

# Modeling and Simulating the Excited-State Dynamics of a System with Condensed Phases: A Machine Learning Approach

Seiji Ueno\* and Yoshitaka Tanimura\*

Cite This: *J. Chem. Theory Comput.* 2021, 17, 3618–3628

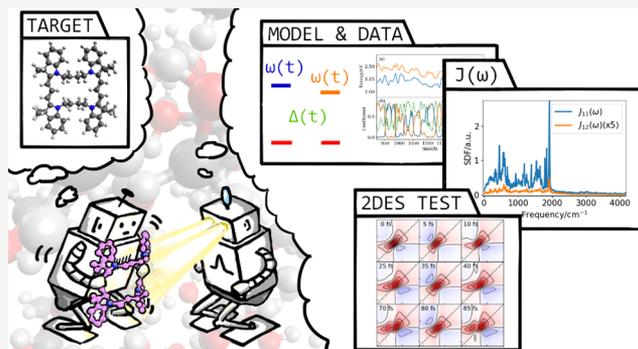
Read Online

ACCESS |

Metrics & More

Article Recommendations

**ABSTRACT:** Simulating the irreversible quantum dynamics of exciton- and electron-transfer problems poses a nontrivial challenge. Because the irreversibility of the system dynamics is a result of quantum thermal activation and dissipation caused by the surrounding environment, it is necessary to include infinite environmental degrees of freedom in the simulation. Because the capabilities of full quantum dynamics simulations that include the surrounding molecular degrees of freedom are limited, employing a system–bath model is a practical approach. In such a model, the dynamics of excitons or electrons are described by a system Hamiltonian, while the other degrees of freedom that arise from the environmental molecules are described by a harmonic oscillator bath (HOB) and system–bath interaction parameters. By extending on a previous study of molecular liquids [*J. Chem. Theory Comput.* 2020, 16, 2099], here, we construct a system–bath model for exciton- and electron-transfer problems by means of a machine learning approach. We determine both the system and system–bath interaction parameters, including the spectral distribution of the bath, using the electronic excitation energies obtained from a quantum mechanics/molecular mechanics (QM/MM) simulation that is conducted as a function of time. Using the analytical expressions of optical response functions, we calculate linear and two-dimensional electronic spectra (2DES) for indocarbocyanine dimers in methanol. From these results, we demonstrate the capability of our approach to elucidate the nonequilibrium exciton dynamics of a quantum system in a nonintuitive manner.



## 1. INTRODUCTION

Quantum dynamics play a significant role in many chemical physics and biochemical physics problems. Frequently studied problems of this kind include exciton- and electron-transfer processes<sup>1,2</sup> that are involved in photosynthetic systems,<sup>3–20</sup> charge-transfer,<sup>21–25</sup> DNA,<sup>26–28</sup> and photovoltaic systems.<sup>29–34</sup> In these problems, the environments (baths), for example, proteins and solvents, play a central role; these baths are complex and strongly coupled to a molecular system of interest at finite temperatures. Recent theoretical studies have demonstrated that such systems and baths are quantum mechanically entangled (bath entanglement), and an understanding of these baths is essential to properly elucidate the quantum dynamics displayed by the system.<sup>35,36</sup> For example, it has been shown that the optimal condition for excitation energy transfer in light-harvesting complexes is realized under non-Markovian system–bath interactions in a strong coupling regime, in which the noise correlation time of the bath is comparable to the time scale of the system dynamics.<sup>10</sup> To conduct high-accuracy simulations with reduced computational costs, some approaches have utilized machine learning methods to develop models that reproduce open quantum dynamics,<sup>37–40</sup> analyze two-dimensional spectroscopy images,<sup>41,42</sup> and

estimate chemical properties for classical molecular dynamics.<sup>43–47</sup>

Although irreversibility of the system dynamics results from quantum thermal activation and dissipation caused by the surrounding environment, it is difficult to conduct a quantum molecular dynamics simulation that exhibits such a characteristic feature arising from macroscopic degrees of freedom. Thus, we introduce a system–bath model in which the dynamics of excitons or electrons are described by a system Hamiltonian, while the other degrees of freedom that arise from environmental molecules are described by a harmonic oscillator bath (HOB). The HOB, whose distribution takes a Gaussian form, exhibits wide applicability in simulating bath effects, despite its simplicity; this is because the influence of the environment can, in many cases, be approximated by a Gaussian process due to the

Received: January 29, 2021

Published: May 17, 2021



cumulative effect of a large number of environmental interactions. In such a situation, the ordinary central limit theorem is applicable, and hence, the Gaussian distribution function is appropriate.<sup>35,48</sup> The distinctive features of the HOB model are determined by the spectral distribution function (SDF) of the coupling strength between the system and the bath oscillators for various frequency values. By choosing the appropriate form of the SDF, the properties of the bath can be adjusted to represent various environments consisting of, for example, solid-state materials<sup>49,50</sup> and protein molecules.<sup>3–5</sup> Because the SDF can be different for different forms of a system Hamiltonian and system–bath coupling, it is difficult to find an optimized Hamiltonian associated with an optimized SDF, in particular, for a bath describing a fluctuation in site–site interaction energy.

In a previous study,<sup>51</sup> we employed a machine learning approach to construct a system–bath model for the intermolecular and intramolecular modes of molecular liquids using atomic trajectories obtained from molecular dynamics (MD) simulations. In this study, we extend the previous approach to investigate an exciton- or electron-transfer problem that is characterized by electronic states embedded in the molecular environment using quantum mechanics/molecular mechanics (QM/MM) calculations to determine the atomic coordinates of molecules. In particular, we focus on the exciton-transfer process of the photosynthesis antenna system to investigate how natural systems can realize such highly efficient yields, presumably by manipulating quantum mechanical processes. As a demonstration, we consider a molecular dimer made of two dipole-coupled dye monomers as a model system that is often studied experimentally and theoretically.<sup>52–54</sup> Then, we construct a model Hamiltonian of an indocarbocyanine dimer compound.<sup>52</sup> The accuracy of this model is examined by calculating linear and two-dimensional electronic spectra.

This paper is organized as follows. In Section 2, we introduce a model that can be used for either exciton or electron transfer and is coupled to a harmonic heat bath. We then describe the machine learning approach that we use to determine the system parameters, the system–bath interactions, and the SDFs on the basis of QM/MM simulations. In Section 3, we present results for an indocarbocyanine dimer model constructed from the analysis of QM/MM trajectories. Linear absorption and two-dimensional spectra are calculated from analytical linear and nonlinear response functional expressions. Section 4 is devoted to concluding remarks.

## 2. THEORY

**2.1. Hamiltonian.** We consider the situations in which an exciton- or electron-transfer system interacts with molecular environments that give rise to dissipation and fluctuation in the system. The Hamiltonian of the system is expressed as

$$\hat{H}_S = \sum_j \hbar\omega_j |j\rangle\langle j| + \sum_{j \neq k} \hbar\Delta_{jk} |j\rangle\langle k| \quad (1)$$

where the  $j$ th exciton or electron states with energies  $\hbar\omega_j$  are represented by bra and ket vectors as  $|j\rangle$  and  $\langle j|$ . The interaction energy between the  $j$ th and  $k$ th states is described by  $\hbar\Delta_{jk}$ . In our model, each state is coupled to a different molecular environment (labeled  $a$ ) that is treated as  $N_a$  harmonic oscillators. The total Hamiltonian is then given by

$$H_{\text{tot}} = H_S - \sum_a \sum_{l=1}^{N_a} \alpha_l^a \hat{V}^a \hat{x}_l^a + \sum_a \sum_{l=1}^{N_a} \left[ \frac{(\hat{p}_l^a)^2}{2m_l^a} + \frac{1}{2} m_l^a (\omega_l^a)^2 (\hat{x}_l^a)^2 \right] \quad (2)$$

where the momentum, position, mass, and frequency of the  $l$ th oscillator in the  $a$ th bath are given by  $\hat{p}_l^a$ ,  $\hat{x}_l^a$ ,  $m_l^a$ , and  $\omega_l^a$ , respectively. The system part of the system–bath interaction is expressed as

$$\hat{V}^a = \sum_{j,k} V_{jk}^a |j\rangle\langle k| \quad (3)$$

where  $V_{jk}^a$  is the coupling constant for the  $a$ th bath between the  $j$  and  $k$  states. The  $a$ th heat bath can be characterized by the spectral distribution function (SDF), defined as

$$J_a(\omega) \equiv \sum_{l=1}^{N_a} \frac{\hbar(\alpha_l^a)^2}{2m_l^a \omega_l^a} \delta(\omega - \omega_l^a) \quad (4)$$

and the inverse temperature is  $\beta \equiv 1/k_B T$ , where  $k_B$  is the Boltzmann constant. Various environments, for example, those consisting of nanostructured materials, solvents, and protein molecules, can be modeled by adjusting the form of the SDF. For the heat bath to act as an unlimited heat source possessing an infinite heat capacity, the number of heat-bath oscillators  $N_a$  is effectively made infinitely large by replacing  $J_a(\omega)$  with a continuous distribution. The above model has been frequently used in the analysis of photosynthetic systems,<sup>6–20</sup> electron-transfer,<sup>21–25</sup> DNA,<sup>26–28</sup> and solar battery systems.<sup>29–34</sup>

**2.2. Learning Data: QM/MM Simulations.** We next consider the pigments in a molecular system, whose electric excitation or exciton states are described by eq 1. The electric states of the pigments depend on the configurations of the surrounding atoms at time  $t$ . The time evolution of the excited states of the system and environmental molecules is described by QM/MM simulations. Because our goal in constructing a system–bath model is to perform a full quantum simulation of the entire system, we should use quantum molecular dynamics (MD) simulations to provide data on the basis of all atomic coordinates. In practice, however, it is impossible to consider large environmental degrees of freedom accurately from a quantum mechanical perspective. Fortunately, we expect that we already have reasonable SDFs for quantum simulation, even though we evaluated them using the classical MD simulation. Such evaluations were conducted utilizing an ensemble of molecular trajectories that exhibit a Gaussian distribution in which the difference between the quantum and classical trajectories is expected to be minor. Further, the dynamics of harmonic oscillators are identical in both the classical and quantum Liouvillian for the  $l$ th oscillator in the  $a$ th bath are expressed as  $\hat{L}_l^a = -(p_l^a/m_l^a)(\partial/\partial x_l^a) - (m_l^a(\omega_l^a)^2)(\partial/\partial p_l^a)$ . We thus use the classical MD simulation technique to acquire the atomic coordinates of the pigments and the molecular environment. We then conduct quantum chemistry calculations to obtain the desired electronic states, typically the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) states of the pigments as a function of time. The excited energy of the  $j$ th pigment is denoted by  $\epsilon_{jj}(t)$ , and the interaction energy between the  $j$ th and  $k$ th pigments that includes the bath-induced fluctuation is denoted by  $\epsilon_{jk}(t)$ ; these

values can be obtained using any kind of numerical program for quantum chemistry calculations. If the main system is too large to enable the evaluation of all electronic states, we evaluate the site energy  $\epsilon_{jj}(t)$  and the interaction energy  $\epsilon_{jk}(t)$  separately. From the calculated  $\epsilon_{jj}(t)$  and  $\epsilon_{jk}(t)$ , we evaluate the system–bath coupling strength in  $\hat{V}_{jk}^a$  and its SDF, in addition to the excitation energy  $\hbar\omega_j$  and the interaction energy  $\Delta_{jk}$ , based on the machine learning approach.

While the SDFs evaluated based on the MD simulations are temperature dependent, the SDFs for the HOB are temperature independent; we therefore eliminate the temperature dependence of optimized parameters, assuming that the sampled MD trajectories exhibit canonical ensembles at finite temperatures.

**2.3. Machine Learning.** For  $n$  exciton or electronic excitation sites, we express the simulated data in terms of  $\epsilon_{jk}(t)$  when describing the excited and site–site interaction energies of interest obtained from the QM/MM simulation. The learning Hamiltonian is then expressed as

$$H(t) = \sum_{j,k=1}^n \epsilon_{jk}(t) |j\rangle\langle k| \quad (5)$$

We then attempt to reproduce the trajectories of  $\epsilon_{jk}(t)$  for the total Hamiltonian, eq 2, with eqs 1 and 3. Although the system–bath model considers an infinite number of degrees of freedom, here, we employ a finite number of bath oscillators to estimate the SDFs. Then, the sampling used for machine learning training is considered the average of the classical bath oscillators for a certain selection of the system and system–bath parameters. The site energy and interaction energy can be expressed as

$$\epsilon_{jj}(t) = \hbar\omega_j - \delta_{jj}(t) \quad (6)$$

and

$$\epsilon_{jk}(t) = \hbar\Delta_{jk} - \delta_{jk}(t) \quad (7)$$

respectively, where  $\delta_{jk}(t)$  is expressed in terms of the linear function of the bath coordinates as

$$\delta_{jk}(t) = \sum_a \alpha_{jk}^a x_{jk}^a(t) \quad (8)$$

Here, the  $a$ th bath coordinate for the  $jk$  site is described as a function of time as

$$x_{jk}^a(t) = A_{jk}^a \sin(\phi_{jk}^a + \omega_{jk}^a t) \quad (9)$$

where  $A_{jk}^a$  and  $\phi_{jk}^a$  are the amplitude and phase of the  $a$ th bath oscillator for the  $jk$  site, respectively. The phase  $\phi_{jk}^a$  is randomly chosen to avoid recursive oscillator motion. Although we can consider such correlated modes separately by introducing additional baths, here, we assume that the influences of the individual bath modes are all independent and that the correlations between the fluctuations among different modes can be ignored.

From eqs 8 and 9,  $\delta_{jk}(t)$  can be expressed as

$$\delta_{jk}(t) = \sum_a c_{jk}^a \sin(\phi_{jk}^a + \omega_{jk}^a t) \quad (10)$$

where

$$c_{jk}^a = \alpha_{jk}^a A_{jk}^a \quad (11)$$

and we treat the system–bath coupling parameters as the product of these two variables. In the machine learning context,

the bath parameters and the system–bath interactions are expressed as a set of latent variables, defined as

$$\theta = (\{\omega_j\}, \{\Delta_{jk}\}, \{c_{jk}^a\}) \quad (12)$$

where  $\{\dots\}$  is the set of system and bath parameters. The trajectories of  $\epsilon_{jj}(t)$  and  $\epsilon_{jk}(t)$ , obtained from the QM/MM calculations, are described as the vibrational motions of the pigment molecule and the surrounding molecules. We then assume that the probability distribution of the pure state energy  $\lambda_i$  is determined based on a Gaussian process and is described by a set of bath parameters  $\alpha_{jk}^a$  and  $\phi_{jk}^a$  by optimizing the probability distribution defined as

$$P(\lambda_i|\theta) = \int \prod_{k,j,a} d\phi_{jk}^a P(\lambda_i|\theta; \phi_{jk}^a) P(\phi_{jk}^a) \quad (13)$$

which represents the marginalization of the phase of the oscillators  $\phi_{jk}^a$  that is introduced to avoid trapping in a local minimal state due to the gradient method. Here,  $P(\phi_{jk}^a)$  is the uniform distribution of  $[0, 2\pi)$  and

$$P(\lambda_i|\theta; \phi_{jk}^a) \propto \exp[-\sigma(\lambda_i - E_i)^2] \quad (14)$$

where  $E_i \equiv E_i(\theta; \phi_{jk}^a)$  is the predicted energy as a function of the parameter set  $\theta$  and initial phase  $\phi_{jk}^a$  for the model Hamiltonian, eq 5, and  $\sigma$  is the error width. Our goal in employing a machine learning method is to choose the optimal parameter set in  $E_i(\theta; \phi_{jk}^a)$  that maximizes the probability distribution for given data  $\lambda_i$ . Among several optimization methods, we use the maximum likelihood method (MLE), where the loss function is expressed in terms of the negative log of the probability as

$$L = \sum_i (\lambda_i - E_i)^2 \quad (15)$$

To find the maximum value of  $L$ , we employ the Adam gradient method for optimization of the parameter set as

$$\theta \leftarrow \theta + \gamma \frac{\partial L}{\partial \theta} \quad (16)$$

where  $\gamma$  is the learning rate. In this way, we obtain the  $J_{jk}$  element of the SDF for the  $jk$  site. Because the energy distribution of each bath oscillator  $E_{jk}^a = (1/2)m(\omega)^2$  is assumed to obey a canonical ensemble, the oscillator amplitude can be expressed as

$$\langle A_{jk}^a \rangle = \frac{1}{\sqrt{\pi\beta m_{jk}^a (\omega_{jk}^a)^2}} \quad (17)$$

Integrating eqs 11 and 17 into eq 4, we obtain

$$J_{jk}(\omega) = \sum_{a=1}^{N_a} \frac{1}{2} \pi \beta \hbar \omega_{jk}^a (c_{jk}^a)^2 \delta(\omega - \omega_{jk}^a) \quad (18)$$

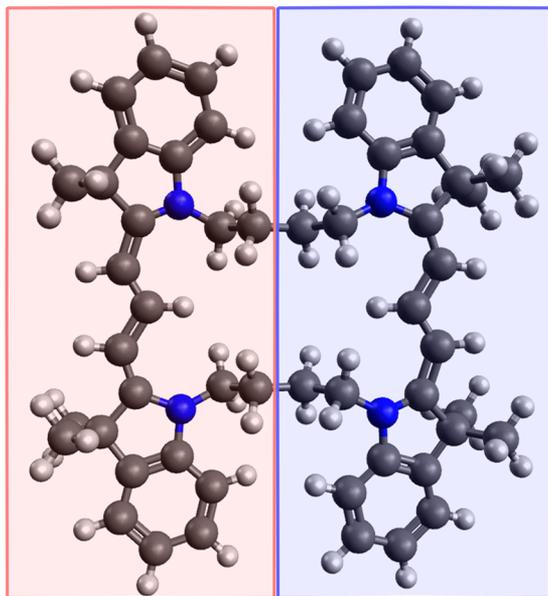
Because  $J_{jk}(\omega)$  rapidly changes over time in accordance with the structural changes in the pigment molecules and environments, we evaluate  $J_{jk}(\omega)$  by averaging the different sample trajectories. From a mathematical perspective,  $c_{jk}^a$  is the frequency domain expression of the time domain data and  $J_{jk}(\omega)$  can be obtained by averaging the power spectra  $c_{jk}^2$  using the Wiener–Khinchin theorem.

It should be noted that the absolute intensity of  $J_{jk}(\omega)$  cannot be determined in the framework of the present study because, for simplicity, we do not evaluate the dipole moment of this complex material; we evaluate the intensity of  $J_{jk}(\omega)$  from the

width of the experimentally obtained linear absorption spectrum.

### 3. NUMERICAL DEMONSTRATION

**3.1. Indocarbocyanine Dimer.** We now demonstrate our numerical approach for a dimer of identical indocarbocyanine molecules.<sup>52</sup> Figure 1 displays the structure of the pigment



**Figure 1.** Molecular structure of the indocarbocyanine dimer. Two pigments are connected by methylene chains. The gray/blue/white atoms represent carbon/nitrogen/hydrogen, respectively. The red square represents pigment 1, whereas the blue square represents pigment 2.

molecule. The ground and excited states of each pigment are expressed as  $|0\rangle_j$  and  $|1\rangle_j$  for  $j = 1$  and  $2$ , respectively. The ground-state energies are each set to zero. The system Hamiltonian is then expressed as

$$\hat{H} = \omega_0(|1\rangle_{11}\langle 1| + |1\rangle_{22}\langle 1|) + \Delta(|0\rangle_{12}\langle 1| + |1\rangle_{12}\langle 0|) \quad (19)$$

where  $\omega_0$  is the excitation energy of a pigment and  $\Delta$  is the interaction energy between the dimers. By diagonalizing  $H$ , we obtained the eigenvalues  $\omega_k$  for the  $k = +$  and  $-$  eigenstates of  $|1+\rangle = (|1\rangle_1|0\rangle_2 + |0\rangle_1|1\rangle_2)/\sqrt{2}$  and  $|1-\rangle = (|1\rangle_1|0\rangle_2 - |0\rangle_1|1\rangle_2)/\sqrt{2}$ , respectively, as

$$\omega_{\pm} = \omega_0 \pm \Delta \quad (20)$$

The excitation energy and interaction energy fluctuations as functions of time, arising from intramolecular motions of the pigment and intermolecular motions of surrounding molecules, are expressed as  $\delta\omega_{\pm}(t)$  and  $\delta\Delta(t)$ , respectively. These functions are evaluated based on the quantum chemistry calculations for given atomic trajectories of the entire molecular system determined by MD simulations.

In our model, because each exciton state is delocalized and the effects of the environmental modes are site specific, we employ an individual heat bath expressed as the sum of site-specific oscillators to describe the energy fluctuation at each exciton site. The distribution of the exciton–oscillator coupling strength is then evaluated based on the machine learning approach. Although it is possible to introduce a global heat bath to induce

low-frequency environmental modes that are coupled to multiple exciton states, we find that such effects are not significant in the present case. Therefore, the excitation energy and interaction energy fluctuations are expressed as

$$\delta\omega_{\pm}(t) = \sum_m^{1,2} w_{\pm,m}(t) \sum_a c_{\omega_{0m}}^a \sin(\omega^a t + \phi_{\omega_{0m}}^a) \quad (21)$$

and

$$\delta\Delta(t) = w_{\Delta}(t) \sum_a c_{\Delta}^a \sin(\omega^a t + \phi_{\Delta}^a) \quad (22)$$

where  $c_{bm}^a$  is the amplitude (scaled by  $\alpha$ , as described in eq 11) and  $\phi_{bm}^a$  is the initial phase of the  $a$ th oscillator for the state indices  $b = 11$  (or  $22$ ) and  $12$ . We introduce the localization weight functions  $w_{\pm,m}(t)$  and  $w_{\Delta}(t)$ , as obtained from the diagonalization of the pigment-based Hamiltonian, expressed in eq 19, to describe the pigment-specific environment effects in the delocalized exciton-state representation. These localization weight functions are evaluated based on the electronic states of the pigment  $m = 1$  and  $2$  established by the atomic orbitals (AO) obtained from quantum chemistry calculations.

Thus, the targeting eigenenergies to be described by the system–bath model,  $\lambda_{\pm}(t; \theta)$ , are expressed as

$$\lambda_{\pm}(t; \theta) = \omega_0 + \delta\omega_{\pm}(t) \pm (\Delta + \delta\Delta(t)) \quad (23)$$

where  $\theta$  is a set of parameters  $\theta = (\omega_0, \{c_{\pm,m}^a\}, \Delta, \{c_{\Delta}^a\})$ . As learning data, we compute the exciton energy  $E_{\pm}(t)$ , the molecular orbital (MO) coefficients for each exciton state, and wave functions (atomic orbital (AO) coefficients for each MO) from quantum chemistry calculations for the given atomic coordinates as a function of time. Additionally, the movements of all atoms in the system are evaluated from the classical MD simulation. Using these data, we optimize the set of parameters  $\theta$ . To evaluate the weight function  $w_{k,m}(t)$ , we calculate the exciton and hole populations  $p_{k,m}^{\text{ex}}(t)$  and  $p_{k,m}^{\text{h}}(t)$  that are obtained as the summation of the absolute square of the AO coefficients, which are evaluated from the AO coefficients involved in the MO in pigment  $m$  for excited state  $k$ . The weight function is then evaluated as  $w_{k,m}(t) = p_{k,m}^{\text{ex}}(t)p_{k,m}^{\text{h}}(t)$  and  $w_{\Delta}(t) = \sum_{k=\pm} (p_{k,1}^{\text{h}}(t)p_{k,2}^{\text{ex}}(t) + p_{k,2}^{\text{h}}(t)p_{k,1}^{\text{ex}}(t))$ . As these definitions indicate, the exciton states are localized when  $w_{\pm,m}$  is close to 1, whereas the exciton states are distributed among the pigments when  $w_{\Delta}$  is close to 1.

To optimize the system and bath parameter set, we minimize the loss function

$$\begin{aligned} L &= \sum_n \sum_t L^n(t) \\ &= \sum_n \sum_t [(\lambda_{-}(t; \theta^n) - E_{-}^n(t))^2 \\ &\quad + (\lambda_{+}(t; \theta^n) - E_{+}^n(t))^2] \end{aligned} \quad (24)$$

where  $E_{-}^n(t)$  and  $E_{+}^n(t)$  are the lowest ( $|1-\rangle$ ) and second lowest ( $|1+\rangle$ ) excitation energies, and the index  $n$  indicates the  $n$ th sample. Using the MLE method, we optimize  $c_{\omega_{0m}}^a$  and  $c_{\Delta}^a$  for each time series as a sample set. To apply the machine learning algorithm, the time series of the tuple  $(E_{-}^n(t), E_{+}^n(t), w_{k,m}^n(t))$  are regarded as the input feature variables. In the indocarbocyanine case, the two pigments are symmetric, and the bath SDFs for each pigment are considered to be identical. Therefore, we use

the averaged value  $c_{\omega_0 a} = (c_{\omega_0 1}^a + c_{\omega_0 2}^a)/2$ . We then evaluate  $J_{ij}(\omega)$  ( $j = 1, 2$ ) and  $J_{12}(\omega)$ , namely,  $J(\omega)$  for  $\omega_0$  and  $\Delta$ , from  $c_{\omega_0}^a$  and  $c_{\Delta}^a$ , respectively, using eq 18.

**3.2. Fourier-Based Approach versus Machine Learning Approach.** A commonly used approach for evaluating the SDFs of  $\epsilon_{ij}(t)$  utilizes the Fourier transformation of the autocorrelation function expressed as  $\mathcal{F}[\langle \delta\epsilon_{ij}(0)\delta\epsilon_{ij}(t) \rangle]$ , where  $\delta\epsilon_{ij}(t) \equiv \epsilon_{ij}(t) - \langle \epsilon_{ij} \rangle$ . In the actual calculation, the time series  $\epsilon_{ij}^n(t)$ , where  $n$  is the sample index, is evaluated as the average of the autocorrelation function expressed as

$$C_{ij}(t) = \frac{1}{N} \sum_n \langle \delta\epsilon_{ij}^n(0)\delta\epsilon_{ij}^n(t) \rangle \quad (25)$$

where  $N$  is the total sample number. We then obtain the SDF as

$$J_{ij}(\omega) = \mathcal{F}[C_{ij}(t)] \quad (26)$$

Alternatively, using the Wiener–Khinchin theorem for stationary random processes, we can obtain the SDF as an average of power spectrum  $P_{ij}^n(\omega) = |\mathcal{F}[\epsilon_{ij}^n(t)]|^2$  as

$$J_{ij}(\omega) = \frac{1}{N} \sum_n P_{ij}^n(\omega) \quad (27)$$

Although this Fourier-based approach is simple and straightforward, for the system–bath Hamiltonian, the obtained SDFs are not necessarily the optimal choice for describing the QM/MM data because the exciton and interaction energies are mutually dependent on each other; thus,  $J_{ij}(\omega)$  and  $J_{ik}(\omega)$  cannot be evaluated separately. In the machine learning approach, however, it is possible to optimize not only  $J_{ij}(\omega)$  and  $J_{ik}(\omega)$  but also  $\omega_0$  and  $\Delta$  without assuming explicit relationships between the SDFs and the system parameters. Moreover, if necessary, we can introduce additional conditions for optimization of the SDFs and system parameters because we employed  $w_{k,m}(t)$  to account for the effects of the indocyanine dimer exciton localization.

**3.3. Calculation Details.** **3.3.1. Step 1: Classical MD.** We prepared a system consisting of an indocyanine dimer molecule with 1024 methanol molecules as the solvent. The classical MD simulations were carried out with the GROMACS software package.<sup>55–57</sup> The conditions for preparation MD simulations were set as 1 atm and 300 K with an NPT ensemble. The equilibrium MD run was carried out for 20 ps in an NVT ensemble, followed by a sampling MD run for 5 ps in an NVE ensemble. These equilibrium MD runs and sampling MD runs were repeated 100 times. The entire MD simulation was performed with a time step of 0.1 fs.

**3.3.2. Step 2: Data Preparation Using Quantum Chemistry Calculations.** To obtain the sample trajectories of the excitation energies, we conducted ZINDO calculations<sup>58,59</sup> and natural transition orbital analysis<sup>60</sup> for a 1 fs period in one sample using the ORCA software package.<sup>61</sup> We then obtained 100 ( $E_-(t)$ ,  $E_+(t)$ ,  $w(t)$ ) samples that were 5 ps in length.

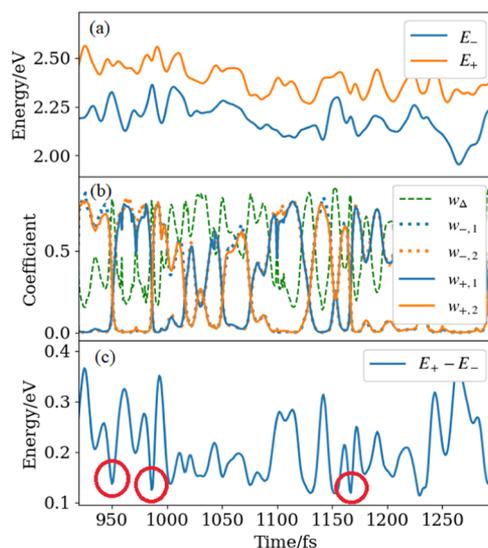
**3.3.3. Step 3: Parameter Optimization for the Machine Learning Approach.** We arranged the data with 5 ps lengths obtained from step 2 according to the starting time in each of 175 steps. We then extracted 604 trajectories containing 1000 data points in an interval of 4 fs. These sampling data were used as the input feature values in the machine learning calculations. To perform learning calculations, we developed Python codes using the TensorFlow library.<sup>62</sup> The training was performed with the learning rate  $\alpha = 1 \times 10^{-4}$  for the first 200 steps and then

the rate was reduced to  $\alpha = 1 \times 10^{-5}$  for the next 200 steps. The number of epochs was chosen to avoid the overfitting problem arising from the MLE that occurs with a gradient method. In the present case, this effect appears in the very low frequency region below  $10 \text{ cm}^{-1}$  of  $J(\omega)$  (see the Appendix B). Because such slow dynamics of the environment are not important in the present exciton-transfer problem, we avoided this effect by simply choosing a shorter epoch known as the early-stopping technique. To minimize the loss function, we employed the Adam algorithm. The bath oscillator number  $N$  is 600. The frequency of the  $a$ th bath oscillator  $\omega^a$  is  $a\Delta\omega$  for  $a = 1, 2, \dots, N$ , where  $\Delta\omega$  is approximately  $8.3391 \text{ cm}^{-1}$ .

The initial values of the target optimization variables for the SDF amplitudes were set as  $c_{\omega_0 m}^a = 1 \times 10^{-5}$  and  $c_{\Delta}^a = 1 \times 10^{-5}$ , and the exciton and interaction energies were set as  $\omega_0 = (\langle E_+ \rangle + \langle E_- \rangle)/2$  and  $\Delta = (\langle E_+ \rangle - \langle E_- \rangle)/2$ . The initial phases  $\phi_b^a$  were randomized 5 times for each series of samples. The loss functions were averaged over each set of 64 samples as a minibatch, while the parameters were optimized for every minibatch. For the 604 samples, each epoch contained 9 iterations.

**3.3.4. Step 4: Calculations of Optical Spectra.** We assumed that the dipole operator for the indocyanine dimer was given by  $\hat{\mu}_1 + \hat{\mu}_2 = \mu(|0\rangle_{11}\langle 1| + |1\rangle_{11}\langle 0| + |0\rangle_{22}\langle 1| + |1\rangle_{22}\langle 0|)$ , which created a transition between the ground state  $|00\rangle$  and the excitation states,  $|1+\rangle$ , and  $|11\rangle$ , while optical transitions from these states to the state  $|1-\rangle$  were forbidden. Thus, the optical transitions in the present system were modeled by a three-level system with eigenenergies of 0,  $\Omega_{\pm}$ , and  $2\omega_0$ . This allowed us to apply analytical expressions of the linear and nonlinear response functions, as presented in Appendix A. We then calculated the linear absorption and two-dimensional (2D) electronic spectroscopy signals using line shape functions.

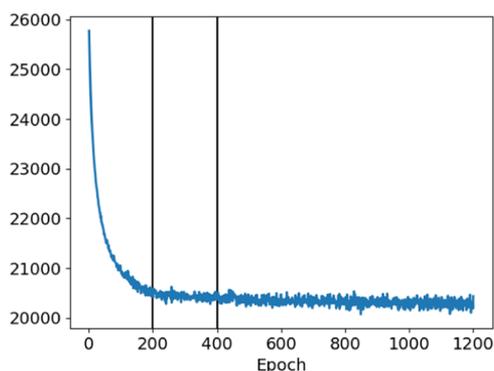
**3.4. Results and Discussion.** Representative examples of the prepared data set are plotted in Figure 2. The abrupt change in the exciton energies in Figure 2b occurs due to the exciton



**Figure 2.** Samples of the data used in learning calculations for (a) the excitation energy  $E_k$  for  $k = \pm$  and (b) the weight functions  $w_{\Delta}$  (green dashed curve),  $w_{k,m}(t)$  for pigments  $m = 1$  (blue), and  $m = 2$  (orange) for  $k = +$  (solid line) and  $k = -$  (dotted line), respectively. (c) Plot of the differences between the energy levels  $E_{\pm}$  to illustrate the relationship between the energies and the weight functions.

transfer between pigments 1 and 2 that takes place in the time period of 10–100 fs. As illustrated in Figure 2c, the difference in the exciton energies  $E_+ - E_-$  exhibits minima with respect to the exciton-transfer processes. As depicted by the red circles in Figure 2c, although such minimal points are significantly narrower and deeper than the minimal point caused by energetic fluctuation, it is difficult to separate the effects of exciton transfer from the energy fluctuation due to environmental motions. By introducing the localization weight functions  $w_{\pm,m}(t)$  and  $w_{\Delta}(t)$  in eqs 21 and 22 to eliminate the effects of the nonenvironmental origin involved in the learning trajectories, we can stabilize and enhance the efficiency of the machine learning process.

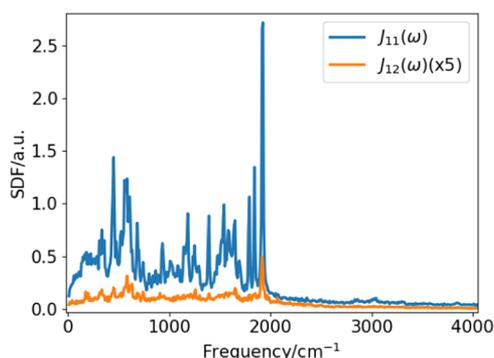
In Figure 3, we depict the learning curve of the loss function, as defined in eq 24. Upon gathering random samplings of  $\phi$ , the



**Figure 3.** Learning curve of the loss function for the indocarbocyanine dimer model. The vertical line at epoch 200 indicates the epoch where the learning rate changed, and the vertical line at epoch 400 indicates early stopping.

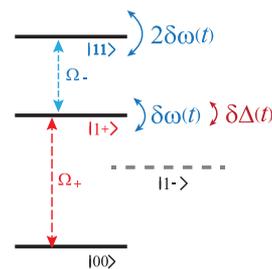
loss function converged monotonically to a certain positive value, which demonstrated the efficiency of the present algorithm. The initial parameter values of the excitation energy and the interaction energy were set as  $\omega_0 = 17\,736\text{ cm}^{-1}$  and  $\Delta = 1004\text{ cm}^{-1}$ , whereas the optimized values of the excitation energy and the interaction energy were given by  $\omega_0 = 17\,794\text{ cm}^{-1}$  and  $\Delta = 963\text{ cm}^{-1}$ , which are closer to the values that fit the experimentally obtained spectra.<sup>52</sup> Here, to avoid overfitting problems, we employed the early-stopping technique (see the Appendix B).

In Figure 4, we present the results of SDFs for the excitation energy  $J_{11}(\omega)$  ( $= J_{22}(\omega)$ ) and the interaction energy  $J_{12}(\omega)$ .



**Figure 4.** SDFs of the indocarbocyanine dimer in the methanol environment for the excitation energy  $J_{11}(\omega)$  ( $= J_{22}(\omega)$ ) (blue) and the interaction energy  $J_{12}(\omega)$  (orange) obtained with the machine learning approach.

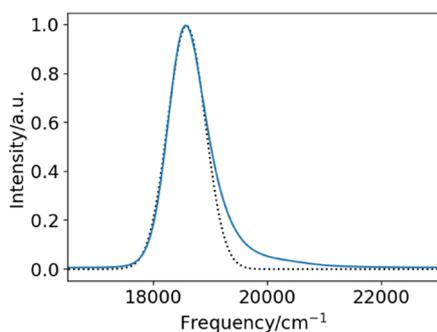
Various intermolecular modes below  $2000\text{ cm}^{-1}$  are observed as prominent sharp peaks near 450, 570, 1185, 1393, 1541, 1791, 1842, and  $1923\text{ cm}^{-1}$ . In the region above  $2000\text{ cm}^{-1}$ , only two tiny peaks are observed at approximately  $3000$  and  $3850\text{ cm}^{-1}$ . The normal mode analysis (B3LYP/def-SV(P)) indicates that these peaks under  $3300\text{ cm}^{-1}$  arise from the intramolecular modes of the indocarbocyanine dimer, whereas the peak at  $3850\text{ cm}^{-1}$  arises from a molecular vibration of the solvent methanol molecules. We found that each sharp peak can be fitted by the Brownian spectral distribution,<sup>24,25</sup> whereas the broadened background peak in the range from 0 to  $2000\text{ cm}^{-1}$  corresponds to the intramolecular modes fitted by the Drude–Lorentz distribution.<sup>63</sup> The intensities of the peaks in  $J_{12}(\omega)$  are considerably weaker than those in  $J_{11}(\omega)$ : only the peaks near 456, 562, 1840, and  $1920\text{ cm}^{-1}$  are identified. As we expected, the intermolecular peak positions are governed by the classical MD simulation, whereas the heights of these peaks are predominately governed by the quantum chemistry calculation (Figure 5).



**Figure 5.** Energy-level diagram for a dimer system that undergoes random fluctuations in the excited energy and coupling strength described by  $\delta\omega(t)$  and  $\delta\Delta(t)$ , respectively. For the description of pure dephasing, only the difference between the energies involved in the optical excitation is important: the frequency fluctuation between  $|00\rangle$  and  $|1+\rangle$  is given by  $\delta\omega(t) + \delta\Delta(t)$ , whereas that between  $|1+\rangle$  and  $|11\rangle$  is given by  $\delta\omega(t) - \delta\Delta(t)$ . For perfectly uncorrelated fluctuations, we consider  $\delta\omega(t)$  and  $\delta\Delta(t)$  independently. The dashed line represents the energy level of the forbidden state  $|1-\rangle$ .

To verify the descriptions of the obtained SDFs and system parameters, we computed the linear absorption and two-dimensional electronic spectra (2DES) for the cases in which the experimentally obtained spectra were available.<sup>52</sup> In general, these spectra should be calculated in the framework of open quantum dynamics that considers the complex interactions between the exciton sites. However, for demonstration purposes, here, we employ the analytical expressions for response functions, ignoring the transitions to the state that are usually forbidden. The details of these calculations are presented in Appendix A.

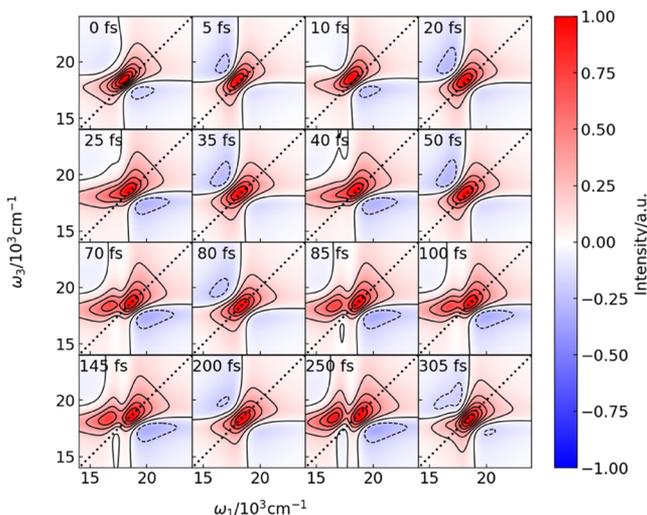
The linear absorption spectrum calculated from eqs A.1 and A.2 is presented in Figure 6. Here, the calculated peak is fitted by the Gaussian function  $\lambda \exp[-((\omega - \omega_c)/\gamma)^2]$ , where the amplitude, central frequency, and width are  $\lambda = 351$ ,  $\omega_c = 18\,583\text{ cm}^{-1}$ , and  $\gamma = 464\text{ cm}^{-1}$ , respectively. Note that we could not determine the absolute SDF intensities because, for simplicity, we did not calculate the amplitude of the dipole operator. Here, we chose to use the intensity of  $J_{11}(\omega)$  to fit the experimentally obtained signal. As shown in Figure 6, we observe a single broadened absorption peak at  $\omega_0 + \Delta$  corresponding to the transition between  $|00\rangle$  and  $|1+\rangle$ , while the transition between  $|00\rangle$  and the state  $|1-\rangle$  is forbidden (see Figure 5). Although the experimentally observed linear absorption spectrum exhibits a



**Figure 6.** Linear absorption spectrum of an indocarbocyanine dimer, as calculated with eqs A.1 and A.2 and the line shape function eq A.3 for the system parameters and SDFs obtained with the machine learning approach. The dotted line shows the fitted Gaussian peak centered at 18 583  $\text{cm}^{-1}$ , indicating that the calculated peak is asymmetric due to the 0–1 phonon transition near 19 500  $\text{cm}^{-1}$ .

0–1 phonon sideband peak near  $\omega = 19\,500\text{ cm}^{-1}$ , here, we observe this phenomenon only as an asymmetry of the Gaussian peak in the high-frequency region.

The 2D correlation electronic spectra calculated using the analytical expressions of the response function (eqs A.5–A.7)<sup>64</sup> are presented in Figure 7. At  $t_2 = 0\text{ fs}$ , only one peak stretched till



**Figure 7.** Two-dimensional electronic spectra of an indocarbocyanine dimer, as calculated with eqs A.5–A.7 and the line shape function eq A.3 for the system parameters and SDFs obtained with the machine learning approach. The waiting time  $t_2$  for each signal is displayed at the top left of each panel. The peak intensity of the signal was normalized for each  $t_2$ . The waiting time  $t_2$  was chosen to illustrate the maximal/minimal points of the oscillating feature of the peak elongation (see text).

near the  $\omega_1 = \omega_3$  line, arising from the  $|00\rangle \rightarrow |1+\rangle$  transition, is observed. At  $t_2 = 10, 25,$  and  $40\text{ fs}$ , the peak is elongated in the low-frequency  $\omega_1$  direction due to a shift in the eigenenergy caused by the heat-bath-induced exciton–exciton interaction described by  $J_{12}(\omega)$ . Because the system–bath interaction we considered here is non-Markovian and its effects appear only after a period longer than the inverse correlation time of noise, we do not observe such heat-bath effects for a small  $t_2$ . Then, at approximately  $t_2 = 70\text{ fs}$ , the off-diagonal peak near  $(\omega_1, \omega_3) = (17\,800, 20\,000)$ , in units of  $\text{cm}^{-1}$ , corresponding to the transition of  $|1+\rangle \rightarrow |11\rangle$  is observed, whereas the peak along the

$\omega_1 = \omega_3$  line shifts to  $(\omega_1, \omega_3) = (21\,000, 20\,000)$  due to the transition of  $|00\rangle \rightarrow |1+\rangle$  that arises from the exciton–exciton interaction described by  $\Delta$  and  $J_{12}(\omega)$ . As  $t_2$  increases, the intensities of these two peaks oscillate as a result of the population transitions among  $|10\rangle, |1+\rangle,$  and  $|11\rangle$  caused by  $\Delta$  and  $J_{12}(\omega)$ . This phenomenon was also observed experimentally.<sup>52</sup> The appearance of this oscillatory feature at a finite period in  $t_2$  indicates the importance of the off-diagonal heat bath, whose modeling is not easy in the framework of the existing approach.

While the off-diagonal peak still exhibits oscillatory motion at  $t_2 \geq 100\text{ fs}$ , the peak profile gradually elongates in the  $\omega_1 = \omega_3$  direction due to the inhomogeneous broadening that arises from the diagonal bath modulation described by  $J_{11}(\omega)$  and  $J_{22}(\omega)$ .<sup>35</sup>

#### 4. CONCLUSIONS

We introduced a machine learning approach for constructing a model that can be used to analyze the dynamics of exciton- or electron-transfer processes in a complex environment on the basis of considering the energy eigenstates evaluated from QM/MM simulations as functions of time. The key feature of the present study is the system–bath model, in which the primary exciton/electron dynamics are described by a system Hamiltonian expressed in terms of discretized energy states, while the other degrees of freedom are described by harmonic heat baths that are characterized by SDFs. An optimized system–bath Hamiltonian obtained from the machine learning approach allows us to conduct time-irreversible quantum simulations that are not feasible with a full quantum MD simulation approach.

Here, we demonstrated the above features by calculating linear and nonlinear optical spectra for the indocarbocyanine dimer system in a methanol environment in which the quantum entanglement between the system and bath plays a central role.<sup>35,36</sup> The calculated results can be used to explain the experimental results reasonably well; we found that the heat bath plays a key role in describing the exciton-transfer process for the exciton–exciton interaction in this system. Although here we ignore the transitions to the state that are usually forbidden due to applicability of the analytical expression, if necessary, we can explicitly consider such transitions using the HEOM formalism.<sup>34–36,65,66</sup>

Finally, we briefly discuss possible extensions of this study. As shown in a previous paper,<sup>51</sup> the machine learning approach can be applied to a system described by reaction coordinates, which is useful for investigating chemical reaction processes characterized by potential energy surfaces. By combining the previous and present approaches, we can further investigate systems described by not only electronic states but also molecular configuration space, for example, photoisomerization,<sup>67</sup> molecular motor,<sup>68</sup> and nonadiabatic transition problems,<sup>69</sup> with frameworks based on the system–bath model. In this way, we may construct a system–bath model for entire photosynthesis reaction processes consisting of photoexcitation,<sup>1,2</sup> exciton-transfer,<sup>3–20</sup> electron-transfer,<sup>21–25</sup> and proton-transfer processes,<sup>70–72</sup> including conversion processes, such as exciton-coupled electron-transfer<sup>73</sup> and electron-coupled proton-transfer processes.<sup>74</sup>

Further theoretical and computational efforts must be initiated that include providing learning data based on accurate and large quantum simulations, improving learning algorithms, and developing an accurate and efficient open quantum dynamics theory to treat a complex system–bath model. We

leave such additional endeavors to future studies in accordance with recent progress in theoretical techniques.

## ■ A: ONE-DIMENSIONAL AND TWO-DIMENSIONAL SPECTRA

Linear and nonlinear optical spectra can be expressed in the Fourier transformation of the response functions.<sup>75</sup> In the present dimer case, we can analytically express the response functions in terms of a line shape function including the contribution from an exciton–exciton interaction.<sup>64</sup>

The linear absorption spectrum is given by<sup>76</sup>

$$S(\omega) = \int_0^\infty dt e^{i\omega t} R^{(1)}(t) - c. c. \quad (\text{A.1})$$

where  $R^{(1)}(t) = \langle [\mu(t), \mu(0)] \rangle$  is the one-dimensional (1D) response function expressed in terms of the transition dipole moment  $\mu(t)$ . For a coupled dimer system, the analytical expression for the response function for eq A.1 is expressed as

$$R^{(1)}(t_1) = \frac{i\mu}{\hbar} \exp[i\Omega_+ t_1 - g_-^{11}(t_1) - g_-^{12}(t_1)] - c. c. \quad (\text{A.2})$$

where the line shape function,  $g_\pm^a(t)$ , for the SDF,  $J_a(\omega)$ , with  $a = 11$  and  $12$  is given by

$$g_\pm^a(t) \equiv \int_0^t dt' \int_0^{t'} dt \int \frac{d\omega}{2\pi} J_a(\omega) \left[ \coth\left(\frac{\beta\hbar\omega}{2}\right) \cos(\omega t) \pm i \sin(\omega t) \right] \quad (\text{A.3})$$

For the coupled dimer system, the third-order response function is

$$R^{(3)}(t_3, t_2, t_1) = \langle [\mu(t_3), [\mu(t_2), [\mu(t_1), \mu(0)]]] \rangle \quad (\text{A.4})$$

and can also be evaluated in the analytical form as<sup>64</sup>

$$R^{(3)}(t_3, t_2, t_1) \equiv \frac{i\mu^4}{\hbar^3} \sum_{\alpha=1}^8 \exp[Q_\alpha(t)] - c. c. \quad (\text{A.5})$$

where

$$\begin{aligned} Q_1(t) &= -i\Omega_+ t_1 - i\Omega_+ t_3 - f_1^1(t_1, t_2, t_3) - f_1^2(t_1, t_2, t_3), \\ Q_2(t) &= i\Omega_+ t_1 - i\Omega_+ t_3 - f_2^1(t_1, t_2, t_3) - f_2^2(t_1, t_2, t_3), \\ Q_3(t) &= i\Omega_+ t_1 - i\Omega_+ t_3 - f_3^1(t_1, t_2, t_3) - f_3^2(t_1, t_2, t_3), \\ Q_4(t) &= -i\Omega_+ t_1 - i\Omega_+ t_3 - f_4^1(t_1, t_2, t_3) - f_4^2(t_1, t_2, t_3), \\ Q_5(t) &= -i\Omega_+ t_1 + i\Omega_- t_3 - f_5^1(t_1, t_2, t_3) - f_5^2(t_1, t_2, t_3), \\ Q_6(t) &= i\Omega_+ t_1 + i\Omega_- t_3 - f_6^1(t_1, t_2, t_3) - f_6^2(t_1, t_2, t_3), \\ Q_7(t) &= i\Omega_+ t_1 + 2i\omega_0 t_2 + i\Omega_- t_3 - f_7^1(t_1, t_2, t_3) \\ &\quad - f_7^2(t_1, t_2, t_3), \\ Q_8(t) &= -i\Omega_+ t_1 - 2i\omega_0 t_2 - i\Omega_+ t_3 - f_8^1(t_1, t_2, t_3) \\ &\quad - f_8^2(t_1, t_2, t_3) \end{aligned} \quad (\text{A.6})$$

with

$$\begin{aligned} f_1^a(t_1, t_2, t_3) &= -g_-^a(t_1) - g_+^a(t_3) - [g_+^a(t_2) - g_+^a(t_{23}) \\ &\quad - g_-^a(t_{12}) + g_-^a(t_{123})], \\ f_2^a(t_1, t_2, t_3) &= -g_+^a(t_1) - g_+^a(t_3) \\ &\quad + [g_-^a(t_2) - g_-^a(t_{23}) - g_+^a(t_{12}) + g_+^a(t_{123})], \\ f_3^a(t_1, t_2, t_3) &= -g_+^a(t_1) - g_-^a(t_3) + [g_+^a(t_2) - g_+^a(t_{23}) \\ &\quad - g_+^a(t_{12}) + g_+^a(t_{123})], \\ f_4^a(t_1, t_2, t_3) &= -g_-^a(t_1) - g_-^a(t_3) - [g_-^a(t_2) - g_-^a(t_{23}) \\ &\quad - g_-^a(t_{12}) + g_-^a(t_{123})], \\ f_5^a(t_1, t_2, t_3) &= -g_-^a(t_1) - g_+^a(t_3) - [g_+^a(t_2) - g_+^a(t_{23}) \\ &\quad - g_-^a(t_{12}) + g_-^a(t_{123})], \\ f_6^a(t_1, t_2, t_3) &= -g_+^a(t_1) - g_+^a(t_3) \\ &\quad + [g_+^a(t_2) - g_-^a(t_{23}) - g_+^a(t_{12}) + g_+^a(t_{123})], \\ f_7^a(t_1, t_2, t_3) &= -g_+^a(t_1) - g_-^a(t_3) + [g_+^a(t_2) - g_+^a(t_{23}) \\ &\quad - g_+^a(t_{12}) + g_+^a(t_{123})], \\ f_8^a(t_1, t_2, t_3) &= -g_-^a(t_1) - g_-^a(t_3) - [g_-^a(t_2) - g_-^a(t_{23}) \\ &\quad - g_-^a(t_{12}) + g_-^a(t_{123})] \end{aligned} \quad (\text{A.7})$$

Here,  $t_{12} \equiv t_1 + t_2$ ,  $t_{23} \equiv t_2 + t_3$ , and  $t_{123} \equiv t_1 + t_2 + t_3$ . Because the fluctuations in the  $|1+\rangle$  and  $|11\rangle$  states are described by  $J_{11}(\omega) + J_{12}(\omega)$  and  $J_{11}(\omega) + J_{22}(\omega) = 2J_{11}(\omega)$ , respectively, the line shape function  $g_\pm^a(t)$  in eq A.7 is now expressed as  $g_\pm^1(t) = g_\pm^{11}(t) + g_\pm^{12}(t)$  and  $g_\pm^2(t) = 2g_\pm^{11}(t)$ . By using third-order diagrams, the pump–probe spectrum and photon echo spectra are, for example, calculated from the  $Q_1(t)$ ,  $Q_4(t)$ , and  $Q_5(t)$  elements and the  $Q_2(t)$ ,  $Q_3(t)$ , and  $Q_6(t)$  elements, respectively (Figure 8).

Although the change in the exciton population can be explored by pump–probe spectroscopy, if we wish to investigate not only population dynamics but also system–bath coherence, two-dimensional electronic correlation spectroscopy is a better choice. This spectrum can be calculated from

$$I^{(\text{corr})}(\omega_3, t_2, \omega_1) = I^{(\text{NR})}(\omega_3, t_2, \omega_1) + I^{(\text{R})}(\omega_3, t_2, \omega_1) \quad (\text{A.8})$$

where the nonrephasing and rephasing parts of the signal are defined by

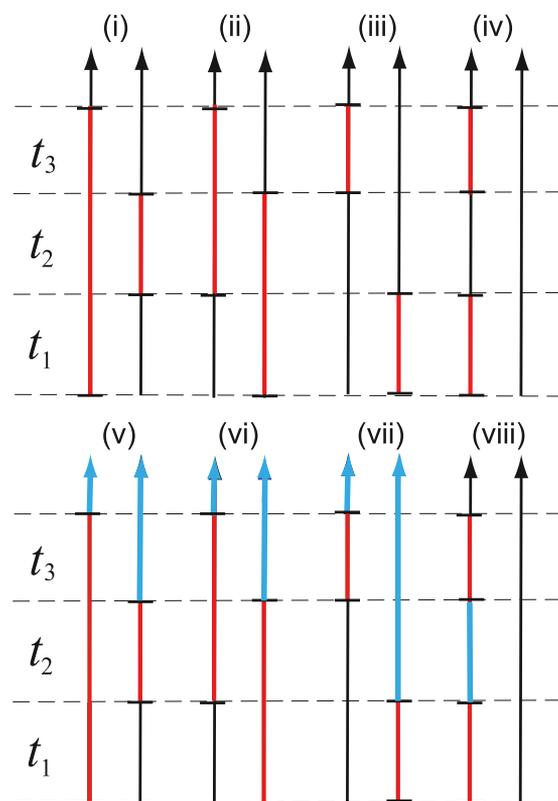
$$I^{(\text{NR})}(\omega_3, t_2, \omega_1) = \text{Im} \int_0^\infty dt_3 \int_0^\infty dt_1 e^{i\omega_3 t_3} e^{i\omega_1 t_1} R^{(3)}(t_3, t_2, t_1) \quad (\text{A.9})$$

and

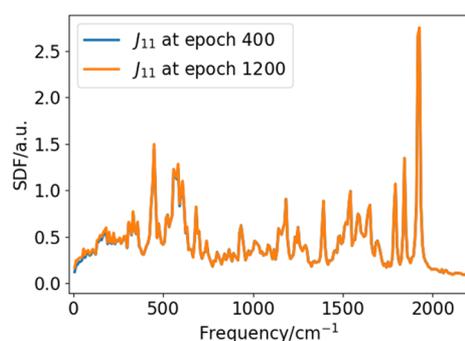
$$I^{(\text{R})}(\omega_3, t_2, \omega_1) = \text{Im} \int_0^\infty dt_3 \int_0^\infty dt_1 e^{i\omega_3 t_3} e^{-i\omega_1 t_1} R^{(3)}(t_3, t_2, t_1) \quad (\text{A.10})$$

## ■ B: OVERFITTING PROBLEM OF MLE

To illustrate the overfitting problem effect, we present the optimized results evaluated in the early-stopping case (epoch 400) and overfitting case (epoch 1200). The learning curve of the loss function for these two cases is presented in Figure 3. As illustrated in Figure 9, the results are similar and overlap everywhere, except the low-frequency region,  $0\text{--}400 \text{ cm}^{-1}$ . The optimized system parameters in the early-stopping case are  $\omega_0 =$



**Figure 8.** Double-sided Feynman diagrams for the third-order response functions  $R^{(3)}(t_3, t_2, t_1)$ . In each diagram, time increases from the bottom to the top and  $t_i$  represents the time intervals for the  $i$ th sequence between successive laser–system interactions. The left line represents the time evolution of the ket, whereas the right line represents that of the bra. The black, red, and blue lines indicate that the system is in the  $|00\rangle$ ,  $|1+\rangle$ , and  $|11\rangle$  states, respectively. For these systems, the complex conjugate paths, which can be obtained by interchanging the ket and bra diagrams, are not shown here.



**Figure 9.** SDF  $J_{11}(\omega)$  ( $= J_{22}(\omega)$ ) obtained in the early-stopping case (blue) and overfitting case (orange).

17 794  $\text{cm}^{-1}$  and  $\Delta = 963 \text{ cm}^{-1}$ , whereas those in the overfitting case are  $\omega_0 = 17 795 \text{ cm}^{-1}$  and  $\Delta = 960 \text{ cm}^{-1}$ . These results indicate that the learning process works very well even in epoch 400. We then found that the accuracy of the obtained SDF, particularly in the region below 400  $\text{cm}^{-1}$ , decreases for the case of epoch 1200, because the fluctuation of the loss function in a larger epoch period suppresses the convergence of the SDF in the low-frequency region, as illustrated in Figure 3. This phenomenon is known as the overfitting (or overtraining) problem of the MLE for the gradient method. We may avoid this problem by regularizing the model, for example, by adopting an

L2 regularization, or by using the Bayesian inference method to account for the physical knowledge as a prior probability.<sup>77</sup> Nevertheless, here, we use an early-stopping method to simply reduce the numerical cost because the low-frequency region  $100 \leq \omega \leq 400 \text{ cm}^{-1}$  is no longer significant in the ultrafast dynamics of the exciton-transfer problem, whereas the region below 10  $\text{cm}^{-1}$  may alter the signal profile significantly due to the quantum thermal factor,  $\coth(\beta\hbar\omega/2)$ , in the line shape function defined as eq A.3.

## AUTHOR INFORMATION

### Corresponding Authors

Seiji Ueno – HPC Systems Inc., Kyoto 600-8412, Japan;  
Email: [se-ueno@hpc.co.jp](mailto:se-ueno@hpc.co.jp)

Yoshitaka Tanimura – Department of Chemistry, Kyoto University, Kyoto 606-8502, Japan; [orcid.org/0000-0002-7913-054X](https://orcid.org/0000-0002-7913-054X); Email: [tanimura.yoshitaka.5w@kyoto-u.jp](mailto:tanimura.yoshitaka.5w@kyoto-u.jp)

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acs.jctc.1c00104>

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors are thankful to Professor Yuki Kurashige for helpful discussions concerning the QM/MM simulations and for providing the indocarbocyanine dimer system. Financial support from HPC Systems Inc. is acknowledged.

## REFERENCES

- (1) May, V.; Kühn, O. *Charge and Energy Transfer Dynamics in Molecular Systems*; Wiley-VCH, 2003.
- (2) Schröter, M.; Ivanov, S.; Schulze, J.; Polyutov, S.; Yan, Y.; Pullerits, T.; Kühn, O. Excitonvibrational coupling in the dynamics and spectroscopy of Frenkel excitons in molecular aggregates. *Phys. Rep.* **2015**, *567*, 1–78.
- (3) Kreisbeck, C.; Kramer, T.; Aspuru-Guzik, A. Scalable High-Performance Algorithm for the Simulation of Exciton Dynamics. Application to the Light-Harvesting Complex II in the Presence of Resonant Vibrational Modes. *J. Chem. Theory Comput.* **2014**, *10*, 4045–4054.
- (4) Lee, M. K.; Coker, D. F. Modeling Electronic-Nuclear Interactions for Excitation Energy Transfer Processes in Light-Harvesting Complexes. *J. Phys. Chem. Lett.* **2016**, *7*, 3171–3178.
- (5) Kreisbeck, C.; Kramer, T. Long-Lived Electronic Coherence in Dissipative Exciton Dynamics of Light-Harvesting Complexes. *J. Phys. Chem. Lett.* **2012**, *3*, 2828–2833.
- (6) Strümpfer, J.; Schulten, K. Light harvesting complex II B850 excitation dynamics. *J. Chem. Phys.* **2009**, *131*, No. 225101.
- (7) Strümpfer, J.; Schulten, K. The effect of correlated bath fluctuations on exciton transfer. *J. Chem. Phys.* **2011**, *134*, No. 095102.
- (8) Strümpfer, J.; Schulten, K. Excited state dynamics in photo-synthetic reaction center and light harvesting complex I. *J. Chem. Phys.* **2012**, *137*, No. 065101.
- (9) Olbrich, C.; Strümpfer, J.; Schulten, K.; Kleinekathöfer, U. Quest for Spatially Correlated Fluctuations in the FMO Light-Harvesting Complex. *J. Phys. Chem. B* **2011**, *115*, 758–764.
- (10) Ishizaki, A.; Fleming, G. R. On the adequacy of the Redfield equation and related approaches to the study of quantum dynamics in electronic energy transfer. *J. Chem. Phys.* **2009**, *130*, No. 234110.
- (11) Raszewski, G.; Saenger, W.; Renger, T. Theory of Optical Spectra of Photosystem II Reaction Centers: Location of the Triplet State and the Identity of the Primary Electron Donor. *Biophys. J.* **2005**, *88*, 986–998.

- (12) Renger, T.; Klinger, A.; Steinecker, F.; Schmidt am Busch, M.; Numata, J.; Müh, F. Normal Mode Analysis of the Spectral Density of the FennaMatthewsOlson Light-Harvesting Protein: How the Protein Dissipates the Excess Energy of Excitons. *J. Phys. Chem. B* **2012**, *116*, 14565–14580.
- (13) Müh, F.; Plöckinger, M.; Ortmyer, H.; Schmidt am Busch, M.; Lindorfer, D.; Adolphs, J.; Renger, T. The quest for energy traps in the CP43 antenna of photosystem II. *J. Photochem. Photobiol., B* **2015**, *152*, 286–300.
- (14) Müh, F.; Plöckinger, M.; Renger, T. Electrostatic Asymmetry in the Reaction Center of Photosystem II. *J. Phys. Chem. Lett.* **2017**, *8*, 850–858.
- (15) Gelzinis, A.; Abramavicius, D.; Ogilvie, J. P.; Valkunas, L. Spectroscopic properties of photosystem II reaction center revisited. *J. Chem. Phys.* **2017**, *147*, No. 115102.
- (16) Fujihashi, Y.; Fleming, G. R.; Ishizaki, A. Impact of environmentally induced fluctuations on quantum mechanically mixed electronic and vibrational pigment states in photosynthetic energy transfer and 2D electronic spectra. *J. Chem. Phys.* **2015**, *142*, No. 212403.
- (17) Novoderezhkin, V. I.; Romero, E.; Dekker, J. P.; van Grondelle, R. Multiple Charge-Separation Pathways in Photosystem II: Modeling of Transient Absorption Kinetics. *ChemPhysChem* **2011**, *12*, 681–688.
- (18) Novoderezhkin, V. I.; Romero, E.; van Grondelle, R. How exciton-vibrational coherences control charge separation in the photosystem II reaction center. *Phys. Chem. Chem. Phys.* **2015**, *17*, 30828–30841.
- (19) Gelzinis, A.; Valkunas, L.; Fuller, F. D.; Ogilvie, J. P.; Mukamel, S.; Abramavicius, D. Tight-binding model of the photosystem II reaction center: application to two-dimensional electronic spectroscopy. *New J. Phys.* **2013**, *15*, No. 075013.
- (20) Adolphs, J.; Renger, T. How Proteins Trigger Excitation Energy Transfer in the FMO Complex of Green Sulfur Bacteria. *Biophys. J.* **2006**, *91*, 2778–2797.
- (21) Garg, A.; Onuchic, J. N.; Ambegaokar, V. Effect of friction on electron transfer in biomolecules. *J. Chem. Phys.* **1985**, *83*, 4491.
- (22) Wolynes, P. G. Dissipation tunneling, and adiabaticity criteria for curve crossing problems in the condensed phase. *J. Chem. Phys.* **1987**, *86*, 1957.
- (23) Yan, Y. J.; Sparpaglione, M.; Mukamel, S. Solvation dynamics in electron-transfer, isomerization, and nonlinear optical processes: a unified Liouville-space theory. *J. Phys. Chem. A* **1988**, *92*, 4842–4853.
- (24) Tanaka, M.; Tanimura, Y. Quantum Dissipative Dynamics of Electron Transfer Reaction System: Nonperturbative Hierarchy Equations Approach. *J. Phys. Soc. Jpn.* **2009**, *78*, No. 073802.
- (25) Tanaka, M.; Tanimura, Y. Multistate electron transfer dynamics in the condensed phase: Exact calculations from the reduced hierarchy equations of motion approach. *J. Chem. Phys.* **2010**, *132*, No. 214502.
- (26) Sim, E.; Makri, N. Path Integral Simulation of Charge Transfer Dynamics in Photosynthetic Reaction Centers. *J. Phys. Chem. B* **1997**, *101*, 5446–5458.
- (27) Sim, E. Determination of the Electron Transfer Mechanism through Decomposition of the Density Matrix. *J. Phys. Chem. B* **2004**, *108*, 19093–19095.
- (28) Dijkstra, A. G.; Tanimura, Y. Correlated fluctuations in the exciton dynamics and spectroscopy of DNA. *New J. Phys.* **2010**, *12*, No. 055005.
- (29) Gélinas, S.; Rao, A.; Kumar, A.; Smith, S. L.; Chin, A. W.; Clark, J. Ultrafast Long-Range Charge Separation in Organic Semiconductor Photovoltaic Diodes. *Science* **2014**, *343*, 512–516.
- (30) Zirzmeier, J.; Lehnher, D.; Coto, P. B.; Chernick, E. T.; Casillas, R.; Basel, B. S.; Thoss, M.; Tykwinski, R. R.; Guldi, D. M. Singlet fission in pentacene dimers. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 5325–5330.
- (31) Tamura, H.; Burghardt, I. Ultrafast Charge Separation in Organic Photovoltaics Enhanced by Charge Delocalization and Vibronically Hot Exciton Dissociation. *J. Am. Chem. Soc.* **2013**, *135*, 16364–16367.
- (32) Huix-Rotllant, M.; Tamura, H.; Burghardt, I. Concurrent Effects of Delocalization and Internal Conversion Tune Charge Separation at Regioregular PolythiopheneFullerene Heterojunctions. *J. Phys. Chem. Lett.* **2015**, *6*, 1702–1708.
- (33) Oviedo-Casado, S.; Urbina, A.; Prior, J. Magnetic field enhancement of organic photovoltaic cells performance. *Sci. Rep.* **2017**, *7*, No. 4297.
- (34) Cainelli, M.; Tanimura, Y. Exciton Transfer in Organic Photovoltaic Cells: A Role of Local and Nonlocal Electron-Phonon Interactions in a Donor Domain. *J. Chem. Phys.* **2021**, *154*, No. 034107.
- (35) Tanimura, Y. Stochastic Liouville Langevin, Fokker-Planck, and Master Equation Approaches to Quantum Dissipative Systems. *J. Phys. Soc. Jpn.* **2006**, *75*, No. 082001.
- (36) Tanimura, Y. Numerically exact approach to open quantum dynamics: The hierarchical equations of motion (HEOM). *J. Chem. Phys.* **2020**, *153*, No. 020901.
- (37) Dral, P. O.; Barbatti, M.; Thiel, W. Nonadiabatic Excited-State Dynamics with Machine Learning. *J. Phys. Chem. Lett.* **2018**, *9*, 5660–5663.
- (38) Hartmann, M. J.; Carleo, G. Neural-Network Approach to Dissipative Quantum Many-Body Dynamics. *Phys. Rev. Lett.* **2019**, *122*, No. 250502.
- (39) Flurin, E.; Martin, L.; Hacoen-Gourgy, S.; Siddiqi, I. Using a Recurrent Neural Network to Reconstruct Quantum Dynamics of a Superconducting Qubit from Physical Observations. *Phys. Rev. X* **2020**, *10*, No. 011006.
- (40) Zheng, F.; Gao, X.; Eisfeld, A. Excitonic Wave Function Reconstruction from Near-Field Spectra Using Machine Learning Techniques. *Phys. Rev. Lett.* **2019**, *123*, No. 163202.
- (41) Rodríguez, M.; Kramer, T. Machine learning of two-dimensional spectroscopic data. *Chem. Phys.* **2019**, *520*, 52–60.
- (42) Namuduri, S.; Titze, M.; Bhansali, S.; Li, H. Machine Learning Enabled Lineshape Analysis in Optical Two-Dimensional Coherent Spectroscopy. *J. Opt. Soc. Am. B* **2020**, *37*, 1587.
- (43) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192–3203.
- (44) Saucedo, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *J. Chem. Phys.* **2019**, *150*, No. 114102.
- (45) Sifain, A. E.; Lubbers, N.; Nebgen, B. T.; Smith, J. S.; Lokhov, A. Y.; Isayev, O.; Roitberg, A. E.; Barros, K.; Tretiak, S. Discovering a Transferable Charge Assignment Model Using Machine Learning. *J. Phys. Chem. Lett.* **2018**, *9*, 4495–4501.
- (46) Chmiela, S.; Tkatchenko, A.; Saucedo, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3*, No. e1603015.
- (47) Chmiela, S.; Saucedo, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, No. 3887.
- (48) Kampen, N. V. *Stochastic Processes in Physics and Chemistry*; Elsevier, 1981.
- (49) Chen, L.; Zhao, Y.; Tanimura, Y. Dynamics of a One-Dimensional Holstein Polaron with the Hierarchical Equations of Motion Approach. *J. Phys. Chem. Lett.* **2015**, *6*, 3110–3115.
- (50) Dunn, I. S.; Tempelaar, R.; Reichman, D. R. Removing instabilities in the hierarchical equations of motion: Exact and approximate projection approaches. *J. Chem. Phys.* **2019**, *150*, No. 184109.
- (51) Ueno, S.; Tanimura, Y. Modeling intermolecular and intramolecular modes of liquid water using multiple heat baths: Machine learning approach. *J. Chem. Theory Comput.* **2020**, *16*, 2099–2108.
- (52) Halpin, A.; Johnson, P. J. M.; Tempelaar, R.; Murphy, R. S.; Knoester, J.; Jansen, T. L. C.; Miller, R. J. D. Two-dimensional spectroscopy of a molecular dimer unveils the effects of vibronic coupling on exciton coherences. *Nat. Chem.* **2014**, *6*, 196–201.
- (53) Tempelaar, R.; Halpin, A.; Johnson, P. J. M.; Cai, J.; Murphy, R. S.; Knoester, J.; Miller, R. J. D.; Jansen, T. L. C. Laser-Limited Signatures of Quantum Coherence. *J. Phys. Chem. A* **2016**, *120*, 3042–3048.

- (54) Duan, H.-G.; Nalbach, P.; Prokhorenko, V. I.; Mukamel, S.; Thorwart, M. On the origin of oscillations in two-dimensional spectra of excitonically-coupled molecular systems. *New J. Phys.* **2015**, *17*, No. 072002.
- (55) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (56) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (57) Bekker, H.; Berendsen, H.; Dijkstra, E.; Achterop, S.; Van Drunen, R.; Van der Spoel, D.; Sijbers, A.; Keegstra, H.; Reitsma, B.; Renardus, M. Gromacs: A parallel computer for molecular dynamics simulations. *Phys. Comput.* **1993**, 252–256.
- (58) Ridley, J.; Zerner, M. An intermediate neglect of differential overlap technique for spectroscopy: Pyrrole and the azines. *Theor. Chim. Acta* **1973**, *32*, 111–134.
- (59) Thompson, M. A.; Zerner, M. C. A theoretical examination of the electronic structure and spectroscopy of the photosynthetic reaction center from *Rhodospseudomonas viridis*. *J. Am. Chem. Soc.* **1991**, *113*, 8210–8215.
- (60) Martin, R. L. Natural transition orbitals. *J. Chem. Phys.* **2003**, *118*, 4775.
- (61) Neese, F. The ORCA program system. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73–78.
- (62) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mane, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viegas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, arXiv:1603.04467. arXiv.org e-Print archive. <https://arxiv.org/abs/1603.04467> (accessed Mar 14, 2016).
- (63) Tanimura, Y. Reduced hierarchy equations of motion approach with Drude plus Brownian spectral distribution: Probing electron transfer processes by means of two-dimensional correlation spectroscopy. *J. Chem. Phys.* **2012**, *137*, No. 22A550.
- (64) Tanimura, Y.; Mukamel, S. Femtosecond pump-probe spectroscopy of intermolecular vibrations in molecular dimers. *J. Chem. Phys.* **1995**, *103*, 1981–1984.
- (65) Tanimura, Y. Reduced hierarchical equations of motion in real and imaginary time: Correlated initial states and thermodynamic quantities. *J. Chem. Phys.* **2014**, *141*, No. 044114.
- (66) Tanimura, Y. Real-time and imaginary-time quantum hierarchical Fokker-Planck equations. *J. Chem. Phys.* **2015**, *142*, No. 144110.
- (67) Ikeda, T.; Tanimura, Y. Probing photoisomerization processes by means of multi-dimensional electronic spectroscopy: The multi-state quantum hierarchical Fokker-Planck equation approach. *J. Chem. Phys.* **2017**, *147*, No. 014102.
- (68) Ikeda, T.; Dijkstra, A.; Tanimura, Y. Modeling and analyzing a photo-driven molecular motor system: Ratchet dynamics and non-linear optical spectra. *J. Chem. Phys.* **2019**, *150*, No. 114103.
- (69) Ikeda, T.; Tanimura, Y. Phase-space wavepacket dynamics of internal conversion via conical intersection: Multi-state quantum Fokker-Planck equation approach. *Chem. Phys.* **2018**, *515*, 203.
- (70) Chen, L.; Shi, Q. Quantum rate dynamics for proton transfer reactions in condensed phase: The exact hierarchical equations of motion approach. *J. Chem. Phys.* **2009**, *130*, No. 134505.
- (71) Shi, Q.; Zhu, L.; Chen, L. Quantum rate dynamics for proton transfer reaction in a model system: Effect of the rate promoting vibrational mode. *J. Chem. Phys.* **2011**, *135*, No. 044505.
- (72) Zhang, J.; Borrelli, R.; Tanimura, Y. Proton tunneling in a two-dimensional potential energy surface with a non-linear system-bath interaction: Thermal suppression of reaction rate. *J. Chem. Phys.* **2020**, *152*, No. 214114.
- (73) Sakamoto, S.; Tanimura, Y. Exciton-Coupled Electron Transfer Process Controlled by Non-Markovian Environments. *J. Phys. Chem. Lett.* **2017**, *8*, 5390–5394.
- (74) Zhang, J.; Borrelli, R.; Tanimura, Y. Probing photoinduced proton coupled electron transfer process by means of two-dimensional electronic-vibrational spectroscopy. *J. Chem. Phys.* **2021**, *154*, No. 144104.
- (75) Mukamel, S. *Principles of Nonlinear Optical Spectroscopy*; Oxford University Press, 1999.
- (76) Tanimura, Y.; Mukamel, S. Real-time path-integral approach to quantum coherence and dephasing in nonadiabatic transitions and nonlinear optical response. *Phys. Rev. E* **1993**, *47*, 118–136.
- (77) Bishop, C. M. *Pattern Recognition and Machine Learning*; Information Science and Statistics; Springer, 2006.