Cost-Efficient High-Resolution Linear Absorption Spectra Through Extrapolating the Dipole Moment from Real-Time Time-Dependent Electronic-Structure Theory

Eirill Hauge,[†,‡] Håkon Emil Kristiansen,[†] Lukas Konecny,[¶,§] Marius Kadek,[¶,‖] Michal Repisky,[¶,⊥] and Thomas Bondo Pedersen[*,†]

†*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Oslo, P.O. Box 1033, Blindern, 0315 Oslo, Norway*

‡*Department of Numerical Analysis and Scientific Computing, Simula Research Laboratory, Kristian Augusts gate 23, 0164 Oslo, Norway*

¶*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Tromsø—The Arctic University of Norway, N-9037 Tromsø, Norway*

§*Max Planck Institute for the Structure and Dynamics of Matter, Center for Free Electron Laser Science, Luruper Chaussee 149, 22761 Hamburg, Germany*

‖*Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA*

⊥*Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University, Ilkovicova 6, SK-84215 Bratislava, Slovakia*

E-mail: t.b.pedersen@kjemi.uio.no

Abstract

We present a novel function fitting method for approximating the propagation of the time-dependent electric dipole moment from real-time electronic structure calculations.

Real-time calculations of the electronic absorption spectrum require discrete Fourier transforms of the electric dipole moment. The spectral resolution is determined by the total propagation time, i.e. the trajectory length of the dipole moment, causing a high computational cost. Our developed method uses function fitting on shorter trajectories of the dipole moment, achieving arbitrary spectral resolution through extrapolation. Numerical testing shows that the fitting method can reproduce high-resolution spectra using short dipole trajectories. The method converges with as little as 100 a.u. dipole trajectories for some systems, though the difficulty converging increases with the spectral density. We also introduce an error estimate of the fit, reliably assessing its convergence and hence the quality of the approximated spectrum.

1 Introduction

The rapid advancement of laser technology in the past decades allows us to probe matter on spatiotemporal scales that approach the characteristic time and length scales of the electron, opening the field of attosecond science.[1,2] This development has forced quantum chemists to shift their attention from the time-independent to the time-dependent Schrödinger and Dirac equations.[3–5] Numerical approaches to laser-driven electron dynamics are often labelled *real-time* methods to distinguish them from the response-theoretical methods to the time-dependent Schrödinger/Dirac equation, the latter solving the equations of motion perturbatively in the frequency domain.[6,7]

Even without explicit reference to results derived from perturbation theory such as, e.g., Fermi's golden rule, it is still possible to extract linear and low-order nonlinear optical properties from nonperturbative real-time simulations, including electronic absorption spectra and time-resolved pump-probe absorption spectra that would be hard or impossible to compute using response theory—see Refs. 8–12 for recent examples.

In this work, we focus on electronic absorption spectra extracted from electron-dynamics simulations driven by a Dirac-delta impulse, which excites the molecule into all dipole-allowed

excited states simultaneously.[8] Due to the nonperturbative nature of real-time methods, the resulting spectrum contains nonlinear (e.g., two-photon) as well as linear absorption lines.[11] For weak pulses, the nonlinear effects are small, and the absorption spectrum is dominated by linear lines. In practice, the induced electric dipole moment is recorded in the course of the simulation and subsequently transformed to the frequency domain to yield the absorption cross section. When using the conventional discrete Fourier transform to process the signal, the spectral resolution is inversely proportional to the number of time steps N and the time-step length Δt, as $\Delta\omega = 2\pi/(N\Delta t)$. Obtaining sufficient spectral resolution typically requires tens to hundreds of thousands of time steps since Δt cannot be increased beyond a certain limit if rapid oscillations of the electron density are to be captured. Moreover, increasing Δt reduces the accuracy and stability of the numerical integration scheme used to propagate the electronic degrees of freedom. As the computational effort in each time step requires multiple rebuilds of the Hamiltonian matrix, it is comparable to several iterations of a ground-state optimization within the chosen electronic-structure model.[13] Hence, there is considerable interest in decreasing the number of time steps required to achieve sufficient spectral resolution.

In addition to reducing the number of time steps, it is possible to increase the computational efficiency of real-time electronic structure methods by disregarding negligible basis functions,[14] basis-function pairs and quartets.[15] As a result, in a large molecule, although there are $\mathcal{O}(N^4)$ electron repulsion integrals (ERIs) in total, it can be shown that only $\mathcal{O}(N^2)$ of them are significant, where N refers to the number of basis functions. As shown within RT-TDDFT, a large prefactor associated with the evaluation of non-negligible ERIs can further be reduced by using, e.g., the resolution-of-the-identity method.[16] By applying a spatial truncation radius upon the time-dependent density matrix, RT-TDDFT can approach the linear $\mathcal{O}(N)$ scaling.[17,18]

Previous efforts to improve the spectral resolution have been made by estimating excitation energies through various signal processing techniques.[19–21] More recently, Bruner et al.[22]

investigated the use of Padé approximants to interpolate the discrete Fourier transforms used for the absorption spectrum. These are all methods operating in the frequency domain, leaving no other validation options than comparison with a fully propagated spectrum.

The original periodic signal is typically damped using a decaying exponential function to reduce unwanted artefacts arising when the discrete Fourier transform is applied to oscillating functions in simulations with finite trajectory length. In the time domain, the number of time points can be increased by padding the damped signal with zeros, leading to finer spectra. However, this artificial extension of the trajectory length can only be applied on sufficiently damped signals.

In this work, we investigate a more sophisticated and powerful alternative: the extrapolation of a short signal. The discrete Fourier transform of an extrapolated signal achieves increasingly higher spectral resolution as the extrapolation length increases. This requires the development of a stable and reliable method for time-series forecasting. The inherently harmonic character of the time-dependent wave function in the absence of an external field suggests that such forecasting of molecular properties should be possible. Importantly, the forecasted signal can be verified in the time domain by comparing it with relatively few additional time steps. To the best of our knowledge, no published work exists on improving the spectral resolution by such extrapolation of the time-dependent dipole moment.

The current success and popularity of machine learning is undeniable, including use cases in chemistry,[23–29] and one might be tempted to leverage artificial neural networks for forecasting the time-dependent electric-dipole moment. However, while artificial neural networks are powerful tools for pattern detection in large data sets, they struggle with precise and reliable extrapolations.[30,31] Although the universal approximation theorem[32] tells us that an excellent *interpolation* can be achieved, it does not guarantee a stable *extrapolation*. In order to achieve a stable extrapolation, over-fitting must be avoided by enforcing sufficient restrictions.

In this paper, we present a novel approach for obtaining high-resolution absorption spec-

tra from real-time simulations of laser-driven electron dynamics by exploiting *a priori* knowledge of the form of the dipole function from quantum mechanics in a finite-dimensional Hilbert space. The form of the dipole function thus is motivated by the underlying physics, with unknown parameters to be determined by fitting a short dipole trajectory from a real-time simulation. The fitted function may be evaluated at any point in time, meaning that it can be extrapolated in the time domain to arbitrary future time. This further implies that we can achieve arbitrary spectral resolution. For sufficiently weak Dirac-delta impulse, the evaluation of absorption spectra based on these fitted functions may use analytical expressions for the linear response function.[6]

Working in the time domain, a quantitative error measure of the fitted dipole function can be monitored during the course of the real-time simulation and used to evaluate convergence. This way, an unnecessarily long real-time propagation can be avoided by automatically terminating the propagation upon the convergence of the fit. The developed method is independent of the quantum mechanical model and is tested with several molecular systems using mainly real-time time-dependent density-functional theory (RT-TDDFT).[8,33–38] Despite certain flaws arising mainly from the almost universally adopted adiabatic density-functional approximation,[3,39] RT-TDDFT is the far most widely used electronic-structure method for laser-driven electron dynamics. With computational costs comparable to (or *below*) time-dependent Hartree-Fock theory,[40] RT-TDDFT takes into account electron-correlation effects that would otherwise require advanced and computationally expensive wave-function theories.[4,5] To demonstrate the independence of the underlying electronic-structure theory, we also present results obtained from real-time time-dependent configuration interaction singles (RT-TDCIS)[41–43] theory.

We will start with a short presentation of the electric-dipole approximation within real-time simulations before introducing the proposed method for fitting the time-dependent electric-dipole moment. After briefly laying out the simulation details for the real-time simulation of a selection of systems, the results of the fitting method on these systems are

presented and discussed. Finally, we reflect on the performance of the fitting method and discuss potential future improvements.

2 Theory

In this work, we employ the following conventions: Subscripts u, v denote Cartesian components, vectors are typed in boldface, and quantum-mechanical operators are denoted by a hat. Following the convention of response theory by Olsen and Jørgensen,[6] we define the Fourier transform and its inverse according to

$$\tilde{f}(\omega) = \mathcal{F}[f(t)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)\mathrm{e}^{\mathrm{i}\omega t}\mathrm{d}t, \tag{1}$$

$$f(t) = \mathcal{F}^{-1}[\tilde{f}(\omega)] = \int_{-\infty}^{\infty} \tilde{f}(\omega)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega, \tag{2}$$

where the transformed function is denoted by a tilde. Atomic units are used throughout unless otherwise specified.

2.1 Real-time simulations of absorption spectra

Within the clamped-nucleus Born-Oppenheimer approximation, real-time simulations of electronic absorption spectra typically assume the electric-dipole approximation, where a molecule is subjected to a time-dependent spatially uniform electric field, $F(t)$. The time-dependent Hamiltonian reads

$$\hat{H}(t) = \hat{H}_0 + \hat{V}(t), \tag{3}$$

where $\hat{H}_0$ is the time-independent electronic Hamiltonian, and the interaction operator is given by

$$\hat{V}(t) = -\hat{\boldsymbol{\mu}} \cdot \mathbf{u}F(t), \tag{4}$$

where $\hat{\boldsymbol{\mu}}$ is the electric dipole moment operator. The linear polarization direction of the electric field is determined by the real unit vector $\mathbf{u}$, such that the field aligns with one of the Cartesian axes. This implies the form $\hat{V}(t) = -\hat{\mu}_u F(t)$, where $\hat{\mu}_u$ is the component of $\hat{\boldsymbol{\mu}}$ along the polarization direction.

We assume that the electronic system is in the ground state $|0\rangle$ at time $t < 0$, and that the external field $F(t)$ is only active between $t = 0$ and time $t_0 \geq 0$. At time t_0, the Hamiltonian reduces to the time-independent Hamiltonian such that Schrödinger's equation for $t \geq t_0$ becomes

$$\hat{H}_0 \left|\Psi(t)\right\rangle = \mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\left|\Psi(t)\right\rangle . \tag{5}$$

The time-dependent wave function in the absence of the external field oscillates around the solution at time t_0, $|\Psi(t_0)\rangle = \sum_n k_n(t_0)\,|n\rangle$, as given by

$$\left|\Psi(t)\right\rangle = \mathrm{e}^{-\mathrm{i}\hat{H}_0(t-t_0)}\left|\Psi(t_0)\right\rangle = \sum_n k_n(t_0)\mathrm{e}^{-\mathrm{i}E_n(t-t_0)}\left|n\right\rangle , \tag{6}$$

where $|n\rangle$ denotes a normalized eigenfunction of the unperturbed Hamiltonian, $\hat{H}_0\,|n\rangle = E_n\,|n\rangle$.[44,45] This formulation is exact when the electronic continuum is excluded, e.g., by choosing a fixed, finite basis as commonly done in quantum chemistry. Actual simulations are not performed in the energy eigenbasis but in, e.g., a basis of Slater determinants, implying that the coefficients $k_n(t_0)$ are not known.

In order to obtain the electronic absorption spectrum averaged over all molecular orientations relative to the electric field, the time-dependent electric dipole moment $\mu_u(t) = \langle\Psi(t)|\hat{\mu}_u|\Psi(t)\rangle$ is calculated in three separate simulations with the electric field polarized in each of the three Cartesian directions ($u = x, y, z$). The absorption cross-section is then obtained from the Fourier transform of the dipole moments, $\tilde{\mu}_u(\omega)$, as[46]

$$S(\omega) = \frac{4\pi\omega}{3c}\,\mathrm{Im}\left[\frac{\tilde{\mu}_x(\omega)\tilde{F}^*(\omega)}{|\tilde{F}(\omega)|^2} + \frac{\tilde{\mu}_y(\omega)\tilde{F}^*(\omega)}{|\tilde{F}(\omega)|^2} + \frac{\tilde{\mu}_z(\omega)\tilde{F}^*(\omega)}{|\tilde{F}(\omega)|^2}\right] , \tag{7}$$

where c is the speed of light. The resulting spectrum contains both linear (one-photon transitions between the ground and excited states) and nonlinear (multi-photon transitions between the ground and excited states, and one- and multi-photon transitions between excited states) absorptions, as recently stressed by Guandalini et al.[11] We note that only the *induced* dipole moment, that is the total dipole moment with the static ground-state part subtracted, contributes to the absorption cross section but, for notational convenience, we will only distinguish between that and the total dipole moment when it is strictly required.

Since the dipole moment is calculated on a finite discrete time grid, the Fourier transforms are replaced by discrete Fourier transforms, thus introducing artificial periodic boundary conditions. To avoid artefacts from these, the dipole moment is multiplied by a damping factor before the discrete Fourier transform, i.e.,

$$\tilde{\mu}_u(\omega) = \mathcal{F}[\mu_u(t)\mathrm{e}^{-\gamma|t|}] = \frac{1}{2\pi}\int_0^\infty \mu(t)\mathrm{e}^{\mathrm{i}(\omega+\mathrm{i}\gamma)t}\mathrm{d}t, \tag{8}$$

where we have used that the induced dipole moment vanishes for $t < 0$. The Fourier transform thus becomes a Laplace transform. The parameter $\gamma \in \mathbb{R}_+$ can be interpreted as a common (inverse) lifetime of all excited states, giving rise to Lorentzian line shapes in the simulated absorption spectra.[47] The discrete Fourier transform, however, requires a very large, often prohibitive, number of time steps to achieve sufficient spectral resolution. In the following sections, we will describe an extrapolation technique aiming at high resolution with short simulation time.

2.2 The expected form of the electric dipole moment

Once the external field is turned off, the time-dependent electric dipole moment evolves according to $\mu_u(t) = \langle\Psi(t)|\,\hat{\mu}_u\,|\Psi(t)\rangle$, where $|\Psi(t)\rangle$ is defined in Eq. (6). The dipole moment

oscillates with the Bohr frequencies $\omega_{nm} = E_n - E_m$ according to

$$\begin{aligned}\mu_u(t) =& 2\sum_{n>m}\Big\{ \mathrm{Re}\left[\langle n|\,\hat{\mu}_u\,|m\rangle\, k_n^*(t_0)k_m(t_0)\right]\cos(\omega_{nm}(t-t_0)) \\ &- \mathrm{Im}\left[\langle n|\,\hat{\mu}_u\,|m\rangle\, k_n^*(t_0)k_m(t_0)\right]\sin(\omega_{nm}(t-t_0))\Big\} \\ &+ \sum_n |k_n(t_0)|^2\,\langle n|\,\hat{\mu}_u\,|n\rangle\,,\end{aligned} \tag{9}$$

for time $t \geq t_0$.[48] The function form of the approximated dipole moment $\bar{\mu}_u(t) \approx \mu_u(t)$ will therefore be given by

$$\bar{\mu}_u(t) = c_0^u + \sum_{i=1}^{N_\omega^u}\left[c_i^u \sin(\omega_i^u(t-t_0)) + c_{N_\omega^u+i}^u \cos(\omega_i^u(t-t_0))\right], \tag{10}$$

where N_ω^u is the number of participating frequencies ω_i^u, each frequency with two independent linear coefficients c_i^u and $c_{N_\omega^u+i}^u$. If we can determine these frequencies and their corresponding real coefficients from a short dipole time series, we obtain a *continuous* dipole function and, hence, infinite spectral resolution.

As will be described in detail below, we estimate the participating frequencies using the poles of a Fourier-Padé approximant, while the linear coefficients are determined using linear regression in a subsequent step.

2.3 Estimating Bohr frequencies

In order to estimate the frequencies of the dipole moment, we will investigate the singular points of the Fourier-Padé approximant, originally introduced in real-time quantum simulations by Bruner et al.[22] In general, the Padé approximant is used to accelerate the convergence of a truncated power series. The discrete Fourier transform can be written as the power series

$$\tilde{\mu}_u(\omega) = \frac{\Delta t}{2\pi}\sum_{n=0}^{N_t-1}\mu_u(t_n)z^n, \tag{11}$$

where z depends on the frequency according to

$$z \equiv z(\omega) = \mathrm{e}^{(\mathrm{i}\omega-\gamma)\Delta t}. \tag{12}$$

The diagonal Fourier-Padé approximates the Fourier transform using two polynomials $P_u(z)$ and $Q_u(z)$ of degree $M = (N_t - 1)/2$,

$$[M/M]_{\mu_u}(\omega) = \frac{\Delta t}{2\pi} \frac{P_u(z(\omega))}{Q_u(z(\omega))}, \tag{13}$$

where the coefficients of the polynomials create a Toeplitz linear system. For details see Ref. 22. The Fourier-Padé poles, denoted z_p^u, are found by

$$Q_u(z_p^u) = 0, \tag{14}$$

where the damping parameter γ is set to zero, as the damping parameter removes the singularities of the spectrum. The Bohr frequencies are positive and real-valued, while the frequencies corresponding to roots of $Q_u(z)$ will be complex. The number of roots of $Q_u(z)$, (which amounts to M roots), should also significantly exceed the number of Bohr frequencies. The potential frequencies are given by

$$\omega_p^u = \left| \frac{\ln\left(z_p^u\right)}{\Delta t} \right|, \tag{15}$$

where $\ln\left(z_p^u\right)$ returns the principal value of the logarithm. Only roots $\mathrm{Im}(z_p^u) > 0$ are considered, as the complex conjugate root theorem states that complex roots will form conjugate pairs. These conjugate pairs yield duplicates of the real-valued frequencies. Real-valued roots z_p^u yield purely imaginary frequencies, and are therefore also excluded. The potential frequencies ω_p^u discard the imaginary component and should represent extrema of the Fourier-Padé spectrum, not singular points like z_p^u.

Estimating the potential frequencies uses the Python NumPy[49] library to compute the eigenvalues of the companion matrix[50] of the polynomial $Q_u(z)$ to determine its roots. This method exhibits poor scaling with respect to the number of time points N_t, representing a computational bottleneck of the dipole-moment $\bar{\mu}_u(t)$ fitting procedure. In real-time simulations using very small time steps, one may safely increase the step length on the dipole data used to create the Fourier-Padé approximant to alleviate the computational cost. As shown by Mattiat and Luber,[51] the convergence of the Fourier-Padé approximant is mostly impacted by the trajectory length $N_t\Delta t$, and not the time step itself. However, the discrete Fourier transform, and hence also the Fourier Padé, is periodic with a cycle length of $2\pi/\Delta t$. Peaks above $\pi/\Delta t$ will *fold back* due to anti-symmetry and appear as negative duplicates polluting the spectrum. Therefore, it is crucial to keep the time step $\Delta t < \pi/\omega_{\max}$, where $\omega_{\max}$ is the largest significant frequency in the signal.

The potential frequencies ω_p^u must be classified as either an estimated frequency or a redundant root. The classification is based on the assumption that $\ln\big(z_p^u\big)/(\mathrm{i}\Delta t)$ should have a significant imaginary component if ω_p^u is a redundant root, while it should lie close to the real axis if ω_p^u corresponds to an actual Bohr frequency. This further means that $Q_u(z(\omega_p^u))$ should be close to zero and that $[M/M]_{\mu_u}(\omega_p^u)$ should be large for estimated frequencies. Hence, we create a two-dimensional representation $\mathbf{r}_p^u$ of the prospective frequencies ω_p^u given by

$$[r_p^u]_x = 1 - \frac{X(\omega_p^u) - \min_{\omega_q^u}\left[X(\omega_q^u)\right]}{\max_{\omega_q^u}\left[X(\omega_q^u)\right] - \min_{\omega_q^u}\left[X(\omega_q^u)\right]}, \tag{16}$$

$$[r_p^u]_y = \frac{Y(\omega_p^u) - \min_{\omega_q^u}\left[Y(\omega_q^u)\right]}{\max_{\omega_q^u}\left[Y(\omega_q^u)\right] - \min_{\omega_q^u}\left[Y(\omega_q^u)\right]}, \tag{17}$$

where the unnormalized features are defined as

$$X(\omega_p^u) = \log_{10}\Big(\big|[M/M]_{\mu_u}(\omega_p^u)\big|\Big), \tag{18}$$

$$Y(\omega_p^u) = \log_{10}\Big(\big|Q_u(z(\omega_p^u))\big|\Big). \tag{19}$$

The base-10 logarithm is used to manage the extreme scaling of both features, as prospective frequencies should cause $Q_u(z(\omega_p^u))$ to approach zero and hence be a nearly singular point of $[M/M]_{\mu_u}(\omega_p^u)$. The features are constructed such that estimated frequencies should be close to $\mathbf{r}_p^u = (0,0)$, while redundant roots should be closer to $\mathbf{r}_p^u = (1,1)$.

We use the K-means clustering algorithm (see e.g. 52,53), implementation from the Python SciKit-Learn[54] library, with $K = 2$ to classify prospective frequencies. The 2-means clustering algorithm is a computationally inexpensive way to separate a set into two categories. The *centroid* for the cluster of potential frequencies should be closer to $(0,0)$, whereas the centroid for the redundancy cluster should be closer to $(1,1)$.

2.4 Determining the linear coefficients

Once the frequencies are estimated, the linear coefficients are determined using linear regression. The coefficients are optimized by minimizing the cost function,[55]

$$R(\mathbf{c}^u) = \sum_{n=0}^{N_t-1} \left[\mu_u(t_n) - \bar{\mu}_u(t_n; \mathbf{c}^u)\right]^2. \tag{20}$$

Using the general form of the dipole moment in Eq. (9), the linear coefficients may be optimized using a simple least squares optimization. The only restraint on the optimization of these coefficients is that they are real. This fitting procedure is general for any type of external field, $F(t)$. However, as shown by Hauge,[48] restricting the coefficients is crucial to avoid over-fitting the dipole moment.

The form $\bar{\mu}_u(t)$ in Eq. (10) is based on the full dipole moment, correct through all orders in perturbation theory and is independent of the electric field. In this work, we will use a Dirac delta-type impulse[8] of strength κ,

$$F(t) = \kappa\delta(t), \tag{21}$$

which has an infinitely wide frequency distribution and thus generates the full absorption spectrum for the given polarization direction. This implies that $t_0 = 0$. Further, we assume that the electric field strength is sufficiently weak, such that we may regard the interaction operator $\hat{V}(t)$ as a time-dependent perturbation and assume that the interaction only induces one-photon transitions from the ground state—i.e., a linear absorption spectrum. The electric dipole moment should then be of the form $\mu_u(t) \approx \mu_u^{(0)} + \mu_u^{(1)}(t)$, where the zeroth order dipole moment corresponds to the ground-state value, $\mu_u^{(0)} = \mu_u(t = 0)$. We will now investigate an analytical expression for the first-order correction to the dipole moment induced by a weak Dirac delta impulse.

We start with the exact expression for the linear response function,[6]

$$\langle\langle \hat{\mu}_u; \hat{V}_u(\omega) \rangle\rangle_{\omega+\mathrm{i}\gamma} = -2\tilde{F}(\omega) \sum_{n\neq 0} |\langle 0| \hat{\mu}_u |n\rangle|^2 \frac{\omega_{n0}}{(\omega + \mathrm{i}\gamma)^2 - \omega_{n0}^2}, \tag{22}$$

where we have used the Fourier transform of the interaction operator $\hat{V}(t)$,

$$\hat{V}_u(\omega) = \hat{\mu}_u \tilde{F}(\omega), \qquad F(\omega) = \frac{\kappa}{2\pi}. \tag{23}$$

The linear response function and the first-order correction to the dipole moment are related by $\langle\langle \hat{\mu}_u; \hat{V}_u(\omega) \rangle\rangle_{\omega+\mathrm{i}\gamma} = \mathcal{F}[\mu_u^{(1)}(t)\mathrm{e}^{-\gamma|t|}]$. Since $\mu_u^{(1)}(t < 0) = 0$, we get the relation

$$\int_0^\infty \mu_u^{(1)}(t)\mathrm{e}^{-(\gamma-\mathrm{i}\omega)t}\mathrm{d}t = -2\kappa \sum_{n\neq 0} |\langle 0| \hat{\mu}_u |n\rangle|^2 \frac{\omega_{n0}}{(\omega + \mathrm{i}\gamma)^2 - \omega_{n0}^2}. \tag{24}$$

Using well-known Laplace transforms, it is readily verified that the first-order dipole correction must be a linear combination of sine waves[48]

$$\mu_u^{(1)}(t) = \sum_{n\neq 0} B_n^u \sin(\omega_{n0} t), \qquad B_n^u = 2\kappa |\langle 0| \hat{\mu}_u |n\rangle|^2. \tag{25}$$

We have also used that the first-order perturbation correction to the dipole moment should

only include one-photon transitions. This further means that the approximated dipole moment, when using a weak Dirac delta impulse, should have the form

$$\bar{\mu}_u(t) = c_0^u + \sum_{i=1}^{N_\omega^u} c_i^u \sin(\omega_i^u t), \tag{26}$$

where all sine coefficients are positive.

The coefficients of $\bar{\mu}_u(t)$, approximating the dipole moment from the Dirac delta impulse, are optimized using the *least absolute shrinkage and selection operator (LASSO)*[56] method. The coefficients are determined according to

$$\mathbf{c}_{\text{LASSO}}^u = \underset{\mathbf{c}^u}{\text{argmin}} \left\{ \frac{1}{2} R(\mathbf{c}^u) + \lambda \sum_i |c_i^u| \right\}, \tag{27}$$

where λ is the *shrinkage parameter* restricting the magnitude of the coefficients, c_i^u. In contrast to the ordinary linear least-squares algorithm, the LASSO method is iterative and therefore somewhat less computationally efficient. In return, this makes it possible to enforce positive coefficients, as in the implementation by SciKit-Learn.[54] This makes the method less prone to over-fitting.

2.5 Molecular orbital decomposition

The electric dipole moment can be written as a sum of contributions from elementary molecular orbital (MO) transitions,[8]

$$\mu_u(t) = \sum_{ia} \mu_u^{ia}(t), \tag{28}$$

where i and a label occupied and virtual MOs, respectively. The components $\mu_u^{ia}(t)$ are then approximated separately. This MO decomposition can divide a dense spectrum into a series of sparser spectra and aid in the assignment of absorption lines.[8] Clustering the MO components into groups can be used to offset the increased memory consumption.[57] For the fitting method, creating clusters with well-separated frequencies could also reduce the

accumulation of errors when summing the component fits.

When fitting the individual components, the assumptions on the sign of the linear coefficients are no longer valid. As is clear from the underlying theory and as demonstrated in practice by Bruner et al.,[22] the same frequencies may be found in several components μ_u^{ia}, and their corresponding partial spectra may contain negative peaks. Only the full spectrum, i.e., the sum of the components, is guaranteed to contain positive peaks exclusively. The ordinary least squares method must therefore be used when optimizing the linear coefficients of the individual components, which may introduce additional errors due to over-fitting in each component.

Alternatively, the fitting algorithm may estimate the frequencies of each component separately and then optimize the linear coefficients for the full dipole moment. This way, the additional coefficient restrictions can be used in the optimization. In our experience, however, this produces a vast number of estimated frequencies leading to problems with over-fitting even when enforcing positive linear coefficients.

2.6 Convergence criterion

The goal of the fitting method is to accurately construct the function $\bar{\mu}(t)$ using the shortest possible dipole trajectory. A given trajectory is divided into two parts, a fitting domain and a verification domain. The linear coefficients are optimized using only the fitting domain, while the error is calculated on the verification domain. When estimating the frequencies, however, the entire available trajectory is used. Measuring the error in the fitting domain gives the *interpolation* error, which is artificially low in cases of over-fitting, whereas the error in the verification window indicates the reliability of the extrapolated dipole moment. The error of the fit is estimated using one minus the coefficient of determination, R^2, i.e.,

$$E_u = 1 - R^2 = \frac{\sum_n \left[\mu_u(t_n) - \bar{\mu}_u(t_n)\right]}{\sum_n \left[\mu_u(t_n) - \mu_u^m\right]}, \tag{29}$$

where μ_u^m is the mean value of the induced dipole moment. The error measure is unitless and independent of the magnitude of the dipole moment. The fitting method can be run in parallel with real-time simulations, which are terminated once E_u drops below a pre-defined threshold value. The computational cost of the fitting method is not insignificant, and we recommend that it is run once per time intervals of $50 - 100$ a.u. when used to automatically terminate the real-time simulation.

Computing the error according to Eq. (29) provides an error estimate of the fit as a whole. In our experience from testing the algorithm with ideal multi-sinusoidal signals, the error of the fit depends primarily on the frequency estimation. Significant deviations in the estimated linear coefficients were only observed if there were frequencies missing and/or poorly estimated. In the case of ideal signals, there would only be a significant error in the fit if the frequency estimation failed. Real dipole data contains noise introduced by numerically integrating in time. How this affects the distribution of error is unknown, though it is reasonable to believe that the main source of error still lies in the frequency estimation.

The error measure E_u cannot distinguish error contributions from different parts of the spectrum, preventing termination once the desired frequency region is converged. In order to focus on valence excitations in the low-frequency region, we apply a low-pass (smoothing) filter to remove frequencies above a cut-off frequency $\omega_{\max}$ from the dipole moment in the time domain. We used a Butterworth filter, implemented by SciPy,[58] which removes the high-energy part of the spectrum while leaving the lower-energy part almost unchanged. If bound core excitations are the main targets, a high-pass filter must be used instead.

3 Computational details

We test the dipole extrapolation scheme using RT-TDDFT simulations, supplemented by a few RT-TDCIS simulations to demonstrate its applicability to wave-function-based theories. The RT-TDDFT simulations are performed using the ReSpect program,[59] while the

RT-TDCIS calculations are performed using the Hylleraas Quantum Dynamics (HyQD) software.[60] The RT-TDDFT and RT-TDCIS simulations are performed with analytic integration at $t = 0$ a.u., as described in Ref. 8. The subsequent time steps are performed numerically using the Magnus integrator for the RT-TDDFT simulations[8] and the three-stage Gauss-Legendre integrator[61] as described in Ref. 62 with the residual norm convergence criterion 10^{-14} a.u. for the implicit equations for the RT-TDCIS simulations.

The RT-TDCIS simulations are performed with time step $\Delta t = 0.01$ a.u. and field strength $\kappa = 10^{-3}$ a.u. The RT-TDDFT simulations for the organic molecules CH_4, CH_2O, CH_3OH, C_2H_6, and C_6H_6 are performed with time step $\Delta t = 0.01$ a.u. and field strength $\kappa = 10^{-4}$ a.u., while $\Delta t = 0.01$ a.u. and $\kappa = 10^{-3}$ a.u. are used for CO_2, H_2O, and NH_3. For the smallest systems, He, H_2, Be, and LiH, $\Delta t = 0.1$ a.u. and $\kappa = 10^{-3}$ a.u. are used.

Molecular geometries are found in the Supporting Information. The simulations were performed in Dunning's cc-pVXZ and aug-cc-pVXZ, X = D,T, basis sets[63–65] (uncontracted in the case of RT-TDDFT calculations). The RT-TDDFT simulations were performed using the PBE0 exchange–correlation potential[66–69] in the adiabatic approximation.

Simulations are performed for all three Cartesian directions for all systems, even in cases where point-group symmetry could have been easily exploited to reduce the computational effort to one or two directions. While this is mainly done to make automation simple (identical treatment for all systems), it also provides a simple check that the extrapolation algorithm does not significantly break point-group symmetry due to numerical noise in the input dipole trajectories.

Our implementation of the dipole-extrapolation algorithm is freely available at `https://github.com/HyQD/absorption-spectrum`.

4 Results

All reference spectra in this paper are produced from low-pass filtered electric dipole moments with a trajectory length of 4000 a.u., such that the spectral resolution becomes $\Delta\omega = 1.6 \cdot 10^{-3}$ a.u.. In this paper, the resolution of the fitted spectrum $\bar{S}(\omega)$ is the same as its reference spectrum. This is to allow direct comparisons of the two spectra, though the resolution of $\bar{S}(\omega)$ could be made arbitrarily fine. The Fourier transform of the approximated dipole moment $\bar{\mu}_u$ is calculated according to

$$\tilde{\mu}_u(\omega) = -\frac{1}{2\pi} \sum_i c_i^u \frac{\omega_i^u}{(\omega + \mathrm{i}\gamma)^2 - (\omega_i^u)^2}. \tag{30}$$

The half-life parameter was always set to $\gamma = 0.5 \cdot 10^{-3}\pi$, and the spectra were cut at an estimated ionization energy of 0.5 a.u. $- \epsilon_{\mathrm{HOMO}}$, where the energy of the highest occupied molecular orbital ϵ_{HOMO} of all systems are listed in the Supporting Information.

Using the low-pass filter will leave the lower energy part of the absorption spectrum unaltered, while the higher energy part is removed and set to zero. Differences between filtered and unfiltered spectra are shown in the Supporting Information. The low-pass filter does not give a clean cut-off at the cut-off frequency $\omega_{\max}$ but rather a gradual lowering of the peak intensity around $\omega_{\max}$. The cut-off frequency should therefore be set somewhat higher than the desired range of the spectrum. We have used $\omega_{\max} = 4$ a.u. for all systems.

When fitting the dipole moment, the available trajectory is from when the external field is turned off at $t = 0$ to time $t = T_{\mathrm{ver}}^u$. The linear coefficients are determined on the time interval $[0, T_{\mathrm{fit}}^u]$, where $T_{\mathrm{fit}}^u = 0.75 T_{\mathrm{ver}}^u$. The frequencies are estimated on the entire available trajectory $[0, T_{\mathrm{ver}}^u]$ but with a limit on the total number of data points supplied to the Padé, set to $5 \cdot 10^3$. The reduction in points, if exceeding the limit, is done by effectively increasing the time step Δt used (by an integer factor) when creating the Padé. The error E_u is only evaluated on $(T_{\mathrm{fit}}^u, T_{\mathrm{ver}}^u]$. The error in the spectrum E_S is calculated the same way as in the time domain, as given in Eq. (29).

4.1 Performance on a selection of systems

For each spatial direction, the convergence of the fit is tested every 50 a.u. in time of the trajectory length, starting from $T_{\min} = 100$ a.u.. The simulation is terminated when the fit has converged below a given threshold or when the trajectory length reaches $T_{\max} = 1000$ a.u., which corresponds to a target minimum spectral resolution of 0.006 a.u. The convergence criterion was set to $E_u < 10^{-3}$, a strict threshold corresponding to a near-perfect fit. The criterion was set based on preliminary investigations.[48] Since the real-time calculations on a given system using three spatial directions of the external field are independent, the trajectory length needed for a converged fit might vary between the three simulations.

The required trajectory length of each spatial direction T^u_{ver} and their corresponding verification error E_u as well as the error in the spectrum E_S are listed in Tables 1 and 2. The fitting of the dipole moment from RT-TDCIS calculations is shown in Table 1, while the fitting of RT-TDDFT data is found in Table 2. Figures of the approximated spectra of all systems can be found in the Supporting Information.

Table 1: Convergence times and corresponding errors of systems from RT-TDCIS calculations.

	basis	T^x_{ver} [a.u.]	T^y_{ver} [a.u.]	T^z_{ver} [a.u.]	E_x	E_y	E_z	E_S
CH_2O	aug-cc-pVDZ	450	600	650	$8 \cdot 10^{-6}$	$8 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$
CO_2	cc-pVDZ	100	100	100	$8 \cdot 10^{-5}$	$8 \cdot 10^{-5}$	$3 \cdot 10^{-6}$	$2 \cdot 10^{-4}$
	aug-cc-pVDZ	250	250	200	$6 \cdot 10^{-5}$	$6 \cdot 10^{-5}$	$3 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
	aug-cc-pVTZ	300	300	250	$2 \cdot 10^{-4}$	$2 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
H_2O	aug-cc-pVDZ	150	200	300	$1 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$7 \cdot 10^{-5}$	$3 \cdot 10^{-4}$
NH_3	aug-cc-pVDZ	350	300	300	$4 \cdot 10^{-5}$	$7 \cdot 10^{-4}$	$7 \cdot 10^{-6}$	$3 \cdot 10^{-3}$

The fitting method reached the strict threshold for most systems, with a maximum spectral error of $E_S \leq 3 \cdot 10^{-3}$. For all converged systems, the approximated functions for the dipole moment $\bar{\mu}_u(t)$ reliably reproduce its reference spectrum. The systems with very sparse spectra (He, H_2, and Be) converged instantly ($T_{\text{ver}} = 100$ a.u.), providing approximated spectra indistinguishable from their reference spectra. In these cases, the fitting method

Table 2: Convergence times and corresponding errors of systems from RT-TDDFT calculations.

	basis	T^x_{ver} [a.u.]	T^y_{ver} [a.u.]	T^z_{ver} [a.u.]	E_x	E_y	E_z	E_S
Be	aug-ucc-pVTZ	100	100	100	$1 \cdot 10^{-8}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-8}$	$6 \cdot 10^{-6}$
C_2H_6	aug-ucc-pVDZ	750	800	550	$6 \cdot 10^{-4}$	$7 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$8 \cdot 10^{-4}$
	aug-ucc-pVTZ	1000	1000	950	$5 \cdot 10^{-2}$	$3 \cdot 10^{-2}$	$2 \cdot 10^{-5}$	$6 \cdot 10^{-3}$
C_6H_6	aug-ucc-pVDZ	1000	1000	550	$3 \cdot 10^{-2}$	$4 \cdot 10^{-2}$	$7 \cdot 10^{-4}$	$4 \cdot 10^{-2}$
CH_2O	aug-ucc-pVDZ	450	650	700	$4 \cdot 10^{-6}$	$8 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$5 \cdot 10^{-4}$
	aug-ucc-pVTZ	650	900	1000	$2 \cdot 10^{-4}$	$4 \cdot 10^{-4}$	$2 \cdot 10^{-3}$	$8 \cdot 10^{-4}$
CH_3OH	aug-ucc-pVDZ	1000	1000	1000	$2 \cdot 10^{-1}$	$8 \cdot 10^{-1}$	$4 \cdot 10^{-2}$	$8 \cdot 10^{-2}$
	aug-ucc-pVTZ	1000	1000	1000	$3 \cdot 10^{-1}$	$8 \cdot 10^{-1}$	$4 \cdot 10^{-1}$	$2 \cdot 10^{-1}$
CH_4	aug-ucc-pVDZ	200	200	200	$8 \cdot 10^{-4}$	$7 \cdot 10^{-4}$	$6 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
	aug-ucc-pVTZ	350	350	350	$4 \cdot 10^{-4}$	$3 \cdot 10^{-4}$	$3 \cdot 10^{-4}$	$3 \cdot 10^{-3}$
CO_2	aug-ucc-pVDZ	350	350	250	$4 \cdot 10^{-4}$	$4 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$5 \cdot 10^{-4}$
H_2O	aug-ucc-pVDZ	200	250	300	$2 \cdot 10^{-6}$	$3 \cdot 10^{-6}$	$3 \cdot 10^{-4}$	$3 \cdot 10^{-4}$
H_2	aug-ucc-pVTZ	100	100	100	$1 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$8 \cdot 10^{-8}$	$5 \cdot 10^{-6}$
He	aug-ucc-pVTZ	100	100	100	$9 \cdot 10^{-8}$	$9 \cdot 10^{-8}$	$1 \cdot 10^{-7}$	$9 \cdot 10^{-6}$
LiH	aug-ucc-pVDZ	100	100	300	$1 \cdot 10^{-5}$	$3 \cdot 10^{-5}$	$3 \cdot 10^{-4}$	$3 \cdot 10^{-4}$
NH_3	aug-ucc-pVDZ	350	400	300	$2 \cdot 10^{-4}$	$4 \cdot 10^{-4}$	$7 \cdot 10^{-4}$	$1 \cdot 10^{-3}$

achieved a speedup of 10 times compared to the max trajectory length of $T_{\max} = 1000\,\text{a.u.}$, or 40 times compared to computing the reference spectra (using 4000 a.u.). The reduction in computational cost achieved by using the fitting method is relative to the desired spectral resolution. We would argue that the least unambiguous way to assess the speedup is to compare the convergence times T^u_{ver} with the simulation time that would have been used if the fitting method was not used, i.e. the max trajectory length. Should the method converge at the preset max trajectory length, one may argue that no simulation time was spared. In this case, one still achieves arbitrary improvement in the spectral resolution.

Systems with relatively sparse spectra (CH_4, CO_2, H_2O, LiH, and NH_3) also converged nicely with short dipole trajectories ($T^u_{\text{ver}} \leq 350\,\text{a.u.}$). As the spectral density increases, the fitting method struggles to converge. Systems like C_2H_6 and CH_2O only converged when using a double-zeta basis set, while the fitting of C_6H_6 and CH_3OH did not achieve errors below the low threshold.

The CH_2O molecule with a double-zeta basis set converged for both real-time methods. The spectra of the fit in both cases are nearly indistinguishable from their reference spectra. A comparison between the approximated and reference spectrum from RT-TDDFT calculations is shown in Fig. 1. The RT-TDDFT triple-zeta case nearly reached the error threshold ($E_z = 2 \cdot 10^{-3}$), also providing a very low spectral error ($E_S = 8 \cdot 10^{-4}$).

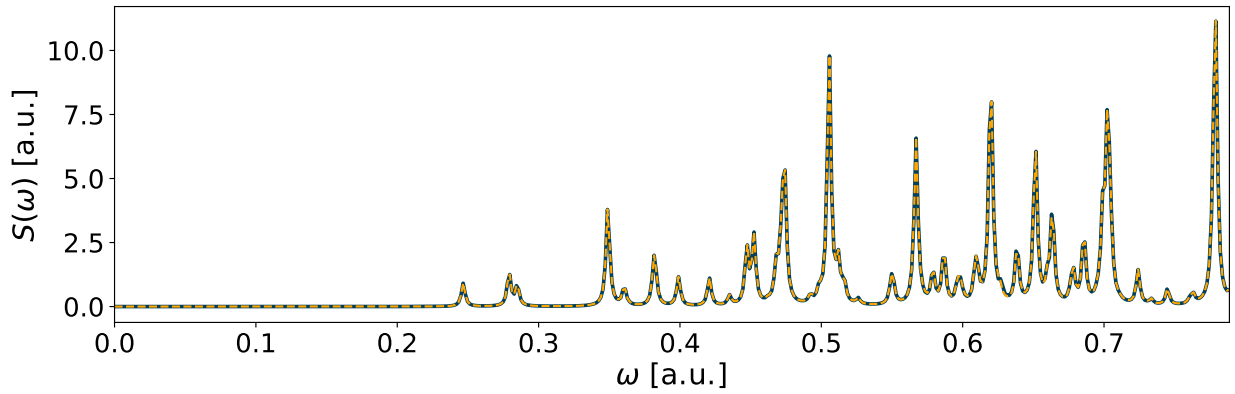

Figure 1: Spectrum of CH_2O using the aug-ucc-pVDZ basis in a RT-TDDFT simulation. The reference spectrum is in solid blue, while the yellow dashed line shows the spectrum of the fitted functions. The fitting error was $E_x = 4 \cdot 10^{-6}$, $E_y = 8 \cdot 10^{-4}$, and $E_z = 1 \cdot 10^{-4}$.

Among the converged systems, NH_3 from RT-TDCIS calculations showed the largest error compared to its reference spectrum ($E_S = 3 \cdot 10^{-3}$). Its spectrum is shown in Fig. 2, and was the approximated spectrum with the most visible deviation from its reference spectrum among the converged systems. The approximated spectrum shows a deviation in a peak at $\omega \approx 0.75\,\text{a.u.}$, but the rest of the peaks correspond well to the reference spectrum.

The fitting method only partially converged for C_6H_6, as well as C_2H_6 and CH_2O with triple-zeta basis, meaning that error of the fit was below the set threshold in only one or two of the spatial directions. Still, the spectral error in all three cases is quite low. The result of the fitting of benzene is shown in Fig. 3, which had the largest spectral error ($E_S = 4 \cdot 10^{-2}$) of the three.

The trajectory length needed for the fitting method to converge strongly depends on the spectral density. We observed a trend in that the fitting becomes increasingly difficult as the spectral density increases. Increasing either the number of electrons in the system or

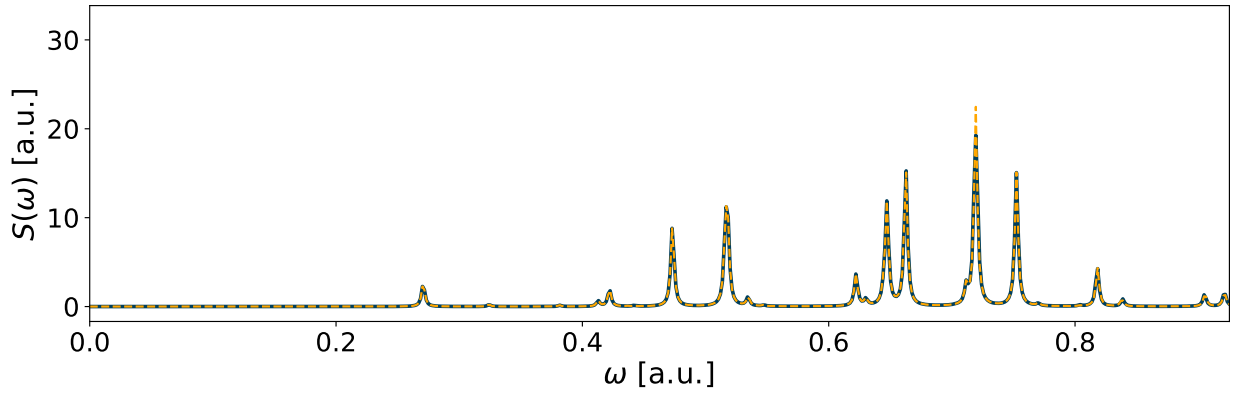

Figure 2: Spectrum of NH_3 using the aug-ucc-pVDZ basis in a RT-TDDFT simulation. The reference spectrum is in solid blue, while the yellow dashed line shows the spectrum of the converged fitted functions. This was the poorest approximated spectrum of all converged cases. The fitting error was $E_x = 2 \cdot 10^{-4}$, $E_y = 4 \cdot 10^{-4}$, and $E_z = 7 \cdot 10^{-4}$.

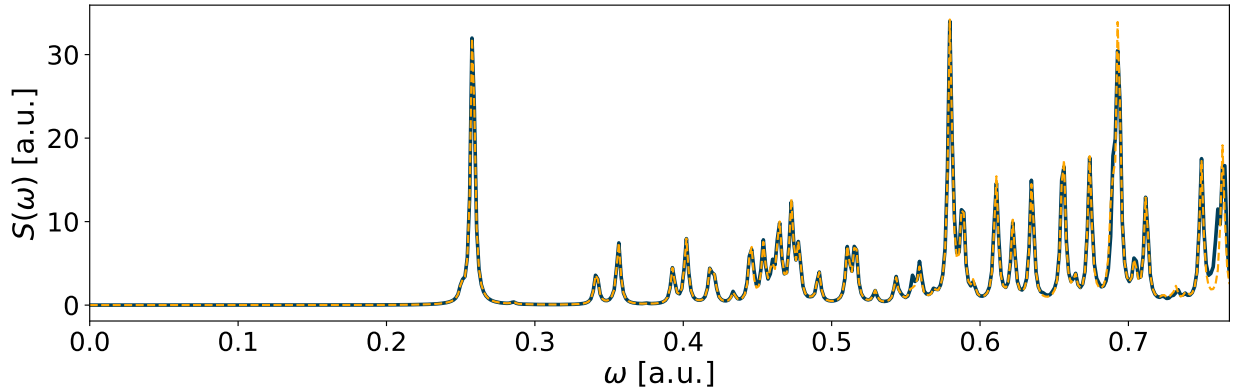

Figure 3: Spectrum of C_6H_6 using the aug-ucc-pVDZ basis in a RT-TDDFT simulation. The reference spectrum is in solid blue, while the yellow dashed line shows the spectrum of the fitted functions. The fitting error was $E_x = 3 \cdot 10^{-2}$, $E_y = 4 \cdot 10^{-2}$, and $E_x = 7 \cdot 10^{-4}$.

the size of the basis set will in general require longer real-time simulations before the fitting method converges. The trend with increasing basis set size is clearly seen from the fitting of CO_2 from RT-TDCIS calculations. The simulation using the cc-pVDZ basis set converges faster ($T^u_{\text{ver}} = 100$ a.u.) than when using the larger basis sets like the aug-cc-pVDZ basis set ($T^u_{\text{ver}} \leq 250$ a.u.) or aug-cc-pVTZ basis set ($T^u_{\text{ver}} \leq 300$ a.u.).

Only the fit of CH_3OH did not get errors below the convergence threshold in any of the spatial directions. This was true for both the double and tripe-zeta basis (from RT-TDDFT calculations). The result using a triple-zeta basis set is shown in Fig. 4 and is the case with

the highest error in the time domain, $E_u \sim 10^{-1}$. There is a significant deviation from the reference spectrum, though the main features are intact. Of all systems in this paper, this gave the worst approximation to the reference spectrum. Despite this, the spectrum $\bar{S}(\omega)$ still provides a reasonable coarse approximation.

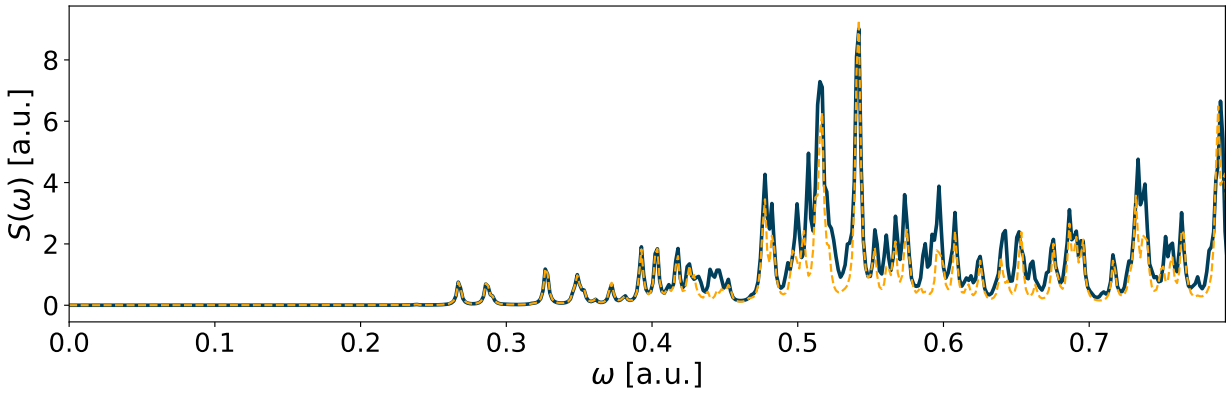

Figure 4: Spectrum of CH_3OH using the aug-ucc-pVTZ basis in a RT-TDDFT simulation. The reference spectrum is in solid blue, while the yellow dashed line shows the spectrum of the fitted functions. The fitting error was $E_x = 3 \cdot 10^{-1}$, $E_y = 8 \cdot 10^{-1}$, and $E_y = 4 \cdot 10^{-1}$.

These results are promising in all cases as the converged fit seems to reproduce its reference spectrum reliably with only minor deviations in the peak intensities. The error of the fit E_u also correlates with the spectral error, E_S. This predictability is crucial if the convergence criterion is used to automatically terminate real-time simulations. Our results also indicate that the convergence criterion used in this study is stricter than necessary. A slight relaxation in the criterion might lead to faster convergence without significantly impacting the quality of the approximated spectrum.

For the estimated dipole moment, the frequencies and their corresponding linear coefficients are known. For a successful fit, one may therefore obtain the transition probability $|\langle 0|\,\hat{\mu}_u\,|n\rangle|^2$ of a transition with energy $E_n - E_0$ directly from the linear coefficient, as $|\langle 0|\,\hat{\mu}_u\,|n\rangle|^2 = B_n^u/(2\kappa)$. This could be used to calculate the oscillator strength and create stick spectra. However, estimated frequencies in different spatial directions but corresponding to the same transition will have a small error associated with the frequencies. In order to compute the oscillator strength, one would therefore have to assess which estimated fre-

quencies across spatial directions correspond to the same transition.

The convergence of the dipole moment fitting depends primarily on the frequency estimation. When the fit does not converge, it follows that the Fourier-Padé approximant is not sufficiently converged to accurately capture the Bohr frequencies. The quality of the Fourier-Padé depends on the dipole trajectory length t_N rather than the number of steps or step length.[51] However, there is no given final time t_N ensuring convergence, the necessary trajectory length depends on the spectral density. High spectral density can cause the Fourier-Padé to fail, even for relatively long simulations. The general Padé approximant is prone to instabilities due to problems with near-degeneracy of the linear system. As pointed out by Cooper et al.,[70] the Fourier-Padé used in real-time spectroscopy is known to struggle with dense spectra. The fitting method introduces a measure of the error E_u which does not rely on any reference spectrum. This introduces a more reliable way of estimating the error in the approximated spectrum.

4.2 Fitting using MO decomposition

We assessed the performance of the fitting procedure used to extrapolate the components of the dipole moment of C_6H_6 decomposed to MO pairs $\bar{\mu}_u^{ia}$ in the RT-TDDFT calculation. Instead of creating a fitting function for each individual MO pair, which would increase the memory overhead, we clustered the components $\bar{\mu}_u^{ia}$ into groups of ten. These groups are formed so that the overall sparsity of spectra obtained for each cluster is maintained. This is accomplished by spreading the individual constituents of the cluster across the energy range. A fitted function of each cluster was then created from the sum over the MO pairs that the cluster contains. The total error is measured for the full dipole moment, $\mu_u(t)$. Some of the MO pairs were omitted entirely, making the MO decomposition work as a low-pass filter. When fitting components, the low-pass filter is therefore not needed.

Fitting the decomposed signal, however, did not improve the convergence compared to when the full dipole moment was used for extrapolation. Simulations for both directions

μ_x and μ_y reached the max trajectory length ($t = 1000\,\text{a.u.}$) without the fitting error going below the error threshold. The errors ($E_x = 1 \cdot 10^{-2}$ and $E_y = 1 \cdot 10^{-2}$) were only slightly lower compared to fitting without the MO decomposition. The last spatial direction μ_z converged at $T_{\text{ver}} = 650\,\text{a.u.}$ ($E_z = 4 \cdot 10^{-4}$), which is somewhat slower than without MO decomposition. The spectral error was $E_S = 9 \cdot 10^{-3}$, which corresponds to a low spectral error.

Although the MO decomposition did not lead to accelerated convergence of the fitting method, we still observed improvements. For example the simulation with $T^u_{\text{ver}} = 600\,\text{a.u.}$ has a lower error of the fit for the decomposed dipole moment ($E_x = 6 \cdot 10^{-2}$, $E_y = 4 \cdot 10^{-2}$ and $E_z = 1 \cdot 10^{-3}$) for all spatial directions compared to the fit on the full dipole moment, ($E_x = E_y = 3 \cdot 10^{-1}$ and $E_z = 2 \cdot 10^{-3}$). The decomposed fit in Fig. 5 ($E_S = 3 \cdot 10^{-2}$) is visibly improved compared to the fit using the full dipole moment in Fig. 6 ($E_S = 2 \cdot 10^{-1}$).

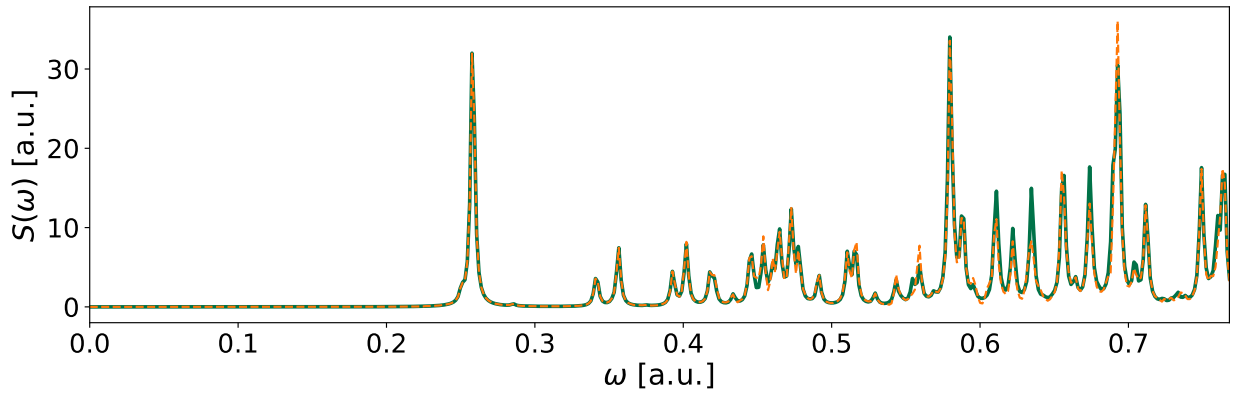

Figure 5: Spectrum of C_6H_6 using the aug-ucc-pVDZ basis in a RT-TDDFT simulation. The reference spectrum is in solid green, while the orange dashed line shows the spectrum of the fitted functions using molecular orbital decomposition from $T^u_{\text{ver}} = 600\,\text{a.u.}$.

It is important to note that the scope of our testing of the fitting method using MO decomposition was limited. Previous success using the Fourier-Padé approximant in combination with the MO decomposition on RT-TDDFT data suggests that this in many cases is very effective.[22] Our study, however, raises cause for caution regarding the use of the Fourier-Padé approximant: The Padé *can* struggle, even when using MO decomposition. The unknown amount of error introduced to the final spectrum by this procedure remains

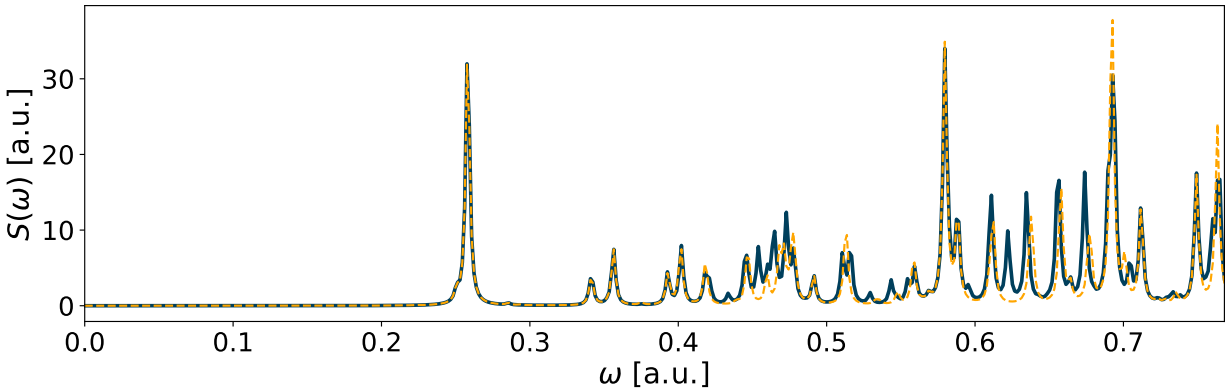

Figure 6: Spectrum of C_6H_6 using the aug-ucc-pVDZ basis in a RT-TDDFT simulation. The reference spectrum is in solid blue, while the yellow dashed line shows the spectrum of the fitted functions from $T^u_{\text{ver}} = 600\,\text{a.u.}$.

an open problem that the user should be aware of when analyzing spectra with the Fourier-Padé method with the MO decomposition.

The particular form of the components μ_u^{pq} is also very dependent on the quantum mechanical method used to compute the time-dependent wave function. Using MO decomposition on the electric dipole moment from RT-TDCCSD calculations leads to large overlaps in frequencies among different components.[48] The usefulness of such decomposition might vary between the different quantum mechanical frameworks.

5 Conclusion

We have developed a novel method for creating functions approximating the electric dipole moment from real-time calculations. The fitted functions for the dipole moment in the three spatial directions can then be used to produce absorption spectra with arbitrary high resolution. Real-time calculations of absorption spectra require the use of the discrete Fourier transforms, demanding long simulation times to obtain high spectral resolution. In our work, we have shown that the length of the real-time simulations, and hence the computational cost, can be greatly reduced by the developed fitting method.

We introduced a quantitative error measure to evaluate the convergence of the fit. For

all systems in this work, a converged fit reliably reproduced the reference spectrum from long real-time calculations. A convergence criterion of 10^{-3} seems to be quite strict, and further studies should be conducted to investigate the impact of slightly higher errors on the estimated spectrum. In order to reduce the computational cost of calculating absorption spectra, the real-time calculations should be automatically terminated once the convergence criterion is reached.

In this work, we set the verification window to be 25% of the available dipole trajectory. The critical step of the method is determining the frequencies, which always uses all available data. For the linear optimization, the verification window should include an entire period of the smallest frequency in the signal, as an insufficiently large verification window may lead to misleading error estimates. In future work, the verification window should depend on an estimate of the smallest frequency in the signal based on differences in the molecular orbital energies.

The fitting method converged with as little as 100 a.u. long trajectories in time for systems with sparse spectra. Convergence slows down as spectral density increases, even leading to failure of convergence in some cases. The current version of the fitting method shows encouraging results for smaller systems, although aspects of the method require further investigation.

Our testing of the fitting method using molecular orbital decomposition of a single system gave mixed results. The decomposition did not enable the fit to meet the convergence criterion, although we observed improvements in the approximated spectrum. This motivates the need for further investigations.

An apparent weakness of the current implementation is the way of estimating frequencies. Future versions should not rely on the Fourier-Padé but rather investigate other methods of estimating frequencies. This could include other methods for harmonic inversion or letting the function form of the fitted dipole moment be a truncated Fourier series based on an estimation of the fundamental frequency. The same frequencies can appear in all spatial

directions, which could be exploited to improve the frequency estimation. In particular, in cases where the frequencies are successfully estimated in one spatial direction, knowledge of these existing frequencies could be used to alleviate the search in the other spatial directions with potentially higher spectral density. Improving the frequency estimation is crucial for stabilizing the fitting method for systems with high spectral density.

This work has focused on the Dirac delta impulse, though the general fitting algorithm may be used on systems with any type of external field. Using a laser pulse targeting a specific spectral region may provide both an upper and lower bound when estimating the frequencies. Any *a priori* information about the frequencies should be exploited by the fitting algorithm. Additionally, the Dirac delta impulse targets *all* excitation energies, maximizing the spectral density. It is not unlikely that a narrow-band laser pulse would somewhat alleviate the fitting process by reducing the spectral density.

Acknowledgement

This work was supported by the Research Council of Norway through its Centres of Excellence scheme, project number 262695, and its Mobility Grant scheme (project nos. 301864 and 314814). The simulations were performed on resources provided by Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway, Grant No. NN4654K. T. B. P. acknowledges the support of the Centre for Advanced Study in Oslo, Norway, which funded and hosted the CAS research project *Attosecond Quantum Dynamics Beyond the Born-Oppenheimer Approximation* during the academic year 2021-2022. In addition, this project received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 945478 (SASPRO2), and the Slovak Research and Development Agency (Grant No. APVV-21-0497).

Supporting Information Available

Molecular geometries and HOMO energies for all systems. Comparisons between the spectra of filtered versus unfiltered dipole moments. Spectra of the fitted functions of the dipole moments compared with the corresponding reference spectrum.

References

(1) Corkum, P. B.; Krausz, F. Attosecond science. *Nat. Phys.* **2007**, *3*, 381–387.

(2) Nisoli, M.; Decleva, P.; Calegari, F.; Palacios, A.; Martín, F. Attosecond Electron Dynamics in Molecules. *Chem. Rev.* **2017**, *117*, 10760–10825.

(3) Li, X.; Govind, N.; Isborn, C.; DePrince, A. E.; Lopata, K. Real-Time Time-Dependent Electronic Structure Theory. *Chem. Rev.* **2020**, *120*, 9951–9993.

(4) Lode, A. U.; Lévêque, C.; Madsen, L. B.; Streltsov, A. I.; Alon, O. E. Multiconfigurational time-dependent Hartree approaches for indistinguishable particles. *Rev. Mod. Phys.* **2020**, *92*, 011001.

(5) Ofstad, B. S.; Aurbakken, E.; Schøyen, Ø. S.; Kristiansen, H. E.; Kvaal, S.; Pedersen, T. B. Time-Dependent Coupled-Cluster Theory. *WIREs Comput. Mol. Sci.* **2023**, (in press. DOI: 10.1002/wcms.1666).

(6) Olsen, J.; Jørgensen, P. Linear and nonlinear response functions for an exact state and for an MCSCF state. *J. Chem. Phys.* **1985**, *82*, 3235–3264.

(7) Helgaker, T.; Coriani, S.; Jørgensen, P.; Kristensen, K.; Olsen, J.; Ruud, K. Recent Advances in Wave Function-Based Methods of Molecular-Property Calculations. *Chem. Rev.* **2012**, *112*, 543–631.

(8) Repisky, M.; Konecny, L.; Kadek, M.; Komorovsky, S.; Malkin, O. L.; Malkin, V. G.; Ruud, K. Excitation Energies from Real-Time Propagation of the Four-Component Dirac–Kohn–Sham Equation. *J. Chem. Theory Comput.* **2015**, *11*, 980–991.

(9) Nascimento, D. R.; DePrince, A. E. A general time-domain formulation of equation-of-motion coupled-cluster theory for linear spectroscopy. *J. Chem. Phys.* **2019**, *151*, 204107.

(10) Skeidsvoll, A. S.; Balbi, A.; Koch, H. Time-dependent coupled-cluster theory for ultrafast transient-absorption spectroscopy. *Phys. Rev. A* **2020**, *102*, 023115.

(11) Guandalini, A.; Cocchi, C.; Pittalis, S.; Ruini, A.; Rozzi, C. A. Nonlinear light absorption in many-electron systems excited by an instantaneous electric field: a non-perturbative approach. *Phys. Chem. Chem. Phys.* **2021**, *23*, 10059–10069.

(12) Moitra, T.; Konecny, L.; Kadek, M.; Rubio, A.; Repisky, M. Accurate Relativistic Real-Time Time-Dependent Density Functional Theory for Valence and Core Attosecond Transient Absorption Spectroscopy. *J. Phys. Chem. Lett.* **2023**, *14*, 1714–1724.

(13) The requirement of multiple rebuilds of the Hamiltonian matrix per time step depends on the choice of algorithm and time step. Some authors report that self-consistency is not strictly necessary for RT-TDDFT calculations with very small time steps.[18]

(14) Han, R.; Mattiat, J.; Luber, S. Automatic purpose-driven basis set truncation for time-dependent Hartree–Fock and density-functional theory. *Nat. Commun.* **2023**, *14*.

(15) Black, E. K. G.; Gilbert, A. T. B.; McKenzie, S. C.; Gill, P. M. W. Avoiding Negligible Shell Pairs and Quartets in Electronic Structure Calculations. *J. Phys. Chem. A* **2023**, *127*, 842–850.

(16) Konecny, L.; Kadek, M.; Komorovsky, S.; Ruud, K.; Repisky, M. Resolution-of-identity

accelerated relativistic two- and four-component electron dynamics approach to chiroptical spectroscopies. *J. Chem. Phys.* **2018**, *149*, 204104.

(17) Yokojima, S.; Chen, G. Time domain localized-density-matrix method. *Chem. Phys. Lett.* **1998**, *292*, 379–383.

(18) O'Rourke, C.; Bowler, D. R. Linear scaling density matrix real time TDDFT: Propagator unitarity and matrix truncation. *J. Chem. Phys.* **2015**, *143*, 102801.

(19) Mandelshtam, V. A.; Taylor, H. S. Harmonic inversion of time signals and its applications. *J. Chem. Phys.* **1997**, *107*, 6756–6769.

(20) Roy, R.; Sumpter, B.; Pfeffer, G.; Gray, S.; Noid, D. Novel methods for spectral analysis. *Phys. Rep.* **1991**, *205*, 109–152.

(21) Wall, M. R.; Neuhauser, D. Extraction, through filter-diagonalization, of general quantum eigenvalues or classical normal mode frequencies from a small number of residues or a short-time segment of a signal. I. Theory and application to a quantum-dynamics model. *J. Chem. Phys.* **1995**, *102*, 8011–8022.

(22) Bruner, A.; LaMaster, D.; Lopata, K. Accelerated Broadband Spectra Using Transition Dipole Decomposition and Padé Approximants. *J. Chem. Theory Comput.* **2016**, *12*, 3741–3750.

(23) Dral, P. O. Quantum Chemistry in the Age of Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11*, 2336–2347.

(24) Häse, F.; Fdez. Galván, I.; Aspuru-Guzik, A.; Lindh, R.; Vacher, M. How machine learning can assist the interpretation of ab initio molecular dynamics simulations and conceptual understanding of chemistry. *Chem. Sci.* **2019**, *10*, 2298–2307.

(25) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555.

(26) Kitchin, J. R. Machine learning in catalysis. *Nat. Catal.* **2018**, *1*, 230–232.

(27) Chen, W.-K.; Liu, X.-Y.; Fang, W.-H.; Dral, P. O.; Cui, G. Deep learning for nonadiabatic excited-state dynamics. *J. Phys. Chem. Lett.* **2018**, *9*, 6702–6708.

(28) Secor, M.; Soudackov, A. V.; Hammes-Schiffer, S. Artificial neural networks as propagators in quantum dynamics. *J. Phys. Chem. Lett.* **2021**, *12*, 10654–10662.

(29) Ullah, A.; Dral, P. O. One-shot trajectory learning of open quantum systems dynamics. *J. Phys. Chem. Lett.* **2022**, *13*, 6037–6041.

(30) Xu, K.; Zhang, M.; Li, J.; Du, S. S.; ichi Kawarabayashi, K.; Jegelka, S. How Neural Networks Extrapolate: From Feedforward to Graph Neural Networks. `https://arxiv.org/abs/2009.11848`, access date: 2023-02-04.

(31) Oleinik, A. What are neural networks not good at? On artificial creativity. *Big Data Soc.* **2019**, *6*, 2053951719839433.

(32) Pinkus, A. Approximation theory of the MLP model in neural networks. *Acta Numer.* **1999**, *8*, 143–195.

(33) Runge, E.; Gross, E. K. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.

(34) van Leeuwen, R. Mapping from densities to potentials in time-dependent density-functional theory. *Phys. Rev. Lett.* **1999**, *82*, 3863–3866.

(35) Theilhaber, J. Ab initio simulations of sodium using time-dependent density-functional theory. *Phys. Rev. B* **1992**, *46*, 12990–13003.

(36) Yabana, K.; Bertsch, G. F. Time-dependent local-density approximation in real time. *Phys. Rev. B* **1996**, *54*, 4484–4487.

(37) Lopata, K.; Govind, N. Modeling Fast Electron Dynamics with Real-Time Time-Dependent Density Functional Theory: Application to Small Molecules and Chromophores. *J. Chem. Theory Comput.* **2011**, *7*, 1344–1355.

(38) Wibowo, M.; Irons, T. J.; Teale, A. M. Modeling Ultrafast Electron Dynamics in Strong Magnetic Fields Using Real-Time Time-Dependent Electronic Structure Methods. *J. Chem. Theory Comput.* **2021**, *17*, 2137–2165.

(39) Provorse, M. R.; Isborn, C. M. Electron dynamics with real-time time-dependent density functional theory. *Int. J. Quantum Chem.* **2016**, *116*, 739–749.

(40) Li, X.; Smith, S. M.; Markevitch, A. N.; Romanov, D. A.; Levis, R. J.; Schlegel, H. B. A time-dependent Hartree–Fock approach for studying the electronic optical response of molecules in intense fields. *Phys. Chem. Chem. Phys.* **2005**, *7*, 233–239.

(41) Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. Toward a systematic molecular orbital theory for excited states. *J. Phys. Chem.* **1992**, *96*, 135–149.

(42) Rohringer, N.; Gordon, A.; Santra, R. Configuration-interaction-based time-dependent orbital approach for ab initio treatment of electronic dynamics in a strong optical laser field. *Phys. Rev. A* **2006**, *74*, 043420.

(43) Greenman, L.; Ho, P. J.; Pabst, S.; Kamarchik, E.; Mazziotti, D. A.; Santra, R. Implementation of the time-dependent configuration-interaction singles method for atomic strong-field processes. *Phys. Rev. A* **2010**, *82*, 023406.

(44) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; John Wiley & Sons: Chichester, 2000.

(45) Griffiths, D. J. *Introduction to quantum mechanics*, 2nd ed.; Cambridge University Press: Cambridge, 2017.

(46) Tannor, D. J. *Introduction to quantum mechanics*; University Science Books: Sausalito, California, 2007; pp 411–412.

(47) Norman, P.; Bishop, D. M.; Jensen, H. J. A.; Oddershede, J. Near-resonant absorption in the time-dependent self-consistent field and multiconfigurational self-consistent field approximations. *J. Chem. Phys.* **2001**, *115*, 10323–10334.

(48) Hauge, E. S. Extrapolating the Electric Dipole Moment. M.Sc. thesis, University of Oslo, 2021; access date: 2023-02-04.

(49) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array programming with NumPy. *Nature* **2020**, *585*, 357–362.

(50) Horn, R. A.; Johnson, C. R. *Matrix Analysis*; Cambridge University Press, 1985; p 257–342.

(51) Mattiat, J.; Luber, S. Efficient calculation of (resonance) Raman spectra and excitation profiles with real-time propagation. *J. Chem. Phys.* **2018**, *149*, 174108.

(52) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer New York: New York, NY, 2009; pp 509–510.

(53) Kaufman, L.; Rousseeuw, P. J. *Finding Groups in Data. An Introduction to Cluster Analysis*; John Wiley & Sons: Hoboken, 2005.

(54) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Courna-

peau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(55) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer New York: New York, NY, 2009; pp 43–99.

(56) Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Statist. Soc. B* **1996**, *58*, 267–288.

(57) Ghosh, S.; Asher, J. C.; Gagliardi, L.; Cramer, C. J.; Govind, N. A semiempirical effective Hamiltonian based approach for analyzing excited state wave functions and computing excited state absorption spectra using real-time dynamics. *J. Chem. Phys.* **2019**, *150*, 104103.

(58) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272.

(59) Repisky, M.; Komorovsky, S.; Kadek, M.; Konecny, L.; Ekström, U.; Malkin, E.; Kaupp, M.; Ruud, K.; Malkina, O. L.; Malkin, V. G. ReSpect: Relativistic spectroscopy DFT program package. *J. Chem. Phys.* **2020**, *152*, 184101.

(60) Aurbakken, E. and Kristiansen H. E., and Kvaal, S. and Ofstad, B. S. and Pedersen, T. B. and Schøyen, Ø. S., HyQD. `https://github.com/HyQD`, access date: 2023-02-04.

(61) Hairer, E.; Wanner, G.; Lubich, C. *Geometric Numerical Integration*, 2nd ed.; Springer: Berlin, Heidelberg, 2006.

(62) Pedersen, T. B.; Kvaal, S. Symplectic integration and physical interpretation of time-dependent coupled-cluster theory. *J. Chem. Phys.* **2019**, *150*, 144106.

(63) Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(64) Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(65) Woon, D. E.; Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(66) Slater, J. C. A simplification of the Hartree-Fock method. *Phys. Rev.* **1951**, *81*, 385.

(67) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(68) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.

(69) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(70) Cooper, B. C.; Koulias, L. N.; Nascimento, D. R.; Li, X.; DePrince, A. E. I. Short Iterative Lanczos Integration in Time-Dependent Equation-of-Motion Coupled-Cluster Theory. *The Journal of Physical Chemistry A* **2021**, *125*, 5438–5447.

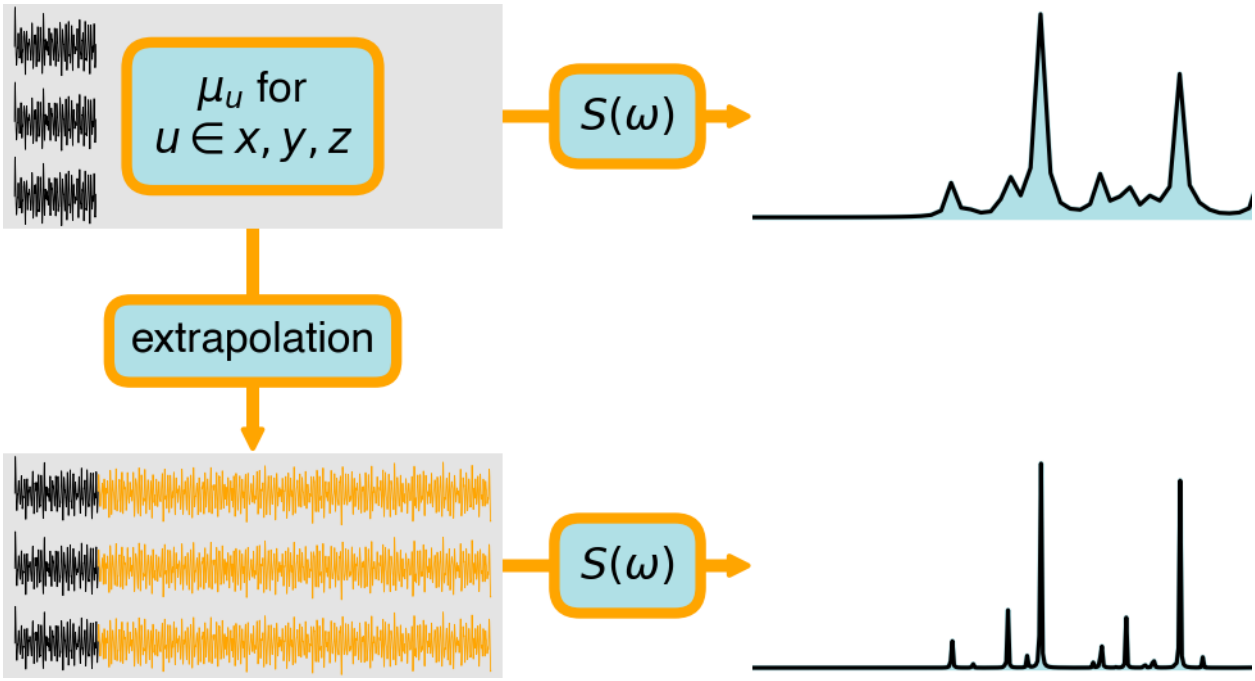

Figure 7: For Table of Contents Only.