



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2023

Automatic purpose-driven basis set truncation for time-dependent Hartree–Fock and density-functional theory

Han, Ruocheng ; Mattiat, Johann ; Lubner, Sandra

DOI: <https://doi.org/10.1038/s41467-022-35694-4>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-239538>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Han, Ruocheng; Mattiat, Johann; Lubner, Sandra (2023). Automatic purpose-driven basis set truncation for time-dependent Hartree–Fock and density-functional theory. *Nature Communications*, 14(1):106.


DOI: <https://doi.org/10.1038/s41467-022-35694-4>

Automatic purpose-driven basis set truncation for time-dependent Hartree–Fock and density-functional theory

Received: 4 July 2022

Accepted: 14 December 2022

Published online: 06 January 2023

 Check for updatesRuocheng Han¹, Johann Mattiat¹ & Sandra Luber¹✉

Real-time time-dependent density-functional theory (RT-TDDFT) and linear response time-dependent density-functional theory (LR-TDDFT) are two important approaches to simulate electronic spectra. However, the basis sets used in such calculations are usually the ones designed mainly for electronic ground state calculations. In this work, we propose a systematic and robust scheme to truncate the atomic orbital (AO) basis set employed in TDDFT and TD Hartree–Fock (TDHF) calculations. The truncated bases are tested for both LR- and RT-TDDFT as well as RT-TDHF approaches, and provide an acceleration up to an order of magnitude while the shifts of excitation energies of interest are generally within 0.2 eV. The procedure only requires one extra RT calculation with 1% of the total propagation time and a simple modification on basis set file, which allows an instant application in any quantum chemistry package supporting RT-/LR-TDDFT calculations. Aside from the reduced computational effort, this approach also offers valuable insight into the effect of different basis functions on computed electronic excitations and further ideas on the design of basis sets for special purposes.

Electronically excited states and their properties are among the central topics of quantum chemistry research. The utilized theoretical methods for excited state calculations typically require equivalent or higher computational resources compared to analogous ground state calculations. Highly accurate multiconfigurational methods are computationally demanding and thus can only be applied to small systems.

Time-dependent density-functional theory (TDDFT), due to its good compromise between accuracy and efficiency, has been employed in a wide range of applications, especially for spectroscopy^{1–3}. Real-time propagation (RTP) has become an appealing technique for the solution of the time-dependent Kohn–Sham calculations, namely, real-time time-dependent density-functional theory (RT-TDDFT), or general approximations to the time-dependent Schrödinger equation^{4–9}. It is based on the evolution of molecular orbitals under the influence of an external field, often with only a δ -pulse (field) applied at the beginning. In the weak field limit within the adiabatic approximation, the spectroscopy simulations using RT-TDDFT and linear response time-dependent density-

functional theory (LR-TDDFT) should provide comparable results⁴. During each time step, one needs to construct the Hamiltonian given by the new molecular orbital (MO) coefficients, which is the most time-consuming part of the RTP. Depending on the quantum chemistry method employed in RTP, the construction of the Hamiltonian may scale to $\mathcal{O}(N^4)$, e.g., for HF Coulomb and exchange matrices calculated with 2-electron integrals, where N refers to the number of AO basis functions. Therefore, reducing the number of basis functions or finding a proper smaller basis set can potentially save a large amount of computational time and memory.

Previous studies on the topic of basis set truncation/reduction have mainly followed three strategies: (1) decreasing the size of the virtual space for frozen natural orbital approximations used in perturbation based methods^{10,11} (e.g., Møller–Plesset perturbation theory, coupled cluster single-double and perturbative triple, complete active space perturbation theory), (2) reducing the number of functions in correlation consistent basis sets^{12,13}, (3) reducing the number of basis functions of subsystems (which apply expensive wavefunction-based

¹Department of Chemistry, University of Zurich, Zurich, Switzerland. ✉e-mail: sandra.luber@chem.uzh.ch

methods) for embedding calculations^{14,15}. However, these works focus on the electronic ground state. Multiple embedding techniques have been applied to accelerate RT-TDDFT calculations by treating subsystems with different level of theories^{16–20}. The idea of a decomposition of the electric dipole moment into molecular orbital pairs was also proposed in recent works for the acceleration or analyses of spectra^{21–23}. This work explores the contribution of a fundamental ingredient—basis functions—to the electronic spectra. The truncation of basis functions proposed in this work is designed to check every single component in the basis set (basis function). One can also apply a shell level truncation for general applications. Basis set files can be easily modified for an accelerated simulation of the spectrum and to obtain a better chemical insight into the electric dipole moments contribution. Moreover, a routine to construct complete basis set (CBS) for TDDFT calculations is proposed. The calculations of electronic absorption and ECD spectra in this work take place in a linear response framework within the electric dipole approximation, assuming that the excited states of the system can be well described within the occupied-virtual space spanned by the ground state solution of the system. As standardly done, we assume the adiabatic approximation, discarding the dependence of the exchange-correlation functional on the history of the propagation. The decomposition of electric dipole moments into the contribution of individual AO basis functions and checking the variation of molecular orbitals (in component of basis functions) during the RTP provides a quantitative evaluation of each AO basis function based on its importance for the electronic spectra under study. This further paves the way for a truncation process on the basis set for the computational speed-up and a way to generate complete basis set for TDDFT calculations.

In this work, we propose a basis set truncation scheme for TDDFT calculations. The method is tested for small molecules up to a highly conjugated system and a metal cluster, and achieves an acceleration up to an order of magnitude in RT-TDDFT or LR-TDDFT calculations with negligible change in the region of interests (e.g., valence-shell transitions) of the computed spectra.

Results

Electric dipole moment

In the context of this work, the electronic part of the electric dipole moment \vec{d} is defined as the trace of the product of the density matrix and integrals of the electric dipole moment operator $-e\vec{r}$ ($\vec{r} = (x, y, z)$, e is elementary charge) in the AO basis with basis functions $\{\chi_\mu\}$. For calculating the time-dependent electric dipole moment $\vec{d}(t)$, we use the AO basis representation for both density matrix $\mathbf{P}^{\text{AO}}(t)$ and the electric dipole moment integrals $\vec{\mathbf{D}}$ as shown in Eq. (1). In this way, only the density matrix $\mathbf{P}^{\text{AO}}(t)$ is time-dependent and $\vec{\mathbf{D}}$ remains the same during the RTP for fixed nuclei. $\mathbf{P}^{\text{AO}}(t)$ can be further expressed in molecular orbital (MO) basis as \mathbf{P}^{MO} (MO density matrix after SCF, see Eq. (3) where f_i is the occupation number of the i th MO) and the time-dependent part is only carried by the MO coefficients $\mathbf{C}(t)$ and its complex conjugate $\mathbf{C}^\dagger(t)$ (see Eq. (2)). In this work, the AO basis functions are all Gaussian-type orbitals.

$$\vec{d}(t) = -e \cdot \sum_{\mu\nu} \mathbf{P}_{\mu\nu}^{\text{AO}} \langle \chi_\mu | \vec{r} | \chi_\nu \rangle = -e \cdot \text{Tr}(\mathbf{P}^{\text{AO}}(t) \vec{\mathbf{D}}) \quad (1)$$

$$\mathbf{P}^{\text{AO}}(t) = \mathbf{C}(t) \mathbf{P}^{\text{MO}} \mathbf{C}^\dagger(t) \quad (2)$$

$$\mathbf{P}^{\text{MO}} = (p_{ij}) = \begin{cases} f_i, & i=j \\ 0, & i \neq j \end{cases} \quad (3)$$

Real-time propagation

In our implementation, the MO coefficients $\mathbf{C}(t)$ are propagated for a small timestep Δt (see Eq. (4)) using the “enforced time-reversal symmetry” (ETRS)²⁴ scheme. $\mathbf{U}(t + \Delta t)$ represents the propagator at time $t + \Delta t$ and is calculated with Eq. (5), where \mathbf{S} is the overlap matrix in AO basis and $\mathbf{F}(t)$ is the Fock matrix or Kohn–Sham (KS) matrix in AO basis at time t . $\mathbf{C}(t + \Delta t)$, $\mathbf{U}(t + \Delta t)$, and $\mathbf{F}(t + \Delta t)$ are computed self-consistently²⁴.

$$\mathbf{C}(t + \Delta t) = \mathbf{U}(t + \Delta t) \mathbf{C}(t) \quad (4)$$

$$\mathbf{U}(t + \Delta t) = \exp \left[-\frac{i}{2} \mathbf{S}^{-1} (\mathbf{F}(t) + \mathbf{F}(t + \Delta t)) \Delta t \right] \quad (5)$$

$\mathbf{F}(t)$ needs to be constructed for each time step and usually contributes most to the computational time in RTP. For example, the elements of the HF exchange matrix $K_{\mu\nu}(t)$ are given in Eq. (6) ($\langle \mu\lambda | \sigma\nu \rangle$ are the two-electron repulsion integral (ERIs) in AO basis expressed in Eq. (7)), and the computation of the exchange matrix $\mathbf{K}(t)$, which is required for the construction of $\mathbf{F}(t)$, scales as $\mathcal{O}(N_{\text{AO}}^4)$ (N_{AO} is the number of AO basis functions).

$$K_{\mu\nu}(t) = \sum_{\lambda\sigma} \mathbf{P}_{\lambda\sigma}^{\text{AO}}(t) \langle \mu\sigma | \lambda\nu \rangle \quad (6)$$

$$\langle \mu\sigma | \lambda\nu \rangle = \int \chi_\mu^*(\vec{r}_1) \chi_\sigma^*(\vec{r}_2) \frac{1}{r_{12}} \chi_\lambda(\vec{r}_1) \chi_\nu(\vec{r}_2) d\vec{r}_1 d\vec{r}_2 \quad (7)$$

AO basis truncation

In order to decrease N_{AO} , we first analyse Eq. (1) for the electric dipole contribution from each AO basis function. For the sake of simplicity, $\vec{\mathbf{D}}_{\mu\mu}(t)$ is used to represent the μ th diagonal element of $\mathbf{P}^{\text{AO}}(t) \vec{\mathbf{D}}$, and thus $\vec{d}(t)$ can then be rewritten as in Eq. (8). Taking a detailed look at the construction of $\vec{\mathbf{D}}_{\mu\mu}(t)$ in Eq. (9), one can find that it provides a decomposed form of electric dipole moments of each basis function. Therefore, we use $\vec{\mathbf{D}}_{\mu\mu}(t)$ to represent the electric dipole contribution from the μ th basis function.

$$\vec{d}(t) = -e \cdot \sum_{\mu}^{N_{\text{AO}}} \vec{\mathbf{D}}_{\mu\mu}(t), \text{ where } \vec{\mathbf{D}}_{\mu\mu}(t) = (\mathbf{P}^{\text{AO}}(t) \vec{\mathbf{D}})_{\mu\mu} \quad (8)$$

$$\vec{\mathbf{D}}_{\mu\mu}(t) = \sum_{\nu} \mathbf{P}_{\mu\nu}^{\text{AO}}(t) \langle \chi_\mu | \vec{r} - \vec{R} | \chi_\nu \rangle \quad (9)$$

However, $\vec{\mathbf{D}}_{\mu\mu}(t)$ is not translational invariant because the value of $\vec{\mathbf{D}}_{\mu\nu}$ (element in $\vec{\mathbf{D}}$) depends on the choice of reference points \vec{R} (see Eq. (10)). Note that \vec{r} and \vec{R} are referenced to the origin of coordinate system. Though \vec{R} does not affect the full spectrum after Fourier transform (because \vec{d} is translational invariant for neutral systems as $\vec{R} S_{\mu\nu}$ cancels with the nuclear electric dipole contribution), it can change the relative contribution of electric dipole moments from each AO basis function ($\vec{\mathbf{D}}_{\mu\mu}(t)$). We can further split $\vec{\mathbf{D}}_{\mu\nu}$ into a reference point (\vec{R})-independent term $\langle \chi_\mu | \vec{r} | \chi_\nu \rangle$ and a reference point-dependent term $\vec{R} S_{\mu\nu}$, where $S_{\mu\nu}$ is the element of the overlap matrix in AO basis.

$$\vec{\mathbf{D}}_{\mu\nu} = \langle \chi_\mu | \vec{r} - \vec{R} | \chi_\nu \rangle = \langle \chi_\mu | \vec{r} | \chi_\nu \rangle - \vec{R} S_{\mu\nu}, \text{ where } S_{\mu\nu} = \langle \chi_\mu | \chi_\nu \rangle \quad (10)$$

In $\langle \chi_\mu | \vec{r} | \chi_\nu \rangle$, the relative position of atoms can cause different values of elements in the matrix, which we would like to avoid. To explain the reason, we can think about a toy system consisting of only

two hydrogen atoms with Cartesian coordinates H1 $\vec{r}_1 = (0, 0, -a)$ and H2 $\vec{r}_2 = (0, 0, a)$ where $a \neq 0$. It is obvious that e.g., diagonal matrix elements $\langle \chi_{H1s} | \vec{r} | \chi_{H1s} \rangle \neq \langle \chi_{H2s} | \vec{r} | \chi_{H2s} \rangle$ (see Eq. (11)) even though, by symmetry, we expect the same “contribution” of electric dipole from the two atoms. In Eq. (11), each H atom has a Slater-type 1s orbital Ae^{-r} where A is normalization constant and r is the distance from the center of the atom, and we change the integration variable from \vec{r} to \vec{s} using $\vec{s} = \vec{r} - \vec{r}_1$ and $\vec{s} = \vec{r} - \vec{r}_2$.

$$\begin{aligned} \langle \chi_{H1s} | \vec{r} | \chi_{H2s} \rangle - \langle \chi_{H1s} | \vec{r} | \chi_{H1s} \rangle &= \int A^2 \vec{r} e^{-2r} \vec{r} - \vec{r}_2 d\vec{r} - \int A^2 \vec{r} e^{-2r} \vec{r} - \vec{r}_1 d\vec{r} \\ &= \int A^2 (\vec{s} + \vec{r}_2) e^{-2|\vec{s}|} d\vec{s} - \int A^2 (\vec{s} + \vec{r}_1) e^{-2|\vec{s}|} d\vec{s} \quad (11) \\ &= (\vec{r}_2 - \vec{r}_1) \int A^2 e^{-2|\vec{s}|} d\vec{s} \\ &= \vec{r}_2 - \vec{r}_1 \end{aligned}$$

One way of minimizing the effect of $\langle \chi_{\mu} | \vec{r} | \chi_{\nu} \rangle$ is to shift the molecular system far from (0, 0, 0), which is equivalent to set a large \vec{R} . It is worth noting that we do not need to formally “move” the molecule, and this is just an assumption made in the derivation from Eq. (9) to Eq. (12). In this method, we only care about the relative value of $\vec{O}_{\mu}(t)$ when determining the basis function(s) to be truncated, and \vec{R} provides the same factor to $\vec{D}_{\mu\nu}$ and later $\vec{O}_{\mu}(t)$. Therefore, it is safe to substitute $\vec{D}_{\mu\nu}$ with $S_{\mu\nu}$ in the expression of $\vec{O}_{\mu}(t)$, and thus we have a scalar $O_{\mu}(t)$ as shown in Eq. (12). It is worth noting that $O_{\mu}(t)$ does not explicitly contain any electric dipole information, which makes sense because the electric dipole cannot be formally defined on a single atomic centered orbital. However, AOs do contribute to the electric dipole by forming MOs across different atomic centers, and the information of such contribution, which we call density matrix contribution, is contained in $\mathbf{P}^{\text{AO}}(t)$. The expression in Eq. (12) is also known in Mulliken population analysis, as the number of electrons associated with χ_{μ} . The following computations are all based on the scalar form $O_{\mu}(t)$.

$$O_{\mu}(t) = (\mathbf{P}^{\text{AO}}(t)\mathbf{S})_{\mu\mu} \quad (12)$$

To quantify the contribution, we use the formula in Eq. (13). x_{μ}^{DC} is an indicator measuring the variation of the Density matrix Contribution (DC) of the μ th AO basis function. $S_t[O_{\mu}(t)]$ computes the standard deviation of $O_{\mu}(t)$ for the total simulation time for each μ . The numerator of Eq. (13) indicates variation (along the RTP) of the electric dipole moment contribution from each AO basis function. The dimensionless quantity x_{μ}^{DC} is then constructed by dividing the numerator with its mean value of all AO basis functions. A small x_{μ}^{DC} value means the change of electric dipole moments contributed from the μ th basis function is comparatively small among all AO basis functions, and removing this basis function should not change the spectrum (a constant value vanishes after Fourier transform for RT-TDHF/TDDFT) significantly. One might point out that x_{μ}^{DC} cannot distinguish the pulse from different directions because $O_{\mu}(t)$ in Eq. (12) is no longer direction dependent like in Eq. (8). However, it is found that $\mathbf{P}^{\text{AO}}(t)$ still varies according to the direction of the pulse because its action is coded in the MO coefficients.

$$x_{\mu}^{\text{DC}} = \frac{S_t[O_{\mu}(t)]}{\frac{1}{N_{\text{AO}}} \sum_{\mu}^N S_t[O_{\mu}(t)]} \quad (13)$$

It is worth noting that we have considered applying basis transformation regarding to $\mathbf{P}^{\text{AO}}(t)\vec{D}$, namely using the eigenvectors of $\mathbf{P}^{\text{AO}}(t)$ (transform to natural orbital basis) or \vec{D} . However, the former one is time-dependent so it is hard to choose a transformation matrix for all time steps, and the latter one is reference point-dependent (see Eq. (10)) and one cannot obtain the consistent truncation choice under translation (note that nuclei do not move in this study as opposed to,

e.g., Ehrenfest dynamics). Also, an AO basis is a common choice in most molecular simulations and some solid state simulations (e.g., Gaussian and Plane Waves²⁵ method in CP2K package (CP2K version 7.0 (Development Version), the CP2K developers group. CP2K is freely available from <https://www.cp2k.org/>)). Therefore, the truncation on the AO basis has broad application prospects and can be easily applied by a simple modification of basis set file.

During practical tests of basis truncation, we observed that using only x_{μ}^{DC} as indicator is not enough for obtaining an accurate spectrum. Another indicator x_{μ}^{IP} is introduced which measures the Importance of Propagation stability (IP) of the μ th AO basis function (see Eq. (14)). $C_{\mu j}(t)$ denotes an element in transformation matrix (from AO to MO basis), S_t computes the standard deviation along the time for each μ and j , and $\sum_j^{N_{\text{MO}}}$ sums over all standard deviations in MOs originating from the μ th AO basis function. The numerator of Eq. (14) indicates the variation (along RTP) of the contribution from each AO basis function to all MOs in transformation matrix $\mathbf{C}(t)$. As for x_{μ}^{DC} , the dimensionless quantity x_{μ}^{IP} is also constructed by dividing the numerator with its mean value of all AO basis functions. Small x_{μ}^{IP} value means that the contributions to MOs from the μ th basis function do not change much compared to the contribution of all AO basis functions, and removing this basis function should not affect the propagation of the density matrix (remaining part) significantly. Note that both $O_{\mu}(t)$ and $C_{\mu j}(t)$ are usually complex numbers for RTP, and the standard deviation of a set of complex numbers is calculated as in Supplementary Eq. (1).

$$x_{\mu}^{\text{IP}} = \frac{\sum_j^{N_{\text{MO}}} S_t[C_{\mu j}(t)]}{\frac{1}{N_{\text{AO}}} \sum_{\mu}^{N_{\text{AO}}} \sum_j^{N_{\text{MO}}} S_t[C_{\mu j}(t)]} \quad (14)$$

In practice, an empirical parameter x^{thr} is chosen as threshold for both x_{μ}^{DC} and x_{μ}^{IP} , where the AO basis functions with both indicators smaller than x^{thr} can be removed. The remaining basis set $\{\chi_{\mu}\}_{\text{trunc}}$ (truncated AO basis set) is then defined as in Eq. (15), given the original AO basis set $\{\chi_{\mu}\}$ (of which the cardinality $|\{\chi_{\mu}\}|$ is N_{AO}). Sometimes $\{\chi_{\mu}\}_{\text{trunc}}$ includes only part of the given shell, e.g., for a p -shell, only χ^{p_x} and χ^{p_y} are in $\{\chi_{\mu}\}_{\text{trunc}}$ and χ^{p_z} is truncated. Such symmetry breaking is mainly due to the utilization of polarized field (δ -pulse) in the RT-TDDFT calculations, and the rotational invariance requires a shell level truncation. Considering the truncated basis set used in any computational chemistry package, we also recommend shell level truncation for the general application of the basis set file. In most cases, the majority rule can be applied for a truncation at the shell level, namely, the shells containing more than half of their original basis functions remain in $\{\chi_{\mu}\}_{\text{trunc}}$ while others are fully discarded. This has to be checked with the x_{μ}^{DC} indicators of the basis functions in the same shell to ensure that there are no strong contribution to electric dipole transitions arising from some basis functions. In this study, the basis sets of (S)-methyloxirane, (-)- α -pinene, ZnPC, and Ag₂₀ systems are truncated at the shell level.

The schematic view of the truncation process is shown in Fig. 1.

$$\{\chi_{\mu}\}_{\text{trunc}} = \left\{ \chi_{\mu} \mid x_{\mu}^{\text{DC}} > x^{\text{thr}} \vee x_{\mu}^{\text{IP}} > x^{\text{thr}}, \forall i \in \{1, \dots, N_{\text{AO}}\} \right\} \quad (15)$$

Using $\{\chi_{\mu}\}_{\text{trunc}}$, namely, reducing number of basis functions from N_{AO} to $N_{\text{trunc}} = |\{\chi_{\mu}\}_{\text{trunc}}|$, can ideally decrease the total computational time to $(N_{\text{trunc}}/N_{\text{AO}})^4$ for a RT-TDHF calculation or a RT-TDDFT calculation with hybrid exchange-correlation functional. Also, this truncated basis set can be transferred to LR-TDHF/TDDFT calculations.

The procedure for carrying out RT-TDHF/TDDFT calculations with truncated AO basis set for the examples studied in this work is described as follows:

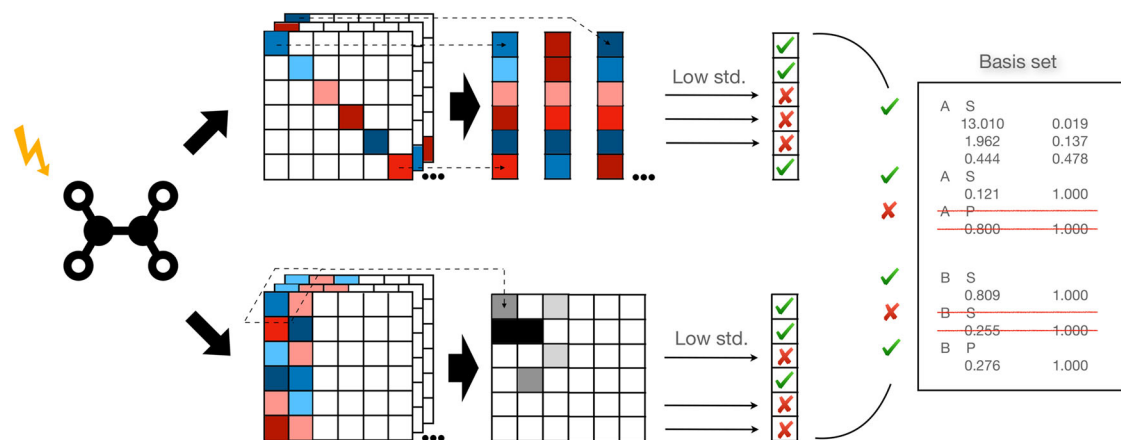


Fig. 1 | Schematic diagram of the basis set truncation process. First, a real-time propagation run of 1% (e.g., 100 steps) of the total simulation time is performed. Then the information of AO density matrix $P^{AO}(t)$ and MO coefficient $C(t)$ at every step, and overlap matrix S , is collected. Basis functions to be truncated are selected

based on the low standard deviation (std. in the figure) of $O_{\mu}(t)$ and $C_{\mu}(t)$. Eventually, one can directly modify the basis set file for a complete RT-TDHF/TDDFT calculation or a LR-TDHF/TDDFT calculation.

- Run 100 (400) steps (1% of the total simulation time) of RT-TDHF/RT-TDDFT simulation with the timestep 0.2 (0.05) atomic units using a preliminarily chosen basis set $\{\chi_{\mu}\}$, and collect the information regarding S , $P^{AO}(t)$, $C(t)$ of every step.
- Calculate x_{μ}^{DC} and x_{μ}^{IP} via Eqs. (13) and (14), respectively, and select truncated AO basis set $\{\chi_{\mu}\}_{trunc}$ based on the criteria in Eq. (15).
- Run 10'000 (40'000) steps of RT-TDHF/TDDFT (full) simulation with the same timestep using $\{\chi_{\mu}\}_{trunc}$. Note that a ground state SCF calculation should be carried out with the truncated basis set before the RT-TDHF/RT-TDDFT simulation in order to apply the perturbation to a converged ground state.

It is worth noting that the truncation procedure by construction eliminates the transitions that are not (or very weakly) electric dipole allowed, and thus this approach focuses more on the overall spectrum rather than the types of transitions.

Complete basis set limit

In addition to the analyses of truncated AO basis functions, we introduce an algorithm to construct basis sets towards the CBS limit for RT(LR)-TDHF/TDDFT calculations.

The idea of CBS limit employed here is to add diffuse functions (see examples (S)-methyloxirane and (-)- α -pinene for the reason) to all types of AO basis functions (s, p, d, f, g, ...) representing different orbital angular momenta l . These functions are added in an even-tempered manner^{26,27} by a geometric progression of the orbital exponents in the original basis set: $\alpha_{l,k} = \alpha_l \beta_l^k$, $\forall k \in \mathbb{N}$. $\alpha_{l,k}$ is an exponent of the l -shell with k th power, and α_l and β_l^k are two parameters to be optimized for the basis set. Since most basis sets available (Pople²⁸, Dunning²⁹, Jensen³⁰, Ahlrichs^{31,32}, etc.) provide more than one exponent for each type of shell, we can directly extrapolate from these values to get the additional exponent $\alpha_{l,k+1} = \alpha_{l,k}^2 / \alpha_{l,k-1}$. One can increase the k value until significant linear dependencies are found in the basis set (sometimes also referred to as basis set overcompleteness³³).

This CBS scheme usually requires a quite large basis set for the calculation, and it is usually unclear which basis function(s) should be removed once overcompleteness is reached. Therefore, we combine it with AO truncation and propose an “Add-While-Truncate” algorithm (see Algo. 1) to construct the CBS specifically designed for RT(LR)-TDHF/TDDFT calculations. Firstly, a preliminarily chosen basis set $\{\chi_{\mu}\}$ is used for a short period of RT-TDHF/TDDFT calculation and $\{\chi_{\mu}\}_{trunc}$ is selected. An additional basis set containing diffuse functions ($\{\chi_{\mu}\}_{diffuse}$) is constructed in an even-tempered manner. $\{\chi_{\mu}\}_{diffuse}$ may

contain some basis functions truncated in previous steps (combined as $\{\chi_{\mu}\}_{deleted}$), which should be removed. Then $\{\chi_{\mu}\}_{diffuse}$ is combined with $\{\chi_{\mu}\}_{trunc}$ to form a new basis set $\{\chi_{\mu}\}$. In order to check the overlap-completeness of the newly created AO basis set, we calculate the overlap matrix $S_{\{\chi_{\mu}\}}$ and solve for its eigenvalues λ . If the minimal absolute eigenvalue $|\lambda|_{min}$ is smaller than a user-defined small value ϵ or $\{\chi_{\mu}\}$ remains the same as in the last cycle (namely, basis functions are neither truncated nor added), $\{\chi_{\mu}\}$ is regarded as the CBS under such ϵ -condition ($\{\chi_{\mu}\}_{CBS-\epsilon}$), otherwise the new $\{\chi_{\mu}\}$ is used to repeat the previous steps until the final condition is fulfilled. In practice, one can also manually remove some newly added diffuse functions within $\{\chi_{\mu}\}$ in the iteration to satisfy the given ϵ -condition. In this case, in order to minimize the total number of basis functions, we first remove the diffuse functions corresponding to higher orbital angular momentum, which is the same idea as the one applied in calendar basis sets³⁴. For the sake of simplicity, we use the term “basis functions” for “AO basis functions” in the remaining part of this manuscript.

Algorithm 1. Add-While-Truncate CBS Algorithm

- 1: **repeat**
- 2: $\{\chi_{\mu}\}_{old} \leftarrow \{\chi_{\mu}\}$
- 3: Run a RT-TDHF/TDDFT simulation with $\{\chi_{\mu}\}$ for 100 (400) steps.
- 4: Construct $\{\chi_{\mu}\}_{trunc}$ by Eqs. (13)–(15)
- 5: Construct additional even-tempered basis set $\{\chi_{\mu}\}_{diffuse}$ for $\{\chi_{\mu}\}_{trunc}$
- 6: $\{\chi_{\mu}\}_{deleted} \leftarrow \{\chi_{\mu}\}_{deleted} \cup (\{\chi_{\mu}\} \setminus \{\chi_{\mu}\}_{trunc})$
- 7: $\{\chi_{\mu}\}_{diffuse} \leftarrow \{\chi_{\mu}\}_{diffuse} \setminus \{\chi_{\mu}\}_{deleted}$
- 8: $\{\chi_{\mu}\} \leftarrow \{\chi_{\mu}\}_{trunc} \cup \{\chi_{\mu}\}_{diffuse}$
- 9: Solve for eigenvalues λ of overlap matrix $S_{\{\chi_{\mu}\}}$
- 10: **until** $|\lambda|_{min} < \epsilon$ **or** $\{\chi_{\mu}\} = \{\chi_{\mu}\}_{old}$
- 11: $\{\chi_{\mu}\}_{CBS-\epsilon} \leftarrow \{\chi_{\mu}\}_{trunc}$

Example: H₂ dimer

The H₂ dimer is used as the first test system, with the δ -pulse applied along z direction (see the geometry in Fig. 2d, z axis is parallel to the H–H bond, and y axis is perpendicular to the plane formed by the four atoms). Four different basis sets, 6-31G, 6-31G**, 6-31++G, and 6-31++G**^{28,35,36}, are utilized for RT-TDHF calculations.

For each H atom, 6-31G contains two s-type basis functions (noted as 2s for convenience), 6-31G** contains 2s1p (1p as extra polarization function, we use *italic* form to represent specific basis function(s), e.g., 1p means the first p-type basis function), 6-31++G contains 3s (3s as an extra diffuse function), and 6-31++G** contains 3s1p. Note that the abbreviations we use here refer to basis functions but not specific

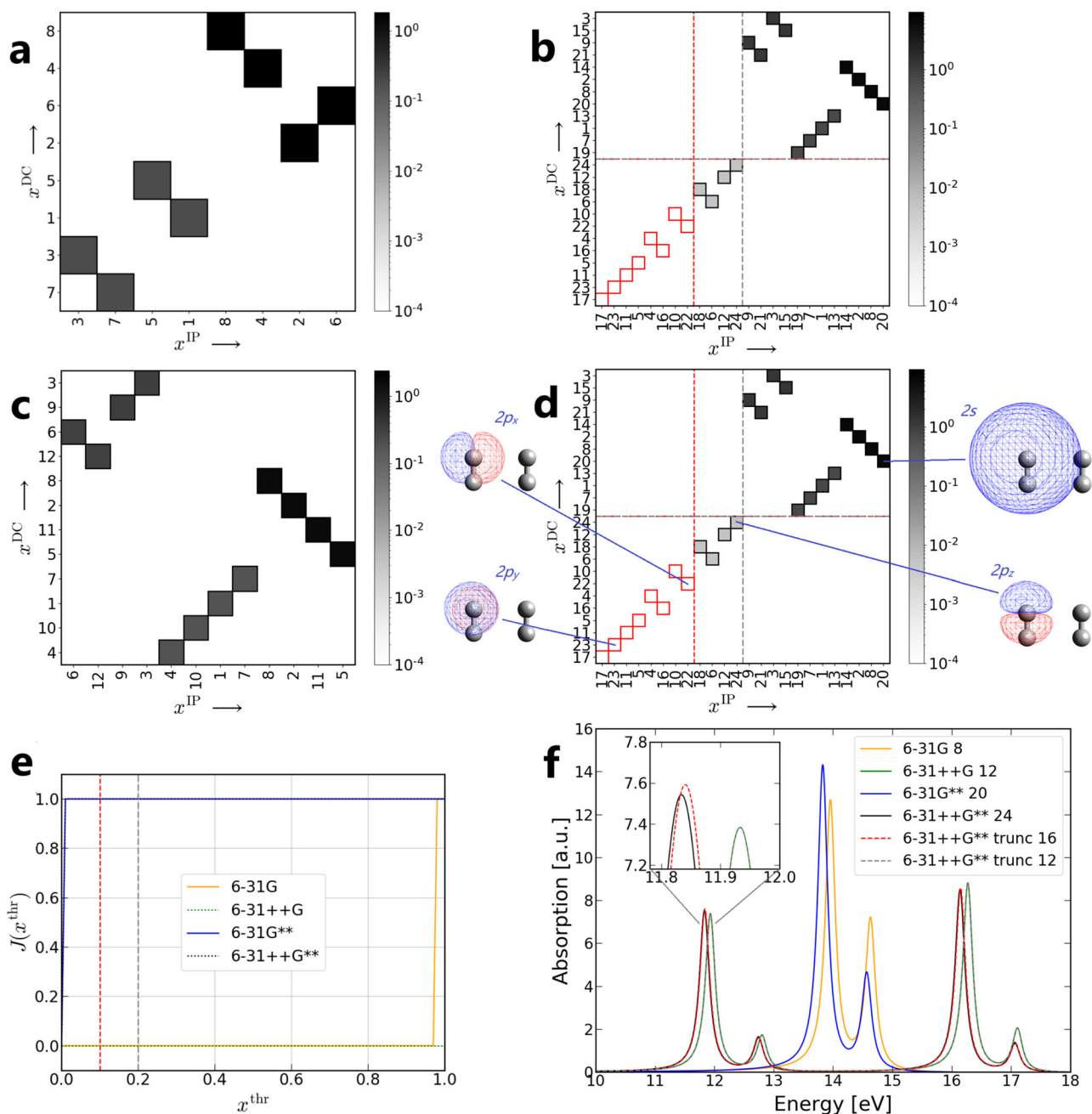


Fig. 2 | RT-TDHF calculations of the H_2 dimer. $x^{\text{DC}}\text{-}x^{\text{IP}}$ map obtained with **a** 6-31G, **b** 6-31G**, **c** 6-31++G, and **d** 6-31++G** basis set. Each square in the map represents a basis function with its numbering on x and y axes. The color of the squares is based on the value calculated as $x^{\text{DC}}\cdot x^{\text{IP}}$, so the deeper the color the more important the basis function is in the TDHF calculation. **d** includes the visualization of related basis functions. Axis labels (basis functions) are sorted in ascending order according to their x^{DC} or x^{IP} values. The red dashed line represents $x^{\text{thr}} = 0.1$ and the

gray dashed line represents $x^{\text{thr}} = 0.2$. **e** Jaccard indices for 6-31G, 6-31G**, 6-31++G, and 6-31++G** basis sets. The red dashed line represents $x^{\text{thr}} = 0.1$ and the gray dashed line represents $x^{\text{thr}} = 0.2$. **f** Electronic absorption spectra using 6-31G, 6-31G**, 6-31++G, 6-31++G**, and truncated 6-31++G** basis sets with $x^{\text{thr}} = 0.1$ (6-31++G** trunc 16) and $x^{\text{thr}} = 0.2$ (6-31++G** trunc 12). The number after the basis set label indicates the number of basis functions used for the RTP. *a.u.* arbitrary units.

electron shells. The same convention of basis/orbital notations is utilized for all examples in this study. E.g. for a truncation from 5s4p to 5s3p ($-3p$), $-3p$ means the 3rd p-type basis function is removed but the 1st, 2nd, and 4th p-type basis functions and all s-type basis functions remain.

After 100 steps of RT-TDHF calculations, x^{DC} and x^{IP} of each basis function are computed and can be visualized in Fig. 2a–d as $x^{\text{DC}}\text{-}x^{\text{IP}}$ map. Basis functions are represented by colored squares with their value $x^{\text{DC}}\cdot x^{\text{IP}}$. The electric dipole contribution and importance of propagation stability of basis functions are sorted in two axes. The red dashed line and the gray dashed line represent $x^{\text{thr}} = 0.1$ and $x^{\text{thr}} = 0.2$,

respectively. x^{thr} splits the $x^{\text{DC}}\text{-}x^{\text{IP}}$ map into four quadrants: important for both electric dipole contribution and propagation stability (top right), important for only electric dipole contribution (top left), important for only propagation stability (bottom right), and important for neither one (bottom left). The basis functions locate inside the left below region (red borders for $x^{\text{thr}} = 0.1$) are the ones recommended to be deleted from the basis set.

One can find that no basis function is to be deleted in the case of 6-31G (Fig. 2a) and 6-31++G (Fig. 2c) basis sets, and there are 8 basis functions to be deleted in the case of 6-31G** (Fig. 2b) and 6-31++G** (Fig. 2d) basis sets with $x^{\text{thr}} = 0.1$. These 8 basis functions are the same

for 6-31G** and 6-31++G**: $2p_x$ and $2p_y$ of each H atom, which belong to the polarization functions. If x^{thr} is set to 0.2 (gray dashed lines in Fig. 2b, d), extra four basis functions are to be deleted in both cases: $2p_z$ of each H atom. Therefore, a setting of $x^{\text{thr}} = 0.2$ essentially truncates 6-31G** to 6-31G and 6-31++G** to 6-31++G for the H₂ dimer. Note that the pulse causes differences between atoms with different nuclear Cartesian coordinates, leading to different $x^{\text{DC}}\text{-}x^{\text{IP}}$ values of the same basis functions in different H atoms. In Supplementary Fig. 1, we further provide a more intuitive view of O_μ from each basis function in 6-31++G and 6-31++G** basis sets. One can easily distinguish small contribution components from the large contribution ones.

Let us take a closer look at the $x^{\text{DC}}\text{-}x^{\text{IP}}$ map in the case of 6-31++G** basis set (see basis functions shown in Fig. 2d, we focus only on one H atom here). Actually, we can explain qualitatively that $2p_x$ and $2p_y$ are the least important basis functions for the simulation of the electronic absorption spectrum. The electric dipole transitions from σ_{1s-1s} to $\pi_{2p_x-2p_x}^*$, $\pi_{2p_x-2p_x}^*$, $\pi_{2p_y-2p_y}^*$, and $\pi_{2p_y-2p_y}^*$ are almost (considering the effect from the other H₂ molecule close-by) forbidden due to symmetry reasons. The $x^{\text{DC}}\text{-}x^{\text{IP}}$ map shows that $2p_x$ is slightly more important than $2p_y$, which may be explained by a stronger interaction on the x direction between H₂ molecules. The electric dipole transition $\pi_{2p_z-2p_z}^* \leftarrow \sigma_{1s-1s}$ is allowed, and thus the $2p_z$ basis function is considered to be more important than $2p_x$ and $2p_y$ for the electronic absorption spectrum. The electric dipole transition $\sigma_{2s-2s}^* \leftarrow \sigma_{1s-1s}$ is also allowed and σ_{2s-2s}^* has a lower energy than $\pi_{2p_z-2p_z}^*$, which leads to a higher occupation probability. Therefore, $2s$ in 6-31++G** basis set is one of the “dominant” basis functions in the RTP for H₂ with the computational settings used.

In more complex systems, such an energetic analysis in terms of “static” wavefunctions (e.g., wavefunctions after SCF) is not enough to give a reasonable truncated basis set since MO coefficients $C(t)$ are time-dependent, which explains our choice of using first 100 (400) steps for the analysis.

Besides, a Jaccard index³⁷ ($J(x^{\text{thr}})$) is applied to analyze the similarity of deleted basis functions suggested by our two criteria x^{DC} and x^{IP} (see Eq. (16)). A high Jaccard index indicates more basis functions in common between two sets and vice versa. This information provides an intuitive view of the truncation along x^{thr} for a given basis set.

$$J(x^{\text{thr}}) = \frac{|\{X_\mu | x_\mu^{\text{DC}} < x^{\text{thr}}\} \cap \{X_\mu | x_\mu^{\text{IP}} < x^{\text{thr}}\}|}{|\{X_\mu | x_\mu^{\text{DC}} < x^{\text{thr}}\} \cup \{X_\mu | x_\mu^{\text{IP}} < x^{\text{thr}}\}|} \quad (16)$$

For H₂ dimer system, it is found that $J(x^{\text{thr}})$ of 6-31G** and 6-31++G** remains at a value of 1.0 from $x^{\text{thr}} = 0.01$ to $x^{\text{thr}} = 1.0$ (see Fig. 2e). This shows that RTP has clear “preference” for some basis functions within the given basis sets. In the case of 6-31G and 6-31++G basis sets, on the other hand, $J(x^{\text{thr}})$ remains at a value of 0.0 up to $x^{\text{thr}} = 0.9$, which means that no redundant basis functions are found for such basis sets.

The spectra using different basis sets and truncated 6-31++G** basis sets are in Fig. 2f. The spectra using 6-31G and 6-31G** basis sets look very similar, which matches the truncation suggestion given in Fig. 2b. The same situation is also found in the spectra using 6-31++G and 6-31++G** basis sets. Truncated 6-31++G** basis set with $x^{\text{thr}} = 0.2$ (noted as 6-31++G** trunc 12, 12 basis functions left) leads to the same basis set as 6-31++G, with an error of -0.1 eV of corresponding excitation energies. With a tighter threshold $x^{\text{thr}} = 0.1$, truncated 6-31++G** basis set (noted as 6-31++G** trunc 16, 16 basis functions left) achieves more accurate spectra compared to 6-31++G basis set, with an error at the level of -0.01 eV. These results are in accordance with the $x^{\text{DC}}\text{-}x^{\text{IP}}$ map introduced before. For the sake of completeness, the cases of δ -pulse from x or y direction are included in Supplementary Fig. 2.

Example: H₂O dimer

Four different basis sets, def2-TZVP, def2-TZVPP, def2-TZVPD, and def2-TZVPPD^{32,38} are utilized for RT-TDHF calculations of H₂O dimer

system (see Fig. 3g for the nuclear structure). Again, these four basis sets are chosen based on the addition of polarization functions and/or diffuse functions. The $x^{\text{DC}}\text{-}x^{\text{IP}}$ map of this system is shown in Fig. 3a–d after 100 steps of RT-TDHF calculations. Compared to H₂ dimer, H₂O dimer is a more complicated system and thus the $x^{\text{DC}}\text{-}x^{\text{IP}}$ map is more involved. Nevertheless, it is clear that the distribution of basis functions in all plots shows a “dumbbell” shape (from left below to right top), namely, more dispersed in low and high $x^{\text{DC}}/x^{\text{IP}}$ region compared to the middle range. This provides a rough idea of the range of truncation, and the cross point of $x^{\text{thr}} = 0.1$ generally locates at the neck of the “dumbbell”. The Jaccard indices for the four basis sets are shown in Fig. 3e. In Supplementary Figure 3, some visualizations of orbitals are shown in the $x^{\text{DC}}\text{-}x^{\text{IP}}$ map of the def2-TZVP basis set.

A natural question of basis set truncation is whether one needs to do it recursively until the basis set does not change anymore (which we refer to as “recursive truncation”). Therefore, we test the recursive truncation of def2-TZVPD and def2-TZVPPD basis sets, and provide the number of basis functions together with Jaccard indices (see Fig. 3f). One-time truncation decreases the number of basis functions from 148 to 103 and from 116 to 92 for the two original basis sets, while recursive truncation only decreases the number further from 103 to 98 and from 92 to 91, respectively. It is worth mentioning that the two basis sets are very similar as one can see from the number of basis functions after the recursive truncation. The Jaccard indices give visual evidence that “def2-TZVPD trunc 91” and “def2-TZVPPD trunc 98” do not change after another truncation process since $J(x) = 0$ for $x \in [0, 0.1]$. It also shows that one-time truncation is good enough to significantly decrease $J(x^{\text{thr}})$ value (see two dashed lines). Considering the time and computational resources spent on recursive truncation process (usually needs several rounds of RT-TDHF/TDDFT calculation), we only focus on one-time truncation in the following.

The spectra of H₂O dimer using def2-TZVP, def2-TZVPP, def2-TZVPD, and def2-TZVPPD basis sets give a similar conclusion as in H₂ dimer case, namely, that extra diffuse functions have large impact on the absorption spectra while extra polarization functions have limited impact on the absorption spectra (see Supplementary Figure 4). In addition, we provide the spectra after a one-time truncation and recursive truncation processes (see Fig. 3g). All spectra in this figure are very close to each other up to an excitation energy of 20 eV, indicating def2-TZVPPD includes many redundant basis functions for RT-TDHF calculations in this case. For the usage of computational resources, in the case of $O(N^4)$ scaling (HF Coulomb and exchange matrices calculated with 2-electron integrals, without real-space gridding or density-fitting), a RT-TDHF run with “def2-TZVPD trunc 91” basis set only consumes $(91/148)^4 = 14\%$ of the time compared to the original def2-TZVPPD basis set. For this system, a special interest is the effect of hydrogen bonds on the truncation process. However, we do not find any dependence of deleted basis functions on the distance (up to 10 Å) between two water molecules, and the suggested truncated basis sets are very similar. This may indicate that the basis functions needed for the description of hydrogen bonds are also important for the electronic absorption spectrum of the water monomers themselves.

Moreover, the CBS scheme is tested for the H₂O dimer system. Two basis sets, def2-TZVPPD and def2-QZVPPD are used as the starting point for the CBS scheme, with $\epsilon = 10^{-6}$ (CBS- 10^{-6}). We directly modify the basis set file every time when truncating or adding basis functions. The detailed steps of the CBS scheme for def2-TZVPPD basis set are shown in Table 1. The original def2-TZVPPD basis set of the H atom and the O atom is 3s3p1d and 6s4p3d1f, respectively, with 148 basis functions for the H₂O dimer system. After the first RT-TDHF run, the first d-subshell (1d) of H and the first d-subshell and f-subshell of O (1d1f) are truncated (shown in the bracket). The diffuse functions are then added to the remaining subshells (shown in the bracket), resulting in 4s4p for H and 7s5p3d for O. This is followed by a second RT-TDHF run with

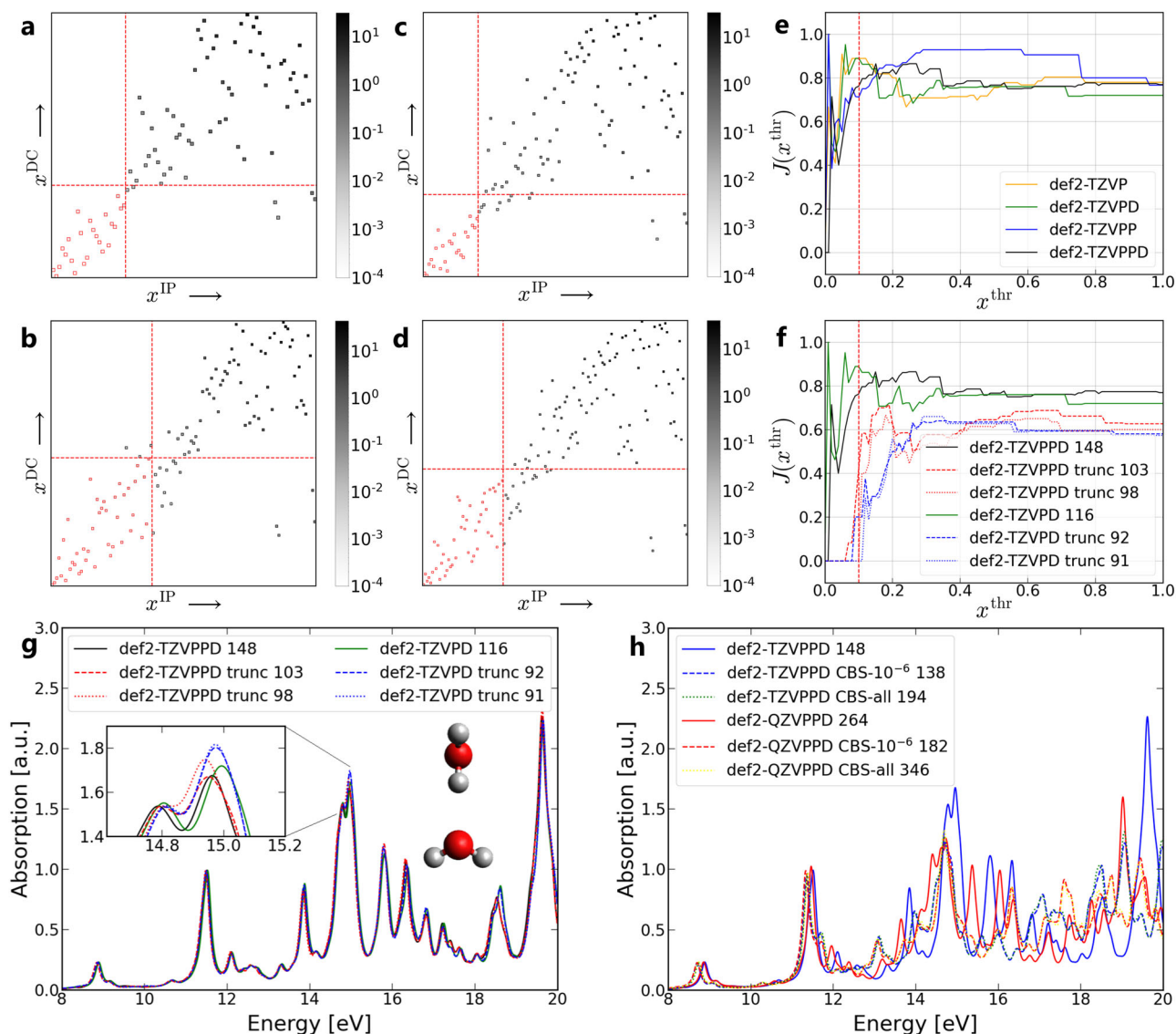


Fig. 3 | RT-TDHF calculations of the H₂O dimer. x^{DC} - x^{IP} map obtained with a def2-TZVP, **b** def2-TZVPP, **c** def2-TZVPD, and **d** def2-TZVPPD basis set. Each square in the map represents a basis function with its numbering on x and y axes. The color of the squares is based on the value calculated as $x^{DC} \cdot x^{IP}$, so the deeper the color the more important the basis function is in the TDHF calculation. The red dashed line represents $x^{thr} = 0.1$. **e** Jaccard indices for def2-TZVP, def2-TZVPP, def2-TZVPD, and def2-TZVPPD basis sets, and **f** def2-TZVPD, def2-TZVPPD, and corresponding truncated/iteratively truncated basis sets. The vertical red dashed line represents

$x^{thr} = 0.1$. **g** Electronic absorption spectra using original, one-time truncation, and recursive truncation of def2-TZVPD, def2-TZVPPD basis sets. The truncation threshold is $x^{thr} = 0.1$. **h** Electronic absorption spectra using original, CBS-10⁻⁶, and CBS-all (combined basis set of the original one and CBS-10⁻⁶) of def2-TZVPPD and def2-QZVPPD basis sets. The truncation threshold used in the CBS scheme is $x^{thr} = 0.1$. The number after the basis set label indicates the number of basis functions used for the RTP. *a.u.* arbitrary units.

truncating and adding basis function. With a basis set of 5s4p for H and 8s6p4d for O, we find $|\lambda|_{min}$ is smaller than the threshold we set (10^{-6}), thus the newly added subshell with the highest angular momentum is removed, say, 4p in H and 4d in O. Finally, we obtain a basis set of 5s3p for H and 8s6p3d for O, with 138 basis functions and $|\lambda|_{min} = 3.7 \times 10^{-6}$. In Supplementary Table 1, we do it analogously for def2-QZVPPD basis set.

RT-TDHF calculations are carried out with these two CBSs, and the resulting spectra are shown in Fig. 3g. Choosing either the def2-TZVPPD or the def2-QZVPPD basis set (blue and red solid lines) can result in differences of absorption peaks with an excitation energy larger than 12 eV, while their corresponding CBSs-10⁻⁶ (blue and red dashed lines) match until 15 eV. Also, CBSs-10⁻⁶ leads to a red shift of 0.2 eV compared to the two original basis sets, which usually indicates the behavior of a larger basis set according to the observations in this work, while in our case this is achieved by less basis functions. In

addition, we utilize combined basis sets of original basis set and its CBS-10⁻⁶ (noted as CBS-all) for the same calculation, and we can see that each spectrum (dotted lines) also agrees with the corresponding CBS-10⁻⁶ one.

Examples: (S)-methyloxirane and (-)- α -pinene

As other examples, (S)-methyloxirane and (-)- α -pinene molecules are tested with truncated basis sets. The def2-TZVPP^{32,38} basis set is adopted as a reference basis set and the B3LYP functional is selected as the exchange-correlation functional for these two systems. $x^{thr} = 0.1$ and $x^{thr} = 0.2$ are used as the truncation threshold. Supplementary Tables 2 and 3 give information about the original and truncated basis sets for (S)-methyloxirane and (-)- α -pinene, respectively. The corresponding x^{DC} - x^{IP} maps are in Supplementary Figs. 5 and 6.

The resulting absorption spectra are shown in Fig. 4a and Supplementary Fig. 8a. The truncated basis sets, both $x^{thr} = 0.1$ and $x^{thr} =$

Table 1 | The steps of the CBS scheme for def2-TZVPPD basis set in the H₂O dimer system

Basis set	H	O	No. basis function
Original	3s3p1d	6s4p3d1f	148
Run 1			
Truncate	3s3p (-1d)	6s4p2d (-1d1f)	104
Add	4s4p (+4s4p)	7s5p3d (+7s5p3d)	138
Run 2			
Truncate	4s3p (-1p)	7s5p3d	126
Add	5s4p (+5s4p)	8s6p4d (+8s6p4d)	160
$ \lambda _{\min} < 10^{-6}$			
Remove	5s3p (-4p)	8s6p3d (-4d)	138
Run 3			
Truncate	5s3p (none)	8s6p3d (none)	138
$ \lambda _{\min} = 3.7 \times 10^{-6}$			

The truncation threshold used in the CBS scheme is $x^{\text{thr}} = 0.1$. Basis set information of H and O atoms is noted as *non-italic* form, and the added/truncated/removed basis functions are in *italic* form in the brackets. "Run" refers to a short time RT-TDHF calculation with 100 steps. In the last column, the total number of basis functions of the system is listed.

0.2, provide a good approximation to the absorption spectra compared to the original def2-TZVPP basis set, while using as few as half of the basis functions (in the case of $x^{\text{thr}} = 0.2$). Apart from the usage in RT-TDDFT, the truncation process is found to be robust in LR-TDDFT as well. In LR-TDDFT calculations, 500 and 2000 roots are solved for (S)-methyloxirane and (-)- α -pinene systems, respectively. From these results and the results in H₂ dimer and H₂O dimer systems, one may find that most basis functions truncated are polarization functions, e.g., p/d-subshell for H and d/f-subshell for C/O, while diffuse functions are usually not removed. This explains why the CBS scheme we propose only considers additional diffuse functions.

Furthermore, we use the same basis sets for ECD spectra calculations, considering that the two quantities x^{DC} and x^{IP} do not explicitly depend on the electric dipole operator. The ECD spectra of (S)-methyloxirane and (-)- α -pinene are shown in Fig. 4b and Supplementary Fig. 8b, respectively.

Table 2 gives the benchmark of basis sets used for (S)-methyloxirane and (-)- α -pinene. RT-TDDFT calculations of these two systems are carried out using CP2K. Because Coulomb and exchange and correlation (XC) terms are evaluated on grids, we do not observe a significant time-saving using the proposed truncated basis sets. Nevertheless, computational resources can be reduced as much as one order of magnitude in LR-TDDFT calculations (Gaussian09³⁹) using truncated basis sets. The corresponding memory usage of (-)- α -pinene system is also shown in the table. The memory cost of Coulomb and exchange matrices scale as $\mathcal{O}(N^4)$ (2-electron integrals, without real-space griding or density-fitting), and one may easily encounter a memory bottleneck with large basis sets (e.g., using def2-TZVPP for (-)- α -pinene, maximal memory set to 200 GB), which, however, can be alleviated with truncated basis sets. In Supplementary Tables 4 and 5, we show the scaling information of (-)- α -pinene using HF/def2-TZVPP and its truncated basis set, and computational time to calculate Coulomb and exchange matrices, respectively. To assess the contribution from HF exchange term, we further show the difference between B3LYP/def2-TZVPP and BLYP/def2-TZVPP in the calculation of RT-TDDFT spectrum in Supplementary Fig. 7.

Example: ZnPc

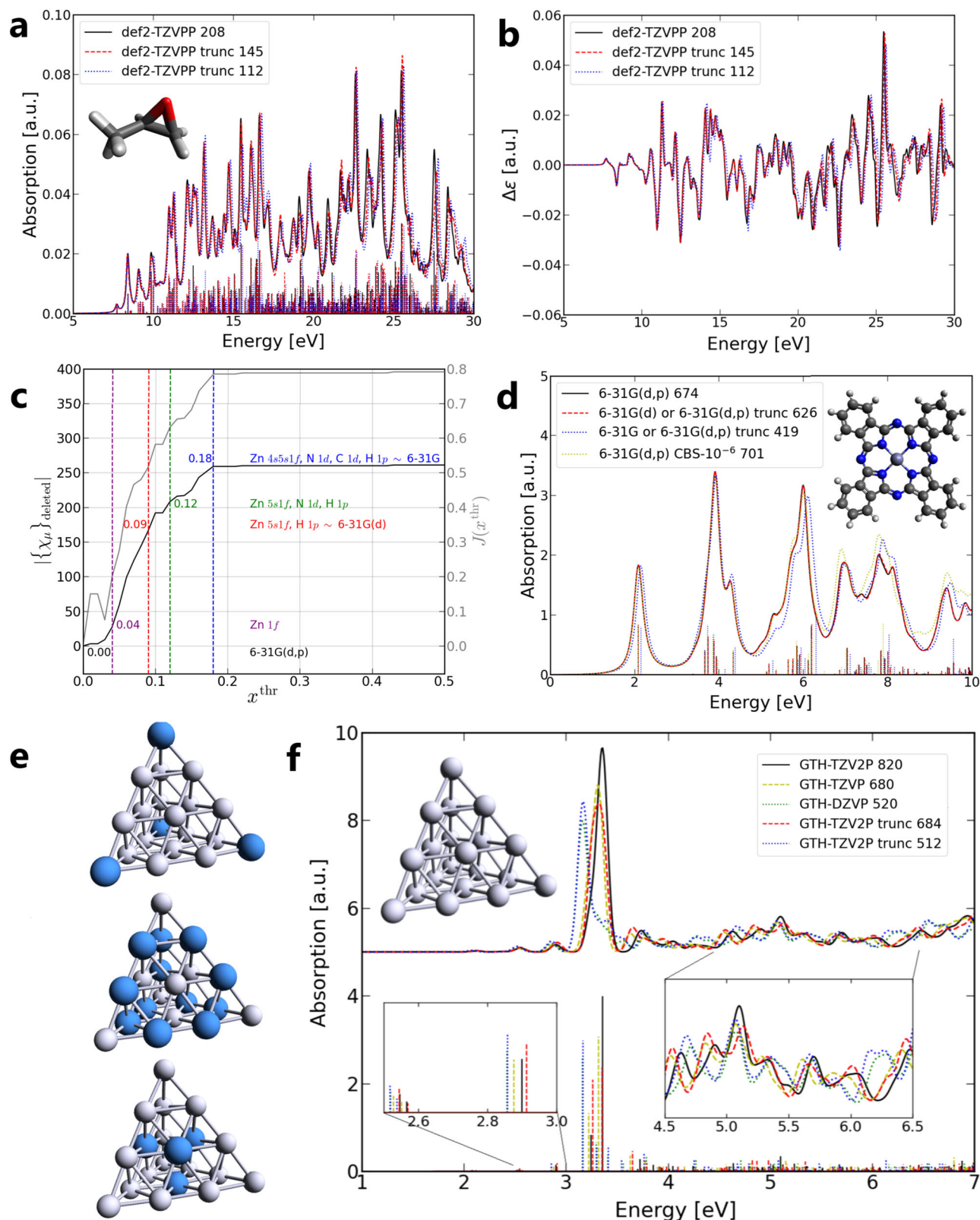
ZnPc is a popular example for excited-state calculations^{40–45}. This example is mainly utilized to demonstrate a step-by-step truncation from 6-31G(d,p) to 6-31G. Here we use the nuclear geometry of ZnPc from a previous study⁴⁵ with the B3LYP functional and 6-31G(d,p)^{28,35,36}

as the reference basis set. Figure 4c provides the information of deleted basis functions and Jaccard indices. The numbers close to the dashed lines are x^{thr} values, and the notations on the right side are details of deleted basis functions in *italic* form. Note that the number of deleted basis functions ($|\{\chi_{\mu}\}_{\text{deleted}}|$) can be higher than the number calculated from the subshell notations on the right. This is because extra basis functions might also be deleted but not the corresponding full subshells, e.g., Zn 1f $x^{\text{thr}}=0.04$ corresponds to a $|\{\chi_{\mu}\}_{\text{deleted}}|$ value larger than 7 because some other basis functions like H 1p_z (but not the full 1p) are deleted. In addition, this planar system, which we place in $x-y$ plane in the simulation, shows some preferences for $1d_{xy}$ and $1d_{x^2-y^2}$ of C/N elements, and thus $1d_{yz}$, $1d_{xz}$, and $1d_{z^2}$ basis functions are the first to be deleted in the range of $x^{\text{thr}} = 0.09 - 0.18$, indicating that the truncation scheme can provide the information of preference on the orientation of basis functions (or AOs with different magnetic quantum number). It is worth mentioning that $x^{\text{thr}} = 0.09$ and $x^{\text{thr}} = 0.18$ truncation leads basically to the 6-31G(d) and 6-31G basis sets, except for the additional truncation on the Zn atom. Also, it is found that $|\{\chi_{\mu}\}_{\text{deleted}}|$ and $J(x^{\text{thr}})$ show a very similar trend. After $x^{\text{thr}} = 0.18$, both lines reach a plateau where seldom further basis functions can be removed, indicating 6-31G as a good truncated basis set. Actually, we can see this from $x^{\text{DC}}-x^{\text{IP}}$ map of the same system (see Supplementary Fig. 9) in which the truncated and remaining basis functions almost form two blocks with $x^{\text{thr}} = 0.18$ (dashed blue line).

The corresponding RT-TDDFT and LR-TDDFT (1000 roots) spectra are given in Fig. 4d. It is clear that 6-31G(d,p), 6-31G(d), and 6-31G basis sets all provide similar results, which match our truncation suggestions. This shows a practical usage of our truncation scheme on the selection of basis set. In addition, the CBS-10⁻⁶ with 6-31G(d,p) reference is constructed (see Supplementary Table 6 for the CBS process). Nevertheless, it does not change much in the RT-TDDFT/LR-TDDFT spectra compared to the original 6-31G(d,p) basis set.

Example: Ag₂₀

Ag₂₀ is a metal cluster with tetrahedral structure (T_d symmetry), which has been investigated with TDDFT calculations^{46–48}. Here we use the nuclear geometry of Ag₂₀ from a previous study⁴⁸ together with PBE0⁴⁹ functional and GTH^{50,51} Gaussian-type pseudopotential basis sets⁵² GTH-DZVP, GTH-TZVP, and GTH-TZV2P. GTH-TZV2P is used as the reference basis set for the truncation process. The Ag atoms in Ag₂₀ cluster are categorized into 3 groups: vertex (v), edge (e), and face (f) (see Fig. 4e). The atoms in the same group are equivalent in space and should have the same contribution to the electronic absorption spectrum. Table 3 shows the truncated basis functions versus increasing x^{thr} values. As one can see from the table, atoms in different groups generally have different suggested basis set truncations. More basis functions are truncated for atoms at vertex position, and less for atoms at face position. This is reasonable because vertex Ag atoms have a limited space angle "bonded" with other atoms, while face Ag atoms have half of their surrounding space occupied with 9 nearest neighbors, and complex surroundings often require more basis functions to describe the interactions. For comparison, basis set information of GTH-TZVP and GTH-DZVP is also listed. The truncation scheme provides quite different basis sets from GTH-TZVP and GTH-DZVP basis set, e.g., GTH-TZVP can be regarded as $-2f$ truncated basis set of GTH-TZV2P, but $2f$ basis functions are the last choice of truncation from our scheme (up to $x^{\text{thr}} = 0.6$). This means that the corresponding standardly available smaller basis sets, e.g., GTH-DZVP or GTH-TZVP, do not always contain the most important basis functions (for TDDFT calculations) from the larger ones, e.g., GTH-TZV2P, which is different from what we have found for the ZnPc example system. We select $x^{\text{thr}} = 0.3$ and $x^{\text{thr}} = 0.5$ truncated basis sets for LR-TDDFT calculations, under the consideration that the numbers of basis functions are close to GTH-TZVP and GTH-DZVP basis sets, respectively.



The LR-TDDFT spectra (2000 roots) of the Ag_{20} cluster using 5 different basis sets are shown in Fig. 4f. In general, “GTH-TZV2P trunc 684” ($x^{\text{thr}} = 0.3$) gives better agreement with the reference GTH-TZV2P basis set than the one with the GTH-TZV2P basis set. This can be seen as follows: 1. for the first several peaks (at -2.6 eV, 2.9 eV, and 3.4 eV), red dashed peaks all locate closer to black peaks than yellow dashed peaks; 2. for the peaks up to 7 eV, the red dashed line follows closer to the

black line than the yellow dashed line. However, the difference between the spectra calculated using “GTH-TZV2P trunc 512” ($x^{\text{thr}} = 0.5$) and GTH-DZVP basis sets is limited, which can be explained by their similar composition in terms of basis functions shown in Table 3. The Ag_{20} example demonstrates that the proposed truncation scheme has the ability of assigning different basis sets to the atoms, according to their “interaction” with the full system.

Fig. 4 | RT-TDDFT calculations of (S)-methyloxirane, ZnPc, and the Ag₂₀ structures. (S)-methyloxirane **a** electronic absorption spectra and **b** electronic circular dichroism spectra using original and truncated basis sets of def2-TZVPP (stick spectra below correspond to LR-TDDFT ones). **c** Number of deleted basis functions ($\{|\chi_{\mu}^{\text{deleted}}|\}$) and corresponding Jaccard indices of 6-31G(d,p) basis set for ZnPc system. Purple, red, green, and blue dashed lines represent different x^{thr} values, and the corresponding colored notations give the deleted basis functions in *italic* form. Note that the red dashed line with $x^{\text{thr}} = 0.09$ and blue dashed line with $x^{\text{thr}} = 0.18$ essentially reduce 6-31G(d,p) to 6-31G(d) and 6-31G basis set, respectively

(except for the addition truncation on the Zn atom). **d** RT-TDDFT/LR-TDDFT electronic absorption spectra using 6-31G(d,p), 6-31G(d), 6-31G, and 6-31G(d,p) CBS-10⁻⁶ basis sets for ZnPc system (stick spectra below correspond to LR-TDDFT ones). **e** Ag₂₀ geometry with colored vertex atoms (top), colored edge atoms (middle), and colored face atoms (bottom). **f** Ag₂₀ LR-TDDFT electronic absorption spectra using GTH-TZV2P, GTH-TZVP, GTH-DZVP, and truncated GTH-TZV2P basis sets. Below are the stick spectra and above are the Gaussian broadened spectra with FWHM = 0.12 eV. The number after the basis set label indicates the number of basis functions. *a.u.* arbitrary units.

Additionally, we also provide some testing calculations to demonstrate: 1. 1% of the total propagation time is sufficient to show the contribution from each basis function, and the same truncation suggestion has been obtained using 1%, 10%, and 100% of the RTP steps (see Supplementary Fig. 10) 2. Indicator x^{tr} is necessary in the truncation scheme, and the truncation using only indicator x^{bc} can lead to a different spectrum (see Supplementary Fig. 11) 3. ERIs Schwarz screening does not affect the truncation scheme, and they can be used together for the acceleration in RT-TDHF/TDDFT calculations (see Supplementary Figs. 12-13 and Supplementary Tables 7-9).

Discussion

We have introduced an AO basis set truncation scheme for TDDFT calculations, based on the analysis of a short period of real time propagation of MO coefficients. Two quantities – density matrix contribution and importance of propagation stability – are constructed as indicators for the truncation process. The truncated basis sets are found to reproduce the electronic absorption spectra obtained with the original basis sets well. In some cases, truncated basis sets can serve as intermediate basis sets between two levels of available basis sets, or are found to be very close to lower level basis sets available, in which the truncation process works as a means to help in basis set selection. Two intuitive graphs, $x^{\text{bc}}-x^{\text{tr}}$ map and Jaccard index, are introduced for the analysis of basis functions. These graphs also provide a guide for the choice of the truncation threshold x^{tr} (e.g., see diagrams in Fig. 5).

As opposed to basis sets constructed mainly for the purpose of energy minimization and geometry optimization, the truncation scheme proposed provides a task-, system-, and chemical environment-specific basis set. It has reduced number of basis functions and accelerates the calculations involving the construction of Coulomb and/or exchange matrices iteratively in every propagation step, potentially with the scaling of $\mathcal{O}(N^4)$. ERIs usually benefit from evaluating all components from the same shell. However, they are only computed once before the propagation. Because

the truncation process is carried out on the original AO basis set without any rotation or reconstruction of basis functions, the truncated basis sets can be easily employed in any quantum chemistry package using Gaussian-type (or Slater-type) basis, with a simple modification of the basis set file. Additionally, we have tested recursive truncation to show that the process is robust and will result in a “truncation consistent” basis set given a certain x^{tr} . Though the truncation is based on the analysis of real-time propagation, the basis sets produced can also be used for LR-TDDFT calculations and provide equally good spectra. Nevertheless, the acceleration of LR-TDDFT calculations depends on the the systems and purposes of the research, e.g., limited number of excitations in LR-TDDFT for a small system may not be worth an additional RT-TDDFT calculation to determine a truncated basis set, while a highly conjugated or a large system with excitations of higher energies should benefit from the truncation scheme. How the truncation scheme and truncated basis set might be transferable to more accurate yet expensive methods like *GW*/Bethe-Salpeter equation⁵³⁻⁵⁵, time-dependent coupled cluster/configurational interaction⁵⁶⁻⁵⁸, or other type of any excited-state calculations might be explored in the future.

Furthermore, an “Add-While-Truncate” algorithm has been proposed to construct basis sets towards the complete basis set limit. The additional basis functions are added as diffuse functions in an even-tempered manner, and no extra polarization functions are added. The neglect of polarization functions is primarily based on truncation experiences we have got from this study (e.g., in H₂O dimer, (S)-methyloxirane, and (-)- α -pinene systems). There are some discussions about the use of polarization and diffuse functions used for electric dipole moment, polarizability, and TDDFT calculations in previous works⁵⁹⁻⁶³. Nevertheless, as shown in test examples, the truncation process provides the possibility to select polarization and diffuse functions quantitatively. The proposed CBS scheme can construct basis sets to arbitrary accuracy, depending on a pre-defined parameter limited by the linear dependency between basis functions.

The truncation scheme might reveal some intrinsic knowledge for the better description of electronic excitations between ground state and excited states, and offer a thought for the design of basis sets in TDDFT calculations. Future work can be on both basis set constructions and migration to other excited-state calculations or properties. In this work, all original basis sets employed are ground state energy-optimized, however, there is another group of completeness-optimized basis sets⁶⁴⁻⁶⁶, with which one may also test the efficiency and validate the accuracy towards CBS-limit⁶⁵. Auxiliary density matrix methods⁶⁷ provide an alternative way to accelerate HF exchange calculation via auxiliary basis set, and have been found to yield highly accurate results in energies and response properties⁶³. Considering the computational demanding HF exchange calculation employed in hybrid functionals, it is possible to further assess the truncation scheme for auxiliary basis sets. In addition, the idea of decomposing the electric dipole contributions into the contribution from individual basis functions can be migrated to other properties and produce different task specific basis sets. One may be interested in a truncated basis for dynamic calculations,

Table 2 | Benchmark of the original and truncated basis sets of def2-TZVPP used in (S)-methyloxirane and (-)- α -pinene systems

System	Basis set	Time/Step (RTP)	Time [Memory] (LR)
	Def2-TZVPP	3.56 ± 0.22 s	767 s
(S)-methyloxirane	Trunc $x^{\text{tr}} = 0.1$	2.19 ± 0.19 s	214 s
	Trunc $x^{\text{tr}} = 0.2$	1.74 ± 0.16 s	111 s
	Def2-TZVPP	4.30 ± 0.54 s	38.4 h [198 GB]
(-)- α -pinene	Trunc $x^{\text{tr}} = 0.1$	2.27 ± 0.19 s	9.1 h [165 GB]
	Trunc $x^{\text{tr}} = 0.2$	1.44 ± 0.18 s	3.1 h [81 GB]

The truncation threshold used are $x^{\text{tr}} = 0.1$ and $x^{\text{tr}} = 0.2$. The RT-TDDFT time is noted as time per step (40'000 steps in total) and LR-TDDFT time is noted as full computational time. The memory usage of LR-TDDFT calculations of (-)- α -pinene system is also presented in square brackets. LR-TDDFT calculations are processed with 8 CPU cores@3.3 GHz, and RT-TDDFT calculations are processed with 2 × 12 CPU cores@2.6 GHz and 32 × 12 CPU cores@2.6 GHz for (S)-methyloxirane and (-)- α -pinene, respectively. s second, h hour, GB gigabyte

which, however, might require further investigations on the consistency in the truncation for each nuclear configuration. Apart from basis set truncation, a direct basis set optimization algorithm (e.g., on exponents of Gaussian-type basis functions) is also possible given a proper loss function based on x^{DC} and x^{IP} parameters. While we have only tested the truncation process on neutral molecules in this study, charged systems could also be investigated. This would be an interesting topic since it may demonstrate the dependence of necessary basis functions on different charges for excited state calculations (e.g., effect of diffuse function on anions which is known for ground state cases).

In summary, our basis set truncation scheme provides a robust process for decreasing the number of basis functions and speeding up

Table 3 | The truncation process of TZV2P basis set in Ag₂₀ system

x^{thr}	Truncated basis functions	No. truncated basis function	No. basis function [basis set]
0.0	– <i>v: -3p</i>	0	820 [GTH-TZV2P]
0.0-0.1	<i>e: -3p</i> <i>f: -3p</i> <i>v: -3p1f</i>	60	760
0.1-0.3	<i>e: -3p</i> <i>f: -3p</i> <i>v: -3p3d1f2f</i>	88	732
0.3	<i>e: -3p</i> <i>f: -3p</i>	136	684
–	<i>v,e,f: -2f</i> <i>v: -3p3d1f2f</i>	140	680 [GTH-TZVP]
0.3-0.5	<i>e: -3p1f</i> <i>f: -3p</i>	220	600
–	<i>v,e,f: -3p3d2f</i> <i>v: -3p3d1f2f</i>	300	520 [GTH-DZVP]
0.5	<i>e: -3p3d1f</i> <i>f: -3p1f</i> <i>v: -3p3d1f2f</i>	308	512
0.5-0.6	<i>e: -3p3d1f2f</i> <i>f: -3p1f</i>	392	428

Due to the T_d symmetry of the system, we can categorize Ag atoms into 3 groups: vertex (4 atoms), edge (12 atoms), and face (4 atoms), with the notation “v”, “e”, and “f”, respectively. Different truncation thresholds x^{thr} are employed and GTH-TZVP and GTH-DZVP basis set information are also listed for comparison. Truncated basis functions are in *italic* form.

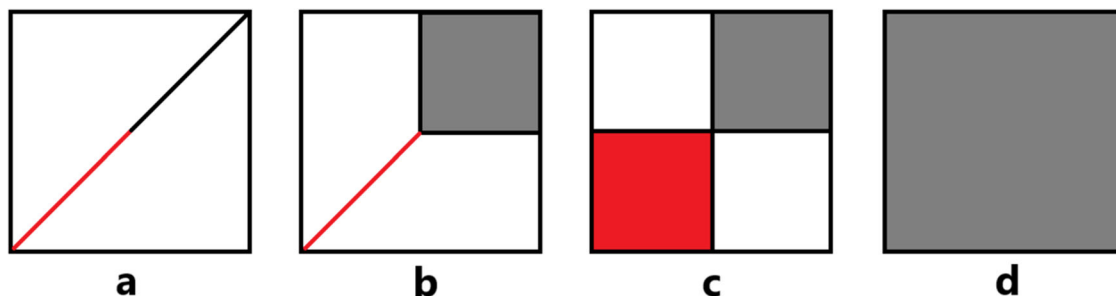


Fig. 5 | Diagrams of four different types of x^{DC} - x^{IP} maps. Red/gray denote basis functions that can/cannot be removed. From left to right: **a** high correlation between x^{DC} and x^{IP} , basis set can be truncated at any accuracy (e.g., H₂ dimer 6-31G**, pulse z direction); **b** high correlation between x^{DC} and x^{IP} until certain s (where s is boundary of high and low correlation regions), basis set can be truncated at

TDDFT calculations, while preserving the high accuracy of the spectra. The quantitative basis set analysis allows a profound understanding of the basis functions employed and opens up a broad area for potential research in excited state calculations.

Methods

The systems H₂ dimer, H₂O dimer, (*S*)-methyloxirane, (-)- α -pinene, zinc phthalocyanine (ZnPc), and Ag₂₀ have been investigated. Information about the applied computational methods, basis sets, and codes are listed in Table 4. For the H₂ dimer and the H₂O dimer, we utilize an in-house version of the PySCF^{68,69} RT-TDHF module⁷⁰ to test truncation and CBS scheme. Calculations are carried out with 6-31G series^{28,35,36} w/o additional polarization/diffuse functions, and def2-TZVP series^{32,38} w/o additional polarization/diffuse functions. No Schwarz screening is used for the RT-TDHF calculations of H₂ dimer and H₂O dimer systems. For (*S*)-methyloxirane, (-)- α -pinene, and ZnPc, the CP2K (CP2K version 7.0 (Development Version), the CP2K developers group. CP2K is freely available from <https://www.cp2k.org/>) package and the Gaussian09³⁹ package is used for RT-TDDFT and LR-TDDFT (B3LYP⁷¹) calculations, respectively. For Ag₂₀, Goedecker-Teter-Hutter (GTH) pseudopotential^{50,51} with the corresponding Gaussian-type pseudopotential basis sets⁵² GTH-DZVP, GTH-TZVP, and GTH-TZV2P, and PBE0⁴⁹ hybrid functional are employed for time-dependent density functional perturbation theory (TDDFT, up to the first order of the perturbation we use the term LR-TDDFT in this work) calculations using CP2K (CP2K version 7.0 (Development Version), the CP2K developers group. CP2K is freely available from <https://www.cp2k.org/>) package. Schwarz screening threshold 10⁻¹⁰ (default in CP2K) is used in the RT-TDDFT calculations of (*S*)-methyloxirane, (-)- α -pinene, ZnPc, and Ag₂₀ systems. All basis set files used in this work are from Basis Set Exchange⁷², visualization of molecular structures and orbitals uses Avogadro⁷³ software, and graphs are generated with Matplotlib⁷⁴.

A δ -pulse is chosen as the electric field perturbation to excite the molecules in RT-TDHF/TDDFT calculations. The application of the δ -pulse can be thought of as being applied instantly to the converged ground states MOs $|\psi^0\rangle$ between a time $t = 0^-$ and $t = 0^+$. It corresponds to an impulse⁷⁵

$$|\psi^\delta(\vec{r}, t = 0^+)\rangle = e^{-\frac{i}{\hbar}\vec{k}\cdot\vec{r}} |\psi^0(\vec{r}, t = 0^-)\rangle, \quad (17)$$

where \hbar is reduced Planck constant. The vector \vec{k} indicates the direction and amplitude of the perturbation. The propagation is then started from the perturbed MOs $|\psi_i^\delta\rangle$.

accuracy with $x^{\text{thr}} \leq s$ (e.g., H₂ dimer 6-31++G**, pulse z direction); **c** two blocks with low correlation regions, basis set is recommended to be truncated at connected part between two blocks (e.g., all other test examples), which is also the most common case; **d** fully low correlation region, basis set can hardly be truncated (e.g., H₂ dimer 6-31G/6-31++G, pulse z direction).

Table 4 | Systems and corresponding computational details

System	Methods	Basis sets	Codes
H ₂ dimer	RT-TDHF	6-31G,	PySCF ^{68,69}
		6-31G**,	
		6-31++G,	
		6-31++G**	
H ₂ O dimer	RT-TDHF	def2-TZVP,	PySCF ^{68,69}
		def2-TZVPP,	
		def2-TZVPD,	
		def2-TZVPPD,	
		CBS- ϵ	
(S)-methyloxirane	RT-TDDFT,	def2-TZVPP	CP2K (CP2K version 7.0 (Development Version), the CP2K developers group. CP2K is freely available from https://www.cp2k.org/),
(-)- α -pinene	LR-TDDFT		Gaussian09 ³⁹
ZnPc	RT-TDDFT,	6-31G,	CP2K (CP2K version 7.0 (Development Version), the CP2K developers group. CP2K is freely available from https://www.cp2k.org/),
		6-31G(d),	Gaussian09 ³⁹
	LR-TDDFT	6-31G(d,p),	
		CBS- ϵ	
Ag ₂₀	LR-TDDFT	GTH-DZVP,	CP2K (CP2K version 7.0 (Development Version), the CP2K developers group. CP2K is freely available from https://www.cp2k.org/)
		GTH-TZVP,	
		GTH-TZV2P	

The hybrid functional B3LYP is employed for the calculations of (S)-methyloxirane, (-)- α -pinene, and ZnPc, and PBE0 for the calculations of Ag₂₀. Pseudopotentials are applied for Ag₂₀, and all electron calculations for the other systems. Timestep: 0.05 atomic units for (S)-methyloxirane, (-)- α -pinene, ZnPc, 0.2 atomic units for the others.

Data availability

The data generated in this study have been deposited at <https://gitlab.uzh.ch/lubergroup/ao-truncation>. Source data are provided in this paper.

Code availability

Python codes for carrying out basis set truncation and RT-TDSCF calculations are available at <https://gitlab.uzh.ch/lubergroup/ao-truncation>.

References

- Casida, M. & Huix-Rotllant, M. Progress in time-dependent density-functional theory. *Annu. Rev. Phys. Chem.* **63**, 287–323 (2012).
- Laurent, A. D. & Jacquemin, D. TD-DFT benchmarks: a review. *Int. J. Quant. Chem.* **113**, 2019–2039 (2013).
- Adamo, C. & Jacquemin, D. The calculations of excited-state properties with time-dependent density functional theory. *Chem. Soc. Rev.* **42**, 845–856 (2013).
- Provorse, M. R. & Isborn, C. M. Electron dynamics with real-time time-dependent density functional theory. *Int. J. Quant. Chem.* **116**, 739–749 (2016).
- Goings, J. J., Lestranger, P. J. & Li, X. Real-time time-dependent electronic structure theory. *WIREs Comput Mol Sci.* **8** (2017). <https://doi.org/10.1002/wcms.1341>.
- Li, X., Govind, N., Isborn, C., DePrince, A. E. & Lopata, K. Real-time time-dependent electronic structure theory. *Chem. Rev.* **120**, 9951–9993 (2020).
- Mattiat, J. & Lubner, S. Efficient calculation of (resonance) Raman spectra and excitation profiles with real-time propagation. *J. Chem. Phys.* **149**, 174108 (2018).
- Mattiat, J. & Lubner, S. Vibrational (resonance) Raman optical activity with real time time dependent density functional theory. *J. Chem. Phys.* **151**, 234110 (2019).
- Mattiat, J. & Lubner, S. Time domain simulation of (resonance) Raman spectra of liquids in the short time approximation. *J. Chem. Theory Comput.* **17**, 344–356 (2020).
- Aquilante, F., Todorova, T. K., Gagliardi, L., Pedersen, T. B. & Roos, B. O. Systematic truncation of the virtual space in multiconfigurational perturbation theory. *J. Chem. Phys.* **131**, 034113 (2009).
- Nagy, P. R., Gyevi-Nagy, L. & Kállay, M. Basis set truncation corrections for improved frozen natural orbital CCSD(t) energies. *Mol. Phys.* **119** (2021). <https://doi.org/10.1080/00268976.2021.1963495>.
- Mintz, B. & Wilson, A. K. Truncation of the correlation consistent basis sets: extension to third-row (ga-kr) molecules. *J. Chem. Phys.* **122**, 134106 (2005).
- Feller, D. & Dixon, D. A. Density functional theory and the basis set truncation problem with correlation consistent basis sets: elephant in the room or mouse in the closet? *J. Phys. Chem. A* **122**, 2598–2603 (2018).
- Barnes, T. A., Goodpaster, J. D., Manby, F. R. & Miller, T. F. Accurate basis set truncation for wavefunction embedding. *J. Chem. Phys.* **139**, 024103 (2013).
- Claudino, D. & Mayhall, N. J. Simple and efficient truncation of virtual spaces in embedded wave functions via concentric localization. *J. Chem. Theory Comput.* **15**, 6085–6096 (2019).
- Ding, F., Manby, F. R. & Miller, T. F. Embedded mean-field theory with block-orthogonalized partitioning. *J. Chem. Theory Comput.* **13**, 1605–1615 (2017).
- Koh, K. J., Nguyen-Beck, T. S. & Parkhill, J. Accelerating realtime TDDFT with block-orthogonalized manby–miller embedding theory. *J. Chem. Theory Comput.* **13**, 4173–4178 (2017).
- Krishtal, A., Ceresoli, D. & Pavanello, M. Subsystem real-time time dependent density functional theory. *J. Chem. Phys.* **142**, 154116 (2015).
- Santis, M. D. et al. Environmental effects with frozen-density embedding in real-time time-dependent density functional theory

- using localized basis functions. *J. Chem. Theory Comput.* **16**, 5695–5711 (2020).
20. Sharma, M. & Sierka, M. Efficient implementation of density functional theory based embedding for molecular and periodic systems using gaussian basis functions. *J. Chem. Theory Comput.* **18**, 6892–6904 (2022).
21. Repisky, M. et al. Excitation energies from real-time propagation of the four-component dirac–kohn–sham equation. *J. Chem. Theory Comput.* **11**, 980–991 (2015).
22. Bruner, A., LaMaster, D. & Lopata, K. Accelerated broadband spectra using transition dipole decomposition and padé approximants. *J. Chem. Theory Comput.* **12**, 3741–3750 (2016).
23. Wibowo, M., Irons, T. J. P. & Teale, A. M. Modeling ultrafast electron dynamics in strong magnetic fields using real-time time-dependent electronic structure methods. *J. Chem. Theory Comput.* **17**, 2137–2165 (2021).
24. Castro, A., Marques, M. A. L. & Rubio, A. Propagators for the time-dependent kohn–sham equations. *J. Chem. Phys.* **121**, 3425–3433 (2004).
25. Lippert, G., Hutter, J. & Parrinello, M. A hybrid gaussian and plane wave density functional scheme. *Mol. Phys.* **92**, 477–487 (1997).
26. Bardo, R. D. & Ruedenberg, K. Even-tempered atomic orbitals. VI. optimal orbital exponents and optimal contractions of gaussian primitives for hydrogen, carbon, and oxygen in molecules. *J. Chem. Phys.* **60**, 918–931 (1974).
27. Cherkes, I., Klaiman, S. & Moiseyev, N. Spanning the hilbert space with an even tempered gaussian basis set. *Int. J. Quant. Chem.* **109**, 2996–3002 (2009).
28. Ditchfield, R., Hehre, W. J. & Pople, J. A. Self-consistent molecular-orbital methods. IX. an extended gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **54**, 724–728 (1971).
29. Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007–1023 (1989).
30. Jensen, F. Polarization consistent basis sets: principles. *J. Chem. Phys.* **115**, 9113–9125 (2001).
31. Weigend, F., Furche, F. & Ahlrichs, R. Gaussian basis sets of quadruple zeta valence quality for atoms h–kr. *J. Chem. Phys.* **119**, 12753–12762 (2003).
32. Weigend, F. & Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).
33. Lehtola, S. Curing basis set overcompleteness with pivoted cholesky decompositions. *J. Chem. Phys.* **151**, 241102 (2019).
34. Papajak, E., Zheng, J., Xu, X., Leverentz, H. R. & Truhlar, D. G. Perspectives on basis sets beautiful: seasonal plantings of diffuse basis functions. *J. Chem. Theory Comput.* **7**, 3027–3034 (2011).
35. Hariharan, P. C. & Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chem. Acc.* **28**, 213–222 (1973).
36. Clark, T., Chandrasekhar, J., Spitznagel, G. W. & Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. III. the 3-21+g basis set for first-row elements, li–f. *J. Comput. Chem.* **4**, 294–301 (1983).
37. Jaccard, P. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1. *New Phytol.* **11**, 37–50 (1912).
38. Rappoport, D. & Furche, F. Property-optimized gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **133**, 134105 (2010).
39. Frisch, M. J. et al. Gaussian 09 Revision D.01. Gaussian Inc. Wallingford CT (2009).
40. Theisen, R. F., Huang, L., Fleetham, T., Adams, J. B. & Li, J. Ground and excited states of zinc phthalocyanine, zinc tetrabenzoporphyrin, and azaporphyrin analogs using DFT and TDDFT with franck-condon analysis. *J. Chem. Phys.* **142**, 094310 (2015).
41. Wang, C., Shao, J., Chen, F. & Sheng, X. Excited-state absorption for zinc phthalocyanine from linear-response time-dependent density functional theory. *RSC Adv.* **10**, 28066–28074 (2020).
42. Martynov, A. G. et al. Methodological survey of simplified TD-DFT methods for fast and accurate interpretation of UV–vis–NIR spectra of phthalocyanines. *ACS Omega* **4**, 7265–7284 (2019).
43. Zhang, L., Qi, D., Zhao, L., Bian, Y. & Li, W. Substituent effects on the structure–property relationship of unsymmetrical methoxy and methoxycarbonyl phthalocyanines: DFT and TDDFT theoretical studies. *J. Mol. Graph. Model.* **35**, 57–65 (2012).
44. Wallace, A. J., Williamson, B. E. & Crittenden, D. L. Coupled cluster calculations provide a one-to-one mapping between calculated and observed transition energies in the electronic absorption spectrum of zinc phthalocyanine. *Int. J. Quant. Chem.* **117**, e25350 (2017).
45. Tussupbayev, S., Govind, N., Lopata, K. & Cramer, C. J. Comparison of real-time and linear-response time-dependent density functional theories for molecular chromophores ranging from sparse to high densities of states. *J. Chem. Theory Comput.* **11**, 1102–1109 (2015).
46. Sánchez-González, Á., Muñoz-Losa, A., Vukovic, S., Corni, S. & Mennucci, B. Quantum mechanical approach to solvent effects on the optical properties of metal nanoparticles and their efficiency as excitation energy transfer acceptors. *J. Phys. Chem. C* **114**, 1553–1561 (2010).
47. Kuda-Singappulige, G. U. & Aikens, C. M. Excited-state absorption in silver nanoclusters. *J. Phys. Chem. C* **125**, 24996–25006 (2021).
48. Chen, M., Dyer, J. E., Li, K. & Dixon, D. A. Prediction of structures and atomization energies of small silver clusters, (ag)_n, n < 100. *J. Phys. Chem. A* **117**, 8298–8313 (2013).
49. Adamo, C. & Barone, V. Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.* **110**, 6158–6170 (1999).
50. Goedecker, S., Teter, M. & Hutter, J. Separable dual-space gaussian pseudopotentials. *Phys. Rev. B* **54**, 1703–1710 (1996).
51. Krack, M. Pseudopotentials for h to kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.* **114**, 145–152 (2005).
52. VandeVondele, J. & Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **127**, 114105 (2007).
53. Rohlfing, M. & Louie, S. G. Electron-hole excitations and optical spectra from first principles. *Phys. Rev. B* **62**, 4927–4944 (2000).
54. Deslippe, J. et al. BerkeleyGW: A massively parallel computer package for the calculation of the quasiparticle and optical properties of materials and nanostructures. *Comput. Phys. Commun.* **183**, 1269–1289 (2012).
55. Bruneval, F. et al. molgw 1: Many-body perturbation theory software for atoms, molecules, and clusters. *Comput. Phys. Commun.* **208**, 149–161 (2016).
56. Pedersen, T. B. & Kvaal, S. Symplectic integration and physical interpretation of time-dependent coupled-cluster theory. *J. Chem. Phys.* **150**, 144106 (2019).
57. Koulias, L. N., Williams-Young, D. B., Nascimento, D. R., DePrince, A. E. & Li, X. Relativistic real-time time-dependent equation-of-motion coupled-cluster. *J. Chem. Theory Comput.* **15**, 6617–6624 (2019).
58. Sonk, J. A., Caricato, M. & Schlegel, H. B. TD-CI simulation of the electronic optical response of molecules in intense fields: comparison of RPA, CIS, CIS(d), and EOM-CCSD. *J. Phys. Chem. A* **115**, 4678–4690 (2011).
59. Darling, C. L. & Schlegel, H. B. Dipole moments, polarizabilities, and infrared intensities calculated with electric field dependent functions. *J. Phys. Chem.* **98**, 5855–5861 (1994).

60. Elliott, P., Furche, F. & Burke, K. Excited states from time-dependent density functional theory. In *Reviews in Computational Chemistry*, 91-165 (John Wiley & Sons, Inc., 2009). <https://doi.org/10.1002/9780470399545.ch3>.
61. Pescitelli, G. & Bruhn, T. Good computational practice in the assignment of absolute configurations by TDDFT calculations of ECD spectra. *Chirality* **28**, 466–474 (2016).
62. Barboza, C. A., Vazquez, P. A. M., Carey, D. M.-L. & Arratia-Perez, R. A TD-DFT basis set and density functional assessment for the calculation of electronic excitation energies of fluorene. *Int. J. Quant. Chem.* **112**, 3434–3438 (2012).
63. Kumar, C. et al. Accelerating kohn-sham response theory using density fitting and the auxiliary-density-matrix method. *Int. J. Quant. Chem.* **118**, e25639 (2018).
64. Chong, D. P. Completeness profiles of one-electron basis sets. *Can. J. Chem.* **73**, 79–83 (1995).
65. Manninen, P. & Vaara, J. Systematic gaussian basis-set limit using completeness-optimized primitive sets. a case for magnetic properties. *J. Comput. Chem.* **27**, 434–445 (2006).
66. Lehtola, S. Automatic algorithms for completeness-optimization of gaussian basis sets. *J. Comput. Chem.* **36**, 335–347 (2014).
67. Guidon, M., Hutter, J. & VandeVondele, J. Auxiliary density matrix methods for hartree-fock exchange calculations. *J. Chem. Theory Comput.* **6**, 2348–2364 (2010).
68. Sun, Q. et al. PySCF: the python-based simulations of chemistry framework. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8** (2017). <https://doi.org/10.1002/wcms.1340>.
69. Sun, Q. et al. Recent developments in the PySCF program package. *J. Chem. Phys.* **153**, 024109 (2020).
70. Nguyen, T. S. & Parkhill, J. Nonadiabatic dynamics for electrons at second-order: real-time TDDFT and OSCF2. *J. Chem. Theory Comput.* **11**, 2918–2924 (2015).
71. Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Chem. Phys.* **98**, 11623–11627 (1994).
72. Pritchard, B. P., Altarawy, D., Didier, B., Gibson, T. D. & Windus, T. L. New basis set exchange: an open, up-to-date resource for the molecular sciences community. *J. Chem. Inf. Model.* **59**, 4814–4820 (2019).
73. Hanwell, M. D. et al. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4** (2012). <https://doi.org/10.1186/1758-2946-4-17>.
74. Hunter, J. D. Matplotlib: A 2d graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
75. Yabana, K., Nakatsukasa, T., Iwata, J.-I. & Bertsch, G. F. Real-time, real-space implementation of the linear response time-dependent density-functional theory. *Phys. Status Solidi B* **243**, 1121–1138 (2006).

Acknowledgements

Funding by the University of Zurich and the Swiss National Science Foundation (grant no: PP00P2_170667) is gratefully acknowledged. We thank the Swiss National Supercomputing Center for computing resources (project ID: pr119, s1001, and s1036).

Author contributions

R.H. conceived the method. R.H., J.M., and S.L. designed the method. R.H. and J.M. performed the research. S.L. provided the resources and supervised the research. R.H., J.M., and S.L. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-35694-4>.

Correspondence and requests for materials should be addressed to Sandra Lubser.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023