Kyushu University

(Fukuoka)


Aoki Yuriko
青木 百合子
Blue tree Lily

Keep in mind to be
soft, modest, elegant,…

**I'm doing very basic science in theoretical chemistry, but not in computer science.**


Theoretical Chemistry

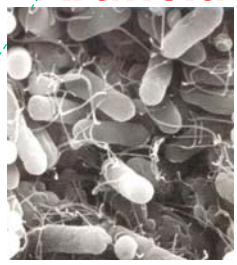
Macro
1m

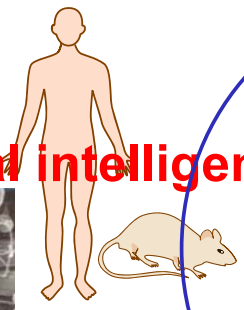
Micro
$10^{-6}$m

Nano
$10^{-9}$m

Pico
$10^{-12}$m

Artificial intelligence



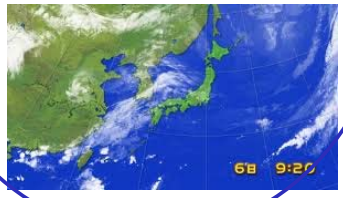


O-157

SYNAPSE

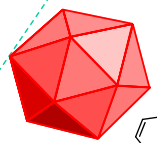
Orbit

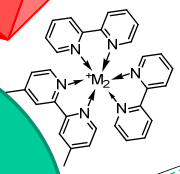
116,464 km



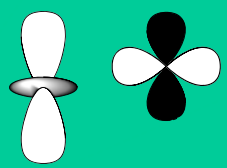
HPC

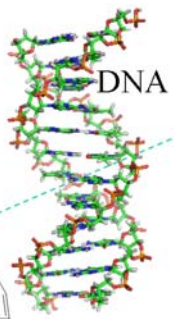
HPC



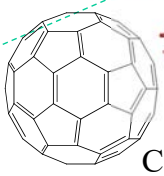


Orbital



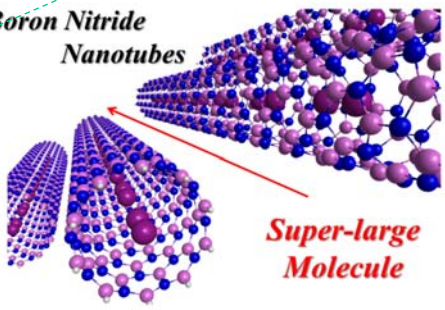


DNA



C60

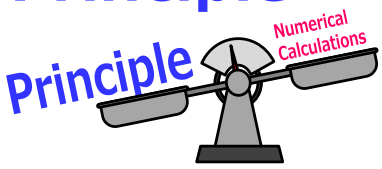
Boron Nitride
Nanotubes

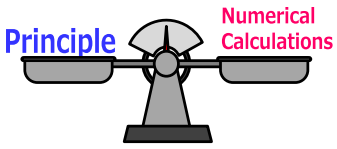


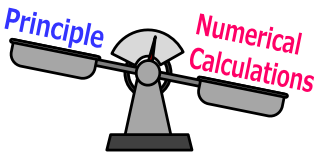
Super-large
Molecule

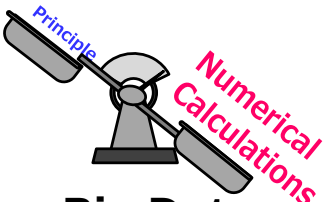
## Basic Science

Principle




Big Data

Applied Science

## 1926's Schrödinger equation $\hat{H}\Psi = E\Psi$

Hartree-Fock-Roothaan eq. Hückel method


Schrödinger


Dirac

1933's Nobel Prize in Physics


Pauli


Slater 1929

1945's Nobel Prize in Physics


Hartree 1928


Fock 1930


Roothaan 1951


Hückel 1936

1954's Nobel Prize in Chemistry


Pauling

1966's Nobel Prize in Chemistry


Mulliken

1981's Nobel Prize in Chemistry


Hoffmann +



Fukui

1991's Nobel Prize in Chemistry

Kohn +

Pople

~Supercomputer~

2013's Nobel Prize in Chemistry


QM/MM Karplus

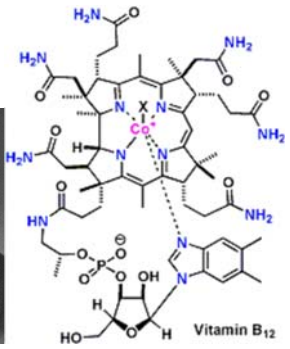

Levitt


Warshel

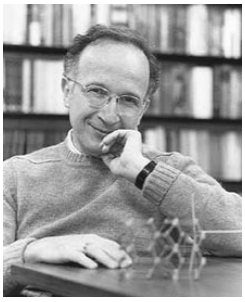

## Robert Burns Woodward → 1981 Nobel Prize in Chemistry

## 1965 Nobel Prize in Chemistry

*for his outstanding achievements in the art of organic synthesis*




## Roald Hoffmann



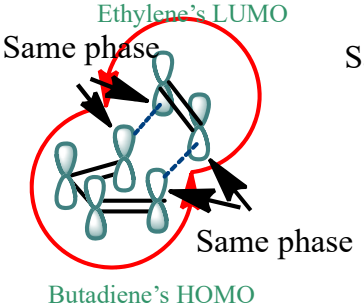
*for their theories, developed independently, concerning the course of chemical reactions*

## Kenichi Fukui

**Woodward-Hoffmann rule** ← Cited Fukui's paper

Reaction occurs

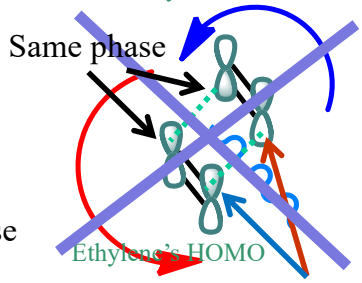
Ethylene's LUMO




Butadiene's HOMO

Reaction NOT occurs

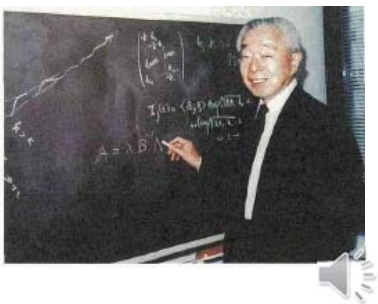
Ethylene's LUMO




Ethylene's HOMO

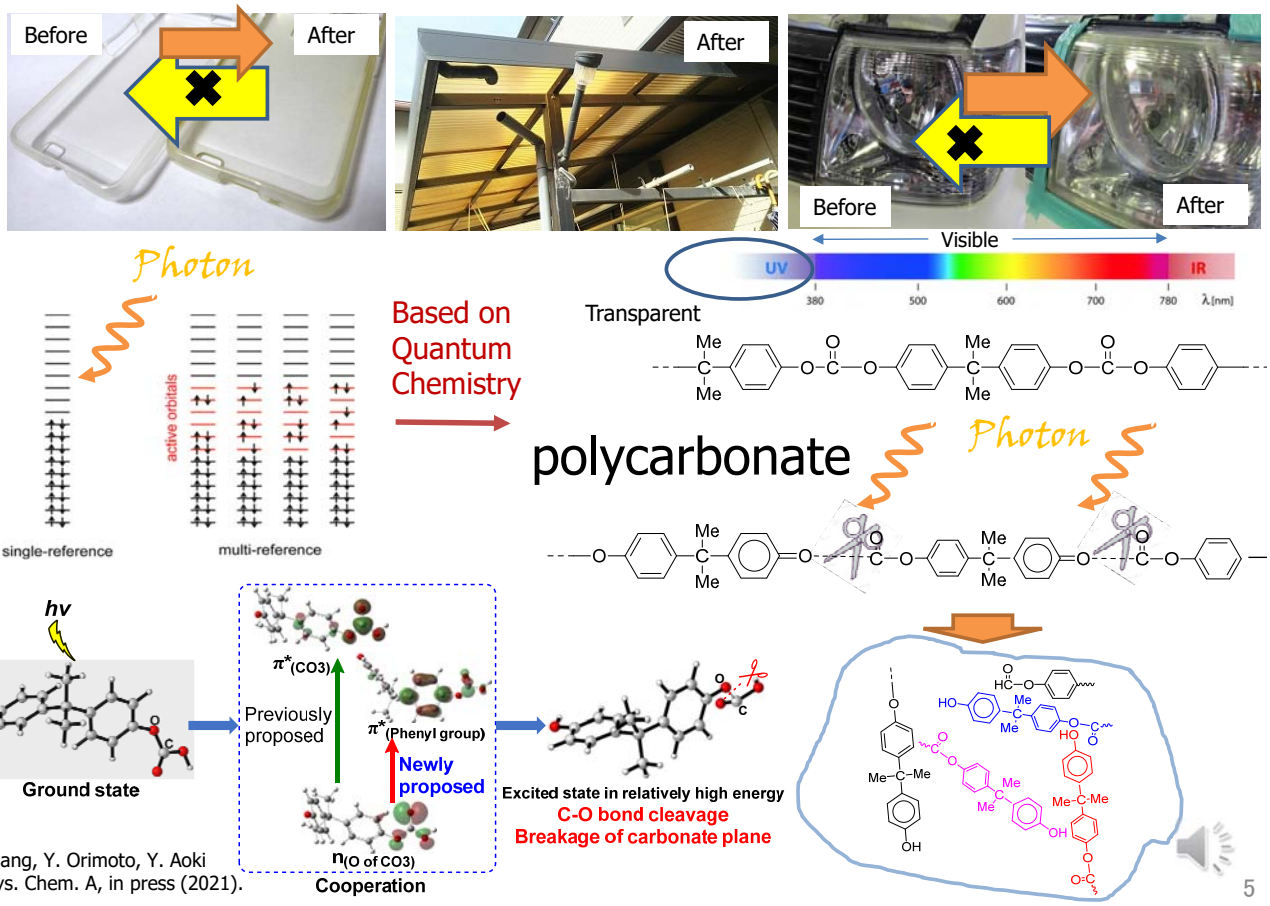
Different phase

Fukui got a phone call on October 18, from Nobel committee



Transparent

# Why polycarbonates get **UV** photo-degradation




X. Huang, Y. Orimoto, Y. Aoki
J. Phys. Chem. A, in press (2021).



# Typical Software for Electronic states calculations

## **Molecules** by basis sets (Quantum chemistry)

Gaussian: http://www.gaussian.com/
GAMESS: http://www.msg.ameslab.gov/gamess/
NWChem: http://www.nwchem-sw.org/index.php/Main_Page
NTChem: https://www.rccs.riken.jp/software_center/jp/success/ntchem/
TURBOMOLE: http://www.turbomole-gmbh.com/
etc.

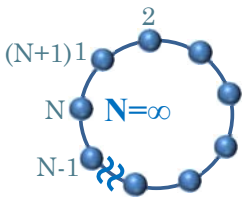
## **Solids** by Plane waves (Periodic boundary condition)

Crystal: http://www.crystal.unito.it/index.php
VASP: https://www.vasp.at/
WIEN2k : http://www.wien2k.at/
siesta : http://www.icmab.es/sieasta/index.php
CPMD : http://www.cpmd.org/
ABINIT : http://www.abinit.org/
PWscf : http://www.quantum-espresso.org/
Phase : http://www.ciss.iis.u-tokyo.ac.jp/dl/
etc.

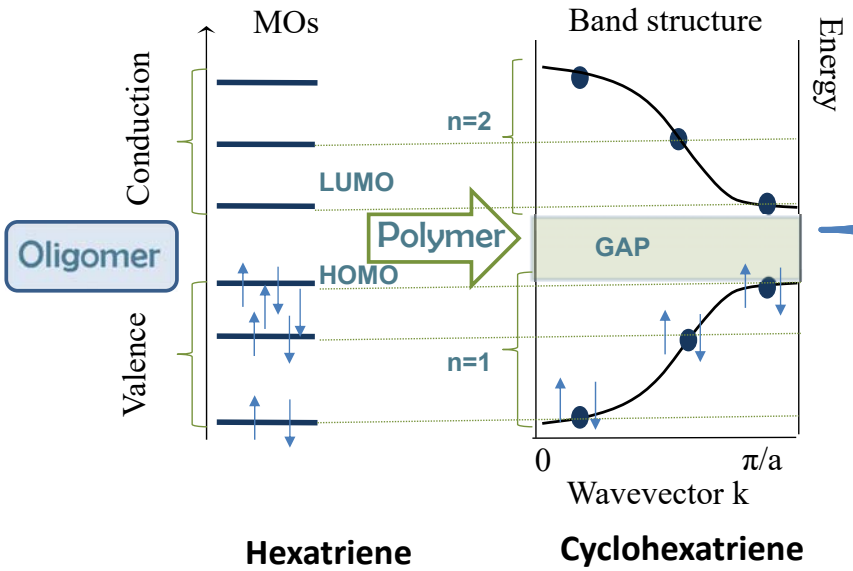
# Parallelization in material science computations were established under plane wave basis but limited


## Finite


## Infinite




Periodic boundary conditions

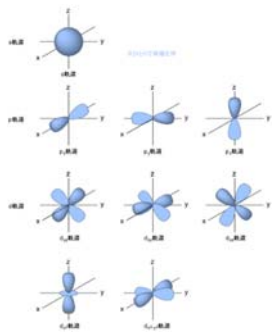



**Uncertainty principle**

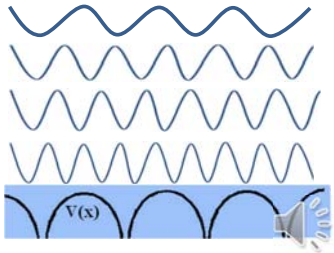
$$\sigma_x \sigma_p \geq \frac{\hbar}{2}$$

Heisenberg

1) Basis functions



2) Plane waves



# Finite

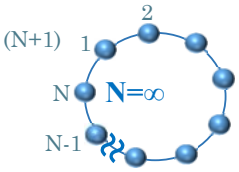
(a)


QC, DFT

M⇒∞

# Infinite

(b)




QC, DFT

## Hückel approach

$$\varepsilon_i = \alpha + 2\beta\cos\left(\frac{\pi i}{M+1}\right), \quad i = 1,2,\cdots,M$$

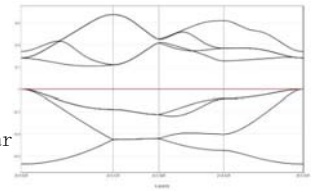
$$\varphi_i = \sqrt{\frac{2}{M+1}}\sum_{r=1}^{M}\sin\left(\frac{\pi r}{M+1}i\right)\chi_r$$

Orbital energy

$$\varepsilon_i = \alpha + 2\beta\cos\left(\frac{2\pi i}{N}\right), \quad i = 1,2,\cdots,N$$

$$\varphi_i = \sqrt{\frac{1}{N}}\sum_{r=0}^{N-1}\exp\left(\tilde{i}\frac{2\pi r}{N}i\right)\chi_r \qquad \tilde{i}:\text{imaginar}$$

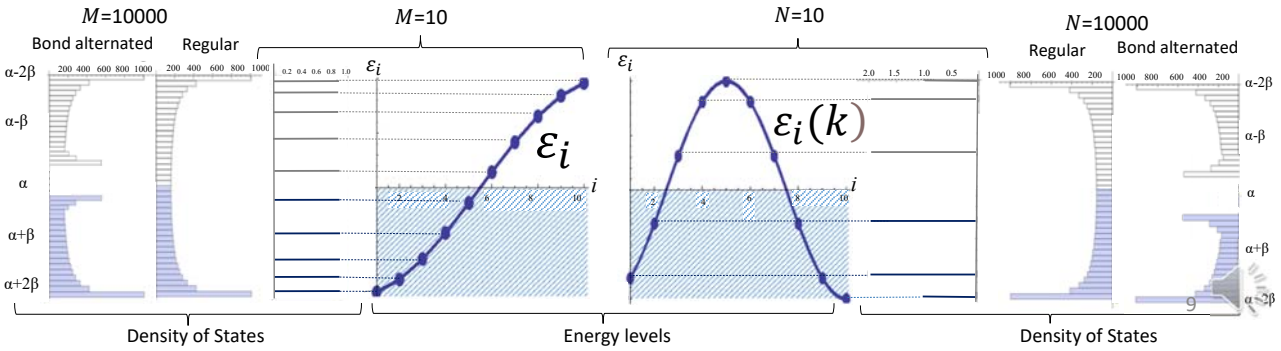
Energy Band



## Hartree-Fock equations

$$\sum_s F_{rs}C_{is} = \sum_s S_{rs}C_{is}\varepsilon_i$$

$$\sum_s F_{rs}(k)C_{is}(k) = \sum_s S_{rs}(k)C_{is}(k)\varepsilon_i(k)$$

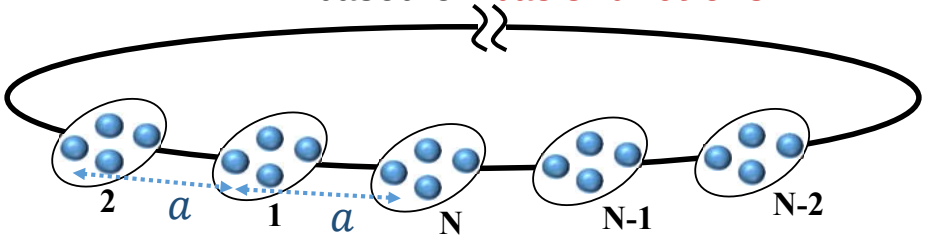



9

# Hartree-Fock Equation for energy band

**under periodic boundary condition**

**based on basis functions**




$$\sum_s F_{rs}(k)C_{is}(k) = \sum_s S_{rs}(k)C_{is}(k)E_i(k)$$

$k$: wavevector

$$X_{rs}(k) = X_{rs}^{(0,0)} + \sum_{l=1}^{L}\left\{\exp(\tilde{i}kl)X_{rs}^{(0,l)} + \exp(-\tilde{i}kl)X_{rs}^{(0,-l)}\right\} \quad X = F, S$$

2-electron integrals

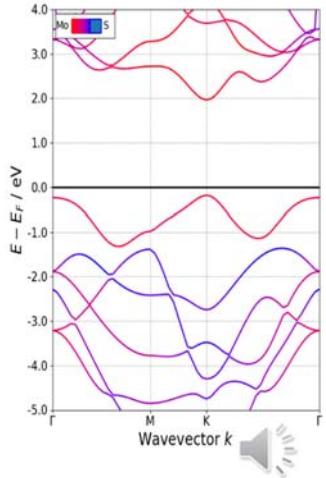
$$F_{rs}^{(0,l)} = I_{rs}^{(0,l)} + \sum_h\sum_j\sum_t\sum_u P_{tu}^{(h,j)}\left\{2\left\langle \begin{matrix} 0 & l \\ r & s \end{matrix}\middle|\begin{matrix} h & j \\ t & u \end{matrix}\right\rangle - \left\langle \begin{matrix} 0 & h \\ r & t \end{matrix}\middle|\begin{matrix} l & j \\ s & u \end{matrix}\right\rangle\right\}$$

Time-consuming

How to calculate in efficient parallel

$$P_{tu}^{(h,j)} = N^{-1}\sum_{k'=0}^{2\pi}\sum_{N=1}^{occ}\exp\left\{\tilde{i}k'(j-h)\right\}C_{it}^{*}(k')C_{iu}(k')$$

$$\phi_i^k = (1/N)^{1/2}\sum_{j=1}^{n}\sum_{u=1}^{n}\exp(\tilde{i}kj)C_{iu}(k)\chi_u(r-ja)$$




10

# Basic formula in quantum chemistry (QC)

## based on basis functions

Schrödinger equation $\hat{H}_{\text{elec}}\Psi = E_{\text{elec}}\Psi$


Schrödinger

Hamiltonian $\hat{H}_{elec}$ , WavefunctionΨ and Energy $E_{\text{elec}}$

$$\hat{H}_{\text{elec}} = -\frac{1}{2}\sum_{i}^{n}\nabla_i^2 - \sum_{i}^{n}\sum_{A}^{N}\frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i}^{n}\sum_{j<i}^{n}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

Hartree-Fock-Roothaan (HFR) approximation

$$\hat{H}_{\text{elec}}\tilde{\Psi}^{\text{HF}} = E_{\text{elec}}^{\text{HF}}\tilde{\Psi}^{\text{HF}} \qquad \psi_i(\mathbf{r}) = \sum_{\mu=1}^{N} C_{i\mu}\varphi_\mu(\mathbf{r})$$

$$\tilde{\Psi}^{\text{HF}} = \frac{1}{\sqrt{n!}}\begin{vmatrix} \psi_1(\mathbf{r}_1) & \psi_2(\mathbf{r}_1) & \cdots & \psi_n(\mathbf{r}_1) \\ \psi_1(\mathbf{r}_2) & \psi_2(\mathbf{r}_2) & \cdots & \psi_n(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{r}_n) & \psi_2(\mathbf{r}_n) & \cdots & \psi_n(\mathbf{r}_n) \end{vmatrix}$$

Slater determinant

$$\int d\mathbf{r}\,\psi_i^*(\mathbf{r})\psi_j(\mathbf{r}) = \delta_{ij}$$


Slater
1929


Hartree
1928


Fock
1930


Roothaan
1951

Roothaan, C. C. J. New Developments in Molecular Orbital Theory. Rev. Mod. Phys.1951, 23, 69-89



# Hartree-Fock-Roothaan equation

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon} \Rightarrow \mathbf{FC} = \mathbf{C}\boldsymbol{\varepsilon}$$

Fock matrix  Overlap matrix

Non-linear Equation

$$\mathbf{F} = \mathbf{H}^{\text{core}} + \mathbf{T} + \mathbf{V} + \mathbf{G}$$

$$T_{\mu\nu} = \int d\mathbf{r}_1\,\varphi_\mu(\mathbf{r}_1)\left(-\frac{1}{2}\nabla_1^2\right)\varphi_\nu(\mathbf{r}_1)$$

$$V_{\mu\nu} = \int d\mathbf{r}_1\,\varphi_\mu(\mathbf{r}_1)\left(\sum_{A}^{N}\frac{-Z_A}{|\mathbf{r}_1 - \mathbf{R}_A|}\right)\varphi_\nu(\mathbf{r}_1)$$

$$G_{\mu\nu} = \sum_{\sigma}^{n}\sum_{\lambda}^{n} D_{\sigma\lambda}\{2(\mu\nu|\sigma\lambda) - (\mu\lambda|\sigma\nu)\}$$

$$(\mu\nu|\sigma\lambda) = \int d\mathbf{r}_1\int d\mathbf{r}_2\,\varphi_\mu(\mathbf{r}_1)\varphi_\nu(\mathbf{r}_1)\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\varphi_\sigma(\mathbf{r}_2)\varphi_\lambda(\mathbf{r}_2)$$

$$D_{\mu\nu} = \sum_{i}^{n/2} C_{\mu i}C_{\nu i}$$

$$\psi_i = \sum_{\mu=1}^{N} C_{\mu i}\varphi_\mu(\mathbf{r})$$

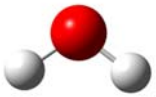
Electron-electron repulstion

Self Consistent Field

( SCF) method

# Process of calculations

| x | y | z |
|---|---|---|
| 0.0000000 | 0.0000000 | 0 |
| 0.7575408 | -0.5870796 | 0 |
| -0.7575408 | -0.5870796 | 0 |








# Difficulty in making G matrix

$$(\mathrm{rs}\,|\,\mathrm{tu}) = \int d\mathbf{r}_1 \int d\mathbf{r}_2\, \varphi_r(\mathbf{r}_1)\varphi_s(\mathbf{r}_1)\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}\varphi_t(\mathbf{r}_2)\varphi_u(\mathbf{r}_2)$$

$$G_{rs} = \sum_t^N \sum_u^N D_{tu}\left\{2(\mathrm{rs}\,|\,\mathrm{tu}) - (\mathrm{ru}\,|\,\mathrm{ts})\right\}$$

For large systems
Number of Basis functions
$N$ = 1000 ~ 10000

**Density matrix**

**2e-integrals**

$N = 5000 \rightarrow \boldsymbol{G}, \boldsymbol{D}$ ~ 120 MB ($10^8$)
(~500atoms) $\langle\varphi_i\varphi_j\,|\,\varphi_k\varphi_l\rangle$ ~600 TB ($10^{15}$)

**Density Functional Theory (DFT)**

- ***G*** is composed of double summations of enormous number of 2e-integrals

Kyushu University
RESEARCH INSTITUTE FOR INFOMATION TECHNOLOGY
Dr. Hiroaki Honda

# Dr. Ishimura at IMS is developing

"SMASH" --- Linear scaling at 10,000 cores ~ OK

(Scalable Molecular Analysis Solver for High performance computing systems)

## Hybrid of OpenMP and MPI

Fock matrix $F_{\mu\nu} = H_{\mu\nu} + \frac{1}{2}\sum_{\lambda=1}^{N}\sum_{\sigma=1}^{N} D_{\lambda\sigma}\{2(\mu\nu|\lambda\sigma) - (\mu\lambda|\nu\sigma)\}$

```
!$OMP parallel do schedule(dynamic,1) reduction(+:Fock)
do μ=n, 1, -1                          <----- OpenMP
  do ν=1, μ
    μν=μ*(μ+1)/2+ν
   λstart=mod(μν+mpi_rank,nproc)+1
   do λ=λstart, μ ,nproc               <----- MPI
     do σ=1, λ
       AO 2e-integrals (μν|λσ)   The most time consuming
     enddo
   enddo
  enddo
enddo
!$OMP end parallel do
call mpi_allreduce(Fock)
```

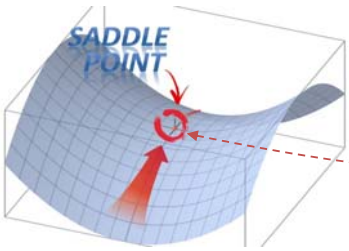


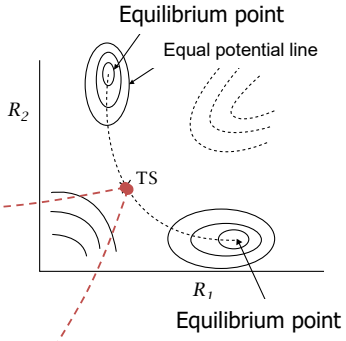
# Solid catalysis---Quantum chemistry is necessary

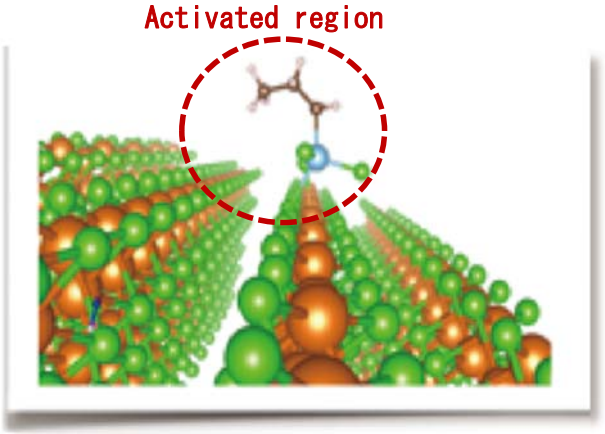
**Fukui's theory (1981, Nobel prize)**
**Intrinsic Reaction Coordinate (IRC)**

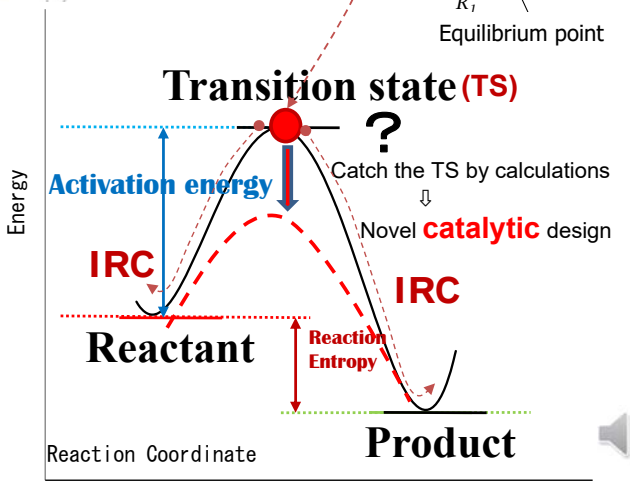
$$m_\alpha \frac{d}{dt}\left(\frac{dX_\alpha}{dt}\right) = -\frac{\partial U}{\partial X_\alpha}$$

$$x_1 = m_1^{\frac{1}{2}} X_1, \quad x_2 = m_1^{\frac{1}{2}} Y_1, \quad x_3 = m_1^{\frac{1}{2}} Z_1$$
$$x_4 = m_2^{\frac{1}{2}} X_2, \quad x_5 = m_2^{\frac{1}{2}} Y_2, \quad x_6 = m_2^{\frac{1}{2}} Z_2, \quad \cdots$$

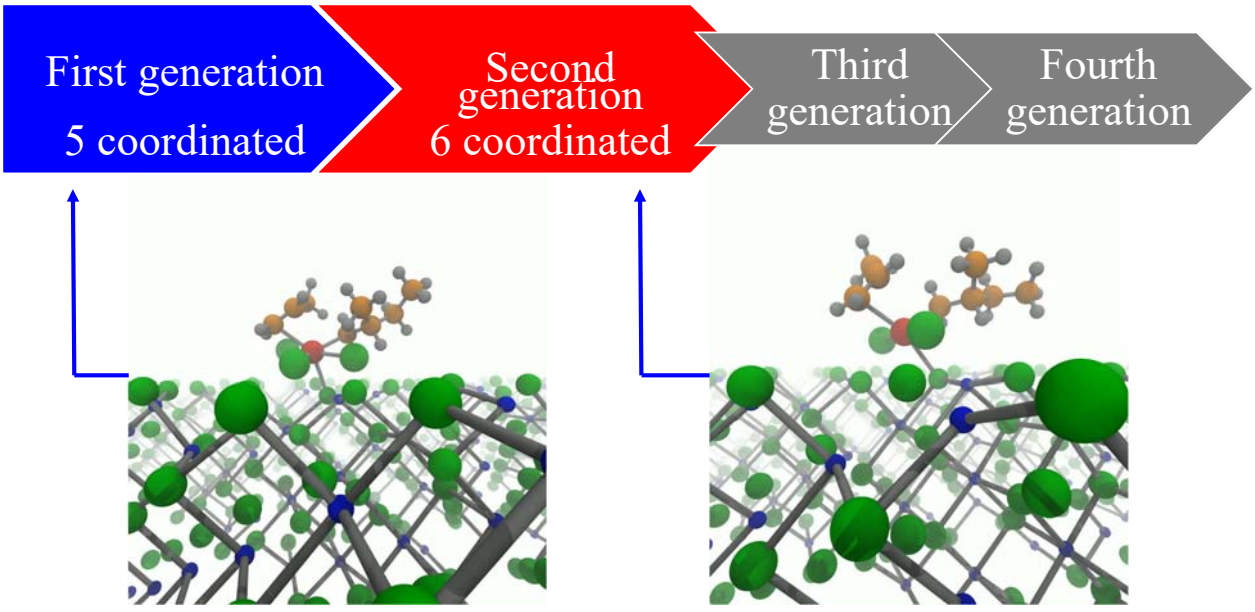


# Ziegler-Natta catalytic reaction on the plane wave bases




Propylene polymerization activity : 5 coordinated < 6 coordinated

VASP , EDIFFG －0.01（TS・Convergency criterion in geometry optimization）, Cutoff 400 eV



## Toward large systems by QC approach for finite size

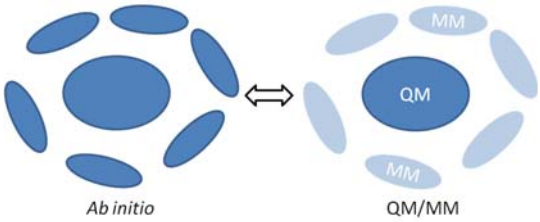
## Group I hybrid methods

2013's Nobel Prize in Chemistry

QM/MM method

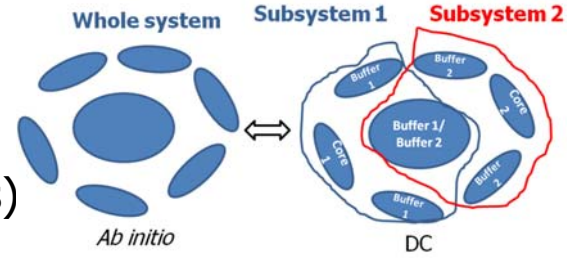
$$H = H_{QM} + H_{MM} + H_{QM/MM}$$
$$E = E_{QM} + E_{MM} + E_{QM/MM}$$

+MD



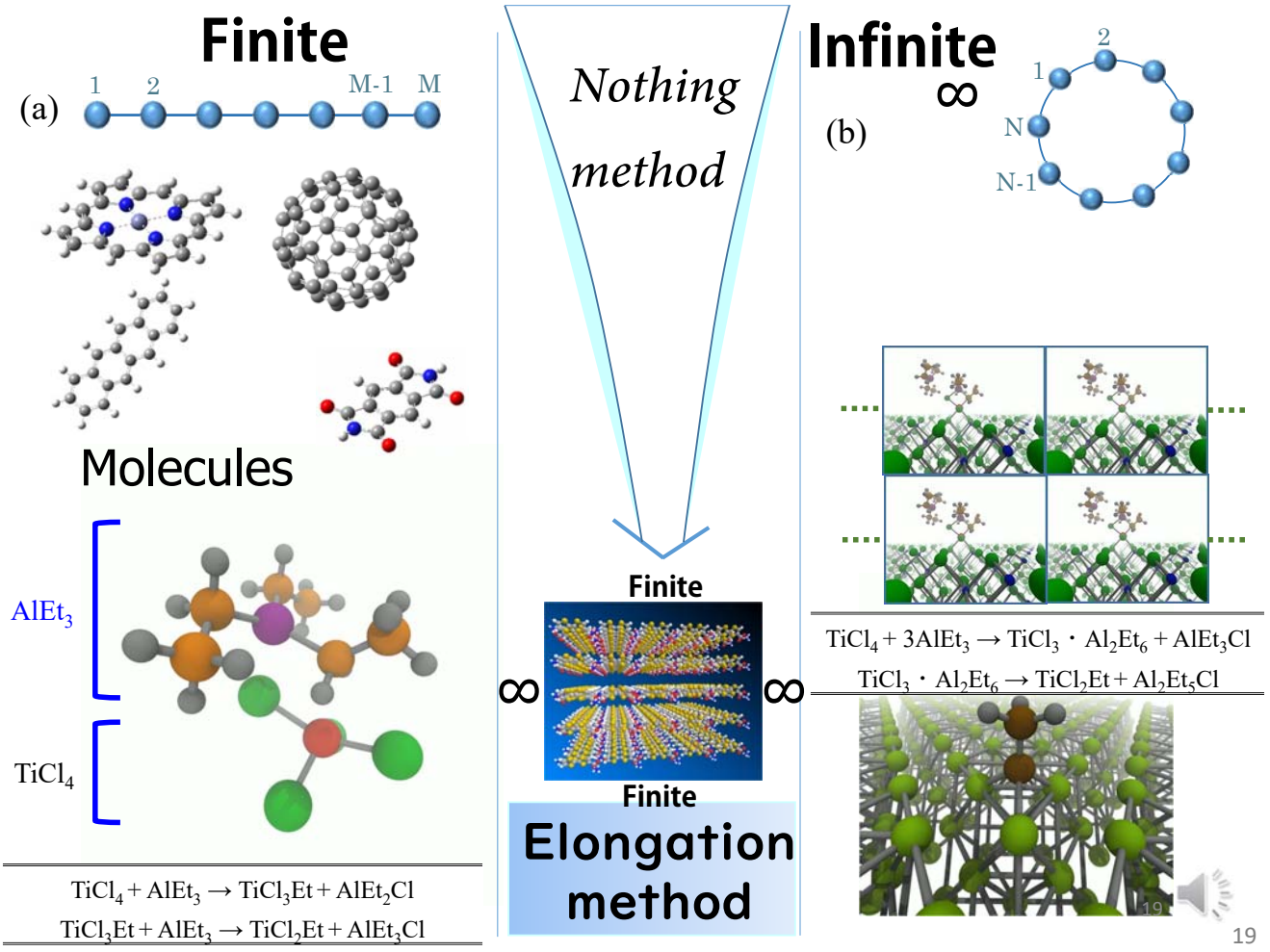

## Group II concentrates on the whole system

$E_{TOTAL}$=?




Methods of Group II: (for Bio systems)

A. Divide and conquer (DC) method
B. Fragment molecular orbital (FMO) method
C. Systematic molecular fragmentation (SMF) method
D. Molecular fractionation with conjugate caps (MFCC) method





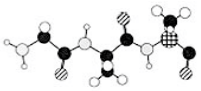

## Elongation Method [O(*N*)] (←$N^4$) with only one-core

(Theoretical Synthesis of Random Polymers)

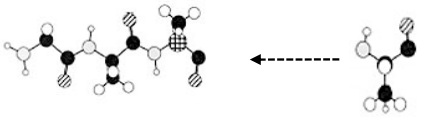
J. Chem. Phys. 95 (1991) 5419.
Imamura, Aoki, Maekawa

***Polymerization reaction***

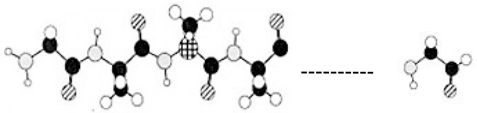
**Initiation**

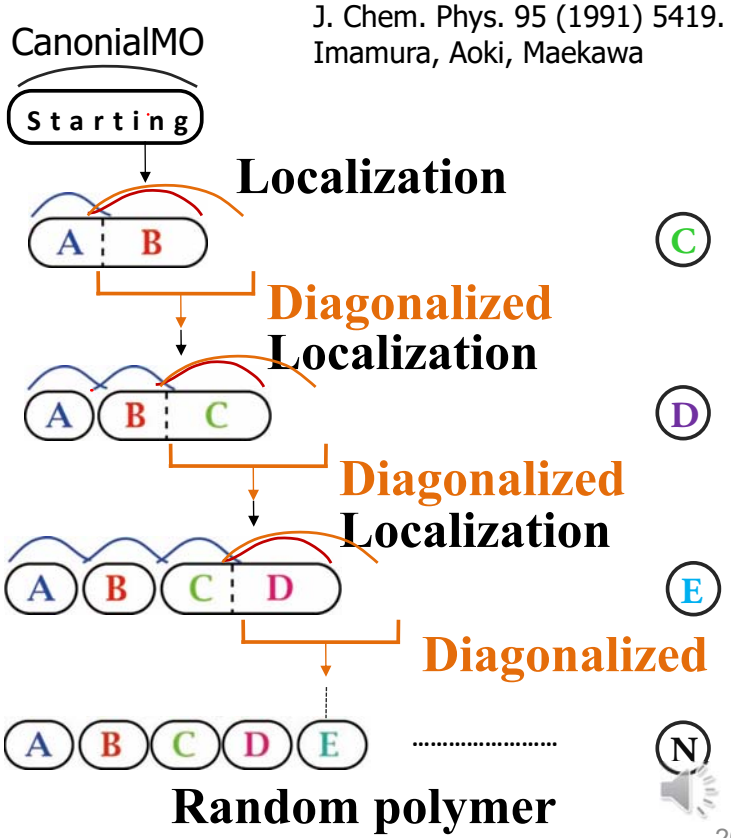


**Propagation**

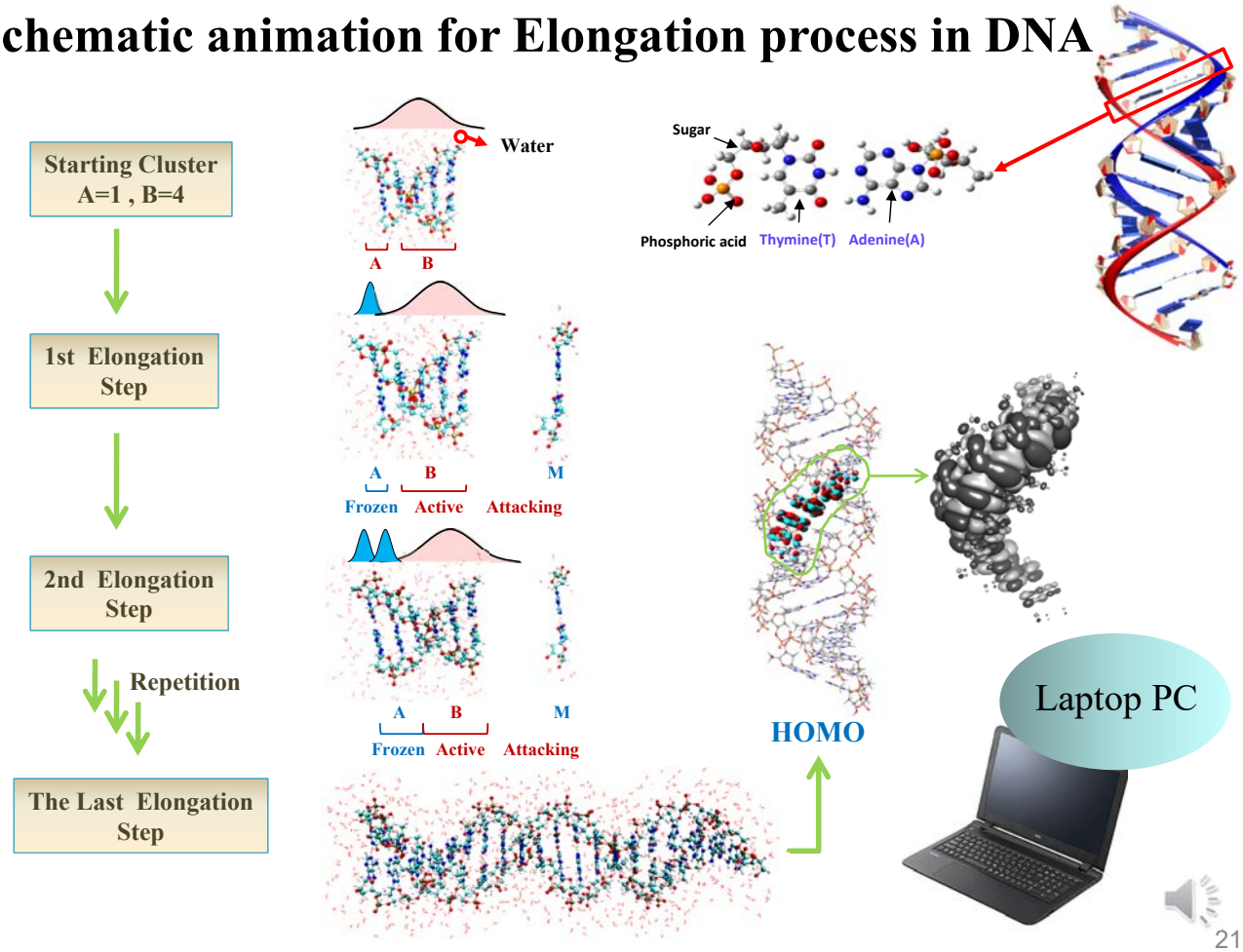


**Termination**





# Schematic animation for Elongation process in DNA



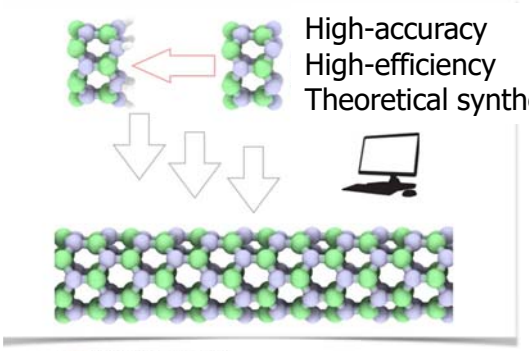



# Linear Scaling Method for Large Systems
# - Elongation method –

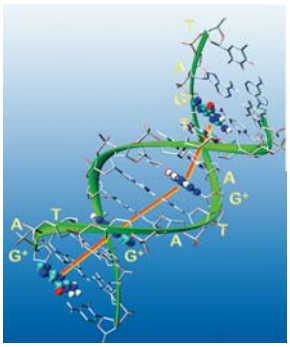
Our own codes since1990's

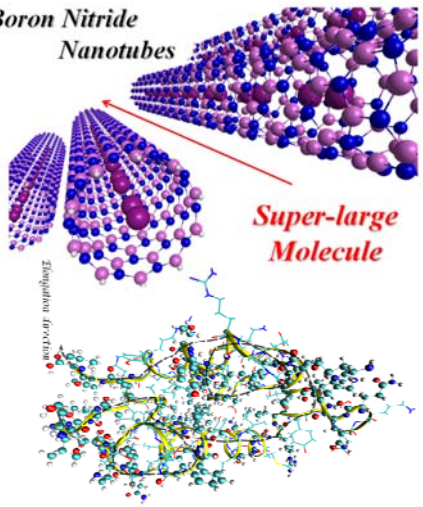
## New methodology

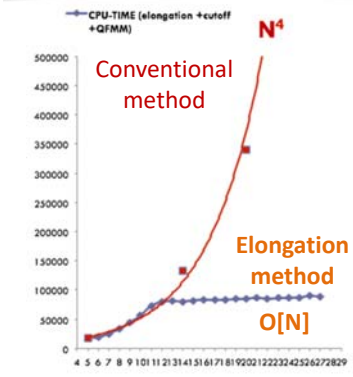
Ex) Our Elongation method



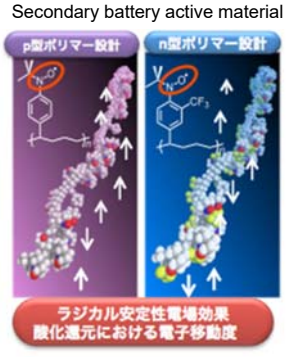
High-accuracy
High-efficiency
Theoretical synthesis





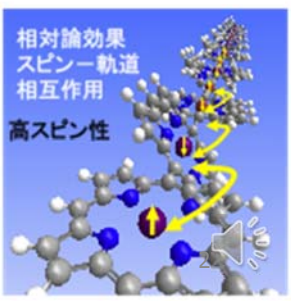




Secondary battery active material



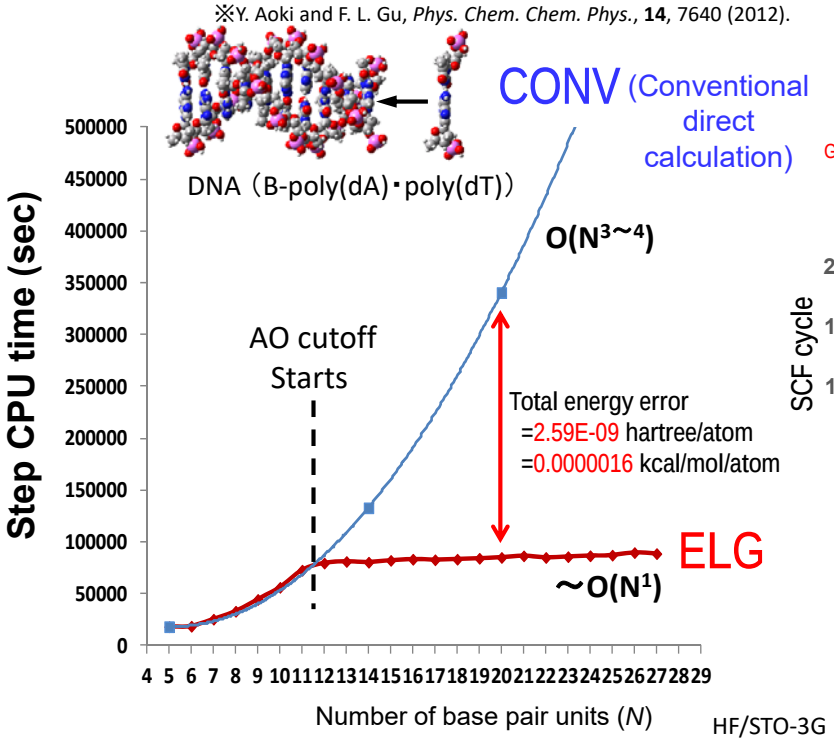

Light collection antenna material

Organic magnetic polymer

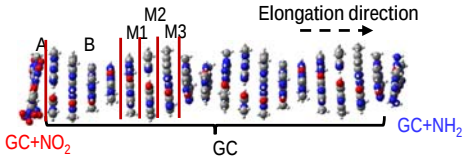


# Efficiency and accuracy of ELG method

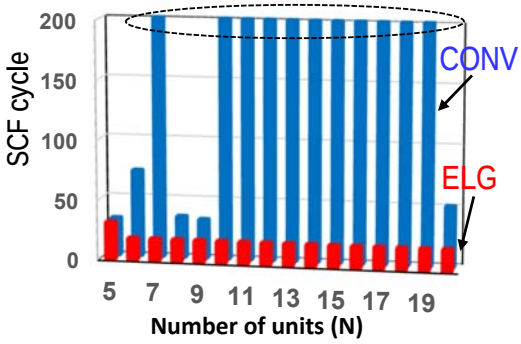
※Y. Aoki and F. L. Gu, *Phys. Chem. Chem. Phys.*, **14**, 7640 (2012).




## SCF convergence




## SCF was not converged




ELG shows better SCF convergence
➔ Advantage for small band gap system

**High SCF convergency for the systems for which conventional method cannot treat**

## High efficiency & high accuracy

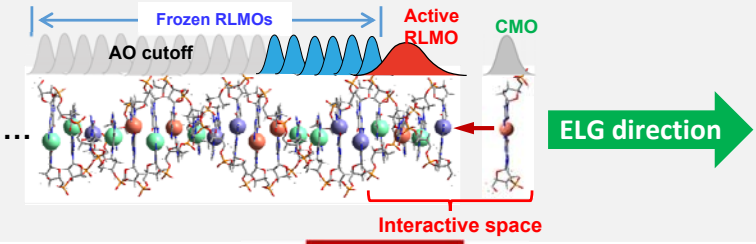
- Linear scaling even by one-core
- High chemical accuracy even with 1,000,000atoms



**Conventional QC calc. --- More than $N^{3\sim4}$ order cost**

## One-way

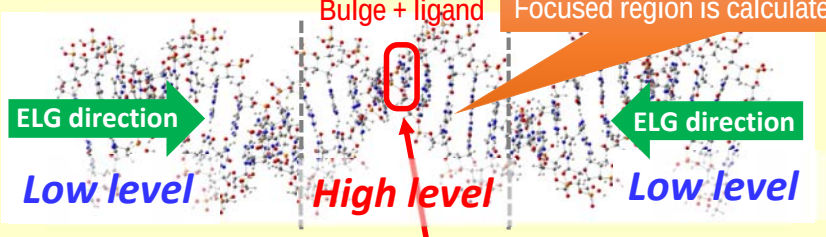
**O(N) Elongation method --- High efficient & accurate calculations for huge molecules**

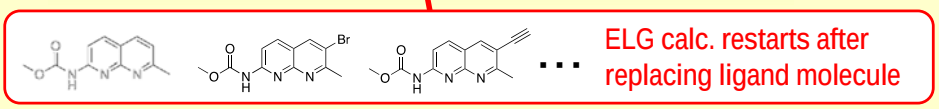



## Two-way

## Multi-layered ELG-OPT ab initio method

アニメーション



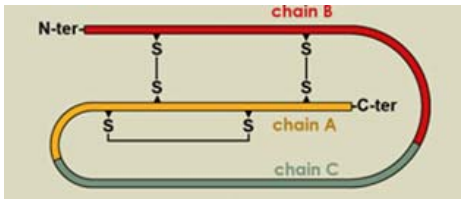




➔ Potential for high throughput molecular design/exploration for DNA bulge/mismatch recognition

# Three Dimensional(3D)-ELG method

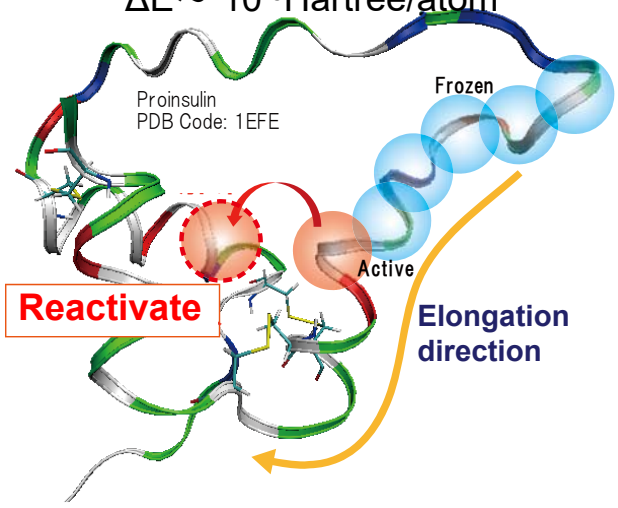
94864

**Proinsulin**




Total energy error

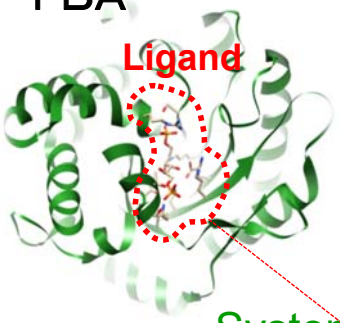
ΔE～ $10^{-8}$Hartree/atom




**Proteins**

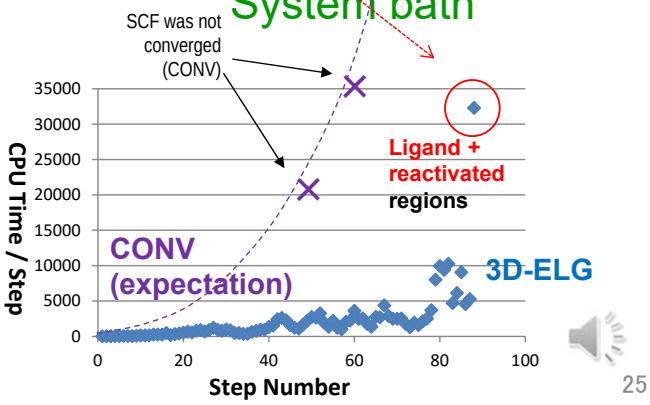
ΔE～$10^{-6}$a.u./atom

FBA




アニメーション

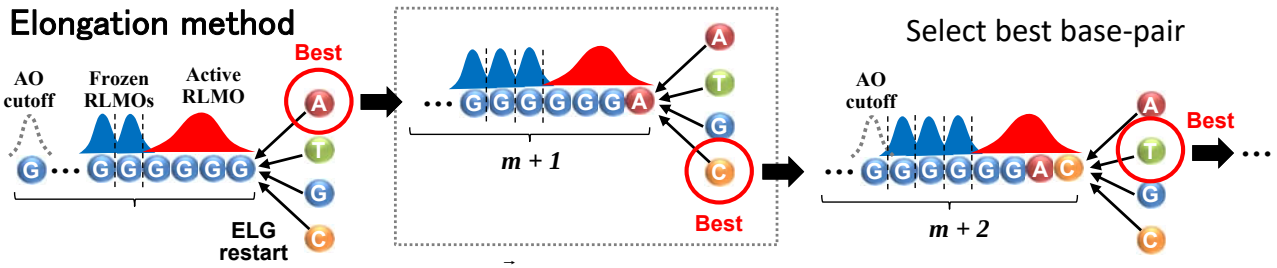
System bath






# The Best NLO functional automatic search

Elongation method

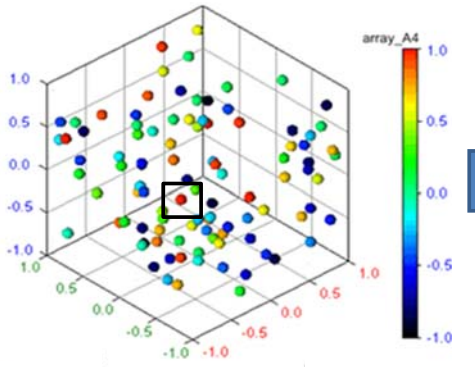



NLO: Nonlinear optical perperty

Total energies are differentiated by electric Field

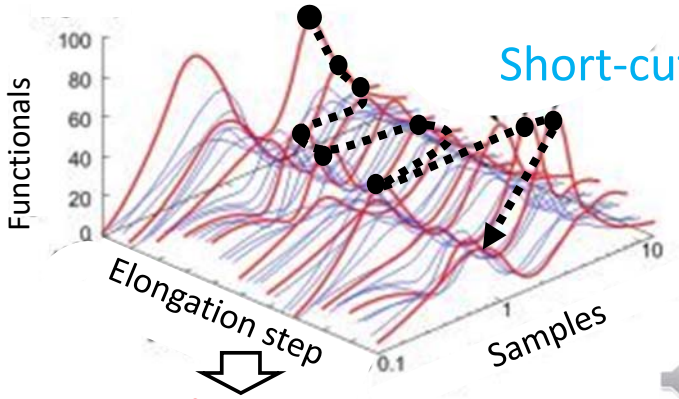
# Automatical finding of best functionals

(a)Data base (Conventional)




All the points must be calculated

(b)Taylor made(Elongation method)

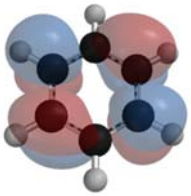



Nice fit to Genetic algorism

# Combination between ELG – Machine Learning

*Small data → Big data*

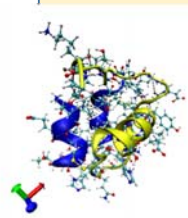
***Finite***
***Non-periodic***
*ex.* Gaussian, etc



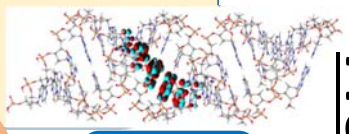
***Intermediates***
***Non-periodic***
**3D-ELG method**
1core × thousands

- High accurate physical properties
- Order-N method even with 1 CPU
- Enormous physical values for samples can be obtained at a time with 100 scalability

***Infinite***
***periodic***
*ex.* VASP, etc



**Problem**
No input data !



***No data for electronic states in polymers and materials***

ELG

DATA

Machine Learning

NN

Prediction

Experimental data

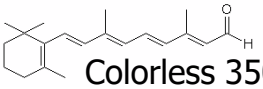
Experimental data
Low accuracy

High accuracy



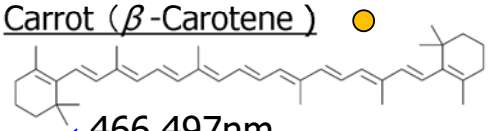
# How can you recognize the color of materials ?

Retinal Vitamin A ○



Colorless 350 nm

Carrot（β-Carotene） ●

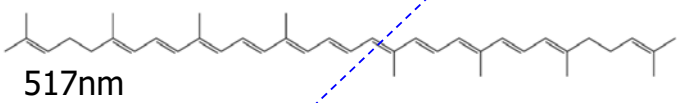


Orange

466 497nm

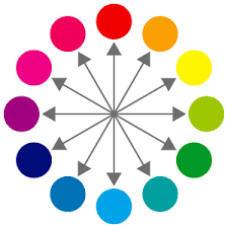
Red

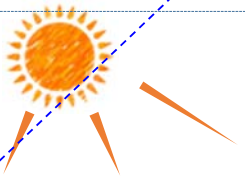
Tomato（Lycopene） ●



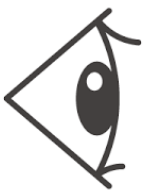
517nm

Complementary colors

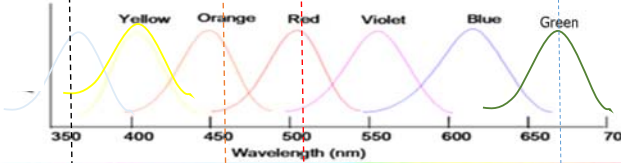




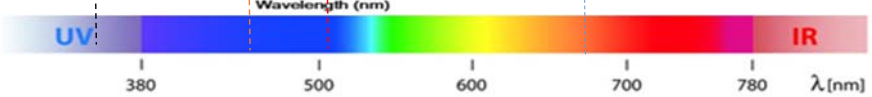
670nm



Reflected colors are seen



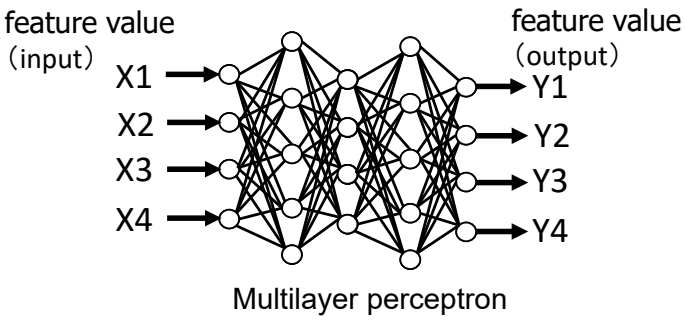

Absorption Spectrum

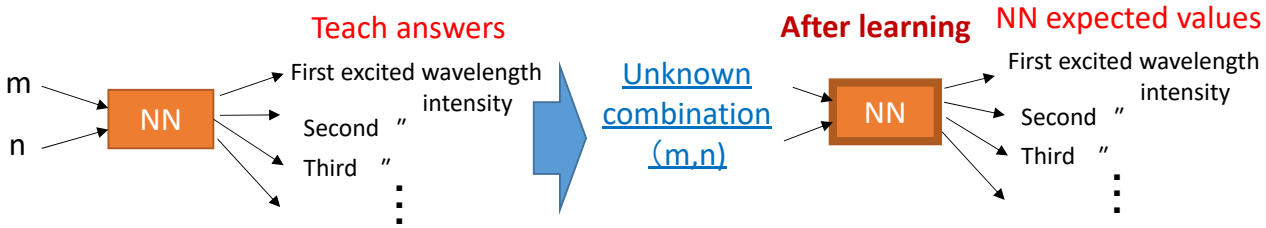


# Polyacetylene block polymer, Photo-absorption property
# Neural network (NN) ELG- method

**Polyacetylene block polymer**

$H\text{-}(CR_1{=}CR_1)_m\text{-}(CR_2{=}CR_2)_n\text{-}H$



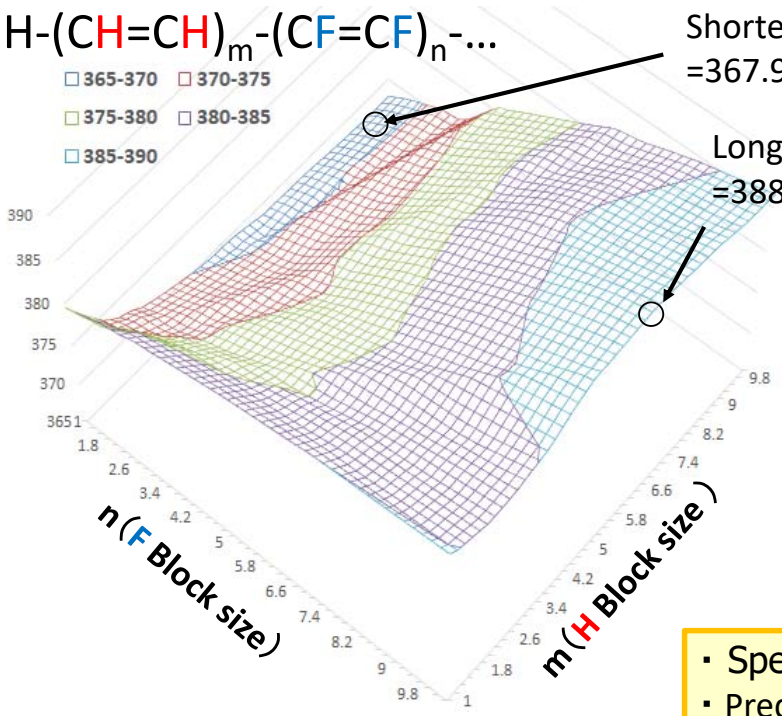

Teach NN about known relationship between m,n and excitation wawe-lengths, intensities

Predict the excitation wawe-lengths, intensities unknown for NN


29

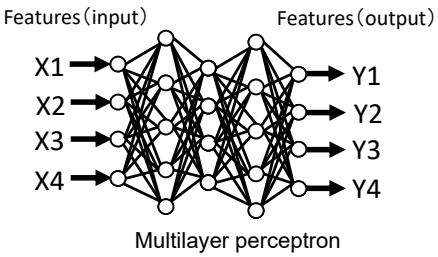
# Excitation energies by neural network Landscape

$H\text{-}(CH{=}CH)_m\text{-}(CF{=}CF)_n\text{-}\ldots$




Shortest absorption wave-length =367.95 nm **(m=9.6, n=1)**

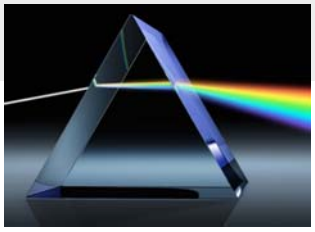
Longest absorption wave-length =388.56nm **(m=6.6, n=10)**

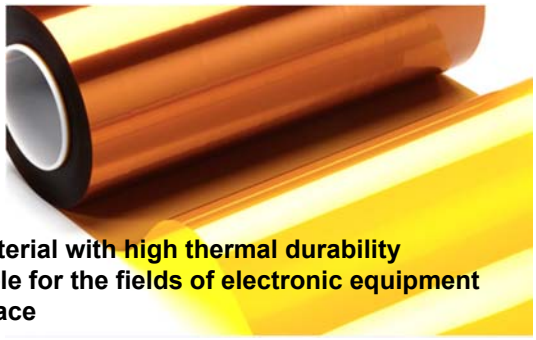



- Specific feature values
- Prediction of Max. and Min.

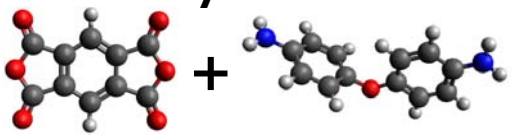
30

# Refractive index

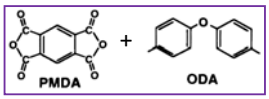




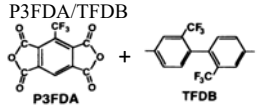
## Acid anhydride + Diamide



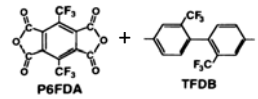
-Polymer material with high thermal durability
-Indispensable for the fields of electronic equipment and aerospace

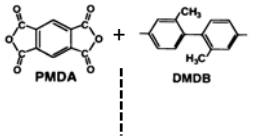



P3FDA/TFDB

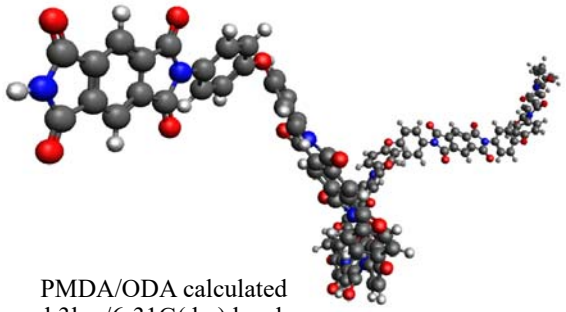



P6FDA/TFDB




Polymerization



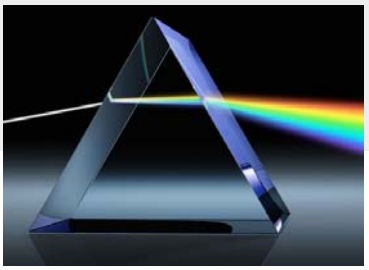
PMDA/ODA calculated
b3lyp/6-31G(d,p) level

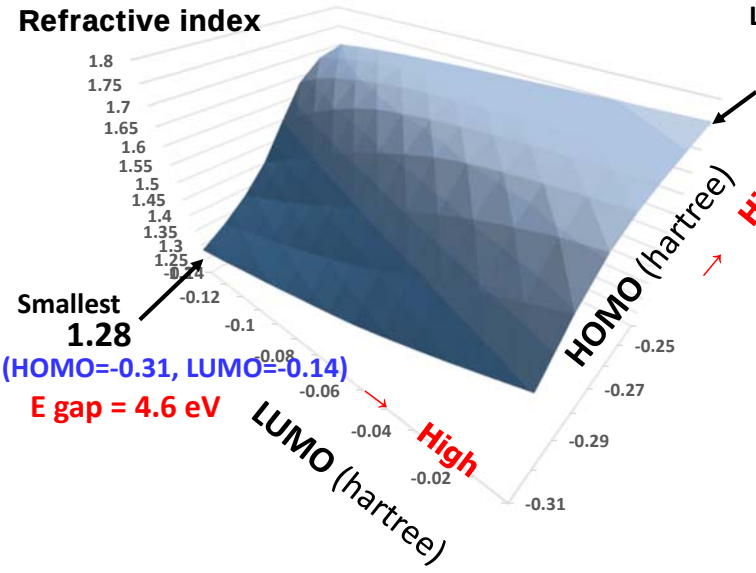
**A lot of variety of combinations**
**➔ Difficulty in material control**



## N=16 Land scape for HOMO, LUMO

# Refractive index





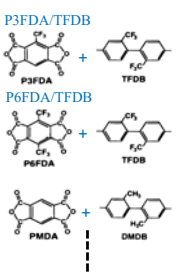

Largest **1.76**
(HOMO=-0.24, LUMO=-0.01)
E gap = 6.3 eV

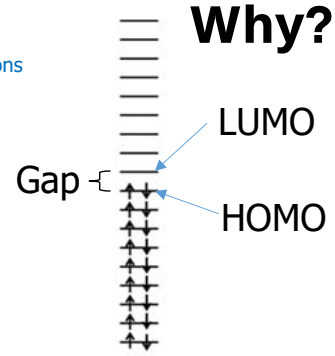
Smallest **1.28**
(HOMO=-0.31, LUMO=-0.14)
E gap = 4.6 eV

● High HOMO + High LUMO
No matter about gap

**Why?**

Plenty of combinations



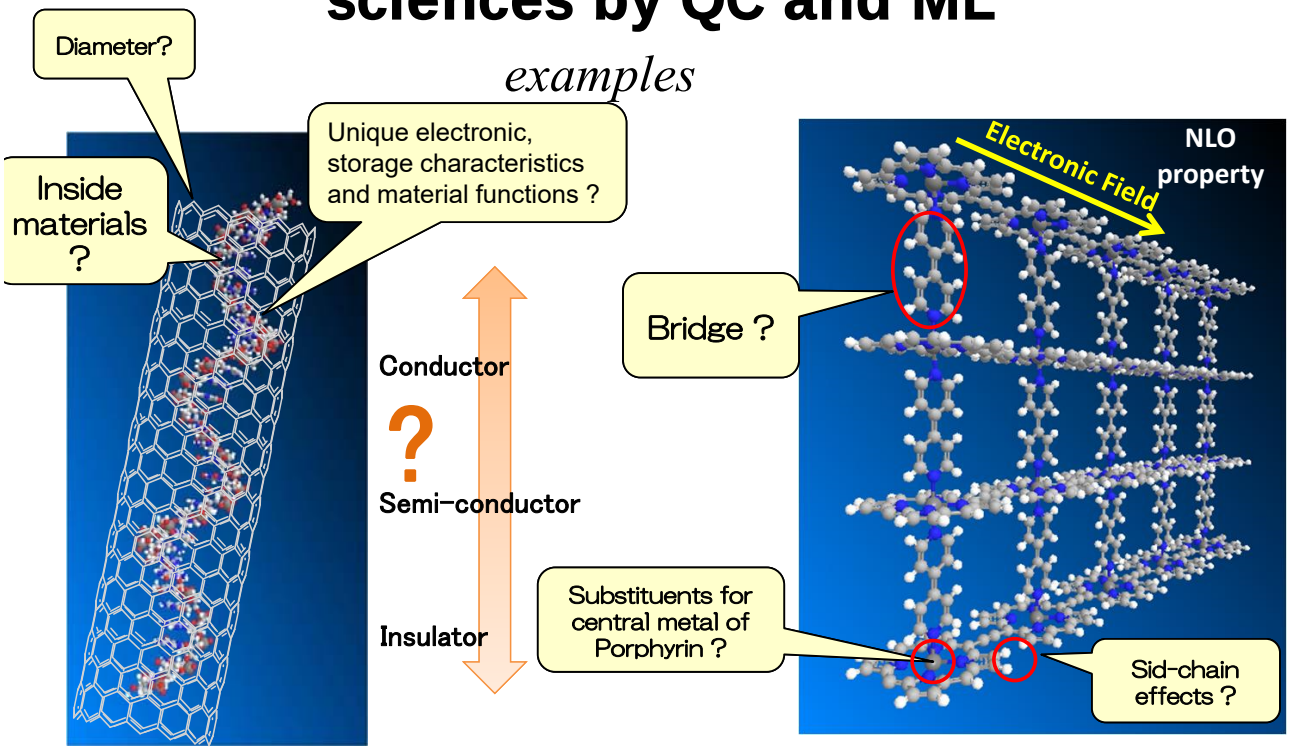

single-reference

Energy levels of 「HOMO」・「LUMO」 control refractive index

**ELG method + NN➔ fast evaluation of feature values**

# Molecular design for material sciences by QC and ML

*examples*

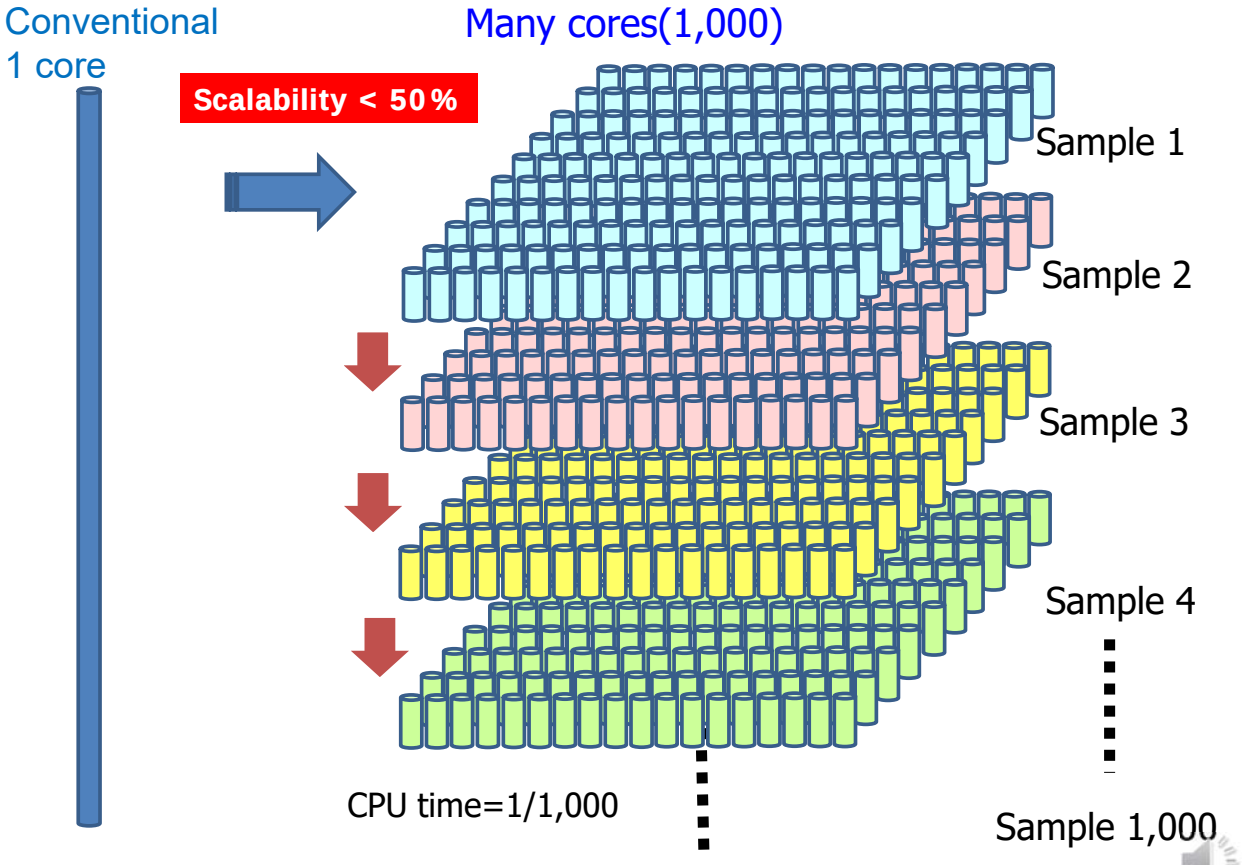



**ML: Thousands possibility for molecular designs**

**QM: Scientific analysis for WHY ?**

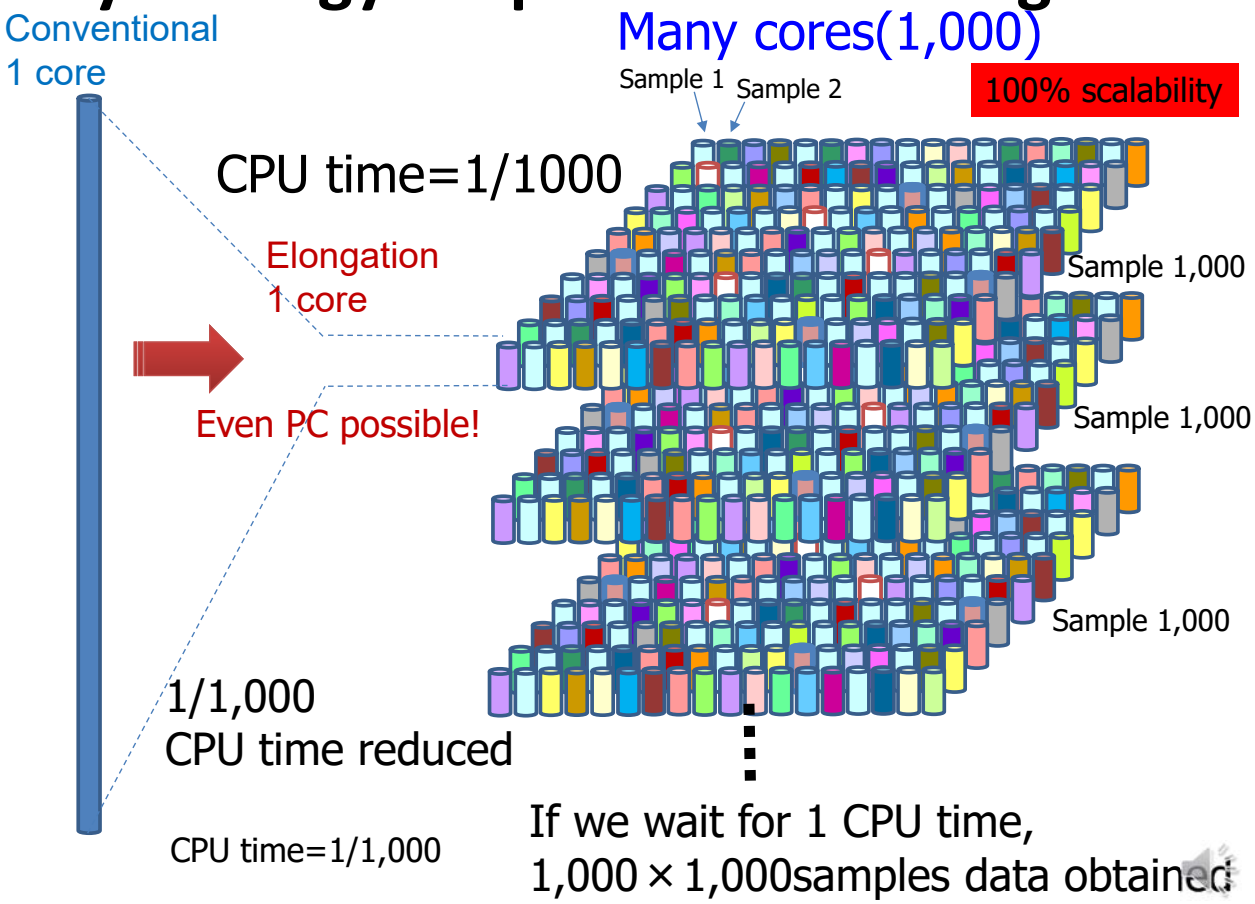


# General Parallelization Concept




CPU time=1/1,000 to get 1,000 samples=1

# My strategy for parallelization big data






## Large scale calculations in Quantum Chemistry

- High accuracy
- A few CPUs
- Fast
- **100% scalability**


Super-computers

Laboratory base
Cluster system

Laptop PC

Smart-phone

Wrong
direction?

***Order (N) calculation***
***"w or w/o supercomputer "***

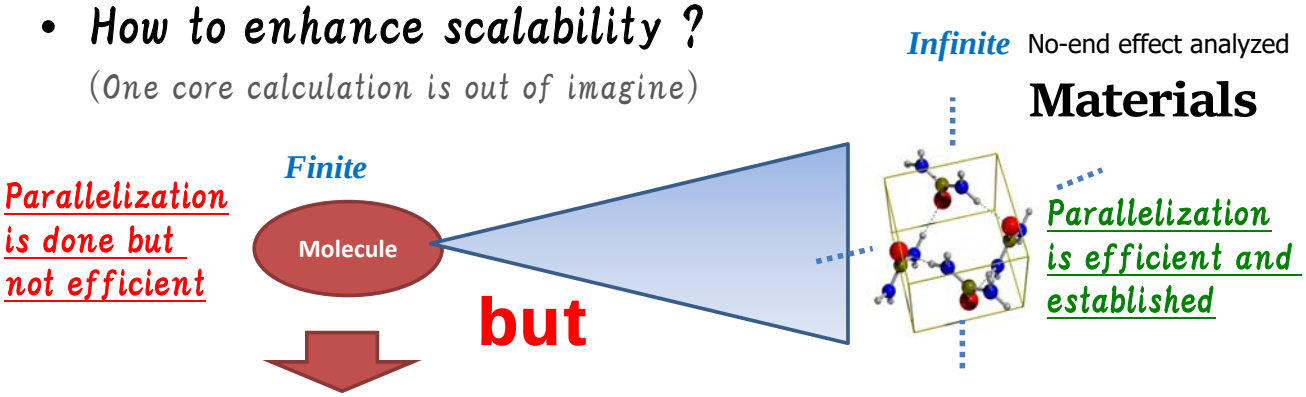
***Thansands samples***

***Machine Learning***

***Efficient Material Science***

# Summary

- How to use plenty of cores by QC material science ?
  (Most of O(N) method depends on hardware of parallel computer)
- How to enhance scalability ?
  (One core calculation is out of imagine)




- One-core calculation must be accelerated first
  (Theory and method in science must be developed)
- Each job is independent using full cores → 100% scalability
  (independent from ability of middleware and number of cores)
- Plenty of samples in parallel are possible → Big data analysis



# Messages to young researchers in material sciences

• Even quantum computer will be developed, one can catch only numerical numbers in each material and don't know sustainable guiding principles in material science. Please don't forget the thinking way using your head even HPC can calculate efficitently.

• Please find some scientific purpose even in the field of computer science, otherwise no meaningful applications found even though you produced HPC technology.

• Please study basic science (mathematics, quantum mechanics, biophysics, earth science, environment chemistry, … etc ), when you are young, then your future in computer science will also be more fruitful for applications.