Large-scale electronic-structure calculations for nanomaterials in density functional theory

Atsushi Oshiyama
Department of Applied Physics, The University of Tokyo
Hongo, Tokyo 113-8656, Japan

Jun-ichi Iwata
Center for Computational Sciences, University of Tsukuba
Tennodai, Tsukuba 305-8577, Japan

E-mail: oshiyama@ap.t.u-tokyo.ac.jp

Abstract. We present our efforts to develop a real-space scheme in calculating total energies and electron states within the local density approximation and the generalized gradient approximation in the density functional theory, with an intention to perform large-scale computations on massively parallel many-core architecture next-generation computers in order to reveal and predict properties of nanometer scale materials. Our real-space scheme is applied to Si nanodots containing more than 10,000 atoms. The calculated ionization energy, the electron affinity and the energy gap reveal the origin of the underestimate of the energy gap in the local approximation.

1. Introduction
Computational condensed matter science has been initiated in 80s of the previous century and contributed tremendously to the progress of physics and chemistry of materials. Density functional theory (DFT) has been a powerful tool to reveal atom-scale mechanisms for a variety of phenomena in real materials. Yet target systems which DFT has treated in the past are relatively small: calculations for 100-atom systems are typical and those for 1000-atom systems are rare at present. On the other hand, the nanometer-scale system contains more than 10,000 atoms. There the nano-scale shape in the systems is now recognized as an important factor which is decisive in physical properties [1]. Hence it is highly demanded to perform DFT calculations with the levels of local density approximation (LDA) or the generalized gradient approximation (GGA) for 10,000-atom systems to clarify physics and chemistry of nano-structures. In the standard scheme using the plane-wave basis set, Fast Fourier Transform (FFT) is frequently used to compute the necessary quantities in either real space or in \( k \) space. On next-generation massively parallel multi-core architecture computers, however, FFT causes a heavy burden in communication tasks so that the large-scale calculations may be formidable. Our choice to overcome this difficulty is to compute the necessary quantities in the real space. This makes the scheme FFT-free and the efficiency of the computation is expected to be high.

Accuracy is another important factor to make simulations reliable. Both LDA and GGA are known to underestimate energy gaps of materials substantially although they predict relatively accurate structural properties. Improvement of the approximation to the exchange-correlation energy is therefore necessary.
In this article, we present our real space DFT (RSDFT) scheme which allows us to perform LDA/GGA calculations for more than 10,000-atom systems [2]. The calculated results for Si nanodots are presented to elucidate the reason for the underestimate of the energy gap.

2. RSDFT Scheme

To numerically minimize the energy functional in DFT, we introduce a 3-dimensional spatial grid and consider the minimization problem within the discrete space of \( M_L \) grid points. Then the orbitals, the density, and the potentials are expressed as the corresponding vectors whose elements are the value at each grid point. For example, the Kohn-Sham orbital \( \phi(r) \) and the charge density \( \rho(r) \) are expressed as vectors,

\[
\vec{\phi} = \begin{pmatrix} \phi^1 \\ \vdots \\ \phi^i \\ \vdots \\ \phi^{M_L} \end{pmatrix}, \quad \text{and} \quad \vec{\rho} = \begin{pmatrix} \rho^1 \\ \vdots \\ \rho^i \\ \vdots \\ \rho^{M_L} \end{pmatrix},
\]

respectively, where the \( i \)-th elements are the values at the grid point \( r_i \): \( \phi^i = \phi(r_i) \) and \( \rho^i = \rho(r_i) \). In this discrete scheme, the energy functional can be written as

\[
E[\vec{\phi}_1, \vec{\phi}_2, \ldots, \vec{\phi}_{M_L}] = -\frac{1}{2} \sum_{n=1}^{M_n} f_n \sum_{i,j=1}^{M_L} \phi^n_i \phi^n_j L_{ij} \Delta \Omega - 2\pi \sum_{i,j=1}^{M_L} \rho^i L_{ij} \rho^j \Delta \Omega + E_{XC}[\rho^1, \rho^2, \ldots, \rho^{M_L}] + \sum_{i=1}^{M_L} \rho_i \Delta \Omega + \sum_{i=1}^{M_L} f_n \sum_{i,j=1}^{M_L} \phi^n_i V^{NL}_{ij} \phi_j \Delta \Omega
\]

where \( f_n, M_n \) and \( \Delta \Omega \) are the occupation number of \( n \)-th Kohn-Sham orbital, the total number of the orbitals and the volume element, respectively. The matrix \( L_{ij} \) is the Laplacian in discrete space; i.e., the finite difference (FD) operator. The formula for the higher-order FD coefficients, from which the matrix \( L_{ij} \) is constructed, is given in the reference [2]. \( V^{NL} \) is the matrix for the non-local pseudopotential. The second term in the right-hand side of Eq. (2) represents the Hartree energy. This form can be derived from that the Hartree potential \( \tilde{v}_h \) satisfies the Poisson equation (in discrete space): \( L\tilde{v}_h = -4\pi \tilde{\rho} \). The density is defined as \( \rho^i = \sum_{n=1}^{M_n} f_n |\phi^n_i|^2 \). \( E_{XC}[\rho^1, \rho^2, \ldots, \rho^{M_L}] \) is the exchange-correlation energy functional. Minimizing the energy functional (2) under the orthonormalization condition, \( \sum_{i=1}^{M_L} \phi^n_i \phi^n_m \Delta \Omega = \delta_{nm} \), we reach the eigenvalue equation,

\[
\left[ -\frac{1}{2} L + V + V^{NL} \right] \vec{\phi}_n = \varepsilon_n \vec{\phi}_n
\]

Where \( L \) is a sparse matrix of the FD operator, \( V^{NL} \) is the matrix of the non-local operator, and \( V \) is the diagonal matrix whose elements are the sum of the local potentials. The Helmann-Feynman forces acting on atoms are also expressed in this real-space scheme.

2.1. A technique: Gram-Schmit orthogonalization

There are a lot of important details which make our RSDFT code run excellently on parallel architecture computers. We here explain just one of such details, Gram-Schmit (GS)
orthonormalization, which is a typical order $N^3$ operations ($N$: system size) and a heavy burden in large-scale calculations. In solving (3), we use iterative techniques such as conjugate gradient and residual minimization methods. After the iterative minimization, it is usually necessary to recover the orthonormalization relations among the orbitals through the GS procedure. Since the GS procedure is widely applied in many fields in the computational sciences, several algorithms have been proposed to perform these calculations efficiently. We have developed a GS algorithm suitable for the RSDFT, and the computational cost of the GS is thereby substantially reduced.

The usual GS procedure is as follows:

\[
\phi_1 = \tilde{\phi}_1 ,
\]

\[
\phi_2 = \tilde{\phi}_2 - \phi_1 (\tilde{\phi}_1^*, \phi_1) ,
\]

\[
\phi_3 = \tilde{\phi}_3 - \phi_1 (\tilde{\phi}_1^*, \phi_2) - \phi_2 (\tilde{\phi}_2^*, \phi_2) ,
\]

\[
\phi_4 = \tilde{\phi}_4 - \phi_1 (\tilde{\phi}_1^*, \phi_3) - \phi_2 (\tilde{\phi}_2^*, \phi_3) - \phi_3 (\tilde{\phi}_3^*, \phi_4) ,
\]

\[
\phi_5 = \tilde{\phi}_5 - \phi_1 (\tilde{\phi}_1^*, \phi_4) - \phi_2 (\tilde{\phi}_2^*, \phi_4) - \phi_3 (\tilde{\phi}_3^*, \phi_4) - \phi_4 (\tilde{\phi}_4^*, \phi_4) ,
\]

\[
\phi_6 = \tilde{\phi}_6 - \phi_1 (\tilde{\phi}_1^*, \phi_5) - \phi_2 (\tilde{\phi}_2^*, \phi_5) - \phi_3 (\tilde{\phi}_3^*, \phi_5) - \phi_4 (\tilde{\phi}_4^*, \phi_5) - \phi_5 (\tilde{\phi}_5^*, \phi_6) ,
\]

This procedure consists of inner products and scalar-by-vector products. These operations could be done using linear-algebra library, e.g., level 1 of Basic Linear Algebra Subprograms (BLAS). BLAS-1 is, however, extremely inefficient when treating large dimension vectors because of high ratio of the data-load tasks to the cpu operations. However, a part of the GS operations can be grouped and regarded as matrix-by-matrix products: e.g., the last three operations in the above are written as,

\[
\begin{pmatrix}
\phi_2^* \\
\phi_3^* \\
\phi_4^*
\end{pmatrix}
- \begin{pmatrix}
\phi_1 \\
\phi_2 \\
\phi_3
\end{pmatrix}
\begin{pmatrix}
\phi_1^* \\
\phi_2^* \\
\phi_3^*
\end{pmatrix}
- \begin{pmatrix}
\phi_1 \\
\phi_2 \\
\phi_3
\end{pmatrix}
\begin{pmatrix}
\phi_4^* \\
\phi_5^* \\
\phi_6^*
\end{pmatrix}.
\]

This is the matrix-matrix product which can be performed efficiently on modern computers by employing a highly tuned linear-algebra library, e.g., level 3 of BLAS. We may tune the size of the matrix block to the computer hardware.

The present RSDFT scheme has been applied to several issues in condensed matter sciences, including microscopic identifications of vacancies in Si crystal [3] and in carbon nanotubes [4], and of dislocation cores at the mismatched interfaces [5]. On the supercomputer developed at Kobe, Japan, electron-state calculations of Si nanowires containing more than 100,000 atoms now become feasible.

2.2. Energy gap of Si nanodots

Si nanodots with the dimension of tens of nanometers have been already fabricated and used for electron devices such as memory units. Reading or writing the data on the memory units corresponds to charging or discharging electrons to the nanodots. This technologically fundamental quantity is related to the energy gap (band gap) of the dot. The energy gap is rigorously defined using the ionization energy and the electron affinity as

\[
E_g = E(N_e + 1) + E(N_e - 1) - 2E(N_e) \quad (4)
\]

where $E(N_e + 1)$, $E(N_e - 1)$ and $E(N_e)$ represent the total energies of singly negative, singly positive, and neutral charge-state systems, respectively. However, it is impossible to calculate the total energy of a
singly charged solid which is of infinite size. Then the energy gaps of solids are usually evaluated by the difference in the orbital energy between the highest occupied state and the lowest unoccupied state. This quantity, occasionally called HOMO-LUMO gap, is obviously different from $E_g$ in (4). The calculated HOMO-LUMO gaps in LDA or in GGA are known to be substantially smaller than the experimental gaps. In order to develop exchange-correlation energy functionals which are capable of describing the energy gaps, it is imperative to examine the difference between $E_g$ and the HOMO-LUMO gap.

We have examined it by performing large-scale electronic-structure calculations for large enough Si nanodots [6]. Figure 1 shows the calculated density of states (DOS) of the Si nanodots consisting of about 6000 Si atoms surrounded by 1000 H atoms. The calculated DOSs are essentially identical to each other. Furthermore they are almost same with DOS of bulk crystalline Si. Hence we are able to simulate infinite-size Si crystal by large enough Si dots. In Figure 2, we show the energy gap $E_g$ calculated by (4) and the HOMO-LUMO gap as a function of the dot size. In the infinite size limit, those quantities should converge to the bulk values. As shown in Figure 2, the energy gap decreases as the system size increases. For the largest cluster Si$_{10701}$H$_{1996}$ the energy gap $E_g$ is 1.07 eV and the HOMO-LUMO gap is 0.66 eV. The former is close to the experimental band gap 1.17 eV in Si crystal, and the latter is close to the HOMO-LUMO gap in bulk Si.

In the DFT, the following relation between $E_g$ and the Kohn-Sham levels holds:

$$E_g = \varepsilon_{N_e+1}(N_e+1) - \varepsilon_{N_e}(N_e).$$  \hfill (5)
Here $\varepsilon_{N_e} (M)$ depicts the M-th Kohn-Sham level in the $N_e$-electron system. Hence the difference between $E_g$ (squares in Figure 2) and the HOMO-LUMO gap (triangles in Figure 2) is

$$E_g - (\varepsilon_{N_e} (N_e + 1) - \varepsilon_{N_e} (N_e)) = \varepsilon_{N_e + 1} (N_e + 1) - \varepsilon_{N_e} (N_e + 1).$$  \hspace{1cm} (6)

We have analyzed the difference between $E_g$ and the HOMO-LUMO gap computed in LDA (Figure 2) and found that the difference corresponds to Coulomb energy of the additional electron distributed over the dot. Then the difference is likely to disappear in the infinite-size limit. This is consistent with the expression (6). Hence we reach a conclusion that the exact expressions (5) and (6) hold even in LDA and that $E_g$ defined in terms of the total-energy difference for the crystalline Si underestimate the experimental energy gap substantially.

Figure 3: (a) Calculated total energy of Si$_{1006}$H$_{1038}$ nanodot as a function of fractional number of electrons $N_e$ in LDA. The neutral state corresponds to $N_e=N_0=17062$. Each contribution to the total energy is also shown as a function of $N_e$: (b) the kinetic energy, (c) the ionic pseudopotential energy, (d) the exchange energy, and (e) the correlation energy. The unit of the energy is the atomic unit: i.e., 27.21 eV.
From the expression (5) combined with the fractional charge states, it is deduced that the energy gap comes from the discontinuity in the total energy at the integer value of electrons [7-9]. It is thus useful to appropriately define the total energy as a function of fractional electron number $N_e$ [10], and compute it. Figure 3 shows the calculated total energy of Si$_{4006}$H$_{1038}$ nanodot along with each contribution to the total energy. In LDA or GGA, the Kohn-Sham levels generally shift upwards with increasing the number of fractional electrons (the self interaction error). This causes concave behaviour of the total energy $E_{\text{tot}}(N_e)$ as a function of the number of the fractional electrons $N_e$ since Janak theorem $\partial E_{\text{tot}} / \partial N_e = \varepsilon(N_e)$ holds [10]. Although the concave behaviour in Figure 3(a) is not very prominent since the dot is so large that the upward shifts of the Kohn-Sham levels are small, the discontinuity of $E_{\text{tot}}(N_e)$ at $N_e=N_0$ (the neutral state) corresponds to the energy gap $\varepsilon$. As for each contribution to the total energy, the kinetic energy and the ionic pseudopotential energy clearly show the discontinuity at $N_e=N_0$. This corresponds to the energy gap between the bonding and the anti-bonding states, being referred as the chemical gap. On the other hand, the exchange energy shows faint discontinuity at $N_e=N_0$ and the correlation energy is almost continuous. This infers that many body effects are not included enough in LDA, causing the underestimate of the energy gap.

3. Summary
We have presented our RSDFT scheme which allows us to perform unprecedentedly large-scale density-functional-theory-based electronic-structure calculations. At present, calculations of electron states of Si nano-structures containing 10,000 – 100,000 atoms are feasible. The calculated energy gaps of Si nanodots using the total energy difference among charged systems have elucidated the reason for well-known underestimate of the energy gaps in the local density approximation in the density functional theory.

Acknowledgments
This work is partly supported by the Grant-in-Aid project from MEXT, Scientific Research on Innovative Areas, “Materials Design through Computics: Complex Correlation and Non-equilibrium Dynamics” (http://computics-material.jp/).

References
[1] Oshiyama A, and Okada S 2009, Roles of Shape and Space in Electronic Properties of Carbon Nanomaterials, The Oxford Handbook of Nanoscience and Technology Volumes II (Oxford University Press) pp94 – pp140.
[2] Iwata J.-I, Takahashi D, Oshiyama A, Boku B, Shiraishi K, Okada S and Yabana K 2010, J. Comp. Phys. 229 2339.
[3] Iwata J.-I, Shiraishi K, and Oshiyama A 2008, Phys. Rev. B 77, 115208.
[4] Nishikawa M, Iwata J.-I, and Oshiyama A 2010, Phys. Rev. B 81, 205542.
[5] Fujimoto Y, and Oshiyama A 2010, Phys. Rev. B 81, 205309.
[6] Calculational parameters such as grid spacing are carefully examined to assure the necessary accuracy.
[7] Perdew J. P, Parr R. G, Levy M and Balduz Jr. J. L 1982, Phys. Rev. Lett. 49, 1691.
[8] Perdew J. P, Levy M 1983, Phys. Rev. Lett. 51, 1884.
[9] Sham L. J and Schlüter M 1983, Phys. Rev. Lett. 51, 1888.
[10] Janak J. F 1978, Phys. Rev. B 18, 7165.