

## FULL PAPER

# Energy-based automatic determination of buffer region in the divide-and-conquer second-order Møller–Plesset perturbation theory

Toshikazu Fujimori<sup>1</sup> | Masato Kobayashi<sup>2,3,4</sup>  | Tetsuya Taketsugu<sup>2,3,4</sup> 

<sup>1</sup>Graduate School of Chemical Sciences and Engineering, Hokkaido University, Sapporo, Japan

<sup>2</sup>Department of Chemistry, Faculty of Science, Hokkaido University, Sapporo, Japan

<sup>3</sup>WPI-ICReDD, Hokkaido University, Sapporo, Japan

<sup>4</sup>ESICB, Kyoto University, Kyoto, Japan

## Correspondence

Masato Kobayashi, Department of Chemistry, Faculty of Science, Hokkaido University, Sapporo 060-0810, Japan.

Email: k-masato@sci.hokudai.ac.jp

## Funding information

Core Research for Evolutional Science and Technology, Grant/Award Number: JPMJCR1902; Japan Society for the Promotion of Science, Grant/Award Number: KAKENHI JP17K05738; Ministry of Education, Culture, Sports, Science and Technology, Grant/Award Number: Elements Strategy Initiative JPMXP0112101003

## Abstract

In the linear-scaling divide-and-conquer (DC) electronic structure method, each subsystem is calculated together with the neighboring buffer region, the size of which affects the energy error introduced by the fragmentation in the DC method. The DC self-consistent field calculation utilizes a scheme to automatically determine the appropriate buffer region that is as compact as possible for reducing the computational time while maintaining acceptable accuracy (*J. Comput. Chem.* **2018**, *39*, 909). To extend the automatic determination scheme of the buffer region to the DC second-order Møller–Plesset perturbation (MP2) calculation, a scheme for estimating the subsystem MP2 correlation energy contribution from each atom in the buffer region is proposed. The estimation is based on the atomic orbital Laplace MP2 formalism. Based on this, an automatic buffer determination scheme for the DC-MP2 calculation is constructed and its performance for several types of systems is assessed.

## KEYWORDS

divide-and-conquer method, electron correlation, fragmentation, Laplace transformed MP2, linear-scaling calculation

## 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<sup>1</sup> or density functional theory (DFT),<sup>2</sup> owing to the diagonalization of the Hamiltonian matrix. Furthermore, in case of post-HF calculations, such as the second order Møller–Plesset perturbation (MP2)<sup>3–5</sup> and coupled cluster (CC) theories,<sup>4,5</sup> 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<sup>6–10</sup> 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.,<sup>11</sup>  $R$ -goodness parameter is used to determine the quality of each fragment.<sup>12,13</sup> In the generalized energy-based fragmentation approach,<sup>14,15</sup> each fragment is constructed with the distance threshold ( $\xi$ ). The cluster-in-molecule local correlation method also adopts the distance threshold  $\xi$  to control the size of the cluster,<sup>16</sup> while a simple correction scheme to account for the distant-pair correlation has recently been proposed.<sup>17</sup> The accuracy of the fragment molecular orbital method<sup>18</sup> can be systematically improved by increasing the order

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Journal of Computational Chemistry* published by Wiley Periodicals LLC.

of many-body expansion from the original two-body to three-body<sup>19,20</sup> and four-body<sup>21</sup> expansions. The pair natural orbital (PNO) electron correlation approach<sup>22,23</sup> 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.<sup>24,25</sup>

Yang and coworkers introduced a linear-scaling approach called the divide-and-conquer (DC) method.<sup>26,27</sup> The DC method has been applied to the HF or DFT self-consistent field (SCF),<sup>26,28</sup> density-functional tight-binding,<sup>29–32</sup> and post-HF (MP2<sup>33–36</sup> or CC<sup>37–39</sup>) energy calculations as well as the SCF<sup>40</sup> and MP2<sup>41</sup> 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<sup>42</sup> and the thermally-assisted occupation (finite temperature) scheme.<sup>43</sup> 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, we<sup>44</sup> 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.<sup>45</sup> This estimation scheme can successfully be applied to automatically determine the appropriate buffer region based on the estimated energy error.<sup>44</sup>

This study attempts to export the idea of the previous automated DC-HF scheme to the DC-MP2 calculation. Kobayashi et al.<sup>36</sup> 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. We 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<sup>46–50</sup> is used as well as the Schwarz inequality. Based on this estimated energy contribution, we established an algorithm to automatically determine the appropriate buffer region in the DC-MP2 calculation.

This paper consists of four sections. Section 2 gives a brief summary of the linear-scaling DC electron correlation method with a fixed buffer region as well as the present procedure to estimate the energy contribution from each buffer atom and the automated DC-MP2 algorithm. Numerical assessments are described in Section 3. Finally, we provide concluding remarks in Section 4.

## 2 | METHODS

### 2.1 | The DC-MP2 electron correlation calculation

We 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,  $\mu, \nu, \dots$ . 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  $\alpha$  are referred to as  $\mathbf{S}(\alpha)$  and  $\mathbf{B}(\alpha)$ , respectively.

In the DC-MP2 method, the MOs in the subsystem  $\alpha$ ,

$$\psi_p^\alpha(\mathbf{r}) = \sum_{\mu \in \mathbf{L}(\alpha)} C_{\mu p}^\alpha \phi_\mu(\mathbf{r}) \quad (1)$$

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

$$\mathbf{F}^\alpha [\mathbf{D}^{\text{SCF}}] \mathbf{C}_p^\alpha = \epsilon_p^\alpha \mathbf{S}^\alpha \mathbf{C}_p^\alpha \quad (2)$$

where  $\mathbf{F}^\alpha[\mathbf{D}^{\text{SCF}}]$  is the subsystem Fock matrix constructed with the density matrix  $\mathbf{D}^{\text{SCF}}$ , and  $\mathbf{S}^\alpha$  is the subsystem overlap matrix with the element  $S_{\mu\nu}^\alpha = (\phi_\mu | \phi_\nu)$  for  $\mu, \nu \in \mathbf{L}(\alpha)$ . Note that the subsystem Roothaan equation (2) has to be solved not self-consistently but just once using predetermined  $\mathbf{D}^{\text{SCF}}$ . The density matrix,  $\mathbf{D}^{\text{SCF}}$ , can be constructed from the standard or approximate HF calculation, such as the DC-HF one. If  $\mathbf{D}^{\text{SCF}}$  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:

$$D_{\mu\nu}^{\text{SCF}} \approx \sum_{\alpha}^{\text{subsystem}} p_{\mu\nu}^\alpha D_{\mu\nu}^\alpha \quad (3)$$

$$D_{\mu\nu}^\alpha = \sum_p f_\beta(\epsilon_F - \epsilon_p^\alpha) C_{\mu p}^\alpha C_{\nu p}^\alpha \quad (4)$$

$$p_{\mu\nu}^\alpha = \begin{cases} 1 & (\mu \in \mathbf{S}(\alpha) \wedge \nu \in \mathbf{S}(\alpha)) \\ 1/2 & ((\mu \in \mathbf{S}(\alpha) \wedge \nu \in \mathbf{B}_{\text{SCF}}(\alpha)) \vee (\mu \in \mathbf{B}_{\text{SCF}}(\alpha) \wedge \nu \in \mathbf{S}(\alpha))) \\ 0 & (\text{otherwise}) \end{cases} \quad (5)$$

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

Before the evaluation of the subsystem correlation energy, the subsystem MOs of Equation (1) must be classified into occupied  $\{\psi_i^\alpha, \psi_j^\alpha, \dots\}$  and virtual ones  $\{\psi_a^\alpha, \psi_b^\alpha, \dots\}$ . 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,  $\Delta E_{\text{corr}}^{(2)}$ , can be approximated as the sum of the subsystem MP2 correlation energies,  $\{\Delta E_{\text{corr}}^{\alpha(2)}\}$ ,

$$\Delta E_{\text{corr}}^{(2)} \approx \sum_{\alpha} \Delta E_{\text{corr}}^{\alpha(2)} \quad (6)$$

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

$$\Delta E_{\text{corr}}^{\alpha(2)} = \sum_{\tilde{i}^{\alpha}, \tilde{j}^{\alpha}}^{\text{occ}(\alpha) \text{vir}(\alpha)} \sum_{a^{\alpha}, b^{\alpha}, \mu \in S(\alpha)} \frac{C_{\mu}^{\alpha} (\mu a^{\alpha} | \tilde{j}^{\alpha} b^{\alpha})}{\epsilon_i^{\alpha} + \epsilon_j^{\alpha} - \epsilon_a^{\alpha} - \epsilon_b^{\alpha}} [2(a^{\alpha} \tilde{i}^{\alpha} | b^{\alpha} \tilde{j}^{\alpha}) - (a^{\alpha} \tilde{j}^{\alpha} | b^{\alpha} \tilde{i}^{\alpha})] \quad (7)$$

with the two-electron integral notation

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

## 2.2 | Estimation of the DC-MP2 energy

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

$$\Delta E_{\text{corr}}^{\alpha(2)} = \sum_{B \in L(\alpha)} \Delta E_B^{\alpha(2)} \quad (8)$$

$$\Delta E_B^{\alpha(2)} = \sum_{\tilde{i}^{\alpha}, \tilde{j}^{\alpha}}^{\text{occ}(\alpha) \text{vir}(\alpha)} \sum_{a^{\alpha}, b^{\alpha}, \mu \in S(\alpha) \nu \in B} \frac{C_{\mu}^{\alpha} C_{\nu}^{\alpha} (\mu \nu | \tilde{j}^{\alpha} b^{\alpha})}{\epsilon_i^{\alpha} + \epsilon_j^{\alpha} - \epsilon_a^{\alpha} - \epsilon_b^{\alpha}} [2(a^{\alpha} \tilde{i}^{\alpha} | b^{\alpha} \tilde{j}^{\alpha}) - (a^{\alpha} \tilde{j}^{\alpha} | b^{\alpha} \tilde{i}^{\alpha})] \quad (9)$$

According to the local correlation philosophy for dynamical electron correlation,<sup>52–54</sup> it is expected that  $\Delta E_B^{\alpha(2)}$  rapidly decreases as the distance between atom  $B$  and central region  $\alpha$  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,<sup>36</sup> 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  $\alpha$  is expected to be small. By applying the AO-Laplace MP2 technique to Equation (9),  $\Delta E_B^{\alpha(2)}$  can be expressed as

$$\Delta E_B^{\alpha(2)} = - \int_0^{\infty} \sum_{\mu \in S(\alpha) \nu \in B} \sum_{\lambda \sigma} \sum_{\gamma \kappa \delta \epsilon} X_{\mu\gamma}^{\alpha}(\tau) Y_{\nu\kappa}^{\alpha}(\tau) X_{\lambda\delta}^{\alpha}(\tau) Y_{\sigma\epsilon}^{\alpha}(\tau) (\mu\nu | \lambda\sigma) [2(\kappa\gamma | \epsilon\delta) - (\kappa\delta | \epsilon\gamma)] d\tau \quad (10)$$

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

$$X_{\mu\nu}^{\alpha}(\tau) = \sum_i C_{\mu}^{\alpha} C_{\nu}^{\alpha} e^{(\epsilon_i^{\alpha} - \epsilon_F)\tau} \quad (11)$$

$$Y_{\mu\nu}^{\alpha}(\tau) = \sum_a C_{\mu}^{\alpha} C_{\nu}^{\alpha} e^{-(\epsilon_a^{\alpha} - \epsilon_F)\tau} \quad (12)$$

Here, the Fermi level,  $\epsilon_F$ , 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, we drastically approximate the integral in Equation (10) by the one-point Gauss-Laguerre quadrature, namely,

$$\Delta E_B^{\alpha(2)} \sim -e \sum_{\mu \in S(\alpha) \nu \in B} \sum_{\lambda \sigma} \sum_{\gamma \kappa \delta \epsilon} X_{\mu\gamma}^{\alpha} Y_{\nu\kappa}^{\alpha} X_{\lambda\delta}^{\alpha} Y_{\sigma\epsilon}^{\alpha} (\mu\nu | \lambda\sigma) [2(\kappa\gamma | \epsilon\delta) - (\kappa\delta | \epsilon\gamma)] \quad (13)$$

$$X_{\mu\nu}^{\alpha} = \sum_i C_{\mu}^{\alpha} C_{\nu}^{\alpha} e^{(\epsilon_i^{\alpha} - \epsilon_F)} \quad (14)$$

$$Y_{\mu\nu}^{\alpha} = \sum_a C_{\mu}^{\alpha} C_{\nu}^{\alpha} e^{-(\epsilon_a^{\alpha} - \epsilon_F)} \quad (15)$$

Assuming that the rhs of Equation (13) gives the upper limit of  $\Delta E_B^{\alpha(2)}$ , its absolute value can be bounded by adopting the Schwarz inequality

$$|(ij|kl)| \leq \sqrt{|(ij|ij)|} \sqrt{|(kl|kl)|} \quad (16)$$

as

$$\begin{aligned} |\Delta E_B^{\alpha(2)}| &\leq e \sum_{\mu \in S(\alpha) \nu \in B} \sum_{\lambda \sigma} \sum_{\gamma \kappa \delta \epsilon} |X_{\mu\gamma}^{\alpha}| |Y_{\nu\kappa}^{\alpha}| |X_{\lambda\delta}^{\alpha}| |Y_{\sigma\epsilon}^{\alpha}| |(\mu\nu | \lambda\sigma)| [2|(\kappa\gamma | \epsilon\delta)| + |(\kappa\delta | \epsilon\gamma)|] \\ &\leq e \sum_{\mu \in S(\alpha) \nu \in B} \sum_{\lambda \sigma} \sum_{\gamma \kappa \delta \epsilon} |X_{\mu\gamma}^{\alpha}| |Y_{\nu\kappa}^{\alpha}| |X_{\lambda\delta}^{\alpha}| |Y_{\sigma\epsilon}^{\alpha}| A_{\mu\nu}^{\alpha} A_{\lambda\sigma}^{\alpha} [2A_{\kappa\gamma}^{\alpha} A_{\epsilon\delta}^{\alpha} + A_{\kappa\delta}^{\alpha} A_{\epsilon\gamma}^{\alpha}] \\ &\leq e \sum_{\mu \in S(\alpha) \nu \in B} \sum_{\lambda \sigma} \sum_{\gamma \kappa \delta \epsilon} |X_{\mu\gamma}^{\alpha}| |Y_{\nu\kappa}^{\alpha}| |X_{\lambda\delta}^{\alpha}| |Y_{\sigma\epsilon}^{\alpha}| A_{\mu\nu}^{\alpha} A_{\lambda\sigma}^{\alpha} [2A_{\kappa\gamma}^{\alpha} \max(\mathbf{A}^{\alpha}) + A_{\kappa\delta}^{\alpha} A_{\epsilon\gamma}^{\alpha}] \\ &\sim e \sum_{\mu \in S(\alpha) \nu \in B} \sum_{\lambda \sigma} \sum_{\gamma \kappa \delta \epsilon} |X_{\mu\gamma}^{\alpha}| |Y_{\nu\kappa}^{\alpha}| |X_{\lambda\delta}^{\alpha}| |Y_{\sigma\epsilon}^{\alpha}| A_{\mu\nu}^{\alpha} A_{\lambda\sigma}^{\alpha} [2A_{\kappa\gamma}^{\alpha} \max(\mathbf{A}^{\alpha})] \\ &= e \left( \sum_{\lambda \sigma} \sum_{\delta \epsilon} |X_{\lambda\delta}^{\alpha}| |Y_{\sigma\epsilon}^{\alpha}| A_{\lambda\sigma}^{\alpha} \right) \sum_{\mu \in S(\alpha) \nu \in B} \sum_{\gamma \kappa} |X_{\mu\gamma}^{\alpha}| |Y_{\nu\kappa}^{\alpha}| A_{\mu\nu}^{\alpha} [2A_{\kappa\gamma}^{\alpha} \max(\mathbf{A}^{\alpha})] \end{aligned} \quad (17)$$

where  $A_{\mu\nu}^{\alpha} = \sqrt{|(\mu\nu | \mu\nu)|}$ . Here, on the analogy to the scaled opposite-spin MP2 method,<sup>55</sup> the term  $A_{\kappa\delta}^{\alpha} A_{\epsilon\gamma}^{\alpha}$  was omitted owing to its smaller contribution. Because the summation in parentheses in Equation (17) is constant for subsystem  $\alpha$ , the following index can be considered as the magnitude of the contribution from atom  $B$ :

$$e_B^{\alpha} = e \sum_{\mu \in S(\alpha) \nu \in B} \sum_{\gamma \kappa} |X_{\mu\gamma}^{\alpha}| |Y_{\nu\kappa}^{\alpha}| A_{\mu\nu}^{\alpha} [2A_{\kappa\gamma}^{\alpha} \max(\mathbf{A}^{\alpha})] \quad (18)$$

Using the above  $e_B^{\alpha}$  index, we propose the following automatic determination scheme for the buffer region in the DC-MP2 method:

1. Assignment of the initial DC-MP2 buffer region for each subsystem. This may be determined by prior DC-HF calculation.
2. Evaluation of  $e_B^{\alpha}$  from Equation (18).
3. The exclusion of atom  $B$  from the buffer region of subsystem  $\alpha$  if  $e_B^{\alpha}$  is smaller than the energy threshold.

4. Reconstruction of subsystem molecular orbitals  $\{C_p^\alpha\}$  and  $\{e_p^\alpha\}$ , using Equation (2).
5. Evaluation of the subsystem correlation energy,  $\Delta E_{\text{corr}}^{\alpha(2)}$ , from Equation (7).

The additional computational cost for the evaluation of all necessary  $e_B^\alpha$  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  $e_B^\alpha$  of Equation (18) scales with  $O(m^2)$  owing to the summation over  $\gamma$  and  $\kappa$  and the number of  $e_B^\alpha$  to be evaluated scales with  $O(Nm)$ .

### 3 | NUMERICAL ASSESSMENTS

#### 3.1 | Computational details

We implemented the above-mentioned automatically controlled DC-MP2 method to the GAMESS package<sup>56,57</sup> and evaluated its accuracy and efficiency for the different types of systems. In the DC-HF calculations, the inverse temperature parameter,  $\beta$ , 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  $e_{\text{thresh}}^{\text{SCF}} = 0.1 \mu E_h$  and  $r_{\text{ext}} = 3.0 \text{ \AA}$ , the definitions of which are given in our previous paper.<sup>44</sup> The 6-31G(d) basis set<sup>58</sup> was adopted throughout this paper. We introduced the major axis radii of the HF and MP2 localization regions for subsystem  $\alpha$ ,  $I_{\text{local}}^{\text{SCF},\alpha}$  and  $I_{\text{local}}^{\text{corr},\alpha}$ , respectively, to discuss the size of the localization regions determined by the automated DC method.  $I_{\text{local}}^{\text{SCF},\alpha}$  (or  $I_{\text{local}}^{\text{corr},\alpha}$ ) was defined as half of the maximum atom-pair distance in the HF (or MP2) localization region for subsystem  $\alpha$ . The two-electron AO integrals,  $(\mu\nu|\lambda\sigma)$ , were treated in so-called “direct algorithm” manner, that is, the same integrals were calculated repeatedly for every subsystem.

#### 3.2 | Estimation of the atomic MP2 energy contributions

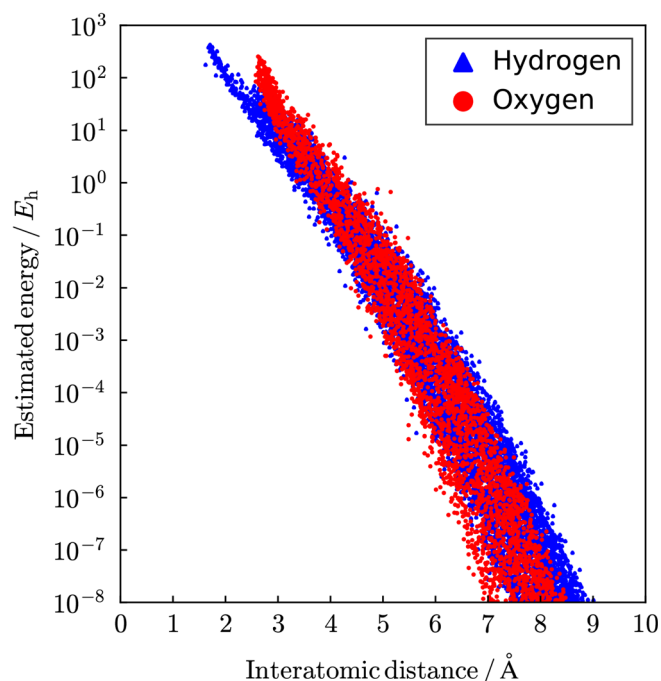
We first applied the present automated DC-MP2 method to a cubic system containing 100 water molecules with weight density of  $1.0 \text{ g cm}^{-3}$ . Each water molecule was adopted as a central region in the DC calculation. To assess the performance of the automated DC-MP2 calculation, the entire system was selected as the initial localization region for every subsystem in the DC-MP2 calculation. Figure 1 shows the estimated MP2 energy contributions from buffer atom  $B$  ( $e_B^\alpha$ ) 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 Figure 1 is probably due to the fact that the summation over AOs at the buffer atom in Equation (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.<sup>59</sup> Note that the estimated energies in Figure 1 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  $\sim 19 E_h$ . This is because that the estimated energy ( $e_B^\alpha$ ) 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,  $e_{\text{thresh}}^{\text{corr}}$ , was set to  $0.1 \mu E_h$  unless otherwise noted.

Next, the dependence of the computational time of  $e_B^\alpha$  on the system size was examined, as shown in Figure 2. 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  $r_b^{\text{in}} = 4.5 \text{ \AA}$  and  $r_b^{\text{out}} = 5.5 \text{ \AA}$ , respectively. The scaling analysis with the double logarithmic plot indicates that the computational time for the evaluation of  $e_B^\alpha$  scales as  $O(N_{\text{water}}^{1.5})$ , which maintains an almost linear-scaling behavior with respect to the entire system size, as was discussed in Section 2.2.

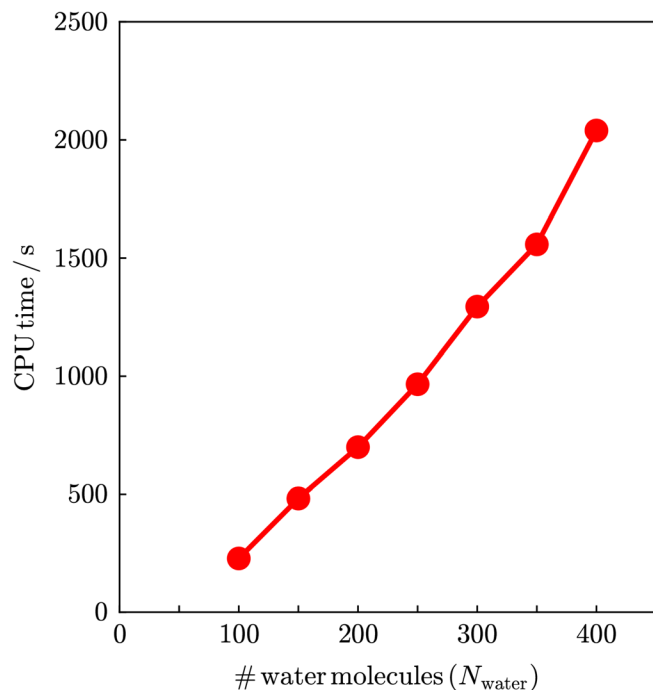
#### 3.3 | Accuracy and computational time of the present method

The accuracy and the computational time of the automated DC-MP2 method were investigated for the cubic water system in Section 3.2. Table 1 shows the energy-threshold ( $e_{\text{thresh}}^{\text{corr}}$ ) dependence of the DC-MP2 correlation energy. Following Section 3.2, each water molecule



**FIGURE 1** 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

was adopted as a central region and the entire system was selected as the initial localization region. The average and standard deviation (SD) of major axis radii ( $\langle r_{\text{local}}^{\text{corr}} \rangle$  and  $\sigma[r_{\text{local}}^{\text{corr}}]$ , respectively) are also given in Table 1. For  $e_{\text{thresh}}^{\text{corr}} = 100 \mu E_h$ , the actual correlation energy error per atom is  $18.37 \mu E_h$ , which is sufficiently smaller than  $e_{\text{thresh}}^{\text{corr}}$ . It should



**FIGURE 2** System-size dependence of the CPU time of the evaluation of  $e_b^{\alpha}$  for the model system containing  $N_{\text{water}}$  water molecules. The initial buffer size for the DC Hartree–Fock (DC-HF) calculation was fixed to  $r_b^{\text{in}} = 4.5 \text{ \AA}$  and  $r_b^{\text{out}} = 5.5 \text{ \AA}$

$e_{\text{thresh}}^{\text{corr}} (\mu E_h)$	$E_{\text{corr}}^{(2)} (E_h)$	(Diff.) ( $\mu E_h \text{ atom}^{-1}$ )	$\langle r_{\text{local}}^{\text{corr},\alpha} \rangle (\text{\AA})$	$\sigma[r_{\text{local}}^{\text{corr},\alpha}] (\text{\AA})$
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			

**TABLE 1**  $e_{\text{thresh}}^{\text{corr}}$  dependences of the DC-MP2 correlation energy and the major axis radius for 100 water cluster system

**TABLE 2** Initial DC-HF buffer-size dependence of the automated DC-MP2 correlation energy and the major axis radius for 100 water cluster system

$r_b^{\text{in}} (\text{\AA})$	$r_b^{\text{out}} (\text{\AA})$	HF Energy ( $E_h$ )	MP2 Energy ( $E_h$ )	(Diff.) ( $\mu E_h \text{ atom}^{-1}$ )	$\langle r_{\text{local}}^{\text{SCF},\alpha} \rangle (\text{\AA})$	$\sigma[r_{\text{local}}^{\text{SCF},\alpha}] (\text{\AA})$	$\langle r_{\text{local}}^{\text{corr},\alpha} \rangle (\text{\AA})$	$\sigma[r_{\text{local}}^{\text{corr},\alpha}] (\text{\AA})$
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					

Note: The energy threshold in the automated DC-HF calculation is  $0.1 \mu E_h$ . Abbreviation: DC-HF, DC Hartree–Fock.

be noted that the MP2 energy error decreases systematically as  $e_{\text{thresh}}^{\text{corr}}$  decreases, while the dependence is not proportional but rather logarithmic to  $e_{\text{thresh}}^{\text{corr}}$ . As with the  $e_{\text{thresh}}$  parameter in automated DC-SCF method,<sup>44</sup> the smaller  $e_{\text{thresh}}^{\text{corr}}$  parameter leads to a larger localization region, which can be confirmed from the average of the major axis radii of all localization regions,  $\langle r_{\text{local}}^{\text{corr}} \rangle$ . Interestingly, the SD of the major axis radii,  $\sigma[r_{\text{local}}^{\text{corr}}]$ , also tends to increase systematically as  $e_{\text{thresh}}^{\text{corr}}$  decreases, except for  $e_{\text{thresh}}^{\text{corr}} = 0.1 \mu E_h$ . 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.

Next, we examined the combination of the present automated DC-MP2 method with the automated DC-HF calculation. Table 2 shows the dependence of the automated DC-MP2 energy on the initial DC-HF inner and outer buffer sizes,  $r_b^{\text{in}}$  and  $r_b^{\text{out}}$ , the definitions of which are given in our previous paper.<sup>44</sup> The averages ( $\langle r_{\text{local}}^{\text{HF}} \rangle$  and  $\langle r_{\text{local}}^{\text{corr}} \rangle$ ) and the SDs ( $\sigma[r_{\text{local}}^{\text{HF}}]$  and  $\sigma[r_{\text{local}}^{\text{corr}}]$ ) 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 ( $\sim 8.5 \mu E_h \text{ atom}^{-1}$ ). The average radius of the DC-HF localization region,  $\langle r_{\text{local}}^{\text{HF}} \rangle$ , is  $7.0\text{--}7.2 \text{ \AA}$ , which is larger than the average radius,  $6.761 \text{ \AA}$ , of the DC-MP2 localization region for  $e_{\text{thresh}}^{\text{corr}} = 0.1 \mu E_h$  given in Table 1. A smaller initial DC-HF buffer size leads to a larger  $\langle r_{\text{local}}^{\text{HF}} \rangle$ , as was also confirmed in the previous study.<sup>44</sup> When combined with the automated DC-HF method,  $\langle r_{\text{local}}^{\text{corr}} \rangle$  becomes smaller than its value when the initial localization region is set to be the entire system. Similarly,  $\sigma[r_{\text{local}}^{\text{corr}}]$  is  $\sim 0.14 \text{ \AA}$  smaller than  $\sigma[r_{\text{local}}^{\text{HF}}]$ .

Next, we applied the proposed method to a covalently bound system, namely, the chignolin protein with 10 amino

**TABLE 3**  $e_{\text{thresh}}^{\text{corr}}$  dependences of the DC-MP2 correlation energy and the major axis radius for chignolin

$e_{\text{thresh}}^{\text{corr}}$ ( $\mu E_h$ )	$E_{\text{corr}}^{(2)}$ ( $E_h$ )	(Diff.) ( $\mu E_h \text{ atom}^{-1}$ )	$\langle I_{\text{local}}^{\text{corr},\alpha} \rangle$ ( $\text{\AA}$ )	$\sigma [I_{\text{local}}^{\text{corr},\alpha}]$ ( $\text{\AA}$ )
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			

**TABLE 4** Initial DC-HF buffer-size dependence of the automated DC-MP2 correlation energy and the major axis radius for chignolin

$r_b^{\text{in}}$ ( $\text{\AA}$ )	$r_b^{\text{out}}$ ( $\text{\AA}$ )	HF Energy ( $E_h$ )	MP2 Energy ( $E_h$ )	(Diff.) ( $\mu E_h \text{ atom}^{-1}$ )	$\langle I_{\text{local}}^{\text{SCF},\alpha} \rangle$ ( $\text{\AA}$ )	$\sigma [I_{\text{local}}^{\text{SCF},\alpha}]$ ( $\text{\AA}$ )	$\langle I_{\text{local}}^{\text{corr},\alpha} \rangle$ ( $\text{\AA}$ )	$\sigma [I_{\text{local}}^{\text{corr},\alpha}]$ ( $\text{\AA}$ )
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					

Note: The energy threshold in the automated DC-HF calculation is  $0.1 \mu E_h$ .

Abbreviation: DC-HF, DC Hartree-Fock.

**TABLE 5** Initial DC-HF buffer-size dependence of the automated DC-MP2 energy and the major axis radius for the  $\beta$ -strand glycine oligomer (GLY)<sub>20</sub>

$r_b^{\text{in}}$ ( $\text{\AA}$ )	$r_b^{\text{out}}$ ( $\text{\AA}$ )	HF Energy ( $E_h$ )	MP2 Energy ( $E_h$ )	(Diff.) ( $\mu E_h \text{ atom}^{-1}$ )	$\langle I_{\text{local}}^{\text{SCF},\alpha} \rangle$ ( $\text{\AA}$ )	$\sigma [I_{\text{local}}^{\text{SCF},\alpha}]$ ( $\text{\AA}$ )	$\langle I_{\text{local}}^{\text{corr},\alpha} \rangle$ ( $\text{\AA}$ )	$\sigma [I_{\text{local}}^{\text{corr},\alpha}]$ ( $\text{\AA}$ )
$e_{\text{thresh}}^{\text{corr}} = 100 \mu E_h$								
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
$e_{\text{thresh}}^{\text{corr}} = 0.1 \mu E_h$								
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					

Note: The energy threshold in the automated DC-HF calculation is set to  $0.1 \mu E_h$  and that in the automated DC-MP2 calculation is set to 100 or  $0.1 \mu E_h$ .

Abbreviation: DC-HF, DC Hartree-Fock.

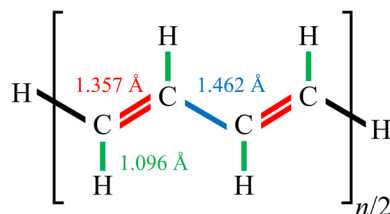
acids. The geometry of chignolin was obtained from the protein data bank (PDBID: 1UAO). Hydrogen atoms were added using the Discovery Studio 2017 R2 software.<sup>60</sup> In the DC calculation, the entire system was divided between the carbonyl C and  $\alpha$ -C atoms, and each of the divided systems was treated as a central region. Table 3 shows the  $e_{\text{thresh}}^{\text{corr}}$  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  $e_{\text{thresh}}^{\text{corr}} = 100 \mu E_h$ , the actual correlation energy error per atom is

$2.82 \mu E_h$ , which is sufficiently smaller than  $e_{\text{thresh}}^{\text{corr}}$ . As was also confirmed in the case of the water system, the MP2 energy error decreases systematically as  $e_{\text{thresh}}^{\text{corr}}$  decreases. Again, the dependence of the error on  $e_{\text{thresh}}^{\text{corr}}$  is rather logarithmic. The smaller  $e_{\text{thresh}}^{\text{corr}}$  leads to the larger  $\langle I_{\text{local}}^{\text{corr},\alpha} \rangle$ , while it leads to the smaller  $\sigma [I_{\text{local}}^{\text{corr},\alpha}]$ , contrary to the case of water system. Comparing Table 3 with Table 1,  $\langle I_{\text{local}}^{\text{corr},\alpha} \rangle$  of chignolin is about  $1.0 \text{\AA}$  larger than that of the water system for the same  $e_{\text{thresh}}^{\text{corr}}$  parameter, reflecting the delocalized electronic nature in the covalently bound system.



Next, we combined this with the automated DC-HF calculation. Table 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 ( $\sim 7.5$  Å) is about 1 Å greater than for the water system ( $\sim 6.5$  Å). Subsequently, the DC-MP2 energy error is also almost constant ( $\sim 0.7 \mu E_h \text{ atom}^{-1}$ ). For this small protein system, in contrast to the result in Table 2 for the water system, the SD 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  $\beta$ -strand glycine oligomer (GLY)<sub>20</sub>, and the result of the calculation are given in Table 5. In Table 5, the DC-MP2 calculations with different  $e_{\text{thresh}}^{\text{corr}}$  were performed to confirm that the present automated DC-MP2 energy error depends primarily on  $e_{\text{thresh}}^{\text{corr}}$  and hardly on the initial buffer radii. For this stretched system, the SD 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. As well as the case of water system, the smaller  $e_{\text{thresh}}^{\text{corr}}$  leads to the larger  $\langle f_{\text{local}}^{\text{corr}} \rangle$  and  $\sigma[f_{\text{local}}^{\text{corr}}]$ .

Finally, the present method was applied to the conjugated polyacetylene chain  $C_{2n}H_{2n+2}$ , shown in Figure 3. All atoms were placed in a plane and the C–C, C=C, and C–H bond lengths were fixed at



**FIGURE 3** Structure of polyacetylene chain system,  $C_{2n}H_{2n+2}$

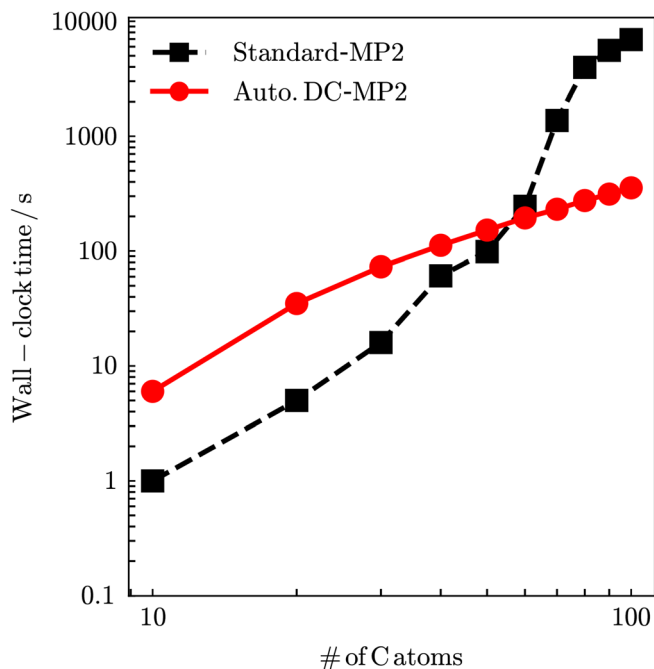
1.462, 1.357, and 1.096 Å, respectively. Each  $C_2H_2$  (or  $C_2H_3$  for edges) unit divided at the C–C single bond was treated as a central region. Table 6 shows the system-size dependence of the standard and DC-MP2 energies. For the automated DC calculations, the initial sizes of the inner and outer buffer regions in the automated DC-HF calculation were set to  $r_b^{\text{in}} = 5.0$  Å and  $r_b^{\text{out}} = 6.5$  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,  $\{e_{\beta}^{\alpha}\}$ , for the atoms in the unit were smaller than the threshold,  $e_{\text{thresh}}^{\text{corr}}$  (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.

For this conjugated system, the dependence of the computational time on the system size was also examined, as shown in Figure 4. 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<sup>61</sup> 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.

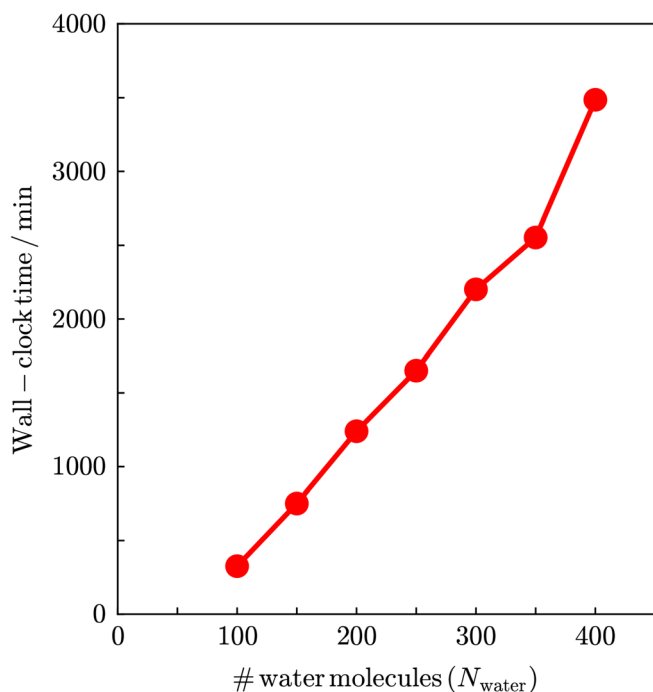
The scaling analysis was also conducted for three-dimensional water cluster systems. Figure 5 shows the dependence of the wall-clock computational time for the DC-MP2 calculation on the number of water molecules,  $N_{\text{water}}$ . The times were measured

**TABLE 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 Energy ( $E_h$ )	Auto. DC-MP2	
		Energy ( $E_h$ )	(Diff.) ( $\mu E_h \text{ atom}^{-1}$ )
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)



**FIGURE 4** 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



**FIGURE 5** System-size dependence of the Wall-clock time of the automated DC-MP2 calculations for the model system containing  $N_{\text{water}}$  water molecules. The initial sizes of the inner and outer buffer regions in the automated DC Hartree-Fock (DC-HF) calculation were set to  $r_b^{\text{in}} = 4.5 \text{ \AA}$  and  $r_b^{\text{out}} = 5.5 \text{ \AA}$ , respectively. The energy threshold in the automated DC-MP2 method,  $e_{\text{thresh}}^{\text{corr}}$ , was set to  $10 \mu E_h$

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  $r_b^{\text{in}} = 4.5 \text{ \AA}$  and  $r_b^{\text{out}} = 5.5 \text{ \AA}$ , respectively. The energy threshold in the automated DC-MP2 method,  $e_{\text{thresh}}^{\text{corr}}$ , was set to  $10 \mu E_h$ . The scaling analysis with the double logarithmic plot indicates that the computational time for the present automated DC-MP2 method scales as  $O(N_{\text{water}}^{1.6})$ , which indicates that the present method also achieves near-linear scaling computational time even for three-dimensional systems.

## 4 | CONCLUDING REMARKS

In this study, we 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,<sup>36</sup> the present scheme reduces the buffer region from the prior DC-HF calculation. We 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,  $e_{\text{thresh}}^{\text{corr}}$ , decreases. We also confirmed that the linear-scaling behavior of the DC-MP2 method is preserved even with the present automation scheme.

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.<sup>37-39</sup> 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.<sup>62</sup> instead of the Schwarz inequality would be one way to provide this improvement.

## ACKNOWLEDGMENTS

Some of the present calculations were performed using the computer facilities at the Research Center for Computational Science, Okazaki, and at the Research Institute for Information Technology, Kyushu University, Japan. This work was supported in part by JSPS KAKENHI Grant Number JP17K05738, Elements Strategy Initiative of MEXT (Ministry of Education, Culture, Sports, Science and Technology) grant number JPMXP0112101003, the Photo-excitonix Project in Hokkaido University, and JST CREST grant number JPMJCR1902, Japan. One of the authors (TF) was supported by MEXT through the Program for Leading Graduate Schools (Hokkaido University "Ambitious Leader's Program"). The Institute for Chemical Reaction Design and Discovery (ICReDD) was established by the World Premier International Research Initiative (WPI), MEXT, Japan. We would like to thank Editage (www.editage.com) for English language editing.



## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Masato Kobayashi  <https://orcid.org/0000-0002-4001-3581>

Tetsuya Taketsugu  <https://orcid.org/0000-0002-1337-6694>

## REFERENCES

- [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*, 3rd ed., John Wiley & Sons, New York 2017.
- [6] S. Goedecker, *Rev. Mod. Phys.* **1999**, *71*, 1085.
- [7] S. Y. Wu, C. S. Jayanthi, *Phys. Rep.* **2002**, *358*, 1.
- [8] R. Zalesny, M. G. Papadopoulos, P. G. Mezey, J. Leszczynski, *Linear-Scaling Techniques in Computational Chemistry and Physics*, Springer, Dordrecht, The Netherlands 2011.
- [9] D. R. Bowler, T. Miyazaki, *Rep. Prog. Phys.* **2012**, *75*, 036503.
- [10] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, 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] Ni Z., Guo Y., Neese F., Li W., Li S., *J. Chem. Theory Comput.* **2021**, <http://dx.doi.org/10.1021/acs.jctc.0c00831>.
- [18] D. G. Fedorov, K. Kitaura, *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*, CRC Press, Boca Raton, FL 2009.
- [19] D. G. Fedorov, K. Kitaura, *J. Chem. Phys.* **2004**, *120*, 6832.
- [20] Y. Nishimoto, 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, M. Dornbach, *J. Chem. Theory Comput.* **2015**, *11*, 484.
- [23] Q. Ma, H.-J. Werner, *J. Chem. Theory Comput.* **2015**, *11*, 5291.
- [24] P. Baudin, P. Ettenhuber, S. Reine, K. Kristensen, T. Kjærgaard, *J. Chem. Phys.* **2016**, *144*, 054102.
- [25] T. Kjærgaard, P. Baudin, D. Bykov, K. Kristensen, P. Jørgensen, *WIREs Comput. Mol. Sci.* **2017**, *7*, e1319.
- [26] W. Yang, T.-S. Lee, *J. Chem. Phys.* **1995**, *103*, 5674.
- [27] M. Kobayashi, H. Nakai, in *Linear-Scaling Techniques in Computational Chemistry and Physics* (Eds: R. Zalesny, M. G. Papadopoulos, P. G. Mezey, J. Leszczynski), Springer, Dordrecht, The Netherlands 2011, p. 97 Ch.5.
- [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, H. Nakai, *J. Chem. Phys.* **2008**, *129*, 044103.
- [38] M. Kobayashi, 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.UTOI, 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* (Eds: C. Dykstra, G. Frenking, K. Kim, G. Scuseria), Elsevier, Amsterdam 2005, p. 1167.
- [57] M. W. Schmidt, K. K. Baldrige, 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*, Dassault Systèmes, San Diego, CA 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.

**How to cite this article:** Fujimori T, Kobayashi M, Taketsugu T. Energy-based automatic determination of buffer region in the divide-and-conquer second-order Møller-Plesset perturbation theory. *J Comput Chem.* 2021;42:620–629. <https://doi.org/10.1002/jcc.26486>

## APPENDIX: MP2 CORRELATION ENERGY DENSITY FOR BONDS

Here, we 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.<sup>63</sup>

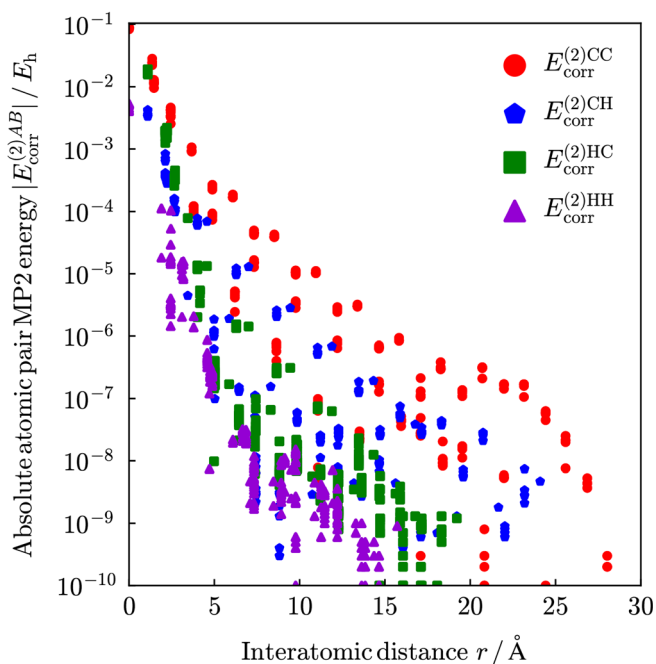
The MP2 correlation energy can be divided into contributions from the atomic pair, expressed as

$$\Delta E_{\text{corr}}^{(2)} = \sum_{A,B} \Delta E_{\text{corr}}^{(2)AB} \quad (\text{A1})$$

$$\Delta E_{\text{corr}}^{(2)AB} = \sum_{i,j}^{\text{occ}} \sum_{a,b}^{\text{vir}} \sum_{\mu \in A} \sum_{\nu \in B} \frac{C_{\mu i} C_{\nu a} (\mu\nu|jb)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} [2(a_i|b_j) - (a_j|b_i)] \quad (\text{A2})$$

Here, we have adopted the overlap separation instead of the electron-pair separation<sup>52-54</sup> to exploit the local nature of the MP2 correlation. This form is also consistent with  $\Delta E_B^{a(2)}$ , Equation (9). Note that  $\Delta E_{\text{corr}}^{(2)AB}$  is different from  $\Delta E_{\text{corr}}^{(2)BA}$  because atoms A and B in  $\Delta E_{\text{corr}}^{(2)AB}$  are associated with the occupied and virtual orbitals, respectively.

The atom-pair MP2 correlation energies,  $\Delta E_{\text{corr}}^{(2)AB}$ , were evaluated for  $\text{C}_{30}\text{H}_{32}$  polyene system with 6-31G(d) basis set. Figure A1 shows the dependence of  $\Delta E_{\text{corr}}^{(2)AB}$  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  $\Delta E_{\text{corr}}^{(2)CC}$  has small hump around  $r = 20 \text{ \AA}$ . Reflecting the small number of correlated electrons around H atom,  $\Delta E_{\text{corr}}^{(2)HH}$  has the smallest contribution at the same distance  $r$ .  $\Delta E_{\text{corr}}^{(2)CH}$  is larger than  $\Delta E_{\text{corr}}^{(2)HC}$ ,



**FIGURE A1** 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

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 3.2).