

# Linear scaling Krylov subspace method implemented in OpenMX

1. Introduction
2. Theory of the Krylov subspace method
3. Benchmark calculations
4. How to start the  $O(N)$  calculations
5. Summary

# Towards large-scale simulations

$10^3 - 10^4$  atom

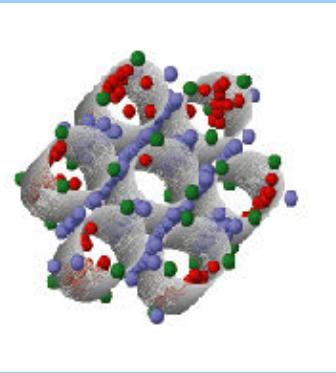
DFT calculations of thousands atoms  
is still a grand challenge.

$$O(N^3) \rightarrow O(N)$$

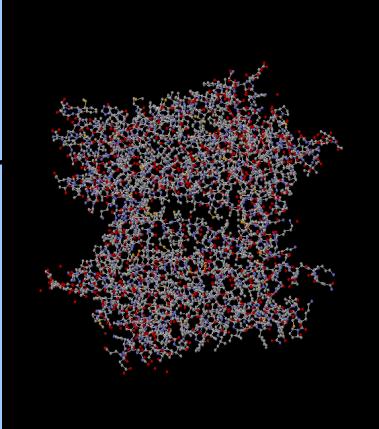
$10^2$  atom

Time scale

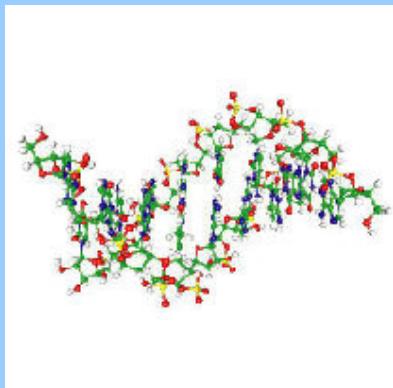
Carbon nanotube



Functional protein



DNA



Lots of successes.  
Even material design  
is attempted with success.

System size

# Possible? $10^5$ atoms by $O(N^3)$ DFT

$O(N^3)$

0.1 days/ $100$  atoms/1 MD/10 GFLOPS



**1000** days/ **$10^5$**  atoms/1 MD/PFLOPS

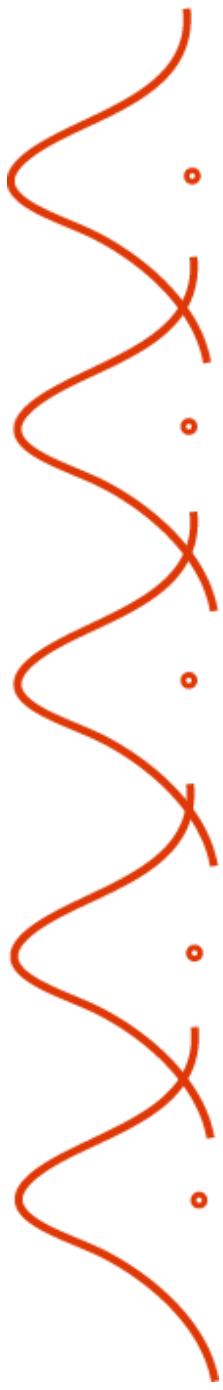
$O(N)$

**0.1** days/ **$10^5$**  atoms/1 MD/PFLOPS

Assuming that  $O(N)$  is faster than  $O(N^3)$  above 1000 atoms.

# How can $O(N)$ be achieved in OpenMX

## 1. Use of local pseudoatomic basis functions



- Non-zero matrix elements, memory size  $\rightarrow O(N)$
- Compatibility between the local basis and  $O(N)$  methods
- Good parallel efficiency

## 2. Solving the eigenvalue problem in an $O(N)$ operation

Three  $O(N)$  methods are available:  
divide-conquer (DC), recursion, **Krylov subspace** methods.

# Methods behind O(N) methods in OpenMX

- Norm-conserving pseudopotentials
- Pseudoatomic basis functions
- LDA, GGA, LDA+U
- Fully self-consistent calculations

# Two routes to linear-scaling calculations

Conventional expression



$\psi$ : one-particle wave function

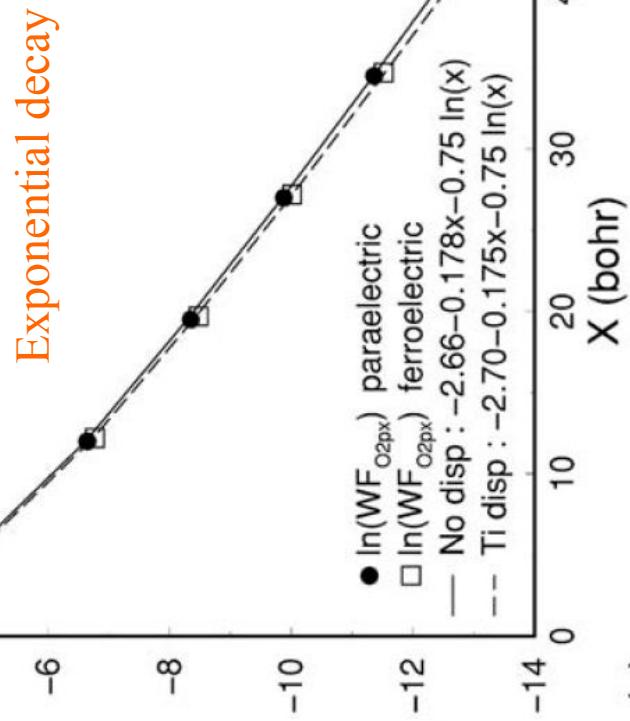
$\rho$ : charge density

$\phi$ : Wannier function

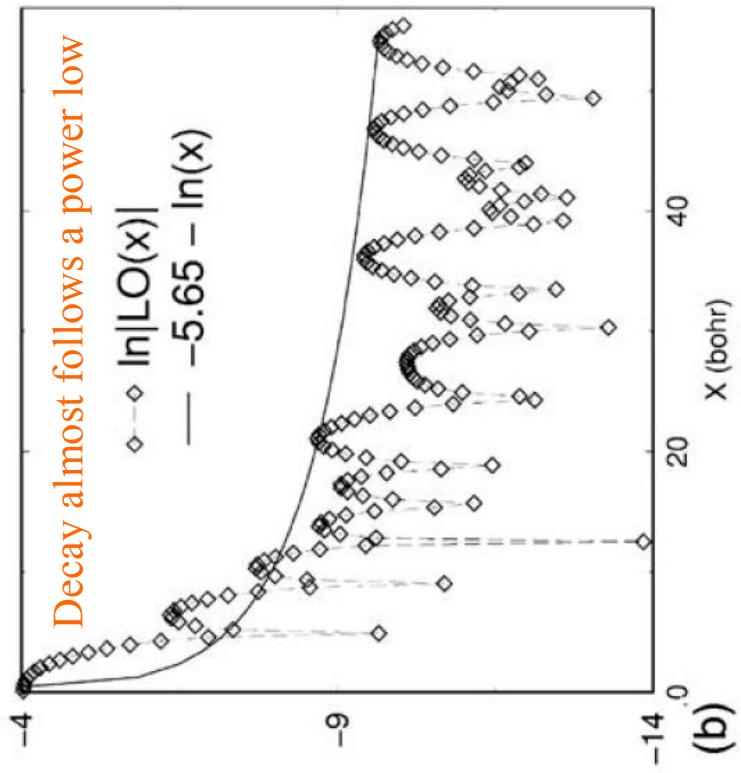
$n$ : density matrix

# Locality of Wannier functions

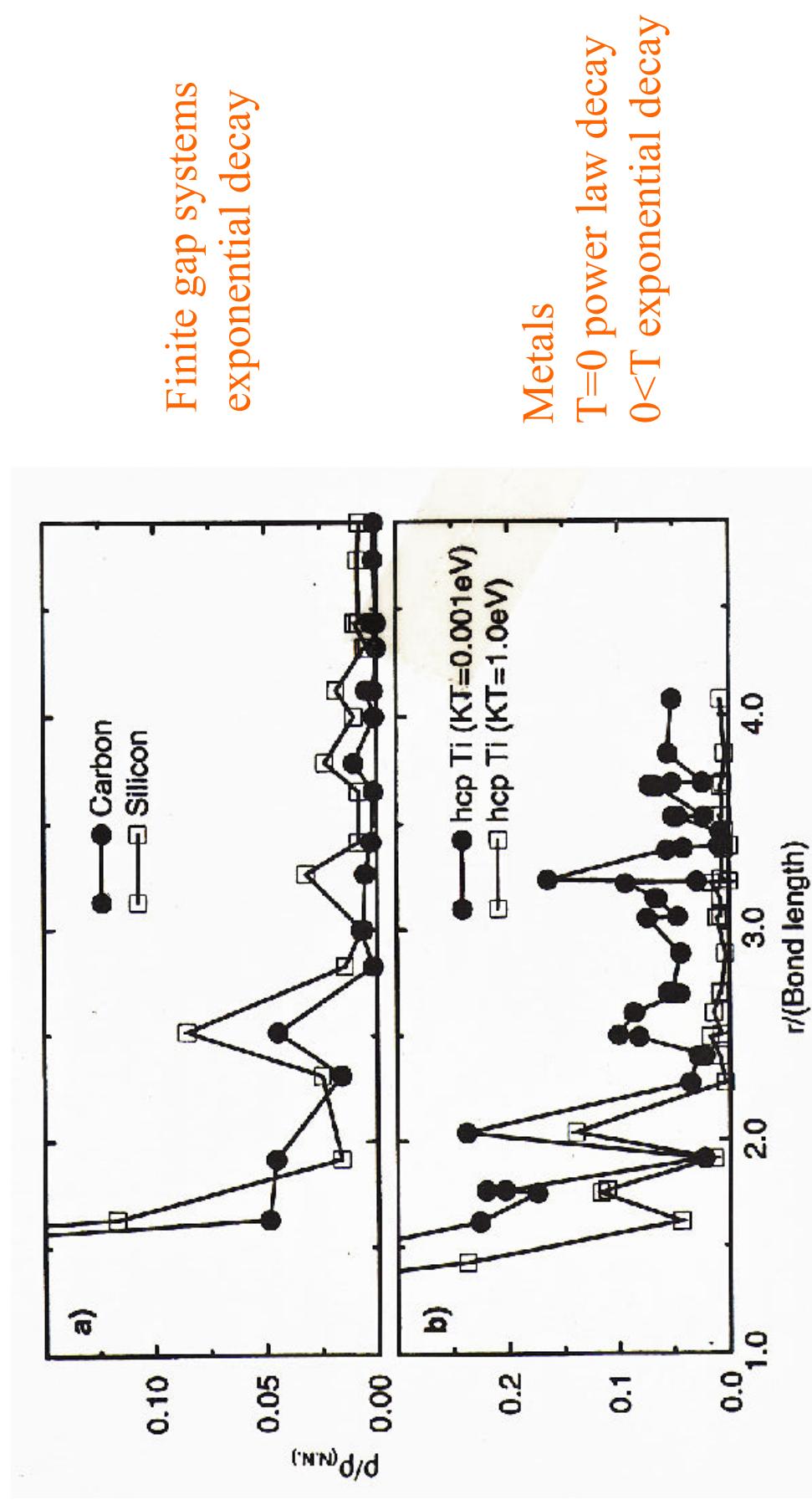
O-2px in PbTiO<sub>3</sub>



An orbital in Alumium



# Locality of density matrix in real space



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

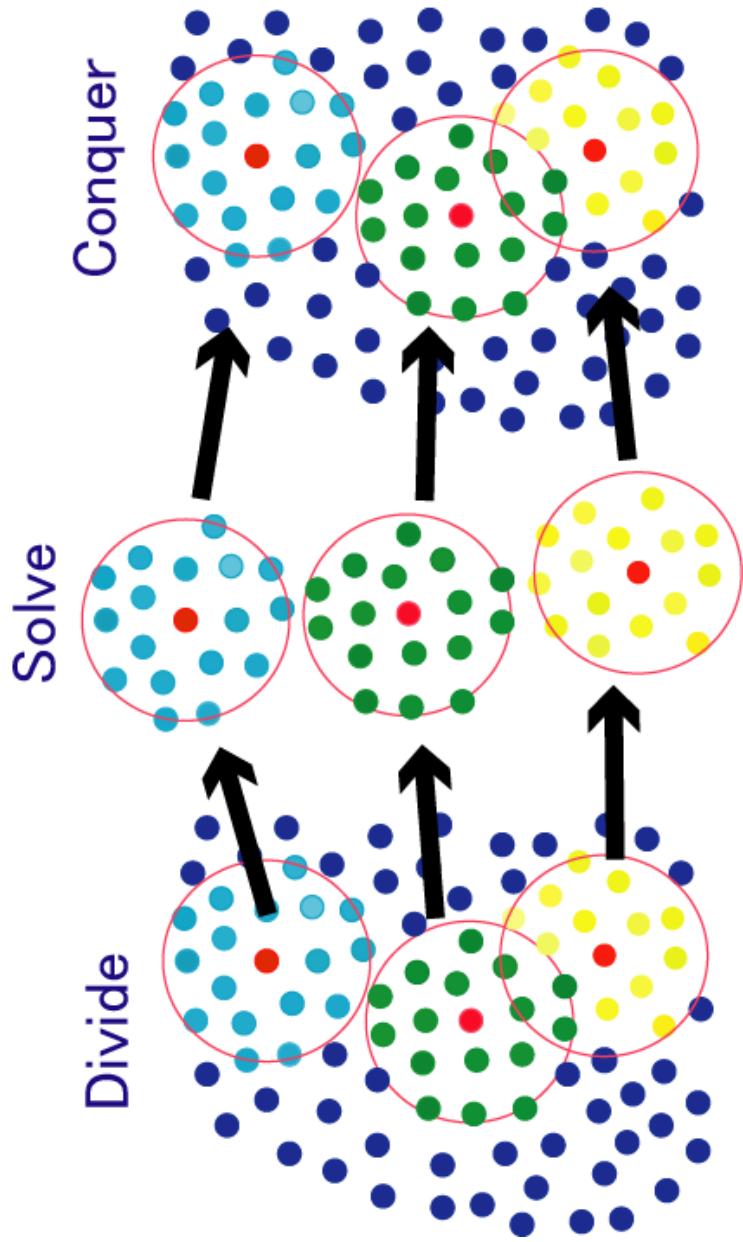
# Various linear scaling methods

$$\left. \begin{array}{c} \text{Wannier functions (WF)} \\ \text{Density matrix (DM)} \end{array} \right\} \times \left. \begin{array}{c} \text{Variational (V)} \\ \text{Perturbative (P)} \end{array} \right\}$$

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

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

# Basic idea behind our $O(N)$ method

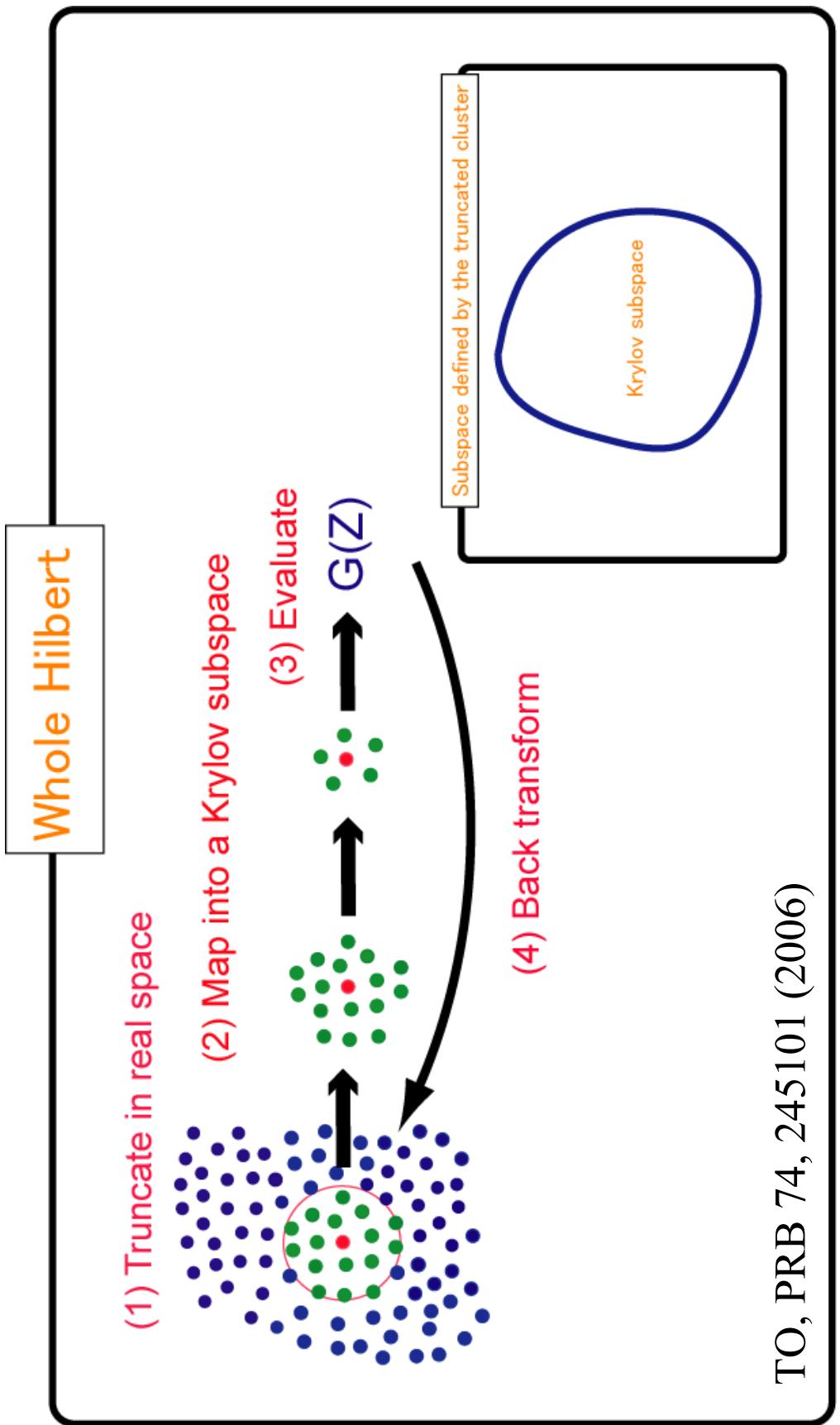


## Assumption

The total Green's function can be approximated by the sum of local Green's functions associated with the central atom in each truncated clusters.

# $O(N)$ Krylov subspace method

Two steps mapping of the whole Hilbert space into subspaces



# Relation between the Krylov subspace and Green's function

A Krylov subspace is defined by

$$U_K = \{|W_0\rangle, (S^{-1}H)|W_0\rangle, (S^{-1}H)^2|W_0\rangle, \dots, (S^{-1}H)^q|W_0\rangle\}$$

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

$$\begin{aligned} H_{mn}^K &= (W_0|(A^\dagger)^m H A^n|W_0) \\ &= (W_0|S(S^{-1}H)^{m+n+1}|W_0), \\ &= (W_0|S\mu^{(m+n+1)}S|W_0) \end{aligned} \quad \text{Definition of moments}$$
$$\begin{aligned} \mu^{(p)} &= c\varepsilon^p c^\dagger, \\ &= cc^\dagger Hcc^\dagger Hc \cdots c^\dagger Hcc^\dagger, \\ &= (S^{-1}H)^p S^{-1} \end{aligned}$$

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.**

# Generation of Krylov subspaces

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

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

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

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

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:

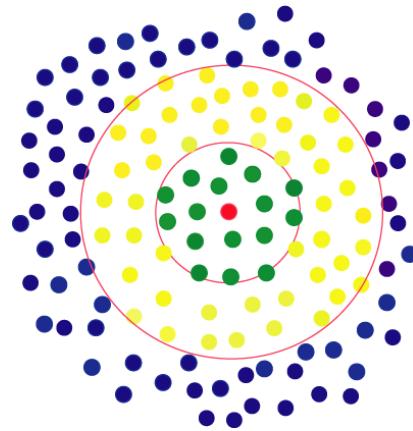
$$H_{C_\mu} = \varepsilon_\mu S_{C_\mu} \longrightarrow H^K b_\mu = \varepsilon b_\mu$$

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

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

**Green:** core region

**Yellow:** buffer region



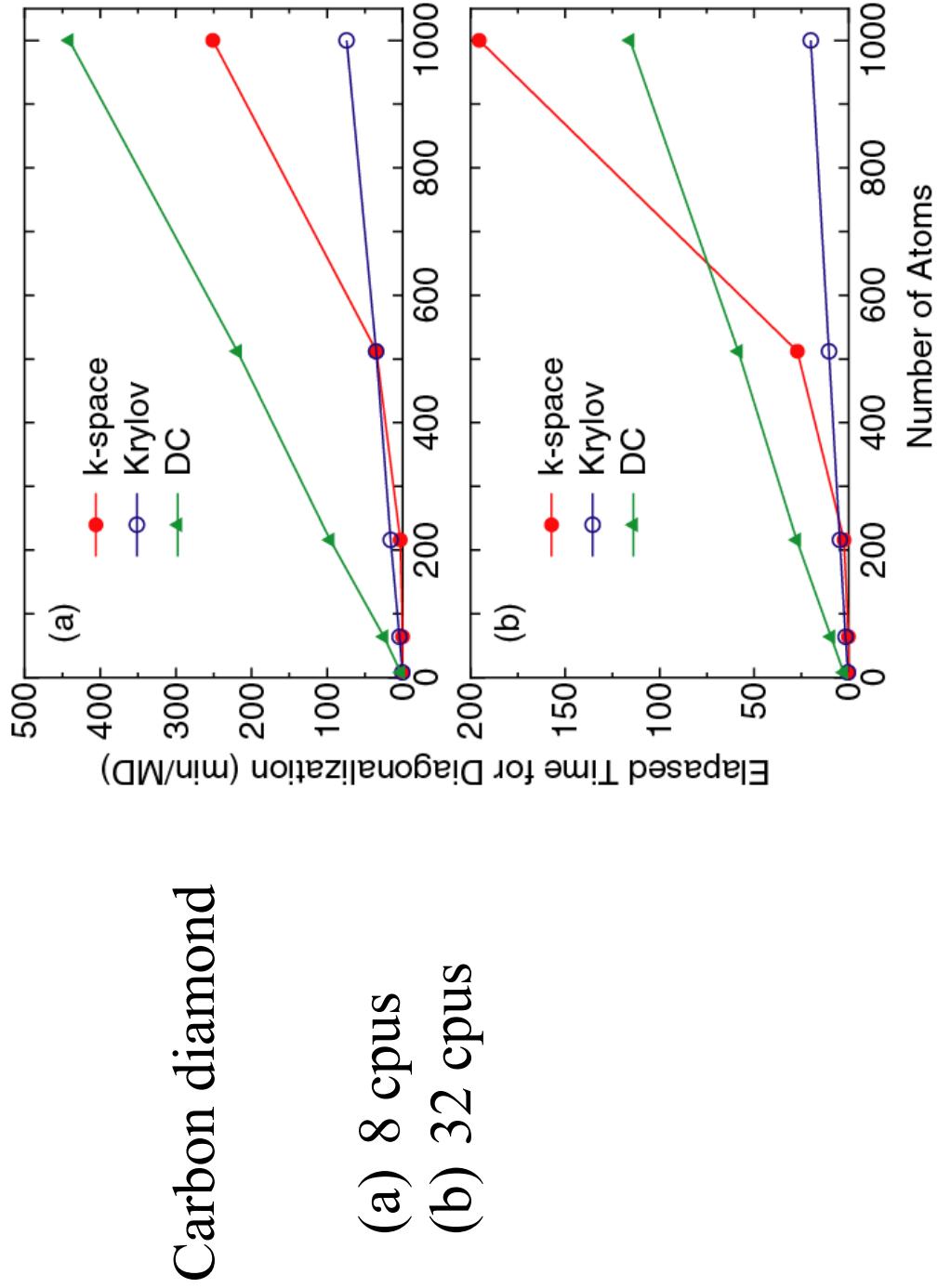
- The embedded cluster is under the Coulomb interaction from the other parts.
- The charge flow from one embedded cluster to the others is allowed.

# Algorithm

1. Construct the truncated clusters
2. Generate the Krylov subspaces
3. Solve the embedded cluster problems
4. Find a common chemical potential
5. Generate the whole charge density
6. Solve Poisson's eq. for the whole system
7. Go to 3 until convergence

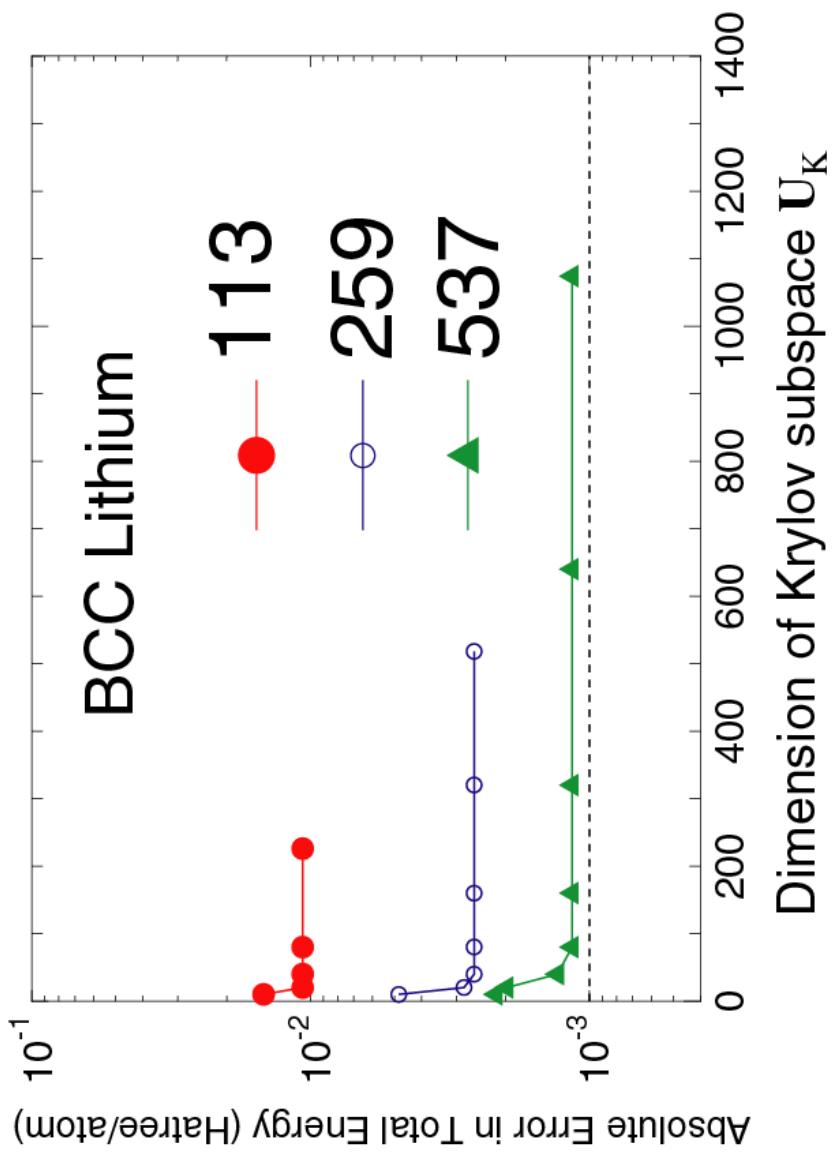
# Comparison of computational time

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



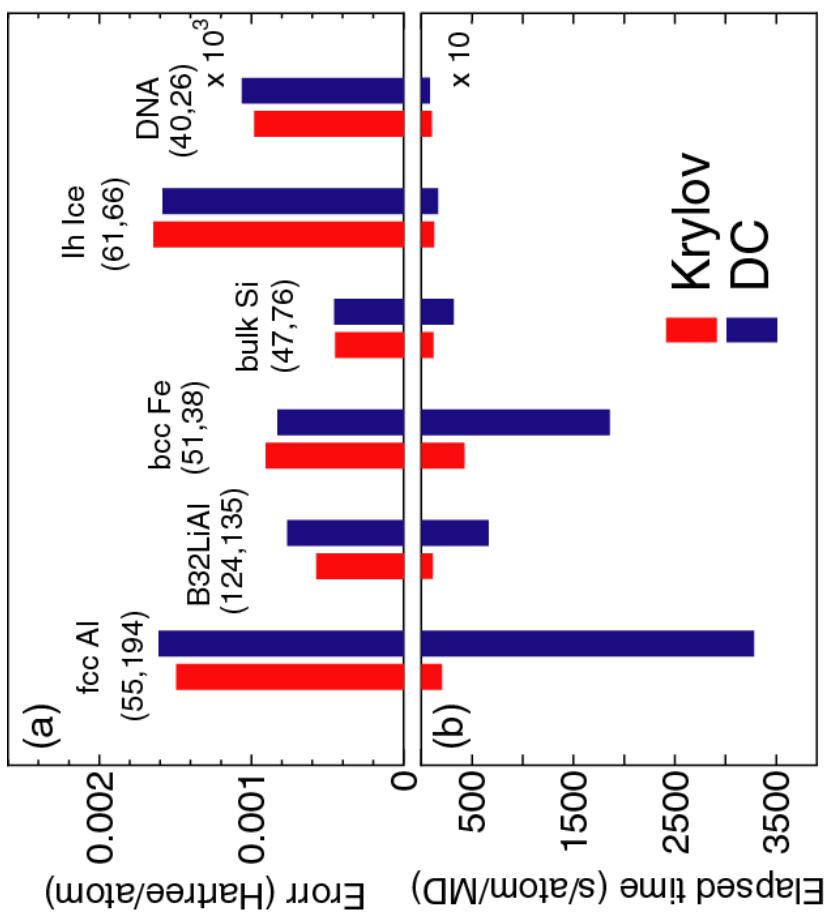
# Convergence property

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



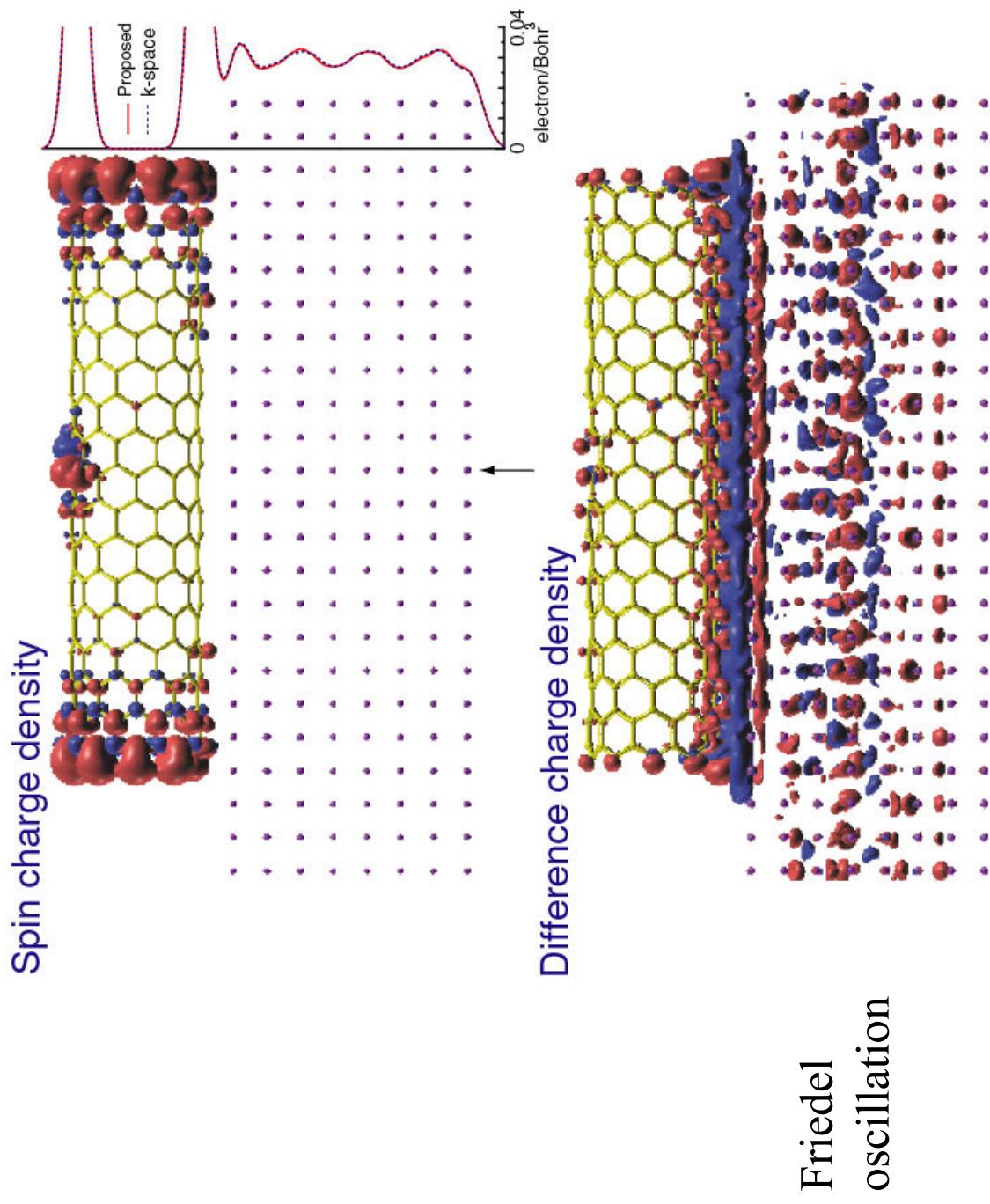
In general, the convergence property is **more complicated**.  
See PRB 74, 245101 (2006).

## Comparison between the DC and Krylov subspace methods



To achieve the almost same accuracy, the Krylov subspace method is much faster than the DC method in metals, while they are comparable for covalent systems.

# Carbon nanotube on Al surface



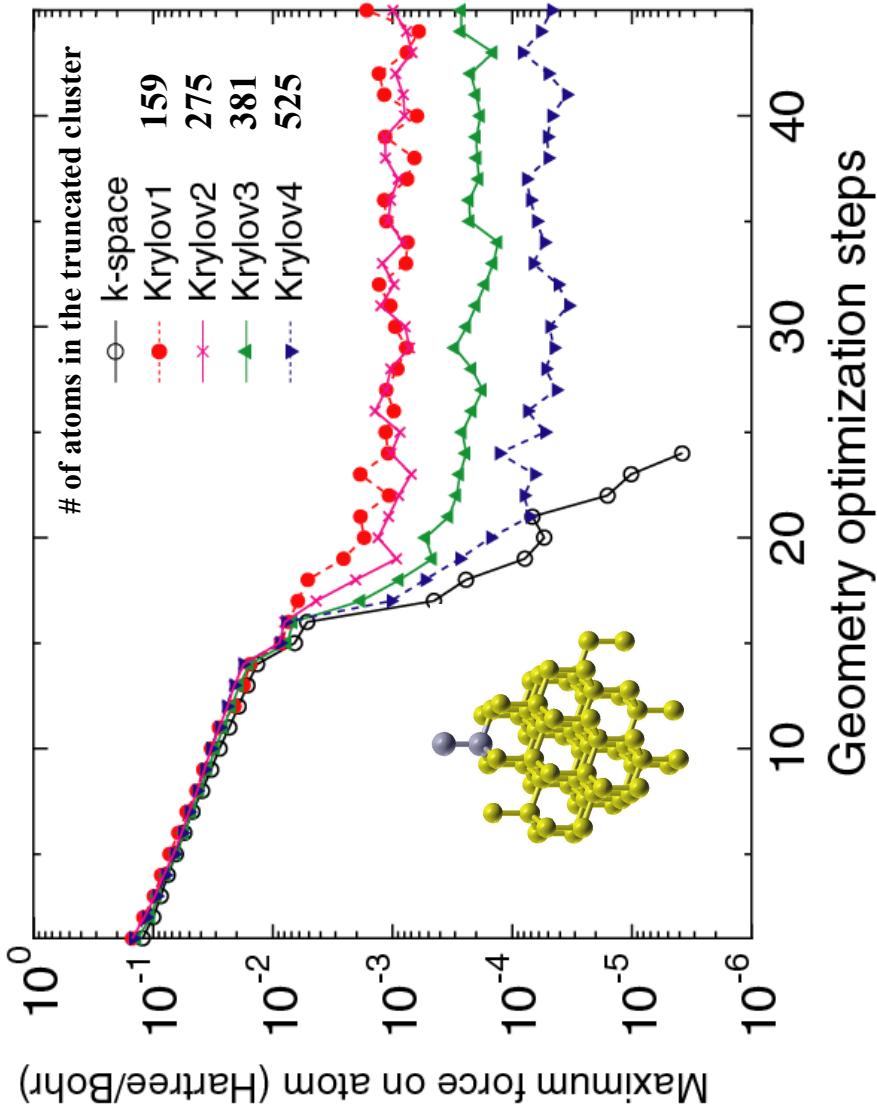
# Comparison of Mulliken populations

Carbon site	Charge		Spin moment	
	$O(N)$	Conventional	$O(N)$	Conventional
C <sub>1A</sub>	4.032	4.029	1.264	1.280
C <sub>2A</sub>	4.028	4.033	1.259	1.224
C <sub>3A</sub>	4.036	4.038	1.178	1.168
C <sub>4A</sub>	4.047	4.051	1.115	1.121
C <sub>5A</sub>	4.147	4.142	0.367	0.512
C <sub>1B</sub>	4.005	4.005	-0.148	-0.149
C <sub>2B</sub>	4.004	4.004	-0.146	-0.144
C <sub>3B</sub>	4.003	4.003	-0.135	-0.133
C <sub>4B</sub>	4.008	4.007	-0.121	-0.122
C <sub>5B</sub>	4.005	4.005	-0.091	-0.105
C <sub>6B</sub>	4.145	4.144	-0.069	-0.081
C <sub>1V</sub>	4.039	4.036	0.656	0.659
C <sub>2V</sub>	4.040	4.036	0.645	0.657
C <sub>3V</sub>	3.997	3.995	-0.364	-0.373

Charge and spin moments  
can be well reproduced.

# Accuracy in geometry optimization No.1

Boron doped diamond of which super cell contains 62 carbon atoms and two boron atoms being nearest neighbors



The accuracy of  $10^{-4}$  to  $10^{-3}$  Hartree/bohr can be achieved as the calculation condition becomes severe.

# Accuracy in geometry optimization No.2

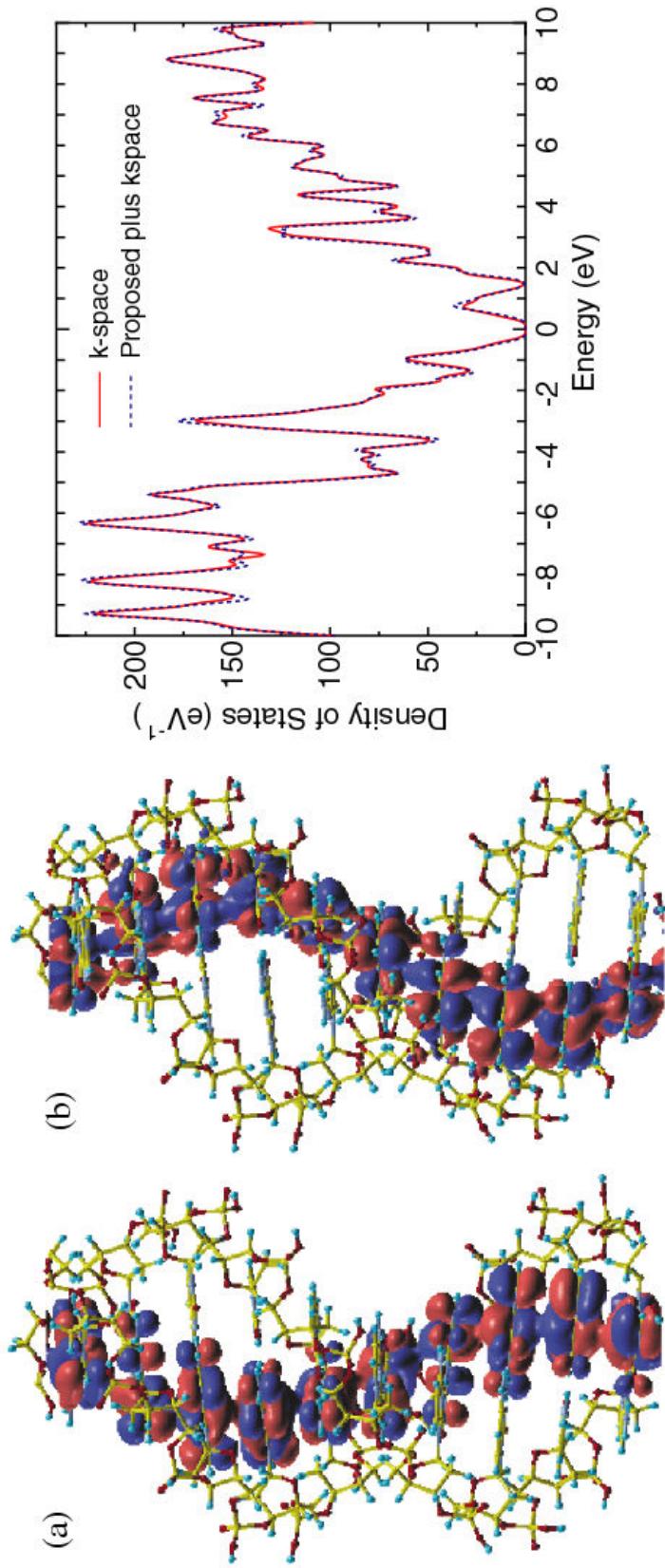
Boron doped diamond of which super cell contains 62 carbon atoms and two boron atoms being nearest neighbors

	Krylov1	Krylov2	Krylov3	Krylov4	k-space
r(B-B)	1.975	1.977	1.983	1.988	1.990
r(B-C <sub>1</sub> )	1.544	1.544	1.543	1.544	1.543
r(C <sub>1</sub> -C <sub>2</sub> )	1.567	1.568	1.568	1.568	1.568
∠(BBC <sub>1</sub> )	100.60	100.57	100.52	100.40	100.37
∠(C <sub>1</sub> BC <sub>1</sub> )	116.64	116.72	116.74	116.81	116.83
Δ E	0.0644	0.0625	0.0576	0.0615	0.0626
Time (s)/MD	<b>139</b>	<b>926</b>	<b>1890</b>	<b>2841</b>	<b>1421</b>

The Krylov3 provides a compromise balancing the accuracy and efficiency.

# A way to calculate wave functions

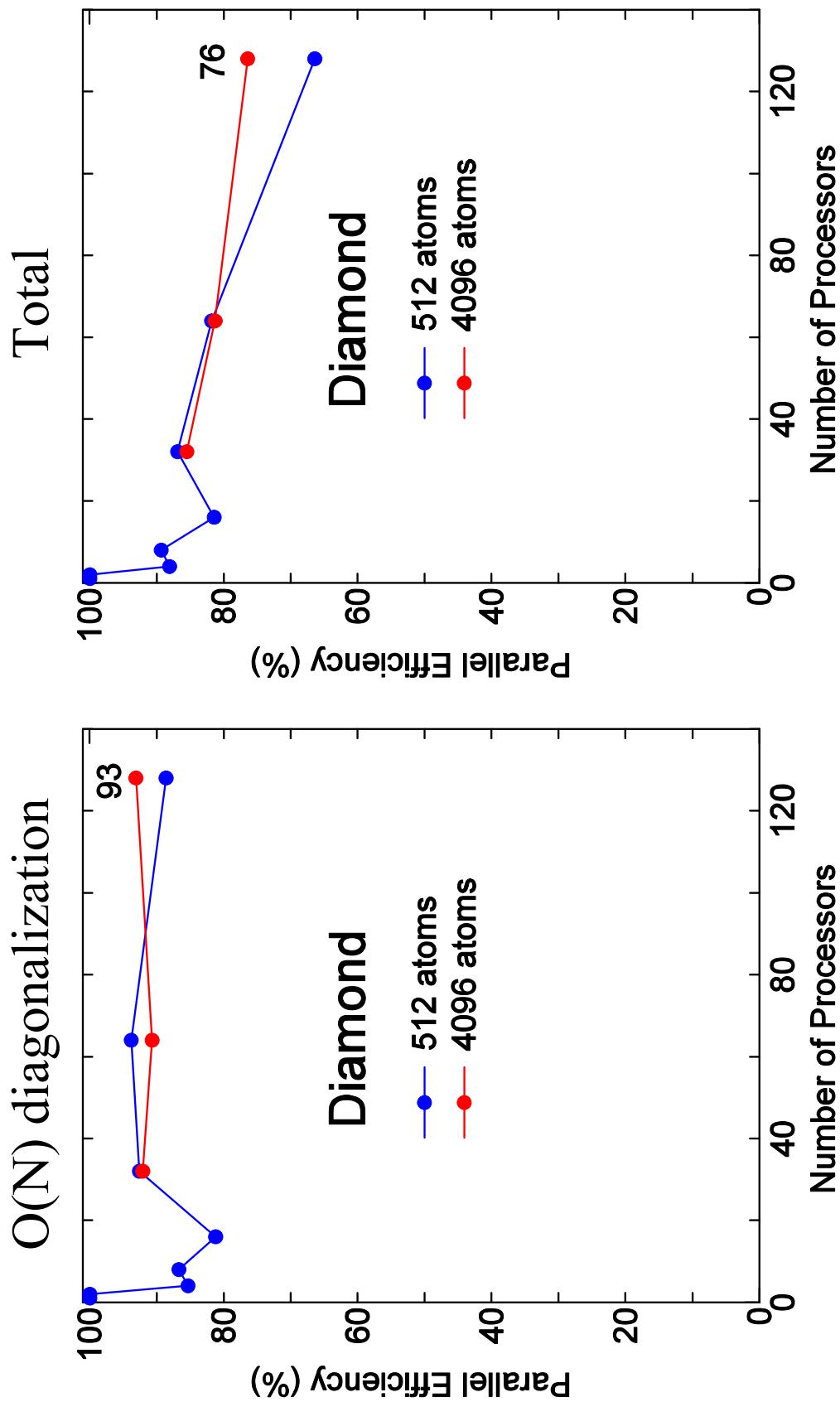
1. Perform the  $O(N)$  calc.
2. Perform an one-shot diagonalization by the conventional scheme with the SCF charge.



HO (guanine)      LU (cytosine)

The comparison of the DOS supports  
the validity of the alternative scheme.

# Parallel efficiency



by Cray-XT3 in JAIST

# How to start the O(N) calc. No.1

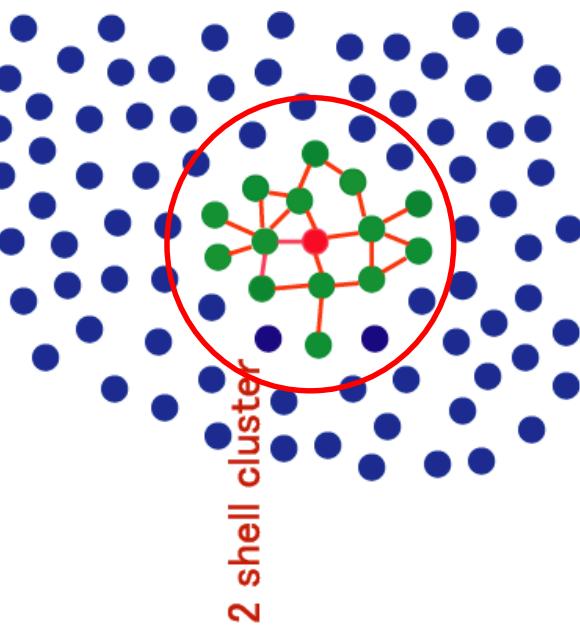
## 1. Choose the Krylov subspace method as

```
scf.EigenvalueSolver      Krylov
```

## 2. Control the size of the truncated cluster by

```
orderN.HoppingRanges    6.8          # default=5.0 (Ang)
orderN.NumHoppings        2            # default=2
```

Logically truncated cluster



The **first** keyword determines the atoms within a sphere with the radius.

The **second** keyword picks up atoms within the atoms determined by the first keyword by considering the connectivity.

The connectivity is defined by the cutoff radii of basis functions on two atoms as.

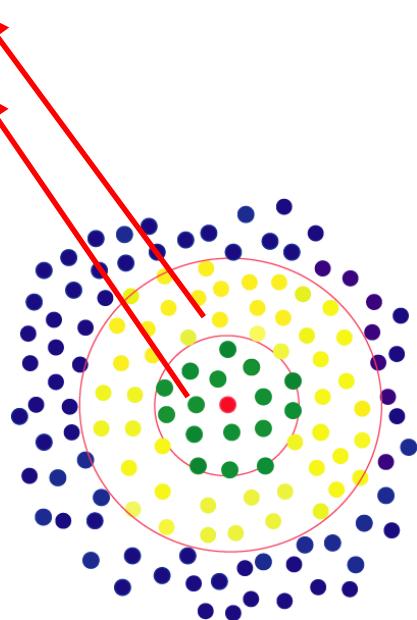
$$\text{Distance}_{12} < (\mathbf{r}_{c1} + \mathbf{r}_{c2})$$

# How to start the O(N) calc. No.2

In the **standard output**, you can confirm the number of atoms in the core and buffer regions as.

```
*****  
Truncation and setting of grids  
*****
```

```
<truncation> Logically truncation of the whole system  
TNaN= 25620 Average NaN= 39.41538  
TSNaN= 24418 Average NaN= 37.56615  
<logical truncation> myid= 0 CpuCell= 1 ct_AN= 1 FNAN NaN 42 9  
<logical truncation> myid= 0 CpuCell= 1 ct_AN= 2 FNAN NaN 21 21  
<logical truncation> myid= 0 CpuCell= 1 ct_AN= 3 FNAN NaN 26 30  
<logical truncation> myid= 0 CpuCell= 1 ct_AN= 4 FNAN NaN 40 27  
<logical truncation> myid= 0 CpuCell= 1 ct_AN= 5 FNAN NaN 79 77
```



Green: core region  
Yellow: buffer region

# How to start the O(N) calc. No.3

3. The dimension of the Krylov subspace can be controlled by

orderN.KrylovH.order  
350

The maximum value of the parameter is determined by the number of basis functions in the truncated cluster. Even if the larger value exceeding the total basis functions in the truncated cluster is used, it does not improve the accuracy.

4. It would be better to start with the default value for the following keywords:

orderN.Exact.Inverse.S      on  
orderN.Expand.Core      on  
orderN.Recalc.Buffer      on

# How to start the O(N) calc. No.4

## 5. Test your O(N) calculations on small systems.

Since the O(N) methods are approximate to the numerical exact diagonalization, it is highly important to find a set of proper parameters using a smaller system with the same chemical environment as the system you are interested in. Otherwise, the accuracy of the calculation is unpredictable.

# Summary

1. The Krylov subspace method is **applicable to a wide variety of systems** including metals, insulators, semiconductors, molecules with any dimensionality.
2. When the crossing point with the conventional diagonalization is set to around 400 atoms, **the typical accuracy** in the total energy is
  - 0.000001 for molecules
  - 0.0001 for semiconductors
  - 0.001 for metalsin Hartree/atom
3. The careful **test calculation** is **indispensable** before performing large-scale calculations.