Predicting quantum evolutions of excitation energy transfer in a light-harvesting complex using multi-optimized recurrent neural networks

Authors: Shun-Cai Zhao, Yi-Meng Huang, Zi-Ran Zhao, Shun-Cai Zhao

Date: 2024-01-20T00:00:00+00:00

Abstract

Constructing models to discover the physics underlying massive datasets is a traditional strategy in data mining which has proven to be powerful and successful. In this work, a multi-optimized recurrent neural network (MRNN) is utilized to predict the dynamics of photosynthetic excitation energy transfer (EET) in a light-harvesting complex. The original dataset produced by the master equation was used to train the model to forecast EET evolution. Agreement between our predictions and theoretical deduction is found with an accuracy exceeding 99.26%, showing the validity of the proposed MRNN. A time-segmented polynomial fitting multiplied by a unit step function results in striking consistency with analytical formulations for photosynthetic EET. This work sets a precedent for accurate EET prediction from large datasets by establishing analytical descriptions for the underlying physics, through minimizing the processing cost during the evolution of weak-coupling EET.

Full Text

Preamble

Predicting quantum evolutions of excitation energy transfer in a light-harvesting complex using multi-optimized recurrent neural networks

Shun-Cai Zhao,[1],[2],* Yi-Meng Huang,[1],[2],† and Zi-Ran Zhao[1],[2]
[1]Department of Physics, Faculty of Science, Kunming University of Science and Technology, Kunming, 650500, PR China
[2]Center for Quantum Materials and Computational Condensed Matter Physics, Faculty of Science, Kunming University of Science and Technology, Kunming, 650500, PR China
(Dated: 12:06, Saturday 20th January, 2024)

Abstract

Constructing models to discover the physics underlying large datasets is a traditional and powerful strategy in data mining. In this work, a multi-optimized recurrent neural network (MRNN) is utilized to predict the dynamics of photosynthetic excitation energy transfer (EET) in a light-harvesting complex. The original dataset produced by the master equation is used to train and forecast the EET evolution. We find agreement between our predictions and theoretical deductions with an accuracy exceeding 99.26%, demonstrating the validity of the proposed MRNN. A time-segment polynomial fitting multiplied by a unit step function achieves a loss rate on the order of 10^{-5}, showing striking consistency with analytical formulations for photosynthetic EET. This work establishes a precedent for accurate EET prediction from large datasets by deriving analytical descriptions of the underlying physics, while minimizing processing costs during the evolution of weak-coupling EET.

Keywords: Quantum evolution, excitation energy transfer, multi-optimized recurrent neural networks, quantum dynamics

Introduction

Many physical laws were discovered through data analysis, such as Kepler's three laws derived from Tycho's extensive observations of celestial motion. Understanding the temporal evolution of photosynthetic excitation energy transfer (EET) in light-harvesting complexes is a topic of broad interest due to its nearly 100% photosynthetic conversion efficiency, offering an ideal solution for mitigating the energy crisis [**?**, **?**, **?**]. However, exact numerical simulations of EET dynamics in light-harvesting complexes require enormous computational resources [**?**, **?**], with costs growing exponentially with the number of simulated time steps and system size.

Although many techniques are available—such as the hierarchy of equations of motion (HEOM) [**?**, **?**], multi-configurational time-dependent Hartree (MCTDH) [**?**], stochastic Liouville-von Neumann equation [**?**], quasi-adiabatic propagator path-integral (QUAPI) [**?**], and path-integral Monte Carlo [**?**]—they are inappropriate for examining long-time quantum dynamical evolution.

The current state of a quantum system is largely determined by and rooted in its early evolution stages, enabling us to learn long-time EET evolution in light-harvesting complexes from short-time dynamics without costly direct long-time simulations. Once a memory kernel is acquired, the Nakajima-Zwanzig generalized quantum master equation (GQME) [**?**] provides a broad and formally reliable prescription for achieving this goal [**?**]. Nonetheless, solving the GQME and directly computing the memory kernel for an arbitrary system remains challenging. The transfer tensor method (TTM) can solve the GQME to some extent, but it requires an external numerical methodology to provide a set of dynamical mappings [**?**, **?**, **?**]. The interplay between machine learning and quantum physics has altered this situation [**?**, **?**, **?**] by providing new concepts

for modeling EET evolution in light-harvesting complexes [**?**], enabling intuitive and direct learning from datasets without explicit theoretical construction. Artificial neural networks (ANNs) [**?**] have demonstrated that complex functional dependencies in time series can be learned directly from data [**?**, **?**], avoiding the great efforts required for theoretical analyses that sometimes become unjustified (e.g., the weak coupling limit) or difficult to derive based solely on phenomenological observations. ANNs have been shown capable of solving the master equations governing the dynamics of long-time dissipative open quantum systems [**?**].

Recurrent neural networks (RNNs) in particular have exceptional capacity to interpret intricate temporal behavior. They preserve historical information for future prediction by building a feedback loop that receives both the current stage's input and the previous step's output [**?**]. However, concerns about gradient vanishing or exploding behavior due to multiple iterating operations limit the use of RNNs in long-time scale applications [**?**]. To address this flaw, this work employs multi-optimized recurrent neural networks (MRNNs) rather than long short-term memory recurrent neural networks (LSTM-RNN) [**?**, **?**] to model the long-term dependencies of the time-series dataset, storing key information and predicting future data not currently available. MRNNs are also used as propagators of time-dependent master equations to regulate the light-harvesting complex across a range of time scales.

The rest of the paper is organized as follows. The quantum processes and common master equations for the photosystem II reaction center (PSII-RC) are described in Sec. II.A, and a sample RNN architecture with optimized hyperparameters is introduced in Sec. II.B. Results are discussed in Sec. III, where we validate the learning model in the interval [0, 80] fs (Sec. III.A) and predict the EET evolution process in the range [80, 500] fs (Sec. III.B), using polynomial fitting and analytical expressions (Sec. III.C). Finally, conclusions and outlook for this work are summarized in Sec. IV.

II. THEORETICAL MODEL AND MULTI-OPTIMIZED RECURRENT NEURAL NETWORKS

A. Theoretical model for Photosystem II reaction center (PSII-RC)

A typical photosystem II reaction center (PSII-RC), commonly found in purple bacteria and oxygen-evolving organisms (cyanobacteria, algae, and higher plants), comprises six pigment molecules located at the core of the complex with two symmetric branches of protein matrix [**?**]. Four chlorophylls (special pair PD1, PD2 and accessory ChlD1, ChlD2) and two pheophytins (PheD1, PheD2) are distributed in parallel in these two branches of protein matrix [**?**]. The pair of chlorophylls, PD1 and PD2, located at the center of the PSII-RC act as the primary electron donors, forming two excited states denoted as $|e_1\rangle$ and $|e_2\rangle$ in Fig. 1. Two pheophytin pigments, PheD1 and PheD2, couple to the rest of the molecules and act as the electron acceptors $|\alpha_1\rangle$ and $|\alpha_2\rangle$, respectively [**?**], as

shown in Fig. 1. This is an energy-level framework abstracted from the PSII-RC [**?**], where $|\beta_1\rangle$ ($|\beta_2\rangle$) is a positively charged state after an electron is released and $|g\rangle$ is a ground state.

After absorbing a solar photon, an electron is excited from $|g\rangle$ to $|e_2\rangle$ with a transition rate γ, where the excited electron may transit to the $|e_1\rangle$ state at a rate γ. Then the excited electron is transferred to the acceptors by emitting a phonon via two pathways: $|e_1\rangle$ ($|e_2\rangle$) $\rightarrow |\alpha_1\rangle$ or $|e_1\rangle$ ($|e_2\rangle$) $\rightarrow |\alpha_2\rangle$ at emission rates $\gamma_1 c$ and $\gamma_2 c$, respectively, with $|\alpha_1\rangle$ and $|\alpha_2\rangle$ being the ion-pair states in these two pathways. This process is accompanied by spatial separation of positive and negative charges induced by the release of excited electrons to the plastoquinone molecule, leaving a hole in the dimer. The electrons are then released from the two acceptors PheD1 and PheD2, denoted by $|\alpha_1\rangle \rightarrow |\beta_1\rangle$ (Path 1) and $|\alpha_2\rangle \rightarrow |\beta_2\rangle$ (Path 2) with respective rates Γ (i = 1, 2), providing energies for possible work. Finally, the electron returns to the primary electron donor via $|\beta_1\rangle$ ($|\beta_2\rangle$) $\rightarrow |g\rangle$. As an alternative pathway to the two-step processes $|\alpha_1\rangle$ ($|\alpha_2\rangle$) $\rightarrow |\beta_1\rangle$ ($|\beta_2\rangle$) $\rightarrow |g\rangle$, the acceptor-to-donor charge recombination can also bring the system directly back to the ground state $|g\rangle$ without producing any current, with rates $\delta\Gamma$ (i = 1, 2), where δ is a dimensionless fraction [**?**] describing the radiative recombination rate of the two pathways.

Before making predictions, it is essential to gather a dataset for the learning model to start with. In this work, the data used for machine learning are collected from the initial stage of the PSII-RC mentioned above, which is generated by weak contact of the system with its environment.

The unitary evolution of electron transfer, as usual, can be described by an electronic Hamiltonian:

$$\hat{H}_e = E_g|g\rangle\langle g| + \sum_{i=1,2}(E_{\alpha i}|\alpha_i\rangle\langle\alpha_i| + E_{\beta i}|\beta_i\rangle\langle\beta_i| + |e_i\rangle\langle e_i|) + (|e_1\rangle\langle e_2| + |e_2\rangle\langle e_1|);$$

via a Lindblad-type master equation:

$$\frac{d\hat{\rho}}{dt} = -i[\hat{H}_e, \hat{\rho}] + \mathcal{L}_H\hat{\rho} + \mathcal{L}_{1c}\hat{\rho} + \mathcal{L}_{2c}\hat{\rho} + \mathcal{L}_{3c}\hat{\rho} + \mathcal{L}_{\Gamma}\hat{\rho} + \mathcal{L}_{\delta\Gamma}\hat{\rho};$$

where the strength of the interaction with the environment is comparable with the internal interactions inside the system. The last term in Eq. (1) shows that the dipole-dipole coupling between $|e_1;_2\rangle$ is set as one unit. The Lindblad-type superoperators in Eq. (2) are listed as:

[Figure 1: see original paper]

FIG. 1. (Color online) Energy-level framework for the photosynthetic RC with two load-transitions $|\alpha\rangle \rightarrow |\beta\rangle$ (i = 1, 2). The electronic transitions from ground state $|g\rangle$ to two coupled dipoles $|e_1\rangle$ and $|e_2\rangle$ are induced by the high-temperature photon bath, while the low-temperature phonon bath drives charge

transfer from $|e_2$ to $|\alpha$ (i = 1, 2), and $|\beta$ (i = 1, 2) to |g with termination of electronic circulation.

Here Eq. (3) describes the effect of the high-temperature reservoir with n denoting its average photon number. The low-temperature reservoir has the average phonon number $n_1c = [\exp((E_2 - E_1)/kT) - 1]^{-1}$ in Eq. (4), where γ c $= \gamma$ c (γ c $= \gamma$ c) are the spontaneous decay rates from level $|e_2$ to level $|\alpha$ (i = 1, 2), respectively, and the cross-coupling γ c with γ c $= \gamma$ c describes the effect of Fano interference. Similarly, another low-temperature reservoir is described by Eq. (5) with $n_2c = [\exp((E - E)/kT) - 1]^{-1}$ being the cold reservoir phonon number. Here Γ c $= \Gamma$ c is defined as Γ c $= {}_1\sqrt{(\Gamma c \Gamma c)}$ with $_1$ denoting the quantum interference robustness to describe the Fano interference induced by the spontaneous decay rates, where Γ c $= \Gamma$ c (Γ c $= \Gamma$ c) (i, j = 1, 2). In Eq. (6), n $= [\exp((E_1 - E_2)/kT) - 1]^{-1}$ is the corresponding thermal occupation number of photons at temperature T . The operator $\hat{\sigma}_{\alpha i \alpha i} = |\alpha_i\rangle\langle\alpha_i|$ (i = 1, 2) in Eqs. (7) and (8) is defined as the projection operator. Next, 1 million training datasets within 100 fs are collected from the density matrix element equations (see the Appendix) with parameters [**?**] listed in Table I.

B. Multi-optimized recurrent neural networks (MRNN)

In this work, the dynamic evolution of populations on each energy level in Fig. 1 is utilized to illustrate EET evolution, and the long-term behavior in the proposed PSII-RC will be predicted using a recurrent neural network (RNN) tuned by four hyperparameters. The simple RNN [**?**, **?**] is a type of neural network designed to learn data sequences such as time series, as illustrated in Fig. 2 [Figure 2: see original paper]. To understand their function, it is reasonable to compare them to more traditional feedforward neural networks, in which the input dataset propagates step-by-step through multiple intermediary layers, with training performed by updating the weight matrices and vectors consecutively until the final output layer is reached. In this way, the neural network learns the possible input-output correlations hidden within the data, which can theoretically be used to handle temporal data.

[Figure 2: see original paper]

FIG. 2. The architecture of the simple recurrent neural network (RNN) model.

However, a feedforward neural network is hardly the optimal solution since the number of free parameters grows rapidly with the number of time steps. As an alternative, RNNs address this issue by adopting a cyclic connection design in which the update rule for the hidden layers at time t is decided not only by the current state S but also by the previous state (t - 1), as shown in Fig. 2, where x (i = t - 1, t, t + 1), S , and O represent the input sample data, the memory of the input sample, and the state information stored at time t , respectively. The RNN update rule is given by two activation functions f and g:

$$S_t = f(Ux_t + WS_{t-1});$$

$$O_t = g(VS_t);$$

where the input sample weight U, the recurrent weight W, and the output sample weight V are all time-independent matrices as the dataset propagates forward. The activation function f can be tanh, ReLU, sigmoid, or other functions, while g is usually softmax or similar functions. As shown in Eq. (10), the previous state S $_1$ participates in predicting the current state S . In practice, using the three weight matrices, both the input data and the hidden state at the previous time step can be considered when generating the hidden state at the current time step [**?**, **?**].

Due to these repeating operations, the Markovian assumption of independent data points at multiple time steps becomes problematic in RNN applications. If a machine learning model favors noise and minutiae while ignoring general trends and patterns in the training dataset, it will perform well on training data but poorly on generating new data—a phenomenon termed overfitting. To avoid this, we use a simple RNN with a multi-optimizer to predict the quantum evolutions of EET in light-harvesting complexes. The multi-optimized hyperparameters are discussed below.

1. Learning rate regulator (LRR) In each iteration, a learning rate regulator (LRR) is typically applied to the simple RNN to update model parameters and determine how far down the gradient the parameters move with each update [**?**, **?**]. Commonly used LRRs include exponential attenuation regulators, cosine LRR, preheating regulators, and LR attenuation regulators. A successful LRR should optimize the model while avoiding both overfitting and underfitting. If the LRR is set too high or too low, the model may diverge or converge slowly, requiring more training rounds to obtain the optimal solution.

To simplify the task, we define an LRR at the starting stage of training. During training, the model computes the gradient of the loss function to determine the parameter updating direction, and this procedure is repeated throughout the dynamic process. Model performance is evaluated using both training and verification sets. Because the time series for quantum evolutions of EET are projected to be long, a faster decaying LRR—specifically, an exponentially decaying LRR, LRR = 0.001 × exp(-epoch/50)—is employed for this RNN. Here, epoch is the parameter reflecting the iterations during training.

2. Early stop function In addition to the aforementioned anti-overfitting strategies, an early stop function [**?**] is employed to decrease overfitting. As implied by its name, the early stop technique completes training before the algorithm overfits, obtaining the optimal global outcome and resulting in robust generalization performance, as shown by:

$$E_{\mathrm{opt}}(t) := \min_{t' \leq t} E_{\mathrm{va}}(t');$$

$$\mathrm{GL}(t) = 100 \times \left[\frac{E_{\mathrm{va}}(t)}{E_{\mathrm{opt}}(t)} - 1 \right];$$

where E_{opt}(t) is the ideal verification error set as a function of the number of repetitions t, and GL(t) is the generalization loss evaluating the rate at which the generalization error grows compared to the previous lowest error. When the generalization error is large, an early stop is preferable since it indicates that the model has been overfitted. Such termination of training is judged by a threshold of GL(t). The early stop technique's halting criterion is classified into three types. In this work, the first type of stop rule is employed to determine the loss function and accuracy on the verification set.

3. Regularization Regularization via adding penalty terms to the model's loss function is becoming a popular strategy for reducing overfitting and improving model generalization in machine learning. There are two types of regularization, namely L1 and L2 [**?**, **?**], evaluated by weight parameters w* and w following and before the update, respectively. They are given by:

$$\mathrm{L1} = \arg\min_{w} \{\mathrm{MSE}(y, \hat{y}, w) + \lambda \sum_{i} |w_i|\};$$

$$\mathrm{L2} = \arg\min_{w} \{\mathrm{MSE}(y, \hat{y}, w) + \lambda \sum_{i} w_i^2\};$$

where the sum of absolute values and the square of parameters have been added to the loss function, respectively [**?**]. The first terms on the right side represent the original loss functions, the second terms represent the regularization parameters λ, and the summation is carried out over numerous model ownership parameters. L1 regularization in Eq. (14) creates sparse solutions and is more suitable for feature selection, whereas L2 regularization in Eq. (15) tends to produce many small but non-zero weights. In this study, L2 regularization is used to prevent overfitting in the long-term training dataset.

4. Dropout In this work, the Dropout regularization technique [**?**, **?**, **?**] is also applied to the simple RNN. As training of the input data proceeds forward through the neural network, the estimated loss propagates backward, and parameters are updated using the gradient descent approach, as shown in Fig. 3(a). In Fig. 3(b), the Dropout parameter P is introduced to deactivate certain neurons with a specific probability during forward propagation, and parameters are updated using the gradient descent approach.

[Figure 3: see original paper]

FIG. 3. The structural diagram of Recurrent Neural Network (a) without and (b) with Dropout regularization techniques.

III. RESULTS AND DISCUSSIONS

A. Training the multi-optimized recurrent neural network (MRNN)

After progressively incorporating several hyperparameters—such as early stop function, L2 regularization, LRR, Dropout, and Bayesian optimizer—a multi-optimized RNN model is constructed. As a distinguishing feature, this model utilizes the Lindblad-type master equation Eq. (2) to gather data from the proposed PSII-RC, with parameters listed in Table I. Using Eq. (2), we produce a dataset of 1 million data points in 100 fs, divided into training and test sets at a 4:1 ratio. The first 800,000 data points serve as the training set, while the remaining 200,000 act as the test set, using the hyperparameters listed in Table II.

[Figure 4: see original paper]

FIG. 4. (Color online) (a) Evolutions of the training population at all excited states within the interval [0, 80] fs. (b) A comparison of learning quality between our proposed learning model and the collected testing set over the range [80, 100] fs. The hyperparameters listed in Table II and the original codes in SM1, SM2, and SM3 are used for the calculation.

Figure 4(a) displays the evolution of the training set fed into this proposed multi-optimized RNN learning model (original codes in SM1) during the interval [0, 80] fs. To assess the validity of the proposed learning model, we predict the evolution of EET (solid curves) over a time range from 80 to 100 fs and compare it with the test set (dotted curves) collected from the aforementioned PSII-RC within the same range, indicating good agreement between them, as illustrated in Fig. 4(b). This ensures the high accuracy of this learning model composed of the simple RNN and some optimizers (original codes in SM1). Supported by these results, the proposed multi-optimized RNN learning model is expected to anticipate EET evolution from 80 to 500 fs.

TABLE II. Hyperparameters used in the RNNs training.

Parameter	Value
Learning rate	0.01
L2 regularization	0.0000001
Dropout	0.01
Epochs	1000

B. EET Predicted by the MRNN

In Fig. 5 [Figure 5: see original paper], the vertical dashed black line at the 80th fs shows the temporal starting point of our prediction achieved by gradually incorporating hyperparameters into the calculation. The cutoff at 80 fs is used to evaluate prediction accuracy by checking how precisely the generated

data coincide with the training dataset. After examining Figs. 5(a)-(e), the evolutionary history of each population indicates that our predictions vary against hyperparameters in subsequent periods. L2 regularization, on the other hand, has little effect on EET prediction, as shown by the nearly identical curves in Figs. 5(a)-(e). When all hyperparameters are added to the simple RNN, there is notable consistency between training and prediction values at 80 fs, corresponding to the red curves in Figs. 5(a)-(e) (original codes in SM2). The network architecture parameters were employed to forecast |b , and the inset in Fig. 5(e) clearly shows the role each hyperparameter plays in predicting the evolution of _{bb}.

Consider the prediction of EET on the excited state $|e_1$ in Fig. 5(a), where hyperparameters and optimizers are added one by one within the RNN architecture. Three layers with 128, 64, and 32 neurons, respectively, constitute the essential neural network architecture. In the absence of optimizers and hyperparameters, the simple RNN timing prediction model fails to generate substantial physical results, as the prediction (black dotted solid line) is not consistent with the training curve at 80 fs in Fig. 5(a). The time forecast equipped with the subsequently added early stop function does not resolve this inconsistency, as seen by the sky blue dotted solid curve overlapping that for the simple RNN. The overlapping purple and green curves in Fig. 5(a) indicate that EET evolution is not sensitive to L2 regularization on the excited state $|e_1$, and almost identical evolutionary properties over [80, 500] fs are exhibited in the insets of Fig. 5(a). The evolution curve of $|e_1$ tends to level off when the Dropout parameter P = 0.32499999999999996 is applied for further optimization, as demonstrated by the red curve.

In Fig. 5(c), docking with the red curve achieves better training at 80 fs with the Dropout parameter set as P = 0.11747474747474748 and the initial LR = 0.001. As it takes time to manually adjust the hyperparameters of the neural network when dynamically optimizing the prediction of population on $|\alpha_2$, the Bayesian optimizer is implemented into the neural network to find the ideal number of layers and neurons. Finally, a neural network composed of 6.23223034545932 layers with 124.19253861421566 neurons per layer was employed, corresponding to the red curve shown in Fig. 5(d). The same neural network architecture parameters were used to forecast |b , and the inset in Fig. 5(e) clearly shows the roles each hyperparameter plays in predicting the evolution of _{bb}.

[Figure 5: see original paper]

FIG. 5. (Color online) Prediction of the evolutive EET with different combinations of hyperparameters shown in (a) to (e) via the multi-optimized RNN in the interval [80, 500] fs, and the predictive and theoretical (within [0, 500] fs) EET evolutions are demonstrated in (f) with all identical parameters to those in Fig. 4 included in the calculation.

To validate the predictive accuracy of the proposed MRNN, the evolution behaviors of EET are theoretically simulated using Eq. (2) in the temporal range

of [0, 500] fs, as shown by the dotted lines in Fig. 5(f), while the optimal prediction results are replotted as dotted solid curves in the time range 80 to 500 fs, as shown by the red curves in Figs. 5(a)-(e) for comparison (original codes in SM1). The theoretical calculation and the curves predicted by MRNN for photosynthetic EET perfectly coincide, indicating the validity of our proposed MRNN.

To quantitatively evaluate the prediction precision of EET for the system, an accuracy rate is defined as the normalized population sum during [80, 500] fs. Although the prediction precision decreases over time for this photosynthetic system, it reaches a minimum of 0.9926 at 500 fs, as shown by an arrow in Fig. 6 [Figure 6: see original paper] (original codes in SM1). Nevertheless, this accuracy outperforms most RNN learning models, such as those reported in Refs. [**?**, **?**]. At the same time, because our MRNN learning model is derived from the simple RNN, long-term prediction inevitably reveals the inherent flaw of short-term prediction, resulting in a progressive drop in accuracy, as seen in the decreasing trend of Fig. 6.

[Figure 6: see original paper]

FIG. 6. (Color online) Prediction accuracy defined by the normalized population sum in the interval [80, 500] fs.

C. Polynomial fitting and analytical expression for the predictive EET

Using a polynomial to fit the predicted findings is a useful mathematical method for gaining analytical knowledge of EET. Such fitting typically relies on the least squares method and minimizing the error sum of squares [**?**, **?**]. In Python, the Polynomial Features function is a utility in the scikit-learn library that generates polynomial features to form a new feature matrix. The feature matrix is then returned for further model training and fitting. When a polynomial of order 2 is provided for a one-dimensional feature A, Polynomial Features generates a new feature matrix containing A to the first and second powers. If the original feature has many dimensions, the resulting feature matrix will have power combinations for each of them.

Figure 7 [Figure 7: see original paper] displays the differences between the predicted curves and the fitting curves using Polynomial Features in the scikit-learn library. A maximum degree of order 5 for all polynomial fittings is adopted in Fig. 7 (original codes in SM3). The analytical polynomial fittings corresponding to Figs. 7(a)-(e) are found to be:

$$\rho_{e1e1}(t) = -0.0022184223655202433t+1.3936336816327624\times10^{-5}t^2-4.3658376701560287\times10^{-8}t^3+6.6945069$$

$$\rho_{e2e2}(t) = 0.0055891350945578t-3.50638450439065\times10^{-5}t^2+1.091124267760133\times10^{-7}t^3-1.6659867620310589$$

$$\rho_{\alpha 1\alpha 1}(t) = 4.317228690621229\times10^{-10}t-3.0372753727222955\times10^{-12}t^2+1.0210132754732106\times10^{-14}t^3-1.64722$$

$$\rho_{\alpha 2\alpha 2}(t) = -1.8447722039325343\times10^{-6}t+1.0562606143359824\times10^{-8}t^2-3.040707800147938\times10^{-11}t^3+4.35655$$

$$\rho_{bb}(t) = -0.003319998227995093t+2.0476938571365984\times10^{-5}t^2-6.313410066086038\times10^{-8}t^3+9.606627816067$$

where $_{1\ 1}$, $_{2\ 2}$, $_{2\ 2}$, and $\_\{bb\}$ agree very well with the predicted curves, as shown in Fig. 7. In contrast, the fitting result for $_{1\ 1}$ is rather poor, as shown in Fig. 7(c), where overfitting is visible in its dynamic population progress, indicating that an alternative fitting technique should be found.

Because the characteristic features of an RNN lie in its internal (hidden) loop memory, it is understandable that a dynamic state contains all information about previous inputs as it evolves through the data sequence, as shown in Fig. 2. If the polynomial order is increased, the polynomial features will over-adapt to past information but perform poorly on fresh data. This enlarges both the amplitude and rate of change, leading to overfitting. An effective solution is to derive information based not just on the previous step, but on all previous inputs, by implementing a time-segment strategy.

Here, polynomial fittings in different time ranges are carried out and a unit step function is incorporated into the final polynomial expression in each time segment:

$$f_1(t) = \begin{cases} 1 & t < 250 \\ 0 & t \geq 250 \end{cases}; \quad f_2(t) = \begin{cases} 0 & t < 250 \\ 1 & t \geq 250 \end{cases}$$

The fitting time is divided into two intervals: [80, 250] fs and [250, 500] fs, and the polynomial fitting in each interval is multiplied by the unit step function given by Eq. (21). Their final fitting is the sum of the two polynomial fittings:

$$\rho_{\alpha 1\alpha 1}(t_{80-500}) = \rho_{\alpha 1\alpha 1}(t_{80-250}) \times f_1(t) + \rho_{\alpha 1\alpha 1}(t_{250-500}) \times f_2(t).$$

The curves in Fig. 8 [Figure 8: see original paper] show the fitting results for the population on state $|\alpha_1$ using the aforementioned procedure (original codes in SM3). The two insets in Fig. 8 compare polynomial fitting and predictions in the intervals [80, 250] fs and [250, 500] fs, respectively. The overlap between the dotted lines (red) and dash-dotted lines (blue) demonstrates the high precision of polynomial fitting at various intervals. Furthermore, the analytical functions derived from the time-segment polynomial fittings are given by:

$$\rho_{\alpha 1\alpha 1}(t_{80-250}) = 1.7388382452293818\times10^{-9}t - 1.8538908793974676\times10^{-11}t^2 + 9.883817236997048\times10^{-14}t^3 - 2.6$$

$$\rho_{\alpha 1\alpha 1}(t_{250-500}) = 2.500773638209706\times10^{-12}t - 1.2744394675654058\times10^{-14}t^2 + 3.227678944102217\times10^{-17}t^3 - 4.$$

The comparison between the whole-period fitting and evolution prediction in [80, 500] fs for the dynamic populations on state $|\alpha_1$ demonstrates that the time-segment polynomial fitting can effectively overcome the overfitting seen in Fig. 7(c).

Figure 9 [Figure 9: see original paper] depicts the total fitting loss rate compared to the predicted results, a physical quantity assessing the precision of the fitting technique employed in this work. Even though a jumpy loss rate can be seen at both the initial and final stages of the time interval, the loss rate exhibits oscillating behavior on the order of 10^{-5}, ensuring high accuracy for this MRNN when utilized in this work.

Thus far, the above fitting findings have accurately described EET evolution, revealing the physical rules behind the data, which may be utilized to purposely control EET and construct artificial photosynthetic devices in the future.

[Figure 7: see original paper]

FIG. 7. (Color online) Polynomial fittings (red dotted curves) in comparison with EET evolution predictions (blue dash-dotted curves) within the interval of [80, 500] fs.

[Figure 8: see original paper]

FIG. 8. (Color online) A comparison between the fittings by the time-segment polynomial multiplied by unit step function (red dotted curves) and the evolution predictions (blue dash-dotted curves) in [80, 500] fs. Time-segment polynomial fittings in comparison with the evolution predictions shown by the insets (a1) in [80, 250] fs and (a2) in [250, 500] fs.

[Figure 9: see original paper]

FIG. 9. (Color online) Total loss rate of the polynomial fitting versus the prediction within the range of [80, 500] fs.

IV. CONCLUSION AND OUTLOOK

In summary, fed by the original PSII-RC dataset, an MRNN strategy is proposed to forecast EET evolution with an accuracy exceeding 99.26% within 500 fs when compared to theoretical deduction. Polynomial fitting is also implemented for EET evolutions to obtain analytical results. The predicted EET evolutions are further subjected to time-segment polynomial fitting multiplied by a unit

step function, and the analytical formulations show high precision with a loss rate on the order of 10^{-5}, demonstrating closeness to physical law. The results reveal that the proposed MRNN is a valid and powerful data mining tool for forecasting EET evolution in a light-harvesting complex. A comparison with experimental data in the future is expected to further assess the validity of this learning model.

V. ACKNOWLEDGMENTS

S. C. Zhao is grateful for significant discussions and language polishing from Prof. Chen-Xu Wu of Xiamen University, and for funding from the National Natural Science Foundation of China (grants 62065009 and 61565008) and the Foundation for Personnel Training Projects of Yunnan Province, China (grant 2016FB009).

VI. APPENDIX

The density matrix dynamic element equations are given by:

$$\dot{\rho}_{e1e1} = -\gamma_e[(n_e+1)\rho_{e1e1} - n_e\rho_{e2e2}] - \gamma_h[(n_h+1)\rho_{e1e1} - n_h\rho_{gg}];$$

$$\dot{\rho}_{e2e2} = \gamma_e[(n_e{+}1)\rho_{e1e1}{-}n_e\rho_{e2e2}]{-}\gamma_{1c}[(n_{1c}{+}1)\rho_{e2e2}{-}n_{1c}\rho_{\alpha1\alpha1}]{-}\gamma_{2c}[(n_{1c}{+}1)\rho_{e2e2}{-}n_{1c}\rho_{\alpha2\alpha2}]{+}2\gamma_{12c}n_{1c}\mathrm{Re}[\rho_{\alpha1\alpha2}$$

$$\dot{\rho}_{\alpha1\alpha1} = \gamma_{1c}[(n_{1c}+1)\rho_{e2e2} - n_{1c}\rho_{\alpha1\alpha1}] - \gamma_{12c}n_{1c}\mathrm{Re}[\rho_{\alpha1\alpha2}] - (1+\delta)\Gamma_1\rho_{\alpha1\alpha1};$$

$$\dot{\rho}_{\alpha2\alpha2} = \gamma_{1c}[(n_{1c}+1)\rho_{e2e2} - n_{1c}\rho_{\alpha2\alpha2}] - \gamma_{12c}n_{1c}\mathrm{Re}[\rho_{\alpha1\alpha2}] - (1+\delta)\Gamma_2\rho_{\alpha2\alpha2};$$

$$\dot{\rho}_{\alpha1\alpha2} = -i\Delta_1\rho_{\alpha1\alpha2}{+}(\gamma_{1c}{+}\gamma_{2c})n_{1c}\rho_{\alpha1\alpha2}{+}\gamma_{12c}[2(n_{1c}{+}1)\rho_{e2e2}{-}n_{1c}\rho_{\alpha2\alpha2}{-}n_{1c}\rho_{\alpha1\alpha1}]{-}\Gamma_{1c}[(n_{2c}{+}1)\rho_{\beta1\beta1}{-}n_{2c}\rho_{gg}]$$

$$\dot{\rho}_{\beta1\beta1} = \Gamma_1\rho_{\alpha1\alpha1} - \Gamma_{1c}[(n_{2c}+1)\rho_{\beta1\beta1} - n_{2c}\rho_{gg}] - \Gamma_{12c}(n_{2c}+1)\mathrm{Re}[\rho_{\beta1\beta2}];$$

$$\dot{\rho}_{\beta2\beta2} = \Gamma_2\rho_{\alpha2\alpha2} - \Gamma_{2c}[(n_{2c}+1)\rho_{\beta2\beta2} - n_{2c}\rho_{gg}] - \Gamma_{12c}(n_{2c}+1)\mathrm{Re}[\rho_{\beta1\beta2}];$$

$$\dot{\rho}_{\beta 1\beta 2} = -i\Delta_2\rho_{\beta 1\beta 2}+(\Gamma_{1c}+\Gamma_{2c})(n_{2c}+1)\rho_{\beta 1\beta 2}-\Gamma_{12c}[(n_{2c}+1)\rho_{\beta 1\beta 1}+(n_{2c}+1)\rho_{\beta 2\beta 2}-2n_{2c}\rho_{gg}];$$

$$\rho_{gg} = 1-\rho_{e1e1}-\rho_{e2e2}-\rho_{\alpha 1\alpha 1}-\rho_{\alpha 2\alpha 2}-\rho_{\beta 1\beta 1}-\rho_{\beta 2\beta 2};$$

where Δ_1 = E_α_1 - E_α_2 and Δ_2 = E_β_1 - E_β_2 are the energy splittings of the states $|\alpha_1\rangle$ ($|\alpha_2\rangle$) and $|\beta_1\rangle$ ($|\beta_2\rangle$).

We utilize these equations to simulate EET dynamics in PSII-RC.

DATA AVAILABILITY STATEMENT

This manuscript has associated data in a data repository. [Authors' comment: All data included in this manuscript are available upon reasonable request by contacting the corresponding author]. The Supporting Information is available free of charge at: Supplements to MRNN.

References

[1] F. H. Alharbi and S. Kais. Theoretical limits of photovoltaics efficiency and possible improvements by intuitive approaches learned from photosynthesis and quantum coherence. Renewable & Sustainable Energy Reviews, 43:1073, 2015.

[2] L. F. Li and S. C. Zhao. Influence of the coupled-dipoles on photosynthetic performance in a photosynthetic quantum heat engine. Chin. Phys. B, 30(4):044215, 2021.

[3] L. F. Li, S. C. Zhao, and L. X. Xu. Charge-transport enhanced by the quantum entanglement in the photosystem II reaction center. Eur. Phys. J. Plus, 136, 10 2021.

[4] A. Kimura and Y. Fujihashi. Quantitative correction of the rate constant in the improved variational master equation for excitation energy transfer. J Chem. Phys., 141(19), 2014.

[5] P. F. Huo and T. F. Miller III. Electronic coherence and the kinetics of inter-complex energy transfer in light-harvesting systems. Phys. Chem. Chem. Phys., 17(46):30914, 2015.

[6] J. Strümpfer and K. Schulten. Open quantum dynamics calculations with the hierarchy equations of motion on parallel computers. J Chem. Theor. and Comput., 8:2808, 2012.

[7] Y. Tanimura. Numerically exact approach to open quantum dynamics: The hierarchical equations of motion (HEOM). J. Chem. Phys., 153:020901, 7 2020.

[8] J. T. Stockburger and H. Grabert. Exact c-number representation of non-Markovian quantum dissipation. page 170407.

[9] H. J. Meyer, U. Manthe, and L. S. Cederbaum. The multi-configurational time-dependent Hartree approach. Chem. Phys. Lett., 165:73, 1990.

[10] N. Makri. Quantum dissipative dynamics: A numerically exact methodology. J. Phys. Chem. A, 102:4414, 05 1998.

[11] D. Kast and J. Ankerhold. Persistence of coherent quantum dynamics at strong dissipation. Phys. Rev. Lett., 110:010402, 2013.

[12] S. Nakajima. On quantum theory of transport phenomena steady diffusion. Progress of Theoretical Physics, (20):948–959, 1958.

[13] A. Kelly and T. E. Markland. Efficient and accurate surface hopping for long time nonadiabatic quantum dynamics. J Chem. Phys., 139:014104, 2013.

[14] J. Cerrillo and J. S. Cao. Non-Markovian dynamical maps: Numerical processing of open quantum trajectories. Phys. Rev. Lett., 112:110401, 2014.

[15] A. A. Kananenka, C. Y. Hsieh, J. S. Cao, and E. Geva. Accurate long-time mixed quantum-classical Liouville dynamics via the transfer tensor method. J. Phys. Chem. Lett., 7:4809, 2016.

[16] Y. Q. Chen, K. L. Ma, Y. C. Zheng, J. Allcock, S. Y. Zhang, and C. Y. Hsieh. Non-Markovian noise characterization with the transfer tensor method. Phys. Rev. Appl., 13:034045, 2020.

[17] K. Naicker, I. Sinayskiy, and F. Petruccione. Machine learning for excitation energy transfer dynamics. Phys. Rev. Research, 4(3):033175, 2022.

[18] Y. Han, I. Ali, Z. Wang, J. Cai, and J. Li. Machine learning accelerates quantum mechanics predictions of molecular crystals. Physics Reports, 934:1–71, 2021.

[19] K. N. Lin, J. W. Peng, F. L. Gu, and Z. G. Lan. Automatic evolution of machine-learning-based quantum dynamics with uncertainty analysis. J. Chem. Theor. and Comput., 18(10):5837, 2022.

[20] A. Ullah and P. O. Dral. Predicting the future of excitation energy transfer in light-harvesting complex with artificial intelligence-based quantum dynamics. Nat. Commun., 13(1930), 2022.

[21] L. Banchi, E. Grant, A. Rocchetto, and S. Severini. Modelling non-Markovian quantum processes with recurrent neural networks. New J. Phys., 20(12), 2018.

[22] S. T. Tsai, E. J. Kuo, and P. Tiwary. Learning molecular dynamics with simple language model built upon long short-term memory neural network. Nat. Comm., 11:5115, 2020.

[23] L. E. H. Rodriguez and A. A. Kananenka. Convolutional neural networks for long-time dissipative quantum dynamics. J. Phys. Chem. Lett., 12(2476), 2021.

[24] M. J. Hartmann and G. Carleo. Neural-network approach to dissipative quantum many-body dynamics. Phys. Rev. Lett., 122:250502, 2019.

[25] D. Lemm, G. F. Von Rudorff, and O. A. Von Lilienfeld. Machine learning based energy-free structure predictions of molecules, transition states, and solids. Nat. Commun., 12:4468, 2021.

[26] S. Nangia, A. W. Jasper, T. F. Miller III, and D. G. Truhlar. Army ants algorithm for rare event sampling of delocalized nonadiabatic transitions by trajectory surface hopping and the estimation of sampling errors by the bootstrap method (vol 120, 3586, 2004). J. Chem. Phys., 120(8):3586, 2004.

[27] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural Comput., 9:1735, 1997.

[28] K. N. Lin, J. W. Peng, F. L. Gu, and Z. G. Lan. Simulation of open quantum dynamics with bootstrap-based long short-term memory recurrent neural network. J. Phys. Chem. Lett., 12(41):10225, 2021.

[29] V. I. Novoderezhkin, J. P. Dekker, and R. Van Grondelle. Mixing of exciton and charge-transfer states in photosystem II reaction centers: Modeling of stark spectra with modified Redfield theory. Biophysical Journal, 93(4):1293, 2007.

[30] R. V. Elisabet, I. H. M. Stokkum, V. Novoderezhkin, J. P. Dekker, and R. V. Grondelle. Two different charge separation pathways in photosystem II. Biochemistry, 49(20):4300, 2010.

[31] V. I. Novoderezhkin, E. Romero, J. P. Dekker, and R. V. Grondelle. Multiple charge-separation pathways in photosystem II: modeling of transient absorption kinetics. Chemphyschem A European Journal of Chemical Physics and Physical Chemistry, 12(3):681, 2015.

[32] Y. Qu, M. Zhou, E. K., N. F. Yu, and Z. F. Yu. Resonance for analog recurrent neural network. ACS Photonics, 9:1647, 04 2022.

[33] M. Schuster and K. K. Paliwal. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11):2673, 1997.

[34] J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, and D. Brunner. Reinforcement learning in a large scale photonic recurrent neural network. Optica, 5(6):756, 2018.

[35] S. H. Ning and K. N. Du. Research on intelligent fault diagnosis of rolling bearing based on adaptive resource allocation deep neural network. IEEE Access, 10:62920, 2022.

[36] Y. Zhang and L. L. Shen. Automatic learning rate adaption for memristive deep learning systems. IEEE Transactions on Neural Networks and Learning Systems, (2):1–12, 2023.

[37] Á. L. Muñoz Castañeda, N. DeCastro-García, and D. Escudero García. Rhoaso: An early stop hyper-parameter optimization algorithm. Mathematics,

9(18):2334, 2021.

[38] D. Krueger, T. Maharaj, J. Kramár, M. Pezeshki, N. Ballas, N. R. Ke, A. Goyal, Y. Bengio, A. Courville, and C. Pal. Zoneout: Regularizing rnns by randomly preserving hidden activations. arXiv preprint arXiv:1606.01305, 2016.

[39] S. Merity, B. McCann, and R. Socher. Revisiting activation regularization for language rnns. arXiv preprint arXiv:1708.01009, 2017.

[40] A. Gajbhiye, S. Jaf, N. A. Moubayed, A. S. McGough, and S. Bradley. An exploration of dropout with rnns for natural language inference. In Artificial Neural Networks and Machine Learning—ICANN 2018: 27th International Conference on Artificial Neural Networks, (Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27), page 157. Springer, 2018.

[41] A. Sarma, S. Singh, H. Jiang, R. Zhang, M. T. Kandemir, and C. Das. Structured in space, randomized in time: Leveraging dropout in rnns for efficient training. Advances in Neural Information Processing Systems, 34:24545, 2021.

[42] I. Salehin and D. K. Kang. A review on dropout regularization approaches for deep neural networks within the scholarly domain. Electronics, 12(14):3106, 2023.

[43] B. Ada and M. Florescu. Modelling non-Markovian dynamics in photonic crystals with recurrent neural networks. Optical Mater. Express, 11(7).

[44] Y. Li, Z. Wang, R. Han, S. Shi, J. Li, R. Shang, H. Zheng, G. Zhong, and Y. Gu. Quantum recurrent neural networks for sequential learning. Neural Networks, 166:148–161, 2023.

[45] J. S. Pei, J. P. Wright, and A. W. Smyth. Mapping polynomial fitting into feedforward neural networks for modeling nonlinear dynamic systems and beyond. Computer Methods in Applied Mechanics and Engineering, 194(42):4481, 2005.

[46] B. A. Escobedo-Trujillo, D. Colorado, W. Rivera, and F. A. Alaffita-Hernández. Neural network and polynomial model to improve the coefficient of performance prediction for solar intermittent refrigeration system. Solar energy, 129:28, 2016.

Note: Figure translations are in progress. See original paper for figures.