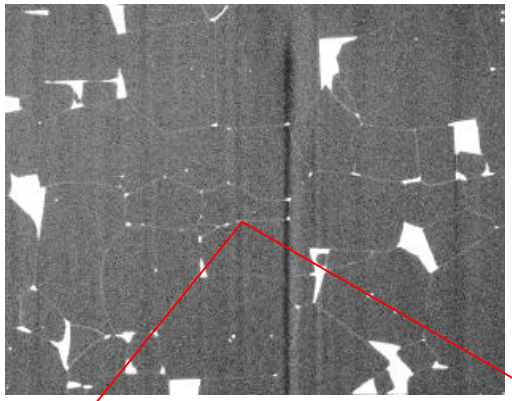


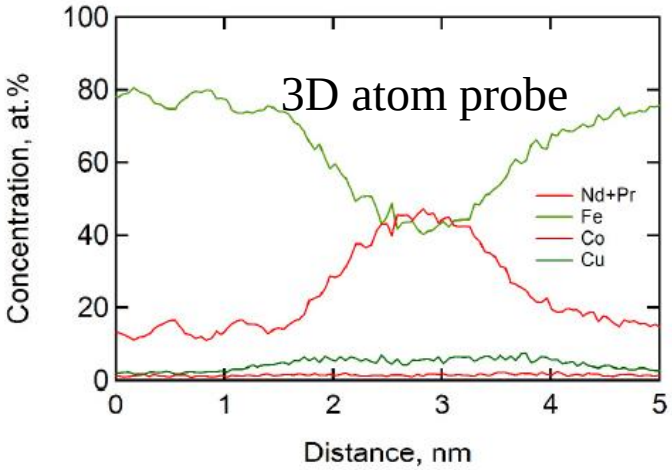
Experimental analysis of inhomogeneous materials

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

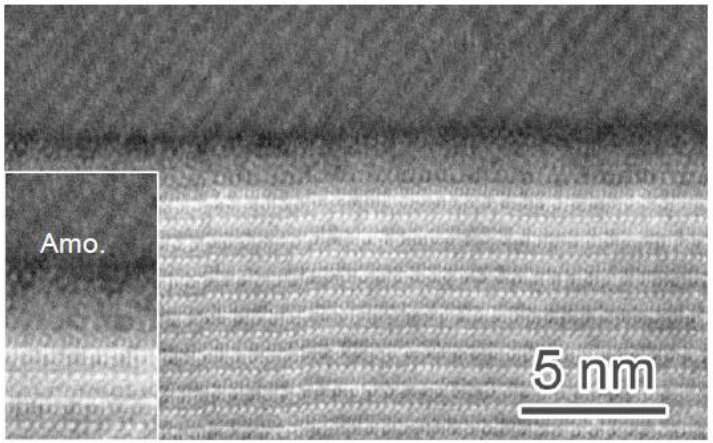
Hono@NIMS

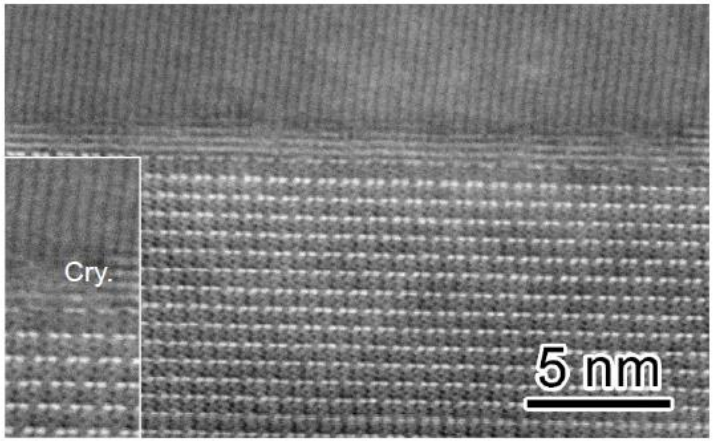
TEM





Around grain boundary

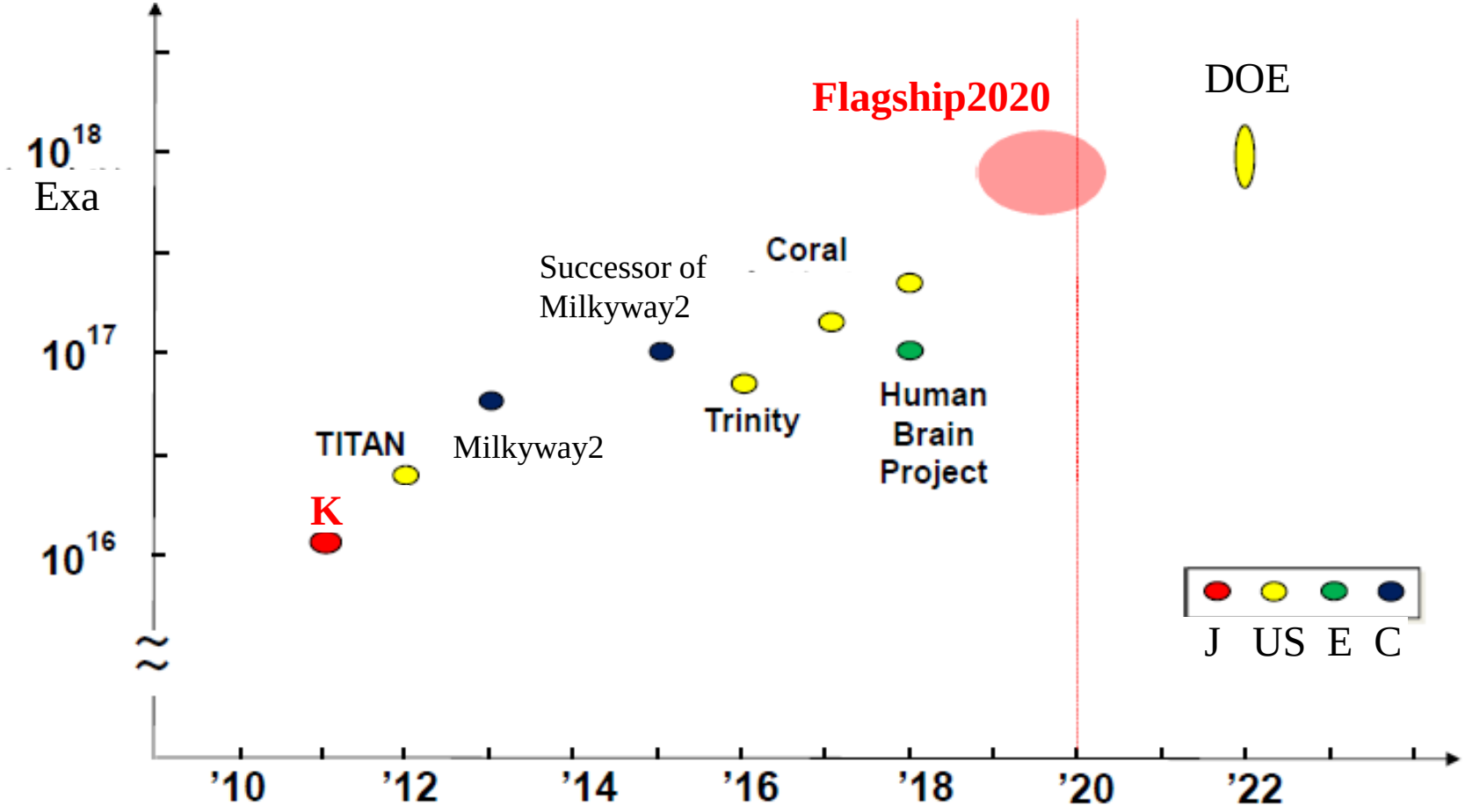




K(京)-Computer: 10 Peta flops machine

CPU: SPARC64 VIIIfx,
2GHz quad cores (128 GFLOPS)
2 processors/node
80000 nodes = **640,000 cores**
Memory: 1 Peta Byte

Development of computing power



How large systems can be treated by Exa machines?

The performance increase is only 100 times.

K-computer → Exa machine

10 PFLOPS → 1000 PFLOPS

Computational Scaling $O(N^p)$	Computable size
7	1.7
6	2.2
5	2.5
4	3.2
3	4.6
2	10
1	100

← DFT

The applicability of the $O(N^3)$ DFT method is extended to only 5 times larger systems.

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

Density functionals as a functional of ρ

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

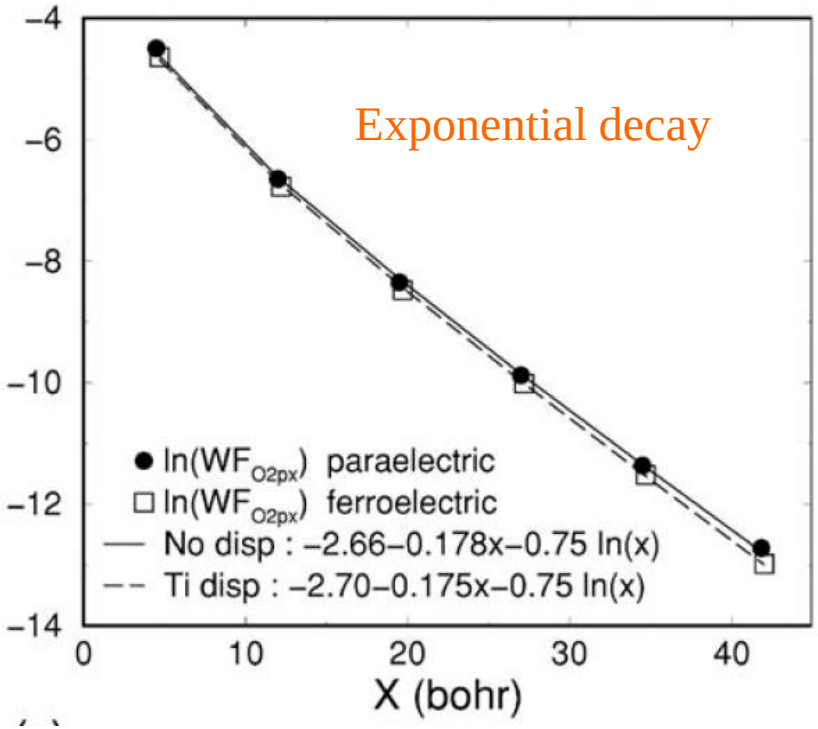
$$E_{\mathrm{tot}}[n,\rho] = \mathrm{Tr}(\rho H_{\mathrm{kin}}) + \int d\mathbf{r} n(\mathbf{r}) v_{\mathrm{ext}}(\mathbf{r}) + \int\int d\mathbf{r} d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\mathrm{xc}}[n]$$

where the electron density is given by ρ

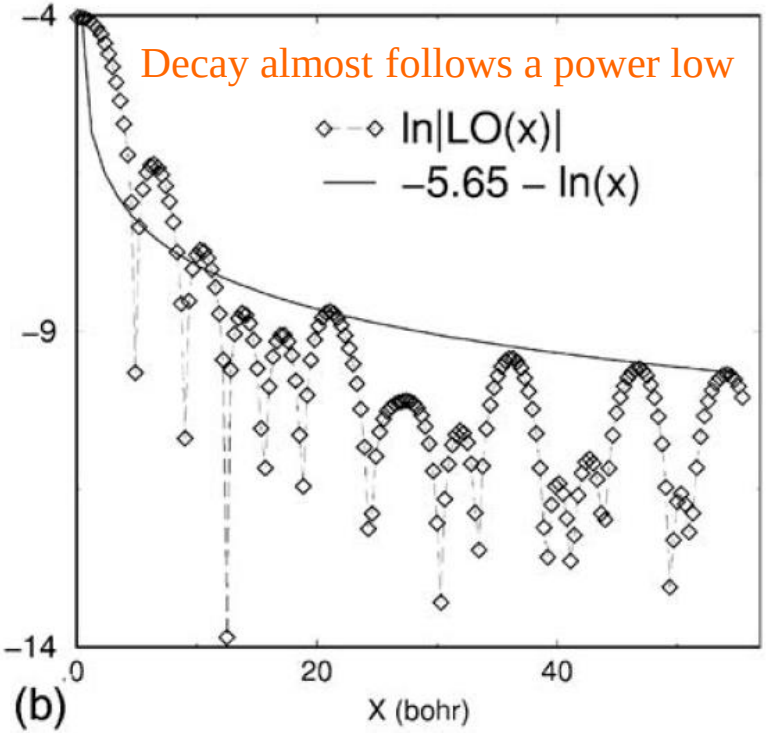
$$n(\mathbf{r}) = \sum_{i,j} \rho_{ij} \chi_j(\mathbf{r}) \chi_i(\mathbf{r})$$

Locality of Wannier functions

O-2px in $PbTiO_3$

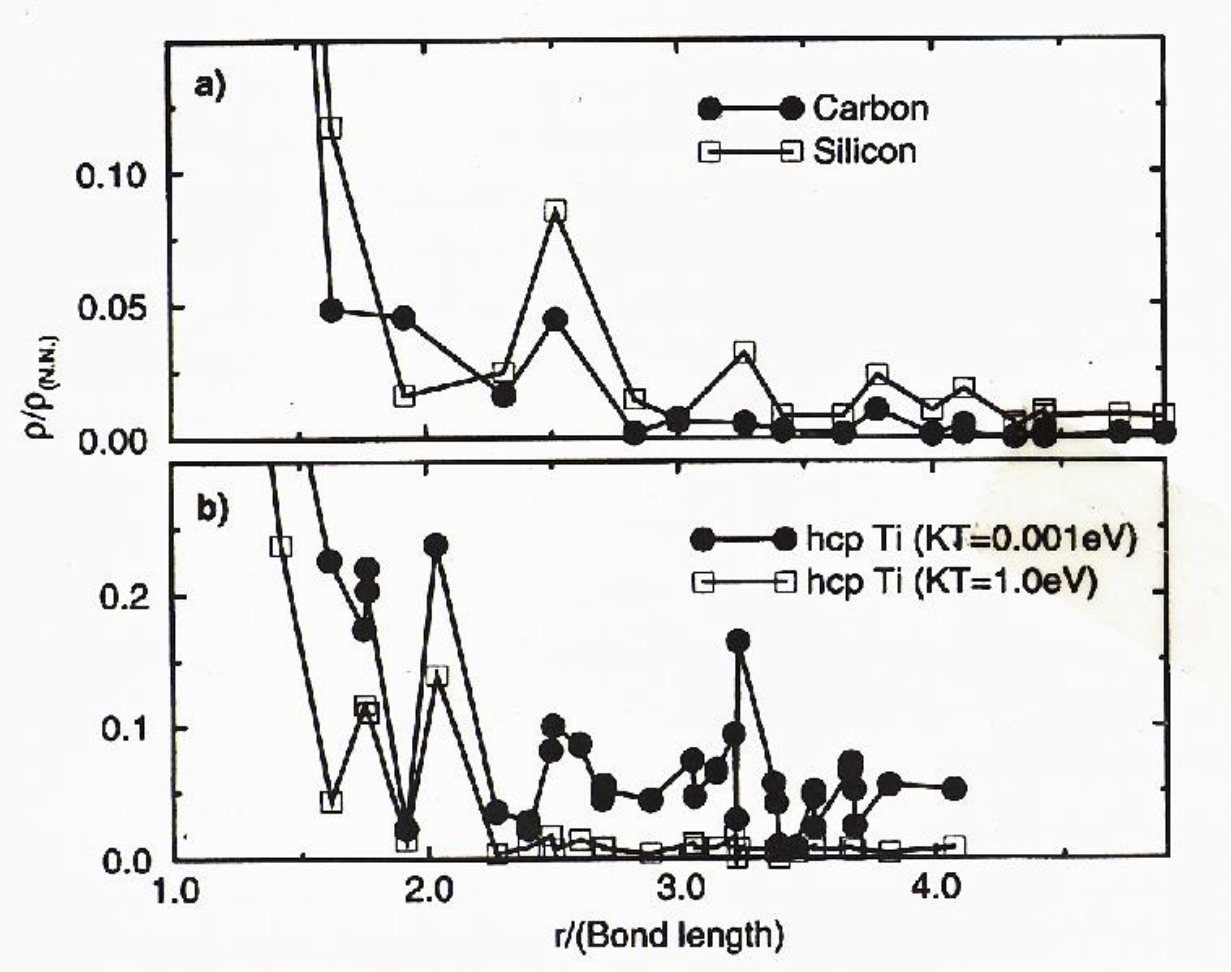


An orbital in Aluminum



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)

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)

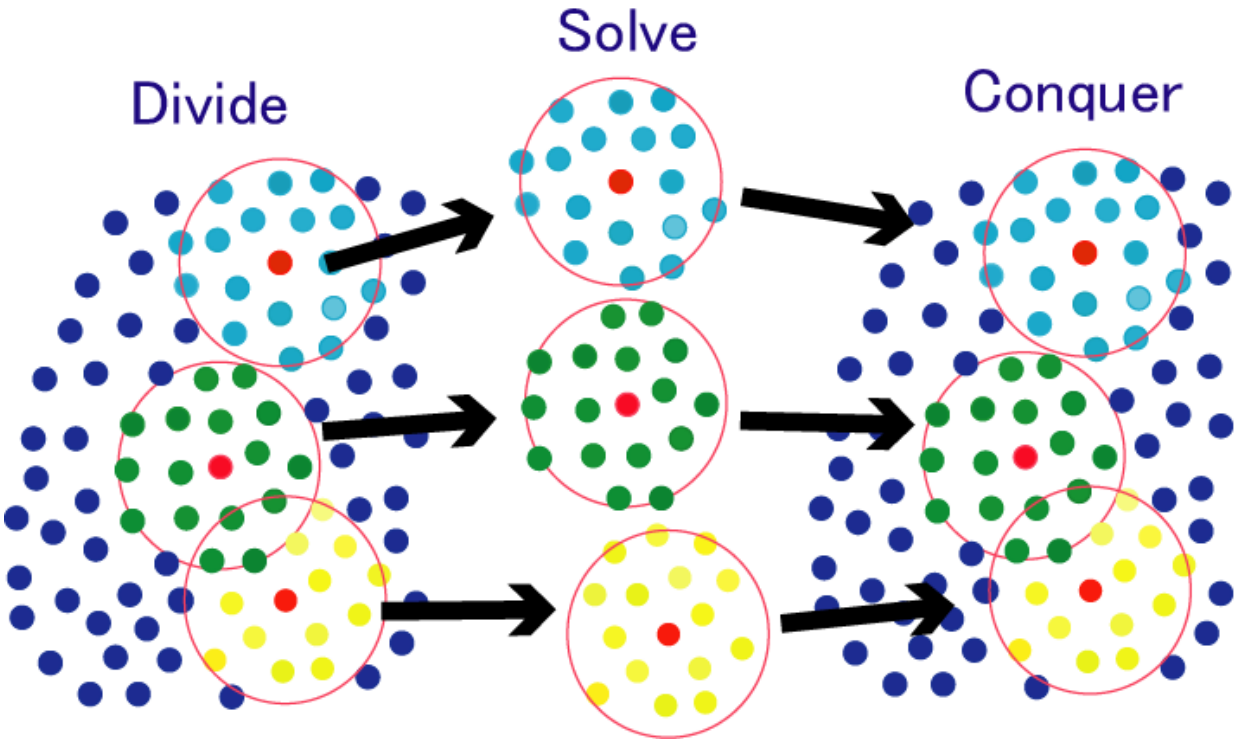
Siesta: (OM) Ordejon et al.(Spain)

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

FEMTECK: (OM) Tsuchida (AIST)

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

Basic idea behind the O(N) method



Assumption

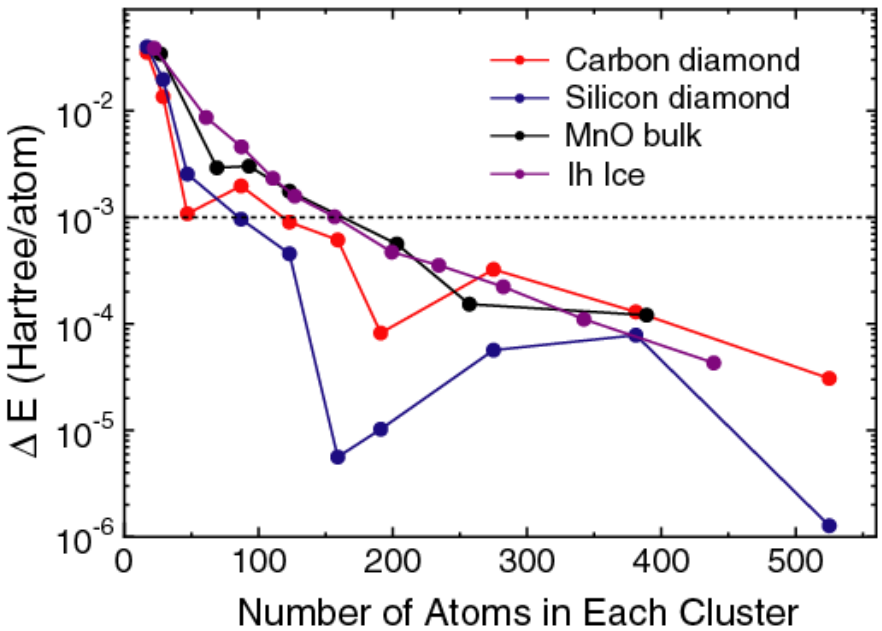
Local electronic structure of each atom is mainly determined by neighboring atomic arrangement producing chemical environment.

Convergence by the DC method

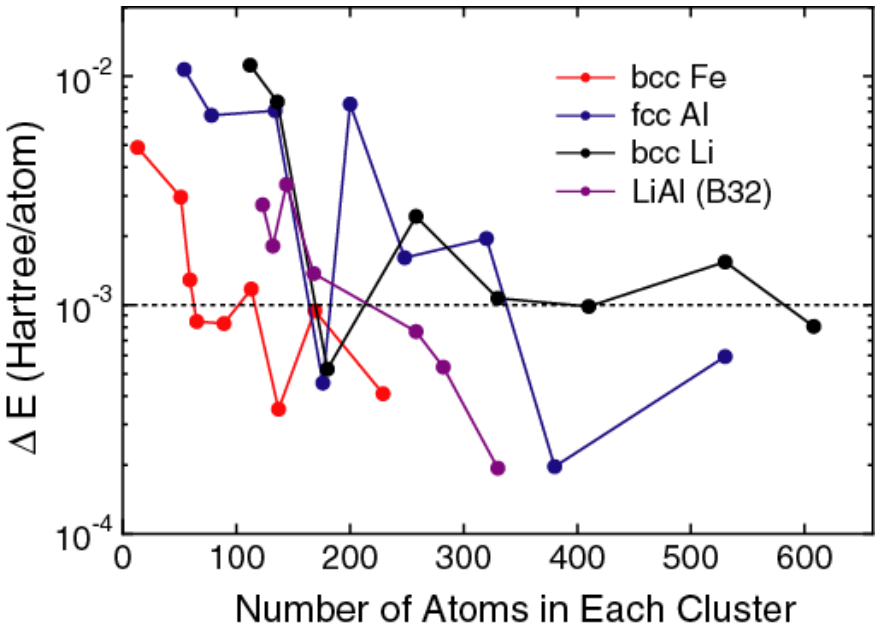
Just solve the truncated clusters → Divide-Conquer method

W.Yang, PRL 66, 1438 (1991)

Insulators, semi-conductors



Metals

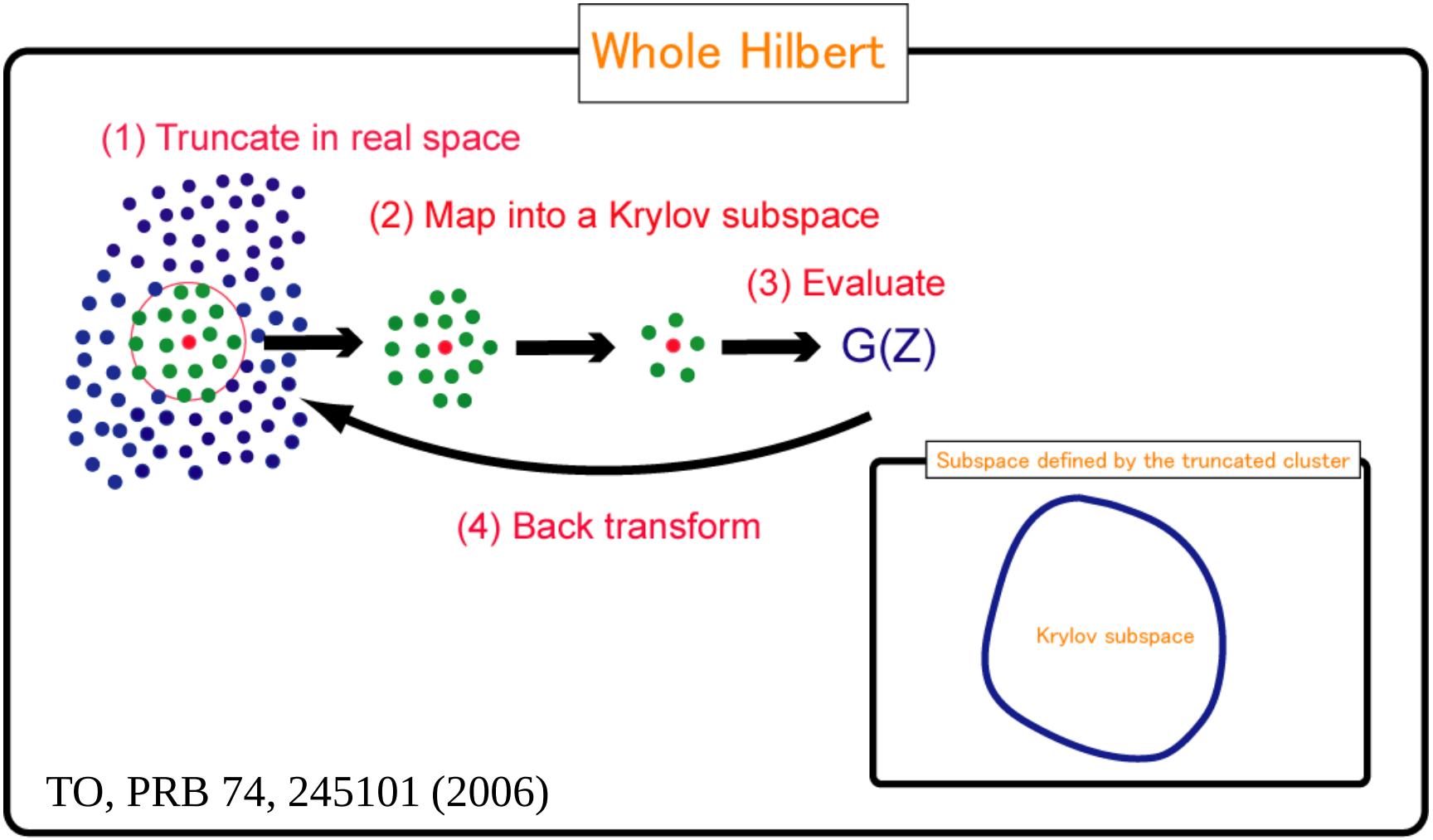


For metals, a large cluster size is required for the convergence.

→ 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

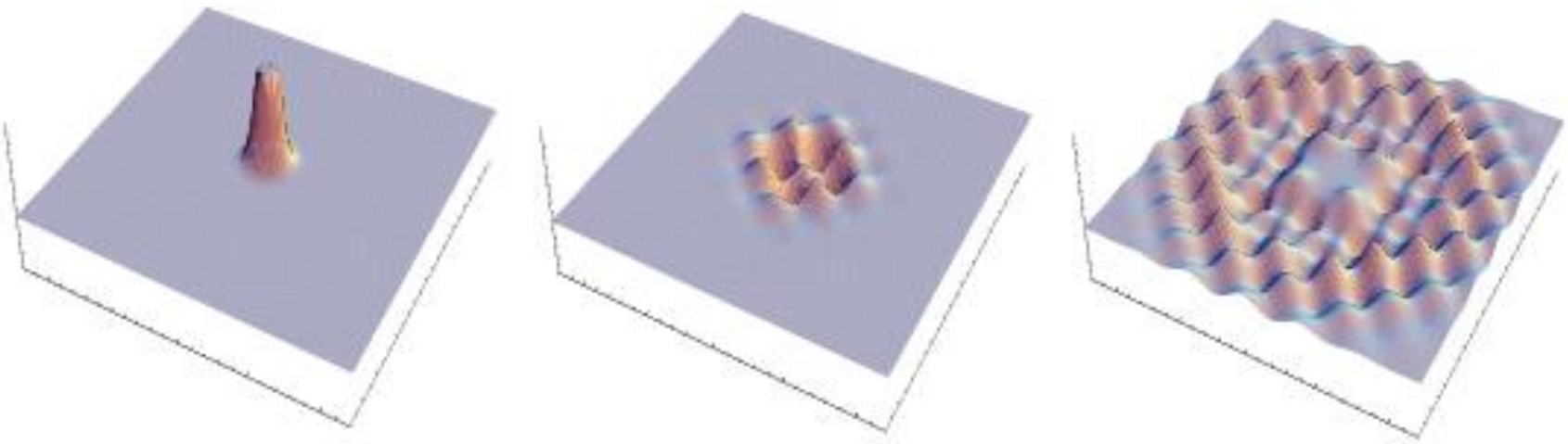


TO, PRB 74, 245101 (2006)

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:

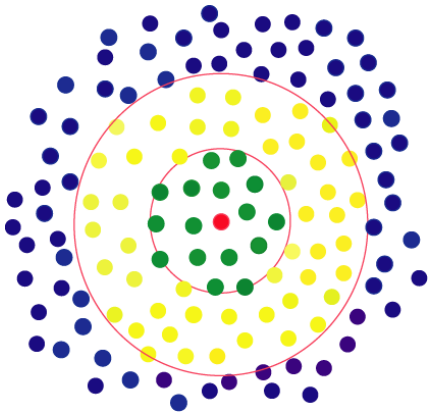
$$H c_\mu = \varepsilon_\mu S c_\mu \longrightarrow H^{\mathrm{K}} b_\mu = \varepsilon b_\mu$$

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

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

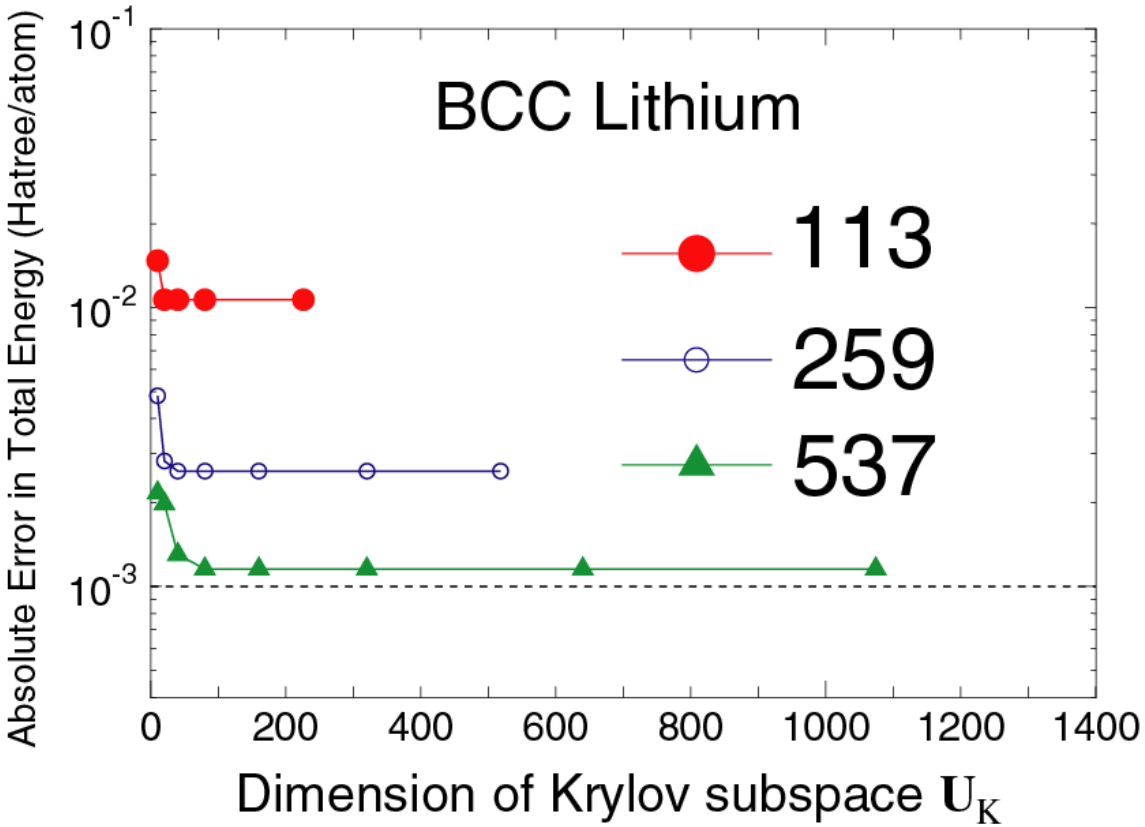
The moment representation of G(Z) gives us the relation.

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

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

Convergence property

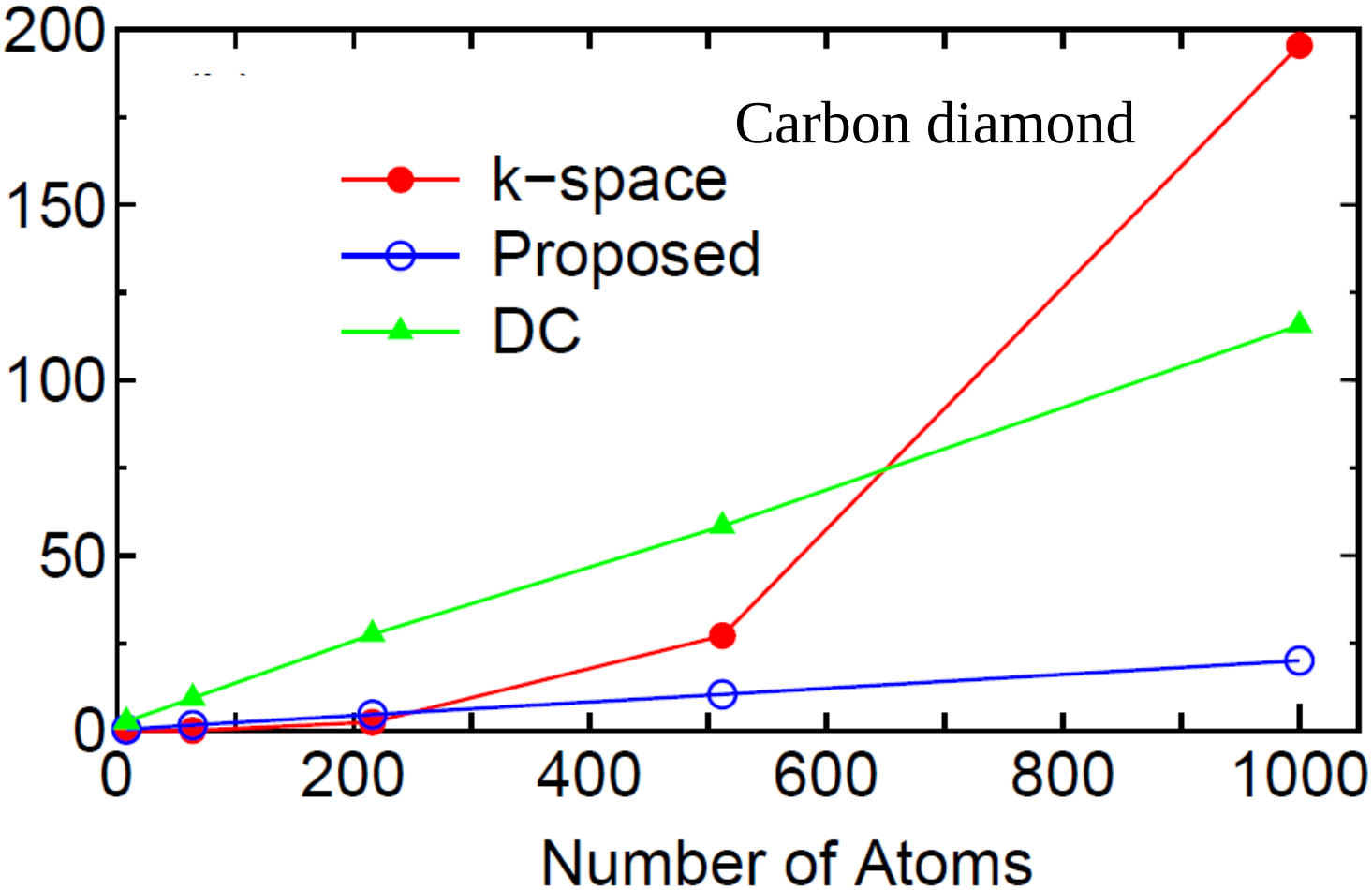
The accuracy and efficiency can be controlled by the size of truncated cluster and dimension of Krylov subspace.



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.



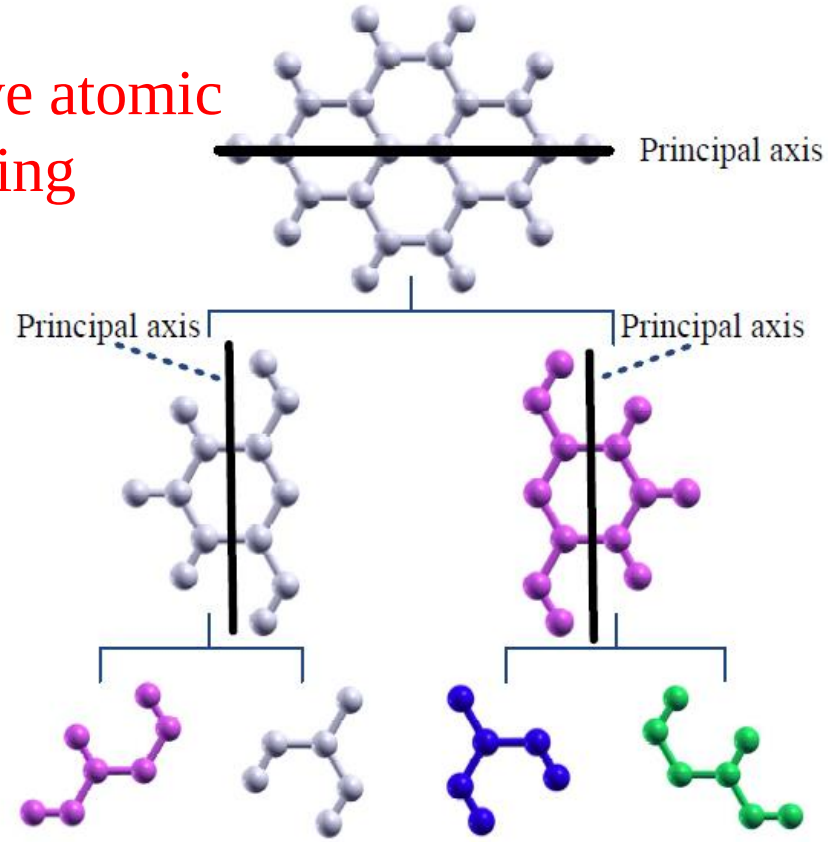
Parallelization

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

Requirement:

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

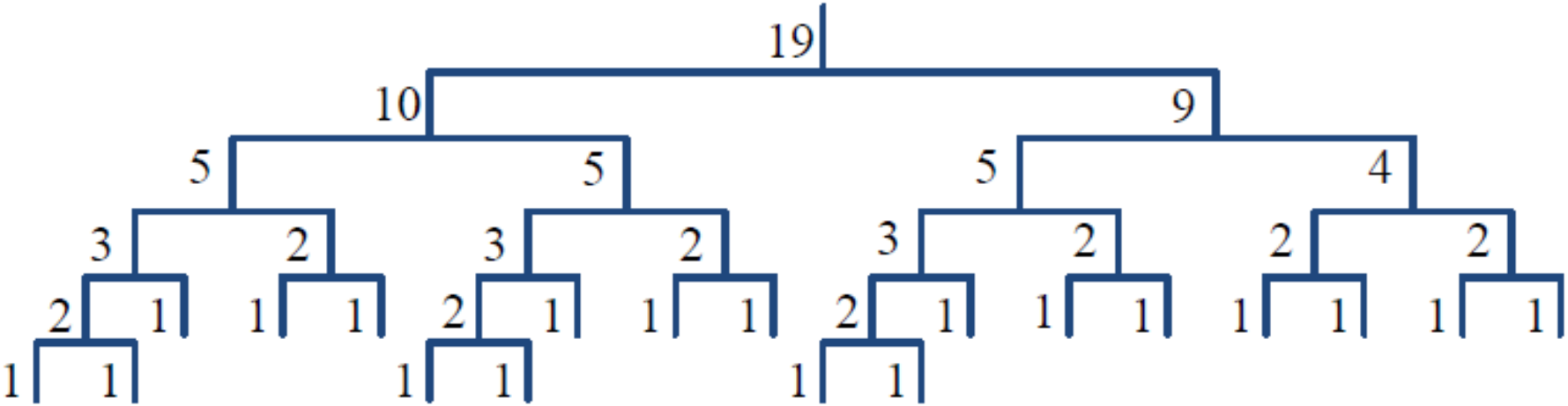
Recursive atomic partitioning



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

Modified recursive bisection

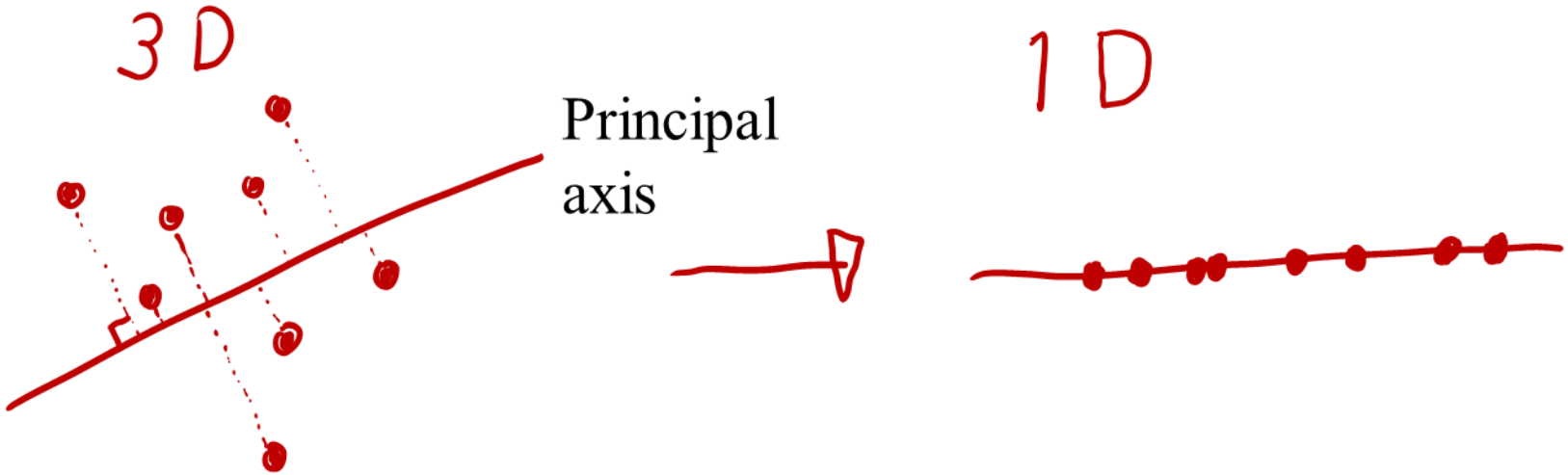
If the number of MPI processes is 19, then the following binary tree structure is constructed.



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

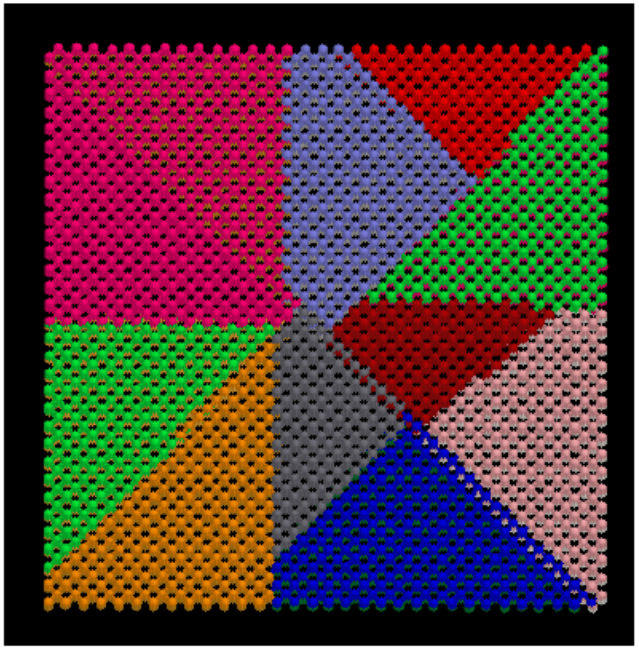
Atoms in an interested region are reordered by projecting them onto a principal axis calculated by an inertia tensor.



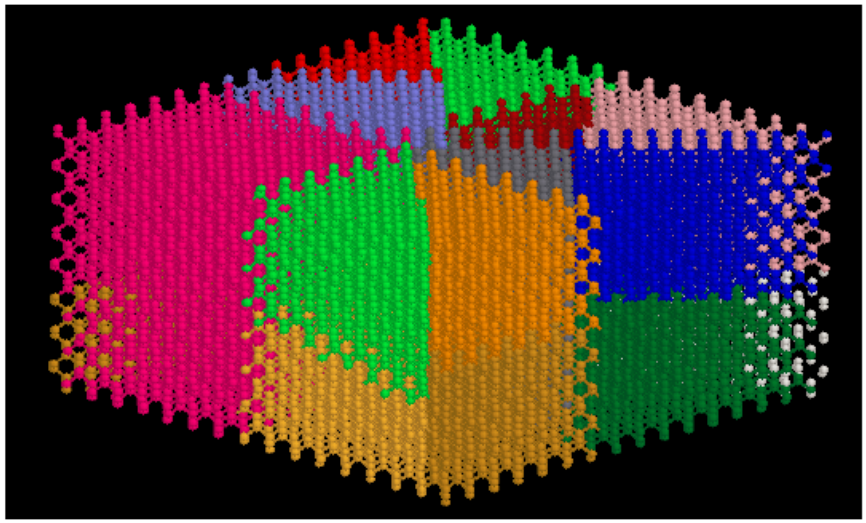
The principal axis is calculated by solving an eigenvalue problem with an inertia tensor:

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

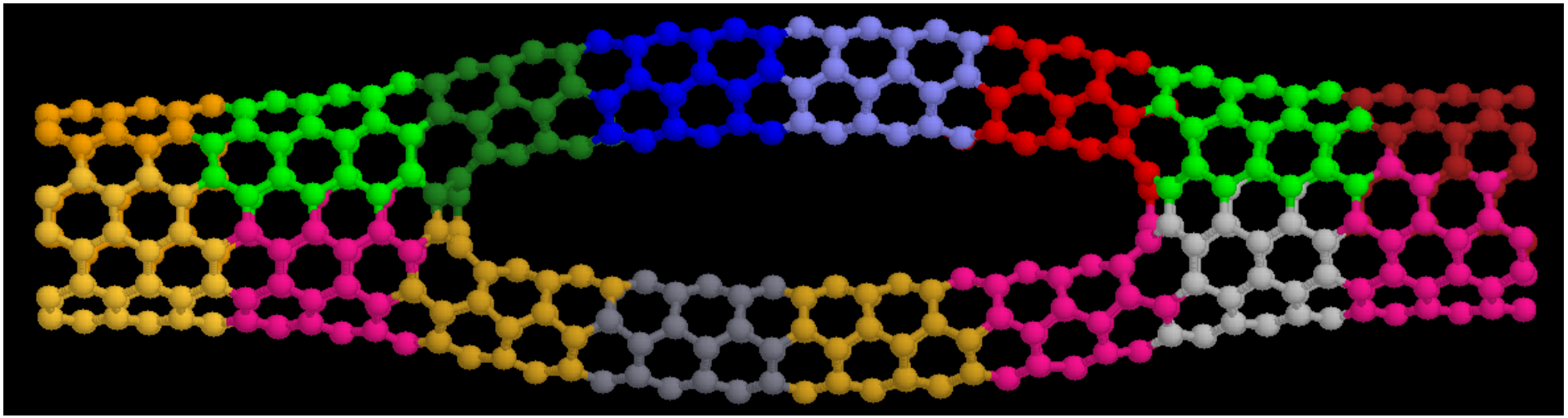
Allocation of atoms to processes



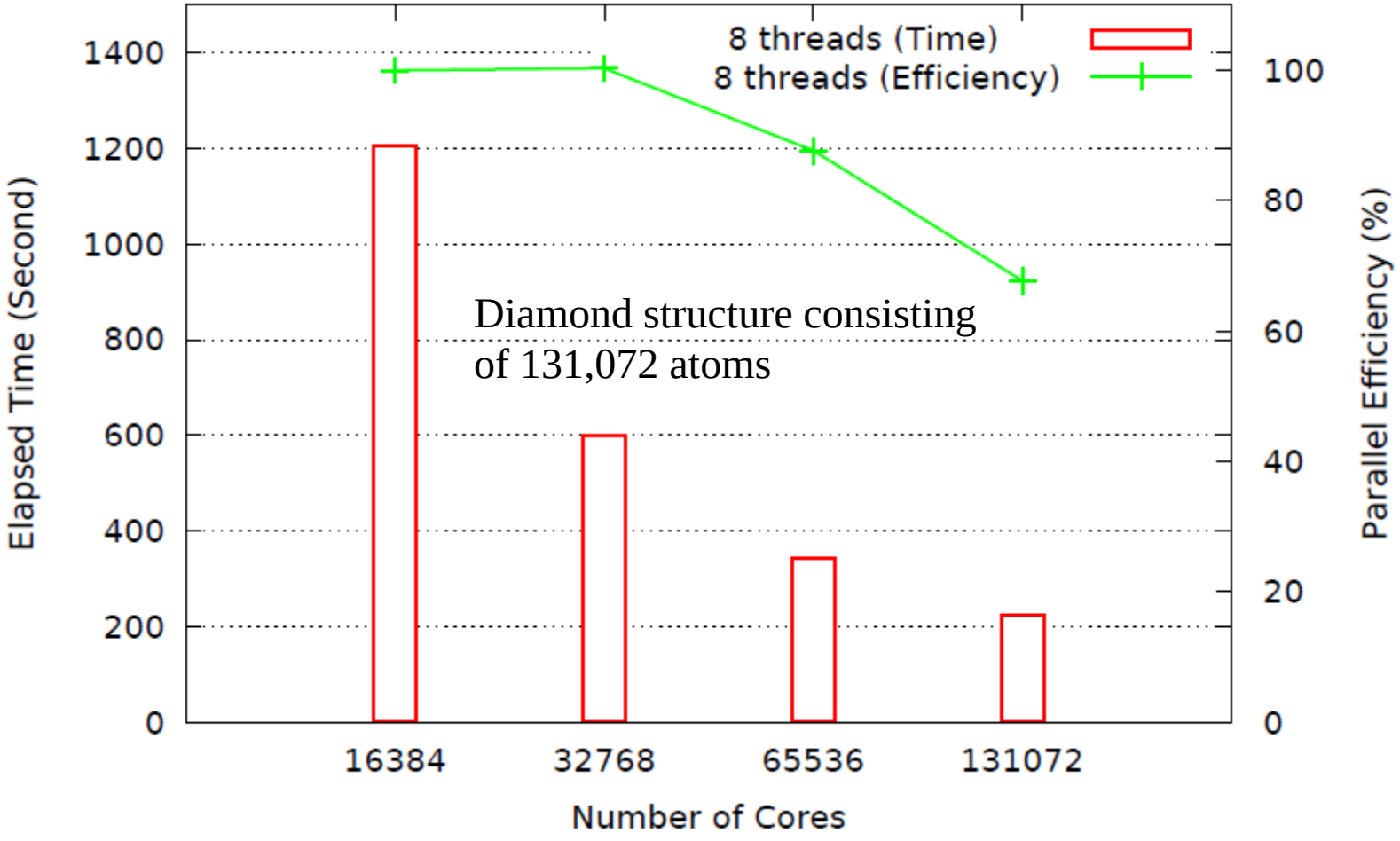
Diamond 16384 atoms, 19 processes



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

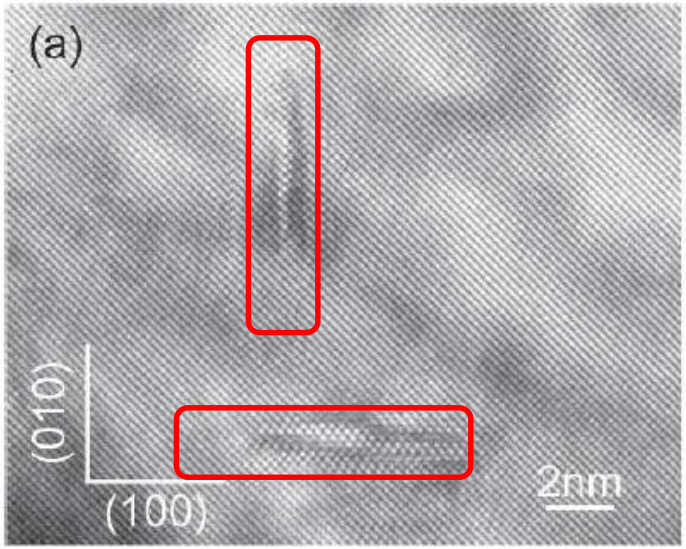
   M. Ohfuchi et al., Appl. Phys. Express 7, 025101 (2014).
   H Jippo, T Ozaki, S Okada, M Ohfuchi, J. Appl. Phys. 120, 154301 (2016).

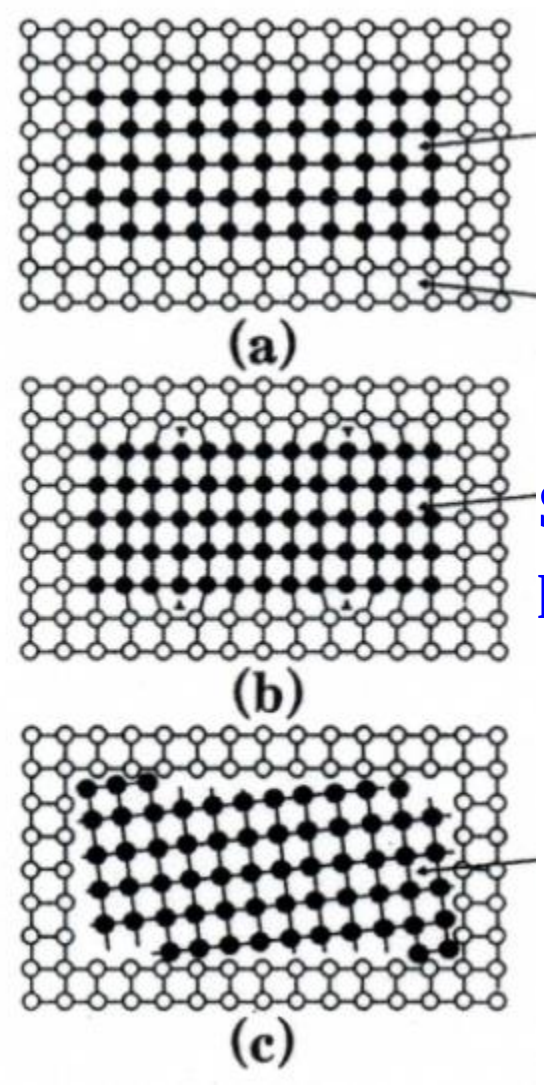
Precipitation in bcc-Fe

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





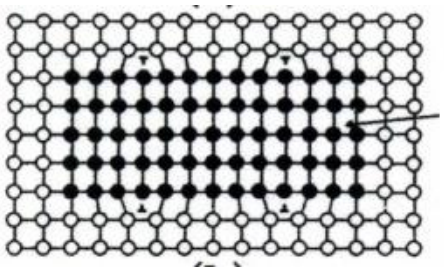
Coherent precipitation

Semicoherent precipitation

Incoherent precipitation

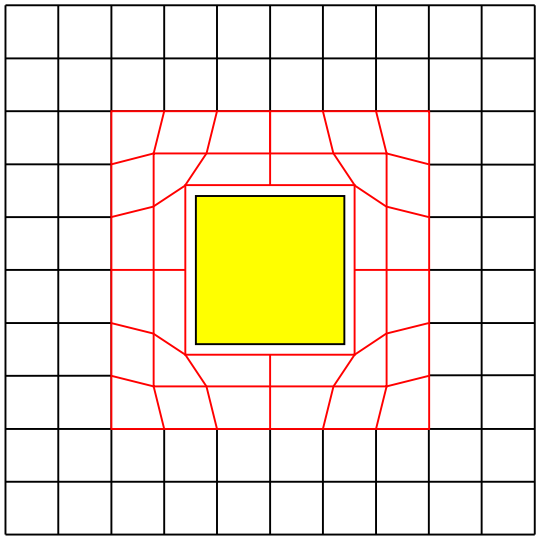
Precipitating materials:
TiC, VC, NbC

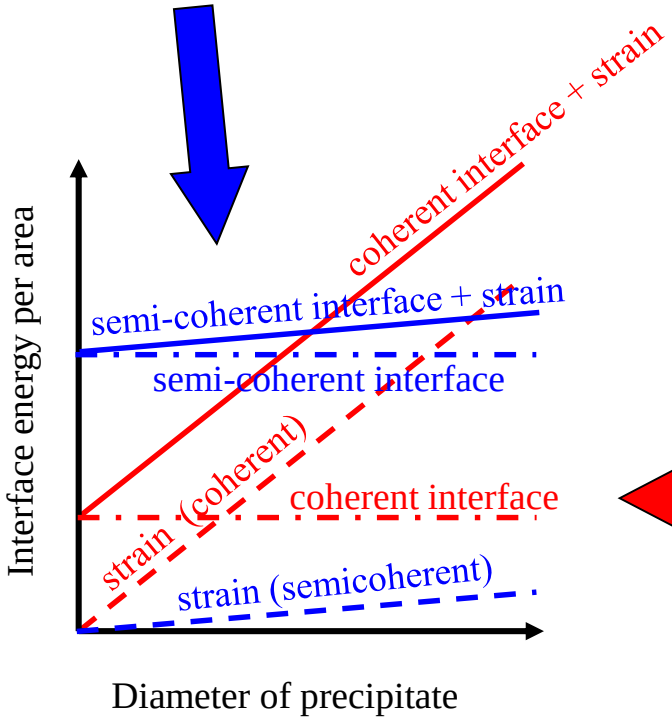
Interface and strain energies

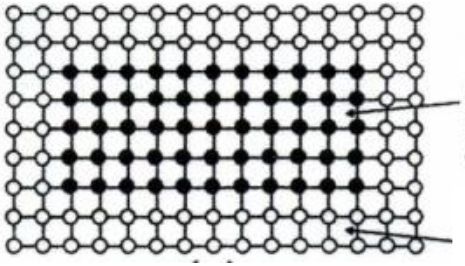


Semi-coherent case

Strain field





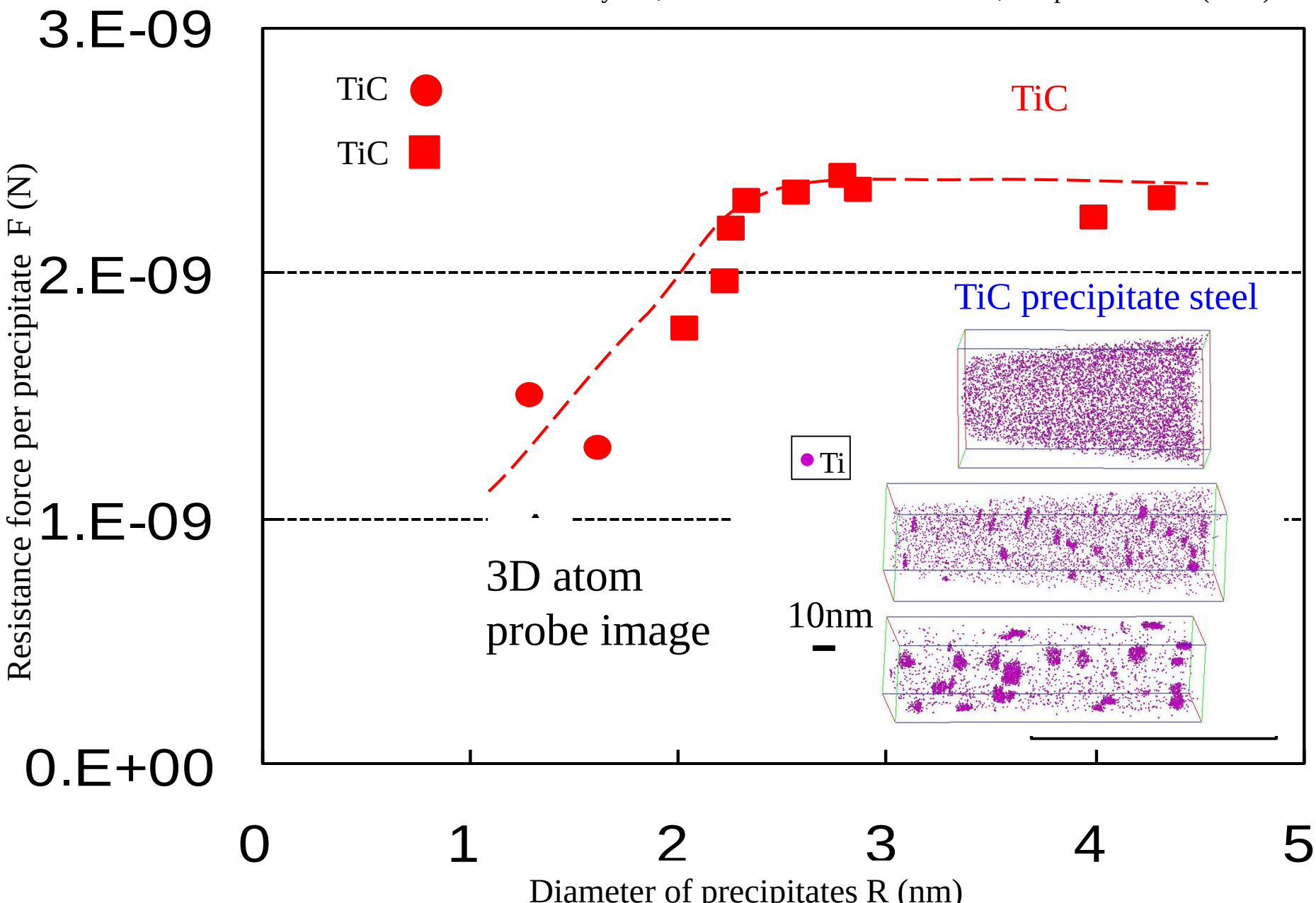


coherent case

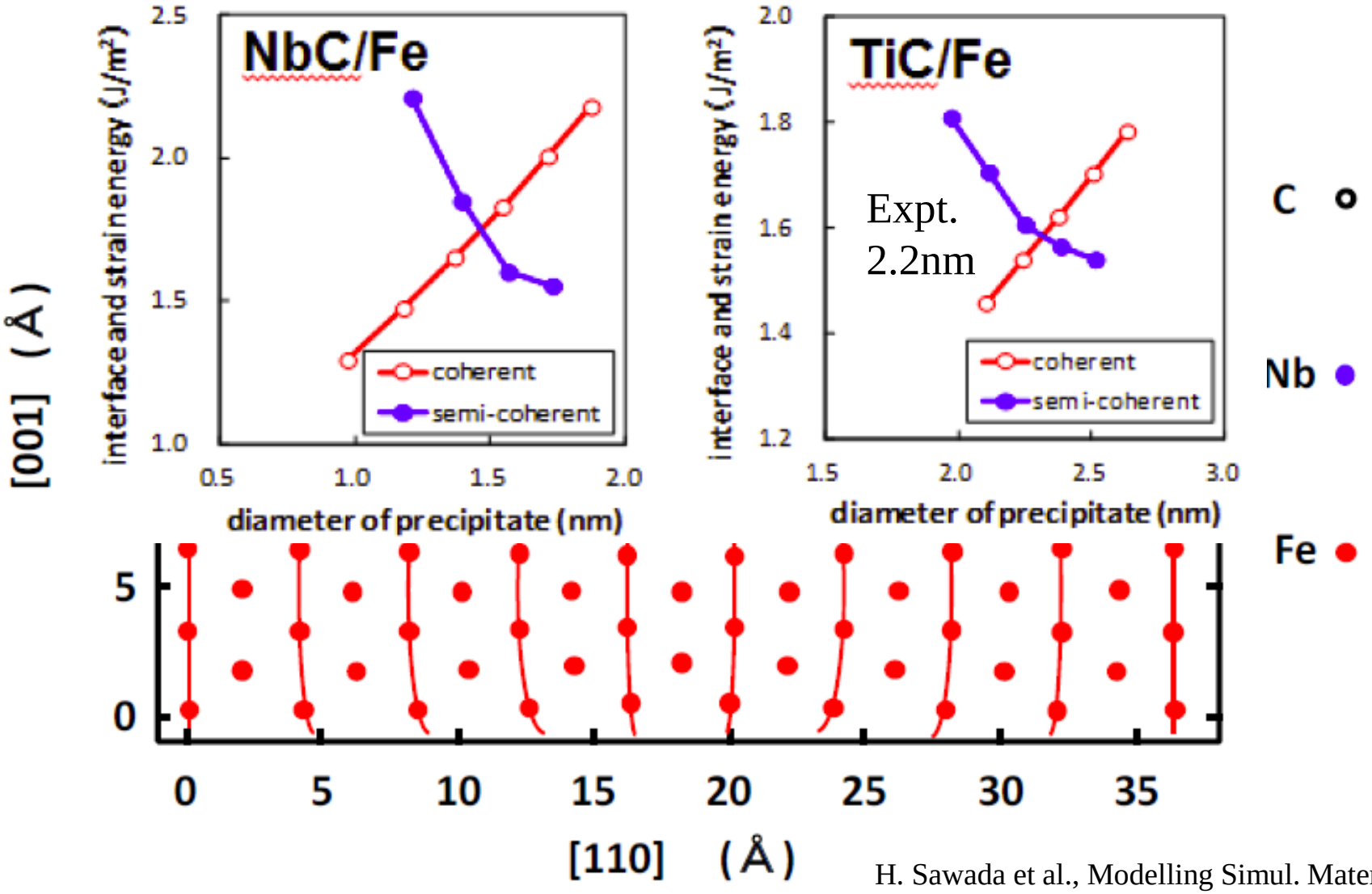
Iron

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).

Free-Energy Analysis on Desolvation of Li^+

Objective

Design of interfaces for fast-charge Li-ion batteries

Free-energy barrier of desolvation of Li^+ in anode-electrolyte interface under charging condition is one of key factors.

→ Theoretical calculation of free-energy barrier of the desolvation is a powerful tool for interface design.

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

Simulations of desolvation of Li^+

Method

Simulation

O(N)-MD calc. with bias imposed by ESM method（constant-N）:

→ Observation of desolvation process

Analysis

Blue-moon ensemble method:

→ O(N)-MD calc. with constraint on z-coordinate of Li^+

→ Mean force along z-axis

→ Free energy profile (barrier) for solvated Li^+ approaching the surface

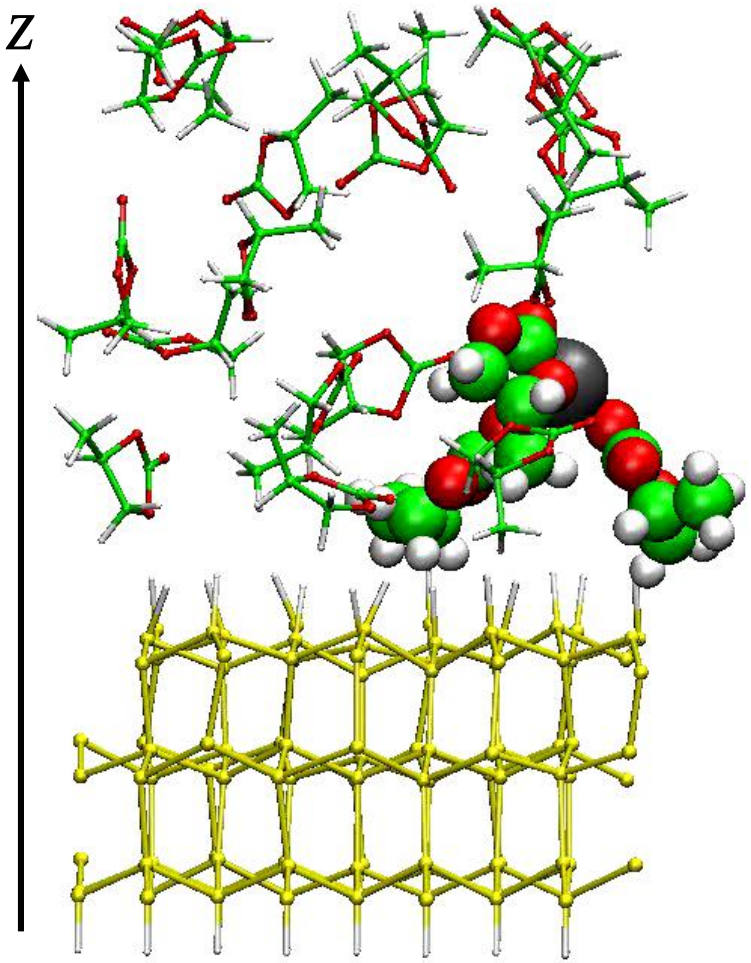
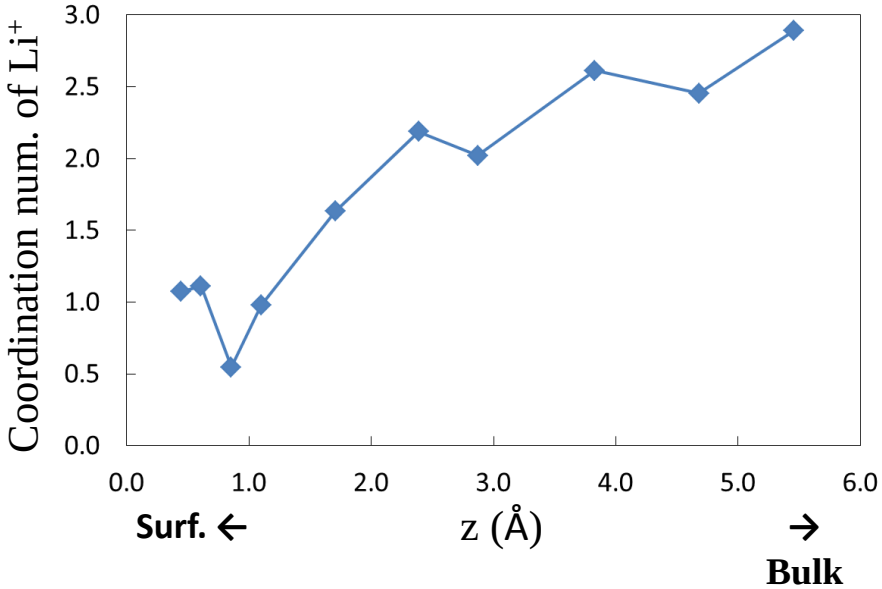
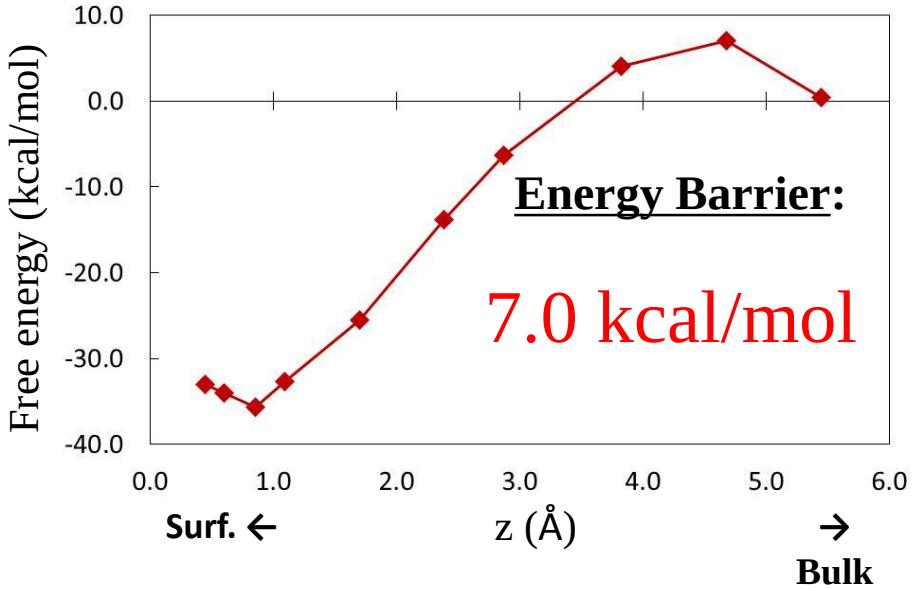
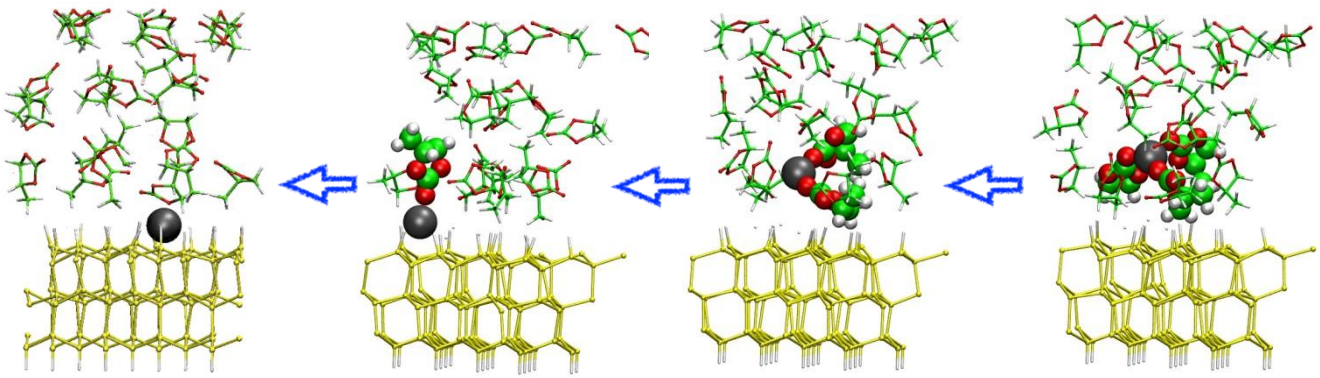


FIG. Calculation model of H-Si(111) anode-PC solvent with Li^+ interface (389 atoms).

Free-Energy Analysis on Desolvation of Li+







adsorbed **1-fold** **2-fold** **3-fold**

Desolvation of Li+:
gradually decreasing coordination num. as approaching the surface

Energy Barrier:
in change of coordination num. from 3 to 2

Height of Energy Barrier:
7.0 kcal/mol

- Desolvation process of Li^+ and its free-energy barrier obtained from O(*N*)-FPMD calculations with bias-control technique.

Summary

- To investigate realistic materials such as Li ion battery, magnets, and structural materials, development of low-order scaling DFT methods is crucial even in the Exa FLOPS era.

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

- We have developed a linear scaling method, low-order scaling method, and O(N) nearly exact exchange functional based on the quantum nearsightedness.

- We expect that such low-order scaling methods will be widely used to address realistic problems on massively parallel computers.