# Development of Divide-and-Conquer Quantum Chemical Code for Biomolecules and Nano Materials

Masato Kobayashi [1], Patchreenart Saparpakorn [2,3], and Hiromi Nakai [2,4,5]

[1] Waseda Institute for Advanced Study (WIAS), Waseda University, Japan
[2] School of Advanced Science and Engineering, Waseda University, Japan
[3] Faculty of Science, Kasetsart University, Thailand
[4] Research Institute for Science and Engineering, Waseda University, Japan
[5] CREST, Japan Science and Technology Agency, Japan

## 1 Introduction

Straightforward applications of standard quantum chemical methods to biomolecules and nano materials are essentially impossible because of the undesired computational time required for the calculations, scaling at least cubically with the system size $N$. Up to now, many types of accelerating techniques for the quantum chemical calculations, which reduce the complexity to $O(N)$, have been developed. The divide-and-conquer (DC) quantum chemical method firstly proposed by Yang and Lee [1] is a pioneering method that diminishes the computational cost for the Hartree–Fock (HF) and density functional theory (DFT) calculations. The authors and coworkers have extended the DC method to the post-HF electron correlation calculations (e.g., the second-order Møller–Plesset (MP2) theory [2]) with the help of the energy density analysis (EDA) [3]. In 2009, they released the DC quantum chemistry program as a part of GAMESS package [4]. The DC-based $O(N)$ methods are summarized in the review articles [5,6]

In this Paper, we present recent developments in the DC method and show their applications to biomolecules and nano materials.

## 2 DC-MP2 Energy and Its Gradient

The DC-MP2 method [2,5] utilizes the subsystem molecular orbitals (MOs) obtained by solving the DC-HF equation:

$$\boldsymbol{F}^{\alpha}\boldsymbol{C}_{p}^{\alpha} = \varepsilon_{p}^{\alpha}\boldsymbol{S}^{\alpha}\boldsymbol{C}_{p}^{\alpha}. \tag{1}$$

Here, the superscript $\alpha$ denotes the spatially partitioned subsystem. The subsystem MOs are constructed by the linear combination of atomic orbitals corresponding to the disjoint subsystem (referred to as the *central region*) $\alpha$ [$\boldsymbol{S}(\alpha)$] and its neighboring region called the *buffer region* [$\boldsymbol{B}(\alpha)$]. $\boldsymbol{F}^{\alpha}$ and $\boldsymbol{S}^{\alpha}$ are the subsystem Fock and overlap matrices, respectively, which are the submatrices of the entire Fock and overlap matrices $\boldsymbol{F}$ and $\boldsymbol{S}$.

The total DC-MP2 correlation energy is estimated by summing up the correlation energies corresponding to individual subsystems as follows

$$\Delta E_{\text{DC-MP2}} = \sum_{\alpha}^{\text{subsystem}} \Delta E_{\text{MP2}}^{\alpha}. \tag{2}$$

Here, the correlation energy of subsystem $\alpha$ is evaluated with subsystem MOs with the assistance of EDA as follows:

$$\Delta E_{\text{MP2}}^{\alpha} = \sum_{i,j}^{\text{occ}(\alpha)} \sum_{a,b}^{\text{vir}(\alpha)} \sum_{\mu\in\boldsymbol{S}(\alpha)} C_{\mu i}^{\alpha *} \left\langle \mu j^{\alpha} \middle| a^{\alpha} b^{\alpha} \right\rangle \left( 2\tilde{t}_{ij,ab}^{\alpha} - \tilde{t}_{ij,ba}^{\alpha} \right), \tag{3}$$

where $\{i^{\alpha}, j^{\alpha}, \ldots\}$ and $\{a^{\alpha}, b^{\alpha}, \ldots\}$ represent the occupied and virtual MOs of subsystem $\alpha$, respectively. $\tilde{t}_{ij,ab}^{\alpha}$ is the effective two-electron excitation amplitude for subsystem $\alpha$, which is given in the MP2 case by

$$\tilde{t}_{ij,ab}^{\alpha} = -\left\langle a^{\alpha} b^{\alpha} \middle| i^{\alpha} j^{\alpha} \right\rangle \Big/ \left( \varepsilon_{a}^{\alpha} + \varepsilon_{b}^{\alpha} - \varepsilon_{i}^{\alpha} - \varepsilon_{j}^{\alpha} \right). \tag{4}$$

Recently, we have derived the gradient of the DC-MP2 energy. The conventional MP2 energy gradient is expressed as

$$\frac{\partial \Delta E_{\text{MP2}}}{\partial Q} = \sum_{\mu\nu} D_{\mu\nu}^{(2)} \frac{\partial H_{\mu\nu}^{\text{core}}}{\partial Q} + \sum_{\mu\nu} W_{\mu\nu}^{(2)} \frac{\partial S_{\mu\nu}}{\partial Q} + \sum_{\mu\nu\lambda\sigma} \Gamma_{\mu\nu\lambda\sigma}^{(2)} \frac{\partial \langle \mu\lambda | \nu\sigma \rangle}{\partial Q}. \tag{5}$$

Here, $\boldsymbol{D}^{(2)}$ and $\boldsymbol{W}^{(2)}$ are standard and energy-weighted perturbed one-body density matrices, respectively, and $\boldsymbol{\Gamma}^{(2)}$ is the perturbed two-body density matrix. In the present DC-MP2 gradient method, the one-body density matrices, $\boldsymbol{D}^{(2)}$ and $\boldsymbol{W}^{(2)}$, are evaluated in the same way as the unperturbed density matrix:

$$D_{\mu\nu}^{(2)} = \sum_{\alpha} P_{\mu\nu}^{\alpha} D_{\mu\nu}^{(2),\alpha}, \tag{6}$$

where $\boldsymbol{P}$ is the partition matrix that is introduced to avoid multiple counting of electronic contribution in the buffer region. When constructing subsystem perturbed density matrices, we applied the Z-vector method [7] to solve the coupled perturbed HF (CPHF) equations of subsystems. The two-body matrix, $\boldsymbol{\Gamma}^{(2)}$, can be divided into the separable (S) and non-separable (NS) terms. The S term is expressed as the product of one-body matrices, which can be evaluated in the DC manner as Eq. (6). In the present method, we evaluated the NS term as follows:

$$\Gamma_{\mu\nu\lambda\sigma}^{\text{NS,DC}} = \sum_{\alpha} P_{\mu\nu}^{\alpha} \Gamma_{\mu\nu\lambda\sigma}^{\text{NS},\alpha}, \tag{7}$$

$$\Gamma_{\mu\nu\lambda\sigma}^{\text{NS},\alpha} = \sum_{i,j} \sum_{a,b} \tilde{t}_{ij,ab}^{\alpha} C_{\mu i}^{\alpha} C_{\nu a}^{\alpha} C_{\lambda j}^{\alpha} C_{\sigma b}^{\alpha}. \tag{8}$$

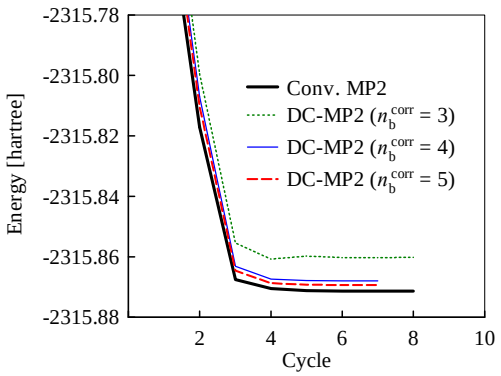


Fig. 1. Energy convergence process in the geometry optimization of $C_{60}H_{62}$ in DC and conventional MP2 calculations with the 6-31G** basis set.

Table 1. Errors of optimized geometrical parameters of $C_{60}H_{62}$ at the DC-MP2/6-31G** level. The energy errors are given at the bottom.

| | | $n_b^{corr}=3$ | $n_b^{corr}=4$ | $n_b^{corr}=5$ |
|---|---|---|---|---|
| Bond length | MAE | 0.202 | 0.022 | 0.015 |
| [pm] | MaxE | 0.522 | 0.071 | 0.059 |
| Bond angle | MAE | 1.003 | 0.062 | 0.020 |
| [°] | MaxE | 2.506 | 0.186 | 0.043 |
| Energy [m$E_h$] | | 11.142 | 3.354 | 2.023 |

We assessed the performance of the present DC-MP2 energy gradient in the geometry optimization of an all-trans polyene chain, $C_{60}H_{62}$. A $C_2H_2$ (or $C_2H_3$ for the edges) unit was adopted as the central region and the adjacent left and right $n_b$ units were adopted as the corresponding buffer region. Figure 1 shows the energy convergence process in DC and conventional MP2 calculations. Here, the buffer size for the preceding DC-HF calculation was fixed at $n_b^{HF}=6$, and that for the MP2 calculation was varied between $n_b^{corr}=3-5$. As the buffer size increases in the DC-MP2 method, the optimization process becomes closer to the conventional MP2 calculation and yields lower converged energy. Table 1 gives the mean absolute error (MAE) and the maximum error (MaxE) of the optimized geometrical parameters and the energy error of $C_{60}H_{62}$ from the conventional results. The geometrical errors also decrease as the buffer size increases. Note that the computational time for one optimization cycle by DC-MP2 method with $n_b^{corr}=5$ was around 15 times as fast as that by conventional MP2.

## 3 Application to the Interaction between HIV-1 Reverse Transcriptase and MK-4965 Inhibitor

Reverse transcriptase (RT) of HIV-1 is an essential enzyme converting the single-stranded viral RNA genome into double-stranded proviral DNA prior to its integration into the host genomic DNA. MK-4965 (Fig. 2) is a novel non-nucleoside reverse transcriptase inhibitor (NNRTI) containing diaryl ether and indazole moieties, which reveals high levels of potency against wild-type (WT) and some important HIV-1 RT mutants, that is, 0.23 and 0.39 nM in $IC_{50}$ for WT and Y181C HIV-1 RTs, respectively.

Recently, the authors applied the spin-component scaled (SCS) version of DC-MP2 to elucidate the binding of MK-4965 in the WT and Y181C HIV-1 RT binding pockets [8]. The DC-MP2 method not only reduces the computational cost but also enables to obtain the partial energy distributions by EDA, which was applied to estimate the partial interaction energies between HIV-1 RTs and MK-4965 subsystems. Figure 3 illustrates the interaction energy components of each amino acid residue in WT and Y181C HIV-1 RTs. It clearly shows the vital interaction between Lys102 and SL3 in both WT and Y181C. The second important interaction is seen between Lys103 and SL3. The more negative interaction energy about 3 kcal/mol is found in Y181C. The aforementioned distribution of interaction energies in each amino acid was clarified in detail by using the analysis of the geometry and atomic charges.

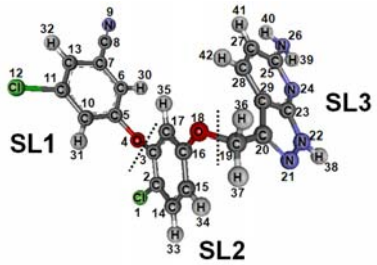


Fig. 2. Structure and subsystems of MK-4965.

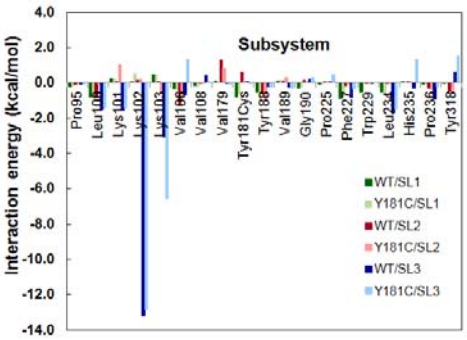


Fig. 3. Interaction energies (in kcal/mol) between SLs and each amino acid obtained by the DC-SCS-MP2/6-31G** calculation.

## 4 Concluding Remarks

After the first release of our DC program in GAMESS, we have continuously developed the DC method and its code for effectively treating biomolecules and nano materials. Especially for MP2 calculations, we have proposed the two-level parallelization method [9] that enables to treat huge systems with high accuracy by using massively parallel computers.


## Acknowledgements

Some of the present calculations were performed at the Research Center for Computational Science (RCCS), Okazaki Research Facilities, National Institutes of Natural Sciences (NINS). This study was supported in part by Grants-in-Aid for Young Scientists (B) "KAKENHI 22750016" and for Challenging Exploratory Research "KAKENHI 22655008" from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan; by the Strategic Program for Innovative Research (SPIRE), MEXT; and by the Computational Materials Science Initiative (CMSI), Japan.


## References


[1] W. Yang and T.-S. Lee, A density-matrix divide-and-conquer approach for electronic structure calculations of large molecules, J. Chem. Phys., 103, 5674, 1995.

[2] M. Kobayashi, Y. Imamura, and H. Nakai, Alternative linear-scaling methodology for the second-order Møller–Plesset perturbation calculation based on the divide-and-conquer method, J. Chem. Phys., 127, 074103, 2007.

[3] H. Nakai, Energy density analysis with Kohn–Sham orbitals, Chem. Phys. Lett., 363, 73, 2002.

[4] http://www.msg.chem.iastate.edu/gamess/

[5] M. Kobayashi and H. Nakai, Divide-and-conquer approaches to quantum chemistry: Theory and implementation, Linear-scaling techniques in computational chemistry and physics: Methods and applications, Springer, Dordrecht, 97–127, 2011.

[6] M. Kobayashi and H. Nakai, How does it become possible to treat delocalized and/or open-shell systems in fragmentation-based linear-scaling electronic structure calculations? The case of the divide-and-conquer method, Phys. Chem. Chem. Phys., *in press* (DOI: 10.1039/c2cp40153c).

[7] N.C. Handy and H.F. Schaefer, On the evaluation of analytic energy derivatives for correlated wave functions, J. Chem. Phys., 81, 5031, 1984.

[8] P. Saparpakorn, M. Kobayashi, S. Hannongbua, and H. Nakai, Divide-and-conquer based quantum chemical study for interaction between HIV-1 reverse transcriptase and MK-4965 inhibitor, Int. J. Quantum Chem., *in press*.

[9] M. Katouda, M. Kobayshi, H. Nakai, and S. Nagase, Two-level hierarchical parallelization of second-order Møller–Plesset perturbation calculations in divide-and-conquer method, J. Comput. Chem., 32, 2756, 2011.