The Journal of Chemical Physics

Numerical assessment for accuracy and GPU acceleration of TD-DMRG time evolution schemes

Cite as: J. Chem. Phys. **152**, 024127 (2020); https://doi.org/10.1063/1.5135363
Submitted: 06 November 2019 . Accepted: 26 December 2019 . Published Online: 14 January 2020

Weitang Li, Jiajun Ren, and Zhigang Shuai

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Numerical assessment for accuracy and GPU acceleration of TD-DMRG time evolution schemes

Cite as: J. Chem. Phys. 152, 024127 (2020); doi: 10.1063/1.5135363
Submitted: 6 November 2019 • Accepted: 26 December 2019 •
Published Online: 14 January 2020

Weitang Li, Jiajun Ren,[a] and Zhigang Shuai

AFFILIATIONS

MOE Key Laboratory of Organic Optoelectronics and Molecular Engineering, Department of Chemistry, Tsinghua University, Beijing 100084, People's Republic of China

[a]Electronic mail: renjj@mail.tsinghua.edu.cn

ABSTRACT

The time dependent density matrix renormalization group (TD-DMRG) has become one of the cutting edge methods of quantum dynamics for complex systems. In this paper, we comparatively study the accuracy of three time evolution schemes in the TD-DMRG, the global propagation and compression method with the Runge-Kutta algorithm (P&C-RK), the time dependent variational principle based methods with the matrix unfolding algorithm (TDVP-MU), and with the projector-splitting algorithm (TDVP-PS), by performing benchmarks on the exciton dynamics of the Fenna-Matthews-Olson complex. We show that TDVP-MU and TDVP-PS yield the same result when the time step size is converged and they are more accurate than P&C-RK4, while TDVP-PS tolerates a larger time step size than TDVP-MU. We further adopt the graphical processing units to accelerate the heavy tensor contractions in the TD-DMRG, and it is able to speed up the TDVP-MU and TDVP-PS schemes by up to 73 times.

I. INTRODUCTION

The time dependent density matrix renormalization group (TD-DMRG) has emerged as a powerful tool to deal with many-body chemical and physical problems.[1,2] Although the DMRG is initially designed to solve the ground state of one-dimensional strongly correlated systems,[3,4] its applications are successfully extended to dynamical properties both in the time and frequency domains, such as the linear and nonlinear optical response of polyenes,[5] polaron formation and diffusion,[6–8] interconversion dynamics of pyrazine,[9–11] exciton dissociation,[12] spectra of molecular aggregates,[13] *ab initio* electron dynamics,[14,15] and many other topics.[16–20] One of the key components in the TD-DMRG is the time evolution scheme, which is essential to the numerical accuracy and efficiency. The available schemes could be roughly classified into three groups. The first group is based on globally approximating the formal propagator e^{-iHt} or $e^{-iHt}|\Psi\rangle$, including time-evolving block decimation (TEBD),[21–23] $W^{I,II}$ method,[24] Runge-Kutta,[13,25] Chebyshev expansion,[26] Krylov subspace[25,27] methods, and split operator method on the grid basis.[9] The same feature shared in these schemes is that in each time step the wavefunction is first propagated as a whole globally and then compressed. The second group is more inspired by the original DMRG, which is formulated in the local renormalized space, and the basis is adapted by the averaged reduced density matrix. The representatives are the time step targeting method (TST)[28] and some related variants.[14,29] The third group is based on the time dependent variational principle (TDVP).[30] Depending on the different ways to derive the equations of motion (EOMs), this group includes the original method with fixed gauge freedom[31] and the more recent projector-splitting method (PS) from a tangent space view.[32] Among the above evolution schemes, all schemes can be directly applied to models with long-range interactions except TEBD, which requires modifications such as the unitary transformation of the Hamiltonian in the system-reservoir quantum models[33] and introduction of swap gates.[34] In addition, the global evolution scheme is the most straightforward one when the modern framework of the matrix product state/matrix product operator (MPS/MPO) is investigated, while the PS scheme seems to have become the most popular choice as it has

been widely employed in the recent articles[10,11,17–20,35] and implemented in a number of TD-DMRG packages.[36,37] Although many evolution schemes have been applied extensively to the simulation of quantum dynamics in both chemistry and physics, a pragmatic analysis of their relative accuracy and efficiency has not been reported yet.

High performance computing for DMRG algorithms has attracted much interest in recent years, including parallelization strategies with the message passing interface (MPI),[38] open multiprocessing (OpenMP), and hybrid MPI/OpenMP[39] on multiple central processing units (CPUs). Meanwhile, the application of graphical processing units (GPUs) in computational chemistry has also drawn much attention in the past decade due to the tardy improvement of CPUs and the more and more vibrant GPU software ecosystem, including electronic structure calculation,[40–43] classical/*ab initio* molecular dynamics,[44,45] and open system quantum dynamics.[46,47] To the best of our knowledge, the only attempt to employ GPUs in the DMRG was made by Nemes *et al.* in 2014.[48] They came up with a smart implementation exploiting both the CPU and GPU which speeds up the Davidson diagonalization part in the DMRG algorithm by 2–5 times. Despite their efforts, how GPUs can accelerate TD-DMRG algorithms and which evolution scheme is more suitable remain unclear.

To benchmark the performance of different TD-DMRG time evolution schemes, a proper model should be chosen. In this work, we focus on the vibronic coupling system represented by the Fenna-Matthews-Olson (FMO) complex from green sulfur bacteria. It is an extremely popular system for both experimental and theoretical studies of energy transfer in photosynthesis.[49,50] The FMO model has become a "guinea pig" for comparing different computational methods,[18] and its exciton dynamics has been studied by many numerically exact methods, including the quasiadiabatic propagator path integral (QUAPI),[51] hierarchical equations of motion (HEOM),[52,53] multilayer multiconfiguration time-dependent Hartree (ML-MCTDH),[54,55] TD-DMRG,[16,18] etc.

In this paper, we select three time evolution schemes in the TD-DMRG to simulate the exciton dynamics of the 7-site FMO model, including the global propagation and compression scheme with the classical 4th order Runge-Kutta algorithm (P&C-RK4), TDVP with the advanced matrix unfolding regularization scheme (TDVP-MU), and TDVP with the projector-splitting algorithm (TDVP-PS). We first study the relative accuracy of the three schemes on the FMO model and then set out to explore the CPU-GPU heterogeneous computing to accelerate the TD-DMRG algorithms.

II. METHODOLOGICAL APPROACHES

A. MPS representation and TD-DMRG algorithms

In the language of matrix product states (MPSs),[56] a quantum state $|\Psi\rangle$ under a certain basis $|\sigma_1\sigma_2\cdots\sigma_N\rangle$, where $|\sigma_i\rangle$ is the basis for each degree of freedom (DOF) and is assumed to be orthonormal, can be represented as the product of a matrix chain, known as an MPS, as follows:

$$|\Psi\rangle = \sum_{\{a\},\{\sigma\}} A^{\sigma_1}_{a_1} A^{\sigma_2}_{a_1 a_2} \cdots A^{\sigma_N}_{a_{N-1}} |\sigma_1\sigma_2\cdots\sigma_N\rangle. \tag{1}$$

$A^{\sigma_i}_{a_{i-1}a_i}$ are the matrices in the chain connected by indices a_i. $\{\cdot\}$ in the summation represents the contraction of the respective connected indices, and N is the total number of DOFs in the system. The dimension of a_i is called the (virtual) bond dimension denoted as M_S or $|a_i|$, while the dimension of σ_i is called the physical bond dimension denoted as d. The graphical representation of an MPS is shown in Fig. 1(a). The many-body renormalized basis is defined as

$$|a_i[1:i]\rangle = \sum_{\{a\},\{\sigma\}} A^{\sigma_1}_{a_1} A^{\sigma_2}_{a_1 a_2} \cdots A^{\sigma_i}_{a_{i-1}a_i} |\sigma_1\cdots\sigma_i\rangle, \tag{2}$$

$$|a_j[j+1:N]\rangle = \sum_{\{a\},\{\sigma\}} A^{\sigma_{j+1}}_{a_j a_{j+1}} \cdots A^{\sigma_N}_{a_{N-1}} |\sigma_{j+1}\cdots\sigma_N\rangle, \tag{3}$$

where $[1: i]$ and $[j + 1: N]$ represent that the renormalized bases are defined from the left and right side, respectively. Hence, $A^{\sigma_n}_{a_{n-1}a_n}$ could be regarded as the coefficient matrix on the space spanned by $|a_{n-1}[1: n-1]\rangle \otimes |\sigma_n\rangle \otimes |a_n[n+1: N]\rangle$. However, in general, the renormalized basis is not necessary to be orthonormal, and the overlap matrix between them is defined as

$$S[1:i]_{a_i a_i'} = \langle a_i[1:i] | a_i'[1:i]\rangle, \tag{4}$$

$$S[j+1:N]_{a_j a_j'} = \langle a_j[j+1:N] | a_j'[j+1:N]\rangle. \tag{5}$$

The matrix product representation for a wavefunction is not unique in that inserting an identity matrix $I = GG^{-1}$ into the neighboring matrices $\cdots A^{\sigma_i} A^{\sigma_{i+1}} \cdots$ will obtain the same wavefunction but with different local matrices $\cdots A'^{\sigma_i} A'^{\sigma_{i+1}} \cdots = \cdots (A^{\sigma_i} G)(G^{-1} A^{\sigma_{i+1}}) \cdots = \cdots A^{\sigma_i} A^{\sigma_{i+1}} \cdots$. Therefore, gauge conditions could be applied to eliminate the parameterization redundancy of an MPS. Among them, the "mixed/left/right-canonical" gauge condition is usually adopted for convenience. A "mixed-canonical" MPS with the gauge center at site n is written as

$$|\Psi\rangle = \sum_{\{l\},\{r\}\{\sigma\}} L^{\sigma_1}_{l_1} L^{\sigma_2}_{l_1 l_2} \cdots L^{\sigma_{n-1}}_{l_{n-2} l_{n-1}} C^{\sigma_n}_{l_{n-1} r_n} \times R^{\sigma_{n+1}}_{r_n r_{n+1}} \cdots R^{\sigma_N}_{r_{N-1}} |\sigma_1\sigma_2\cdots\sigma_N\rangle, \tag{6}$$

where $L^{\sigma_i}_{l_{i-1}l_i}$ and $R^{\sigma_j}_{r_{j-1}r_j}$ satisfy

$$\sum_{\sigma_i, l_{i-1}} L^{\sigma_i *}_{l_{i-1} l_i'} L^{\sigma_i}_{l_{i-1} l_i} = \delta_{l_i' l_i}, \tag{7}$$

$$\sum_{\sigma_j, r_j} R^{\sigma_j *}_{r_{j-1}' r_j} R^{\sigma_j}_{r_{j-1} r_j} = \delta_{r_{j-1}' r_{j-1}}. \tag{8}$$

Here, $L^{\sigma_i}(R^{\sigma_j})^*$ represents the conjugate of $L^{\sigma_i}(R^{\sigma_j})$.

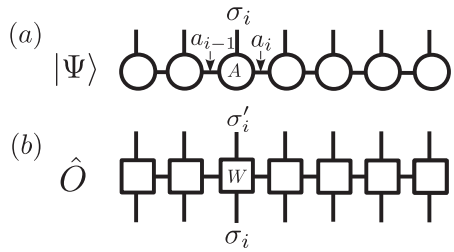

FIG. 1. The graphical representation of (a) an MPS in Eq. (1) and (b) an MPO in Eq. (11) with $N = 7$.

With Eqs. (7) and (8), the renormalized basis fulfills the orthonormal relations $S[1:i]_{l_i l_i'} = \delta_{l_i l_i'}$ ($i = 1, 2, \ldots, n-1$) and $S[j+1:N]_{r_j r_j'} = \delta_{r_j r_j'}$ ($j = n, n+1, \ldots, N-1$). When the gauge center n is at the right(left)-most site, the MPS is called the left(right)-canonical state. With the "mixed/left/right-canonical" condition, the only redundancy left is that G is unitary. If we further decompose C^{σ_n} by QR decomposition,

$$C^{\sigma_n}_{l_{n-1}r_n} \overset{\text{reshape}}{\rightarrow} C_{\sigma_n l_{n-1}, r_n} \overset{\text{QR}}{=} \sum_{l_n} L_{\sigma_n l_{n-1}, l_n} D_{l_n, r_n} \overset{\text{reshape}}{\rightarrow} \sum_{l_n} L^{\sigma_n}_{l_{n-1} l_n} D_{l_n r_n}, \tag{9}$$

L^{σ_n} fulfills the relation in Eq. (7) and $D_{l_n r_n}$ is also a coefficient matrix defined between sites n and $n+1$ on space $|l_n[1{:}\,n]\rangle \otimes |r_n[n+1{:}\,N]\rangle$. Afterwards, $D_{l_n r_n}$ is combined with $R^{\sigma_{n+1}}_{r_n r_{n+1}}$ to obtain $C^{\sigma_{n+1}}_{l_n r_{n+1}} = \sum_{r_n} D_{l_n r_n} R^{\sigma_{n+1}}_{r_n r_{n+1}}$, and apparently, the gauge center is moved one site to the right. The reverse process to move the gauge center to the left could be carried out in a similar way as Eq. (9) while QR is replaced with RQ decomposition. More generally, a canonical MPS with the gauge center at n as Eq. (6) could be prepared by performing QR decomposition from site 1 to $n-1$ sequentially and RQ decomposition from site N to $n+1$ sequentially on any MPS, which is called canonicalization. In the following paper, QR and RQ are not distinguished for simplicity.

If the QR decomposition in Eq. (9) is replaced with singular value decomposition (SVD),

$$C^{\sigma_n}_{l_{n-1}r_n} \overset{\text{reshape}}{\rightarrow} C_{\sigma_n l_{n-1}, r_n} \overset{\text{SVD}}{=} \sum_{s=1}^{k} U_{\sigma_n l_{n-1}, s} \Lambda_{ss} V^{\dagger}_{s, r_n} \overset{\text{reshape}}{\rightarrow} \sum_{s=1}^{k} U^{\sigma_n}_{l_{n-1} s} \Lambda_{ss} V^{\dagger}_{s r_n} \approx \sum_{l_n=1}^{M_S<k} L^{\sigma_n}_{l_{n-1} l_n} \Lambda_{l_n l_n} V^{\dagger}_{l_n r_n}. \tag{10}$$

Here, Λ is a real diagonal matrix with elements ordered from large to small $\Lambda_{11} \geq \Lambda_{22} > \cdots \geq \Lambda_{kk} \geq 0$ ($k = \min[|\sigma_n| \cdot |l_{n-1}|, |r_n|]$) and $\sum_s \Lambda^2_{ss} = 1$ if $|\Psi\rangle$ is normalized. U^{σ_n} fulfills the relation in Eq. (7). If we retain all Λ_{ss} ($M_S = k$) and move the gauge center to the right with $C^{\sigma_{n+1}}_{l_n r_{n+1}} = \sum_{l_n, r_n} \Lambda_{l_n l_n} V^{\dagger}_{l_n r_n} R^{\sigma_{n+1}}_{r_n r_{n+1}}$, the wavefunction is not changed and SVD plays the same role as QR in canonicalization but with a little bit higher cost. However, if we only retain the largest M_S terms ($M_S < k$) to obtain $\tilde{C}^{\sigma_{n+1}}_{l_n r_{n+1}}$, $\tilde{C}^{\sigma_{n+1}}_{l_n r_{n+1}}$ could be a good approximation to $C^{\sigma_{n+1}}_{l_n r_{n+1}}$ with a smaller bond dimension. The loss of accuracy due to this truncation could be measured by $\varepsilon = \sum_{s=M_S+1}^{k} \Lambda^2_{ss}$ if $|\Psi\rangle$ is normalized. With this algorithm, a left(right)-canonical MPS $|\Psi\rangle$ could be "compressed" to $|\tilde{\Psi}\rangle$ by successive approximate SVDs from site N to 1 (site 1 to N). At each local step of the compression, two truncation criteria are commonly used (i) a fixed predefined M_S and (ii) adaptive M_S with all Λ_{ss} larger than predefined ζ retained.

Apart from canonicalization and compression, several other operations are usually met within the MPS algorithms, including $\hat{O}|\Psi\rangle$ and $|\Psi\rangle + |\Phi\rangle$. Similar to the MPS, a common quantum operator could be expressed as a matrix product operator (MPO) as follows:[56,57]

$$\hat{O} = \sum_{\{w\},\{\sigma\},\{\sigma'\}} W^{\sigma_1', \sigma_1}_{w_1} W^{\sigma_2', \sigma_2}_{w_1 w_2} \cdots W^{\sigma_N', \sigma_N}_{w_{N-1}} \times |\sigma_1' \sigma_2' \cdots \sigma_N'\rangle \langle \sigma_N \sigma_{N-1} \cdots \sigma_1|. \tag{11}$$

The graphical representation of an MPO is shown in Fig. 1(b). With the local matrix representation of the wavefunction and operator in Eqs. (1) and (11), $\hat{O}|\Psi\rangle$ (MPO × MPS) could be calculated by contracting $W^{\sigma_i', \sigma_i} A^{\sigma_i}$ locally as follows:

$$\hat{O}|\Psi\rangle = \sum_{\{w,a\},\{\sigma'\}} A'^{\sigma_1'}_{\{w,a\}_1} A'^{\sigma_2'}_{\{w,a\}_1 \{w,a\}_2} \cdots A'^{\sigma_N'}_{\{w,a\}_{N-1}} |\sigma_1' \sigma_2' \cdots \sigma_N'\rangle, \tag{12}$$

where

$$A'^{\sigma_i'}_{\{w,a\}_{i-1}\{w,a\}_i} = \sum_{\sigma_i} W^{\sigma_i', \sigma_i}_{w_{i-1} w_i} A^{\sigma_i}_{a_{i-1} a_i}. \tag{13}$$

Suppose the bond dimensions of the original MPS and the MPO are M_S and M_O, respectively, then the new state $|\hat{O}\Psi\rangle$ has a bond dimension $M_O M_S$. $|\Psi\rangle + |\Phi\rangle$ (MPS + MPS) is constructed by merging the local matrices $[A^{\sigma_i}, B^{\sigma_i}]$ block-diagonally,

$$|\Psi\rangle + |\Phi\rangle = \sum_{\{a,b\},\{\sigma\}} A'^{\sigma_1}_{\{a,b\}_1} A'^{\sigma_2}_{\{a,b\}_1 \{a,b\}_2} \cdots A'^{\sigma_N}_{\{a,b\}_{N-1}} |\sigma_1 \sigma_2 \cdots \sigma_N\rangle, \tag{14}$$

where

$$A'^{\sigma_1} = \begin{bmatrix} A^{\sigma_1} & B^{\sigma_1} \end{bmatrix}, \quad A'^{\sigma_i} = \begin{bmatrix} A^{\sigma_i} & \mathbf{0} \\ \mathbf{0} & B^{\sigma_i} \end{bmatrix} (i = 2, 3 \cdots N-1),$$

$$A'^{\sigma_N} = \begin{bmatrix} A^{\sigma_N} \\ B^{\sigma_N} \end{bmatrix}. \tag{15}$$

Suppose the bond dimensions of the original two MPSs are M_{S_A} and M_{S_B}, respectively, then the new state $|\Psi + \Phi\rangle$ has a bond dimension $M_{S_A} + M_{S_B}$.

In Subsections II A 1 and II A 2, we present the ideas and the algorithms of three different TD-DMRG time evolution schemes adopted in our benchmark calculations afterwards: P&C-RK4, TDVP-MU, and TDVP-PS.

1. Schemes based on global propagation and compression

The propagation and compression (P&C) scheme is a global time evolution scheme benefiting from the modern MPS/MPO representation of DMRG wavefunctions and operators. If $|\Psi(t)\rangle$ and the time derivative $|\dot{\Psi}(t)\rangle$ are explicitly known (according to the Schrödinger equation, the time derivative is $-i\hat{H}\Psi$ in atomic units), any ordinary differential equation (ODE) integrator for the initial value problem (IVP), such as the classical 4th order Runge-Kutta algorithm (RK4) we used in our former work,[13] could be applied to obtain the MPS of the next time step $|\Psi(t+\tau)\rangle$. The equations of RK4 are

$$\begin{aligned} |k_1\rangle &= -i\hat{H}(t)|\Psi(t)\rangle, \\ |k_2\rangle &= -i\hat{H}(t+\tau/2)(|\Psi(t)\rangle + \frac{1}{2}\tau|k_1\rangle), \\ |k_3\rangle &= -i\hat{H}(t+\tau/2)(|\Psi(t)\rangle + \frac{1}{2}\tau|k_2\rangle), \\ |k_4\rangle &= -i\hat{H}(t+\tau)(|\Psi(t)\rangle + \tau|k_3\rangle), \\ |\Psi(t+\tau)\rangle &= |\Psi(t)\rangle + \frac{1}{6}\tau(|k_1\rangle + 2|k_2\rangle + 2|k_3\rangle + |k_4\rangle). \end{aligned} \tag{16}$$

Each $|k_i\rangle$ is represented by an MPS, and $\hat{H}(t_i)$ is represented by an MPO. Therefore, only two types of operations including MPO × MPS [Eq. (13)] and MPS + MPS [Eq. (14)] exist in Eq. (16).

After each operation, the new MPS such as $|k_1\rangle = -i\hat{H}(t)|\Psi(t)\rangle$ will have a larger virtual bond dimension, which should be compressed using the algorithm based on Eq. (10) for the further operations. It is worth mentioning that as the new MPS is not canonical, canonicalization should be carried out before the actual compression.

The procedure of P&C-RK4 described above involves $\hat{H}|\Psi\rangle$ ("MPO × MPS") scaling at $\mathcal{O}(M_S^2 M_O^2 d^2)$ for each site (labeled as "$\hat{H}|\Psi\rangle$"), "MPS + MPS," which does not require intensive computation and MPS compression. The compression consists of four kinds of operations [two for canonicalization in Eq. (9) and two for compression in Eq. (10)], namely, QR decomposition of tensors on each site (labeled as "QR"), matrix multiplication to absorb the decomposed coefficients (labeled as "MatMul-QR"), and the SVD counterparts (labeled as "SVD" and "MatMul-SVD"). "QR" and "MatMul-QR" both scale at $\mathcal{O}(M_S^3 M_O^3 d)$, while their SVD counterparts scale at $\mathcal{O}(\min(M_S^3 M_O^2 d, M_S^3 M_O d^2))$ for "SVD" and $\mathcal{O}(M_S^3 M_O^2 d)$ for "MatMul-SVD" because compression reduces the bond dimension from $M_O M_S$ to M_S during the sweep. Thus, the bottleneck of P&C-RK4 is the canonicalization, and the scaling of the method in a single time step is $\mathcal{O}(N M_S^3 M_O^3 d)$.

Compared to the local evolution scheme such as TST,[28] P&C could exactly calculate the operation of a high order Hamiltonian such as $\hat{H}^n|\Psi\rangle$ and it is natural to increase the bond dimension with entanglement during time evolution. However, though P&C is in principle a general evolution scheme for the TD-DMRG, it prefers Hamiltonian whose MPO is straightforward to construct and has a small bond dimension M_O, such as the Frenkel-Holstein type model whose M_O is in a linear relationship to the number of electronic states.[13] With regard to the time step size, the error of P&C does not have a monotonic relationship with the time step and there is an optimal time step depending on the system as discussed in the previous studies.[13,14] A smaller time step will improve the accuracy of the RK4 integrator but lead to more MPS compressions and then deteriorate the whole accuracy.

2. Schemes based on the time dependent variational principle

The Rayleigh-Ritz variational principle is widely used in finding an approximate ground state in the time independent Schrödinger equation. Similarly, the time dependent variational principle (TDVP) also provides a strong tool to find an optimal time dependent wavefunction if the wavefunction ansatz and the initial state are known. The Dirac-Frenkel TDVP is[30,58]

$$\langle \delta\Psi | i\frac{\partial}{\partial t} - \hat{H}|\Psi\rangle = 0. \tag{17}$$

It has been proved that the TDVP could strictly conserve the norm of the wavefunction and the total energy in the real-time propagation,[58] which is believed to be essential for long-time dynamics. In a geometric fashion, the TDVP could be understood as an orthogonal projection of $-i\hat{H}|\Psi\rangle$ onto the tangent space of $|\Psi(t)\rangle$ at the current time,

$$\frac{\partial|\Psi\rangle}{\partial t} = -i\hat{P}\hat{H}|\Psi\rangle, \tag{18}$$

where $\hat{P}$ is the projector constructed by the orthonormal vectors in the tangent space. For a general MPS in Eq. (1), the tangent space projector could be defined as

$$\hat{P} = \sum_{i=1}^{N} \hat{P}[1:i-1]\otimes\hat{I}_i\otimes\hat{P}[i+1:N] - \sum_{i=1}^{N-1}\hat{P}[1:i]\otimes\hat{P}[i+1:N], \tag{19}$$

where

$$\hat{P}[1:i] = \sum_{a_i,a_i'} |a_i'[1:i]\rangle S[1:i]^{-1}_{a_i' a_i}\langle a_i[1:i]|, \tag{20}$$

$$\hat{P}[i+1:N] = \sum_{a_i,a_i'} |a_i'[i+1:N]\rangle S[i+1:N]^{-1}_{a_i' a_i}\langle a_i[i+1:N]|, \tag{21}$$

$$\hat{I}_i = \sum_{\sigma_i} |\sigma_i\rangle\langle\sigma_i|, \tag{22}$$

$$\hat{P}[1:0] = \hat{P}[N+1:N] = 1. \tag{23}$$

The inversion of the overlap matrix S^{-1} accounts for the nonorthogonality of the renormalized basis, and the "−" terms are to eliminate the parameterization redundancy.[31,59] The graphical representation of the projector is shown in Fig. 2. In the literature, there are two different time evolution schemes based on the TDVP. They differ in choosing the specific gauge condition of the MPS and in solving Eq. (18), which will be discussed in detail in the following.

a. TDVP-MU In the first TDVP evolution scheme, the gauge freedom of the MPS is fixed. For convenience, the projector in Eq. (19) could be transformed to the following equation [Eq. (24)] by combining the neighboring "+" term and "−" term together except one "+" term with $i = n$:

$$\hat{P} = \hat{P}[1:n-1]\otimes\hat{I}_n\otimes\hat{P}[n+1:N] + \sum_{i=1}^{n-1}\hat{Q}[1:i]\otimes\hat{P}[i+1:N] + \sum_{i=n+1}^{N}\hat{P}[1:i-1]\otimes\hat{Q}[i:N], \tag{24}$$

where

$$\hat{Q}[1:i] = \sum_{a_{i-1},a_{i-1}',\sigma_i,\sigma_i'} |a_{i-1}'[1:i-1]\sigma_i'\rangle\langle a_{i-1}[1:i-1]\sigma_i| \cdot \Big(S[1:i-1]^{-1}_{a_{i-1}' a_{i-1}}\delta_{\sigma_i'\sigma_i} - \sum_{a_i',a_i} A^{\sigma_i'}_{a_{i-1}' a_i'} S[1:i]^{-1}_{a_i' a_i} A^{\sigma_i *}_{a_{i-1} a_i}\Big), \tag{25}$$

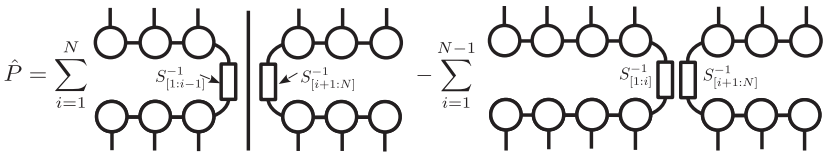

FIG. 2. The graphical representation of the tangent space projector in Eq. (19) with $N = 7$.

$$\hat{Q}[i:N] = \sum_{a_i,a_i',\sigma_i,\sigma_i'} |a_i'[i+1:N]\sigma_i'\rangle\langle a_i[i+1:N]\sigma_i| \cdot \Big(S[i+1:N]^{-1}_{a_i'a_i}\delta_{\sigma_i'\sigma_i} - \sum_{a_{i-1}',a_{i-1}} A^{\sigma_i'}_{a_{i-1}'a_i'} S[i:N]^{-1}_{a_{i-1}'a_{i-1}} A^{\sigma_i *}_{a_{i-1}a_i}\Big). \tag{26}$$

This type of projector or the corresponding tangent space vectors were first proposed by Haegeman *et al.*[31] in the TD-DMRG and restated by Wouters *et al.*[59] to derive several post-DMRG methods for the ground state and excited states.

Equations (25) and (26) could be further simplified by adopting a specific gauge condition, and then, some overlap matrices turn to identity. Assuming that the MPS is left-canonical with the gauge center at site N, $S[1{:}\,i]$ is reduced to I and it is most convenient to set $n = N$ in Eq. (24). Inserting the simplified projector into Eq. (18) yields

$$i\frac{\partial C^{\sigma_N'}_{l_{N-1}'}}{\partial t} = \sum_{\sigma_N'',l_{N-1}''} H[N]_{l_{N-1}'\sigma_N',l_{N-1}''\sigma_N''} C^{\sigma_N''}_{l_{N-1}''}, \tag{27}$$

$$i\frac{\partial L^{\sigma_i'}_{l_{i-1}'l_i'}}{\partial t} = \sum_{l_{i-1},\sigma_i} (\delta_{l_{i-1}'l_{i-1}}\delta_{\sigma_i'\sigma_i} - p[i]_{l_{i-1}'\sigma_i',l_{i-1}\sigma_i}) \times \sum_{l_i} S[i+1:N]^{-1}_{l_i'l_i} \times \sum_{l_{i-1}'',\sigma_i'',l_i''} H[i]_{l_{i-1}\sigma_i l_i,l_{i-1}''\sigma_i''l_i''} L^{\sigma_i''}_{l_{i-1}''l_i''}, \tag{28}$$

where

$$H[i]_{l_{i-1}'\sigma_i'l_i',l_{i-1}\sigma_i l_i} = \sum_{\{w\}} h[1:i-1]_{\{l',w,l\}_{i-1}} W^{\sigma_i',\sigma_i}_{w_{i-1}w_i} \times h[i+1:N]_{\{l',w,l\}_i}, \tag{29}$$

$$h[1:i-1]_{\{l',w,l\}_{i-1}} = \sum_{\{l'\},\{w\},\{l\}} h[1]_{\{l',w,l\}_1}\cdots \times h[i-1]_{\{l',w,l\}_{i-2},\{l',w,l\}_{i-1}}, \tag{30}$$

$$h[i+1:N]_{\{l',w,l\}_i} = \sum_{\{l'\},\{w\},\{l\}} h[i+1]_{\{l',w,l\}_i,\{l',w,l\}_{i+1}}\cdots \times h[N]_{\{l',w,l\}_{N-1}}, \tag{31}$$

$$h[i]_{\{l',w,l\}_{i-1},\{l',w,l\}_i} = \sum_{\sigma_i,\sigma_i'} A^{\sigma_i'*}_{l_{i-1}'l_i'} W^{\sigma_i',\sigma_i}_{w_{i-1}w_i} A^{\sigma_i}_{l_{i-1}l_i} \quad (A = L \text{ or } C), \tag{32}$$

$$p[i]_{l_{i-1}'\sigma_i',l_{i-1}\sigma_i} = \sum_{l_i} L^{\sigma_i'}_{l_{i-1}'l_i} L^{\sigma_i *}_{l_{i-1}l_i}. \tag{33}$$

Similar equations can be derived for the right/mixed-canonical MPS. With Eqs. (28) and (33), it is straightforward to prove that $\sum_{l_{i-1},\sigma_i} L^{\sigma_i *}_{l_{i-1}l_i}\frac{\partial}{\partial t}L^{\sigma_i}_{l_{i-1}l_i'} = 0$ and then

$$\frac{\partial}{\partial t}\Big(\sum_{l_{i-1},\sigma_i} L^{\sigma_i *}_{l_{i-1}l_i} L^{\sigma_i}_{l_{i-1}l_i'}\Big) = 0, \tag{34}$$

which ensures that the left-canonical condition preserves during the time evolution formally. Though in practical numerical calculation with a finite time step the relation in Eq. (34) would not be rigorously fulfilled, this problem is not severe with a proper time step in our experience. Otherwise, the more general equations in Eq. (24) which have already considered the nonorthogonality of the left and right renormalized basis should be used. In addition, it could be proved that $\frac{\partial}{\partial t}S[1:i] = 0\,(i = 1,2,\ldots,n-1)$ and $\frac{\partial}{\partial t}S[j:N] = 0\,(j = n+1, n+2,\ldots,N)$.

Equations (27) and (28) together form a set of coupled nonlinear equations that are very similar to the standard EOMs of (ML-) MCTDH.[60–62] To integrate these equations, we borrow ideas called the variable mean field (VMF) and constant mean field (CMF) from the MCTDH community.[61,63] The VMF employs an all-purpose integrator to directly solve the coupled equations. While in the CMF, it is assumed that $H[i]$ and $S[i+1{:}\,N]$ generally change much slower in time than the local matrices C^{σ_N} and L^{σ_i}. As a result, during the integration of Eqs. (27) and (28), one may hold the "mean field" $H[i]$ and $S[i+1{:}\,N]$ constant for τ and evolve only the local matrix with a time step smaller than τ. Hence, the CMF is more efficient yet less accurate than the VMF. In this work, we use a second order CMF with a midpoint scheme in which $L^{\sigma_i}(0)$ and $C^{\sigma_N}(0)$ are integrated to $L^{\sigma_i}(\tau)$ and $C^{\sigma_N}(\tau)$ with the "mean field" constructed by $L^{\sigma_i}(\tau/2)$ and $C^{\sigma_N}(\tau/2)$. In the VMF, the adaptive Dormand-Prince's 5/4 Runge-Kutta method (RK45) is adopted to integrate Eqs. (27) and (28), while in CMF, since Eq. (27) is linear, it is integrated by the Krylov subspace method instead.

Another aspect should be concerned is that the inversion of S would be unstable numerically if some eigenvalues of S are very small. This problem will be severe when the state is weakly correlated (such as a Hartree product state which is usually an initial state) and M_S is much larger than what is required. To some extent, this instability problem makes this evolution scheme paradoxical in that large M_S should in principle push the result to a numerically exact limit, but in fact, it deteriorates it. The same problem also arises in (ML)-MCTDH, where in order to make the EOMs more "well-behaved," S is usually replaced with a regularized overlap matrix $\tilde{S}$ as follows:[61]

$$\tilde{S} = S + \varepsilon e^{-S/\varepsilon}. \tag{35}$$

Here, ε is a small scalar commonly from 10^{-8} to 10^{-14}. More recently, an improved regularization scheme based on the matrix unfolding (MU) of the coefficient matrix by SVD in (ML-)MCTDH is proposed by Meyer and Wang, which has been proved to make the time integration more accurate and robust.[64,65] The same idea is adopted here to integrate Eq. (28), giving the name of the scheme as "TDVP-MU." When calculating the overlap matrix $S[i+1:N]^{-1}_{a_i'a_i}$, the gauge center is moved to the $(i+1)$th site and the matrix at this site is further decomposed by SVD,

$$|l_i[i+1:N]\rangle = \sum_{\{l\},\{\sigma\}} L^{\sigma_{i+1}}_{l_il_{i+1}}\cdots L^{\sigma_{N-1}}_{l_{N-2}l_{N-1}} C^{\sigma_N}_{l_{N-1}} |\sigma_{i+1}\cdots\sigma_N\rangle = \sum_{\{r\},\{\sigma\}} U_{l_ir_i}\Lambda_{r_ir_i} R^{\sigma_{i+1}}_{r_ir_{i+1}}\cdots R^{\sigma_{N-1}}_{r_{N-2}r_{N-1}} R^{\sigma_N}_{r_{N-1}} \times |\sigma_{i+1}\cdots\sigma_N\rangle, \tag{36}$$

where $|r_i|$ equals $|l_i|$. Thus, the overlap matrix $S[i+1{:}\,N]$ and its inversion could be expressed as $U^*\Lambda^2U^T$ and $U^*\Lambda^{-2}U^T$, respectively. The Hamiltonian matrix in Eq. (29) is also reconstructed. For the site from $i+1$ to N, the matrix A^{σ_i} in Eq. (32) is replaced with matrix R^{σ_i} in Eq. (36), and then, Eq. (28) with the matrix unfolding

algorithm becomes

$$i\frac{\partial L^{\sigma_i'}_{l'_{i-1}l'_i}}{\partial t} = \sum_{l_{i-1},\sigma_i}(\delta_{l'_{i-1}l_{i-1}}\delta_{\sigma_i'\sigma_i} - p[i]_{l'_{i-1}\sigma_i',l_{i-1}\sigma_i}) \times \sum_{r_i,r_i',l_i}\left[U^*_{l'_ir'_i}\Lambda^{-1}_{r'_ir'_i}\underline{\Lambda^{-1}_{r'_ir'_i}U^T_{r'_il_i}}\right]U^*_{l_ir_i}\Lambda_{r_ir_i} \times \sum_{l''_{i-1},\sigma_i'',l''_i,r''_i} H[i]_{l_{i-1}\sigma_ir_i,l''_{i-1}\sigma_i''r_i''}\Lambda_{r_i''r_i''}U^T_{r_i''l''_i}L^{\sigma_i''}_{l''_{i-1}l''_i}. \quad (37)$$

The expression inside "[···]" is $S[i+1:N]^{-1}$. The key point of this new regularization scheme is that the underlined part could be contracted first, which is $\delta_{r'_ir_i}$. Thus, only the singular matrix $\Lambda_{r'_ir'_i}$ instead of $\Lambda^2_{r'_ir'_i}$ should be regularized as follows:

$$\tilde{\Lambda}_{r'_ir'_i} = \Lambda_{r'_ir'_i} + \varepsilon^{1/2}e^{-\Lambda_{r'_ir'_i}/\varepsilon^{1/2}}. \quad (38)$$

The power 1/2 here is for consistency with the original regularization scheme [Eq. (35)], and $\Lambda_{r_i''r_i''}$ in Eq. (37) is untouched in order to be minimally invasive, as stated in Refs. 64 and 65. In this work, we choose $\varepsilon = 10^{-10}$ unless otherwise stated. Although in the matrix unfolding scheme it is necessary to perform canonicalization on the environmental part $|l_i[i+1:N]\rangle$, the gauge condition of the MPS that is evolved remains unchanged. In principle, the tangent space is strictly the same no matter where the gauge center is, as shown in Eq. (24) where n is arbitrary. Practically, the difference comes from the regularization on different parts of the wavefunction, but as we will show below, the regularization only introduces a minor error in the short time regime with a proper regularization parameter. Therefore, in our implementation, the MPS is always left-canonical and we expect the effect of the changing gauge center to be negligible.

The complexity analysis of the TDVP-MU scheme with the VMF and CMF is as follows: Note that in the TD-DMRG, it is necessary to contract small matrices one by one instead of explicitly constructing the large tensors.[38,56] The VMF and CMF both require the calculation of environment matrices $h[1:i-1]$ and $h[i+1:N]$ scaling at $\mathcal{O}(M_S^2M_O^2d^2 + M_S^3M_Od)$ (labeled as "Get Env"), SVD of the gauge center with the subsequent matrix multiplication scaling at $\mathcal{O}(M_S^3d)$ in Eq. (36) (labeled as "SVD" and "MatMul-SVD"), and the calculation of the time derivative in Eqs. (27) and (37) (labeled as "Deriv"). Evaluation of the time derivatives contains a number of matrix multiplications, and the overall scaling is $\mathcal{O}(M_S^2M_O^2d^2 + M_S^3M_Od + M_S^3d^2)$, where the first two terms and the last term correspond to the contraction of $H[i]$ and $(I - p[i])$, respectively. Thus, the total scaling in the VMF is $\mathcal{O}(N(M_S^2M_O^2d^2 + M_S^3M_Od + M_S^3d^2))$, while in the CMF, the total scaling for a single step is $\mathcal{O}(Nf(M_S^2M_O^2d^2 + M_S^3M_Od + M_S^3d^2))$, where f is the number of steps required for the propagation of each local site within τ. Although at first glance the CMF is more time-consuming than the VMF, a much larger step size is possible for the CMF which reduces the time cost spent on decomposing the matrix and constructing the environment, compensating the effect of the factor f. This time saving in the CMF is important in MCTDH in that constructing the environment involves a contraction of the high order coefficient matrix.[58,61,63] However, in the MPS context, such an advantage is not prominent because all matrices including the coefficient matrix C^{σ_N} are of the same size. As a result, constructing the environment has a similar complexity as calculating the derivatives.

b. TDVP-PS The second evolution scheme based on the TDVP is called projector-splitting (PS). The idea of PS is that the tangent space projector in Eq. (19) is invariant under different gauge conditions. More specifically, after canonicalization of a general MPS in Eq. (1) from site N to $i+1$, $|r_i[i+1:N]\rangle$ becomes the right-hand orthonormal renormalized basis, which is related to $|a_i[i+1:N]\rangle$, by

$$|a_i[i+1:N]\rangle = \sum_{r_i} D_{a_ir_i}|r_i[i+1:N]\rangle. \quad (39)$$

The matrix D is an upper triangular matrix in RQ decomposition after the canonicalization described in Eq. (9). Therefore, the overlap matrix $S[i+1:N]$ equals D^*D^T and the projector $\hat{P}[i+1:N]$ in Eq. (21) defined for a general noncanonical MPS is transformed to

$$\hat{P}[i+1:N] = \sum_{r'_i,r_i}|r'_i[i+1:N]\rangle\left[\sum_{a'_i,a_i}D^T_{r'_ia'_i}(D^*D^T)^{-1}_{a'_ia_i}D^*_{a_ir_i}\right]_{=\delta_{r'_ir_i}} \times\langle r_i[i+1:N]| = \sum_{r_i}|r_i[i+1:N]\rangle\langle r_i[i+1:N]|. \quad (40)$$

A similar result can be obtained for $\hat{P}[1:i]$ as follows:

$$\hat{P}[1:i] = \sum_{l_i}|l_i[1:i]\rangle\langle l_i[1:i]|. \quad (41)$$

On the one hand, this definition of the tangent space projector does not contain any inversion operations of the overlap matrix, which seems to be a remarkable improvement over the first definition in Eqs. (20) and (21). On the other hand, since the gauge is not fixed in different terms of this projector, the integration algorithm described in Sec. II A 2 a could not be directly applied. Lubich and Haegeman *et al.* proposed to use a symmetric second order Trotter decomposition to split the formal propagator into the individual terms,[32,66]

$$e^{-i\hat{P}\hat{H}\tau} = \left[\prod_{i=1}^{N-1}e^{-i\hat{P}[1:i-1]\otimes\hat{I}_i\otimes\hat{P}[i+1:N]\hat{H}\tau/2}\cdot e^{i\hat{P}[1:i]\otimes\hat{P}[i+1:N]\hat{H}\tau/2}\right] \cdot e^{-i\hat{P}[1:N-1]\otimes\hat{I}_N\hat{H}\tau} \cdot\left[\prod_{i=N-1}^{1}e^{i\hat{P}[1:i]\otimes\hat{P}[i+1:N]\hat{H}\tau/2}\cdot e^{-i\hat{P}[1:i-1]\otimes\hat{I}_i\otimes\hat{P}[i+1:N]\hat{H}\tau/2}\right] + \mathcal{O}(\tau^3). \quad (42)$$

Based on the propagator in Eq. (42), a single step of time evolution consists of a left-to-right sweep and a subsequent right-to-left sweep each with step size $\tau/2$. Taking the left-to-right sweep as an example, the matrix at the gauge center $C^{\sigma_i}_{l_{i-1}r_i}$ is first evolved forward in time by applying the projector $\hat{P}[1:i-1]\otimes\hat{I}_i\otimes\hat{P}[i+1:N]$ as follows:

$$i\frac{\partial C^{\sigma_i'}_{l'_{i-1}r'_i}}{\partial t} = \sum_{l_{i-1},\sigma_i,r_i}H[i]_{l'_{i-1}\sigma_i'r'_i,l_{i-1}\sigma_ir_i}C^{\sigma_i}_{l_{i-1}r_i}, \quad (43)$$

where $H[i]$ and the ingredients $h[1: i-1]$, $h[i+1: N]$, and $h[i]$ all have the same definitions as in Eqs. (29)–(32) except that the A^{σ_i} in Eq. (32) is replaced with L^{σ_i} or R^{σ_i} accordingly. Then, the evolved matrix $C^{\sigma_i}_{l_{i-1}r_i}$ is decomposed by QR as Eq. (9) to obtain the left-canonical matrix $L^{\sigma_i}_{l_{i-1}l_i}$ and the coefficient matrix $D_{l_ir_i}$. $D_{l_ir_i}$ is evolved backward in time by applying the projector $\hat{P}[1:i] \otimes \hat{P}[i+1:N]$ as follows:

$$i\frac{\partial D_{l'_ir'_i}}{\partial t} = \sum_{l_i,w_i,r_i} h[1:i]_{\{l',w,l\}_i} h[i+1:N]_{\{r',w,r\}_i} D_{l_ir_i}. \quad (44)$$

Afterwards, the gauge center is moved to site $i+1$ by contracting the evolved $D_{l_ir_i}$ and $R^{\sigma_{i+1}}_{r_ir_{i+1}}$ together to obtain $C^{\sigma_{i+1}}_{l_ir_{i+1}} = \sum_{r_i} D_{l_ir_i} R^{\sigma_{i+1}}_{r_ir_{i+1}}$. Following the procedure above, the sweep continues until all the individual projectors in Eq. (42) are applied. Inspired by the original two-site DMRG algorithm, it is also possible to formulate TDVP-PS into a two-site algorithm so that the bond dimension could grow up adaptively.[11,32] However, considering the two-site algorithm is much more expensive than the one-site algorithm both in the tensor contraction and QR decomposition, we use the one-site algorithm described above in this paper.

In principle, solving Eqs. (43) and (44) could be accomplished by any ODE integrator such as the RK45 algorithm we use in TDVP-MU. However, since they are linear equations, the Krylov subspace method (the Lanczos algorithm for the Hermitian operator here) is preferred as it is unitary and is considered to be better than the explicit time-stepping integrators for the matrix exponential operator.[32,67] In our calculation, the dimension of the Krylov subspace is adaptive and the Lanczos iteration continues until $|\Psi(t+\tau)\rangle$ converges.

TDVP-PS has a similar computational complexity as TDVP-MU (CMF) except for the calculation of derivatives. With simpler EOMs in Eqs. (43) and (44), the calculation of the derivative for forward evolution scales at $\mathcal{O}(M_S^2 M_O^2 d^2 + M_S^3 M_O d)$ (labeled as "Deriv-Forward") and for backward evolution it scales at $\mathcal{O}(M_S^3 M_O)$ (labeled as "Deriv-Backward"). Therefore, the total scaling for a single step is $\mathcal{O}(Nf(M_S^2 M_O^2 d^2 + M_S^3 M_O d))$, where f represents the number of Krylov subspace vectors. Besides calculating the derivatives, the Krylov subspace method also includes other matrix operations of small size such as calculating the $\{\alpha\}$ and $\{\beta\}$ matrix elements with the Lanczos three-term recurrence relation[68] and diagonalizing the tridiagonal Hamiltonian in the subspace (labeled as "Krylov"). Other labels for TDVP-PS are the same with that of TDVP-MU.

In summary, since the TDVP-MU scheme and the TDVP-PS scheme are both based on the time dependent variational principle, $|\Psi(t)\rangle$ should be the same if not considering the numerical error. Additionally, both the schemes require defining a fixed M_S *a priori*, and an additional renormalized basis should be constructed smartly to complement the empty MPS space if the initial state is weakly correlated. It is worth noting that the TDVP provides a deterministic wavefunction path during the time evolution determined by tangent space projections with an infinitesimal time step, which is locally optimal. As discussed in a recent review,[2] the TDVP path suffers from the projection error due to the restricted bond dimension M_S and it may not be the globally optimal path in some models. If this TDVP path is the target in the time integration or the projection error is smaller than the time integration error with a sufficiently large M_S, the error of these two TDVP-based schemes will have a monotonic relationship with the time step size. The main difference lies in that TDVP-MU would introduce a minor artificial regularization, while TDVP-PS is inherently free of it. We also note that TDVP-PS has already been implemented in MCTDH recently.[69–71]

B. CPU-GPU heterogeneous computing

In the TD-DMRG algorithms described above, the hotspots are usually the tensor contractions and the matrix decompositions. GPUs are well suitable for the former not for the latter. Therefore, we adopt a CPU-GPU heterogeneous computing strategy. We store the matrices in the GPU memory instead of the host memory, and in most cases, we call cuBLAS to manipulate them. When doing matrix decomposition, we first transfer the matrix from the GPU memory to the host memory and then use a single CPU core to complete the decomposition, and in the end, we copy the matrix back. The overhead caused by data transfer is not negligible, which we believe could be avoided by a smarter implementation because GPUs are able to run computation and data transfer at the same time.

For comparison, multicore CPU calculation is also benchmarked. Here, we only consider the simplest strategy of CPU parallelism, which is breaking up the dense matrix computations into sub-blocks performed automatically by the standard linear algebra libraries such as the Intel® Math Kernel Library in our case.

C. Computational details of the FMO model

The widely used Frenkel-Holstein Hamiltonian[72,73] to describe the 7-site FMO model is that

$$\hat{H} = \sum_m E_m a_m^\dagger a_m + \sum_{m\neq n} J_{mn} a_m^\dagger a_n + \sum_{m\lambda} \omega_{m\lambda}\left(b_{m\lambda}^\dagger b_{m\lambda} + \frac{1}{2}\right) + \sum_{m\lambda} g_{m\lambda}\omega_{m\lambda}\left(b_{m\lambda}^\dagger + b_{m\lambda}\right) a_m^\dagger a_m, \quad (45)$$

where $a_m^\dagger(a_m)$ is the exciton creation (annihilation) operator on the mth site whose local excitation energy is E_m, J_{mn} is the Coulomb interaction between the mth and nth site, and $b_{m\lambda}^\dagger(b_{m\lambda})$ is the phonon creation (annihilation) operator of vibration mode λ of the mth site with vibration frequency $\omega_{m\lambda}$ and dimensionless electron-phonon coupling strength $g_{m\lambda}$. It has been reported that the harmonic bath models are a reasonable approximation for describing the energy transfer dynamics of the FMO complex.[74] Figure 3 shows a diagram of the model.

We obtain E_m and J_{mn} from the previous literature,[55,75] and each site has the same environment with $\omega_{m\lambda}$ and $g_{m\lambda}$ calculated by equally spaced discretization of the bath spectral density[76] from

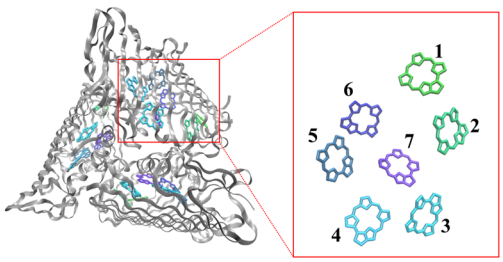

FIG. 3. A diagram of the 7-site FMO model in our calculations.

experiments.[55,77] For each site, 35 vibration modes are discretized, giving 252 DOFs in total (245 vibrational DOFs + 7 electronic DOFs). The dimension of the local basis (harmonic potential eigenbasis) for each mode varies with the frequency, and most modes have 8 or 4 phonon occupation levels. For the vibration mode with the Huang-Rhys factor $S_{m\lambda} = g_{m\lambda}^2$, we choose the number of levels $d_{m\lambda}$ such that for $d'_{m\lambda} > d_{m\lambda}$ the Franck-Condon factor $\mathrm{FC}(0, d'_{m\lambda}) = \frac{S_{m\lambda}^{d'_{m\lambda}}}{d'_{m\lambda}!} e^{-S_{m\lambda}} < 1 \times 10^{-10}$. In principle, numerically exact TD-DMRG calculation should check the convergence of the number of vibration modes discretized from the continuous spectral density and the number of the local basis for each vibration mode. However, such strict convergence is not necessary for the present work as we are more interested in the relative error due to different time evolution schemes rather than the absolute error. To minimize the entanglement, the electronic sites in the MPS chain are arranged in the order of [7, 5, 3, 1, 2, 4, 6]. All the local vibrational DOFs are arranged next to the local electronic site as in our previous work.[13]

We simulate the exciton dynamics at zero temperature. At $t = 0$, an excitation at the electronic site 1 is prepared with all vibrations at their lowest energy level, which composes a Hartree product state. Therefore, for simulation with $M_S > 1$, TDVP-based methods (TDVP-MU and TDVP-PS) should somehow find the $M_S - 1$ renormalized basis to complement the empty matrix elements in the initial MPS. To solve this, in all the simulations, we use P&C-RK4 to propagate the state with $M_{\max} = 256$ and time step size $\tau = 80$ a.u. in the first 10 steps to $t = 800$ a.u. Thereafter, the state is compressed to the predefined bond dimension M_S and then evolves with different evolution schemes. Throughout this paper, we use atomic unit (a.u.) as the unit of time unless otherwise stated and omit the unit of t and τ for simplicity.

III. RESULTS AND DISCUSSIONS

A. Accuracy and time step size

To quantitatively evaluate the relative accuracy in our calculations, the mean cumulative deviation of exciton populations at time t is used,

$$\mathrm{error}(t) = \frac{\sum_{n=1}^{7} \int_0^t |P^{(n)}(t') - P_{\mathrm{ref}}^{(n)}(t')| dt'}{7t}, \tag{46}$$

where $P^{(n)}$ represents the exciton population at the nth site of the 7-site FMO model. $P_{\mathrm{ref}}^{(n)}$ is the reference result obtained from TD-DMRG calculation with an appropriate MPS bond dimension M_S and evolution time step size τ. The standard for choosing M_S and τ is that the exciton populations have converged on both of them. Such convergence on M_S for the TDVP-PS method is illustrated in Fig. 4 where all curves have already converged on τ. The population obtained with $M_S = 32$ has already been semiquantitatively correct although the deviation is visible after $t > 20\,000$. The curves for $M_S = 64$ and $M_S = 128$ almost overlap with each other. To be more discreet, we choose $M_S = 256$ with $\tau = 10$ as the reference for better accuracy. A more comprehensive and quantitative demonstration for the validity of our reference is included in the Appendix. Note that using the reference calculated by TDVP-MU (VMF) with a sufficiently small regularization parameter rather than TDVP-PS has little effect on the results presented in this paper.

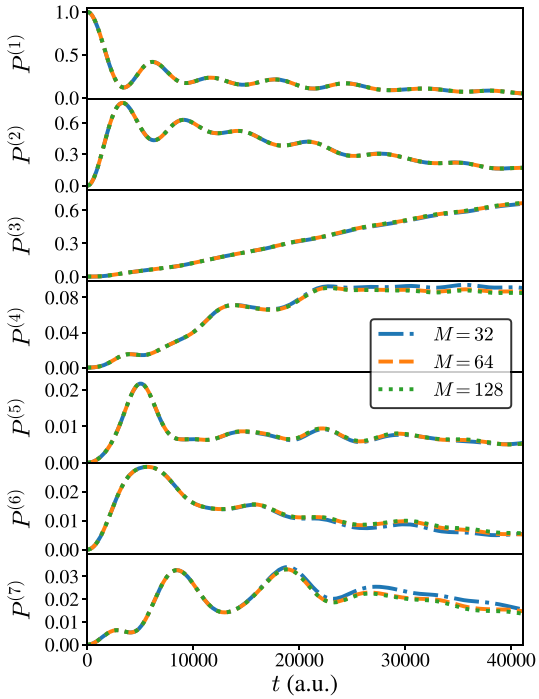

FIG. 4. The exciton population dynamics ($P^{(n)}$ vs t) obtained from $M_S = 32$, $M_S = 64$, and $M_S = 128$ by TDVP-PS. The curves have already converged on step size τ.

The mean cumulative deviation of the three schemes as a function of t is plotted in Fig. 5 with $M_S = 64$. For P&C-RK4, the time step size is 160, which strikes a balance between the RK4 integrator error and the MPS wavefunction compression error. The step size in TDVP-based schemes is chosen such that the error due to the integrator is negligible compared to the restricted M_S. For TDVP-MU (VMF), the step size controlled by the adaptive RK45 integrator (the relative tolerance and absolute tolerance are 10^{-5} and 10^{-8},

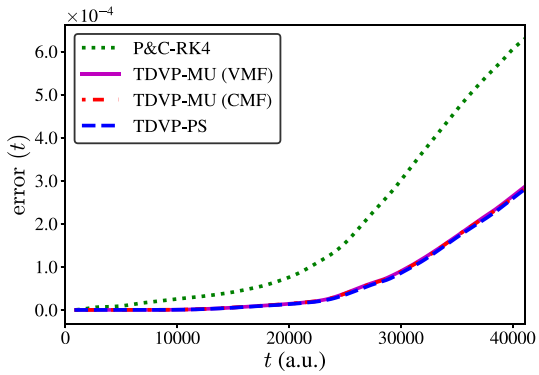

FIG. 5. Mean cumulative deviation as a function of evolution time t of P&C-RK4, TDVP-MU with VMF or CMF, and TDVP-PS to $t = 41\,120$. The bond dimensions M_S are set to 64, and the step sizes are chosen to minimize the error for each scheme.

respectively) spans from 5 to 30. For TDVP-MU (CMF) and TDVP-PS, we choose $\tau = 2$, which is a relatively small value to reduce the error introduced by the CMF approximation or Trotter decomposition. We find that at a long time scale the TDVP based methods exhibit almost the same error, as they share the same starting point during their derivation, while P&C-RK4 guided by a different philosophy shows a larger error than TDVP based methods.

Although TDVP-MU and TDVP-PS are similar in the long time limit, subtle differences between the two schemes when $t < 15\,000$ are observed, which is illustrated in Fig. 6. TDVP-MU (CMF) is not shown in Fig. 6 as it produces the same result with the VMF approach at a small step size limit. Figure 6 shows that TDVP-MU (VMF) is not as accurate as TDVP-PS at short time scales, which is probably caused by the artificial regularization in Eq. (38). Using TDVP-MU (VMF) with a sufficiently small regularization parameter ε as the reference will give the same tendency shown here. When the regularization parameter ε is decreased from 10^{-7} to 10^{-16}, which reduces the invasion to the original EOMs, the accuracy of TDVP-MU (VMF) is improved, but a smaller time step size is required. However, it should be noted that the error caused by the regularization at the short time regime is several orders of magnitude smaller than the error at the long time regime controlled by the MPS approximation of the wavefunction with a restricted M_S, as illustrated in Fig. 5. Thus, in this case, setting ε to be 10^{-10} is more reasonable in terms of practical computation because this setup costs less simulation time than setting ε to be 10^{-13} or 10^{-16} without significant precision deterioration. We also compare the new MU regularization scheme with the original regularization scheme and find that the new MU scheme is more accurate with the same ε as the former studies in (ML-)MCTDH.[64,65]

Despite the fact that TDVP based methods will obtain the same results in the small step size τ limit and small regularization parameter ε limit, they behave differently to the adjustment of τ because their EOMs and integration algorithms differ from each other. Since TDVP-MU (CMF) and TDVP-PS are both second order methods, we investigate how large the time step τ can be used in these two schemes with $M_S = 64$. The upper panel of Fig. 7 illustrates the error at $t_1 = 6560$, which is at the early stage of the evolution. The error of

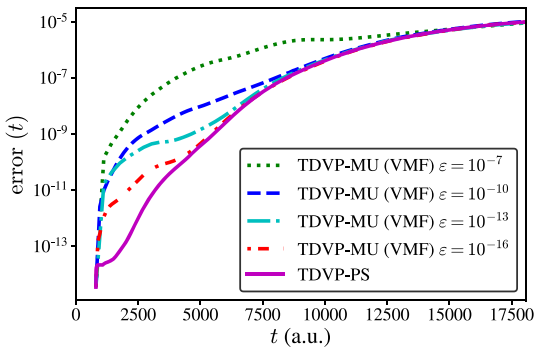

FIG. 6. Mean cumulative deviation as a function of evolution time t of TDVP-MU (VMF) and TDVP-PS to $t = 17\,500$. The y axis is in logarithmic scale. For TDVP-MU (VMF), results with the regularization parameter ε ranging from 10^{-7} to 10^{-16} are shown. The bond dimensions M_S of the two schemes are set to 64, and the step sizes are chosen to minimize the error for each scheme.

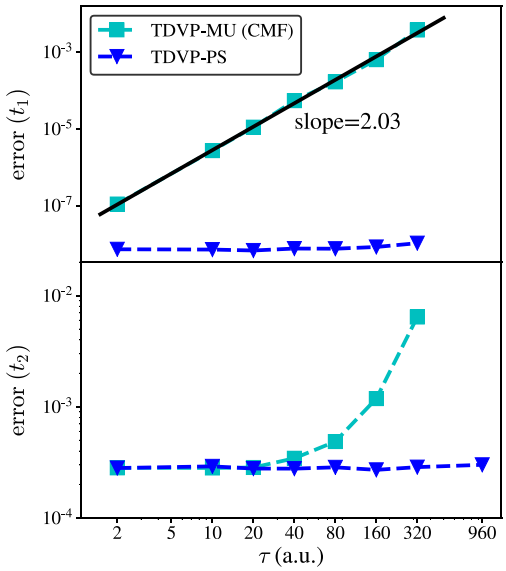

FIG. 7. Mean cumulative deviation evaluated at $t_1 = 6560$ (upper panel) and $t_2 = 41\,120$ (lower panel) for second order evolution schemes TDVP-MU (CMF) and TDVP-PS with various step sizes τ. The bond dimensions of the two schemes are set to 64. In the upper panel, the relation between mean cumulative deviations of TDVP-MU (CMF) and τ is linear in logarithmic scale and the slope of linear fitting is 2.03.

TDVP-MU (CMF) shows a linear relationship with τ under logarithmic scale, and its slope of linear fitting is 2.03, which is in accordance with the second order approximation of TDVP-MU (CMF). It also indicates that for CMF the integrator error is dominant in this time scale. To the contrary, the error of TDVP-PS is insensitive to τ within the range of τ shown in Fig. 7, which means that the error due to second order Trotter decomposition is smaller than the error due to the projection error of the MPS ansatz with $M_S = 64$. The small fluctuations in the error-τ relationship of TDVP-PS is considered to be an artifact of our incomplete error metric defined in Eq. (46) because the style of the fluctuation becomes different if another error metric is used, such as the mean cumulative derivation of a particular site rather than the average of all sites. In the lower panel, the time to measure the error is at $t_2 = 41\,120$ and we observe generally the same trend as in the upper panel, except that when $\tau \leq 20$ the errors of TDVP-MU (CMF) and TDVP-PS are close to each other, indicating that the time step has already converged at this time point and the error is controlled by the MPS ansatz. However, as the upper panel, TDVP-PS allows a time step at least 16 times larger than TDVP-MU (CMF) for a converged result. One of the major reasons is that the widely known stiffness problem in the EOMs of TDVP-MU due to the inverse of the ill-conditioned overlap matrix prevents a larger time step.[64,65] Though the matrix unfolding regularization scheme could partly relieve this problem, it still exists especially when the regularization parameter ε is small. Besides, TDVP-PS combined with the Krylov subspace integrator is both unitary and symplectic and ensures the conservation of the wavefunction norm and total energy regardless of the time step. The same is not true for TDVP-MU (CMF); thus, the conservation properties will be violated during the time evolution, especially with a large time step size. We

have ruled out the possibility that the superiority of TDVP-PS over TDVP-MU (CMF) is a special case of the FMO complex by performing additional benchmarks on the spin-boson model. The details of the result can be found in the supplementary material. Therefore, though both TDVP-PS and TDVP-MU (CMF) implemented here are second order methods, the prefactor of the error term of TDVP-PS is much smaller than that of TDVP-MU (CMF).

For results presented above, TDVP-MU seems to have little advantages against TDVP-PS, but in terms of practical computation, TDVP-MU (VMF) has an appealing advantage that the EOMs derived can be integrated using any general ODE integrator such as RK45 in this work. Therefore, the step size in TDVP-MU (VMF) can be adaptively modified with robust algorithms based on a rigorous mathematical derivation. On the contrary, only a brute-force step size adjustment algorithm is proposed in TDVP-PS,[70] which compares the states obtained from the time step τ to $\tau/2$ to estimate the local error. Unlike the case of adaptive RK or other adaptive ODE integrator where only a small additional cost is required to estimate the local error, this adaptive algorithm in TDVP-PS takes almost twice as much time. The TDVP-MU (CMF) approach does not gain much from the constant mean field approximation in our implementation; however, the CMF approximation results in decoupled EOMs and makes distributed evolution of local sites much easier for large-scale high-performance computing.

B. GPU acceleration

According to the analysis of the computational complexity of the intensive steps in the three time evolution schemes in Sec. II (a summary is shown in Table I), we explore how these steps can be accelerated by a multicore CPU and particularly CPU-GPU heterogeneous computing. Although the time cost of each algorithm should vary with different implementations and there certainly are rooms for optimizations in our codes, we believe that the data presented in this section reflect the correct tendency. Our benchmark platforms are Intel Xeon® CPU E5-2680 v4 at 2.40 GHz for CPU-only calculations and Intel Xeon Gold 5115 CPU at 2.40 GHz with NVIDIA® Tesla® V100-PCIE-32 GB for CPU-GPU heterogeneous calculations. Only one GPU is used in the benchmark.

For the three schemes discussed in this paper with $M_S = 128$, the wall time cost of a single evolution step ($\tau = 160$) as well as the composed substeps is shown in Fig. 8. For TDVP-MU (VMF), there are several steps within τ. We first focus on the total time cost of each scheme in Fig. 8 from which we can learn that for all schemes the parallelization implemented in the standard linear algebra library on the CPU is not optimal for the TD-DMRG. Using 4 cores can merely double the computational speed, and further improvement is not significant up to using all 28 cores of the CPU. A more inspiring fact is that the CPU-GPU heterogeneous computing is able to boost the speed of time evolution up to 40 times for TDVP based algorithms although for P&C-RK4 the effect of the GPU is comparable with 4 cores of the CPU. We note that when doing calculations on the GPU, the GPU usage is not always 100%, so more drastic speed up is expected with even larger bond dimensions. Indeed, when $M_S = 256$, the time costs of TDVP-MU (CMF) on a single core CPU and the heterogeneous CPU-GPU are 4572 s and 63 s, respectively, indicating a 73-fold acceleration.

A closer inspection of the substeps shown in Fig. 8 reveals more details about the acceleration. For P&C-RK4, the bottlenecks of the algorithm are "QR" and "MatMul-QR," as discussed in Sec. II A 1. Although "MatMul-QR" can be efficiently accelerated by a multicore CPU and CPU-GPU heterogeneous computing, the same is not true for "QR" and "SVD," limiting the overall acceleration efficiency. When the calculation runs on the GPU, the matrix multiplication virtually costs no time while "QR" and "SVD" become the bottlenecks. When the GPU is incorporated in TDVP-MU (VMF), the SVD for regularization becomes prominent even though it only scales at $\mathcal{O}(NM_S^3 d)$ in a single step. Thus, it is suggested to return back to the original equation in Eq. (28) for a higher efficiency once the regularization does not take effect anymore, owing to the fact that calculating the inverse of S directly only costs $\mathcal{O}(NM_S^3)$. Even better acceleration efficiency can be achieved in TDVP-MU (CMF) because SVD is carried out less frequently. The "Get Env" part also takes less time in the CMF than VMF due to the same reason. The time costs of bottleneck substeps of TDVP-PS are similar with that of TDVP-MU (CMF) except that the integrator seems to take a large fraction of time especially when running on the GPU. This is considered to be a demonstration of the GPU latency which makes the total time cost of lots of operations on small matrices

TABLE I. Summary of the computational complexity of P&C-RK4, TDVP-MU, and TDVP-PS time evolution schemes.

Substep[a]	P&C-RK4	TDVP-MU	TDVP-PS
$\hat{H}\|\Psi\rangle$	$M_S^2 M_O^2 d^2$	—	—
QR	$M_S^3 M_O^3 d$	—	$M_S^3 d$
MatMul-QR	$M_S^3 M_O^3 d$	—	$M_S^3 d$
SVD	$\min(M_S^3 M_O^2 d, M_S^3 M_O d^2)$	$M_S^3 d$	—
MatMul-SVD	$M_S^3 M_O^2 d$	$M_S^3 d$	—
Get Env	—	$M_S^2 M_O^2 d^2 + M_S^3 M_O d$	$M_S^2 M_O^2 d^2 + M_S^3 M_O d$
Deriv	—	$M_S^2 M_O^2 d^2 + M_S^3 M_O d + M_S^3 d^2$	$M_S^2 M_O^2 d^2 + M_S^3 M_O d$ (forward) $M_S^3 M_O$ (backward)
Overall	$NM_S^3 M_O^3 d$	$N(M_S^2 M_O^2 d^2 + M_S^3 M_O d + M_S^3 d^2)$ (VMF) $Nf(M_S^2 M_O^2 d^2 + M_S^3 M_O d + M_S^3 d^2)$ (CMF)	$Nf(M_S^2 M_O^2 d^2 + M_S^3 M_O d)$

[a]See the text in Sec. II for the definitions and labels of the sub-steps.

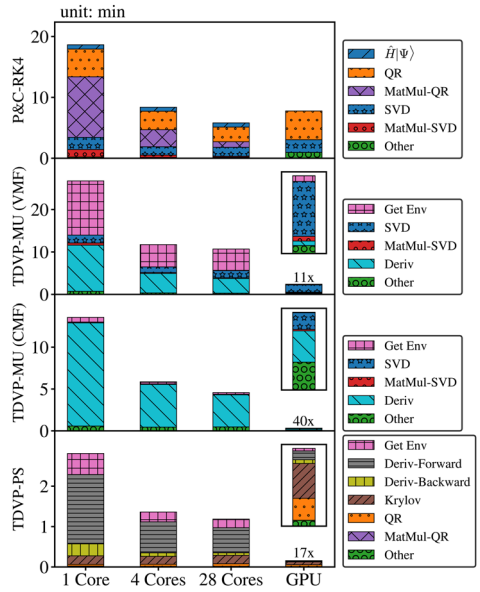

FIG. 8. The time cost of a single evolution step and its intensive substeps with M_S = 128 and τ = 160. The bars for CPU-GPU heterogeneous computation are further shown in the insets for clarity. (See the text in Sec. II for the definitions and labels of the substeps.)

larger than the total time cost of a few operations on large matrices. When M_S is increased to 256, the relative time cost of the "Krylov" part is diminished. The exact abnormality with GPU acceleration is also manifested to a lesser extent in TDVP-MU (CMF) where the time-consuming "other" part is primarily composed of operations in the Runge-Kutta integrator. Hence, if a smaller bond dimension is employed, devoting more resource on the computation will not gain much benefit. In fact, for TDVP based schemes, the absolute wall time with the GPU is roughly the same between M_S = 128 and M_S = 32. For example, the time costs for a single step of evolution in TDVP-PS with M_S = 128 and M_S = 32 are 10.8 s and 7.9 s, respectively. This phenomenon implies that when $M_S \leq 128$ the latency of GPU calculation is of the same order with the actual computational time cost. Therefore, with the CPU-GPU heterogeneous computation, TD-DMRG simulation with $M_S > 100$ for a large-scale system would become a routine task in the future.

IV. CONCLUSION

To summarize, in this paper, we carry out numerical benchmarks on three different TD-DMRG time evolution schemes, which are P&C-RK4, TDVP-MU, and TDVP-PS, in terms of accuracy based on the vibronic coupling model represented by the FMO complex. After defining the tangent space projector for a general noncanonical MPS, the EOMs of TDVP-MU and TDVP-PS are rederived from the same starting point but with different gauge conditions. The numerical results demonstrate that TDVP-MU and TDVP-PS could indeed obtain similar accuracy with a converged time step size τ, while P&C-RK4 is not as accurate as them. However, surprisingly, the converged time step size of TDVP-PS is at least 16 times larger than that of TDVP-MU (CMF) although both of them implemented in this work are second order methods. Regarding the computational efficiency, we first analyze the complexity of each evolution scheme. To further accelerate the intensive tensor computations in the TD-DMRG, we adopt a CPU-GPU heterogeneous computing strategy in which the CPU and GPU are, respectively, responsible for the tensor decomposition and contraction. We find that for the TDVP-based schemes where the tensor contraction is the main bottleneck, this heterogeneous computing approach is able to speed up these two schemes by up to 73 times (M_S = 256). Taking both the accuracy and efficiency into consideration, TDVP-PS is highly recommended for the vibronic coupling problems.

Though in this work we mainly focus on the vibronic model, the TD-DMRG has also been used in the ultrafast electron dynamics with *ab initio* quantum chemistry Hamiltonian[14,15] and is highly possible to be used in *ab initio* quantum molecular dynamics in the future. These problems are more complicated than the model problem, for example, the dimensions of the MPS and MPO are usually much larger in the former problem and even an explicit MPO could hardly be constructed in the latter problem. However, we still speculate that TDVP-PS should be the better evolution scheme and GPU acceleration would be more essential in these problems.

Finally, we conclude that with the accurate and efficient TDVP-based evolution schemes (especially TDVP-PS) and CPU-GPU heterogeneous algorithm/hardware, the TD-DMRG has been a promising method for the large-scale quantum dynamics simulation of real chemical and physical problems.

SUPPLEMENTARY MATERIAL

See the supplementary material for the benchmark results between TDVP-MU (CMF) and TDVP-PS based on the spin-boson model from the weak to strong coupling regime.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China through the project "Science CEnter for Luminescence from Molecular Aggregates (SCELMA)," Grant No. 21788102, as well as by the Ministry of Science and Technology of China through the National Key R&D Plan, Grant No. 2017YFA0204501. J.R. is also supported by the Shuimu Tsinghua Scholar Program. The authors are indebted to Professor Garnet Chan for the stimulating discussions. The authors also gratefully thank Mr. Hongde Yu for support on the crystal structure of the FMO complex.

APPENDIX: THE VALIDITY OF THE REFERENCE STATE

Here, we present the data for the numerical convergence of the reference obtained by TDVP-PS with M_S = 256 and τ = 10. The mean cumulative deviation with various M_S and τ = 10 at t_1 = 6560 and t_2 = 41 120 is shown in Table II, while the same data for M_S = 256 and various τ are shown in Table III.

From Tables II and III, we can see that the reference is accurate to a level of approximately 10^{-11} for a short time scale (t_1 = 6560) and 10^{-6} for a long time scale (t_2 = 41 120), which is considered to be accurate enough for the error scales discussed in the main text.

TABLE II. The mean cumulative deviation with various M_S and τ = 10 by TDVP-PS at t_1 = 6560 and t_2 = 41 120 (reference: M_S = 256, τ = 10).

	Error (t)	
M_S	$t_1 = 6560$	$t_2 = 41\,120$
32	6×10^{-7}	1×10^{-3}
64	8×10^{-9}	3×10^{-4}
128	3×10^{-10}	3×10^{-5}
192	3×10^{-11}	7×10^{-6}

TABLE III. The mean cumulative deviation with M_S = 256 and various τ by TDVP-PS at t_1 = 6560 and t_2 = 41 120 (reference: M_S = 256, τ = 10).

	Error (t)	
τ	$t_1 = 6560$	$t_2 = 41\,120$
320	4×10^{-10}	8×10^{-7}
160	7×10^{-11}	7×10^{-7}
80	9×10^{-13}	8×10^{-7}
40	7×10^{-13}	6×10^{-7}

REFERENCES

[1] H. Ma, Z. Luo, and Y. Yao, "The time-dependent density matrix renormalisation group method," Mol. Phys. **116**, 854–868 (2018).
[2] S. Paeckel, T. Köhler, A. Swoboda, S. R. Manmana, U. Schollwöck, and C. Hubig, "Time-evolution methods for matrix-product states," Ann. Phys. **411**, 167998 (2019).
[3] S. R. White, "Density matrix formulation for quantum renormalization groups," Phys. Rev. Lett. **69**, 2863 (1992).
[4] S. R. White, "Density-matrix algorithms for quantum renormalization groups," Phys. Rev. B **48**, 10345 (1993).
[5] Z. Shuai, J. Brédas, A. Saxena, and A. Bishop, "Linear and nonlinear optical response of polyenes: A density matrix renormalization group study," J. Chem. Phys. **109**, 2549–2555 (1998).
[6] C. Zhang, E. Jeckelmann, and S. R. White, "Dynamical properties of the one-dimensional Holstein model," Phys. Rev. B **60**, 14092 (1999).
[7] Y. Yao, "Polaronic quantum diffusion in dynamic localization regime," New J. Phys. **19**, 043015 (2017).
[8] W. Barford and J. R. Mannouch, "Torsionally induced exciton localization and decoherence in π-conjugated polymers," J. Chem. Phys. **149**, 214107 (2018).
[9] S. M. Greene and V. S. Batista, "Tensor-train split-operator fourier transform (TT-SOFT) method: Multidimensional nonadiabatic quantum dynamics," J. Chem. Theory Comput. **13**, 4034–4042 (2017).
[10] A. Baiardi and M. Reiher, "Large-scale quantum dynamics with matrix product states," J. Chem. Theory Comput. **15**, 3481–3498 (2019).
[11] X. Xie, Y. Liu, Y. Yao, U. Schollwöck, C. Liu, and H. Ma, "Time-dependent density matrix renormalization group quantum dynamics for realistic chemical systems," J. Chem. Phys. **151**, 224101 (2019).
[12] Y. Yao, K.-W. Sun, Z. Luo, and H. Ma, "Full quantum dynamics simulation of a realistic molecular system using the adaptive time-dependent density matrix renormalization group method," J. Phys. Chem. Lett. **9**, 413–419 (2018).
[13] J. Ren, Z. Shuai, and G. K.-L. Chan, "Time-dependent density matrix renormalization group algorithms for nearly exact absorption and fluorescence spectra of molecular aggregates at both zero and finite temperature," J. Chem. Theory Comput. **14**, 5027–5039 (2018).
[14] E. Ronca, Z. Li, C. A. Jimenez-Hoyos, and G. K.-L. Chan, "Time-step targeting time-dependent and dynamical density matrix renormalization group algorithms with ab initio Hamiltonians," J. Chem. Theory Comput. **13**, 5560–5571 (2017).
[15] L.-H. Frahm and D. Pfannkuche, "Ultrafast ab initio quantum chemistry using matrix product states," J. Chem. Theory Comput. **15**, 2154–2165 (2019).
[16] A. Chin, J. Prior, R. Rosenbach, F. Caycedo-Soler, S. Huelga, and M. B. Plenio, "The role of non-equilibrium vibrational structures in electronic coherence and recoherence in pigment–protein complexes," Nat. Phys. **9**, 113 (2013).
[17] F. A. Schröder, D. H. Turban, A. J. Musser, N. D. Hine, and A. W. Chin, "Tensor network simulation of multi-environmental open quantum dynamics via machine learning and entanglement renormalisation," Nat. Commun. **10**, 1062 (2019).
[18] R. Borrelli and M. F. Gelin, "Simulation of quantum dynamics of excitonic systems at finite temperature: An efficient method based on thermo field dynamics," Sci. Rep. **7**, 9127 (2017).
[19] B. Kloss and Y. B. Lev, "Spin transport in a long-range-interacting spin chain," Phys. Rev. A **99**, 032114 (2019).
[20] B. Kloss, D. R. Reichman, and R. Tempelaar, "Multiset matrix product state calculations reveal mobile Franck-Condon excitations under strong Holstein-type coupling," Phys. Rev. Lett. **123**, 126601 (2019).
[21] G. Vidal, "Efficient simulation of one-dimensional quantum many-body systems," Phys. Rev. Lett. **93**, 040502 (2004).
[22] S. R. White and A. E. Feiguin, "Real-time evolution using the density matrix renormalization group," Phys. Rev. Lett. **93**, 076401 (2004).
[23] A. J. Daley, C. Kollath, U. Schollwöck, and G. Vidal, "Time-dependent density-matrix renormalization-group using adaptive effective Hilbert spaces," J. Stat. Mech. Theory Exp. **2004**, P04005.
[24] M. P. Zaletel, R. S. Mong, C. Karrasch, J. E. Moore, and F. Pollmann, "Time-evolving a matrix product state with long-ranged interactions," Phys. Rev. B **91**, 165112 (2015).
[25] J. J. García-Ripoll, "Time evolution of matrix product states," New J. Phys. **8**, 305 (2006).
[26] J. C. Halimeh, F. Kolley, and I. P. McCulloch, "Chebyshev matrix product state approach for time evolution," Phys. Rev. B **92**, 115130 (2015).
[27] M. Wall and L. D. Carr, "Out-of-equilibrium dynamics with matrix product states," New J. Phys. **14**, 125015 (2012).
[28] A. E. Feiguin and S. R. White, "Time-step targeting methods for real-time dynamics using the density matrix renormalization group," Phys. Rev. B **72**, 020404 (2005).
[29] T. Dutta and S. Ramasesha, "Double time window targeting technique: Real-time DMRG dynamics in the Pariser-Parr-Pople model," Phys. Rev. B **82**, 035115 (2010).
[30] P. A. Dirac, "Note on exchange phenomena in the Thomas atom," in *Mathematical Proceedings of the Cambridge Philosophical Society* (Cambridge University Press, 1930), Vol. 26, pp. 376–385.
[31] J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pižorn, H. Verschelde, and F. Verstraete, "Time-dependent variational principle for quantum lattices," Phys. Rev. Lett. **107**, 070601 (2011).
[32] J. Haegeman, C. Lubich, I. Oseledets, B. Vandereycken, and F. Verstraete, "Unifying time evolution and optimization with matrix product states," Phys. Rev. B **94**, 165116 (2016).
[33] J. Prior, A. W. Chin, S. F. Huelga, and M. B. Plenio, "Efficient simulation of strong system-environment interactions," Phys. Rev. Lett. **105**, 050404 (2010).
[34] E. M. Stoudenmire and S. R. White, "Minimally entangled typical thermal state algorithms," New J. Phys. **12**, 055026 (2010).
[35] Y. Kurashige, "Matrix product state formulation of the multiconfiguration time-dependent Hartree theory," J. Chem. Phys. **149**, 194114 (2018).
[36] J. Hauschild and F. Pollmann, "Efficient numerical simulations with tensor networks: Tensor Network Python (TeNPy)," SciPost Phys. Lect. Notes **2018**, 5.
[37] C. B. Mendl, "PyTeNet: A concise Python implementation of quantum tensor network algorithms," J. Open Source Software **3**, 948 (2018).
[38] G. K.-L. Chan, "An algorithm for large scale density matrix renormalization group calculations," J. Chem. Phys. **120**, 3172–3178 (2004).

[39]Y. Kurashige and T. Yanai, "High-performance *ab initio* density matrix renormalization group method: Applicability to large-scale multireference problems for metal compounds," J. Chem. Phys. **130**, 234114 (2009).
[40]C. Song and T. J. Martínez, "Analytical gradients for tensor hyper-contracted MP2 and SOC-MP2 on graphical processing units," J. Chem. Phys. **147**, 161723 (2017).
[41]I. A. Kaliman and A. I. Krylov, "New algorithm for tensor contractions on multi-core CPUs, GPUs, and accelerators enables CCSD and EOM-CCSD calculations with over 1000 basis functions on a single compute node," J. Comput. Chem. **38**, 842–853 (2017).
[42]J. Shee, E. J. Arthur, S. Zhang, D. R. Reichman, and R. A. Friesner, "Phaseless auxiliary-field quantum Monte Carlo on graphical processing units," J. Chem. Theory Comput. **14**, 4109–4121 (2018).
[43]F. Liu, D. M. Sanchez, H. J. Kulik, and T. J. Martínez, "Exploiting graphical processing units to enable quantum chemistry calculation of large solvated molecules with conductor-like polarizable continuum models," Int. J. Quantum Chem. **119**, e25760 (2019).
[44]T. Penfold, "Accelerating direct quantum dynamics using graphical processing units," Phys. Chem. Chem. Phys. **19**, 19601–19608 (2017).
[45]T.-S. Lee, D. S. Cerutti, D. Mermelstein, C. Lin, S. LeGrand, T. J. Giese, A. Roitberg, D. A. Case, R. C. Walker, and D. M. York, "GPU-accelerated molecular dynamics and free energy methods in Amber18: Performance enhancements and new features," J. Chem. Inf. Model. **58**, 2043–2050 (2018).
[46]C. Kreisbeck, T. Kramer, M. Rodríguez, and B. Hein, "High-performance solution of hierarchical equations of motion for studying energy transfer in light-harvesting complexes," J. Chem. Theory Comput. **7**, 2166–2174 (2011).
[47]M. Tsuchimoto and Y. Tanimura, "Spins dynamics in a dissipative environment: Hierarchal equations of motion approach using a graphics processing unit (GPU)," J. Chem. Theory Comput. **11**, 3859–3865 (2015).
[48]C. Nemes, G. Barcza, Z. Nagy, Ö. Legeza, and P. Szolgay, "The density matrix renormalization group algorithm on kilo-processor architectures: Implementation and trade-offs," Comput. Phys. Commun. **185**, 1570–1581 (2014).
[49]G. S. Engel, T. R. Calhoun, E. L. Read, T.-K. Ahn, T. Mančal, Y.-C. Cheng, R. E. Blankenship, and G. R. Fleming, "Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems," Nature **446**, 782 (2007).
[50]Y.-C. Cheng and G. R. Fleming, "Dynamics of light harvesting in photosynthesis," Annu. Rev. Phys. Chem. **60**, 241–262 (2009).
[51]P. Nalbach, D. Braun, and M. Thorwart, "Exciton transfer dynamics and quantumness of energy transfer in the Fenna-Matthews-Olson complex," Phys. Rev. E **84**, 041926 (2011).
[52]A. Ishizaki and G. R. Fleming, "Theoretical examination of quantum coherence in a photosynthetic system at physiological temperature," Proc. Natl. Acad. Sci. U. S. A. **106**, 17255–17260 (2009).
[53]Q. Shi, Y. Xu, Y. Yan, and M. Xu, "Efficient propagation of the hierarchical equations of motion using the matrix product state method," J. Chem. Phys. **148**, 174102 (2018).
[54]J. Schulze and O. Kuhn, "Explicit correlated exciton-vibrational dynamics of the FMO complex," J. Phys. Chem. B **119**, 6211–6216 (2015).
[55]J. Schulze, M. F. Shibl, M. J. Al-Marri, and O. Kühn, "Multi-layer multi-configuration time-dependent Hartree (ML-MCTDH) approach to the correlated exciton-vibrational dynamics in the fmo complex," J. Chem. Phys. **144**, 185101 (2016).
[56]U. Schollwöck, "The density-matrix renormalization group in the age of matrix product states," Ann. Phys. **326**, 96–192 (2011).
[57]G. K.-L. Chan, A. Keselman, N. Nakatani, Z. Li, and S. R. White, "Matrix product operators, matrix product states, and *ab initio* density matrix renormalization group algorithms," J. Chem. Phys. **145**, 014102 (2016).
[58]F. Gatti, B. Lasorne, H.-D. Meyer, and A. Nauts, *Applications of Quantum Dynamics in Chemistry* (Springer, 2017), Vol. 98.
[59]S. Wouters, N. Nakatani, D. Van Neck, and G. K.-L. Chan, "Thouless theorem for matrix product states and subsequent post density matrix renormalization group methods," Phys. Rev. B **88**, 075122 (2013).
[60]H.-D. Meyer, U. Manthe, and L. S. Cederbaum, "The multi-configurational time-dependent Hartree approach," Chem. Phys. Lett. **165**, 73–78 (1990).
[61]M. H. Beck, A. Jäckle, G. A. Worth, and H.-D. Meyer, "The multiconfiguration time-dependent Hartree (MCTDH) method: A highly efficient algorithm for propagating wavepackets," Phys. Rep. **324**, 1–105 (2000).
[62]H. Wang and M. Thoss, "Multilayer formulation of the multiconfiguration time-dependent Hartree theory," J. Chem. Phys. **119**, 1289–1299 (2003).
[63]M. H. Beck and H.-D. Meyer, "An efficient and robust integration scheme for the equations of motion of the multiconfiguration time-dependent Hartree (MCTDH) method," Z. Phys. D: At., Mol. Clusters **42**, 113–129 (1997).
[64]H.-D. Meyer and H. Wang, "On regularizing the MCTDH equations of motion," J. Chem. Phys. **148**, 124105 (2018).
[65]H. Wang and H.-D. Meyer, "On regularizing the ML-MCTDH equations of motion," J. Chem. Phys. **149**, 044119 (2018).
[66]C. Lubich, I. V. Oseledets, and B. Vandereycken, "Time integration of tensor trains," SIAM J. Numer. Anal. **53**, 917–941 (2015).
[67]M. Hochbruck and C. Lubich, "On Krylov subspace approximations to the matrix exponential operator," SIAM J. Numer. Anal. **34**, 1911–1925 (1997).
[68]W. E. Arnoldi, "The principle of minimized iterations in the solution of the matrix eigenvalue problem," Q. Appl. Math. **9**, 17–29 (1951).
[69]C. Lubich, "Time integration in the multiconfiguration time-dependent Hartree method of molecular quantum dynamics," Appl. Math. Res. Express **2015**, 311–328.
[70]B. Kloss, I. Burghardt, and C. Lubich, "Implementation of a novel projector-splitting integrator for the multi-configurational time-dependent Hartree approach," J. Chem. Phys. **146**, 174107 (2017).
[71]M. Bonfanti and I. Burghardt, "Tangent space formulation of the multi-configuration time-dependent Hartree equations of motion: The projector–splitting algorithm revisited," Chem. Phys. **515**, 252–261 (2018).
[72]T. Holstein, "Studies of polaron motion: Part II. The "small" polaron," Ann. Phys. **8**, 343–389 (1959).
[73]M. Schröter, S. D. Ivanov, J. Schulze, S. P. Polyutov, Y. Yan, T. Pullerits, and O. Kühn, "Exciton–vibrational coupling in the dynamics and spectroscopy of Frenkel excitons in molecular aggregates," Phys. Rep. **567**, 1–78 (2015).
[74]C. W. Kim, B. Choi, and Y. M. Rhee, "Excited state energy fluctuations in the Fenna–Matthews–Olson complex from molecular dynamics simulations with interpolated chromophore potentials," Phys. Chem. Chem. Phys. **20**, 3310–3319 (2018).
[75]J. Moix, J. Wu, P. Huo, D. Coker, and J. Cao, "Efficient energy transfer in light-harvesting systems, III: The influence of the eighth bacteriochlorophyll on the dynamics and efficiency in FMO," J. Phys. Chem. Lett. **2**, 3045–3052 (2011).
[76]I. de Vega, U. Schollwöck, and F. A. Wolf, "How to discretize a quantum bath for real-time evolution," Phys. Rev. B **92**, 155126 (2015).
[77]M. Wendling, T. Pullerits, M. A. Przyjalgowski, S. I. Vulto, T. J. Aartsma, R. van Grondelle, and H. van Amerongen, "Electron–vibrational coupling in the Fenna–Matthews–Olson complex of prosthecochloris a estuarii determined by temperature-dependent absorption and fluorescence line-narrowing measurements," J. Phys. Chem. B **104**, 5825–5831 (2000).