Theoretical study on the ground state of water clusters

Doctoral Thesis

Theoretical Study on the Ground State of Water Clusters

Division of Material Chemistry

Graduate School of Natural Science and Technology

Kanazawa University

Student ID: 1924022003

Hiroki Sugisawa

Chief supervisor: Tomonori Ida

Submitted: January 2021

CONTENTS

1. INTRODUCTION

1.1. Anomalous Properties of Protons and Hydroxide Ions

Protons (H^+) and hydroxide ions (OH^-) in liquid water have abnormally higher mobility compared with other ions. In general, the ionic mobility in water decreases with a decrease in the ionic radius. This occurs because ions with a smaller radius tend to be more easily hydrated by neighboring water molecules and finally have a larger radius; for example, the ionic mobilities [1] are 5.19, 7.62, 7.91, and 8.09 $10^{-8}\ m^2s^{-1}V^{-1}$ for Na^+, K^+, Cl^-, and Br, respectively. However, the mobilities of protons and hydroxide ions do not follow this trend, with mobilities [1] of $36.23 \times 10^{-8}\ m^2s^{-1}V^{-1}$ for H^+ and $20.64\times 10^{-8}\ m^2s^{-1}V^{-1}$ for OH^-. This abnormal behavior is explained by the Grotthuss mechanism [2], in which protons diffuse by connection and disconnection of the O–H chemical bond, rather than by the usual diffusion process. This mechanism plays important roles in the fuel cells and proton exchange membranes [3–7]; and thus, a thorough understanding analysis is warranted.

In the research of the Grotthuss mechanism [8–15], the particularly difficult part would be that the water molecules in the liquid state constantly change their microstructure [9,15]; this property is thought to complicate the understanding of the anomalous phenomena at the microscale. For example, although previous studies [16,17]

have tried to isolate the IR spectrum of acidic water from that of pure water, this was difficult because the broad absorptions in these spectra did not present assignable features. In contrast, the clusters with a relatively small number of degrees of freedom, such as $(H_2O)_nH^+$ and $(H_2O)_{n-1}OH^-$, contained distinct IR features that help isolation from the convoluted liquid spectrum associated with these anomalous phenomena [13]. Therefore, extracting a core structure of the target phenomena as a cluster, and studying that in place of liquid water is a reasonable strategy; it is clearly easier than performing a direct analysis of liquid water. This is evidenced by a large number of research papers for these clusters [9,13,15,18–39]. However, the spectroscopic experiment, such as the IR, measures the excitation energy from the ground state and cannot observe the molecular behavior at the ground state in which the abnormal phenomena are thought to occur; In addition, experimental approaches generally cannot observe the ground state because of strong quantum effects. Therefore, in the context of interpreting such experimental results, it is not surprising that there has been extensive theoretical research on protonated and deprotonated water clusters [10,15,40–68].

1.2. Theoretical Approaches for the Ground State

As shown in previous works [10,15,40–68], theoretical approaches have an important role in the context of providing an interpretation to the experimental results. The simplest theoretical approach perhaps is the harmonic approximation (HA), which allows an affordable determination of ground-state vibrational energies using the curvature around the potential energy minimum. However, protonated and deprotonated water clusters have been shown to be strongly affected by the anharmonicity of the potential energy surface

[51,54,57,68–77]. For example, the shared proton in the Zundel cation has a flat potential energy surface, such as a fourth-order function [70], and the ground-state energy by the harmonic approximation was adjusted by 1.5–5.5 mHartree by the anharmonic effects [51,68]. Therefore, it is important to take into account the anharmonic effects for the precise examination of the quantum properties of water clusters.

Two approaches can achieve the above requirements. The first approach is a method of optimizing the trial wavefunction based on variational principles [68,78–84]: for example, the vibrational self-consistent field (VSCF) method [78,79] or a variational Monte-Carlo/molecular dynamics (VMC/VMD) method [85–87] combined with Newton or steepest-descent methods [68,80–84]. These approaches can partly describe the anharmonic effects within the representing ability of the chosen trial wavefunction. For example, in a VMD study of protonated water clusters [68], we clarified that the anharmonic effects strongly affected the vibrational ground-state because the wavefunction is delocalized so as to cover multiple potential energy minima. However, as indicated above, these approaches have a weakness relative to variational methods. This weakness is that the accuracy is strongly dependent on the representation ability of the trial wavefunction: naturally, although the exact solution can be obtained if it is possible to prepare and optimize infinite trial wavefunctions, it is explicitly impossible.

Another approach [85,86] is the diffusion Monte-Carlo (DMC) or the variational path integral (VPI), which can compute the exact ground state for quantum many-body systems, numerically. Although these methods are essentially comparable, for the VPI only, a molecular dynamics algorithm has been developed, which is referred to as the variational path integral molecular dynamics (VPIMD) method [87]. As molecular dynamics sampling is expected to be more efficient than Monte-Carlo sampling, the VPI

method is considered to be better than DMC in terms of computational efficiency. Therefore, in this research, we utilized the VPIMD method to analyze the nature of the ground state of water clusters.

1.3. Potential Energy Surface for Water

The accuracy of the VPIMD calculations is decided by the accuracy of the given potential energy surface: thus, the most important procedure in ground-state calculations is the selection of an adequate potential energy surface for the target system. The potential energy surface is computed from electronic structure calculations as the eigenvalue of the electronic Hamiltonian based on the Born-Oppenheimer approximation. However, high computational cost remains a problem even if the target system is relatively small, because many potential energy samples are required in VPIMD simulations. For example, in our estimation, the VPIMD simulation for $H_5O_3^-$ takes approximately 10,000 days for MP2 and approximately 1,000,000 days for CCSD(T). Therefore, it is important to model the potential energy surface to minimize computational time.

In view of these demands, many potential energy surfaces for water clusters have been proposed [44,45,88,89]: the anisotropic site potentials (ASPs) for H_3O^+ and H_2O [45]; the neural network potential energy surface for H_4O_2 [88]; and the OSS series for $(H_2O)_nH^+$ up to $n = 5$ [44]; and the Huang, Braams, and Bowman potential energy surface for $H_3O_2^-$ [89]. The OSS3 PES [44] has been demonstrated to predict the formation energies and structures of the small protonated water clusters with close accuracy to ab initio MP2 results; in contrast, for deprotonated water clusters, to the best of our knowledge, no model potential energy surface has been proposed to describe the

interaction for over eight atoms. Therefore, we employed the OSS3 PES for protonated water clusters and have constructed a new potential energy surface by machine learning.

Regression by a machine learning approach is certainly easier than the traditional method of least squares regression because there is no need to fit the target system to the analytical functions. In previous studies [88,90–92], neural networks and Gaussian process regression have been utilized as machine learning approaches to construct the potential energy surface. The use of these machine learning methods is dependent on the size of the target system: for a system with a relatively small number of degrees of freedom, Gaussian process regression has been reported to be a more efficient tool than the neural network [90,91,93]. In this work, we employed Gaussian process regression because the deprotonated water clusters can be regarded as a relatively small system. However, two problems remained when constructing the potential energy surface using Gaussian process regression: the exponential increase in computational cost the number of learning materials increased; and molecular symmetry. Although the potential energy must be invariant with respect to any symmetric operations, traditional Gaussian process regression does not satisfy this condition. Therefore, we developed two algorithms to improve the accuracy of machine learning of the potential energy surface with a limited amount of training data [91] and to adapt the molecular symmetry to the Gaussian process regression algorithm.

1.4. Goals of the Thesis

In summary, the aim of this thesis was to clarify the nature of the ground state of protonated and deprotonated water clusters, instead of liquid water, as it is easier to

interpret. To calculate the nature of the ground state precisely, the variational path integral molecular dynamics method was selected. For the interatomic interactions, the OSS3 potential energy surface was employed for protonated water clusters, and the machine learning potential energy surface (MLPES) was adopted for deprotonated water clusters to obtain high-quality results. However, before constructing the MLPES, we proposed solutions to the two problems associated with Gaussian process regression.

2. THEORY for NUCLEAR GROUND STATE

2.1. Theory for Nuclear Ground State

In Chapter 2, we explain two theories, which can precisely simulate the nuclear ground state for the many-body atomic system. We first derive a target Hamiltonian using the Born Oppenheimer approximation. Then, section 2.3 explains the variational molecular dynamics (VMD) method combined with the steepest-descent or Newton method, which is a method [68] to minimize the variational energy of the given trial wavefunction based on the variational principle in quantum mechanics. Finally, we briefly introduce another theory, known as the variational path integral molecular dynamics (VPIMD) method, which is an exact numerical method to extract the exact ground state from the arbitrary trial wavefunction.

2.2. Born-Oppenheimer approximation

As expressed above, the purpose of this section is to derive the Hamiltonian for the nuclear ground state. The starting point is the exact non-relativistic, time-independent Hamiltonian $\hat{H}$ for the system consisting of nucleus and electrons:

$$\hat{H}(\boldsymbol{r},\boldsymbol{R}) = \hat{H}_{\mathrm{e}}(\boldsymbol{r},\boldsymbol{R}) + \hat{T}_{\mathrm{n}}(\boldsymbol{R}), \tag{2.1}$$

with the electronic Hamiltonian,

$$\hat{H}_{\mathrm{e}}(\boldsymbol{r},\boldsymbol{R}) = \hat{T}_{\mathrm{e}}(\boldsymbol{r}) + \hat{U}(\boldsymbol{r},\boldsymbol{R}), \tag{2.2}$$

where $\boldsymbol{r}$ is the coordinate vector for the electron, $\boldsymbol{R}$ is the coordinate vector for the nuclei, $\hat{T}_{\mathrm{n}}$ is the nuclear kinetic energy operator, $\hat{T}_{\mathrm{e}}$ is the electronic kinetic energy operator, $\hat{U}$ is the potential energy operator as a coupling term between electrons and nuclei. Moreover, please note that the electronic Hamiltonian includes the nuclear position, $\boldsymbol{R}$.

As the mass of the electron is much smaller than one of the nuclei (the proton has 1800 times bigger mass comparing with electron), we can put the following assumption:

$$\hat{T}_e(\boldsymbol{r}) \gg \hat{T}_n(\boldsymbol{R}) \tag{2.3}$$

Therefore, the Hamiltonian on Equation 2.1 can be rewritten as follow:

$$\hat{H}(\boldsymbol{r},\boldsymbol{R}) \approx \hat{H}_{\mathrm{e}}(\boldsymbol{r},\boldsymbol{R}) \tag{2.4}$$

where nuclear coordinate is no longer a variable and are regarded to a constant. The Schrödinger equation for the electronic Hamiltonian $\hat{H}_{\mathrm{e}}$ can be described by:

$$\hat{H}_{\mathrm{e}}(\boldsymbol{r},\boldsymbol{R})|\Psi_{\mathrm{e}}\rangle = V(\boldsymbol{R})|\Psi_{\mathrm{e}}\rangle. \tag{2.5}$$

where the eigenfunction $|\Psi_{\mathrm{e}}\rangle$ and the eigenvalue V correspond the electronic wavefunction and energy at the electronic ground state, respectively.

Finally, replacing the electronic Hamiltonian on Equation 2.1 with the resulting eigenvalue $V(\boldsymbol{R})$, we can obtain the following Hamiltonian:

$$\hat{H}(\boldsymbol{r},\boldsymbol{R}) \approx \hat{H}(\boldsymbol{R}) = \hat{T}_n(\boldsymbol{R}) + V(\boldsymbol{R}). \tag{2.6}$$

This is referred to as the nuclear Hamiltonian and depends only on the nuclear coordinates

$\boldsymbol{R}$. Moreover, the associated Schrödinger equation can be expressed by:

$$\hat{H}(\boldsymbol{R})|\Psi_0(\boldsymbol{R})\rangle = E_0|\Psi_0(\boldsymbol{R})\rangle. \tag{2.7}$$

where the eigenfunction $|\Psi_0(\boldsymbol{R})\rangle$ and eigenvalue E_0 indicate the nuclear wavefunction and energy at the ground state, respectively. Consequently, in the present thesis, Equations 2.6 and 2.7 are utilized to calculate nuclear ground-state wavefunctions and energies for protonated/deprotonated water clusters.

2.3. Variational Monte-Carlo/Molecular Dynamics Method

In Section 2.3, we explain the methods for minimizing the variational energy of the trial wavefunction based on the variational principle in quantum mechanics. The importance of minimizing the variational energy is discussed in Section 2.3.1 in terms of the variational principle in quantum mechanics. Then, we show the concrete expression of the trial wavefunction in Section 2.3.2. As a method to compute the variational energy for the trial wavefunction, we introduce the variational Monte-Carlo/molecular dynamics (VMC/MD) methods in the present work. Finally, in Section 2.3.4, the steepest-descent method is explained to minimize the variational energy.

2.3.1. Variational Principle in Quantum Mechanics

We derive the variational principle in quantum mechanics to provide the importance of minimizing the variational energy. Let us start from eigenvalues E_i and eigenfunctions $|\Psi_i\rangle$ of the nuclear Hamiltonian on Equation 2.6:

$$\hat{H}|\Psi_i\rangle = E_i|\Psi_i\rangle, \qquad \langle\Psi_j|\Psi_i\rangle = \delta_{ji} \tag{2.8}$$

where δ_{ij} is Kronecker delta and the eigenvalues are numbered in ascending order:

$$E_0 \leq E_1 \leq E_2 \leq \cdots. \tag{2.9}$$

Then, using the closure $1 = \sum_{i=0}^{\infty} |\Psi_i\rangle\langle\Psi_i|$ enables the expansion of a normalized trial wavefunction $|\Phi_{\mathrm{T}}\rangle$ as:

$$|\Phi_{\mathrm{T}}\rangle = \sum_{i=0}^{\infty} c_i \, |\Psi_i\rangle, \tag{2.10}$$

where c_i is the expansion constant. The expectation value of the nuclear Hamiltonian $\hat{H}$ can be evaluated by:

$$\begin{aligned} E_{\mathrm{v}} &= \langle\Phi_{\mathrm{T}}|\hat{H}|\Phi_{\mathrm{T}}\rangle = \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} c_j^* c_i \langle\Psi_j|\hat{H}|\Psi_i\rangle \\ &= \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} c_j^* c_i E_i \delta_{ij} \\ &= \sum_{i=0}^{\infty} |c_i|^2 E_i. \end{aligned} \tag{2.11}$$

Here, by the large/small relation on Equation 2.9, the following relation is derived:

$$E_{\mathrm{v}} \geq E_0 \tag{2.12}$$

This fact shows that the trial wavefunction $|\Phi_{\mathrm{T}}\rangle$ becomes the ground state wavefunction $|\Psi_0\rangle$ when $E_0 = E_{\mathrm{v}}$. Therefore, minimizing the variational energy can make the trial wavefunction closer to the ground state wavefunction.

2.3.2. Trial Wavefunction for Nuclear Ground States

We start by considering a quantum system consisting of N atoms whose coordinates

are denoted by $\boldsymbol{R}_i,\ i = 1, \cdots N$. The local mode, S_ν, for this system is defined as: [94]

$$S_\nu = \left|R_i - R_j\right|, \qquad 1 \le i < j \le N, \qquad 1 \le \nu \le N_{pair}, \tag{2.13}$$

where $N_{pair} = N(N-1)/2$ is the total number of the local modes and $\left|R_i - R_j\right|$ is the interatomic distance between ith and jth atoms. The ground state is modeled by the following trial wavefunction of the Gaussian form:

$$\Phi_{\mathrm{T}}(\boldsymbol{R}, \mathbf{A}) = \exp\left(\sum_{\mu}^{N_{pair}} \sum_{\nu}^{N_{pair}} \Delta S_\mu A_{\mu\nu} \Delta S_\nu\right), \tag{2.14}$$

where $A_{\mu\nu}$ denotes an element of a variational parameter matrix $\mathbf{A}$ and

$$\Delta S_\nu = S_\nu - S_\nu^0. \tag{2.15}$$

The constant S_ν^0 in the trial wavefunction associated with the νth local mode is set to the equilibrium interatomic distance in the potential energy surface.

2.3.3. Variational Monte-Carlo/Molecular Dynamics method

The goal in this section is to transform the integral into a suitable form for the Monte-Carlo or the molecular dynamics simulations to evaluate the variational energy E_{V} for the given trial wavefunction $|\Phi_{\mathrm{T}}\rangle$. Following Section 2.3.2, we consider the N atomic system described by a trial wavefunction $\Phi_{\mathrm{T}}(\boldsymbol{R}, \mathbf{A})$ that consists of the vector of the atomic coordinates $\boldsymbol{R}$ and the matrix of variational parameters $\mathbf{A}$. The expectation value of the Hamiltonian is written by the following integral:

$$E_{\mathrm{V}} = \frac{\int d\boldsymbol{R}\ \Phi_{\mathrm{T}}^*(\boldsymbol{R}) \widehat{H} \Phi_{\mathrm{T}}(\boldsymbol{R})}{\int d\boldsymbol{R}\ |\Phi_{\mathrm{T}}(\boldsymbol{R})|^2} = \frac{\int d\boldsymbol{R}\ |\Phi_{\mathrm{T}}(\boldsymbol{R})|^2 E_L(\boldsymbol{R})}{\int d\boldsymbol{R}\ |\Phi_{\mathrm{T}}(\boldsymbol{R})|^2}, \tag{2.16}$$

where $E_L(\boldsymbol{R})$ denotes a local energy defined by:

$$E_L(\boldsymbol{R}) = \frac{\widehat{H}\Phi_{\mathrm{T}}(\boldsymbol{R})}{\Phi_{\mathrm{T}}(\boldsymbol{R})}. \tag{2.17}$$

Equation 2.16 indicates that the variational energy E_{v} can be computed as the statistical average of the local energy $E_L(\boldsymbol{R}_i)$:

$$E_{\mathrm{v}} = \frac{1}{N}\sum_{i}^{N} E_L(\boldsymbol{R}_i), \tag{2.18}$$

where the nuclear coordinates $\boldsymbol{R}_i$ are generated based on the distribution function, $\rho(\boldsymbol{R}) = |\Phi_{\mathrm{T}}(\boldsymbol{R})|^2 / \int d\boldsymbol{R}\, |\Phi_{\mathrm{T}}(\boldsymbol{R})|^2$. Then, we introduce the following effective potential $V_{\mathrm{VMC}}(\boldsymbol{R})$:

$$\rho(\boldsymbol{R}) \equiv e^{-V_{VMC}(\boldsymbol{R})/k_B T}. \tag{2.19}$$

Here, k_B is the Boltzmann constant and T is an arbitrary parameter that can be regarded as temperature. The distribution $\rho(\boldsymbol{R})$ is generated by canonical molecular dynamics method in accordance with the following classical Hamiltonian H_{VMD}:

$$H_{\mathrm{VMD}} = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i'} + V_{\mathrm{VMC}}(\boldsymbol{R}) \tag{2.20}$$

where $\boldsymbol{p}_i$ and m_i are the momentum and the associated mass of the ith atom, respectively. A single Nosé–Hoover chain thermostat [95] is attached to the system to control the parameter T. Finally, the expectation values of physical quantities are evaluated by taking the average of suitable estimators along the molecular dynamics trajectory.

As explained above, the classical Hamiltonian on Equation 2.20 enables to solve the integral on Equation 2.16 by the molecular dynamics algorithms; thus, this method is referred to as the variational molecular dynamics (VMD) method. In contrast, if the Monte-Carlo algorithms are employed, its method is referred to as the variational Monte

Carlo (VMC) method. Both VMC and VMD methods are expected to have the following great advantages [96] for obtaining quantum expectation values. It is possible to describe many-body systems far more compactly than possible with other methods such as vibrational self-consistent field methods [78,79] even in complex systems including explicit two-body and higher-order correlation terms. The computational cost increases only between N^2 and N^3 even if systems being treated increase in size, unlike vibrational SCF methods that scale with larger powers of system size.

2.3.4. Optimization of Variational Parameters

We finally describe the method [91] for optimizing variational parameters in the trial wavefunctions using the VMD method to minimize the variational energy, E_{v}. Two methods are currently known to optimize the variational parameters: variance minimization [97,98] and energy minimization [80–84]. However, in the present study, we have adopted the energy minimization approach that has been reported to be numerically more efficient than the variance optimization approach [99,100].

In order to find the variational parameters corresponding to the minimum variational energy, the energy gradient $\boldsymbol{g}$ and energy Hessian $\mathbf{h}$ regarding the variational parameters $A_{\mu\nu}$ are summarized below. Using the hermiticity of the Hamiltonian $\widehat{H}$, the energy gradient g_α is written as follows: [101]

$$g_\alpha = \langle \Upsilon_\alpha E_L \rangle - \langle \Upsilon_\alpha \rangle \langle E_L \rangle, \qquad \left(\alpha = 1, \dots, N_{pair}^2\right), \tag{2.21}$$

where

$$\Upsilon_\alpha = \frac{2}{\Phi_T(\mathbf{A})} \frac{\partial \Phi_T(\mathbf{A})}{\partial A_\alpha}. \tag{2.22}$$

Here and hereafter, the combination $\mu\nu$ of two subscripts is expressed by a single symbol

α or β for simplicity. It is worthwhile to note that $E_L(\boldsymbol{R})$ becomes constant everywhere in the configuration space if the trial wavefunction $\Phi_T(\boldsymbol{R}, \mathbf{A})$ is the exact ground state wavefunction. This is referred to be the zero-variance principle [100]; the energy gradients g_α in Equation 2.21 would vanish in the zero-variance condition.

On the other hand, the energy Hessian $\mathbf{h}$ is evaluated by the following form: [83]

$$h_{\alpha\beta} = a_{\alpha\beta} + b_{\alpha\beta} + c_{\alpha\beta}, \tag{2.23}$$

where

$$a_{\alpha\beta} = \langle \Upsilon_{\alpha\beta} E_L \rangle - \langle \Upsilon_{\alpha\beta} \rangle \langle E_L \rangle, \tag{2.24}$$

$$b_{\alpha\beta} = \langle \Upsilon_\alpha \Upsilon_\beta E_L \rangle - \langle \Upsilon_\alpha \Upsilon_\beta \rangle \langle E_L \rangle - \langle \Upsilon_\alpha \rangle g_\beta - \langle \Upsilon_\beta \rangle g_\alpha, \tag{2.25}$$

$$c_{\alpha\beta} = \frac{1}{2}\left[\left\langle \Upsilon_\alpha \frac{\partial E_L}{\partial A_\beta} \right\rangle - \langle \Upsilon_\alpha \rangle \left\langle \frac{\partial E_L}{\partial A_\beta} \right\rangle + \left\langle \Upsilon_\beta \frac{\partial E_L}{\partial A_\alpha} \right\rangle - \langle \Upsilon_\beta \rangle \left\langle \frac{\partial E_L}{\partial A_\alpha} \right\rangle \right], \tag{2.26}$$

and

$$\Upsilon_{\alpha\beta} = \frac{\partial \Upsilon_\alpha}{\partial A_\beta} = 2\left[\frac{1}{\Phi_T(\mathbf{A})} \frac{\partial^2 \Phi_T(\mathbf{A})}{\partial A_\alpha \partial A_\beta} - \frac{1}{\Phi_T^2(\mathbf{A})} \frac{\partial \Phi_T(\mathbf{A})}{\partial A_\alpha} \frac{\partial \Phi_T(\mathbf{A})}{\partial A_\beta} \right]. \tag{2.27}$$

Though a conventional expression of Equation 2.26 is given by $c_{\alpha\beta} = \langle \Upsilon_\alpha \cdot \partial E_L / \partial A_\beta \rangle$ [81], the use of the Hessian computed by the conventional $c_{\alpha\beta}$ is known to make the numerical computation unstable. Therefore, we here introduced Equation 2.26, which is another expression for reducing the numerical instability proposed by Umrigar and Filippi in Ref. [83]. Please note that $\Upsilon_\alpha = 2\Delta S_\nu \Delta S_\mu$ and $\Upsilon_{\alpha\beta} = 0$ when the trial wavefunction on Equation 2.14 is adopted.

Here, we introduce methods of optimizing the trial wavefunction using two quantities explained above. One is the steepest descent method. In this case, the variational parameters are updated as follows:

$$A_\alpha^{\mathrm{new}} = A_\alpha^{\mathrm{old}} - w g_\alpha\left(A_\alpha^{\mathrm{old}}\right), \tag{2.28}$$

where w is a positive constant that determines the weight of the value to be updated in one iteration. Another is the Newton method, which update the variational parameters as follows:

$$\boldsymbol{A}^{\mathrm{new}} = \boldsymbol{A}^{\mathrm{old}} - \mathbf{h}^{-1}\left(\boldsymbol{A}^{\mathrm{old}}\right) \cdot \boldsymbol{g}\left(\boldsymbol{A}^{\mathrm{old}}\right), \tag{2.29}$$

where $\mathbf{h}^{-1}\left(\boldsymbol{A}^{\mathrm{old}}\right)$ is the inverse matrix of the Hessian $\mathbf{h}\left(\boldsymbol{A}^{\mathrm{old}}\right)$. Since the Newton method is known to be stabilized by adding a positive constant w_{diag} to the diagonal elements of the Hessian $\mathbf{h}\left(\boldsymbol{A}^{\mathrm{old}}\right)$ [83]: that is, the Hessian is usually modified as follows: $h_{ij} \rightarrow h_{ij} + w_{diag}\delta_{ij}$. This operation rotates the direction of energy optimization from the Newtonian-direction to the steepest descent-direction. In the case of the Newton method, the total energy is expected to quadratically converge and faster than in the case of the steepest descent method; however, the evaluation of the Hessian $\mathbf{h}$ (Equation 2.23) requires the huge computational cost that the square of the evaluation of the gradient $\boldsymbol{g}$. In addition, it should be noted that all the quantities E_V, $\boldsymbol{g}$, and $\mathbf{h}$ evaluated by the VMC/VMD method include statistical error. That is, there is a possibility that the Newton method is more computationally unstable than the steepest-descent method because more parameters are updated in the Newton method. Therefore, in this thesis, we employed the steepest descent method to minimize the variational energy of the trial wavefunction.

2.4. Variational Path Integral Molecular Dynamics Method

2.4.1. Variational Path Integral

In this section, we explain the variational path integral molecular dynamics (VPIMD) method which is one of the exact numerical methods for many-body quantum ground states. The starting point is the following relation, which projects out the exact ground state $|\Psi_0\rangle$ from the trial wavefunction $|\Phi_T\rangle$ for a target system [85,86]:

$$|\Psi_0\rangle = \lim_{\beta\to\infty} e^{-\frac{\beta}{2}\hat{H}}\,|\Phi_T\rangle, \tag{2.30}$$

where β is a real number parameter referred to as the total projection time. Then, taking a scalar product of the exact ground state $|\Psi_0\rangle$ yields the following pseudo partition function Z_0 [102]:

$$\begin{aligned} Z_0 &= \langle\Psi_0|\Psi_0\rangle \\ &= \lim_{\beta\to\infty}\left\langle\Phi_T\middle|e^{-\beta\hat{H}}\middle|\Phi_T\right\rangle \\ &= \lim_{\beta\to\infty}\int\int d\boldsymbol{R}\,d\boldsymbol{R}'\,\langle\Phi_T|\boldsymbol{R}\rangle\langle\boldsymbol{R}\middle|e^{-\beta\hat{H}}\middle|\boldsymbol{R}'\rangle\langle\boldsymbol{R}'|\Phi_T\rangle, \end{aligned} \tag{2.31}$$

where the closure relation for the coordinate basis, $\int d\boldsymbol{R}\,|\boldsymbol{R}\rangle\langle\boldsymbol{R}| = 1$, is utilized. A matrix element $\langle\boldsymbol{R}|e^{-\beta\hat{H}}|\boldsymbol{R}'\rangle$ is found to be the same as a density matrix at the inverse temperature β, $\rho(\boldsymbol{R},\boldsymbol{R}';\beta)$ [103,104]. For a general nuclear Hamiltonian on Equation 2.6, the matrix element $\langle\boldsymbol{R}|e^{-\beta\hat{H}}|\boldsymbol{R}'\rangle$ cannot be evaluated analytically. However, if the argument of the exponential can be regarded small, the matrix element could be evaluated approximately. But since β need not be small, the argument of the exponential can be made arbitrarily small by writing:

$$e^{-\beta H} = \left[e^{-\beta H/M}\right]^M = e^{-\beta H/M} e^{-\beta H/M} \dots e^{-\beta H/M}. \quad (2.32)$$

Substituting Equation 2.32 into 2.31 yields for the pseudo partition function,

$$Z_0 = \lim_{M\to\infty} \int\int d\boldsymbol{R}\, d\boldsymbol{R}' \, \langle \Phi_T|\boldsymbol{R}\rangle\langle \boldsymbol{R}\left|\left[e^{-\beta H/M}\right]^M\right|\boldsymbol{R}'\rangle\langle \boldsymbol{R}'|\Phi_T\rangle. \quad (2.33)$$

In addition, we introduce the second order Suzuki-Trotter decomposition [105], which states that:

$$e^{-\beta H} = e^{-\beta(\hat{T}+\hat{V})} = e^{-\beta\hat{V}/2} e^{-\beta\hat{T}} e^{-\beta\hat{V}/2} + O(\beta^3). \quad (2.34)$$

For the higher order expansion of this operator [106–108], please confirm Appendix 7.1. Thus, the pseudo partition function can be written as

$$Z_0 = \lim_{M\to\infty} \int\int d\boldsymbol{R}\, d\boldsymbol{R}' \times$$

$$\langle \Phi_T|\boldsymbol{R}\rangle\langle \boldsymbol{R}\left|\left[e^{-\beta\hat{V}/2M} e^{-\beta\hat{T}/M} e^{-\beta\hat{V}/2M} + O((\beta/M)^3)\right]^M\right|\boldsymbol{R}'\rangle\langle \boldsymbol{R}'|\Phi_T\rangle. \quad (2.35)$$

The Trotter theorem guarantees that, at the limit $M \to \infty$, the integral in Equation 2.35 will converge to the exact pseudo partition function at the ground state. Matrix elements of the individual factors $\left[e^{-\beta\hat{V}/2M} e^{-\beta\hat{T}/M} e^{-\beta\hat{V}/2M} + O((\beta/M)^3)\right]$ are obtained by inserting the closure relation into the coordinate basis, $\int d\boldsymbol{R}|\boldsymbol{R}\rangle\langle\boldsymbol{R}| = 1$ between each of the $M-1$ pairs of factors in Equation 2.35, giving the following expression,

$$Z_0 = \int d\boldsymbol{R}\, d\boldsymbol{R}'\, d\boldsymbol{R}^1 \dots d\boldsymbol{R}^{M-1} \, \langle \Phi_T|\boldsymbol{R}\rangle\langle \boldsymbol{R}|\left[e^{-\beta\hat{V}/2M} e^{-\beta\hat{T}/M} e^{-\beta\hat{V}/2M}\right]|\boldsymbol{R}^1\rangle$$

$$\times \langle \boldsymbol{R}^1|\left[e^{-\beta\hat{V}/2M} e^{-\beta\hat{T}/M} e^{-\beta\hat{V}/2M}\right] \dots |\boldsymbol{R}^{M-1}\rangle$$

$$\times \langle \boldsymbol{R}^{M-1}|\left[e^{-\beta\hat{V}/2M} e^{-\beta\hat{T}/M} e^{-\beta\hat{V}/2M}\right]|\boldsymbol{R}'\rangle\langle \boldsymbol{R}'|\Phi_T\rangle$$

$$= \int d\boldsymbol{R}^0 \dots d\boldsymbol{R}^M \, \langle \Phi_T | \boldsymbol{R}^0 \rangle \left\{ \prod_{s=0}^{M-1} \left\langle \boldsymbol{R}^{(s)} \middle| e^{-\beta \hat{V}/2M} e^{-\beta \hat{T}/M} e^{-\beta \hat{V}/2M} \middle| \boldsymbol{R}^{(s+1)} \right\rangle \right\} \langle \boldsymbol{R}^M | \Phi_T \rangle, \qquad (2.36)$$

where, in Equation 2.36, the integration variables $\boldsymbol{R}$ and $\boldsymbol{R}'$ are renamed to $\boldsymbol{R}^0$ and $\boldsymbol{R}^M$, respectively. The matrix elements appearing in Equation 2.36 can now be evaluated analytically: The sought-after matrix element becomes:

$$\left\langle \boldsymbol{R}^{(s)} \middle| e^{-\beta \hat{V}/2M} e^{-\beta \hat{T}/M} e^{-\beta \hat{V}/2M} \middle| \boldsymbol{R}^{(s+1)} \right\rangle$$

$$= e^{-\beta \left(V(\boldsymbol{R}^{(s)}) + V(\boldsymbol{R}^{(s+1)}) \right)/2M} \left\langle \boldsymbol{R}^{(s)} \middle| e^{-\beta \hat{T}/M} \middle| \boldsymbol{R}^{(s+1)} \right\rangle, \qquad (2.37)$$

which follows from the fact that the potential is only a function of position, and thus its action on the coordinate eigenstate is equivalent to multiplication by the eigenvalue. The kinetic energy operator $\hat{T}$ is diagonal in the momentum rather than coordinate representation, and hence a basis change is required in order to evaluate the remaining matrix element in Equation 2.37:

$$\begin{aligned} \left\langle \boldsymbol{R}^{(s)} \middle| e^{-\beta \hat{T}/M} \middle| \boldsymbol{R}^{(s+1)} \right\rangle &= \int d\boldsymbol{p} \, d\boldsymbol{p}' \, \left\langle \boldsymbol{R}^{(s)} \middle| \boldsymbol{p} \right\rangle \langle \boldsymbol{p} | e^{-\frac{\beta \hat{\boldsymbol{p}}^2}{M2m}} | \boldsymbol{p}' \rangle \left\langle \boldsymbol{p}' \middle| \boldsymbol{R}^{(s+1)} \right\rangle \\ &= \frac{1}{(2\pi\hbar)^3} \int d\boldsymbol{p} \, d\boldsymbol{p}' \, e^{i\boldsymbol{R}^{(s)} \cdot \boldsymbol{p}/\hbar} e^{-\frac{\beta \boldsymbol{p}^2}{M2m}} e^{-i\boldsymbol{R}^{(s+1)} \cdot \boldsymbol{p}'/\hbar} \delta(\boldsymbol{p} - \boldsymbol{p}') \\ &= \frac{1}{(2\pi\hbar)^3} \int d\boldsymbol{p} \, e^{i\boldsymbol{p} \cdot (\boldsymbol{R}^{(s)} - \boldsymbol{R}^{(s+1)})/\hbar} e^{-\frac{\beta \boldsymbol{p}^2}{M2m}}. \end{aligned} \qquad (2.38)$$

The momentum integral in Equation 2.38 can be carried out by completing the square, yielding:

$$\left\langle \boldsymbol{R}^{(s)} \middle| e^{-\beta \hat{T}/M} \middle| \boldsymbol{R}^{(s+1)} \right\rangle = \left(\frac{mM}{2\pi\beta\hbar^2} \right)^{3/2} e^{-\frac{mM}{2\beta\hbar} \left(\boldsymbol{R}^{(s)} - \boldsymbol{R}^{(s+1)} \right)^2}. \qquad (2.39)$$

Substituting Equations 2.39 and 2.37 into Equation 2.36 and taking the limit $M \to \infty$ gives the following expression for the canonical pseudo partition function:

$$Z_0 = \lim_{M\to\infty}\left(\frac{mM}{2\pi\beta\hbar^2}\right)^{3M/2}\int d\boldsymbol{R}^0 \ldots d\boldsymbol{R}^M \times$$

$$\langle\Phi_T|\boldsymbol{R}^0\rangle \exp\left\{-\sum_{s=0}^{M-1}\frac{mM}{2\beta\hbar^2}\left(\boldsymbol{R}^{(s)}-\boldsymbol{R}^{(s+1)}\right)^2+\frac{\beta}{M}V\left(\boldsymbol{R}^{(s)}\right)\right\}\langle\boldsymbol{R}^M|\Phi_T\rangle. \quad (2.40)$$

For finite M, Equation 2.40 is equivalent to a configurational integral for a linear polymer with nearest neighbor harmonic couplings interacting with an external potential $V(\boldsymbol{r}_i)/M$: however, the end points of the linear polymer, $\boldsymbol{R}^0$ and $\boldsymbol{R}^M$ are affected by the trial wavefunctions $\langle\Phi_T|\boldsymbol{R}^0\rangle$ and $\langle\boldsymbol{R}^M|\Phi_T\rangle$. The pseudo partition function is rewritten as

$$Z_0 \propto \int d\boldsymbol{R}^{(0)} \ldots d\boldsymbol{R}^{(M)} \times$$

$$\Phi_T\left(\boldsymbol{R}^{(0)}\right)\exp\left\{-\Delta\tau\left[\sum_{s=0}^{M-1}\frac{m}{2\hbar^2\Delta\tau^2}\left(\boldsymbol{R}^{(s)}-\boldsymbol{R}^{(s+1)}\right)^2+V\left(\boldsymbol{R}^{(s)}\right)\right]\right\}\Phi_T\left(\boldsymbol{R}^{(M)}\right). \quad (2.41)$$

where $\Delta\tau = \beta/M$ and S indicate the imaginary time increment and the discretized imaginary time action, respectively.

2.4.2. Molecular Dynamics Algorithm for Variational Path Integral

We next explain a wat to sample configurations of the linear polymers. To do this, it would be useful to introduce the following frequency to Equation 2.41.

$$\omega = \frac{1}{\hbar\Delta\tau}, \quad (2.42)$$

Moreover, it is also helpful to introduce a set of momentum integrations into Equation 2.41:

$$Z_0 \propto \int d\boldsymbol{R}^{(0)} \dots d\boldsymbol{R}^{(M)}\, d\boldsymbol{p}_1 \dots d\boldsymbol{p}_M \times$$

$$\Phi_T\left(\boldsymbol{R}^{(0)}\right) \exp\left\{-\Delta\tau \left[\sum_{s=0}^{M-1} \frac{\boldsymbol{p}^{(s)^2}}{2m'^{(s)}} + \frac{1}{2} m\omega^2 \left(\boldsymbol{R}^{(s)} - \boldsymbol{R}^{(s+1)}\right)^2 + V\left(\boldsymbol{R}^{(s)}\right)\right]\right\} \Phi_T\left(\boldsymbol{R}^{(M)}\right). \quad (2.43)$$

where $\boldsymbol{p}_i^{(s)}$ and $m_i'^{(s)}$ are a fictitious momentum and the associated fictitious mass of an ith atom at an sth time slice, respectively. Equation 2.43 has the form of a phase space integral for a $(M+1)$-particle system with two interactions, the harmonic interaction between $\boldsymbol{R}^{(s)}$ and $\boldsymbol{R}^{(s+1)}$ and the potential, $V(\boldsymbol{R}^{(s)})$ and can be solved using the molecular dynamics algorithms. We define the following Hamiltonian:

$$H_{\mathrm{VPIMD}} = \sum_{s=0}^{M} \left[\sum_{i=1}^{N} \frac{\boldsymbol{p}_i^{(s)^2}}{2m_i'^{(s)}} + S\left(\left\{\boldsymbol{R}^{(s)}\right\}\right)\right] - \ln \Phi_T\left(\boldsymbol{R}^{(0)}\right) - \ln \Phi_T\left(\boldsymbol{R}^{(M)}\right), \quad (2.44)$$

where

$$S\left(\left\{\boldsymbol{R}^{(s)}\right\}\right) = \frac{1}{2} m\omega^2 \left(\boldsymbol{R}^{(s)} - \boldsymbol{R}^{(s+1)}\right)^2 + V\left(\boldsymbol{R}^{(s)}\right) \quad (2.45)$$

Then, we derive equations of motion via Hamilton's canonical equations. Single Nosé-Hoover chain thermostat [95] is attached to each degree of freedom [109,110] to generate the canonical distribution. In the present study, the staging coordinates are utilized to enhance the sampling efficiency of the linear polymer configurations [111]; definitions for the staging variables can be found in Ref. [110,112].

3. POTENTIAL ENERGY SURFACE

As demonstrated in Chapter 2, the variational path integral molecular dynamics (VPIMD) method can compute exactly molecular properties for the many-body ground state according to the given Hamiltonian, $\hat{H}(\boldsymbol{R}) = \hat{T}(\boldsymbol{R}) + \hat{V}(\boldsymbol{R})$. The accuracy of the VPIMD simulations is clearly dependent on the quality of the potential energy surface (PES), $\hat{V}(\boldsymbol{R})$. Electronic structure calculations are often utilized to obtain a high-quality PES. However, the VPIMD simulation requires many potential energy calculations and, when using electronic structure calculations, the high computational cost becomes a serious problem. Therefore, the purpose of this chapter is to describe the reduced computational cost in the VPIMD simulation by modeling the potential energy surface as a simple function of the nuclear coordinates.

3.1. Electronic Structure Calculation

Electronic structure calculations give numerical solutions of the Schrödinger equation based on the electronic Hamiltonian on Equation 2.2. In this section, we discuss the electronic structure calculations appropriate for describing the interatomic interactions of water clusters. These calculations are important because their results are used to model the PES. For non-periodic molecular systems such as water clusters, the

following theories using a coordinate basis may be effective: molecular orbital theory (MOT), density functional theory (DFT), and valence bond theory (VBT). However, VBT is for conditioning the shape of the electronic wavefunction to chemically understand the molecular orbitals and is not an efficient tool for high-accuracy PESs. Therefore, we mainly discuss the MOT and DFT.

First, we shall explain MOT. In general, this theory consists of three approximations: The Born-Oppenheimer approximation in Section 2.2, the independent electron approximation (also called the orbital approximation), and the linear combination of atomic orbitals (LCAO) approximation. The most famous MOT is the Hartree-Fock (HF) method [113]: although the HF method generally gives more than 99% of the experimental value and an overlap of 95% with electronic wavefunctions obtained using more accurate MOTs, many important chemical phenomena occur in the remaining 1% or 5%. Such errors can be described by using post–Hartree-Fock approaches, such as the second-order Møller–Plesset perturbation theory (MP2) [114], coupled-cluster singles and doubles augmented by a perturbative treatment of triple excitations (CCSD(T)) [115,116], and configuration interaction (CI) methods.

Next, we shall introduce DFT. The accuracy of DFT relies strongly on the specific functional used to approximate the electronic exchange and correlation contributions. The most famous is the Becke hybrid functional for exchange and the Lee−Yang−Parr functional for correlation (B3LYP) [117], which is evidenced by the number of citations. However, the DFT calculations using B3LYP often underestimate the hydrogen bonding energy of the water clusters with respect to MP2 or CC results [118,119]. In contrast, for water clusters, MP2 produces results that are in excellent agreement with CC theory, despite their low computational cost [120]. Therefore, in the present research, MP2 was

employed mainly as the ab initio method; its validity is discussed in Section 5.2.2.

3.2. Regression using Least Squares Method

3.2.1. Preparation

First, we consider fitting the PES computed by the electronic structure calculations, V, using arbitrary model functions. For this, we assume that the potential energy is a function dependent only on the D-dimensional vector of the nuclear coordinate, $\boldsymbol{R}$:

$$V = f(\boldsymbol{R}), \qquad R = [R_1, \dots, R_D]^T \tag{3.1}$$

We prepare a training set $\mathcal{D}$ of m training points:

$$\mathcal{D} = \{(V_i, \boldsymbol{R}_i) | i = 1, \dots, m\}, \tag{3.2}$$

And rewrite it as the following:

$$\mathcal{D} = \{\boldsymbol{V}, \mathbf{X}\}, \tag{3.3}$$

using vector $\boldsymbol{V}$ and matrix $\mathbf{X}$ notations, where:

$$\boldsymbol{V} = [V_1, \dots, V_m]^T, \qquad \mathbf{X} = [\boldsymbol{R}_1, \dots, \boldsymbol{R}_m]^T. \tag{3.4}$$

The $m \times D$ matrix $\mathbf{X}$ is called the design matrix.

3.2.2. Linear Regression

In the least squares method, the simplest regression model – multiple regression – is given by the following equation:

$$\widehat{\boldsymbol{V}} = \begin{pmatrix} \hat{V}_1 \\ \hat{V}_2 \\ \vdots \\ \hat{V}_m \end{pmatrix} = \boldsymbol{w}^T \begin{pmatrix} \boldsymbol{R}_1 \\ \boldsymbol{R}_2 \\ \vdots \\ \boldsymbol{R}_m \end{pmatrix} = \mathbf{X}\boldsymbol{w}, \tag{3.5}$$

where $\hat{V}$ is the predicted value of V,

$$\hat{V} = \boldsymbol{w}^T \boldsymbol{R}, \tag{3.6}$$

$\boldsymbol{w}^T \boldsymbol{R} = \boldsymbol{R}^T \boldsymbol{w}$, and $\boldsymbol{w} = [w_1, \dots, w_D]^T$ is the weight constant vector.

A clear goal of the multiple regression is to minimize the difference between the predicted value, $\hat{V}$, and the observed value, V. To do this, we define the prediction error by following relationship,

$$\mathcal{J} = \sum_{i=1}^{m} \left(V_i - \hat{V}_i\right)^2 = \sum_{i=1}^{m} (V_i - \boldsymbol{w}^T \boldsymbol{R}_i)^2, \tag{3.7}$$

and its expression in terms of the matrices by the expression:

$$\begin{aligned} \mathcal{J} &= (\boldsymbol{V} - \mathbf{X}\boldsymbol{w})^T (\boldsymbol{V} - \mathbf{X}\boldsymbol{w}) \\ &= \boldsymbol{V}^T \boldsymbol{V} - 2\boldsymbol{w}^T \mathbf{X}^T \boldsymbol{V} + \boldsymbol{w}^T \mathbf{X}^T \mathbf{X}\boldsymbol{w} \end{aligned} \tag{3.8}$$

Here, when $\mathcal{J}$ is zero, the regression model can be regarded as a complete predictor for the observations. Then, we optimize the weight constants, $\boldsymbol{w}$, to minimize the least squares $\mathcal{J}$; this is equivalent to finding the stationary point at which the gradient with respect to the weight constants is equal to zero:

$$\frac{d}{d\boldsymbol{w}} \mathcal{J} = \frac{d}{d\boldsymbol{w}} \boldsymbol{V}^T \boldsymbol{V} - 2 \frac{d}{d\boldsymbol{w}} \boldsymbol{w}^T \mathbf{X}^T \boldsymbol{V} + \frac{d}{d\boldsymbol{w}} \boldsymbol{w}^T \mathbf{X}^T \mathbf{X}\boldsymbol{w} = 0 \tag{3.9}$$

Although the first term is clearly zero, the second and third terms remain, as

$$\frac{\partial}{\partial \boldsymbol{w}} \boldsymbol{w}^T \mathbf{X}^T \boldsymbol{V} = \mathbf{X}^T \boldsymbol{V} \tag{3.10}$$

and

$$\frac{\partial}{\partial \boldsymbol{w}} \boldsymbol{w}^T \mathbf{X}^T \mathbf{X} \boldsymbol{w} = (\mathbf{X}^T \mathbf{X} + (\mathbf{X}^T \mathbf{X})^T) \boldsymbol{w} = \mathbf{2} \mathbf{X}^T \mathbf{X} \boldsymbol{w}, \tag{3.11}$$

respectively. Finally, we obtain:

$$\frac{\partial}{\partial \boldsymbol{w}} \mathcal{J} = -2\mathbf{X}^T \boldsymbol{V} + \mathbf{2} \mathbf{X}^T \mathbf{X} \mathbf{w} = 0 \tag{3.12}$$

Transforming Equation 3.12 results in the analytical solutions for the multiple regression,

$$\mathbf{X}^T \mathbf{X} \boldsymbol{w} = \mathbf{X}^T \boldsymbol{V} \tag{3.13}$$

$$\boldsymbol{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{V} \tag{3.14}$$

Here, it is important to state that the solution exists only when $\mathbf{X}^T \mathbf{X}$ has an inverse matrix, $(\mathbf{X}^T \mathbf{X})^{-1}$. The equation $\mathbf{X}^T \mathbf{X} \boldsymbol{w} = \mathbf{X}^T \boldsymbol{V}$ is called the normal equation in multiple regression.

We have explained the case of using the simple linear equation,

$$\hat{V} = w_1 R_1 + w_2 R_2 + \cdots + w_D R_D, \tag{3.15}$$

but this model is expected to give a poor prediction when the relationship between potential energy and nuclear coordinate is not linear and more complex. Therefore, a more flexible regression model is required. To reach this purpose, the design matrix, $\mathbf{X}$, is redefined using arbitrary basis functions $\phi(\boldsymbol{R})$ to:

$$\widehat{\boldsymbol{V}} = \begin{pmatrix} \hat{V}_1 \\ \hat{V}_2 \\ \vdots \\ \hat{V}_m \end{pmatrix} = \begin{pmatrix} \phi_0(\boldsymbol{R}_1) & \phi_1(\boldsymbol{R}_1) & \cdots & \phi_h(\boldsymbol{R}_1) \\ \phi_0(\boldsymbol{R}_2) & \phi_1(\boldsymbol{R}_2) & \cdots & \phi_h(\boldsymbol{R}_2) \\ \vdots & & & \vdots \\ \phi_0(\boldsymbol{R}_m) & \phi_1(\boldsymbol{R}_m) & \cdots & \phi_h(\boldsymbol{R}_m) \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_h \end{pmatrix} = \boldsymbol{\Phi} \boldsymbol{w}, \tag{3.16}$$

where $\phi_0 = 1$ for any nuclear coordinates $\boldsymbol{R}$, h is the number of the basis functions, and the vector by basis functions,

$$\boldsymbol{\phi} = [\phi_0(\boldsymbol{R}), \phi_1(\boldsymbol{R}), \dots, \phi_h(\boldsymbol{R})]^T, \tag{3.17}$$

is referred to as the feature vector. Then, we investigate the least-square solution in the case of using the basis functions. The least square solution is expressed as:

$$\boldsymbol{w} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\boldsymbol{V} \tag{3.18}$$

as with the multiple regression (Equation 3.14) as it does not depend on whether basis functions are utilized. This regression model is called linear regression and a specific case is the multiple regression model when $\mathbf{X} = \boldsymbol{\Phi}$. Here, the reason why this is not nonlinear regression is that the predicted values from Equation 3.16 are linear for the parameter, $\boldsymbol{w}$.

3.2.3. LASSO and Ridge Regression

We introduced the linear regression model as a regression model with high expressivity. However, the problem remains that there is no analytical solution when the matrix $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ does not have an inverse matrix $(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}$. A simple solution is to add a small value to elements of the matrix $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ so that the matrix $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ has an inverse matrix; however, this operation raises the problem that the weight vector $\boldsymbol{w}$ contains extremely large values [121]. In general, such models are quite sensitive regarding small changes in the nuclear coordinate $\boldsymbol{R}$ and this may become a source of overfitting.

Least absolute shrinkage and selection operator (LASSO) regression [122] and ridge regression [121,123–126] are some simple techniques that can be used to prevent overfitting. These techniques comprise the following operations: minimizing the least square difference $\mathcal{J}$ and the magnitude of the coefficient vector $\boldsymbol{w}$.

$$\mathcal{J}_1 = (\boldsymbol{V} - \boldsymbol{\Phi}\boldsymbol{w})^2 + \alpha|\boldsymbol{w}|, \tag{3.19}$$

$$\mathcal{J}_2 = (\boldsymbol{V} - \boldsymbol{\Phi w})^2 + \alpha \boldsymbol{w}^T \boldsymbol{w}, \tag{3.20}$$

where regularizations $\mathcal{J}_1$ and $\mathcal{J}_2$, which add the penalty terms $\alpha|\boldsymbol{w}|$ and $\alpha \boldsymbol{w}^T \boldsymbol{w}$, are called L1 and L2 regularizations, respectively, and the parameter α is the arbitrary coefficient to control the penalty of the regularization. If $\alpha = 0$, the above equation agrees with the general method of the least squares; if $\alpha = \infty$, all elements of the wight constant vector $\boldsymbol{w}$ become zero.

Then, the analytical solutions for L1 and L2 regularizations must be considered. For the L1 regularization, there is no analytical solution owing to the inclusion of a non-differentiable point. However, the L2 regularization has an analytical solution. As shown in Section 3.2.2, the stationary point where the derivative of the least-squares difference with respect to the coefficient vector is zero can be found from the following:

$$\frac{\partial}{\partial \boldsymbol{w}} \mathcal{J}_2 = \frac{\partial}{\partial \boldsymbol{w}} (\boldsymbol{V} - \boldsymbol{\Phi w})^2 + \frac{\partial}{\partial \boldsymbol{w}} \alpha \boldsymbol{w}^T \boldsymbol{w}, \tag{3.21}$$

The solution for the former term is already presented in Section 3.2.2. For the latter term, its solution is easily obtained from:

$$\frac{\partial}{\partial \boldsymbol{w}} \alpha \mathbf{w}^T \mathbf{w} = 2\alpha \mathbf{w} \tag{3.22}$$

Therefore, using Equations 3.12 and 3.22, the stationary point in the difference $\mathcal{J}_2$ can be expressed by:

$$\frac{\partial}{\partial \boldsymbol{w}} \mathcal{J}_2 = -2\boldsymbol{\Phi}^T \boldsymbol{y} + \mathbf{2}\boldsymbol{\Phi}^T \boldsymbol{\Phi w} + 2\alpha \mathbf{w} = 0 \tag{3.23}$$

and, from the perspective of the weight constant vector,

$$\boldsymbol{w} = (\boldsymbol{\Phi}^T \boldsymbol{\Phi} + \alpha \mathbf{I})^{-1} \boldsymbol{\Phi}^T \boldsymbol{y}. \tag{3.24}$$

The regression models using the L1 and L2 regularizations are named LASSO regression and ridge regression, respectively. Further, the elastic net [127] has been proposed as a

compromise between the LASSO and ridge regressions.

3.3. Gaussian Process Regression

3.3.1. Regression with the Machine Learning

If it is possible to guess basis functions to completely reproduce the PES, linear regression can generate the complete fitting model. However, in general, it is difficult to guess appropriate basis functions and hard to design a set of basis functions for every problem. Using machine learning (ML) approaches instead of basis functions helps to overcome these disadvantages in linear regression models.

Given these advantages, neural network (NN) [128–130] and Gaussian process regression (GPR) [131–133] have attracted attention as ML approaches to fitting the PES: the first NN and GPR PESs were presented in 1995 [134] and 2009 [135], respectively, and many types of PESs have since been successfully developed [90–93,136–142]. As each of these approaches has different advantages, it is important to choose the appropriate ML technique according to the nature of the target PES. Therefore, to determine the appropriate ML for the small water clusters, we briefly discuss the advantages of NN and GPR [90–93,140,141]:

Training cost: A NN is generally faster than GPR, although it is strongly dependent on the number of hidden layers N_{hl}, artificial neurons N_{an}, dimensionality D, and training points m. For a single-hidden-layer NN ($N_{\mathrm{hl}} = 1$), the training cost is dominated by the inner product $\mathbf{W}\boldsymbol{R}$ between the $N_{\mathrm{an}} \times D$ weight constant matrix $\mathbf{W}$ and the D-dimensional coordinate vector $\boldsymbol{R}$. In the case of a multi-hidden-layer NN, if the number of artificial neurons is unchanged, the cost depends mostly on

$N_{\mathrm{an}}^2 \times N_{\mathrm{hl}} \times D$. The rate-determining step in GPR is the evaluation of the inverse matrix $\mathbf{K}^{-1}$ of the $m \times m$ matrix $\mathbf{K}$, although this is explained further in Section 3.3.6. Finally, the training costs are $m \times N_{\mathrm{an}} \times D$ for the single-hidden-layer NN, $m \times N_{\mathrm{an}}^2 \times N_{\mathrm{hl}} \times D$ for the multi-hidden-layer NN, and $m^3 \times D$ for the GPR.

Cost of the prediction: The prediction cost for each ML approach is $N_{\mathrm{an}} \times D$ in the single-hidden-layer NN, $N_{\mathrm{an}}^2 \times N_{\mathrm{hl}} \times D$ in the multi-hidden-layer NN, and $m \times D$ for GPR. A simple comparison is not possible because the cost of NN does not include the number of training points m. However, with the exception of the multi-hidden-layer NN, the costs are of a similar order. Therefore, there is no major difference in the cost of the prediction between simple NN and GPR methods.

Overfitting problem: As explained in Section 3.2, general regression often causes an overfitting problem. Indeed, for the NN, overfitting is a serious problem. In contrast, the GPR methodology is based on the ridge regression and usually does not create an overfitting problem; thus, in GPR, more accurate results are guaranteed when more training points are used.

Accuracy: Both NN and GPR methods provide a more accurate PES if it is possible to use more training points. When there are fewer training points, GPR generally produces more accurate result than the NN.

Based on these features, the appropriate ML approach to describe the high-accuracy PES should be considered. The deprotonated water trimer, $H_5O_3^-$, can be considered to be a small system as it only has 18 degrees of freedom ($3N - 6 = 18$). Thus, relatively few training points are expected to be required to obtain a high-accuracy PES. As GPR can provide high-accuracy results with fewer training points, we decided that GPR is a more appropriate approach for the target system.

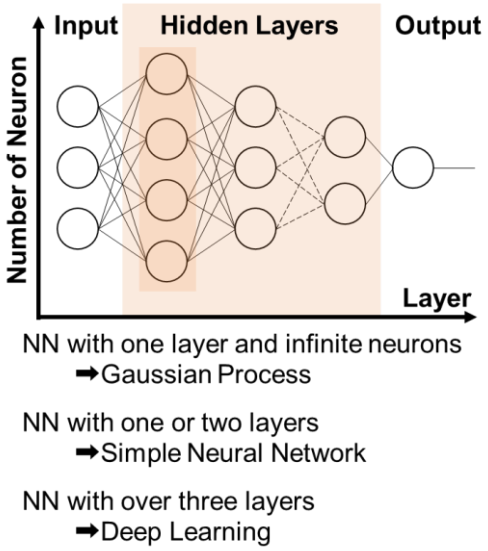


Fig. 3.1 Conceptual diagram for the cross point between NN and GPR.

Finally, it would be useful to clarify the crossing point between the NN and the GPR. The neural network with one hidden layer ($N_{\mathrm{hl}} = 1$) completely agrees with the Gaussian process with the limit that the number of artificial neurons is infinity ($N_{\mathrm{an}} = \infty$) [143,144]; proof of this is given in Appendix 7.2. Fig. 3.1 shows the conceptual diagram.

3.3.2. Probability Model for Linear Regression

We here explain the probability model for the linear regression, which would be the most basic stating point of the GPR. The linear regression model obtained above is expected to be incomplete because it do not consider observation noise. To take into account such noise ϵ, the Gaussian (normal) distribution with the mean μ and variance σ^2 is often used:

$$p(\epsilon) = \mathcal{N}(\epsilon|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(\epsilon-\mu)^2}{2\sigma^2}\right) \tag{3.25}$$

where the constant $1/\sqrt{2\pi}\sigma$ is the normalizing constant. The distribution with the mean 0 and variance 1, $\mathcal{N}(0,1)$, is a special case, referred to as the standard normal distribution. According to the central limit theorem, even if observation noise is not normally distributed, their sum tends toward a normal distribution. Thus, the normal distribution is utilized widely in situations in which the distribution of noise is unknown.

Now, the construction of probabilistic models for linear regression must be considered. For this, we assume that the potential energy V is always generated with the noise: that is, the PES $V(\boldsymbol{R})$ follows the normal distribution with a mean of $\boldsymbol{w}^T\boldsymbol{R}$ and a variance of σ^2:

$$p(V|\boldsymbol{R}) = \mathcal{N}(V|\boldsymbol{w}^T\boldsymbol{R}, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(V-\boldsymbol{w}^T\boldsymbol{R})^2}{2\sigma^2}\right) \tag{3.26}$$

In addition, the probability of predicting the PES vector $\boldsymbol{V}$ using the coordinate matrix $\mathbf{X}$ can be expressed by:

$$p(\boldsymbol{V}|\mathbf{X}) = \prod_{i=1}^{m} p(V_i|\boldsymbol{R}_i) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(V_i-\boldsymbol{w}^T\boldsymbol{R}_i)^2}{2\sigma^2}\right) \tag{3.27}$$

As taking the logarithm for both side of equation does not change the maximum point in $p(\boldsymbol{V}|\mathbf{X})$, we obtain:

$$-\log p(\boldsymbol{V}|\mathbf{X}) = \sum_{i=1}^{m} \frac{(V_i-\boldsymbol{w}^T\boldsymbol{R}_i)^2}{2\sigma^2} + m\log\sqrt{2\pi}\sigma \tag{3.28}$$

where the right-hand side is clearly equivalent to the least squares $\mathcal{J}$. Therefore, maximizing this probabilistic model is equivalent to solving the linear regression using the method of least squares.

Next, we extend the probabilistic model to ridge regression to avoid overfitting. The purpose of the ridge regression is to prevent anomalous increase in the elements w_i of

the weight vector $\boldsymbol{w}$; this operation can also be performed by controlling the following normal distribution with zero mean and λ^2 variance:

$$p(\boldsymbol{w}) = \prod_{i=1}^{D} \mathcal{N}(w_i|0, \lambda^2). \tag{3.29}$$

Taking the logarithm for both sides yields:

$$\begin{aligned} \log p(\boldsymbol{w}) &= \sum_{i=1}^{D} \log \mathcal{N}(w_i|0, \lambda^2) \\ &= -(D+1)\log\left(\sqrt{2}\lambda\right) - \frac{1}{2\lambda^2}\sum_{i=1}^{D} w_i^2. \end{aligned} \tag{3.30}$$

Here, the simultaneous probability of $\boldsymbol{V}$ and $\boldsymbol{w}$, when the coordinate matrix $\mathbf{X}$ is given, can be described by:

$$p(\boldsymbol{V}, \boldsymbol{w}|\mathbf{X}) = p(\boldsymbol{V}|\boldsymbol{w}, \mathbf{X})p(\boldsymbol{w}|\mathbf{X}) = p(\boldsymbol{V}|\boldsymbol{w}, \mathbf{X})p(\boldsymbol{w}), \tag{3.31}$$

using the independency of $p(\boldsymbol{w})$. Finally, the probabilistic model of the ridge regression can be expressed by:

$$\begin{aligned} -\log p(\boldsymbol{V}, \boldsymbol{w}|\boldsymbol{X}) &= -\log(\boldsymbol{V}|\boldsymbol{w}, \boldsymbol{X}) - \log p(\boldsymbol{w}) \\ &= m\log\sqrt{2\pi}\sigma + \sum_{i=1}^{m}\frac{(V_i - \mathbf{w}^T\boldsymbol{R}_i)^2}{2\sigma^2} + (D+1)\log\sqrt{2}\lambda + \frac{1}{2\lambda^2}\sum_{i=1}^{D} w_i^2. \end{aligned} \tag{3.32}$$

Ignoring the constant terms leads to:

$$\frac{1}{\sigma^2}\sum_{i=1}^{m}(V_i - \boldsymbol{w}^T\boldsymbol{R}_i)^2 + \frac{1}{\lambda^2}\sum_{i=1}^{D} w_i^2 \propto (\boldsymbol{V} - \boldsymbol{X}\boldsymbol{w}^T)^2 + \alpha\boldsymbol{w}^T\boldsymbol{w}, \tag{3.33}$$

where

$$\alpha = \frac{\sigma^2}{\lambda^2}. \tag{3.34}$$

This is in complete agreement with the equation for the ridge regression (Equation 2.20) and shows that ridge regression is equivalent to maximizing the simultaneous probability $p(\boldsymbol{V}, \boldsymbol{w}|\boldsymbol{X})$. For clarification, the parameter α indicates the ratio of observation noise σ^2 and variance λ^2 of the coefficient $\boldsymbol{w}$. In general, the parameter α should be quite small as the observation noise is smaller than the variance λ^2 of the weight constants. However, if an extremely small α is selected, the accuracy of the regression model is expected to be reduced owing to underestimation of the observation noise. Therefore, it is important to select an appropriate value for α according to the properties of the observation noise.

3.3.3. Gaussian Process Model

In this section, we present briefly how to construct the fitting model using the Gaussian process, starting from the general linear regression model,

$$\widehat{\boldsymbol{V}} = \boldsymbol{\Phi}\boldsymbol{w}. \tag{3.35}$$

Here, we assume that the weight parameter $\boldsymbol{w}$ follows a Gaussian distribution with zero mean and variance $\lambda^2\mathbf{I}$:

$$\boldsymbol{w} \sim \mathcal{N}(0, \lambda^2\mathbf{I}) \tag{3.36}$$

Then, the vector $\widehat{\boldsymbol{V}}$ can be regarded to be a new the Gaussian distribution projected out from the Gaussian distribution $\boldsymbol{w}$ by the matrix $\boldsymbol{\Phi}$. For the distribution $\widehat{\boldsymbol{V}}$, the expected value is given by:

$$\mathbb{E}\left[\widehat{\boldsymbol{V}}\right] = \mathbb{E}[\boldsymbol{\Phi}\boldsymbol{w}] = \boldsymbol{\Phi}\mathbb{E}[\boldsymbol{w}] = 0 \tag{3.37}$$

and the covariance matrix is evaluated by:

$$\begin{aligned}\Sigma &= \mathbb{E}\left[\widehat{\boldsymbol{V}}\widehat{\boldsymbol{V}}^T\right] - \mathbb{E}\left[\widehat{\boldsymbol{V}}\right]\mathbb{E}\left[\widehat{\boldsymbol{V}}^T\right] \\ &= \mathbb{E}[(\boldsymbol{\Phi w})(\boldsymbol{\Phi w})^T] \\ &= \boldsymbol{\Phi}\mathbb{E}[\boldsymbol{w w}^T]\boldsymbol{\Phi}^T \\ &= \lambda^2\boldsymbol{\Phi\Phi}^T \end{aligned} \tag{3.38}$$

where $\mathbb{E}[\boldsymbol{w w}^T] = \lambda^2\mathbf{I}$ is used. As a result, the distribution for $\widehat{\boldsymbol{V}}$ is found to be the multivariate normal distribution:

$$\widehat{\boldsymbol{V}} \sim \mathcal{N}(0, \lambda^2\boldsymbol{\Phi\Phi}^T). \tag{3.39}$$

As shown in Equation 3.39, the weight vector $\boldsymbol{w}$ for the linear regression is eliminated, leaving only the design matrix $\boldsymbol{\Phi\Phi}^T$. In this case, unlike general linear regression, there is no need to optimize the weight vector $\boldsymbol{w}$; therefore, if the weight constant vector $\boldsymbol{w}$ is the infinite-dimensional vector, it may be possible to obtain the distribution of $\boldsymbol{V}$ with a low cost. Here, the distribution of $\boldsymbol{V}$ is dependent on only the covariance matrix $\lambda^2\boldsymbol{\Phi\Phi}^T$ and is therefore expected to be a function of the coordinate vector $\boldsymbol{R}$.

Consequently, we can now define the Gaussian process model. As demonstrated above, if the joint distribution $p(V)$ for any finite subsets $\mathbf{X} = [\boldsymbol{R}_1, \boldsymbol{R}_2, \dots, \boldsymbol{R}_m]^T$ and $\boldsymbol{V} = [V_0, V_1, \dots, V_m]^T$ follows a multivariate normal distribution, the relationship between $\boldsymbol{R}$ and V is defined to follow the Gaussian process.

$$V \sim \mathrm{GP}(\mu, \lambda^2\boldsymbol{\phi\phi}^T) \tag{3.40}$$

Next, the features of the Gaussian process model is considered. Putting the covariance matrix as:

$$\mathbf{K} = \lambda^2 \mathbf{\Phi}\mathbf{\Phi}^T, \tag{3.41}$$

its elements are described by:

$$K_{ij} = \lambda^2 \phi(\boldsymbol{R}_i)^T \phi(\boldsymbol{R}_j) \tag{3.42}$$

The multivariate normal distribution basically takes similar values when the covariance between two variables is large. In other words, if the value of inner product $\phi(\boldsymbol{R}_i)^T \phi(\boldsymbol{R}_j)$ is large, the predicted values $\widehat{V}_i$ and $\widehat{V}_j$ are expected to have similar values.

3.3.4. Kernel Trick

The distribution of $\boldsymbol{V}$ is determined by the covariance matrix $\mathbf{K}$ with elements $\phi(\boldsymbol{R}_i)^T$ and $\phi(\boldsymbol{R}_j)$. As this matrix is clearly a function of $\phi(\boldsymbol{R}_i)^T$, it appears to require explicit calculation of $\phi(\boldsymbol{R}_i)^T$. However, there is a way to avoid this, known as the kernel trick. We must then consider the direct calculation of the element K_{ij} of the covariance matrix, where the function that gives K_{ij} is defined as the kernel function $k(\boldsymbol{R}_i, \boldsymbol{R}_j)$ of $\boldsymbol{R}_i$ and $\boldsymbol{R}_j$. In addition, the covariance matrix constructed by the kernel function is called the kernel matrix.

To explain the usefulness of this operation, the following example using the radial basis function (RBF) kernel can be examined. Let start with considering the situation that the following basis function is put in the coordinate space,

$$\phi_i(R) = A \exp\left(-\frac{(R - i/h)^2}{\theta^2}\right) \tag{3.43}$$

with the center on $i/h (i = -h^2, \ldots, h^2)$, every $1/h$, in the range $-h < R < h$ on R-axis. The function that consists of $2h^2 + 1$ functions is defined to be the feature vector

corresponding the linear regression model.

$$\boldsymbol{\phi}(R) = [\phi_{-h^2}(R) \ \ldots \ \phi_0(R) \ \ldots \ \phi_{h^2}(R)] \tag{3.44}$$

Here, the kernel function is expressed by:

$$k(R, R') = \sum_{i=-h^2}^{h^2} \phi_i(R)\phi_i(R') \tag{3.45}$$

Assuming the infinite small grid $h \to \infty$, the kernel function is also the RBF, as shown below:

$$\begin{aligned} k(R, R') &= \lim_{h\to\infty} \sum_{i=-h^2}^{h^2} \phi_i(R)\phi_i(R') \\ &= \int_{-\infty}^{\infty} A^2 \exp\left(-\frac{(R-i)^2}{\theta^2}\right) \exp\left(-\frac{(R'-i)^2}{\theta^2}\right) di \\ &= A^2 \int_{-\infty}^{\infty} \exp\left(-\frac{2}{\theta^2}\left(i - \frac{R+R'}{2}\right)^2\right) di \times \exp\left(-\frac{1}{2\theta^2}(R-R')^2\right) \\ &= A^2 \sqrt{\frac{\pi\theta^2}{2}} \exp\left(-\frac{1}{2\theta^2}(R-R')^2\right) \end{aligned} \tag{3.46}$$

Therefore, the Gaussian process model using the RBF can be regarded as a linear regression model using the infinite number of the RBF as a feature vector. As a result, although the feature vector has an infinite-dimensional degree of freedom in the linear regression model, the Gaussian process model can be described by only the finite number of the input vector as the weight vector is eliminated.

We have mentioned that utilizing the kernel function instead of the explicit feature vector can simplify the optimization for the infinite number of weights. This is a quite powerful property, known as the kernel trick [145–147]. However, please note that the

kernel matrix $\mathbf{K}$ in the Gaussian process model need to be a symmetric matrix and to have an inverse matrix, as $\mathbf{K}$ corresponds to the covariance matrix. For example, the Matérn kernel [132] is well-known as the generalized model of the RBF kernel (confirm Reference 132 for more examples of the kernel function).

3.3.5. Observation Noise

In a real-world situation, the observed values would include the noise. That is, we consider the noise by the following model for $i = 1, \dots, m$,

$$V_i = f(\boldsymbol{R}_i) + \epsilon_i. \tag{3.47}$$

Assuming that above noise follows the Gaussian distribution

$$\epsilon_i = \mathcal{N}(0, \sigma^2), \tag{3.48}$$

with zero mean and variance σ^2, the probability distribution for the observed value V can be written as

$$p(\boldsymbol{V}|\boldsymbol{f}) = \mathcal{N}(\boldsymbol{f}|\sigma^2\mathbf{I}), \tag{3.49}$$

where we assume $\boldsymbol{f} = [f(\boldsymbol{R}_1)\, f(\boldsymbol{R}_2)\ \dots\ f(\boldsymbol{R}_m)]^T$. Therefore, how can be the prior distribution after giving the inputs $\mathbf{X} = [\boldsymbol{R}_1\ \boldsymbol{R}_2\ \dots\ \boldsymbol{R}_N]$ be described? As the output depend on the function f having input $\mathbf{X}$ as a variable, taking the expectation value respect with f leads to

$$\begin{aligned} p(\boldsymbol{V}|\boldsymbol{X}) &= \int p(\boldsymbol{V}, \boldsymbol{f}|\boldsymbol{X}) d\boldsymbol{f} \\ &= \int p(\boldsymbol{V}|\boldsymbol{f}) p(\boldsymbol{f}|\boldsymbol{X}) d\boldsymbol{f} \end{aligned}$$

$$= \int \mathcal{N}(\boldsymbol{f}|\sigma^2\boldsymbol{I})\mathcal{N}(\boldsymbol{f}|\boldsymbol{\mu},\boldsymbol{K})d\boldsymbol{f}, \tag{3.50}$$

where the chain rule for the probability is utilized. This is the simple convolution of two Gaussian distributions and its covariance matrix is expressed as follows:

$$p(\boldsymbol{V}|\boldsymbol{X}) = \mathcal{N}(\boldsymbol{f}|\boldsymbol{\mu}, \mathbf{K} + \sigma^2\mathbf{I}). \tag{3.51}$$

In other words, the probability distribution considering the noise can be simply described by the kernel matrix to which its diagonal terms are added by the variance σ^2 of the noise distribution. Therefore, the kernel in this case is newly determine as:

$$k(\boldsymbol{R}_i, \boldsymbol{R}_j) = k(\boldsymbol{R}_i, \boldsymbol{R}_j) + \sigma^2\delta_{ij}, \tag{3.52}$$

where δ_{ij} is a delta function that returns 1 at $i = j$ and 0 in other cases.

3.3.6. Gaussian Process Regression

Now, solving the regression problem based on the Gaussian process can be considered. As shown in Section 3.2.1, on the assumption of m pairs, which consist of input $\boldsymbol{R} \in \mathcal{X}$ and output V, we obtain:

$$\mathcal{D} = \{(V_i, \boldsymbol{R}_i)|i = 1, \dots, m\}, \tag{3.53}$$

where there is the following relation between $\boldsymbol{R}$ and V,

$$V = f(\boldsymbol{R}), \tag{3.54}$$

and this function f is generated from the Gaussian process with mean μ and variance $k(\boldsymbol{R}, \boldsymbol{R}')$.

$$f \sim \mathrm{GP}(\mu, k(\boldsymbol{R}, \boldsymbol{R}')) \tag{3.55}$$

The noise here is included in the kernel function as shown in Section 3.3.5. Here, the outputs also follow a Gaussian distribution and are expressed as:

$$\boldsymbol{V} \sim \mathrm{GP}(\boldsymbol{\mu}, \mathbf{K}). \tag{3.56}$$

Next, we consider the method to predict the unknown V_+ for a new input $\boldsymbol{R}_+$, where the new vector adding V_+ to $\boldsymbol{V}$ is defined as $\boldsymbol{V}_+$ and the new matrix adding $\boldsymbol{R}_+$ to $\boldsymbol{X}$ is defined as $\mathbf{X}_+$: in other words, $\boldsymbol{V}_+ = [V_1 \ \ldots \ V_m \ V_+]^T$ and $\mathbf{X}_+ = [\boldsymbol{R}_1 \ \ldots \ \boldsymbol{R}_m \ \boldsymbol{R}_+]^T$. The new vector and matrix also are also subject to the Gaussian process.

$$\boldsymbol{V}_+ \sim \mathrm{GP}(\boldsymbol{\mu}, \mathbf{K}_+) \tag{3.57}$$

and

$$\begin{pmatrix} \boldsymbol{V} \\ V_+ \end{pmatrix} \sim \mathcal{N}\left(\boldsymbol{\mu}, \begin{pmatrix} \mathbf{K} & \boldsymbol{k}_+ \\ \boldsymbol{k}_+^T & k_{++} \end{pmatrix}\right). \tag{3.58}$$

As the above equation is the joint distribution of $\boldsymbol{V}$ and V_+, the conditional probability of V_+ given $\boldsymbol{V}$ is obtained from:

$$p(V_+ | \boldsymbol{R}_+, \boldsymbol{X}, \boldsymbol{V}) = \mathcal{N}(\boldsymbol{k}_+^T \mathbf{K}^{-1} \boldsymbol{V}, k_{++} - \mathbf{k}_+^T \mathbf{K}^{-1} \mathbf{k}_+) \tag{3.59}$$

The expectation value of the distribution V_+ becomes:

$$\hat{V}_+ = \mathbb{E}[V_+ | \boldsymbol{R}_+, \mathbf{X}, \boldsymbol{V}] = \boldsymbol{k}_+^T \mathbf{K}^{-1} \boldsymbol{V}, \tag{3.60}$$

where the hat ^ over the symbol denotes the predicted value and

$$\boldsymbol{k}_+^T = [k(\boldsymbol{R}_+, \boldsymbol{R}_1) \ \ldots \ k(\boldsymbol{R}_+, \boldsymbol{R}_m)]. \tag{3.61}$$

In Equation 3.60, the parameters $\mathbf{X}$, $\mathbf{K}^{-1}$, and $\boldsymbol{V}$ can be estimated using the given data (training data), $\mathcal{D}$. Therefore, Equation 3.60 indicates that the prediction value $\hat{V}_+$ is a simple function of $\boldsymbol{R}_+$ and is the regression model of the GP. Here, it should be noted that Equation 3.60 uses the training points $\boldsymbol{V}$ directly, which means that the prediction

accuracy is sensitive to the spatial distribution of points $\mathbf{X}$ in D-dimensional space.

Finally, the optimization of the GP model is explained. The probability that the GP precisely predicts the training points $\boldsymbol{V}$ is expressed by:

$$p(\boldsymbol{V}|\mathbf{X},\mathbf{K}) = \mathcal{N}(\boldsymbol{V}|\mathbf{X},\mathbf{K}) = \frac{1}{(2\pi)^{m/2}|\mathbf{K}|^{1/2}}\exp(-\frac{1}{2}\boldsymbol{V}^T\mathbf{K}^{-1}\boldsymbol{V}) \tag{3.62}$$

and taking the logarithm of the above gives

$$\log p(\boldsymbol{V}|\mathbf{X},\mathbf{K}) = -\frac{m}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{K}| - \frac{1}{2}\boldsymbol{V}^T\mathbf{K}^{-1}\boldsymbol{V} \tag{3.63}$$

Therefore, the GP model is optimized so that the kernel parameters give this equation its maximum value. For this, we need to compute the partial derivatives of Equation 3.63 with respect to the hyperparameters as:

$$\frac{\partial}{\partial\theta_i}\log p(\boldsymbol{V}|\mathbf{X},\mathbf{K}) = \frac{1}{2}\mathrm{tr}\left((\boldsymbol{\alpha}\boldsymbol{\alpha}^T - \mathbf{K}^{-1})\frac{\partial\mathbf{K}}{\partial\theta_i}\right), \tag{3.64}$$

where

$$\boldsymbol{\alpha} = \mathbf{K}^{-1}\boldsymbol{V}, \tag{3.65}$$

and the Cholesky decomposition is typically utilized to efficiently solve Equation 3.64. However, this approach scales as $\mathcal{O}(m^3)$ and causes the problem that the computational cost increases exponentially with increasing the number of training data as explained in Chapter 1.

3.3.7. Improvement of Gaussian Process Regression

As discussed in Chapter 1, we here suggest two algorithms to solve two problems occurring during the construction of the MLPES by GPR. The first is to generate the MLPES with high accuracy using only a limited number of training points. The second is

to satisfy the molecular symmetry.

Accuracy Improvement: fragmentation approach [91]: We propose a method to improve the accuracy of the potential energy surface given by the GPR without increasing the number of training points. We follow the methods described in Ref. 148 and 149, which split the full configuration space into smaller parts $\boldsymbol{Q}$, and represent the energy of the entire molecular system as:

$$\hat{V}_{\text{Exact}}(\boldsymbol{R}) = \hat{V}_1(\boldsymbol{Q}_1) + \hat{V}_2(\boldsymbol{Q}_2) + \hat{V}_{12}(\boldsymbol{Q}_{12}), \tag{3.66}$$

where $\boldsymbol{R}$ is the full-dimensional vector, $\boldsymbol{Q} \subset \boldsymbol{R}$, $\hat{V}_1$ and $\hat{V}_2$ are independent GP models depending on vectors of lower dimensionality, $\hat{V}_{12}$ is a GP model that brings fragments $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ together into the full surface and that depends on the vector $\boldsymbol{Q}_{12}$ with the dimensionality to be determined.

The representation (Equation 3.66) essentially reduces the problem of constructing the D-dimensional PES to build GP models of potential energy for smaller molecular fragments and constructing a D-dimensional GP model of the difference between the global surface and these lower-dimensional GPs. In Chapter 5, we will demonstrate the gain in accuracy resulting from this approach by comparing GP models (3.66) with those obtained directly by fitting energy in the 28-dimensional space for the deprotonated water trimer $H_5O_3^-$. This approach is motivated by Ref. 148, which introduced a hierarchy of molecular fragmentations to approximate the total electronic energy from the energies of the fragments. In general, the molecular fragmentation in Equation 3.66 should be performed to ensure that the energy of the fragments and of the full system can be computed using the same ab initio method.

Molecular Symmetry Adaption: In this subsection, we determine the condition for which the MLPES has the same value at both $\boldsymbol{R}$ and $\hat{O}\boldsymbol{R}$, where $\hat{O}$ is the arbitrary

symmetric operator that satisfies the following relationship:

$$\hat{V}(\boldsymbol{R}) = \hat{V}\big(\hat{O}\boldsymbol{R}\big). \tag{3.67}$$

where

$$\boldsymbol{R} = [R_1, \dots, R_D]^T \tag{3.68}$$

$$\hat{O}\boldsymbol{R} = \left[R_1^{\mathrm{inv}}, \dots, R_D^{\mathrm{inv}}\right]^T \tag{3.69}$$

As shown in Equation 3.60, the MLPES by GPR is described by:

$$\hat{V}(\boldsymbol{R}) = \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{R}, \boldsymbol{R}')\mathbf{K}'^{-1}\boldsymbol{V}' = \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{R}, \boldsymbol{R}')\boldsymbol{\alpha} = \sum_{i=1}^{m} k_i(\boldsymbol{R}, \boldsymbol{R}_i')\,\alpha_i, \tag{3.70}$$

where the training dataset is given by:

$$\mathcal{D} = \{\boldsymbol{R}_i', V_i' | i = 0, \dots, m\}, \tag{3.71}$$

$$\boldsymbol{V}' = [V_1', \dots, V_m']^T \tag{3.72}$$

$\boldsymbol{k}$ and $\mathbf{K}$ are the vector and the $m \times m$ matrix of the kernel, respectively, and $\mathbf{K}'^{-1}\boldsymbol{V}' = \boldsymbol{\alpha}$, for simplicity, as the number of training points can be regarded as a constant. Therefore, we employ the RBF as the kernel function:

$$\begin{aligned} k_i(\boldsymbol{R}, \boldsymbol{R}_i') &= \exp\left(\sum_{j=1}^{D} -\frac{\left(R_j - R_{ij}'\right)^2}{2\theta_j}\right) \\ &= \exp\left(-\sum_{j=1}^{D} \frac{R_j^2 - 2R_jR_{ij}' + {R_{ij}'}^2}{2\theta_j}\right), \end{aligned} \tag{3.73}$$

where the D indicates the number of dimensions of the nuclear coordinates and θ_j is the anisotropic length parameter for the jth nuclear coordinate. In another words, adapting the molecular symmetry to GPR is equivalent to finding the conditions that

satisfy the following relationship:

$$\hat{V}(\boldsymbol{R}) = \sum_{i=1}^{m} k_i(\boldsymbol{R}, \boldsymbol{R}_i')\,\alpha_i = \sum_{i=1}^{m} k_i\big(\hat{O}\boldsymbol{R}, \boldsymbol{R}_i'\big)\,\alpha_i = V\big(\hat{O}\boldsymbol{R}\big), \tag{3.74}$$

and, that is,

$$\sum_{i=1}^{m} \exp\left(-\sum_{j=1}^{D} \frac{R_j^2 - 2R_jR_{ij}' + {R_{ij}'}^2}{2\theta_j}\right)\alpha_i = \sum_{i=1}^{m} \exp\left(-\sum_{j=1}^{D} \frac{{R^{\mathrm{inv}}}_j^2 - 2{R^{\mathrm{inv}}}_jR_{ij}' + {R_{ij}'}^2}{2\theta_j^{\mathrm{inv}}}\right)\alpha_i. \tag{3.75}$$

First, if θ_j is equivalent to θ_j^{inv},

$$\boldsymbol{\theta} = \hat{O}\boldsymbol{\theta}, \qquad \boldsymbol{\theta} = [\theta_1, \dots, \theta_D]^T, \tag{3.76}$$

the above equation can be rewritten as the following:

$$\sum_{i=1}^{m} \exp\left(\sum_{j=1}^{D} \frac{2R_jR_{ij}' - {R_{ij}'}^2}{2\theta_j}\right)\alpha_i = \sum_{i=1}^{m} \exp\left(\sum_{j=1}^{D} \frac{2{R^{\mathrm{inv}}}_jR_{ij}' - {R_{ij}'}^2}{2\theta_j}\right)\alpha_i \tag{3.77}$$

Here, we prepared the new training dataset $\boldsymbol{R}^{\mathrm{inv}'}$ and $\hat{V}\big(\boldsymbol{R}^{\mathrm{inv}'}\big)$ and expanded it as with above equation:

$$\sum_{i}^{m} \exp\left(\sum_{j=1}^{D} \frac{2R_j{R^{\mathrm{inv}'}}_{ij} - {R^{\mathrm{inv}'}}_{ij}^{\,2}}{2\theta_j}\right)\alpha_i = \sum_{i}^{m} \exp\left(\sum_{j=1}^{D} \frac{2R_j^{\mathrm{inv}}{R^{\mathrm{inv}'}}_{ij} - {R^{\mathrm{inv}'}}_{ij}^{\,2}}{2\theta_j}\right)\alpha_i \tag{3.78}$$

Finally, adding the two equations gives:

$$\sum_{i}^{m} \exp\left(\sum_{j=1}^{D} \frac{2R_jR_{ij}' - {R_{ij}'}^2}{2\theta_j}\right)\alpha_i + \sum_{i}^{m} \exp\left(\sum_{j=1}^{D} \frac{2R_j{R^{\mathrm{inv}'}}_{ij} - {R^{\mathrm{inv}'}}_{ij}^{\,2}}{2\theta_j}\right)\alpha_i$$

$$= \sum_{i}^{m} \exp\left(\sum_{j=1}^{D} \frac{2R^{\mathrm{inv}}{}_{j} R'_{ij} - {R'_{ij}}^{2}}{2\theta_j} \right) \alpha_i + \sum_{i}^{m} \exp\left(\sum_{j=1}^{D} \frac{2R_j^{\mathrm{inv}} R^{\mathrm{inv}'}{}_{ij} - {R^{\mathrm{inv}'}{}_{ij}}^{2}}{2\theta_j} \right) \alpha_i \quad (3.79)$$

where the following relations are clearly established:

$$\sum_{i}^{m} \exp\left(\sum_{j=1}^{D} \frac{2R_j R'_{ij} - {R'_{ij}}^{2}}{2\theta_j} \right) \alpha_i = \sum_{i}^{m} \exp\left(\sum_{j=1}^{D} \frac{2R_j^{\mathrm{inv}} R^{\mathrm{inv}'}{}_{ij} - {R^{\mathrm{inv}'}{}_{ij}}^{2}}{2\theta_j} \right) \alpha_i \quad (3.80)$$

and

$$\sum_{i}^{m} \exp\left(\sum_{j=1}^{D} \frac{2R^{\mathrm{inv}}{}_{j} R'_{ij} - {R'_{ij}}^{2}}{2\theta_j} \right) \alpha_i = \sum_{i}^{m} \exp\left(\sum_{j=1}^{D} \frac{2R_j R^{\mathrm{inv}'}{}_{ij} - {R^{\mathrm{inv}'}{}_{ij}}^{2}}{2\theta_j} \right) \alpha_i \quad (3.81)$$

In Equation 3.79, the left-hand side is clearly equal to the right-hand side. Therefore, two conditions in which the MLPES by GPR satisfies the molecular symmetry are: (1) the kernel parameters are immutable for the arbitrary symmetrical operation, $\boldsymbol{\theta} = \hat{O}\boldsymbol{\theta}$; and (2) the training dataset $\mathcal{D}$ consists of $\{\boldsymbol{R}'_i, V_i | i = 1, \dots, m\}$ and $\{\boldsymbol{R}^{\mathrm{inv}'}{}_i, V_i | i = 1, \dots, m\}$. This approach confers the following advantages: (I) the symmetry of the PES is guaranteed for all the configuration space; (II) the variables to be optimized can be reduced; and (III) it can be easily combined with other GPR algorithms, such as fragmentation approaches. Although we explain the case of the system with only one axis of symmetry, can easily extend to system with multiple axes of symmetry, such as NH_3 and CH_4.

4. COMPUTATIONAL DETAILS

4.1. Potential Energy Surface

4.1.1. OSS Potential Energy Surfaces

To describe the interatomic interaction for the protonated water clusters, we adopted a polarizable and dissociable model developed by Ojamae, Singer, and Shavitt [44] that is dubbed as OSS potential. A series of the OSS potential has been parameterized using the extensive *ab initio* MP2 results for the Zundel dimer $H_5O_2^+$; the resulting OSS potentials have been reported to be able to predict formation energies and structures with close accuracy to *ab initio* MP2 results for small protonated water clusters [44,51]. Among the OSS potentials, the following OSS3 potential,

$$\begin{aligned} V_{\mathrm{OSS3}}(\boldsymbol{R}) = \; & V_{\mathrm{el}}(\boldsymbol{R}) + \sum_{i}^{n_{\mathrm{O}}} \sum_{j}^{n_{\mathrm{H}}} V_{\mathrm{OH}}(R_{ij}) + \sum_{i}^{n_{\mathrm{O}}-1} \sum_{j>i}^{n_{\mathrm{O}}} V_{\mathrm{OO}}(R_{ij}) \\ & + \sum_{i}^{n_{\mathrm{H}}-1} \sum_{k>i}^{n_{\mathrm{H}}} \sum_{j}^{n_{\mathrm{O}}} V_{\mathrm{HOH}}(R_{ij}, R_{kj}, \theta_{ijk}) \\ & + \sum_{i}^{n_{\mathrm{H}}-1} \sum_{k>i}^{n_{\mathrm{H}}} \sum_{j}^{n_{\mathrm{O}}} V_{\mathrm{HOH}\mu}(R_{ij}, R_{kj}, \theta_{ijk}, \mu_j), \end{aligned} \tag{4.1}$$

was adopted in the present study, which is expected to be most accurate in the OSS potentials; V_{el} is the total electrostatic energy, V_{OH} and V_{OO} are the two-body potential energies for H–O and O–O pairs, respectively, V_{HOH} and $V_{\mathrm{HOH}\mu}$ are three-body potential and dipole-three-body coupling energies for H–O–H triplets, respectively, $R_{ij} = \left|\boldsymbol{R}_i - \boldsymbol{R}_j\right|$, θ indicates the H–O–H angle, μ is the induced dipole moment of the oxygen ion. The explicit expression of the potential energy terms can be found in Ref. [44].

TABLE 4.1. The minimum potential energy V_{min} for protonated water clusters $(\mathrm{H_2O})_n\mathrm{H}^+$ up to $n = 5$ calculated by the OSS3 potential. Previously obtained minimum potential energy using the OSS3 potential [44] is also presented for comparison. Energy is given in units of hartree.

		V_{min}	
	cluster size n	Ref. [51]	this work
(a)	1	-1.51534	-1.51534
(b)	2	-2.91762	-2.91762
(c)	3	-4.30319	-4.30319
(d)	4 branched	-5.68131	-5.68131
(e)	4 cyclic	-5.67798	-5.67798
(f)	4 linear	-5.67791	-5.67791
(g)	5 kitelike	-7.05296	-7.05296
(h)	5 branched	-7.05133	-7.05163
(i)	5 cyclic[a]	-7.04996	-7.04996
(j)	5 cyclic[b]	-7.04982	-7.04991
(k)	5 pentagonal	-7.04923	-7.04923
(l)	5 linear	-7.04841	-7.04928

We next present stable structures for the protonated water clusters $(\mathrm{H_2O})_n\mathrm{H}^+$ on the OSS3 potential energy surface. Up to the size $n = 2$, one stable structure is known for each system. For the trimer, $n = 3$, several stable structures characterized by the same hydrogen bonding topology have been found in the present study; the structure presented in Figure 4.1 is the most stable one among them. For the tetramer, $n = 4$, there exist three stable isomers respectively called branched, cyclic, and linear structures. For the pentamer, $n = 5$, six stable isomers are known: kitelike, branched, cyclic[a], cyclic[b], pentagonal, and linear structures. In the present thesis, we find slightly lower energy structures than those found previously for branched, cyclic[b], and linear pentamers in Ref. 51; details

are described in Section 5.1.3. Energies for the stable structures are collected in Table 4.1. These structures have been used to construct the trial wavefunction Equation 2.14 for the protonated water clusters.

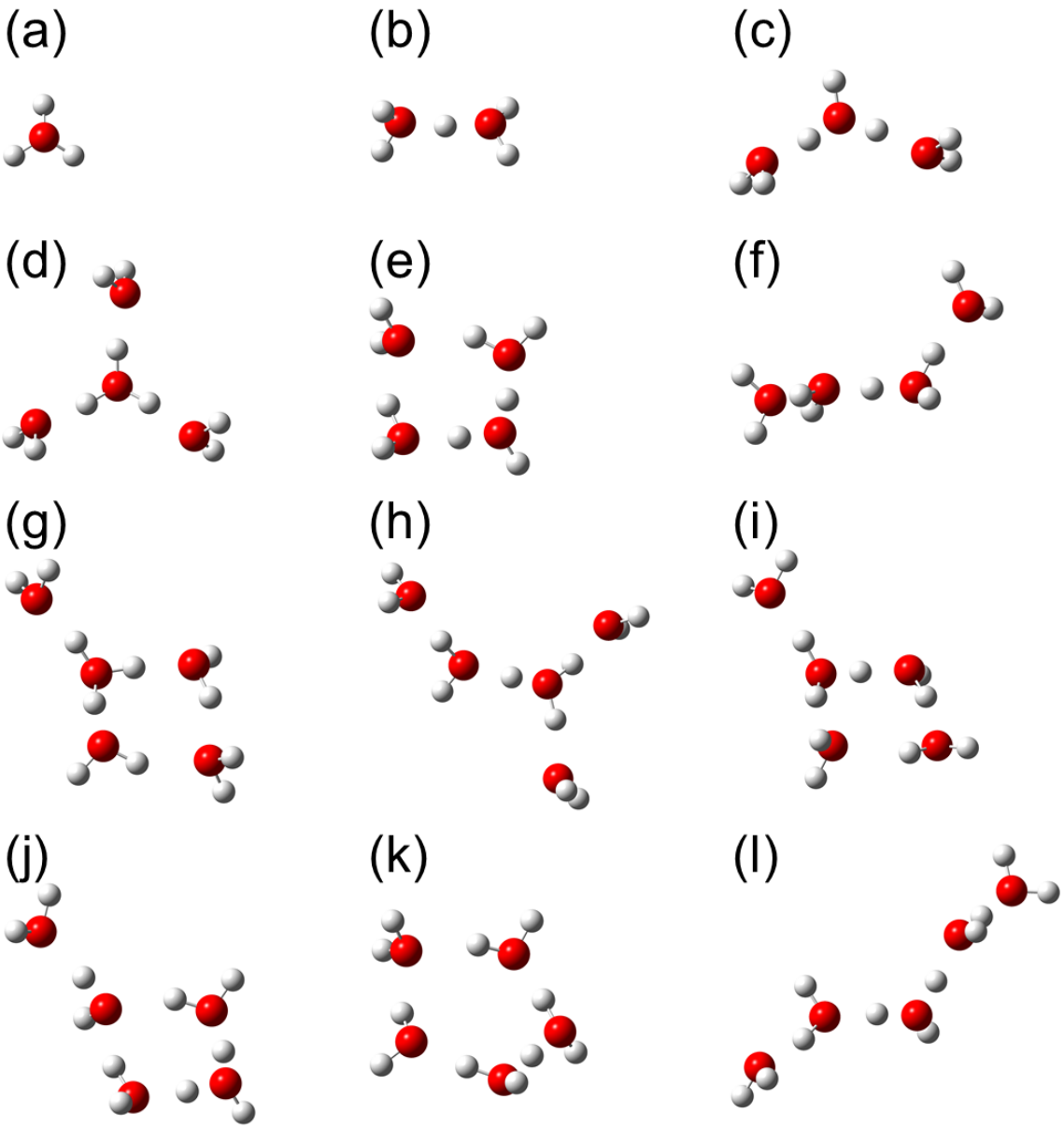


FIG. 4.1. The stable structures of the protonated water clusters on the OSS3 potential energy surface. The labels from (a) to (l) are defined in Table 4.1.

4.1.2. Machine Learning Potential Energy Surface

In this section, we generate the training points to construct the machine learning potential energy surface (MLPES) for deprotonated water clusters OH^-, $H_3O_2^-$, and $H_5O_3^-$. The potential energies for the training points are calculated at ab initio MP2/6-31++G (d, p) level. We start with the known geometry of deprotonated water clusters in Ref [150] and reoptimized it to obtain the most stable structure (Figure 4.2). Then, centering on these structures, 51 points were randomly sampled for OH^-, 5001 points for $H_3O_2^-$, and 6000 points for $H_5O_3^-$ in the range shown in Figure 4.2. In this thesis, as a random number generator, the Latin hypercube sampling method is adopted to avoid clustering.

Then, we condition the training dataset so that deprotonated water clusters have the same energy at the coordinates $\boldsymbol{R}$ and $\hat{O}\boldsymbol{R}$, where the symmetric operator $\hat{O}$ are defined in given by the symmetric in Figure 4.3. As discussed above, the restriction to satisfy the molecular symmetry is that the training dataset always includes the pair of $\boldsymbol{R}$ and $\hat{O}\boldsymbol{R}$ for all the training points. Therefore, we generated symmetry-operated coordinates $\hat{O}\boldsymbol{R}$ paired with the coordinate $\boldsymbol{R}$ and added $\hat{O}\boldsymbol{R}$ to the training data. Finally, 51 points for OH^-, 10,001 points for $H_3O_2^-$, and 12,000 points for $H_5O_3^-$ were used as training data. Here, please note that the number of training points is not simply double when the structure is included in the training points so that $\boldsymbol{R} = \hat{O}\boldsymbol{R}$. Accuracy of the constructed PES will be discussed in Section 5.2.1.

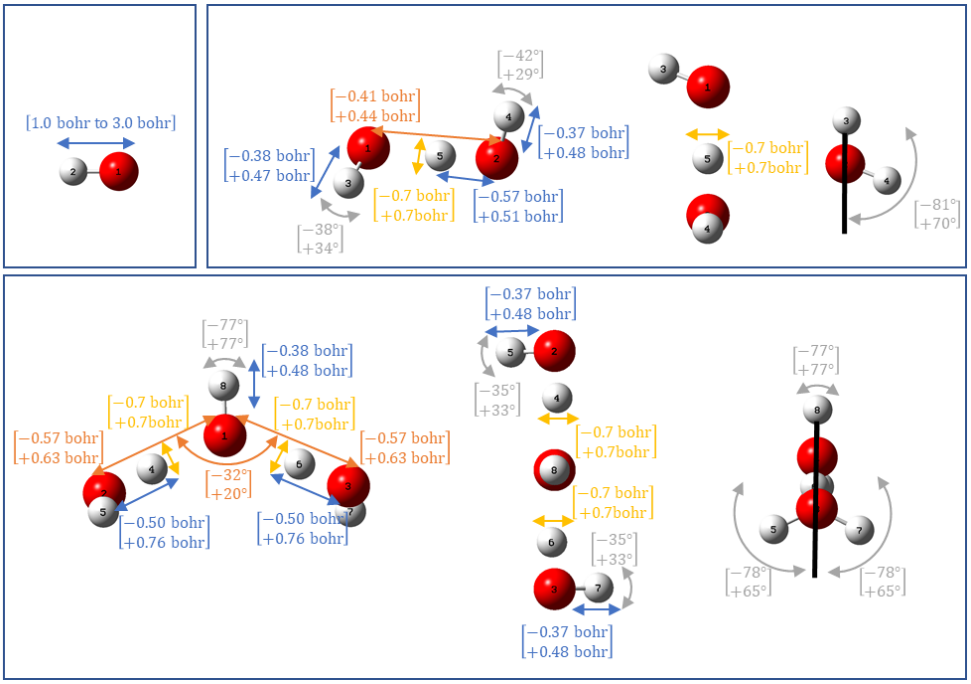


FIG. 4.2. Sampling ranges for the 1 degree-of-freedom of OH^-, the 9 degree-of-freedoms of $H_3O_2^-$, and the 18 degree-of-freedoms of H_5O_3.

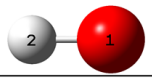
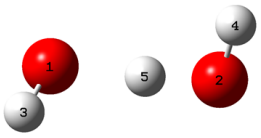
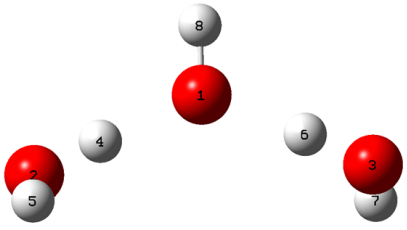
	$\hat{O}\mathbf{r} = \hat{O}[r_1]^T = [r_1]^T$ $\{\, r_1 = r_{12} \,\}$
	$\hat{O}\mathbf{r} = \hat{O}[r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8, r_9, r_{10}]^T$ $= [r_1, r_6, r_5, r_7, r_3, r_2, r_4, r_8, r_{10}, r_9]^T$ $\begin{pmatrix} r_1 = r_{1,2}, & r_2 = r_{1,3}, & r_3 = r_{1,4}, & r_4 = r_{1,5}, \\ r_5 = r_{2,3}, & r_6 = r_{2,4}, & r_7 = r_{2,5}, & r_8 = r_{3,4}, \\ r_9 = r_{3,5}, & r_{10} = r_{4,5} & & \end{pmatrix}$
	$\hat{O}\mathbf{r} = \hat{O}[r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8, r_9, r_{10}, r_{11}, r_{12}, r_{13}, r_{14}, r_{15}, r_{16}, r_{17}, r_{18}, r_{19}, r_{20}, r_{21}, r_{22}, r_{23}, r_{24}, r_{25}, r_{26}, r_{27}, r_{28}]^T$ $= [r_2, r_1, r_5, r_6, r_3, r_4, r_7, r_8, r_{16}, r_{17}, r_{14}, r_{15}, r_{18}, r_{11}, r_{12}, r_9, r_{10}, r_{13}, r_{26}, r_{20}, r_{23}, r_{27}, r_{21}, r_{24}, r_{28}, r_{19}, r_{22}, r_{25}]^T$ $\begin{pmatrix} r_1 = r_{1,2}, r_2 = r_{1,3}, r_3 = r_{1,4}, r_4 = r_{1,5}, r_5 = r_{1,6}, \\ r_6 = r_{1,7}, r_7 = r_{1,8}, r_8 = r_{2,3}, r_9 = r_{2,4}, r_{10} = r_{2,5}, \\ r_{11} = r_{2,6}, r_{12} = r_{2,7}, r_{13} = r_{2,8}, r_{14} = r_{3,4}, r_{15} = r_{3,5}, \\ r_{16} = r_{3,6}, r_{17} = r_{3,7}, r_{18} = r_{3,8}, r_{19} = r_{4,5}, r_{20} = r_{4,6}, \\ r_{21} = r_{4,7}, r_{22} = r_{4,8}, r_{23} = r_{5,6}, r_{24} = r_{5,7}, r_{25} = r_{5,8}, \\ r_{26} = r_{6,7}, r_{27} = r_{6,8}, r_{28} = r_{7,8} \end{pmatrix}$

FIG. 4.3. The definition of the symmetrical operator.

4.2. Variational Parameter Optimization

We numerically optimized the trial wavefunctions on Equation 2.14, where the number of the variational parameters is 36 for H_3O^+, 441 for $H_5O_2^+$, 2025 for $H_7O_3^+$, 6084 for $H_9O_4^+$, 14400 for $H_{11}O_5^+$, 1 for OH^-, 100 for $H_3O_2^-$, and 748 for $H_5O_3^-$, respectively. In order to evaluate the energy gradient $\boldsymbol{g}$ for given variational parameters, the VMD calculations have been performed 200,000 steps with time increment $\Delta t = 0.2$ fs. Then, the variational parameters have been updated according to Equation 2.28; the iteration was terminated if the absolute values of all components of the energy gradient $\boldsymbol{g}$ become a value less than 1.0×10^{-3} hartree bohr^2. The inverse temperature of the system $\beta = 1/k_B T$ was set to be $T = 100$ K where k_B is the Boltzmann constant.

Please note that, in the present study, the trial wavefunctions have been optimized for $(H_2O)_n H^+$ and $(H_2O)_{n-1}OH^-$, which are not isotope-substituted; the resulting variational parameters have also been utilized for variational path integral calculations of isotopologues substituted by deuterium D or tritium T.

4.3. Variational Path Integral Molecular Dynamics

We here determine the computational conditions for the VPIMD method. We first investigate the imaginary time increment $\Delta\tau$ and total projection time β to extract the exact ground state from the trial wavefunction, which has been optimized in Section 4.2. Figure 4.4 shows the total energy of the $H_5O_2^+$ as a function of the imaginary time increment $\Delta\tau$ for a given projection time $\beta = 0.02$ K^{-1}. The total energy is demonstrated to show quadratic convergence on $\Delta\tau$; an approximate density operator with $\Delta\tau$ less than or equal to 0.0001 K^{-1} is found to yield numerically converged results.

Figure 4.5 shows the total energy as a function of the projection time β with $\Delta\tau = 0.0001\ \mathrm{K}^{-1}$. We find that the projection time β more than or equal to 0.01 K^{-1} attains convergence on the total energy. In this thesis, we adopted the imaginary time increment $\Delta\tau = 0.0001\ \mathrm{K}^{-1}$ and the total projection time $\beta = 0.02\ \mathrm{K}^{-1}$ for all the VPIMD calculations; these parameters correspond to the number of imaginary time slices $M = \beta/\Delta\tau = 200$. For all the clusters studied in the present study, VPIMD calculations have been performed from 1,000,000 to 5,000,000 steps with MD time increment $\Delta t = 0.2$ fs. The fictitious masses for the staging variables were set to be equal to the corresponding standard staging masses except for end-point coordinates (at $s = 0$ and M) where $m'^{(0)} = m'^{(M)} = \gamma_{\mathrm{ep}} m$ with $\gamma_{\mathrm{ep}} = 4/M$ and physical masses of particles m.

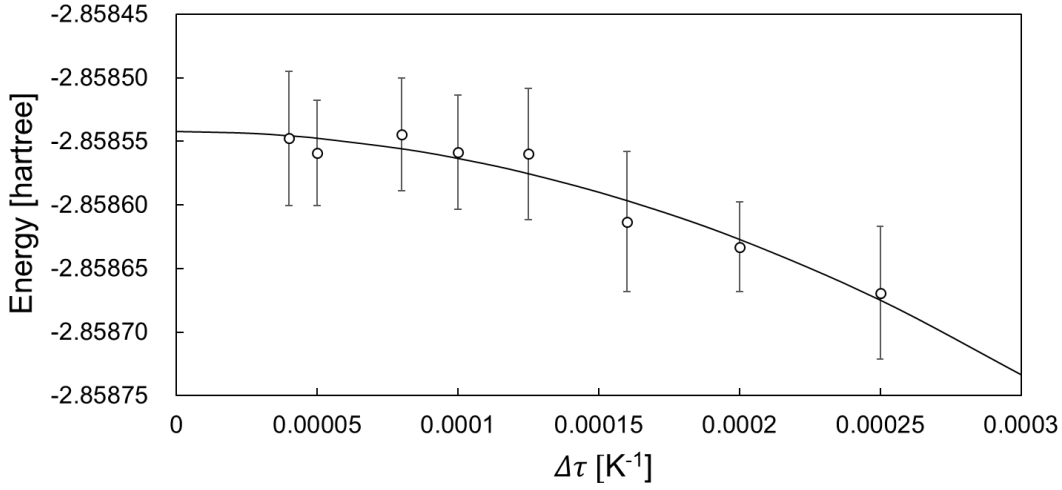


FIG. 4.4. Total energy of $H_5O_2^+$ as a function of the imaginary time increment $\Delta\tau$. Open circles indicate the total energies calculated by the VPIMD method. The black curve represents a fitted curve by a quadratic function of $\Delta\tau$. The total projection time β is set to be 0.02 K^{-1} for all the calculations. Error bars are expressed at 95% confidence level. Energy is given in units of hartree.

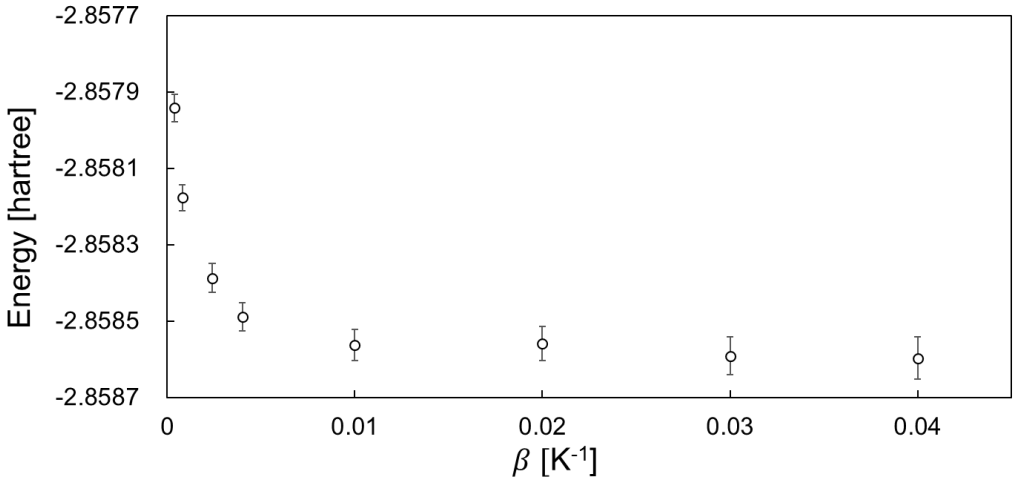


FIG. 4.5. Total energy of $H_5O_2^+$ as a function of the total projection time β. Open circles indicate the total energies calculated by the VPIMD method. The imaginary time increment $\Delta\tau$ is set to be 0.0001 K^{-1} for all the calculations. Error bars are expressed at 95% confidence level. Energy is given in units of hartree.

5. GROUND STATES of WATER CLUSTERS

5.1. Protonated Water Clusters $(H_2O)_nH^+$

5.1.1. Energy

In this section, we present the energy for the protonated water clusters, $(H_2O)_nH^+$. Table 5.1 shows that the total energy E_{HA} by the harmonic approximation (HA) and the total energy E_0 by the VPIMD method. The VPIMD energy E_0 has been calculated using the following mixed estimator, [151,152]

$$E_0 = \frac{\langle \Phi_T | \hat{H} e^{-\beta \hat{H}} | \Phi_T \rangle}{\langle \Phi_T | e^{-\beta \hat{H}} | \Phi_T \rangle}. \quad (5.1)$$

In the beginning, we compare the total energy by our VPIMD method with that computed by the diffusion Monte Carlo (DMC) method [51]; the DMC calculations have been performed using the same OSS3 potential. The DMC energies are collected in Table 5.1 where the entry "R.O." indicates that the cyclic isomer was converted to the linear or branched isomer by spontaneous ring opening (R.O.) during the DMC calculations. The VPIMD energies are found to be in good agreement with the available DMC energies. This demonstrates the reliability of our calculations. One exception is given by the branched tetramer; the VPIMD energy is 2 mhartree lower than the DMC energy. As demonstrated in Section 5.1.3, our VPIMD calculation seems to have sampled more

widely in the configuration space than the previous DMC calculation; this could be ascribed to the source of the small discrepancy in energy.

TABLE 5.1. The total energies for protonated water clusters $(H_2O)_nH^+$, $(D_2O)_nD^+$, and $(T_2O)_nT^+$ up to $n = 5$ calculated by the harmonic approximation, E_{HA} and the variational path integral molecular dynamics (VPIMD) method, E_0. The total energies by the diffusion Monte Carlo (DMC) method [51] are also presented for comparison. Energy is given in units of hartree. Statistical error in the last digits is indicated in parentheses (1 hartree = 627.51 kcal/mol). The labels from (a) to (l) are defined in Table 4.1.

	E_{HA}			E_0 (DMC)		E_0 (VPIMD)	
	$(H_2O)_nH^+$	$(D_2O)_nD^+$	$(T_2O)_nT^+$	$(H_2O)_nH^+$		$(D_2O)_nD^+$	$(T_2O)_nT^+$
(a)	-1.48083	-1.49017	-1.49418	-1.482253(7)	-1.48223(3)	-1.49097(4)	-1.49478(3)
(b)	-2.85641	-2.87261	-2.87960	-2.85862(3)	-2.85856(5)	-2.87390(4)	-2.88055(6)
(c)	-4.21517	-4.23854	-4.24862	-4.21879(4)	-4.21883(5)	-4.24069(6)	-4.25020(8)
(d)	-5.56869	-5.59863	-5.61156	-5.57188(9)	-5.57244(5)	-5.60090(7)	-5.61321(8)
(e)	-5.56348	-5.59389	-5.60702	R.O.	-5.56756(5)	-5.59635(7)	-5.60883(7)
(f)	-5.56498	-5.59487	-5.60777	-5.56962(9)	-5.56930(6)	-5.59754(8)	-5.60969(9)
(g)	-6.91408	-6.95100	-6.96695	R.O.	-6.91829(6)	-6.95340(9)	-6.96881(10)
(h)	-6.91389	-6.95047	-6.96627	-6.9165(1)	-6.91848(6)	-6.95321(8)	-6.96830(9)
(i)	-6.91086	-6.94777	-6.96371	R.O.	-6.91527(11)	-6.95044(13)	-6.96573(14)
(j)	-6.91068	-6.94758	-6.96353	R.O.	-6.91553(9)	-6.95057(9)	-6.96578(11)
(k)	-6.91010	-6.94692	-6.96283	R.O.	-6.91438(7)	-6.94953(8)	-6.96483(10)
(l)	-6.91046	-6.94724	-6.96311	-6.9156(2)	-6.91570(12)	-6.95039(9)	-6.96541(13)

We next discuss the structural stability for three isomers of the tetramer and six isomers of the pentamer. Regarding the tetramer, the minimum potential energy V_{min} indicates the following stability order:

$$\text{linear} \approx \text{cyclic} > \text{branched}.$$

Here, the energy is lowered to the right. The spatially compact isomer, branched one, has the lowest potential energy. Other isomers, cyclic and linear, have almost the same energy, the difference is found to be 7×10^{-5} hartree (= 0.04 kcal/mol). Considering the nuclear quantum effect by the standard harmonic approximation, the order changes to be

$$\text{cyclic} > \text{linear} > \text{branched}.$$

The cyclic isomer is predicted to have higher total energy, although both isomers have almost the same $V_{\min}$. The higher total energy of the cyclic isomer is due to the larger kinetic energy arising from the larger curvature around the potential energy minimum in comparison to the linear one. Regarding the VPIMD results, the total energy E_0, which fully includes the anharmonic effect, is found to be the following order:

$$\text{cyclic} > \text{linear} > \text{branched}.$$

The trend is the same with the HA prediction, E_{HA}. It is worth noting that the VPIMD energy is systematically lower than the HA counterpart for all the $H_9O_4^+$ isomers. This difference comes from the anharmonicity of the potential energy surface; as expected, the harmonic approximation is not enough to quantitatively describe the floppy molecular systems. More importantly, as demonstrated in Section 5.1.3, the nuclear wavefunction of the clusters is delocalized in the configuration space so as to cover multiple local minima; this type of effect must be included for quantitative prediction in addition to the anharmonicity around the single potential energy minimum.

Next, the structural stability of six isomers of the pentamer is discussed. Regarding the minimum potential energy $V_{\min}$, the isomers are sorted in the following order:

$$\text{pentagonal} \approx \text{linear} > \text{cyclic[b]} \approx \text{cyclic[a]} > \text{branched} > \text{kitelike}.$$

The energy differences between the pentagonal and linear structures and between the cyclic[b] and cyclic[a] structures are both quite small. Taking account of the nuclear

quantum effect by the HA, the order becomes as follows:

pentagonal > linear > cyclic[b] > cyclic[a] > branched > kitelike.

Nearly energetically degenerate isomers are well separated by reflecting the curvature around the potential energy minimum. Then, the VPIMD energies give the following order:

pentagonal > cyclic[a] > cyclic[b] > linear > kitelike > branched.

As in the case of the tetramer, the VPIMD energy is systematically lower than the HA prediction. Unlike the tetramer, however, the order of the structural stability partly changes for the pentamer. Although the HA results predict that the kitelike isomer is the most stable, the VPIMD results show that the branched isomer is the most stable. In addition, the spatially loose linear structure is found to has lower total energy than the compact structures, the cyclic[a], cyclic[b], and pentagonal. Thus, the simple harmonic approximation cannot predict the properties of the pentamer even in a qualitative level for the pentamer.

Finally, we discuss the isotope effects on the structural stability of the protonated water clusters. Since substituting the hydrogen with heavier isotopes generally reduces the spatial delocalization of the nuclear wavefunction, the total energy is expected to be lower with increasing mass of the isotope. Indeed, as shown in Table 5.1, both the HA and the VPIMD results show this trend for all the clusters studied. The harmonic approximation predicts that the stability order of the isomers does not change for the tetramer and the pentamer by substituting the hydrogen atom with the deuterium or tritium. Likewise, the stability order regarding the isomers of the tetramer does not change by the isotopic substitution even when the anharmonic effect is fully accounted for by the VPIMD method. On the other hand, the deuteration of the pentamer changes the stability

order of the isomer into

$$\text{pentagonal} > \text{linear} > \text{cyclic[b]} = \text{cyclic[a]} > \text{branched} = \text{kitelike},$$

where the symbol = indicates that both structures have the same total energy within the statistical error. The tritiation also changes the stability order as follows:

$$\text{pentagonal} > \text{linear} > \text{cyclic[b]} = \text{cyclic[a]} > \text{branched} > \text{kitelike}.$$

These behaviors demonstrate that the stability order approaches that of the minimum potential energy by the isotopic substitution.

5.1.2. Hydrogen Bonds

It is widely known that there are two important structural motifs for the protonated water clusters. One is given by complexes including the hydronium ion, H_3O^+; the Eigen cation, $H_3O^+(H_2O)_3$ [153,154], provides an important example. In the present study, the structure containing the motif of the H_3O^+ cation is called the Eigen structure. The other is given by the Zundel cation, $H_5O_2^+$, where a proton is shared by two neighboring water molecules; we call the structure containing the motif of the $H_5O_2^+$ cation the Zundel structure [155,156]. It is worth mentioning that the hydrogen bond associated with the shared proton in the Zundel cation is known to be shorter than the standard hydrogen bond, which is categorized into the low-barrier hydrogen bond [157]. The average hydrogen bond length calculated by the VPIMD method is summarized in Figure 5.1. The structures (b), (f), and (k), which are presented in Figure 4.1, are classified into the Zundel structure and the others (c), (d), (e), (g), (h), (i), (j), and (l) in Figure 4.1 into the Eigen one.

To precisely discuss the hydrogen bond, two types of the hydrogen bond are introduced; one is defined between the structural motif of H_3O^+ or $H_5O_2^+$ and water molecules coordinated to the motif. For the Zundel structure, an additional type of the hydrogen bond is defined between the shared proton and neighboring oxygens bonded to the proton. The former bond length is denoted by R_1 and the latter by R_2. While the

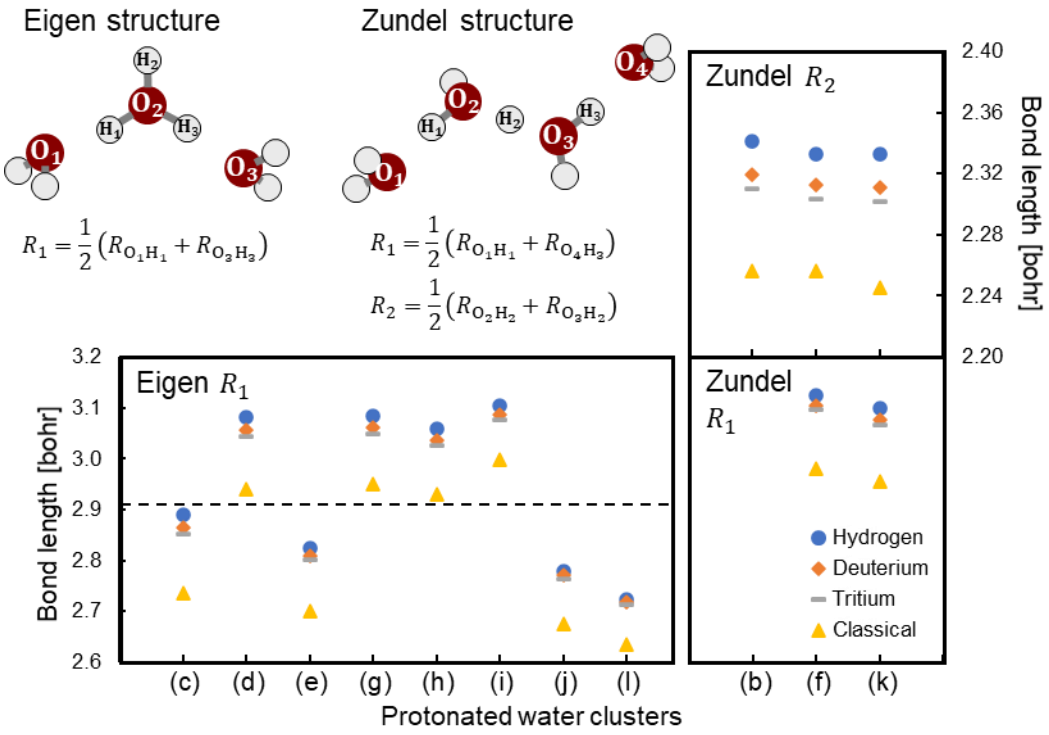


FIG. 5.1. Average bond lengths of protonated water clusters evaluated by the VPIMD calculations. Two types of the bond, R_1 and R_2 are defined. The structures (b), (f), and (k) in Figure 4.1 are classified into the Zundel structure and others (c), (d), (e), (g), (h), (i), (j), and (l) in Figure 4.1 are classified into the Eigen structure. Blue circles, orange rhombus, and gray bars indicate the bond lengths calculated for $(H_2O)_nH^+$, $(D_2O)_nD^+$, and $(T_2O)_nT^+$, respectively. Yellow triangles indicate the associated bond lengths evaluated for the potential energy minimum structure. Dashed line in the lower-left panel is drawn for the bond length = 2.91 bohr as an aid to the eye. Length is given in units of bohr (1 bohr = 0.5292 Å).

hydrogen bonds contained in the Eigen structures are characterized by R_1, those in the Zundel structures by R_1 and R_2. Definition of the bond lengths is given pictorially in Figure 5.1. We first discuss nuclear quantum effects on the structure (b), the Zundel cation, $H_5O_2^+$. The hydrogen bond is found to become longer compared with the minimum potential energy structure; in addition, substituting heavier isotopes of the hydrogen shortens the hydrogen bonds. This trend is consistent with the observation in the previous study [158]. Regarding the larger size Zundel structures (f) and (k), the hydrogen bond length evaluated by the VPIMD calculations is found to be longer about 0.07~0.15 bohr than the minimum potential energy structure. Similar trend is found for the hydrogen bond in the Eigen structures. This clearly shows the importance of the anharmonic effect for all the clusters studied. It is interesting to mention that the hydrogen bond length R_1 for the Eigen structures can be classified into two ranges $2.72 < R_1 < 2.89$ and $3.06 < R_1 < 3.12$ bohr. It is found that the former corresponds to the hydronium ion coordinated by two water molecules and the latter to the ion coordinated by three water molecules. On the other hand, the hydrogen bond lengths R_1 and R_2 are mostly independent for the Zundel structures studied. Regarding the isotope effect, as shown in Figure 5.1, substituting heavier hydrogen isotopes is found to make the hydrogen bonds of the Eigen and Zundel structures shorter, as expected.

5.1.3. Nuclear Wavefunction in Configuration Space

In order to characterize the multidimensional nuclear wavefunction in the configuration space, we have applied the inherent structure analysis [159–161] to the VPIMD trajectories. As seen in the Section 2.4, configurations at the imaginary time $\tau = \beta/2$ are distributed according to the exact nuclear wavefunction when the projection time

β is long enough. Each structure at the imaginary time $\tau = \beta/2$ along the VPIMD trajectory can be mapped onto the nearest local potential energy minimum in the configuration space using the steepest descent minimization method. The obtained structures are classified by the topology of the hydrogen bonds and summarized in Table 5.2.

TABLE 5.2. Number of types of inherent structures belonging to each isomer for $(H_2O)_nH^+$, $(D_2O)_nD^+$, and $(T_2O)_nT^+$.

cluster size n	$(H_2O)_nH^+$	$(D_2O)_nD^+$	$(T_2O)_nT^+$
1	1	1	1
2	1	1	1
3	3	3	3
4 branched	3	3	3
4 cyclic	3	2	2
4 linear	8	7	4
5 kitelike	4	3	3
5 branched	7	7	5
5 cyclic[a]	4	4	3
5 cyclic[b]	4	3	3
5 pentagonal	3	3	2
5 linear	10	8	7

We first discuss the nuclear wavefunction of the clusters containing lightest hydrogen isotopes. Both for the H_3O^+ and $H_5O_2^+$, a single inherent structure is found, indicating the nuclear wavefunction is well localized around the minimum energy structure. For the trimer, three inherent structures are found. Each structure is characterized by the same topology of the hydrogen bonds; the obtained inherent structures are interconverted by rotating the torsion angle around the hydrogen bond. As shown in Figure 4.1, several topologically distinguishable isomers exist for the larger clusters. Regarding the tetramer, there are three isomers; each isomer is characterized by several inherent structures. The linear isomer has a larger number of inherent structures than the cyclic and branched isomers, due to the structural flexibility along the hydrogen bonding chain. For the pentamer, the wavefunction increases the complexity in the configuration space. As in the case of the tetramer, the linear isomer has higher flexibility than other spatially compact isomers and

its nuclear wavefunction is delocalized to cover many local minima in the configuration space.

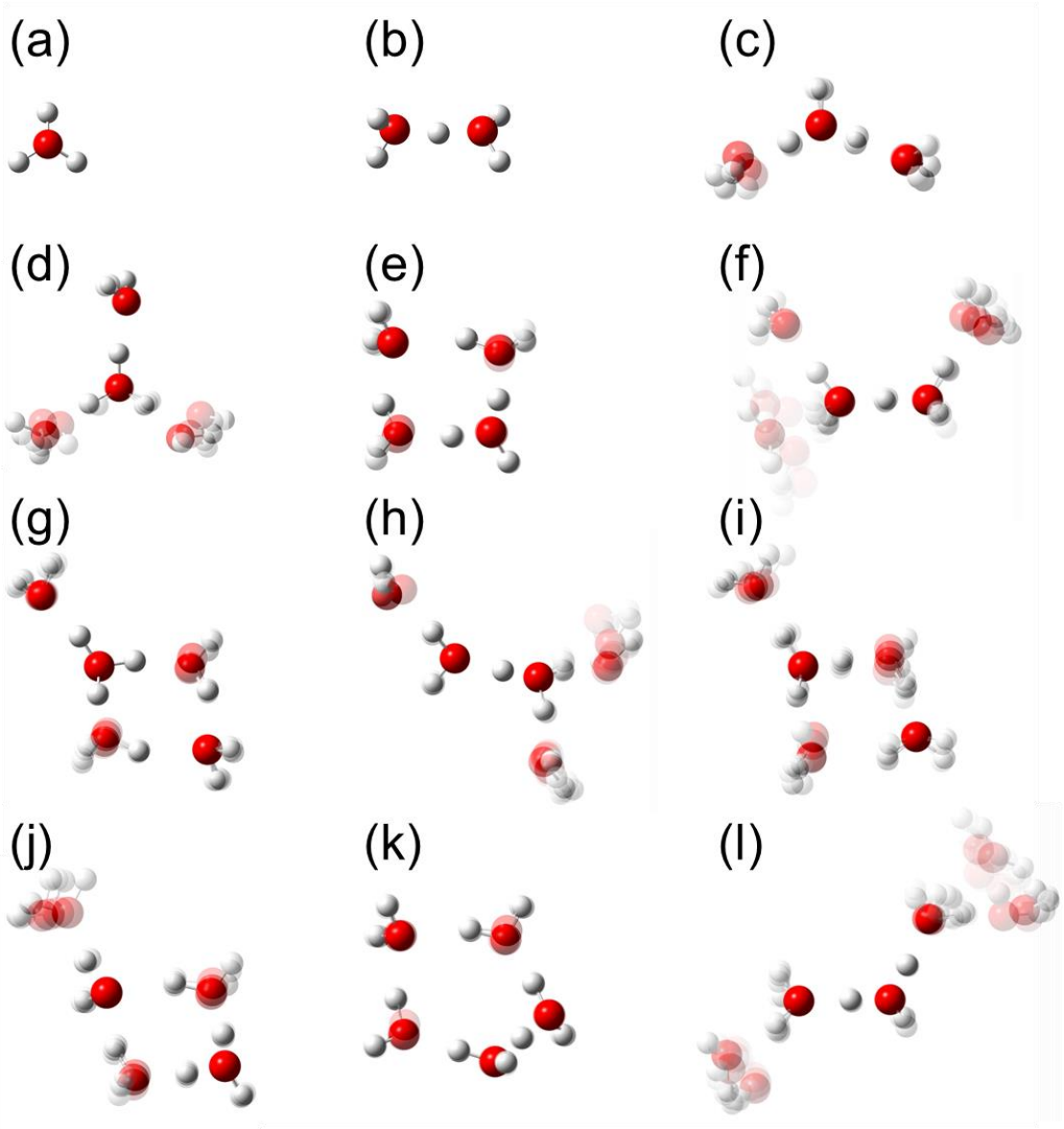


FIG. 5.2. Overlaying snapshots of inherent structures belonging to the same isomer. The labels from (a) to (l) are defined in Table I. The inherent structures were obtained using instantaneous structures of the $(H_2O)_nH^+$ at the imaginary time $\beta/2$ along the variational path integral molecular dynamics trajectory.

In order to visualize the difference between the inherent structures, we overlay all the inherent structures belonging to the same class of the hydrogen bonding topology to show in Figure 5.2. In the case of spatially compact isomers, molecules at the edge of the clusters take different torsion angles around hydrogen bonds to form a group of inherent structures. On the other hand, the nuclear wavefunctions of spatially flexible structures have more diversity in inherent structures than those of the compact structures.

We next examine the population of the inherent structures reflecting the probability of finding the system in the corresponding basin in the potential energy landscape. In Figures 5.3 and 5.4, we present the population visiting in each potential energy basin. The inherent structures belonging to each isomer are labeled by Roman numbers; the smaller number is assigned to lower energy inherent structure. For the trimer, the lowest energy structure has the highest population; other two local minima are occupied by less than 15 %. For the tetramer, the branched and cyclic isomers show a similar trend as seen in the trimer; the lowest energy structure has the highest population, and the population decreases with increasing the energy. On the other hand, the linear isomer shows more diverse behavior. The overall trend is similar with the branched and cyclic isomers, however, the populations of the structures (III), (IV), (VI), (VII), and (VIII) are vanishingly small. This indicates that quantum mechanical tunneling from the neighboring stable structures in the configuration space would be suppressed; this could be ascribed to the complexity of the tunneling path from the neighbors to the above structures. For the pentamer, the kitelike, branched, cyclic[a], cyclic[b], and pentagonal isomers show a similar trend with the trimer and the branched and cyclic tetramers. On the other hand, the linear pentamer obeys the similar trend seen in the linear tetramer; several inherent structures (II), (III), (VI), (VII), (VIII), (IX), and (X) are occupied in

vanishingly small rates. It is worth noting that for the branched, cyclic[b], and liner pentamers, lower energy inherent structures are found compared with the previous DMC results. This indicates that our VPIMD calculations could sample more widely in the configuration space than the DMC calculations, though the energy difference between the present and previous stable structures is rather small, about sub mhartree as seen in Table 4.1.

Finally, we discuss the isotope effects on the wavefunction. Table 5.2 shows that the variety of the inherent structure decreases with increasing the hydrogen mass for all the cluster sizes; the nuclear wavefunction tends to be localized for the heavier isotopes. Figures 5.3 and 5.4 show that while the deuteration or the tritiation increases the population of the most stable structure, the populations for other structures are decreased or vanished.

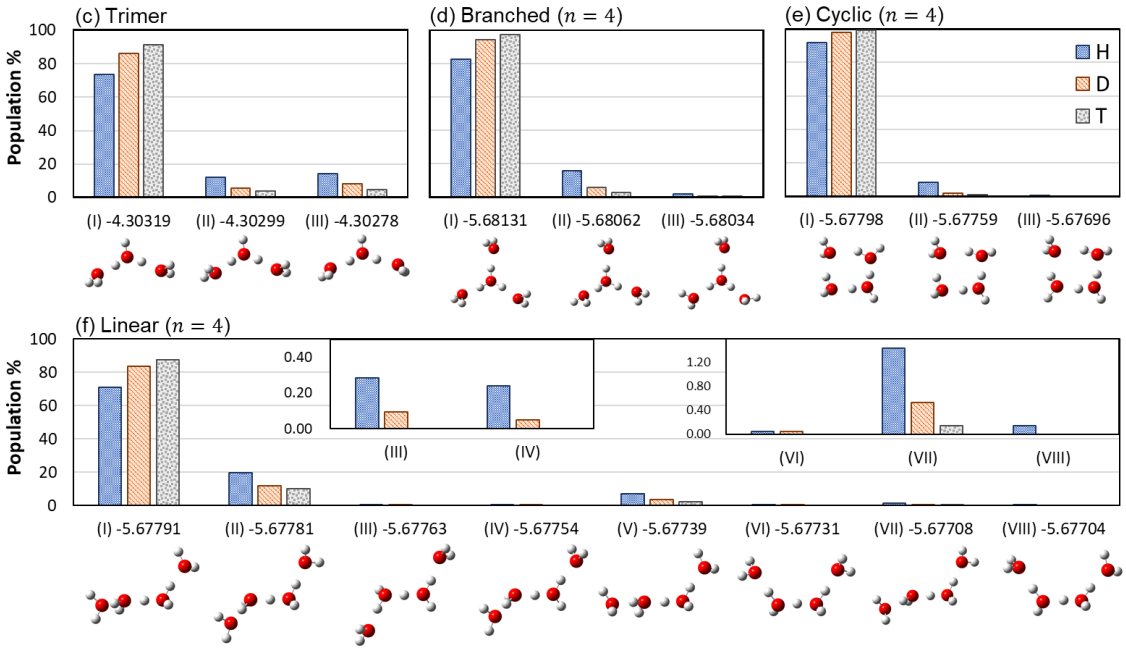


FIG. 5.3. Population of inherent structures for $H_7O_3^+$ and each isomer of $H_9O_4^+$ and associated isotope effect by deuterium (D) or tritium (T). The labels from (c) to (f) are defined in Table 4.1. The inherent structures belonging to the trimer or each isomer of the tetramer are labeled by Roman numbers; the smaller number is assigned to lower energy inherent structure. Snapshot of each inherent structure is presented (1 hartree = 627.51 kcal/mol).

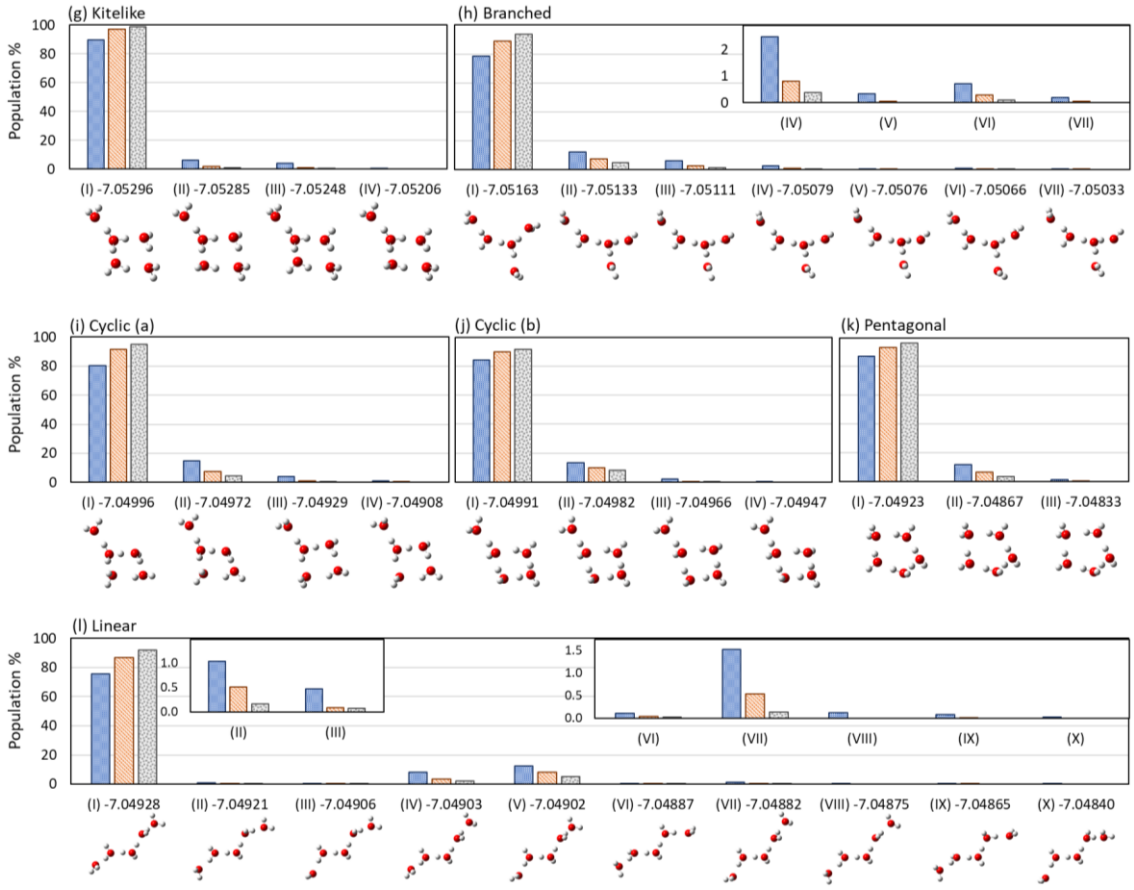


FIG. 5.4. Population of inherent structures for each isomer of $H_{11}O_5^+$ and associated isotope effect by deuterium (D) or tritium (T). The labels from (g) to (l) are defined in Table 4.1. The inherent structures belonging to each isomer of the pentamer are labeled by Roman numbers; the smaller number is assigned to lower energy inherent structure. Snapshot of each inherent structure is presented (1 hartree = 627.51 kcal/mol).

5.2. Machine Learning Potential Energy Surface

5.2.1. Accuracy vs Number of Training Points

We investigated the prediction accuracy of a machine-learning PES for potential energies and forces. First, the variational path integral molecular dynamics (VPIMD) simulation was preliminarily executed 10^6 times to generate a trajectory. Then, the coordinate data were extracted from the trajectory for every 100 points, and the potential energy and force were calculated with ab initio MP2 at the 6-31 ++ G (d, p) level. The prediction accuracy (RMSE) of machine learning PES for the generated 10^4 test data is shown in Table 5.3. The units are kcal/mol for energy and kcal/mol/Å for force. Here, the force was calculated analytically from the coordinate derivative of the kernel function in Equation 3.60.

TABLE 5.3. RMSEs for 10,000 testing points extracted from the VPIMD trajectory computed for the deprotonated water clusters OH^-, $H_3O_2^-$, $H_5O_3^-$, and fragmented $H_5O_3^-$.

Molecule	Number of training points N	RMSE for energy		RMSE for force	
		cm^{-1}	kcal/mol	cm^{-1}/Å	kcal/mol/Å
OH^-	51	0.4	0.001	6.3	0.018
$H_3O_2^-$	10001	4.2	0.012	28.8	0.082
$H_5O_3^-$	12000	83.5	0.239	451.6	1.3
Fragmented $H_5O_3^-$	12000	18.3	0.052	80.4	0.23

The RMSEs for energy and force are 0.001 kcal/mol and 0.02 kcal/mol/Å for OH^-, and 0.01 kcal/mol and 0.08 kcal/mol/Å for $H_3O_2^-$, respectively, as shown in Table 5.3; therefore, the accuracy can be regarded as sufficiently high. However, the RMSE for $H_5O_3^-$ exceeds 1.0 kcal/mol/Å for the force and is therefore expected not to be sufficient to describe the chemical phenomena precisely. To solve this problem, we applied our fragmentation approach to the potential energy surface of the protonated water trimer,

$H_5O_3^-$. In this approach, as shown in Figure 5.5, the potential energy is represented by the sum of the energies of two fragments $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ and one connecting layer $\boldsymbol{Q}_3$, as described in Equation 3.66. The RMSEs for energy and force were then improved from 0.24 to 0.05 kcal/mol and from 1.3 to 0.23 kcal/mol/Å, respectively, without changing the number of training points. Therefore, the VPIMD simulations were performed using the most accurate machine-learning PES for 51, 10001, and 12000 points.

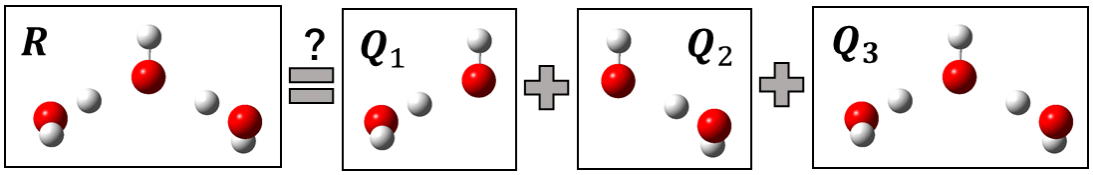


FIG. 5.5. Conceptual diagram of the fragmentation approach.

5.2.2. Accuracy vs Electronic Structure Calculations

We employed the MP2/6-31++G (d, p) to generate training points for the GPR. However, it is unknown if the MP2/6-31++G (d, p) can describe the PES for the deprotonated water clusters with high accuracy. Thus, to prove the validity of our simulation, we investigate the dependency between the VPIMD results and the electronic structure calculations. GPR PESs for HF, B3KYP, and CCSD(T) were constructed and the VPIMD simulations were performed on these PESs. The results are shown in Figure 5.6, where the VPIMD results indicate the ground state vibrational energy. The MP2 method is close in value to the CCSD(T) and is therefore sufficiently accurate. In contrast, HF overestimates the ground state energy and DFT (B3LYP) underestimates the ground state energy. Here, it is interesting that the ground-state energy given by the HF method

has an extremely higher value than by the CCSD(T) method even though it is generally considered to give more than 99% of the experimental value as introduced in Section 3.1. As introduced in Section 3.1, the HF method generally gives more than 99% of the experimental value. It is also interesting results that the underestimation by the DFT(B3LYP) was as we expected in Section 3.1. Therefore, we conclude that it is difficult to utilize the PESs generated by the low-level electronic structure calculations to quantitively estimate the nuclear ground-state properties. In addition, we also examined the dependence on the basis set. As shown in Figure 5.7, our choice, 6-31++G(d, p), results in high accuracy, similar to the results from the higher-level basis sets.

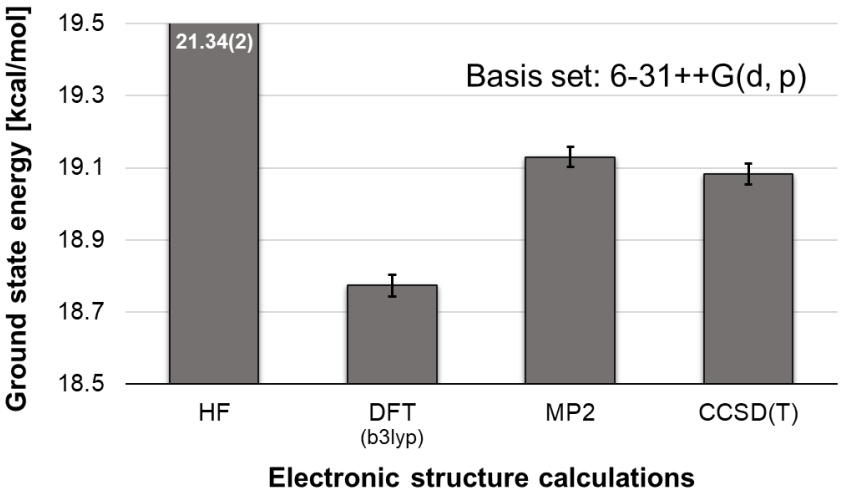


FIG. 5.6. Dependency between the variational path integral molecular dynamics (VPIMD) results and the electronic structure calculations.

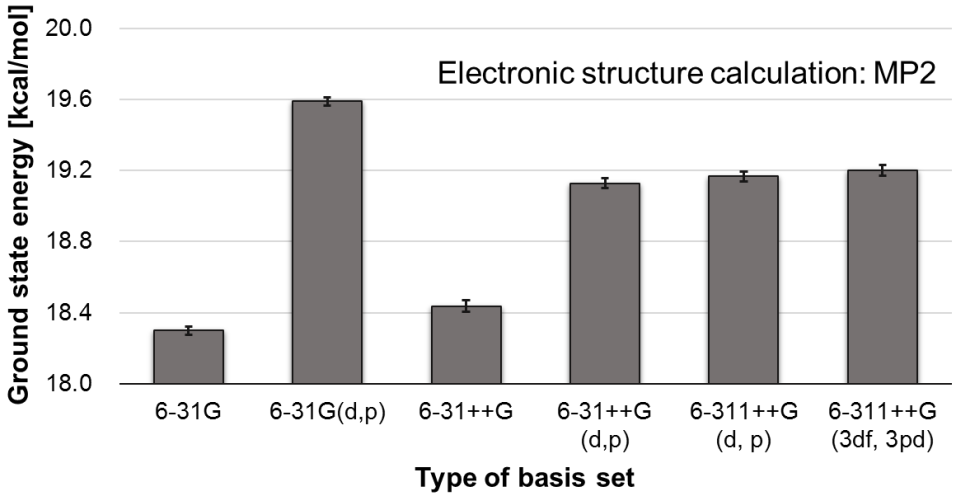


FIG. 5.7. Dependency of the variational path integral molecular dynamics (VPIMD) results on the quality of the basis set.

5.2.3. Computational Cost

We compared the computational time between our GP model and the ab initio calculations. In the present study, as the VPIMD sampling was performed 10^6 times for 200 beads, so the calculations of potential energy and force were required 2×10^8 times in total. Here, we estimated the computational time when these calculations were performed by using 8-cores of the Intel Core i7-9700K: the computational cost for ab initio MP2/6-31++G(d, p) was found to be 200 days for OH^-, 1,700 days for $H_3O_2^-$, and 10,000 days for $H_5O_3^-$. In contrast, the prediction cost for the GP model is $\mathcal{O}(n)$ and was found to require only 0.5 hours for OH^-, 5.0 hours for $H_3O_2^-$, and 20 hours for $H_5O_3^-$. This showed that the GP model enabled approximately 10,000 times faster VPIMD simulation.

5.3. Deprotonated Water Clusters $(H_2O)_{n-1}OH^-$

5.3.1. Energy

In this section, we discuss the nuclear ground-state energy of the deprotonated water clusters $(H_2O)_{n-1}OH^-(n = 1 - 3)$. The total energies for deprotonated water clusters $(H_2O)_{n-1}OH^-$, $(D_2O)_{n-1}OD^-$, and $(T_2O)_{n-1}OT^-$ up to $n = 3$ calculated by the harmonic approximation, the Numerov method, and the VPIMD method are summarized in Table 5.4, where the VPIMD energy E_0 has been calculated using the mixed estimator as expressed in Section 5.1.1. First, the energies computed by the Numerov method are found to be equivalent to by the VPIMD method within the error margin. This fact clearly shows that our VPIMD calculations can project out the exact ground-state according to Equation 2.30.

Table 5.4. The total energies for deprotonated water clusters $(H_2O)_{n-1}OH^-$, $(D_2O)_{n-1}OD^-$, and $(T_2O)_{n-1}OT^-$ up to $n = 3$ calculated by the harmonic approximation, E_{HA}, the Numerov method, E_0, and the variational path integral molecular dynamics (VPIMD) method, E_0. The unit of energy is the kcal/mol (1 hartree = 627.51 kcal/mol). The statistical error of the last digit is indicated in parentheses.

	E_{HA}	E_0 by Numerov	E_0 by VPIMD
$OH^-(n = 1)$	5.45	5.40	5.40(1)
$OD^-(n = 1)$	3.96	3.94	3.94(1)
$OT^-(n = 1)$	3.33	3.31	3.31(1)
$H_3O_2^-(n = 2)$	20.11	-	19.13(3)
$D_3O_2^-(n = 2)$	14.78	-	14.13(2)
$T_3O_2^-(n = 2)$	12.48	-	11.97(2)
$H_5O_3^-(n = 3)$	36.57	-	35.64(2)
$D_5O_3^-(n = 3)$	26.85	-	26.38(2)
$T_5O_3^-(n = 3)$	22.65	-	22.32(1)

Then, for all the clusters, the VPIMD calculation gives the lower energy than the harmonic approximation, $E_{\mathrm{HA}} > E_0$. Therefore, we expect that deprotonated water clusters are strongly affected by the anharmonic effects and the harmonic approximation cannot analyze the quantum properties of these clusters, quantitatively. Moreover, we focus on the isotope effects on the total energy. As expected in Section 5.1.1, the total energies are reduced by replacing the hydrogen atoms with the Deuteriums or Tritiums. Therefore, it is considered that isotope substitution with a heavy atom suppresses the nuclear quantum effects.

Here, the other important result is that the ground state energies are much higher than the thermodynamic energy. For example, the total energies for $(\mathrm{H_2O})_{n-1}\mathrm{OH}^-$ are 5.40(1) kcal/mol (1,359 K/atom) for OH^-, 19.13(3) kcal/mol (1,925 K/atom) for $\mathrm{H_3O_2^-}$, and 35.64(2) kcal/mol (2,242 K/atom) for $\mathrm{H_5O_3^-}$. These facts show that the PESs we constructed in Section 5.2 have effective ranges for not only our VPIMD simulation but also global applications.

5.3.2. Hydrogen bonds and Nuclear wavefunctions

In this section, we investigate the hydrogen bond length and the nuclear wavefunction of deprotonated water clusters. However, in contrast to the protonated water clusters, it is considered that nuclear wavefunctions of protonated water clusters include only the small number of inherent structures because the molecular degree of freedom is small. Thus, other analysis would be required to visualize the spread of the nuclear wavefunctions. In the present thesis, we employed the simplest approach plotting the nuclear wavefunction in the Cartesian space. Here, the hydrogen bond lengths and the nuclear wavefunctions

of deprotonated water clusters $H_3O_2^-$ and $H_5O_3^-$ are shown in Table 5.5 and Figure 5.8, respectively.

For $H_3O_2^-$ in Figure 5.9 (a), the wavefunction of the gray hydrogen is delocalized widely around the O–O axis. In addition, we examine the hydrogen atom shared between two oxygens. As shown in Table 5.5, the hydrogen atom is biased to one oxygen when the nuclear quantum effect is not considered: the distances are 1.397 Å and 1.094 Å, respectively. However, the hydrogen bond lengths calculated from the nuclear wavefunction are 1.257 and 1.255 Å, respectively, which indicates that the central hydrogen is shared equally between the two oxygen atoms. Moreover, the nuclear quantum effect increases R_1 from 1.245 to 1.256 Å, similar to the case of protonated water clusters. In contrast, the average bond length between two oxygens R_2 did not change.

We next discuss the nuclear wavefunctions and bond lengths for $H_5O_3^-$. The nuclear wavefunction in Figure 5.9 (c) indicates that the hydrogen atom is widely delocalized around the O–O axis; this trend is the same as $H_3O_2^-$. On the other hand, unlike $H_3O_2^-$, the hydrogen between the oxygen atoms is unevenly distributed to the outer oxygen even if the nuclear quantum effect is considered. Interestingly, the hydrogen bond length R_1 is reduced by the consideration of the nuclear quantum effect.

TABLE 5.5. Bond lengths of deprotonated water clusters by the VPIMD calculations, where the label "min" indicates the bond lengths evaluated for the potential energy minimum structure. The label H, D, and T indicate the bond lengths calculated for $(H_2O)_{n-1}OH^-$, $(D_2O)_{n-1}OD^-$, and $(T_2O)_{n-1}OT^-$, respectively. Four types of the bond, R_{OH_1}, R_{OH_2}, R_1 and R_2 are defined in Figure 5.8. Length is given in units of Å (1 bohr = 0.5292 Å).

cluster size		min	H	D	T
	R_{OH_1}	1.397	1.257	1.254	1.256
	R_{OH_2}	1.094	1.255	1.248	1.239
$H_3O_2^-$	R_1	1.245	1.256	1.251	1.247
$(n = 2)$	R_2	2.489	2.489	2.486	2.481
$H_5O_3^-$	R_1	1.548	1.535	1.543	1.546
$(n = 3)$	R_2	2.575	2.575	2.579	2.583

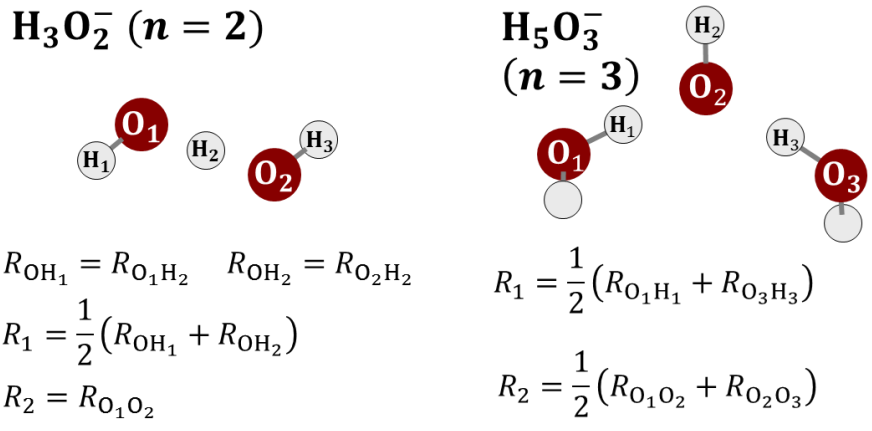


FIG 5.8. Definition for four types of the bond length, where R_{OH_1}, R_{OH_2}, and R_1 correspond to the hydrogen bond length and R_2 is the bond length between oxygens.

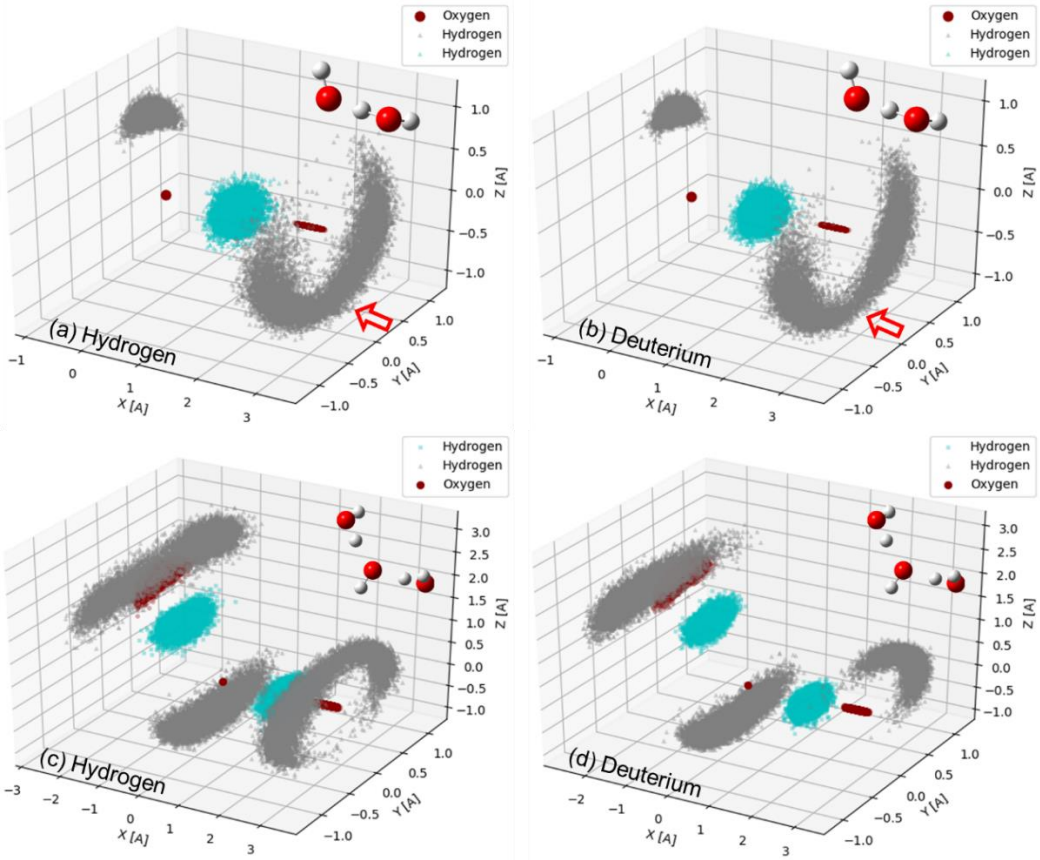


FIG. 5.9. Quantum fluctuations in the $H_3O_2^-$ and $H_5O_3^-$ structures, where red, gray, and blue dots indicate oxygen, hydrogen, and hydrogen-bonding hydrogen, respectively. For $H_3O_2^-$, one of oxygens in the origin, the two oxygen atoms sharing the proton on the X axis, and a hydrogen atom adjacent to the atom at the origin are in the XY plane. For $H_5O_3^-$, a third oxygen is in the XY plane.

Then, we considered the isotopic substitution. As expected in Section 5.3.1, it is considered that isotope substitution with a heavy atom suppresses the spread of the nuclear wavefunction. To clarify this, we have also analyzed the nuclear wavefunction after the isotopic substitution. For $D_3O_2^-$, the isotope substitution does not change the trend for the delocalization of hydrogen around the O-O axis. However, with a decrease in the total energy, the distribution of the hydrogen around the activation state (indicated by the arrow in Figure 5.9 (a) and (b)) is decreased. In addition, the bond length between

the shared hydrogen and the oxygen atoms, R_1, was decreased to 1.251 Å by isotope effects. We further confirmed that the tritiation reduced the bond length R_1 to 1.247 Å. Then, when investigating the distance between the two oxygen atoms, the bond lengths R_2 were 2.489 Å for hydrogen, 2.486 Å for deuterium, and 2.481 Å for tritium, and were reduced by the isotope effects as seen for the bond length between the central hydrogen and oxygen.

For $H_5O_3^-$, decreasing the total energy suppress the spread of the hydrogen around the O–O axis and make the rotation more difficult; for example, Figure 5.9 (d) shows that the gray hydrogens on the right side are only distributed in the range over zero (> 0 Å). In addition, the bond length R_1 is increased to 1.543 Å by the deuteration and to 1.546 Å by the tritiation in contrast to the case of $H_3O_2^-$. Then, we also investigated the bond length between oxygens: the lengths were 2.575 Å for $H_5O_3^-$, 2.579 Å for $D_5O_3^-$, and 2.580 Å for $T_5O_3^-$, and followed an increasing trend.

5.4. Difference between $(H_2O)_nH^+$ and $(H_2O)_{n-1}OH^-$

We finally discuss the differences between protonated and deprotonated water clusters only up to $n = 3$. For the total energy, both protonated and deprotonated water clusters are strongly affected by the anharmonic effect and there was no explicit difference. In contrast, some differences were found in the results of the bond length and the nuclear wavefunction.

The first difference is that the hydrogen bond length R_1 of the $H_5O_2^-$ is decreased by the nuclear quantum effects. As demonstrated in Sections 5.1.2 and 5.3.2, the hydrogen bond length for protonated water clusters is increased in the same situation. This reason can be considered to be in the number of protons. As shown in Figure 5.10, the number

of protons related to hydrogen bond length is higher in protonated water clusters than in deprotonated water clusters: In other words, the protonated water clusters are expected to be easily affected by the spatial spread of the proton's wavefunction. Here, as the protons clearly repel each other, it is expected that the hydrogen bond of protonated water clusters was lengthened by considering nuclear quantum effects. In contrast, in the case of deprotonated water clusters, it can be predicted that the hydrogen bond was shortened by the nuclear quantum effect because the attraction between the proton and oxygen worked stronger than the repulsion between protons.

The second one is that the ionic core of the $H_3O_2^-$ was changed from OH^- to $H_3O_2^-$ by the nuclear quantum effects. Here, it is important to note that the ionic core in all other clusters was kept even if the nuclear quantum effect is considered, where H_3O^+, $H_5O_2^+$, and OH^- are known as the typical ionic core. This result clearly indicates the possibility that the nuclear quantum effects change the structure of the ionic core.

The third point is in the spread of nuclear wavefunctions. To visualize this, we plot the nuclear wavefunctions of $H_5O_2^+$, $H_7O_3^+$, $H_3O_2^-$, and $H_5O_3^-$ in Figure 5.11. Comparing Figures 5.11 (a) and (b), the distributions of the gray hydrogen are different: its range is approximately 360° for $H_5O_2^+$ and 180° for $H_3O_2^-$. In contrast, the gray hydrogens of $H_7O_3^+$ and $H_5O_3^-$ are distributed in the range of about 180° and 360°, respectively. In addition, for $H_7O_3^+$, the distribution for gray hydrogen neighboring to the origin's oxygen is clearly restricted when comparing with the $H_5O_3^-$. The reason is expected to be in the most stable structure at the potential energy minima: indeed, Figure 5.12 shows that the gray hydrogen of $H_7O_3^+$ is unstable on the O-O axis.

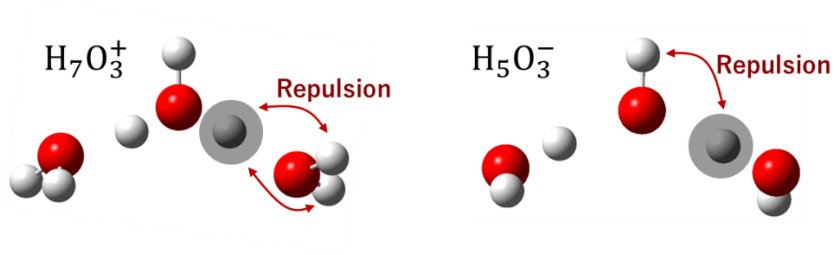


FIG. 5.10. Conceptual figure for the proton related to the hydrogen bond length.

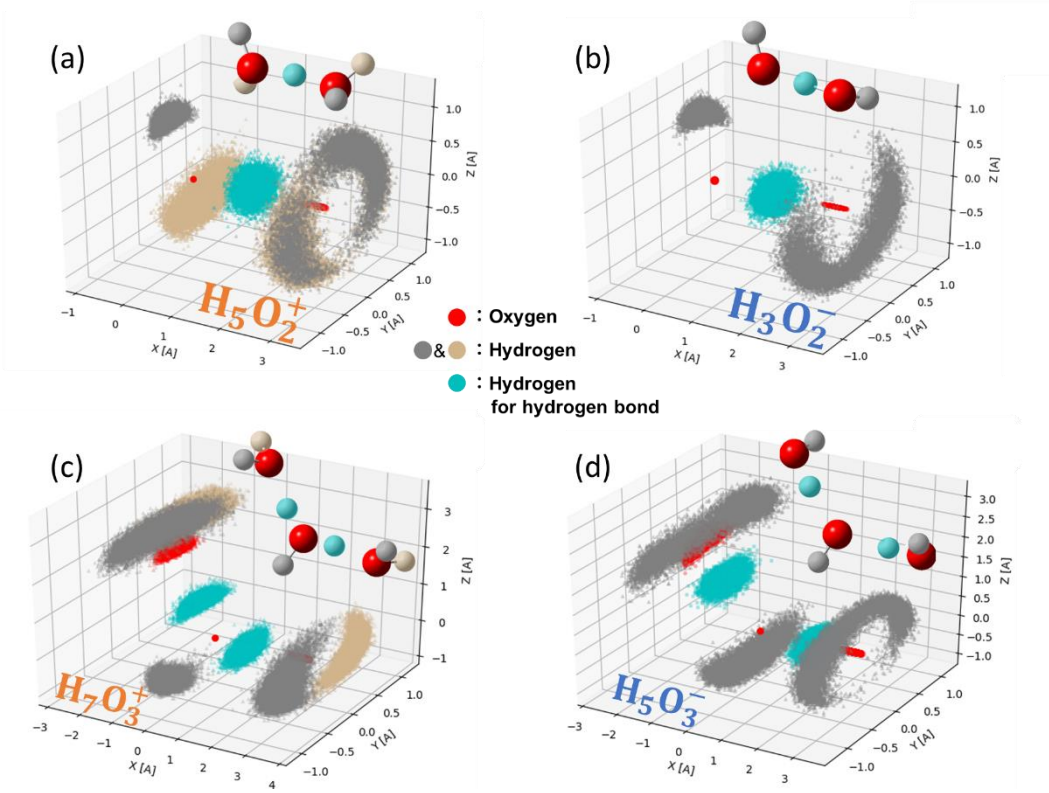


FIG. 5.11. Nuclear wavefunctions in (a) $H_5O_2^+$, (b) $H_3O_2^-$, (c) $H_7O_3^+$, and (d) $H_5O_3^-$ structures, where red, gray (orange), and blue dots indicate oxygen, hydrogen, and hydrogen-bonding hydrogen, respectively. For $H_5O_2^+$ and $H_3O_2^-$, one of oxygens in the origin, the two oxygen atoms sharing the proton on the X axis, and a hydrogen atom adjacent to the atom at the origin are in the XY plane. For $H_7O_3^+$ and $H_5O_3^-$, a third oxygen is in the XY plane.

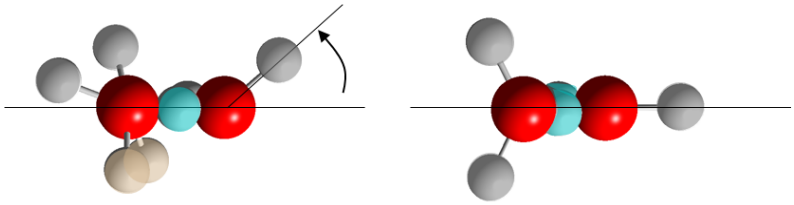


FIG. 5.12. Views of the side of the stable structures for $H_7O_3^+$ and $H_5O_3^-$ at the potential energy minima, where the atomic colors correspond to Figure 5.11.

6. CONCLUSION

6.1. Protonated Water Clusters

We have applied the variational path integral molecular dynamics (VPIMD) method to the protonated water clusters $(H_2O)_nH^+$ $(n = 1 - 5)$. Ground state energetics and structural fluctuation of the clusters were analyzed. The harmonic approximation was also applied to the clusters to extract the anharmonic effect using the numerically exact total energy obtained by the VPIMD method. Remarkable anharmonic effect has been found for the protonated water clusters studied. The anharmonicity for the clusters for $n \geq 3$ is characterized not only by a single potential energy minimum, but also by multiple potential energy minima, as revealed by the inherent structure analysis. Indeed, the nuclear wavefunction for the clusters was found to be delocalized in the configuration space so as to cover multiple local minima around the specified potential energy minimum. Regarding the isotope effect, it has been demonstrated that substituting the hydrogen atom with a heavier isotope spatially localizes the nuclear wavefunction in the configuration space and reduces the quantum kinetic energy, changing the stability order for the isomers of the pentamer.

6.2. Machine Learning Potential Energy Surface

For deprotonated water clusters, we constructed potential energy surfaces to describe the interatomic interactions using Gaussian process regression (GPR) in advance of the VPIMD simulations. Here, for the construction of the machine-learning PES (ML-PES), two algorithms we developed were utilized in order to solve two problems on the GPR.

Problem (1) is that the computational cost is exponentially increased with increasing the number of training data. For this problem, we have developed the fragmentation approach and succeed to improve the RMSE without the use of much training data.

Problem (2) is that the ML-PES constructed by the GPR does not satisfy the molecular symmetry. To solve this problem, we have found two conditions to corporate the molecular symmetry to the GPR algorithm.

As a result, we succeed to obtain sufficiently accurate MLPESs with an RMSE of less than 1.0 kcal/mol (/Å). In addition, we found that our MLPES can calculate potential energy 10,000 times faster than ab initio MP2 level calculations.

6.3. Deprotonated Water Clusters

We applied the VPIMD method and MLPESs to the deprotonated water clusters, $(H_2O)_{n-1}OH^-$ $(n = 1 - 3)$, to analyze their nuclear quantum properties such as energy, hydrogen bond, and nuclear wavefunction. For the total energy, it was found that deprotonated water clusters are also affected strongly by the anharmonic effects similar to the case of protonated water clusters. However, as discussed in Section 5.4, we found some differences between protonated and deprotonated water clusters in hydrogen bonds and nuclear wavefunctions. For the hydrogen bond, follows were mainly suggested: (1)

the relationship between the number of protons and the hydrogen bond length and (2) the possibility that nuclear quantum effects change the ionic core. For the nuclear wavefunction, it was clarified that the spread of nuclear wavefunctions is different between protonated and deprotonated water clusters. Regarding isotope effects, it has been demonstrated that isotopic substitution spatially localizes the nuclear wavefunction similar to protonated water clusters; in contrast to the trend of the protonated water clusters, the hydrogen bond length for $H_5O_2^-$ was reduced.

6.4. Summary of This Thesis

Finally, we conclude this thesis. We have exactly investigated the ground-state natures of protonated and deprotonated water clusters using the VPIMD method. Regarding the interatomic interaction, the OSS3 PES was employed for the protonated water clusters; for deprotonated water clusters, we newly suggested two machine learning algorithms to solve problems on conventional GPR and applied them to construct the MLPES. The resulting VPIMD simulation suggested the following possibilities for the ground state: (1) the structural stability is changed by anharmonic effects, (2) anharmonic effects are characterized not only by a single potential energy minimum but also by multiple potential energy minima, (3) nuclear quantum effects increase or decrease the hydrogen bond length depending on the number of near protons, (4) the structure of the ionic core is changed by the nuclear quantum effects, (5) isotopic substitution spatially localizes the nuclear wavefunction, and (6) isotopic substitution makes the hydrogen bond length closer to the structure with the minimum potential energy. We believe that our results help the interpretation of the ground state nature of protonated and deprotonated water clusters.

7. APPENDIX

7.1. Forth order Trotter expansion

The fourth-order approach for the Trotter decomposition is one of the most common technics to construct the short-time propagators [106–108]. However, the computation of the commutator between $\hat{V}$ and $\hat{T}$ is required in contrast to the second-order approach as following form:

$$e^{-2\beta\hat{H}} = e^{-2\beta(\hat{T}+\hat{V})} = e^{-\beta\hat{V}_e/3}e^{-\beta\hat{T}}e^{-4\beta\hat{V}_O/3}e^{-\beta\hat{T}}e^{-\beta\hat{V}_e/3} + O(\beta^5), \tag{7.1}$$

where

$$\hat{V}_e = \hat{V} + \frac{\alpha}{6}\beta^2\left[\hat{V},\left[\hat{T},\hat{V}\right]\right], \tag{7.2}$$

$$\hat{V}_O = \hat{V} + \frac{(1-\alpha)}{12}\beta^2\left[\hat{V},\left[\hat{T},\hat{V}\right]\right], \tag{7.3}$$

and α is an arbitrary parameter within the range [0, 1] and generally conditioned so that the variational energy is minimized.

7.2. Neural Network and Gaussian Process

A neural network (NN) with a single hidden layer ($N_{\mathrm{hl}} = 1$), in principle, agrees with Gaussian process (GP) in the limit of infinity of the artificial neuron ($N_{\mathrm{an}} = \infty$) [143,144]. For the detailed understanding, we briefly explain the proof of this in this section. An

output of the NN, $\hat{V}$, corresponding to an input vector $\boldsymbol{R}$ is assumed to be given by:

$$\hat{V} = \sum_{i=1}^{N_{\mathrm{an}}} v_i \phi_i(\boldsymbol{R}), \tag{7.4}$$

$$\phi_i(\boldsymbol{R}) = f\left(b + \sum_{j=1}^{D} W_{ij} R_j \right), \tag{7.5}$$

where v_i and W_{ij} denote elements of weight constant vector and matrix, respectively, b denotes the bias parameter, and the function $f(\cdot)$ is the nonlinear transformation. Moreover, let W, v, and b are independent and identically distributed with zero mean and σ_W, σ_v, and σ_b variance, respectively.

$$W_{ij} \sim \mathcal{N}(0, \sigma_W), \qquad v_i \sim \mathcal{N}(0, \sigma_v / N_{\mathrm{an}}), \qquad b = \mathcal{N}(0, \sigma_b) \tag{7.6}$$

Then, in order to investigate the relationship between the NN and GP, we discuss the mean and variance of the distribution that the output of NN follows. The mean for distribution $\hat{V}$ of Equation 7.4 is given by:

$$\mathbb{E}\left[\hat{V}\right] = \mathbb{E}\left[\sum_{i=1}^{N_{\mathrm{an}}} v_i \phi_i(\boldsymbol{R})\right] = \sum_{i=1}^{N_{\mathrm{an}}} \mathbb{E}[v_i \phi_i(\boldsymbol{R})] = \sum_{i=1}^{N_{\mathrm{an}}} \mathbb{E}[v_i]\mathbb{E}[\phi_i(\boldsymbol{R})] = 0 \tag{7.7}$$

using $\mathbb{E}[v_i] = 0$ and the variance is evaluated by:

$$\begin{aligned}
\mathbb{E}\left[\hat{V}\hat{V}'\right] - \mathbb{E}\left[\hat{V}\right]\mathbb{E}\left[\hat{V}'\right] &= \mathbb{E}\left[\hat{V}\hat{V}'\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{N_{\mathrm{an}}} v_i \phi_i(\boldsymbol{R})\right)\left(\sum_{j=1}^{N_{\mathrm{an}}} v_j \phi_j(\boldsymbol{R}')\right)\right] \\
&= \sum_{i=1}^{N_{\mathrm{an}}} \sum_{j=1}^{N_{\mathrm{an}}} \mathbb{E}\left[v_i v_j\right] \mathbb{E}\left[\phi_i(\boldsymbol{R}) \phi_j(\boldsymbol{R}')\right] \\
&= \sum_{i=1}^{N_{\mathrm{an}}} \sum_{j=1}^{N_{\mathrm{an}}} \delta_{ij} \frac{\sigma_v}{N_{\mathrm{an}}} \mathbb{E}\left[\phi_i(\boldsymbol{R}) \phi_j(\boldsymbol{R}')\right]
\end{aligned}$$

$$= \frac{\sigma_v}{N_{\mathrm{an}}} \sum_{i=1}^{N_{\mathrm{an}}} \mathbb{E}[\phi_i(\boldsymbol{R})\phi_i(\boldsymbol{R}')] \tag{7.8}$$

$$= k(\boldsymbol{R}, \boldsymbol{R}'), \tag{7.9}$$

where

$$\mathbb{E}\left[v_i v_j\right] = \delta_{ij} \frac{\sigma_v}{N_{\mathrm{an}}}. \tag{7.10}$$

Therefore, the output of the NN, $\hat{V}$, follows the Gaussian distribution with zero mean and $k(\boldsymbol{R}, \boldsymbol{R}')$ variance:

$$\hat{V} \sim \mathcal{N}(0, k). \tag{7.11}$$

AKNOWLEDGEMENTS

First of all, I would like to thank my supervisor, Assistant Professor Tomonori Ida. His openness to new ideas has made my student life fruitful.

I am also grateful to Professor Motohiro Mizuno. His advice from the experimental perspective has made my field of view wider. In addition, I am fortunate to have been part of the Mizuno and Ida research group where we have the chance to touch only theoretical research but also experimental research. I would particularly like to thank Assistant Professor Yuta Hori for his helpful, kind, and fruitful comments on my research and student life. My thanks also go to the present and previous members, Shogo Amemori, Yasuhiro Shigeta, Takuya Kurihara, Lim-Fang Han, Manami Nishida, Masato Mogi, Toshiya Suetake, Masashi Annen, Tomoya Fukata, and Ryota Watanabe.

I am especially pleased to thank Roman V. Krems, who accepted my research visiting and gave me much knowledge on the Gaussian process regression, Bayesian optimization, etc. Without his encouragement, help, and guidance, the project on the GPR would not have materialized.

I also express our sincere thanks to Professor Shinichi Miura. He has contributed much to this thesis through many discussions and e-mail exchanges, which have ranged from the techniques of the variational path integral molecular dynamics to detailed comments on the presentation.

This work was supported by the TOBITATE! Young Ambassador Program (No. S191N133010001; Japan)

8. REFERENCE

[1] P. Atkins and J. D. Paula, Atkins Physical Chemistry 8th Edition (Oxford University Press, Oxford, 2006), Chap. 2, pp. 765.

[2] V. Grotthuss, *C. J. D. Ann. Chim.* **58**, 54 (1806).

[3] B. C. H. Steele and A. Heinzel, *Nature* **414**, 345 (2001).

[4] L. Carrette, K. A. Friedrich, and U. Stimming, *ChemPhysChem* **1**, 162 (2001).

[5] K.-D. Kreuer, S. J. Paddison, E. Spohr and M. Schuster, *Chem. Rev*. **104**, 4637 (2004).

[6] J.-H. Wee, *Renewable Sustainable Energy Rev.* **11**, 1720 (2007).

[7] E. Fontananova, A. Brunetti, F. Trotta, M. Biasizzo, E. Drioli, and G. Barbieri, *Fuel Cells* **13**, 86 (2013).

[8] K. D. Kreuer, *Chem. Mater.* **8**, 610 (1996).

[9] G. N.-Schatteburg and V. E. Bondybey, *Chem. Rev.* **100**, 11, 4059 (2000).

[10] D. Marx, *Chem. Phys. Chem.* **7**, 1848 (2006).

[11] F. Paesani and G. A. Voth, *J. Phys. Chem. B* **113**, 5702 (2009).

[12] D. Marx, A. Chandra, and M. E. Tuckerman, *Chem. Rev.* **110**, 2174 (2010).

[13] J. A. Fournier, C. T. Wolke, M. A. Johnson, T. T. Odbadrakh, K. D. Jordan, S. M. Kathmann, and S. S. Xantheas, *J. Phys. Chem. A* **119**, 9425 (2015).

[14] M. Ceriotti, W. Fang, P. G. Kusalik, R. H. McKenzie, A. Michaelides, M. A. Morales, and T. E. Markland, *Chem. Rev.* **116**, 7529 (2016).

[15] N. Agmon, H. J. Bakker, R. K. Campen, R. H. Henchman, P. Pohl, S. Roke, M. Thämer, and A. Hassanali, *Chem. Rev.* **116**, 7642 (2016).

[16] S. Woutersen and H. J. Bakker, *Phys. Rev. Lett.* **96**, 138305 (2006).

[17] M. Thamer, L. D. Marco, K. Ramasesha, A. Mandal, and A. Tokmakoff, *Science* **350**, 78 (2015).

[18] S. -S. Lin, *Rev. Sci. Instrum.* **44**, 516 (1973).

[19] H. A. Schwarz, *J. Chem. Phys.* **67**, 5525, (1977).

[20] M. Okumura, L. I. Yeh, J. D. Myers, and Y. T. Lee, *J. Chem. Phys.* **85**, 2328 (1986).

[21] L. I. Yeh, M. Okumura, J. D. Myers, J. M. Price, and Y. T. Lee, *J. Chem. Phys.* **91**, 7319 (1989).

[22] L. I. Yeh, Y. T. Lee, and J. T. Hougen, *J. Mol. Spectrosc.* **164**, 473 (1994).

[23] J.-C. Jiang, Y.-S. Wang, H.-C. Chang, S. H. Lin, Y. T. Lee, G. Niedner-Schatteburg, and H.-C. Chang, *J. Am. Chem. Soc.* **122**, 1398 (2000).

[24] C. Chaudhuri, Y. S. Wang, J. C. Jiang, Y. T. Lee, H. C. Chang, and G. N-Schatteburg, *Mol. Phys.* **99**, 1161 (2001).

[25] K. R. Asmis, N. L. Pivonka, G. Santambrogio, M. Brümmer, C. Kaposta, D. M. Neumark, and L. Wöste, *Science* **299**, 1375 (2003).

[26] W. H. Robertson, E. G. Diken, E. A. Price, J. W. Shin, and M. A. Johnson, *Science* **299**, 1367 (2003).

[27] T. D. Fridgen, T. B. McMahon, L. MacAleese, J. Lemaire, and P. Maitre, *J. Phys. Chem. A* **108**, 9008 (2004).

[28] M. Miyazaki, A. Fujii, T. Ebata, and N. Mikami, *Science* **304**, 1134 (2004).

[29] J.-W. Shin, N. I. Hammer, E. G. Diken, M. A. Johnson, R. S. Walters, T. D. Jaeger, M. A. Duncan, R. A. Christie, and K. D. Jordan, *Science* **304**, 1137 (2004).

[30] J. M. Headrick, E. G. Diken, R. S. Walters, N. I. Hammer, R. A. Christie, J. Cui, E. M. Myshakin, M. A. Duncan, M. A. Johnson, and K. D. Jordan, *Science* **308**, 1765 (2005).

[31] G. E. Douberly, A. M. Ricks, and M. A. Duncan, *J. Phys. Chem. A* **113**, 8449 (2009).

[32] G. E. Douberly, R. S. Walters, J. Cui, K. D. Jordan, and M. A. Duncan, *J. Phys. Chem. A* **114**, 4570 (2010).

[33] K. Mizuse and A. Fujii, *J. Phys. Chem. A* **116**, 4868 (2012).

[34] N. Heine, M. R. Fagiani, M. Rossi, T. Wende, G. Berden, V. Blum, and K. R. Asmis, *J. Am. Chem. Soc.* **135**, 8266 (2013).

[35] N. Heine, M. R. Fagiani, and K. R. Asmis, *J. Phys. Chem. Lett.* **6**, 2298 (2015).

[36] C. T. Wolke, J. A. Fournier, L. C. Dzugan, M. R. Fagiani, T. T. Odbadrakh, H. Knorke,

K. D. Jordan, A. B. McCoy, K. R. Asmis, and M. A. Johnson, *Science* **354**, 1131 (2016).

[37] M. R. Fagiani, H. Knorke, T. K. Esser, N. Heine, C. T. Wolke, S. Gewinner, W. Schöllkopf, M.-P. Gaigeot, R. Spezia, M. A. Johnsonc, and K. R. Asmis, *Phys. Chem. Chem. Phys*. **18**, 26743 (2016).

[38] T. K. Esser, H. Knorke, K. R. Asmis, W. Schöllkopf, Q. Yu, C. Qu, J. M. Bowman, and M. Kaledin, *J. Phys. Chem. Lett.* **9**, 798 (2018).

[39] D. C. McDonald, J. P. Wagner, A. B. McCoy, and M. A. Duncan, *J. Phys. Chem. Lett.* **9**, 5664 (2018).

[40] D. Wei and D. R. Salahub, *J. Chem. Phys.* **101**, 7633 (1994).

[41] D. Wei and D. R. Salahub, *J. Chem. Phys.* **106**, 6086 (1997).

[42] M. E. Tuckerman, D. Marx, M. L. Klein, and M. Parrinello, *Science* **275**, 817 (1997).

[43] H.-P. Cheng, *J. Phys. Chem. A* **102**, 6201 (1998).

[44] L. Ojamäe, I. Shavitt, and S. J. Singer, *J. Chem. Phys.* **109**, 5547 (1998).

[45] M. P. Hodges and A. J. Stone, *J. Chem. Phys.* **110**, 6766 (1999).

[46] M. P. Hodges and D. J. Wales, *Chem. Phys. Lett.* **324**, 279 (2000).

[47] S. J. Singer, S. McDonald, and L. Ojamäe, *J. Chem. Phys.* **112**, 710 (2000).

[48] C. V. Ciobanu, L. Ojamäe, I. Shavitt, and S. J. Singer, *J. Chem. Phys.* **113**, 5321 (2000).

[49] J. R. Pliego and J. M. Riveros, *J. Chem. Phys.* **112**, 4045 (2000).

[50] R. A. Christie and K. D. Jordan, *J. Phys. Chem. A* **105**, 7551 (2001).

[51] M. Mella and D. C. Clary, *J. Chem. Phys.* **119**, 10048 (2003).

[52] M. Tachikawa and M. Shiga, *J. Chem. Phys.* **121**, 5985 (2004).

[53] M. Tachikawa and M. Shiga, *J. Am. Chem. Soc.* **127**, 11908 (2005).

[54] M. Mella, J. L. Kuo, D. C. Clarya, and M. L. Kleinc, *Phys. Chem. Chem. Phys.* **7**, 2324 (2005).

[55] O. Vendrell, F. Gatti, and H.-D. Meyer, *J. Chem. Phys.* **131**, 034308 (2009).

[56] O. Vendrell, F. Gatti, and H.-D. Meyer, *Angew. Chem. Int. Ed.* **48**, 352 (2009).

[57] M. Kaledin, A.L. Kaledin, J. M. Bowman, J. Ding, and K. D. Jordan, *J. Phys. Chem. A* **113**, 7671 (2009).

[58] C. Knight and G. A. Voth, *Acc. Chem. Res.* **45**, 101 (2012).

[59] Q. Yu and J. M. Bowman, *J. Chem. Phys.* **146**, 121102 (2017).

[60] Q. Yu and J. M. Bowman, *J. Am. Chem. Soc.* **139**, 10984 (2017).

[61] K. Yagi and B. Thomsen, *J. Phys. Chem. A* **121**, 2386 (2017).

[62] H. Wang and N. Agmon, *J. Phys. Chem. A* **121**, 3056 (2017).

[63] C. H. Duong, O. Gorlova, N. Yang, P. J. Kelleher, M. A. Johnson, A. B. McCoy, Q. Yu, and J. M. Bowman, *J. Phys. Chem. Lett.* **8**, 3782 (2017).

[64] C. H. Duong, N. Yang, P. J. Kelleher, M. A. Johnson, R. J. DiRisio, A. B. McCoy, Q. Yu, J. M. Bowman, B. V. Henderson, and K. D. Jordan, *J. Phys. Chem. A* **122**, 9275 (2018).

[65] L. C. Dzugan, R. J. DiRisio, L. R. Madison, and A. B. McCoy, *Faraday Discuss.* **212**, 443 (2018).

[66] J. P. Heindel, Q. Yu, J. M. Bowman, and S. S. Xantheas, *J. Chem. Theory Comput.* **14**, 4553 (2018).

[67] Q. Yu and J. M. Bowman, *J. Phys. Chem. A* **123**, 1399 (2019).

[68] H. Sugisawa, T. Ida, and S. Miura, *J. Mol. Liq.* **284**, 157 (2019).

[69] L. Ojamäe, I. Shavitt, and S. J. Singer, *Int. J. Quantum Chem. Symp.* **29**, 657 (1995).

[70] R. Pomès and B. Roux, *J. Phys. Chem.* **100**, 2519 (1996).

[71] E. F. Valeev and H. F. Schaefer III, *J. Chem. Phys.* **108**, 7197 (1998).

[72] G. M. Chaban, J. O. Jung, and R. B. Gerber, *J. Phys. Chem. A* **104**, 2772 (2000).

[73] M. V. Vener and O. Kühn, J. Sauer, *J. Chem. Phys.* **114**, 240 (2001).

[74] J. M. Headrick, J. C. Bopp, and M. A. Johnson, *J. Chem. Phys.* **121**, 11523 (2004).

[75] N. I. Hammer, E. G. Diken, J. R. Roscioli, M. A. Johnson, E. M. Myshakin, K. D. Jordan, A. B. McCoy, X. Huang, J. M. Bowman, and S. Carter, *J. Chem. Phys*. **122**,

244301 (2005).

[76] M. Park, I. Shin, N. J. Singh, and K. S. Kim, *J. Phys. Chem. A* **111**, 10692 (2007).

[77] T. L. Guasco, M. A. Johnson, and A. B. McCoy, *J. Phys. Chem. A* **115**, 5847 (2011).

[78] J. M. Bowman, *Acc. Chem. Res.* **19**, 202 (1986).

[79] M. A. Ratner and R. B. Gerber, *J. Phys. Chem.* **90**, 20 (1986).

[80] S. Huang, Z. Sun, and W. A. Lester Jr., *J. Chem. Phys.* **92**, 597 (1990).

[81] X. Lin, H. Zhang, and A. M. Rappe, *J. Chem. Phys.* **112**, 2650 (2000).

[82] M. W. Lee, M. Mella, and A. M. Rappe, *J. Chem. Phys.* **122**, 244103 (2005).

[83] C. J. Umrigar and C. Filippi, *Phys. Rev. Lett.* **94**, 150201 (2005).

[84] J. Toulouse and C. J. Umrigar, *J. Chem. Phys.* **126**, 084102 (2007).

[85] D. M. Ceperley, *Rev. Mod. Phys.* **67**, 279 (1995).

[86] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).

[87] S. Miura, *Chem. Phys. Lett.* **482**, 165 (2009).

[88] T. Morawietz, V. Sharma, and J. Behler, *J. Chem. Phys.* **136**, 064103 (2012).

[89] X. Huang, B. J. Braams, S. Carter, and J. M. Bowman, *J. Am. Chem. Soc.* **126**, 5042 (2004).

[90] A. Kamath, R. A. Vargas-Hernández, R. V. Krems, T. Carrington, and S. Manzhos, *J. Chem. Phys.* **148**, 241702 (2018).

[91] H. Sugisawa, T. Ida, and R. V. Krems, *J. Chem. Phys.* **153**, 114101 (2020).

[92] S. Manzhos and T. Carrington Jr., *Chem. Rev.* (published online, 2020).

[93] R. V. Krems, *Phys. Chem. Chem. Phys.* **21**, 13392 (2019).

[94] B. Bernu and D. M. Ceperley, *J. Chem. Phys.* **93**, 552 (1990).

[95] G. J. Martyna, M. L. Klein, and M. Tuckerman, *J. Chem. Phys.* **97**, 2635 (1992).

[96] D. Bressanini and P. J. Reynolds, *QMC, Adv. Chem. Phys*. **105**, 37 (1999).

[97] R. L. Coldwell, *Int. J. Quantum Chem., Quantum Chem. Symp.* **11**, 215 (1977).

[98] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, *Phys. Rev. Lett.* **60**, 1719 (1988).

[99] M. Snajdr and S. M. Rothstein, *J. Chem. Phys.* **112**, 4935 (2000).

[100] F. J. Gálvez, E. Buendía, and A. Sarsa, *J. Chem. Phys.* **115**, 1166 (2001).

[101] D. Ceperley, G. V. Chester, and M. H. Kalos, *Phys. Rev. B* **16**, 3081 (1977).

[102] S. Baroni and S. Moroni, *Phys. Rev. Lett.* **82**, 4745 (1999).

[103] R. P. Feynman, and A. R. Hibbs, Quantum Mechanics and Path Integrals (McGraw-Hill, New York, 1965).

[104] R. P. Feynman, Statistical Mechanics (Addison-Wesley, Redwood City, 1972).

[105] T. Kashiwa, Y. Ohnuki, and M. Suzuki, Path Integral Methods (Oxford University Press, Oxford, 1997).

[106] S. A. Chin, *Phys. Lett. A* **226**, 344 (1997).

[107] S. Jang, S. Jang, and G. A. J. Voth, *J. Chem. Phys.* **115**, 7832 (2001).

[108] Y. Kamibayashi and S. Miura, *J. Chem. Phys.* **145**, 074114 (2016).

[109] M. E. Tuckerman, B. J. Berne, G. J. Martyna, and M. L. Klein, *J. Chem. Phys*. **99**, 2796 (1993).

[110] M. E. Tuckerman and A. Hughes, in Classical and Quantum Dynamics in Condensed Phase Simulations, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, Singapore, 1999).

[111] M. E. Tuckerman, Statistical Mechanics: Theory and Molecular Simulation (Oxford University Press, New York, 2010).

[112] S. Miura, J. Tanaka, *J. Chem. Phys.* **120**, 2161 (2004).

[113] J. Almlijf, K. Faegri Jr., K. Korsell, J. Comput. Chem. 3, 385 (1982).

[114] J. A. Pople, J. S. Binkley, R. Seeger, Intern. J. Quantum Chem. Symp. 10, 1 (1976).

[115] G. D. Purvis, R. J. J. Bartlett, Chem. Phys. 76, 1910 (1982).

[116] K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, Chem. Phys. Lett.,

157, 479 (1989).

[117] A. Becke, D. J. Chem. Phys. 98, 5648 (1993).

[118] S. S. Xantheas, *J. Chem. Phys.* 102, 4505 (1995).

[119] K. Kim and K. D. Jordan, *J. Phys. Chem.* **98**, 10089 (1994).

[120] S. S. Xantheas, C. J. Burnham, and R. J. Harrison, *J. Chem. Phys.* **116**, 1493 (2002).

[121] A. E. Hoerl and R. W. Kennard. *Technometrics* **12**, 55 (1970).

[122] R. Tibshirani, *J. Roy. Statist. Soc. B* **58**, 267 (1996).

[123] A. E. Hoerl, *Chem. Eng. Prog.* **58**, 54 (1962).

[124] N. R. Draper, *Technometrics* **5**, 469 (1963).

[125] L. Crone, *J. Franklin Inst.* **294**, 133 (1972).

[126] G. C. McDonald, *WIREs Comput. Stat.* **1**, 93 (2009).

[127] H. Zou and T. Hastie, *J. Roy. Statist. Soc. B* **67**, 301 (2005).

[128] B. Widrow and M. A. Lehr, *Proceedings of the IEEE* **78**, 1415 (1990).

[129] J. Zupan and J. Gasteiger, *Analytica Chimica Acta* **248**, 1 (1991).

[130] M. Rocha, P. Cortez, and J. Neves, *Neurocomputing* **70**, 2809 (2007).

[131] D. J. C. MacKay, Information Theory, Inference, and Learning Algorithms (Cambridge University Press, Cambridge, England, 2003).

[132] C. K. I Williams and C. E. Rasmussen. Gaussian processes for machine learning. (MIT Press, Cambridge, MA, 2006).

[133] K. P. Murphy, Machine Learning: A Probabilistic Perspective; (MIT Press, Cambridge, MA, 2012).

[134] T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, *J. Chem. Phys.* **103**, 4129 (1995).

[135] C. M. Handley, G. I. Hawe, D. B. Kellab, and P. L. A. Popelier, *Phys. Chem. Chem. Phys.* **11**, 6365 (2009).

[136] A. P. Bartok, M. C. Payne, R. Kondor, and G. Csanyi, *Phys. Rev. Lett.* **104**, 136403 (2010).

[137] A. P. Bartok and G. Csanyi, *Int. J. Quant. Chem.* **115**, 1051 (2015).

[138] M. Caccin, Z. Li, J. R. Kermode, and A. De Vita, *Int. J. Quant. Chem.* **115**, 1129 (2015).

[139] P. L. A. Popelier, *Int. J. Quant. Chem.* **115**, 1005 (2015).

[140] J. Cui, Z. Li, R. V. Krems, *J. Chem. Phys.* **143**, 154101 (2015).

[141] J. Cui, R. V. Krems, *Phys. Rev. Lett.* **115**, 073202 (2015).

[142] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, *Chem. Rev.* (published online, 2020).

[143] R. M. Neal. *Bayesian Learning for Neural Networks, Lecture Notes in Statistics*, **118**, 29 (Springer, New York, 1996)

[144] J. Lee, Y. Bahri, R. Novak, S. S. Schoenholz, J. Pennington, J. S.-Dickstein, arXiv:1711.00165 (2017)

[145] K.-R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, *Artif. Neural Networks* **1327**, 999 (1997).

[146] B. Schölkopf, A. Smola, and K.-R. Müller, *Neural Comput.* **10**, 1299 (1998).

[147] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, *IEEE Trans. Neural Netw*. **12**, 181 (2001).

[148] V. Deev and M. A. Collins, *J. Chem. Phys.* **122**, 154102 (2005).

[149] A. P. Bartók and G. Csányi, *Int. J. Quantum Chem.* **115**, 1051 (2015).

[150] V. S. Bryantsev, M. S. Diallo, A. C. T. van Duin, and W. A. Goddard, *J. Chem. Theory Comput.* **5**, 1016 (2009).

[151] P. A. Whitlock, D. M. Ceperley, G. V. Chester, and M. H. Kalos, *Phys. Rev. B* **19**, 5598 (1979).

[152] A. Sarsa, K. E. Schmidt, and W. R. Magro, *J. Chem. Phys.* **113**, 1366 (2000).

[153] M. Eigen and L. De Maeyer, Proc. R. Soc. London, Ser. A 247, 505 (1958).

[154] M. Eigen, Angew. Chem. Int. Ed. 3, 1 (1964).

[155] M. L. Huggins, J. Phys. Chem. 40, 723 (1936).

[156] G. Zundel and H. Z. Metzer, Phys. Chem. 58, 225 (1968).

[157] W. W. Cleland, M. M. Kreevoy, Science 264, 1887 (1994).

[158] M. Tachikawa, M. Shiga, J. Am. Chem. Soc. 127, 11908 (2005).

[159] F.H. Stillinger, T.A. Weber, Phys. Rev. A 25, 978 (1982).

[160] F.H. Stillinger, J. Chem. Phys. 89, 4180 (1988).

[161] A.B. Finnila, M.A. Gomez, C. Sebenik, C. Stenson, J.D. Doll, Chem. Phys. Lett. 219, 343 (1994).