HOKKAIDO UNIVERSITY

Title	Development of Automated Error Control Scheme Based on Divide-and-Conquer Method for Large-Scale Quantum Chemical Calculation
Author(s)	Fujimori, Toshikazu; 藤森, 俊和
Degree Grantor	北海道大学
Degree Name	博士(理学)
Dissertation Number	甲第14462号
Issue Date	2021-03-25
DOI	https://doi.org/10.14943/doctoral.k14462
Doc URL	https://hdl.handle.net/2115/84523
Type	doctoral thesis
File Information	FUJIMORI_Toshikazu.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Dissertation

Development of Automated Error Control Scheme Based on Divide-and-Conquer Method for Large-Scale Quantum Chemical Calculation

（大規模量子化学計算に対する
分割統治法に基づいた誤差自動制御スキームの開発）

Toshikazu Fujimori

Hokkaido University

Graduate School of Chemical Sciences and Engineering

Quantum Chemistry Laboratory

(北海道大学大学院総合化学院　量子化学研究室)

March 2021

Contents

1. General introduction 1

2. Theoretical backgrounds 6

2.1. Hartree-Fock theory 6

2.2. Second order Møller-Plesset perturbation theory 10

2.3. Fragment-based quantum chemical calculation method 13

2.3.1. Molecular tailoring approach 13

2.3.2. Fragment molecular orbital method 15

2.3.3. Elongation method 18

2.3.4. Divide and conquer method 21

2.4. Reference 24

3. Automatic error control in DC-SCF calculation 28

3.1. Introduction 28

3.2. DC-SCF scheme with two-layer buffer region 31

3.3. Estimation of DC-SCF energy 34

3.4. Numerical assessment 38

3.4.1. Computational details 38

3.4.2. Accuracy and computational time of the present method 39

3.5. Concluding remarks 56

3.6. Reference 58

4. Energy based automatic determination of buffer region in DC-MP2 calculation 63

4.1. Introduction 63

4.2. DC-MP2 scheme 66

4.3. Estimation of DC-MP2 energy based on AO-Laplace MP2 method 69

4.4. Numerical assessment 72

4.4.1. Computational details 72

4.4.2. Accuracy and computational time of the present method 73

4.5. Concluding remarks 89

4.6. Appendix 90

4.7. Reference 92

5. Approach to automatic error control for DC-HF energy gradient calculation 97

5.1. Introduction 97

5.2. DC-HF energy gradient expression 98

5.3. Estimation of DC-HF energy gradient 103

5.4. Numerical assessment 106

5.5. Concluding remarks 109

5.6. Reference 110

6. General conclusion 112

1. General introduction

The electronic structure of a molecule governs its all chemical properties. The electronic structure can be calculated by solving the Schrödinger equation. By virtue of the recent improvement of computer performance, *ab initio* calculations are applied to the systems that were previously only applicable to the classical mechanics calculation. However, it is still very difficult to apply the standard quantum chemical method to the large-scale systems such as proteins. In the standard quantum chemical calculation, the computational time increases drastically with respect to the system size. For example, the computational time in Hartree-Fock (HF) method and Density functional theory (DFT) increase cubically with respect to the system size, due to the diagonalization in the Self-consistent-field (SCF) calculation. In the post-HF theories, e.g. MP2 and coupled cluster method (CC), the computational time increases $O(N^5)$ or more, where N is the number of atoms in the target system.

To overcome this problem, the construction of the new theory for large systems has been actively studied since 1990s. Owing to their effort, a lot of linear or low scaling methods have been developed. For example, in the ProteinDF program, the entire system can be treated straightforwardly by means of the efficient parallelization and the cut off scheme for two-electron integrals. A method which focuses on the active cite has been developed by means of the hierarchical computational level. In the Quantum mechanics / molecular mechanics (QM/MM) method, a part of entire system (e.g. the active site) is only treated QM region and the other part is treated MM region. In the Our own *n*-layered

integrated molecular orbital and molecular mechanics (ONIOM) method, the large-scale system can be treated with 2 or 3 different computational levels.

To calculate electronic structure of the entire system, many fragment-based quantum chemical calculation methods have been developed. In these methods, the entire system is divided into several fragments and the electronic structure of the entire system is approximately calculated by combining the properties of all fragments. The divide and conquer (DC) method proposed by Yang and Lee is one of the fragment-based quantum chemical calculation methods. In the DC method, each subsystem (fragment) is composed of two regions, namely the central region and buffer region. The central region is the separated region from each other. To consider the environment around the central region, the buffer region is added for each central region. In the DC method, the size of the buffer region plays an important role of the accuracy of the DC method. The error introduced by the fragmentation can be improved systematically by increasing the size of the buffer region. In the DC-SCF calculation, the density matrix for the entire system can be approximately constructed by using those of all subsystems and the total electronic energy is calculated with the approximate entire density matrix. Nakai and coworkers extended the DC method to post-HF (MP2 or CC) energy calculations as well as the HF and MP2 energy gradient calculations. In the DC-MP2 calculation, because the buffer region in each subsystem overlaps with the other subsystem, the MP2 correlation energy of the entire system is approximated as the sum of the MP2 correlation energy corresponding to the central region of all subsystems.

Except for the DC method, many fragment-based methods have been developed and applied to wide variety of large systems. However, to the best of my knowledge, the

appropriate fragment shape that is as compact as possible for reducing the computational time while keeping the acceptable accuracy depends on the target system. In addition, in most of these methods, the fragment shape is determined by the distance parameter. Consequently, it is difficult to evaluate the error directly and the preliminary assessment is required to determine the appropriate fragment shape. In particular, the appropriate size of the buffer region in the DC method is different for different quantum chemical calculation method.

In my study, to control automatically the errors between the standard and the DC methods, I have proposed a method which determines the appropriate size of the buffer region based on the error estimation. Especially, for the DC-SCF and DC-MP2 calculations, automatic energy error control methods with the estimated energy error were proposed. In addition, an automatic energy gradient error control scheme for the DC-HF energy gradient calculation was proposed. In the DC-SCF calculation, an energy estimation scheme with a two-layered buffer region was constructed. the estimated energy in the DC-MP2 calculation was formulated with the idea of atomic orbital (AO)-Laplace MP2 method. In addition, the estimation of the DC-SCF energy gradient was proposed by means of two inequalities.

This dissertation consists of six Chapters, including this Chapter. An overview of each Chapter is given below.

Chapter 2 gives the theoretical backgrounds of this dissertation. Specifically, the HF theory and the MP2 perturbation theory are summarized briefly. In addition, several fragment-based methods are explained, namely, Molecular tailoring approach (MTA), Fragment molecular orbital (FMO)method, Elongation (ELG) method and DC method.

In Chapter 3, an automatic error control method in the DC-SCF calculation is proposed.

In this method, the atomic energy variation in each subsystem is estimated from the density matrix change between the two-layer buffer regions introduced by Dixon and Merz. Based on this estimated energy, the method to automatically determine the appropriate size of the buffer region is developed. In numerical assessment, it is confirmed that the present method works effectively for water cluster and protein. In addition, the present method achieves the linear scaling as well as the conventional DC method.

In Chapter 4, an extension of the automatic error control method to the DC-MP2 calculation is proposed. By using the idea of the AO-Laplace MP2 method proposed by Häser, the atomic DC-MP2 energy variations in each subsystem can be estimated. The appropriate size of the buffer region can be determined automatically with the estimated atomic energy contributions.

In Chapter 5, an automatic error control scheme for the DC-SCF energy gradient error calculation is proposed. From the philosophy of the automated DC-HF method, I propose the error estimation of the energy gradient with respect to the nuclear coordinate. This estimation is formulated with the DC-HF energy gradient expression proposed by Kobayashi *et al.* and two inequalities. In my research, the estimated energy gradient tends to be consistent with the actual error between the standard and DC energy gradient calculations in the Pulay term. However, it tends to be underestimated the actual error in the Hellmann-Feynman term.

In Chapter 6, the overall summary about this dissertation and the prospects of the DC method is mentioned.

2. Theoretical backgrounds

In this Chapter, I briefly summarize the Hartree-Fock theory, which is the simplest *ab initio* calculation, in Section 2.1 and the second order Møller-Plesset perturbation theory in Section 2.2. In Section 2.3, several fragment-based quantum chemical methods for treating the large-scale systems are explained.

2.1. Hartree-Fock theory

A variety of physical properties of molecular structure are dominated by the electronic structure. The electronic structure can be determined by solving the Schrödinger equation. The non-relativistic and time-independent Schrödinger equation is expressed as

$$\hat{H}\left|\Psi\right\rangle = E\left|\Psi\right\rangle. \tag{2.1}$$

The Hamiltonian consists of the kinetic energy of the nucleus and electrons and their potentials by the coulomb interaction (Eq. 2.2).

$$\hat{H} = -\sum_{l=1}^{n_e}\frac{1}{2}\nabla_l^2 - \sum_{A=1}^{N_N}\frac{1}{2M_A}\nabla_A^2 + \sum_{l<m}^{n_e}\frac{1}{r_{lm}} - \sum_{l=1}^{n_e}\sum_{A=1}^{N_N}\frac{Z_A}{r_{lA}} + \sum_{A<B}^{N_N}\frac{Z_A Z_B}{R_{AB}} \tag{2.2}$$

r_{lm}, r_{lA} and R_{AB} represent the electron-electron, the electron-nucleus and the nucleus-nucleus distances. n_e and N_N are the total numbers of electrons and nuclei, respectively. M_A and Z_A are the nuclear mass and charge, respectively. The Schrödinger equation can be solved rigorously for systems with only one electron, such as hydrogen. it is very difficult to obtain an exact solution for systems with two or more electrons such as helium,

which is called the many-body problem. The Born-Oppenheimer approximation that the motions of the nucleus and the electron are separated approximately is introduced. Under this approximation, the wave function of the nuclei is not treated explicitly and the kinetic energy of the nuclei is neglected in the calculation of its electronic structure. Therefore, Eq. (2.2) is rewritten as the electronic Hamiltonian by

$$\hat{H}_{\mathrm{elec}} = -\sum_{l=1}^{n_{\mathrm{e}}} \frac{1}{2}\nabla_l^2 + \sum_{l<m}^{n_{\mathrm{e}}} \frac{1}{r_{lm}} - \sum_{l=1}^{n_{\mathrm{e}}} \sum_{A=1}^{N_{\mathrm{N}}} \frac{Z_A}{r_{lA}}. \tag{2.3}$$

To solve the electronic Schrödinger equation feasibly, many computational methods have been developed. I briefly summarize the Hartree-Fock (HF) method,[1-3] which is the simplest *ab initio* method. In the HF method, the potential for one electron is approximated by the averaged potential from the other electrons, that is, the many-electron problem is replaced with the one-electron problem. In the HF method, the wave function of the ground state for the n_{e} electron system, Ψ^{HF}, is represented as the Slater determinant of the one-electron wave functions , shown as

$$\Psi^{\mathrm{HF}}(x_1, x_2, \cdots, x_{n_{\mathrm{e}}}) = \frac{1}{\sqrt{n_{\mathrm{e}}!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \cdots & \chi_{n_{\mathrm{e}}}(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \cdots & \chi_{n_{\mathrm{e}}}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_{n_{\mathrm{e}}}) & \chi_2(x_{n_{\mathrm{e}}}) & \cdots & \chi_{n_{\mathrm{e}}}(x_{n_{\mathrm{e}}}) \end{vmatrix}, \tag{2.4}$$

where x_n is the n th electron coordinate and $\chi_l(x)$ is the l th orthonormalized molecular orbital (MO). $\{\chi_l\}$ can be determined variationally by the HF equation, given as

$$\hat{F}\chi_l = \varepsilon_l \chi_l, \tag{2.5}$$

where $\hat{F}$ is the Fock operator shown in Eq. (2.6) and ε_l is the orbital energy of the HF method.

$$\hat{F} = \hat{h} + \sum_{m=1}^{N_e} \left(\hat{J}_m - \hat{K}_m \right) \tag{2.6}$$

In Eq. (2.6), $\hat{h}$ is the one electron operator, shown as

$$\hat{h}\chi_l(x) = \left[-\frac{1}{2}\nabla^2 - \sum_{A=1}^{N_N} \frac{Z_A}{|r - R_A|} \right] \chi_l(x), \tag{2.7}$$

and $\hat{J}_m$ and $\hat{K}_m$ are the Coulomb and exchange operators, shown as

$$\hat{J}_m \chi_l(x) = \int \chi_m^*(x') \frac{1}{|r - r'|} \chi_m(x') dx_2 \chi_l(x), \tag{2.8}$$

$$\hat{K}_m \chi_l(x) = \int \chi_m^*(x') \frac{1}{|r - r'|} \chi_l(x') dx_2 \chi_m(x). \tag{2.9}$$

$\chi_l(x)$ is composed of the spatial orbital, $\psi_l(\mathbf{r})$, and the spin function. Because the spin function is not treated explicitly, Eq. (2.5) can be expressed as the HF equation for the spatial region. In the closed shell case, Eq. (2.5) is rewritten as

$$\hat{F}\psi_l(\mathbf{r}) = \varepsilon_l \psi_l(\mathbf{r}). \tag{2.10}$$

MO can be described by the linear combination of the AOs, described as

$$\psi_l(\mathbf{r}) = \sum_{\nu=1} C_{\nu l} \phi_\nu(\mathbf{r}), \tag{2.11}$$

where $\phi_\nu(\mathbf{r})$ is the AO and $C_{\nu l}$ is the MO coefficient. By substituting Eq. (2.11) for Eq. (2.10) and integrating after multiplying $\phi_\mu^*(\mathbf{r})$ from left side, the Roothaan-Hall equation is obtained, show in

$$\sum_{\nu=1} C_{\nu l} F_{\mu\nu} = \varepsilon_l \sum_{\nu=1} C_{\nu l} S_{\mu\nu}, \tag{2.12}$$

where $F_{\mu\nu}$ and $S_{\mu\nu}$ are the Fock and the overlap matrices. These are expressed as

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + \sum_{\lambda\sigma} D_{\lambda\sigma}\left[\left(\mu\sigma \mid \nu\lambda\right) - \frac{1}{2}\left(\mu\sigma \mid \lambda\nu\right)\right] \tag{2.13}$$

$$S_{\mu\nu} = \int \phi_\mu^*(\mathbf{r})\phi_\nu(\mathbf{r})\mathrm{d}\mathbf{r} = \left\langle \phi_\mu \middle| \phi_\nu \right\rangle, \tag{2.14}$$

where $\left(\mu\nu \mid \lambda\sigma\right) = \iint \phi_\mu^*(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)\mathrm{r}_{12}^{-1}\phi_\lambda^*(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2$, **D** is the density matrix, described as

$$D_{\mu\nu} = 2\sum_{l=1}^{n_e/2} C_{\mu l}^* C_{\nu l} \ . \tag{2.15}$$

Eq. (2.12) is solved iteratively from the initial guess because this is nonlinear equation. This procedure is called Self-Consistent-Field (SCF). Eq. (2.12) is expressed as the matrix form, shown as

$$\mathbf{FC}=\boldsymbol{\varepsilon}\mathbf{SC}. \tag{2.16}$$

The total electronic energy in the HF method can be calculated with the MOs obtained from above equation:

$$E_{\text{elec}}^{\text{HF}} = \frac{1}{2}\mathrm{Tr}\left[\mathbf{D}(\mathbf{H}^{\text{core}} + \mathbf{F})\right]. \tag{2.17}$$

The total energy is calculated by the sum of the electronic energy of Eq. (2.17) and the internuclear repulsion energy with the fixed the nucleus-nucleus distances. In the HF method, the computational time for the diagonalization in the SCF procedure increases cubically with respect to system size.

2.2. Second order Møller-Plesset perturbation theory

The HF energy, E_{HF}, represents about 99% of the exact energy, E_{exact}, where E_{exact} is the exact ground state energy of the non-relativistic Schrödinger equation under the Born-Oppenheimer approximation. In the above Schrödinger equation, the electrons that are close to each other are excluded because of Coulomb holes. Consequently, the exact ground state energy is lower than the HF energy because the Coulomb interaction is reduced. The energy difference between E_{HF} and E_{exact} is called the electron correlation energy, which is given as

$$E_{corr} = E_{exact} - E_{HF}. \tag{2.18}$$

To consider electron correlation, the post-HF theories, which is based on the HF method, has been developed to account for the electron correlation. In the configuration interaction (CI) method,[1,2] the wavefunction is constructed by the linear combination of the electron configurations. Although all configurations are considered in the full-CI method, to reduce computational cost, the truncated CI methods are often used. In the coupled cluster (CC) method,[4,5] the higher order contributions can be considered by the product of the lower order configurations. The Møller-Plesset perturbation theory[1-2,4-7] is a method to treat electron correlations. Particularly, the second order Møller-Plesset method is called MP2 method, which is the simplest molecular orbital method that can deal with the electron correlation. I briefly summarize the second order Møller-Plesset perturbation theory. The exact Hamiltonian, $\hat{H}$, can be constructed $\hat{H}_0$ and $\hat{V}$;

$$\hat{H} = \hat{H}_0 + \lambda\hat{V}, \tag{2.19}$$

where $\hat{V}$ is called the perturbation term and $\hat{V} << \hat{H}_0$. λ is the perturbation parameter to only specify the corresponding order, thus λ=1. In the Møller-Plesset perturbation theory, the ground state eigenfunction and eigenvalue of $\hat{H}$ are obtained by the successive approximation with the eigenfunctions and eigenvalues of $\hat{H}_0$. Assume that $\hat{H}_0 \left|\psi_n^{(0)}\right\rangle = E_n^{(0)} \left|\psi_n^{(0)}\right\rangle$ is already solved. Because $\hat{V} << \hat{H}_0$, the ground state eigenfunction, Ψ_0, and energy, E_0, of $\hat{H}$ lie near $\psi_0^{(0)}$ and $E_0^{(0)}$ of $\hat{H}_0$ and these are expanded by the power series of λ, described as

$$\Psi_0 = \psi_0^{(0)} + \lambda \psi_0^{(1)} + \lambda^2 \psi_0^{(2)} + \cdots \tag{2.20}$$

$$E_0 = E_0^{(0)} + \lambda E_0^{(1)} + \lambda^2 E_0^{(2)} + \cdots . \tag{2.21}$$

Eq. (2.20) and Eq. (2.21) are substituted for the exact eigenvalue problem, $\hat{H}\left|\Psi_0\right\rangle = E_0 \left|\Psi_0\right\rangle$, and thus, the following equations can be obtained by summarizing with the order of λ.

$$\lambda^0 : \hat{H}_0 \left|\psi_0^{(0)}\right\rangle = E_0^{(0)} \left|\psi_0^{(0)}\right\rangle \tag{2.22}$$

$$\lambda^1 : \hat{H}_0 \left|\psi_0^{(1)}\right\rangle + \hat{V} \left|\psi_0^{(0)}\right\rangle = E_0^{(1)} \left|\psi_0^{(0)}\right\rangle + E_0^{(0)} \left|\psi_0^{(1)}\right\rangle \tag{2.23}$$

$$\begin{array}{l} \lambda^2 : \hat{H}_0 \left|\psi_0^{(2)}\right\rangle + \hat{V} \left|\psi_0^{(1)}\right\rangle = E_0^{(2)} \left|\psi_0^{(0)}\right\rangle + E_0^{(1)} \left|\psi_0^{(1)}\right\rangle + E_0^{(0)} \left|\psi_0^{(2)}\right\rangle \\ \vdots \qquad\qquad\qquad\qquad \vdots \end{array} \tag{2.24}$$

Because $\psi_0^{(0)}$ and $E_0^{(0)}$ have been already obtained, $\psi_0^{(1)}$ and $E_0^{(1)}$ can be obtained from Eq. (2.23) with $\psi_0^{(0)}$ and $E_0^{(0)}$. In addition, $\psi_0^{(2)}$ and $E_0^{(2)}$ can be also obtained from Eq. (2.24) with $\psi_0^{(0)}$, $E_0^{(0)}$, $\psi_0^{(1)}$ and $E_0^{(1)}$. Similarly, the higher order terms in Eqs. (2.20) and (2.21) can be obtained by the successive approximation. In the Møller-Plesset perturbation theory, $\hat{H}_0$ can be expressed as the sum of the HF operator;

$$\hat{H}_0 = \sum_{l=1}^{n_e} \hat{F}(l), \tag{2.25}$$

where $\hat{F}(l)$ is the Fock operator for the electron *l*. The eigenfunction and eigenvalue of $\hat{H}_0$ are the HF wave function and the sum of the orbital energy, shown in Eqs. (2.26) and (2.27).

$$\hat{H}_0 \left| \Psi^{\mathrm{HF}} \right\rangle = E_0 \left| \Psi^{\mathrm{HF}} \right\rangle \tag{2.26}$$

$$E_0 = \sum_l \varepsilon_l \tag{2.27}$$

$\hat{V}$ is given by the difference between the exact electron-electron interaction and one-electron potential in the HF method. Consequently, $\hat{V}$ is expressed as

$$\begin{aligned} \hat{V} &= \sum_{l<m} r_{lm}^{-1} - \sum_l v^{\mathrm{HF}}(l) \\ &= \sum_{l<m} r_{lm}^{-1} - \sum_l \sum_m \left[\hat{J}_m(l) - \hat{K}_m(l) \right], \end{aligned} \tag{2.28}$$

where $\hat{J}_m(l)$ and $\hat{K}_m(l)$ are the Coulomb and exchange operators, respectively. The HF energy can be considered in corrections up to the first order perturbation term. Therefore, the electron correlation energy can be corrected from the second order perturbation term. In the closed shell case, the second order Møller-Plesset perturbation energy, $E_{\mathrm{corr}}^{(2)}$, are described as

$$E_{\mathrm{corr}}^{(2)} = \sum_{i,j}^{\mathrm{occ}} \sum_{a,b}^{\mathrm{vir}} \frac{(ia \mid jb)}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \left[2(ia \mid jb) - (ib \mid ja) \right], \tag{2.29}$$

where *i*, *j* and *a*,*b* are the occupied and virtual orbitals, respectively. To improve the accuracy in the Møller-Plesset perturbation theory, the methods which increase the order of the perturbation terms have been developed, which are called MP3, MP4 and so on.

2.3. Fragment-based quantum chemical calculation method

For treating large-scale systems such as proteins, a lot of fragment-based quantum chemical calculation methods have been developed.[8-12] In the fragment-based method, there are several classifications for each fragmentation. Li *et al.* classified these methods into two categories.[13] One is the energy-based approach and the other is the density-based approach. In the energy-based approach, the total energy for the entire system can be calculated directly from the energies for all fragments. On the other hand, in the density-based approach, the total energy for the entire system can be calculated with the approximate density matrix for the entire system, which is constructed from the density matrix or molecular orbital in each fragment or local region. In this Section, I summarize about several methods in the energy-based and the density-based approaches.

2.3.1. Molecular tailoring approach

The Molecular tailoring approach (MTA) proposed by Gadre *et al* is one of the energy-based approach.[14-16] In the MTA, the entire system is divided into several initial fragments with the distance-based parameter, which is called *R*-goodness (R_g) parameter and the detail of R_g is given below. A sphere of radius R_g centered at each atom in the entire system is constructed and all atoms within the sphere belong to the initial fragment. It is noted that the sensitive regions, such as multiple bonds, aromatic rings and functional

group, are not divided but are kept intact during fragmentation. The neighboring initial fragments are merged while the number of atoms in the fragment is less than the maximum number of atoms set initially. After merging of these fragments, hydrogen atom caps are added for the broken bonds at the edge of each fragment. These dummy atoms are added at the standard bond length of X-H, where X is the connected atom.

In the MTA, the total energy of the entire system is calculated by means of the inclusion-exclusion principle in the set theory. Therefore, the intersection regions of all fragment pairs are calculated. The intersection regions are also calculated for three, four, etc. fragments. The total energy for the entire system can be expressed as

$$E = \sum_{I} E^{f_I} - \sum_{I,J} E^{f_I \cap f_J} + \cdots + (-1)^{N-1} \sum_{I,J,\cdots,N} E^{f_I \cap f_J \cap \cdots \cap f_N} , \tag{2.30}$$

where E^{f_I} is the energy for the fragment I and $E^{f_I \cap f_J}$ is the energy for the intersection region of fragment I and J, and so on. It is noted that the energy of the hydrogen atom caps is cancelled formally in this equation.

I summarize about the determination of R_g.[17,18] The R_g is introduced to mimic the chemical environment of each atom in the fragment. For a certain fragment, a sphere of centered at the atom included in this fragment is constructed. The radius of this sphere is enlarged until it touches an atom which belongs to other fragments. The radius of the enlarged sphere is called R-goodness (R_g) value for that atom. If an atom belongs to several fragments, the maximum of the sphere radii which are obtained from each fragment is taken as the R-goodness (R_g) value for that atom. This procedure is repeated for every atom in the entire system and the minimum of R_g for every atom is considered as the R_g of the scheme. The larger R_g gives the more accurate accuracy for the MTA calculations in general. From the earlier research, R_g is greater than or equal to 4 Å to be accurate enough for normal chemical structures.

2.3.2. Fragment molecular orbital method

The Fragment molecular orbital (FMO) method proposed by Kitaura and Fedorov *et al*. is also the energy-based approach.[19-21] In the FMO method, the entire system is divided into several fragments in real space. The electrons are assigned to each fragment so as not to break the bond electron pair, in other words, the bonds which exist on the border are assigned to the fragments. It is important that the nuclei are not assigned to the fragments. At the boundaries of each fragment, the atoms to which the bond electron pair is assigned are defined as the bond attached atoms (BAAs). On the other hand, the atoms on the opposite side of BAAs is called the bond detached atoms (BDAs). To describe the bond between BAA and BDA, the basis function of BDA is transformed into the hybrid orbitals with the projection operator and its hybrid orbitals is assigned to the fragments. In addition, the atomic charge for BDA is split and one side of its atomic charge is transferred to the fragment including BDA. It is noted that atomic charge is split to keep the original fragment charge.

The total energy in the original FMO method can be obtained from the energies for one fragment and the fragment pair. The Hamiltonian for the fragment I is described as

$$\hat{H}_I = \sum_i^{n_I} \left\{ -\frac{1}{2}\nabla_i^2 - \sum_A^{\text{all atoms}} \frac{Z_A}{|r_i - r_A|} + \sum_{J \neq I}^{N_F} \int dr' \frac{\rho_J(r')}{|r_i - r'|} \right\} + \sum_{i>j}^{n_I} \frac{1}{|r_i - r_j|}, \tag{2.31}$$

where n_I and N_F are the total numbers of electrons in the fragment I and total numbers of the fragments, respectively. In addition, Z_A and $\rho_J(r)$ are the atomic charge of atom A and the electron density for fragment J, respectively. In the FMO method, the Hamiltonian for one fragment includes the coulomb interaction from all nuclei in the entire system and

the electrostatic interaction from the electrons in the surrounding (N-1) fragments. These effects are the characteristic in the FMO method. The Schrödinger equation for the fragment is

$$\hat{H}_I \Psi_I = E_I \Psi_I . \tag{2.32}$$

The SCF procedure is performed for each fragment. By solving the above equation, the fragment energy for each fragment, E_I, and the electron density can be obtained. The SCF procedure for each fragment is carried out iteratively until the electron density of all fragments converges. Note that the iterative procedure for the electron density is called Self-Consistent-Charge (SCC). Next, the SCF procedure for the fragment pair is performed. The Hamiltonian for the fragment pair IJ is also described as

$$\hat{H}_{IJ} = \sum_{i}^{n_I+n_J} \left\{ -\frac{1}{2}\nabla_i^2 - \sum_{A}^{\text{all atoms}} \frac{Z_A}{|r_i - r_A|} + \sum_{K \neq I,J}^{N_F} \int dr' \frac{\rho_K(r')}{|r_i - r'|} \right\} + \sum_{i>j}^{n_I+n_J} \frac{1}{|r_i - r_j|}, \tag{2.33}$$

and the Schrödinger equation for the fragment pair is expressed as

$$\hat{H}_{IJ} \Psi_{IJ} = E_{IJ} \Psi_{IJ} . \tag{2.34}$$

The Hamiltonian for the fragment pair also includes the contributions from the surrounding (N-2) fragments. By solving Eq. (2.34), the fragment energy for each fragment pair, E_{IJ}, can be obtained. It is noted that the SCC loop is not performed in the fragment pair calculation because the Hamiltonian for the fragment pair is constructed with the converged electron density, $\rho_J(r)$. Therefore, the total energy for the entire system can be calculated with E_I and E_{IJ}, described as

$$E = \sum_{I} E_I - \sum_{I>J} \left(E_{IJ} - E_I - E_J \right). \tag{2.35}$$

The accuracy of the FMO method can be systematically improved by increasing the order of many-body expansion from the original two-body to three-body[22,23] and four-body[24] expansions.

2.3.3. Elongation method

The Elongation (ELG) method proposed by Aoki and Imamura *et. al.* is one of the density-based approach.[25-27] The ELG method has been developed to investigate the electronic structure of polymers effectively. In the fragmentation of the ELG method, an initial cluster is constructed, which is called the oligomer. It is composed of several monomers. By interacting the monomers with the oligomer one after another, the electronic structure for the entire system can be calculated approximately. In the ELG method, the canonical molecular orbital (CMO) for the oligomer, $\{\psi_i\}$, is obtained as well as the standard quantum chemical calculation, shown as

$$\psi_i = \sum_{\mu=1} C_{\mu i} \chi_\mu , \tag{2.36}$$

where χ_μ represents the μ th atomic orbital (AO). In the ELG method, $\{\psi_i\}$ is localized in two regions. The interaction region with the added monomer is called the active localized molecular orbitals (active LMOs) and the others is called the frozen localized molecular orbitals (frozen LMOs). It is noted that the frozen LMOs do not interact with the added monomer because it is far away from the added monomer. For convenience, the active LMOs is defined as region *B* and the frozen LMOs is defined as region *A*.

From here, I explain how to construct LMOs for each region. It is important that the localization procedure in the ELG method is applied to the occupied orbitals and the virtual orbitals separately in order to preserve the invariance of the density matrix. The LMO in the occupied orbitals is obtained by means of a 2×2 unitary transformation between CMOs, which is described as

$$\begin{bmatrix} \phi_i \\ \phi_j \end{bmatrix} = \begin{bmatrix} \sin\theta & \cos\theta \\ -\cos\theta & \sin\theta \end{bmatrix} \begin{bmatrix} \psi_i \\ \psi_j \end{bmatrix}. \tag{2.37}$$

By substituting Eq. (2.36) into Eq. (2.37) and dividing the AOs of the oligomer into regions A and B, the following equation is given:

$$\begin{aligned} \phi_i &= \left(\sum_{\mu=1}^{A} + \sum_{\mu=1}^{B} \right) \left(\sin\theta C_{\mu i} + \cos\theta C_{\mu j} \right) \chi_\mu, \\ &= \phi_i(A) + \phi_j(B) \end{aligned} \tag{2.38}$$

$$\begin{aligned} \phi_j &= \left(\sum_{\nu=1}^{A} + \sum_{\nu=1}^{B} \right) \left(-\cos\theta C_{\nu i} + \sin\theta C_{\nu j} \right) \chi_\nu. \\ &= \phi_j(A) + \phi_i(B) \end{aligned} \tag{2.39}$$

$\phi_i(A)$ and $\phi_i(B)$ are the LMOs which are expanded by only the sum of AOs in the region A or the region B, respectively. ϕ_i and ϕ_j can be determined with $\theta_{\max}$ that gives the maximum value of L_{ij}, shown as

$$L_{ij} = \langle \phi_i(A) | \phi_i(A) \rangle + \langle \phi_j(B) | \phi_j(B) \rangle. \tag{2.40}$$

This procedure is performed iteratively until the CMO pair does not give further localization and the same procedure is also performed for the virtual orbitals. By the above localization procedure, the occupied and virtual LMOs, $\phi_i^{\mathrm{occ}}(A)$, $\phi_i^{vir}(A)$,$\phi_j^{\mathrm{occ}}(B)$ and $\phi_j^{\mathrm{vir}}(B)$ can be obtained. Next, I explain about the MO calculation at the interaction region of the added monomers with the oligomers. As mentioned above, the added monomer does not interact with region A (frozen LMOs). Therefore, after removing the matrix in region A, the eigenvalue problem is solved for the between region B (active LMOs) and the added monomer. For the new system obtained by adding the monomer, region A and region B are newly distinguished and the molecular orbitals are localized in each region in the same manner. The above procedure for the orbital localization and eigenvalue problem is repeated until the last monomer is added.

Since the molecular orbitals in the frozen LMOs are conserved, the density matrix for the entire system can be constructed approximately with LMOs obtained in the last step and the conserved LMOs. In addition, the total energy for the entire system can be calculated with the approximate density matrix.

In the ELG method, the localization procedure to construct the active and frozen LMOs plays an important role in the accuracy. To improve the accuracy in the ELG method, Gu *et. al.* proposed the new localization procedure, where the MOs on a certain special region are localized.[28,29]

2.3.4. Divide and conquer method

The Divide and conquer method, which is also one of the density-based approach, is proposed by Yang and Li.[30-32] In the DC method, the entire system is divided into several disjointed region in the real space, which is called the central region. To consider the interaction around the central region, the buffer region is added at each central region. The fragment which is composed of the central region and the buffer region is called subsystem or localization region. In the DC method, the atomic centered basis function is used. Therefore, it is called an atomic orbital (AO) and referred to the Gleek index (μ, ν, …). A molecular orbital (MO) is expanded by the linear combination of the AOs and referred to (p, q, …). $\boldsymbol{S}(\alpha)$ is the set of AOs for the central region in subsystem α and the set of AOs for the entire system, $\boldsymbol{T}$, is expressed as

$$\boldsymbol{T} = \coprod_{\alpha} \boldsymbol{S}(\alpha) . \tag{2.41}$$

In addition, $\boldsymbol{B}(\alpha)$ is the set of AOs for the buffer region in subsystem α and the set of AOs for the localization region, $\boldsymbol{L}(\alpha)$, is expressed as

$$\boldsymbol{S}(\alpha) \coprod \boldsymbol{B}(\alpha) = \boldsymbol{L}(\alpha) . \tag{2.42}$$

In the DC-SCF calculation, the Fock matrix for the entire system is constructed with the density matrix of the entire system as well as the standard SCF calculation, shown in

$$F_{\mu\nu}[\mathbf{D}^{\mathrm{DC}}] = H_{\mu\nu}^{\mathrm{core}} + \sum_{\lambda\sigma} D_{\lambda\sigma}^{\mathrm{DC}} \left[2\left(\mu\sigma \mid \nu\lambda\right) - \left(\mu\sigma \mid \lambda\nu\right) \right], \tag{2.43}$$

where $\left(\mu\sigma \mid \nu\lambda\right) = \iint \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \phi_\mu^*(\mathbf{r}_1) \phi_\nu^*(\mathbf{r}_2) r_{12}^{-1} \phi_\sigma(\mathbf{r}_1) \phi_\lambda(\mathbf{r}_2)$ and $\left\{\phi_\mu(\mathbf{r})\right\}$ is the AO. The detail of the total density matrix, $\mathbf{D}^{\mathrm{DC}}$, is noted later. Although Eq. (2.43) is the typical Fock matrix, the effective Hamiltonian generally depends on the density matrix. The Fock

matrix for subsystem α is constructed the submatrix of the effective Hamiltonian for the entire system. The MOs for subsystem α are expanded with the linear combination of AOs, which is expressed as

$$\psi_p^\alpha(\mathbf{r}) = \sum_{\mu \in \boldsymbol{L}(\alpha)} C_{\mu p}^\alpha \phi_\mu(\mathbf{r}) . \tag{2.44}$$

The MO coefficient, $\left\{C_{\mu p}^\alpha\right\}$, and the orbital energy, $\left\{\varepsilon_p^\alpha\right\}$, are obtained from the Roothaan equation in each subsystem, which is described as

$$\mathbf{F}^\alpha[\mathbf{D}^{\mathrm{DC}}]\mathbf{C}^\alpha = \boldsymbol{\varepsilon}^\alpha \mathbf{S}^\alpha \mathbf{C}^\alpha, \tag{2.45}$$

where $\mathbf{S}^\alpha$ is the overlap matrix corresponding to the subsystem α, shown in

$$S_{\mu\nu} = \left\langle \phi_\mu \middle| \phi_\nu \right\rangle . \tag{2.46}$$

The $\mathbf{D}^{\mathrm{DC}}$ can be constructed approximately by the sum of the density matrix for subsystem, expressed as

$$D_{\mu\nu}^{\mathrm{DC}} \approx \sum_{\alpha}^{\text{subsystem}} P_{\mu\nu}^\alpha D_{\mu\nu}^\alpha . \tag{2.47}$$

In the closed shell case, the density matrix for subsystem α, $\mathbf{D}^\alpha$, is constructed with Fermi distribution function with the inverse temperature parameter, β :

$$D_{\mu\nu}^\alpha = \sum_p f_\beta\left(\varepsilon_{\mathrm{F}} - \varepsilon_p^\alpha\right) C_{\mu p}^\alpha C_{\nu p}^\alpha . \tag{2.48}$$

In the DC method, the universal common Fermi level, ε_{F}, is determined to preserve the total number of electrons by means of the following nonlinear equation;

$$n_{\mathrm{e}} = 2\mathrm{Tr}(\mathbf{D}^{\mathrm{DC}}\mathbf{S}) , \tag{2.49}$$

where n_{e} is the total number of electrons. It is noted that ε_{F} is located between the occupied and virtual orbitals in principle. The $P_{\mu\nu}^\alpha$ in Eq. (2.47) is the element of the partition matrix, $\mathbf{P}^\alpha$, which is expressed as

$$P_{\mu\nu}^{\alpha} = \begin{cases} 1 & \text{for } \mu \in \boldsymbol{S}(\alpha) \wedge \nu \in \boldsymbol{S}(\alpha) \\ 1/2 & \text{for } \left(\mu \in \boldsymbol{S}(\alpha) \wedge \nu \in \boldsymbol{B}(\alpha)\right) \text{or vice versa} \\ 0 & \text{otherwise} \end{cases}. \tag{2.50}$$

The value of the buffer region which overlaps each other can be averaged by using the $\mathbf{P}^{\alpha}$. The total density matrix and the effective Hamiltonian matrix is determined self-consistently. The total electronic energy can be calculated as the functional of $\mathbf{D}^{\mathrm{DC}}$.

$$E[\mathbf{D}^{\mathrm{DC}}] = \mathrm{Tr}[\mathbf{D}^{\mathrm{DC}}(\mathbf{H}^{\mathrm{core}} + \mathbf{F}[\mathbf{D}^{\mathrm{DC}}])]. \tag{2.51}$$

In the standard calculation, the computational cost for the diagonalization is increased cubically with respect to the system size N. On the other hand, in the DC method, it is increased cubically for the subsystem size n. In the large-scale system, the computation cost for the diagonalization can be decreased significantly because $n \ll N$ and thus the linear scaling can be achieved in the DC method.

2.4. Reference

[1] A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory;* Dover: New York, **1996**.

[2] 原田義也、量子化学　下、裳華房（2007）.

[3] 小林正人（2019）、理論化学会誌 『フロンティア』、第 1 巻、2 号、pp. 5-20.

[4] T. Helgaker, P. Jørgensen, J. Olsen, *Molecular Electronic-Structure Theory*, John Wiley & Sons, New York, **2000**.

[5] F. Jensen, *Introduction to Computational Chemistry*, *Third Edition*, John Wiley & Sons, New York, **2017**.

[6] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618.

[7] 中谷直輝（2019）、理論化学会誌 『フロンティア』、第 1 巻、2 号、pp. 21-38.

[8] S. Goedecker, *Rev. Mod. Phys.* **1999**, *71*, 1085.

[9] S. Y. Wu and C. S. Jayanthi, *Phys. Rep.* **2002**, *358*, 1.

[10] R. Zaleśny, M. G. Papadopoulos, P. G. Mezey, J. Leszczynski, *Linear-Scaling Techniques in Computational Chemistry and Physics*; Springer: Dordrecht, **2011**.

[11] D. R. Bowler and T Miyazaki, *Rep. Prog. Phys.* **2012**, *75*, 036503.

[12] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, L. V. Slipchenko, *Chem. Rev.* **2012**, *112*, 632.

[13] W. Li, S. Li, Y. Jiang, *J. Phys. Chem. A* **2007**, *111*, 2193.

[14] A. P. Rahalkar, S. D. Yeole, V. Ganesh and S. R. Gadre, In *Linear-Scaling Techniques in Computational Chemistry and Physics,* Chapter 5; R. Zaleśny, M.

G. Papadopoulos, P. G. Mezey, J. Leszczynski, Eds.; Springer: Dordrecht, **2011**; pp. 199–225.

[15] K. Babu, S. R. Garde, *J. Comput. Chem.* **2003**, *24*, 484.

[16] S. D. Yeole, S. R. Gadre, *J. Chem. Phys.* **2010**, *132*, 094102.

[17] N. Sahu, S. D. Yeole, S. R. Gadre, *J. Chem. Phys*. **2013**, *138*, 104101.

[18] V. Ganesh, R. K. Dongare, P. Balanarayan, S. R. Gadre, *J. Chem. Phys.* **2006**, *125*, 104109.

[19] D. G. Fedorov, K. Kitaura, The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems, CRC Press, Boca Raton, **2009**.

[20] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, *Chem. Phys. Lett.* **1999**, *313*, 701.

[21] T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, K. Kitaura, *Chem. Phys. Lett.* **2002**, *351*, 475.

[22] D. G. Fedorov, K. Kitaura, *J. Chem. Phys.* **2004**, *120*, 6832.

[23] Y. Nishimoto, D. G. Fedorov, *J. Comput. Chem*. **2017**, *38*, 406.

[24] T. Nakano, Y. Mochizuki, K. Yamashita, C. Watanabe, K. Fukuzawa, K. Segawa, Y. Okiyama, T. Tsukamoto, S. Tanaka, *Chem. Phys. Lett.* **2012**, *523*, 128.

[25] A. Imamura, Y. Aoki, K. Maekawa, *J. Chem. Phys.* **1991**, *95*, 5419.

[26] Y. Aoki, A. Imamura, *J. Chem. Phys.* **1992**, *97*, 8432.

[27] G. Räther, Y. Aoki, A. Imamura, *Int. J. Quantum Chem.* **1999**, *74*, 35.

[28] F. L. Gu, Y. Aoki, J. Korchowiec, A. Imamura, B. Kirtman, *J. Chem. Phys.* **2004**, *121*, 10385.

[29] Y. Aoki, F. L. Gu, *Phys. Chem. Phys. Chem.* **2012**, *14*, 7640.

[30] W. Yang, T.-S. Lee, *J. Chem. Phys.* **1995**, *103*, 5674.

[31] M. Kobayashi and H. Nakai, In *Linear-Scaling Techniques in Computational Chemistry and Physics,* Chapter 5; R. Zaleśny, M. G. Papadopoulos, P. G. Mezey, J. Leszczynski, Eds., Springer, Dordrecht, **2011**, pp. 97–127.

[32] T. Akama, M. Kobayashi, H. Nakai, *J. Comput. Chem.* **2007**, *28*, 2003.

3. Automatic error control in DC-SCF calculation

3.1. Introduction

A Since the advent of computational quantum chemistry, the rapid increase in computational power has allowed the electronic structure calculation of ever-larger systems. In variational quantum chemical methods, the major computational task is the diagonalization of the Hamiltonian matrix that scales cubically with respect to the number of basis functions. To enable the electronic structure calculation of very large systems, many types of approximate electronic structure methods have been proposed in the last two decades that show linear-scaling computational time with respect to the system size. Almost all linear-scaling methods are approximations of existing matured electronic structure methods, such as Hartree–Fock (HF),[1] Kohn–Sham density functional theory (DFT),[2] and post-HF correlation calculations.[3] The results of linear-scaling methods bear two types of errors, i.e., those derived from the methodology and those from the linear-scaling approximation itself, the latter of which is desired to be controlled by the linear-scaling method itself.

In many linear-scaling methods, procedures for distance-based control and/or accuracy evaluation have been introduced. For example, in the density matrix minimization method,[4] a cutoff distance was introduced for the construction of an auxiliary density matrix of the support function.[5,6] In the molecular tailoring approach,[7] Gadre and

coworkers defined the *R*-goodness parameter,[8,9] which indicates the quality of a fragmentation scheme based on the distance. The generalized energy-based fragmentation (GEBF) approach[10] can also employ a distance-based accuracy control scheme. [11] In the divide-and-conquer (DC) method,[12–14] the size of the buffer region controls primarily the accuracy of the approximation. Although methods with distance-based control parameters can systematically improve the accuracy of the approximations, it is still difficult to estimate the error in energy, which is the most important property in electronic structure calculations. For some linear-scaling methods, density-based or energy-based error estimation schemes have also been developed. For example, for the density matrix purification method,[15,16] Rubensson and coworkers proposed a scheme to control the density-matrix error derived from iterative purification.[17,18] Niklasson *et al.* proposed a graph-based Fermi-operator expansion scheme, in which the accuracy was controlled by thresholded sparse matrix algebra.[19] However, in fragment-based linear-scaling approaches, such as those with DC and molecular tailoring methods, it can be difficult to control the accuracy without careful prior testing.[19] Another example of accuracy control can be found in the fragment molecular orbital method,[20] in which the results can be improved by increasing the order of the many-body expansion.[21,22]

In this Chapter, a scheme to estimate the energy error introduced in DC-HF or DC-DFT calculations[23] is proposed. Nakai and coworkers extended the DC method to open-shell systems[13,24] and proposed an energy gradient.[25] Recently, they have also applied this method to the density-functional tight-binding (DFTB) theory,[26,27] which has enabled us to perform quantum mechanical calculation of one million atom systems within one minute with the Japanese K supercomputer.[28,29] In the present error estimation method, the two-layer buffer region scheme originally introduced by Dixon

and Merz[30] was adopted. Guided by this error estimation scheme, an algorithm to automatically determine the appropriate buffer size was established.

3.2. DC-SCF scheme with two-layer buffer region

Before summarizing the DC-SCF method with a two-layer buffer region, I note that, in the DC method, each basis function should be connected to an atom. Therefore, it is simply called an atomic orbital (AO) and denoted with a Greek letter index, $\{\mu, \nu, \ldots\}$. In the DC method, the entire system is first divided into N_{sub} disjoint subsystems, each of which is referred to as the central region. A set of basis functions connected to the central region of subsystem α is denoted by $\boldsymbol{S}(\alpha)$. For each subsystem, the buffer region is added to the central region to construct a localization region, where the subsystem molecular orbitals (MOs) are constructed. In the two-layer buffer scheme introduced by Dixon and Merz,[30] the buffer region is hierarchically divided into two sub-regions, denoted as the inner and outer buffer regions (Fig. 3-1). The inner buffer region, in which the set of AOs is denoted by $\boldsymbol{B}_{\mathrm{i}}(\alpha)$, is used to construct the subsystem MOs as well as to contribute to the density matrix; while the outer buffer region, in which the set of AOs is denoted by $\boldsymbol{B}_{\mathrm{o}}(\alpha)$, is only used to construct the subsystem MOs.

According to the DC-SCF scheme, the one-body density matrix of the entire system is approximated by the sum of subsystem contributions:

$$D_{\mu\nu} \approx D_{\mu\nu}^{\mathrm{DC}} = \sum_{\alpha}^{\mathrm{subsystem}} P_{\mu\nu}^{\alpha} D_{\mu\nu}^{\alpha} , \tag{3.1}$$

where $\mathbf{D}^{\alpha}$ expresses the density matrix of subsystem α, which is given in closed-shell case by

$$D_{\mu\nu}^{\alpha} = \sum_{p} f_{\beta}(\varepsilon_{\mathrm{F}} - \varepsilon_{p}^{\alpha}) C_{\mu p}^{\alpha} C_{\nu p}^{\alpha *} . \tag{3.2}$$

The subsystem MOs, $\{\psi_p^\alpha\}$, are expanded in the two-layer buffer scheme with the AOs, $\{\phi_\mu\}$, in the outer localization region, $\boldsymbol{L}_{\mathrm{o}}(\alpha) \equiv \boldsymbol{S}(\alpha) \cup \boldsymbol{B}_{\mathrm{i}}(\alpha) \cup \boldsymbol{B}_{\mathrm{o}}(\alpha)$:

$$\psi_p^\alpha(\mathbf{r}) = \sum_{\mu \in \boldsymbol{L}_{\mathrm{o}}(\alpha)} C_{\mu p}^\alpha \phi_\mu(\mathbf{r}). \tag{3.3}$$

The MO coefficients, $\{\mathbf{C}_p^\alpha\}$, and MO energies, $\{\varepsilon_p^\alpha\}$, are obtained by solving the following subsystem Roothaan equation:

$$\mathbf{F}^\alpha[\mathbf{D}^{\mathrm{DC}}]\mathbf{C}_p^\alpha = \varepsilon_p^\alpha \mathbf{S}^\alpha \mathbf{C}_p^\alpha. \tag{3.4}$$

$\mathbf{F}^\alpha[\mathbf{D}^{\mathrm{DC}}]$ and $\mathbf{S}^\alpha$ are the subsystem effective Hamiltonian and overlap matrices, respectively, which are the submatrices of the entire effective Hamiltonian and overlap matrices,

$$F_{\mu\nu}[\mathbf{D}^{\mathrm{DC}}] = H_{\mu\nu}^{\mathrm{core}} + \sum_{\lambda\sigma} D_{\lambda\sigma}^{\mathrm{DC}} \left[2\langle \mu\sigma | \nu\lambda \rangle - \langle \mu\sigma | \lambda\nu \rangle \right], \tag{3.5}$$

$$S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle, \tag{3.6}$$

for $\boldsymbol{L}_{\mathrm{o}}(\alpha)$ with two-electron integral notation of $\langle \mu\sigma | \nu\lambda \rangle = \iint d\mathbf{r}_1 d\mathbf{r}_2 \phi_\mu^*(\mathbf{r}_1)\phi_\sigma^*(\mathbf{r}_2) r_{12}^{-1} \phi_\nu(\mathbf{r}_1)\phi_\lambda(\mathbf{r}_2)$. Although the Fock matrix is shown in Eq. (3.5) as a typical example, the effective Hamiltonian generally depends on the density matrix. $\mathbf{P}^\alpha$ in Eq. (3.1) is the partition matrix, which is defined in the two-layer buffer scheme by

$$P_{\mu\nu}^\alpha = \begin{cases} 1 & \text{for } \mu \in \boldsymbol{S}(\alpha) \wedge \nu \in \boldsymbol{S}(\alpha) \\ 1/2 & \text{for } (\mu \in \boldsymbol{S}(\alpha) \wedge \nu \in \boldsymbol{B}_{\mathrm{i}}(\alpha)) \text{ or } \textit{vice versa}, \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

and $f_\beta(x) = [1 + \exp(-\beta x)]^{-1}$ is the Fermi distribution function with the inverse temperature parameter β. ε_{F} represents the universal Fermi level, which is determined

by solving the following non-linear equation to conserve the total number of electrons, n_e, in the entire system:

$$n_e = 2\mathrm{Tr}(\mathbf{D}^{\mathrm{DC}}\mathbf{S}) . \tag{3.8}$$

The density matrix of Eq. (3.1) and the effective Hamiltonian matrix of Eq. (3.5) are determined self-consistently. The electronic energy can be obtained as the functional of the density matrix:

$$E[\mathbf{D}^{\mathrm{DC}}] = \mathrm{Tr}\left[\mathbf{D}^{\mathrm{DC}}(\mathbf{H}^{\mathrm{core}} + \mathbf{F}[\mathbf{D}^{\mathrm{DC}}])\right], \tag{3.9}$$

when the effective Hamiltonian is linear with respect to the density matrix, which is satisfied in HF and semiempirical MO calculations, but is not in typical DFT calculations.

3.3. Estimation of DC-SCF energy

If the outer buffer region is transferred into the inner buffer region, the density matrix changes by

$$\Delta D_{\mu\nu} = \sum_{\alpha}\left[P'^{\alpha}_{\mu\nu} \sum_{p} f_{\beta}(\varepsilon'_{\mathrm{F}} - \varepsilon^{\alpha}_{p}) C^{\alpha}_{\mu p} C^{\alpha *}_{\nu p} - P^{\alpha}_{\mu\nu} \sum_{p} f_{\beta}(\varepsilon_{\mathrm{F}} - \varepsilon^{\alpha}_{p}) C^{\alpha}_{\mu p} C^{\alpha *}_{\nu p} \right], \tag{3.10}$$

where relaxation of the subsystem MOs is neglected. $\mathbf{P}'^{\alpha}$ is the auxiliary partition matrix

$$P'^{\alpha}_{\mu\nu} = \begin{cases} 1 & \text{for } \mu \in \boldsymbol{S}(\alpha) \wedge \nu \in \boldsymbol{S}(\alpha) \\ 1/2 & \text{for } [\mu \in \boldsymbol{S}(\alpha) \wedge \nu \in (\boldsymbol{B}_{\mathrm{i}}(\alpha) \cup \boldsymbol{B}_{\mathrm{o}}(\alpha))] \text{ or } \textit{vice versa}, \\ 0 & \text{otherwise} \end{cases} \tag{3.11}$$

and $\varepsilon'_{\mathrm{F}}$ is the auxiliary Fermi level. The first-order energy variation can be estimated with the density matrix correction, $\Delta\mathbf{D}$, as

$$\Delta E = 2\mathrm{Tr}\left[\Delta\mathbf{D}\mathbf{F}[\mathbf{D}^{\mathrm{DC}}] \right], \tag{3.12}$$

where the effective Hamiltonian is assumed to be linear with respect to the density matrix. There are two ways of obtaining the auxiliary Fermi level, $\varepsilon'_{\mathrm{F}}$. The first one is to consider $\varepsilon'_{\mathrm{F}} = \varepsilon_{\mathrm{F}}$, which simplifies Eq. (3.10) to

$$\Delta D_{\mu\nu} = \sum_{\alpha} (P'^{\alpha}_{\mu\nu} - P^{\alpha}_{\mu\nu}) \sum_{p} f_{\beta}(\varepsilon_{\mathrm{F}} - \varepsilon^{\alpha}_{p}) C^{\alpha}_{\mu p} C^{\alpha *}_{\nu p} = \sum_{\alpha} \Delta P^{\alpha}_{\mu\nu} D^{\alpha}_{\mu\nu}, \tag{3.13}$$

where

$$\Delta P^{\alpha}_{\mu\nu} = \begin{cases} 1/2 & \text{for } \left(\mu \in \boldsymbol{S}(\alpha) \wedge \nu \in \boldsymbol{B}_{\mathrm{o}}(\alpha)\right) \text{ or } \textit{vice versa} \\ 0 & \text{otherwise} \end{cases}. \tag{3.14}$$

Substituting Eq. (3.13) into Eq. (3.12) gives

$$\Delta E = 2\sum_{\mu\nu}\sum_{\alpha}\Delta P^{\alpha}_{\mu\nu}D^{\alpha}_{\mu\nu}F^{\alpha}_{\nu\mu}[\mathbf{D}^{\mathrm{DC}}] = \sum_{\alpha}\sum_{\mu\in\boldsymbol{S}(\alpha)}\sum_{\nu\in\boldsymbol{B}_{\mathrm{o}}(\alpha)} 2D^{\alpha}_{\mu\nu}F^{\alpha}_{\nu\mu}[\mathbf{D}^{\mathrm{DC}}]\,. \tag{3.15}$$

According to the energy density analysis (EDA),[31] which is analogous to the Mulliken population analysis, the variation in energy can be separated into the contributions from the atoms in the outer buffer regions:

$$\Delta E = \sum_{\alpha}\sum_{\mu\in\boldsymbol{S}(\alpha)}\sum_{A\in\boldsymbol{B}_{\mathrm{o}}(\alpha)}\sum_{\nu\in A} 2D^{\alpha}_{\mu\nu}F^{\alpha}_{\nu\mu}[\mathbf{D}^{\mathrm{DC}}] = \sum_{\alpha}\sum_{A\in\boldsymbol{B}_{\mathrm{o}}(\alpha)}\Delta E^{\alpha}_{A}\,, \tag{3.16}$$

where

$$\Delta E^{\alpha}_{A} = \sum_{\mu\in\boldsymbol{S}(\alpha)}\sum_{\nu\in A} 2D^{\alpha}_{\mu\nu}F^{\alpha}_{\nu\mu}[\mathbf{D}^{\mathrm{DC}}], \tag{3.17}$$

and index A designates an atom.

The other way to obtain the auxiliary Fermi level relies on the electron number constraint, i.e., $\varepsilon'_{\mathrm{F}}$ is found by solving the following equation:

$$n_{\mathrm{e}} = 2\mathrm{Tr}\left[\left(\mathbf{D}^{\mathrm{DC}} + \Delta\mathbf{D}\right)\mathbf{S}\right]. \tag{3.18}$$

Note that, in semiempirical MO calculations with a zero differential overlap (ZDO) approximation, the solution of Eq. (3.18) is $\varepsilon'_{\mathrm{F}} = \varepsilon_{\mathrm{F}}$ (as in the first case), since $\mathbf{S} = \mathbf{I}$ and the diagonal elements of $\Delta\mathbf{D}$ with $\varepsilon'_{\mathrm{F}} = \varepsilon_{\mathrm{F}}$ [i.e., Eq. (3.13)] are zero.

If one chooses $\varepsilon'_{\mathrm{F}} = \varepsilon_{\mathrm{F}}$ for DC calculations with a two-layer buffer region, the energy error introduced by the DC method can be estimated as the sum of contributions from the outer buffer atoms in each subsystem according to Eq. (3.16). Also, it is known that the density matrix $\rho(\mathbf{r}_1,\mathbf{r}_2) = D_{\mu\nu}\phi_{\mu}(\mathbf{r}_1)\phi_{\nu}(\mathbf{r}_2)$ decays exponentially with the distance $|\mathbf{r}_1 - \mathbf{r}_2|$ in the case of an insulator.[32] Based on these facts, the following automatic extension scheme for the buffer region was developed:

i). Evaluation of ΔE_A^α according to Eq. (3.16) after constructing $\mathbf{F}[\mathbf{D}^{\mathrm{DC}}]$ for each SCF cycle.

ii). Transferring all atoms in the outer buffer region of subsystem α to its inner buffer region.

iii). Inclusion of the atoms in the sphere with radius r_{ext} centered on atom A with $\Delta E_A^\alpha \geq e_{\mathrm{thresh}}$ into the new outer buffer region of subsystem α.

iv). Calculation of the subsystem MOs with Eq. (3.4), construction of the density matrices with Eqs. (3.1) and (3.13), and back to step i).

The above procedure is illustrated in Fig. 3-2. After several cycles, the outer buffer region automatically vanishes when all ΔE_A^α become less than the threshold. Following this scheme, it may become possible to choose the appropriate buffer region for each subsystem while preserving the energy error per atom. In the actual implementation, the subsystem density matrix element required in Eq. (3.16) is approximated as $D_{\mu\nu}^\alpha \sim \Delta D_{\mu\nu}$ to avoid the need for storing the density matrices of all subsystems. This approximation can be validated because $\Delta D_{\mu\nu} = \Delta P_{\mu\nu}^{\alpha_1} D_{\mu\nu}^{\alpha_1} + \Delta P_{\mu\nu}^{\alpha_2} D_{\mu\nu}^{\alpha_2} = 1/2\left(D_{\mu\nu}^{\alpha_1} + D_{\mu\nu}^{\alpha_2}\right) \sim D_{\mu\nu}^{\alpha_1}$ for $\mu \in \boldsymbol{S}(\alpha_1) \wedge \mu \in \boldsymbol{B}_{\mathrm{o}}(\alpha_2)$ and $\nu \in \boldsymbol{S}(\alpha_2) \wedge \nu \in \boldsymbol{B}_{\mathrm{o}}(\alpha_1)$, where $D_{\mu\nu}^{\alpha_2}$ is considered to be

Fig. 3-2. Illustration of the automatic extension scheme for the buffer region.

similar to $D_{\mu\nu}^{\alpha_1}$. Here, $\mu \in \boldsymbol{B}_{\mathrm{o}}(\alpha_2)$ is not always the case. Therefore, $D_{\mu\nu}^{\alpha} \sim 2\Delta D_{\mu\nu}$ is the other choice of the approximation, while it is equivalent to halve e_{thresh}.

3.4. Numerical assessment

3.4.1. Computational details

The automatically controlled DC method was implemented to the GAMESS package[33,34] and assessed its accuracy and efficiency for different types of systems. In the DC method, the inverse temperature parameter, β, in Eq. (3.2) was set to 200 a.u. The parameters for the automated DC method were set to e_{thresh} = 0.1 μE_{h} and r_{ext} = 3.0 Å unless otherwise noted.

To discuss quantitatively the size of the localization region determined in the present scheme, I defined the major axis radius of localization region α, $l_{\mathrm{local}}(\alpha)$, as half of the maximum atom pair distance in localization region α. The major axis radius at the initial SCF step, $l_{\mathrm{local}}^{\mathrm{ini}}$, where the outer buffer region is excluded from the localization region, should strongly correlate with the initial buffer size, while that at the final SCF step, $l_{\mathrm{local}}^{\mathrm{fin}}$, is expected to be barely dependent on the initial buffer size.

3.4.2. Accuracy and computational time of the present method

I first compared two estimation schemes of the DC-HF energy error with Eq. (3.12): (a) where $\varepsilon'_{\mathrm{F}} = \varepsilon_{\mathrm{F}}$ and (b) where $\varepsilon'_{\mathrm{F}}$ was determined for $(\mathbf{D}^{\mathrm{DC}} + \Delta\mathbf{D})$. The estimated and actual energy errors were obtained for calculations of the crambin protein, as summarized in Table 3-1. Here, the 6-31G basis set[35] was adopted. The geometry of crambin was obtained via the protein data bank (PDB, identification number 1CRN) and the hydrogen atoms were then added with the FU program.[36] The estimated energy errors obtained from the second and final SCF steps are given for both estimation schemes. The initial guess density, which affects the estimation at the second SCF step, was obtained by the DC extended Hückel method implemented in GAMESS. In the DC calculations, the entire protein was cut between the carbonyl C and the α-C, and each fragment was treated as a central region. The buffer size was defined by $r_{\mathrm{b}}^{\mathrm{in}}$ and $r_{\mathrm{b}}^{\mathrm{out}}$, where the unions of the spherical regions with radius $r_{\mathrm{b}}^{\mathrm{in}}$ and $r_{\mathrm{b}}^{\mathrm{out}}$ centered on each atom in the central region were considered as the inner and outer localization regions, respectively. As expected, the actual energy error decreased with the increasing buffer size, except for the smallest buffer size where an error cancellation seems to have occurred. The two estimation schemes did not display significant differences. At both the second and final SCF steps, the difference in the errors estimated by the two schemes was less than 10% for $r_{\mathrm{b}}^{\mathrm{in}} \geq 4.0$ Å. The order of the estimated energy error at the final SCF step was consistent with that of the actual error. This estimation scheme worked reasonably even at the early SCF step, although the estimated error at the second SCF step was two or

more times larger than that at the final step for $r_b^{in} \geq 4.0$ Å. The method was also tested in calculations of delocalized polyene system and the similar results were obtained (see Table 3-2). From the following Section on, I will mainly focus on the semiempirical PM3 method,[37,38] which adopts the ZDO approximation.

Table 3-1. Buffer size dependence of the actual and estimated DC-HF energy errors for crambin protein. Standard HF energy is –17,996.926754 E_h.

			$-\Delta E$ by scheme (a) /E_h		$-\Delta E$ by scheme (b) /E_h	
r_b^{in} /Å	r_b^{out} /Å	Actual error /E_h	2nd step	Final step	2nd step	Final step
3.5	4.5	–0.144241	–0.768550	–0.886510	–0.682481	–0.890704
4.0	5.0	–0.348118	–1.005190	–0.532129	–1.071929	–0.526987
4.5	5.5	–0.067862	–0.504153	–0.126115	–0.507108	–0.125820
5.0	6.0	–0.017408	–0.123892	–0.038836	–0.124559	–0.038845
5.5	6.5	–0.005229	–0.084293	–0.016880	–0.084118	–0.016881

Table 3-2. Buffer size dependence of the actual and estimated DC-HF energy errors for the following polyene derivative. The cc-pVDZ basis set was adopted. Standard HF energy is –2314.676893 E_h.

			$-\Delta E$ by scheme (a) /E_h		$-\Delta E$ by scheme (b) /E_h	
r_b^{in} /Å	r_b^{out} /Å	Actual error /E_h	2nd step	Final step	2nd step	Final step
4.5	7.0	–5.095390	–0.896898	–10.330447	–0.896490	–10.330270
7.0	9.5	–0.137146	–0.177043	–0.293687	–0.177043	–0.293687
9.5	12.0	–0.021289	–0.012954	–0.047777	–0.012954	–0.047777
12.0	14.5	–0.001921	–0.004375	–0.005201	–0.004375	–0.005201
14.5	17.0	–0.000091	–0.000386	–0.000648	–0.000386	–0.000648

The accuracy of the present method and its computational time requirements were examined in the calculation of a cubic system containing N_{water} randomly oriented water molecules with a weight density of 1.0 g cm^{-3}. In the DC calculations, each water molecule was treated as the central region. The initial buffer size was determined by r_b^{in} and r_b^{out}, the definitions of which are the same as in the previous Section. Table 3-3 summarizes the initial buffer-size dependence of the automated DC-PM3 energy, the wall-clock computational time, and the number of SCF cycles for N_{water} = 1000. The

computational time for the SCF calculations was measured using a computer node equipped with an Intel Xeon E5-1650 CPU (6 cores, 3.50 GHz), and the average of three measurements was calculated. The energy difference from the standard PM3 results divided by the number of atoms (3000) is also shown in parentheses. For $r_b^{in} \leq 6.0$ Å, the energy difference values are comparable: ~0.5 μE_h atom^{-1}. For $r_b^{in} \geq 6.5$ Å, the

Table 3-3. Initial buffer-size dependence of the total energy, the wall-clock computational time, and the number of SCF cycles for the automated DC-PM3 calculation of the model system containing 1000 water molecules.

r_b^{in} /Å	r_b^{out} /Å	Energy /E_h	(Diff. /μE_h atom^{-1})	Time /s	# cycles
3.5	4.5	–11945.190938	(+0.48)	250	14
4.0	5.0	–11945.190942	(+0.48)	246	14
4.5	5.5	–11945.190837	(+0.51)	233	13
5.0	6.0	–11945.190719	(+0.55)	209	12
5.5	6.5	–11945.190414	(+0.65)	209	13
6.0	7.0	–11945.190229	(+0.72)	202	12
6.5	7.5	–11945.191077	(+0.43)	246	12
7.0	8.0	–11945.191791	(+0.19)	325	12
Standard-PM3		–11945.192376		2443	11

energy difference gradually decreases to zero because the estimation for the initial buffer size is smaller than the threshold for most of the subsystems. In fact, the energy error at a single fixed buffer size of $r_b = 7.5$ Å is 0.50 μE_h atom^{-1}, in good agreement with the result for $r_b^{out} = 7.5$ Å. Although the number of SCF cycles is slightly larger than that for standard PM3 calculations, the computational time is ~10 times shorter for $r_b^{in} \leq 6.5$ Å. It is also suggested that a smaller initial buffer size results in the deterioration of the SCF convergence, which in turn leads to longer computational times.

Table 3-4 summarizes the average ($<l_{local}>$) and standard deviation ($\sigma[l_{local}]$) of the major axis radii among all localization regions in the automated DC-PM3 calculations of the water system (N_{water} = 1000). As expected, $<l_{local}^{ini}>$ increased linearly with the initial buffer size, and $\sigma[l_{local}^{ini}]$ was found to be relatively small. Interestingly, $<l_{local}^{fin}>$ was found to be larger for small initial buffer sizes up to $r_b^{in} = 6.0$ Å, although the difference was fairly small. Accordingly, $\sigma[l_{local}^{fin}]$ displayed smaller values for larger initial buffer sizes. It was thus suggested that large initial buffer sizes efficiently aid the selection of the appropriate buffer region and hence may reduce the computational time, although this effect does not largely affect the energy error. For readers with particular interest, the behavior of $<l_{local}>$ during the SCF iteration is given in Fig. 3-3.

Table 3-4. Average and standard deviation of the major axis radii of all localization regions at the initial and final SCF steps in the automated DC-PM3 calculation of the model system containing 1000 water molecules.

r_b^{in} /Å	r_b^{out} /Å	$<l_{local}^{ini}>$ /Å	$\sigma[l_{local}^{ini}]$ /Å	$<l_{local}^{fin}>$ /Å	$\sigma[l_{local}^{fin}]$ /Å
3.5	4.5	3.577	0.306	8.105	0.626
4.0	5.0	4.105	0.314	8.030	0.604
4.5	5.5	4.659	0.319	8.031	0.581
5.0	6.0	5.194	0.305	8.176	0.635
5.5	6.5	5.703	0.313	7.981	0.766
6.0	7.0	6.213	0.318	7.640	0.647
6.5	7.5	6.708	0.342	7.832	0.507
7.0	8.0	7.218	0.375	8.228	0.430

Fig. 3-3. Behavior of $< l_{\mathrm{local}} >$ during the SCF iteration in the automated DC-PM3 calculations of the water system (N_{water} = 1000).

Next, the dependence of the computational time on the system-size was determined, as shown in Fig. 3-4. The initial buffer size was set to $r_b^{in} = 5.0$ Å and $r_b^{out} = 6.0$ Å. Even for N_{water} = 400, the time for the automated DC-PM3 calculation (54 s) was around four times shorter than that for the standard PM3 calculation (204 s). Furthermore, the time required for the standard PM3 calculations increased steeply with the system size. The scaling analysis with the double logarithmic plot indicated that the times for the standard and automated DC-PM3 calculations scaled as $O(n^{2.7})$ and $O(n^{1.6})$, respectively. For all systems, the DC energy error per atom was within a narrow range: 0.44–0.57 μE_h. It was thus confirmed that the present method is able to control the accuracy of the DC method while maintaining an almost linear-scaling computational cost.

Fig. 3-4. System-size dependence of the wall-clock computational time of standard PM3 and automated DC-PM3 calculations for the model system containing N_{water} water molecules. The initial buffer size for the DC calculations was fixed to $r_{\mathrm{b}}^{\mathrm{in}} = 5.0$ and $r_{\mathrm{b}}^{\mathrm{out}} = 6.0$ Å.

Table 3-5 summarizes the dependence of the energy error, computational time, and average of the major axis radii at the final step ($<l_{\mathrm{local}}^{\mathrm{fin}}>$) on the energy-based threshold, e_{thresh}, in the automated DC-PM3 calculations of the water system (N_{water} = 1000). The initial buffer size was set to $r_{\mathrm{b}}^{\mathrm{in}} = 3.5$ and $r_{\mathrm{b}}^{\mathrm{out}} = 4.5$ Å. The result confirmed that the energy error is almost proportional to the energy threshold, e_{thresh}. As expected, $<l_{\mathrm{local}}^{\mathrm{fin}}>$ decreases gradually as the energy threshold increases. Accordingly, the computational time decreases as the energy threshold increases, while it shows more significant

dependence than $<l_{\mathrm{local}}^{\mathrm{fin}}>$ does. Therefore, it is important to set e_{thresh} appropriately to enjoy both of good accuracy and less computational time.

Table 3-5. Energy threshold (e_{thresh}) dependence of the energy error, computational time, and average of the major axis radii of all localization regions at the final SCF steps in the automated DC-PM3 calculation of the model system containing 1000 water molecules. The initial buffer size was set to $r_{\mathrm{b}}^{\mathrm{in}} = 3.5$ and $r_{\mathrm{b}}^{\mathrm{out}} = 4.5$ Å.

e_{thresh} /μE_{h}	Energy error /μE_{h} atom^{-1}	Time /s	$<l_{\mathrm{local}}^{\mathrm{fin}}>$ /Å
0.01	0.103	451	8.994
0.05	0.295	319	8.396
0.10	0.479	243	8.105
0.50	1.360	179	7.421
1.00	2.153	173	7.156
1.50	2.953	144	7.024

Finally, the parallel efficiency of the program was examined, although the present source code is not optimized for the parallelization. Table 3-6 summarizes the dependence of the wall-clock times (t) for the automated DC-PM3 SCF calculation of the system with N_{water} = 1000 on the number of CPU cores (N_{core}). The times were measured using a computer node equipped with two Intel Xeon E5-2667 CPU (8 cores, 3.20 GHz) and the average of three measurements was calculated. The initial buffer size was set to $r_b^{in} = 5.0$ Å and $r_b^{out} = 6.0$ Å. The parallel scalability S, given at the last column of the Table, is defined as the wall-clock time ratio $S = t(N_{core} = 1) / [N_{core} \times t(N_{core})]$. Up to $N_{core} = 4$, the scalability is higher than 0.7, while it rapidly decreases for $N_{core} > 4$. There are two main reasons for the deterioration: (i) the reordering of the processing subsystem, which is effective for minimizing load imbalance, is not optimized for the varying subsystem size in the present automated DC method, and (ii) the semiempirical Hamiltonian matrix construction is not efficiently parallelized in GAMESS. Although there is room for improvement, the present automated DC implementation is moderately parallelized, which especially works better for larger systems.

Table 3-6. Parallelization efficiency of the automated DC-PM3 calculations of the model system containing 1000 water molecules

N_{core}	Time (t) /s	Scalability (S)
1	710	1.000
2	381	0.933
4	252	0.706
8	229	0.387
16	272	0.163

Then the method was applied to covalently bound systems. Table 3-7 shows the initial buffer-size dependence of the automated DC-PM3 energy for the crambin system treated in the previous Section. The energy difference from the standard PM3 results is also shown in parentheses. Again, it was confirmed that the energy difference was suppressed to small enough values: <1.4 μE_h atom^{-1}. The results for the crambin system did not show a systematic decrease of the energy difference up to $r_b^{in} = 5.5$ Å, as the initial buffer size was sufficiently smaller than the major axis radius of the final localization region, as summarized in Table 3-8. From these data, it was again confirmed that $< l_{local}^{fin} >$ and $\sigma[l_{local}^{fin}]$ tend to be smaller for larger initial buffer sizes. In comparison with Table 3-4, the $< l_{local}^{fin} >$ value for the crambin system is ~1 Å longer than that of the water system as the decay rate of the density matrix elements through covalent bonds is slower than that through hydrogen bonds.

Table 3-7. Initial buffer-size dependence of the automated DC-PM3 energy for the crambin system.

r_b^{in} /Å	r_b^{out} /Å	Energy /E_h	(Diff. /μE_h atom^{-1})
3.5	4.5	–2117.084675	(+0.10)
4.0	5.0	–2117.084601	(+0.21)
4.5	5.5	–2117.084647	(+0.14)
5.0	6.0	–2117.083858	(+1.37)
5.5	6.5	–2117.084462	(+0.43)
Standard-PM3		–2117.084739	

Table 3-8. Average and standard deviation of the major axis radii of all localization regions at the initial and final SCF steps in the automated DC-PM3 calculation of the crambin system.

r_b^{in} /Å	r_b^{out} /Å	$<l_{local}^{ini}>$ /Å	$\sigma[l_{local}^{ini}]$ /Å	$<l_{local}^{fin}>$ /Å	$\sigma[l_{local}^{fin}]$ /Å
3.5	4.5	5.348	0.890	9.082	1.204
4.0	5.0	5.835	0.981	9.158	1.208
4.5	5.5	6.338	0.948	9.043	1.178
5.0	6.0	6.781	0.969	9.072	1.198
5.5	6.5	7.204	1.065	8.899	1.191

Next, the present method was examined in calculations of the conjugated graphene system depicted in Fig. 3-5 ($C_{180}H_{48}$). All atoms were placed on a plane and the C–C and C–H bond lengths were fixed to 1.42 and 1.09 Å, respectively. Table 3-9 shows the initial buffer-size dependence of the DC-PM3 energy for $C_{180}H_{48}$. In the DC calculation, the entire system was divided by a lattice spacing of 3.5 Å and each fragment was treated as a central region. The definitions of the initial buffer sizes, $r_{\mathrm{b}}^{\mathrm{in}}$ and $r_{\mathrm{b}}^{\mathrm{out}}$, were the same as those in the previous Sections. The energies obtained with a fixed buffer size are given in Table 3-9, together with the estimated energy errors at the final SCF step and $< l_{\mathrm{local}}^{\mathrm{fin}} >$. Unlike the results for the water and crambin systems, the present automated DC method afforded in some cases a large energy deviation of >10 μE_{h} atom^{-1}. The estimated energy error with the fixed buffer size was found to be about one order of magnitude smaller than the actual error. Due to the significantly slow decay of the density matrix for conjugated systems, the energy error estimated in the outer buffer region may be insufficient to reproduce the actual energy error. In addition, the energy error does not converge to the standard PM3 result due to the finite temperature approximation in the DC method. Actually, the finite-temperature PM3 energy with $\beta = 200$ a.u. is -810.643352 E_{h}, which is much closer to the converged DC-PM3 energy.

Fig. 3-5. Structure of the graphene system, $C_{180}H_{48}$.

Table 3-9. Initial buffer-size dependence of the buffer-size fixed and automated DC-PM3 energies for the graphene system, $C_{180}H_{48}$. The estimated energy error at the final SCF step in the buffer-size fixed calculation is also given. Standard PM3 energy is –810.650309 E_h

		Buffer size fixed		Automated		
r_b^{in} /Å	r_b^{out} /Å	Actual error /E_h	$-\Delta E$ /E_h	Energy /E_h	(Diff. /μE_h atom^{-1})	$<l_{local}^{fin}>$ /Å
3.5	5.0	+0.466022	+0.141160	–810.650931	(–2.73)	10.672
4.0	5.5	+0.163445	+0.023277	–810.651058	(–3.29)	10.820
4.5	6.0	+0.107509	+0.024784	–810.644480	(+25.56)	12.251
5.0	6.5	+0.137531	+0.013554	–810.650325	(–0.07)	11.458
5.5	7.0	+0.065487	+0.008674	–810.648093	(+9.72)	11.427
6.0	7.5	+0.041446	+0.008377	–810.651029	(–3.16)	10.926
6.5	8.0	+0.044827	+0.004561	–810.644494	(+25.50)	12.033
7.0	8.5	+0.039109	+0.003055	–810.643122	(+31.52)	11.796
7.5	9.0	+0.016598	+0.001494	–810.655281	(–21.81)	10.846
8.0	9.5	+0.011964	+0.001004	–810.644095	(+27.25)	12.199
8.5	10.0	+0.014299	+0.001242	–810.642821	(+32.84)	11.992
9.0	10.5	+0.017775	+0.000638	–810.643328	(+30.61)	12.310
9.5	11.0	+0.010530	+0.000513	–810.644435	(+25.76)	12.543
10.0	11.5	+0.008625	+0.000339	–810.644620	(+24.95)	12.211

Finally, the dependence of the energy error on the energy-based threshold, e_{thresh}, was assessed. Fig. 3-6 shows the dependence of the final energy error on e_{thresh} for the automated DC-PM3 calculation of 1000 water molecules and the crambin and graphene systems. The initial buffer size was set to $r_{\mathrm{b}}^{\mathrm{in}} = 3.5$ and $r_{\mathrm{b}}^{\mathrm{out}} = 4.5$ Å (or $r_{\mathrm{b}}^{\mathrm{out}} = 5.0$ Å for the graphene system). For the water and crambin systems, which were adequately treated by the automated DC method, the energy error increased proportionally to e_{thresh}, as expected. For the graphene system, however, the energy error did not show a systematic trend but oscillated throughout the e_{thresh} value range, even at low e_{thresh} values. Although there is still some room for improvement in the present automated DC scheme, it has been demonstrated that the energy error can be suppressed with the present method even for conjugated systems.

Fig. 3-6. Dependence of the energy error on the energy-based threshold, e_{thresh}, in the automated DC-PM3 calculations of 1000 water molecules and the crambin and graphene systems.

The present scheme was applied to the HF method and DFT with the pure BLYP[39,40] and hybrid B3LYP[41.42] functionals. Here, the DC energy error is estimated with Eq. (3.9) even for DFT calculations, where the Hamiltonian matrix is not linear with the density matrix. The option to use the HF Hamiltonian (Fock) matrix at the early SCF stage of the DFT calculation, which is adopted in the default setting of the GAMESS program, was switched off. Table 3-10 shows the initial buffer-size dependence of the DC-HF, DC-B3LYP, and DC-BLYP energies for a *n*-alkane ($C_{150}H_{302}$) with the 6-31G* basis set.[43] In the DC calculations, a C_2H_4 (or C_2H_5 for the edges) group is adopted as a central region. For the DC-HF and DC-B3LYP calculations, the energy error could be controlled within 0.7 μE_h atom^{-1}, while that for the pure DFT (DC-BLYP) calculation is one order larger in magnitude. The final localization region for central subsystems contains $(C_2H_4)_{10-12}$ in the DC-HF and DC-B3LYP calculations, while that does $(C_2H_4)_{7-8}$ in the DC-BLYP calculations. This result suggests that the magnitude of ΔE with Eq. (3.9) is underestimated without the HF exchange term. Actually, the energy error of the DC-BLYP calculation with the DC-HF final localization region was 0.09 μE_h atom^{-1} for $r_b^{in} = 3.0$ and $r_b^{out} = 4.5$ Å. Therefore, in practical DC-DFT calculation, it is recommended to determine the appropriate buffer size with the early DC-HF SCF cycles, which can be performed in GAMESS by switching back on the default option to use the Fock matrix at the early SCF stage.

Table 3-10. Initial buffer-size dependence of the automated DC-HF and DC-DFT energies for the *n*-alkane system, $C_{150}H_{302}$.

		HF		B3LYP		BLYP	
r_b^{in} /Å	r_b^{out} /Å	Energy /E_h	(Diff.) /$\mu E_h \cdot atom^{-1}$	Energy /E_h	(Diff.) /$\mu E_h \cdot atom^{-1}$	Energy /E_h	(Diff.) /$\mu E_h \cdot atom^{-1}$
3.0	4.5	–5855.829391	(–0.20)	–5897.795934	(+0.64)	–5893.578891	(+4.98)
4.0	5.5	–5855.829339	(–0.08)	–5897.796218	(+0.02)	–5893.578904	(+4.95)
5.0	6.5	–5855.829321	(–0.04)	–5897.796217	(+0.02)	–5893.578904	(+4.95)
6.0	7.5	–5855.829329	(–0.06)	–5897.795958	(+0.59)	–5893.578954	(+4.84)
7.0	8.5	–5855.829347	(–0.10)	–5897.796192	(+0.07)	–5893.579524	(+3.58)
Standard		–5855.829302		–5897.796225		–5893.581143	

3.5. Concluding remarks

In this study, I have proposed an energy-based error estimation scheme for the linear-scaling DC quantum chemical method with the help of two-layer buffer regions. Exploiting the fact that the estimated energy error can be divided into contributions from the atoms in the outer buffer region of each subsystem, our error estimation scheme was utilized for the automatic determination of the appropriate buffer region for the DC method. The present automated DC method worked satisfactorily in calculations of water, protein, and alkane systems, although its performance was insufficient in the calculation of a delocalized graphene system. Improvement of the present scheme for delocalized systems will be the scope of future studies. Furthermore, in the present scheme, the buffer region was gradually extended during the SCF cycles. There is an alternative approach to reduce the buffer region from a large initial buffer size, which may be preferably used when the method is applied to a series of quantum chemical calculations, such as geometry optimizations, where the appropriate buffer region of the previous step is available.

An energy-based error control scheme such as the present method will be indispensable for quantum chemical molecular dynamics simulations, especially for microcanonical ensembles, where the total energy conservation is rigorously examined. Recently, Nakai and coworkers have published a series of studies performing quantum chemical molecular dynamics simulations with the DC-DFTB method.[28,29] The present automated DC method can straightforwardly be extended to the so-called DFTB2 Hamiltonian, which is linear with respect to the density matrix. Furthermore, the present

error estimation scheme is expected to work even for non-linear Hamiltonians such as DFTB3. The development of an automated DC-DFTB molecular dynamics program is desirable not only to reduce the effort of preliminary assessments before the production runs but also to guarantee the accuracy of the results. The application of the present scheme to the DC Hartree–Fock–Bogoliubov method,[44] which can effectively treat the static electron correlation of large systems,[45] is also straightforward, as well as that to the open-shell DC unrestricted HF method.[13,24] However, the present method cannot be combined with DC post-HF correlation methods such as the second-order Møller–Plesset perturbation (MP2)[46–48] and coupled cluster[49–51] theories. As pointed out by Kobayashi and Nakai,[52] the appropriate buffer size for DC post-HF correlation calculations is generally smaller than that for DC-HF calculations. Furthermore, especially in DC-MP2 calculations, the appropriate buffer size should be determined before carrying out the MP2 calculations as the procedure is not iterative. The solution to this issue will pave the way toward the development of an automated DC-MP2 scheme in the near future.

3.6. Reference

[1] A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory;* Dover: New York, **1996**.

[2] T. Tsuneda, *Density Functional Theory in Quantum Chemistry;* Springer: Tokyo, **2014**.

[3] T. Helgaker, P. Jørgensen, J. Olsen, *Molecular Electronic-Structure Theory;* John Wiley & Sons: New York, **2000**.

[4] X.-P. Li, R. W. Nunes, D. Vanderbilt, *Phys. Rev. B* **1993**, *47*, 10891.

[5] R. W. Nunes, D. Vanderbilt, *Phys. Rev. B* **1994**, *50*, 17611.

[6] D. R. Bowler, R. Choudhury, M. J. Gillan, T. Miyazaki, *Phys. Stat. Sol. B* **2006**, *243*, 989.

[7] N. Sahu, S. R. Gadre, *Acc. Chem. Res.* **2014**, *47*, 2739.

[8] N. Sahu, S. D. Yeole, S. R. Gadre, *J. Chem. Phys*. **2013**, *138*, 104101.

[9] V. Ganesh, R. K. Dongare, P. Balanarayan, S. R. Gadre, *J. Chem. Phys.* **2006**, *125*, 104109.

[10] S. Li, W. Li, J. Ma, *Acc. Chem. Res.* **2014**, *47*, 2712.

[11] W. Li, S. Li, Y. Jiang, *J. Phys. Chem. A* **2007**, *111*, 2193.

[12] W. Yang, T.-S. Lee, *J. Chem. Phys.* **1995**, *103*, 5674.

[13] M. Kobayashi, H. Nakai, *Phys. Chem. Chem. Phys.* **2012**, *14*, 7629.

[14] M. Kobayashi, H. Nakai, In *Linear-Scaling Techniques in Computational Chemistry and Physics,* Chapter 5; R. Zaleśny, M. G. Papadopoulos, P. G. Mezey, J. Leszczynski, Eds.; Springer: Dordrecht, **2011**; pp. 97–127.

[15] A. M. N. Niklasson, *Phys. Rev. B* **2002**, *66*, 155115.

[16] J. Kim, Y. Jung, *Int. J. Quantum Chem.* **2016**, *116*, 563.

[17] E. H. Rubensson, *SIAM J. Sci. Comput.* **2012**, *34*, B1.

[18] E. H. Rubensson, E. Rudberg, P. Salek, *J. Chem. Phys.* **2008**, *128*, 074106.

[19] A. M. N. Niklasson, S. M. Mniszewski, C. F. A. Negre, M. J. Cawkwell, P. J. Swart, J. Mohd-Yusof, T. C. Germann, M. E. Wall, N. Bock, E. H. Rubensson, H. Djidjev, *J. Chem. Phys.* **2016**, *144*, 234101.

[20] D. G. Fedorov, K. Kitaura, *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems;* CRC: Boca Raton, **2009**.

[21] D. G. Fedorov, K. Kitaura, *J. Chem. Phys.* **2004**, *120*, 6832.

[22] T. Nakano, Y. Mochizuki, K. Yamashita, C. Watanabe, K. Fukuzawa, K. Segawa, Y. Okiyama, T. Tsukamoto, S. Tanaka, *Chem. Phys. Lett.* **2012**, *523*, 128.

[23] T. Akama, M. Kobayashi, H. Nakai, *J. Comput. Chem.* **2007**, *28*, 2003.

[24] M. Kobayashi, T. Yoshikawa, H. Nakai, *Chem. Phys. Lett.* **2010**, *500*, 172.

[25] M. Kobayashi, T. Kunisada, T. Akama, D. Sakura, H. Nakai, *J. Chem. Phys.* **2011**, *134*, 034105.

[26] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai, G. Seifert, *Phys. Rev. B* **1998**, *58*, 7260.

[27] M. Gaus, Q. Cui, M. Elstner, *J. Chem. Theory Comput.* **2011**, *7*, 931.

[28] H. Nishizawa, Y. Nishimura, M. Kobayashi, S. Irle, H. Nakai, *J. Comput. Chem.* **2016**, *37*, 1983.

[29] H. Nakai, A. W. Sakti, Y. Nishimura, *J. Phys. Chem. B* **2016**, *120*, 217.

[30] S. L. Dixon, K. M. Merz Jr., *J. Chem. Phys.* **1997**, *107*, 879.

[31] H. Nakai, *Chem. Phys. Lett.* **2002**, *363*, 73.

[32] S. Goedecker, *Rev. Mod. Phys.* **1999**, *71*, 1085.

[33] M. S. Gordon, M. W. Schmidt, In *Theory and Applications of Computational Chemistry: The First Forty Years;* C. Dykstra, G. Frenking, K. Kim, G. Scuseria, Eds.; Elsevier: Amsterdam, **2005**; pp 1167–1189.

[34] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery Jr., *J. Comput. Chem.* **1993**, *14*, 1347.

[35] W. J. Hehre, R. Ditchfield, J.A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.

[36] K. Kitaura, FU. https://sourceforge.net/projects/fusuite/ (accessed Feb 6, 2017).

[37] J. J. P. Stewart, *J. Comput. Chem.* **1989**, *10*, 209.

[38] J. J. P. Stewart, *J. Comput. Chem.* **1989**, *10*, 221.

[39] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.

[40] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.

[41] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.

[42] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.

[43] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213.

[44] V. N. Staroverov, G. E. Scuseria, *J. Chem. Phys.* **2002**, *117*, 11107.

[45] M. Kobayashi, T. Taketsugu, *Chem. Lett.* **2016**, *45*, 1268.

[46] M. Kobayashi, Y. Imamura, H. Nakai, *J. Chem. Phys.* **2007**, *127*, 074103.

[47] M. Kobayashi, T. Taketsugu, *Theor. Chem. Acc.* **2015**, *134*, 107.

[48] M. Kobayashi, H. Nakai, *J. Chem. Phys.* **2013**, *138*, 044102.

[49] M. Kobayashi, H. Nakai, *J. Chem. Phys.* **2008**, *129*, 044103.

[50] M. Kobayashi, H. Nakai, *J. Chem. Phys.* **2009**, *131*, 114108.

[51] T. Yoshikawa, M. Kobayashi, H. Nakai, *Int. J. Quantum Chem.* **2013**, *113*, 218.

[52] M. Kobayashi, H. Nakai, *Int. J. Quantum Chem.* **2009**, *109*, 2227.

4. Energy based Automatic determination of buffer region in DC-MP2 calculation

4.1. Introduction

By virtue of recent advances in quantum chemical theory as well as the improvements in computer performance, electronic structure calculations of large-scale systems such as proteins have now become technically feasible. Such theoretical advances include the development of linear-scaling (or low-scaling) electronic structure methods. In the standard formalism of electronic structure methods, the computational time increases cubically [$O(N^3)$] with respect to the system size N, even with the simplest Hartree-Fock (HF) method[1] or density functional theory (DFT),[2] owing to the diagonalization of the Hamiltonian matrix. Furthermore, in case of post-HF calculations, such as the second order Møller-Plesset perturbation (MP2)[3-5] and coupled cluster (CC) theories,[4,5] their time scalings deteriorate as $O(N^5)$ or more. Therefore, the standard formalisms of electronic structure methods cannot be applied to large-scale systems. By introducing approximations to the standard formalisms, many low-scaling electronic structure methods[6-10] have been proposed for treating such systems. Many of these methods equip some schemes to adjust the errors derived from the low-scaling approximations based on the distance parameter. For example, in the molecular tailoring approach proposed by Garde *et al.*,[11] R-goodness parameter is used to determine the quality of each

fragment.[12,13] In the generalized energy-based fragmentation approach,[14,15] each fragment is constructed with the distance threshold (ξ). The cluster-in-molecule local correlation method also adopts the distance threshold ξ to control the size of the cluster,[16] while a simple correction scheme to account for the distant-pair correlation has recently been proposed.[17] The accuracy of the fragment molecular orbital method[18] can be systematically improved by increasing the order of many-body expansion from the original two-body to three-body[19,20] and four-body[21] expansions. The pair natural orbital (PNO) electron correlation approach[22,23] adopts several truncation schemes for construction of correlated virtual orbitals (i.e., PNOs) for each occupied local molecular orbital (MO) pair, where the bond-based (so-called IEXT) or distance-based (so-called REXT) truncation is used to determine the local virtual orbital region to construct PNOs. Since molecular energy is the most important property in quantum chemical calculations, an energy-based parameter is more desirable than a distance-based one. For example, the divide-expand-consolidate method utilizes the energy-based fragment optimization threshold to determine the atomic occupied and virtual orbital spaces in each fragment.[24,25]

Yang and coworkers introduced a linear-scaling approach called the divide-and-conquer (DC) method.[26,27] The DC method has been applied to the HF or DFT self-consistent field (SCF),[26,28] density-functional tight-binding,[29-32] and post-HF (MP2[33-36] or CC[37-39]) energy calculations as well as the SCF[40] and MP2[41] energy gradient calculations. For treating static electron correlation in large-scale systems, the DC method has also been combined with the Hartree-Fock-Bogoliubov method[42] and the thermally-assisted occupation (finite temperature) scheme.[43] In the DC method, the size of the buffer region plays the role of the distance parameter to adjust the approximation error; a

larger buffer size leads to a smaller approximation error. However, it is still difficult to estimate the error in energy based on the distance-based adjustment parameter. Recently, I[44] proposed a scheme to estimate the energy error introduced in the DC-HF and DC-DFT calculations using a two-layer buffer region scheme introduced by Dixon and Merz.[45] This estimation scheme can successfully be applied to automatically determine the appropriate buffer region based on the estimated energy error.[44]

This Chapter attempts to export the idea of the previous automated DC-HF scheme to the DC-MP2 calculation. Kobayashi *et al.*[36] reported that the buffer region used for the MP2 correlation calculation can be contracted from that for the HF one to achieve the same energy accuracy as the DC-HF calculation because of the short-range nature of the MP2 dynamical electron correlation. I first develop a method to estimate the subsystem MP2 correlation energy contribution from each atom in the buffer region. Here, the idea of the atomic orbital (AO) Laplace MP2 method[46-50] is used as well as the Schwarz inequality. Based on this estimated energy contribution, I established an algorithm to automatically determine the appropriate buffer region in the DC-MP2 calculation.

This Chapter consists of five Sections. Section 4.2 gives a brief summary of the linear-scaling DC electron correlation method with a fixed buffer region. In Section 4.3, the present procedure to estimate the energy contribution from each buffer atom and the automated DC-MP2 algorithm is explained. Numerical assessments are described in Section 4.4. Finally, I provide concluding remarks in Section 4.5.

4.2. DC-MP2 scheme

I first outline the DC-MP2 electron correlation calculation scheme. The DC-MP2 method is applicable only with atom-centered basis functions. Each basis function, $\phi_\mu(\mathbf{r})$, called an AO, is denoted by a Greek letter index, μ, ν, …. In the DC method, the entire system is divided into several subsystems, each of which consists of the central and buffer regions. Each central region is mutually exclusive with the other central regions. The sets of AOs belonging to the central and buffer regions of subsystem α are referred to as $\boldsymbol{S}(\alpha)$ and $\boldsymbol{B}(\alpha)$, respectively.

In the DC-MP2 method, the MOs in the subsystem α,

$$\psi_p^\alpha(\mathbf{r}) = \sum_{\mu \in \boldsymbol{L}(\alpha)} C_{\mu p}^\alpha \phi_\mu(\mathbf{r}), \tag{4.1}$$

are used to evaluate the correlation energy of subsystem α, where $\boldsymbol{L}(\alpha) = \boldsymbol{S}(\alpha) \cup \boldsymbol{B}(\alpha)$ represents the set of AOs in the localization region and p refers to an arbitrary MO. The MO coefficients, $\{\mathbf{C}_p^\alpha\}$, and the MO energies, $\{\varepsilon_p^\alpha\}$, of subsystem α are obtained by solving the Roothaan equation for each subsystem:

$$\mathbf{F}^\alpha[\mathbf{D}^{\mathrm{SCF}}]\mathbf{C}_p^\alpha = \varepsilon_p^\alpha \mathbf{S}^\alpha \mathbf{C}_p^\alpha, \tag{4.2}$$

where $\mathbf{F}^\alpha[\mathbf{D}^{\mathrm{SCF}}]$ is the subsystem Fock matrix constructed with the density matrix $\mathbf{D}^{\mathrm{SCF}}$, and $\mathbf{S}^\alpha$ is the subsystem overlap matrix with the element $S_{\mu\nu}^\alpha = \left(\phi_\mu \mid \phi_\nu\right)$ for $\mu, \nu \in \boldsymbol{L}(\alpha)$.

The density matrix, $\mathbf{D}^{\mathrm{SCF}}$, can be constructed from the standard or approximate HF calculation, such as the DC-HF one. Note that the subsystem SCF equation (4.2) has to be solved not self-consistently but just once using predetermined $\mathbf{D}^{\mathrm{SCF}}$. If $\mathbf{D}^{\mathrm{SCF}}$ is obtained

from the DC-HF calculation, it is constructed with the local density matrices, $\{\mathbf{D}^{\alpha}\}$, and the partition matrices, $\{\mathbf{p}^{\alpha}\}$, as the following:

$$D_{\mu\nu}^{\mathrm{SCF}} \approx \sum_{\alpha}^{\mathrm{subsystem}} p_{\mu\nu}^{\alpha} D_{\mu\nu}^{\alpha} , \tag{4.3}$$

$$D_{\mu\nu}^{\alpha} = \sum_{p} f_{\beta}(\varepsilon_{\mathrm{F}} - \varepsilon_{p}^{\alpha}) C_{\mu p}^{\alpha} C_{\nu p}^{\alpha} , \tag{4.4}$$

$$p_{\mu\nu}^{\alpha} = \begin{cases} 1 & (\mu \in \boldsymbol{S}(\alpha) \wedge \nu \in \boldsymbol{S}(\alpha)) \\ 1/2((\mu \in \boldsymbol{S}(\alpha) \wedge \nu \in \boldsymbol{B}_{\mathrm{SCF}}(\alpha)) \vee (\mu \in \boldsymbol{B}_{\mathrm{SCF}}(\alpha) \wedge \nu \in \boldsymbol{S}(\alpha))) \\ 0 & (\text{otherwise}) \end{cases} , \tag{4.5}$$

where $f_{\beta}(x) = [1 + \exp(-\beta x)]^{-1}$ is the Fermi distribution function with the inverse temperature, β, and ε_{F} is the universal Fermi level. The details of the DC-HF procedure can be found in Refs. 26 and 27.

Before the evaluation of the subsystem correlation energy, the subsystem MOs of Eq. (4.1) must be classified into occupied $\{\psi_i^{\alpha}, \psi_j^{\alpha}, ...\}$ and virtual ones $\{\psi_a^{\alpha}, \psi_b^{\alpha}, ...\}$. This can be accomplished by, for example, using the Fermi level determined in the prior DC-HF calculations. The MP2 correlation energy for the entire system, $\Delta E_{\mathrm{corr}}^{(2)}$, can be approximated as the sum of the subsystem MP2 correlation energies, $\{\Delta E_{\mathrm{corr}}^{\alpha(2)}\}$,

$$\Delta E_{\mathrm{corr}}^{(2)} \approx \sum_{\alpha} \Delta E_{\mathrm{corr}}^{\alpha(2)} . \tag{4.6}$$

Because the buffer region in each localization region overlaps with the other localization regions, $\Delta E_{\mathrm{corr}}^{\alpha(2)}$ is obtained as the MP2 correlation energy corresponding to the central region of the localization region α by means of energy density analysis (EDA).[51] The subsystem correlation energy is then evaluated by

$$\Delta E_{\mathrm{corr}}^{\alpha(2)} = \sum_{i^\alpha, j^\alpha}^{\mathrm{occ}(\alpha)} \sum_{a^\alpha, b^\alpha}^{\mathrm{vir}(\alpha)} \sum_{\mu \in \mathbf{S}(\alpha)} \frac{C_{\mu i}^{\alpha} (\mu a^\alpha \mid j^\alpha b^\alpha)}{\varepsilon_i^\alpha + \varepsilon_j^\alpha - \varepsilon_a^\alpha - \varepsilon_b^\alpha} [2(a^\alpha i^\alpha \mid b^\alpha j^\alpha) - (a^\alpha j^\alpha \mid b^\alpha i^\alpha)], \tag{4.7}$$

with the two-electron integral notation

$$(i^\alpha a^\alpha \mid j^\alpha b^\alpha) = \iint d\mathbf{r}_1 d\mathbf{r}_2 \psi_i^{\alpha *}(\mathbf{r}_1) \psi_a^\alpha(\mathbf{r}_1) r_{12}^{-1} \psi_j^{\alpha *}(\mathbf{r}_2) \psi_b^\alpha(\mathbf{r}_2).$$

4.3. Estimation of DC-MP2 energy based on AO-Laplace MP2 method

Based on EDA, the MP2 correlation energy for subsystem α, $\Delta E_{\mathrm{corr}}^{\alpha(2)}$, can be further divided into contributions from the atoms in the localization region α, $\Delta E_{B}^{\alpha(2)}$, as

$$\Delta E_{\mathrm{corr}}^{\alpha(2)} = \sum_{B \in \boldsymbol{L}(\alpha)} \Delta E_{B}^{\alpha(2)}, \tag{4.8}$$

$$\Delta E_{B}^{\alpha(2)} = \sum_{i^{\alpha}, j^{\alpha}}^{\mathrm{occ}(\alpha)} \sum_{a^{\alpha}, b^{\alpha}}^{\mathrm{vir}(\alpha)} \sum_{\mu \in \boldsymbol{S}(\alpha)} \sum_{\nu \in B} \frac{C_{\mu i}^{\alpha} C_{\nu a}^{\alpha} (\mu\nu \mid j^{\alpha} b^{\alpha})}{\varepsilon_{i}^{\alpha} + \varepsilon_{j}^{\alpha} - \varepsilon_{a}^{\alpha} - \varepsilon_{b}^{\alpha}} [2(a^{\alpha} i^{\alpha} \mid b^{\alpha} j^{\alpha}) - (a^{\alpha} j^{\alpha} \mid b^{\alpha} i^{\alpha})]. \tag{4.9}$$

According to the local correlation philosophy for dynamical electron correlation,[52-54] it is expected that $\Delta E_{B}^{\alpha(2)}$ rapidly decreases as the distance between atom B and central region α increases. The exponential decay of the MP2 energy contribution with respect to the interatomic distance is discussed in the Appendix. As pointed out by Kobayashi and Nakai,[36] the appropriate size of the buffer region for the DC-MP2 calculation can be smaller than that for the DC-HF calculation because of the locality of the dynamical electron correlation. Therefore, if the absolute value of $\Delta E_{B}^{\alpha(2)}$ is estimated to be smaller than some criterion, the energy change by excluding atom B from the buffer region of subsystem α is expected to be small. By applying the AO-Laplace MP2 technique to Eq. (4.9), $\Delta E_{B}^{\alpha(2)}$ can be expressed as

$$\Delta E_{B}^{\alpha(2)} = -\int_{0}^{\infty} \sum_{\mu \in \boldsymbol{S}(\alpha)} \sum_{\nu \in B} \sum_{\lambda\sigma} \sum_{\gamma\kappa\delta\varepsilon} X_{\mu\gamma}^{\alpha}(\tau) Y_{\nu\kappa}^{\alpha}(\tau) X_{\lambda\delta}^{\alpha}(\tau) Y_{\sigma\varepsilon}^{\alpha}(\tau) (\mu\nu \mid \lambda\sigma) [2(\kappa\gamma \mid \varepsilon\delta) - (\kappa\delta \mid \varepsilon\gamma)] d\tau, \tag{4.10}$$

where $\mathbf{X}^{\alpha}(\tau)$ and $\mathbf{Y}^{\alpha}(\tau)$ are the energy-weighted density matrices expressed as

$$X_{\mu\nu}^{\alpha}(\tau)=\sum_{i}C_{\mu i}^{\alpha}C_{\nu i}^{\alpha}e^{(\varepsilon_i^{\alpha}-\varepsilon_{\mathrm{F}})\tau}\ , \tag{4.11}$$

$$Y_{\mu\nu}^{\alpha}(\tau)=\sum_{a}C_{\mu a}^{\alpha}C_{\nu a}^{\alpha}e^{-(\varepsilon_a^{\alpha}-\varepsilon_{\mathrm{F}})\tau}\ . \tag{4.12}$$

Here, the Fermi level, ε_{F}, may be already determined in the prior DC-HF calculation, or may be the midpoint energy between HOMO and LUMO in the prior HF calculation. For estimation purpose, I drastically approximate the integral in Eq. (4.10) by the one-point Gauss-Laguerre quadrature, namely,

$$\Delta E_B^{\alpha(2)}\sim -e\sum_{\mu\in\mathbf{S}(\alpha)}\sum_{\nu\in B}\sum_{\lambda\sigma}\sum_{\gamma\kappa\delta\varepsilon}X_{\mu\gamma}^{\alpha}Y_{\nu\kappa}^{\alpha}X_{\lambda\delta}^{\alpha}Y_{\sigma\varepsilon}^{\alpha}(\mu\nu\,|\,\lambda\sigma)[2(\kappa\gamma\,|\,\varepsilon\delta)-(\kappa\delta\,|\,\varepsilon\gamma)]\,, \tag{4.13}$$

$$X_{\mu\nu}^{\alpha}=\sum_{i}C_{\mu i}^{\alpha}C_{\nu i}^{\alpha}e^{(\varepsilon_i^{\alpha}-\varepsilon_{\mathrm{F}})}\ , \tag{4.14}$$

$$Y_{\mu\nu}^{\alpha}=\sum_{a}C_{\mu a}^{\alpha}C_{\nu a}^{\alpha}e^{-(\varepsilon_a^{\alpha}-\varepsilon_{\mathrm{F}})}\ . \tag{4.15}$$

Assuming that the rhs of Eq. (4.13) gives the upper limit of $\Delta E_B^{\alpha(2)}$, its absolute value can be bounded by adopting the Schwarz inequality

$$|(ij\,|\,kl)|\le\sqrt{|(ij\,|\,ij)|}\sqrt{|(kl\,|\,kl)|} \tag{4.16}$$

as

$$\begin{aligned}
|\Delta E_B^{\alpha(2)}| &\le e\sum_{\mu\in\mathbf{S}(\alpha)}\sum_{\nu\in B}\sum_{\lambda\sigma}\sum_{\gamma\delta\kappa\varepsilon}|X_{\mu\gamma}^{\alpha}||Y_{\nu\kappa}^{\alpha}||X_{\lambda\delta}^{\alpha}||Y_{\sigma\varepsilon}^{\alpha}||(\mu\nu\,|\,\lambda\sigma)|[2|(\kappa\gamma\,|\,\varepsilon\delta)|+|(\kappa\delta\,|\,\varepsilon\gamma)|]\\
&\le e\sum_{\mu\in\mathbf{S}(\alpha)}\sum_{\nu\in B}\sum_{\lambda\sigma}\sum_{\gamma\delta\kappa\varepsilon}|X_{\mu\gamma}^{\alpha}||Y_{\nu\kappa}^{\alpha}||X_{\lambda\delta}^{\alpha}||Y_{\sigma\varepsilon}^{\alpha}|A_{\mu\nu}^{\alpha}A_{\lambda\sigma}^{\alpha}[2A_{\kappa\gamma}^{\alpha}A_{\varepsilon\delta}^{\alpha}+A_{\kappa\delta}^{\alpha}A_{\varepsilon\gamma}^{\alpha}]\\
&\le e\sum_{\mu\in\mathbf{S}(\alpha)}\sum_{\nu\in B}\sum_{\lambda\sigma}\sum_{\gamma\delta\kappa\varepsilon}|X_{\mu\gamma}^{\alpha}||Y_{\nu\kappa}^{\alpha}||X_{\lambda\delta}^{\alpha}||Y_{\sigma\varepsilon}^{\alpha}|A_{\mu\nu}^{\alpha}A_{\lambda\sigma}^{\alpha}[2A_{\kappa\gamma}^{\alpha}\max(\mathbf{A}^{\alpha})+A_{\kappa\delta}^{\alpha}A_{\varepsilon\gamma}^{\alpha}]\quad ,\\
&\sim e\sum_{\mu\in\mathbf{S}(\alpha)}\sum_{\nu\in B}\sum_{\lambda\sigma}\sum_{\gamma\delta\kappa\varepsilon}|X_{\mu\gamma}^{\alpha}||Y_{\nu\kappa}^{\alpha}||X_{\lambda\delta}^{\alpha}||Y_{\sigma\varepsilon}^{\alpha}|A_{\mu\nu}^{\alpha}A_{\lambda\sigma}^{\alpha}[2A_{\kappa\gamma}^{\alpha}\max(\mathbf{A}^{\alpha})]\\
&=e\left(\sum_{\lambda\sigma}\sum_{\delta\varepsilon}|X_{\lambda\delta}^{\alpha}||Y_{\sigma\varepsilon}^{\alpha}|A_{\lambda\sigma}^{\alpha}\right)\sum_{\mu\in\mathbf{S}(\alpha)}\sum_{\nu\in B}\sum_{\gamma\kappa}|X_{\mu\gamma}^{\alpha}||Y_{\nu\kappa}^{\alpha}|A_{\mu\nu}^{\alpha}[2A_{\kappa\gamma}^{\alpha}\max(\mathbf{A}^{\alpha})]
\end{aligned} \tag{4.17}$$

where $A_{\mu\nu}^{\alpha}=\sqrt{|(\mu\nu\,|\,\mu\nu)|}$. Here, the analogy to the scaled opposite-spin MP2 method,[55] the term $A_{\kappa\delta}^{\alpha}A_{\varepsilon\gamma}^{\alpha}$ was omitted owing to its smaller contribution. Because the

summation in parentheses in Eq. (4.17) is constant for subsystem α, the following index can be considered as the magnitude of the contribution from atom B:

$$e_B^\alpha = e \sum_{\mu \in \boldsymbol{S}(\alpha)} \sum_{\nu \in B} \sum_{\gamma\kappa} \left| X_{\mu\gamma}^\alpha \right| \left| Y_{\nu\kappa}^\alpha \right| A_{\mu\nu}^\alpha [2A_{\kappa\gamma}^\alpha \max(\mathbf{A}^\alpha)]. \tag{4.18}$$

Using the above e_B^α index, the following automatic determination scheme for the buffer region in the DC-MP2 method is proposed:

i. Assignment of the initial DC-MP2 buffer region for each subsystem. This may be determined by prior DC-HF calculation.

ii. Evaluation of e_B^α from Eq. (4.18).

iii. The exclusion of atom B from the buffer region of subsystem α if e_B^α is smaller than the energy threshold.

iv. Reconstruction of subsystem molecular orbitals $\{\mathbf{C}_p^\alpha\}$ and $\{\varepsilon_p^\alpha\}$, using Eq. (4.2).

v. Evaluation of the subsystem correlation energy, $\Delta E_{\mathrm{corr}}^{\alpha(2)}$, from Eq. (4.7).

The above procedure is illustrated in Fig. 4-1. The additional computational cost for the evaluation of all necessary e_B^α scales as $O(Nm^3)$, where N and m represent the sizes of the entire system and buffer region, respectively, since the evaluation of each e_B^α of Eq. (4.18) scales with $O(m^2)$ owing to the summation over γ and κ and the number of e_B^α to be evaluated scales with $O(Nm)$.

Fig. 4-1. Illustration of determination scheme for the buffer region.

4.4. Numerical assessment

4.4.1. Computational details

I implemented the above-mentioned automatically controlled DC-MP2 method to the GAMESS package[56,57] and evaluated its accuracy and efficiency for the different types of systems. In the DC-HF calculations, the inverse temperature parameter, β, was set to 125 a.u. and the Fermi function cutoff factor (the FTOL option of $DANDC input group in GAMESS program) was set to 20. In addition, the parameters in the automated DC-HF method were set to $e_{\mathrm{thresh}}^{\mathrm{SCF}}$ = 0.1 μE_{h} and r_{ext} = 3.0 Å, the definitions of which are given in our previous paper.[44] The 6-31G(d) basis set[58] was adopted throughout this paper. I introduced the major axis radii of the HF and MP2 localization regions for subsystem α, $l_{\mathrm{SCF}}^{\alpha}$ and $l_{\mathrm{corr}}^{\alpha}$, respectively, to discuss the size of the localization regions determined by the automated DC method. $l_{\mathrm{local}}^{\mathrm{SCF},\alpha}$ (or $l_{\mathrm{local}}^{\mathrm{corr},\alpha}$) was defined as half of the maximum atom-pair distance in the HF (or MP2) localization region for subsystem α. The two-electron AO integrals, $(\mu\nu|\lambda\sigma)$, are treated in so-called "direct algorithm" manner, i.e., the same integrals are calculated repeatedly for every subsystem.

4.4.2. Accuracy and computational time of the present method

I first applied the present automated DC-MP2 method to a cubic system containing 100 water molecules with weight density of 1.0 g cm^{-3}. Each water molecule was adopted as a central region in the DC calculation. To assess the performance of the automated DC-MP2 calculation, the entire system was selected as the initial localization region for every subsystem in the DC-MP2 calculation. Fig. 4-2 shows the estimated MP2 energy contributions from buffer atom B (e_B^α) with respect to its distance from the O atom in the central region. The blue plot represents the value for B being an H atom, and the red plot that for B being an O atom. The estimated energy contribution decays exponentially as the distance from the central region increases. The slight difference in the slope for H and O atoms in Fig. 4-2 is probably due to the fact that the summation over AOs at the buffer atom in Eq. (4.9) runs for the virtual orbital, that is, the charge-transfer excited configurations from O atoms in donor water to H atoms in acceptor water are more significant than those from acceptor to donor. This behavior was also confirmed for the water dimer system using the intermolecular interaction energy decomposition with the local PNO method.[59] Note that the estimated energies in Fig. 4-2 for the interatomic distance of 2-3 Å are up to several hundred E_h, which are significantly larger than the total MP2 energy of ~19 E_h. This is because that the estimated energy (e_B^α) is derived as the upper limit of the atomic MP2 energy contribution. From the following section, the energy threshold in the automated DC-MP2 method, $e_\mathrm{thresh}^\mathrm{corr}$, was set to 0.1 μ$E_\mathrm{h}$ unless otherwise noted.

Fig. 4-2. Estimated atomic MP2 energy contributions with respect to the interatomic distance. The blue plots represent the estimated MP2 energy of H atom and the red plots represent of O atom in the buffer region.

Next, the dependence of the computational time of e_B^α on the system-size was examined, as shown in Fig. 4-3. These were measured using a computer node equipped with two Intel Xeon Gold 5118 CPUs (12 cores, 2.30 GHz) and the average of three measurements was plotted. The initial sizes of the inner and outer buffer regions in the automated DC-HF calculation were set to $r_b^{in} = 4.5\,\text{Å}$ and $r_b^{out} = 5.5\,\text{Å}$, respectively. The scaling analysis with the double logarithmic plot indicates that the computational

time for the evaluation of e_B^α scales as $O(N_{\mathrm{water}}^{1.5})$, which maintains an almost linear-scaling behavior.

Fig. 4-3. System-size dependence of the CPU time of the evaluation of e_B^α for the model system containing N_{water} water molecules. The initial buffer size for the DC-HF calculations was fixed to $r_{\mathrm{b}}^{\mathrm{in}} = 4.5\,\text{Å}$ and $r_{\mathrm{b}}^{\mathrm{out}} = 5.5\,\text{Å}$.

The accuracy and the computational time of the automated DC-MP2 method were investigated for the cubic water system. Table 4-1 shows the energy-threshold ($e_{\mathrm{thresh}}^{\mathrm{corr}}$) dependence of the DC-MP2 correlation energy. Following Section, each water molecule was adopted as a central region and the entire system was selected as the initial localization region. The average and standard deviation of major axis radii ($\langle l_{\mathrm{local}}^{\mathrm{corr}} \rangle$ and $\sigma[l_{\mathrm{local}}^{\mathrm{corr}}]$, respectively) are also given in Table 4-1. For $e_{\mathrm{thresh}}^{\mathrm{corr}} = 100\,\mu E_{\mathrm{h}}$, the actual

correlation energy error per atom is 18.37 μE_h, which is sufficiently smaller than $e_{\mathrm{thresh}}^{\mathrm{corr}}$. It should be noted that the MP2 energy error decreases systematically as $e_{\mathrm{thresh}}^{\mathrm{corr}}$ decreases, while the dependence is not proportional but rather logarithmic to $e_{\mathrm{thresh}}^{\mathrm{corr}}$. As with the e_{thresh} parameter in automated DC-SCF method,[44] the smaller $e_{\mathrm{thresh}}^{\mathrm{corr}}$ parameter leads to a larger localization region, which can be confirmed from the average of the major axis radii of all localization regions, $\langle l_{\mathrm{local}}^{\mathrm{corr}} \rangle$. Interestingly, the standard deviation of the major axis radii, $\sigma[l_{\mathrm{local}}^{\mathrm{corr}}]$, also tends to increase systematically as $e_{\mathrm{thresh}}^{\mathrm{corr}}$ decreases, except for $e_{\mathrm{thresh}}^{\mathrm{corr}} = 0.1\mu E_{\mathrm{h}}$. This fact suggests that the present scheme can effectively aid the selection of the appropriate buffer region for each subsystem in the DC-MP2 calculation.

Table 4-1. $e_{\mathrm{thresh}}^{\mathrm{corr}}$ dependences of the DC-MP2 correlation energy and the major axis radius for 100 water cluster system.

$e_{\mathrm{thresh}}^{\mathrm{corr}}$ /μE_h	$E_{\mathrm{corr}}^{(2)}$ /E_h	(Diff.) /μE_h atom^{-1}	$\langle l_{\mathrm{local}}^{\mathrm{corr},\alpha} \rangle$ /Å	$\sigma\left[l_{\mathrm{local}}^{\mathrm{corr},\alpha}\right]$ /Å
100.000	–19.102140	(+18.37)	5.596	0.569
10.000	–19.103891	(+12.54)	6.038	0.589
1.000	–19.104999	(+8.84)	6.380	0.677
0.100	–19.105661	(+6.64)	6.761	0.659
0.010	–19.106160	(+4.97)	7.131	0.681
Standard-MP2	–19.107652			

Next, I examined the combination of the present automated DC-MP2 method with the automated DC-HF calculation. Table 4-2 shows the dependence of the automated DC-MP2 energy on the initial DC-HF inner and outer buffer sizes, $r_{\mathrm{b}}^{\mathrm{in}}$ and $r_{\mathrm{b}}^{\mathrm{out}}$, the definitions of which are given in our previous paper.[44] The averages ($\langle l_{\mathrm{local}}^{\mathrm{HF}} \rangle$ and $\langle l_{\mathrm{local}}^{\mathrm{corr}} \rangle$) and the standard deviations ($\sigma[l_{\mathrm{local}}^{\mathrm{HF}}]$ and $\sigma[l_{\mathrm{local}}^{\mathrm{corr}}]$) of the major axis radii among all localization regions in the DC-HF and DC-MP2 calculations are also shown. Similar to the results in Ref. 44, the DC-HF energy error is sufficiently small and almost independent of the initial DC-HF buffer region. Subsequently, the DC-MP2 energy error is almost constant (~8.5 μE_{h} atom^{-1}). The average radius of the DC-HF localization region, $\langle l_{\mathrm{local}}^{\mathrm{HF}} \rangle$, is 7.0–7.2 Å, which is larger than the average radius, 6.761 Å, of the DC-MP2 localization region for $e_{\mathrm{thresh}}^{\mathrm{corr}} = 0.1 \mu E_{\mathrm{h}}$ given in Table 4-1. A smaller initial DC-HF buffer size leads to a larger $\langle l_{\mathrm{local}}^{\mathrm{HF}} \rangle$, as was also confirmed in the previous study.[44] When combined with the automated DC-HF method, $\langle l_{\mathrm{local}}^{\mathrm{corr}} \rangle$ becomes smaller than its value when the initial localization region is set to be the entire system. Similarly, $\sigma[l_{\mathrm{local}}^{\mathrm{corr}}]$ is approximately 0.14 Å smaller than $\sigma[l_{\mathrm{local}}^{\mathrm{HF}}]$.

Table 4-2. Initial DC-HF buffer-size dependence of the automated DC-MP2 correlation energy and the major axis radius for 100 water cluster system. The energy threshold in the automated DC-HF calculation is 0.1 μE_h.

r_b^{in} /Å	r_b^{out} /Å	HF Energy /E_h	MP2 Energy /E_h	(Diff.) /μE_h atom^{-1}	$\langle l_{local}^{SCF,\alpha} \rangle$ /Å	$\sigma[l_{local}^{SCF,\alpha}]$ /Å	$\langle l_{local}^{corr,\alpha} \rangle$ /Å	$\sigma[l_{local}^{corr,\alpha}]$ /Å
3.5	4.5	–7601.504443	–19.105142	(+8.37)	7.233	0.840	6.564	0.744
4.0	5.0	–7601.504613	–19.105141	(+8.37)	7.238	0.903	6.538	0.767
4.5	5.5	–7601.504342	–19.105031	(+8.74)	7.161	0.885	6.522	0.743
5.0	6.0	–7601.504417	–19.105000	(+8.84)	7.161	0.905	6.480	0.726
5.5	6.5	–7601.504467	–19.105185	(+8.23)	7.000	0.806	6.427	0.682
Standard		–7601.504673	–19.107652					

Next, the proposed method was applied to a covalently bound system, namely, the chignolin protein with 10 amino acids. The geometry of chignolin was obtained from the protein data bank (PDBID: 1UAO). Hydrogen atoms were added using the Discovery Studio 2017 R2 software.[60] In the DC calculation, the entire system was divided between the carbonyl C and α-C atoms, and each of the divided systems was treated as a central region. Table 4-3 shows the $e_{\mathrm{thresh}}^{\mathrm{corr}}$ dependence of the DC-MP2 energy for chignolin. The entire system was selected as the initial localization region for every subsystem in the DC-MP2 calculation. For $e_{\mathrm{thresh}}^{\mathrm{corr}} = 100\ \mu E_{\mathrm{h}}$, the actual correlation energy error per atom is 2.82 μE_{h}, which is sufficiently smaller than $e_{\mathrm{thresh}}^{\mathrm{corr}}$. As was also confirmed in the case of the water system, the MP2 energy error decreases systematically as $e_{\mathrm{thresh}}^{\mathrm{corr}}$ decreases. Again, the dependence of the error on $e_{\mathrm{thresh}}^{\mathrm{corr}}$ is rather logarithmic. The smaller $e_{\mathrm{thresh}}^{\mathrm{corr}}$ leads to the larger $\langle l_{\mathrm{local}}^{\mathrm{corr}} \rangle$, while it leads to the smaller $\sigma[l_{\mathrm{local}}^{\mathrm{corr}}]$, contrary to the case of water system. Comparing Table 4-3 with Table 4-1, $\langle l_{\mathrm{local}}^{\mathrm{corr}} \rangle$ of chignolin is about 1.0 Å larger than that of the water system for the same $e_{\mathrm{thresh}}^{\mathrm{corr}}$ parameter, reflecting the delocalized electronic nature in the covalently bound system.

Table 4-3. $e_{\mathrm{thresh}}^{\mathrm{corr}}$ dependences of the DC-MP2 correlation energy and the major axis radius for chignolin.

$e_{\mathrm{thresh}}^{\mathrm{corr}}$ /μE_{h}	$E_{\mathrm{corr}}^{(2)}$ /E_{h}	(Diff.) /μE_{h} atom^{-1}	$\left\langle l_{\mathrm{local}}^{\mathrm{corr},\,\alpha} \right\rangle$/Å	$\sigma\left[l_{\mathrm{local}}^{\mathrm{corr},\,\alpha} \right]$/Å
100.000	–11.194529	(+2.82)	7.003	0.671
10.000	–11.194689	(+1.67)	7.185	0.598
1.000	–11.194770	(+1.08)	7.530	0.614
0.100	–11.194828	(+0.66)	7.629	0.597
0.010	–11.194847	(+0.52)	7.726	0.564
Standard-MP2	–11.194919			

Next, I combined this with the automated DC-HF calculation. Table 4-4 shows the dependence of the DC-MP2 energy on the initial DC-HF buffer size. The automated DC-HF energy error for chignolin is smaller than that for the water system and almost independent of the initial DC-HF buffer region, while the radius of the DC-HF localization region (~7.5 Å) is about 1 Å greater than for the water system (~6.5 Å). Subsequently, the DC-MP2 energy error is also almost constant (~0.7 μE_{h} atom^{-1}). For this small protein system, in contrast to the result in Table 4-2 for the water system, the standard deviation of the sizes of the localization regions for the MP2 calculation is larger than that for the HF calculation. This is because the entire size of the chignolin system is so small that the localization region for every subsystem is close to the entire system. The present method was also tested on the β-strand glycine oligomer $(\mathrm{GLY})_{20}$, and the result

of the calculation are given in Table 4-5. In Table 4-5, the DC-MP2 calculations with different $e_{\mathrm{thresh}}^{\mathrm{corr}}$ were performed to confirm that the present automated DC-MP2 energy error depends primarily on $e_{\mathrm{thresh}}^{\mathrm{corr}}$ and hardly on the initial buffer radii. For this stretched system, the standard deviation of the localization region sizes for the MP2 calculation is smaller than that for the HF calculation, while the energy error is similar to the result in Table 4-4. As well as the case of water system, the smaller $e_{\mathrm{thresh}}^{\mathrm{corr}}$ leads to the larger $\left\langle l_{\mathrm{local}}^{\mathrm{corr}} \right\rangle$ and $\sigma[l_{\mathrm{local}}^{\mathrm{corr}}]$.

Table 4-4. Initial DC-HF buffer-size dependence of the automated DC-MP2 correlation energy and the major axis radius for chignolin. The energy threshold in the automated DC-HF calculation is 0.1 μE_h.

r_b^{in} /Å	r_b^{out} /Å	HF Energy /E_h	MP2 Energy /E_h	(Diff.) /μE_h atom^{-1}	$\langle l_{local}^{SCF,\alpha} \rangle$ /Å	$\sigma[l_{local}^{SCF,\alpha}]$ /Å	$\langle l_{local}^{corr,\alpha} \rangle$ /Å	$\sigma[l_{local}^{corr,\alpha}]$ /Å
3.5	4.5	–3799.529116	–11.194860	(+0.43)	8.248	0.550	7.606	0.620
4.0	5.0	–3799.528978	–11.194810	(+0.79)	8.121	0.626	7.606	0.620
4.5	5.5	–3799.528977	–11.194825	(+0.68)	8.174	0.562	7.582	0.598
5.0	6.0	–3799.528978	–11.194820	(+0.71)	8.304	0.625	7.606	0.620
5.5	6.5	–3799.528978	–11.194814	(+0.76)	8.151	0.508	7.606	0.620
Standard		–3799.528980	–11.194919					

Table 4-5. Initial DC-HF buffer-size dependence of the automated DC-MP2 energy and the major axis radius for the β-strand glycine oligomer $(GLY)_{20}$. The energy threshold in the automated DC-HF calculation is set to 0.1 μE_h and that in the automated DC-MP2 calculation is set to 100 and 0.1 μE_h.

r_b^{in} /Å	r_b^{out} /Å	HF Energy /E_h	MP2 Energy /E_h	(Diff.) /μE_h atom^{-1}	$\langle l_{local}^{SCF,\alpha} \rangle$ /Å	$\sigma[l_{local}^{SCF,\alpha}]$ /Å	$\langle l_{local}^{corr,\alpha} \rangle$ /Å	$\sigma[l_{local}^{corr,\alpha}]$ /Å
$e_{thresh}^{corr} = 100\ \mu E_h$								
3.5	4.5	–4211.847790	–11.932344	(+1.01)	10.469	1.534	7.488	0.910
4.0	5.0	–4211.847790	–11.932345	(+1.00)	10.501	1.450	7.488	0.910
4.5	5.5	–4211.847790	–11.932345	(+1.01)	10.469	1.534	7.488	0.910
5.0	6.0	–4211.847790	–11.932345	(+1.01)	10.469	1.534	7.488	0.910
5.5	6.5	–4211.847790	–11.932345	(+1.01)	10.469	1.534	7.488	0.910
$e_{thresh}^{corr} = 0.1\ \mu E_h$								
3.5	4.5	–4211.847790	–11.932431	(+0.40)	10.469	1.534	8.603	1.136
4.0	5.0	–4211.847790	–11.932433	(+0.39)	10.501	1.450	8.603	1.136
4.5	5.5	–4211.847790	–11.932432	(+0.40)	10.469	1.534	8.603	1.136
5.0	6.0	–4211.847790	–11.932432	(+0.39)	10.469	1.534	8.603	1.136
5.5	6.5	–4211.847790	–11.932432	(+0.40)	10.469	1.534	8.603	1.136
Standard		–4211.847819	–11.932489					

Finally, the present method was applied to the conjugated polyacetylene chain $C_{2n}H_{2n+2}$, shown in Fig. 4-4. All atoms were placed in a plane and the C–C, C=C, and C–H bond lengths were fixed at 1.462, 1.357, and 1.096 Å, respectively. Each C_2H_2 (or C_2H_3 for edges) unit divided at the C–C single bond was treated as a central region. Table 4-6 shows the system-size dependence of the standard and DC-MP2 energies. For the automated DC calculations, the initial sizes of the inner and outer buffer regions in the automated DC-HF calculation were set to $r_b^{in} = 5.0$ Å and $r_b^{out} = 6.5$ Å, respectively. To avoid division of the localization region at C=C double bond, each C_2H_2 (or C_2H_3) unit was treated as one piece, that is, a unit was extracted from the DC-MP2 localization region only when all the estimated MP2 correlation energies, $\{e_B^\alpha\}$, for the atoms in the unit were smaller than the threshold, e_{thresh}^{corr} (analogous to the BUFTYP=RADSUB option of $DANDC input group in GAMESS program). The DC-MP2 energy error per atom is almost constant for $n \geq 30$. It was demonstrated that the correlation energy error can be controlled with the present method, even for conjugated systems.

Fig. 4-4. Structure of polyacetylene chain system, $C_{2n}H_{2n+2}$.

Table 4-6. The system-size dependence of the MP2 electron correlation energy in the standard MP2 and automated DC-MP2 calculations for polyacetylene chain system, $C_{2n}H_{2n+2}$

# of C atoms	Standard-MP2	Auto.DC-MP2	
	Energy /E_h	Energy /E_h	(Diff.) /μE_h atom^{-1}
10	–1.266346	–1.266346	(+0.00)
20	–2.533020	–2.532799	(+5.25)
30	–3.799773	–3.799303	(+7.58)
40	–5.066529	–5.065806	(+8.81)
50	–6.333285	–6.332309	(+9.56)
60	–7.600041	–7.598813	(+10.06)
70	–8.866797	–8.865319	(+10.40)
80	–10.133553	–10.131822	(+10.68)
90	–11.400309	–11.398327	(+10.89)
100	–12.667065	–12.664831	(+11.06)

For this conjugated system, the dependence of the computational time on the system size was also examined, as shown in Fig. 4-5. The computational time for the MP2 calculation was measured using a computer node equipped with two Intel Xeon E5–2667 CPUs (8 cores, 3.20 GHz), and the average of three measurements was plotted. For comparison, the time required for the standard MP2 calculation was also plotted. The CODE=IMS program[61] specified in the $MP2 input group implemented in the GAMESS package was used. The automated DC-MP2 calculation shows a faster computational time than that of the standard MP2 calculation for $n \geq 30$. The scaling analysis with the double logarithmic plot for $n \geq 40$ indicates that the computational time for the standard MP2 scales as $O(n^{2.5})$, while that for the present automated DC-MP2 method scales as $O(n^{1.1})$. It is confirmed that the linear-scaling behavior of the DC-MP2 method is preserved even with the present automation scheme. In the present paper, the scaling analysis is examined for one-dimensional systems.

Fig. 4-5. System-size dependence of the Wall-clock time of the standard MP2 and the automated DC-MP2 calculations for polyacetylene chain system containing *n* carbon atoms $C_{2n}H_{2n+2}$. Black dashed line: standard-MP2; solid red line: automated DC-MP2.

The scaling analysis was also conducted for three-dimensional water cluster systems. Figure 4-6 shows the dependence of the wall-clock computational time for the DC-MP2 calculation on the number of water molecules, N_{water}. The times were measured using a computer node equipped with two Intel Xeon Gold 5118 CPUs (12 cores, 2.30 GHz), and the average of three measurements was plotted. The initial sizes of the inner and outer buffer regions in the automated DC-HF calculation were set to $r_{\mathrm{b}}^{\mathrm{in}} = 4.5\,\text{Å}$ and $r_{\mathrm{b}}^{\mathrm{out}} = 5.5\,\text{Å}$, respectively. The energy threshold in the automated DC-MP2 method, $e_{\mathrm{thresh}}^{\mathrm{corr}}$, was set to 10 μ$E_{\mathrm{h}}$. The scaling analysis with the double logarithmic plot indicates that the

computational time for the present automated DC-MP2 method scales as $O(N_{\mathrm{water}}^{1.6})$, which indicates that the present method also achieves near-linear scaling computational time even for three-dimensional systems.

Fig. 4-6. System-size dependence of the Wall-clock time of the automated DC-MP2 calculations for the model system containing N_{water} water molecules. The initial sizes of the inner and outer buffer regions in the automated DC-HF calculation were set to $r_{\mathrm{b}}^{\mathrm{in}} = 4.5\,\text{Å}$ and $r_{\mathrm{b}}^{\mathrm{out}} = 5.5\,\text{Å}$, respectively. The energy threshold in the automated DC-MP2 method, $e_{\mathrm{thresh}}^{\mathrm{corr}}$, was set to 10 μE_{h}.

4.5. Concluding remarks

In this study, I have proposed an automatic determination scheme for the buffer region in the DC-MP2 calculation. This method is based on a subsystem MP2 correlation energy contribution from each atom in the buffer region, which is estimated with the help of the AO-Laplace MP2 method and the Schwarz inequality. Because the appropriate size of the buffer region in the DC-MP2 calculation can be smaller than that in the DC-HF calculation, as suggested in a previous paper,[36] the present scheme reduces the buffer region from the prior DC-HF calculation. I applied the present method to a 100 water cluster system and the chignolin system, and confirmed that the estimated DC-MP2 energy error can be systematically reduced as the energy threshold, $e_{\mathrm{thresh}}^{\mathrm{corr}}$, decreases. I also confirmed that the linear-scaling behavior of the DC-MP2 method is preserved even with the present automation scheme, from a calculation of linear polyene system.

Since the MP2 amplitude is known to provide a good guess for the CC method in many cases, the proposed automation scheme is straightforwardly applicable to the DC-CC method.[37-39] Improvements in the accuracy of the correlation energy contributions from buffer atoms are also desirable, especially for delocalized systems. The use of the inequality test proposed by Thompson *et al.* [62] instead of the Schwarz inequality would be one way to provide this improvement.

4.6. Appendix

Here, I propose a scheme to partition the standard MP2 energy into atom-pair (bond) contributions to demonstrate the local character of the MP2 correlation. The scheme is related to the bond EDA proposed by Nakai and coworkers.[63]

The MP2 correlation energy can be divided into contributions from the atomic pair, expressed as

$$\Delta E_{\mathrm{corr}}^{(2)} = \sum_{A,B} \Delta E_{\mathrm{corr}}^{(2)AB} , \tag{4.19}$$

$$\Delta E_{\mathrm{corr}}^{(2)AB} = \sum_{i,j}^{\mathrm{occ}} \sum_{a,b}^{\mathrm{vir}} \sum_{\mu \in A} \sum_{\nu \in B} \frac{C_{\mu i} C_{\nu a} (\mu\nu \mid jb)}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} [2(ai \mid bj) - (aj \mid bi)] . \tag{4.20}$$

Here, I have adopted the electron coordinate separation instead of the electron pair separation[52-54] to exploit the local nature of the MP2 correlation. This form is also consistent with $\Delta E_B^{\alpha(2)}$, Eq. (4.9). Note that $\Delta E_{\mathrm{corr}}^{(2)AB}$ is different from $\Delta E_{\mathrm{corr}}^{(2)BA}$ because atoms A and B in $\Delta E_{\mathrm{corr}}^{(2)AB}$ are associated with the occupied and virtual orbitals, respectively. The atom-pair MP2 correlation energies, $\Delta E_{\mathrm{corr}}^{(2)AB}$, were evaluated for $C_{30}H_{32}$ polyene system with 6-31G(d) basis set. Fig. 4-7 shows the dependence of $\Delta E_{\mathrm{corr}}^{(2)AB}$ on the distance between the A and B atoms, r. Different color plots indicate different combinations of elements for atoms A and B. Overall, the atom-pair contribution decreases exponentially with respect to the distance, although that for $\Delta E_{\mathrm{corr}}^{(2)\mathrm{CC}}$ has small hump around r = 20 Å. Reflecting the small number of correlated electrons around H atom, $\Delta E_{\mathrm{corr}}^{(2)\mathrm{HH}}$ has the smallest contribution at the same distance r. $\Delta E_{\mathrm{corr}}^{(2)\mathrm{CH}}$ is larger

than $\Delta E_{\mathrm{corr}}^{(2)\mathrm{HC}}$, probably due to more significant contribution of the charge-transfer excitation configurations from the electron-rich C atoms to the electron-deficient H atoms, similar to the discussion on the water system (see Section 4.4.2).

Fig. 4-7. The absolute atomic pair MP2 correlation energy contribution with respect to the interatomic distance. The circle, pentagon, square and triangle plots represent the MP2 correlation energy contribution for C-C, C-H, H-C and H-H atomic pairs, respectively.

4.7. Reference

[1] A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover, New York, **1996**.

[2] T. Tsuneda, *Density Functional Theory in Quantum Chemistry*, Springer, Tokyo, **2014**.

[3] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618.

[4] T. Helgaker, P. Jørgensen, J. Olsen, *Molecular Electronic-Structure Theory*, John Wiley & Sons, New York, **2000**.

[5] F. Jensen, *Introduction to Computational Chemistry*, *Third Edition*, John Wiley & Sons, New York, **2017**.

[6] S. Goedecker, *Rev. Mod. Phys.* **1999**, *71*, 1085.

[7] S. Y. Wu and C. S. Jayanthi, *Phys. Rep.* **2002**, *358*, 1.

[8] R. Zaleśny, M. G. Papadopoulos, P. G. Mezey, J. Leszczynski, *Linear-Scaling Techniques in Computational Chemistry and Physics*; Springer: Dordrecht, **2011**.

[9] D. R. Bowler and T Miyazaki, *Rep. Prog. Phys.* **2012**, *75*, 036503.

[10] M. S. Gordon, D. G. Fedorov, S. R. Pruitt and L. V. Slipchenko, *Chem. Rev.* **2012**, *112*, 632.

[11] N. Sahu, S. R. Gadre, *Acc. Chem. Res.* **2014**, *47*, 2739.

[12] N. Sahu, S. D. Yeole, S. R. Gadre, *J. Chem. Phys.* **2013**, *138*, 104101.

[13] V. Ganesh, R. K. Dongare, P. Balanarayan, S. R. Gadre, *J. Chem. Phys.* **2006**, *125*, 104109.

[14] W. Li, S. Li, Y. Jiang, *J. Phys. Chem. A* **2007**, *111*, 2193.

[15] S. Li, W. Li, J. Ma, *Acc. Chem. Res.* **2014**, *47*, 2712.

[16] Z. Ni, W. Li, S. Li, *J. Comput. Chem.* **2019**, *40*, 1130.

[17] Z. Ni, Y. Guo, F. Neese, W. Li, S. Li, arXiv:2008:02057 [physics.chem-ph].

[18] D. G. Fedorov, K. Kitaura, The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems, CRC Press, Boca Raton, **2009**.

[19] D. G. Fedorov, K. Kitaura, *J. Chem. Phys.* **2004**, *120*, 6832.

[20] Y. Nishimoto and D. G. Fedorov, *J. Comput. Chem*. **2017**, *38*, 406.

[21] T. Nakano, Y. Mochizuki, K. Yamashita, C. Watanabe, K. Fukuzawa, K. Segawa, Y. Okiyama, T. Tsukamoto, S. Tanaka, *Chem. Phys. Lett.* **2012**, *523*, 128.

[22] H.-J. Werner, G. Knizia, C. Krause, M. Schwilk and M. Dornbach, *J. Chem. Theory Comput.* **2015**, *11*, 484.

[23] Q. Ma and H.-J. Werner, J. *J. Chem. Theory Comput.* **2015**, *11*, 5291.

[24] P. Baudin, P. Ettenhuber, S. Reine, K. Kristensen and T. Kjærgaard, *J. Chem. Phys.* **2016**, *144*, 054102.

[25] T. Kjærgaard, P. Baudin, D. Bykov, K. Kristensen and P. Jørgensen, *WIREs Comput. Mol. Sci.* **2017**, *7*, e1319.

[26] W. Yang, T.-S. Lee, *J. Chem. Phys.* **1995**, *103*, 5674.

[27] M. Kobayashi and H. Nakai, In *Linear-Scaling Techniques in Computational Chemistry and Physics,* Chapter 5; R. Zaleśny, M. G. Papadopoulos, P. G. Mezey, J. Leszczynski, Eds., Springer, Dordrecht, **2011**, pp. 97–127.

[28] T. Akama, M. Kobayashi, H. Nakai, *J. Comput. Chem.* **2007**, *28*, 2003.

[29] H. Nishizawa, Y. Nishimura, M. Kobayashi, S. Irle , H. Nakai, *J. Comput. Chem.* **2016**, *37*, 1983.

[30] H. Nakai, A. W. Sakti, Y. Nishimura, *J. Phys. Chem. B* **2016**, *120*, 217.

[31] N. Komoto, T. Yoshikawa, J. Ono, Y. Nishimura, H. Nakai, *J. Chem. Theory Comput.* **2019**, *15*, 1719.

[32] N. Komoto, T. Yoshikawa, Y. Nishimura, H. Nakai, *J. Chem. Theory Comput.* **2020**, *16*, 2369.

[33] M. Kobayashi, Y. Imamura, H. Nakai, *J. Chem. Phys.* **2007**, *127*, 074103.

[34] M. Kobayashi, T. Taketsugu, *Theor. Chem. Acc.* **2015**, *134*, 107.

[35] M. Kobayashi, T. Akama , H. Nakai, *J. Chem. Phys.* **2006**, *125*, 204106.

[36] M. Kobayashi, H. Nakai, *Int. J. Quantum Chem.* **2009**, *109*, 2227.

[37] M. Kobayashi and H.Nakai, *J. Chem. Phys.* **2008**, *129*, 044103.

[38] M. Kobayashi and H. Nakai, *J. Chem. Phys*. **2009**, *131*, 114108.

[39] T. Yoshikawa, M. Kobayashi, H. Nakai, *Int. J. Quantum Chem.* **2013**, *113*, 218.

[40] M. Kobayashi, T. Kunisada, T. Akama, D. Sakura, H. Nakai, *J. Chem. Phys.* **2011**, *134*, 034105.

[41] M. Kobayashi, H. Nakai, *J. Chem. Phys.* **2013**, *138*, 044102.

[42] M. Kobayashi, T. Taketsugu, *Chem. Lett.* **2016**, *45*, 1268.

[43] T. Yoshikawa, T. Doi, H. Nakai, *Chem. Phys. Lett.* **2019**, *725*, 18.

[44] M. Kobayashi, T. Fujimori, T. Taketsugu, *J. Comput. Chem.* **2018**, *39*, 909.

[45] S. L. Dixon, K. M. Merz Jr., *J. Chem. Phys.* **1997**, *107*, 879.

[46] J. Almlöf, *Chem. Phys. Lett.* **1991**, *181*, 319.

[47] M. Häser, J. Almlöf, *J. Chem. Phys.* **1992**, *96*, 489.

[48] M. Häser, *Theor. Chim. Acta* **1993**, *87*, 147.

[49] P. Y. Ayala, G. E. Scuseria, *J. Chem. Phys.*, **1999**, *110*, 3660.

[50] M. Kobayashi, H. Nakai, *Chem. Phys. Lett.* **2006**, *420*, 250.

[51] H. Nakai, *Chem. Phys. Lett.* **2002**, *363*, 73.

[52] P. Pulay, *Chem. Phys. Lett.* **1983**, *100*, 151.

[53] S. Saebo, P. Pulay, *Annu. Rev. Phys. Chem.* **1993**, *44*, 213.

[54] Y. Guo, U. Becker, F. Neese, *J. Chem. Phys.* **2018**, *148*, 124117.

[55] Y. Jung, R. C. Lochan, A. D. Dutoi, M. Head-Gordon, *J. Chem. Phys.* **2004**, *121*, 9793.

[56] M. S. Gordon, M. W. Schmidt, In *Theory and Applications of Computational Chemistry: The First Forty Years*, C. Dykstra, G. Frenking, K. Kim, G. Scuseria, Eds., Elsevier, Amsterdam, **2005**, pp 1167–1189.

[57] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery Jr., *J. Comput. Chem.* **1993**, *14*, 1347.

[58] P. C. Hariharan, J. A. Pople, *Chem. Phys. Lett.*, **1972**, *16*, 217.

[59] W. B. Schneider, G. Bistoni, M. Sparta, M. Saitow, C. Riplinger, A. A. Auer, F. Neese, *J. Chem. Theory Comput.* **2016**, *12*, 4778.

[60] Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016.

[61] K. Ishimura, P.Pulay, S. Nagase, *J. Comput. Chem.* **2007**, *28*, 2034.

[62] T. H. Thompson, C. Ochsenfeld, *J. Chem. Phys.* **2017**, *147*, 144101.

[63] H. Nakai, Y. Kikuchi, *J. Theor. Comput. Chem.* **2005**, *4*, 317.

5. Approach to automatic error control for DC-HF energy gradient calculation

5.1. Introduction

Equilibrium and transition structures are one of the most important information to the chemical reaction analyses. Therefore, the quantum chemical calculation method for predicting such optimized molecular structures, in which the entire system can be considered, is needed also in large-scale systems. The structure optimization calculation with fragment-based electronic structure calculation methods have been developed. However, in these methods, the energy error is introduced by the fragmentation. Recently, an automated energy error control scheme in divide and conquer (DC) Hartree-Fock (HF) calculation was proposed in a previous paper by Kobayashi *et al*.[1] In this method, the size of the buffer region, which considerably affects the accuracy of DC-HF calculation, can be controlled properly in the SCF cycles. In this Chapter, I'm aiming to extend this automatic control method to the energy gradient calculation that is essential for the structure optimization calculation and the energy gradient error estimation with Schwarz and Thompson inequalities are proposed.

5.2. DC-HF energy gradient expression

I briefly summarize the DC-HF energy gradient expression. Firstly, the DC-HF procedure is mentioned. In this Section, the Gleek letters $\{\mu, \nu, \ldots\}$ refer to atomic orbitals (AOs). $\{p, q, \ldots\}$ refer to molecular orbitals (MOs) expanded by the linear combination of AOs, which is described as $\psi_p = \sum_\mu C_{\mu p}\phi_\mu$. In the DC method,[2-4] the entire system is divided into several subsystems. Each subsystem of the DC method is composed of the central region and the buffer region. The central region is separated from each other and the buffer region is added for each central region. In the DC method, the density matrix for the entire system can be obtained approximately by sum of the subsystem density matrix, expressed as

$$D_{\mu\nu} \approx D_{\mu\nu}^{\mathrm{DC}} = \sum_{\alpha}^{\text{subsystem}} P_{\mu\nu}^{\alpha} D_{\mu\nu}^{\alpha} . \tag{5.1}$$

Shown in Eq. (5.1), the partition matrix, $\mathbf{P}^{\alpha}$, is introduced to average the value of the buffer regions because the buffer regions overlap each other.

$$P_{\mu\nu}^{\alpha} = \begin{cases} 1 & \text{for } \mu \in \boldsymbol{S}(\alpha) \wedge \nu \in \boldsymbol{S}(\alpha) \\ 1/2 & \text{for } \left(\mu \in \boldsymbol{S}(\alpha) \wedge \nu \in \boldsymbol{B}(\alpha)\right) \text{or vice versa} \\ 0 & \text{otherwise} \end{cases} , \tag{5.2}$$

where $\boldsymbol{S}(\alpha)$ is the set of AOs in the central region and $\boldsymbol{B}(\alpha)$ is the set of AOs in the buffer region.

The density matrix for the subsystem α is constructed with the Fermi function: $f_\beta(x) = [1+\exp(-\beta x)]^{-1}$ (β is the inverse temperature), which is described as

$$D_{\mu\nu}^{\alpha} = \sum_p f_\beta\left(\varepsilon_{\mathrm{F}} - \varepsilon_p^{\alpha}\right) C_{\mu p}^{\alpha} C_{\nu p}^{\alpha} , \tag{5.3}$$

$\{\boldsymbol{C}^{\alpha}\}$ and $\{\boldsymbol{\varepsilon}^{\alpha}\}$ represent the MO coefficients and orbital energies for subsystem α. These are obtained from the Roothaan equation for each subsystem, which is described as

$$\mathbf{F}^{\alpha}[\mathbf{D}^{\mathrm{DC}}]\mathbf{C}_{p}^{\alpha} = \boldsymbol{\varepsilon}_{p}^{\alpha}\mathbf{S}^{\alpha}\mathbf{C}_{p}^{\alpha}, \tag{5.4}$$

where $\mathbf{F}^{\alpha}$ and $\mathbf{S}^{\alpha}$ are the Fock matrix and the overlap matrix for subsystem α. ε_{F} represent the universal Fermi level and is determined by the following constraint for the total number of electrons, n_{e}:

$$n_{\mathrm{e}} = 2\mathrm{Tr}\left[\mathbf{D}^{\mathrm{DC}}\mathbf{S}\right]. \tag{5.5}$$

In the DC method, the total energy for the entire system is calculated with converged Fock and density matrices, which is expressed as

$$E[\mathbf{D}^{\mathrm{DC}}] = \mathrm{Tr}\left[\mathbf{D}^{\mathrm{DC}}(\mathbf{H}^{\mathrm{core}} + \mathbf{F}[\mathbf{D}^{\mathrm{DC}}])\right], \tag{5.6}$$

where $\mathbf{H}^{\mathrm{core}}$ is the core Hamiltonian matrix.

The standard HF energy gradient[5] is expressed as

$$\begin{aligned}\frac{\partial E_{\mathrm{HF}}}{\partial Q} = &\sum_{\mu\nu} D_{\mu\nu}\left\{\frac{\partial H_{\nu\mu}^{\mathrm{core}}}{\partial Q} + \sum_{\lambda\sigma} D_{\lambda\sigma}\left[2\frac{\partial(\nu\mu\,|\,\lambda\sigma)}{\partial Q} - \frac{\partial(\nu\lambda\,|\,\mu\sigma)}{\partial Q}\right]\right\}, \\ &+\sum_{\mu\nu}\frac{\partial D_{\mu\nu}}{\partial Q}F_{\nu\mu}\end{aligned} \tag{5.7}$$

where Q is the atomic coordinate. In Eq. (5.7), the first term is called the Hellmann-Feynman force and the second term is called Pulay force. Because MOs are normalized, the following equation is constructed.

$$\mathbf{C}^{\dagger}\mathbf{S}\mathbf{C} = \mathbf{1} \tag{5.8}$$

Therefore, the derivative of Eq. (5.8) is described as

$$\frac{\partial \mathbf{C}^{\dagger}}{\partial Q}\mathbf{S}\mathbf{C} + \mathbf{C}^{\dagger}\frac{\partial \mathbf{S}}{\partial Q}\mathbf{C} + \mathbf{C}^{\dagger}\mathbf{S}\frac{\partial \mathbf{C}}{\partial Q} = 0. \tag{5.9}$$

By using Eq. (5.9) and the Roothaan-Hall equation: $\mathbf{F}\mathbf{C}_p = \boldsymbol{\varepsilon}_p \mathbf{S}\,\mathbf{C}_p$, the pulay term in Eq. (5.7) can be rewritten as

$$\sum_{\mu\nu} \frac{\partial D_{\mu\nu}}{\partial Q} F_{\nu\mu} = -\sum_{\mu\nu} W_{\mu\nu} \frac{\partial S_{\nu\mu}}{\partial Q}, \tag{5.10}$$

where **W** is the energy-weighted density matrix, expressed as

$$W_{\mu\nu} = \sum_{i}^{\text{occ}} \varepsilon_i C_{\mu i}^{*} C_{\nu i} . \tag{5.11}$$

Consequently, the standard HF energy gradient is given as

$$\begin{aligned} \frac{\partial E_{\text{HF}}}{\partial Q} = & \sum_{\mu\nu} D_{\mu\nu} \left\{ \frac{\partial H_{\nu\mu}^{\text{core}}}{\partial Q} + \sum_{\lambda\sigma} D_{\lambda\sigma} \left[2 \frac{\partial (\nu\mu \mid \lambda\sigma)}{\partial Q} - \frac{\partial (\nu\lambda \mid \mu\sigma)}{\partial Q} \right] \right\} \\ & - \text{Tr}\left[\mathbf{W} \frac{\partial \mathbf{S}}{\partial Q} \right] \end{aligned}, \tag{5.12}$$

In the DC-HF energy gradient calculation, the Hellmann-Feynman term can be calculated in the same way as the standard calculation. However, the two expressions exist for the Pulay term. One is proposed by Yang *et al.*[6] and the other is Kobayashi *et al.*[7] In the DC-HF energy gradient proposed by Yang and Lee, the Pulay term is approximated by means of the density matrix obtained from the DC method. The energy-weighted density matrix for the entire system is constructed approximately by the sum of the density matrix for subsystems, described as

$$W_{\mu\nu} \approx W_{\mu\nu}^{\text{DC}} = \sum_{\alpha}^{\text{subsystem}} P_{\mu\nu}^{\alpha} W_{\mu\nu}^{\alpha}, \tag{5.13}$$

$$W_{\mu\nu}^{\alpha} = \sum_{p} f_{\beta} \left(\varepsilon_{\text{F}} - \varepsilon_{p}^{\alpha} \right) \varepsilon_{p}^{\alpha} C_{\mu p}^{\alpha} C_{\nu p}^{\alpha} . \tag{5.14}$$

Therefore, the DC-HF energy gradient can be approximately expressed by using the standard Pulay term formula in Eq. (5.12), shown as

$$\frac{\partial E_{\mathrm{HF}}^{\mathrm{DC(Yang)}}}{\partial Q} = \sum_{\mu\nu} D_{\mu\nu}^{\mathrm{DC}} \left\{ \frac{\partial H_{\nu\mu}^{\mathrm{core}}}{\partial Q} + \sum_{\lambda\sigma} D_{\lambda\sigma}^{\mathrm{DC}} \left[2\frac{\partial(\nu\mu\,|\,\lambda\sigma)}{\partial Q} - \frac{\partial(\nu\lambda\,|\,\mu\sigma)}{\partial Q} \right] \right\} - \mathrm{Tr}\left[\mathbf{W}^{\mathrm{DC}} \frac{\partial \mathbf{S}}{\partial Q} \right] . \tag{5.15}$$

On the other hand, the partition matrix in Eq. (5.2) is rewritten as

$$P_{\mu\nu}^{\alpha} = \begin{cases} P_{\mu}^{\alpha} + P_{\nu}^{\alpha} & \text{for } \mu \in \boldsymbol{L}(\alpha) \wedge \nu \in \boldsymbol{L}(\alpha) \\ 0 & \text{otherwise} \end{cases} ,$$

(5.16)

$$P_{\mu}^{\alpha} = \begin{cases} 1/2 & \text{for } \mu \in \boldsymbol{S}(\alpha) \\ 0 & \text{otherwise} \end{cases} , \tag{5.17}$$

where $\boldsymbol{L}(\alpha) \equiv \boldsymbol{S}(\alpha) \cup \boldsymbol{B}(\alpha)$. The exact Pulay term in the DC-HF energy gradient calculation can be constructed by means of the definition of Eq. (5.16) and the natures of Hermitian in $\mathbf{D}^{\alpha}$ and $\mathbf{F}^{\alpha}$, which is given as

$$\sum_{\mu\nu} \frac{\partial D_{\mu\nu}^{\mathrm{DC}}}{\partial Q} F_{\mu\nu} = \sum_{\alpha} \sum_{\mu\nu} P_{\mu\nu}^{\alpha} \frac{\partial D_{\mu\nu}^{\alpha}}{\partial Q} F_{\mu\nu} = \sum_{\alpha} \sum_{\mu \in \boldsymbol{S}(\alpha)} \left(\frac{\partial \mathbf{D}^{\alpha}}{\partial Q} \mathbf{F}^{\alpha} \right)_{\mu\mu} . \tag{5.18}$$

Eq. (5.18) cannot be rewritten like Eq. (5.10) because of the existence of the partition matrix. Kobayashi *et al.* proposed the different DC-HF energy gradient expression with Eq. (5.16). In their approach, it is assumed that $\mathbf{D}^{\alpha}$ has the idempotence, which is described as

$$\mathbf{D}^{\alpha} \approx \mathbf{D}^{\alpha}\mathbf{S}^{\alpha}\mathbf{D}^{\alpha} \tag{5.19}$$

from the Pulay term at the finite electronic temperature proposed by Niklasson. In addition, it is assumed that $\mathbf{D}^{\alpha}$ is determined variationally. Under the condition that $\mathbf{D}^{\alpha}$ has the idempotence and is the variational solution, the relationship equation is given as

$$\frac{\partial \mathbf{D}^{\alpha}}{\partial Q} \sim -\mathbf{D}^{\alpha} \frac{\partial \mathbf{S}^{\alpha}}{\partial Q} \mathbf{D}^{\alpha} . \tag{5.20}$$

By substituting Eq. (5.20) for Eq. (5.18) and using the Roothaan equation for subsystem, the Pulay term can be described as

$$\sum_{\alpha} \sum_{\mu \in \boldsymbol{S}(\alpha)} \left(\frac{\partial \mathbf{D}^{\alpha}}{\partial Q} \mathbf{F}^{\alpha} \right)_{\mu\mu} = -\sum_{\alpha} \sum_{\mu \in \boldsymbol{S}(\alpha)} \sum_{\nu\lambda\sigma} D^{\alpha}_{\mu\nu} \frac{\partial S^{\alpha}_{\nu\lambda}}{\partial Q} W^{\alpha}_{\lambda\sigma} S^{\alpha}_{\sigma\mu} . \tag{5.21}$$

Therefore, instead of the formula proposed by Yang and Lee, the alternative DC-HF energy gradient can be expressed as

$$\begin{aligned} \frac{\partial E_{\mathrm{HF}}^{\mathrm{DC(Kobayashi)}}}{\partial Q} = & \sum_{\mu\nu} D^{\mathrm{DC}}_{\mu\nu} \left\{ \frac{\partial H^{\mathrm{core}}_{\nu\mu}}{\partial Q} + \sum_{\lambda\sigma} D^{\mathrm{DC}}_{\lambda\sigma} \left[2 \frac{\partial (\nu\mu \mid \lambda\sigma)}{\partial Q} - \frac{\partial (\nu\lambda \mid \mu\sigma)}{\partial Q} \right] \right\} \\ & - \sum_{\alpha} \mathrm{Tr} \left[\frac{\partial \mathbf{S}^{\alpha}}{\partial Q} \mathbf{X}^{\alpha} \right] \end{aligned} , \tag{5.22}$$

where $\mathbf{X}^{\alpha} = \mathbf{W}^{\alpha} \mathbf{S}^{\alpha} \mathbf{D}^{\alpha}$. When $[n \times m]$ represents the size of matrix (n is row and m is column), the size of matrix for $\mathbf{S}^{\alpha}$ and $\mathbf{D}^{\alpha}$ in $\mathbf{X}^{\alpha}$ are $[\boldsymbol{L}(\alpha) \times \boldsymbol{S}(\alpha)]$ and $[\boldsymbol{S}(\alpha) \times \boldsymbol{L}(\alpha)]$, respectively.

5.3. Estimation of DC-HF energy gradient

I propose a method which estimates the DC-HF energy gradient based on the philosophy of the automated DC-HF method. As mentioned in Chapter 3, the automated DC-HF method[1] is introduced the two-layered buffer region proposed by Dixon and Merz[8], which is called the inner buffer region [$\boldsymbol{B}_{\mathrm{i}}(\alpha)$] and the outer buffer region [$\boldsymbol{B}_{\mathrm{o}}(\alpha)$], respectively. In the automated DC-HF calculation, the density matrix change, $\Delta\mathbf{D}$, between the inner and outer buffer regions is constructed, described as

$$\Delta D_{\mu\nu} = \sum_{\alpha}\left[P'^{\alpha}_{\mu\nu}\sum_{p} f_{\beta}(\varepsilon'_{\mathrm{F}} - \varepsilon^{\alpha}_{p}) C^{\alpha}_{\mu p} C^{\alpha *}_{\nu p} - P^{\alpha}_{\mu\nu}\sum_{p} f_{\beta}(\varepsilon_{\mathrm{F}} - \varepsilon^{\alpha}_{p}) C^{\alpha}_{\mu p} C^{\alpha *}_{\nu p} \right]. \tag{5.23}$$

$\mathbf{P}'^{\alpha}$ is the auxiliary partition matrix, expressed as

$$P'^{\alpha}_{\mu\nu} = \begin{cases} 1 & \text{for } \mu \in \boldsymbol{S}(\alpha) \wedge \nu \in \boldsymbol{S}(\alpha) \\ 1/2 & \text{for } [\mu \in \boldsymbol{S}(\alpha) \wedge \nu \in (\boldsymbol{B}_{\mathrm{i}}(\alpha) \cup \boldsymbol{B}_{\mathrm{o}}(\alpha))] \text{ or } \textit{vice versa} \\ 0 & \text{otherwise} \end{cases} \tag{5.24}$$

$\mathbf{P}'^{\alpha}$ and $\varepsilon'_{\mathrm{F}}$ the auxiliary Fermi level in transferring the outer buffer region to inner buffer region. In the present method, I consider $\varepsilon'_{\mathrm{F}} = \varepsilon_{\mathrm{F}}$ and Eq. (5.23) is rewritten as

$$\Delta D_{\mu\nu} \approx \Delta D^{\mathrm{DC}}_{\mu\nu} = \sum_{\alpha}(P'^{\alpha}_{\mu\nu} - P^{\alpha}_{\mu\nu})\sum_{p} f_{\beta}(\varepsilon_{\mathrm{F}} - \varepsilon^{\alpha}_{p}) C^{\alpha}_{\mu p} C^{\alpha *}_{\nu p} = \sum_{\alpha} \Delta P^{\alpha}_{\mu\nu} D^{\alpha}_{\mu\nu}, \tag{5.25}$$

where

$$\Delta P^{\alpha}_{\mu\nu} = \begin{cases} 1/2 & \text{for } \left(\mu \in \boldsymbol{S}(\alpha) \wedge \nu \in \boldsymbol{B}_{\mathrm{o}}(\alpha)\right) \text{ or } \textit{vice versa} \\ 0 & \text{otherwise} \end{cases}. \tag{5.26}$$

If the density matrix is changed only $\Delta\mathbf{D}$ in Eq. (5.25), the Pulay term proposed by Kobyashi *et al*. can be described as

$$
\begin{aligned}
g_Q^{\text{Pulay}} &= -\sum_{\alpha}\sum_{\mu}\left(\frac{\partial \mathbf{S}^{\alpha}}{\partial Q}\mathbf{W}^{\alpha}\mathbf{S}^{\alpha[\boldsymbol{L}_{\text{o}}(\alpha)\times\boldsymbol{S}(\alpha)]}\Delta\mathbf{D}^{\alpha[\boldsymbol{S}(\alpha)\times\boldsymbol{L}_{\text{o}}(\alpha)]}\right)_{\mu\mu} \\
&= -\sum_{\alpha}\sum_{\mu}\left(\frac{\partial \mathbf{S}^{\alpha[\boldsymbol{B}_{\text{o}}(\alpha)\times\boldsymbol{L}_{\text{o}}(\alpha)]}}{\partial Q}\mathbf{W}^{\alpha}\mathbf{S}^{\alpha[\boldsymbol{L}_{\text{o}}(\alpha)\times\boldsymbol{S}(\alpha)]}\mathbf{D}^{\alpha[\boldsymbol{S}(\alpha)\times\boldsymbol{B}_{\text{o}}(\alpha)]}\right)_{\mu\mu}
\end{aligned}
, \tag{5.27}
$$

where $\boldsymbol{L}_o(\alpha) \equiv \boldsymbol{S}(\alpha) \cup \boldsymbol{B}_{\text{i}}(\alpha) \cup \boldsymbol{B}_{\text{o}}(\alpha)$. Similarly, the one electron part of the Hellmann-Feynman term in Eq. (5. 22) can be described with $\Delta\mathbf{D}$ as

$$
g_Q^{\text{OEI}} = \sum_{\mu\nu}\Delta D_{\mu\nu}^{\text{DC}}\frac{\partial H_{\nu\mu}^{\text{core}}}{\partial Q}. \tag{5.28}
$$

To reduce the computational cost of the two-electron integrals, the Schwarz inequality and the Thompson inequality[9] are employed:

$$
\left|(ij\,|\,kl)\right| \le \sqrt{\left|(ij\,|\,ij)\right|}\sqrt{\left|(kl\,|\,kl)\right|} = A_{ij}A_{kl}\,, \tag{5.29}
$$

$$
\left|(ij\,|\,kl)\right| \le \sqrt{\left|(ii\,|\,jj)\right|}\sqrt{\left|(kk\,|\,ll)\right|} = M_{ij}M_{kl}\,. \tag{5.30}
$$

Therefore, the upper limit of the two-electron part of the Hellmann-Feynman term can be described as

$$
\begin{aligned}
\left|g_Q^{\text{TEI}}\right| &= \left|\sum_{\mu\nu}\Delta D_{\mu\nu}^{\text{DC}}\sum_{\lambda\sigma}D_{\lambda\sigma}^{\text{DC}}\left[2\frac{\partial(\nu\mu\,|\,\lambda\sigma)}{\partial Q} - \frac{\partial(\nu\lambda\,|\,\mu\sigma)}{\partial Q}\right]\right| \\
&\le \sum_{\mu\nu}\left|\Delta D_{\mu\nu}^{\alpha}\right|\sum_{\lambda\sigma}\left|D_{\lambda\sigma}^{\text{DC}}\right|\left[2\left|\frac{\partial(\nu\mu\,|\,\lambda\sigma)}{\partial Q}\right| + \left|\frac{\partial(\nu\lambda\,|\,\mu\sigma)}{\partial Q}\right|\right]
\end{aligned}
. \tag{5.31}
$$

By introducing Eqs. (5.29) and (5.30) into the two-electron integrals term in Eq. (5.31), the estimated DC-HF energy gradient is expressed as

$$
\begin{aligned}
\left|g_Q^{\text{DC-HF}}\right| \le &\left|\sum_{\mu\nu}\Delta D_{\mu\nu}^{\text{DC}}\frac{\partial H_{\nu\mu}^{\text{core}}}{\partial Q}\right| + \sum_{\mu\nu}\left|\Delta D_{\mu\nu}^{\alpha}\right|\left[2\left(A_{\nu^Q\mu} + A_{\nu\mu^Q}\right) + 2A_{\nu\mu} + \left(M_{\nu^Q\mu} + M_{\nu\mu^Q}\right) + M_{\nu\mu}\right]\bullet\sum_{\lambda\sigma}C_{\lambda\sigma} \\
&+\left|\sum_{\alpha}\sum_{\mu}\left(\frac{\partial \mathbf{S}^{\alpha[\boldsymbol{B}_{\text{o}}(\alpha)\times\boldsymbol{L}_{\text{o}}(\alpha)]}}{\partial Q}\mathbf{W}^{\alpha}\mathbf{S}^{\alpha[\boldsymbol{L}_{\text{o}}(\alpha)\times\boldsymbol{S}(\alpha)]}\mathbf{D}^{\alpha[\boldsymbol{S}(\alpha)\times\boldsymbol{B}_{\text{o}}(\alpha)]}\right)_{\mu\mu}\right|
\end{aligned}
, \tag{5.32}
$$

where $A_{i^Q j}$ is the first derivative of the Schwarz inequality[10] and $M_{i^Q j}$ is the first derivative of the Thompson inequality:

$$A_{i^Q j} = \sqrt{\left(\frac{\partial i}{\partial Q} j \,|\, \frac{\partial i}{\partial Q} j \right)}, \tag{5.33}$$

$$M_{i^Q j} = \sqrt{\left(\frac{\partial i}{\partial Q} \frac{\partial i}{\partial Q} \,|\, jj \right)}. \tag{5.34}$$

Because the summation of $C_{\lambda\sigma}$ in Eq. (5.32) is constant, the following index can be considered as

$$\begin{aligned} e^{\mathrm{grad}} \le & \left| \sum_{\mu\nu} \Delta D_{\mu\nu}^{\mathrm{DC}} \frac{\partial H_{\nu\mu}^{\mathrm{core}}}{\partial Q} \right| + \sum_{\mu\nu} \left| \Delta D_{\mu\nu}^{\alpha} \right| \left[2\left(A_{\nu^Q \mu} + A_{\nu\mu^Q} \right) + 2A_{\nu\mu} + \left(M_{\nu^Q \mu} + M_{\nu\mu^Q} \right) + M_{\nu\mu} \right] \\ & + \left| \sum_{\alpha} \sum_{\mu} \left(\frac{\partial \mathbf{S}^{\alpha[\boldsymbol{B}_{\mathrm{o}}(\alpha)\times\boldsymbol{L}_{\mathrm{o}}(\alpha)]}}{\partial Q} \mathbf{W}^{\alpha} \mathbf{S}^{\alpha[\boldsymbol{L}_{\mathrm{o}}(\alpha)\times\boldsymbol{S}(\alpha)]} \mathbf{D}^{\alpha[\boldsymbol{S}(\alpha)\times\boldsymbol{B}_{\mathrm{o}}(\alpha)]} \right)_{\mu\mu} \right| \end{aligned} . \tag{5.35}$$

The above e^{grad} can be further divided into the contribution from atom A, e_A^{grad}. Consequently, I propose the following automatic determination scheme for the buffer region in the DC-HF energy gradient method:

i. Evaluation of e_A^{grad} after constructing $\mathbf{F}\left[\mathbf{D}^{\mathrm{DC}}\right]$ for each SCF cycle.

ii. Transferring all atoms in the outer buffer region of subsystem α to its inner buffer region.

iii. Inclusion of the atoms in the sphere with radius $r_{\mathrm{ext}}^{\mathrm{grad}}$ centered on atom A with $e_A^{\mathrm{grad}} \ge e_{\mathrm{thresh}}^{\mathrm{grad}}$ into the new outer buffer region of subsystem α.

Note that the above procedure has not been implemented yet in the present paper.

5.4. Numerical assessment

In this work, the estimation of the Pulay term (g_Q^{Pulay}) and the one electron part of the Hellmann-Feynman term in Eq. (5.35) (g_Q^{OEI}) were implemented to the GAMESS package[11,12] and assessed its accuracy for α-helix glycine oligomer $(\mathrm{GLY})_{10}$. The structure was constructed by using the Discovery Studio 2017 R2 software.[13] In the DC method, the inverse temperature parameter, β, in Eq. (5.25) was set to 200 a.u. To examine the estimated and actual energy errors, the automated DC-SCF method is not performed throughout this Section.

Firstly, the buffer size dependence of the estimated Pulay term in the final SCF cycles are shown in Table 5-1. In Table 5-1, the maximum absolute deviations (MaxADs) and mean absolute deviations (MADs) between the DC-HF energy gradient and the standard HF energy gradient for each buffer size are shown. In addition, the actual and the estimated DC-HF energy errors are also given. Here, the 6-31G basis set[14] was adopted.

As expected, in both the DC-HF energy and DC-HF energy gradient calculations, the actual errors decreased with the increasing buffer size. In the energy calculation, the order of the estimated energy error at the final SCF step was consistent with that of the actual error. In the energy gradient calculation, the order of the estimated Pulay term error at the final SCF step tended to be also consistent with that of the actual error.

For the same system, I also examined the estimated one electron part of the Hellmann-Feynman term, as shown in Table 5-2. In this term, except for $r_{\mathrm{b}}^{\mathrm{in}} = 3.5\ \text{Å}$, the actual errors also decreased with the increasing buffer size. The MaxAD in $r_{\mathrm{b}}^{\mathrm{in}} = 3.5\ \text{Å}$ is

smaller than that in $r_b^{in} = 4.0$ Å because it may be "lucky cancellation". In contrast to the Pulay term, the order of the estimated MaxAD and MAD at the final SCF step was two or three orders of magnitude smaller than that of the actual errors. The estimated error decreased drastically than the actual error. It is because that the energy gradient may be sensitive to the density matrix change, which is affected widely. Consequently, the size of the outer buffer region in the energy error estimation may be insufficient in the energy gradient error estimation.

Table 5-1. Buffer size dependence of the actual and estimated Pulay term for α-helix glycine oligomer $(GLY)_{10}$. Standard HF energy is –2142.6879 E_h.

r_b^{in} /Å	r_b^{out} /Å	$E_{HF}^{DC} - E_{HF}^{Stand}$ /E_h	$-\Delta E$ /E_h		$\nabla E_{Pulay}^{DC} - \nabla E_{Pulay}^{Stand}$ /E_h · bohr^{-1}	$\nabla\left(\Delta E_{Pulay}^{DC}\right)$ /E_h · bohr^{-1}
3.5	4.5	+0.0553	+0.0540	MaxAD	0.0265	0.0317
				MAD	0.0043	0.0027
4.0	5.0	–0.0397	–0.0511	MaxAD	0.0188	0.0209
				MAD	0.0026	0.0014
4.5	5.5	+0.0027	+0.0022	MaxAD	0.0021	0.0007
				MAD	0.0003	0.0002
5.0	6.0	+0.0002	–0.0007	MaxAD	0.0007	0.0007
				MAD	0.0001	0.0001
5.5	6.5	+0.0001	–0.0003	MaxAD	0.0006	0.0003
				MAD	0.0001	0.0000

Table 5-2. Buffer size dependence of the actual and estimated one electron part of the Hellmann-Feynman term for α-helix glycine oligomer $(GLY)_{10}$. Standard HF energy is –2142.6879 E_h.

r_b^{in} /Å	r_b^{out} /Å	$E_{HF}^{DC} - E_{HF}^{Stand}$ /E_h	$-\Delta E$ /E_h		$\nabla E_{Pulay}^{DC} - \nabla E_{Pulay}^{Stand}$ /E_h · bohr^{-1}	$\nabla\left(\Delta E_{Pulay}^{DC}\right)$ /E_h · bohr^{-1}
3.5	4.5	+0.0553	+0.0540	MaxAD	0.5021	0.2697
				MAD	0.1032	0.0208
4.0	5.0	–0.0397	–0.0511	MaxAD	0.5795	0.0998
				MAD	0.0808	0.0072
4.5	5.5	+0.0027	+0.0022	MaxAD	0.0664	0.0076
				MAD	0.0107	0.0007
5.0	6.0	+0.0002	–0.0007	MaxAD	0.0201	0.0010
				MAD	0.0037	0.0001
5.5	6.5	+0.0001	–0.0003	MaxAD	0.0160	0.0002
				MAD	0.0031	0.0000

5.5. Concluding remarks

In this study, I have proposed an approach to determine the appropriate size of the buffer region in the DC-HF energy gradient calculation. This estimation approach is based on the philosophy of the automated DC-HF method. From the energy gradient expression proposed by Kobayashi *et. al*., the Pulay term in the estimated DC-HF energy gradient is derived. In addition, Schwarz and Thompson inequalities were introduced to reduce the computational cost of the two-electron integrals. In this paper, the estimation of the Pulay term and the one electron part of the Hellmann-Feynman term were implemented to the GAMESS package.

I applied the above terms to α-helix glycine oligomer $(GLY)_{10}$ and confirmed that MaxAD and MAD for the estimated DC-HF energy gradient tended to be consistent with the actual values in the Pulay term. However, MaxAD and MAD in the one electron part of the Hellmann-Feynman term tended to be smaller than the actual value.

In the geometry optimization calculation, the energy gradient for each atom is calculated in each optimization step. Therefore, the approach which determine the size of the buffer region based on the energy gradient differentiation between the optimization steps is proposed alternatively in the future work.

5.6. Reference

[1] M. Kobayashi, T. Fujimori, T. Taketsugu, *J. Comput. Chem.* **2018**, *39*, 909.

[2] W. Yang, T.-S. Lee, *J. Chem. Phys.* **1995**, *103*, 5674.

[3] M. Kobayashi and H. Nakai, In *Linear-Scaling Techniques in Computational Chemistry and Physics,* Chapter 5; R. Zaleśny, M. G. Papadopoulos, P. G. Mezey, J. Leszczynski, Eds., Springer, Dordrecht, **2011**, pp. 97–127.

[4] T. Akama, M. Kobayashi, H. Nakai, *J. Comput. Chem.* **2007**, *28*, 2003.

[5] 原田義也、量子化学　下、裳華房(2007).

[6] W. Yang and T.-S. Lee, J. Chem. Phys. 1995, 103, 5674.

[7] M. Kobayashi, T. Kunisada, T. Akama, D. Sakura and H. Nakai, *J. Chem. Phys.* **2011**, *134*, 034105.

[8] S. L. Dixon, K. M. Merz Jr., *J. Chem. Phys.* **1997**, *107*, 879.

[9] T. H. Thompson, C. Ochsenfeld, *J. Chem. Phys.* **2017**, *147*, 144101.

[10] H. Horn, H. Weiβ, M. Häser, M. Ehrig, R. Ahlrichs, *J. Comput. Chem.* **1991**, *12*, 1058.

[11] M. S. Gordon, M. W. Schmidt, In *Theory and Applications of Computational Chemistry: The First Forty Years;* C. Dykstra, G. Frenking, K. Kim, G. Scuseria, Eds.; Elsevier: Amsterdam, **2005**; pp 1167–1189.

[12] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery Jr., *J. Comput. Chem.* **1993**, *14*, 1347.

[13] Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016.

[14] W. J. Hehre, R. Ditchfield, J.A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.

6. General conclusion

In this study, to control automatically the errors between the standard and the DC methods, I have developed the method which determines the appropriate size of the buffer region automatically with the error estimation scheme for DC-SCF, DC-MP2 and DC-HF energy gradient calculations.

In Chapter 3, the automatic error control method in the DC-SCF calculation was developed. In this method, the two-layer buffer regions proposed by Dixon and Merz was introduced and the density matrix change between the layers was constructed with it. The first order atomic energy variation can be calculated from the density matrix change. The appropriate size of the buffer region in the DC-SCF calculation can be constructed with the atomic energy contribution as a criterion. The present method was applied to the water cluster, protein and alkane systems and worked effectively for these molecule systems. In addition, the present method achieved linear scaling as well as the conventional DC method.

In Chapter 4, an extension of the automatic error control method to the DC-MP2 calculation was developed. With the help of the idea of atomic orbital (AO)-Laplace MP2 method proposed by Häser, the atomic energy variation in the buffer region can be estimated from the DC-MP2 electron correlation energy in each subsystem. Because of the short-range nature of the MP2 dynamical electron correlation, the appropriate size of the buffer region in the DC-MP2 calculation can be smaller than that in the DC-HF calculation. Therefore, the appropriate size of the buffer region in the DC-MP2 calculation can be determined automatically by reducing from the size of the buffer region

in the prior DC-HF calculation with the estimated atomic energy. The present method was applied to the 100 water cluster system and the chignolin system. It was confirmed that the DC-MP2 energy error can be systematically controlled by the estimated atomic energy. It was also confirmed that the linear-scaling behavior of the DC-MP2 method is preserved even with the present automation scheme, from a calculation of linear polyene system.

In Chapter 5, I have proposed the automatic error control scheme for the DC-HF energy gradient calculation. Based on the philosophy of the automated DC-HF method, the Pulay term in the estimated energy gradient was constructed with the idea of the DC-HF energy gradient expression proposed by Kobayashi *et. al*. In addition, to reduce the computational cost of the two-electron integrals, the upper limit of the Hellmann-Feynman term was constructed with Schwarz and Thompson inequalities. The automatic determination of the buffer region in the DC-HF energy gradient calculation was proposed with the above estimation. In this paper, the Pulay term and the one-electron part of the Hellmann-Feynman term in the estimated DC-HF energy gradient were applied to the α-helix glycine oligomer $(GLY)_{10}$. It was confirmed that the maximum absolute deviation (MaxAD) and the mean absolute deviation (MAD) in the Pulay term tended to be consistent with the actual errors. However, MaxAD and MAD in the one electron part of the Hellmann-Feynman term tended to be smaller than the actual ones. Improvement of the present scheme or the alternative approach will be the scope of future studies.

The present study makes it possible to perform the chemical reaction analysis for large-scale systems with the high accuracy and short computational time. In addition, because the effort of preliminary assessment for determination of the size of the buffer region can be reduced, the convenience of the DC method was improved. In the future, it is expected that people from other fields such as biochemistry perform the quantum chemical

calculation and I hope that this study can contribute to a theoretical understanding of chemical phenomena for large-scale systems.

Acknowledgment

This research was carried out under the supervision of Prof. Tetsuya Taketsugu and Associate Prof. Masato Kobayashi of the Faculty of Science, Hokkaido University. I received a lot of advice for this dissertation from them. I would like to express my deepest gratitude to them.

I would like to express my sincere gratitude to Prof. Junya Hasegawa of Institute for Catalysis, Hokkaido University, Prof. Satoshi Maeda of the Faculty of Science, Hokkaido University, and Associate Prof. Shinichiro Sato of the Faculty of Engineering, Hokkaido University, to assist as chief reviewer and associate reviewers of this dissertation and their constructive comments.

In addition, I would like to also express my deepest gratitude to all members of the quantum chemistry laboratory. Especially, I'm grateful to Assistant Prof. Takeshi Iwasa, Assistant Prof. Yuriko Ono, Ph.D. Tomoko Akama, and the students in M. Kobayashi group for the helpful discussions and advices on my research. I'm also grateful to my batchmates, B.S. Hiroshi Morita, M.S. Masashi Sanpei, M.S. Maki Nakahara, Ph.D. Masato Takenaka and Ph.D. Takuro Tsutsumi for their support both my working life and private life.

I also thank to support by Ministry of Education, Culture, Sports, Science and Technology through the Program for Leading Graduate Schools (Hokkaido University "Ambitious Leader's Program").

Finally, I deeply appreciate my family for understanding my decision to enter the doctoral course.

Research achievements

1. Paper

1) ○"Automated error control in divide-and-conquer self-consistent field calculations: Automated Error Control"
M. Kobayashi, T. Fujimori, T. Taketsugu,
J. Comput. Chem. **2018**, *39*, 909.

2) ○"Energy-based automatic determination of buffer region in the divide-and-conquer second-order Møller-Plesset perturbation theory
T. Fujimori, M. Kobayashi, T. Taketsugu,
J. Comput. Chem. **2021**, *42*, 620.

2. Conference presentation

International conference

1) "Automation of large-scale quantum chemical calculations based on the divide and conquer method"
The 4th International Symposium on AMBITIOUS LEADER'S PROGRAM, Japan, November 2016,
Toshikazu Fujimori, Masato Kobayashi, Tetsuya Taketsugu.

2) "Automatic error control in the divide and conquer quantum chemical calculation"
CPMD2017 Workshop, Japan, October 2017,
Toshikazu Fujimori, Masato Kobayashi, Tetsuya Taketsugu.

3) “Automatic energy-error control in the large-scale quantum chemical calculations based on the divide-and-conquer method”
The 13th Nanjing University-Hokkaido University-NIMS-Jiaxing University Joint Symposium, China, October 2017,
Toshikazu Fujimori, Masato Kobayashi, Tetsuya Taketsugu.

4) “Automation of Energy-Error Control in the Large-Scale Quantum Chemical Calculations Based on the Divide and Conquer Method”
The 5th International Symposium on AMBITIOUS LEADER’S PROGRAM, Japan, November 2017,
Toshikazu Fujimori, Masato Kobayashi, Tetsuya Taketsugu.

5) “Automatic error control scheme for the divide-and-conquer electronic structure calculation”
ICQC2018, France, June 2018,
Toshikazu Fujimori, Masato Kobayashi, Tetsuya Taketsugu.

6) “Automatic Fragment Size Determination Scheme in Divide and Conquer Calculation”
Hokkaido-Strasbourg Symposium Biotic and Abiotic Molecular Machines and Motors, France, November 2018,
Toshikazu Fujimori, Masato Kobayashi, Tetsuya Taketsugu.

7) “Automated Error Control and Static Electron Correlation Treatment in Divide-and-Conquer Method”
QRS2019, Japan, July 2019,
M.Kobayashi, T.Fujimori, R.Kodama, T.Taketsugu.

8) “Energy-based automatic fragmentation in the linear-scaling divide-and-conquer electron correlation calculations”
APATCC2019, Australia, September 2019,
Toshikazu Fujimori, Masato Kobayashi, Tetsuya Taketsugu.

国内学会

1) “分割統治（DC）法による大規模近似量子化学計算：誤差の自動制御手法開発”
化学系学協会北海道支部 2016 年冬季研究会，北海道大学, 2016 年 1 月,
藤森俊和・小林正人・武次徹也.

2) “階層型バッファ領域を用いた分割統治（DC）法における誤差の自動制御化”
第 19 回理論化学討論会，早稲田大学, 2016 年 5 月,
藤森俊和・小林正人・武次徹也.

3) ”大規模量子化学計算を簡便化する自動制御型分割統治法の開発”
日本化学会北海道支部 2016 年夏季研究会，室蘭工業大学, 2016 年 7 月,
藤森俊和・小林正人・武次徹也.

4) “分割統治（DC）法に基づいた大規模近似量子化学計算における誤差の自動制御化”
日本化学会第 97 春季年会，慶応義塾大学, 2017 年 3 月,
藤森俊和・小林正人・武次徹也.

5) “分割統治 SCF 計算における誤差の自動制御手法の開発”
第 20 回理論化学討論会，京都大学, 2017 年 5 月,
藤森俊和・小林正人・武次徹也.

6) “分割統治 MP2 計算における相関バッファ領域の自動決定”
第 21 回理論化学討論会，岡崎コンファレンスセンター, 2018 年 5 月,
藤森俊和・小林正人・武次徹也.

7) “分割統治 MP2 計算における誤差自動制御スキームの開発”
第 12 回分子科学討論会，福岡国際会議場, 2018 年 9 月,
藤森俊和・小林正人・武次徹也.

8) “分割統治電子相関計算に対する誤差の自動制御化”
Symposium for Reaction Path Search, 九州大学, 2018 年 9 月,
藤森俊和・小林正人・武次徹也.

9) “分割統治エネルギー勾配計算に対するバッファ領域自動制御法の開発”
第 22 回理論化学討論会, 北海道大学, 2019 年 5 月,
藤森俊和・小林正人・武次徹也.