LARGE-SCALE FIRST-PRINCIPLES ELECTRONIC STRUCTURE CALCULATIONS IN PETASCALE AND EXASCALE SUPERCOMPUTERS: A REAL-SPACE DENSITY FUNCTIONAL THEORY CODE

Graduate School of Engineering, The University of Tokyo  J.-I. Iwata
Graduate School of Engineering, The University of Tokyo  A. Oshiyama

J.-I. Iwata
R&D center of Computational Materials Science, Advance Soft Corp.  A. Oshiyama
Institute of Materials and Systems for Sustainability, Nagoya University

https://github.com/j-iwata/RSDFT
http://rsdft.jp
FIRST-PRINCIPLES SIMULATIONS IN REAL NANO WORLD

Large-scale DFT calculations and experiments meet together in Nano World!

Realistic modeling → large number of atoms many times, a long time simulation → shorter computation time
FROM K COMPUTER TO POST-K COMPUTER
( FROM PFLOPS TO EXAFLOPS )
( FROM MULTI-CORE TO MANY-CORE )

K computer
SPARC64 VIIIfx
8 cores/CPU ( 128 GFLOPS )
88,964 nodes ( > 10 PFLOPS )

FX100
SPARC64 XIfx
32 + 2 cores/CPU ( >1 TFLOPS )
( ~10 PFLOPS )

Oakforest-PACS
Intel Xeon Phi 7250 (Knights Landing)
68 cores/CPU ( 3.0464 TFLOPS )
1 CPU/node
8208 nodes ( 25 PFLOPS )
**FROM K COMPUTER TO POST-K COMPUTER**

( FROM PFLOPS TO EXAFLOPS )

( FROM MULTI-CORE TO MANY-CORE )

---

**K computer**

SPARC64 VIII Ifx

8 cores/CPU (128 GFLOPS)

**FX100**

SPARC64 XIfx

32 + 2 cores/CPU (>1 TFLOPS)

**K computer**

SPARC64 VIIIfx

8 cores/CPU (128 GFLOPS)

88,964 nodes (> 10 PFLOPS)

**post-K computer**

( ~EXAFLOPS )

- The next flagship machine in Japan is being developed and will appear around 2021

- Conducted by RIKEN R-CCS and FUJITSU

- ARM processor (instruction set architecture)

---

1 CPU/node
8208 nodes (25 PFLOPS)
KOHN-SHAM EQUATION and PLANE-WAVE EXPANSION METHOD

Kohn-Sham equation

\[
\left( -\frac{1}{2} \nabla^2 + v_{ion} + \int dr' \frac{\rho(r')}{|r-r'|} + \frac{\delta E_{XC}[\rho]}{\delta \rho(r)} \right) \phi_j(r) = \varepsilon_j \phi_j(r)
\]

Self-consistent iteration

\[
\rho(r) = \sum_j |\phi_j(r)|^2
\]

Electron density

Plane Waves - a standard (and traditional) basis set for “FP” calculations -

\[
\chi_n (r) = e^{i(k+G_n) \cdot r} \quad \phi_j (r) = \sum_n c_{jn} \chi_n (r)
\]

Frequent call of FFT (a communication burden) is inevitable!

Other drawbacks of the plane-wave basis

• Periodic Boundary Condition
  → super-cell models must be used for 0-D, 1-D, and 2-D systems
• Delocalization in real space
  → incompatible with the linear-scaling computation methods
CONTENTS

- K computer and post-K computer
- Traditional first-principles calculation method (Plane-wave method)
- RSDFT
  - Kohn-Sham equation, Eigenvalue problems
  - \(O(N^3)\)
  - Real-Space, Finite-Difference, Pseudopotential
  - FFT free
- Benchmark test at K computer
- Results with K computer
  - Twisted bilayer graphene
  - Nanofacet on the SiC vicinal surface
  - Multi-atom vacancies in SiC
  - Silicene on Ag(111) substrate
- Real-Space Car-Parrinello Molecular Dynamics (RS-CPMD)
- Summary
A REAL-SPACE GRID DFT CODE
REAL-SPACE FINITE-DIFFERENCE PSEUDOPOTENTIAL METHOD

\[
\left\{ \begin{array}{l}
\left( -\frac{1}{2} \nabla^2 + \hat{v}_{\text{ion}} + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta E_{XC}[\rho]}{\delta \rho(\mathbf{r})} \right) \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}) \\
\rho(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2 \quad (N << \# \text{ of grid points } M)
\end{array} \right.
\]

Kohn-Sham equation is solved as a finite-difference equation in discretized space

- **Derivatives** → (higher-order) finite difference
  \[
  \frac{\partial^2}{\partial x^2} \phi(\mathbf{r}) \approx \sum_{m=-6}^{6} C_m \phi(x + m\Delta, y, z)
  \]

- **Integrals** → summation over grid points
  \[
  \int \phi^*_{m}(\mathbf{r}) \phi_{n}(\mathbf{r}) d\mathbf{r} \approx \sum_{i=1}^{M} \phi^*_{m}(\mathbf{r}_i) \phi_{n}(\mathbf{r}_i) \Delta V
  \]

- **Ionic potentials** → **Pseudopotentials**
  \[
  \hat{v}_{\text{ion}} \phi(\mathbf{r}) = v_{\text{ion}}^{\text{local}}(\mathbf{r}) \phi(\mathbf{r}) + \int d\mathbf{r}' v_{\text{ion}}^{\text{nonlocal}}(\mathbf{r}, \mathbf{r}') \phi(\mathbf{r}')
  \]

Smaller
Valence only

HISTORY OF RSDFT

Three-dimensional time-dependent Hartree-Fock calculation: Application to $^{16}\text{O} + ^{16}\text{O}$ collisions
H. Flocard, S. E. Koonin and M. S. Weiss

Higher-order finite-difference pseudopotential method: An application to diatomic molecules
J. R. Chelikowsky, N. Troullier, K. Wu, and Y. Saad

Time-dependent local-density approximation in real time
K. Yabana and G. F. Bertsch

Real-space, real-time method for the dielectric function

http://github.com/j-iwata/RSDFT


AWARD
ACM Gordon Bell Prize (2011)
(CG) Conjugate-Gradient method

\[ H = \left( -\frac{1}{2} \nabla^2 + \hat{V}_{ion}^P + \int dr' \frac{\rho(r')}{|r-r'|} + \frac{\delta E_{xc}[\rho]}{\delta \rho(r)} \right) \]

\[ \mathcal{E}_n = \frac{\langle \phi_n | H | \phi_n \rangle}{\langle \phi_n | \phi_n \rangle} \]

minimize

( GS ) Ortho-normalization by Gram-Schmidt method

\( O(MN) \)

( SD ) Subspace Diagonalization (Ritz procedure)

Diagonalize \( H \) within a small \((N \times N)\) dimensional space

spanned by \( \{ \phi_1(r), \phi_2(r), \ldots, \phi_N(r) \} \)

\[ H_{mn} = \langle \phi_m | H | \phi_n \rangle \]

\( O(MN^2) \)

( dgemm/zgemm )

\( O(MN^2) \)

( dgemm/zgemm )

\( O(N^3) \)

( ScaLAPACK/Eigen )
Almost all $O(N^3)$ computations can be performed with LEVEL 3 BLAS.

### Gram-Schmidt procedure

- $\psi_1 = \varphi_1$
- $\psi_2 = \varphi_2 - \psi_1 \langle \psi_1 | \varphi_2 \rangle$
- $\psi_3 = \varphi_3 - \psi_1 \langle \psi_1 | \varphi_3 \rangle - \psi_2 \langle \psi_2 | \varphi_3 \rangle$
- $\psi_4 = \varphi_4 - \psi_1 \langle \psi_1 | \varphi_4 \rangle - \psi_2 \langle \psi_2 | \varphi_4 \rangle - \psi_3 \langle \psi_3 | \varphi_4 \rangle$
- $\psi_5 = \varphi_5 - \psi_1 \langle \psi_1 | \varphi_5 \rangle - \psi_2 \langle \psi_2 | \varphi_5 \rangle - \psi_3 \langle \psi_3 | \varphi_5 \rangle - \psi_4 \langle \psi_4 | \varphi_5 \rangle$
- $\psi_6 = \varphi_6 - \psi_1 \langle \psi_1 | \varphi_6 \rangle - \psi_2 \langle \psi_2 | \varphi_6 \rangle - \psi_3 \langle \psi_3 | \varphi_6 \rangle - \psi_4 \langle \psi_4 | \varphi_6 \rangle - \psi_5 \langle \psi_5 | \varphi_6 \rangle$
- $\psi_7 = \varphi_7 - \psi_1 \langle \psi_1 | \varphi_7 \rangle - \psi_2 \langle \psi_2 | \varphi_7 \rangle - \psi_3 \langle \psi_3 | \varphi_7 \rangle - \psi_4 \langle \psi_4 | \varphi_7 \rangle - \psi_5 \langle \psi_5 | \varphi_7 \rangle - \psi_6 \langle \psi_6 | \varphi_7 \rangle$
- $\psi_8 = \varphi_8 - \psi_1 \langle \psi_1 | \varphi_8 \rangle - \psi_2 \langle \psi_2 | \varphi_8 \rangle - \psi_3 \langle \psi_3 | \varphi_8 \rangle - \psi_4 \langle \psi_4 | \varphi_8 \rangle - \psi_5 \langle \psi_5 | \varphi_8 \rangle - \psi_6 \langle \psi_6 | \varphi_8 \rangle - \psi_7 \langle \psi_7 | \varphi_8 \rangle$

### Similar algorithm is applied in the subspace diagonalization

- Matrix $\times$ Matrix product
- Recursive algorithm

### AdvanceSoft
**PARALLELIZATION IN RSDFT**

- **Real Space Grids**
- **CPU Space**

- **MPI (Message-Passing Interface) library**
  - `MPI_ISEND, MPI_IRecv` → finite-difference calc.
  - `MPI_Allreduce` → global summation

- **OpenMP**
  - Further grid parallelization (within each node, CPU, or CMG) is performed by thread parallelization

- **spin, k-point, and band index parallelization** (orbital parallelization) is also implemented

---

We were awarded Gordon Bell Prize in 2011

\[ \text{system: Si nanowire (107,292 atoms)} \]
\[ \text{grid: } 576 \times 576 \times 180 = 59,719,680 \]
\[ \text{band: } 230,400 \]

nodes: 82,944 (10.6 PFLOPS peak)
cores: 663,552
parallelization: 27,648 (grid)

Elapsed time for 1 step of SCF iteration

<table>
<thead>
<tr>
<th>.</th>
<th>SCF</th>
<th>CG (O(N^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>time (s)</td>
<td>2903</td>
<td>160</td>
</tr>
<tr>
<td>PFLOPS</td>
<td>5.48</td>
<td>0.06</td>
</tr>
<tr>
<td>peak%</td>
<td>51.67</td>
<td>0.60</td>
</tr>
</tbody>
</table>

\[ \text{Execution time (sec.)} \]

<table>
<thead>
<tr>
<th>.</th>
<th>SCF</th>
<th>SD</th>
<th>CG</th>
<th>GS</th>
</tr>
</thead>
<tbody>
<tr>
<td>38 steps for SCF convergence \rightarrow 45 hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \# \text{ of atoms } = 39,696 \]
\[ \text{Grid } = 348 \times 348 \times 288 \]
\[ \text{Orbitals } = 82,176 \]
\[ \# \text{ of CPUs } = 32 \times 22 \times 6 = 4224 \text{ (541 TFLOPS)} \]

ATOMIC & ELECTRONIC STRUCTURES OF TWISTED BILAYER GRAPHENE

TWISTED BILAYER GRAPHENE


Elapsed time for 1 step of SCF loop at K computer (841 nodes) (sec.)

<table>
<thead>
<tr>
<th>atoms</th>
<th>SCF</th>
<th>CG</th>
<th>GS</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>22,708</td>
<td>2980</td>
<td>446</td>
<td>710</td>
<td>1452</td>
</tr>
</tbody>
</table>
Fermi-velocity Reduction in tBLG

- (8,3) tBLG
  - \(\theta = 29.4^\circ\)
  - Natom = 388
  - L = 24.1 Å

- (12,11) tBLG
  - \(\theta = 2.88^\circ\)
  - Natom = 1588
  - L = 48.7 Å

- (34,33) tBLG
  - \(\theta = 0.99^\circ\)
  - Natom = 13468
  - L = 143 Å
  - d = 3.42 Å, without corrugation

Energy Level (eV)

K \rightarrow \Gamma \rightarrow K

Fermi level

Dirac cone

2 bands

\(U_F^{\text{far}}\)

\(U_F^{\text{near}}\)

\(\theta_e \sim 5^\circ\)

\(\theta_M\)

Calculation (this work)

Experiment [8]
SiC(0001) VICINAL SURFACE STRUCTURE
MAGIC ANGLE IN NANOFACETS ON SiC(0001) VICINAL SURFACE

TEM image
M. Fujii and . Tanaka

K. Sawada, J.I, A. Oshiyama

Elapsed time for 1 step of SCF loop at K computer (960 nodes)

<table>
<thead>
<tr>
<th>atoms</th>
<th>SCF</th>
<th>CG</th>
<th>GS</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>968</td>
<td>29.55</td>
<td>6.38</td>
<td>4.08</td>
<td>6.41</td>
</tr>
</tbody>
</table>
Appearance of magnetism at the step edge of the nanofacet on the SiC surface

Energy gain for hydrogen termination on Si DB is larger (∼0.2 eV)

→ Only Si DBs are expected to be terminated by hydrogens


Ferro

Antiferro (most stable 5~30 meV)
MULTI-ATOM VACANCIES IN 4H-SiC

Multi-atom vacancy in SiC

- Extended Dangling Bond Counting Model
  Prediction of the stability of multivacancies in SiC from the formation energies of monovacancies (Vc and Vs) and the energy gain of the paring relaxation.

- In addition to the well known Vs-Vc divacancy, we have newly found Vc-Vc can also be exist!

\[ V_{\text{s}}-V_{\text{c}} & V_{\text{c}}-V_{\text{c}} \ (\text{GGA}) \]

\[ V_{\text{s}}-V_{\text{c}} & V_{\text{c}}-V_{\text{c}} \ (\text{HSE}) \]

BAND-UNFOLDING ANALYSIS OF THE ELECTRONIC STRUCTURE OF Silicene ON Ag(111) SUBSTRATE

Dirac electrons are absent due to the $3 \times 3$ reconstruction of silicene and hybridization with Ag substrate.
Im \[ G_k(\varepsilon) \] \( \propto \sum_{g} |C_{NK}(g + k - K)|^2 \)

\[ \Psi_{NK}(\mathbf{r}) = \sum_{G} C_{NK}(G)e^{i(K+G)\mathbf{r}} \]

When \( k_z = 0 \)

\[ P_N(k_x, k_y) = \sum_{G_{PC}, G'_{PC}, \varepsilon_{PC}} \int dz \int dz' C_{NK}(k^x - K^x + G^x_{PC}, k^y - K^y + G^y_{PC}, z) \times C_{NK}(k^y - K^y + G^y_{PC}, k^x - K^x + G^x_{PC}, z') e^{-iG_{PC}(z-z')} \]

\[ = \sum_{G^x_{PC}, G^y_{PC}} \int dz \int dz' C_{NK}(k^x - K^x + G^x_{PC}, k^y - K^y + G^y_{PC}, z) \times C_{NK}(k^y - K^y + G^y_{PC}, k^x - K^x + G^x_{PC}, z') \sum_{G_z_{PC}} e^{-iG_{PC}(z-z')} \]

\[ \propto \sum_{G^x_{PC}, G^y_{PC}} \int dz |C_{NK}(k^x - K^x + G^x_{PC}, k^y - K^y + G^y_{PC}, z)|^2. \]

\[ P_N(k_x, k_y, z) = \sum_{G^x_{PC}, G^y_{PC}} |C_{NK}(k^x - K^x + G^x_{PC}, k^y - K^y + G^y_{PC}, z)|^2. \]
(a) $E_B = 0.0$ eV  
(b) $E_B = 0.2$ eV  
(c) $E_B = 0.4$ eV  
(d) $E_B = 0.6$ eV  
(e) $E_B = 0.8$ eV  
(f) $E_B = 1.0$ eV

HYPERBOLIC PARABOLOID AND ISO-VALUE LINES
ARPES EXPERIMENTS

Y. Feng et al., PNAS 113, 14656 (2016).


D. Tsoutsou et al., APL103, 201604 (2013).
REAL-SPACE IMPLEMENTATION OF CPMD

RS-CPMD - Another target application for post-K -
CAR-PARRINELLO MOLECULAR DYNAMICS

Car-Parrinello Lagrangian

\[ L = \mu \sum_{i=1}^{N} \int dr \left| \dot{\phi}_i(r) \right|^2 + \frac{1}{2} \sum_{I=1}^{N_{\text{atom}}} M_I \dot{R}_I^2 - E_{\text{DFT}}[\phi_i, R_I] + \sum_{i=1}^{N} \sum_{j=1}^{N} \Lambda_{ij} \left( \int dr \phi_i^*(r) \phi_j(r) - \delta_{ij} \right) \]

Equation of Motion

\[ M_I \frac{d^2 R_I(t)}{dt^2} = - \frac{\partial E_{\text{DFT}}(R_1, R_2, \ldots, R_{N_{\text{atom}}})}{\partial R_I} \]

\[ \mu \frac{d^2 \phi_i(t)}{dt^2} = - \frac{\delta E_{\text{DFT}}}{\delta \phi_i^*(t)} + \sum_{j=1}^{N} \Lambda_{ij} \phi_j(t) = -\hat{H}_{\text{KS}} \phi_i(t) + \sum_{j=1}^{N} \Lambda_{ij} \phi_j(t) \]

Fictitious dynamics of wave functions are introduced instead of the self-consistent calculation at each atomic configuration.

( In the early days of the first-principles calculations, the CP method was also used as an eigensolver.)
Algorithm of CPMD

\[
\tilde{V}_i(t + \Delta t) = V_i(t) + \frac{\Delta t}{2} \mathbf{F}_i(t) \\
\mathbf{R}_i(t + \Delta t) = \mathbf{R}_i(t + \Delta t) + \Delta t \tilde{V}_i(t + \Delta t) \\
\ddot{\psi}_i(t + \Delta t) = \dot{\psi}_i(t) + \frac{\Delta t}{2} \hat{H}(t) \psi_i(t) \\
\ddot{\psi}_i(t + \Delta t) = \dot{\psi}_i(t) + \Delta t \ddot{\psi}_i(t + \Delta t)
\]

Velocity Verlet method

**overall scaling** $O(N^3)$

**rotorb** $O(N^3)$ (BLAS 3)

**Calculate X**

\[
\psi_i(t + \Delta t) = \ddot{\psi}_i(t + \Delta t) + \sum_{j=1}^{N} X_{ij} \psi_j(t)
\]

$X_{ij}$ is constructed from

\[
A_{ij} = \langle \ddot{\psi}_i(t + \Delta t) | \psi_j(t) \rangle \\
B_{ij} = \langle \ddot{\psi}_i(t + \Delta t) | \ddot{\psi}_j(t + \Delta t) \rangle
\]

through some iterative calculation

**Update Hamiltonian, and calculate HF force**

\[
\tilde{V}_i(t + \Delta t) = \tilde{V}_i(t + \Delta t) + \frac{\Delta t}{2} \mathbf{F}_i(t + \Delta t) \\
\ddot{\psi}_i(t + \Delta t) = \dot{\psi}_i(t + \Delta t) + \frac{\Delta t}{2} \hat{H}(t + \Delta t) \psi_i(t + \Delta t)
\]

Hamiltonian operation $O(N^2)$

**Calculate Y**

\[
\dot{\psi}_i(t + \Delta t) = \ddot{\psi}_i(t + \Delta t) + \sum_{j=1}^{N} Y_{ij} \psi_j(t + \Delta t)
\]

$Y_{ij}$ is constructed from

\[
C_{ij} = \langle \psi_i(t + \Delta t) | \ddot{\psi}_j(t + \Delta t) \rangle
\]

**rotorb2** $O(N^3)$ (BLAS 3)
## Strong scaling @K computer

<table>
<thead>
<tr>
<th>RS-CPMD</th>
<th>1 MD step</th>
<th>rotorb</th>
<th>Potential &amp; Force construction</th>
<th>Hamiltonian operation</th>
<th>rotorb2</th>
</tr>
</thead>
<tbody>
<tr>
<td>64 MPI × 8 OMP (512 cores)</td>
<td>38.98 (54.3 %)</td>
<td>20.46 (61.4 %)</td>
<td>2.28 (8.79 %)</td>
<td>2.23 (2.43 %)</td>
<td>14.03 (68.5 %)</td>
</tr>
<tr>
<td>512 MPI × 8 OMP (4096 cores)</td>
<td>5.54 (37.5 %)</td>
<td>2.69</td>
<td>0.52</td>
<td>0.44</td>
<td>1.83</td>
</tr>
<tr>
<td>1024 MPI × 8 OMP (8192 cores)</td>
<td>3.23</td>
<td>1.50</td>
<td>0.42</td>
<td>0.28</td>
<td>0.99</td>
</tr>
<tr>
<td>2048 MPI × 8 OMP (16,384 cores)</td>
<td>2.18</td>
<td>0.93</td>
<td>0.43</td>
<td>0.20</td>
<td>0.59</td>
</tr>
</tbody>
</table>

### Graph

- **Parallel Efficiency**

![Graph showing Parallel Efficiency vs. MPI processes](image)

- **RS-CPMD**

- **rotorb**

- **rotorb2**

- **total**

- **Hop**

- **H&F**

**propylene carbonate electrolyte (128 molecules, 1664 atoms)**
SUMMARY

- **RSDFT**
  - A first-principles electronic structure calculation code based on the real-space finite-difference pseudopotential method
  - FFT free
  - Suitable for massively-parallel computations (cf. traditional PW methods)

- **K computer (10 PFLOPS / 88,000 nodes)**
  - LDA, GGA
    - several thousand atoms with atomic-structure optimization ( ~ 4000 atoms )
    - a few tens of thousand atoms for single-point calc. ( ~ 40,000 atoms )
  - Hybrid-XC
    - several hundreds to one thousand atoms with opt. ( 100 ~ 1000 atoms )

- **post-K computer (~1 EFLOPS)**
  - tens of thousands of atoms with structure optimization
  - a large-scale calculation that is feasible only with the whole resource of the K computer is expected to be feasible with a small resources of the post-K computer

Thank you for your attention!