JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2018, 14, 5645−5652

pubs.acs.org/JCTC

Fast Real-Time Time-Dependent Density Functional Theory Calculations with the Parallel Transport Gauge

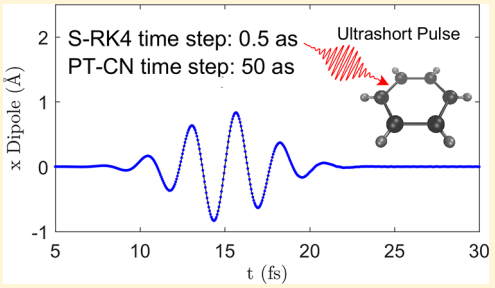
Weile Jia,[†] Dong An,[†] Lin-Wang Wang,[‡] and Lin Lin*[,†,§]

[†]Department of Mathematics, University of California, Berkeley, California 94720, United States

[‡]Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States

[§]Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States

ABSTRACT: Real-time time-dependent density functional theory (RT-TDDFT) is known to be hindered by the very small time step (attosecond or smaller) needed in the numerical simulation, because of the fast oscillation of electron wave functions, which significantly limits its range of applicability for the study of ultrafast dynamics. In this paper, we demonstrate that such oscillation can be considerably reduced by optimizing the gauge choice using the parallel transport formalism. RT-TDDFT calculations can thus be significantly accelerated using a combination of the parallel transport gauge and implicit integrators, and the resulting scheme can be used to accelerate any electronic structure software that uses a Schrödinger representation. Using absorption spectrum, ultrashort laser pulse, and Ehrenfest dynamics calculations for example, we show that the new method can utilize a time step that is on the order of 10−100 attoseconds using a planewave basis set. Thanks to the significant increase of the size of the time step, we also demonstrate that the new method is more than 10 times faster, in terms of the wall clock time, when compared to the standard explicit fourth-order Runge−Kutta time integrator for silicon systems ranging from 32 to 1024 atoms.



1. INTRODUCTION

Recent developments of ultrafast laser techniques have enabled a large number of excited-state phenomena to be observed in real time. One of the most widely used techniques for studying ultrafast properties is the real-time time-dependent density functional theory (RT-TDDFT),[1−5] which has achieved successes in many fields, including, e.g., nonlinear optical response[6] and the collision of an ion with a substrate.[7] Nonetheless, the range of applicability of RT-TDDFT is often hindered by the very small time step needed to propagate the Schrödinger equation. Many numerical propagators used in practice are explicit time integrators,[8−10] which require a small time step size satisfying $\Delta t \lesssim \|H\|^{-1}$, because of the stability restriction. For H discretized under a flexible basis set such as planewaves, the required time step is often <1 as. On the other hand, ultrafast properties often need to be observed on the order of $10-10^3$ fs. This requires 10^4-10^6 time steps to be performed and is often prohibitively expensive. Given the recent emphasis on ultrafast physics, this is, thus, an urgent problem to be solved.

However, physical observables such as the electron density are squared quantities of the wave functions, and they often oscillate much slower. In this paper, we find that such a gap is largely due to the nonoptimal gauge choice of the Schrödinger dynamics, which is irrelevant to the computation of physical observables. We propose that the optimal gauge choice is given by a parallel transport formulation. Compared to the Schrödinger representation, the orbitals with the parallel transport gauge can often be "flattened" into an approximate straight line over a much longer time interval. When combined with implicit time integrators to propagate the parallel transport dynamics, it is possible to significantly increase the time step size without sacrificing accuracy. The parallel transport formulation only introduces one extra term to the Schrödinger equation, and thus can be easily applied to any electronic structure software packages for RT-TDDFT calculations, which is unlike other methods, where approximations and some significant amount of rewriting of the code are needed.[11,12] We remark that RT-TDDFT calculations with local and semilocal exchange correlation functionals can achieve quadratic scaling,[5,8,13] with respect to the system size, while the computational cost associated with the extra term in the parallel transport formulation scales cubically. This term will eventually dominate the computational cost. Nonetheless, we demonstrate that the new method can be more than 10 times faster than the standard explicit fourth-order Runge−Kutta time integrator for silicon systems ranging from 32 to 1024 atoms, and we estimate that the crossover point is ∼4000 atoms.

2. THEORY

In order to derive the parallel transport gauge, let us first consider the RT-TDDFT equation:

$$\imath\partial_t\psi_i(t) = H(t, P(t))\psi_i \quad i = 1, ..., N_e \tag{1}$$

Received: June 11, 2018
Published: October 23, 2018

DOI: 10.1021/acs.jctc.8b00580
J. Chem. Theory Comput. 2018, 14, 5645−5652

Here, $\Psi(t) = [\psi_1, ..., \psi_{N_e}]$ are the electron orbitals, and the Hamiltonian can be dependent explicitly on t and nonlinearly on the density matrix $P(t) = \Psi(t)\Psi^*(t)$ or the electron density $\rho(t) = \sum_{i=1}^{N_e}|\psi_i(t)|^2$. Equation 1 can be equivalently written using a set of transformed orbitals $\Phi(t) = \Psi(t)U(t)$, where the gauge matrix $U(t)$ is a unitary matrix of size N_e. An important property of the density matrix is that it is gauge-invariant: $P(t) = \Psi(t)\Psi^*(t) = \Phi(t)\Phi^*(t)$, and always satisfies the von Neumann equation (or quantum Liouville equation)

$$\imath\partial_t P = [H, P] = HP - PH \tag{2}$$

Our goal is to optimize the gauge matrix, so that the transformed orbitals $\Phi(t)$ vary *as slowly as possible*, without altering the density matrix. This results in the following variational problem:

$$\min_{U(t)} \|\dot{\Phi}\|_F^2 \tag{3}$$

subject to

$$\Phi(t) = \Psi(t)U(t),\ U^*(t)U(t) = I_{N_e}$$

Here, $\|\dot{\Phi}\|_F^2$ measures the Frobenius norm of the time derivative of the transformed orbitals ($\|\dot{\Phi}\|_F^2 := \mathrm{Tr}[\dot{\Phi}^*\dot{\Phi}]$). The minimizer of eq 3, in terms of Φ, satisfies the expression

$$P\dot{\Phi} = 0 \tag{4}$$

(see Appendix A for its derivation).

Equation 4 implicitly defines a gauge choice for each $U(t)$, and this gauge is called the *parallel transport gauge*. The governing equation of each transformed orbital φ_i can be concisely written as

$$\imath\partial_t\varphi_i = H\varphi_i - \sum_{j=1}^{N_e}\varphi_j\langle\varphi_j|H|\varphi_i\rangle \qquad i = 1, ..., N_e \tag{5}$$

or, more concisely, in the matrix form,

$$\imath\partial_t\Phi = H\Phi - \Phi(\Phi^*H\Phi) \qquad P(t) = \Phi(t)\Phi^*(t) \tag{6}$$

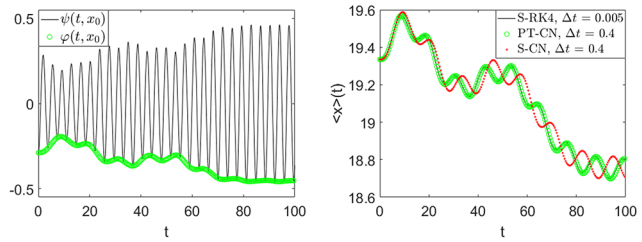
The right-hand side of eq 6 is analogous to the residual vectors of an eigenvalue problem in the time-independent setup. Hence, $\Phi(t)$ follows the dynamics driven by residual vectors and is expected to oscillate less frequently than $\Psi(t)$.

The name "parallel transport gauge" originates from the parallel transport formulation associated with a family of density matrices $P(t)$, which generates a parallel transport evolution operator[14,15] $\mathcal{T}(t)$ as

$$\imath\partial_t\mathcal{T} = [\imath\partial_t P, P]\mathcal{T} \qquad \mathcal{T}(0) = I \tag{7}$$

We demonstrate that, starting from an initial set of orbitals Ψ_0, the solution to the parallel transport dynamics (eq 6) is simply evolved by the parallel transport evolution operator. Mathematically speaking, we need to show that $\Phi(t) = \mathcal{T}(t)\Psi_0$, where Φ is defined according to eq 6. The detailed derivation, together with its theoretical analysis in the *near adiabatic regime* (i.e., when the right-hand side of (eq 6) approximately vanishes) is reported a separate publication.[16]

Figure 1a demonstrates a simple example with one electron, and a time-dependent Hamiltonian, $H(t) = -\frac{1}{2}\partial_x^2 + V(x, t)$ in one dimension. The initial state is the ground state of $H(0)$. Here, the time-dependent potential is chosen to be $V(x,t) = -2\exp[-0.1(x-R(t))^2] - 2\exp[-0.1(x-12.5)^2]$, which is a double-well potential with one fixed center at 12.5 and one moving center $R(t) = 25 + 1.5\exp(-0.0025(t-10)^2) + \exp(-0.0025(t-50)^2)$. Figure 1a shows that, while $\psi(t)$ oscillates rapidly, the oscillation of the parallel transport orbital $\varphi(t)$ is significantly slower, and therefore can be approximated by a straight line over a much larger interval. We remark that efficient numerical methods based on the construction of instantaneous adiabatic states have also been recently developed for the near adiabatic regime.[11,12] The advantage of the parallel transport dynamics is that it only operates on N_e orbitals as in the original Schrödinger dynamics. Even outside the near adiabatic regime, eq 6 always yields the slowest possible dynamics, because of the *variational principle* in eq 3. However, when the frequency of the external field becomes sufficiently high, the oscillation of the density matrix becomes fast as well, which will limit the size of the time step that can be utilized in the simulation.



(a) Real parts of $\psi(x_0, t)$ and $\varphi(x_0, t)$ at $x_0 = 25.0$ (b) Dipole moment

Figure 1. Comparison of the Schrödinger and parallel transport dynamics for the 1D example.

In order to propagate the parallel transport dynamics numerically, all the RT-TDDFT propagation methods can be used since eq 6 only differs from eq 1 by one extra term, $\Phi(\Phi^*H\Phi)$. As an example, the parallel transport Crank–Nicolson scheme (PT-CN) gives rise to the following scheme:

$$\Phi_{n+1} + \imath\frac{\Delta t}{2}\{H_{n+1}\Phi_{n+1} - \Phi_{n+1}(\Phi_{n+1}^*H_{n+1}\Phi_{n+1})\} = \Phi_n - \imath\frac{\Delta t}{2}\{H_n\Phi_n - \Phi_n(\Phi_n^*H_n\Phi_n)\} \tag{8}$$

Here, $H_n = H(t_n, P_n)$ is the Hamiltonian at the time step t_n, and $t_{n+1} = t_n + \Delta t$. Other time integrators can be straightforwardly generalized to the parallel transport dynamics as well (see Appendix B). In eq 8, the solution Φ_{n+1} must be solved self-consistently. This is a set of nonlinear equations, with respect to the unknowns Φ_{n+1}, and can be efficiently solved, e.g., by the preconditioned Anderson mixing scheme.[17] The propagation of $\Phi(t)$ can also be naturally combined with the motion of nuclei discretized, e.g., by the Verlet scheme for the simulation of Ehrenfest dynamics.[18]

Since the parallel transport dynamics yields the optimal gauge choice, it can be used to improve the accuracy and efficiency of any propagator currently applied to the Schrödinger dynamics. For example, let us first consider again the one-dimensional Schrödinger equation. Table 1

Table 1. Error at $T = 100$ for the 1D Example

method	Δt	error
S-RK4	0.01	3.76×10^{-7}
PT-RK4	0.01	4.36×10^{-10}
S-RK4	0.005	2.35×10^{-8}
PT-RK4	0.005	2.73×10^{-11}

DOI: 10.1021/acs.jctc.8b00580
J. Chem. Theory Comput. 2018, 14, 5645–5652

compares the explicit fourth-order Runge−Kutta scheme for the Schrödinger dynamics (S-RK4) and the parallel transport dynamics (PT-RK4), which indicates that the error of the latter is considerably smaller. When combined with implicit integrators, we can further significantly increase the time step size. We compare the PT-CN scheme with the standard CN scheme for the Schrödinger dynamics (S-CN), and use S-RK4 with a small time step $\Delta t = 0.005$ as the benchmark. The accuracy is measured by the dipole moment $\langle x \rangle(t) = \mathrm{Tr}[xP(t)]$ along the trajectory. Figure 1b indicates that PT-CN can use a much larger time step than S-RK4 without losing accuracy. On the other hand, while S-CN can still be numerically stable with the same step size, it becomes significantly less accurate after the first peak at $t \approx 10$.

We remark that the computational complexity of standard RT-TDDFT calculations can achieve $O(N_e^2)$ scaling.[5,8,13] This assumes local and semilocal exchange-correlation functionals and certain explicit time integrators are used, and no orbital reorthogonalization step is needed throughout the simulation. The $O(N_e^2)$ complexity allows RT-TDDFT calculations to be performed for very large systems consisting of more than 6000 atoms.[19] The PT dynamics requires the evaluation of the residual in eq 6 and, thus, the term $\Phi(\Phi^* H\Phi)$, which inherently scales as $O(N_e^3)$. Therefore, we may expect that for sufficiently large system sizes, the $O(N_e^2)$ methods will be faster. On the other hand, numerical results in section 3 indicate that, for systems up to 1000 atoms (with ∼2000 orbitals, which covers a vast range of practical applications of RT-TDDFT), the cubic scaling components are still not yet dominating the computational cost, and the PT dynamics can be significantly faster than standard explicit time integrators. Furthermore, when hybrid exchange-correlation functionals are used, both standard RT-TDDFT calculations and the PT dynamics will scale as $O(N_e^3)$, because of the Fock exchange term, and we expect that the advantage of the PT formulation will become even more evident.

3. NUMERICAL RESULTS

We demonstrate the performance of the PT-CN scheme for RT-TDDFT calculations representing several prototypical usages of RT-TDDFT. Our method is implemented in PWDFT code, which uses the planewave basis set and is a self-contained module in the massively parallel Discontinuous Galerkin Density Functional Theory (DGDFT) software package[20,21] (parallelized at the MPI level). We use the Perdew−Burke−Ernzerhof (PBE) exchange correlation functional,[22] and the SG15 Optimized Norm-Conserving Vanderbilt (ONCV) pseudopotentials.[23,24]

PWDFT is mainly parallelized along the orbital direction, i.e., the maximum number of cores is equal to the number of occupied orbitals. In order to evaluate the additional term $\Phi(\Phi^* H\Phi)$ in the PT dynamics, the data are first transformed and partitioned along the grid direction via the MPI_Alltoallv operation, and then are efficiently implemented in parallel as matrix−matrix multiplications. The Anderson mixing for each occupied orbital can also be done in parallel in a straightforward fashion. We remark that the use of the Anderson mixing scheme may increase the memory footprint due to the storage of a few copies of the orbitals. If the memory consumption becomes a concern, one may employ other methods to achieve self-consistency, such as those based on the mixing of the charge density.

We measure the efficiency by comparing the wall clock time of the PT-CN scheme with that of the S-RK4 scheme, as well as the number of matrix-vector multiplications per orbital involving the Hamiltonian operator. The latter can be seen as a measure of efficiency that is relatively independent of the implementation. We remark that certain components of PWDFT, such as the solution of the Hartree potential, are currently performed on a single computational core. This is consistent with the choice of parallelizing along the orbital direction, where each $H\psi$ is performed on a single core. However, as will be shown below, S-RK4 typically requires many more evaluations of the Hartree potential than the PT-CN scheme, and the comparison of using the wall clock time is then slightly biased toward PT-CN. Hence, in section 3.4, we also report the wall clock time of the "scalable components" in PT-CN and S-RK4 separately. We report other computational details in Appendix C.

3.1. Absorption Spectrum. The first example is the computation of the absorption spectrum of an anthracene molecule. We set the time step size of PT-CN to be 12 as, and that of S-RK4 was determined to be 1 as (the scheme becomes unstable when the step size is larger). The total simulation time is 4.84 fs per polarization direction. Figure 2 compares the

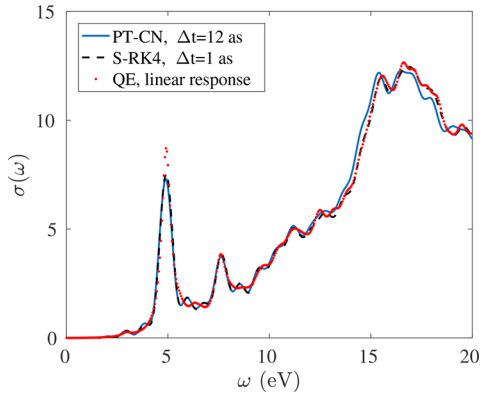


Figure 2. Absorption spectrum for anthracene.

absorption spectrum obtained from PT-CN and S-RK4 with PWDFT. This result is benchmarked against the linear response time-dependent density functional theory calculation using the turboTDDFT module[25] from the Quantum ESPRESSO (QE) software package,[26] which performs 3000 Lanczos steps along each perturbation direction to evaluate the polarization tensor. A Lorentzian smearing of 0.27 eV is applied to all calculations. Both QE and PWDFT use the same SG15 ONCV pseudopotential, and the kinetic energy cutoff is set to 20 hartree. No empty state is used in calculating the spectrum in PWDFT. We find that the absorption spectrum calculations from the three methods agree very well. The spectrum obtained from PT-CN and that from S-RK4 are nearly indistinguishable below 10 eV, and becomes slightly different above 15 eV. Note that the δ-pulse simultaneously excites all eigenstates from the entire spectrum, and $\omega = 15$ eV already amounts to the time scale of 40 as, which is approaching the step size of the PT-CN method. Since the computational cost of RT-TDDFT calculations is mainly dominated by the cost of applying the Hamiltonian operator to orbitals, we measure the numerical efficiency using the number

DOI: 10.1021/acs.jctc.8b00580
J. Chem. Theory Comput. 2018, 14, 5645−5652

of such matrix-vector multiplications per orbital. The PT-CN method requires, on average, 4.9 matrix-vector multiplications for each orbital. This is comparable to the S-RK4 method, which requires 4 matrix-vector multiplications per time step. Hence for this example, the PT-CN method is ∼10 times faster than the S-RK4 method, in terms of the number of matrix-vector multiplications. Using 33 computational cores, the total clock time for PT-CN and S-RK4 is 6.0 and 48.2 h, respectively.

3.2. Ultrafast Dynamics under a Laser Pulse. The second system is a benzene molecule driven by an ultrashort laser pulse. We apply two lasers with its wavelength being 800 and 250 nm, respectively. We measure the accuracy using the dipole moment along the x-direction, as well as the energy difference $E(t) - E(0)$ along the trajectory, as shown in Figure 3. For the case of the 800 nm laser, Figure 3 confirms that the

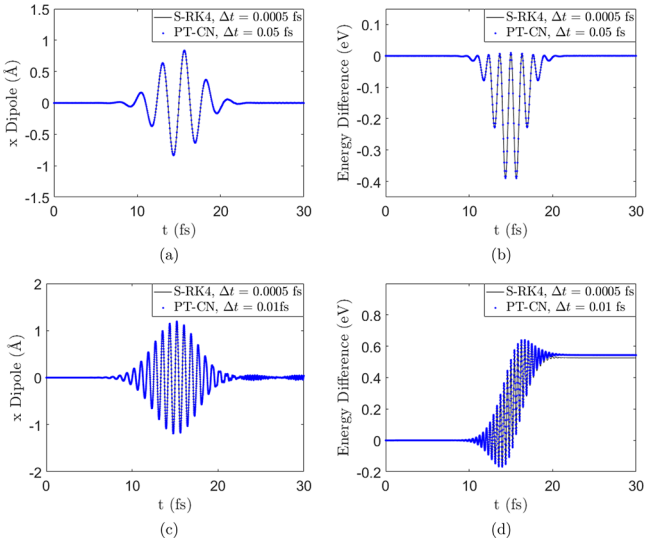


Figure 3. Electronic dynamics of benzene driven by laser with (a, b), λ = 800 nm and (c, d) λ = 250 nm.

results of PT-CN with a time step size of 50 as fully match those obtained from the S-RK4 method with a time step that is 100 times smaller. Again, S-RK4 becomes unstable when the time step size is >1 as. After 25.0 fs, the increase of the total energy for S-RK4 and PT-CN is 2.00×10^{-4} eV and 2.44×10^{-4} eV, respectively, indicating negligible energy absorption, because of the band gap of the system. During the time interval for which the laser is active (from 5.5 fs to 24.5 fs), the average number of matrix-vector multiplications per orbital in each PT-CN time step is 12.6, and the total number of matrix-vector multiplications per orbital is 4798. The total number of matrix-vector multiplications per orbital for S-RK4 within the same time interval is 152 000, and the speedup of PT-CN over RK4, measured by the number of matrix-vector multiplications, is 31.7. For the entire simulation of 30.0 fs, the wall clock times for PT-CN and S-RK4 are 0.53 and 16.82 h, respectively.

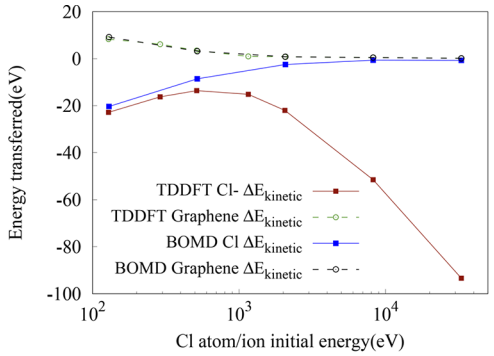
The 250 nm laser has a higher photon energy above the benzene band gap, thus resulting in significant energy absorption. In this case, even physical observables such as dipole moments become fast oscillating, and PT-CN must adopt a smaller step size of 10 and still yields very good approximation to the electron dynamics, compared to S-RK4. The increase of the total energy after 25.0 fs for S-RK4 and PT-CN is 0.526 and 0.544 eV, respectively. The average number of matrix-vector multiplications per orbital in each PT-CN time step is 8.3 due to the reduced step size, and the total number of matrix-vector multiplications per orbital is 15 817. Therefore, in this case, PT-CN achieves 9.6 times speedup over S-RK4, in terms of the $H\psi$ calculation. For the entire simulation up to 30.0 fs, the wall clock time of PT-CN and S-RK4 are 1.7 and 16.9 h, respectively.

3.3. Ehrenfest Dynamics for Ion Collision. In the third example, we use the RT-TDDFT based Ehrenfest dynamics to study the process of a chlorine ion (Cl^-) colliding to a graphene nanoflake consisting of 112 atoms. This models the ion implantation procedure for doping a substrate. At the beginning of the simulation, the Cl^- is placed 6 Å away from the graphene and is given an initial velocity perpendicular to the plane of the graphene pointing toward the center of one hexagonal ring formed by the C atoms. The initial velocities of Cl^- are set to be 0.5, 0.75, 1.0, 1.5, 2.0, 4.0, and 8.0 Bohr/fs, and the kinetic energy carried by the ion ranges from 128 eV to 32 926 eV, respectively. The simulation is terminated before the ion reaches the boundary of the supercell. For instance, we set T = 10 fs when the velocity is 2.0 Bohr/fs. In such cases, the time step size for PT-CN and S-RK4 is set to be 50 as and 0.5 as, respectively. Each PT-CN step requires, on average, 28 matrix-vector multiplication operations per orbital, and the overall speedup of PT-CN over S-RK4 is 14.2. The wall clock times of PT-CN and S-RK4 are 5.2 and 78 h, respectively. We note that, in this example, the speedup predicted by the number of matrix-vector multiplication operations is 14.3, while the speedup measured by the wall clock time is 15.0. This is because, in Ehrenfest dynamics, PT-CN only requires one evaluation of the pseudopotential every 50 as, while S-RK4 requires 400 such steps within the same period of simulation. This makes PT-CN even more favorable, in the context of Ehrenfest dynamics.

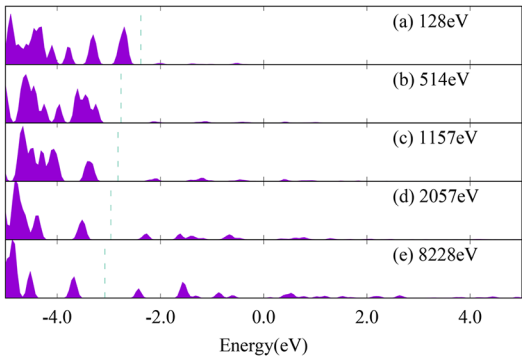
We compare the result obtained from the Ehrenfest dynamics with that from the Born−Oppenheimer Molecular Dynamics (BOMD). In the BOMD simulation, since the extra electron of Cl^- will localize on the conduction band of the graphene conduction rather than on Cl during the self-consistent field iteration, we replace the Cl^- ion by the Cl atom. Figure 4a illustrates the energy transfer with different initial kinetic energies. As the Cl/Cl^- initial kinetic energy increases, the gain of the kinetic energy by the graphene atoms decreases, because of the fact that Cl/Cl^- can pass through the system faster. When the initial kinetic energy of Cl/Cl^- is smaller than 500 eV, the losses of the kinetic energy for Cl/Cl^- are similar between RT-TDDFT and BOMD. However, when the initial kinetic energy of Cl/Cl^- further increases, the RT-TDDFT predicts an increase of the loss of the Cl/Cl^- kinetic energy, while the gain of the graphene kinetic energy remains decreasing. This is a consequence of the electron excitation, which is absent in the BOMD simulation. Such excitation is illustrated in Figure 4b for the occupied electron density of states in the higher-energy regimes. The occupied density of states is calculated as $\rho(\varepsilon) := \sum_{j=1}^{N_e} \sum_{i=1}^{\infty} |\langle \psi_i(T) | \varphi_j(T) \rangle|^2 \tilde{\delta}(\varepsilon - \varepsilon_i(T))$. Here, $\phi_j(T)$ is the j-th orbital obtained at the end of the RT-TDDFT simulation at time T, and $\varepsilon_i(T)$,$\psi_i(T)$ are the eigenvalues and wave functions corresponding to the Hamiltonian at time T. $\tilde{\delta}$ is a Dirac-δ function with a Gaussian broadening of 0.05 eV.

Figure 5 presents further details of the energy transfer along the trajectory of the RT-TDDFT and BOMD simulation when the initial velocity is 2.0 Bohr/fs (2057 eV). When the collision

DOI: 10.1021/acs.jctc.8b00580
J. Chem. Theory Comput. 2018, 14, 5645−5652



(a) BOMD and RT-TDDFT energy transfer with different initial kinetic energies.



(b) Density of state after the ion collision. Green dashed line: Fermi energy.

Figure 4. Energy transfer and density of states.

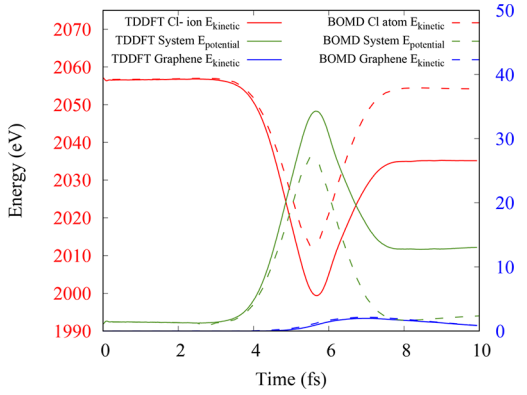


Figure 5. BOMD and RT-TDDFT energy transfer with time; the projectile speed is 2.0 Bohr/fs.

occurs at $T \approx 6$ fs, the loss of the Cl/Cl$^-$ kinetic energy is 44 and 58 eV under RT-TDDFT and BOMD, respectively. However, after the collision, Cl regains almost all the kinetic energy in BOMD, and the final kinetic energy is only 2.5 eV less than the initial one. Correspondingly, the kinetic energy of the graphene increases by 0.86 eV and the potential energy increases by 1.63 eV. On the other hand, RT-TDDFT predicts that the Cl$^-$ ion should lose 22.5 eV of kinetic energy, which is mostly transferred to the potential energy of the excited electrons. The increase in the kinetic energy of the graphene is 0.84 eV and is similar to the BOMD result. Therefore, in RT-TDDFT, the Cl$^-$ loses its kinetic energy to electron excitation in graphene.

3.4. Computational Scaling for Silicon Systems.

Finally, we demonstrate the performance of the PT-CN method for silicon systems ranging from 32 to 1024 atoms, driven by a 800 nm laser pulse, as shown in Figure 6a. The

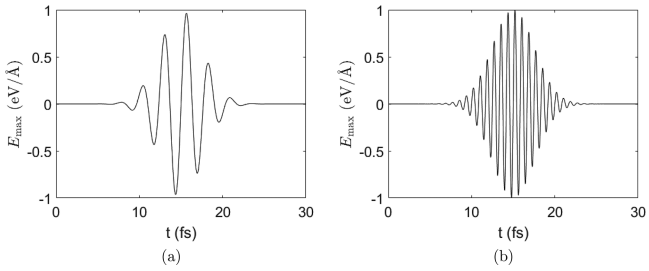
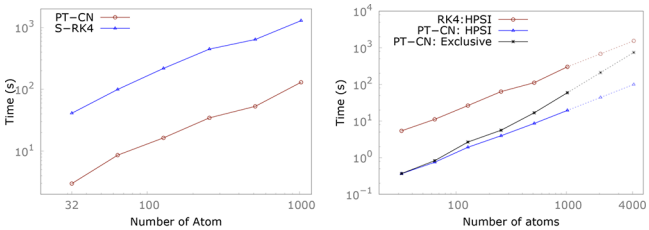


Figure 6. External fields of the lasers. The peak electric field is 1.0 eV/Å, occurring at t_0 = 15.0 fs, and the fwhm pulse width is 6.0 fs. The wavelength is (a) 800 nm and (b) 250 nm.

supercells are constructed from 2 × 2 × 1 to 8 × 4 × 4 unit cells, respectively, and each simple cubic unit cell consists of 8 Si atoms with a lattice constant of 5.43 Å. The kinetic energy cutoff is set to 10 hartree. In order to validate the scalability of our parallel implementation, in all tests, the number of computational cores is equal to the number of occupied orbitals, i.e., we use 64 cores for the system consisting of 32 atoms, and 2048 cores for the system with 1024 atoms. The time step for PT-CN is 50, whereas the time step for S-RK4 is 0.5 as. We report the total wall clock time of both methods per 50 as. In the case of PT-CN, the number of mixing iterations averaged over the simulation of 40 fs is 28, and the maximum mixing dimension per step is set to 10. First, we collect the total wall clock time for both S-RK4 and PT-CN method in Figure 7a. We find that PT-CN is 13 times faster than S-RK4



(a) Total wall clock time (b) Scalable components

Figure 7. Comparison of the computational time between S-RK4 and PT-CN methods for a time period of 50 as: (left) total wall clock time and (right) components of the wall clock time that are scalable to large number of processors in our implementation (matrix-vector multiplication, and operations exclusive to PT-CN). The dashed lines extrapolate the computational time to system sizes up to 4000 atoms.

for the smallest system with 32 atoms, and is still 10 times faster for the 1024-atom system. On the other hand, as discussed at the beginning of this section, certain components, such as the solution of the Hartree potential, are currently implemented on a single computational core. Note that S-RK4 requires 400 evaluations of the Hartree potential and PT-CN only requires 28 such evaluations per unit time. Hence, in the following, we also present the breakdown of the wall clock time, in terms of the "scalable components" in Figure 7b. In particular, this means the application of the Hamiltonian operator (denoted by the label "HPSI") for S-RK4 and PT-CN, as well as the additional components exclusively needed

DOI: 10.1021/acs.jctc.8b00580
J. Chem. Theory Comput. 2018, 14, 5645−5652

for PT-CN (denoted by the label "Exclusive"). This includes the evaluation of the residual, the orbital orthogonalization, and the Anderson mixing steps. We also assume that S-RK4 does not require the orthogonalization step, so the computational cost should scale as $O(N_e^2)$, with respect to the system size. We find that, when measured by such a metric, PT-CN still can be 7.4 times faster for the 32-atom system, and the speedup is reduced to 3.8 for the 1024-atom system. The reduction of the speedup is mainly due to the increase of the computation of the residual, as well as orbital orthogonalization needed in the PT dynamics, which scales as $O(N_e^3)$. In Figure 7b, we also extrapolate the computational cost of different components to systems of larger sizes, which indicates that the crossover between the S-RK4 and PT-CN method measured in this metric is ∼4000 atoms.

4. CONCLUSION

In this work, we demonstrate that one significant factor leading to the very small time step size in RT-TDDFT calculations is the nonoptimal gauge choice in the Schrödinger dynamics. Since all physical observables should be gauge-independent, we may optimize the gauge choice to improve the numerical efficiency without sacrificing accuracy. The resulting scheme can be beneficial to any RT-TDDFT integrator, and can even be nearly symplectic. With the increased time step size, we hope that RT-TDDFT can be used to study many ultrafast problems unamenable today. Furthermore, the parallel transport formulation can be naturally extended to RT-TDDFT calculations with hybrid exchange-correlation functionals. In such context, the PT dynamics is expected to significantly reduce the number of applications of the Fock exchange operator, which dominates the computational cost. We will present numerical results along this direction in the near future.

APPENDIX A: DERIVATION OF THE PARALLEL TRANSPORT GAUGE

We follow two steps to solve the minimization problem described in eq 3. We first derive eq 4 for which the minimizer should satisfy, and then derive the parallel transport dynamics in eq 6.

(1) We first split $\dot{\Phi}$ into two orthogonal components:

$$\dot{\Phi} = P\dot{\Phi} + (I - P)\dot{\Phi} \tag{A1}$$

Then, we have

$$\|\dot{\Phi}\|_F^2 = \|P\dot{\Phi}\|_F^2 + \|(I - P)\dot{\Phi}\|_F^2 \tag{A2}$$

To reformulate the second term, we take the time derivative on the equation $P\Phi = \Phi$ (this is because Φ and Ψ just differ by a gauge choice, thus $P = \Psi\Psi^* = \Phi\Phi^*$, and so the condition is implied). We obtain $\dot{P}\Phi = \dot{\Phi} - P\dot{\Phi} = (I - P)\dot{\Phi}$. Therefore, eq A2 becomes

$$\|\dot{\Phi}\|_F^2 = \|P\dot{\Phi}\|_F^2 + \|\dot{P}\Phi\|_F^2 = \|P\dot{\Phi}\|_F^2 + \|\dot{P}\Psi\|_F^2 \tag{A3}$$

where the last equality again comes from the fact that Φ and Ψ only differ by a unitary gauge matrix.

Equation A3 gives an orthogonal decomposition. The second term,

$$\|\dot{P}\Psi\|_F^2 = \mathrm{Tr}[\Psi^*\dot{P}^2\Psi] = \mathrm{Tr}[\dot{P}^2\Psi\Psi^*] = \mathrm{Tr}[\dot{P}^2P] \tag{A4}$$

is defined solely from the density matrix and, thus, is gauge-invariant. Therefore, the variation of Φ is minimized when

$$P\dot{\Phi} = 0 \tag{A5}$$

which is exactly the parallel transport condition.

(2) Now we would like to directly write down the governing equation of Φ. First, the equation $\Phi = P\Phi$ and the parallel transport condition described by eq 4 implies that

$$\dot{\Phi} = \partial_t(P\Phi) = \dot{P}\Phi + P\dot{\Phi} = \dot{P}\Phi \tag{A6}$$

Together with the von Neumann equation, we have

$$\imath\dot{\Phi} = \imath\dot{P}\Phi = [H, P]\Phi = HP\Phi - PH\Phi = H\Phi - \Phi(\Phi^*H\Phi) \tag{A7}$$

This is exactly the parallel transport dynamics.

APPENDIX B: TIME DISCRETIZATION SCHEMES

We list several propagation schemes used in this paper, but the parallel transport dynamics can be discretized with any propagator. Here, all the $H_n = H(t_n, P_n)$ are the Hamiltonians at step t_n, and $t_{n+1/2} = t_n + \frac{1}{2}\Delta t$, $t_{n+1} = t_n + \Delta t$. For implicit time integrators, Ψ_{n+1} or Φ_{n+1} must be solved self-consistently.

The standard explicit fourth-order Runge−Kutta scheme for the Schrödinger dynamics (S-RK4):

$$\begin{aligned}
k_1 &= -\imath\Delta t H_n\Psi_n \\
\Psi_n^{(1)} &= \Psi_n + \frac{1}{2}k_1, \quad H_n^{(1)} = H(t_{n+1/2}, \Psi_n^{(1)}\Psi_n^{(1)*}) \\
k_2 &= -\imath\Delta t H_n^{(1)}\Psi_n^{(1)} \\
\Psi_n^{(2)} &= \Psi_n + \frac{1}{2}k_2, \quad H_n^{(2)} = H(t_{n+1/2}, \Psi_n^{(2)}\Psi_n^{(2)*}) \\
k_3 &= -\imath\Delta t H_n^{(2)}\Psi_n^{(2)} \\
\Psi_n^{(3)} &= \Psi_n + k_3, \quad H_n^{(3)} = H(t_{n+1}, \Psi_n^{(3)}\Psi_n^{(3)*}) \\
k_4 &= -\imath\Delta t H_n^{(3)}\Psi_n^{(3)} \\
\Psi_{n+1} &= \Psi_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)
\end{aligned} \tag{B1}$$

The standard explicit fourth-order Runge−Kutta scheme for the parallel transport dynamics (PT-RK4):

$$\begin{aligned}
k_1 &= -\imath\Delta t\{H_n\Phi_n - \Phi_n(\Phi_n^*H_n\Phi_n)\} \\
\Phi_n^{(1)} &= \Phi_n + \frac{1}{2}k_1, \quad H_n^{(1)} = H(t_{n+1/2}, \Phi_n^{(1)}\Phi_n^{(1)*}) \\
k_2 &= -\imath\Delta t\{H_n^{(1)}\Phi_n^{(1)} - \Phi_n^{(1)}(\Phi_n^{(1)*}H_n^{(1)}\Phi_n^{(1)})\} \\
\Phi_n^{(2)} &= \Phi_n + \frac{1}{2}k_2, \quad H_n^{(2)} = H(t_{n+1/2}, \Phi_n^{(2)}\Phi_n^{(2)*}) \\
k_3 &= -\imath\Delta t\{H_n^{(2)}\Phi_n^{(2)} - \Phi_n^{(2)}(\Phi_n^{(2)*}H_n^{(2)}\Phi_n^{(2)})\} \\
\Phi_n^{(3)} &= \Phi_n + k_3, \quad H_n^{(3)} = H(t_{n+1}, \Phi_n^{(3)}\Phi_n^{(3)*}) \\
k_4 &= -\imath\Delta t\{H_n^{(3)}\Phi_n^{(3)} - \Phi_n^{(3)}(\Phi_n^{(3)*}H_n^{(3)}\Phi_n^{(3)})\} \\
\Phi_{n+1} &= \Phi_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)
\end{aligned} \tag{B2}$$

DOI: 10.1021/acs.jctc.8b00580
J. Chem. Theory Comput. 2018, 14, 5645−5652

Table C1. Accuracy and Efficiency of PT-CN for the Electron Dynamics with the 250 nm Laser, Compared to S-RK4[a]

method	Δt (fs)	AEI (eV)	AOE (eV)	MVM	speedup	WCT (h)	speedup
S-RK4	0.0005	0.5260	/	152000	/	16.9	/
PT-CN	0.005	0.5340	0.0080	28610	5.3	3.2	5.3
PT-CN	0.0065	0.5347	0.0087	22649	6.7	2.5	6.8
PT-CN	0.0075	0.5362	0.0102	21943	6.9	2.4	7.0
PT-CN	0.01	0.5435	0.0175	15817	9.6	1.7	10.1
PT-CN	0.02	0.5932	0.0672	12110	12.6	1.3	12.5

[a]The accuracy is measured using the average energy increase (AEI) after 25.0 fs and the average overestimated energy (AOE) after 25.0 fs. The efficiency is measured using the total number of matrix-vector multiplications per orbital (MVM) during the time interval from 5.5 fs to 24.5 fs, the wall clock time (WCT) of the entire simulation, and the computational speedup.

The implicit Crank–Nicolson scheme for the Schrödinger dynamics (S-CN) is given as

$$\left(I + \imath\frac{\Delta t}{2}H_{n+1}\right)\Psi_{n+1} = \left(I - \imath\frac{\Delta t}{2}H_n\right)\Psi_n \tag{B3}$$

The implicit Crank-Nicolson scheme for the parallel transport dynamics (PT-CN) is given as

$$\begin{aligned}&\Phi_{n+1} + \imath\frac{\Delta t}{2}\{H_{n+1}\Phi_{n+1} - \Phi_{n+1}(\Phi_{n+1}^* H_{n+1}\Phi_{n+1})\}\\ &= \Phi_n - \imath\frac{\Delta t}{2}\{H_n\Phi_n - \Phi_n(\Phi_n^* H_n\Phi_n)\}\end{aligned} \tag{B4}$$

APPENDIX C: DETAILS OF RT-TDDFT CALCULATIONS

For the example of absorption spectrum of anthracene ($C_{14}H_{10}$), the simulation is performed using a cubic supercell of size (20 Å^3), and the kinetic energy cutoff is 20 hartree. In order to compute the absorption spectrum, a δ-pulse of strength 0.005 a.u. is applied to the x,y,z directions to the ground state wave functions, respectively, and the system is then propagated for 4.8 fs along each direction. This gives the polarization tensor $\chi(\omega)$, and the optical absorption cross-section is evaluated as

$$\sigma(\omega) = \left(\frac{4\pi\omega}{c}\right)\mathrm{Im}\mathrm{Tr}[\chi(\omega)]$$

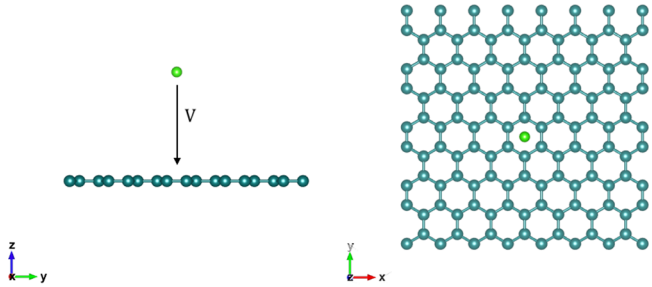
For the example of the benzene molecule driven by an ultrashort laser pulse, the electric field takes the form

$$\mathbf{E}(t) = \hat{\mathbf{k}}E_{\max}\exp\left[-\frac{(t-t_0)^2}{2a^2}\right]\sin[\omega(t-t_0)] \tag{C1}$$

where $\hat{\mathbf{k}}$ is a unit vector defining the polarization of the electric field. The parameters a, t_0, $E_{\max}$, and ω define the width, the initial position of the center, the maximum amplitude of the Gaussian envelope, and the frequency of the laser, respectively. In practice, ω and a are related to the wavelength λ and the full width at half-maximum (fwhm) pulse width[27] as $\lambda\omega = 2\pi c$ and FWHM $= 2a\sqrt{2\log 2}$, where c is the speed of the light. In this example, the peak electric field ($E_{\max}$) is 1.0 eV/Å, occurring at t_0 = 15.0 fs. The fwhm pulse width is 6.0 fs, and the polarization of the laser field is aligned along the x-axis (the benzene molecule is in the x–y plane). We consider one relatively slow laser with a wavelength of 800 nm, and another faster laser with a wavelength of 250 nm, respectively (see Figure 6). The electron dynamics for the first laser is in the near-adiabatic regime, where the system remains near the ground state after the active time interval of the laser, while the second laser drives a significant amount of electrons to excited states. We propagate RT-TDDFT to T = 30.0 fs. For the parameters in the Anderson mixing, the step length α is 0.2, the mixing dimension is 10, and the tolerance is 10^{-6}.

Even though the electron dynamics varies rapidly under the laser with a wavelength of 250 nm, PT-CN can still be stable within a relatively large range of time steps. Table C1 measures the accuracy of PT-CN with Δt = 5, 6.5, 7.5, 10, and 20 as, respectively. We find that the number of matrix-vector multiplications per orbital systematically decreases as the step size increases. When Δt = 20 as, the speedup over S-RK4 is 12.6, and this is at the expense of overestimating the energy by 0.0672 eV after the active interval of the laser. Hence, one may adjust the time step size to obtain a good compromise between efficiency and accuracy, and use PT-CN to quickly study the electron dynamics with a large time step. This is not possible using an explicit scheme such as S-RK4.

For the ion collision example, the system is shown in Figure C1. The supercell has 113 atoms (including Cl^-) sampled at the Γ-point. The size of the single-layer graphene sheet is 17.28 Å along both the x- and y-directions. The kinetic energy cutoff is 30 hartree. The system has 228 states and each state is occupied with 2 electrons, and the calculation is performed with 228 computational cores.



(a) x direction (b) z direction

Figure C1. Model for the collision of Cl/Cl^- and a graphene nanoflake.

AUTHOR INFORMATION

Corresponding Author

*E-mail: linlin@math.berkeley.edu.

ORCID

Weile Jia: 0000-0002-7049-2587

Lin Lin: 0000-0001-6860-9566

Notes

The authors declare no competing financial interest.

DOI: 10.1021/acs.jctc.8b00580
J. Chem. Theory Comput. 2018, 14, 5645–5652

■ ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation (under Grant Nos. 1450372 and DMS-1652330) (D.A., W.J., and L.L.), the Department of Energy (under Grant Nos. DE-SC0017867 and DE-AC02-05CH11231) (L.L.), and the Department of Energy Theory of Materials (KC2301) program under Contract No. DE-AC02-05CH11231 (L.W.). We thank the National Energy Research Scientific Computing (NERSC) Center and the Berkeley Research Computing (BRC) Program at the University of California, Berkeley for making computational resources available. We thank Stefano Baroni, Roberto Car, Zhanghui Chen, Wei Hu, Christian Lubich, and Chao Yang for helpful discussions.

■ REFERENCES

(1) Runge, E.; Gross, E. K. U. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* **1984**, *52*, 997.
(2) Onida, G.; Reining, L.; Rubio, A. Electronic excitations: density-functional versus many-body Green's-function approaches. *Rev. Mod. Phys.* **2002**, *74*, 601.
(3) Yabana, K.; Bertsch, G. F. Time-dependent local-density approximation in real time. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 4484−4487.
(4) Ullrich, C. A. *Time-Dependent Density-Functional Theory: Concepts and Applications*; Oxford University Press: Oxford, U.K., 2011.
(5) Andrade, X.; Alberdi-Rodriguez, J.; Strubbe, D. A.; Oliveira, M. J. T.; Nogueira, F.; Castro, A.; Muguerza, J.; Arruabarrena, A.; Louie, S. G.; Aspuru-Guzik, A.; Rubio, A.; Marques, M. A. L. Time-dependent density-functional theory in massively parallel computer architectures: the octopus project. *J. Phys.: Condens. Matter* **2012**, *24*, 233202.
(6) Takimoto, Y.; Vila, F. D.; Rehr, J. J. Real-time time-dependent density functional theory approach for frequency-dependent nonlinear optical response in photonic molecules. *J. Chem. Phys.* **2007**, *127*, 154114.
(7) Krasheninnikov, A. V.; Miyamoto, Y.; Tománek, D. Role of electronic excitations in ion collisions with carbon nanostructures. *Phys. Rev. Lett.* **2007**, *99*, 016104.
(8) Castro, A.; Marques, M.; Rubio, A. Propagators for the time-dependent Kohn-Sham equations. *J. Chem. Phys.* **2004**, *121*, 3425−33.
(9) Schleife, A.; Draeger, E. W.; Kanai, Y.; Correa, A. A. Plane-wave pseudopotential implementation of explicit integrators for time-dependent Kohn-Sham equations in large-scale simulations. *J. Chem. Phys.* **2012**, *137*, 22A546.
(10) Gómez Pueyo, A.; Marques, M. A.; Rubio, A.; Castro, A. Propagators for the time-dependent Kohn-Sham equations: multistep, Runge-Kutta, exponential Runge-Kutta, and commutator free Magnus methods. *J. Chem. Theory Comput.* **2018**, *14*, 3040.
(11) Jahnke, T.; Lubich, C. Numerical integrators for quantum dynamics close to the adiabatic limit. *Numer. Math.* **2003**, *94*, 289−314.
(12) Wang, Z.; Li, S.-S.; Wang, L.-W. Efficient Real-Time Time-Dependent Density Functional Theory Method and its Application to a Collision of an Ion with a 2D Material. *Phys. Rev. Lett.* **2015**, *114*, 1−5.
(13) Alonso, J. L.; Andrade, X.; Echenique, P.; Falceto, F.; Prada-Gracia, D.; Rubio, A. Efficient formalism for large-scale ab initio molecular dynamics based on time-dependent density functional theory. *Phys. Rev. Lett.* **2008**, *101*, 096403.
(14) Nakahara, M. *Geometry, Topology and Physics*; CRC Press: Boca Raton, FL, 2003.
(15) Cornean, H. D.; Monaco, D.; Teufel, S. Wannier functions and Z2 invariants in time-reversal symmetric topological insulators. *Rev. Math. Phys.* **2017**, *29*, 1730001.
(16) An, D.; Lin, L. Quantum dynamics with the parallel transport gauge. *arXiv:1804.02095*.
(17) Anderson, D. G. Iterative procedures for nonlinear integral equations. *J. Assoc. Comput. Mach.* **1965**, *12*, 547−560.
(18) Lubich, C. *From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis*; European Mathematical Society: Zurich, Switzerland, 2008.
(19) Jornet-Somoza, J.; Alberdi-Rodriguez, J.; Milne, B. F.; Andrade, X.; Marques, M.; Nogueira, F.; Oliveira, M.; Stewart, J.; Rubio, A. Insights into colour-tuning of chlorophyll optical response in green plants. *Phys. Chem. Chem. Phys.* **2015**, *17*, 26599−26606.
(20) Lin, L.; Lu, J.; Ying, L.; E, W. Adaptive local basis set for Kohn-Sham density functional theory in a discontinuous Galerkin framework I: Total energy calculation. *J. Comput. Phys.* **2012**, *231*, 2140−2154.
(21) Hu, W.; Lin, L.; Yang, C. DGDFT: A massively parallel method for large scale density functional theory calculations. *J. Chem. Phys.* **2015**, *143*, 124110.
(22) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(23) Hamann, D. R. Optimized norm-conserving Vanderbilt pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *88*, 085117.
(24) Schlipf, M.; Gygi, F. Optimization algorithm for the generation of ONCV pseudopotentials. *Comput. Phys. Commun.* **2015**, *196*, 36−44.
(25) Malcıoğlu, O. B.; Gebauer, R.; Rocca, D.; Baroni, S. turboTDDFT−A code for the simulation of molecular spectra using the Liouville−Lanczos approach to time-dependent density-functional perturbation theory. *Comput. Phys. Commun.* **2011**, *182*, 1744−1754.
(26) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Dal Corso, A.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter* **2009**, *21*, 395502−395520.
(27) Russakoff, A.; Li, Y.; He, S.; Varga, K. Accuracy and computational efficiency of real-time subspace propagation schemes for the time-dependent density functional theory. *J. Chem. Phys.* **2016**, *144*, 204125.

DOI: 10.1021/acs.jctc.8b00580
J. Chem. Theory Comput. 2018, 14, 5645−5652